Paper 209-2012

Investigating Host Plant Resistance to Aphid Feeding through SAS® Text Miner

Ning Song, Jiawen Liu, Goutam Chakraborty, Oklahoma State University, Stillwater, OK, U.S.
James A. Anstead, Pennsylvania State University, University Park, PA, U.S.

ABSTRACT

The processes of host plant resistance to insect feeding and pathogen attack involve several complicated plant defense pathways comprising numerous regulations of pathogen-related gene expressions. The aim of this study is to examine the ethylene signaling defense pathway of melon plants. We present a novel way of applying text mining in plant resistance research literature reviews. SAS® Text Miner is employed to analyze current literature emphasis with the purpose of identifying interesting and important research trends in the field of host plant resistance to insect attacks. We show that ethylene, jasmonic acid, salicylic acid, and calcium signaling pathways are major emphases in the plant-pathogen interaction field. Additionally, SAS® Enterprise Guide® is used to analyze gene expression changes in ethylene signaling pathway.

INTRODUCTION

Aphids are one of the most harmful non-indigenous threats to agriculture in the United States. The consequences directly caused by aphids include production losses, quality decreases, and agricultural risk increases (Miller et al., 2009). For example, in the western U.S., the yield losses caused by Russian wheat aphids reached $500-$900 million in early 1990 (Bernal et al., 1993). In Iowa, a significant soybean aphid outbreak resulted in a 32% reduction of soybean yields in 2003 (Marlin et al., 2007). Aphids harm crop plants by direct feeding on plant phloem sap and spreading pathogenic viruses during feeding processes. These behaviors significantly disrupt normal plant physiology. More important, pesticides are not effective in protecting plants from aphid infection because pesticides kill aphid predators together with aphids. Therefore, it is necessary for crop plants to develop host plant resistance in response to aphid attacks.

Host plant resistance responses contain many physiologic processes that will activate and express plant pathogen-related genes to defend plants from pest damages. Some chemical compounds are commonly produced in these processes, such as ethylene (ET) (Moran et al., 2002). ET is a gaseous compound that is recognized as an important pathogen defense mediator in plant-pathogen interactions. ET participates in the plant response process in two ways: plants produce ET by activating and expressing ET generation genes, and then they activate other pathogen defense pathways by inducing the expression of related genes (Anstead et al., 2009; Thompson and Goggin, 2006). However, these defense pathways are very complex, and ET's impact on the outcome of pathogen related genes is still unclear.

In our research, we investigated host plant defense and aphid attack interactions with two kinds of host plants: virus aphid transmission (Vat) resistant melon plants, named AR 5; and virus aphid transmission susceptible melon plants, named PMR 5. The aphids we studied in this research are cotton-melon aphids named *Aphis gossypii* (*A. gossypii*), one of the most destructive aphids in their family. AR 5 melon plants contain Vat genes; therefore, they are resistant to both *A. gossypii* feeding and viruses transmitted by *A.gossypii*. On the other hand, PMR 5 melon plants do not contain Vat genes, so they are susceptible to *A. gossypii* feeding and transmission viruses brought by *A. gossypii*. The experimental approach in this study is to compare the plant pathogenic defense gene expression levels during aphid feeding in both AR 5 and PMR 5 plants. Three groups of pathogenic defense genes are studied: ET perceived genes (ETR1, ETR2, EIN2, EIN3, EIL1, ERF1, and ERS1), ET synthesis genes (ACS1, ACS2, ACS3, ACO1, and ACO2), and some other pathogen-related genes induced by ET (SSA-13, SAG-12, type1 PI, CCH, and CAM). The full name and function of each gene is listed in Table 1.

The technique we used to detect gene expression level is real-time polymerase chain reaction (RT-PCR). The complete data for gene expression levels of AR 5 and PMR 5 plants are subjected to two-sample t-tests. In the previous paper by Anstead et al (2009), equality of variance was assumed as equal; however, this is a biased assumption. Here, we present a more accurate method using SAS® Enterprise Guide®.

Moreover, the literature review is considered a significant part in research because it is important for researchers to keep up with the most current knowledge and discoveries. However, hundreds of new papers come out every week, and it is challenging for researchers to review all informative papers. Here, we present a novel way of applying text mining to plant pathogen science literature reviews by using SAS® Enterprise Miner™ 6.1. Text mining is a very efficient way to find out the patterns of current research trends and the most informative papers in the field of plant resistance to aphid attack.

