Application Note

METABOLOMICS GENOMICS INFORMATICS PROTEOMICS

Metabolomic Profiling of Bacterial Leaf Blight in Rice

Authors

Steven Fischer
Senior Applications Chemist
Agilent Technologies
Santa Clara, California U.S.A

Theodore Sana
Senior Research Scientist
Agilent Laboratories
Santa Clara, California U.S.A

Abstract

Rice (*Oryza sativa* and *Oryza glaberrima*) is one of the world's most important staple crops, providing food for more than 3 billion people. Bacterial leaf blight (BLB) of rice, caused by the *Xanthomonas oryzae pv. oryzae* (*Xoo*) bacteria, leads to crop losses of up to 50%. Currently, the use of resistant rice cultivars is the most economical and effective way to combat BLB. The interaction between *Xoo* and rice is governed by resistance genes in rice and corresponding pathogenic avirulence genes in *Xoo*.

In order to better understand the mechanism of infection and immunity of rice to BLB, a study was undertaken to identify metabolites that are related to infection and resistance. Two rice lines were studied: TP309 which is susceptible to infection by the *Xoo* bacterial strain PXO99, and TP309-Xa21, a resistant transgenic line. In addition, the effect of a *raxST* gene knock-out in PXO99 was evaluated for its effect on resistance in TP309-Xa21. Appropriate controls were included.

A two-step LC/MS approach was employed. Rapid differential expression analysis of samples using time-of-flight (TOF) mass spectrometry (MS) was followed by targeted identification of differentially expressed metabolites using quadrupole time-of-flight (Q-TOF) MS/MS. Clear differences in the metabolite profiles of the different rice/bacteria conditions were detected. Based on relatively few metabolites, the rice lines and state of infection were clearly distinguishable.

Introduction

Rice (*Oryza sativa* and *Oryza glaberrima*) is the primary food for more than 3 billion people worldwide. Over 600 million people derive more than half of their calories from rice. It is the third largest commercial crop behind wheat and corn. In 2005, 700 million metric tons were produced world wide with a market value of US$ 120 billion. It is estimated that 50% of the potential yield of the world rice crop is lost to diseases caused by bacteria, fungi and viruses. In 2005, 300 million metric tons were lost due to disease.

One of the most serious bacterial diseases of rice in Africa and Asia is bacterial leaf blight (BLB) caused by *Xanthomonas oryzae pv. oryzae* (*Xoo*) (Figure 1). BLB is one of the oldest recorded rice diseases and has been problematic for over a century. *Xoo* spreads rapidly from rice plant to rice plant and from field to field in water droplets. Infected leaves develop lesions, yellow, and wilt in a matter of days. In severely affected fields, bacterial blight can wipe out half a farmer's rice crop.

Breeding and deployment of resistant cultivars carrying major resistance genes has been the most effective approach to combating BLB. One such gene, *Xa21,* was successfully cloned into a rice variety Taipei 309 (TP309).[1] Once cloned, *Xa21* can be passed on to the next generation through self-fertilization.

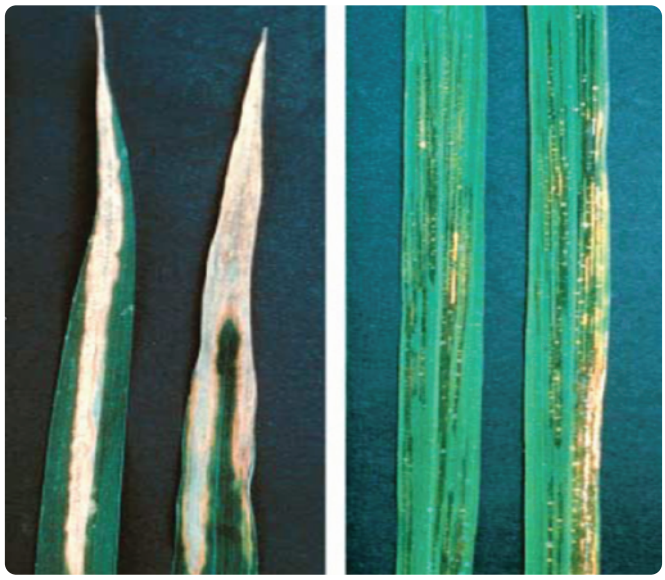


Figure 1. Rice leaves infected by bacterial leaf blight (left) and uninfected rice leaves (right).

The protein product of the gene *Xa21* carries both a leucine-rich repeat motif and a serine-threonine kinase-like domain, suggesting a role in cell-surface recognition of a pathogen ligand and subsequent activation of an intracellular defense response. This receptor directly or indirectly 'recognizes' a signal generated via a corresponding avirulence (*avr*) gene product encoded by the pathogen, in this case the AvrXa21 peptide of *Xoo*. The formation of this putative receptor–ligand complex is postulated to initiate a signaling cascade culminating in defense responses that halt the pathogen's progress.[2]

AvrXa21 must be present for the resistant rice line TP309-Xa21 to elicit an immune response and not be infected by *Xoo*. The *raxST* gene in *Xoo* encodes for a sulfotransferase-like protein that is necessary for the production of the AvrXa21 peptide. Two bacterial strains were used in this experiment. The pathogenic bacterial *Xoo* strain PXO99 includes the *raxST* gene and produces the AvrXa21 peptide (Figure 2). The *raxST* knock-out strain, PXO99-raxST^{-}, does not produce the AvrXa21 peptide.

