PlnTFDB: updated content and new features of the plant transcription factor database

Paulino Pérez-Rodríguez1,2, Diego Mauricio Riaño-Pachón1,3,*, Luiz Gustavo Guedes Corrêa1,4, Stefan A. Rensing5, Birgit Kersten1,3 and Bernd Mueller-Roeber1,4

1Department of Molecular Biology, Institute of Biochemistry and Biology, GoFORSYS, University of Potsdam, Karl-Liebknecht-Str. 24-25, Haus 20, 14476 Potsdam-Golm, Germany, 2Colegio de Postgraduados, Km. 36.5 Carretera México, Texcoco, Montecillo, Estado de México. C.P. 56230, Mexico, 3GabiPD Team, Bioinformatics Group, Max Planck Institute of Molecular Plant Physiology, 4Cooperative Research Group, Max Planck Institute of Molecular Plant Physiology, Wissenschaftspark Golm, Am Mühlenberg 1, 14476 Potsdam - Golm and 5FRISYS, Faculty of Biology, University of Freiburg, Hauptstr. 1, D-79104 Freiburg, Germany

Received July 10, 2009; Accepted September 13, 2009

ABSTRACT

The Plant Transcription Factor Database (PlnTFDB; http://plntfdb.bio.uni-potsdam.de/v3.0/) is an integrative database that provides putatively complete sets of transcription factors (TFs) and other transcriptional regulators (TRs) in plant species (sensu lato) whose genomes have been completely sequenced and annotated. The complete sets of 84 families of TFs and TRs from 19 species ranging from unicellular red and green algae to angiosperms are included in PlnTFDB, representing >1.6 billion years of evolution of gene regulatory networks. For each gene family, a basic description is provided that is complemented by literature references, and multiple sequence alignments of protein domains. TF or TR gene entries include information of expressed sequence tags, 3D protein structures of homologous proteins, domain architecture and cross-links to other computational resources online. Moreover, the different species in PlnTFDB are linked to each other by means of orthologous genes facilitating cross-species comparisons.

DATA SOURCES, ANALYSES AND IMPLEMENTATION

Species and proteomes covered

In order to identify putatively complete sets of TFs and TRs, we applied our previously established analysis pipeline to the proteomes of species whose genomes have been completely sequenced and annotated (4). The PlnTFDB v3.0 covers 19 different plant species ranging from unicellular red and green algae to angiosperms,

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

*To whom correspondence should be addressed. Tel: +49-(0)331-567-8752; Fax: +49-(0)331-567-89-8750; Email: riano@mpimp-golm.mpg.de

© The Author(s) 2009. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Identification of protein domains and new domain models

The identification of TFs and TRs and their classification into families exploits the presence of protein domains and their combination within proteins (4). To generate the current release of PlnTFDB, domains were identified using the Pfam protein families database v23.0 (21) and the software package HMMER v2.3.2 (http://hmmer.janelia.org/). Domain hits with a score higher than or equal to the gathering cut-off (–cut_ga) defined for each hidden Markov model (HMM) were kept for further analyses.

For some families, there is no domain represented in the Pfam database; in such cases we developed profile HMMs based on sequence alignments of the respective domains. For the current version of PlnTFDB, we established HMMs for the characteristic domains of the families NOZZLE and VARL. An HMM for the NOZZLE family is available in the Pfam database; however, this model only recovers members from the Brassicaceae family (e.g. Arabidopsis sp.). Hence we used the Arabidopsis thaliana sequences to perform a PSI-BLAST search against the non-redundant protein database at NCBI (http://www.ncbi.nlm.nih.gov/). This allowed us building a multiple sequence alignment and HMM of NOZZLE proteins from several angiosperms, i.e. A. thaliana, Brassica juncea, Medicago truncatula and Vitis vinifera.

The HMM for the VARL family was built by using the alignment reported in Duncan et al. (22), with sequences from Chlamydomonas reinhardtii and Volvox carteri. The alignments used to create the new HMMs are available through the database web interface.

