Published online 11 November 2006 *Nucleic Acids Research, 2007, Vol. 35, Database issue* ***D841–D845***
doi:10.1093/nar/gkl835

PathoPlant®: a platform for microarray expression data to analyze co-regulated genes involved in plant defense responses

Lorenz Bülow*, Martin Schindler[1] and Reinhard Hehl

Institut für Genetik, Technische Universität Braunschweig, Spielmannstraße 7, D-38106 Braunschweig, Germany and [1]Software Systems Engineering Institute, Technische Universität Braunschweig, Mühlenpfordtstraße 23, D-38106 Braunschweig, Germany

Received August 15, 2006; Revised October 5, 2006; Accepted October 6, 2006

ABSTRACT

Plants react to pathogen attack by expressing specific proteins directed toward the infecting pathogens. This involves the transcriptional activation of specific gene sets. PathoPlant®, a database on plant–pathogen interactions and signal transduction reactions, has now been complemented by microarray gene expression data from *Arabidopsis thaliana* subjected to pathogen infection and elicitor treatment. New web tools enable identification of plant genes regulated by specific stimuli. Sets of genes co-regulated by multiple stimuli can be displayed as well. A user-friendly web interface was created for the submission of gene sets to be analyzed. This results in a table, listing the stimuli that act either inducing or repressing on the respective genes. The search can be restricted to certain induction factors to identify, e.g. strongly up- or down-regulated genes. Up to three stimuli can be combined with the option of induction factor restriction to determine similarly regulated genes. To identify common *cis*-regulatory elements in co-regulated genes, a resulting gene list can directly be exported to the AthaMap database for analysis. PathoPlant is freely accessible at http://www.pathoplant.de.

INTRODUCTION

To counteract pathogen attacks, plants have evolved strategies that comprise pathogen perception, signal transduction and induction of appropriate defense responses (1,2). Regulation of these responses is mediated by a network of signal transduction pathways in which classical signal transmitters such as receptors and MAP kinases are triggered by signals from elicitors and signal molecules such as ethylene, salicylic acid (SA) and jasmonic acid (JA) to activate defense-related gene and protein expression (3–5). Infection of plants with distinct pathogens results in specific accumulation rates of ethylene, SA and JA, and in distinct sets of activated genes representing individual signal molecule signatures and gene expression profiles for different pathogens (6–9). Although the significance of signal molecules is evident from extensive experiments with transgenic and mutant plants altered in ethylene, SA and JA signaling (10–13), it has been shown that cross-communication between their signaling pathways exists (7,14–16). Gene expression analysis in *Arabidopsis thaliana* revealed that orchestrated regulation results in specific gene induction patterns for distinct signal molecules and pathogens with a considerable degree of overlapping genes (6,9).

The PathoPlant® database was developed to display signal perception and signal transduction pathways on a molecular level during plant pathogenesis as well as the corresponding interactions between plants and pathogens on the organism level (17). Only experimentally proven direct molecular interactions have been annotated so far which lead to a limited number of regulated genes being covered in PathoPlant. In order to display all other genes regulated independently of the underlying molecular mechanism, gene expression data from microarray experiments represent an ideal source of information. Therefore, PathoPlant has now been complemented by *A.thaliana* microarray gene expression data. The datasets chosen are plant pathogenesis related and represent not only endogenous plant signal molecules, such as SA and JA, but also include treatments with elicitors and infections with different pathogens. This enables comparative studies of gene expression patterns.

Several web-based services harbor gene expression data from *A.thaliana* microarray experiments and allow recovery of information for individual genes or gene sets such as TAIR (18), NASCArrays tools (19), Stanford Microarray Database (20,21), Botany Array Resource (22), GEO (23) and Genevestigator (24). ACT (25,26), Botany Array Resource (22), CSB.DB (27) and Genevestigator (28) offer comparative gene analysis services to detect clusters of

*To whom correspondence should be addressed. Tel: +49 531 391 5783; Fax: +49 531 391 5765; Email: l.buelow@tu-braunschweig.de

genes with similar expression patterns across selected or the complete set of stimuli. These tools start with a given gene of interest to determine similarities in expression patterns to other genes. In contrast, the PathoPlant gene expression function was designed to start with combinations of up to three different stimuli to determine all overlapping genes being up-, down- or not regulated by these stimuli. An additional valuable feature of PathoPlant is the integration of AthaMap (29–31) for subsequent *cis*-regulatory element identification. A similar way of analysis is offered by Promomer at the Botany Array Resource (22). Promomer is a web tool to discover over-represented sequence motifs in regulatory regions from sets of *A.thaliana* genes. In contrast, AthaMap identifies putative functional *cis*-regulatory elements based on binding site specificities of transcription factors (29). The possibility of stimulus combinations to find overlapping genes and the interplay with the AthaMap database are unique features of PathoPlant.

