Published online 2 November 2010 *Nucleic Acids Research, 2011, Vol. 39, Database issue* ***D1123–D1129***
doi:10.1093/nar/gkq1066

AHD2.0: an update version of Arabidopsis Hormone Database for plant systematic studies

Zhiqiang Jiang[1], Xiaochuan Liu[2], Zhiyu Peng[3], Yinan Wan[1], Yusi Ji[1], Wenrong He[1], Wen Wan[1], Jingchu Luo[2,*] and Hongwei Guo[1,*]

[1]National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, [2]Center for Bioinformatics, National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871 and [3]Beijing Genomics Institute at Shenzhen, Shenzhen 518083, P. R. China

Received August 14, 2010; Revised October 9, 2010; Accepted October 13, 2010

ABSTRACT

Phytohormone studies enlightened our knowledge of plant responses to various changes. To provide a systematic and comprehensive view of genes participating in plant hormonal regulation, an online accessible database Arabidopsis Hormone Database (AHD) has been developed, which is a collection of hormone related genes of the model organism *Arabidopsis thaliana* (AHRGs). Recently we updated our database from AHD to a new version AHD2.0 by adding several pronounced features: (i) updating our collection of AHRGs based on most recent publications as well as constructing elaborate schematic diagrams of each hormone biosynthesis and signaling pathways; (ii) adding orthologs of sequenced plants listed in OrthoMCL-DB to each AHRG in the updated database; (iii) providing predicted miRNA splicing site(s) for each AHRG; (iv) integrating genes that genetically interact with each AHRG according to literatures mining; (v) providing links to a powerful online analysis platform WebLab for the convenience of in-time bioinformatics analysis and (vi) providing links to widely used protein databases and integrating more expression profiling information that would facilitate users for a more systematic and integrative analysis related to phytohormone research.

INTRODUCTION

Phytohormones play essential roles throughout the lifespan of plants. During the past decades, plant researchers took efforts to elucidate biological roles and functioning mechanisms of phytohormones in various plant responses (1). And thus, phytohormones had been revealed to mediate a whole range of developmental processes, as well as to interact with environmental factors by extensive physiological and genetic studies. To summarize the advance made on phytohormone research in recent decades, an Arabidopsis Hormone Database (AHD, http://ahd.cbi.pku.edu.cn) was first released in 2008 (2). The database includes a large collection of Arabidopsis hormone related genes (AHRGs), which are defined as genes involving in the biosynthesis, metabolism, transport, perception or signaling pathways of eight types of phytohormones. The AHD also integrates detailed gene information (largely manually curated) and a phenotype ontology that is developed to precisely describe myriad hormone-regulated morphological processes with standardized vocabularies in the model organism *Arabidopsis*. The database aims to provide systematic and comprehensive view of Arabidopsis hormone related genes. Free and friendly interface is available for online access to the database, which benefits searching particular hormone related gene or retrieve bulk data by hormone category or specific phenotypic trait. Meanwhile, cross-links to PPI and KEGG were also provided for molecular biochemical studies.

In the past 2 years, a large body of advances had been made in phytohormone studies. Many new genes have been identified to function in plant hormone production or signaling pathways. For example the discovery of ETP1/2 that regulate EIN2 proteolysis filled the gap of ethylene signaling pathway (3). Meanwhile, unequivocal evidence had been accumulated to solve some long-sought questions or clarify long-standing controversies, for instance, the recent identification of PYR/PYL/PCAR

*To whom correspondence should be addressed. Tel: +86 10 6276 7823; Email: hongweig@pku.edu.cn
Correspondence may also be addressed to Jingchu Luo. Email: luojc@mail.cbi.pku.edu.cn

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

© The Author(s) 2010. 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), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

proteins as authentic ABA receptors (4). Besides, wide-range interactions among various components lead to complicated modes of phytohormones action, thus increasing efforts using systems biology approaches have been employed in order to uncover the underlying mechanism (5). As a result, phytohormone studies have been expanded to a much wider scope. For example genomic scale analysis has been commonly used in recent studies. A set of microarray profiling data sets have been released by AtGenexpress group (6–8), including the data sets derived from different developmental stages or various treatments of *Arabidopsis*, providing a systematic view of transcriptional regulation in response to environmental or developmental factors. Meanwhile, next generation sequencing has also enabled us to analyse plant gene functions from different perspectives. On one hand, sequencing information helps us understand features of genomic sequences as well as epigenetic regulations of plants such as methylation or microRNA binding on particular loci (9,10); on the other hand, as more plant species are in the sequencing pipeline, comparative genomic approaches make evolutionary studies of plant pathways available. For example the orthologs of ABI3 proteins in the moss *Physcomitrella patens* were found to be functional in drought tolerance responses (11), indicating that ABA regulation is at least partly conserved between early land plants and seed plants. Recently, theoretical approaches had also been developed for phytohormone research. These approaches are highly relied on the information about genetic and biochemical interactions among various genes or components, and often utilize mathematical methods to integrate and model verified experimental data sets, leading to conceptual understanding of complex regulatory networks. One example is a recently published work using mathematical modeling to illustrate how POLARIS gene functions and interacts with auxin, ethylene and cytokinin in *Arabidopsis* (12).