After building these HMMs, a score threshold had to be defined, beyond which the hits are considered significant. To this end, we run an HMM search with the newly defined, beyond which the hits are considered significant.

| Groups                  | Species                  | Source            | Annotation version | Reference | Total number of proteinsa | Genome size (Mb) | Number of families | Number of classified proteinsa |
|-------------------------|--------------------------|-------------------|-------------------|-----------|--------------------------|-----------------|-------------------|-----------------------------|
| Red algae (Rhodophytes) | Cyanidioschyzon merolae  | 1 20070710        | (8)               |           | 5008                     | 16.52            | 34                | 147                         |
|                        | Galderia sulpharioria    | 9                 |                   |           | 6604                     | 10               | 37                | 201                         |
| Green algae (Prasinophytes) | Micromonas pusilla CCMP1545 | 2                 |                   |           | 10455                    | 15               | 49                | 289                         |
|                        | Micromonas sp. RCC299    | 2                 |                   |           | 10160                    | 15               | 49                | 326                         |
|                        | Ostreococcus tauri       | 2                 |                   |           | 7812                     | 12.56            | 47                | 216                         |
|                        | Ostreococcus lucinarius  | 2                 |                   |           | 7651                     | 13.204           | 46                | 236                         |
| Green algae (Chlorophytes) | Chlamydomonas reinhardtii | 2                 |                   |           | 16460                    | 121              | 52                | 346                         |
|                        | Chlorella sp. NC64A      | 2                 |                   |           | 9762                     | 40               | 48                | 304                         |
|                        | Coccomyxa sp. C-169      | 2                 |                   |           | 10174                    | 120              | 47                | 261                         |
| Bryophyte (Bryopsida)   | Physcomitrella patens    | 2                 |                   |           | 35724                    | 480              | 72                | 1295                        |
| Spike-moss (Lycophyophyte) | Selaginella moellendorfii | 2                 |                   |           | 22138                    | 100              | 74                | 896                         |
| Angiosperms (Eudicots)  | Oryza sativa subsp. indica | 3                 | 20050118          | (14)      | 49643                    | 420              | 79                | 2393                        |
|                        | Oryza sativa subsp. japonica | 4                 |                   | (15)      | 63306                    | 420              | 79                | 2722                        |
|                        | Sorghum bicolor         | 2                 |                   | (16)      | 35682                    | 730              | 78                | 2231                        |
|                        | Zea mays                 | 5                 | 3b.50             |           | 55810                    | 2400             | 79                | 3608                        |
| Angiosperms (Monocots)  | Carica papaya           | 7                 |                   |           | 24852                    | 372              | 81                | 1480                        |
|                        | Arabidopsis lyrata      | 2                 |                   |           | 32324                    | 206.7            | 81                | 2162                        |
|                        | Arabidopsis thaliana    | 6                 | 8                 | (18)      | 30707                    | 125              | 81                | 2451                        |
|                        | Populus trichocarpa     | 2                 | 1.1               | (19)      | 45009                    | 485              | 81                | 2901                        |
|                        | Vitis vinifera          | 8                 | 1                 | (20)      | 30342                    | 500              | 80                | 1725                        |

(1) CME GP, Cyanidioschyzon merolae Genome Project, http://merolae.biol.s.u-tokyo.ac.jp/; (2) JGI/DOE, Joint Genome Institute/Department of Energy, http://www.jgi.doe.gov/; (3) BGI, Beijing Genomics Institute, http://www.genomics.org.cn/; (4) TIGR, The Institute for Genomic Research, http://www.tigr.org/; (5) MaizeSequence.org, http://www.maizesequence.org; (6) TAIR, The Arabidopsis Information Resource, http://www.arabidopsis.org/; (7) The Hawaii Papaya Genome Project, http://asgpb.mhpcc.hawaii.edu/papaya/; (8) Genoscope, Centre National de Séquençage http://www.genoscope.cns.fr/spip/Vitis-vinifera-e.html; (9) Data communicated by Prof. Dr Andreas Weber, University of Duesseldorf, Germany.

aNumber of non-redundant proteins.
with the lowest
interest contains domains characteristic of more than one
meta-rule it is assigned to C3H. (ii) When the protein of
(TF) and SWI/SNF-BAF60b (TR), but according to this
protein AT3G51120.1 could be assigned to families C3H
family, we assigned it to the TF family, e.g.
harbours domains characteristic of a TF family and a TR
single TF or TR family only. The current sets of ‘required’
and ‘forbidden’ domains of each individual family are
lishing a mutually exclusive classification system ensuring
‘forbidden’ domains. The forbidden domains allow estab-
lishing a mutually exclusive classification system ensuring
that each individual protein is classified as a member of a
single TF or TR family only. The current sets of ‘required’
and ‘forbidden’ domains of each individual family are
listed in Supplementary Data, Appendix 1. We included
two meta-rules in our classification scheme: (i) if a protein
harbours domains characteristic of a TF family and a TR
family, we assigned it to the TF family, e.g. *A. thaliana*
protein AT3G51120.1 could be assigned to families C3H
(TF) and SWI/SNF-BAF60b (TR), but according to this
meta-rule it is assigned to C3H. (ii) When the protein of
interest contains domains characteristic of more than one
TF family or more than one TR family, it was assigned
to the family to which its characteristic domains matched
with the lowest *e*-value. For example, protein 425147
from *Selaginella moellendorfii* could be classified as
C2H2 (TF, *e*-value 7.3e-3) or RWP-RK (TF, *e*-value
6.1e-11), according to the meta-rule it was assigned to
the RWP-RK family.