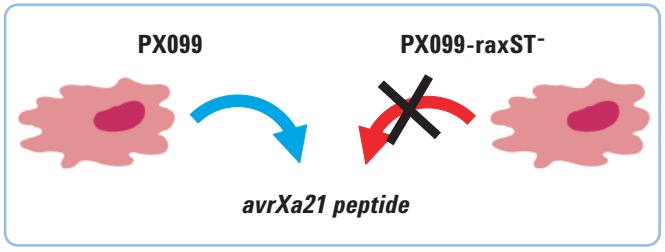


Figure 2. The gene *raxST* in *Xoo* encodes for a sulfotransferase-like protein that is necessary for production of the AvrXa21 peptide, which is what elicits an immune response in the resistant rice line TP309-Xa21. Two bacterial strains were used in this experiment. The wild-type PXO99 includes the *raxST* gene and produces AvrXa21. The *raxST* knock-out, PXO99-raxST^{-}, lacks the *raxST* gene and does not produce AvrXa21.

To gain greater insight into the mechanisms of infection and immunity of rice to BLB, a study employing LC/MS metabolite profiling was undertaken to find and identify metabolites related to infection and resistance. An LC/MS system composed of an Agilent 1200 Series LC and Agilent 6210 Time-of-Flight LC/MS was selected for the profiling work based on its sub-2-ppm mass accuracy, outstanding reproducibility, and robustness. Agilent GeneSpring MS bioinformatics software was used to analyze the complex, multi-class data generated by the study.

An Agilent LC/MS system incorporating an Agilent 1200 Series LC and Agilent 6510 Quadrupole Time-of-Flight LC/MS was chosen to identify metabolites that were found to have statistically significant variations in abundance between the control and experiment samples. This system was selected because of its combination of accurate-mass measurements and MS/MS spectra. The METLIN Personal metabolite database was used to narrow the list of possible identities during the identification process.

Two rice lines (TP309 and transgenic TP309-Xa21) and two bacterial strains (PXO99 and PXO99-raxST$^-$) were included in the study along with controls. As shown in Table 1, a total of seven different classes were studied. Due to natural variations in both the rice and bacteria, multiple biological replicates were necessary. Based on previously demonstrated reproducibility of the LC/MS system, technical replicates were not necessary in this study.

Several possible biomarkers involved in the elicitation of defense to bacterial infection of TP309 and the resistance of TP309-Xa21 were identified. Based on relatively few metabolites, the two rice lines, the state of infection, and state of infection within each rice line were all clearly distinguishable.

Table 1. Conditions tested and number of biological replicates used for each condition.

Condition (Class)	TP309 (Wild-type)	TP309-Xa21 (Transgenic)
PXO99 (wild-type)	6	6
Mock treatment control	6	6
No treatment (NT) control	6	6
PXO99-raxST$^-$ (*raxST* knock-out)	NA	6

Experimental

Inoculation of rice

The *Xoo* bacterial strains (PXO99 and PXO99-raxST$^-$) were grown for 72 hours at 30°C on peptone sucrose agar (Tsuchiya et al., 1982). Six-week-old rice plants (TP309 and TP309-Xa21) were cut approximately 4 cm from the tip of fully expanded leaves with scissors dipped in a bacterial suspension at either 10^9 cells per ml (Kauffman et al.,1973) or just peptone sucrose agar (mock condition). After inoculation, plants were maintained in a growth chamber and allowed to grow.

Rice sample preparation

The samples were processed according to Weckwerth et. al.[3] with the following changes. Approximately 20 mg segments of rice leaf were cut and weighed. They were placed in liquid-nitrogen-cooled 2 mL Eppendorf tubes, each containing a 5 mm stainless steel ball bearing (Figure 3). The tubes were transferred to MM301 Retsch Mill adapter racks that had been pre-cooled with liquid nitrogen. Samples were homogenized for 30 seconds at 25 Hz. 1 mL of solvent—a 2:3:3 v/v/v mixture of water/acetonitrile/isopropanol—at –20°C was used to extract the metabolites from membrane and cell wall components in the homogenized samples. This solvent system was chosen to minimize extraction of waxes and to enable analysis by both LC/MS and GC/MS.

LC/MS analysis

An Agilent 1200 LC equipped with a ZORBAX SB-Aq column 2.1 x 150 mm was used to separate the rice extracts. 2 µL injections were made from 1 mL sample volumes. At a flow rate of 0.3 mL/min., a 2% to 98% linear gradient of water / acetonitrile was employed over 46 minutes followed by a solvent hold until 54.9 minutes, at which time data collection was stopped. 0.1% formic acid was used as a mobile phase modifier.

