Plant Biotechnology 29, 131–135 (2012)
DOI: 10.5511/plantbiotechnology.12.0508a

Note

Development of KaPPA-View4 for omics studies on Jatropha and a database system KaPPA-Loader for construction of local omics databases

Nozomu Sakurai[1,*], Yoshiyuki Ogata[2], Takeshi Ara[1], Ryosuke Sano[1], Nayumi Akimoto[1], Atsushi Hiruta[1], Hideyuki Suzuki[1], Masataka Kajikawa[3], Utut Widyastuti[4], Sony Suharsono[4], Akiho Yokota[3], Kinya Akashi[3], Jun Kikuchi[2], Daisuke Shibata[1]

[1] Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan; [2] RIKEN Plant Science Center, Yokohama, Kanagawa 230-0045, Japan; [3] Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan; [4] Research Center for Bioresources and Biotechnology, Bogor Agricultural University, Gd. PAU, Kampus IPB Darmaga, Bogor 16680, Indonesia
* E-mail: sakurai@kazusa.or.jp Tel: +81-438-52-3590 Fax: +81-438-52-3501

Received January 20, 2012; accepted May 8, 2012 (Edited by T. Demura)

Abstract The publication of the whole genome sequence of *Jatropha curcas* L. has contributed to the study of gene functions of this plant, especially in data-driven investigations such as transcriptome and proteome analyses. Metabolomics analyses of Jatropha have also been reported in recent years. However, the analytical tools for omics data from Jatropha are limited. We prepared a set of pathway maps where the predicted genes of Jatropha were assigned based on KEGG pathway maps, and developed an omics viewer named KaPPA-View4-Jatropha where the pathway maps were implemented. Out of 40,929 predicted genes, 8085 genes were mapped on the KEGG Metabolism maps, other KEGG maps, or gene category maps that were generated from gene classification data of KEGG BRITE. Two transcriptome datasets, four metabolome datasets and one gene co-expression dataset were registered in the viewer. To facilitate data sharing of unpublished omics data among research collaborators, we developed a local database system, KaPPA-Loader. These data mining environments and the supporting database system will help Jatropha researchers to discover key genes such as those involved in oil production, biosynthesis of toxic compounds, and stress resistance. KaPPA-View4-Jatropha and KaPPA-Loader are available from the KaPPA-View portal site (http://kpv.kazusa.or.jp/).

Key words: Correlation analysis, *Jatropha curcas* L., metabolic pathway viewer, metabolomics, transcriptomics.

Jatropha curcas L. is a tropical non-domesticated shrub that has considerable potential for multipurpose uses, including as a biofuel (Abdulla et al. 2011), in animal nutrition (Devappa et al. 2010; King et al. 2009) and for medical applications (Thomas et al. 2008). Several proteomics analyses and large-scale collection of expressed sequence tags (ESTs) have been reported, contributing to understanding the nature of metabolic regulation in cells, to the discovery of key genes involved in stress resistances, and to the development of biomarkers for molecular breeding (Sudhakar Johnson et al. 2011). Most recently, the whole genome of Jatropha was sequenced, and a total of 40,929 complete and partial structures of protein encoding genes were deduced (Sato et al. 2011). Transcriptome analyses based on the Jatropha genome information using next generation sequencers have emerged, from which much progress is expected for the discovery of molecular mechanisms of the production of oil and other materials. In addition, comprehensive analyses of metabolic compounds (metabolome analyses) have been performed in recent years (Sano et al., in this issue).

One way to intuitively understand biological significance in huge "omics" data are to visualize the data in metabolic pathway maps. Many tools and databases are available for such purposes, including MapMan, BioCyc, and Reactome (Gehlenborg et al. 2010). Tools based on the pathway maps provided by the Kyoto Encyclopedia of Genes and Genomes (KEGG) have also been reported (Kanehisa et al. 2012; Kono et al. 2006; Suhre and Schmitt-Kopplin 2008). We have developed an omics viewer system, KaPPA-View, for simultaneous representation of transcriptome (proteome) and metabolome data in metabolic pathway maps (Sakurai et

Abbreviations: EST, Expressed Sequence Tag.

This article can be found at http://www.jspcmb.jp/
Published online June 15, 2012

al. 2011; Tokimatsu et al. 2005). The latest version of the system, KaPPA-View4, uniquely visualizes gene-to-gene and metabolite-to-metabolite relationships such as gene co-expression on the pathway maps. All of these systems require the pathway map data on which the genes for the organisms are mapped. For most model plants and major crops such as *Arabidopsis*, rice, and tomato, the pathway map data are constructed and curated by the database providers or research communities of these plant species. However, as far as we know, pathway map data for Jatropha has not been developed yet.

