BMC Genomics

Open Access

Plant stress RNA-seq Nexus: a stress-specific transcriptome database in plant cells

CrossMark

Jian-Rong Li[1,2], Chun-Chi Liu[1,2,3], Chuan-Hu Sun[3] and Yu-Ting Chen[1,3*]

Abstract

Background: Abiotic and biotic stresses severely affect the growth and reproduction of plants and crops. Determining the critical molecular mechanisms and cellular processes in response to stresses will provide biological insight for addressing both climate change and food crises. RNA sequencing (RNA-Seq) is a revolutionary tool that has been used extensively in plant stress research. However, no existing large-scale RNA-Seq database has been designed to provide information on the stress-specific differentially expressed transcripts that occur across diverse plant species and various stresses.

Results: We have constructed a comprehensive database, the plant stress RNA-Seq nexus (PSRN), which includes 12 plant species, 26 plant-stress RNA-Seq datasets, and 937 samples. All samples are assigned to 133 stress-specific subsets, which are constructed into 254 subset pairs, a comparison between selected two subsets, for stress-specific differentially expressed transcript identification.

Conclusions: PSRN is an open resource for intuitive data exploration, providing expression profiles of coding-transcript/lncRNA and identifying which transcripts are differentially expressed between different stress-specific subsets, in order to support researchers generating new biological insights and hypotheses in molecular breeding or evolution. PSRN is freely available at http://syslab5.nchu.edu.tw/PSRN.

Keywords: RNA-Seq, Plant stress, Database, Differential expression

Background

Abiotic and biotic stresses are the most harmful factors affecting the growth and productivity of crops and leading to food crisis worldwide. Drought, salinity, heat, cold, chilling, freezing, nutrient, high light intensity, ozone (O_3) and anaerobic stresses are the main abiotic stresses [1]. Along with abiotic stresses, plants also have to face biotic stresses caused by pathogens (including bacteria, fungi, viruses, and nematodes) and herbivorous pests [2]. Recent studies showed that abiotic stresses, such as rising temperatures, can facilitate the spread of pathogens and weaken the defense mechanisms of plants [3, 4]. To improve the resistance to stress, several cultivars and mutant lines have been developed [5, 6]. In response to stress, plant physiology and transcriptomes

undergo changes in alarm, resistance, exhaustion, and regeneration phases [7, 8]. Since different tissues and developmental stages present different resistances to stress [9, 10], transcriptome profiling of different tissues, strains, and developmental stages under various environmental stress conditions could provide insights into the molecular mechanisms as how plants respond to stress.

RNA sequencing (RNA-Seq) [11] is a revolutionary tool that can quantify gene/isoform expression levels at a higher resolution than microarray technology and provide coding-transcript profiling as well as long noncoding RNA (lncRNA) profiling. RNA-Seq has been used extensively in plant research. Recently, global transcriptome profiling analysis using RNA-Seq has been reported to identify differentially expressed lncRNAs, coding genes and alternatively spliced isoforms in response to environmental stresses, such as salt, heat, cold, drought, light, ozone, excessive boron, and pathogen infection [12–14]. Meanwhile, there now exist several publicly available databases built for plant stress, such as RiceSRTFDB [15], STIFDB2 [16], PASmiR [17],

* Correspondence: yuting@dragon.nchu.edu.tw
[1]Program in Medical Biotechnology, National Chung Hsing University, 145 Xingda Rd., South Dist, Taichung City 402, Taiwan
[3]Institute of Genomics and Bioinformatics, National Chung Hsing University, 145 Xingda Rd., South Dist, Taichung City 402, Taiwan
Full list of author information is available at the end of the article

Li *et al. BMC Genomics*   (2018) 19:966

Page 2 of 8

QlicRice [18], PhytAMP [19], and PSPDB [20]. Among them, RiceSRTFDB and STIFDB2 are curated with microarray data and focused on transcription factors. The development of PASmiR was based on abiotic stress responsive miRNA. PSPDB was constructed with biotic and abiotic stress protein annotations. Only the PRGdb 2.0 database collected RNA-Seq data, but that data was limited to resistance genes (R-genes). In addition, several data portals contain a vast amount of plant RNA-Seq data, such as the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) [21] and the Sequence Read Archive (SRA) [22]. However, these data portals mainly serve as raw biological data archives. Therefore, a large-scale stress-specific RNA-Seq database that can provide comprehensively visualized transcriptome expression profiles and statistical analysis for differential expression has not been reported for plants. In the postgenomic era, RNA-Seq provides a global transcriptome profile, which could cover lncRNAs, coding genes and their alternatively spliced isoforms in response, and helps plant biologists to expand new insights into molecular mechanisms and responses to biotic and abiotic events. Thus, we developed an extensive genome-wide plant stress RNA-Seq database.

