DNA RESEARCH **20**, 151–162, (2013) doi:10.1093/dnares/dss040
Advance Access publication on 2 January 2013

TreeTFDB: An Integrative Database of the Transcription Factors from Six Economically Important Tree Crops for Functional Predictions and Comparative and Functional Genomics

KEIICHI Mochida[1,2,3,*], TAKUHIRO Yoshida[1], TETSUYA Sakurai[1], KAZUKO Yamaguchi-Shinozaki[4], KAZUO Shinozaki[1,2], and LAM-SON PHAN Tran[1,*]

RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan[1]; RIKEN Biomass Engineering Program, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan[2]; Kihara Institute for Biological Research, Yokohama City University, 641-12 Maioka-cho, Totsuka-ku, Yokohama, Kanagawa 244-0813, Japan[3] and Japan International Research Center for Agricultural Sciences, Ibaraki 305-8686, Japan[4]

*To whom correspondence should be addressed. Tel. +81 45-503-9183. Fax. +81 45-503-9182. Email: mochida@psc.riken.jp (K.M.); Tel. +81 45-503-9593. Fax. +81 45-503-9591. Email: tran@psc.riken.jp (L-S.P.T.).

Edited by Prof. Hiroyuki Toh
(Received 10 September 2012; accepted 10 December 2012)

Abstract

Crop plants, whose productivity is affected by a wide range of growing and environmental conditions, are grown for economic purposes. Transcription factors (TFs) play central role in regulation of many biological processes, including plant development and responses to environmental stimuli, by activating or repressing spatiotemporal gene expression. Here, we describe the TreeTFDB (http://treetfdb.bmep.riken.jp/index.pl) that houses the TF repertoires of six economically important tree crop species: *Jatropha curcas*, papaya, cassava, poplar, castor bean and grapevine. Among these, the TF repertoire of *J. curcas* has not been reported by any other TF databases. In addition to their basic information, such as sequence and domain features, domain alignments, gene ontology assignment and sequence comparison, information on available full-length cDNAs, identity and positions of all types of known *cis*-motifs found in the promoter regions, gene expression data are provided. With its newly designed and friendly interface and its unique features, TreeTFDB will enable research community to predict the functions and provide access to available genetic resources for performing comparative and functional genomics of the crop TFs, either individually or at whole family level, in a comprehensive and convenient manner.

Key words: tree crops; transcription factor; database; expression; *cis*-motif

1. Introduction

Transcription factors (TFs), which bind to *cis*-regulatory sequences and either activate or repress gene expression, control many biological processes, such as development, growth, cell division and responses to environmental stimuli.[1–6] The TFs form complex regulatory networks at the transcriptional level and through protein–protein interactions among themselves or with proteins of other classes such as RING E3 ligases for degradation or chromatin modifying proteins to recruit or block access of RNA polymerases to the DNA template.[7–9] The specific interactions between TFs and consensus *cis*-motifs play a key role in the regulation of proteins to affect spatial and temporal gene expression.[10] Proper characterization of particular TFs often requires study in the biological context of a whole family because functional redundancy is common within the families. Identification, characterization and annotation of TF repertoires from different plant species will provide an insight into TF organization and evolution as well as a basic foundation for comparative and functional genomics of the TFs.[11,12] Additionally, from a biotechnology perspective, identification and functional analyses of TFs are especially important for improvement of plant performance and seed quality under adverse conditions.[3,13]

Given the importance of TFs, in the last decade, the availability of complete plant genome sequences and the development of various software tools have enabled scientists to identify TF sets from various plant species and compile their basic information in a number of species-specific TF databases[14] such as RARTF,[15] DRTF,[16] SoybeanTFDB,[17] SoyDB[18] and TOBFAC[19] or integrative databases such as DATFAP,[20] Grassius,[21] PlnTFDB,[22] LegumeTFDB,[23] GramineaeTFDB[24] and PlantTFDB.[25] Among these databases, the SoybeanTFDB (http://soybeantfdb.psc.riken.jp/), LegumeTFDB (http://legumetfdb.psc.riken.jp/) and GramineaeTFDB (http://gramineaetfdb.psc.riken.jp) have been constructed by our group during the last three years with the focus on economically important legume and grass crops.[26] The recent completion of the genome sequences of several important tree crops has enabled us to develop the TreeTFDB (http://treetfdb.bmep.riken.jp/index.pl) to accelerate biomass engineering researches for development of bioenergy resources through genetic engineering of TFs. TreeTFDB comprises the TF sets from six important tree crops, namely the jatropha (*Jatropha curcas*), papaya (*Carica papaya*), cassava (*Manihot esculenta*), poplar (*Populus trichocarpa*), castor bean (*Ricinus communis*) and grapevine (*Vitis vinifera*). Among these, the TF set of jatropha has not been reported by any other public databases because the genome sequence of jatropha has just been made available to public very recently.[27] In addition to the basic information, such as sequence features, domain alignments and gene ontology (GO) assignment, TreeTFDB also provides information on available full-length cDNA (FL-cDNA), genomic distribution, multiple sequence alignment, gene expression data and the sequences and positions of all types of known *cis*-motifs found in the promoter regions.

2. Materials and methods

2.1. Identification of TF repertoires in six tree crops

The bioinformatics approach established previously was used to identify the complete sets of TFs from the annotated proteomes of jatropha, papaya, cassava, poplar, castor bean and grapevine.[17,28] A predefined threshold of $E < 1e\text{-}5$ was used as the common value cutoff for HMMER search using 61 built HMM profiles.[17,29] The criteria described previously for the classification of each TF family were applied.[30] The TFs identified by initial HMMER search were subsequently subjected to homology search (blastp E-value ≤1e-30) with known TFs of *Arabidopsis* and rice to confirm the HMMER search results based on the results of homology search as described previously.[24] TFs for which the homology search yielded results of blastp E-value between 1e-30 and 1e-5 were inspected manually to exclude false-positive hits and determine the true E-value for each family (TreeTFDB Help page, Statistics).

2.2. Structural and functional annotations of putative TFs

Structural and functional annotations of putative TFs were done as described previously.[17] All the similarity searches using blastn were performed with a threshold E-value < 1e-100 against available nucleotide sequence resources, such as clustered expressed sequence tags and target sequences of microarrays, and the top scoring hit for each query was applied and displayed in ‘Corresponding to other identifiers’. All similarity searches with blastp against protein datasets [National Center for Biotechnology Information (NCBI) nr, *Arabidopsis* annotated proteins from TAIR10, UniProt, rice annotated proteins from Michigan State University rice genome annotation project] were performed with a threshold E-value < 1e-5 to find possible functional descriptions for TF-encoding genes and putative homologs among tree TFs in TreeTFDB. The top scoring hit for each query was applied. The TFs were assigned to possible GO terms based on a blastp similarity search (E-value < 1e-10) using the dataset of *Arabidopsis* of TAIR10 as previously described.[24,31]

2.3. FL-cDNA resources for TF-encoding genes

Hyperlinks linking putative TF-encoding genes of poplar and cassava to their respective FL-cDNA resources publicly available at http://rpop.psc.riken.jp/index.pl (RPOPDB: RIKEN Populus Database) and http://cassava.psc.riken.jp/ver.1/ (Cassava Online Archive) databases were built and supplied on TreeTFDB. FL-cDNA information was also used to verify the annotation of the poplar and cassava TF-encoding genes during the manual inspection.

2.4. Promoter regions and discovery of *cis*-regulatory motifs in promoter regions of TF genes

The −500, −1000 and −3000 bp upstream sequences from the putative transcription start site (TSS) for each TF-encoding gene were extracted from the respective genomic sequence dataset of each tree crop. TSSs were defined as start site of ‘mRNA’ recorded in General Feature Format files that are provided on the Phytozome (http://www.phytozome.net/) and Kazusa (http://www.kazusa.or.jp/) databases. Discovery of *cis*-regulatory motifs located in these promoter regions using all the *cis*-motif sequences collected from the PLACE database (http://www.dna.affrc.go.jp/PLACE/)[32] and 12 major stress- and/or hormone-responsive *cis*-motifs reported previously[10,33] was done as described previously.[17] The *cis*-element search results were implemented into the TreeTFDB as a searchable property. In addition, these search results were also incorporated as an annotation track of Generic Genome Browser (Gbrowse).[34]

2.5. *Expression data for TF-encoding genes*

Gene expression data of putative TF-encoding genes in poplar and grape were obtained based on correspondences between putative TF-encoding genes and GeneChip target sequences. Subsequently, the hyperlinks to PlexDB (http://www.plexdb.org/) and eFP browser (http://bar.utoronto.ca/efp_poplar/cgi-bin/efpWeb.cgi) for poplar TF-encoding genes and those to PlexDB for grapevine TF-encoding genes were built and supplied on TreeTFDB. In addition, gene expression patterns in castor bean obtained through analysis of RNA-sequencing transcriptomes[35] were also compiled, and expression data available for corresponding TFs were extracted and supplied on TreeTFDB.