Database interface and availability
The information about the different regulatory proteins
and their classification into families, as well as sequence
alignments, 3D structures, literature references and
links to other databases are stored in a relational
database, powered by MySQL (http://www.mysql.com;
database schema in Supplementary Data, Appendix 2).

The interface of the database to the World Wide Web
(WWW) was developed by using PHP, JavaScript and
Java applets (Jmol, http://www.jmol.org/; and Jalview,
http://www.jalview.org/) following HTML 4.01 and CCS
v2.1 W3 standards to ensure browser interoperability.

PlnTFDB can be queried using keywords or sequences
(using blastp or blastx), and it is freely accessible through
the WWW via http://plntfdb.bio.uni-potsdam.de/v3.0/
using any modern web browser. The Java Runtime
Environment (JRE) 1.6.0.12 or newer is required in
order to visualize domain alignments and protein 3D
structures.

3D PROTEIN STRUCTURES, EXPRESSED
SEQUENCE TAGS AND ORTHOLOGUES
To widen the information provided for each TF and TR in
PlnTFDB, we have performed similarity-based searches
against the database of sequences with known protein
tertiary structures available from the Protein Data Bank
(PDB) and the expressed sequence tag (EST) databases
available from GenBank. To identify related ESTs, we
used BLAST as search engine, keeping as significant all
hits with an *e*-value ≤ 10^{-10} and an alignment identity
of ≥50% over a length of ≥80 amino acids. For the detection
of homologous 3D protein structures, we used the package
hhsearch (http://toolkit.tuebingen.mpg.de/hhpred) that
employs HMM—HMM comparisons to detect remote
homologues. Hits were considered significant if the
probability of the target being a TP was > 98%. The 3D
structures of proteins similar to entries in PlnTFDB can be
visualized with the Jmol applet (Figure 2), and links are
provided to the PDB web site.

The genomes of some species covered by PlnTFDB, e.g.
*A. thaliana* and *Oryza sativa* ssp. *japonica*, are relatively
well annotated with respect to the biological functions of
the proteins they encode, whereas genomes of others,
including *C. reinhardtii*, are still in a preliminary status
of annotation of biological functions. As orthologous
genes often have the same function in different species
(23), we have used InParanoid (24) to detect clusters of
orthologous genes between pairs of species in PlnTFDB.
This will ease the transfer of functional information and
provide effective cross-references among the species in
PlnTFDB.

QUALITY CONTROL
To evaluate the quality of the putatively complete sets of
TFs and TRs reported in PlnTFDB, we compared our
predictions to published datasets on detailed single-
family phylogenetic studies, and defined the published
analyses as gold standards. We calculated the sensitivity
and the positive predicted value (PPV) as described before
(4). The results of this evaluation are shown in Table 2.
In all cases, both measures are >80%, and for most
families the sensitivity and PPV values are >90%
(shown in bold face in Table 2), evidencing low rates of
false negatives (FNs) and positives (FPs).
MAIN RESULTS

In the current version of PlnTFDB (v3.0), we present a total of 84 different TF and TR families that occur in 19 different plant species and encompass 26,184 distinct proteins. A summary of the content of the database is shown in Table 1; there is a tendency that the number of TFs and TRs per family, as well as the number of families, increases along with the organismic complexity. Correlation analyses support this observation (Supplementary Data, Appendix 3).
The wide spectrum of gene families covered by PlnTFDB has already been exploited by researchers, e.g. for use in genome annotations (12,40,41), functional studies of TFs and TRs (42,43) and detailed phylogenetic studies of TF families in the whole plant lineage (28), among others.

OUTLOOK

As the cost of genome sequencing continues to decrease, the number of newly sequenced genomes will increase dramatically in the near future. The computational analysis pipeline behind PlnTFDB will be applied to these new genomes, increasing even further its wide phylogenetic coverage. We envisage that PlnTFDB will increasingly be exploited in genome annotation projects as a primary repository serving the identification of transcription regulatory proteins.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to express our gratitude to the people and institutions working on the sequencing and annotation of the plant genomes analyzed in this study. We are particularly thankful to Andreas Weber and Detlef Weigel who allowed us to explore plant genome data not published yet.

