D1154–D1160 *Nucleic Acids Research, 2016, Vol. 44, Database issue* *Published online 17 October 2015*
doi: 10.1093/nar/gkv1035

PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants

Chi-Nga Chow[1], Han-Qin Zheng[1], Nai-Yun Wu[1], Chia-Hung Chien[1], Hsien-Da Huang[2], Tzong-Yi Lee[3], Yi-Fan Chiang-Hsieh[1], Ping-Fu Hou[1,4], Tien-Yi Yang[1] and Wen-Chi Chang[1,*]

[1]College of Biosciences and Biotechnology, Institute of Tropical Plant Sciences, National Cheng Kung University, Tainan 701, Taiwan, [2]Department of Biological Science and Technology, Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsin-Chu 300, Taiwan, [3]Department of Computer Science and Engineering, Yuan Ze University, Chung-Li 320, Taiwan and [4]College of Biosciences and Biotechnology, Department of Life Sciences, National Cheng Kung University, Tainan 701, Taiwan

Received August 14, 2015; Revised September 21, 2015; Accepted September 29, 2015

ABSTRACT

Transcription factors (TFs) are sequence-specific DNA-binding proteins acting as critical regulators of gene expression. The Plant Promoter Analysis Navigator (PlantPAN; http://PlantPAN2.itps.ncku.edu.tw) provides an informative resource for detecting transcription factor binding sites (TFBSs), corresponding TFs, and other important regulatory elements (CpG islands and tandem repeats) in a promoter or a set of plant promoters. Additionally, TFBSs, CpG islands, and tandem repeats in the conserve regions between similar gene promoters are also identified. The current PlantPAN release (version 2.0) contains 16 960 TFs and 1143 TF binding site matrices among 76 plant species. In addition to updating of the annotation information, adding experimentally verified TF matrices, and making improvements in the visualization of transcriptional regulatory networks, several new features and functions are incorporated. These features include: (i) comprehensive curation of TF information (response conditions, target genes, and sequence logos of binding motifs, etc.), (ii) co-expression profiles of TFs and their target genes under various conditions, (iii) protein-protein interactions among TFs and their co-factors, (iv) TF-target networks, and (v) downstream promoter elements. Furthermore, a dynamic transcriptional regulatory network under various conditions is provided in PlantPAN 2.0. The PlantPAN 2.0 is a systematic platform for plant promoter analysis and reconstructing transcriptional regulatory networks.

INTRODUCTION

Transcription factors (TFs) are sequence-specific DNA-binding proteins acting as critical regulators of gene expression. TFs regulate target genes through the recognition of specific *cis*-regulatory sequences in promoter regions. In plants, TFs play important roles in cellular physiology, developmental processes and responses to environmental stimuli. Owing to the importance of the transcriptional regulation governed by TFs, bioinformatic prediction of gene regulation on the basis of the presence of TF binding sites in the promoters is of priority concern in systems biology. For this purpose, several public web-based resources were promptly established. For instance, TRANSFAC is the commercial database manually collecting experimentally verified TFs, transcription factor binding sites (TFBSs) and matrix-based target prediction profiles (1,2). Two other well-known TFBS repositories, PLACE and JASPAR, were created to facilitate the identification of TFBSs in input sequences (3,4). However, only information on TFBSs is provided rather than their corresponding TFs. As to the systems developed for plants, AGRIS, AthaMap and AtPAN are the most useful resources for gene regulatory studies (5–7). Among them, AGRIS is dedicated to amassing information on transcriptional regulation (e.g. TFs, *cis*-regulatory elements and experimentally validated gene regulatory networks) in *Arabidopsis*, containing three sub-databases, At*cis*DB, AtTFDB and AtRegNet (8). Moreover, AthaMap covers not only a genome-wide map of predicted TFBSs used to analyze gene regulation but also data on post-transcriptional regulation mediated by small RNAs (5). Furthermore, AtPAN is an system used to analyze high-confidence TFs in promoters and to reconstruct co-expressed transcriptional regulatory networks in *Arabidopsis* (7). Since these resources only supports *Ara-*

*To whom correspondence should be addressed. Tel: +886 6 2757575 (Ext. 57322); Fax: +886 6 2083663; Email: sarah321@mail.ncku.edu.tw

bidopsis, Plant Promoter Analysis Navigator (PlantPAN) was developed to elucidate the gene transcription regulatory networks (TRNs) for 19 plant species (9). It allows users to detect more significant regulatory elements such as CpG islands and tandem repeats in promoters or to conserve regions of similar gene promoters, and it even allows identification of co-occurrence TFBSs in a group of gene promoters. Although the systems mentioned above are effective in identifying TFBSs in promoter regions, users may confront difficulties related to picking candidates out of redundant TFs and their binding sites for further experiments.