2.6. *Construction of a web-accessible database*

The database is implemented in MySQL and the web interface of Perl CGI and Java script run on the Apache Web server. The definition strings used for sequence similarity searches for each database, the domain searches by InterProScan, *cis*-motif names from the PLACE database and the assigned GO terms were assembled as a keyword database, enabling users to specify queries on any keyword and to retrieve relevant information for genes from the TreeTFDB. A BLAST server was implemented to provide a similarity search interface for queried sequences using NCBI BLAST together with sequences of the six tree crops, as well as those from *Arabidopsis*. Gbrowse was also implemented in TreeTFDB to visualize the gene annotations of the putative TF-encoding genes together with *cis*-motifs found on the upstream sequence of the TF genes. The cross-references of corresponding data for each of the entries were also implemented into the TreeTFDB together with the URLs for each of the original referenced data to provide hyperlinks on the web interface with seamless navigations.

3. Results and discussion

3.1. *Identification of putative TFs in jatropha, papaya, cassava, poplar, castor bean and grapevine*

The strategy and bioinformatics pipeline established previously for identification of the TF repertoire of soya bean were used to identify the complete sets of TFs from the annotated proteome of jatropha.[17] To our knowledge, the TF set of jatropha has not been reported by any other databases. Although the TF repertoires of papaya, cassava, poplar, castor bean and grapevine are available in PlantTFDB,[25] with the aim to make the comparative and functional genomics of the tree crop TFs convenient, we also used the same bioinformatics approach to identify the TF repertoires of these five tree crops. The sequence data of jatropha (JAT_r3.0) were downloaded from Kazusa DNA Research Institute (http://www.kazusa.or.jp/jatropha/), whereas for the papaya (Cpapaya_113), cassava (Cassava4.1 gene set, Mesculenta_147), poplar (JGI v2.2 gene annotation, Ptrichocarpa_156), castor bean (TIGR/JCVI release v0.1, Rcommunis_119) and grapevine (Genoscope 12X March 2010 release, Vvinifera_145), the release version of annotated proteome datasets was downloaded from Phytozome (v.7.0) (http://www.phytozome.net/). The results gained from the bioinformatics pipeline using a predefined threshold of $E < 1e\text{-}5$ were subsequently refined by combined automatic and manual inspections of the raw alignments to exclude false-positive hits and determine the true E-value for each TF family. A total of 1548, 1481, 2636, 3106, 1510 and 1492 TF models were identified in papaya, jatropha, cassava, poplar, castor bean and grapevine, respectively. As our pipeline is similar to the method used by PlnTFDB,[36] among the TF repertoires of three tree species reported by both PlnTFDB and TreeTFDB, two species have difference in TF numbers with less than 5% (1483 and 2930 in PlnTFDB versus 1548 and 3106 in TreeTFDB for *C. papaya* and *P. trichocarpa*, respectively). As for the third tree species, *V. vinifera*, the difference is approximately 14% (1728 in PlnTFDB versus 1492 in TreeTFDB), which may be due to the use of different sequence versions for TF identification. PlnTFDB used Genoscope, release version 1.0, whereas we used the newest version from Phytozome.

All the 6 TF sets were classified into 61 TF families based on the presence of their specific signature domains (Table 1). The number of predicted TFs in each family may be changed by future updating of sequence versions and/or fine-tuning of gene annotations. The identified TF sets of the six tree crops together with their basic features, such as Gbrowse to provide the genomic structure and position of each TF entry, gene and protein sequence information

Table 1. Predicted TF models in six tree crop species

TF families	*C. papaya*	*J. curcas*	*M. esculenta*	*P. trichocarpa*	*R. communis*	*V. vinifera*
(R1)R2R3_Myb	98	110	177	204	92	132
ABI3VP1	39	54	79	128	42	30
AP2_EREBP	97	111	198	219	116	103
ARF	12	14	39	50	0	20
ARID	7	5	12	14	9	9
Alfin-like	4	5	7	10	5	6
Aux_IAA	21	26	48	38	38	24
BBR-BPC	2	4	7	20	4	3
BES1	6	7	13	17	7	6
C2C2_Zn-CO-like	23	22	43	48	22	19
C2C2_Zn-Dof	20	22	48	48	23	22
C2C2_Zn-GATA	23	19	40	46	18	19
C2C2_Zn-YABBY	8	7	15	14	6	7
C2H2_Zn	80	86	154	167	100	67
C3H-TypeI	50	59	113	104	54	64
CAMTA	4	3	5	9	5	4
CCAAT_Dr1	1	1	3	2	1	1
CCAAT_HAP2	5	6	17	19	6	7
CCAAT_HAP3	10	9	18	22	13	16
CCAAT_HAP5	3	7	14	15	10	6
CPP	4	6	9	16	6	6
E2F_DP	6	7	10	10	6	7
EIL	3	3	2	8	4	2
GARP_ARRB	8	5	18	22	10	10
GARP_G2-like	33	23	60	76	32	39
GRAS	37	41	75	104	46	39
GRF	2	5	9	12	14	5
GeBP	4	6	7	8	4	1
HB	64	56	135	157	71	78
HMG-box	8	12	20	16	10	9
HRT	2	1	1	1	1	1
HSF	18	21	35	33	19	19
JUMONJI	17	16	21	27	18	19
LFY	1	1	2	1	1	1
LIM	8	7	18	25	9	10
LUG	5	1	8	11	4	3
MADS	229	54	77	99	39	54
MBF1	2	2	3	3	2	2
Myb-related	36	34	83	87	43	43
NAC	80	86	127	183	96	71
NOZZILE	0	0	0	0	0	0
Nin-like	6	8	12	20	10	8
PHD	116	83	176	207	107	103
PLATZ	11	9	21	23	11	13
PcG	28	29	38	59	33	37
S1Fa-like	1	1	1	2	1	2

Continued

Table 1. Continued

TF families	*C. papaya*	*J. curcas*	*M. esculenta*	*P. trichocarpa*	*R. communis*	*V. vinifera*
SAP	1	1	2	1	1	1
SBP	11	14	24	34	15	19
SRS	4	6	11	11	5	5
TCP	22	17	33	35	22	15
TUB	6	6	19	16	7	11
Trihelix	14	12	22	27	13	10
ULT	3	2	3	2	2	1
VOZ	2	1	3	4	2	2
WRKY_Zn	50	59	114	122	59	59
Whirly	2	0	3	3	2	2
ZIM	13	11	31	27	12	16
atypical_MYB	21	19	42	54	23	28
bHLH	96	87	189	220	117	113
bZIP	49	136	91	113	52	50
zf-HD	10	9	24	26	11	10
zf-TAZ	6	7	9	11	5	4
Total	1548	1481	2636	3106	1510	1492