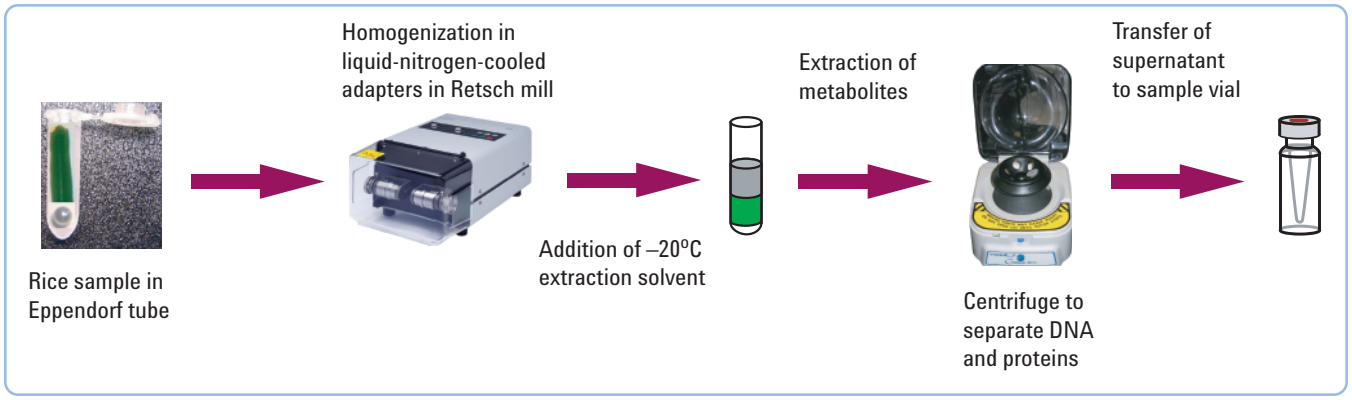


Figure 3. Sample preparation workflow for rice leaf samples.

An Agilent 6210 Time-of-Flight LC/MS equipped with an electrospray (ESI) ion source was used to acquire profiling data. The ESI source featured a separate nebulizer for the continuous, low-level introduction of reference mass compound. The reference mass compound facilitates compensation for instrument drift. Data was collected at a rate of 1 MS spectrum per second in both positive and negative ion modes from *m/z* 50 to 950.

An Agilent 6510 Quadrupole Time-of-Flight LC/MS equipped with an electrospray ion source was used to acquire accurate-mass MS/MS data for metabolite identification. The ESI source featured a separate nebulizer for the continuous, low-level introduction of reference mass compound to maximize mass accuracy.

Data analysis

Initial processing of the accurate-mass MS profiling data was done using Agilent MassHunter Software. The feature extraction and correlation algorithms in the MassHunter software located the groups of co-variant ions in each chromatogram. Each of these groups represented a unique compound. Thus, the algorithm located all the components in a chromatogram, instead of just locating chromatographic peaks, which could have concealed multiple components.

After locating components, background was subtracted. Charge state was set to 1. The algorithm identified salt

Instrument Conditions – LC/TOF MS

LC Conditions

Column: ZORBAX SB-Aq column 2.1 x 150 mm, 3.5 µm
Mobile phase:
- A = 0.1% formic acid in water
- B = 0.1% formic acid in acetonitrile

Gradient:
- 2% B at 0 min
- 98% B at 46 min
- 98% B at 54.9 min
- 2% B at 55 min

MS stop time: 54.9 min
LC stop time: 55 min
Column temperature: 20°C
Flow rate: 0.3 mL/min
Injection volume: 2 µL + 3 sec flush

MS Conditions

Ionization mode: Electrospray
Ionization polarity: Positive ionization*
Drying gas flow: 10 L/min
Drying gas temperature: 250°C
Nebulizer pressure: 35 psi
Scan range: *m/z* 50–950
Fragmentor voltage: 170 V
Capillary voltage: 4000 V
Reference masses: *m/z* 121, 922
Reference mass flow: 10 µL/min

*Both positive- and negative-ionization data were successfully acquired, but this note deals only with processing of the positive ion data.

Instrument Conditions – LC/Q-TOF MS/MS

LC Conditions

Column: ZORBAX SB-Aq column 2.1 x 150 mm, 3.5 µm
Mobile phase:
A = 0.1% formic acid in water
B = 0.1% formic acid in acetonitrile
Gradient:
2% B at 0 min
98% B at 46 min
98% B at 54.9 min
2% B at 55 min
MS stop time: 54.9 min
LC stop time: 55 min
LC post time: 7 min
Column temperature: 20°C
Flow rate: 0.3 mL/min
Injection volume: 2 µL

MS Conditions

Ionization mode: Electrospray
Ionization polarity: Positive ionization
Drying gas flow: 10 L/min
Drying gas temperature: 250°C
Nebulizer pressure: 40 psig
Scan range
MS: *m/z* 100–1000 at 250 ms/spectrum
MS/MS: *m/z* 100–1000 at 250 ms/spectrum
Collision energy: 5 x +10eV
Isolation: medium
Fragmentor voltage: 170 V
Skimmer voltage: 65 V
Octopole RF voltage: 750 V
Capillary voltage: 4000 V
Reference masses: *m/z* 121, 922
Reference mass flow: 10 µL/min
Reference nebulizer pressure: 15 psig

adducts (Na^+ and K^+) and the protonated molecules [$M+H^+$] and associated adduct ions were treated as a single compound. Finally, the algorithm identified isotopes. The monoisotopic mass and retention time was reported for each feature. An empirical formula was calculated for each feature using the monoisotopic mass and isotope ratios.

The retention time/mass pairs generated by the MassHunter Workstation software were then exported for subsequent analysis in Agilent GeneSpring MS software. The workflow used was as follows (see Figure 4):

1. Alignment and normalization of features
2. Hierarchical clustering to check data quality
3. Identify features with differential abundances across classes using 1-way and 2-way analysis of variance (ANOVA)
4. Perform principle component analysis (PCA) to show discriminating classes
5. Visualize fold changes

The result of the analysis to this point was a mass list of metabolites that showed statistically significant variations in abundance between experimental classes.