The feasibility of promoter analysis relies on how abundantly TFBSs and TFs are integrated into reference databases. However, the difficulty related to characterizing TFBSs by either computational or traditional experimental methods restricts the number of available TFBSs. The development of *in vitro* protein-binding microarray (PBM) technology has allowed for DNA-binding specificities (10). In recent years, PBM has been widely used in eukaryotic species to evaluate the binding affinity of TFs. UniPROBE, for example, is a database hosting published data generated using PBM technology and covering 22 highly diverse species (11). Another *in vivo* approach combining chromatin immunoprecipitation with either microarray (ChIP-chip) or NGS technology (ChIP-seq) achieves the mapping of TFBSs in gene promoters (12). Even though plenty of PBM and ChIP data are produced in plants, no promoter analysis databases have incorporated these high-throughput data.

PlantPAN 2.0 serves as a systematic platform for plant promoter analysis and reconstructs transcriptional regulatory networks. In addition to the integration of comprehensive gene regulation data, various new features and advances have been updated in PlantPAN 2.0. Users can access the promoter of interest via four main functions (Gene Search, Gene Group Analysis, Promoter Analysis and Cross Species), and they can search for detailed information on a given TF/TFBS by the new function, TF/TFBS Search. The system flow chart of PlantPAN 2.0 is illustrated in Figure 1.

IMPROVEMENTS

Table 1 lists the advantages and new features provided in PlantPAN 2.0. Besides the update to the annotation information, the addition of experimentally verified TF matrices, and improvements in the visualization of transcriptional regulatory networks, several new features and functions have been incorporated. These features include: (i) comprehensive curation of TF information (response conditions, target genes, and sequence logos of binding motifs, etc.), (ii) co-expression profiles of TFs and their target genes under various conditions (i.e. environmental stresses, hormone treatments, and developmental stages), (iii) protein-protein interactions among TFs and their co-factors, (iv) TF-target networks, and (v) downstream promoter elements.

Experimentally verified TF binding sites

An application of high-throughput *in vitro* techniques is capable of providing comprehensive high-resolution profiling of DNA-binding specificities of given TFs. In recent studies, Jose *et al*. pioneered the use of *in vitro* protein-binding microarray technology to identify the binding sequences of large *Arabidopsis* TFs (13). The 63 Arabidopsis TFs and corresponding 108 matrices were determined. In addition, Matthew *et al*. not only utilized the PBM to broadly survey the DNA-binding affinities of eukaryotic TFs but also integrated their findings with experimentally verified TFBSs from other public resources to establish the CIS-BP database (14). Totally, 12 829 TFs and 615 position weight matrices from these two publications were introduced into PlantPAN 2.0. Furthermore, the IUPAC consensus motifs of plant TFs were extracted manually from Uniprot databases (15). After removing the redundant TF-matrices mentioned above and the up-to-date 548 matrices from JASPAR, TRANSFAC (public release 7.0), and PLACE, 16 960 TFs and 1143 matrices of TF binding sites among 76 plant species were integrated into PlantPAN 2.0 (1,3–4). Table 2 lists the number of curated TF-matrices in each resource.

Identification of TFBSs in promoter sequences

In PlantPAN 2.0, the Match^{TM} program (16) is used to scan putative TFBS with its library of positional weight matrices and cut-off profiles. In order to set up the cut-off profiles of TF-matrices and matrices without TFs curated in PlantPAN 2.0, TRANSFAC®, like a cut-off profile, has to be trained. According to the definition of TRANSFAC® cut-off, Minimize False Negative Matches (minFN), Minimize False Positive Matches (minFP), and Minimize the Sum of Both Error Rates (minSUM) features are essential for Match^{TM}. For minFN computation, random sequences (100 000 sequences, 50 bp/per sequence) were generated and scanned with TFBS weight matrices using Match^{TM} without any cut-off threshold. In each matrix, the core and matrix similar score were estimated by selecting a score that provided recognition of at least 90% of the oligonucleotides. This value is defined as minFN cut-off. To estimate minFP, 30 000 reliable promoters (−5000 to +1000 relative to TSSs) of *Arabidopsis thaliana*, *Oryza sativa* and *Glycine max* were randomly selected from TRANSPro^{TM} for TFBS scanning using the acquired minFN cut-off (17). The score that recognizes the best 1% of the hits is defined as minFP. Then, the minimum sum of both error rates, false positives and false negatives, is defined as minSUM. To provide strict scanning results, the trained minFP is set as the cut-off profile in PlantPAN 2.0.