THE PATHOPLANT GENE EXPRESSION RESOURCE

Microarray data processing and database content

Selected *A.thaliana* microarray expression sets were downloaded from the TAIR microarray experiments resource (18). cDNA microarray experiments with non-transgenic plants treated with pathogens, signal molecules and elicitors were chosen for import into PathoPlant. Print-tip-group lowess normalized data records were downloaded (32). These comprise array element name, AGI locus/gene identifier, induction factor and a tag indicating gene expression. Owing to different array designs, the raw data had to be processed individually prior to database import. In cDNA microarray experiments, cross-hybridizations lead to multiple locus/gene assignments for one array element. For each dataset, a Perl script was used to detect and tag these multiple locus/gene assignments to single array elements. The processed datasets were imported into the PathoPlant database.

The previously described PathoPlant database structure (17) was extended by three new tables for storage of microarray expression data, information on the experiments with corresponding links, and *A.thaliana* gene annotation data (TIGR release 5.0, January 21, 2004). After import, tagged records with very low absolute expression values, i.e. below a signal intensity of 350 in both channels, were excluded from being displayed online as these expression values are near background activity and are commonly considered as genes being not expressed (18). Records derived from multiple locus assignments to one single array element are not displayed as well because expression values may also rely on cross-hybridization to homologous genes. In order to merge records representing replicate hybridizations, geometric mean values and base-10 logarithms of the standard deviations of the individual induction factors were determined using a MS Visual Basic script. In addition to the number of replicates, this provides a common measure for experimental variability.

For import into PathoPlant, *A.thaliana* expression data from cDNA microarray experiments were selected for 19 stimuli. Table 1 displays the current database content including the number of records and the number of genes covered. The selected stimuli cover bacterial, viral and fungal pathogens as well as a fungal elicitor and signal molecules. For treatment with chitin, ethylene and tobacco mosaic virus (TMV), experimental setups with different sample collection time points after inoculation were annotated. Gene expression data for TMV infection discriminate between inoculated lower leaves and non-inoculated systemically infected upper leaves.

Expression data retrieval and analysis

The web interface of PathoPlant was extended to permit easy access to the expression data. Internal and external links were

Table 1. Microarray expression data annotated to PathoPlant

Stimulus	Class	No. of records	No. of genes	TAIR expression set ID
Xanthomonas campestris	Bacterial pathogen	17 846	6850	1005823536
Fusarium virguliforme	Fungal pathogen	18 963	6804	1005823583
Phythophthora infestans	Fungal pathogen	18 643	6719	1005823534
Powdery mildew	Fungal pathogen	10 130	6974	1005823549
TMV infected leaves 3 dpi	Viral pathogen	17 921	6430	1005823504
TMV infected leaves 4 dpi	Viral pathogen	28 369	7835	1005823602
TMV systemic leaves 14 dpi	Viral pathogen	74 320	9289	1005823505, 1005823602
Chitin 10 min	Elicitor	3939	1585	1005823605
Chitin 30 min	Elicitor	3905	1578	1005823605
Chitin 1 h	Elicitor	3837	1563	1005823605
Chitin 3 h	Elicitor	3876	1569	1005823605
Chitin 6 h	Elicitor	3888	1586	1005823605
Chitin 24 h	Elicitor	3655	1522	1005823605
cis-Jasmone	Signal molecule	19 919	8027	1005823574
Ethylene 2 h	Signal molecule	7697	1487	1005823581
Ethylene 24 h	Signal molecule	9200	1423	1005823581
Hydrogen peroxide	Signal molecule	19 047	6880	1005823545
Methyl-jasmonate	Signal molecule	20 401	8160	1005823574
SA analog BTH	Signal molecule	9928	6835	1005823548

Stimuli are categorized into different classes. The number of records and genes being represented is given. TAIR expression set IDs are specified for reference.

incorporated through a data retrieval tool. Two basic query modes are available depending on the question addressed. One consists in submitting a list of genes to retrieve the stimuli that these genes are regulated by, and the other displays all genes regulated by certain selected stimuli.

In the first query mode, a gene list, i.e. the locus ID separated by carriage returns, can be submitted to obtain expression data on these genes for all stimuli annotated in PathoPlant. This search can be restricted to records with induction factors higher or lower than a given value. Either individual induction factors or mean factors from replicates can be chosen. The final result table is a list with those genes and corresponding stimuli matching the search criteria. Since cDNA microarray expression datasets will not cover all genes (Table 1), some genes submitted may not match any stimulus and induction factor. These genes will be identified in a separate list.