Abbreviation	Gene Full Name	Gene Encoded Protein Function
ETR1	ethylene receptor 1	perceive ET presence
ETR2	ethylene receptor 2	perceive ET presence
ERS1	ethylene response sensor 1	perceive ethylene receptor presence
EIN2	protein ethylene insensitive 2	a nuclear transcription factor that initiates downstream transcriptional cascades for ethylene responses
EIN3	protein ethylene insensitive 3	a nuclear transcription factor that initiates downstream transcriptional cascades for ethylene responses
EIL1	ethylene insensitive 3-like 1	a nuclear transcription factor EIN3 like protein
ERF1	ethylene response factor 1	an ET response element binding protein
ACS1	1-aminocyclopropane-1-carboxylate synthase-like 1	ethylene-forming enzyme
ACO1	1-aminocyclopropane-1-carboxylate oxidase 1	ethylene-forming enzyme
Type 1 PI	Type 1 proteinase inhibitor	inhibite proteinase function
SSA-13	senescence-associated protein 13	pathogenesis-related protein 13
SAG-21	senescence-associated gene 21	pathogenesis-related protein 21
ACS2	1-aminocyclopropane-1-carboxylate synthase 2	enzyme to produce 1-aminocyclopropane-1-carboxylic acid
CCH	copper chaperone	copper delivering protein, mainly located in plant pholem cells
ACO2	1-aminocyclopropane-1-carboxylate oxidase 2	ethylene-forming enzyme
ACS3	1-aminocyclopropane-1-carboxylate synthase 3	enzyme to produce 1-aminocyclopropane-1-carboxylic acid
CAMTA1	calmodulin-binding transcription activator 1	transcription activator that mediates host plant responds to aphid stresses

Table 1. Gene full name and encoded protein function.

MATERIALS AND METHODS

Two-Sample T-Test

To examine the results of host plant resistance responses to *A. gossypii* between AR 5 and PMR 5 plants, we conducted experiments with 17 groups of gene expressions and analyzed the data using two-sample t-tests in SAS® Enterprise Guide®. In each group, we have control (AR 5 and PMR 5 plants not treated with aphids) and treatment (AR 5 and PMR 5 plants treated with aphids). For each set of experiments, four replicates were performed and two time points were recorded, 6 hours and 24 hours. The raw data were analyzed using the Pfaffl method (Pfaffl, 2001), for which a relative gene expression value was calculated for each gene's treatment and control samples at both time points.

Anstead et al. paper assumed that the equality of variances were equal. Here, F-tests in SAS® Enterprise Guide® will present correct decisions regarding equality of variances.

Text Mining

Besides the two-sample t-tests, we also investigated the recent research literature on host plant resistance using text mining capabilities available in SAS® Enterprise Miner™ 6.1. The corpus was created by searching current literature with key words "plant resistance to aphids" in Pubmed. Pubmed is an online website comprised of more than 21 million articles from biological or biomedical science journals. Of 389 search results, 249 articles published after 2004 were selected. The abstracts of these papers were collected from online PDF documents, and a sas7bdat textual data file containing these abstracts was created in SAS® Enterprise Guide®. All 249 abstracts were imported and read in SAS® Enterprise Miner™ 6.1. No further data cleaning was performed.

Text mining starts with text parsing, which identifies unique terms in the text variable and identifies parts of speech, entities, synonyms, and punctuation (SAS® Enterprise Miner™ Help). The terms identified from text parsing are used to create a term-by-document matrix with terms as rows and documents as variables. A typical text mining problem has more terms than documents, resulting in a sparse rectangular terms-by-document matrix (SAS® Enterprise Miner™ Help). Stop lists help in reducing the number of rows in the matrix by dropping some of the terms (Miller, 2005).

A stop list is a dictionary of terms that are ignored in the analysis (SAS® Enterprise Miner™ Help). A standard stop list removes words such as the, and, of, etc. However, a user can create custom stop lists to achieve better text mining results (Cerrito, 2006). In this research, we used the standard stop list supplied with SAS® Text Miner. Singular Value Decomposition (SVD) can be used to reduce the dimensionality by transforming the matrix into a lower dimensional and more compact form (SAS® Enterprise Miner™ Help). However, a careful decision must be made on how many SVD high dimensions (*k*) to use. A high number for *k* can give better results, but high computing resources are required. It is customary to try different values for the number of dimensions and compare the results (SAS® Enterprise Miner™ Help). As a general rule, smaller values of *k* (2 to 50) are useful for clustering, and larger values (30 to 200) are useful for prediction or classification (Sanders, 2004). In this study, we used the default options in SAS® Text Miner for SVD dimensions. Each term identified in text parsing was given a weight based on different criteria. The term weights help in identifying important terms (Battioui, 2008). The default setting for this property is Entropy. Using this setting, terms that appear more frequently will be weighted lower compared to terms that appear less frequently (Cerrito, 2006).