In an attempt to cover the above-mentioned advances and extending the functionality of the previous AHD, we have updated our database from its first version to AHD2.0. In this update version, we made tremendous modifications and improvements to the original database and added several new features to cater to nowadays interests. First, we manually updated our collection of AHRG genes based on most recent publications (the latest 2-year literatures) as well as corrected information (e.g. the arguable ABA receptors). Besides, elaborate schematic diagrams of phytohormone biosynthesis and signaling pathways are constructed and provided in AHD2.0, which represents the first effort to integrate a large volume of data sets into a comprehensive regulatory network for each hormone; second, we integrated Orthologs of sequenced plants in OrthoMCL-DB (13) into each gene in the database for comparative genomic or evolution studies of phytohormone related genes; third, we predicted microRNA splicing site of each gene. The sequence of each binding site is now also available for performing further analysis such as primer design; fourth, we provided the genetic relationship of these phytohormone related genes manually curated based on published literatures; fifth, for the convenience of in-time bioinformatics analysis, we provided links to a powerful online analysis platform WebLab, which was recently developed by us. It allows users to readily perform various sequence analysis with these phytohormone related genes retrieved from the database; finally, we provided links to other protein databases as well as more expression profiling information that would facilitate users for a more systematic and integrative analysis for phytohormone research. Based on these newly developed features, AHD2.0 would greatly intensify and expand its capability in the systematic studies of plant hormone responses.

NEW FEATURES AND DETAILS

The update of phytohormone related gene collection

AHD2.0 integrated gene entries extracted from 906 scientific papers published before August 2010. Compared with AHD, gene entries have increased from 1026 to 1318 (Figure 1A). Mutant entries have increased from 575 to 1012 (Figure 1B). The increased entries are all manually curated based on the recent 2-year literatures from September 2008 to August 2010. These newly added gene entries were categorized by hormones, while the mutants related to those genes were classified by their phenotype ontology that was developed in AHD. In AHD, a number of genes were considered as AHRGs

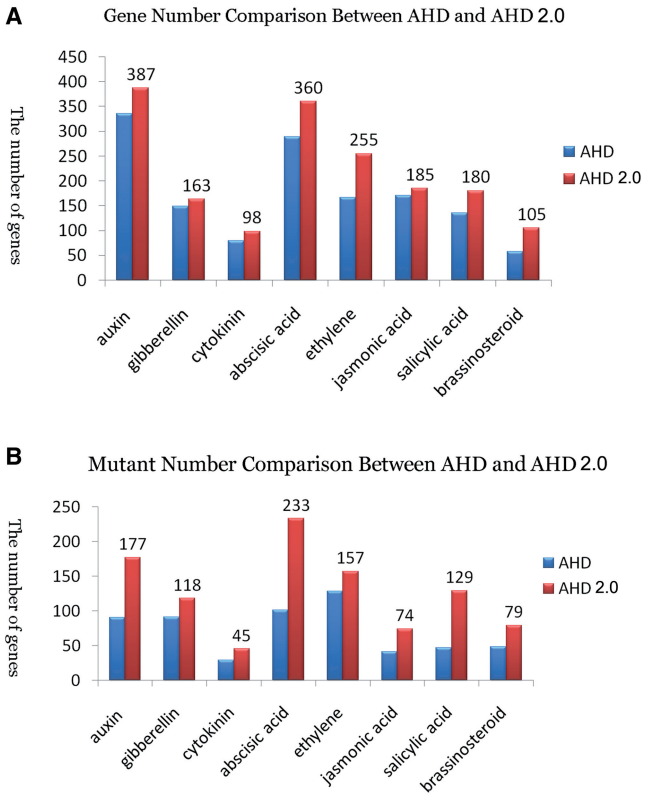

Figure 1. Entries changes in AHD2.0. Histogram illuminates the comparisons of gene number (**A**) and mutant number (**B**) between the two versions of the database. Color in blue represents the numbers in AHD and color in red represents the numbers in AHD2.0.

merely based on annotation from Gene Ontology. While in AHD2.0, with a large number of genetic studies using mutants and transgenically overexpressing or silencing lines had been incorporated to some of these genes, the number of mutants (or genes with verified genetic evidence) in the database was increased by 76%.