TF-target co-expressed networks

In general, TFs are induced by specific conditions and play essential roles in gene regulation. This suggests that the expression level of TF targets may be changed owing to the variations in conditions. Gene co-expression analysis is a powerful approach to systematically infer the transcriptional regulation and functionality of genes. Although, AtPAN provides the co-expression analysis among TFs and the input genes, it is performed only for *Arabidopsis* and across all conditions rather than for sample-specific data sets designated by users (7). The old version of PlantPAN did not incorporate any microarray data. Therefore,

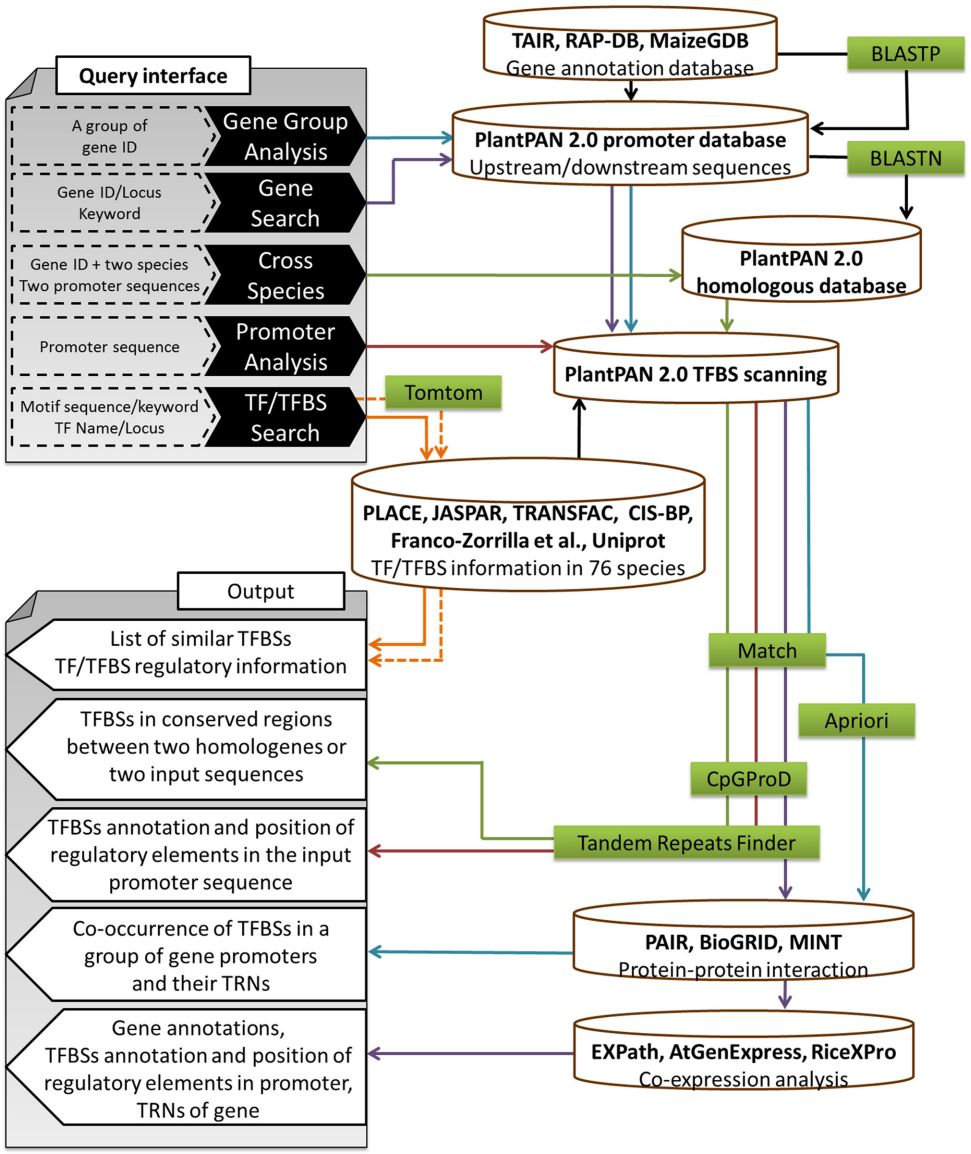

Figure 1. System flow of PlantPAN 2.0. Five major functions (Gene Group Analysis, Gene Search, Cross Species, Promoter Analysis and TF/TFBS Search) are provided in PlantPAN 2.0. The analysis processes (arrowhead) of five functions are marked by different colors. In TF/TFBS Search, the orange dotted arrowheads represent similar TFBSs identified by the input TFBS motifs, whereas the orange arrowheads refer to the regulatory information obtained by searching TF Name/Locus.

