



Prioritization and Evaluation of Flooding Tolerance Genes in Soybean [*Glycine max* (L.) Merr.]

Mu-Chien Lai¹, Zheng-Yuan Lai¹, Li-Hsin Jhan¹, Ya-Syuan Lai¹ and Chung-Feng Kao^{1,2*}

¹ Department of Agronomy, College of Agriculture and Natural Resources, National Chung Hsing University, Taichung, Taiwan, ² Advanced Plant Biotechnology Center, National Chung Hsing University, Taichung, Taiwan

OPEN ACCESS

Edited by:

Nunzio D'Agostino,
University of Naples Federico II, Italy

Reviewed by:

Tuanjie Zhao,
Nanjing Agricultural University, China
Heng Ye,
University of Missouri, United States

*Correspondence:

Chung-Feng Kao
kaoc@nchu.edu.tw;
gcf6@hotmail.com

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 30 September 2020

Accepted: 31 December 2020

Published: 27 January 2021

Citation:

Lai M-C, Lai Z-Y, Jhan L-H,
Lai Y-S and Kao C-F (2021)
Prioritization and Evaluation
of Flooding Tolerance Genes
in Soybean [*Glycine max* (L.) Merr.]
Front. Genet. 11:612131.
doi: 10.3389/fgene.2020.612131

Soybean [*Glycine max* (L.) Merr.] is one of the most important legume crops abundant in edible protein and oil in the world. In recent years there has been increasingly more drastic weather caused by climate change, with flooding, drought, and unevenly distributed rainfall gradually increasing in terms of the frequency and intensity worldwide. Severe flooding has caused extensive losses to soybean production and there is an urgent need to breed strong soybean seeds with high flooding tolerance. The present study demonstrates bioinformatics big data mining and integration, meta-analysis, gene mapping, gene prioritization, and systems biology for identifying prioritized genes of flooding tolerance in soybean. A total of 83 flooding tolerance genes (FTgenes), according to the appropriate cut-off point, were prioritized from 36,705 test genes collected from multidimensional genomic features linking to soybean flooding tolerance. Several validation results using independent samples from SoyNet, genome-wide association study, SoyBase, GO database, and transcriptome databases all exhibited excellent agreement, suggesting these 83 FTgenes were significantly superior to others. These results provide valuable information and contribution to research on the varieties selection of soybean.

Keywords: soybean, flooding tolerance, genomic data meta-analysis, gene prioritization, gene-set enrichment analysis

INTRODUCTION

Soybean [*Glycine max* (L.) Merr] is an important food crop worldwide that provides an essential source of anthocyanins and isoflavones for human beings (Cederroth and Nef, 2009). Previous studies have shown that soybean isoflavones can decelerate at the apoptotic rate of the cerebral cortex of rats and minimize the occurrence of ischemic stroke (Schreihofner et al., 2005; Burguete et al., 2006; Lovekamp-Swan et al., 2007). Some anthocyanin compounds possess novel antioxidant capacity and have neuroprotective effects on the central nervous system (Wang et al., 2010; Lu et al., 2012) that are beneficial for conditions such as Alzheimer's disease and Parkinson's disease (Huang et al., 2009; Zhang et al., 2019). Therefore, maintaining a stable supply of soybean is important for the treatment of complex diseases.

Even though global soybean production has steadily risen in recent years (FAOstat, faostat.fao.org), there is still a shortage of food supply for human beings due to increases in natural disasters. Soybean is stress-sensitive and is particularly affected by flooding (Hou and Thseng, 1991; VanToai et al., 2010), one of the major abiotic stresses that can cause enormous

losses in soybean production (Rosenzweig et al., 2002; Ahmed et al., 2013). In 2011, flooding and drought stress accounted for more than 70% of the reduction of major crops in the United States (Bailey-Serres et al., 2012). According to the degree of damage, flooding stress is classified into waterlogging and submergence, which means the water covers only the root system and the water covers both the shoot and the root system, respectively (Fukao et al., 2019). The roots and shoots of soybeans grow much more slowly if they are submerged in water (Oosterhuis et al., 1990). The nitrogen fixation in their root systems is also impeded (Board, 2008; Youn et al., 2008). As the time of submergence increases, it interferes with photosynthesis, stomatal conductance, and the absorption of nutrients (Jackson et al., 2009). Therefore, breeding work for flooding-tolerant soybean varieties is imperative.

Many studies have examined the selection of flood-tolerant soybean varieties. Shannon et al. (2005) conducted flood-tolerance experiments using 350 soybean germplasm lines and found that six lines (Archer, Misuzudaiz, PI408105A, PI561271, PI567651, and PI567343) were highly related to flooding tolerance during early reproductive stages. Wu et al. (2017) evaluated 722 soybean germplasms through foliar damage score and plant survival rate for accessing flooding tolerance during 2012 and 2016. Eleven flooding-tolerant lines (PI408105A, PI471931, PI471938, RA-452, Walters, R11-6870, R10-4892, R10-230, R07-6669, R07-2001, and R04-342) were identified which showed consistent flooding tolerance during 4- to 5- year continual evaluations. Meanwhile, the genetic information of several of these were involved with flood-tolerant related studies in soybean had been reported and accumulated. For instance, in gene expression studies, Chen et al. (2016) used RNA-Seq transcriptome profiling to identify differentially expressed genes and found that 3,498 of them were significantly associated with flooding tolerance. Since flooding tolerance was a quantitative trait, many previous studies used linkage mapping analysis to find quantitative trait locus (QTLs) (VanToai et al., 2001; Sayama et al., 2009; Nguyen et al., 2012, 2017). Cornelious et al. (2005) used 912 simple sequence repeat (SSR) markers to select QTLs and identified 17 SSR markers that were significantly associated with flooding tolerance in soybean. In the pathway function regulation platform, a total of 31 genes were identified under flooding stress by using quantitative reverse transcription-polymerase chain reaction (qRT-PCR). These genes were linked to the pathways including protein synthesis, nucleotide metabolism, hormone metabolism and glycolysis signaling which were induced by submergence conditions (Yin et al., 2017). Although previous studies were abundant, it was still costly and time-consuming for experimental validation of each of the flooding tolerance candidate genes (Rhee and Mutwil, 2014; Zhai et al., 2016). Furthermore, the data collected from previous studies had large batch effects (Goh et al., 2017), thus making it thorny to unify and integrate. Consequently, a gene prioritization technique, based on the computational method was developed.

The first research applying gene prioritization to plants was published by Xia et al. (2013). They utilized a keyword search in the NCBI PubMed database to collect data related to rice blight and used chaos-algorithm based classifiers to identify 74 blight resistance-related candidate genes in rice. Recently, network

analysis advanced the development of gene prioritization. Zhai et al. (2016) proposed rank aggregation-based data fusion for a gene prioritization (RAP) method that integrated RafSee and AraNet v2 prioritization algorithms, a total of 380 flowering-time genes of *Arabidopsis thaliana* were identified. They found that the prioritized genes identified by the RAP method had a higher ranking in comparison to those that identified by AraNet v2. However, the limitation of the RAP gene prioritization method is that the prediction ability decreases when the number of core genes is deficient. This indicates that the RAP gene prioritization method is not suitable for other crops with fewer known functional genes. In soybean, 59 prioritized genes were identified by using the SoyNet database, which contains 40,812 soybean genes and 1,940,284 links from several data platforms covering 72.8% of the soybean genome (Kim et al., 2017). Nevertheless, the prioritization of flooding tolerance genes (FTgenes) in soybean has not been elucidated to date.

In the present study, we mined and collected genetic information and databases from different data sources on flood-tolerant in soybean. Here we defined flooding tolerance genes (denoted as FTgenes) to be significantly associated with traits that are related to flooding tolerance or responded to the stress after flooding treatment was given during germination and vegetative stages of soybean. The flooding treatment mainly focused on submergence. Genes that regulated physiological mechanisms involving both submergence and waterlogging were also considered to extend the pool of genomic data collected from multiple dimensional data platforms.

To minimize possible selection bias and noise, we considered nine data platforms, including association mapping study (including genome-wide association study, GWAS), linkage mapping analysis, gene expression, pathway regulation, protein-protein interaction networks (PPIN), network analysis, proteomes, and text mining, as well as functional genomic data from model plants. A scoring and weighting scheme was proposed to extract, integrate, and prioritize genomic data across multidimensional data sources using meta-analysis and prioritization system. To avoid false positive results, all positive and negative results were considered and integrated to give a real value for every genetic locus to construct an evidence-based gene pool (denoted as test genes). We computed a weighted combined score summarized from multiple data sources for each of the test genes. Similarly, a set of core genes were established for prioritizing the test genes. A clear separation between the test genes and the core genes was determined to identify prioritized FTgenes. This provides a valuable contribution for subsequent research and soybean variety selection.

MATERIALS AND METHODS

We constructed a comprehensive framework to computationally optimize the test genes to construct prioritized FTgenes in soybean. There were four stages in the framework, including bioinformatic big data mining and integration, meta-analysis and gene prioritization, evaluation, and gene-set enrichment analysis. Detailed methods and materials used in this study were illustrated below (Figure 1).

Test Genes and Core Genes

The test genes and the core genes were generated at the bioinformatics big data mining and integration stage. The collection of the test genes used text/data mining from multiple dimensional data sources. For detailed methods and approaches please refer to our proposed techniques described below in the “Bioinformatics Big Data Mining and Integration” subsection. The establishment of the core genes was to provide an important basis for gene prioritization. The selection of the core genes was based on three criteria as follows. (1) Only genes that were reported in previous studies that significantly related to flooding tolerance or responded to flooding stress. These genes were also ranked in the top 2% of gene prioritization procedures. (2) Only genes that were significantly reported to be associated with flooding tolerance in network data platform. (3) Only genes that were significantly reported on at least four data platforms, where they scored higher than 40 in a weighted score and scored higher than 4 in each of the data platforms. We used the core genes to prioritize test genes according to their magnitude of association or response change to the stress from multiple data sources to search for prioritized FTgenes.

Validation Databases

We used the whole genome expression database (RNA-seq data) of soybean seedling submergence published by Lin et al. (2019) as our validation databases. They investigated and recorded the expression change of soybean roots at four different time-points (3, 6, 12, and 24 h) after given submergence treatments. A total

of 14,772, 17,017, 19,060, and 18,889 gene expression data were contained in the four time-point databases, respectively.