Besides, we modified existed information based on the most recent publications. For instance, GCR2 and FCA were classified into ABA receptors in AHD, but in AHD2.0 these genes have been re-classified into ABA signaling category due to the recent suspicion. Meanwhile GTG1/2 and PYR/PYL/PCAR were added into AHD2.0 and were classified into ABA receptors based on current studies (4,14).

New information added to genes

For epigenetic study users, we now provide predicted microRNA splicing site of each gene. Prediction of the potential interactions between microRNAs and genes utilized RNAhybrid method as described (15). In sum, we have predicted 954 microRNA splicing site targeting on hormone related genes. All details were given in the 'miRNA interaction information for this gene' section with the name of miRNAs and their target genes, the length of mature miRNAs and their target genes and their binding site (Figure 2D). Detailed information of a given miRNA is available in the cross-link to miRBase from each ID of the miRNA.

Genes that were reported to genetically interact with AHRGs are included in AHD2.0. The interactions are manually curated based on literatures. In 'Gene interaction information for this gene' section we provide information including related genes, genetic relationships (represented by positive/negative regulation), regulation type, enzyme activity (if possible) and the accesses to literature evidence from PubMed ID for each interacting gene (Figure 2E). Nodes to describe genes and edges to

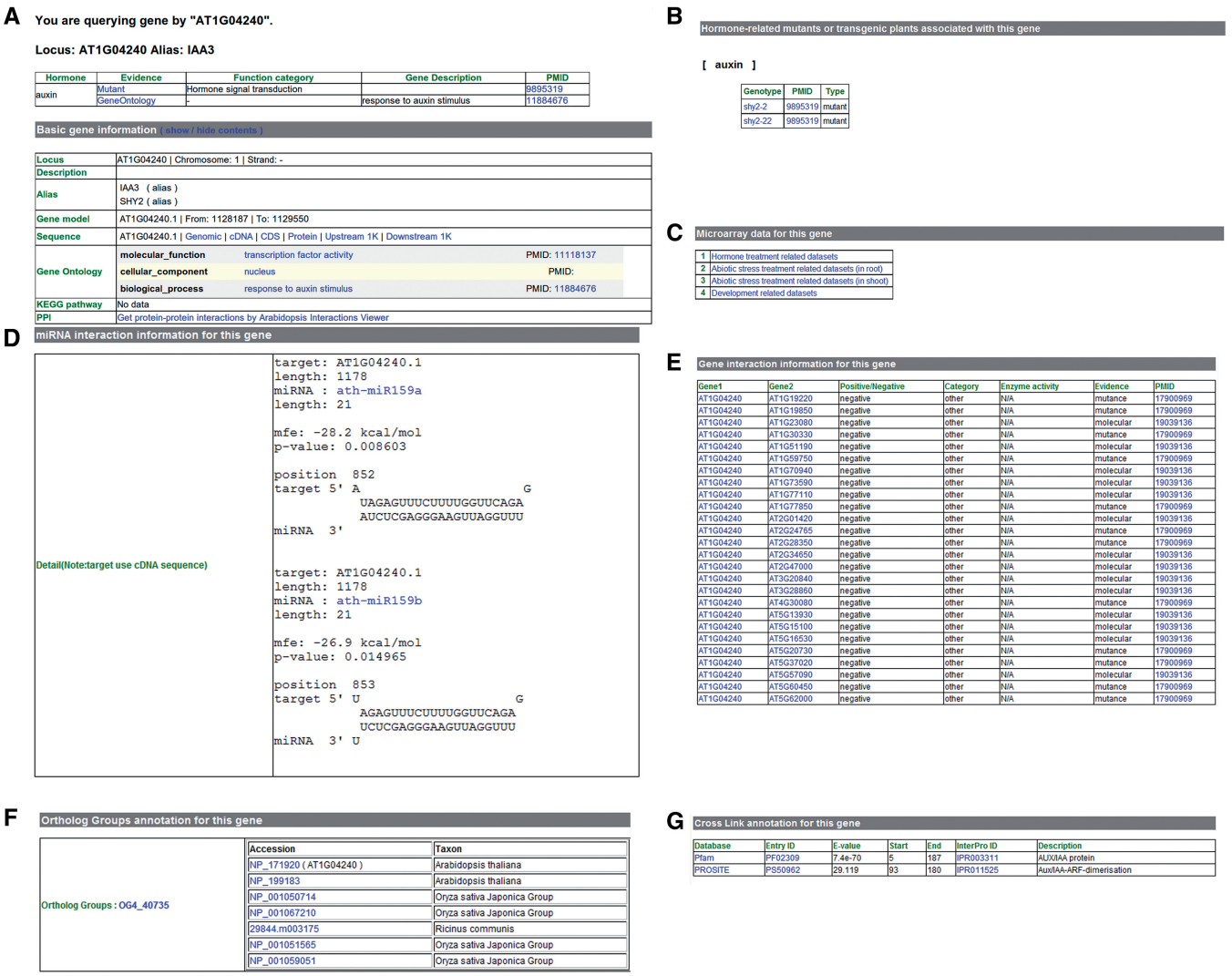

Figure 2. Screenshot of gene information. The information of an updated hormone-related gene that contains seven sections, including (**A**) response hormone(s) and basic gene information retrieved from TAIR, GO and KEGG, (**B**) associated mutants, (**C**) expression data from microarray experiements, (**D**) predicted miRNA splicing site, (**E**) genetic interactions related to the gene, (**F**) Ortholog groups annotation for the gene and (**G**) cross-links to other protein databases.

