PCP
PLANT & CELL PHYSIOLOGY

Short Communication

KAGIANA: An Excel-Based Tool for Retrieving Summary Information on *Arabidopsis* Genes

Yoshiyuki Ogata[1], Nozomu Sakurai[1], Koh Aoki[1], Hideyuki Suzuki[1], Koei Okazaki[1], Kazuki Saito[2] and Daisuke Shibata[1,*]

[1]Kazusa DNA Research Institute, Kazusa-Kamatari 2-6-7, Kisarazu, Chiba, 292-0818 Japan
[2]Graduate School of Pharmaceutical Science, Chiba University, Yayoi-cho 1-33, Inage-ku, Chiba, 263-8522 Japan

Various public databases provide *Arabidopsis* gene information via the internet. It is useful to abstract information obtained from such databases. We have developed the KAGIANA tool, which allows a user to retrieve summary information obtained from selective databases and to access pages for a gene of interest in those databases. The tool is based on Microsoft Excel and provides several macro programs for gene expression analyses. It can assist plant biologists in accessing omics information for plant biology. The KAGIANA tool is freely available at http://pmnedo.kazusa.or.jp/kagiana/.

Keywords: Annotation • *Arabidopsis* • Database • Gene expression • Omics.

Abbreviations: AGI, *Arabidopsis* Genome Initiative; GO, Gene Ontology; NCBI, The National Center for Biotechnology Information; TAIR, The *Arabidopsis* Information Resource.

Since the completion of the genome sequence of the model plant *Arabidopsis thaliana* (*Arabidopsis Genome Initiative 2000*), advances in genome and gene expression analysis have resulted in a vast number of data sets generated for *Arabidopsis*. Data sets of the *Arabidopsis* genome sequence are available at GenBank (Benson et al. 2008; http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide) and DDBJ (Sugawara et al. 2008; http://www.ddbj.nig.ac.jp/). In *Arabidopsis*, the sequence is separated into five chromosomes, which have 33,282 gene loci according to TAIR8 (Swarbreck et al. 2008; http://www.arabidopsis.org/). Amino acid sequences of proteins encoded by genes have been analyzed for various purposes. On the basis of localization signals included in such sequences, subcellular localization of proteins has been predicted using tools such as WoLF PSORT (Horton et al. 2007; http://wolfpsort.org/) and TargetP (Emanuelsson et al. 2007; http://www.cbs.dtu.dk/services/TargetP/). Domain structures, which show functional properties in proteins, have been predicted on the basis of amino acid sequences using tools such as SCOP (Andreeva et al. 2008; http://scop.berkeley.edu/), and can be found in databases such as InterPro (Mulder et al. 2007; http://www.ebi.ac.uk/interpro/). Analysis of transmembrane domains has been performed using tools such as TMHMM (Moller et al. 2001; http://www.cbs.dtu.dk/services/TMHMM/) and that of hydropathy can be found in databases such as SUBA (Heazlewood et al. 2007; http://www.plantenergy.uwa.edu.au/suba2/). Data sets of gene expression are available from databases such as the Gene Expression Omnibus (GEO) database (Barrett et al. 2006; http://www.ncbi.nlm.nih.gov/geo/). Several databases such as ATTED-II (Obayashi et al. 2007; http://atted.jp/) and Genevestigator (Zimmermann et al. 2004; https//www.genevestigator.ethz.ch/) provide a function to perform gene expression analysis. Using a vast number of gene expression data sets, approaches for detecting co-expressed genes, such as the ARACNE tool (Margolin et al. 2006; http://amdec-bioinfo.cu-genome.org/html/ARACNE.htm), the average clustering coefficient index (Gupta et al. 2006) and the Confeito algorithm (http://pmnedo.kazusa.or.jp/kagiana/coexprocess/) have been developed. On the basis of these analyses, molecular function, subcellular localization and biological processes of genes have finally been consistently assigned to 'molecular function', 'cellular component' and 'biological process', respectively, the three aspects of the Gene Ontology (GO) terminology (Gene Ontology Consortium 2008; http://geneontology.org/).

To obtain genomic and transcriptomic information on genes of interest, a user can visit these databases and access these tools via the internet or download them for personal

*Corresponding author: E-mail, shibata@kazusa.or.jp; Fax, +81-438-52-3948.