The second query mode permits the search for genes co-regulated by certain stimuli. Up to three different stimuli can be selected to identify genes that match all stimuli by using the AND operator. Alternatively for displaying all genes that match at least one of the stimuli, the OR operator can be used. Experimental setups for single stimuli comprising diverse incubation conditions can be chosen individually, e.g. Chitin 10 min, or may be selected all at once (Chitin all). Restriction to certain induction factors or mean factors is also applicable in this search mode.

Both search modes result in a list of genes, a short gene description, stimuli and induction factors (Figure 1). Besides the induction factors given for a single experiment, mean induction factors are displayed that cover results from replicate experiments. In order to validate mean and single induction factors, the number of replicates (n) and base-10 logarithm of standard deviation are given. The result list is sorted by genes (locus) by default. It can be resorted by description, stimulus, induction factor and mean factor by selecting the respective column header. For further information on the genes displayed, a short description is provided and all genes are linked to respective entries of the TAIR database on *A.thaliana* genes (Locus links, Figure 1). For additional information on the experimental setups, metadata on the stimuli and experiments are provided via hyperlinks that directly link to the entries in the TAIR microarray experiments resource (Stimulus links, Figure 1). By selecting the locus/gene ID, the presence of the genes on other microarray datasets can be queried. Additionally, the entire list of genes obtained by an expression search can directly be submitted to the 'Search by locus/gene ID' form to perform a 'microarray expression search in PathoPlant' (Figure 1) to identify other stimuli acting on this set of genes. The genomic context of a single gene can be analyzed for regulatory transcription factor binding sites by using the link to the AthaMap resource (Figure 1). Most importantly, by submitting the list of all displayed genes to AthaMap, a comparative transcription factor

Search for expressed *A. thaliana* genes

Search by stimulus:

TMV systemic leaves 14dpi
SA analog BTH

Expression parameter induction factor
mean from replicates
single experiments
>1.5

AND OR

Search Demo Reset

factor<-4
-4 <= factor < -2.5
-2.5 <= factor < -1.5
-1.5 <= factor < 1.5
1.5 <= factor < 2.5
2.5 <= factor < 4
factor >= 4

Display of 197 records representing 23 loci.

Use locus list for a microarray expression search in PathoPlant.
Use locus list for a transcription factor binding site analysis in AthaMap.

Locus	Description	Locus links	Stimulus	Stimulus links	Factor	Mean factor	n (lg std dev)
At5g57560	xyloglucan:xyloglucosyl transferase / xyloglucan e ...	TAIR AthaMap	SA analog BTH	TAIR	4.757	4.757	1 (0)
At5g10760	aspartyl protease family protein contains Pfam dom ...	TAIR AthaMap	SA analog BTH	TAIR	4.488	4.488	1 (0)
At3g57260	glycosyl hydrolase family 17 protein similar to gl ...	TAIR AthaMap	SA analog BTH	TAIR	5.533	4.342	2 (0.105)
At3g57260	glycosyl hydrolase family 17 protein similar to gl ...	TAIR AthaMap	SA analog BTH	TAIR	3.408	4.342	2 (0.105)
At5g47120	Bax inhibitor-1 putative / BI-1 putative SP:Q9LD45 ...	TAIR AthaMap	SA analog BTH	TAIR	4.199	4.183	2 (0.002)
At5g47120	Bax inhibitor-1 putative / BI-1 putative SP:Q9LD45 ...	TAIR AthaMap	SA analog BTH	TAIR	4.167	4.183	2 (0.002)
At1g02930	glutathione S-transferase, putative similar to glu ...	TAIR AthaMap	TMV systemic leaves 14dpi	TAIR	3.761	3.69	4 (0.058)
At1g02930	glutathione S-transferase, putative similar to glu ...	TAIR AthaMap	TMV systemic leaves 14dpi	TAIR	3.189	3.69	4 (0.058)
At1g02930	glutathione S-transferase, putative similar to glu ...	TAIR AthaMap	TMV systemic leaves 14dpi	TAIR	4.547	3.69	4 (0.058)
At1g02930	glutathione S-transferase, putative similar to glu ...	TAIR	TMV systemic leaves	TAIR	3.399	3.69	4 (0.058)

Figure 1. Screenshot of a PathoPlant microarray expression search result after submission of a query using the parameters displayed in the figure. The result table was sorted by Mean factor.

binding site analysis can be performed as described below. This permits the identification of common *cis*-regulatory elements in co-regulated genes.