indicate their reciprocal interactions are used for modeling a regulatory network. In general, we have found a total of 911 nodes and 2608 edges derived from ARHGs listed in AHD2.0.

In addition, a quick description with the schematic pathways of all eight hormones is added in the AHD2.0 introduction. By integrating all available information from literatures, we constructed the schematic diagrams for each type of phytohormones, including their biosynthesis, metabolism, transport (if applicable), perception and signal transduction pathways. The schematic diagrams of these pathways are now available from the introduction page of the database and can also be found in the left bar of the database. By clicking the name of each hormone in the table, users can get access to the graphic pathways separately. These informative and elaborate diagrams would be useful to hormone researchers.

For easy access to comparative or evolution studies, we also provide Ortholog groups in OrthoMCL-DB (13) for each gene. Sequence of Ortholog groups can be retrieved from the ID shown in the section 'Ortholog Groups annotation for this gene' from the cross link to OrthoMCL-DB (Figure 2F).

On-line analysis tool

In AHD2.0, links to WebLab (16) are provided in the sequence page of each gene. WebLab is a data-centric knowledge-sharing bioinformatics platform in which 279 local tools and Web-services and Grid-services are integrated. This platform will allow users to benefit from on-time analysis when sequence is retrieved. With submitted sequence to WebLab, users can help themselves to analysis such as sequence alignment, protein 2D and 3D structure prediction and motif search. Besides, registered users can take benefit from saving and sharing data, which is our highly recommended procedure.

Cross-links and other information

In AHD2.0, cross-links to other protein databases are provided, which include Pfam, SuperFamily, PROSITE, Gene3D, PRINTS, PANTHER, PIR, HAMAP,