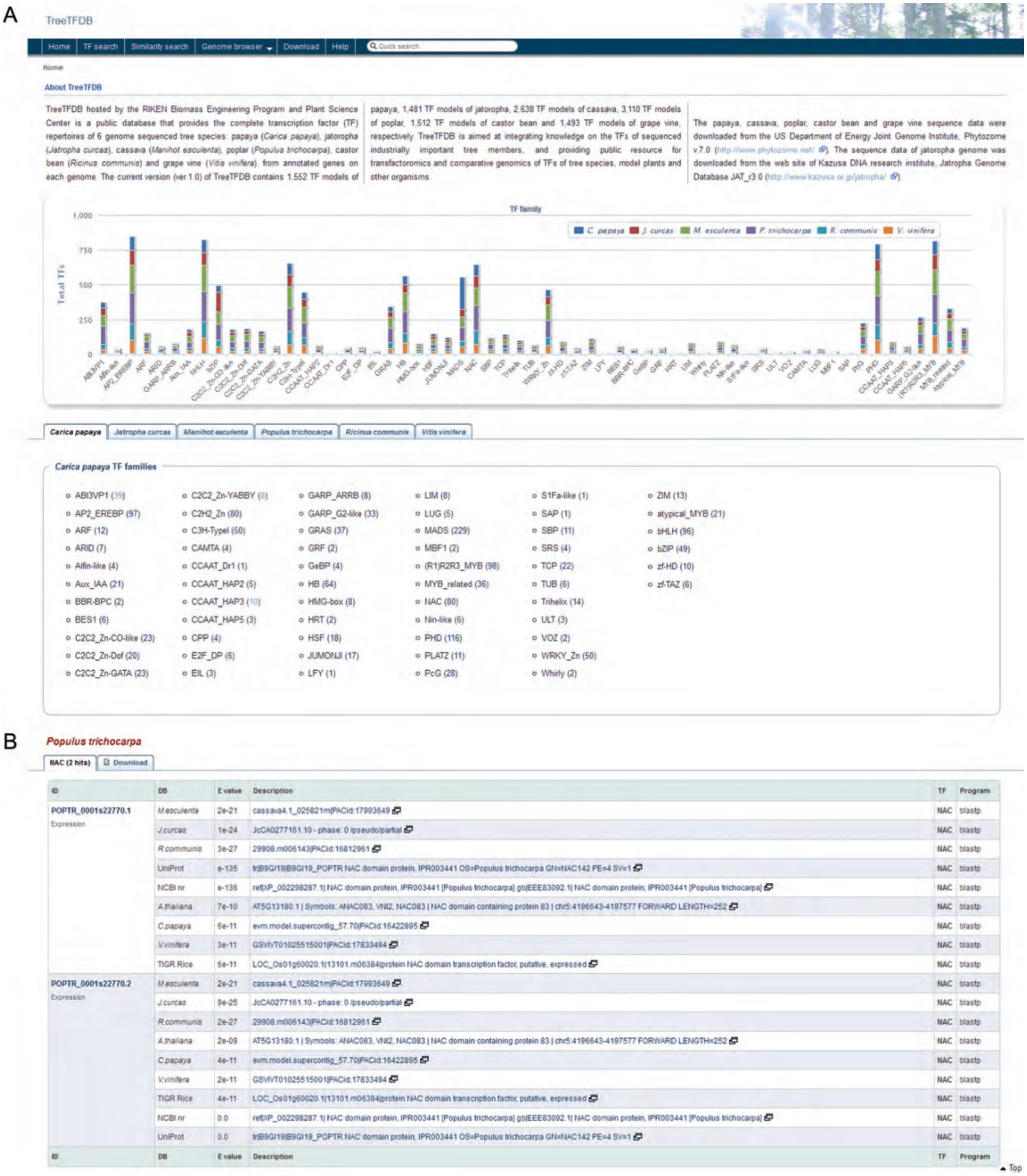
and domain alignments, were integrated into a database named TreeTFDB (http://treetfdb.bmep.riken.jp/index.pl) (Fig. 1). We will continue to update our database with new information, when it becomes available, to enhance the accuracy of TF prediction and annotation.

3.2. *Annotations of the tree crop TFs*

In the next line of the study, extensive annotations at both gene and family levels were carried out to provide comprehensive information on the identified TFs of the six tree crops (for details, see the TreeTFDB Help page). A search for potential functions of the identified TFs of these tree crops by analysis of published papers on PubMed database has revealed that the majority of their TFs are functionally unknown. Thus, as a means to provide a knowledge base regarding their biological function, especially in stress responses and plant development, the putative functions of the identified TFs of the tree crops were assessed via comparative analyses with relevant GO annotations of *Arabidopsis*. All the assigned terms at the biological process level obtained from sequence similarity searches against *Arabidopsis* counterparts that have GO terms in TAIR10 were counted to grasp the overall representation of 11 broad categories (GO slim) in applied entries of crop TFs (Fig. 2). A substantial proportion of the analysed TFs were shown to be related to stresses and/or their stimuli as well as plant development (Fig. 2). These data indicate that regulation of stress-related biological processes can also be considered one of important events as for the functionality of the plant TFs. This GO-based functional annotation overall provides an insight into potential functions of identified TFs of these tree crops, supporting researchers in selection of TFs of interest for further studies.

3.3. *Promoter regions of the identified TFs and discovery of known* cis*-element(s) in these TF promoter regions*

The specific interactions between *cis*-regulatory sequences and a family of TFs play a central part in how genetic regulatory proteins affect spatial and temporal gene expression.[4] Increasing evidence has suggested that the *cis*-motifs are highly conserved among plant species, and defined *cis*-elements can effectively aid in functional prediction of genes[37–40] such as stress-responsive or hormone-responsive genes based on combinatory relationships of *cis*-motifs.[38,41] In plants, extensive promoter analyses have identified a large number of *cis*-motifs that are known to be essential for transcriptional regulation of gene activities controlling various biological processes, including plant development, stress responses and hormone responses.[10,32,33,41] To make the promoter analyses convenient for the geneticists and to facilitate the functional characterization and prediction of the TFs, especially the stress-related TFs, first we retrieved the −500, −1000 and −3000



About TreeTFDB

TreeTFDB hosted by the RIKEN Biomass Engineering Program and Plant Science Center is a public database that provides the complete transcription factor (TF) repertoires of 6 genome sequenced tree species: papaya (*Carica papaya*), jatoropha (*Jatropha curcas*), cassava (*Manihot esculenta*), poplar (*Populus trichocarpa*), castor bean (*Ricinus communis*) and grape vine (*Vitis vinifera*), from annotated genes on each genome. The current version (ver.1.0) of TreeTFDB contains 1,552 TF models of papaya, 1,481 TF models of jatoropha, 2,638 TF models of cassava, 3,110 TF models of poplar, 1,512 TF models of castor bean and 1,493 TF models of grape vine, respectively. TreeTFDB is aimed at integrating knowledge on the TFs of sequenced industrially important tree members, and providing public resource for transfactoromics and comparative genomics of TFs of tree species, model plants and other organisms.

The papaya, cassava, poplar, castor bean and grape vine sequence data were downloaded from the US Department of Energy Joint Genome Institute, Phytozome v.7.0 (http://www.phytozome.net/). The sequence data of jatoropha genome was downloaded from the web site of Kazusa DNA research institute, Jatropha Genome Database JAT_r3.0 (http://www.kazusa.or.jp/jatropha/).

Carica papaya TF families

- ABI3VP1 (39)
- AP2_EREBP (97)
- ARF (12)
- ARID (7)
- Alfin-like (4)
- Aux_IAA (21)
- BBR-BPC (2)
- BES1 (6)
- C2C2_Zn-CO-like (23)
- C2C2_Zn-Dof (20)
- C2C2_Zn-GATA (23)
- C2C2_Zn-YABBY (8)
- C2H2_Zn (80)
- C3H-TypeI (50)
- CAMTA (4)
- CCAAT_Dr1 (1)
- CCAAT_HAP2 (5)
- CCAAT_HAP3 (10)
- CCAAT_HAP5 (3)
- CPP (4)
- E2F_DP (6)
- EIL (3)
- GARP_ARRB (8)
- GARP_G2-like (33)
- GRAS (37)
- GRF (2)
- GeBP (4)
- HB (64)
- HMG-box (8)
- HRT (2)
- HSF (18)
- JUMONJI (17)
- LFY (1)
- LIM (8)
- LUG (5)
- MADS (229)
- MBF1 (2)
- (R1)R2R3_MYB (98)
- MYB_related (36)
- NAC (80)
- Nin-like (6)
- PHD (116)
- PLATZ (11)
- PcG (28)
- S1Fa-like (1)
- SAP (1)
- SBP (11)
- SRS (4)
- TCP (22)
- TUB (6)
- Trihelix (14)
- ULT (3)
- VOZ (2)
- WRKY_Zn (50)
- Whirly (2)
- ZIM (13)
- atypical_MYB (21)
- bHLH (96)
- bZIP (49)
- zf-HD (10)
- zf-TAZ (6)

Populus trichocarpa

ID	DB	E value	Description	TF	Program
POPTR_0001s22770.1 Expression	*M.esculenta*	2e-21	cassava4.1_025821m|PACid:17993649	NAC	blastp
	J.curcas	1e-24	JcCA0277161.10 - phase: 0 /pseudo/partial	NAC	blastp
	R.communis	3e-27	29908.m006143|PACid:16812961	NAC	blastp
	UniProt	e-135	tr|B9GI19|B9GI19_POPTR NAC domain protein, IPR003441 OS=Populus trichocarpa GN=NAC142 PE=4 SV=1	NAC	blastp
	NCBI nr	e-136	ref|XP_002298287.1| NAC domain protein, IPR003441 [Populus trichocarpa] gb|EEE83092.1| NAC domain protein, IPR003441 [Populus trichocarpa]	NAC	blastp
	A.thaliana	7e-10	AT5G13180.1 | Symbols: ANAC083, VNI2, NAC083 | NAC domain containing protein 83 | chr5:4196643-4197577 FORWARD LENGTH=252	NAC	blastp
	C.papaya	6e-11	evm.model.supercontig_57.70|PACid:16422895	NAC	blastp
	V.vinifera	3e-11	GSVIVT01025515001|PACid:17833494	NAC	blastp
	TIGR Rice	6e-11	LOC_Os01g60020.1|13101.m06384|protein NAC domain transcription factor, putative, expressed	NAC	blastp
POPTR_0001s22770.2 Expression	*M.esculenta*	2e-21	cassava4.1_025821m|PACid:17993649	NAC	blastp
	J.curcas	9e-25	JcCA0277161.10 - phase: 0 /pseudo/partial	NAC	blastp
	R.communis	2e-27	29908.m006143|PACid:16812961	NAC	blastp
	A.thaliana	2e-09	AT5G13180.1 | Symbols: ANAC083, VNI2, NAC083 | NAC domain containing protein 83 | chr5:4196643-4197577 FORWARD LENGTH=252	NAC	blastp
	C.papaya	4e-11	evm.model.supercontig_57.70|PACid:16422895	NAC	blastp
	V.vinifera	2e-11	GSVIVT01025515001|PACid:17833494	NAC	blastp
	TIGR Rice	4e-11	LOC_Os01g60020.1|13101.m06384|protein NAC domain transcription factor, putative, expressed	NAC	blastp
	NCBI nr	0.0	ref|XP_002298287.1| NAC domain protein, IPR003441 [Populus trichocarpa] gb|EEE83092.1| NAC domain protein, IPR003441 [Populus trichocarpa]	NAC	blastp
	UniProt	0.0	tr|B9GI19|B9GI19_POPTR NAC domain protein, IPR003441 OS=Populus trichocarpa GN=NAC142 PE=4 SV=1	NAC	blastp
ID	DB	E value	Description	TF	Program