Here, we report the preparation of pathway map data for Jatropha and their implementation in KaPPA-View4-Jatropha for omics analysis (http://kpv.kazusa.or.jp/kpv4-jat/). The predicted genes obtained from the genome sequencing of Jatropha were mapped onto the KEGG pathway maps. Some data of transcriptome, metabolome, and a set of gene-to-gene co-expression correlation data from Jatropha were registered. We also developed a database system named KaPPA-Loader for data sharing of private omics data between research collaborators.

To assign as many Jatropha genes as possible on the metabolic pathway maps, we used pathway map data from KEGG (http://www.genome.jp/kegg/), because KEGG provides a set of unified metabolic pathway maps for various organisms, including plants, animals, and microorganisms. The predicted protein-encoding genes from Jatropha that were not annotated as "/short" or "/TE" genes (40,929 out of 58,720, Sato et al. 2011) were subjected to BLASTP searches against the amino acid sequence data of the KEGG GENES database (Ogata et al. 1999), which was downloaded on December 20, 2011. Thresholds of amino acid sequence identity ≥25% and of length coverage of the query sequence ≥50% with a cut-off (*E*-value≤1e-10) were applied. For each Jatropha gene, a corresponding KEGG gene that had a KEGG Ontology (KO) ID and showed the highest bit score was selected. The Jatropha genes were assigned on the KEGG pathway maps at the places where the enzymes of the KO are drawn (Figure 1A). The gene mapping data provided by KEGG were used to create the pathway maps for *Arabidopsis thaliana*, *Oryza sativa* japonica, *Populus trichocarpa*, and *Ricinus communis*, whose organism codes in KEGG are ath, osa, pop, and rcu, respectively.

To investigate the functional differences of genes within gene families, such as transcription factors and cytochrome P450 monooxygenases, we created "gene category maps" where the gene symbols (squares) for each gene family are arranged and arrayed (Figure 1B). Data from the classification named "Genes and Proteins" in the collection of hierarchical classifications of KEGG (KEGG BRITE) was used to create the gene category maps. The maps were created by an in-house Java program by which the genes of the third lowest level of the hierarchy were drawn in a single map. The Jatropha genes were also assigned on the gene category maps by BLASTP searches and KO annotation as described above. The mapping data of KEGG was also used for the other plant species.

Out of 40,929 predicted genes of Jatropha, 8058 genes were assigned on 401 maps in total (Table 1). The gene numbers assigned on the Metabolism maps (3279 genes on 120 maps) were larger than those estimated by Sato et al. 2011 (2213 genes on 134 maps). As the same procedure was applied for the mapping, the increase of the assigned gene number is probably due to the updating of KEGG data. Larger numbers of Jatropha genes were assigned on both the pathway maps and the gene category maps compared to the other plant species. This could be due mainly to differences in the mapping procedure, i.e. BLASTP searches and assignments based on the KO were conducted for Jatropha, whereas the mapping data annotated and curated by KEGG was used for the other plants. Some Jatropha genes were assigned to maps of non-plant pathways such as "Retinol metabolism" and "Cardiac muscle contraction" (Supplemental Table S1). These might be mis-assignments, but we did not remove them. As it is possible that Jatropha has its own metabolic pathways that are not found in other model plants, providing more diverse possibilities could help researchers to discover novel gene functions.

We registered two gene expression datasets, four metabolome datasets, and one gene co-expression dataset of Jatropha in KaPPA-View4-Jatropha. The gene expression datasets were obtained as follows. Total RNA was prepared from hermaphrodite flowers and the early stage of developing fruits of *Jatropha curcas* L. accession IBP-1 grown in an experimental field at Bogor Agricultural University, Indonesia. RNA-seq analysis was performed by outsourcing to RIKEN GENESIS Co. Ltd., Yokohama, Japan. In the analysis, the RNA samples were pretreated to make library samples of approximately 200 base pairs (bp) and sequenced using a paired-end method with the Illumina GAIIx system (Illumina Inc., San Diego, CA) according to the manufacturer's instructions. Obtained read sequences were 75+75 bp from both sides of each sample sequence. The read sequences were directly (without assembling into contigs) mapped onto the *Jatropha curcas* genome sequence (Sato et al. 2011), and then the numbers of read sequences were counted per gene predicted in the genome dataset. The counts were normalized by the length of the genes to create a dataset of gene expression values. The values transferred to logarithm base 10 were registered to KaPPA-View4-Jatropha. The cosine correlation coefficients of gene co-expression of Jatropha were calculated by 6 transcriptome datasets (unpublished data). The correlation coefficients (≥0.99) between the 6898 genes assigned on the Metabolism pathway maps of