A clustering technique is used for text categorization. Using this technique, documents are classified into groups such that those within any one group are closely related and those in different groups are not closely related (Cerrito,

2006). The terms along with their weights are used for creating these groups. Each group or cluster is represented by a list of terms, and those terms will appear in most of the documents within the group (Battioui, 2008). SAS® Text Miner uses an Expectation-Maximization algorithm for clustering. For all the four decades, we first used 20 for the maximum number of cluster property and subsequently modified this property based on clarity of clustering results.

RESULTS

Two-Sample T-Test

Based on the reports generated by SAS® Enterprise Guide®, 67.6% equality of variances were equal and 32.4% equality of variances were unequal. As shown in Table 2, for 6-hour records, five genes had unequal variances and from 24-hour records, six genes had unequal variance. (A detailed report for each two-sample t-test is not shown with this paper.)

Three out of 17 ethylene-related genes were up-regulated under the treatment of the aphid attacking. For the 6-hour treatment, two genes showed a significant difference between two plants, ERF1 and ACO2. For the 24-hour treatment, only ERS1 showed a significant difference. (Refer to Figure 1, Figure 2, and Figure 3.)

	6-hour Result		24-hour Result	
Gene	Equality of Variance	Significance	Equality of Variance	Significance
ETR1	Equal	Insignificant	Unequal	Insignificant
ETR2	Unequal	Insignificant	Equal	Insignificant
ERS1	Equal	Insignificant	Equal	Significant
EIN2	Equal	Insignificant	Unequal	Insignificant
EIN3	Equal	Insignificant	Equal	Insignificant
EIL1	Equal	Insignificant	Equal	Insignificant
ERF1	Equal	Significant	Unequal	Insignificant
ACS1	Unequal	Insignificant	Equal	Insignificant
ACO1	Unequal	Insignificant	Unequal	Insignificant
TYPE P1	Unequal	Insignificant	Unequal	Insignificant
SSA 13	Unequal	Insignificant	Equal	Insignificant
SAG 21	Equal	Insignificant	Unequal	Insignificant
ACS2	Equal	Insignificant	Equal	Insignificant
CCH	Equal	Insignificant	Equal	Insignificant
ACO2	Equal	Significant	Equal	Insignificant
ACS3	Equal	Insignificant	Equal	Insignificant
CAM	Equal	Insignificant	Equal	Insignificant

Table 2. Summary list of equality of variance for each gene at two time points. The significance results for each two-sample t test are indicated by highlighting.

Plant	N	Mean	Std Dev	Std Err	Minimum	Maximum
ar5	4	4.6332	1.6993	0.8496	3.0665	6.7906
pmr5	4	0.4287	0.5753	0.2876	-0.3761	0.9517
Diff (1-2)		4.2045	1.2686	0.8970		

Plant	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
ar5		4.6332	1.9293	7.3372	1.6993	0.9626	6.3358
pmr5		0.4287	-0.4867	1.3441	0.5753	0.3259	2.1449
Diff (1-2)	Pooled	4.2045	2.0096	6.3994	1.2686	0.8175	2.7935
Diff (1-2)	Satterthwaite	4.2045	1.6259	6.7832			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	6	4.69	0.0034
Satterthwaite	Unequal	3.6787	4.69	0.0115

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	3	3	8.73	0.1084

Figure 1.Two-sample t-test result of Gene ERF1 for 6-hour treatment.

Plant	N	Mean	Std Dev	Std Err	Minimum	Maximum
ar5	4	1.0645	0.7929	0.3965	0.1470	2.0838
pmr5	4	0.00382	0.6431	0.3215	-0.3754	0.9601
Diff (1-2)		1.0607	0.7219	0.5105		

Plant	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
ar5		1.0645	-0.1972	2.3263	0.7929	0.4492	2.9565
pmr5		0.00382	-1.0195	1.0271	0.6431	0.3643	2.3978
Diff (1-2)	Pooled	1.0607	-0.1884	2.3098	0.7219	0.4652	1.5897
Diff (1-2)	Satterthwaite	1.0607	-0.2014	2.3228			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	6	2.08	0.0830
Satterthwaite	Unequal	5.7547	2.08	0.0850

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	3	3	1.52	0.7390

Figure 2. Two-sample t-test result of Gene ACO2 for 6-hour treatment.

Plant	N	Mean	Std Dev	Std Err	Minimum	Maximum
ar5	4	-0.0865	0.5213	0.2606	-0.7865	0.4532
pmr5	4	1.6262	1.3077	0.6539	0.7044	3.5600
Diff (1-2)		-1.7127	0.9955	0.7039		

Plant	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
ar5		-0.0865	-0.9160	0.7430	0.5213	0.2953	1.9436
pmr5		1.6262	-0.4547	3.7071	1.3077	0.7408	4.8759
Diff (1-2)	Pooled	-1.7127	-3.4351	0.00966	0.9955	0.6415	2.1921
Diff (1-2)	Satterthwaite	-1.7127	-3.6809	0.2554			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	6	-2.43	0.0509
Satterthwaite	Unequal	3.9299	-2.43	0.0729

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	3	3	6.29	0.1651

Figure 3. Two-sample t-test result of Gene ERS1 for 24-hour treatment.