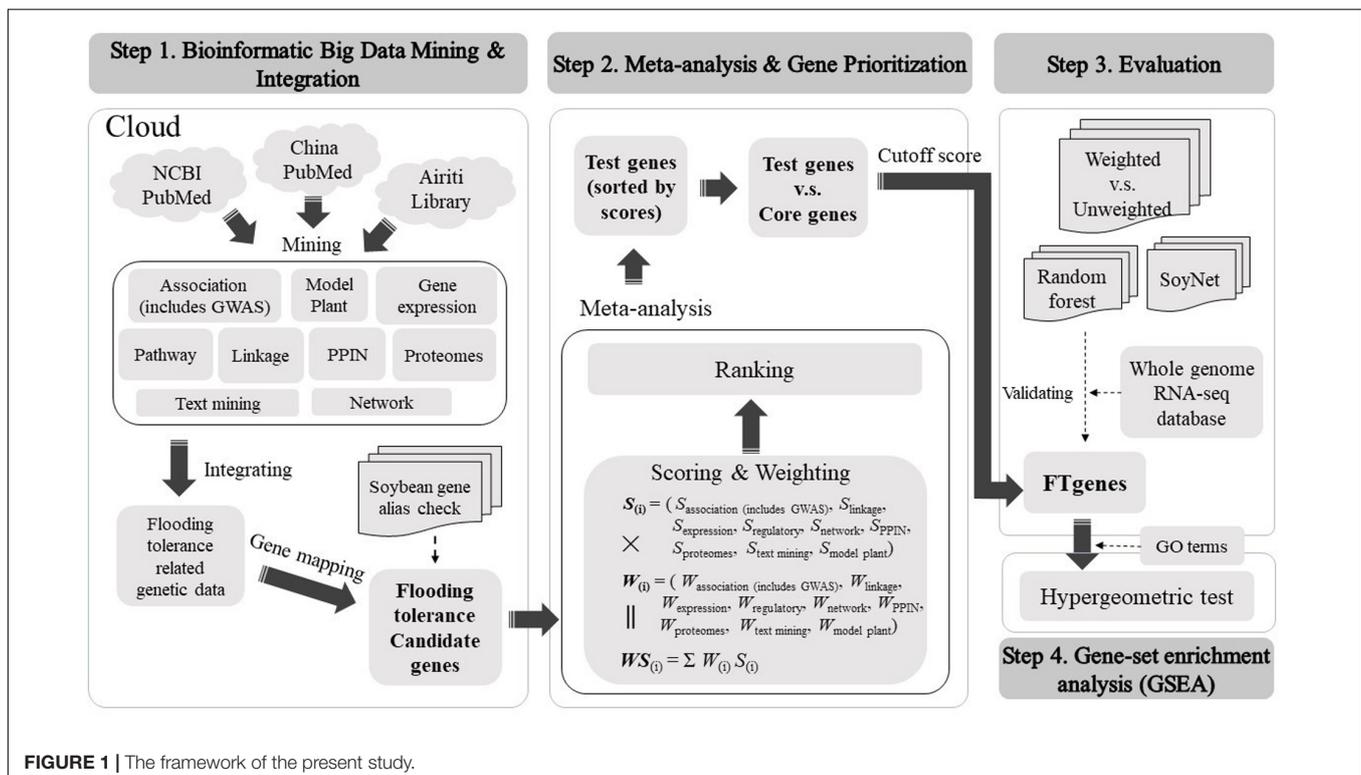
Bioinformatics Big Data Mining and Integration

We collected genomic data that related to flooding tolerance or responded to the stress in soybean from the National Center for Biotechnology Information (NCBI) PubMed¹, Airiti Library², and China PubMed³. A keyword search method was applied to mine genomic data extracted from published journal articles and available open databases that were published before March 2020 that were relevant to our target. The keywords consist of combinations of three terms (crop, trait, and data platform). The keywords for “crop” terms were “soybean,” “cultured soybean,” and “glycine max.” The keywords for “trait” terms were “flooding tolerance,” “flooding stress,” “waterlogging,” and “submergence.” The keywords for “data platform” were “GWAS,” “association mapping,” “linkage mapping,” “gene expression,” “pathway regulation,” “PPIN,” “networks,” “proteomes,” and “text mining.” All genomic data, including journal articles and databases, collected from the cloud were examined and integrated carefully by experienced experts who are well-trained in big data mining for data management and data quality control. Studies that involved human trails, animal experiments, genetically modified organism studies, non-soybean studies, not applicable, and others that

¹<https://www.ncbi.nlm.nih.gov/>

²<https://www.airitilibrary.com/>

³<https://www.pubmedchina.com/>



were irrelevant to the present study were excluded from big data integration.

The selection criteria of genomic data are illustrated below. In the association mapping data platform (including GWAS), only SNP markers having minor allele frequency (MAF) >5%, heterozygous allele calls <10%, call rate >90%, and P -value <0.05 were included. In the gene expression data platform, only genes with a P -value <0.05 and/or fold change (FC) >1.5 were considered. In linkage mapping and proteomes data platforms, only genetic data that P -value < 0.05 and/or logarithm of odds (LOD) >3 were mined. In the networks data platform, only genes that P -value <0.05 were collected. In the pathway regulation data platform, only pathways that were significantly reported to be associated with flooding tolerance were considered and we then screened out the genes involved in the pathways. In text mining data platform, we searched the keyword combinations of “gene symbol + crop + trait” using both MySQL and R package for parallel web crawling in the NCBI PubMed, Airiti Library, and China PubMed. The gene symbols were downloaded from SoyBase⁴. A hit was made if the journal article matched the keyword combinations. In the PPIN data platform, a total of 717,676 protein-protein interactions (regulations) were downloaded from the PlantRegMap⁵ in February 2020. We considered both positive findings and negative results to minimize possible false positive results and false negative results so that a precise evidence-based score for each genomic data could be calculated with fewer biases and noise.

To extend our method to summarize the roles of selected genes in regulating tolerance by integrating available genes from model plants, we included homologous genes from *A. thaliana* and *Medicago truncatula*. For the homologous gene data platform, protein sequences from the *Glycine max* genome were, respectively, aligned to those from the two model plants using BLASTP (Camacho et al., 2009). All protein sequence data, integrated from several studies (Schmutz et al., 2010; Tang et al., 2014; Cheng et al., 2017), were downloaded from the phytozome database⁶ (Goodstein et al., 2012). We searched for sequences with the highest similarity from alignment results to identify homologous proteins or genes corresponding to soybean genes. These homologous genes were further confirmed whether they were significantly reported as flooding related genes in *A. thaliana* and *M. truncatula*.

The genomic data collected in the present study included SNPs, genes, SSRs, and QTLs. We conducted gene mapping to assign all genomic data to a gene region using a window spanning 20 kb upstream to 20 kb downstream of the gene (Ravelombola et al., 2019). The principle of gene mapping was shown in **Supplementary Figure 1**. We notice that regions of QTL with longer than 5 centiMorgan (cM) in length were excluded (Liu et al., 2019) from big data integration to reduce

possible false positive results and noise, in particular, QTLs across two different chromosomes. All genomic data were mapped into gene-level data corresponding to the correct gene version that was used in the original studies. Finally, we conducted gene version correspondence analysis to unambiguously match the above mapped gene-level data using the SoyBase Gene Model Correspondence Lookup tool⁷ to unify various gene versions (Glyma v1.0, Glyma v1.1, and Glyma v2.0) into Glyma v2.0 gene version. As a result, a set of test genes in the Glyma v2.0 version were established.

Meta-Analysis and Gene Prioritization

We developed a scoring scheme according to the corresponding magnitude of association of flooding tolerance or response to flooding stress in soybean to address the issue of different types of data sources. The evidence from multidimensional data sources was in a wide range of different types of values, including association P -value, LOD, FC, score, degree, cluster coefficient, and correlation (r). The scoring systems for different platforms are different. Generally, we applied $-\log_{10}(P\text{-value})$, [FC] and [LOD] to separately convert P -value, FC, and LOD into a score bounded within [0, 10]. The symbol of $\log(\cdot)$, $|\cdot|$, and $[\cdot]$ represents 10-based logarithm, absolute value, and floor function, respectively. We denoted S^M as a scoring system computed by using a transformation function described above, where the superscript letter M is a measurement characterized by a P -value, FC, LOD, degree, and r . In the association mapping data platform, we applied $S^P = -\log_{10}(P\text{-value})$ for scoring. In the linkage mapping and proteomes data platforms, we calculated $\max\{S^P, S^{FC}, S^{degree}\}$, and $\max\{S^P, S^{FC}, S^{degree}\}$ separately for loci, respectively. In gene expression data platform, the $\max\{S^P, S^{FC}\}$ was applied for scoring. In the networks data platform, we quantified $\max\{S^P, S^{FC}, S^r\}$ to obtain a score range [1,6]. In the pathway regulation data platform, we proposed a reported frequency-based algorithm for the scoring of selected genes. In text mining and model plant data platforms, we scored 1 if the keyword search was made or the homologous gene was confirmed and scored 0 otherwise. In the PPIN data platform, we scored the regulation networks according to the strength of degree and cluster coefficient. If two or more types of different data were present for a gene within the same data platform, we computed a score for each data type and selected the maximal value as the score for the gene. A detailed scoring system can be found in **Supplementary Table 1**.

In the stage of weighting, we evaluated each of the selected genomic data using an impact factor corresponding to their published journal articles to access data reliability (**Supplementary Table 2**). For each gene, we calculated a weighted score using score multiply weight across multiple data platforms to quantify the importance of gene linking to flooding tolerance in soybean. Generally, the distribution of weighted scores for the test genes and the core genes are skewed to the right and skewed to the left, respectively. Hence, an optimal cutoff-score can be found to distinctly separate the two distributions of

⁴<https://soybase.org/dlpages/#annot>

⁵<http://plantregmap.cbi.pku.edu.cn/download.php#networks>

⁶<https://phytozome.jgi.doe.gov/pz/portal.html>

⁷<https://soybase.org/correspondence/index.php>

the test genes and the core genes. Only genes that scored greater than the cutoff-score were selected as prioritized FTgenes.

Evaluation

Two approaches were applied to access the reliability and the robustness of the prioritized FTgenes using the RNA-seq data (Lin et al., 2019). First, we compared the prioritized FTgenes using a weighted scheme to examine whether the prioritized FTgenes showed a higher change to obtain smaller *P*-values (or larger FCs) than those using the unweighted scheme. Comparisons of the intersection, the difference of unweighted FTgenes from weighted FTgenes (denoted as FTgenes^{w\^nw}), and the difference of weighted FTgenes from unweighted FTgenes (denoted as FTgenes^{nw\^w}) were also examined.

We also compared the prioritized FTgenes (using weighted scheme) selected by our proposed algorithm with other existing methods, e.g., random forest algorithm (Rafsee) (Zhai et al., 2016) and network-based gene prioritization method (SoyNet) (Kim et al., 2017). Due to the difference of sources and characteristics for data mining between other methods and our study, we modified the random forest algorithm used in Rafsee without changing the idea and concept and applied it to our test genes. We modified the random forest algorithm as follows. (1) Bootstrap: we generated a subset of the same size from the test genes by sampling with replacement, and selected the top 83 genes with the highest weighted score (denoted as bootstrap genes). (2) Permutation: we randomly shuffled (i.e., under the null) the order of genes to break inherent structure of dependence between genes and the scores in each bootstrap subset, and selected the top 83 genes with highest weighted score (denoted as permutation genes). (3) Selection: we selected the bootstrap genes if the weighted score of the top 83 bootstrap gene was greater than the highest weighted score in permutation genes, and discarded them otherwise. (4) Loop: we repeated steps 1–3 until 10,000 sets of bootstrap genes were achieved. (5) Ranking: we counted gene frequencies for each set of 10,000 bootstrap genes, and selected the top 83 bootstrap genes with the highest frequencies as prioritized genes. The second method compared to our algorithm was SoyNet (Kim et al., 2017), which

is a co-functional network webserver⁸. The SoyNet contained 40,812 soybean genes and 1,940,284 links collected from 21 distinct data types, covering 72.8% of the soybean genome. We conducted gene prioritization based on Bayesian statistics using the top 2,000 weighted genes using function II “Find context associated genes.” We selected the top 83 prioritized genes from 502 significant genes (*P*-value < 1.0 × 10⁻⁸). We compared our prioritized FTgenes to those two top genes prioritized using other methods described above for validation using the Wilcoxon rank-sum test, and a *P*-value was calculated based on 100,000 bootstrap samples for each method.