In this study, we constructed the first large-scale plant stress RNA-Seq database, the *plant stress RNA-Seq nexus (PSRN)*, that achieves the following unprecedented features: (i) large-scale and comprehensive data archives and analyses, including coding-transcript profiling and lncRNA profiling, (ii) phenotype-oriented data organization and searching, and (iii) the visualization of expression profiles, as well as differential expression. PSRN was developed with the goal of collecting, processing, analyzing and visualizing publicly available plant RNA-Seq data. Figure 1 presents the framework of PSRN database construction. It resulted in 12 plant species and 26 plant-stress RNA-Seq datasets, including 133 stress-specific subsets and 937 samples (Additional file 1: Table S1). Each subset is a group of plant RNA-Seq samples associated with a specific stress phenotype or genotype. In addition, PSRN provides a user-friendly interface to efficiently organize and visualize the expression profiles of the differential expressed transcripts for any pair of stress-specific subsets (Additional file 2: Table S2). PSRN is freely available at http://syslab5.nchu.edu.tw/PSRN.

Construction and content
Plant stress RNA-Seq dataset collection
The annotation for the RNA-Seq datasets regarding plant stress was collected from NCBI GEO [21], and RNA-Seq reads were downloaded from the SRA [22]. Each dataset has several stress-specific subsets that contain a group of RNA-Seq samples of plants treated with a specific stress condition, e.g., salt, heat cold, drought, and light. We manually created subsets and then systematically assigned samples to subsets according to the descriptions of the datasets and samples. For expression analysis, the reference sequences and annotations of the protein-coding transcriptome were collected from Phytozome v11 [23], Ensembl Plants [24], and PopGenIE [25]. Moreover, the long noncoding RNAs (lncRNAs) of *Arabidopsis thaliana* and *Oryza sativa* were obtained from plant non-coding RNA database (PNRD) [26] for the plant lncRNAs expression profile analysis. We only retained the datasets that had a reference transcriptome for subsequent RNA-Seq analysis. In addition to transcriptome annotations collected from the abovementioned databases, the transcriptome of each species was employed for homology searches against plant sequences retrieved from the NCBI RefSeq database [27] using the NCBI BLAST+ toolkit [28], version 2.3.0, with an E-value threshold of 1E-3. Furthermore, KAAS (http://www.genome.jp/tools/kaas/) [29] was used for homology searches of KEGG and pathway annotation [30] of each plant transcriptome.

Stress-specific differentially expressed transcripts
To estimate the expression profile of each sample, Bowtie [31], version 1.1.2, was used to build the indexer of the reference sequence and align RNA-Seq reads to the reference transcriptome with the indexer. Samples with at least 1 million raw reads and over 15% of reads mapped to the reference transcriptome were retained for subsequent analysis to ensure enough depth of the sequencing coverage [32]. After alignment, transcript abundances of each sample were estimated using eXpress [33], version 1.5.1, with the expression quantification unit 'fragments per kilobase of transcript per million mapped reads (FPKM) [34]. To identify stress-specific, differentially expressed, transcripts in each dataset, we selected the subsets that included at least three samples to fulfill the significance test criteria, and then converted the expression FPKM value to a $\log_2$ scale with an added pseudo-count and performed a *t*-test between two subsets without overlapping samples. To take the type-I error rate into account, we simultaneously used the Benjamini-Hochberg procedure [35] to calculate the false discovery rate (FDR). This resulted in 259 stress-specific subset pairs with differentially expressed transcripts (*P*-value < 0.01).