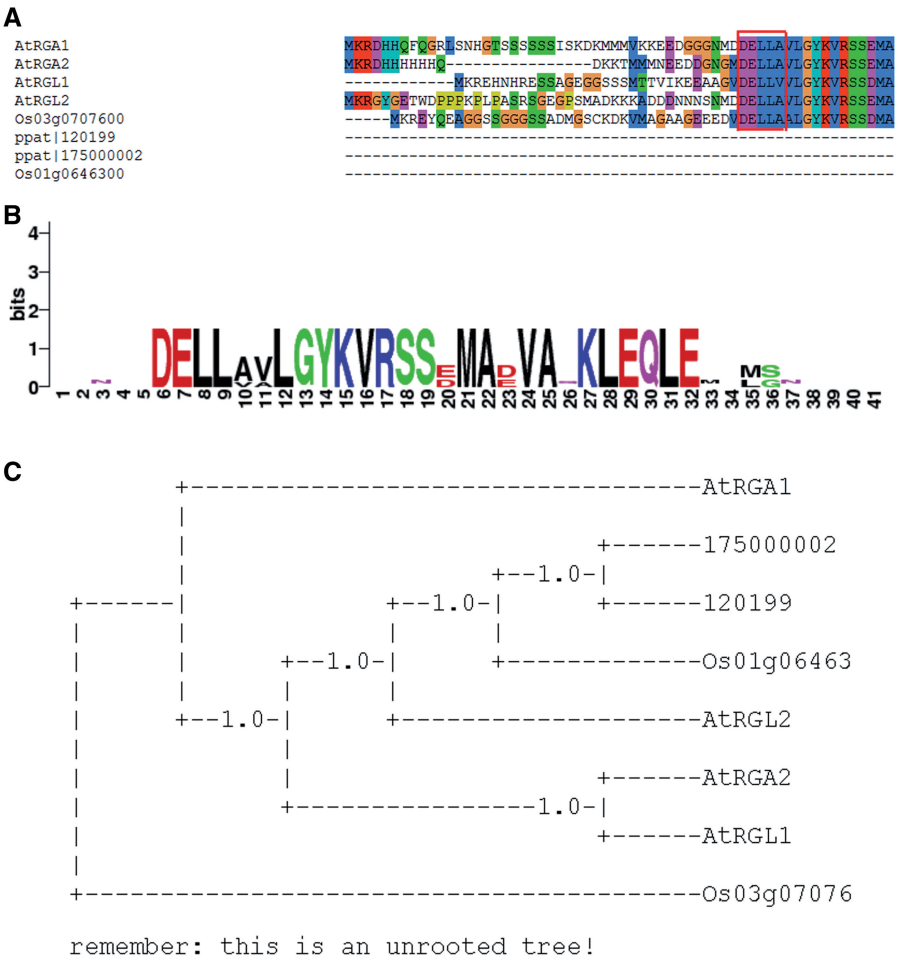

Figure 3. A case for the application of WebLab. DELLA proteins are chosen as queries. Based on Ortholog groups retrieved from AHD2.0, multiple sequence alignment can be performed using clustal software in WebLab (**A**). It reveals that in lower plant such as the moss *Physcomitrella patens* (with the ID 175000002 and 120199), although Ortholog proteins of AtRGA1 had been evolved, DELLA motif is lacked in the moss. DELLA motif can be visualized by using Weblogo in WebLab (**B**). And Phylogenetic tree can be constructed by fneighbor in WebLab based on N-J algorism (**C**), which indicates systematic relationship of these Ortholog genes.

TIGRFAM, SMART, PROSITE and ProDom (Figure 2G). We believe that this protein information will facilitate users in predicting gene functions in the network based on domain analysis.

Besides, we have integrated more expression profiling information in AHD2.0 derived from AtGenexpress (6–8), including different developmental stages and abiotic-stress treatment profiling data sets (Figure 2C). These data sets would facilitate researches to study transcriptional regulation of hormone related genes in various plant responses. To make it user friendly, the normalized expression level was shown in LOG2 transformation and the graphic view of gene expression data was only displayed for those genes influenced >2-fold in at least one experiment within one data set.

DISCUSSION AND FUTURE PLANS

We have updated our previous version of AHD and renamed it as AHD2.0. With the tremendous expansion of gene collection and correction, as well as newly developed features and web tools, we believe that AHD2.0 would be more efficient and powerful for the systems biology studies of plant hormone functions and mechanisms. These kinds of integrative research would allow a better understanding of the phytohormone regulatory network in myriad plant responses and eventually lead to the proficient manipulation of phytohormone action to improve plant fitness, defense and crop yield.

Several applications can be derived from AHD2.0. First, we provided Ortholog groups of each gene. Comparative or evolutionary studies are powerful methods in uncovering essential mechanisms in plant responses due to the fact that each complex network is originated from simple state and is fixed by nature selection (17). Analyzing function of Ortholog genes in ancestors or related species may indicate the origin, adjustment as well as function in particular responses of a pathway. AHD2.0 will benefit users in searching Orthologs of hormone related genes and performing necessary analysis from quick access to WebLab (Figure 3).

Second, statistic methods can be used in analyzing information retrieved from AHD2.0. Take interactions between microRNA and hormone related genes as an example. Based on on-line information, the majority of genes interacted with microRNA are related to auxin and abscisic acid (Figure 4A). Interestingly, our database and analysis revealed that 88% ethylene related genes have interactions with microRNA (Figure 4B), indicating that ethylene response pathways are probably extensively under small RNA regulation, which is supported by previous findings (18–20).