Gene-Set Enrichment Analysis

The GeneOntology (GO; <http://geneontology.org/>) database collected abundant terms that were related to gene functions in soybean. According to the gene products, these terms can be classified into three categories: biological process (7,332 terms), cellular component (2,761 terms), and molecular function (3,199 terms). We applied gene-set enrichment analysis to investigate significantly enriched potential physiological regulation pathways, using the GO database and the prioritized FTgenes. The hypergeometric test was conducted to identify significantly enriched pathways, and adjusted *P*-values were calculated using the Bonferroni correction method to avoid false positive results.

RESULTS

Bioinformatics Big Data Mining and Integration

In bioinformatics, big data mining, 66 journal articles, and 4 databases were qualified. Through quality control, we removed 17 articles irrelevant to our study, there were 49 journal articles, and 4 databases in the end (Table 1). We found 47,227, 2,014, 59, 47,931, and 376 genes, respectively, from gene expression, pathway regulatory, networks, PPIN, and the proteomes data platform. 79 SNPs and 66 SSR

⁸www.inetbio.org/soynet

TABLE 1 | Summary results of big data mining and bioinformatics across different data platforms for flooding tolerance in soybean.

Data platform	Data mining results		Selected data sources		No. of genetic data extracted ^a
	No. of articles	No. of databases	No. of articles	No. of databases	
Association mapping (includes GWAS)	8	0	2	0	79
Linkage mapping	10	0	7	0	66
Gene expression	11	2	9	2	47,227
Pathway regulatory	21	1	15	1	2,014
Gene networks	1	0	1	0	59
PPIN	0	1	0	1	47,931
Proteomes	4	0	4	0	376
Homologous gene	11	0	11	0	8,511
Total	66	4	49	4	106,263

Abbreviation: GWAS, genome-wide association study; PPIN, protein-protein interaction network.^aGenetic data include SNPs, genes, SSRs, and QTLs.

markers were found from association mapping and linkage mapping data platforms, respectively. A total of 106,263 genotype data were included in this study. Two RNA-seq databases (46,938 genes) (Nanjo et al., 2011; Chen et al., 2016) and one GO pathway regulation database (2,014 genes in 14 pathways) were included. In the PPIN database, a total of 717,676 gene pairs were included (PlantRegMap: <http://plantregmap.cbi.pku.edu.cn/>) (**Supplementary Table 3** and **Supplementary Material 1**).

In the stage of bioinformatics big data integration, genetic data including SNPs, SSRs, and QTLs are required to perform gene mapping so that every loci was mapped onto the gene level. A total of 79 SNPs and 66 SSRs were mapped to 117 genes and 296 genes, separately in association mapping and linkage mapping data platform. For the homologous gene data platform, a total of 34,738 and 50,188 potential homologous genes, respectively, from *A. thaliana* and *M. truncatula* were identified using BLASTP. To confirm these genes whether associated with flooding, we collected flooding-related studies in two model plants to extract their candidate genes. In a total of 8,511, *A. thaliana* candidate genes were found in 11 papers. Unfortunately, we did not find any genetic information in *M. truncatula*. Next, the overlap between homologous and candidate genes was verified to compute homo.score for tested genes. Our results showed that 11,275 tested genes were reported in previous studies that related to flooding stress. In total, 36,705 test genes were constructed from multidimensional data platforms for meta-analysis and gene prioritization.

Table 2 shows the summary results of genomic data on flooding tolerance or response to flooding stress indices at a particular growth stage in soybean. In the germination stage, differentially expressed genes in root and/or leaf tissues were measured in gene expression, pathway regulators, and proteomes studies. During germination-reproductive stages, three indices (flooding injury scores, germination rate, and normal seeding rates) were often used to evaluate the degree of flooding tolerance in linkage and association mapping studies.

Meta-Analysis and Gene Prioritization

There are three steps in big data meta-analysis of FTgenes in soybean, including scoring, weighting, and ranking (**Figure 1**). Different scoring schemes were set to score test genes for each data source (**Supplementary Table 1**). In the association mapping data platform, the scores of 117 genes were between 4.01 and 10. In the linkage mapping data platform, 296 genes were scored between 1 and 10. In the gene expression data platform, the scores of 47,227 genes ranged between 1 and 6.45. In the pathway regulation data platform, the scores of 2,014 genes were between 5 and 6. In the PPIN data platform, the scores of 47,931 genes were between 1 and 6. In the networks data platform, the scores of 59 genes were ranged from 0 to 3. In the proteomes data platform, the scores of 376 genes were between 1.3 and 10.

In terms of weighting and weighted score (**Supplementary Table 2**), first of all, the range of weighting in the association mapping data platform is between 0 and 3. In the linkage mapping data platform, the range of weighting is between 0 and 4. In the gene expression and networks data platforms, the ranges

of weighting were between 0 and 6. In pathway regulation and proteomes data platforms, the range of weighting is between 0 and 3. In the PPIN data platform, the range of weighting is between 0 and 2. Finally, after weighting by the impact factor, the weighted score of 36,705 flooding tolerance test genes was between 11.28 and 80.41. Furthermore, to identify FTgenes, a total of 28 test genes that passed the criteria were selected as the core genes (**Supplementary Tables 4, 5**).

In gene prioritization, after calculating the weighted score for 36,705 test genes and 28 core genes, we compared the distribution of these two data sets. A trivial separation was observed between the test gene set and core gene set at a cutoff score of 45 (**Figure 2**), and a total of 83 genes were chosen as FTgenes (**Table 3**). The physical location and numbers of 83 FTgenes in the soybean genome are shown in **Figure 3** and **Supplementary Figure 2**. Among them, chromosome 13 (14 FTgenes), 8 (13 FTgenes), and 7 (9 FTgenes) contain the most FTgenes. However, there are no FTgenes located on chromosomes 15 and 17.

Evaluation of Prioritized Genes

To evaluate the prioritized FTgenes, we compared two sets of FTgenes using a weighted scheme and an unweighted scheme (**Supplementary Table 6**). **Table 4** showed the validation results. First, the 83 weighted FTgenes had significantly smaller *P*-values (higher FCs) than 100,000 random sets (3, 6, and 12 h: *P*-values $<1 \times 10^{-5}$, 24 h: *P*-value <0.01) using an independent whole genome RNA-seq database (Lin et al., 2019). However, the 83 unweighted FTgenes had significantly smaller *P*-values (higher FCs) than random sets at only 3 h and 6 h (*P*-values <0.05). The same situation was also observed in the intersection set (74 out of 83 genes, 89.16%) between weighted FTgenes and unweighted FT genes (**Supplementary Table 10**). Interestingly, we found that the set of FT genes^{w\^nw} had significantly smaller *P*-values (higher FCs) than 100,000 random sets at 3, 6, and 12 h (*P*-values <0.05), but the set of FTgenes^{nw\^w} showed no significant difference at 3, 6, 12, and 24 h. This indicates that the set of FTgenes^{w\^nw} may play roles in regulating mechanisms responding to flooding stress, suggesting that the use of a weighted scheme would provide informative and meaningful results.

The whole genome RNA-seq database was also used to evaluate 83 prioritized genes identified by the SoyNet network method (Kim et al., 2017). The results showed that there was no significant difference in FC between the prioritized set and 100,000 random sets. The proportion of *P*-value less than 0.05 does not exceed 81% in the four time points (**Table 4**). On the other hand, to compare the random forest method with our prioritized algorithm, the modified random forest method was conducted to satisfy our data format (Zhai et al., 2016). As a result, a total of 83 prioritized genes (ranked in the top 83 genes) were also identified based on the modified random forest method. Of which, the 83 prioritized genes were the same as our prioritized 83 FTgenes identified via an algorithm. The result suggested that the performance for our algorithm is as good as the random forest method, indicating our resulting 83 FTgenes are precise and reliable, which means our 83 weighted FTgenes had a

higher chance to be involved in the physiological mechanism of flooding tolerance in soybean.

Gene-Set Enrichment Analysis

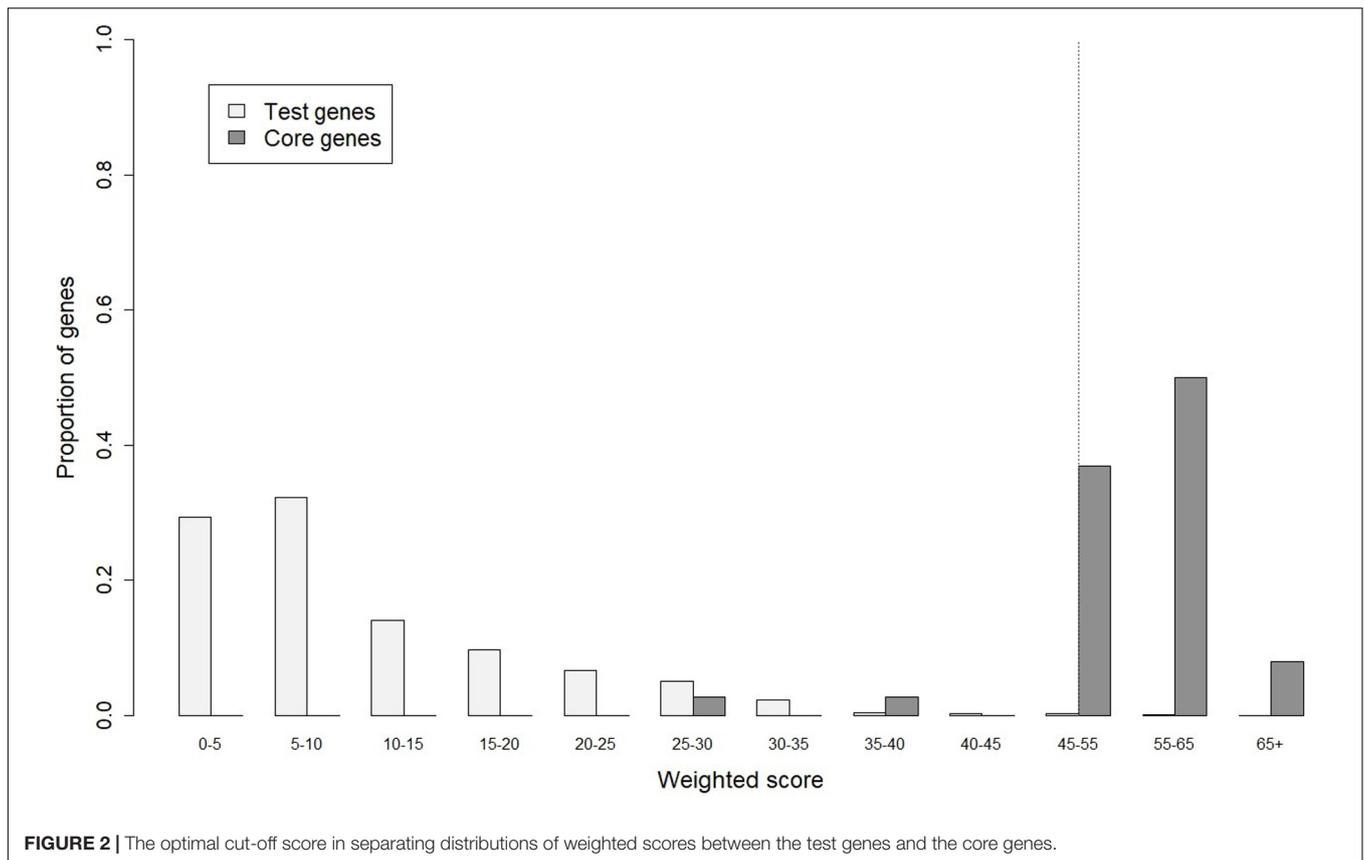
To understand the physiological mechanisms of 83 FTgenes in soybean under flooding tolerance, we used the GO enrichment tool for gene-set enrichment analysis through SoyBase website. As shown in **Table 5**, the 83 FTgenes

are significantly enriched in seven biological processes and molecular functions ($P_{adjusted}$ – value <0.05). Our results suggested that the 83 FTgenes were significantly enriched in ethylene biosynthetic process, abscisic acid (ABA) biosynthetic process, nuclear-transcribed mRNA poly(A) tail shortening, glucan biosynthetic process, ethylene mediated signaling pathway, phosphorylation, and response to hypoxia.