Protein-coding RNA-lncRNA coexpression networks
To investigate the potential stress-specific biological functions of lncRNAs in plant cells, we constructed protein-coding RNA and lncRNA coexpression networks from *Arabidopsis thaliana* and *Oryza sativa* data. We selected coding transcripts and lncRNAs with a relatively
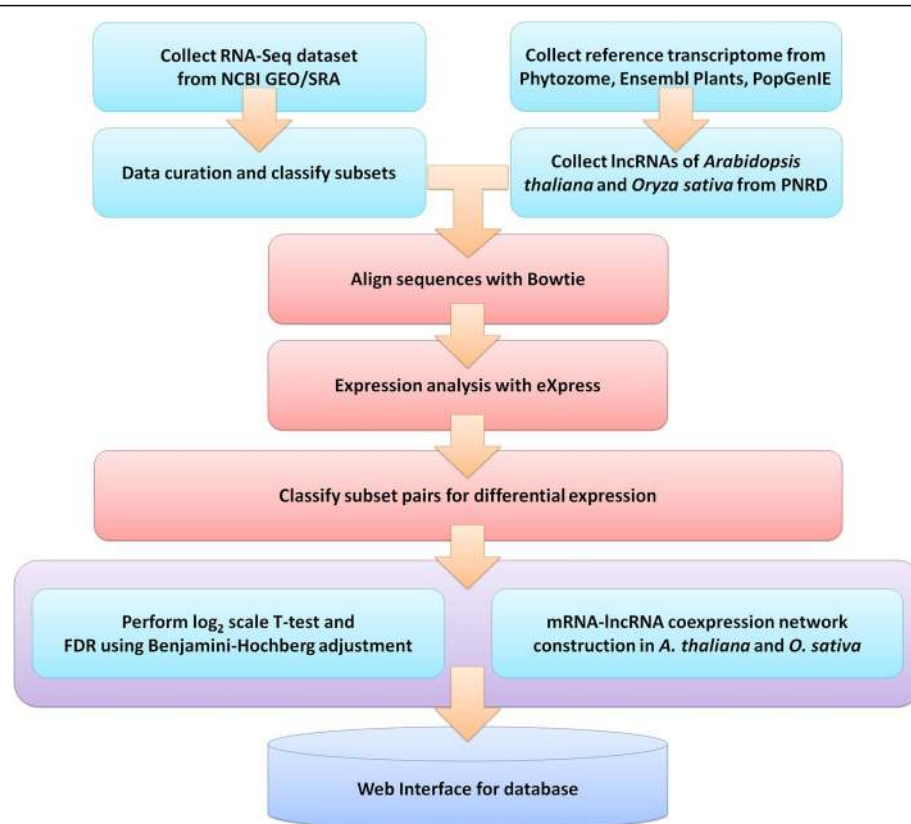
Li *et al. BMC Genomics*   (2018) 19:966

Page 3 of 8

Fig. 1 The framework of the database construction in PSRN. The plant stress RNA-Seq datasets were collected from NCBI GEO and SRA, and then all samples were classified into stress-specific subsets for each dataset. In the RNA-Seq data processing, Bowtie2 and eXpress were used to calculate transcript expressions of each RNA-Seq dataset with references collected from Phytozome, Ensembl Plants, and PopGenIE. Finally, we calculated the log2 scale T-test and FDR between two subsets belonging to the same dataset and then constructed the user interface for PSRN

high expression variance (standard deviation > 0.3 and (standard deviation/mean) > 0.5) within a subset and then calculated Pearson's correlation coefficient between every pair of selected coding RNA and lncRNA expression profiles for a given subset pair. To assess the significance of these connections, the Pearson's correlation coefficient was transferred into the *P*-value and the significant correlations (*P*-value <1E-6) between protein-coding RNAs and lncRNAs were collected for network construction. Cytoscape [36] was used to demonstrate the protein-coding RNA-lncRNA coexpression networks in the web interface of PSRN.

Utility and discussion

The PSRN database includes 12 plant species: *Arabidopsis thaliana, Chlamydomonas reinhardtii, Glycine max, Manihot esculenta, Oryza sativa* indica, *Oryza sativa* Japonica, *Panicum virgatum, Populus tremuloides, Solanum lycopersicum, Sorghum bicolor, Triticum aestivum*, and *Vitis vinifera*, which contain 26 RNA-Seq datasets and 937 samples. All samples were classified and assigned to 133 stress-specific subsets, which were constructed into 254 subset pairs to describe

stress-specific differentially expressed (DE) transcripts from a systematic RNA-Seq analysis. Considering the variants exiting in different analyses, we just compared the expression profile in the same database. We have also checked the consistency between PSRN and the individual analysis in each paper. According to the information of individual papers, we used transcript ID or gene name to find the corresponding expression profile. Most of them displayed similar patterns between PSRN and the individual analyses in different papers.

