

KaPPA-View4: a metabolic pathway database for representation and analysis of correlation networks of gene co-expression and metabolite co-accumulation and omics data

Nozomu Sakurai^{1,*}, Takeshi Ara¹, Yoshiyuki Ogata¹, Ryosuke Sano¹, Takashi Ohno¹, Kenjiro Sugiyama¹, Atsushi Hiruta¹, Kiyoshi Yamazaki¹, Kentaro Yano², Koh Aoki¹, Asaph Aharoni³, Kazuki Hamada², Koji Yokoyama², Shingo Kawamura², Hirofumi Otsuka², Toshiaki Tokimatsu⁴, Minoru Kanehisa⁴, Hideyuki Suzuki¹, Kazuki Saito^{5,6} and Daisuke Shibata¹

¹Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, ²Bioinformatics Laboratory, Department of Life Sciences, School of Agriculture, Meiji University, 1-1-1, Higashi-Mita, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan, ³Department of Plant Sciences, Weizmann Institute of Science, POB 26, Rehovot 76100, Israel, ⁴Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, ⁵Graduate School of Pharmaceutical Sciences, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba, Chiba 263-8522 and ⁶RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Received August 14, 2010; Revised October 4, 2010; Accepted October 5, 2010

ABSTRACT

Correlations of gene-to-gene co-expression and metabolite-to-metabolite co-accumulation calculated from large amounts of transcriptome and metabolome data are useful for uncovering unknown functions of genes, functional diversities of gene family members and regulatory mechanisms of metabolic pathway flows. Many databases and tools are available to interpret quantitative transcriptome and metabolome data, but there are only limited ones that connect correlation data to biological knowledge and can be utilized to find biological significance of it. We report here a new metabolic pathway database, KaPPA-View4 (<http://kpv.kazusa.or.jp/kpv4/>), which is able to overlay gene-to-gene and/or metabolite-to-metabolite relationships as curves on a metabolic pathway map, or on a combination of up to four maps. This representation would help to discover, for example, novel functions of a transcription factor that regulates genes on a metabolic pathway. Pathway maps of the Kyoto Encyclopedia of Genes and Genomes

(KEGG) and maps generated from their gene classifications are available at KaPPA-View4 KEGG version (<http://kpv.kazusa.or.jp/kpv4-kegg/>). At present, gene co-expression data from the databases ATTED-II, COXPRESdb, CoP and MiBASE for human, mouse, rat, *Arabidopsis*, rice, tomato and other plants are available.

INTRODUCTION

A vast amount of DNA microarray data has been accumulated in public repositories, contributing to studies such as investigations of gene expression patterns and discoveries of marker transcripts. Furthermore, correlation coefficients as indicators of co-expression between genes, calculated from transcriptome data from a large variety of tissues and cell conditions, and analyses of gene co-expression networks whose member genes are connected with high correlation coefficients, have contributed to the discovery of novel functions of genes (1–3). For instance, the functions of key genes in the biosynthesis of glucosinolates (4), cell walls (5) and flavonoids (6) in plants were successfully uncovered by co-expression analyses. Recent advances in co-expression analysis include the

*To whom correspondence should be addressed. Tel: +81 438 52 3947; Fax: +81 438 52 3948; Email: sakurai@kazusa.or.jp

introduction of the mutual rank (MR) as a better index of co-expression (7), and new strategies to extract networks whose member genes are densely and specifically co-expressed (8,9). Meanwhile, efforts to obtain comprehensive quantitative data of metabolites in organisms (metabolomes) have also been made by chromatography–mass spectrometry and nuclear magnetic resonance (10,11), and the number of public databases in which metabolome data are compiled has been increasing (12). Correlation analyses of metabolite-to-metabolite co-accumulations have also been made (13–15). Müller-Linow *et al.* (16) suggested that the distance between metabolites on a metabolic pathway was not related to the co-accumulation correlation. Camacho *et al.* (17) reported several factors that relate to how metabolites co-accumulate. Correlations of metabolite co-accumulations are thought to be useful for understanding the regulation of metabolic pathways when they are comparatively analyzed between different sets of data sources (18).

So far, numerous databases and tools have been developed to understand the biological significance of transcriptome and metabolome data. Tools for functional categorization of genes, referred to as gene ontology (GO), have been published (19). To promote a more intuitive understanding, databases and tools that project data onto metabolic pathway maps have been developed (20). As a source for pathway maps, data from the Kyoto Encyclopedia of Genes and Genomes (KEGG) have been used in many tools, including KegArray provided by KEGG itself (<http://www.genome.jp/kegg/download/kegtools.html>) (21–23). We have also developed and upgraded a web-based viewer, KaPPA-View, which can represent transcriptome and metabolome data simultaneously on a single pathway map (24–26), and it has contributed to several plant studies (27–29). In contrast to the quantity of software available for analyzing transcriptome and metabolome quantitative data, only a few tools have so far been developed for studying gene-to-gene and metabolite-to-metabolite correlation data. In ATTED-II (30) and COXPRESdb (2), genes mapped onto KEGG pathways are highlighted on a gene network figure that is represented according to a user's query gene. MapMan can represent the correlation value for each gene on a map against a query gene (31). Both systems only represent correlations for a single gene, and there are no databases or tools that can overview all the gene-to-gene relationships on the pathway maps. Furthermore, no tools for projecting metabolite-to-metabolite correlations onto pathway maps have been developed so far.