Plant Cell Physiol. 50(1): 173–177 (2009) doi:10.1093/pcp/pcn179, available online at www.pcp.oxfordjournals.org

© The Author 2008. Published by Oxford University Press on behalf of Japanese Society of Plant Physiologists. All rights reserved.
The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and the Japanese Society of Plant Physiologists are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

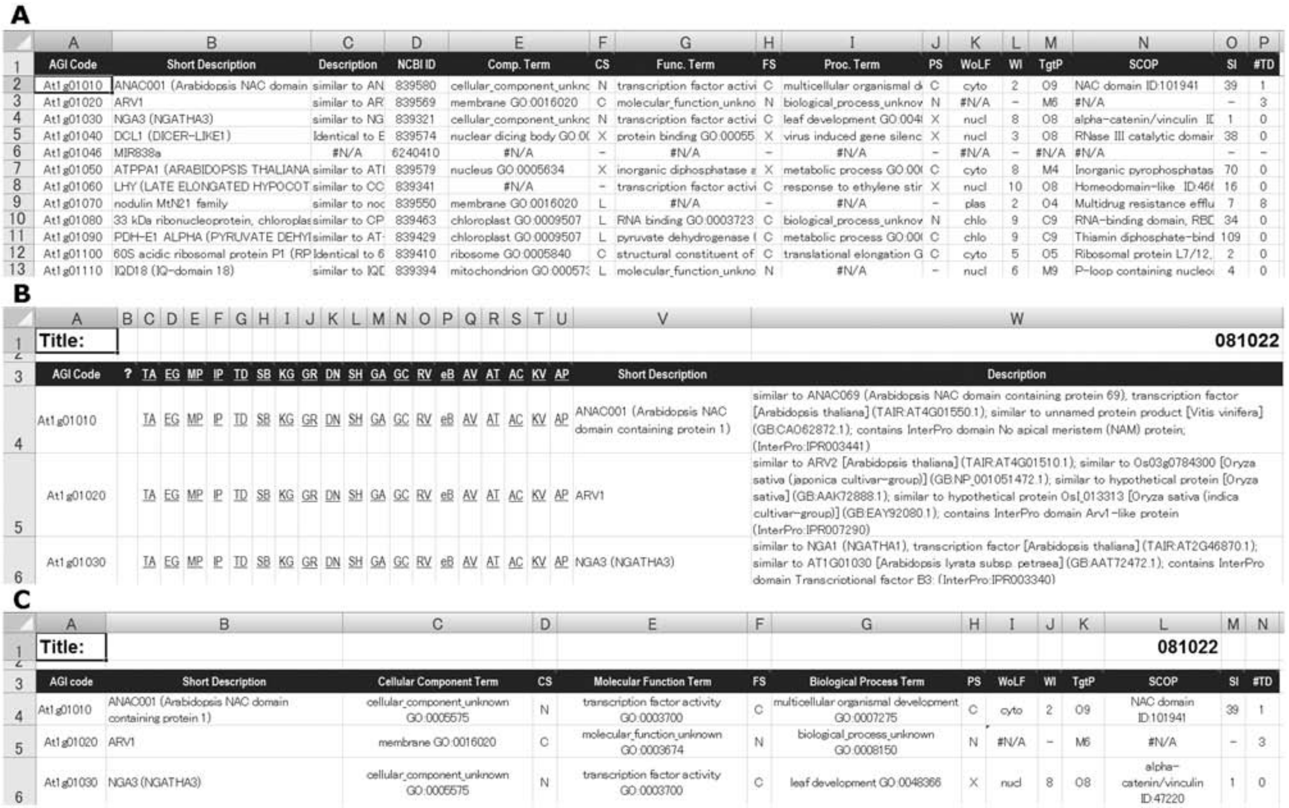
Downloaded from http://pcp.oxfordjournals.org at Institute of Physical and Chemical Research(RIKEN) on April 7, 2010

use. However, to retrieve the information, users generally require knowledge of the omics information published in the databases; for example, how to select an adequate website and how to set an adequate threshold value such as the gene-to-gene correlation coefficient for acquiring data of interest in the website. For biological users, unfamiliar with omics analyses such as genomics and transcriptomics, it is useful to have access to abstracted gene information from such databases and analyses and to use quick links to these databases.