Figure 1. (A) The top page of TreeTFDB displays TF families and number of TFs of each TF family identified in six tree crops: *J. curcas*, *C. papaya*, *M. esculenta*, *P. trichocarpa*, *R. communis* and *V. vinifera*. By clicking on 'Go to TF search', the users will be directed to the search page that provides search queries for the names of TF families, keywords, sequence identifiers, identifiers of domains supported by InterProScan, GO terms and available *cis*-motifs for each tree species. (B) An example of a search result. By clicking on a species, such as *P. trichocarpa*, and a TF family, such as NAC, the users will be navigated to page listing search results for the TF family with a description of corresponding genes based on similarity searches.

promoter regions of all the TF genes (upstream from the TSSs) from the genomic sequences of the tree crops and provided these promoter sequences together with the TF gene and protein sequences on TreeTFDB for convenient downloading. Subsequently, these promoter regions were subjected to an

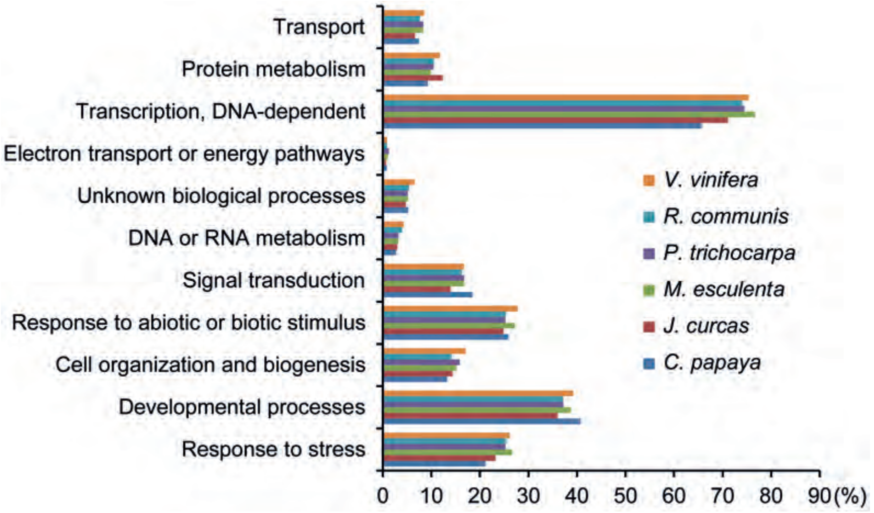
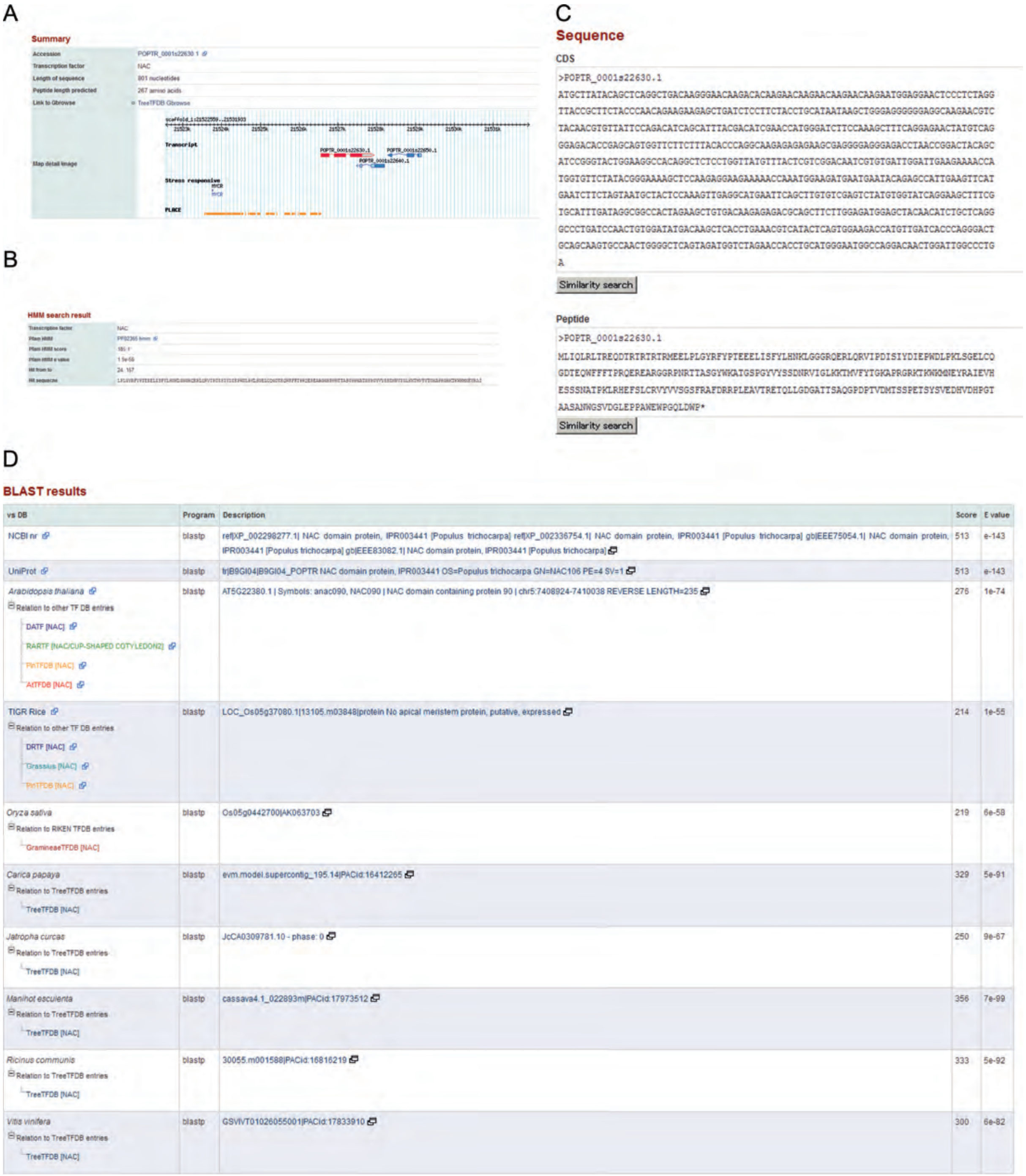


Figure 2. The representative distributions of the broad GO categories associated with TFs from *J. curcas*, *C. papaya*, *M. esculenta*, *P. trichocarpa*, *R. communis* and *V. vinifera*. The GO categories were assigned based on homology searches against annotated *Arabidopsis* genes (blastp homology search with E-value < 1e-10).

extensive computational analysis to search for the existence of a total of 481 putative known *cis*-regulatory motifs, including 12 major abiotic stress- and/or hormone-responsive *cis*-motifs (Supplementary Table S1).[10,32,33] Type, position and sequence of the *cis*-element(s) identified in each promoter region of each TF are provided on the detailed page of each TF gene under '*cis*-motif prediction' keyword. Our database also provides the '*cis*-motif (stress responsive)', '*cis*-motif (hormone responsive)' and '*cis*-motif (PLACE)' search functions that enable the users to search conveniently for all types of analysed *cis*-motifs in promoter region of any TF and/or search for those TFs that contain the *cis*-motif(s) of interest. In combination with GO annotations, these data remarkably facilitate the systematic functional predictions of the tree TFs. For instance, with the help of the TreeTFDB, users can use the '*cis*-motif (stress responsive)' search function to identify TF genes harbouring known stress-responsive *cis*-motif(s) in their promoter regions. Next, they can screen the identified TFs using GO annotation provided for each TF on detailed annotation page (Fig. 3). Thus, we will be able to predict the putative stress-responsive TFs based on combinatory analyses of stress-responsive *cis*-motif(s) and the associated stress-responsive GO terms. Finally, expression profiling approach should be used to verify the predicted stress-responsive function prior to the launching of laborious *in planta* functional studies.