FUNDING

Bundesministerium fuer Bildung und Forschung, Germany (GABI-FUTURE grant 0315046, GoFORSYS grant 0313924 and FRISYS grant 0313921); Subdireccion de Investigacion: Linea 15, Colegio de Postgraduados, Mexico; Deutscher Akademischer Austauschdienst (DAAD). Funding for open access charge: GoFORSYS.

Conflict of interest statement. None declared.

REFERENCES

1. Kummerfeld,S.K. and Teichmann,S.A. (2006) DBD: a transcription factor prediction database. Nucleic Acids Res., 34, D74–D81.
2. Guo,A.-Y., Chen,X., Gao,G., Zhang,H., Zhu,Q.-H., Liu,X.-C., Zhong,Y.-F., Gu,X., He,K. and Luo,J. (2008) PlantTFDB: a comprehensive plant transcription factor database. Nucleic Acids Res., 36, D966–D969.
3. Palaniswamy,S.K., James,S., Sun,H., Lamb,R.S., Davuluri,R.V. and Grotewold,E. (2006) AGRIS and AtRegNet. A platform to link cis-regulatory elements and transcription factors into regulatory networks. Plant Physiol., 140, 818–829.
4. Riano-Pachon,D.M., Ruzicic,S., Dreyer,I. and Mueller-Roeber,B. (2007) PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics, 8, 42.
5. Yilmaz,A., Nishiyama,M.Y. Jr, Fuentes,B.G., Souza,G.M., Janies,D., Gray,J. and Grotewold,E. (2009) GRASSIUS: a platform for comparative regulatory genomics across the grasses. Plant Physiol., 149, 171–180.
6. Richardt,S., Lang,D., Reski,R., Frank,W. and Rensing,S.A. (2007) PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins. Plant Physiol., 143, 1452–1466.
7. Zimmer,A., Lang,D., Richardt,S., Frank,W., Reski,R. and Rensing,S.A. (2007) Dating the early evolution of plants: detection and molecular clock analyses of orthologs. Mol. Genet. Genomics, 278, 393–402.
8. Matsuoka,M., Misumi,O., Shin,I.T., Maruyama,S., Takahara,M., Miyagishima,S.Y., Mori,T., Nishida,K., Yagisawa,F., Nishida,K.

Table 2. Sensitivity and PPV of PlnTFDB predictions

| Species | Family      | Reference | TP/TP + FN | TP/TP + FP | Sensitivity | PPV  |
|---------|-------------|-----------|------------|------------|-------------|------|
| ATH     | AP2-EREBP   | (25)      | 146/147    | 146/146    | 0.99        | 1.00 |
|         | ARF         | (26)      | 21/23      | 21/23      | 0.91        | 0.91 |
|         | AUX/IAA     | (26)      | 26/29      | 28/28      | 0.97        | 1.00 |
|         | bHLH        | (27)      | 125/154    | 125/136    | 0.81        | 0.92 |
|         | bZIP        | (28)      | 70/76      | 70/70      | 0.92        | 1.00 |
|         | C2C2-Dof    | (29)      | 35/36      | 35/36      | 0.97        | 0.97 |
|         | C2C2-GATA   | (30)      | 29/29      | 29/29      | 1.00        | 1.00 |
|         | C3H         | (31)      | 65/67      | 65/68      | 0.97        | 0.96 |
|         | GRAS        | (32)      | 32/32      | 32/33      | 1.00        | 0.97 |
|         | MADS        | (33)      | 97/105     | 98/105     | 0.91        | 0.93 |
|         | MADS        | (34)      | 98/108     | 185/212    | 0.93        | 0.87 |
|         | MYB         | (35)      | 100/100    | 100/104    | 1.00        | 0.96 |
|         | NAC         | (36)      | 16/17      | 16/16      | 0.94        | 1.00 |
|         | SBP         | (37)      | 71/72      | 71/72      | 0.99        | 0.99 |
|         | WRKY        | (38)      | 134/146    | 134/143    | 0.81        | 0.94 |

The sensitivity and the PPV were determined for selected A. thaliana (ATH) and O. sativa ssp. japonica (OSAJ) TF families. For the PPV, a deviation from 1.00 means the inclusion of FPs. For the sensitivity, deviations from 1.00 indicate exclusion of true members (FNS). Families with both values larger than 0.90 appear in bold face. TPs according to gold standard.
et al. (2004) Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. Nature, 428, 653–657.