Table 1. A comparison of PlantPAN 2.0 with the previous version and similar resources

Features	PlantPAN 2.0	PlantPAN (old)	AtPAN	NewPLACE	AGRIS	JASPAR	AthaMAP
No. of plant transcription factors from different species	**16 960**	591	441	56	1770	65	NA
No. of plant species	**76**	19	1	49	1	8	1
No. of TF matrices	**1143**	629	NA	457	NA	65	NA
Experimentally verified TF matrices from PBM	**Yes**	No	No	No	No	Yes	No
Providing high-confidence TFs/TFBSs	**Yes**	No	Yes	No	No	No	No
Co-expression profiles of TFs and their target genes under various conditions	**Yes**	No	No	No	No	No	No
Protein-protein interactions between TFs and their co-factors	**Yes**	No	Yes	No	No	No	No
TF-target networks	**Yes**	No	Yes	No	Yes	No	No
The downstream promoter elements	**Yes**	No	No	No	No	No	No
Comprehensive curation of TF information (response conditions, target genes, activator or repressor, and sequence logos of binding motifs)	**Yes**	No	No	No	Yes (target genes)	No	No
Analysis of customized TFBS	**Yes**	No	No	No	No	No	No

Table 2. The statistics for TFs and Matrices incorporated in six resources

Resources	Number of TFs	Number of Matrices	Number of Species
Uniprot	4666	44	3
CIS-BP (14)	12 802	507	44
Franco-Zorrilla *et al.* (13)	65	108	1
TRANSFAC (public release 7.0)	20	27	6
PLACE	56	457	49
JASPAR	64	64	8
Total	17 673	1207	111
Non-redundant in PlantPAN 2.0	**16 960**	**1143**	**76**

to provide the co-expression profiles of TFs and their target genes under various conditions, microarray expression data derived from samples of environmental stresses, hormone treatments and developmental stages were collected in PlantPAN 2.0. Among them, the data sets for environmental stresses (abiotic stresses and biotic stresses) and hormone treatments were retrieved from EXPath, whereas the expression profiles of different developmental stages were downloaded from AtGenExpress, and RiceXPro (18–20). To sum up, we collected 583, 536, and 162 microarray samples from *Arabidopsis* (*A. thaliana*), rice (*O. sativa*) and maize (*Zea mays*), respectively (Supplementary Table S1). Users can customize the Pearson/Spearman's rank correlation coefficient and the conditions for assessing expression similarities among the TF-gene pairs.

Gene annotation databases and promoter sequences

In PlantPAN 2.0, gene information of *Arabidopsis*, rice, and maize were updated according to the latest version of TAIR (TAIR10_genome_release), RAP-DB and MaizeGDB (21–23). On the other hand, since previous studies revealed that downstream sequences may be the factors that require regulation by vital TFs, PlantPAN 2.0 offers not only upstream sequences but also downstream sequences of genes for TFBS scanning (24). Users can customize either the upstream region from −2000bp to +500bp (or 5′UTR-end) relative to the transcription start site (TSS) or the downstream regions to +2000 relative to the transcription stop site (or full-length 3′UTR) as gene promoters. In total, 41 671, 44 553, and 63 391 gene transcripts were collected from *Arabidopsis*, rice and maize, respectively.

Comprehensive curation of TF information

In plants, numerous TFs co-regulate gene regulatory networks involved in different cell types, developmental stages and responses to environmental stresses. During gene transcriptional regulation, activated TFs bind short DNA segments to activate or repress the expression of target genes. To integrate regulatory information for TFs, the descriptions and functions of TFs were retrieved from the Uniprot database, which contains the most complete annotation data of proteins (15). Each description was carefully reviewed and information such as response conditions, target genes, regulatory types (repressor/activator), and corresponding binding motifs was collected. Through the TF/TFBS search function in PlantPAN 2.0, users can access comprehensive TF information for further analysis.