3.4. *Expression patterns of TF-encoding genes from poplar, grapevine and castor bean*

Gene expression data generated by high-throughput expression profiling experiments, such as microarray analysis, RNA-sequencing, are known to comprise one of the effective genomic resources for functional prediction. A combined analysis of gene expression, GO, *cis*-regulatory elements and coexpression neighbourhood will perfectly enable us to predict gene functions for unknown genes.[42,43] For instance, positive correlation has been found between the existence of *cis*-regulatory motifs and tissue-specific and/or stress-responsive expression patterns by either *in silico* or genetic inspection.[42] To make our database a comprehensive integrated database for functional characterization and selection of stress-responsive TFs, we have searched public resources for available expression datasets of the six tree crops with the aim to link the TF-encoding genes to their respective expression. Among the six tree crops, three species, namely poplar, grapevine and castor bean, were found to have large-scale expression datasets for further analyses. The datasets for poplar could be obtained from PlexDB and eFP browser websites while that for grapevine and castor bean from PlexDB and a RNA-sequencing project,[35] respectively. Hyperlinks linking to PlexDB and eFP browser were then built and supplied on TreeTFDB to provide access to the expression profiles of those poplar and grapevine TF-encoding genes for which data are available. The available expression datasets available at PlexDB, which were obtained through microarray analyses, have enabled us to examine the expression patterns of a significant number of polar and grape TF-encoding genes in response to their pathogens and symbionts as well as abiotic stressors, such as drought, high salinity and high temperature. On the other hand, hyperlinks linking to eFP browser will allow the users to access expression profiles of many poplar TF-encoding genes in several organs, such as roots, leaves and catkins and light-grown or



Summary

Accession	POPTR_0001s22630.1
Transcription factor	NAC
Length of sequence	801 nucleotides
Peptide length predicted	267 amino acids
Link to Gbrowse	TreeTFDB Gbrowse
Map detail image	

HMM search result

Transcription factor	NAC
Pfam HMM	PF02365 hmm
Pfam HMM score	185.1
Pfam HMM e value	1.5e-58
Hit from to	24. 157
Hit sequence	[illegible]

Sequence

CDS

```
>POPTR_0001s22630.1
ATGCTTATACAGCTCAGGCTGACAAGGGAACAAGACACAAGAACAAGAACAAGAACAAGAATGGAGGAACTCCCTCTAGG
TTACCGCTTCTACCCAACAGAAGAAGAGCTGATCTCCTTCTACCTGCATAATAAGCTGGGAGGGGGGAGGCAAGAACGTC
TACAACGTGTTATTCCAGACATCAGCATTTACGACATCGAACCATGGGATCTTCCAAAGCTTTCAGGAGAACTATGTCAG
GGAGACACCGAGCAGTGGTTCTTCTTTACACCCAGGCAAGAGAGAGAAGCGAGGGGAGGGAGACCTAACCGGACTACAGC
ATCCGGGTACTGGAAGGCCACAGGCTCTCCTGGTTATGTTTACTCGTCGGACAATCGTGTGATTGGATTGAAGAAAACCA
TGGTGTTCTATACGGGAAAAGCTCCAAGAGGAAGAAAAACCAAATGGAAGATGAATGAATACAGAGCCATTGAAGTTCAT
GAATCTTCTAGTAATGCTACTCCAAAGTTGAGGCATGAATTCAGCTTGTGTCGAGTCTATGTGGTATCAGGAAGCTTTCG
TGCATTTGATAGGCGGCCACTAGAAGCTGTGACAAGAGAGACGCAGCTTCTTGGAGATGGAGCTACAACATCTGCTCAGG
GCCCTGATCCAACTGTGGATATGACAAGCTCACCTGAAACGTCATACTCAGTGGAAGACCATGTTGATCACCCAGGGACT
GCAGCAAGTGCCAACTGGGGCTCAGTAGATGGTCTAGAACCACCTGCATGGGAATGGCCAGGACAACTGGATTGGCCCTG
A
```

Similarity search

Peptide

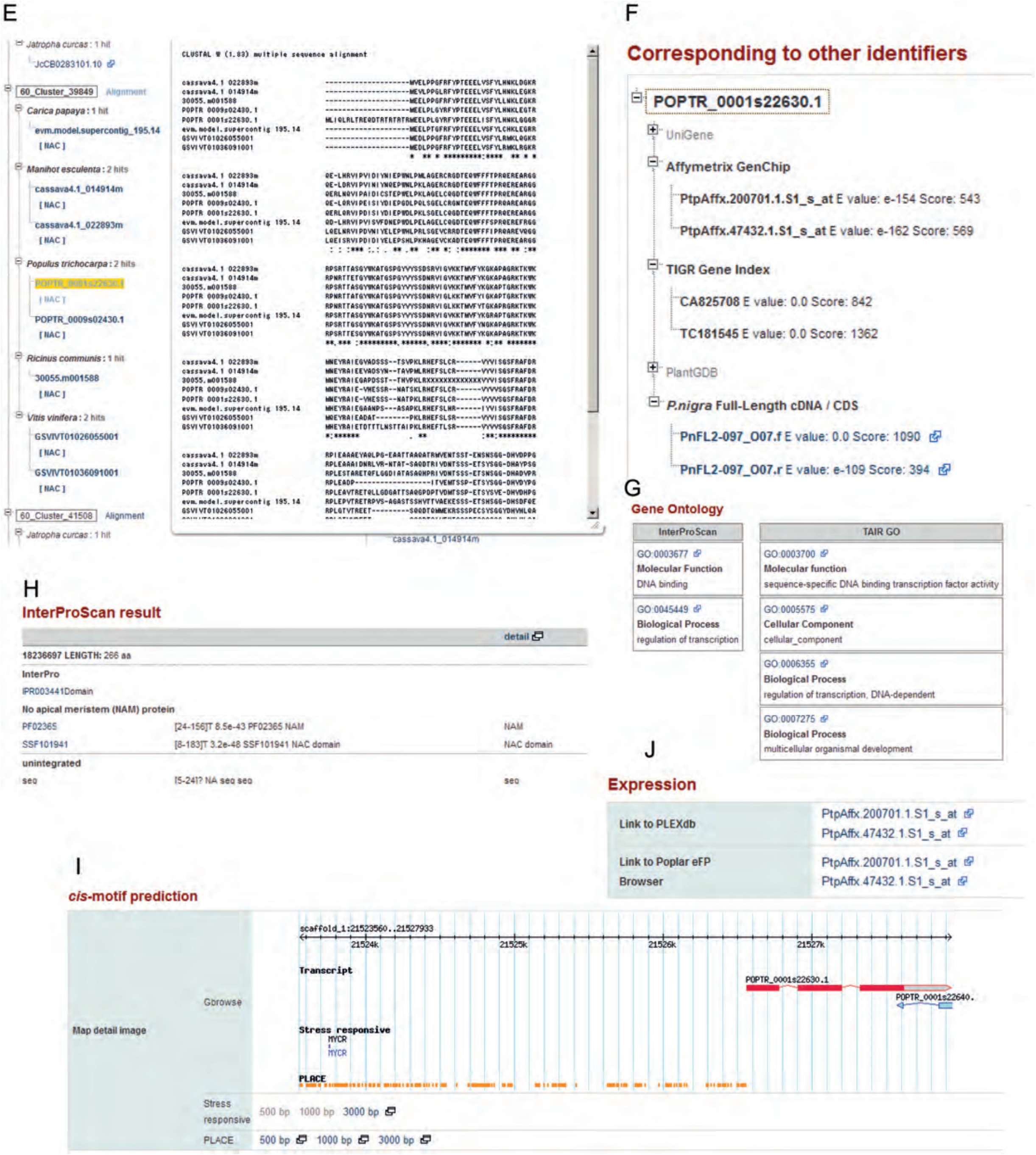
```
>POPTR_0001s22630.1
MLIQLRLTREQDTRTRTRTRMEELPLGYRFYPTEEELISFYLHNKLGGGRQERLQRVIPDISIYDIEPWDLPKLSGELCQ
GDTEQWFFFTPRQEREARGGRPNRTTASGYWKATGSPGYVYSSDNRVIGLKKTMVFYTGKAPRGRKTKWKMNEYRAIEVH
ESSSNATPKLRHEFSLCRVYVVSGSFRAFDRRPLEAVTRETQLLGDGATTSAQGPDPTVDMTSSPETSYSVEDHVDHPGT
AASANWGSVDGLEPPANEWPGQLDWP*
```