In another application, genetic interactions can be used as important resources in systems biology study. Detailed schematic diagrams illustrating biosynthesis, perception as well as transduction process of phytohormones would enable phytohormone researchers to get more comprehensive view of phytohormone actions. Meanwhile, these diagram model (nodes with connections by arrowhead

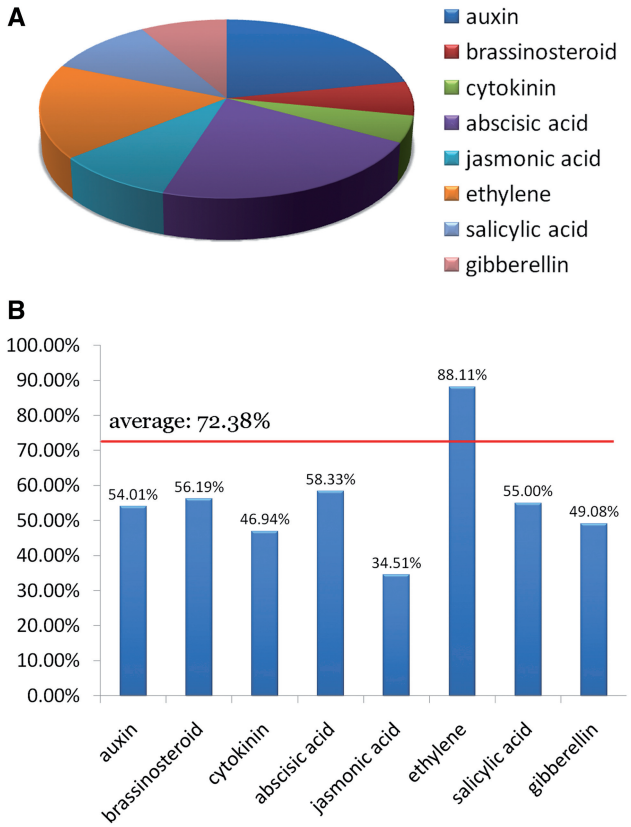

Figure 4. Statistical analysis of the interactions between miRNA and hormone-related genes. Auxin and abscisic acid related genes predicted to have more miRNA targets, as shown in (**A**). Different color represents different hormone categories. The percentages of genes predicted to have miRNA binding sites in each hormone category are shown in (**B**), in which a red line marked out the average percentage of 72.38%.

or perpendicular lines) can be changed into Boolean networks through a simple transformation, which may facilitate for mathematical modeling by employing Boolean rules and binary values as well (21,22). Based on 911 nodes and 2608 edges provided in AHD2.0, a combination of genetic interactions may enable us to draw a relatively complete regulatory network of phytohormones, which may assist systematic users performing theoretical analysis (Figure 5). It should be pointed out that interactions among hormones are too intricate to be concluded by simple up or down regulations, because they are under the influence of tissue and treatment specificity. For example for the reason that some of the ABA related genes are seed-specific genes, the relationship of ethylene and ABA is different in drought tolerance and seed germination (23). The addition of more expression profiling data sets and the powerful on-line analysis platform are believed to be able to facilitate detailed researches. As for systems biology, conclusions could be heuristically derived when data sets from different angles are perfectly integrated.

To date, information in AHD2.0 mainly focuses on Arabidopsis genes, although Ortholog groups are provided to assist phytohormone studies in other plants. In the future, we aim to include more characteristics of