For the TF search, detailed information including the TF family, a short description, the species, regulatory types, inductive condition, involvement of developmental stages/biological pathways, targets, functions, and binding sequence logos are displayed by searching for the TF name/locus or by browsing all available TFs by family or species. Take circadian clock associated 1 (CCA1; AT2G46830) as an example. Figure 2 demonstrates two experimentally verified binding sites and regulatory information for CCA1. As to the TFBS search, users can submit their own motifs in the form of an IUPCA code and can specify the plant species of interest. Similar TFBSs are distinguished after comparing the input motifs with TFBSs in PlantPAN2.0 by using the Tomtom program (25). The threshold of Tomtom program was set at 0.05. Alternatively, users can input keywords (e.g. heat, GCC-box) to search for the TFBSs and their corresponding TFs.

PPIs among TFs and their co-factors

As is already known, most eukaryotic TFs collaborate with various co-factors to dominate biological function under specific conditions. The regulatory types and the DNA-binding specificities of TFs depend on which co-factors they interact with. In addition, the protein complexes formed by TFs and their co-factors promote efficiency related to recruiting the pre-initiation complex and the RNA polymerase. Hence, it is thus critical to identify TFs and their cofactors in transcriptional regulation. To address this issue, the protein-protein interactions (PPIs) from PAIR, MINT, and BioGRID were incorporated in PlantPAN 2.0 for the purpose of reconstructing transcription regulatory networks (26–28). The total number of experimentally verified and computationally predicted PPIs was found to be 16030 and 145 239, respectively.

Enhanced user interface and other new features

The regulatory elements discovered in the promoters were graphically displayed or tabulated. Through the enhanced user interface in PlantPAN 2.0, the TF-target networks were displayed and reconstructed according to input genes, corresponding TFs, and co-factors interacting with the TFs. To ameliorate the visualization of transcriptional regulatory networks, PlantPAN 2.0 uses the web-based network visualization tool, Cytoscape, to visualize either the relations among a input gene and the predicted TFs or the PPIs among the TFs and their co-factors (29). This visualization can help users understand complicated TF-targets networks.

TF ID (Gene ID)	AT2G46830
TF Family	MYB-related
Gene Name	CCA1
Short Description	circadian clock associated 1; Protein CCA1 (MYB-related transcription factor CCA1) (Protein CIRCADIAN CLOCK ASSOCIATED 1) [CCA1] [F19D11]
Pfam family	Accession: PF00249 / Type: Domain / Description: Myb-like DNA-binding domain
Species	Arabidopsis thaliana
Taxonomy	Land plants
Activator / Repressor	Activator / Repressor
Involved condition	involved in the circadian clock and in the phytochrome regulation.
Target	APRR1/TOC1[Repressed] TCP21/CHE[Repressed] CAB2A[Activated] CAB2B[Activated] LHY[Repressed] AT2G46830[Repressed]
Function	Transcription factor involved in the circadian clock and in the phytochrome regulation. Binds to the promoter regions of APRR1/TOC1 and TCP21/CHE to repress their transcription. Binds to the promoter regions of CAB2A and CAB2B to promote their transcription. Represses both LHY and itself.
Method	PBM
Reference	Pubmed: 24477691
External links	www.arabidopsis.org www.uniprot.org
Sequence	Protein Sequence
TF Binding Sequence	1. TFmatrixID_0029 2. TFmatrixID_0030

Figure 2. An example of the output in TF search.

In the latest version of PlantPAN, the similar genes among Arabidopsis, rice, and maize are provided in the Cross Species analysis. Since the number of homologues in maize is much less than that in the two other species from the available resources, the similar genes among *Arabidopsis*, rice, and maize were determined by using BLAST in PlantPAN 2.0 (30). For the description, users have to specify species, gene locus and upstream/downstream regions. The protein sequence of input gene is extracted and aligned to all protein sequences of a specified species to identify the similar genes. In this study, the e-value of the BLAST program was set at 1e-20. After the identification of the paired similar genes, BLASTN was applied to define the conserved regions in the promoters. Finally, users were able to investigate the TFBSs, CpG islands, and tandem repeats of interest in these conserved regions. Additionally, to provide user web-based tool for detecting novel TFBSs in the Promoter Search, user-customized motifs can be employed to search binding sites in a promoter region using the Perl script we created.