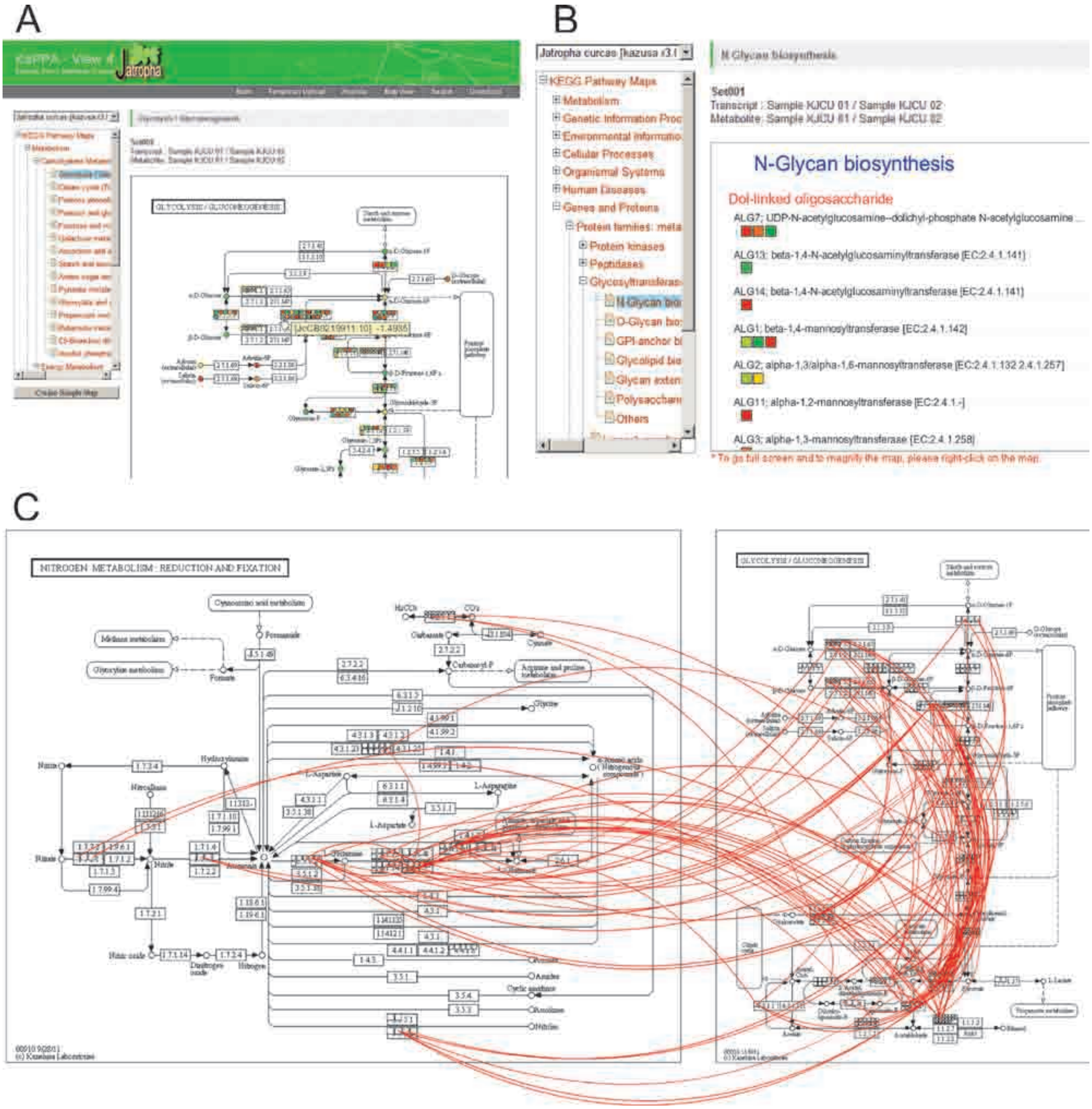
A
B
C
KaPPA - View 4 Jatropha
GLYCOLYSIS / GLUCONEOGENESIS
JcCB0215911.10 -1.4935
Jatropha curcas [kazusa r3.0]
KEGG Pathway Maps
Metabolism
Genetic Information Proc
Environmental Informatio
Cellular Processes
Organismal Systems
Human Diseases
Genes and Proteins
Protein families: meta
Protein kinases
Peptidases
Glycosyltransferas
N-Glycan bio
O-Glycan bio
GPI-anchor bi
Glycolipid bio
Glycan exten
Polysacchari
Others
N-Glycan biosynthesis
Set001
Transcript : Sample KJCU 01 / Sample KJCU 02
Metabolite : Sample KJCU 01 / Sample KJCU 02
N-Glycan biosynthesis
Dol-linked oligosaccharide
ALG7; UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosamine ...
ALG13; beta-1,4-N-acetylglucosaminyltransferase [EC:2.4.1.141]
ALG14; beta-1,4-N-acetylglucosaminyltransferase [EC:2.4.1.141]
ALG1; beta-1,4-mannosyltransferase [EC:2.4.1.142]
ALG2; alpha-1,3/alpha-1,6-mannosyltransferase [EC:2.4.1.132 2.4.1.257]
ALG11; alpha-1,2-mannosyltransferase [EC:2.4.1.-]
ALG3; alpha-1,3-mannosyltransferase [EC:2.4.1.258]
* To go full screen and to magnify the map, please right-click on the map.
NITROGEN METABOLISM : REDUCTION AND FIXATION
GLYCOLYSIS / GLUCONEOGENESIS