Here, we report a novel database system, KaPPA-View4, which enables the overlay of correlation data onto pathway maps. After reporting the first version of KaPPA-View for an *Arabidopsis*-specific 'omics' data viewer (24), we added functions for correlation representation in version 2, and for multiple map representation and for managing multiple organisms in version 3. To meet the demands of dealing with multiple and huge sets of correlation data in recent years, we reviewed the architecture of the database system, totally reconstructed the web application, especially for the pathway display module, and released the newest version, KaPPA-View4,

as an 'omics' database that can immediately display results. The 'omics' viewer functions were extended to accept data uploading from outside systems. As far as we know, this version is a unique database system that can represent all gene-to-gene and metabolite-to-metabolite correlations on metabolic pathway maps.

MATERIALS AND METHODS

Implementation

We are currently providing two versions of the system. One is KaPPA-View4 Classic (<http://kpv.kazusa.or.jp/kpv4/>) whose pathway maps are based on the traditional KaPPA-View maps of *Arabidopsis* (24). Another is KaPPA-View4 KEGG (<http://kpv.kazusa.or.jp/kpv4-kegg/>), where pathway maps are acquired from KEGG. KaPPA-View4 Classic is primarily for plant scientists, implementing information for genome-sequenced plant species: *Arabidopsis*, rice and *Lotus japonicus*; and for plant species whose Affymetrix GeneChips and correlation data are available: tomato, soybean, barley, poplar, wheat, wine grape and maize. KaPPA-View4 KEGG is for general users, including animal, microorganism and also plant scientists. One of the major advantages of KaPPA-View4 KEGG is that we can share the latest results of KEGG's continuous effort of curation of gene descriptions, categorizations and assignments on the maps. At present, we provide information for 15 species, namely human, mouse, rat, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis*, rice, poplar, castor bean, sorghum, wine grape, maize, *Physcomitrella patens* subsp. *patens*, *Escherichia coli* and budding yeast. Details of the basic information of the default installed data are shown in Supplementary Data Section 1. Both versions are freely available.

Preparation of map data

In KaPPA-View4 Classic, the map data is based on approximately 150 leaves of *Arabidopsis* pathway maps generated for the initial version (24). The assignment of genes for the other plant species was done by sequence homology searches by the blastx or blastp program against *Arabidopsis* amino acid sequences TAIR9_pep provided by The Arabidopsis Information Resource (TAIR, <http://www.arabidopsis.org/>). The best hit *Arabidopsis* genes for each gene or target sequence of Affymetrix GeneChip probes of the other species were defined as those having the minimum *e*-value among candidates with an *e*-value threshold of 1×10^{-30} . Then the genes or probes were associated with the enzymatic reactions where the best hit *Arabidopsis* genes are assigned. For *Lotus japonicus*, two extra pathway maps for isoflavonoid biosynthesis (I and II) were added and the gene assignments on the maps are manually performed.

A pipeline was constructed by Java (www.java.com) to download files from the KEGG FTP (file transfer protocol) and generate all the data files required for the KaPPA-View4 KEGG setup including gene, compound and reaction information and pathway map data. The pathway maps were arranged as a tree structure for each

organism according to the KEGG BRITE hierarchy, and bird's eye maps (24) were generated for the nodes of the tree. The hierarchically categorized gene family data 'Genes and Proteins' in the KEGG BRITE was added to the tree where each family was represented as a map and its member genes were arrayed on the map (Figure 2, top-right map). During setup of the version on 14 June 2010, 290 and 379 sheets of metabolic pathway maps and gene family maps, respectively, were installed.

Preparation of gene-to-gene correlation data

Table 1 summarizes the gene co-expression data we provide in the KaPPA-View4 system by default. The data files were downloaded from ATTED-II and COXPRESdb or provided by MiBASE (32) and CoP (33). In the case of the co-expression data described in Pearson's correlation coefficients (PCC), strong positive and negative correlations were filtered at threshold values of 0.6 and -0.6 . In the case of mutual rank (MR), only positive values ≤ 100 were selected. In the case of CoP data, significant gene-to-gene relationships evaluated as 'Co-expressed Genes' (33) were extracted, and the co-expression values were described as cosign theta correlation coefficients.

In KaPPA-View4 KEGG, the gene IDs originated from the entry IDs of KEGG. In the case of rice, KEGG entry IDs were derived from Entrez Gene IDs from the National Center for Biotechnology Information (NCBI,

<http://www.ncbi.nlm.nih.gov/>), while the co-expression data from ATTED-II and CoP were written as gene IDs from The Rice Annotation Project Database (RAP-DB, <http://rapdb.dna.affrc.go.jp/>). Therefore, the gene IDs in the co-expression data were converted to NCBI Entrez Gene IDs according to the gene information file of KEGG for rice.