9. Barbier,G., Oesterhelt,C., Larson,M.D., Halgren,R.G., Wilkerson,C., Garavito,R.M., Benning,C. and Weber,A.P. (2005) Comparative genomics of two closely related unicellular thermoacidophilic red algae, Galderia sulphuraria and Cyanidioschyzon merolae, reveals the molecular basis of the metabolic flexibility of Galderia sulphuraria and significant differences in carbohydrate metabolism of both algae. Plant Physiol., 137, 460–474.

10. Worden,A.Z., Lee,J.-H., Mock,T., Rouze,P., Simmons,M.P., Aerts,A.L., Allen,A.E., Cuveller,M.L., Derelle,E., Everett,M.V. et al. (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. Science, 324, 265–272.

11. Palenik,B., Grimwood,J., Aerts,A., Rouze,P., Salamov,A., Putnam,N., Dupont,C., Jorgensen,R., Derelle,E., Rombouts,S. et al. (2007) The tiny eukaryote Ostreococcus tosilacatus provides genomic insights into the paradox of plankton specialization. Proc. Natl Acad. Sci. USA, 104, 7705–7710.

12. Merchant,S.S., Prochnik,S.E., Vallon,O., Harris,E.H., Karpowitz,S.J., Wilman,G.B., Terry,A., Salamov,A., Fritz-Laylin,L.K., Marechal-Drouard,L. et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science, 318, 245–250.

13. Rensing,S.A., Lang,D., Zimmer,A.D., Terry,A., Salamov,A., Shapiro,H., Nishiyama,Y., Perroud,P.F., Lindquist,E.A., Kamisugi,Y. et al. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science, 319, 64–69.

14. Zhao,W., Wang,J., He,X., Huang,X., Jiao,Y., Dai,M., Wei,S., Fu,J., Chen,Y., Ren,X. et al. (2004) BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. Nucleic Acids Res., 32, D377–D382.

15. Yuan,Q., Ouyang,S., Wang,A., Zhu,W., Maiti,R., Lin,H., Hamilton,J., Haas,B., Sultana,R., Cheung,F. et al. (2005) The institute for genomic research OsA rice genome annotation database. Plant Physiol., 138, 18–26.

16. Paterson,A.H., Bowers,J.E., Bruggmann,R., Dubchak,I., Grimwood,J., Gundlach,H., Haberer,G., Hellsten,U., Mitros,T., Poliakov,A. et al. (2009) The Sorghum bicolor genome and the diversification of grasses. Nature, 457, 551–556.

17. Ming,R., Hou,S., Feng,Y., Yu,Q., Dione-Laporte,A., Saw,J.H., Senin,P., Wang,W., Ly,B.-V., Lewis,K.L. et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya). Nature, 452, 991–996.

18. Swarbreck,D., Wilks,C., Lamesch,P., Berardini,T.Z., Garcia-Hernandez,M., Foerster,H., Li,D., Meyer,T., Muller,R., Pfalle,A. et al. (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res., 36, D1009–D1014.

19. Tuskan,G.A., Difazio,S., Jansson,S., Bohmann,J., Grigoriev,I., Hellsten,U., Putnam,N., Ralph,P., Rombouts,S., Salamov,A. et al. (2006) The genome of black cottonwood, Populus trichocarpa (Torrey & Gray). Science, 313, 1596–1604.

20. The French International Public Consortium for Grapevine Genome Characterization. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature, 449, 463–467.

21. Finn,R.D., Tate,J., Mistry,J., Coggill,P.C., Sammut,S.J., Hotz,H.R.C., Ceric,G., Forslund,K., Eddy,S.R., Sonnhammer,E.L. et al. (2008) The Pfam protein families database. Nucleic Acids Res., 36, D218–D228.

22. Duncan,L., Nishii,L., Harryman,A., Buckley,S., Howard,A., Friedman,N. and Miller,S.M. (2007) The VARl gene family and the evolutionary origins of the master cell-type regulatory gene, regA, in Volvox carteri. J. Mol. Evol., 65, 1–11.

23. Dolinski,K. and Botstein,D. (2007) Orthology and functional conservation in eukaryotes. Annu. Rev. Genet., 41, 465–507.

24. Remm,M., Storm,C.E. and Sonnhammer,E.L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J. Mol. Biol., 314, 1041–1052.