We have developed the KAGIANA (Kazusa *Arabidopsis* Gene Information And Network Analysis) tool to summarize various *Arabidopsis* omics analyses from the above-mentioned databases and tools, and to provide links to pages for genes of interest in the databases. The tool is based on Microsoft Excel (version 2003 or higher) and thus requires only enough skill for basic Excel operation. The implementation of this tool is verified using Windows XP or higher for PC, and OS X or higher for Macintosh. The macro programs of the tools are available only for Windows users as of November 2008. Our goal is to assist plant biologists in accessing information from omics analyses so that they can incorporate it into their plant biology research.

The KAGIANA tool is downloadable as a ZIP-format file at http://pmnedo.kazusa.or.jp/kagiana/. The KAGIANA tool is formatted as a Microsoft Excel workbook file, composed of five worksheets [one database sheet ('Data20080524'), two readme sheets ('ReadMe_1st' and 'ReadMe_Tools') and two retrieval sheets ('Selected_Link' and 'Selected_GO')] and one macro program ('Tools') comprising four analysis tools ('Confeito', 'GX bar chart', 'GO pie chart' and 'ATTED chart'). In KAGIANA, AGI codes (e.g. At1g01010) are used for the retrieval and performance of the tool.

The database sheet is composed of the following information for 33,362 loci (**Fig. 1A**), which was obtained from the TAIR database. First, the A to D columns represent AGI codes, a short description, description, and identifiers for NCBI, respectively. Secondly, the E to J columns display representative GO terms, which certainly accompany the evidence codes, and their Evidence Code categories, which are abbreviated as 'X' (experimental) for EXP, IDA, IPI, IMP, IGI and IEP; 'S' (statement) for TAS and IC; 'C' (computational) for ISS, ISO, ISA, ISM, IGC and RCA; 'L' (electronic) for IEA; and 'N' (not available) for NAS and ND, in the three aspects of GO terminology, i.e. cellular component (the E and F columns), molecular function (the G and H columns) and biological



A

AGI Code	Short Description	Description	NCBI ID	Comp. Term	CS	Func. Term	FS	Proc. Term	PS	WoLF	WI	TgtP	SCOP	SI	#TD
At1g01010	ANAC001 (Arabidopsis NAC domain	similar to AN	839580	cellular_component_unknc	N	transcription factor activi	C	multicellular organismal d	C	cyto	2	O9	NAC domain ID:101941	39	1
At1g01020	ARV1	similar to AR	839569	membrane GO:0016020	C	molecular_function_unkno	N	biological_process_unknov	N	#N/A	–	M6	#N/A	–	3
At1g01030	NGA3 (NGATHA3)	similar to NG	839321	cellular_component_unknc	N	transcription factor activi	C	leaf development GO:004	X	nucl	8	O8	alpha-catenin/vinculin II	1	0
At1g01040	DCL1 (DICER-LIKE1)	Identical to E	839574	nuclear dicing body GO:0(	X	protein binding GO:00055	X	virus induced gene silenc	X	nucl	3	O8	RNase III catalytic domair	38	0
At1g01046	MIR838a	#N/A	6240410	#N/A	–	#N/A	–	#N/A	–	#N/A	–	#N/A	#N/A	–	–
At1g01050	ATPPA1 (ARABIDOPSIS THALIANA	similar to ATI	839579	nucleus GO:0005634	X	inorganic diphosphatase a	X	metabolic process GO:00(	C	cyto	8	M4	Inorganic pyrophosphatas	70	0
At1g01060	LHY (LATE ELONGATED HYPOCOT	similar to CC	839341	#N/A	–	transcription factor activi	C	response to ethylene stir	X	nucl	10	O8	Homeodomain-like ID:46(	16	0
At1g01070	nodulin MtN21 family	similar to noc	839550	membrane GO:0016020	L	#N/A	–	#N/A	–	plas	2	O4	Multidrug resistance efflu	7	8
At1g01080	33 kDa ribonucleoprotein, chloroplas	similar to CP	839463	chloroplast GO:0009507	L	RNA binding GO:0003723	C	biological_process_unknov	N	chlo	9	C9	RNA-binding domain, RBD	34	0
At1g01090	PDH-E1 ALPHA (PYRUVATE DEHY	similar to AT-	839429	chloroplast GO:0009507	L	pyruvate dehydrogenase (	C	metabolic process GO:00(	C	chlo	9	C9	Thiamin diphosphate-bind	109	0
At1g01100	60S acidic ribosomal protein P1 (RP	Identical to 6	839410	ribosome GO:0005840	C	structural constituent of	C	translational elongation G	C	cyto	5	O5	Ribosomal protein L7/12,	2	0
At1g01110	IQD18 (IQ-domain 18)	similar to IQD	839394	mitochondrion GO:00057:	L	molecular_function_unkno	N	#N/A	–	nucl	6	M9	P-loop containing nucleo	4	0