Interplay with AthaMap

A link implemented in PathoPlant permits direct submission of the genes listed in the result table to the AthaMap gene analysis tool for identification of *cis*-regulatory elements (Figure 1). In a selected region relative to the start codon of the submitted genes, AthaMap will identify predicted binding sites of transcription factors. A table in AthaMap can be selected showing common transcription factor binding sites in all genes to identify binding site over-representation.

In the opposite direction, AthaMap has also been linked with PathoPlant. An AthaMap co-localization analysis (http://www.athamap.de/search_colocalization.php) with selected transcription factors will detect binding site co-localizations and all corresponding *A.thaliana* genes. Through a link implemented in AthaMap, this list of genes can be submitted to PathoPlant for the determination of stimuli by which these genes are regulated.

The interplay between PathoPlant gene expression analysis and AthaMap web tools enables easy identification of sets of co-regulated genes that can be further analyzed for common *cis*-regulatory elements. Alternatively, sets of co-regulated genes identified with AthaMap can be submitted to PathoPlant for expression profile analysis.

AVAILABILITY

The PathoPlant resource is a free service accessible via http://www.pathoplant.de. The database content is displayed on the PathoPlant homepage and is being updated on a regular basis.

ACKNOWLEDGEMENTS

We would like to thank Thomas Zobel, Daniela Bruhn and Juliane Scheithauer for data annotation. We thank Claudia Galuschka for critical reading of the manuscript. This project is part of the Intergenomics Center Braunschweig (http://www.intergenomics.de) funded by the German Federal Ministry for Education and Research, BMBF, Grant No. 031U110C/031U210C. Funding to pay the Open Access publication charges for this article was provided by BMBF.

Conflict of interest statement. None declared.

REFERENCES

1. Dangl,J.L. and Jones,J.D. (2001) Plant pathogens and integrated defence responses to infection. *Nature*, **411**, 826–833.
2. Dicke,M. and Hilker,M. (2003) Induced plant defences: from molecular biology to evolutionary ecology. *Basic Appl. Ecol.*, **4**, 3–14.
3. Glazebrook,J. (2001) Genes controlling expression of defense responses in *Arabidopsis*—2001 status. *Curr. Opin. Plant Biol.*, **4**, 301–308.
4. Thomma,B.P., Penninckx,I.A., Broekaert,W.F. and Cammue,B.P. (2001) The complexity of disease signaling in *Arabidopsis*. *Curr. Opin. Immunol.*, **13**, 63–68.
5. Reymond,P. and Farmer,E.E. (1998) Jasmonate and salicylate as global signals for defense gene expression. *Curr. Opin. Plant Biol.*, **1**, 404–411.
6. De Vos,M., Van Oosten,V.R., Van Poecke,R.M., Van Pelt,J.A., Pozo,M.J., Mueller,M.J., Buchala,A.J., Métraux,J.P., Van Loon,L.C. and Dicke,M. (2005) Signal signature and transcriptome changes of *Arabidopsis* during pathogen and insect attack. *Mol. Plant Microbe Interact.*, **18**, 923–937.
7. Glazebrook,J., Chen,W., Estes,B., Chang,H.S., Nawrath,C., Métraux,J.P., Zhu,T. and Katagiri,F. (2003) Topology of the network integrating salicylate and jasmonate signal transduction derived from global expression phenotyping. *Plant J.*, **34**, 217–228.
8. Reymond,P., Bodenhausen,N., Van Poecke,R.M., Krishnamurthy,V., Dicke,M. and Farmer,E.E. (2004) A conserved transcript pattern in response to a specialist and a generalist herbivore. *Plant Cell*, **16**, 3132–3147.
9. Schenk,P.M., Kazan,K., Wilson,I., Anderson,J.P., Richmond,T., Somerville,S.C. and Manners,J.M. (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl Acad. Sci. USA*, **97**, 11655–11660.
10. Pozo,M.J., Van Loon,L.C. and Pieterse,C.M.J. (2005) Jasmonates–signals in plant–microbe interactions. *J. Plant Growth Regul.*, **23**, 211–222.
11. Delaney,T.P., Uknes,S., Vernooij,B., Friedrich,L., Weymann,K., Negrotto,D., Gaffney,T., Gut–Rella,M., Kessmann,H. and Ward,E. (1994) A central role of salicylic acid in plant disease resistance. *Science*, **266**, 1247–1250.
12. Nawrath,C. and Métraux,J.P. (1999) Salicylic acid induction-deficient mutants of *Arabidopsis* express PR-2 and PR-5 and accumulate high levels of camalexin after pathogen inoculation. *Plant Cell*, **11**, 1393–1404.
13. Wildermuth,M.C., Dewdney,J., Wu,G. and Ausubel,F.M. (2001) Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature*, **414**, 562–565.
14. Felton,G.W. and Korth,K.L. (2000) Trade-offs between pathogen and herbivore resistance. *Curr. Opin. Plant Biol.*, **3**, 309–314.
15. Spoel,S.H., Koornneef,A., Claessens,S.M., Korzelius,J.P., Van Pelt,J.A., Mueller,M.J., Buchala,A.J., Métraux,J.P., Brown,R. and Kazan,K. (2003) NPR1 modulates cross-talk between salicylate- and jasmonate-dependent defense pathways through a novel function in the cytosol. *Plant Cell*, **15**, 760–770.
16. Pieterse,C.M. and Van Loon,L.C. (2004) NPR1: the spider in the web of induced resistance signaling pathways. *Curr. Opin. Plant Biol.*, **7**, 456–464.
17. Bülow,L., Schindler,M., Choi,C. and Hehl,R. (2004) PathoPlant®: a database on plant–pathogen interactions. *In Silico Biol.*, **4**, 0044.
18. Rhee,S.Y., Beavis,W., Berardini,T.Z., Chen,G., Dixon,D., Doyle,A., Garcia–Hernandez,M., Huala,E., Lander,G. and Montoya,M. *et al.* (2003) The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res.*, **31**, 224–228.
19. 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.
20. Gollub,J., Ball,C.A. and Sherlock,G. (2006) The Stanford Microarray Database: a user's guide. *Methods Mol. Biol.*, **338**, 191–208.
21. Ball,C.A., Awad,I.A., Demeter,J., Gollub,J., Hebert,J.M., Hernandez–Boussard,T., Jin,H., Matese,J.C., Nitzberg,M., Wymore,F. *et al.* (2005) The Stanford Microarray Database accommodates additional microarray platforms and data formats. *Nucleic Acids Res.*, **33**, D580–D582.
22. Toufighi,K., Brady,S.M., Austin,R., Ly,E. and Provart,N.J. (2005) The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. *Plant J.*, **43**, 153–163.
23. Barrett,T., Suzek,T.O., Troup,D.B., Wilhite,S.E., Ngau,W.C., Ledoux,P., Rudnev,D., Lash,A.E., Fujibuchi,W. and Edgar,R. (2005) NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Res.*, **33**, D562–D566.
24. Zimmermann,P., Hirsch–Hoffmann,M., Hennig,L. and Gruissem,W. (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.*, **136**, 2621–2632.
25. Manfield,I.W., Jen,C.H., Pinney,J.W., Michalopoulos,I., Bradford,J.R., Gilmartin,P.M. and Westhead,D.R. (2006) *Arabidopsis* Co-expression