System development

The web application of the KaPPA-View4 system was developed with Java 1.6 (www.java.com), Tomcat 5.5 (tomcat.apache.org), MySQL 5.0 (dev.mysql.com) and Flex SDK 3 (opensource.adobe.com/wiki/display/flexsdk/). We chose the Adobe Flash Player (ver. 9.0 or higher) to display the map representations on Internet browsers. All the map data on KaPPA-View4 were prepared in the SVG (scalable vector graphics) format, and the map data are converted to Flash objects when displayed. Therefore, the SVG Viewer plug-in that had been required for previous versions is no longer needed for KaPPA-View4. This modification has remarkably improved the processing speeds of map representations, thus avoiding longer rendering time on the SVG Viewer plug-ins, especially for the maps where a lot of objects were drawn, e.g. maps in the Multiple Map mode (mentioned later) with a lot of genes and correlation curves. It also improved compatibility with Internet browsers and operating systems.

Table 1. Gene-to-gene co-expression correlation data installed in KaPPA-View4

Species	Source	No. of Chips ^a	Threshold	Data name in KaPPA-View4
for both KaPPA-View4 Classic and KaPPA-View4 KEGG				
<i>Arabidopsis thaliana</i>	ATTED-II, ver. c4.1	1388	MR ^b ≤ 100 PCC ^c ≥ 0.6 PCC ≤ -0.6	Ath: ATTED-II c4.1 (1388 chips) MR ≤ 100 Ath: ATTED-II c4.1 (1388 chips) PCC ≥ 0.6 Ath: ATTED-II c4.1 (1388 chips) PCC ≤ -0.6
<i>Oryza sativa</i>	CoP	5257	–	Ath: CoP (5257 chips) CCC ^d Co-expressed Genes ^e
	ATTED-II, ver. c1.0 ^f	208	MR ≤ 100	Osa: ATTED-II c1.0 (208 chips) MR ≤ 100
	CoP ^f	432	–	Osa: CoP (432 chips) CCC Co-expressed Genes
for only KaPPA-View4 Classic				
<i>Solanum lycopersicum</i>	MiBASE	50	MR ≤ 100	Sly: MiBASE (50 chips) MR ≤ 100
<i>Glycine max</i>	CoP	2994	–	Gma: CoP (2994 chips) CCC Co-expressed Genes
<i>Hordeum vulgare</i>	CoP	403	–	Hvu: CoP (403 chips) CCC Co-expressed Genes
<i>Populus trichocarpa</i>	CoP	95	–	Ptr: CoP (95 chips) CCC Co-expressed Genes
<i>Triticum aestivum</i>	CoP	368	–	Tae: CoP (368 chips) CCC Co-expressed Genes
<i>Vitis vinifera</i>	CoP	210	–	Vvi: CoP (210 chips) CCC Co-expressed Genes
<i>Zea mays</i>	CoP	263	–	Zma: CoP (263 chips) CCC Co-expressed Genes
for only KaPPA-View4 KEGG				
<i>Homo sapiens</i>	COXPRESdb, ver. c3.1	4401	MR ≤ 100 PCC ≥ 0.6 PCC ≤ -0.6	Hsa: COXPRESdb c3.1 (4401 chips) MR ≤ 100 Hsa: COXPRESdb c3.1 (4401 chips) PCC ≥ 0.6 Hsa: COXPRESdb c3.1 (4401 chips) PCC ≤ -0.6
<i>Mus musculus</i>	COXPRESdb, ver. c2.1	2226	MR ≤ 100 PCC ≥ 0.6 PCC ≤ -0.6	Mmu: COXPRESdb c2.1 (2226 chips) MR ≤ 100 Mmu: COXPRESdb c2.1 (2226 chips) PCC ≥ 0.6 Mmu: COXPRESdb c2.1 (2226 chips) PCC ≤ -0.6
<i>Rattus norvegicus</i>	COXPRESdb, ver. c1.1	632	MR ≤ 100 PCC ≥ 0.6 PCC ≤ -0.6	Rno: COXPRESdb c1.1 (632 chips) MR ≤ 100 Rno: COXPRESdb c1.1 (632 chips) PCC ≥ 0.6 Rno: COXPRESdb c1.1 (632 chips) PCC ≤ -0.6

^aNumber of DNA microarray chips for correlation calculations.

^bMR: Mutual rank defined by Obayashi *et al.* (7).

^cPCC: Pearson's correlation coefficient.

^dCCC: Cosign theta correlation coefficient.

^eAll significant gene-to-gene relationships evaluated as 'Co-expressed Genes' in CoP (33).

^fThe gene IDs in the original data were converted to NCBI Entrez Gene IDs for KaPPA-View4 KEGG (see 'Materials and Methods' section).

Both KaPPA-View4 Classic and KaPPA-View4 KEGG run on a Red Hat Enterprise Linux ES v.4 server equipped with a 3.2 GHz 64 bit Intel Xeon CPU and 2 GB RAM.

WEB APPLICATION

Representation of co-expression data

Gene-to-gene co-expression data are represented as curves on pathway maps (colored in red in default) after selecting the kind of data listed in the control panel placed below the pathway map, displayed in the Map View function, after selecting the species and pathway maps to view (Figure 1a). All correlations filtered by the control panel are seen here. As shown in Figure 1a, the program finds that a lot of the genes in the Calvin Cycle are co-expressed in *Arabidopsis*. Pathway maps having dense correlations are distinguished on the bird's eye map set to correlation mode (Figure 1b). When the species is switched to tomato and the correlation data is set to the tomato, we see that the genes in the Calvin Cycle are densely connected too, implying that this is a general expression tendency across plant species (Figure 1c, see 'Discussion' section). Metabolite-to-metabolite co-accumulations can also be overlaid on pathway maps simultaneously with gene co-expression and 'omics' data (Figure 1d, curves in blue). Users can utilize not only the default correlation data but also user-created ones (as described later).