B

Title: 081022

AGI Code	?	TA	EG	MP	IP	TD	SB	KG	GR	DN	SH	GA	GC	RV	eB	AV	AT	AC	KV	AP	Short Description	Description
At1g01010		TA	EG	MP	IP	TD	SB	KG	GR	DN	SH	GA	GC	RV	eB	AV	AT	AC	KV	AP	ANAC001 (Arabidopsis NAC domain containing protein 1)	similar to ANAC069 (Arabidopsis NAC domain containing protein 69); transcription factor [Arabidopsis thaliana] (TAIR:AT4G01550.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO62872.1); contains InterPro domain No apical meristem (NAM) protein; (InterPro:IPR003441)
At1g01020		TA	EG	MP	IP	TD	SB	KG	GR	DN	SH	GA	GC	RV	eB	AV	AT	AC	KV	AP	ARV1	similar to ARV2 [Arabidopsis thaliana] (TAIR:AT4G01510.1); similar to Os03g0784300 [Oryza sativa (japonica cultivar-group)] (GB:NP_001051472.1); similar to hypothetical protein [Oryza sativa] (GB:AAK72888.1); similar to hypothetical protein OsI_013313 [Oryza sativa (indica cultivar-group)] (GB:EAY92080.1); contains InterPro domain Arv1-like protein (InterPro:IPR007290)
At1g01030		TA	EG	MP	IP	TD	SB	KG	GR	DN	SH	GA	GC	RV	eB	AV	AT	AC	KV	AP	NGA3 (NGATHA3)	similar to NGA1 (NGATHA1), transcription factor [Arabidopsis thaliana] (TAIR:AT2G46870.1); similar to AT1G01030 [Arabidopsis lyrata subsp. petraea] (GB:AAT72472.1); contains InterPro domain Transcriotional factor B3; (InterPro:IPR003340)

C

Title: 081022

AGI code	Short Description	Cellular Component Term	CS	Molecular Function Term	FS	Biological Process Term	PS	WoLF	WI	TgtP	SCOP	SI	#TD
At1g01010	ANAC001 (Arabidopsis NAC domain containing protein 1)	cellular_component_unknown GO:0005575	N	transcription factor activity GO:0003700	C	multicellular organismal development GO:0007275	C	cyto	2	O9	NAC domain ID:101941	39	1
At1g01020	ARV1	membrane GO:0016020	C	molecular_function_unknown GO:0003674	N	biological_process_unknown GO:0008150	N	#N/A	–	M6	#N/A	–	3
At1g01030	NGA3 (NGATHA3)	cellular_component_unknown GO:0005575	N	transcription factor activity GO:0003700	C	leaf development GO:0048366	X	nucl	8	O8	alpha-catenin/vinculin ID:47220	1	0

Fig. 1 Composition of the KAGIANA worksheets. (A) The database sheet, including summary information of *Arabidopsis* genes obtained from the selected databases. (B) The sheet for hyperlinks to the selected public databases, as shown in **Table 1.** (C) The sheet for summary information of the selected omics analyses, i.e. GO terms and results from analyses of WoLF PSORT, TargetP, SCOP and TMHMM.

Downloaded from http://pcp.oxfordjournals.org at Institute of Physical and Chemical Research(RIKEN) on April 7, 2010

process (the I and J columns), respectively. A GO term was selected as the representative term for each aspect for a gene, according to the order of Evidence Code categories, i.e. X, S, C, L and N. The following columns represent information from the analytical tools. The K and L columns represent data from WoLF PSORT, which predicts the subcellular localization of proteins, and the reliability index, whose best score is 14, respectively. The M column represents information from TargetP, which also predicts subcellular localization, and the reliability index, ranging from 0 to 9 at the maximum. The N and O columns represent that from SCOP, which predicts domains of proteins, and the reliability index, which is the negative logarithm of the actual value, respectively. The P column represents TMHMM, which predicts the number of transmembrane domains of proteins.