Web interface

PSRN provides a user-friendly web interface that integrates large-scale stress-specific RNA-Seq datasets of plants. Three main functional units, analysis, tutorial, and download, were included in PSRN. The tutorial unit provides brief instructions about how to use PSRN. From the download unit, users can retrieve all the differential expression analysis results of all datasets in PSRN. Here, we describe the analysis unit, which is the main function unit of the PSRN database. As shown in Fig. 2a, the analysis function provides a tree structure in the species and subset panel, which facilitates searching and
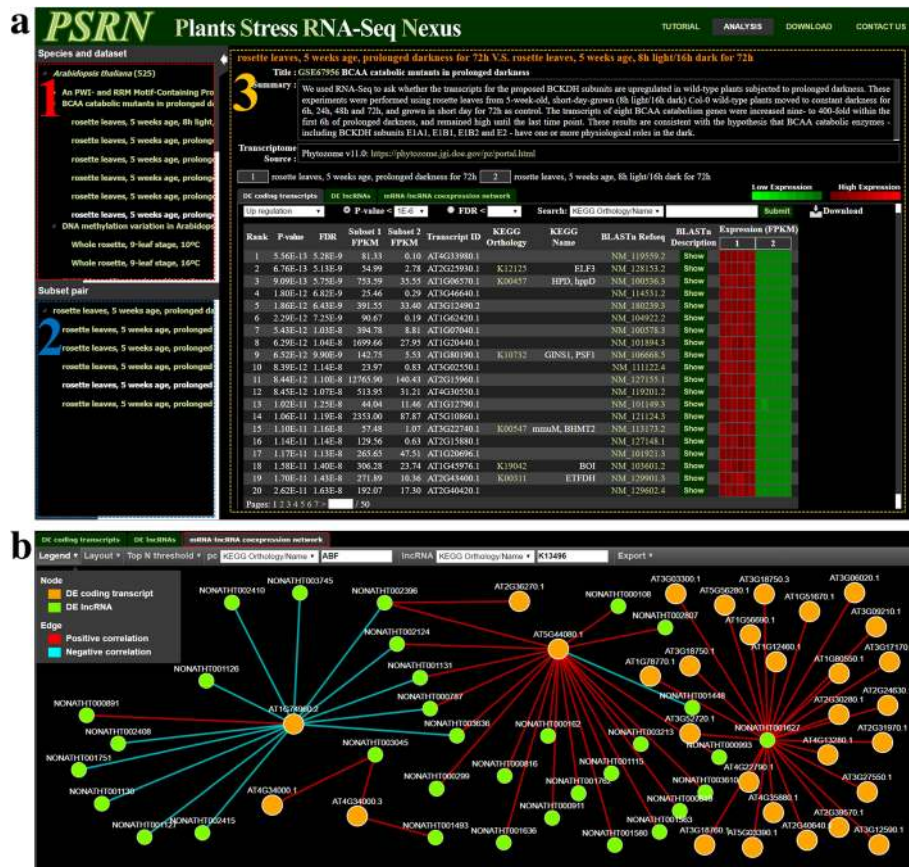
Li *et al. BMC Genomics*   (2018) 19:966

Page 4 of 8

Fig. 2 Screenshots of the web interface of PSRN. (**a**) The PSRN user interface consists of three major panels as follows: **(1) Species and subset panel**: users can browse plant species and phenotype-specific subsets. **(2) Subset-pair panel**: given the phenotype-specific subset, users can select an associated subset pair. **(3) Expression profile panel**: subset-pair information and differentially expressed protein-coding transcripts are shown in the panel and sorted by significance level. When users use Search in the expression panel, they can input a transcript ID, KEGG Orthology/name, or RefSeq ID into the autocomplete field that allows for quickly searching and selecting the partially matched terms. Furthermore, by entering more characters, it will filter down the list to better matches. At last, PSRN generates the expression profiles of all isoforms in the search results. In addition, users can download the result generated from PSRN as a file with PDF format. Moreover, users can select a *P* value or FDR threshold and regulation type to identify significantly differentially expressed transcripts. If a user clicks "DE lncRNAs" in *Arabidopsis thaliana* and *Oryza sativa*, differentially expressed lncRNA transcripts are replaced with protein-coding isoforms. (**b**) The lncRNA regulatory network function in *A. thaliana* and *O. sativa* presents the regulatory network according to the correlation of expression between lncRNAs and protein-coding transcripts

browsing stress-specific subsets. When users select a subset of interest, the associated subset pairs are subsequently listed in the subset-pair panel. When users select a subset pair, the web server shows the detailed information of the DE transcripts of the subset pair, as well as the detailed description of the dataset and subsets, into the right main panel. In the right main panel, there are three subpanel tabs as follows:

(i) *DE coding transcripts:* Given a subset pair, PSRN constructs a heat map to visualize the expression profiles of all DE protein-coding transcripts, sorted by the significance level (*P*-value) between two subsets. The interface displays the rank, the *P*-value, the FDR, the average expression values in the

subset, the transcript ID and the homology annotation for each transcript, such as KEGG and RefSeq annotations. Whenever user clicks transcript ID, the PSRN will show details of its expression in all subsets across all datasets belonging to the same species. In this panel, users can filter DE transcripts based on their *P*-value or FDR threshold and choose to present the sorted transcripts according to whether they are up- or downregulated. Additionally, the search function on the expression profiles allows users to investigate all transcripts associated with the given transcript ID, KEGG Orthology number/Name, or RefSeq annotation, and an autocomplete function provides suggestions for search field as the user types, quickly searching and displaying partially matched

Rank	P-value	FDR	Subset 1 FPKM	Subset 2 FPKM	Transcript ID	KEGG Orthology	KEGG Name	BLASTn Refseq
1	4.41E-78	1.68E-73	62.51	9.26	AT3G50970.1			NM_114957.2
2	1.04E-70	1.97E-66	33.14	3.87	AT4G14690.1			NM_117551.2
3	1.27E-65	1.60E-61	12.14	1.49	AT1G09350.1	K18819	GOLS	NM_100805.1
4	1.79E-60	1.69E-56	218.83	57.85	AT5G52310.1			NM_124610.2
5	3.09E-57	2.34E-53	348.43	126.35	AT1G20440.1			NM_101894.3
6	5.03E-55	3.18E-51	14.61	2.79	AT5G17460.1			NM_121752.3
7	1.58E-54	8.54E-51	35.03	13.46	AT5G20830.1	K00695	E2.4.1.13	NM_122090.3
8	3.03E-54	1.44E-50	1186.55	473.90	AT2G42530.1			NM_129814.2
9	8.04E-54	3.39E-50	33.81	13.75	AT5G35970.1			NM_122988.4
10	5.35E-48	2.03E-44	15.82	5.88	AT1G07180.1	K17871	ndh1	NM_100592.4

Fig. 3 The expression of cold-related transcripts in *Arabidopsis thaliana* treated with different temperature conditions. The top 5 significantly upregulated transcripts in the 10 °C subset are AT3G50970.1, AT4G14690.1, AT1G09350.1, AT5G52310.1, and AT1G20440.1 (*P*-value = 4.41E-78 ~ 3.09E-57; FDR = 1.68E-73 ~ 2.34E-53). According to the TAIR database, AT3G50970.1, AT5G52310.1, and AT1G20440.1 are Low Temperature-Induced 30 (LTI30, XERO2), Low-Temperature-Induced 78 (LTI78, COR78), and Cold-Regulated 47 (COR47, ATCOR47), respectively. In addition, AT4G14690.1 is Early Light-Inducible Protein 2 (ELIP2), and AT1G09350.1 is Galactinol Synthase 3 (GOLS3, ATGOLS3). Limited by the layout width, only parts of the expression profiles are shown

terms. The search function also allows users to investigate the expression profiles of multiple genes at one time. To do so, users simply type their transcript IDs, KEGG Orthology numbers/Names, or RefSeq annotations, and separate them by comma.

(ii) *DE lncRNAs*: The DE lncRNAs tab panel is only provided for *Arabidopsis thaliana* and *Oryza sativa* (both in indica group and japonica group). All functions of this panel are similar to the DE coding transcript panel but visualize the expression profiles of DE lncRNAs sorted by the *P*-value.

(iii) *Protein-coding RNA-lncRNA coexpression network*: As shown in Fig. 2b, PSRN visualizes the coding RNA-lncRNA coexpression network to depict the significant correlations between coding transcripts and lncRNAs. Given the subset pair, users can input a transcript ID/KEGG Orthology/Name/RefSeq annotation of a coding transcript and lncRNA into the autocomplete field. Furthermore, users can select the top 10/15/20/25 most significant connections to show for the given coding transcripts and/or the given lncRNA. Similar to the DE lncRNA panel, this panel is only constructed for *Arabidopsis thaliana* and *Oryza sativa* groups.

Case study

To demonstrate the biological functionality of PSRN, we used cold-related transcripts and HAB1 isoforms of *Arabidopsis thaliana* as examples.