Correlations across the maps

The KaPPA-View4 system enables up to four maps to be displayed in a single browser window (Multiple Map mode) with correlation curves represented, resulting in visualization of correlations across the maps (Figure 2). Users can choose the combination of the pathway maps for the Multiple Map mode. In KaPPA-View4 KEGG, gene family maps can be included too (Figure 2, top-right panel). This greatly facilitates analysis of non-metabolic pathway genes such as protein kinases and transcription factors that are not available in other KEGG-based programs developed so far. Furthermore, user-created SVG maps (Figure 2, bottom-right) and simple maps where genes for user-input gene IDs are arrayed (Figure 2, bottom-left) are included in the representation in the Multiple Map mode. Therefore, correlation representations in the Multiple Map mode help researchers primarily to understand relationships between genes and metabolic pathways, which could also be described as functional modules of biological systems. As exemplified in Supplementary Data Section 2, it is possible to find potential genes regulating certain metabolic pathways by representing correlation curves between the gene family maps and the pathway maps (see 'Discussion' section).

Utilization of user's data

Users are allowed to upload their own correlation data, as well as their experimental data and map data, and utilize them for their own analyses. Therefore, users can analyze not only the correlation data calculated from the large

variety of sample conditions, but also data obtained from sets of specific treatments to focus on treatment-dependent responses. This functional extension should especially help analyze metabolite correlations that are presently of limited public availability. Furthermore, users can freely create their own user accounts to save these data on the KaPPA-View4 server, which allows users to start analyses immediately after login.

'Omics' data comparison between multiple species

As an extra mode for map representations, the KaPPA-View4 system has a 'Universal Map' mode where genes from multiple species are drawn simultaneously on a single map. This mode helps to recognize the differences in gene assignments and orthologous genes between the species selected. In this mode, transcriptome and metabolome data from different species can be compared on the maps (Figure 3). This should help study the diversity of paralogs and their functional differences between species by comparing omics data obtained from different species under similar physiological conditions.

Viewing the data on outside systems

The ability to represent transcriptome and metabolome data on the pathway maps was extended to an API (application programming interface), which is accessible from outside systems. Databases and application programs that store omics data can upload their data to KaPPA-View4 by the API and view it directly from their systems without logging in to KaPPA-View4 through a web user interface. As an example, the microarray data stored in MiBASE are immediately represented on KaPPA-View4 when users click the 'view' button in the data list of MiBASE (Supplementary Data Section 3). Since the session is kept alive as long as the browser is running, the uploaded data are stacked as temporary data on the KaPPA-View4 server. This allows users to set and compare other pairs of data.

DISCUSSION

Correlation analyses on KaPPA-View4

KaPPA-View4 is a unique and novel database system that helps users study the biological significance of correlation data by overlaying it onto pathway maps. ATTED-II (30), COXPRESdb (2) and MapMan (31) can represent gene-to-pathway relationships, but the representations are restricted to a single query gene. One advantage of these systems is that users can get all the information for the query gene. In contrast, in KaPPA-View4, whose subject is pathway information, all the correlations are represented on the pathway maps, although the representations are restricted to the selected pathways. As far as we know, KaPPA-View4 is the only database that can overview all the correlations on pathway maps.

One potential application of representing data in the KaPPA-View4 style is hypothesis generation. We found