The 'Selected_Link' sheet provides hyperlinks to 19 selected public databases for information retrieval of genes of interest and their Short Description and Description (**Fig. 1B**). These hyperlinks lead a user to the pages for individual genes in the individual databases by the following steps: (i) input AGI code(s) in the A column from the A4 cell to the lower cells (e.g. input 'At1g01010' in the A4 cell and 'At1g01020' in the A5 cell); (ii) select the range of the B4 to the W4 cells; and (iii) double-click the right lower corner (a black square) to copy the equations in the fourth row into the lower rows in the same columns (e.g. copy the B4-W4 into the B5-W5). Then, a user can access a database of interest from among the C to U columns (e.g. click the T4 cell for access to the page for the query gene in the KaPPA-View tool). The tool provides access to the databases shown in **Table 1**. The way to use this sheet is also described in the 'ReadMe_1st' sheet.

In the 'Selected_GO' sheet, a user can retrieve information on genes of interest from various omics analyses (**Fig. 1C**), i.e. the three GO term aspects, WoLF PSORT, TargetP, SCOP and TMHMM as mentioned above. Steps for retrieval are similar to those in the 'Selected_Link' sheet. The terms in the third row are the same as those in the database sheet mentioned above, and the 'ReadMe_1st' sheet has the explanation for such retrieval. By selecting the 'Selected_Link' and the 'Selected_GO' sheets, a user can manage to operate them simultaneously, e.g. when inputting AGI codes.

KAGIANA provides 'Tools' macro programs including the four analyses (**Fig. 2A**), i.e. including 'Confeito', 'GX bar chart', 'GO pie chart' and 'ATTED chart'. The 'Confeito' tool allows a user to extract co-expressed genes using the Confeito algorithm on the basis of a co-expression network approach (http://pmnedo.kazusa.or.jp/kagiana/coexprocess/). The way to use the tools is described in the 'ReadMe_Tools' sheet. The 'GX bar chart' tool allows a user to depict bar charts of gene expression profiles for multiple genes of interest (**Fig. 2B**). Bar charts are depicted using 1,245 DNA microarray data from the AtGenExpress project, which are available at http://www.weigelworld.org/resources/microarray/AtGenExpress/. The 'GO pie chart' tool allows a user to depict a pie chart of the distribution of GO-SLIM terms associated with multiple genes of interest (**Fig. 2C**). GO-SLIM terms are available at the TAIR database. This tool counts all multiple GO-SLIM terms assigned to a gene. For this version of KAGIANA, such terms were obtained at May 2008. The 'ATTED chart' tool helps users download the charts of AtGenExpress gene expression profiles for individual genes from the ATTED database onto a worksheet in KAGIANA per gene.

Table 1 Abstract of databases linkable from the KAGIANA tool

Abbreviation	Databases	URL
TA	TAIR (Swarbreck et al. 2008)	http://www.arabidopsis.org/
EG	Entrez Gene (Maglott et al. 2005)	http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene
MP	MPSS (Brenner et al. 2000)	http://mpss.udel.edu/at/
IP	InParanoid (Berglund et al. 2008)	http://inparanoid.sbc.su.se/cgi-bin/index.cgi
TD	SIGnAL (Alonso et al. 2003)	http://signal.salk.edu/cgi-bin/tdnaexpress
SB	SUBA (Heazlewood et al. 2007)	http://www.plantenergy.uwa.edu.au/applications/suba2/
KG	KEGG (Kanehisa et al. 2008)	http://www.genome.jp/kegg/
GR	GRAMENE (Liang et al. 2008)	http://www.gramene.org/
DN	NASCArrays Digital Northern (Craigon et al. 2004)	http://affymetrix.arabidopsis.info/narrays/digitalnorthern.pl
SH	NASCArrays Spot History (Craigon et al. 2004)	http://affymetrix.arabidopsis.info/narrays/spothistory.pl
GA	Genevestigator Gene Atlas (Zimmermann et al. 2004)	https://www.genevestigator.ethz.ch/gv/index.jsp
GC	Genevestigator Gene Chronologer (Zimmermann et al. 2004)	https://www.genevestigator.ethz.ch/gv/index.jsp
RV	Genevestigator Response Viewer (Zimmermann et al. 2004)	https://www.genevestigator.ethz.ch/gv/index.jsp
eB	eFP Browser (Winter et al. 2007)	http://bbc.botany.utoronto.ca/efp/cgi-bin/efpWeb.cgi
AV	AtGenExpress Visualization Tool	http://jsp.weigelworld.org/expviz/expviz.jsp
AT	ATTED-II (Obayashi et al. 2007)	http://atted.jp/
AC	AraCyc (Zhang et al. 2005)	http://www.arabidopsis.org/biocyc/index.jsp
KV	KaPPA-View 3 (Sakurai and Shibata 2006)	http://kpv.kazusa.or.jp/kappa-view3/
AP	AtProteome (Baerenfaller et al. 2008)	http://fgcz-atproteome.unizh.ch/