(i) In the analysis results for the GSE54680 dataset, the top 5 significant upregulated transcripts in 10 °C subset are AT3G50970.1 (Low Temperature-Induced 30, LTI30, XERO2), AT4G14690.1 (Early Light-Inducible Protein 2, ELIP2), AT1G09350.1 (Galactinol Synthase 3, GOLS3, ATGOLS3), AT5G52310.1 (Low- Temperature-Induced 78, LTI78, COR78), and AT1G20440.1 (Cold-Regulated 47, COR47, ATCOR47) (illustrated in Fig. 3). LTI30 is a late embryogenesis abundant protein (LEA)/ dehydrin that can bind to and protect membranes to enhance freezing stress resistance in *Arabidopsis* [37, 38]. ELIP 2, a thylakoid protein, was induced not only by the cold but also by UVB and high light for protective photoinhibition under stresses [39]. Raffinose family oligosaccharides (RFO) were identified as being involved in tolerance to drought, high salinity and cold stresses. GOLS catalyzes the first step in the biosynthesis of RFO. In *Arabidopsis*,

Rank	P-value	FDR	Subset 1 FPKM	Subset 2 FPKM	Transcript ID	KEGG Orthology	KEGG Name	BLASTn Refseq	BLASTn Description	Expression (FPKM) 1	2
1	2.01E-3	2.64E-1	14.32	33.37	AT1G72770.2	K14497	PP2C	NM_001084346.2	Show		
2	9.33E-3	2.94E-1	48.06	19.83	AT1G72770.1	K14497	PP2C	NM_105936.3	Show		
3	2.14E-1	6.41E-1	14.60	10.00	AT1G72770.3	K14497	PP2C	NM_001198456.1	Show		

Fig. 4 The expression of HAB1 isoforms of the seedlings treated with ABA in *Arabidopsis thaliana*. The group A PP2C (protein phosphatases 2C) HAB1 gene has three transcript isoforms: HAB1.1 (AT1G72770.1), HAB1.2 (AT1G72770.2), and HAB1.3 (AT1G72770.3). Based on the *P*-value calculated by PSRN, the HAB1 isoform AT1G72770.1 is significantly downregulated (*P*-value = 9.33E-03) in the rbm25–1 mutation seedlings when compared with the wild type, while AT1G72770.2 is significantly upregulated (*P*-value = 2.01E-3) in rbm25–1 mutant seedlings

AtGOLS1 and AtGOLS12 were induced by drought and high-salinity stresses; in contrast, AtGOLS3 was induced by cold stress only [40]. COR78 [41] and COR47 [42] are well-known cold-regulated (COR) genes, and high expression levels of COR47 and COR78 in response to cold acclimation has been reported [5]. The abovementioned results are consistent with previous studies.

(ii) In the analysis results for the GSE66737 dataset, we used the HAB1 gene, a group A protein phosphatase 2C (PP2C), to demonstrate the biological importance of the level of expression of isoforms provided in PSRN (illustrated in Fig. 4). HAB1 is a coreceptor of abscisic acid (ABA) and an important negative regulator of ABA signaling. In *Arabidopsis*, the HAB1 gene has three alternatively spliced isoforms: HAB1.1 (AT1G72770.1), HAB1.2 (AT1G72770.2), and HAB1.3 (AT1G72770.3). HAB1.1 interacts with the Open Stomata 1 (OST1), inhibiting its kinase activity, which switches the ABA signaling off. In contrast, HAB1.2 encodes a nonfunctional truncated protein, thereby keeping the ABA signaling on. Thus, accurate regulation of the HAB1.1 to HAB1.2 ratio is important for the fine-tuning of ABA signaling and plant adaptation to stress. An RNA-binding motif (RBM)-containing protein RBM 25 is known as a critical key regulator of HAB1 isoforms through its binding to the last intron of HAB1 pre-mRNA in *Arabidopsis* [43]. A loss-of-function mutation in RBM25, rbm25−1, resulted in an increase in the HAB1.2:HAB1.1 ratio and ABA-hypersensitive phenotypes [44]. As shown in Fig. 4, HAB1.1 (At1g72770.1) is significantly downregulated while HAB1.2 (At1g72770.2) is significantly upregulated in rbm25−1 mutation seedlings, which is consistent with previous reports.