Tool (ACT): web server tools for microarray-based gene expression analysis. *Nucleic Acids Res*., **34**, W504–W509.
26. Jen,C.H., Manfield,I.W., Michalopoulos,I., Pinney,J.W., Willats,W.G., Gilmartin,P.M. and Westhead,D.R. (2006) The *Arabidopsis* co-expression tool (ACT): a WWW-based tool and database for microarray-based gene expression analysis. *Plant J*., **46**, 336–348.
27. Steinhauser,D., Usadel,B., Luedemann,A., Thimm,O. and Kopka,J. (2004) CSB.DB: a comprehensive systems-biology database. *Bioinformatics*, **20**, 3647–3651.
28. Zimmermann,P., Hennig,L. and Gruissem,W. (2005) Gene-expression analysis and network discovery using Genevestigator. *Trends Plant Sci*., **10**, 407–409.
29. Steffens,N.O., Galuschka,C., Schindler,M., Bülow,L. and Hehl,R. (2004) AthaMap: an online resource for *in silico* transcription factor binding sites in the *Arabidopsis thaliana* genome. *Nucleic Acids Res*., **32**, D368–D372.
30. Steffens,N.O., Galuschka,C., Schindler,M., Bülow,L. and Hehl,R. (2005) AthaMap web tools for database-assisted identification of combinatorial *cis*-regulatory elements and the display of highly conserved transcription factor binding sites in *Arabidopsis thaliana*. *Nucleic Acids Res*., **33**, W397–W402.
31. Bülow,L., Steffens,N.O., Galuschka,C., Schindler,M. and Hehl,R. (2006) AthaMap: from *in silico* data to real transcription factor binding sites. *In Silico Biol*., **6**, 0023.
32. Yang,Y.H., Dudoit,S., Luu,P., Lin,D.M., Peng,V., Ngai,J. and Speed,T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*., **30**, e15.