In summary, two-sample t-test analysis suggests that current data are not informative enough to draw a strong conclusion about host plant resistance to aphid attack. In order to propose an insightful future step for this research, we will present text mining analysis showing the current research trends in this field in the following section.

Text Mining

The textual data set of 249 article abstracts was read in SAS® Enterprise Miner™ 6.1. All of these abstracts were used to cluster the research trends using the text mining node (refer to Figure 4). As shown in the partial view of the properties used for text mining (refer to Figure 5), default settings of properties in the text mining node were used and a 7-cluster solution was identified (refer to Figure 7). Figure 6 shows part of the Term-Document Matrix calculated based on the discriminative power of the weights. The terms are listed in descending order of frequencies.

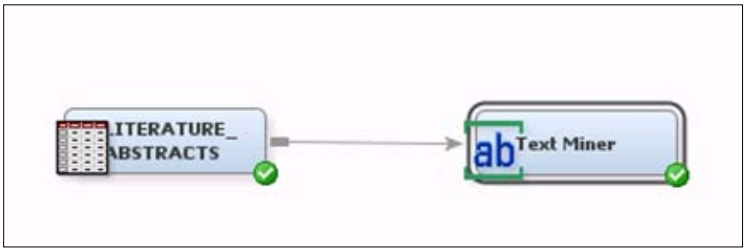

Figure 4. Application of text mining on literature abstract textual data.

Property	Value
Force Run	No
Parse	
Parse Variable	F1
Language	ENGLISH
Stop List	Sashelp.stoplst
Start List	
Stem Terms	Yes
Terms in Single Document	No
Punctuation	No
Numbers	No
Different Parts of Speech	No
Ignore Parts of Speech	
Noun Groups	Yes
Synonyms	Sashelp.engsynms
Find Entities	No
Types of Entities	
Transform	
Compute SVD	Yes
SVD Resolution	Low
Max SVD Dimensions	20
Scale SVD Dimensions	No
Frequency weighting	Log
Term Weight	Entropy
Roll up Terms	No
No. of Rolled-up Terms	100
Drop Other Terms	No
Cluster	
Automatically Cluster	Yes
Exact or Maximum Number	Maximum
Number of Clusters	10
Cluster Algorithm	EXPECTATION-MAXIMIZATION
Ignore Outliers	No
Hierarchy Levels	.
Descriptive Terms	5
What to Cluster	SVD Dimensions

Figure 5. Properties of text mining node.

Terms

TERM	FREQ ▼	# DOCS	KEEP	WEIGHT	ROLE	ATTRIBUTE
be	192	134	☐	0.13		Alpha
plant	150	113	☑	0.159		Alpha
ethylene	113	85	☑	0.214		Alpha
response	103	83	☑	0.211		Alpha
signal	84	57	☑	0.285		Alpha
gene	68	56	☑	0.28		Alpha
expression	57	43	☑	0.335		Alpha
mutant	56	36	☑	0.377		Alpha
acid	54	36	☑	0.368		Alpha
that	52	51	☐	0.289		Alpha
defense	50	38	☑	0.355		Alpha
protein	50	37	☑	0.36		Alpha
ja	48	40	☑	0.35		Alpha
pathway	43	38	☑	0.348		Alpha
feed	42	33	☑	0.379		Alpha
insect	41	36	☑	0.358		Alpha
arabidopsis	38	35	☑	0.361		Alpha
not	35	32	☑	0.377		Alpha
increase	34	29	☑	0.401		Alpha
growth	32	21	☑	0.475		Alpha
sa	32	28	☑	0.403		Alpha
induce	32	28	☑	0.403		Alpha
indicate	31	30	☑	0.386		Alpha
level	30	28	☑	0.4		Alpha
have	28	25	☑	0.423		Alpha
role	26	24	☑	0.429		Alpha
also	25	24	☐	0.427		Alpha
cell	24	16	☑	0.518		Alpha

Figure 6. Partial view of customized stop-list based on Term-Document Matrix.