Figure 1. The KaPPA-View4-Jatropha system. (A) The assignment of Jatropha predicted genes on the KEGG pathway map and the representation of transcriptome and metabolome data on the map of "Glycolysis/Gluconeogenesis". (B) Omics data representation on the "gene category map" generated from the hierarchical classification of "Genes and Proteins" of the KEGG BRITE database. The map for "N-Glycan biosynthesis" is shown. (C) Overlay of gene co-expression correlation data on the pathway maps. The registered gene co-expression data are represented on the maps of "Nitrogen Metabolism: Reduction and fixation" and "Glycolysis/Gluconeogenesis". High-resolution images of Figure 1A, B, C are shown in the Supplemental Figure S1, S2, and S3.

KEGG were registered. From the metabolomics analysis reported by Sano et al. (in this issue), 71 compounds that were annotated and attributed to the corresponding KEGG IDs were selected and registered. The values of the compound accumulations were normalized as described in Sano et al. These data will be useful for practical Jatropha studies. In particular, gene co-expression analysis is a powerful tool for gene function prediction (Usadel et al. 2009). The unique function of KaPPA-View4 that visualizes the co-expression data between the pathway maps will help researchers to discover key genes such as those controlling the production of oil and other materials in Jatropha.

We developed a local database system, KaPPA-Loader, for omics data sharing. KaPPA-Loader will facilitate the utilization of unpublished omics data between research collaborators. Anyone can download the program, set it up on his/her own server, and start the KaPPA-Loader

Table 1. The number of genes and maps (parentheses) assigned onto maps of *Jatropha curcas* and other plant species

	Jatropha curcas		*Arabidopsis thaliana*		*Oryza sativa*		*Populus trichocarpa*		*Ricinus communis*		Map Number	
	genes[a]	maps[b]	genes	maps	genes	maps	genes	maps	genes	maps	assigned[c]	total[d]
KEGG PATHWAY Maps	6898	(282)	4025	(262)	3106	(254)	4322	(260)	2945	(261)	(285)	(328)
Metabolism	3279	(120)	2039	(113)	1594	(104)	2159	(107)	1506	(107)	(123)	(150)
Others	4319	(162)	2255	(149)	1744	(150)	2462	(153)	1656	(154)	(162)	(178)
Gene Category Maps[e]	2996	(119)	2058	(122)	1561	(115)	2166	(123)	1408	(120)	(139)	(365)
Total	8058	(401)	4775	(384)	3626	(369)	5085	(383)	3436	(381)	(424)	(693)

[a] The number of genes assigned on the maps were counted non-redundantly. [b] The number of maps where at least one gene of the species was assigned. [c] The number of maps where at least one gene from any of the five plants was assigned. [d] The total number of maps retrieved from the KEGG PATHWAY and generated from the KEGG BRITE database. [e] The gene category maps were generated from the hierarchical classification of "Gene and Proteins" of the KEGG BRITE database.

service with management of collaboration user accounts. The data deposited in KaPPA-Loader can be visualized and analyzed on the KaPPA-View4 system. Publication of the deposited data can be done easily by setting data access permissions. The KaPPA-Loader was developed with Java 6, Tomcat 5.5, and MySQL 5, and requires a server system that serves this middleware. The KaPPA-Loader is freely available at http://kpv.kazusa.or.jp/.