Downloaded from http://pcp.oxfordjournals.org at Institute of Physical and Chemical Research(RIKEN) on April 7, 2010

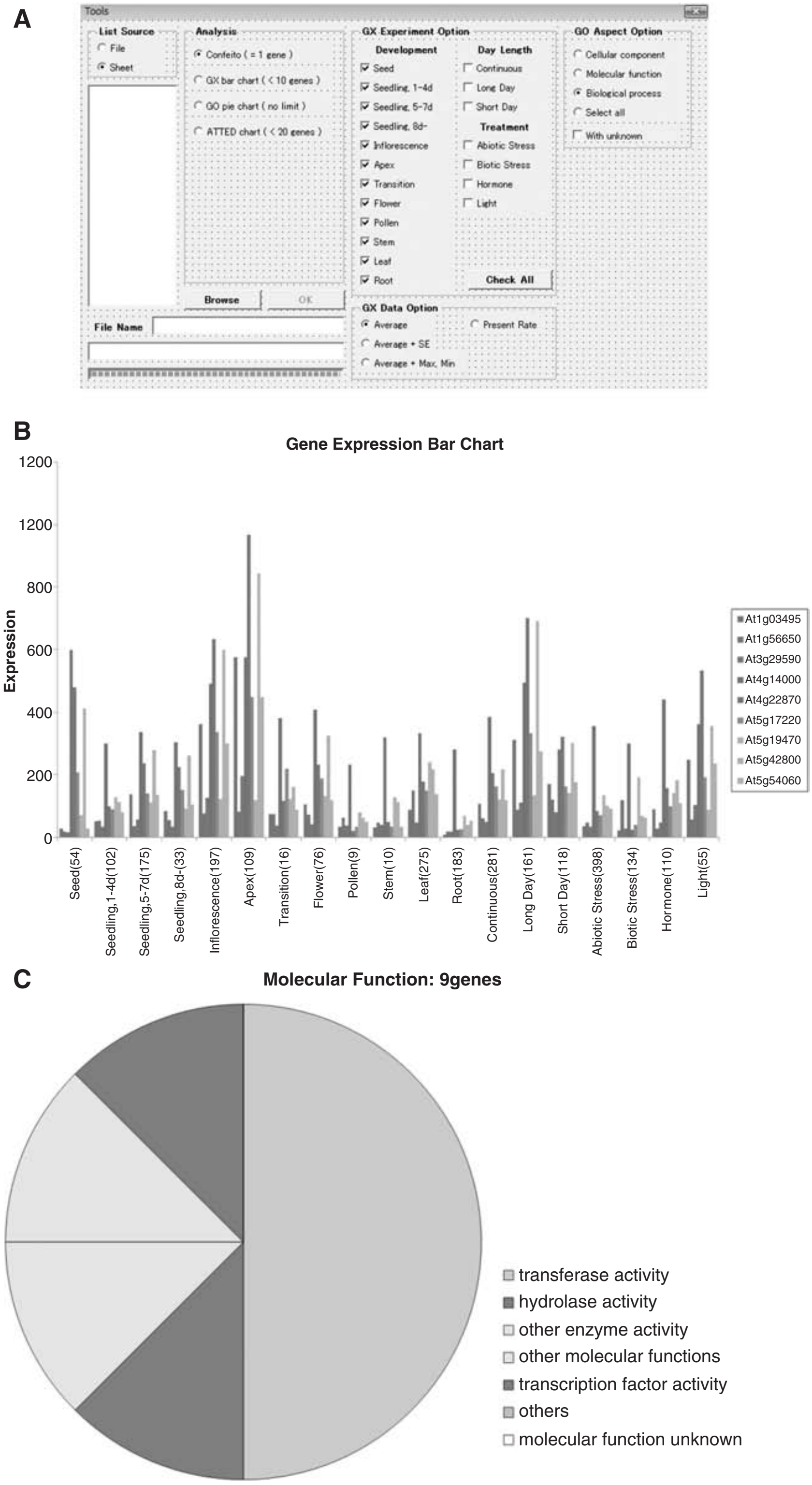


Fig. 2 Composition of the KAGIANA tools. (A) The window of the 'Tools' macro program, including analyses of 'Confeito', 'GX bar chart', 'GO pie chart' and 'ATTED chart'. (B) The result from the 'GX bar chart' analysis. (C) The result from the 'GO pie chart' analysis, showing the aspect of molecular function.