Clusters

#	DESCRIPTIVE TERMS	FREQ	PERCENTAGE	RMS STD.
1	+ heat-induce, hyponastic, + heat, cal cium, oxidative	17	0.0682730923...	0.1622668...
2	sequence, family, + protein, + factor, + have	40	0.1606425702...	0.2103263...
3	+ stem, + fruit, + leaf, + express, exp ression	50	0.2008032128...	0.2137762...
4	oserf3, + suppress, + transcript, + kin ase, + activity	17	0.0682730923...	0.2111363...
5	+ mutant, + mutation, wild-type, + eff ect, + indicate	30	0.1204819277...	0.1845802...
6	jasmonic acid, jasmonic, salicylic acid, s alicylic, sa	16	0.0642570281...	0.1699571...
7	+ signal, + defense, + insect, + pathw ay, + feed	79	0.3172690763...	0.1826201...

Figure 7.Descriptive terms of each cluster

Among the seven clusters (refer to Figure 7), the most common concepts in these articles are grouped in cluster 7. The key words for this cluster are signal, defense, insect, pathway, and feed. This cluster suggests that insect feeding commonly triggers signal pathways in the host plant. The second-largest cluster is cluster 3, in which the key words are stem, fruit, leaf, express, and expression. This cluster indicates that the signal pathways are usually located in these parts of plants. The next largest cluster is cluster 2, in which the key words are sequence, family, protein, factor, and have. This cluster suggests that there are some protein families identified in plant defense responses. The next largest cluster is cluster 5, in which the key words are mutant, mutation, wild-type, effect, and indicate. This cluster shows that mutation study is a common method used in plant resistance research. Moreover, Cluster 1 conveys the information that oxidative related signaling takes up 7% of the current research in host plant and insect interactions.

Documents

F1	CLUSTE...
S-nitrosoglutathione reductase (GSNOR) reduces the nitric oxide (NO) adduct S-nitrosoglutathi	7.0
In plants, GSNOR has been found to be important in resistance to bacterial and fungal pathog	7.0
Using a virus-induced gene silencing (VIGS) system, the activity of GSNOR in a wild tobacco sp	7.0
Furthermore, GSNOR is required for methyl jasmonate (MeJA)-induced accumulation of defenc	7.0
This work highlights the important role of GSNOR in plant resistance to herbivory and jasmonat	7.0
Plant Ca2+ signals are involved in a wide array of intracellular signaling pathways after pest in	7.0
Ca2+-binding sensory proteins such as Ca2+-dependent protein kinases (CPKs) have been pr	7.0
To investigate the roles CPKs play in a herbivore response-signaling pathway, we screened th	7.0
CPK13 strongly phosphorylated only HsfB2a, irrespective of the presence of Ca2+. Furthermo	7.0
These results reveal the involvement of two Arabidopsis CPKs (CPK3 and CPK13) in the herbiv	7.0
This cascade is not involved in the phytohormone-related signaling pathways, but rather direct	7.0
In plants, ethylene and jasmonate control the defense responses to multiple stressors, includin	7.0
Among the defense proteins known to be regulated by ethylene is maize insect resistance 1-cy	7.0
To resolve this discrepancy and elucidate the role of ethylene and jasmonate in the signaling o	7.0
However, these studies are focused on the plant responses to feeding by well-studied caterpill	7.0
Our work clearly shows that JA signaling, but not JA/ET signaling, is involved in plant tolerance	7.0
Pathogen infection, mechanical wounding, and oxidative stress induce expression of TPK1b, a	7.0
TPK1b functions independent of JA biosynthesis and response genes required for resistance to	7.0
Three residues in the activation segment play a critical role in the kinase activity and in vivo sig	7.0
We analyzed the interaction between Arabidopsis and western flower thrips (Frankliniella occid	7.0
Comparative transcriptome analyses suggested a strong relationship between thrip feeding an	7.0
The JA content of WT plants was significantly increased after thrip feeding. Moreover, coi1-1,	7.0
Application of JA to WT plants before thrip feeding enhanced the plants' feeding tolerance. JA	7.0
Our results indicate that JA plays an important role in Arabidopsis in terms of response to, and	7.0
These defenses can protect trees against insect herbivory and fungal colonization. The phytoh	7.0
As a consequence, plant perception of and responses to PFIs differ from plant interactions wit	7.0
Transcriptome-wide analyses of gene expression are currently being applied to characterize pl	7.0
Recent studies indicate that PFIs induce transcriptional reprogramming in their host plants, and	7.0
Plant responses to these insects appear to be regulated in part by the salicylate, jasmonate, a	7.0
As additional transcript profiling data become available, forward and reverse genetic approach	7.0
To understand how plants integrate pathogen- and insect-induced signals into specific defense	7.0
Monitoring the signal signature in each plant-attacker combination showed that the kinetics of	7.0
Comparison of the transcript profiles revealed that consistent changes induced by pathogens	7.0
Notably, although these four attackers all stimulated JA biosynthesis, the majority of the chan	7.0
Plant responses to enemies are coordinated by several interacting signaling systems. Molecular	7.0

Figure 8. Partial result of filtered documents showing literatures grouped in Cluster 7.

