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

Lorenz Bülow*, Martin Schindler¹ and Reinhard Hehl

Institut für Genetik, Technische Universität Braunschweig, Spielmannstraße 7, D-38106 Braunschweig, Germany
and ¹Software 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...
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>
<thead>
<tr>
<th>Stimulus</th>
<th>Class</th>
<th>No. of records</th>
<th>No. of genes</th>
<th>TAIR expression set ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xanthomonas campestris</td>
<td>Bacterial pathogen</td>
<td>17 846</td>
<td>6850</td>
<td>1005823536</td>
</tr>
<tr>
<td>Fusarium virguliforme</td>
<td>Fungal pathogen</td>
<td>18 963</td>
<td>6804</td>
<td>1005823583</td>
</tr>
<tr>
<td>Phytophthora infestans</td>
<td>Fungal pathogen</td>
<td>18 643</td>
<td>6719</td>
<td>1005823534</td>
</tr>
<tr>
<td>Powdery mildew</td>
<td>Fungal pathogen</td>
<td>10 130</td>
<td>6974</td>
<td>1005823549</td>
</tr>
<tr>
<td>TMV infected leaves 3 dpi</td>
<td>Viral pathogen</td>
<td>17 921</td>
<td>6430</td>
<td>1005823504</td>
</tr>
<tr>
<td>TMV infected leaves 4 dpi</td>
<td>Viral pathogen</td>
<td>28 369</td>
<td>7835</td>
<td>1005823602</td>
</tr>
<tr>
<td>TMV systemic leaves 14 dpi</td>
<td>Viral pathogen</td>
<td>74 320</td>
<td>9289</td>
<td>1005823505, 1005823602</td>
</tr>
<tr>
<td>Chitin 10 min</td>
<td>Elicitor</td>
<td>3939</td>
<td>1585</td>
<td>1005823605</td>
</tr>
<tr>
<td>Chitin 30 min</td>
<td>Elicitor</td>
<td>3905</td>
<td>1578</td>
<td>1005823605</td>
</tr>
<tr>
<td>Chitin 1 h</td>
<td>Elicitor</td>
<td>3837</td>
<td>1563</td>
<td>1005823605</td>
</tr>
<tr>
<td>Chitin 3 h</td>
<td>Elicitor</td>
<td>3876</td>
<td>1569</td>
<td>1005823605</td>
</tr>
<tr>
<td>Chitin 6 h</td>
<td>Elicitor</td>
<td>3888</td>
<td>1586</td>
<td>1005823605</td>
</tr>
<tr>
<td>Chitin 24 h</td>
<td>Elicitor</td>
<td>3655</td>
<td>1522</td>
<td>1005823605</td>
</tr>
<tr>
<td>cis-Jasmonate</td>
<td>Signal molecule</td>
<td>19 919</td>
<td>8027</td>
<td>1005823574</td>
</tr>
<tr>
<td>Ethylene 2 h</td>
<td>Signal molecule</td>
<td>7697</td>
<td>1487</td>
<td>1005823581</td>
</tr>
<tr>
<td>Ethylene 24 h</td>
<td>Signal molecule</td>
<td>9200</td>
<td>1423</td>
<td>1005823581</td>
</tr>
<tr>
<td>Hydrogen peroxide</td>
<td>Signal molecule</td>
<td>19 047</td>
<td>6880</td>
<td>1005823545</td>
</tr>
<tr>
<td>Methyl-jasmonate</td>
<td>Signal molecule</td>
<td>20 401</td>
<td>8160</td>
<td>1005823574</td>
</tr>
<tr>
<td>SA analog BTH</td>
<td>Signal molecule</td>
<td>9928</td>
<td>6835</td>
<td>1005823548</td>
</tr>
</tbody>
</table>

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

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 *Arabidopsis* 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**


AthaMap: from in silico data to real transcription factor binding sites. In Silico Biol., 6, 0023.