TABLE 2 | Summary results of flooding tolerance index from each data platform.

Platform	Growth stage	Flooding tolerance related traits	References
Association (including GWAS)	Reproductive stage	Flooding injury scores, level 1–5	Ye et al. (2018)
Association (including GWAS)	Germination stage	Germination rate, normal seedling rate, electric conductivity	Yu et al. (2019)
Association (including GWAS)	Reproductive stage	Foliar damage score	Wu et al. (2019)
Gene expression	Germination stage	Collect root and hypocotyl and detect the expression of RNAs and proteins	Komatsu et al. (2009)
Gene expression	Germination stage	Extract RNA form roots (including hypocotyl)	Nanjo et al. (2011)
Gene expression	Germination stage	Extract RNA form roots (including hypocotyl)	Nishizawa et al. (2013)
Gene expression	Germination stage	Extract RNA form roots (including hypocotyl)	Song et al. (2018)
Gene expression	V1_Vegetative stage	Extract RNA from root tissues	Nakayama et al. (2014)
Gene expression	V1_Vegetative stage	qRT-PCR for roots	Valliyodan et al. (2014)
Gene expression	V4_Vegetative stage	Use leaves to do RNA-sequencing and real-time PCR analysis	Chen et al. (2016)
Gene expression	V5_Vegetative stage	Use leaves to do RNA-sequencing, RT-PCR, and qPCR	Syed et al. (2015)
Gene expression	Vegetative stage	Extract RNA from root tissues	Nakayama et al. (2017)
Linkage mapping	Germination stage	Germination rate, healthy growth rate, damage of roots and shoots	Rizal and Karki (2011)
Linkage mapping	Germination stage	100 seed weight, germination rate, normal seedling rate	Sayama et al. (2009)
Linkage mapping	R1_Reproductive stage	Plant height, seed yield	VanToai et al. (2001)
Linkage mapping	R1_Reproductive stage	Flooding tolerance score, score1–5	Nguyen et al. (2012)
Linkage mapping	R2_Reproductive stage	Plant injury score (score 1–9)	Cornelious et al. (2005)
Linkage mapping	V3_Vegetative stage	Flooding tolerance index (seed weight of treated plants divided by that of control plants), days to flowering, days to maturity, plant height, branch number, pod number, seed number, seed weight, 100-seed weight	Githiri et al. (2006)
Pathway	Germination stage	Hypocotyl	Khan et al. (2015)
Pathway	Germination stage	Roots	Khatoon et al. (2012)
Pathway	Germination stage	Roots	Komatsu et al. (2013)
Pathway	Germination stage	Roots	Mustafa and Komatsu (2014)
Pathway	Germination stage	Roots and cotyledons	Mustafa et al. (2015)
Pathway	Germination stage	Cotyledons	Nanjo et al. (2010)
Pathway	Germination stage	Roots including hypocotyl	Nanjo et al. (2011)
Pathway	Germination stage	5 mm long root tips	Nanjo et al. (2012)
Pathway	Germination stage	Root tips	Yin and Komatsu (2016)
Pathway	Germination stage	Root tips	Yin et al. (2014)
Pathway	Germination stage	Root tips	Yin and Komatsu (2015)
Pathway	Germination stage	Root tips	Yin et al. (2016)
Pathway	Germination stage	Root tips, including hypocotyl and cotyledons	Yin et al. (2017)
Proteomes	Germination stage	Plant survival rate, lacking apparent damage, lateral root development, radicle elongation rate	Nanjo et al. (2014)
Proteomes	Germination stage	Roots, hypocotyls, cotyledons	Oh and Komatsu (2015)
Proteomes	V2_Vegetative stage	Leaf fresh weight, leaf dry weight, leaf turgid weight, chlorophyll content chlorophyll content data	Mutava et al. (2015)
Proteomes	Vegetative stage	Root and leaf samples	Kazemi Oskuei et al. (2017)

qRT-PCR, quantitative reverse transcription polymerase chain reaction; RT-PCR, reverse transcription polymerase chain reaction.



DISCUSSION

In the present study, we proposed a comprehensive framework that consists of bioinformatics big data mining, meta-analysis, and a gene prioritization algorithm to prioritize 83 FTgenes from 36,705 test genes set collected from multidimensional data platforms. We collected bioinformatics information including trait index, genetic data (SNP, gene, SSR, loci, and QTLs), variety, biochemical and statistical value (P -value, LOD, FC, score, R^2 , and keyword hits). The impact factors of journal articles were also collected for determining the reliability and quality of data. Our data was collected from a variety of countries, using different methods, plant materials, and genotype data, which is a diverse and informative database. There were three strengths to our gene prioritization method. First, we used quality control in big data collection to reduce the influence of noises effectively. Second, regardless of the positivity and negativity of the genotype data, we aimed to minimize the impact of publication bias. Third, we checked and unified gene aliases to avoid overestimations or underestimations that may occur in the calculation of the weighted score. Therefore, 83 FTgenes were found to have good characteristics of a more comprehensive, higher accuracy, and less bias and noise. Compared with traditional methods, this gene prioritization algorithm is more informative.

In the process of collecting bioinformatics big data, we found that GWAS had not been popular in the field of flooding tolerance in soybean, only two journal articles were found

in the GWAS data platform (**Table 1**) (Wu et al., 2019; Yu et al., 2019). GWAS enables us to select the molecular markers that are significantly related to target traits. However, the main limitations of GWAS are low effect size and spurious associations. To cope with these problems, our research collected and integrated flooding tolerance candidate genes from different data sources and attempted to find high-ranking genes with a gene prioritization algorithm. Generally, high-ranking FTgenes should be reported in multiple data platforms, suggesting more potential that these are related to flooding tolerance compared to other genes. In our framework of analytic strategy, we can effectively reduced the chance of false positive results and increased the effect size that GWAS may encounter (Tam et al., 2019). In text mining, we searched the title and abstract of keywords (crop, gene, trait) with structured query language and R programming language, and found no results. We searched the full text alternatively, but most of the results were still irrelevant to the target. Therefore, in our opinion, text mining was not suitable for searching in the plant field at present.

The expression level of *Glyma.04g240800*, one of the 28 core genes, was repeatedly used to index the response of flooding stress of soybean in previous studies. *Glyma.04g240800* is one of the alcohol dehydrogenase (ADH) which participated in flooding tolerance. Flooding causes oxygen deprivation and forthwith activates anaerobic respiration in soybean. ADH reduces NAD^+ to NADH in glycolysis, which is the first step of anaerobic

TABLE 3 | Eighty-three prioritized FTgenes (weighted score ≥ 45).

Gene	Weighted score	Flooding treatment	Gene	Weighted score	Flooding treatment
<i>Glyma.13g243800</i>	81.99	Submergence, waterlogging	<i>Glyma.08g119200</i>	49.19	Submergence, waterlogging
<i>Glyma.13g244000</i>	80.41	Submergence, waterlogging	<i>Glyma.14g202300</i>	49.00	Submergence
<i>Glyma.11g055700</i>	66.24	Submergence	<i>Glyma.08g119600</i>	48.65	Submergence, waterlogging
<i>Glyma.13g244100</i>	63.75	Submergence, waterlogging	<i>Glyma.08g119000</i>	48.10	Submergence, waterlogging
<i>Glyma.13g243700</i>	62.84	Submergence, waterlogging	<i>Glyma.13g251100</i>	48.05	Submergence
<i>Glyma.14g121200</i>	62.81	Submergence	<i>Glyma.10g180800</i>	48.02	Submergence
<i>Glyma.03g132700</i>	62.81	Submergence	<i>Glyma.08g083300</i>	47.81	Submergence
<i>Glyma.11g179300</i>	62.81	Submergence	<i>Glyma.14g049000</i>	47.81	Submergence
<i>Glyma.02g148200</i>	62.81	Submergence	<i>Glyma.07g105700</i>	47.81	Submergence
<i>Glyma.12g094100</i>	62.81	Submergence	<i>Glyma.12g093100</i>	47.81	Submergence
<i>Glyma.13g243600</i>	61.21	Submergence, waterlogging	<i>Glyma.09g172500</i>	47.81	Submergence
<i>Glyma.13g243900</i>	61.00	Submergence, waterlogging	<i>Glyma.02g268200</i>	47.81	Submergence
<i>Glyma.05g108900</i>	59.81	Submergence, waterlogging	<i>Glyma.03g173300</i>	47.81	Submergence
<i>Glyma.16g018500</i>	59.81	Submergence, waterlogging	<i>Glyma.18g042100</i>	47.81	Submergence
<i>Glyma.07g036400</i>	59.11	Submergence, waterlogging	<i>Glyma.17g174500</i>	47.81	Submergence
<i>Glyma.08g128500</i>	58.40	Submergence, waterlogging	<i>Glyma.11g181200</i>	47.81	Submergence
<i>Glyma.08g128100</i>	58.19	Submergence, waterlogging	<i>Glyma.06g100900</i>	47.81	Submergence
<i>Glyma.10g048000</i>	57.16	Submergence, waterlogging	<i>Glyma.05g128200</i>	47.81	Submergence
<i>Glyma.09g153900</i>	56.30	Submergence	<i>Glyma.07g049900</i>	47.81	Submergence
<i>Glyma.16g204600</i>	55.96	Submergence	<i>Glyma.04g092100</i>	47.81	Submergence
<i>Glyma.12g150500</i>	54.95	Submergence	<i>Glyma.17g158100</i>	47.81	Submergence
<i>Glyma.13g208000</i>	54.63	Submergence	<i>Glyma.07g253700</i>	47.81	Submergence
<i>Glyma.08g199800</i>	54.11	Submergence	<i>Glyma.09g149200</i>	47.81	Submergence
<i>Glyma.12g222400</i>	53.87	Submergence	<i>Glyma.20g218100</i>	47.81	Submergence
<i>Glyma.10g048100</i>	53.25	Submergence, waterlogging	<i>Glyma.12g187400</i>	47.81	Submergence
<i>Glyma.07g036300</i>	52.67	Submergence, waterlogging	<i>Glyma.08g119100</i>	47.64	Submergence, waterlogging
<i>Glyma.10g047800</i>	52.63	Submergence, waterlogging	<i>Glyma.11g180500</i>	47.52	Submergence
<i>Glyma.10g048200</i>	52.48	Submergence, waterlogging	<i>Glyma.08g119400</i>	47.29	Submergence, waterlogging
<i>Glyma.05g123900</i>	52.14	Submergence	<i>Glyma.13g279900</i>	47.09	Submergence
<i>Glyma.01g118000</i>	52.09	Submergence	<i>Glyma.19g213300</i>	46.87	Submergence
<i>Glyma.07g036200</i>	51.70	Submergence, waterlogging	<i>Glyma.08g176300</i>	46.85	Submergence
<i>Glyma.05g124000</i>	51.54	Submergence	<i>Glyma.08g119500</i>	46.70	Submergence, waterlogging
<i>Glyma.07g036100</i>	51.13	Submergence, waterlogging	<i>Glyma.13g234500</i>	46.47	Submergence
<i>Glyma.04g044900</i>	50.81	Submergence	<i>Glyma.10g073600</i>	46.36	Submergence
<i>Glyma.08g218600</i>	50.81	Submergence	<i>Glyma.13g251300</i>	46.20	Submergence
<i>Glyma.13g250400</i>	50.80	Submergence	<i>Glyma.13g250300</i>	46.13	Submergence
<i>Glyma.07g032900</i>	50.39	Submergence, waterlogging	<i>Glyma.01g037200</i>	46.13	Submergence
<i>Glyma.10g047900</i>	50.24	Submergence, waterlogging	<i>Glyma.19g174200</i>	46.00	Submergence
<i>Glyma.13g270100</i>	50.13	Submergence	<i>Glyma.11g149900</i>	45.96	Submergence
<i>Glyma.07g153100</i>	50.03	Submergence	<i>Glyma.17g020600</i>	45.85	Submergence
<i>Glyma.18g009700</i>	49.70	Submergence	<i>Glyma.17g205000</i>	45.81	Submergence
<i>Glyma.08g139100</i>	49.57	Submergence, waterlogging			