APPLICATION AND DISCUSSION

Case study: the reconstruction of TRNs

PlantPAN 2.0 offers an effective platform for detecting TFBSs in a promoter or a set of promoters in plants. To reconstruct transcription regulatory networks, PlantPAN 2.0 provides not only co-expression profiles of TFs and their target genes under various conditions but also provides the PPIs among TFs and their co-factors. The case study given below demonstrates the application details.

In *A. thaliana*, ascorbate peroxidase 2 (APX2; AT3G09640) is an enzyme that plays a vital role in hydrogen peroxide detoxifying system by catalyzing the conversion of H_2O_2 into H_2O. Previous studies suggest that APX2 may act as heat-responsive gene that protects the cytosol and other cellular compartments from photooxidative damage during heat stress due to activation by heat stress transcription factors A2 (HSFA2; AT2G26150) (31,32). Promoter analysis in transient assays using a GUS reporter gene has revealed that HSFA2 binds directly to the promoter regions of APX2 (32). In this example, the APX2 promoter (upstream of −1000 to +100) was used to search for regulatory elements. Consequently, three HSFA2 binding sites were identified in the APX2 promoter (Supplementary Figure S1A). The result is consistent with that of pervious findings. Moreover, based on the results of the co-expression analysis, HSFA2 was co-expressed with APX2 under environmental stresses rather than during hormone treatments and in the developmental stages. This may imply that the activation of APX2 gene expression by HSFA2 occurs exclusively under environmental stresses. The PPIs among HSFA2 and its co-factors such as HSFA1B and HSP90.1 were illustrated in the TRN of APX2 (Supplementary Figure S1B). In addition to HSFA2, the TRN of APX2 introduces a novel TF, early-phytochrome-responsive1 (EPR1). A previous study showed that EPR1 is regulated by both phytochrome A and phytochrome B and is involved in a regulatory circadian feedback loop (33). These findings might suggest that EPR1 acts as regulator of active gene expression of APX2 and reduces photooxidative damage under environmental stresses. This gives a new avenue for further experimental validation of APX2. The case of APX2 reveals the reliability of the application of PlantPAN 2.0 for reconstruction of transcription regulatory networks.

Case study: TRNs of gene group analysis

As is already known, TFs play vital roles in activating plant defense responses and maintaining normal physiology. The WRKY TF superfamily is generally regarded as being highly specific and effective against plant pathogens (34). In a previous study, Noëllie *et al.* constructed gene regulatory network of WRKY17 (AT2G24570) from microarray data and Q-RT-PCR experiments under *Pseudomonas*

syringae pv *tomato* (*Pst*) infection (35). During the *Pst* infection, WRKY17 activated *Pst*-induced jasmonic acid (JA) -dependent responses. Two genes encoding JA biosynthetic enzymes, AOS (AT5G42650.1) and LOX2 (AT3G45140.1) were promoted by WRKY17, whereas the negative regulator of JA signaling, WRKY70 (AT3G56400.1), was repressed. In addition, two other WRKY genes, WRKY54 (AT2G40750.1) and WRKY11 (AT4G31550.1), were also respectively influenced by WRKY17. To demonstrate the effectiveness of PlantPAN 2.0, the promoter regions (upstream of −1000 to +100) of these five genes were used to analyze the co-occurrence of combinatorial TFBSs in the Gene Group analysis. The support and confidence values were set to the default values. Consequently, the binding sites of WRKY17 in these five promoters are identical to an already-known regulatory network, as shown in Supplementary Figure S2A (35). The TRNs of five genes display PPIs among WRKY17, CAM2, CAM3 and CAM5 from PAIR (Supplementary Figure S2B). Moreover, WRKY17 and most of the TFs show the co-occurrence TFBSs in the promoter regions of the five genes under consideration (Supplementary Figure S2C). Supplementary Figure S2D illustrates that WRKY11 co-regulates the same WRKY17 targets which is consistent with previous findings (35). Accordingly, the other WRKY TFs or abscisic acid responsive TFs such as ABF3and AREB2 might be the new candidates involved in the regulation of *Pst* infection. The PlantPAN 2.0 thus helps users understand transcription regulation networks in plants.