As far as we know, KaPPA-View4-Jatropha is the only tool that provides a pathway analysis environment for the omics data of Jatropha. At present, no gene expression data from Jatropha have been deposited in the public data repositories Gene Expression Omnibus (GEO) (Barrett et al. 2011) or ArrayExpress (Parkinson et al. 2011). However, it is apparent that gene expression data will increasingly become public with the availability of genome sequences, the number of custom microarray services, and the popularization of the RNA-Seq analyses by next generation sequencers. Metabolomics analyses are attracting much attention in studies of the unique metabolic pathways of Jatropha (Sano et al., in this issue). KaPPA-View4-Jatropha will contribute to Jatropha studies based on these omics data.

We will continue to update the data on KaPPA-View4-Jatropha upon the updating of KEGG maps, Jatropha genome annotations, and the release of omics data.

Acknowledgements

This research was supported in part by the New Energy and Industrial Technology Development Organization (NEDO), Japan, in part by the Bilateral Joint Project of the Japan Society of the Promotion of Science (JSPS), and in part by the Kazusa DNA Research Institute Foundation.

References

Abdulla R, Chan ES, Ravindra P (2011) Biodiesel production from *Jatropha curcas*: a critical review. *Crit Rev Biotechnol* 31: 53–64

Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Holko M, Ayanbule O, Yefanov A, Soboleva A (2011) NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res* 39: D1005–D1010

Devappa RK, Makkar HP, Becker K (2010) Nutritional, biochemical, and pharmaceutical potential of proteins and peptides from jatropha: review. *J Agric Food Chem* 58: 6543–6555, review

Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D, Gavin AC (2010) Visualization of omics data for systems biology. *Nat Methods* 7: S56–S68

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40: D109–D114

King AJ, He W, Cuevas JA, Freudenberger M, Ramiaramanana D, Graham IA (2009) Potential of *Jatropha curcas* as a source of renewable oil and animal feed. *J Exp Bot* 60: 2897–2905

Kono N, Arakawa K, Tomita M (2006) MEGU: pathway mapping web-service based on KEGG and SVG. *In Silico Biol* 6: 621–625

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 27: 29–34

Parkinson H, Sarkans U, Kolesnikov N, Abeygunawardena N, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Holloway E, Kurbatova N, Lukk M, Malone J, Mani R, Pilicheva E, Rustici G, Sharma A, Williams E, Adamusiak T, Brandizi M, Sklyar N, Brazma A (2011) ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res* 39: D1002–D1004

Sakurai N, Ara T, Ogata Y, Sano R, Ohno T, Sugiyama K, Hiruta A, Yamazaki K, Yano K, Aoki K, Aharoni A, Hamada K, Yokoyama K, Kawamura S, Otsuka H, Tokimatsu T, Kanehisa M, Suzuki H, Saito K, Shibata D (2011) KaPPA-View4: a metabolic pathway database for representation and analysis of correlation networks of gene co-expression and metabolite co-accumulation and omics data. *Nucleic Acids Res* 39: D677–D684

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N, Takahashi C, Nakayama S, Kishida Y, Kohara M, Yamada M, Tsuruoka H, Sasamoto S, Tabata S, Aizu T, Toyoda A, Shin-i T, Minakuchi Y, Kohara Y, Fujiyama A, Tsuchimoto S, Kajiyama S, Makigano E, Ohmido N, Shibagaki N, Cartagena JA, Wada N, Kohinata T, Atefeh A, Yuasa S, Matsunaga S, Fukui K (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Res* 18: 65–76

Sudhakar Johnson T, Eswaran N, Sujatha M (2011) Molecular approaches to improvement of *Jatropha curcas* Linn. as a sustainable energy crop. *Plant Cell Rep* 30: 1573–1591

Suhre K, Schmitt-Kopplin P (2008) MassTRIX: mass translator into pathways. *Nucleic Acids Res* 36: W481-4

Thomas R, Sah NK, Sharma PB (2008) Therapeutic biology of *Jatropha curcas*: a mini review. *Curr Pharm Biotechnol* 9: 315–324

Tokimatsu T, Sakurai N, Suzuki H, Ohta H, Nishitani K, Koyama T, Umezawa T, Misawa N, Saito K, Shibata D (2005) KaPPA-view: a web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol* 138: 1289–1300

Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ* 32: 1633–1651