respiration (Russell et al., 1990; Komatsu et al., 2009; Tucker et al., 2011; Nakayama et al., 2014; Song et al., 2018).

Of the 83 FTgenes, chromosome 13 contains the most FTgenes (**Supplementary Figure 2**), which reflects the results of a previous study (Yu et al., 2019). In chromosome 13, six FTgenes (*Glyma.13g243800*, *Glyma.13g244000*, *Glyma.13g244100*, *Glyma.13g243700*, *Glyma.13g243600*, and *Glyma.13g243900*) are in the top 12. Furthermore, the SNP marker QTN13 was reported to be remarkably related to flooding resistance (Yu et al., 2019), and these six genes are located within a 1.0 Mb

region where QTN13 has extended the region of 500 kb upstream and downstream on both sides. The third-ranked FTgenes *Glyma.11g055700* was reported to show significant performance under flooding conditions (P -value = 0.00005) (Chen et al., 2016), and also reported participating in the ABA biosynthetic process in SoyBase. The 29th FTgene *Glyma.05g123900* was reported on four data platforms, which were gene expression, pathway regulation, PPIN, and proteomes. Moreover, this gene was also reported to show significant protein expression under flooding conditions (P -value = 5.57×10^{-10}) and participated in

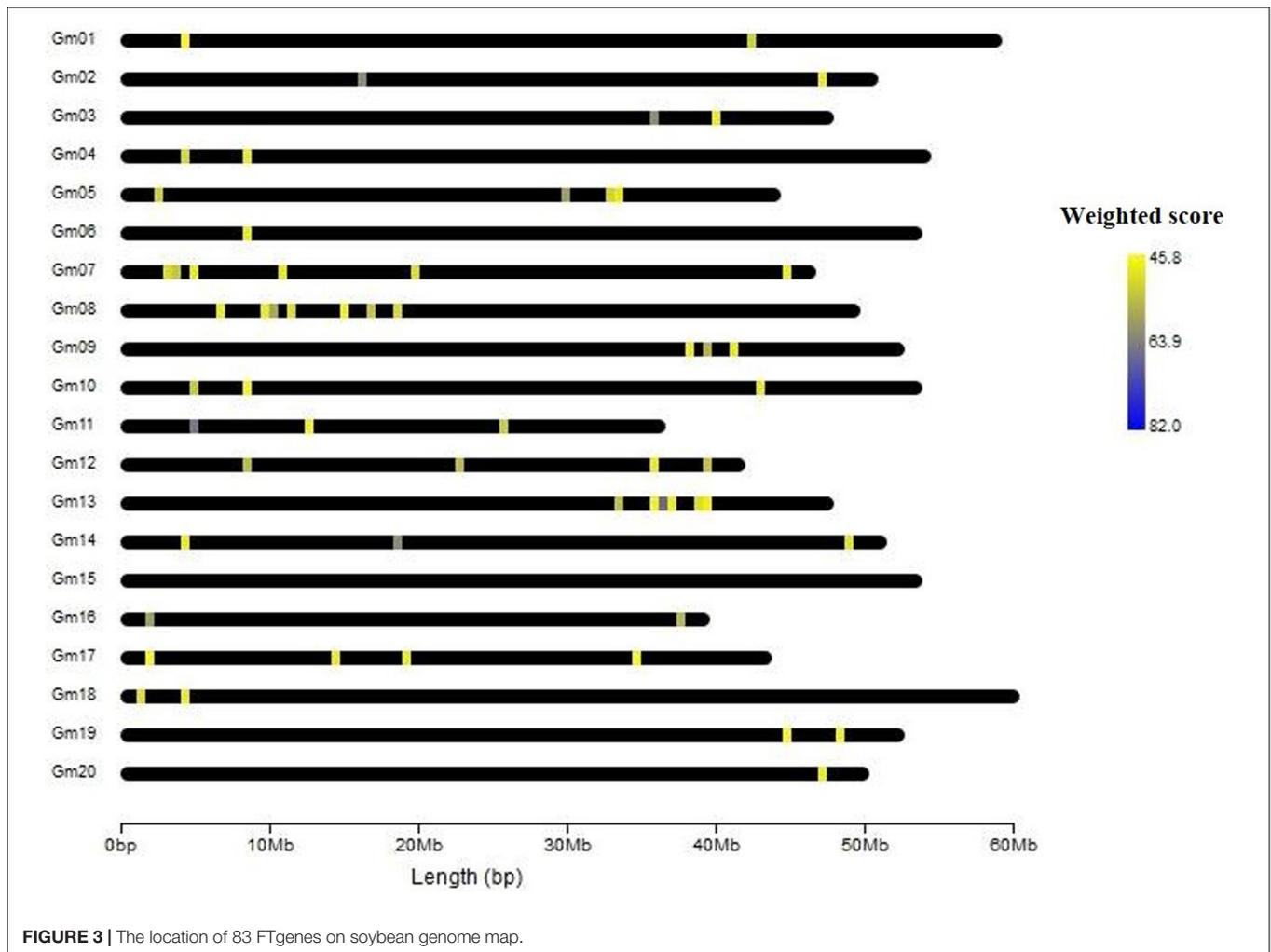


TABLE 4 | The results of comparing fold change after 3, 6, 12, and 24 hours of flooding treatment between the prioritized set and the random sets using the Wilcoxon rank-sum test.

Method	Number of candidate genes	P-value			
		3 h	6 h	12 h	24 h
FTgenes (weighted)	83	$<1 \times 10^{-5}$	$<1 \times 10^{-5}$	$<1 \times 10^{-5}$	0.01
FTgenes (unweighted)	83	0.01	0.04	0.12	0.90
FTgenes (weighted \cap unweighted) ^a	74	0.01	0.04	0.13	0.87
FTgenes (weighted \ unweighted) ^b	9	0.01	0.01	0.05	0.43
FTgenes (unweighted \ weighted) ^c	9	1.00	1.00	1.00	1.00
SoyNet ^d	83	0.60	0.03	0.35	0.99
Random forest	83	$<1 \times 10^{-5}$	$<1 \times 10^{-5}$	$<1 \times 10^{-5}$	$<1 \times 10^{-5}$

^aFTgenes (weighted \cap unweighted) represents the intersection of weighted genes and unweighted genes.

^bFTgenes (weighted \ unweighted) represents the difference of unweighted FTgenes from weighted FTgenes.

^cFTgenes (unweighted \ weighted) represents the difference of weighted FTgenes from unweighted FTgenes.

^dSoyNet represents the network analysis method used in Kim et al. (2017).

the MAPK signaling pathway and plant pathogen interaction regulation pathway (Lin et al., 2019).

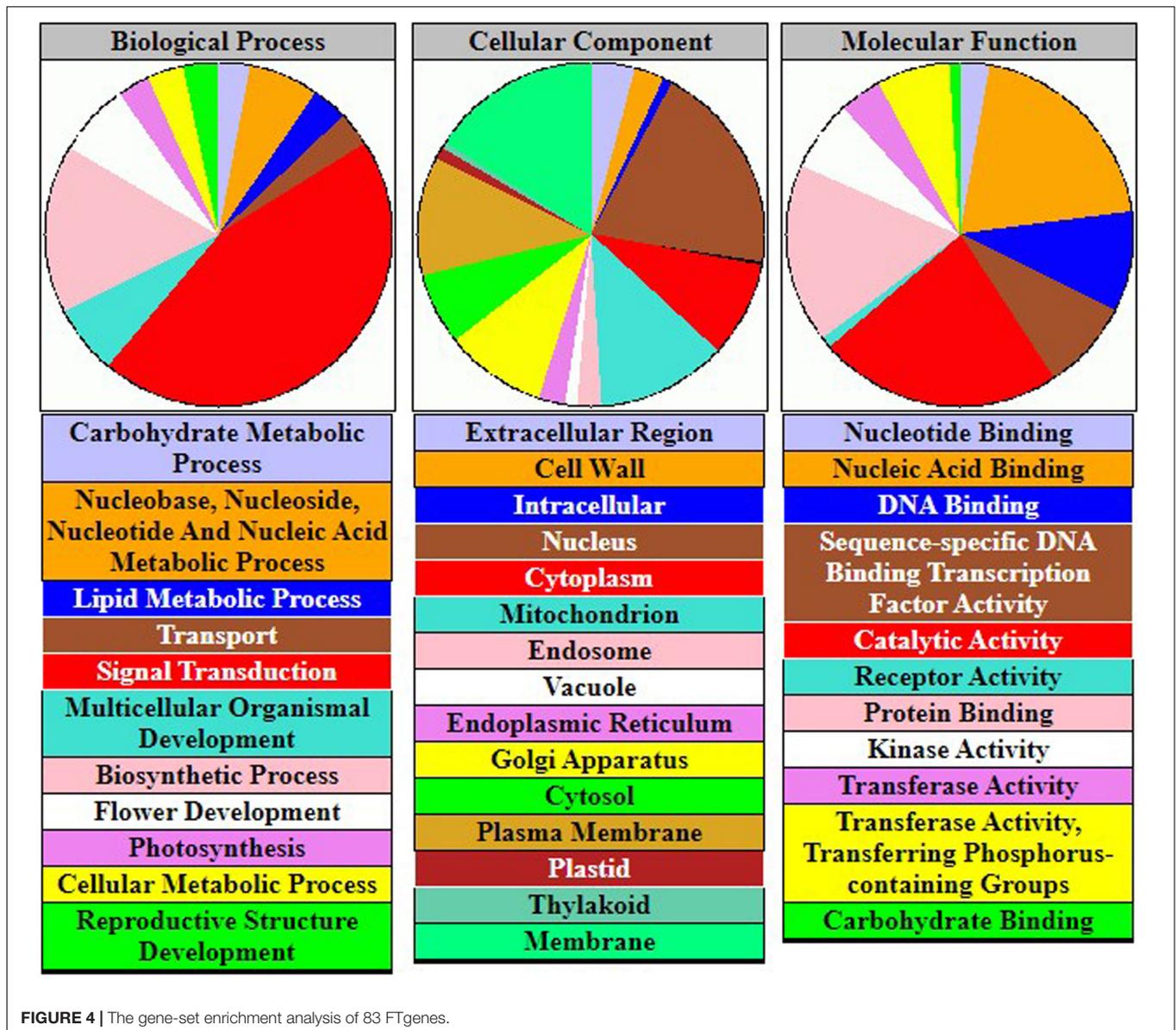
In gene-set enrichment analysis, we found that seven GO pathways are significantly involved in the relevant mechanisms of flooding tolerance (Table 5 and Figure 4). GO:0009688

was the ABA biosynthetic process. Previous studies indicated that the concentration of ABA in hypocotyls will gradually decrease if soybeans are subjected to flooding stress, thereby leading to the growth of secondary aerenchyma (Shimamura et al., 2015). GO:0001666 is the pathway that responds to

TABLE 5 | Gene-set enrichment analysis using 83 FTgenes.