CONCLUSIONS

PlantPAN 2.0 provides an informative resource for detecting TFBSs, corresponding TFs, and other important regulatory elements (CpG islands and tandem repeats) in a promoter or a set of promoters in plants. In the current release, experimentally verified TF binding matrices were curated from previous studies. PlantPAN 2.0 contains 16 960 TFs and 1143 matrices of TF binding sites among 76 plant species covering major lineages of plants. Moreover, the TF/TFBS search function in PlantPAN 2.0 was developed for the purpose of offering integrated regulatory information for TFs. To support high-confidence TF-target relations, we also collected numerous microarray expression data to facilitate co-expression analysis among TFs and their target genes under various conditions. The visualization of PPIs among TFs and co-factors in PlantPAN 2.0 will be helpful for users to understand complicated transcriptional regulatory networks. In a future work, we plan to obtain more detailed information on gene regulation and TFs, such as DNA modifications and chromatin remodeling. The PlantPAN resource will be continuously maintained and updated for these upcoming studies.

AVAILABILITY

The PlantPAN 2.0 is via a web interface, freely available to all interested users, at http://PlantPAN2.itps.ncku.edu.tw/.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Science and Technology (MOST 103–2311-B-006–001, MOST 104–2627-M-009–008, MOST 103–2628-B-009 -001 -MY3, MOST 104–2319-B-400–002 and MOST 103–2911-I-009–101) and Academia Sinica (Innovative Translational Agricultural Research Grant) of the Republic of China for financially supporting this research.

FUNDING

Ministry of Science and Technology [MOST 103-2311-B-006–001, MOST 104-2627-M-009–008, MOST 103-2628-B-009-001-MY3, MOST 104-2319-B-400-002 and MOST 103-2911-I-009-101]; Academia Sinica [Innovative Translational Agricultural Research Grant] of the Republic of China. Funding for open access charge: Academia Sinica [Innovative Translational Agricultural Research Grant].
Conflict of interest statement. None declared.

REFERENCES

1. Matys,V., Kel-Margoulis,O.V., Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M., Hornischer,K. *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
2. Wingender,E. (2008) The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Bioinformatics*, **9**, 326–333.
3. Mathelier,A., Zhao,X., Zhang,A.W., Parcy,F., Worsley-Hunt,R., Arenillas,D.J., Buchman,S., Chen,C.Y., Chou,A., Ienasescu,H. *et al.* (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **42**, D142–D147.
4. Higo,K., Ugawa,Y., Iwamoto,M. and Korenaga,T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.*, **27**, 297–300.
5. Hehl,R. and Bulow,L. (2014) AthaMap web tools for the analysis of transcriptional and posttranscriptional regulation of gene expression in Arabidopsis thaliana. *Methods Mol. Biol.*, **1158**, 139–156.
6. Lescot,M., Dehais,P., Thijs,G., Marchal,K., Moreau,Y., de Peer,Y., Rouze,P. and Rombauts,S. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.*, **30**, 325–327.
7. Chen,Y.A., Wen,Y.C. and Chang,W.C. (2012) AtPAN: an integrated system for reconstructing transcriptional regulatory networks in Arabidopsis thaliana. *BMC Genomics*, **13**, 85.
8. Yilmaz,A., Mejia-Guerra,M.K., Kurz,K., Liang,X., Welch,L. and Grotewold,E. (2011) AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. *Nucleic Acids Res.*, **39**, D1118–D1122.
9. Chang,W.C., Lee,T.Y., Huang,H.D., Huang,H.Y. and Pan,R.L. (2008) PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. *BMC Genomics*, **9**, 561.
10. Berger,M.F., Philippakis,A.A., Qureshi,A.M., He,F.S., Estep,P.W. 3rd and Bulyk,M.L. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.*, **24**, 1429–1435.
11. Hume,M.A., Barrera,L.A., Gisselbrecht,S.S. and Bulyk,M.L. (2015) UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.*, **43**, D117–D122.
12. Geertz,M. and Maerkl,S.J. (2010) Experimental strategies for studying transcription factor-DNA binding specificities. *Brief. Funct. Genomic.*, **9**, 362–373.
13. Franco-Zorrilla,J.M., Lopez-Vidriero,I., Carrasco,J.L., Godoy,M., Vera,P. and Solano,R. (2014) DNA-binding specificities of plant

transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2367–2372.
14. Weirauch,M.T., Yang,A., Albu,M., Cote,A.G., Montenegro-Montero,A., Drewe,P., Najafabadi,H.S., Lambert,S.A., Mann,I., Cook,K. *et al.* (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*, **158**, 1431–1443.
15. The UniProt, Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
16. Kel,A.E., Gößling,E., Reuter,I., Cheremushkin,E., Kel-Margoulis,O.V. and Wingender,E. (2003) MATCHTM: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.*, **31**, 3576–3579.
17. Matys,V., Fricke,E., Geffers,R., Gößling,E., Haubrock,M., Hehl,R., Hornischer,K., Karas,D., Kel,A.E. and Kel-Margoulis,O.V. (2003) TRANSFAC®: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.*, **31**, 374–378.
18. Chien,C.H., Chow,C.N., Wu,N.Y., Chiang-Hsieh,Y.F., Hou,P.F. and Chang,W.C. (2015) EXPath: a database of comparative expression analysis inferring metabolic pathways for plants. *BMC Genomics*, **16** (Suppl. 2), S6.
19. Kilian,J., Whitehead,D., Horak,J., Wanke,D., Weinl,S., Batistic,O., D'Angelo,C., Bornberg-Bauer,E., Kudla,J. and Harter,K. (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.*, **50**, 347–363.
20. Sato,Y., Takehisa,H., Kamatsuki,K., Minami,H., Namiki,N., Ikawa,H., Ohyanagi,H., Sugimoto,K., Antonio,B.A. and Nagamura,Y. (2013) RiceXPro version 3.0: expanding the informatics resource for rice transcriptome. *Nucleic Acids Res.*, **41**, D1206–D1213.
21. Lamesch,P., Berardini,T.Z., Li,D., Swarbreck,D., Wilks,C., Sasidharan,R., Muller,R., Dreher,K., Alexander,D.L., Garcia-Hernandez,M. *et al.* (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.*, **40**, D1202–D1210.
22. Sakai,H., Lee,S.S., Tanaka,T., Numa,H., Kim,J., Kawahara,Y., Wakimoto,H., Yang,C.C., Iwamoto,M., Abe,T. *et al.* (2013) Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.*, **54**, e6.
23. Schaeffer,M.L., Harper,L.C., Gardiner,J.M., Andorf,C.M., Campbell,D.A., Cannon,E.K., Sen,T.Z. and Lawrence,C.J. (2011) MaizeGDB: curation and outreach go hand-in-hand. *Database: J. Biol. Databases Curation*, bar022.
24. Heyndrickx,K.S., Van de Velde,J., Wang,C., Weigel,D. and Vandepoele,K. (2014) A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. *Plant Cell*, **26**, 3894–3910.
25. Gupta,S., Stamatoyannopoulos,J.A., Bailey,T.L. and Noble,W.S. (2007) Quantifying similarity between motifs. *Genome Biol.*, **8**, R24.
26. Chatr-Aryamontri,A., Breitkreutz,B.J., Oughtred,R., Boucher,L., Heinicke,S., Chen,D., Stark,C., Breitkreutz,A., Kolas,N., O'Donnell,L. *et al.* (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res.*, **43**, D470–D478.
27. Licata,L., Briganti,L., Peluso,D., Perfetto,L., Iannuccelli,M., Galeota,E., Sacco,F., Palma,A., Nardozza,A.P., Santonico,E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
28. Lin,M., Shen,X. and Chen,X. (2011) PAIR: the predicted Arabidopsis interactome resource. *Nucleic Acids Res.*, **39**, D1134–D1140.
29. Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
30. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
31. Caverzan,A., Passaia,G., Rosa,S.B., Ribeiro,C.W., Lazzarotto,F. and Margis-Pinheiro,M. (2012) Plant responses to stresses: Role of ascorbate peroxidase in the antioxidant protection. *Genet. Mol. Biol.*, **35**, 1011–1019.
32. Schramm,F., Ganguli,A., Kiehlmann,E., Englich,G., Walch,D. and von Koskull-Doring,P. (2006) The heat stress transcription factor HsfA2 serves as a regulatory amplifier of a subset of genes in the heat stress response in Arabidopsis. *Plant Mol. Biol.*, **60**, 759–772.
33. Kuno,N., Moller,S.G., Shinomura,T., Xu,X., Chua,N.H. and Furuya,M. (2003) The novel MYB protein EARLY-PHYTOCHROME-RESPONSIVE1 is a component of a slave circadian oscillator in Arabidopsis. *Plant Cell*, **15**, 2476–2488.
34. Pandey,S.P. and Somssich,I.E. (2009) The role of WRKY transcription factors in plant immunity. *Plant Physiol.*, **150**, 1648–1655.
35. Journot-Catalino,N., Somssich,I.E., Roby,D. and Kroj,T. (2006) The transcription factors WRKY11 and WRKY17 act as negative regulators of basal resistance in Arabidopsis thaliana. *Plant Cell*, **18**, 3289–3302.