Future developments
Despite numerous RNA-Seq datasets being collected at the beginning of the study, only a minority of them were retained for PSRN construction. Most of the discarded datasets were excluded because of the insufficient number of samples suitable for the criteria of subset creation, and the rest were excluded because of a lack of a reference sequence. To solve the latter problem, we will continue to collect references, annotations, and RNA-Seq datasets to expand the PSRN database and keep it up to date. The next updated version will include: (1) the new plant stress RNA-Seq datasets and the reference transcriptomes collection, data curation, and KEGG and Refseq annotations, (2) performing the expression profile analysis pipeline of RNA-Seq datasets. The DESeq will be performed in this update, and the results of both t-test and DESeq will be reported in the database.

Conclusions
In this study, we collected, processed, analyzed and visualized all publicly available plant stress RNA-Seq data from NCBI GEO and SRA to construct the PSRN database. Although a few databases have previously been developed for plant stress, PSRN is the first comprehensive database using RNA-Seq data to obtain complete transcriptome profiles and differentially expressed protein-coding transcripts and lncRNAs for various plant species and stress types. Moreover, the systematic and user-friendly web interface can assist researchers in accessing the information efficiently. We hope that PSRN may provide a new resource that can facilitate biologists gaining better biological insights into plant responses to stress with a high resolution of transcript isoforms levels.

Additional files

Additional file 1: Table S1. The list of 133 subsets in 26 datasets of PSRN. (XLSX 15 kb)

Additional file 2: Table S2. The stress treatment information and web link of subset pairs in PSRN. (XLSX 37 kb)

Availability of data and materials
Database is available freely at http://syslab5.nchu.edu.tw/PSRN.

Authors' contributions
YTC, CCL, and JRL conceived the study and drafted the manuscript. JRL also collected RNA-Seq datasets and reference transcriptomes and created the work-flow. CHS performed bioinformatics pipelines and contributed to the database and website construction. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Program in Medical Biotechnology, National Chung Hsing University, 145 Xingda Rd., South Dist, Taichung City 402, Taiwan. [2]Advanced Plant Biotechnology Center, National Chung Hsing University, 145 Xingda Rd., South Dist, Taichung City 402, Taiwan. [3]Institute of Genomics and Bioinformatics, National Chung Hsing University, 145 Xingda Rd., South Dist, Taichung City 402, Taiwan.