# GO annotation	Category	# FTgenes/# genes in pathway	Adjusted P-value ^a
Ethylene biosynthetic process	Biological process	18/270	2.57×10^{-11}
Abscisic acid (ABA) biosynthetic process	Biological process	5/53	3.63×10^{-3}
Nuclear-transcribed mRNA poly(A) tail shortening	Biological process	3/11	6.94×10^{-3}
Glucan biosynthetic process	Biological process	3/15	1.88×10^{-2}
Ethylene mediated signaling pathway	Biological process	9/311	2.82×10^{-2}
Phosphorylation	Biological process	5/84	3.34×10^{-2}
Response to hypoxia	Biological process	7/195	4.19×10^{-2}

^aAnalysis was conducted using GO annotation database. Adjusted P-values were computed using the Bonferroni correction method.



hypoxia. Flooding causes hypoxia in plant roots and induces hypoxia-related regulatory pathways, thus, it is intuitive that GO:0001666 was selected. GO:0009693 and GO:0009873 are the ethylene biosynthetic process and ethylene-activated signaling

pathways, respectively. Yin et al. (2014) showed that the fresh weight of waterlogged soybean plants with ethylene application was significantly higher than the control one. This indicates that the presence of ethylene can help soybean plants

resist flooding stress. In addition, the pathway of ethylene biosynthesis had been determined to be significantly related to the response of plants to flooding stress in previous studies (Morgan and Drew, 1997; Kim et al., 2017). The GO results above confirm that all of the 83 FTgenes we identified are reliable.

We compared 83 FTgenes with 83 equal weight prioritized genes for validation and found that there were 74 genes (89.16%) overlapping in both gene-sets (**Supplementary Table 6**). According to Fuller and Hester (1999), we speculated that it was owing to the existence of large sample outliers (LSO) in our big data meta-analysis. LSO will reduce the bias of the unweighted method, thereby narrowing the ranking of weighted and unweighted prioritized genes. Furthermore, three independent databases were also used for verification. In the first database, we included 27 SSR markers associated with flooding tolerance in soybean from SoyBase (Cregan et al., 1999; Cornelious et al., 2005; Githiri et al., 2006; Sayama et al., 2009; Rizal and Karki, 2011). We extended and selected the region 20 kb upstream and downstream of the 27 SSR markers on both sides by gene mapping. As a result, 87 genes were discovered (**Supplementary Table 7**). Only six genes overlapped between these 87 genes and the 83 FTgenes (*Glyma.08g119000*, *Glyma.08g119100*, *Glyma.08g119200*, *Glyma.08g119400*, *Glyma.08g119500*, and *Glyma.08g139100*). Subsequently, we found that the GO pathways which involved these 87 genes were not significantly associated with flooding tolerance. It is likely that the genes located at the region 20 kb upstream and downstream of the 27 SSR markers contain genes that were not related to flooding tolerance.

In another database, we compared 83 FTgenes with 59 prioritized genes which were identified by Kim et al. (2017) (**Supplementary Table 8**). We found that 16 genes (*Glyma.01g118000*, *Glyma.02g148200*, *Glyma.03g132700*, *Glyma.07g153100*, *Glyma.08g199800*, *Glyma.09g153900*, *Glyma.11g149900*, *Glyma.11g179300*, *Glyma.12g094100*, *Glyma.12g150500*, *Glyma.12g222400*, *Glyma.13g208000*, *Glyma.13g270100*, *Glyma.14g121200*, *Glyma.16g204600*, and *Glyma.18g009700*) overlapped in both gene-sets. The other 67 FTgenes without overlapping, were from the association data platform and pathway regulation data platform. The SoyNet database constructed by Kim et al. (2017) did not include these two data platforms. In the other database, we compared 83 FTgenes with the 117 prioritized genes identified by Yu et al. (2019) (**Supplementary Table 9**). Soon we found that 23 genes (*Glyma.13g243800*, *Glyma.13g244000*, *Glyma.11g055700*, *Glyma.13g244100*, *Glyma.13g243700*, *Glyma.13g243600*, *Glyma.13g243900*, *Glyma.07g036400*, *Glyma.08g128500*, *Glyma.08g128100*, *Glyma.10g048000*, *Glyma.10g048100*, *Glyma.07g036300*, *Glyma.10g047800*, *Glyma.10g048200*, *Glyma.07g036200*, *Glyma.07g036100*, *Glyma.13g250400*, *Glyma.10g047900*, *Glyma.14g202300*, *Glyma.13g251100*, *Glyma.13g251300*, and *Glyma.13g250300*) overlapped in both gene-sets. Comparing the fold change of these 117 prioritized genes with 100,000 bootstrap random groups by using the Wilcoxon rank-sum test, we discovered that their fold change did not show a significant difference

(P -value > 0.05). After validating the three independent databases above, we confirmed that our multi-data platforms are superior to a single data platform in terms of the accuracy of identifying FTgenes.

There are six FTgenes (*Glyma.13g243600*, *Glyma.13g243700*, *Glyma.13g243800*, *Glyma.13g243900*, *Glyma.13g244000*, and *Glyma.13g244100*) located in the same region in chromosome 13. However, the six FTgenes were reported in four different data platforms, including GWAS, linkage mapping, gene expression, and PPIN. Of which only linkage mapping data platform reported a QTL. Although the six FTgenes were significantly reported in several data platforms using various experiments or methods, further validation using an independent sample is needed to access which are casual genes or whether they are gene clusters. Another way of finding a casual gene (or an index SNP) in a QTL is to apply an LD-based clumping association test (Marees et al., 2018).

The similarities and differences of molecular mechanisms and responses to submergence and waterlogging have been studied in a wide range of species (Voeselek and Bailey-Serres, 2015). Both types of flooding stress limit oxygen availability in plant cells and produce hypoxia ($< 21\% O_2$) (Sasidharan et al., 2017). There are two survival strategies, low- O_2 escape syndrome (LOES) and low- O_2 quiescence syndrome (LOQS), for flood-tolerant plants (Bailey-Serres and Voeselek, 2008, 2010; Voeselek and Bailey-Serres, 2013). The key responses for root waterlogging include the formation of aerenchyma and barriers in adventitious roots to avoid oxygen loss (Mustroph, 2018), which involves ethylene, reactive oxygen species (ROS), and hormonal signaling pathways including ABA and gibberellin (Voeselek and Bailey-Serres, 2015). The key responses for submergence include escape by elongation of aerial organs (LOES strategy for partial submergence), quiescence of metabolism and growth, protection of meristems or organs (Bailey-Serres and Voeselek, 2008, 2010; Bailey-Serres et al., 2012) (LOQS strategy for prolonged complete submergence), which involve signaling of ethylene, reduced light, low O_2 , nitric oxide, and ROS (Voeselek and Bailey-Serres, 2015). Our results in gene-set enrichment analysis (**Table 5**) are based on 83 prioritized FTgenes supported previous reported mechanisms or pathways including thylene biosynthetic process, ABA biosynthetic process, ethylene mediated signaling pathway, phosphorylation, and response to hypoxia. We conducted pathway analysis based on SUMSTAT with 10,000 permutations (Tintle et al., 2009), using 27 FTgenes that were identified to be associated or responded to both types of flooding stress (**Table 3**). Interestingly, the 27 FTgenes were significantly enriched with N-terminal protein myristoylation, response to hypoxia, defense response, secondary cell wall biogenesis, and biosynthetic process (P -values $< 1 \times 10^{-4}$; data not shown), which again echo previously reported results reviewed by Voeselek and Bailey-Serres (2015).

There were two limitations to our study. The deficiency of a verified experiment of 83 FTgenes, e.g., RNA-seq transcriptome profiling and qRT-PCR, was the first limitation. Thus, the gene expression database from Lin et al. (2019) was adopted to evaluate

the 83 FTgenes. We then obtained a significant difference between the prioritized set and the random set. In addition, 83 FTgenes were verified to be more reliable and robust by using three databases (GWAS, SoyNet, and SoyBase). The other limitation was that potential problems may exist in the results of previous studies, including noises and biases that affect our final prioritized results. To take over this problem, our big data meta-analysis and gene prioritization method can remove the genes with spurious associations. We are looking into the possibility of minimizing the effect of noises and biases from previous studies.

CONCLUSION

To the best of our knowledge, this study is the first to report prioritized FTgenes for soybean. We introduced a comprehensive framework to integrate and prioritize diverse genetic data collected from multiple dimensional data sources to search for important genes that are highly connected to flooding tolerance or responding to stress. In the present study, a total of 83 FTgenes were prioritized, based on their magnitude of association or expression change, from a 36,705 test genes pool of flooding tolerance in soybean. These FTgenes were significantly enriched with a response to hypoxia, ethylene, ABA, and glucan biosynthetic process pathways, which play an important role in the biosynthesis of plant hormones in soybean.

These results provide a basis for breeders to design efficient markers near or within the target locus of the FTgenes, and then marker-assisted selection can be applied to introduce FTgenes into the genome of commercial cultivars, such that these cultivars will be characterized by the ability to adapt to stress caused by flooding. The proposed analytic framework applied in the present study provides a shortcut to overcome a challenge in identifying the most promising genes from a large candidate-gene pool for agricultural traits of interest.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

C-FK: study conception and design. M-CL, L-HJ, and C-FK: acquisition of data. C-FK and M-CL: analysis and interpretation of data. M-CL, C-FK, Z-YL, L-HJ, and Y-SL: draft and manuscript revision. All authors read and approved the final manuscript.

REFERENCES

Ahmed, F., Rafii, M. Y., Ismail, M. R., Juraimi, A. S., Rahim, H. A., Asfaliza, R., et al. (2013). Waterlogging tolerance of crops: breeding, mechanism of tolerance, molecular approaches, and future

FUNDING

This work was supported by grants MOST 107-2313-B-005-046 and MOST 108-2313-B-005-017 from the Taiwan Ministry of Science and Technology, and grant 109AS-1.1.5-ST-aE from the Council of Agriculture, Executive Yuan, Taiwan. This work was financially supported (in part) by the Advanced Plant Biotechnology Center from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

ACKNOWLEDGMENTS

We thank Yi-An Chang and Chiu-Ting Chen for English and format editing. Afterwards, we also appreciated Chih-Min Huang and for the assistance in GWAS technology and genetics.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.612131/full#supplementary-material>

Supplementary Figure 1 | The principle of gene mapping.