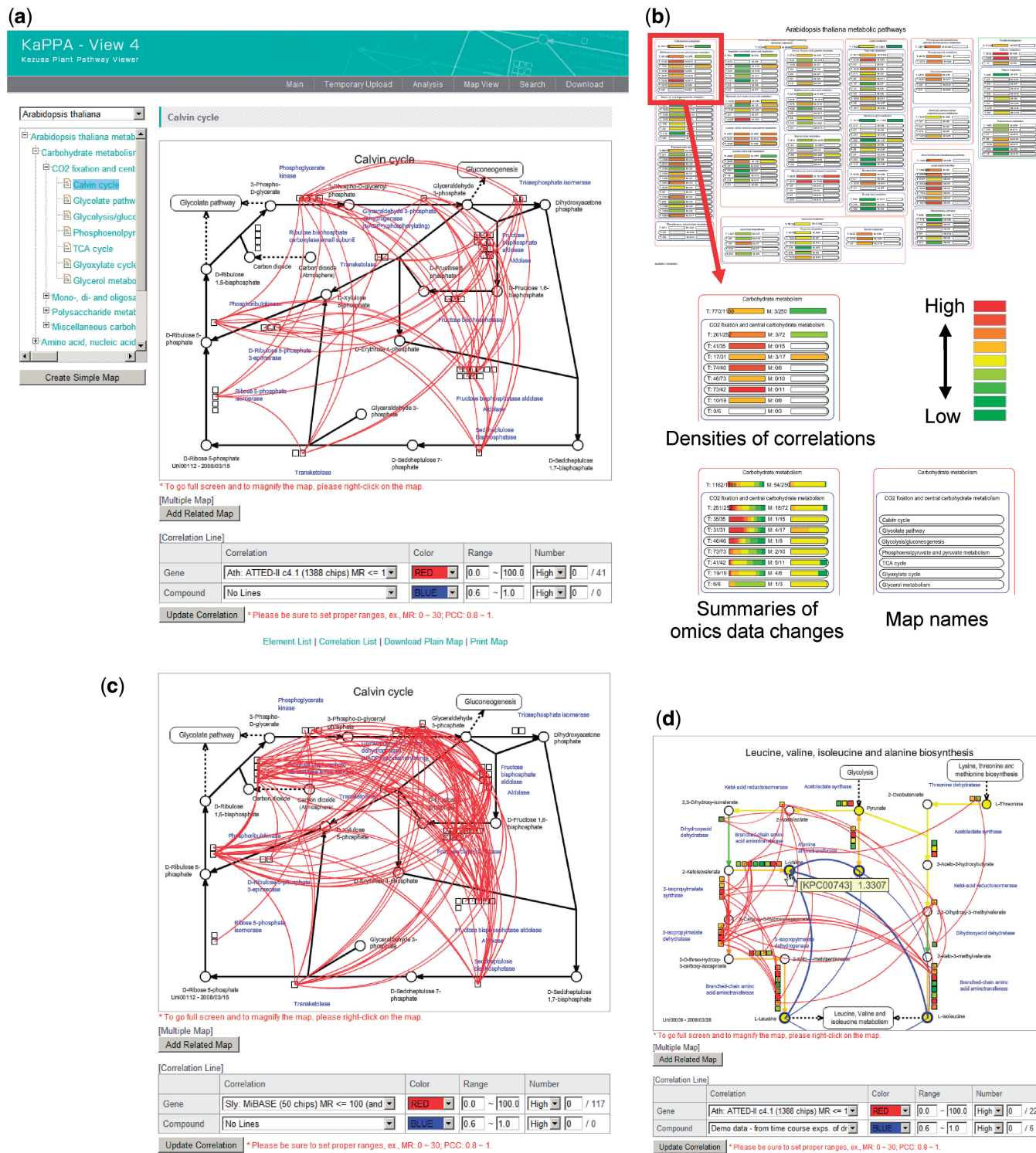


Figure 1. Representation of correlation networks on metabolic pathway maps in KaPPA-View4 Classic. **(a)** Curves representing gene-to-gene correlations (red) are overlaid on the pathway map of the Calvin Cycle in *Arabidopsis*. Co-expression data from ATTED-II ($MR \leq 100$) was selected. The pathway tree for selecting a map to view is placed on the left. **(b)** A Bird's Eye Map of KaPPA-View4 is displayed by clicking a node of the pathway tree. Nodes (pathway groups) are depicted as round rectangles and each metabolic pathway map included in a group is indicated as a round rectangle. The density of the correlation curves for each map is represented by a color scale. Summaries of the omics data and pathway names of the bars can also be shown. **(c)** Map of the Calvin Cycle for tomato overlaid with the gene co-expression data of tomato. Co-expression data from MiBASE ($MR \leq 100$) was selected. One to several genes for most of the reactions were co-expressed with each other as with those in *Arabidopsis* shown in (a). **(d)** Metabolite-to-metabolite correlations can be represented on the maps simultaneously with gene co-expression, transcriptome and metabolome data. The sample data installed in KaPPA-View4 Classic are represented on the map of leucine, valine, isoleucine and alanine biosynthesis. The detail of how the figures are produced on the KaPPA-View4 system is shown in Supplementary Data Section 4.