Similarity search

BLAST results

vs DB	Program	Description	Score	E value
NCBI nr	blastp	ref|XP_002298277.1| NAC domain protein, IPR003441 [Populus trichocarpa] ref|XP_002336754.1| NAC domain protein, IPR003441 [Populus trichocarpa] gb|EEE75054.1| NAC domain protein, IPR003441 [Populus trichocarpa] gb|EEE83082.1| NAC domain protein, IPR003441 [Populus trichocarpa]	513	e-143
UniProt	blastp	tr|B9GI04|B9GI04_POPTR NAC domain protein, IPR003441 OS=Populus trichocarpa GN=NAC106 PE=4 SV=1	513	e-143
Arabidopsis thaliana ⊟Relation to other TF DB entries DATF [NAC] RARTF [NAC/CUP-SHAPED COTYLEDON2] PlnTFDB [NAC] AtTFDB [NAC]	blastp	AT5G22380.1 | Symbols: anac090, NAC090 | NAC domain containing protein 90 | chr5:7408924-7410038 REVERSE LENGTH=235	276	1e-74
TIGR Rice ⊟Relation to other TF DB entries DRTF [NAC] Grassius [NAC] PlnTFDB [NAC]	blastp	LOC_Os05g37080.1|13105.m03848|protein No apical meristem protein, putative, expressed	214	1e-55
Oryza sativa ⊟Relation to RIKEN TFDB entries GramineaeTFDB [NAC]	blastp	Os05g0442700|AK063703	219	6e-58
Carica papaya ⊟Relation to TreeTFDB entries TreeTFDB [NAC]	blastp	evm.model.supercontig_195.14|PACid:16412265	329	5e-91
Jatropha curcas ⊟Relation to TreeTFDB entries TreeTFDB [NAC]	blastp	JcCA0309781.10 - phase: 0	250	9e-67
Manihot esculenta ⊟Relation to TreeTFDB entries TreeTFDB [NAC]	blastp	cassava4.1_022893m|PACid:17973512	356	7e-99
Ricinus communis ⊟Relation to TreeTFDB entries TreeTFDB [NAC]	blastp	30055.m001588|PACid:16816219	333	5e-92
Vitis vinifera ⊟Relation to TreeTFDB entries TreeTFDB [NAC]	blastp	GSVIVT01026055001|PACid:17833910	300	6e-82

Figure 3. Demonstration of a typical example of related annotations for a putative TF-encoding gene. (A) The summary provides basic information on each of the gene models annotated with gene structure. The figure for a gene structure is accessible via a hyperlink to a genome browser, which is browsed together with other sequences allocated onto the grass genome. (B) The HMM search result for the TF is displayed. (C) The sequences of cDNA and protein are provided and all clickable buttons navigate users to the blast search interface directory. (D) The similarity search results for each of the entries against NCBI nr, UniProt and gene models of *Arabidopsis* and other species, including tree species, with detailed search results and hyperlinks to the original data. (E) Resultant hierarchical clustering of homologous TFs can be browsed with multiple alignment of each cluster. (F) Information of other sequence identifiers for representative transcript sequence databases, including UniGene, TIGR Gene Index and PlantGDB as well as FL-cDNA ID and the probe ID of target sequences on the GeneChip, if available, are also accessible. (G) The GO terms assigned to each of the entries based on InterProScan and sequence similarity search against the annotated genes of *Arabidopsis* of TAIR10. (H) The domain structure predicted by InterProScan is provided. (I) The result of a *cis*-motif sequence pattern search of promoter regions for each gene is shown together with genomic gene structure. (J) Hyperlinks to PlexDB and/or eFP browser are provided for those TFs of poplar and grapevine for which expression data are available. For castor bean, expression data, if available, are provided directly on TreeTFDB under the 'Expression'.



Jatropha curcas : 1 hit
JcCB0283101.10
60_Cluster_39849 Alignment
Carica papaya : 1 hit
evm.model.supercontig_195.14
[NAC]
Manihot esculenta : 2 hits
cassava4.1_014914m
[NAC]
cassava4.1_022893m
[NAC]
Populus trichocarpa : 2 hits
POPTR_0001s22630.1
[NAC]
POPTR_0009s02430.1
[NAC]
Ricinus communis : 1 hit
30055.m001588
[NAC]
Vitis vinifera : 2 hits
GSVIVT01026055001
[NAC]
GSVIVT01036091001
[NAC]
60_Cluster_41508 Alignment
Jatropha curcas : 1 hit

```
CLUSTAL W (1.83) multiple sequence alignment

cassava4.1 022893m              --------------------MVELPPGFRFYPTEEELVSFYLHNKLDGKR
cassava4.1 014914m              --------------------MEVLPPGLRFYPTEEELVSFYLHNKLEGKR
30055.m001588                   --------------------MEELPPGFRFYPTEEELVSFYLLNKLEGKR
POPTR 0009s02430.1              --------------------MEELPLGYRFYPTEEELVSFYLHNKLEGTR
POPTR 0001s22630.1              MLIQLRLTREQDTRTRTRTRMEELPLGYRFYPTEEELISFYLHNKLGGGR
evm.model.supercontig 195.14    --------------------MEELPTGFRFYPTEEELVSFYLCHKLEGRR
GSVIVT01026055001               --------------------MEDLPPGFRFYPTEEELVSFYLRMKLQGKR
GSVIVT01036091001               --------------------MEDLPPGFRFYPTEEELVSFYLRMKLRGKR
                                                    *  ** * *********:****  ** * *

cassava4.1 022893m              QE-LHRVIPVIDIYNIEPWNLPMLAGERCRGDTEQWFFFTPRQEREARGG
cassava4.1 014914m              QE-LDRVIPVINIYNQEPWNLPKLAGERCRGDTEQWFFFTPRQEREARGG
30055.m001588                   QERLNQVIPAIDICSTEPWELPKLAGELCQGDTEQWFFFTPRQEREARGG
POPTR 0009s02430.1              QE-LQRVIPEISIYDIEPGDLPQLSGELCRGNTEQWFFFTPRQAREARGG
POPTR 0001s22630.1              QERLQRVIPDISIYDIEPWDLPKLSGELCQGDTEQWFFFTPRQEREARGG
evm.model.supercontig 195.14    QD-LHRVIPVISVFDNEPWDLPELAGELCQGDTEQWFFFSPRQEREFRGG
GSVIVT01026055001               LQELNRVIPDVNIYELEPWHLPRLSGEVCRRDTEQWFFFIPRQAREVQGG
GSVIVT01036091001               LQEISRVIPDIDIYELEPSHLPKHAGEVCKADTEQWFFFTPRQEREARGG
                                 : : :*** :.: . ** .**  :** *: :******* *** ** :**

cassava4.1 022893m              RPSRTTASGYWKATGSPGYVYSSDSRVIGVKKTMVFYKGKAPAGRKTKWK
cassava4.1 014914m              RPNRTTETGYWKATGSPGYVYSSDNRVIGVKKTMVFYKGKAPAGRKTKWK
30055.m001588                   RPSRTTASGYWKATGSPGYVYSSDNRVIGVKKTMVFYKGKAPTGRKTKWK
POPTR 0009s02430.1              RPNRTTATGYWKATGSPGYVYSSDNRVIGLKKTMVFYIGKAPTGRKTKWK
POPTR 0001s22630.1              RPNRTTASGYWKATGSPGYVYSSDNRVIGLKKTMVFYTGKAPRGRKTKWK
evm.model.supercontig 195.14    RPSRTTASGYWKATGSPSYVYSSDNRVIGVKKTMVFYKGRAPTGRKTKWK
GSVIVT01026055001               RPSRTTGSGYWKATGSPSYVYSSDNRVIGVKKTMVFYNGKAPAGRKTKWK
GSVIVT01036091001               RPSRTTESGYWKATGSPSYVYSSDNRVIGVKKTMVFYKGKAPAGRKTKWK
                                **.*** :*********.******.****:******* *:** *******

cassava4.1 022893m              MNEYRAIEGVADSSS--TSVPKLRHEFSLCR------VYVISGSFRAFDR
cassava4.1 014914m              MNEYRAIEEVADSYN--TAVPMLRHEFSLCR------VYVISGSFRAFDR
30055.m001588                   MNEYRAIEGAPDSST--THVPKLRXXXXXXXXXXXXXVYVISGSFRAFDR
POPTR 0009s02430.1              MNEYRAIE-VHESSR--NATSKLRHEFSLCR------VYVVSGSFRAFDR
POPTR 0001s22630.1              MNEYRAIE-VHESSS--NATPKLRHEFSLCR------VYVVSGSFRAFDR
evm.model.supercontig 195.14    MNEYRAIEGAANPS---ASAPKLRHEFSLHR------IYVISGSFRAFDR
GSVIVT01026055001               MQEYRAIEADAT---------PKLRHEFSLSR------VYVISGSFRAFDR
GSVIVT01036091001               MNEYRAIETDTTTLNSTTAIPKLRHEFTLSR------VYVVSGSFRAFDR
                                *:******              . **              :**:*********

cassava4.1 022893m              RPIEAAAEYAQLPG-EAATTAAQATRMVENTSST-ENSNSGG-DHVDPPG
cassava4.1 014914m              RPLEAAAIDNRLVR-NTAT-SAQDTRIVDNTSSS-ETSYSGG-DHAYPSG
30055.m001588                   RPLESTARETQFLGGDIATASAQHPRIVDNTSSS-ETSNSGG-DHADVPR
POPTR 0009s02430.1              RPLEADP-------------------ITVEMTSSP-ETSYSGG-DHVDYPG
POPTR 0001s22630.1              RPLEAVTRETQLLGDGATTSAQGPDPTVDMTSSP-ETSYSVE-DHVDHPG
evm.model.supercontig 195.14    RPLEPVTRETRPVS-AGASTSSHVTTVAEKESSS-ETSHSGG-DHSDFQE
GSVIVT01026055001               RPLGTVTREET---------SQQDTQMMEKRSSSPECSYSGGYDHVHLQA
```