Supplementary Figure 2 | Distribution of the quantity of 83 FTgenes on each soybean chromosome.

Supplementary Table 1 | The scoring scheme of test genes from different data sources.

Supplementary Table 2 | The range of impact factors with corresponding weight.

Supplementary Table 3 | Summary results of big data mining of flooding tolerance in soybean.

Supplementary Table 4 | The sources and association information for the five core genes selected from previous studies.

Supplementary Table 5 | The sources and association information for the 23 core genes selected from multiple platforms.

Supplementary Table 6 | Comparison of weighted and unweighted FTgenes by impact factor.

Supplementary Table 7 | Gene mapping results between ± 20 kb of 27 SSR molecular marker in previous studies.

Supplementary Table 8 | Fifty-nine soybean flooding tolerance prioritized genes identified in Kim et al. (2017).

Supplementary Table 9 | The 117 genes mapped from GWAS data. Detailed information can be found in Yu et al. (2019).

Supplementary Table 10 | The list of 74 overlapped FTgenes between 83 weighted genes and 83 unweighted genes.

Supplementary Material 1 | Raw data of genomic big data mining from multidimensional data platforms.

prospects. *Biomed. Res. Int.* 2013;963525. doi: 10.1155/2013/963525

Bailey-Serres, J., Lee, S. C., and Brinton, E. (2012). Waterproofing crops: effective flooding survival strategies. *Plant Physiol.* 160, 1698–1709. doi: 10.1104/pp.112.208173

- Bailey-Serres, J., and Voesenek, L. A. (2008). Flooding stress: acclimations and genetic diversity. *Annu. Rev. Plant Biol.* 59, 313–339. doi: 10.1146/annurev.arplant.59.032607.092752
- Bailey-Serres, J., and Voesenek, L. A. (2010). Life in the balance: a signaling network controlling survival of flooding. *Curr. Opin. Plant Biol.* 13, 489–494. doi: 10.1016/j.pbi.2010.08.002
- Board, J. E. (2008). Waterlogging effects on plant nutrient concentrations in soybean. *J. Plant Nutr.* 31, 828–838. doi: 10.1080/01904160802043122
- Burguete, M. C., Torregrosa, G., Pérez-Asensio, F. J., Castelló-Ruiz, M., Salom, J. B., Gil, J. V., et al. (2006). Dietary phytoestrogens improve stroke outcome after transient focal cerebral ischemia in rats. *Eur. J. Neurosci.* 23, 703–710. doi: 10.1111/j.1460-9568.2006.04599.x
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421
- Cederroth, C. R., and Nef, S. (2009). Soy, phytoestrogens and metabolism: a review. *Mol. Cell Endocrinol.* 304, 30–42. doi: 10.1016/j.mce.2009.02.027
- 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. *Front. Plant Sci.* 7:1044. doi: 10.3389/fpls.2016.01044
- Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 89, 789–804. doi: 10.1111/tj.13415
- Cornelius, B., Chen, P., Chen, Y., de Leon, N., Shannon, J. G., and Wang, D. (2005). Identification of QTLs underlying water-logging tolerance in soybean. *Mol. Breed.* 16, 103–112. doi: 10.1007/s11032-005-5911-5912
- Cregan, P. B., Jarvik, T., Bush, A. L., Shoemaker, R. C., Lark, K. G., Kahler, A. L., et al. (1999). An integrated genetic linkage map of the soybean genome. *Crop Sci.* 39, 1464–1490. doi: 10.2135/cropsci1999.3951464x
- Fukao, T., Barrera-Figueroa, B. E., Juntawong, P., and Peña-Castro, J. M. (2019). Submergence and waterlogging stress in plants: a review highlighting research opportunities and understudied aspects. *Front. Plant Sci.* 10:340. doi: 10.3389/fpls.2019.00340
- Fuller, J. B., and Hester, K. (1999). Comparing the sample-weighted and unweighted meta-analysis: an applied perspective. *J. Manage.* 25, 803–828.
- Githiri, S. M., Watanabe, S., Harada, K., and Takahashi, R. (2006). QTL analysis of flooding tolerance in soybean at an early vegetative growth stage. *Plant Breed.* 125, 613–618.
- Goh, W. W. B., Wang, W., and Wong, L. (2017). Why batch effects matter in omics data, and how to avoid them. *Trends Biotechnol.* 35, 498–507. doi: 10.1016/j.tibtech.2017.02.012
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Hou, F. F., and Thseng, F. S. (1991). Studies on the flooding tolerance of soybean seed: varietal differences. *Euphytica* 57, 169–173.
- Huang, G., Cao, X., Zhang, X., Chang, H., Yang, Y., Du, W., et al. (2009). Effects of soybean isoflavone on the notch signal pathway of the brain in rats with cerebral ischemia. *J. Nutr. Sci. Vitaminol.* 55, 326–331.
- Jackson, M. B., Ishizawa, K., and Ito, O. (2009). Evolution and mechanisms of plant tolerance to flooding stress. *Ann. Bot.* 103, 137–142. doi: 10.1093/aob/mcn242
- Kazemi Oskuei, B., Yin, X., Hashiguchi, A., Bandehagh, A., and Komatsu, S. (2017). Proteomic analysis of soybean seedling leaf under waterlogging stress in a time-dependent manner. *Biochim Biophys. Acta Proteins Proteom.* 1865, 1167–1177. doi: 10.1016/j.bbapap.2017.06.022
- Khan, M. N., Sakata, K., and Komatsu, S. (2015). Proteomic analysis of soybean hypocotyl during recovery after flooding stress. *J. Proteom.* 121, 15–27. doi: 10.1016/j.jprot.2015.03.020
- Khatoun, A., Rehman, S., Salavati, A., and Komatsu, S. (2012). A comparative proteomics analysis in roots of soybean to compatible symbiotic bacteria under flooding stress. *Amino Acids* 43, 2513–2525. doi: 10.1007/s00726-012-1333-1338
- Kim, E., Hwang, S., and Lee, I. (2017). SoyNet: a database of co-functional networks for soybean *Glycine max*. *Nucleic Acids Res.* 45, D1082–D1089. doi: 10.1093/nar/gkw704
- Komatsu, S., Han, C., Nanjo, Y., Altaf-Un-Nahar, M., Wang, K., He, D., et al. (2013). Label-free quantitative proteomic analysis of abscisic acid effect in early-stage soybean under flooding. *J. Proteome Res.* 12, 4769–4784. doi: 10.1021/pr4001898
- Komatsu, S., Yamamoto, R., Nanjo, Y., Mikami, Y., Yunokawa, H., and Sakata, K. (2009). A comprehensive analysis of the soybean genes and proteins expressed under flooding stress using transcriptome and proteome techniques. *J. Proteome Res.* 8, 4766–4778.
- Lin, Y., Li, W., Zhang, Y., Xia, C., Liu, Y., Wang, C., et al. (2019). Identification of genes/proteins related to submergence tolerance by transcriptome and proteome analyses in soybean. *Sci. Rep.* 9:14688. doi: 10.1038/s41598-019-50757-50751
- Liu, D. L., Chen, S. W., Liu, X. C., Yang, F., Liu, W. G., She, Y. H., et al. (2019). Genetic map construction and QTL analysis of leaf-related traits in soybean under monoculture and relay intercropping. *Sci. Rep.* 9:2716. doi: 10.1038/s41598-019-39110-39118
- Lovekamp-Swan, T., Glendenning, M., and Schreihofer, D. A. (2007). A high soy diet reduces programmed cell death and enhances bcl-xL expression in experimental stroke. *Neuroscience* 148, 644–652. doi: 10.1016/j.neuroscience.2007.06.046
- Lu, J., Wu, D. M., Zheng, Y. L., Hu, B., Cheng, W., and Zhang, Z. F. (2012). Purple sweet potato color attenuates domoic acid-induced cognitive deficits by promoting estrogen receptor-alpha-mediated mitochondrial biogenesis signaling in mice. *Free Radic Biol. Med.* 52, 646–659. doi: 10.1016/j.freeradbiomed.2011.11.016
- Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., et al. (2018). A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* 27:e1608. doi: 10.1002/mpr.1608
- Morgan, W. P., and Drew, C. M. (1997). Ethylene and plant responses to stress. *Physiol. Plantarum* 100, 620–630.
- Mustafa, G., and Komatsu, S. (2014). Quantitative proteomics reveals the effect of protein glycosylation in soybean root under flooding stress. *Front. Plant Sci.* 5:627. doi: 10.3389/fpls.2014.00627
- Mustafa, G., Sakata, K., Hossain, Z., and Komatsu, S. (2015). Proteomic study on the effects of silver nanoparticles on soybean under flooding stress. *J. Proteom.* 122, 100–118. doi: 10.1016/j.jprot.2015.03.030
- Mustroph, A. (2018). Improving flooding tolerance of crop plants. *Agronomy* 8:160. doi: 10.3390/agronomy8090160
- Mutava, R. N., Prince, S. J. K., Syed, N. H., Song, L., Valliyodan, B., Chen, W., et al. (2015). Understanding abiotic stress tolerance mechanisms in soybean: a comparative evaluation of soybean response to drought and flooding stress. *Plant Physiol. Biochem.* 86, 109–120. doi: 10.1016/j.plaphy.2014.11.010
- Nakayama, T. J., Rodrigues, F. A., Neumaier, N., Marcelino-Guimarães, F. C., Farias, J. R. B., de Oliveira, M. C. N., et al. (2014). Reference genes for quantitative real-time polymerase chain reaction studies in soybean plants under hypoxic conditions. *Genet. Mol. Res.* 13, 860–871. doi: 10.4238/2014.February.13.4
- Nakayama, T. J., Rodrigues, F. A., Neumaier, N., Marcolino-Gomes, J., Molinari, H. B. C., Santiago, T. R., et al. (2017). Insights into soybean transcriptome reconfiguration under hypoxic stress: functional, regulatory, structural, and compositional characterization. *PLoS One* 12:e0187920. doi: 10.1371/journal.pone.0187920
- Nanjo, Y., Jang, H. Y., Kim, H. S., Hiraga, S., Woo, S. H., and Komatsu, S. (2014). Analyses of flooding tolerance of soybean varieties at emergence and varietal differences in their proteomes. *Phytochemistry* 106, 25–36. doi: 10.1016/j.phytochem.2014.06.017
- Nanjo, Y., Maruyama, K., Yasue, H., Yamaguchi-Shinozaki, K., Shinozaki, K., and Komatsu, S. (2011). Transcriptional responses to flooding stress in roots including hypocotyl of soybean seedlings. *Plant Mol. Biol.* 77, 129–144. doi: 10.1007/s11103-011-9799-9794
- Nanjo, Y., Skultety, L., Ashraf, Y., and Komatsu, S. (2010). Comparative proteomic analysis of early-stage soybean seedlings responses to flooding by using gel and gel-free techniques. *J. Proteome Res.* 9, 3989–4002.
- Nanjo, Y., Skultety, L., Uvackova, L., Klubicova, K., Hajduch, M., and Komatsu, S. (2012). Mass spectrometry-based analysis of proteomic changes in the root tips of flooded soybean seedlings. *J. Proteome Res.* 11, 372–385. doi: 10.1021/pr200701y