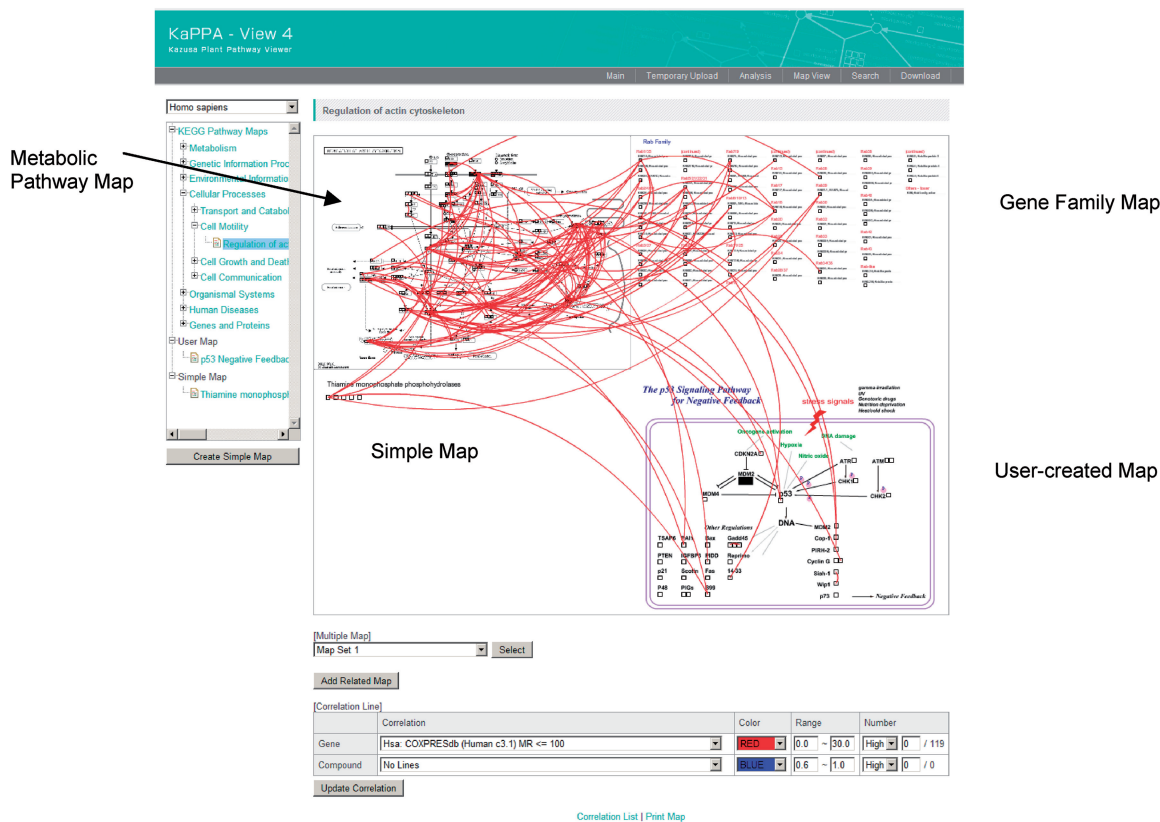


Figure 2. Representation of correlation networks on a combination of four maps (Multiple Map mode) in KaPPA-View4 KEGG. The correlation curves are drawn across the maps. The gene co-expression data from COXPRESdb for human (c3.1, MR ≤ 30) was selected. The map 'Regulation of actin cytoskeleton' from the KEGG Pathway was placed at the top-left as an example of a metabolic pathway map. The map for Rab family proteins is an example of a gene family map generated from the KEGG BRITE classification (top-right). A Simple map was created by the gene IDs for thiamine monophosphate phosphohydrolases (bottom-left). A user map depicting p53 negative feedback (bottom-right) was created by the free SVG drawing software Inkscape 0.47 (www.inkscape.org). The detail of how the figure is produced on the KaPPA-View4 system is shown in Supplementary Data Section 4.

that one of the genes classified as a helix-turn-helix type transcription factor in human is related to many of the genes on the map of Hypertrophic cardiomyopathy (HCM) (Supplementary Data Section 2). The gene, CSRP3 (cysteine and glycine-rich protein 3, cardiac LIM protein, NCBI Entrez Gene ID: 8048), is also highly co-expressed with a series of genes on the maps of Cardiac muscle contraction, Dilated cardiomyopathy (DCM) and Cell Communication/Tight junction (Supplementary Figure S2b). There is a recent report showing that CSRP3 is indeed involved in HCM and DCM (34). Although the function of the CSRP3 gene has already been reported and the gene might not act as transcription factor in this case, this example clearly shows that representing co-expression networks on pathway maps is a powerful approach to discovering novel gene functions.

Another advantage of the KaPPA-View4 style is shown in Figure 1 where a similarity in gene co-expression networks is observed in the Calvin Cycle between *Arabidopsis* and tomato. Differences in networks concerning the genes of ribulose biphosphate carboxylases and triosephosphate isomerases were observed too. To depict the similarities and differences of the networks using

currently existing databases and tools that depend on query genes, we would need to obtain correlation data for all the genes in the Calvin Cycle for each species, then re-categorize the data according to the gene assignments to the reactions, and compile them. As KaPPA-View4 is a pathway map oriented system, we can intuitively understand the nature of the relationships and discuss the significance of the similarities and differences observed. This is also an advantage when comparing the differences between metabolite-to-metabolite co-accumulation data sets, as comparative correlation analyses have been suggested to be useful for understanding the regulation of metabolic flows (18).

As described here, pathway map-oriented representation of correlation networks in KaPPA-View4 helps researchers to understand biological systems and to generate working hypotheses to advance their studies, and powerfully complements existing databases and tools such as ATTED-II, COXPRESdb and MapMan.

Future developments

We have constructed a pipeline to prepare all the information files required for KaPPA-View4 KEGG. However, manual operations are still required for updating the