6. Search of the mass list against the METLIN Personal metabolite database.

Samples were then rerun using targeted MS/MS on the Q-TOF LC/MS system, and further data analysis was performed:

7. Comparison of metabolite database search results against the acquired Q-TOF spectra

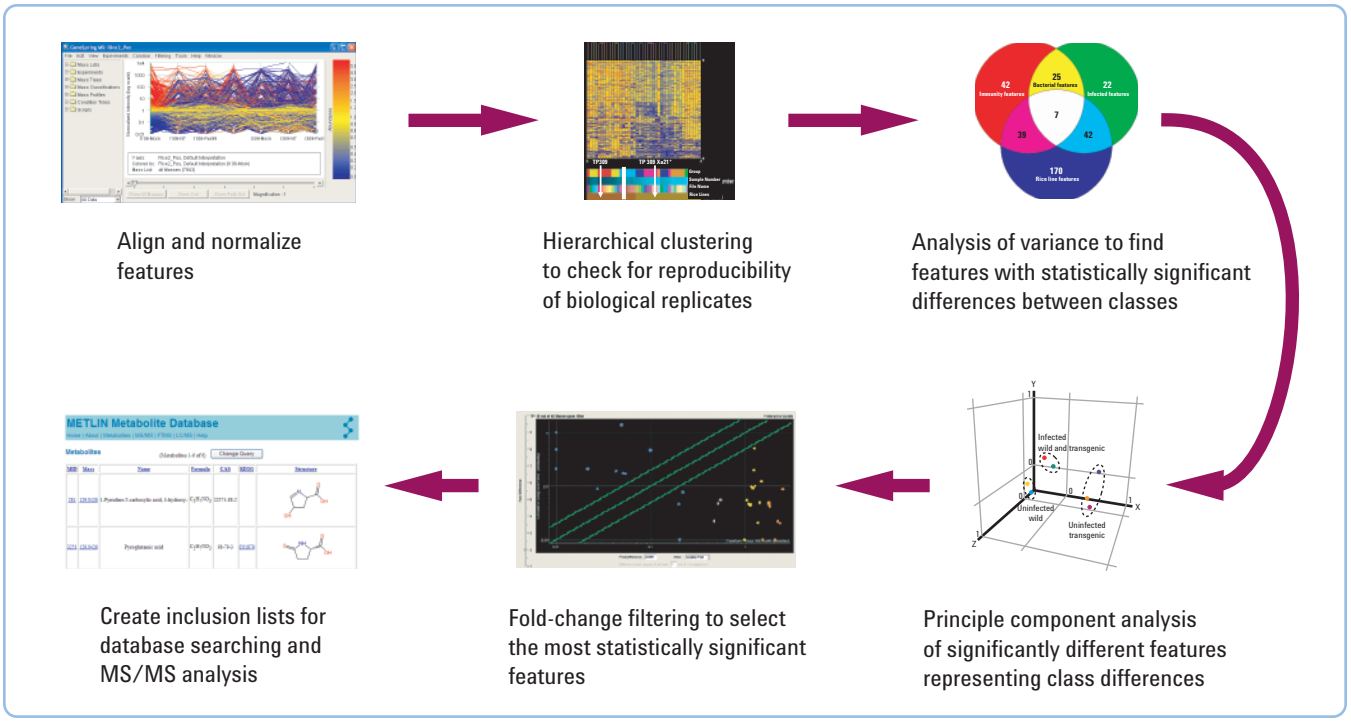


Figure 4. Workflow for GeneSpring MS analysis of the MS profiling data.

Results and Discussion

Morphology of the rice/bacterial interactions

TP309 is not resistant to either PXO99 or the knock-out PXO99-raxST$^-$ (Figure 5). TP309 lacks a mechanism that would recognize either strain of *Xoo* and trigger an immune response.

TP309-Xa21 is resistant to PXO99. The *Xa21* gene confers resistance by producing a receptor protein that recognizes and binds the AvrXa21 peptide produced by PXO99. This triggers an immune response.

TP309-Xa21 is not resistant to PXO99-raxST$^-$. Since PXO99-raxST$^-$ does not produce AvrXa21, TP309-Xa21 has no way to recognize the pathogenic bacteria and does not mount an immune response.

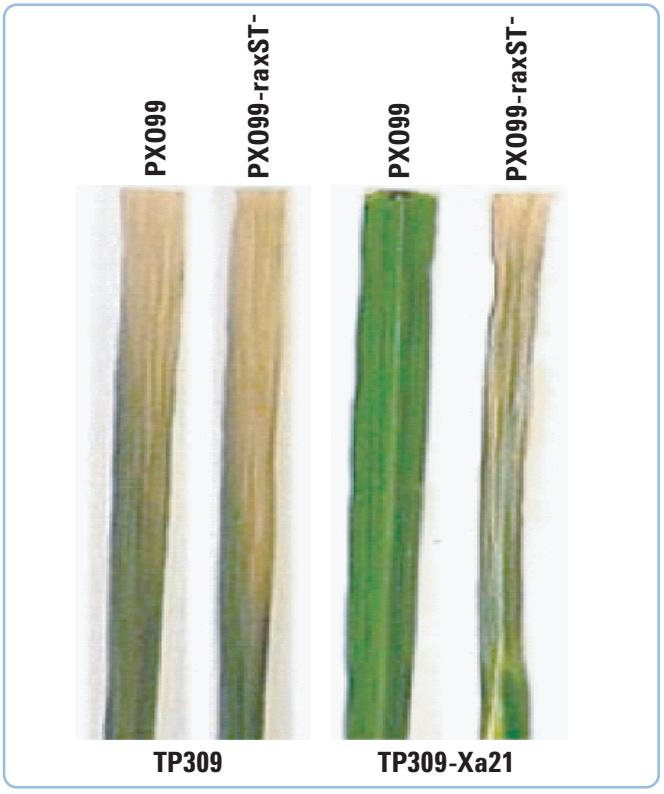


Figure 5. Rice leaves showing both diseased and healthy states.