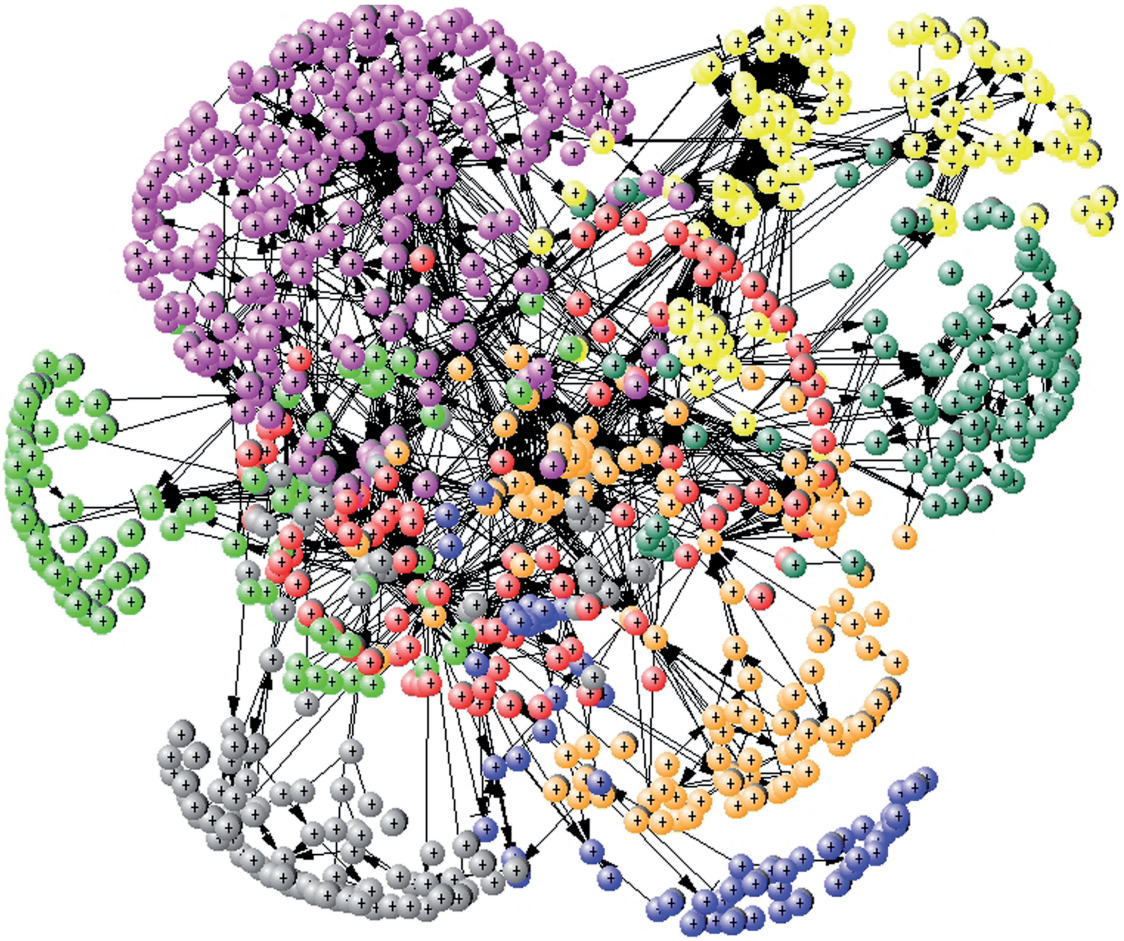

Figure 5. Boolean network of phytohormone. The nodes and edges represent a Boolean Network constructed by VisAnt. Different color of the node represents hormone category the gene (node) belongs to, (auxin: orange, gibberellin: grey, cytokinin: yellow, abscisic acid: purple, ethylene: light green, jasmonic acid: red, salicylic acid: blue, brassinosteroid: dark green).

phenotype ontology for other plant species and to integrate physiological and/or genetic studies from other species. The ultimate aim of our database is to provide an informative resource and analysis platform for phytohormone studies, not only in *Arabidopsis*, but also in other non-model plant species as well. We believe that the new version of AHD will make a broader impact on plant biology community and cater to the demand of more users. Users are also welcome to help us in improving this database via our online platform (http://ahd.cbi.pku.edu.cn/help.php). All the useful information will be appreciated and carefully considered to be integrated into our database.

ACKNOWLEDGEMENTS

The authors thank Bo Qin, a graduate student from The University of North Carolina at Chapel Hill for article revising and Xing Tang from center of bioinformatics in Peking University for web decoration. The authors also appreciate the useful suggestion from Dr Galperin and the anonymous referee.

FUNDING

National Natural Science Foundation of China (30625003 and 30730011 to H.G.); Ministry of Science and Technology of China (2009CB119101 to H.G.); Ministry of Education of China (ED20060047 to H.G.). Funding for open access charge: 111 project.

Conflict of interest statement. None declared.

REFERENCES

1. Wolters,H. and Jurgens,G. (2009) Survival of the flexible: hormonal growth control and adaptation in plant development. *Nat. Rev. Genet.*, **10**, 305–317.
2. Peng,Z.Y., Zhou,X., Li,L., Yu,X., Li,H., Jiang,Z., Cao,G., Bai,M., Wang,X., Jiang,C. *et al.* (2009) Arabidopsis Hormone Database: a comprehensive genetic and phenotypic information database for plant hormone research in Arabidopsis. *Nucleic Acids Res.*, **37**, D975–D982.
3. Qiao,H., Chang,K.N., Yazaki,J. and Ecker,J.R. (2009) Interplay between ethylene, ETP1/ETP2 F-box proteins, and degradation of EIN2 triggers ethylene responses in Arabidopsis. *Genes Dev.*, **23**, 512–521.
4. Miyazono,K., Miyakawa,T., Sawano,Y., Kubota,K., Kang,H.J., Asano,A., Miyauchi,Y., Takahashi,M., Zhi,Y., Fujita,Y. *et al.*