- Nguyen, V. L., Takahashi, R., Githiri, S. M., Rodriguez, T. O., Tsutsumi, N., Kajihara, S., et al. (2017). Mapping quantitative trait loci for root development under hypoxia conditions in soybean [*Glycine max* (L.) Merr.]. *Theor. Appl. Genet.* 130, 743–755. doi: 10.1007/s00122-016-2847-2843
- Nguyen, V. T., Vuong, T. D., VanToai, T., Lee, J. D., Wu, X., Mian, M. A. R., et al. (2012). Mapping of quantitative trait loci associated with resistance to phytophthora sojae and flooding tolerance in soybean. *Crop Sci.* 52, 2481–2493. doi: 10.2135/cropsci2011.09.0466
- Nishizawa, K., Hiraga, S., Yasue, H., Chiba, M., Tougou, M., Nanjo, Y., et al. (2013). The synthesis of cytosolic ascorbate peroxidases in germinating seeds and seedlings of soybean and their behavior under flooding stress. *Biosci. Biotechnol. Biochem.* 77, 2205–2209. doi: 10.1271/bbb.130384
- Oh, M., and Komatsu, S. (2015). Characterization of proteins in soybean roots under flooding and drought stresses. *J. Proteom.* 114, 161–181. doi: 10.1016/j.jprot.2014.11.008
- Oosterhuis, D. M., Scott, H. D., Hampton, R. E., and Wullschleger, S. D. (1990). Physiological response of two soybean [*Glycine max* (L.) Merr.] cultivars to short-term flooding. *Environ. Exp. Botany* 30, 85–92.
- Ravelombola, W. S., Qin, J., Shi, A., Nice, L., Bao, Y., Lorenz, A., et al. (2019). Genome-wide association study and genomic selection for soybean chlorophyll content associated with soybean cyst nematode tolerance. *BMC Genom.* 20:904. doi: 10.1186/s12864-019-6275-z
- Rhee, S. Y., and Mutwil, M. (2014). Towards revealing the functions of all genes in plants. *Trends Plant Sci.* 19, 212–221. doi: 10.1016/j.tplants.2013.10.006
- Rizal, G., and Karki, S. (2011). Alcohol dehydrogenase (ADH) activity in soybean [*Glycine max* (L.) Merr.] under flooding stress. *Electronic J. Plant Breed.* 2, 50–57.
- Rosenzweig, C., Tubiello, F. N., Goldberg, R., Mills, E., and Bloomfield, J. (2002). Increased crop damage in the US from excess precipitation under climate change. *Global Environ. Change* 12, 197–202. doi: 10.1016/S0959-3780(02)00008-0
- Russell, A. D., Wong, M. L. D., and Sachs, M. M. (1990). The anaerobic response of soybean. *Plant Physiol.* 92, 401–407.
- Sasidharan, R., Bailey-Serres, J., Ashikari, M., Atwell, J. B., Colmer, D. T., Fagerstedt, K., et al. (2017). Community recommendations on terminology and procedures used in flooding and low oxygen stress research. *New Phytol.* 214, 1403–1407.
- Sayama, T., Nakazaki, T., Ishikawa, G., Yagasaki, K., Yamada, N., Hirota, N., et al. (2009). QTL analysis of seed-flooding tolerance in soybean [*Glycine max* (L.) Merr.]. *Plant Sci.* 176, 514–521. doi: 10.1016/j.plantsci.2009.01.007
- 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
- Schreihöfer, D. A., Do, K. D., and Schreihöfer, A. M. (2005). High-soy diet decreases infarct size after permanent middle cerebral artery occlusion in female rats. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 289, R103–R108. doi: 10.1152/ajpregu.00642.2004
- Shannon, J. G., Stevens, W. E., Wiebold, W. J., McGraw, R. L., Slepner, D. A., and Nguyen, H. T. (2005). “Breeding soybeans for improved tolerance to flooding,” in *Procedure of 30th Soybean Research Conference* (Chicago, IL: American Seed Trade Association).
- Shimamura, S., Yoshioka, T., Yamamoto, R., Hiraga, S., Nakamura, T., Shimada, S., et al. (2015). Role of abscisic acid in flood-induced secondary aerenchyma formation in soybean [*Glycine max* (L.) Merr.] hypocotyls. *Plant Product. Sci.* 17, 131–137. doi: 10.1626/pss.17.131
- Song, L., Valliyodan, B., Prince, S., Wan, J., and Nguyen, H. (2018). Characterization of the XTH gene family: new insight to the roles in soybean flooding tolerance. *Int. J. Mol. Sci.* 19:2705. doi: 10.3390/ijms19092705
- Syed, N. H., Prince, S. J., Mutava, R. N., Patil, G., Li, S., Chen, W., et al. (2015). Core clock, *SUB1*, and *ABAR* genes mediate flooding and drought responses via alternative splicing in soybean. *J. Exp. Bot.* 66, 7129–7149. doi: 10.1093/jxb/erv407
- Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484. doi: 10.1038/s41576-019-0127-121
- Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., et al. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genom.* 15:312. doi: 10.1186/1471-2164-15-312
- Tintle, N. L., Borchers, B., Brown, M., and Bekmetjev, A. (2009). Comparing gene set analysis methods on single-nucleotide polymorphism data from genetic analysis workshop 16. *BMC Proc.* 3(Suppl. 7):S96. doi: 10.1186/1753-6561-3-s7-s96
- Tucker, M. L., Murphy, C. A., and Yang, R. (2011). Gene expression profiling and shared promoter motif for cell wall-modifying proteins expressed in soybean cyst nematode-infected roots. *Plant Physiol.* 156, 319–329. doi: 10.1104/pp.110.170357
- Valliyodan, B., Van Toai, T. T., Alves, J. D., de Fatima, P. G. P., Lee, J. D., Fritschi, F. B., et al. (2014). Expression of root-related transcription factors associated with flooding tolerance of soybean [*Glycine max* (L.) Merr.]. *Int. J. Mol. Sci.* 15, 17622–17643. doi: 10.3390/ijms151017622
- VanToai, T. T., Hoa, T. T. C., Nguyen, T. N. H., Nguyen, T. H., Shannon, G., and Mohammed, A. R. (2010). Flooding tolerance of soybean [*Glycine max* (L.) Merr.] germplasm from southeast asia under field and screen-house environments. *Open Agr. J.* 4, 38–46.
- VanToai, T. T., Martin, K. S. S., Chase, K., Boru, G., Schnipke, V., Schmitthenner, F. A., et al. (2001). Identification of a QTL associated with tolerance of soybean to soil waterlogging. *Crop Sci.* 41, 1247–1252.
- Voensek, L., and Bailey-Serres, J. (2015). Flood adaptive traits and processes: an overview. *New Phytol.* 206, 57–73. doi: 10.1111/nph.13209
- Voensek, L. A., and Bailey-Serres, J. (2013). Flooding tolerance: O₂ sensing and survival strategies. *Curr. Opin. Plant Biol.* 16, 647–653. doi: 10.1016/j.pbi.2013.06.008
- Wang, Y. J., Zheng, Y. L., Lu, J., Chen, G. Q., Wang, X. H., Feng, J., et al. (2010). Purple sweet potato color suppresses lipopolysaccharide-induced acute inflammatory response in mouse brain. *Neurochem. Int.* 56, 424–430. doi: 10.1016/j.neuint.2009.11.016
- Wu, C., Mozzoni, L. A., Moseley, D., Hummer, W., Ye, H., Chen, P., et al. (2019). Genome-wide association mapping of flooding tolerance in soybean. *Mol. Breed.* 40:4. doi: 10.1007/s11032-019-1086-1080
- Wu, C., Zeng, A., Chen, P., Hummer, W., Mokua, J., Shannon, J. G., et al. (2017). Evaluation and development of flood-tolerant soybean cultivars. *Plant Breed.* 136, 913–923. doi: 10.1111/pbr.12542
- Xia, J., Zhang, X., Yuan, D., Chen, L., Webster, J., and Fang, A. C. (2013). Gene prioritization of resistant rice gene against *Xanthomas oryzae* pv. *oryzae* by using text mining technologies. *Biomed. Res. Int.* 2013:853043. doi: 10.1155/2013/853043
- Ye, H., Song, L., Chen, H., Valliyodan, B., Cheng, P., Ali, L., et al. (2018). A major natural genetic variation associated with root system architecture and plasticity improves waterlogging tolerance and yield in soybean. *Plant Cell Environ.* 41, 2169–2182. doi: 10.1111/pce.13190
- Yin, X., Hiraga, S., Hajika, M., Nishimura, M., and Komatsu, S. (2017). Transcriptomic analysis reveals the flooding tolerant mechanism in flooding tolerant line and abscisic acid treated soybean. *Plant Mol. Biol.* 93, 479–496. doi: 10.1007/s11103-016-0576-572
- Yin, X., and Komatsu, S. (2015). Quantitative proteomics of nuclear phosphoproteins in the root tip of soybean during the initial stages of flooding stress. *J. Proteom.* 119, 183–195. doi: 10.1016/j.jprot.2015.02.004
- Yin, X., and Komatsu, S. (2016). Nuclear proteomics reveals the role of protein synthesis and chromatin structure in root tip of soybean during the initial stage of flooding stress. *J. Proteome Res.* 15, 2283–2298. doi: 10.1021/acs.jproteome.6b00330
- Yin, X., Nishimura, M., Hajika, M., and Komatsu, S. (2016). Quantitative proteomics reveals the flooding-tolerance mechanism in mutant and abscisic acid-treated soybean. *J. Proteome Res.* 15, 2008–2025. doi: 10.1021/acs.jproteome.6b00196
- Yin, X., Sakata, K., and Komatsu, S. (2014). Phosphoproteomics reveals the effect of ethylene in soybean root under flooding stress. *J. Proteome Res.* 13, 5618–5634. doi: 10.1021/pr500621c
- Youn, J. T., Van, K., Lee, J. E., Kim, W. H., Yun, H. T., Kwon, Y. U., et al. (2008). Waterlogging effects on nitrogen accumulation and N₂ fixation of supernodulating soybean mutants. *Crop Sci. Biotechnol.* 11, 111–118.
- Yu, Z., Chang, F., Lv, W., Sharmin, R. A., Wang, Z., Kong, J., et al. (2019). Identification of QTN and candidate gene for seed-flooding tolerance in

- soybean [*Glycine max* (L.) Merr.] using genome-wide association study (GWAS). *Genes (Basel)* 10:957. doi: 10.3390/genes10120957
- Zhai, J., Tang, Y., Yuan, H., Wang, L., Shang, H., and Ma, C. (2016). A meta-analysis based method for prioritizing candidate genes involved in a pre-specific function. *Front. Plant. Sci.* 7:1914. doi: 10.3389/fpls.2016.01914
- Zhang, J., Wu, J., Liu, F., Tong, L., Chen, Z., Chen, J., et al. (2019). Neuroprotective effects of anthocyanins and its major component cyanidin-3-O-glucoside (C3G) in the central nervous system: an outlined review. *Eur. J. Pharmacol.* 858:172500. doi: 10.1016/j.ejphar.2019.172500

Conflict of Interest: 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 © 2021 Lai, Lai, Jhan, Lai and Kao. 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) and the copyright owner(s) 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.