The detailed articles grouped into cluster 7 are displayed by Filter Document function in SAS® Enterprise Miner™ 6.1. As shown in Figure 8, these signal pathways include ethylene signaling, jasmonic acid signaling, and calcium signaling.

Documents

F1	CLUSTE...
Comments	3.0
CPK3 was also suggested to be involved in a negative feedback regulation of the cytosolic Ca2	3.0
This protein is constitutively expressed in the insect-resistant maize (Zea mays) genotype Mp7	3.0
Immunoblot analysis of Mir1-CP accumulation and quantitative reverse-transcriptase polymera	3.0
The results also suggest that jasmonate functions upstream of ethylene in the Mir1-CP expres	3.0
In contrast, we have focused on a minute insect pest, the western flower thrips (Frankliniella o	3.0
TPK1b RNAi seedlings are also impaired in ethylene (ET) responses. Notably, susceptibility to B	3.0
The enzyme 1-aminocyclopropane-1-carboxylate oxidase (ACO) catalyzes the final step in eth	3.0
Using an Arabidopsis anti-ACO antibody we determined that ACO is constitutively expressed in	3.0
These insects have evolved to survive on a nutritionally imbalanced diet of phloem sap, and to	3.0
Transcript profiling studies also suggest that PFIs induce cell wall modifications, reduce photos	3.0
Arabidopsis plants were exposed to a pathogenic leaf bacterium (Pseudomonas syringae pv. t	3.0
Of all consistent changes induced by A. brassicicola, Pieris rapae, and E occidentalis, more tha	3.0
It is interesting that although all seven genes displayed their R-specific patterns in the treated	3.0
Ethylene was not responsible for any of the specific patterns of expression. R collected from di	3.0
The combined pharmacological application of JA and the ET precursor, 1-aminocyclopropane-1-	3.0
Despite the fact that maize actively mounts a defense response to ECB stem feeding, no differ	3.0
The effects of root hypoxia on ethylene biosynthesis and perception have been documented i	3.0
Gravistimulation increased ethylene production in both lower and upper halves of the stems wit	3.0
Expression patterns of three different 1-aminocyclopropane-1-carboxylate (ACC) synthase (A	3.0
One of the ACS genes (Am-ACS3) was abundantly expressed in the bending zone cortex at th	3.0
Am-ACS3 was not expressed in vertical stems or in other parts of (gravistimulated) stems, leav	3.0
Am-ACS3 was strongly induced by indole-3-acetic acid (IAA) but not responsive to ethylene.	3.0
The Am-ACS3 expression pattern strongly suggests that Am-ACS3 is responsible for the obser	3.0
Am-ACS1 also showed increased expression in gravistimulated and IAA-treated stems althoug	3.0
In contrast to Am-ACS3, Am-ACS1 was also expressed in non-bending regions of vertical and	3.0
Expression of both Am-ACO and Am-ETR/ERS was responsive to ethylene, suggesting regulati	3.0
Am-ACO expression and in vivo ACO activity, in addition, were induced by IAA, independent o	3.0
IAA-induced growth of vertical stem sections and bending of gravistimulated flowering stems w	3.0
Ethylene-inducible PR genes are expressed constitutively in roots and cultured cells even when	3.0
Using RNase protection analysis, the mRNAs of LeETR1, LeETR2 and NR were quantified in tiss	3.0
LeETR1 was expressed constitutively in all plant tissues examined. LeETR2 mRNA was express	3.0
NR expression was developmentally regulated in floral ovaries and ripening fruit. Notably, hor	3.0
Furthermore, the abundance of mRNAs for all three LeETR genes remained uniform in multiple	3.0
Expression of the basic chitinase is organ-specific and age-dependent in Arabidopsis. A high co	3.0

Figure 9. Partial results of filtered documents showing literatures grouped in Cluster 3.

The secondary largest cluster indicates that the major defense pathways exist in the stem, fruit, and leaf parts of host plants. As shown in Figure 9, the major genes studied in this cluster are ACO, ACS, ACC, and ETR, etc. Therefore, although our research data shows insignificant gene up-regulation results for most of the genes, 20% of the literature is still focusing on the study of them.