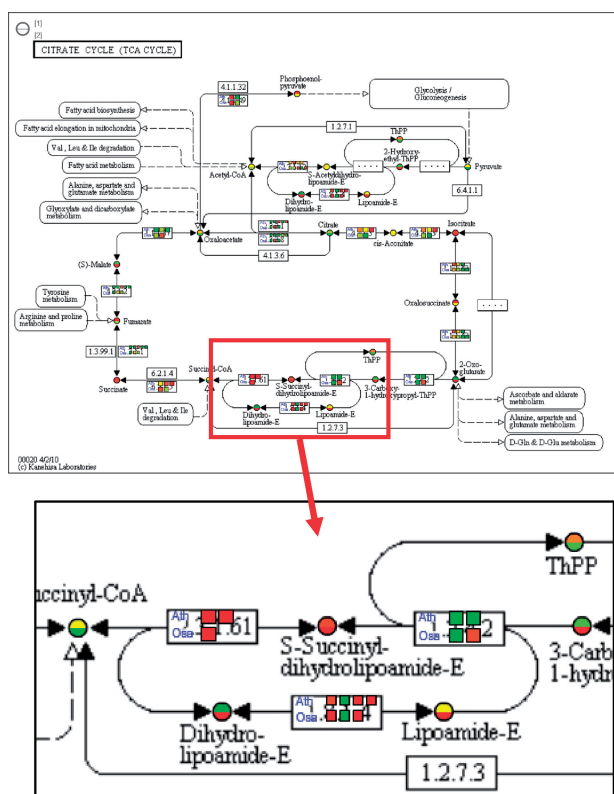


Figure 3. Comparative representation of transcriptome and metabolome data between *Arabidopsis* and rice in the Universal Map mode. The species to be represented on the maps can be selected using the ‘Select Species’ button below the pathway diagram. The data for both species were randomly generated as examples, and the values are represented as colors. In the rectangles for the enzyme reactions, both of the genes from *Arabidopsis* (Ath) and rice (Osa) are represented. The metabolome data from *Arabidopsis* is represented in the upper part of the circles for the metabolites, and the data from rice is in the lower part. The detail of how the figure is produced on the KaPPA-View4 system is shown in Supplementary Data Section 4.

files. As the KEGG data is updated once a week, we are now developing a program to automate this procedure so that the data on KaPPA-View4 KEGG are always kept up to date.

AVAILABILITY

Both KaPPA-View4 Classic and KaPPA-View4 KEGG are freely accessible to the public. User accounts to save the user’s data can be created freely. The KaPPA-View4 system program is freely available for academic users upon request. All the information files for genes, compounds, enzyme reactions and pathway map hierarchies are available from the Download function in the web application. The pathway map data in SVG can be downloaded from the map representation window. We will support the updating and editing of default data according to users’ requirements.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Most of the system programming was done by Axiohelix Co. Ltd. (www.axiohelix.com/en). The authors thank Dr Toshio Aoki of Nihon University for preparation of pathway maps of isoflavonoids biosynthesis of *Lotus japonicus*. We are grateful to all the people who had been engaged in the PMPj for preparation of the pathway maps, testing of the system and making useful comments.

FUNDING

New Energy and Industrial Technology Development Organization (NEDO, Japan) as part of a project entitled the ‘Development of Fundamental Technologies for Controlling the Material Production Process of Plants’ (PMPj) [P02001]; NEDO as part of a project named the ‘New Energy Technology Development’ [P07015]; Japan Society for the Promotion of Science (JSPS), Grants-in-Aid for Scientific Research (B) [2038022 to K.Y.]; JSPS, Grants-in-Aid for Special Research on Priority Areas [19043015 and 21024010 to K.Y.]. Funding for open access charge: Kazusa DNA Research Institute Foundation.

Conflict of interest statement. None declared.

REFERENCES

- Aoki,K., Ogata,Y. and Shibata,D. (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.*, **48**, 381–390.
- Obayashi,T., Hayashi,S., Shibaoka,M., Saeki,M., Ohta,H. and Kinoshita,K. (2008) COXPRESdb: a database of coexpressed gene networks in mammals. *Nucleic Acids Res.*, **36**, D77–D82.
- Usadel,B., Obayashi,T., Mutwil,M., Giorgi,F.M., Bassel,G.W., Tanimoto,M., Chow,A., Steinhäuser,D., Persson,S. and Provar,N.J. (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ.*, **32**, 1633–1651.
- Hirai,M.Y., Sugiyama,K., Sawada,Y., Tohge,T., Obayashi,T., Suzuki,A., Araki,R., Sakurai,N., Suzuki,H., Aoki,K. *et al.* (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl Acad. Sci. USA*, **104**, 6478–6483.
- Persson,S., Wei,H., Milne,J., Page,G.P. and Somerville,C.R. (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc. Natl Acad. Sci. USA*, **102**, 8633–8638.
- Yonekura-Sakakibara,K., Tohge,T., Niida,R. and Saito,K. (2007) Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in *Arabidopsis* by transcriptome coexpression analysis and reverse genetics. *J. Biol. Chem.*, **282**, 14932–14941.
- Obayashi,T. and Kinoshita,K. (2009) Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. *DNA Res.*, **16**, 249–260.
- Ogata,Y., Sakurai,N., Suzuki,H., Aoki,K., Saito,K. and Shibata,D. (2009) The prediction of local modular structures in a co-expression network based on gene expression datasets. *Genome Inform.*, **23**, 117–127.
- Altaf-Ul-Amin,M., Shinbo,Y., Mihara,K., Kurokawa,K. and Kanaya,S. (2006) Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinformatics*, **7**, 207.
- Hall,R.D. (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytol.*, **169**, 453–468.