Downloaded from http://pcp.oxfordjournals.org at Institute of Physical and Chemical Research(RIKEN) on April 7, 2010

Detailed steps for using these tools are described in the 'ReadMe_Tools' sheet in the KAGIANA workbook. Briefly, the steps are (i) click 'Tools' in the menu bar; (ii) select 'Macro' and click 'Macros'; (iii) select 'Tools' in the macro box and click 'Execute' (open the 'Tools' window); (iv) select a tool in the 'Analysis' frame in the window; (v) input AGI codes into different lines in the textbox left of the frame; (vi) select the option frame when selecting 'GX bar chart' and 'GO pie chart' tools; and then (vii) click the 'OK' button if the character color on the button is black (otherwise, there is insufficient information for retrieval).

Funding

The New Energy and Industrial Technology Development (NEDO) program, part of the 'Development of Fundamental Technologies for Controlling the Material Production Process of Plants' project.

References

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., et al. (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* 301: 653–657.

Andreeva, A., Howorth, D., Chandonia, J.M., Brenner, S.E., Hubbard, T.J., Chothia, C., et al. (2008) Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* 36: D419–D425.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.

Baerenfaller, K., Grossmann, J., Grobei, M.A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., et al. (2008) Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. *Science* 320: 938–941.

Barrett, T., Troup, D.B., Willhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., et al. (2006) NCBI GEO: mining tens of millions of expression profiles—database and tools update. *Nucleic Acids Res.* 35: D760–D765.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. (2008) GenBank. *Nucleic Acids Res* 36: D25–D30.

Berglund, A.C., Sjölund, E., Östlund, G. and Sonnhammer, E.L. (2008) InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res.* 36: D263–D266.

Brenner, S., Williams, S.R., Vermaas, E.H., Storck, T., Moon, K., McCollum, C., et al. (2000) In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci. USA* 18: 630–634.

Craigon, D.J., James, N., Okyere, J., Higgins, J., Jotham, J. and May, S. (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res.* 32: D575–D577.

Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2: 953–971.

Gene Ontology Consortium(2008) The Gene Ontology project in 2008. *Nucleic Acids Res.* 36: D440–D444.

Gupta, A., Maranas, C.D. and Albert, R. (2006) Elucidation of directionality for co-expressed genes: predicting intra-operon termination sites. *Bioinformatics* 22: 209–214.

Heazlewood, J.L., Verboom, R.E., Tonti-Filippini, J., Small, I. and Millar, A.H. (2007) SUBA: the Arabidopsis Subcellular Database. *Nucleic Acids Res.* 35: D213–D218.

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., et al. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35: W585–W587.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36: D480–D484.

Liang, C., Jaiswal, P., Hebbard, C., Avraham, S., Buckler, E.S., Casstevens, T., et al. (2008) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.* 36: D947–D953.

Maglott, D., Ostell, J., Pruitt, K.D. and Tatusova, T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* 33: D54–D58.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla-Favera, R., et al. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7: S7.

Möller, S., Croning, M.D.R. and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17: 646–653.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., et al. (2007) New developments in the InterPro database. *Nucleic Acids Res* 35: D224–D228.

Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M., et al. (2007) ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis. *Nucleic Acids Res* 35: D863–D869.

Sakurai, N. and Shibata, D. (2006) KaPPA-View for integrating quantitative transcriptomic and metabolomic data on plant metabolic pathway maps. *J. Pestic. Sci.* 31: 293–295.

Sugawara, H., Ogasawara, O., Okubo, K., Gojobori, T. and Tateno, Y. (2008) DDBJ with new system and face. *Nucleic Acids Res.* 36: D22–D24.

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

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J. (2007) An 'electronic fluorescent pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* 2: e718.

Zhang, P., Foerster, H., Tissier, C.P., Mueller, L., Paley, S., Karp, P.D., et al. (2005) MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiol.* 138: 27–37.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W. (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiol.* 136: 2621–2632.

(Received October 22, 2008; Accepted November 17, 2008)

Downloaded from http://pcp.oxfordjournals.org at Institute of Physical and Chemical Research(RIKEN) on April 7, 2010