Documents

F1	CLUSTE...
Ethylene responsive factors (ERFs) are a large family of plant-specific transcription factors tha	2.0
In vitro kinase assays of CPK3 protein with a suite of substrates demonstrated that the protei	2.0
The tomato protein kinase 1 (TPK1b) gene encodes a receptor-like cytoplasmic kinase localized	2.0
Members of the Pinaceae family have complex chemical defense strategies. Conifer defenses a	2.0
However, very little is known about the genes involved in ethylene formation in conifer defens	2.0
We cloned full-length and near full-length ACO cDNAs from three conifer species, Sitka spruce	2.0
Immunolocalization showed cytosolic ACO is predominantly present in specialized cell types of t	2.0
The relationship between phloem-feeding insects (PFIs) and plants offers an intriguing exampl	2.0
Analysis of global gene expression profiles demonstrated that the signal signature characteristi	2.0
R extensively modified wound-induced responses by suppressing wound-induced transcripts (t	2.0
Ethylene responsive factors (ERFs) are a large family of plant-specific transcription factors tha	2.0
Members of the Pinaceae family have complex chemical defense strategies. Conifer defenses a	2.0
Ethylene receptor family in Arabidopsis consists of five components, ETR1, ERS1, ETR2, ERS2	2.0
Some receptors contain a well-conserved carboxy terminal histidine (His) kinase domain and so	2.0
The ETR1 localizing to endoplasmic reticulum and activating CTR1 negatively regulates the eth	2.0
F box protein EBF1/EBF2 accelerate the EIN3 degeneration. ERF1, one of transcription factor	2.0
Hormones are important regulators of plant growth and development. In Arabidopsis, percepti	2.0
This mechanism may create a novel signal transfer from endoplasmic reticulum-associated ETR	2.0
Ethylene responses in Arabidopsis are controlled by the ETR receptor family. The receptors fu	2.0
OMT was stimulated in LAR but not in SAR(T) and SAR(S). The four classes of acidic and basic	2.0
In glycoprotein-treated plants, expression of the acidic and basic PR proteins in LAR and SAR	2.0
Gene expression of one group of PR proteins is known to be mediated by phytohormone ethyl	2.0
We discuss the mechanisms of this pathogen-independent expression of PR genes and describ	2.0
Genes of PR-1 and -5 proteins have now been identified in the genomes of various species of	2.0
Ethylene perception in plants is co-ordinated by multiple hormone receptor candidates sharing	2.0
Two tomato homologs of the Arabidopsis ethylene receptor ETR1 were cloned from a root cDN	2.0
LeETR1 and LeETR2 contained all the major structural elements of two-component regulators, i	2.0
We have genetically mapped the etr mutation and by chromosome walking have isolated an 18	2.0
These two domains are separated by a single 24 amino acid hydrophobic domain. A model is pr	2.0
Plants synthesize a number of antimicrobial proteins in response to pathogen invasion and envi	2.0
We have cloned and determined the nucleotide sequence of the genes encoding the acidic and	2.0
Both chitinases are encoded by single copy genes that contain introns, a novel feature in chitin	2.0
Exposure of plants to ethylene induced high levels of systemic expression of basic chitinase wit	2.0

Figure 10. Partial results of filtered documents showing literatures grouped in Cluster 2.

Moreover, cluster 2 shows that 16% of articles focused on researching plant defense-related protein families. When the content of these articles is displayed by Filter Document function, it is clear that the target genes in these papers are ERF, ERS, ETR, ACO and other pathogen-related proteins and transcription factors (refer to Figure 10).

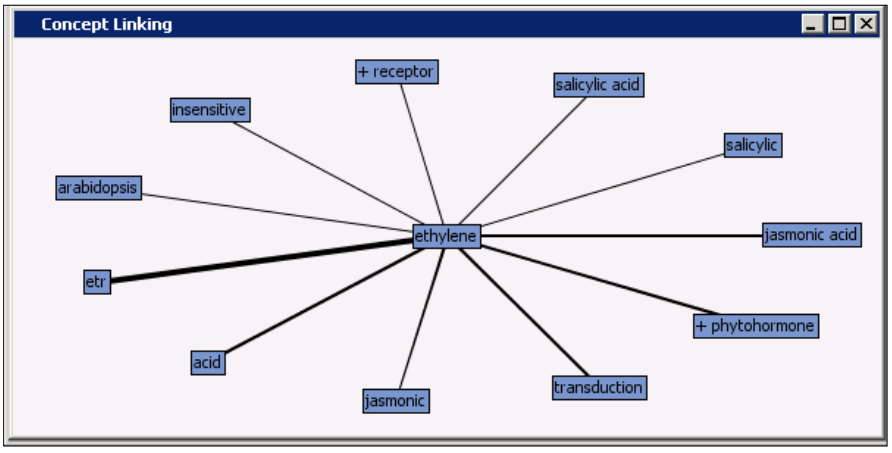

Figure 11. Concept links for ethylene

Furthermore, Figure 11 shows the concept links of ethylene in these documents. The term “ethylene” is strongly associated with the term “ETR”. ETR is a gene that encodes an ethylene receptor, and its expression level was also studied in our research. The two-sample t-test of the ETR gene, however, didn’t show a significant difference between aphid resistance and susceptible plants. Although our current data didn’t show a significant response, text mining results indicated that ETR is still an important and interesting gene to study.