cassava4.1_014914m

Corresponding to other identifiers

⊟ POPTR_0001s22630.1

⊞ UniGene

⊟ Affymetrix GenChip

PtpAffx.200701.1.S1_s_at E value: e-154 Score: 543

PtpAffx.47432.1.S1_s_at E value: e-162 Score: 569

⊟ TIGR Gene Index

CA825708 E value: 0.0 Score: 842

TC181545 E value: 0.0 Score: 1362

⊞ PlantGDB

⊟ *P.nigra* Full-Length cDNA / CDS

PnFL2-097_O07.f E value: 0.0 Score: 1090

PnFL2-097_O07.r E value: e-109 Score: 394

Gene Ontology

InterProScan	TAIR GO
GO:0003677 Molecular Function DNA binding	GO:0003700 Molecular function sequence-specific DNA binding transcription factor activity
GO:0045449 Biological Process regulation of transcription	GO:0005575 Cellular Component cellular_component
	GO:0006355 Biological Process regulation of transcription, DNA-dependent
	GO:0007275 Biological Process multicellular organismal development

InterProScan result

detail

18236697 LENGTH: 266 aa

InterPro

IPR003441Domain

No apical meristem (NAM) protein

PF02365	[24-156]T 8.5e-43 PF02365 NAM	NAM
SSF101941	[8-183]T 3.2e-48 SSF101941 NAC domain	NAC domain

unintegrated

seq	[5-24]? NA seq seq	seq

Expression

Link to PLEXdb	PtpAffx.200701.1.S1_s_at PtpAffx.47432.1.S1_s_at
Link to Poplar eFP Browser	PtpAffx.200701.1.S1_s_at PtpAffx.47432.1.S1_s_at

cis-motif prediction

Map detail image	Gbrowse	
	Stress responsive	500 bp 1000 bp 3000 bp
	PLACE	500 bp 1000 bp 3000 bp

Figure 3. *Continued*

dark-grown seedlings. As for the TF-encoding genes in castor bean, their expression data obtained from analysis of RNA-seq transcriptomes of five tissues, namely endosperm at stages II/III (endosperm free-nuclear stage) and stages V/VI (onset of cellular endosperm development), expanding true leaves, germinating seeds (cotyledon) and developing male flowers are provided in the detailed page of TreeTFDB. Those TF-encoding genes of poplar, grapevine and castor bean, for which expression data are available, are indicated by [Expression] string below the gene ID (Fig. 1B).

3.5. *Web interface of TreeTFDB*

All the data obtained through this study were integrated to create TreeTFDB (http://treetfdb.bmep.riken.jp/index.pl) with the aim to provide a convenient environment for comparative and functional genomics of the tree crop TFs at both individual and family levels. The web-based user interface of TreeTFDB is

Table 2. Comparison of TreeTFDB, PlnTFDB and PlantTFDB

Organisms	Contents	TreeTFDB	PlnTFDB (3.0)	PlantTFDB (v2.0)
C. papaya	Hosted	+	+	+
	Expression	NA	NA	NA
	Promoter sequence	+	−	−
	Internal genome browser	+	−	−
	Cis-element	+	−	−
J. curcas	Hosted	+	−	−
	Expression	NA	NA	NA
	Promoter sequence	+	−	−
	Internal genome browser	+	−	−
	Cis-element	+	−	−
M. esculenta	Hosted	+	−	+
	Expression	NA	NA	NA
	Promoter sequence	+	−	−
	Internal genome browser	+	−	−
	Cis-element	+	−	−
P. trichocarpa	Hosted	+	+	−
	Expression	PLEXdb (GeneChip) eFP browser (GeneChip)	NCBI UniGene (EST profile viewer)	−
	Promoter sequence	+	−	−
	Internal genome browser	+	−	−
	Cis-element	+	−	−
R. communis	Hosted	+	−	+
	Expression	RNA-seq	−	−
	Promoter sequence	+	−	−
	Internal genome browser	+	−	−
	Cis-element	+	−	−
V. vinifera	Hosted	+	+	+
	Expression	PLEXdb (GeneChip)	NCBI UniGene (EST profile viewer)	−
	Promoter sequence	+	−	−
	Internal genome browser	+	−	−
	Cis-element	+	−	−

+, Available; −, Not available; NA, data are not available for analysis.

illustrated in Fig. 1. The first page of TreeTFDB provides general information about TreeTFDB along with ‘TF search’, ‘Similarity search’, ‘Genome browser’, ‘Download’, ‘Help’ and ‘Quick search’ functions (Fig. 1A). The TF search interface for each species provides 7 types of search queries for names of TF families, keywords, gene identifiers, identifiers of domains supported by InterProScan, GO terms and all available *cis*-motifs documented in the PLACE database (http://www.dna.affrc.go.jp/PLACE/) as well as 12 known abiotic stress- and/or hormone-responsive *cis*-motifs published previously[10,33] (for details, see the TreeTFDB Help page). By clicking on a species and then a TF family of interest, such as poplar and NAM-no apical meristem, ATAF-*Arabidopsis* transcription activation factor and CUC-cup-shaped cotyledon TF family, users will be navigated to the page listing all the identified members of the TF family (Fig. 1B). Users can then click on a gene ID and conveniently access the detailed information on gene annotations, including gene structure, cDNA and protein sequences, domain structure predicted by InterProScan, domain alignments, promoter regions, predicted *cis*-regulatory motifs in −500, −1000 and −3000 bp promoter regions of the TF gene as well as clusters of homologous proteins within families and GO terms derived from GO annotation using comparative analysis with their *Arabidopsis* counterparts (Fig. 3). Additionally, the data supplied are available not only for viewing through the web interface but also for immediate downloading via the ‘Download’ search key (Fig. 1A). In terms of similarity (blastp) with TFs from trees or non-tree plants, such as *Arabidopsis*, maize and rice, TreeTFDB supplies not only the results of blastp and clustered protein families but also links to either species-specific TF databases, such as DATF, RARTF, AtTFDB, DRTF or integrative TF database such as PlnTFDB, Grassius and GramineaeTFDB (Fig. 3D and E). This feature provides a unique option for comparative studies of TF repertoires not only among tree species but also between tree and non-tree plants (Fig. 3D and E). For poplar and cassava, hyperlinks linking their TF-encoding genes to available FL-cDNA clones reported in public repositories are also provided in detailed pages under 'Corresponding to other identifiers' (Fig. 3F). Furthermore, we supplied on TreeTFDB hyperlinks linking directly the TFs of poplar and grapevine to their expression patterns documented in public and freely available resources (Fig. 3J). The expression profiles of the castor bean TF-encoding genes, for which data are available, in five different tissues are also provided on the detailed page of the TreeTFDB for immediate viewing. These expression data together with information of *cis*-motif analyses, GO annotations and sequence similarities inferred from comparative analyses of the tree crops can facilitate the systematic functional predictions of identified TFs. Finally, in comparison with PlnTFDB and PlantTFDB, which provide three and five TF repertoires out of six species analysed by TreeTFDB, the TreeTFDB has several major updates, allowing the users to access a number of new information and features that are not available on PlnTFDB and PlantTFDB even for the TF repertoires provided on these databases. Table 2 summarized the major differences among these three databases.