1-way analysis of variance to identify class differences

Analysis of variance (ANOVA) is a powerful tool for analyzing the variation present in an experiment. Unlike a t-test, which can only make pair-wise comparisons, ANOVA can analyze any number of data sets. Multiple t-tests are independent, so their errors are cumulative. In the case of this rice experiment with seven classes, a total of 21 t-tests would have had to be performed. If the probability of a type-I error (false positive) for a single t-test was 0.05, the cumulative probability of error would have been greater than 1.00. A major advantage of 1-way ANOVA is that its probability of error remains the same, no matter how many conditions are included. The chance of a type-I error using 1-way ANOVA for this experiment is much less than with cumulative t-tests.

Figure 6 shows the results of ANOVA ($P < 0.05$) after applying a Tukey post-hoc test for all pair-wise comparisons. In the blue boxes are the number of features with statistically insignificant differences. In the red boxes are the number of features with statistically significant differences. The features with statistically significant differences are potential biomarkers and are potentially of greater biological interest. These were analyzed further.

In order to find metabolites that might account for differences between rice lines, the effect of PXO99 on infectivity was analyzed. The results of three pair-wise comparisons from the ANOVA analysis:

- TP309-Xa21—PXO99 vs. TP309-Xa21—Mock
- TP309—PXO99 vs. TP309—Mock
- TP309—Mock vs. TP309-Xa21—Mock

were combined for further study (Figure 7). The three comparisons had a total of 347 unique, statistically significant features. Analysis found:

- Immunity features (42)—metabolites possibly related to resistance
- Infected features (22)—metabolites possibly related to infection response
- Bacterial features (25)—metabolites possibly produced by, or in response to, the bacteria
- Rice line features (170)—metabolites related to differences in the rice lines

	Xa21-Mock	Xa21-NT	Xa21-Pxo99	Xa21_Pxo99/RaxST-	TP309-Mock	TP309-NT	TP309-Pxo99
Xa21-Mock	564	124	113	167	258	290	219
Xa21-NT	44	564	109	178	236	264	210
Xa21-Pxo99	45	456	564	168	242	289	204
Xa21_Pxo99/RaxST-	39	386	396	564	141	197	111
TP309-Mock	30	328	322	423	564	78	96
TP309-NT	27	300	275	367	486	564	141
TP309-Pxo99	34	354	360	453	469	423	564

Figure 6. ANOVA ($P < 0.05$) of MS profiling data identified features that showed statistically insignificant differences (blue) and statistically significant differences (red) in pair-wise comparisons. For example, comparing the results of a mock challenge of TP309-Xa21 with a PXO99 challenge of TP309-Xa21 found 113 features with statistically significant differences in expression levels.

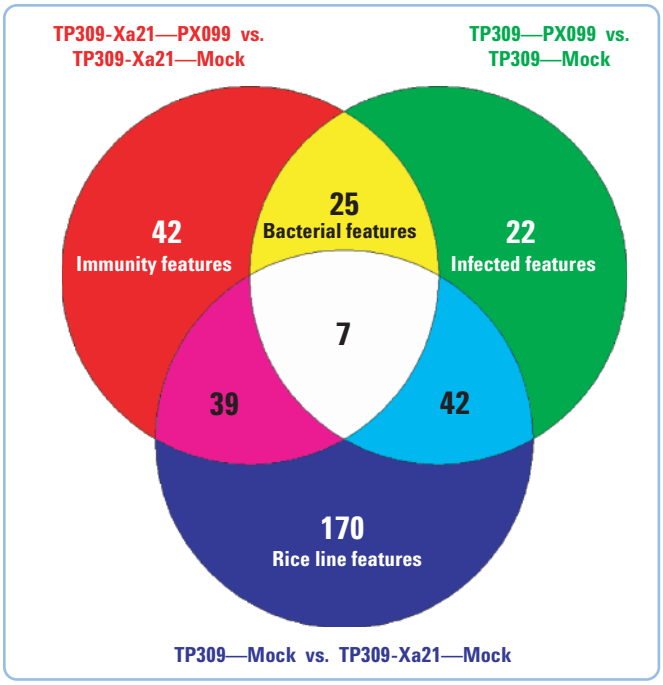


Figure 7. Combination of the ANOVA results from 3 pair-wise comparisons yielded a total of 347 unique, statistically significant features. Of these, 42 were related to immunity, 22 to infection, 25 to the bacteria, and 170 to the rice lines.

Principle component analysis (PCA)

PCA is a mathematical method of compressing complex data into a few variables. The objective is to discover new variables—principle components—which account for the majority of the differences in the data. When PCA was performed in GeneSpring MS with no prefiltering of data, separation of the TP309 and TP309-Xa21 rice lines was observed (Figure 8a).