11. Wishart,D.S. (2007) Current progress in computational metabolomics. *Brief Bioinform.*, **8**, 279–293.
12. Tohge,T. and Fernie,A.R. (2009) Web-based resources for mass-spectrometry-based metabolomics: A user's guide. *Phytochemistry*, **70**, 450–456.
13. Weckwerth,W., Loureiro,M.E., Wenzel,K. and Fiehn,O. (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc. Natl Acad. Sci. USA*, **101**, 7809–7814.
14. Szymanski,J., Jozefczuk,S., Nikoloski,Z., Selbig,J., Nikiforova,V., Catchpole,G. and Willmitzer,L. (2009) Stability of metabolic correlations under changing environmental conditions in *Escherichia coli* - a systems approach. *PLoS One*, **4**, e7441.
15. Bradley,P.H., Brauer,M.J., Rabinowitz,J.D. and Troyanskaya,O.G. (2009) Coordinated concentration changes of transcripts and metabolites in *Saccharomyces cerevisiae*. *PLoS Comput. Biol.*, **5**, e1000270.
16. Müller-Linow,M., Weckwerth,W. and Hütt,M.T. (2007) Consistency analysis of metabolic correlation networks. *BMC Syst. Biol.*, **1**, 44.
17. Camacho,D., De la Fuente,A. and Mendes,P. (2005) The origin of correlations in metabolomics data. *Metabolomics*, **1**, 53–63.
18. Morgenthal,K., Weckwerth,W. and Steuer,R. (2006) Metabolomic networks in plants: transitions from pattern recognition to biological interpretation. *Biosystems*, **83**, 108–117.
19. Khatri,P. and Drăghici,S. (2005) Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, **21**, 3587–3595.
20. Gehlenborg,N., O'Donoghue,S.I., Baliga,N.S., Goesmann,A., Hibbs,M.A., Kitano,H., Kohlbacher,O., Neuweger,H., Schneider,R., Tenenbaum,D. *et al.* (2010) Visualization of omics data for systems biology. *Nat. Methods*, **7**, S56–S68.
21. Kono,N., Arakawa,K. and Tomita,M. (2006) MEGU: pathway mapping web-service based on KEGG and SVG. *In Silico Biol.*, **6**, 621–625.
22. Arakawa,K., Kono,N., Yamada,Y., Mori,H. and Tomita,M. (2005) KEGG-based pathway visualization tool for complex omics data. *In Silico Biol.*, **5**, 419–423.
23. Suhre,K. and Schmitt-Kopplin,P. (2008) MassTRIX: mass translator into pathways. *Nucleic Acids Res.*, **36**, W481–W484.
24. Tokimatsu,T., Sakurai,N., Suzuki,H., Ohta,H., Nishitani,K., Koyama,T., Umezawa,T., Misawa,N., Saito,K. and 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.
25. Tokimatsu,T., Sakurai,N., Suzuki,H. and Shibata,D. (2006) In Saito,K., Dixon,R.A. and Willmitzer,L. (eds), *Biotechnology in Agriculture and Forestry*, Vol. 57. Springer-Verlag, Berlin Heidelberg, pp. 155–163.
26. Sakurai,N. and Shibata,D. (2006) KaPPA-View for integrating quantitative transcriptomic and metabolomic data on plant metabolic pathway maps. *J. Pesticide Sci.*, **31**, 293–295.
27. Nakamura,Y., Kimura,A., Saga,H., Oikawa,A., Shinbo,Y., Kai,K., Sakurai,N., Suzuki,H., Kitayama,M., Shibata,D. *et al.* (2007) Differential metabolomics unraveling light/dark regulation of metabolic activities in *Arabidopsis* cell culture. *Planta*, **227**, 57–66.
28. Harada,H., Fujisawa,M., Teramoto,M., Sakurai,N., Suzuki,H., Shibata,D. and Misawa,N. (2009) Simple functional analysis of key genes involved in astaxanthin biosynthesis using *Arabidopsis* cultured cells. *Plant Biotechnol.*, **26**, 81–92.
29. Zhao,C.R., Sawaki,Y., Sakurai,N., Shibata,D. and Koyama,H. (2010) Transcriptomic profiling of major carbon and amino acid metabolism in the roots of *Arabidopsis thaliana* treated with various rhizotoxic ions. *Soil Sci. Plant Nutr.*, **56**, 150–162.
30. Obayashi,T., Hayashi,S., Saeki,M., Ohta,H. and Kinoshita,K. (2009) ATTED-II provides coexpressed gene networks for *Arabidopsis*. *Nucleic Acids Res.*, **37**, D987–D991.
31. Usadel,B., Nagel,A., Thimm,O., Redestig,H., Blaesing,O.E., Palacios-Rojas,N., Selbig,J., Hannemann,J., Piques,M.C., Steinhäuser,D. *et al.* (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. *Plant Physiol.*, **138**, 1195–1204.
32. Yano,K., Watanabe,M., Yamamoto,N., Tsugane,T., Aoki,K., Sakurai,N. and Shibata,D. (2006) MiBASE: a database of a miniature tomato cultivar Micro-Tom. *Plant Biotechnol.*, **23**, 195–198.
33. Ogata,Y., Suzuki,H., Sakurai,N. and Shibata,D. (2010) CoP: a database for characterizing co-expressed gene modules with biological information in plants. *Bioinformatics*, **26**, 1267–1268.
34. Geier,C., Gehmlich,K., Ehler,E., Hassfeld,S., Perrot,A., Hayess,K., Cardim,N., Wenzel,K., Erdmann,B., Krackhardt,F. *et al.* (2008) *Beyond the sarcomere: CSR3 mutations cause hypertrophic cardiomyopathy*. *Hum. Mol. Genet.*, **17**, 2753–2765.