The term “ethylene” is also linked to some other interesting terms, such as jasmonic acid, salicylic acid, transduction, and receptor. These links recommend that future researches pay attention to the biological effects of these suggested terms.

CONCLUSION

It is clear that the ethylene signaling pathway plays a critical role in host plant resistance to aphid attacks. In addition, jasmonic acid signaling, salicylic acid signaling, calcium signaling, and oxidative-related signaling are also interesting pathways in plant-pathogen interactions. Although only three genes have been demonstrated to be significantly up-regulated by aphid feeding treatment, the study of the other 14 genes is recommended to continue in order to collect more data for further conclusions.

The biological mechanism of ethylene signaling pathways is still a hot topic in the current research field, and its interaction with other plant-pathogen pathways is recommended for the next step of this project.

ACKNOWLEDGEMENTS

We would like to thank Dr. Gary A. Thompson of the Pennsylvania State University for his generous support in providing research data.

REFERENCES

Anstead, J., Samuel, P., Song, N., Wu, C., Thompson, G., and and Goggin, F. (2009). Activation of ethylene-related genes in response to aphid feeding on resistant and suceptible melon and tomato plants. Entomologia Experimentalis et Applicata *134*.

Anstead, J., Samuel P., Song N., Wu C., Thompson G., Goggin F. (2009). Activation of ethylene-related genes in response to aphid feeding on resistant and suceptible melon and tomato plants. Entomologia Experimentalis et Applicata *134*, 11.

Battioui, C. (2008). A Text Miner analysis to compare internet and medline information about allergy medications. SAS Regional Conference.

Bernal, J.D., Gonzalez, E.T., Natwick, J.G., Loya, R., Leon-Lopez, and Bendixen, a.W.E. (1993). Natural enemies of Russian wheat aphid identified in California. California Agriculture *47*.

Cerrito, B.P. (2006). Introduction to Data Mining using SAS® Enterprise Miner. SAS Publishing.

Marlin, E.R., O'Neal, M., and and Pedersen, P. (2007). Soybean Aphids in Iowa.

Miller, G.L., Favret, C., Carmichael, A., and Voegtlin, D.J. (2009). Is there a cryptic species within Aulacorthum solani (Hemiptera: Aphididae)? Journal of economic entomology *102*, 398-400.

Miller, W.T. (2005). Data and Text Mining-A Business Applications Approach. Pearson Pentice Hall.

Moran, P.J., Cheng, Y., Cassell, J.L., and Thompson, G.A. (2002). Gene expression profiling of Arabidopsis thaliana in compatible plant-aphid interactions. Arch Insect Biochem Physiol *51*, 182-203.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic acids research *29*, e45.

Sanders, A., DeVault, C. (2004). Using SAS® at SAS: The Mining of SAS Technical Support. SUGI 29.

Thompson, G.A., and Goggin, F.L. (2006). Transcriptomics and functional genomics of plant defence induction by phloem-feeding insects. Journal of experimental botany *57*, 755-766.

TRADEMARKS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Ning Song, Email: nings@ostatemail.okstate.edu

Jiawen Liu, Email: Jiawen.liu@okstate.edu

Dr. Goutam Chakraborty, Email: goutam.chakraborty@okstate.edu

James A. Anstead, Email: jaa25@psu.edu

Brief Bios:

Ning Song is a Master's student in Management Information Systems at Oklahoma State University. She is a SAS® Certified Base Programmer for SAS® 9 and has one and a half years' experience with SAS® software. She also holds a Master's degree from Oklahoma State University in Biochemistry and Molecular Biology.

Jiawen Liu is a graduate student in Management Information Systems at Oklahoma State University. She is Base SAS® programming certified and has been using SAS® for a year.

Goutam Chakraborty is a Professor of Marketing and founder of the SAS®/OSU Data Mining Certificate and the SAS®/OSU Business Analytics Certificate at Oklahoma State University. He has published in many journals such as *Journal of Interactive Marketing*, *Journal of Advertising Research*, *Journal of Advertising*, *Journal of Business Research*, etc. He chaired the national conference for direct marketing educators in 2004 and 2005 and co-chaired the M2007 Data Mining Conference. He is also a Business Knowledge Series instructor for SAS®.

James A. Anstead is a Research Associate at Pennsylvania State University. His research focuses on plant-insect interactions, in particular those involving phloem feeding insects. He has been a SAS user for a number of years.