25. Feng,J.X., Liu,D., Pan,Y., Gong,W., Ma,L.G., Luo,J.C., Deng,X.W. and Zhu,Y.X. (2005) An annotation update via cDNA sequence analysis and comprehensive profiling of developmental, hormonal or environmental responsiveness of the Arabidopsis AP2/EERBP transcription factor gene family. Plant Mol. Biol., 59, 853–868.

26. Remington,D.L., Vision,T.J., Guilfoyle,T.J. and Reed,J.W. (2004) Contrasting modes of diversification in the Aux/IAA and ARF gene families. Plant Physiol., 135, 1738–1752.

27. Bailey,P.C., Martin,C., Toledo-Ortiz,G., Quail,P.H., Huq,E., Heim,M.A., Jakoby,M., Merker,W. and Weisshaar,B. (2003) Update on the basic helix-loop-helix transcription factor gene family in Arabidopsis thaliana. Plant Cell, 15, 2497–2502.

28. Corrêa,L.G., Riaño-Pachón,D.M., Schragøe,C.G., dos Santos,R.V., Mueller-Roemer,B. and Vincenzi,M. (2008) The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. PLoS ONE, 3, e2944.

29. Lijavetzky,Z., Carbonero,P. and Vicente-Carbajosa,J. (2003) Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. BMC Evol. Biol., 3, 17.

30. Reyes,J.C., Muro-Pastor,M.I. and Florencio,F.J. (2004) The GATA family of transcription factors in Arabidopsis and rice. Plant Physiol., 134, 1718–1732.

31. Wang,D., Guo,Y., Wu,C., Yang,G., Li,Y. and Zheng,C. (2008) Genome-wide analysis of CCHC zinc finger family in Arabidopsis and rice. BMC Genomics, 9, 44.

32. Bolle,C. (2004) The role of GRAS proteins in plant signal transduction and development. Plant, 218, 683–692.

33. Martinez-Castilla,L.P. and Alvarez-Buylla,E.R. (2003) Adaptive transcription of the Arabidopsis papaya (Carcia papaya Linnaeus). Sci. USA, 100, 13407–13412.

34. Parenicova,L., de Folter,S., Klieffer,M., Horner,D.S., Favaelli,L., Busscher,J., Cook,H.E., Ingram,R.M., Kater,M.M., Davies,B. et al. (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell, 15, 1538–1551.

35. Yanhui,C., Xiaoyuan,Y., Kun,H., Meihua,L., Jigang,L., Zhaofeng,G., Zhiqiang,L., Yunfei,Z., Xiaoxiao,W. and Xiaoming,Q. et al. (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol. Biol., 60, 107–124.

36. Ooka,H., Satoh,K., Doi,K., Nagata,T., Otomo,Y., Murakami,K., Matsubara,K., Osato,N., Kawai,J., Carninci,P. et al. (2003) Comprehensive analysis of NAC family genes in Orzya sativa and Arabidopsis thaliana. DNA Res., 10, 239–247.

37. Guo,A.Y., Zhu,Q.H., Gu,X., Ge,S., Yang,J. and Luo,J. (2008) The GATA family of transcription factors in Arabidopsis and rice. BMC Genomics, 9, 44.

38. Ulker,B. and Somssich,I.E. (2004) WRKY transcription factors: from DNA binding towards biological function. Curr. Opin. Plant Biol., 7, 491–498.

39. Li,X., Duan,X., Jiang,H., Sun,Y., Tang,Y., Yuan,Z., Guo,J., Liang,W., Chen,L., Yin,J. et al. (2006) Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. Plant Physiol., 141, 1167–1184.

40. Riaño-Pachón,D.M., Corrêa,L.G., Trejos-Espinosa,R. and Mueller-Roemer,B. (2008) Green transcription factors: a Chlamydomonas overview. Genetics, 179, 31–39.

41. Velasco,R., Zharkikh,A., Troggio,M., Cartwright,D.A., Cestaro,A., Pruss,D., Pindo,M., Fitzgerald,L.M., Vezzulli,S., Reid,J. et al. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One, 2, e1326.

42. Caldana,C., Scheible,W.R., Mueller-Roemer,B. and Ruzicic,S. (2007) A quantitative RT-PCR platform for high-throughput expression profiling of 2500 rice transcription factors. Plant Methods, 3, 7.

43. Street,N.R., Sjödin,A., Bylesjo,M., Gustafsson,P., Trygg,J. and Jansson,S. (2008) A cross-species transcriptionomics approach to identify genes involved in leaf development. BMC Genomics, 9, 889.