However, in this experiment, information regarding the immunity and defense features was the goal, not just differentiation between the two rice lines. Combining 1-way ANOVA with PCA made the rice line differences much clearer, and also made it easier to distinguish infection status, regardless of rice line (Figure 8b).

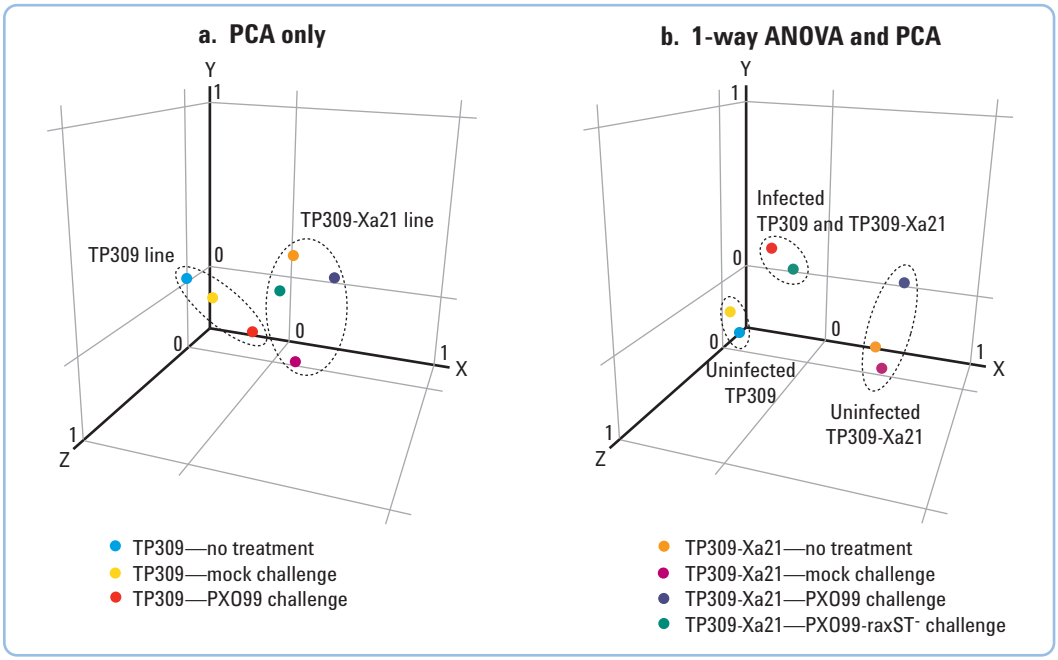


Figure 8. Principal component analysis without prefiltering of data (a) allowed differentiation between rice lines. When PCA was combined with 1-way ANOVA (b) it was much easier to differentiate not only rice line, but infection status regardless of rice line.

2-way ANOVA measures rice line and class differences simultaneously

2-way ANOVA is a powerful tool to study combinations of treatments. Parameters are compared in every combination. A prior knowledge regarding these features is not necessary. After performing 2-way ANOVA, a Venn diagram was plotted (Figure 9). Results included:

- 360 features explained the variance between the two rice lines
- 41 features alone were sufficient to separate all classes
- 30 features separated all rice lines and classes

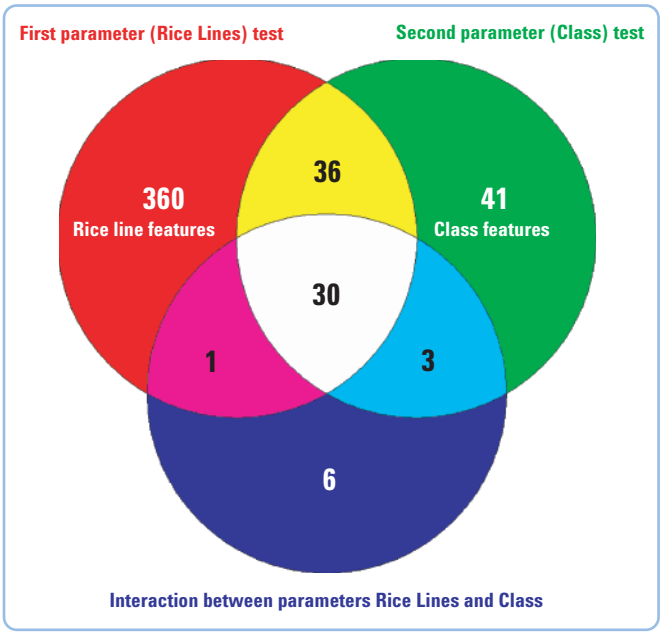


Figure 9. Analysis by 2-way ANOVA identified 30 features that together separated all classes and rice lines.

These 30 significant features were then processed by PCA to demonstrate the separation of rice lines and classes. The TP309-Xa21 classes group together except for the TP309-Xa21 that was challenged by PXO99-raxST⁻. It grouped with the infected class TP309—PXO99 challenge (Figure 10).

These results are consistent with the published results for the genotype and phenotype for these rice lines and bacteria.