References

1. Suzuki N, Rivero RM, Shulaev V, Blumwald E, Mittler R. Abiotic and biotic stress combinations. New Phytol. 2014;203(1):32–43.
2. Atkinson NJ, Urwin PE. The interaction of plant biotic and abiotic stresses: from genes to the field. J Exp Bot. 2012;63(10):3523–43.
3. Bale JS, Masters GJ, Hodkinson ID, Awmack C, Bezemer TM, Brown VK, Butterfield J, Buse A, Coulson JC, Farrar J, et al. Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. Glob Chang Biol. 2002;8(1):1–16.
4. Luck J, Spackman M, Freeman A, Trebicki P, Griffiths W, Finlay K, Chakraborty S. Climate change and diseases of food crops. Plant Pathol. 2011;60(1):113–21.
5. Rajashekar CB, Zhou HE, Zhang Y, Li W, Wang X. Suppression of phospholipase Dalpha1 induces freezing tolerance in Arabidopsis: response of cold-responsive genes and osmolyte accumulation. J Plant Physiol. 2006;163(9):916–26.
6. Kuzuya M, Hosoya K, Yashiro K, Tomita K, Ezura H. Powdery mildew (Sphaerotheca fuliginea) resistance in melon is selectable at the haploid level. J Exp Bot. 2003;54(384):1069–74.
7. Lichtenthaler HK. The stress concept in plants: an introduction. Ann N Y Acad Sci. 1998;851(1):187–98.
8. Robinson SJ, Parkin IA. Differential SAGE analysis in Arabidopsis uncovers increased transcriptome complexity in response to low temperature. BMC Genomics. 2008;9(1):434.
9. Yu L, Ma J, Niu Z, Bai X, Lei W, Shao X, Chen N, Zhou F, Wan D. Tissue-specific transcriptome analysis reveals multiple responses to salt stress in Populus euphratica seedlings. Genes. 2017;8(12):372.
10. Zheng J, Cheng X, Hoffmann AA, Zhang B, Ma C-S. Are adult life history traits in oriental fruit moth affected by a mild pupal heat stress? J Insect Physiol. 2017;102:36–41.
11. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
12. Di C, Yuan J, Wu Y, Li J, Lin H, Hu L, Zhang T, Qi Y, Gerstein MB, Guo Y, et al. Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. Plant J. 2014;80(5):848–61.
13. Moustafa K, Cross JM. Genetic approaches to study plant responses to environmental stresses: an overview. Biology (Basel). 2016;5(2).
14. Staiger D, Brown JW. Alternative splicing at the intersection of biological timing, development, and stress responses. Plant Cell. 2013;25(10):3640–56.
15. Priya P, Jain M. RiceSRTFDB: a database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. Database (Oxford). 2013;2013:bat027.
16. Naika M, Shameer K, Mathew OK, Gowda R, Sowdhamini R. STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in Arabidopsis and rice. Plant Cell Physiol. 2013;54(2):e8.
17. Zhang S, Yue Y, Sheng L, Wu Y, Fan G, Li A, Hu X, Shangguan M, Wei C. PASmiR: a literature-curated database for miRNA molecular regulation in plant response to abiotic stress. BMC Plant Biol. 2013;13:33.
18. Smita S, Lenka SK, Katiyar A, Jaiswal P, Preece J, Bansal KC. QlicRice: a web interface for abiotic stress responsive QTL and loci interaction channels in rice. Database (Oxford). 2011;2011:bar037.
19. Hammami R, Ben Hamida J, Vergoten G, Fliss I. PhytAMP: a database dedicated to antimicrobial plant peptides. Nucleic Acids Res. 2009; 37(Database):D963–8.
20. Kumar SA, Kumari PH, Sundararajan VS, Suravajhala P, Kanagasabai R, Kishor PBK. PSPDB: plant stress protein database. Plant Mol Biol Rep. 2014;32(4): 940–2.
21. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41(Database issue):D991–5.
22. Leinonen R, Sugawara H, Shumway M. International nucleotide sequence database C: the sequence read archive. Nucleic Acids Res. 2011;39(Database issue):D19–21.
23. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012;40(Database issue):D1178–86.
24. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C, et al. Ensembl genomes 2016: more genomes, more complexity. Nucleic Acids Res. 2016;44(D1):D574–80.
25. Sjodin A, Street NR, Sandberg G, Gustafsson P, Jansson S. The Populus genome integrative explorer (PopGenIE): a new resource for exploring the Populus genome. New Phytol. 2009;182(4):1013–25.
26. Yi X, Zhang Z, Ling Y, Xu W, Su Z. PNRD: a plant non-coding RNA database. Nucleic Acids Res. 2015;43(Database issue):D982–9.
27. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1):D733–45.
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
29. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35(Web Server):W182–5.
30. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
31. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
32. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, Chen P, Gertner RS, Gaublomme JT, Yosef N, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature. 2014;510(7505):363–9.
33. Roberts A, Pachter L. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods. 2013;10(1):71–3.
34. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5.
35. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J Roy Stat Soc B Met. 1995;57(1):289–300.
36. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
37. Welin BV, Olson A, Nylander M, Palva ET. Characterization and differential expression of Dhn/lea/Rab-like genes during cold-acclimation and drought stress in Arabidopsis-Thaliana. Plant Mol Biol. 1994;26(1):131–44.
38. Puhakainen T, Hess MW, Makela P, Svensson J, Heino P, Palva ET. Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. Plant Mol Biol. 2004;54(5):743–53.
39. Hayami N, Sakai Y, Kimura M, Saito T, Tokizawa M, Iuchi S, Kurihara Y, Matsui M, Nomoto M, Tada Y, et al. The responses of Arabidopsis early light-induced Protein2 to ultraviolet B, high light, and cold stress are regulated by a transcriptional regulatory unit composed of two elements. Plant Physiol. 2015;169(1):840–55.
40. Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K. Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. Plant J. 2002;29(4):417–26.
41. Nordin K, Vahala T, Palva ET. Differential expression of two related, low-temperature-induced genes in Arabidopsis thaliana (L.) Heynh. Plant Mol Biol. 1993;21(4):641–53.

Li *et al. BMC Genomics*   (2018) 19:966

Page 8 of 8

42. Guo W, Ward RW, Thomashow MF. Characterization of a cold-regulated wheat gene related to Arabidopsis cor47. Plant Physiol. 1992;100(2):915–22.

43. Wang Z, Ji H, Yuan B, Wang S, Su C, Yao B, Zhao H, Li X. ABA signalling is fine-tuned by antagonistic HAB1 variants. Nat Commun. 2015;6:8138.

44. Zhan X, Qian B, Cao F, Wu W, Yang L, Guan Q, Gu X, Wang P, Okusolubo TA, Dunn SL, et al. An Arabidopsis PWI and RRM motif-containing protein is critical for pre-mRNA splicing and ABA responses. Nat Commun. 2015;6:8139.