(2009) Structural basis of abscisic acid signalling. *Nature*, **462**, 609–614.
5. Gutierrez,R.A., Shasha,D.E. and Coruzzi,G.M. (2005) Systems biology for the virtual plant. *Plant Physiol.*, **138**, 550–554.
6. Goda,H., Sasaki,E., Akiyama,K., Maruyama-Nakashita,A., Nakabayashi,K., Li,W., Ogawa,M., Yamauchi,Y., Preston,J., Aoki,K. *et al.* (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J.*, **55**, 526–542.
7. 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.
8. Schmid,M., Davison,T.S., Henz,S.R., Pape,U.J., Demar,M., Vingron,M., Scholkopf,B., Weigel,D. and Lohmann,J.U. (2005) A gene expression map of Arabidopsis thaliana development. *Nat. Genet.*, **37**, 501–506.
9. Lister,R., O'Malley,R.C., Tonti-Filippini,J., Gregory,B.D., Berry,C.C., Millar,A.H. and Ecker,J.R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, **133**, 523–536.
10. Zhang,X., Yazaki,J., Sundaresan,A., Cokus,S., Chan,S.W., Chen,H., Henderson,I.R., Shinn,P., Pellegrini,M., Jacobsen,S.E. *et al.* (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell*, **126**, 1189–1201.
11. Khandelwal,A., Cho,S.H., Marella,H., Sakata,Y., Perroud,P.F., Pan,A. and Quatrano,R.S. (2010) Role of ABA and ABI3 in desiccation tolerance. *Science*, **327**, 546.
12. Liu,J., Mehdi,S., Topping,J., Tarkowski,P. and Lindsey,K. (2010) Modelling and experimental analysis of hormonal crosstalk in Arabidopsis. *Mol. Syst. Biol.*, **6**, 373.
13. Chen,F., Mackey,A.J., Stoeckert,C.J. Jr and Roos,D.S. (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.*, **34**, D363–D368.
14. Pandey,S., Nelson,D.C. and Assmann,S.M. (2009) Two novel GPCR-type G proteins are abscisic acid receptors in Arabidopsis. *Cell*, **136**, 136–148.
15. Alves,L. Jr, Niemeier,S., Hauenschild,A., Rehmsmeier,M. and Merkle,T. (2009) Comprehensive prediction of novel microRNA targets in Arabidopsis thaliana. *Nucleic Acids Res.*, **37**, 4010–4021.
16. Liu,X., Wu,J., Wang,J., Zhao,S., Li,Z., Kong,L., Gu,X., Luo,J. and Gao,G. (2009) WebLab: a data-centric, knowledge-sharing bioinformatic platform. *Nucleic Acids Res.*, **37**, W33–W39.
17. Jiang,Z. and Guo,H. (2010) A comparative genomic analysis of plant hormone related genes in different species. *J. Genet. Genomics*, **37**, 219–230.
18. Olmedo,G., Guo,H., Gregory,B.D., Nourizadeh,S.D., Aguilar-Henonin,L., Li,H., An,F., Guzman,P. and Ecker,J.R. (2006) ETHYLENE-INSENSITIVE5 encodes a $5'\rightarrow3'$ exoribonuclease required for regulation of the EIN3-targeting F-box proteins EBF1/2. *Proc. Natl Acad. Sci. USA*, **103**, 13286–13293.
19. Potuschak,T., Vansiri,A., Binder,B.M., Lechner,E., Vierstra,R.D. and Genschik,P. (2006) The Exoribonuclease XRN4 Is a Component of the Ethylene Response Pathway in Arabidopsis. *Plant Cell*, **18**, 3047–3057.
20. Gregory,B.D., O'Malley,R.C., Lister,R., Urich,M.A., Tonti-Filippini,J., Chen,H., Millar,A.H. and Ecker,J.R. (2008) A Link between RNA Metabolism and Silencing Affecting Arabidopsis Development. *Dev. Cell*, **14**, 854–866.
21. Davidson,E.H., Rast,J.P., Oliveri,P., Ransick,A., Calestani,C., Yuh,C.H., Minokawa,T., Amore,G., Hinman,V., Arenas-Mena,C. *et al.* (2002) A genomic regulatory network for development. *Science*, **295**, 1669–1678.
22. Price,N.D. and Shmulevich,I. (2007) Biochemical and statistical network models for systems biology. *Curr. Opin. Biotechnol.*, **18**, 365–370.
23. Suzuki,M. and McCarty,D.R. (2008) Functional symmetry of the B3 network controlling seed development. *Curr. Opin. Plant Biol.*, **11**, 548–553.