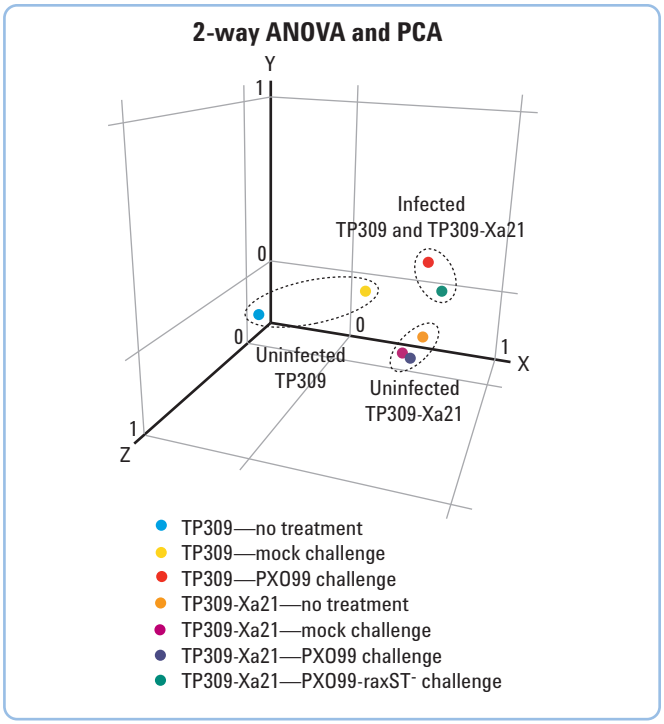


Figure 10. PCA analysis of 30 significant features found by 2-way ANOVA clearly differentiates rice lines, while also grouping infected classes together.

Fold-change filtering and visualization

Fold-change filtering (Figure 11) was another statistical tool used to determine which features (metabolites) were most likely to be relevant. The greater the fold change, relative to mock infection, the more likely it is that a particular metabolite is relevant. In the GeneSpring MS software, fold-change filtering is interactive, so the fold-change threshold was easily altered to determine what effect a change would have on the number of features passing the filter.

Another analysis and visualization option available in the GeneSpring MS software, but not used in this analysis, is the volcano plot. The volcano plot combines the results of fold-change filtering and t-tests in a single visualization. Criteria for both tests can be varied interactively to find the most relevant features.

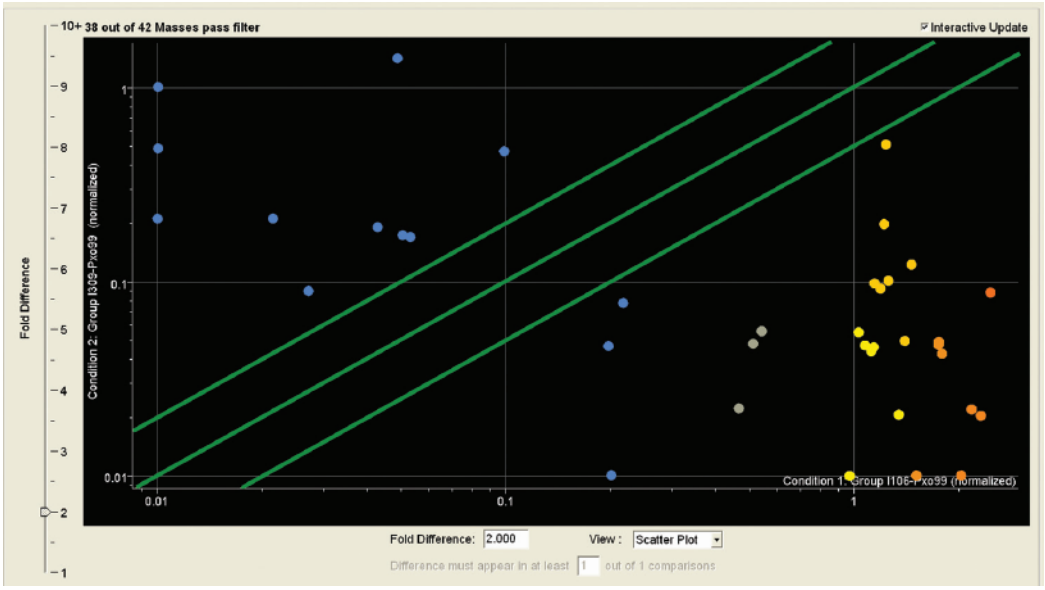


Figure 11. Interactive fold-change filtering of the results of 1-way ANOVA helped determine which features (metabolites) were most likely to be relevant.