3.6. *Conclusions*

The interactions between TFs and *cis*-regulatory DNA sequences control gene expression, constituting the essential functional linkages of gene regulatory networks. Knowledge gained from *cis*-elements, GO annotation and expression data will enable effective functional prediction of the identified TFs. With its friendly user interface, TreeTFDB will, therefore, meet the broad demands of researchers who strive to perform research on tree TFs with the goal of gaining greater understanding of their regulatory roles in different signalling pathways, underlying plant development, differentiation and environmental responses, ultimately leading to development of varieties with improved performance.

Supplementary data: Supplementary Data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was partially supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan [Grant-in-Aid for Scientific Research on Innovative Areas (23119524) to K.M.]. Research in L.-S.P.T.'s lab was supported by a grant (No. AP24-1-0076) from RIKEN Strategic Research Program for R & D.

References

1. Riechmann, J.L., Heard, J., Martin, G., et al. 2000, Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes, *Science*, **290**, 2105–10.
2. Tran, L.S., Nishiyama, R., Yamaguchi-Shinozaki, K. and Shinozaki, K. 2010, Potential utilization of NAC transcription factors to enhance abiotic stress tolerance in plants by biotechnological approach, *GM Crops*, **1**, 32–9.
3. Hadiarto, T. and Tran, L.S. 2011, Progress studies of drought-responsive genes in rice, *Plant Cell Rep.*, **30**, 297–310.
4. Thao, N.P. and Tran, L.S. 2011, Potentials toward genetic engineering of drought-tolerant soybean, *Crit. Rev. Biotechnol.*, **32**, 349–62.
5. Jogaiah, S., Ramsandra Govind, S. and Tran, L.S. 2012, System biology-based approaches towards understanding drought tolerance in food crops, *Crit. Rev. Biotechnol.*, doi:10.3109/07388551.2012.659174.
6. Moreno-Risueno, M.A., Van Norman, J.M. and Benfey, P.N. 2012, Transcriptional switches direct plant organ formation and patterning, *Curr. Top. Dev. Biol.*, **98**, 229–57.
7. Tran, L.S., Nakashima, K., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2007, Plant gene networks in osmotic stress response: from genes to regulatory networks, *Methods Enzymol.*, **428**, 109–28.
8. Qin, F., Sakuma, Y., Tran, L.S., et al. 2008, Arabidopsis DREB2A-interacting proteins function as RING E3 ligases and negatively regulate plant drought stress-responsive gene expression, *Plant Cell*, **20**, 1693–707.
9. Choi, N.M. and Boss, J.M. 2012, Multiple histone methyl and acetyltransferase complex components bind the HLA-DRA gene, *PLoS One*, **7**, e37554.
10. Yamaguchi-Shinozaki, K. and Shinozaki, K. 2005, Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters, *Trends Plant Sci.*, **10**, 88–94.
11. Tran, L.S. and Mochida, K. 2010, Identification and prediction of abiotic stress responsive transcription factors involved in abiotic stress signaling in soybean, *Plant Signal Behav.*, **5**, 255–7.
12. Tran, L.S. and Mochida, K. 2010, A platform for functional prediction and comparative analyses of transcription factors of legumes and beyond, *Plant Signal Behav.*, **5**, 550–2.
13. Tran, L.S. and Mochida, K. 2010, Functional genomics of soybean for improvement of productivity in adverse conditions, *Funct. Integr. Genomics*, **10**, 447–62.
14. Mochida, K. and Shinozaki, K. 2010, Genomics and bioinformatics resources for crop improvement, *Plant Cell Physiol.*, **51**, 497–523.

15. Iida, K., Seki, M., Sakurai, T., et al. 2005, RARTF: database and tools for complete sets of Arabidopsis transcription factors, *DNA Res.*, **12**, 247–56.
16. Gao, G., Zhong, Y., Guo, A., et al. 2006, DRTF: a database of rice transcription factors, *Bioinformatics*, **22**, 1286–7.
17. Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K. and Tran, L.S. 2009, In silico analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean, *DNA Res.*, **16**, 353–69.
18. Wang, Z., Libault, M., Joshi, T., et al. 2010, SoyDB: a knowledge database of soybean transcription factors, *BMC Plant Biol.*, **10**, 14.
19. Rushton, P.J., Bokowiec, M.T., Laudeman, T.W., Brannock, J.F., Chen, X. and Timko, M.P. 2008, TOBFAC: the database of tobacco transcription factors, *BMC Bioinformatics*, **9**, 53.
20. Fredslund, J. 2008, DATFAP: a database of primers and homology alignments for transcription factors from 13 plant species, *BMC Genomics*, **9**, 140.
21. Yilmaz, A., Nishiyama, M.Y. Jr, Fuentes, B.G., et al. 2009, GRASSIUS: a platform for comparative regulatory genomics across the grasses, *Plant Physiol.*, **149**, 171–80.
22. Perez-Rodriguez, P., Riano-Pachon, D.M., Correa, L.G., Rensing, S.A., Kersten, B. and Mueller-Roeber, B. 2010, PlnTFDB: updated content and new features of the plant transcription factor database, *Nucleic Acids Res.*, **38**, D822–7.
23. Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K. and Tran, L.S. 2010, LegumeTFDB: an integrative database of *Glycine max*, *Lotus japonicus* and *Medicago truncatula* transcription factors, *Bioinformatics*, **26**, 290–1.
24. Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K. and Tran, L.S. 2011, In silico analysis of transcription factor repertoires and prediction of stress-responsive transcription factors from six major gramineae plants, *DNA Res.*, **18**, 321–32.
25. Zhang, H., Jin, J., Tang, L., et al. 2011, PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database, *Nucleic Acids Res.*, **39**, D1114–7.
26. Mochida, K. and Shinozaki, K. 2011, Advances in omics and bioinformatics tools for systems analyses of plant functions, *Plant Cell Physiol.*, **52**, 2017–38.
27. Sato, S., Hirakawa, H., Isobe, S., et al. 2011, Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L, *DNA Res.*, **18**, 65–76.
28. Mochida, K., Yoshida, T., Sakurai, T., Ogihara, Y. and Shinozaki, K. 2009, TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics, *Plant Physiol.*, **150**, 1135–46.
29. Sammut, S.J., Finn, R.D. and Bateman, A. 2008, Pfam 10 years on: 10,000 families and still growing, *Brief Bioinform.*, **9**, 210–9.
30. Zhu, Q.H., Guo, A.Y., Gao, G., et al. 2007, DPTF: a database of poplar transcription factors, *Bioinformatics*, **23**, 1307–8.
31. Swarbreck, D., Wilks, C., Lamesch, P., et al. 2008, The Arabidopsis information resource (TAIR): gene structure and function annotation, *Nucleic Acids Res.*, **36**, D1009–14.
32. 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.
33. Wang, Z.Y., Bai, M.Y., Oh, E. and Zhu, J.Y. 2012, Brassinosteroid signaling network and regulation of photomorphogenesis, *Ann. Rev. Genet.*, **46**, 701–24.
34. Donlin, M.J. 2009, Using the Generic Genome Browser (GBrowse), *Curr. Protoc. Bioinform.*, **28**, 9.9.1–25.
35. Brown, A.P., Kroon, J.T., Swarbreck, D., et al. 2012, Tissue-specific whole transcriptome sequencing in castor, directed at understanding triacylglycerol lipid biosynthetic pathways, *PLoS One*, **7**, e30100.
36. Lang, D., Weiche, B., Timmerhaus, G., et al. 2010, Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity, *Genome Biol. Evol.*, **2**, 488–503.
37. Walther, D., Brunnemann, R. and Selbig, J. 2007, The regulatory code for transcriptional response diversity and its relation to genome structural properties in *A. thaliana*, *PLoS Genet.*, **3**, e11.
38. Zhang, W., Ruan, J., Ho, T.H., You, Y., Yu, T. and Quatrano, R.S. 2005, Cis-regulatory element based targeted gene finding: genome-wide identification of abscisic acid- and abiotic stress-responsive genes in *Arabidopsis thaliana*, *Bioinformatics*, **21**, 3074–81.
39. Kim, D.W., Lee, S.H., Choi, S.B., et al. 2006, Functional conservation of a root hair cell-specific cis-element in angiosperms with different root hair distribution patterns, *Plant Cell*, **18**, 2958–70.
40. Won, S.K., Lee, Y.J., Lee, H.Y., Heo, Y.K., Cho, M. and Cho, H.T. 2009, Cis-element- and transcriptome-based screening of root hair-specific genes and their functional characterization in Arabidopsis, *Plant Physiol.*, **150**, 1459–73.
41. Zou, C., Sun, K., Mackaluso, J.D., et al. 2011, Cis-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. USA*, **108**, 14992–7.
42. Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L. and Van de Peer, Y. 2009, Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks, *Plant Physiol.*, **150**, 535–46.
43. Ma, Y., Qin, F. and Tran, L.S. 2012, Contribution of genomics to gene discovery in plant abiotic stress responses, *Mol. Plant*, **5**, 1176–8.