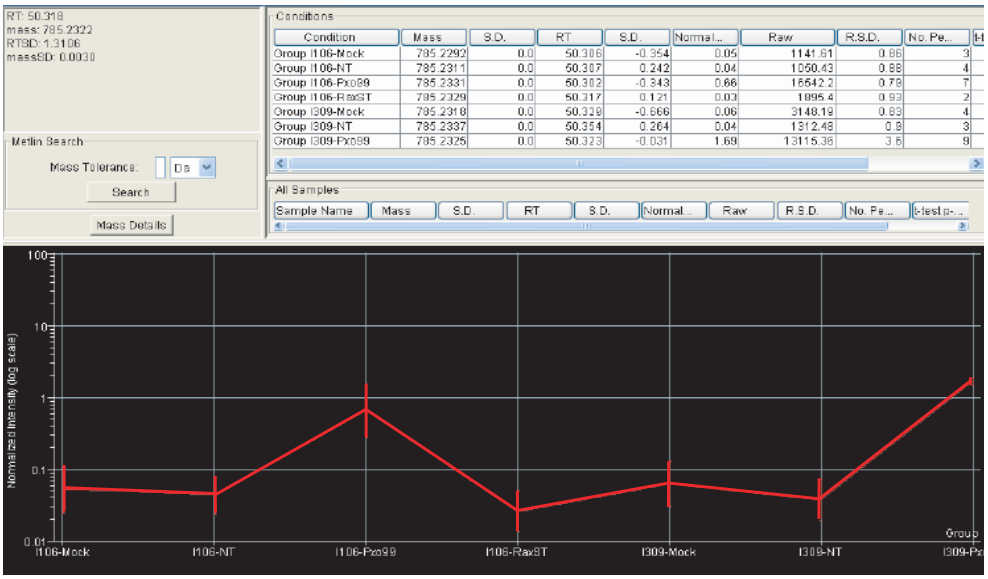
Feature inspection

Further analysis of individual features was performed using two additional functions of the GeneSpring MS software. The first, was a "mass inspector" view that permitted the data on individual features to be examined. The second was the ability to export data related to the features on a mass list in a tab-delimited format that is compatible with Microsoft® Excel.

Mass inspector view

The mass inspector view in the GeneSpring MS software allowed comparison of the relative abundance of a particular feature across all desired classes (see Figure 12). This made it easy to see if a particular feature exhibited, for example, an on/off behavior where it was present in some experimental classes but not present in others. On/off behavior is a valuable indicator of possible relevance. The mass inspector view also provided additional information, such as retention time, molecular weight, standard deviation, and t-test significance, about the feature being examined.

Figure 12. The mass inspector view in the GeneSpring MS software allowed individual feature (metabolite) abundance to be compared across classes. It also displays other useful feature information such as retention time, molecular weight, standard deviation, and t-test significance.



Microsoft is a U.S. registered trademark of Microsoft Corporation.

Export to Excel

Using another GeneSpring MS software feature, all data related to the immunity features that passed the fold-change filter was exported in tab-delimited format. Manipulation of the data in a spreadsheet allowed us to separate the data into lists of both up-regulated and down-regulated features (Tables 2 and 3). Of particular interest were the features (highlighted in the tables) which were absent in one or more classes. These features may be on/off metabolites and were flagged for further investigation.

Table 2. Up-regulated metabolites from the TP309-Xa21—PXO99 class (immunity) that passed the fold-change filter.

Retention time (min)	Mass (u)	Fold change
32.64	771.4705	2.4
1.11	296.9389	2.8
32.76	710.4604	4.2
41.36	167.0575	6.1
43.74	401.3279	9.7
40.60	849.5386	10.7
31.24	295.2517	11.5
25.80	329.2925	11.9
27.20	453.2855	12.4
32.66	739.4514	12.9
47.38	934.5473	18.6
46.41	660.5333	20.3
50.91	565.8811	20.7
49.41	948.5989	23.2
2.38	221.0538	24.8
45.17	817.5082	25.0
37.68	608.2646	27.8
38.53	861.5044	28.0
2.07	129.0414	35.5
2.10	122.0383	36.2
37.25	624.2587	41.8
53.60	945.6066	64.5
53.31	93.0454	97.1
2.06	385.1011	98.6
52.22	157.9583	112.7
51.76	580.4290	151.1
51.35	225.9444	203.8

Targeted metabolite identification

Metabolite profiling provided a list of up-regulated immunity features that were selected for the next step in metabolomic investigation: targeted metabolite identification. Tentative identification of the metabolites was a two-step process.

First, each of the target masses identified in the profiling process was searched in the METLIN Personal metabolite database. The database was searched over a narrow ±10 ppm mass window. This was significantly wider than the ~2 ppm mass accuracy of the 6210 TOF, but it was felt that it was better to review a few extra hits than to possibly exclude the correct match by using a too-narrow window. The empirical formula calculations were set with a mass error of 5 ppm and 100 as the maximum number of empircal formulas. Table 4 shows the settings for elements.

Table 3. Down-regulated metabolites from the TP309-Xa21—PXO99 class (immunity) that passed the fold-change filter.

Retention time (min)	Mass (u)	Fold change
1.09	213.9057	3.2
45.66	452.3297	3.3
32.52	600.4134	3.4
1.09	73.0268	4.4
36.95	281.6063	4.9
1.09	109.1268	10.0
47.21	524.3837	21.3
1.28	103.0648	21.6
46.57	652.4474	29.2
1.43	95.9816	48.3
47.20	540.3579	101.8

The results of the searches of the METLIN Personal metabolite database were incorporated into the spreadsheets with the original mass lists (Tables 5 and 6). That the searches did not produce matches for all of the metabolites was not surprising.

Table 4. Minimum and maximum element settings for METLIN database search.

Element	Minimum Number	Maximum Number
Carbon	1	100
Hydrogen	1	400
Nitrogen	0	20
Oxygen	0	20
Phosphorous	0	1
Sulfur	0	1
Fluorine	0	6

A wider mass window would have increased the number of possible metabolite identities, but it would also have increased the number of incorrect identities. And while the METLIN database, with over 15,000 entries, is probably the most comprehensive commercially available database in the world, that number is still just a small fraction of all possible metabolites. As will be shown, when a database search results in matches, it can be an invaluable aid to identification.

In the second part of the identification process, one of the upregulated immunity features, *m/z* 129.0414 from Table 5,

Table 5. Up-regulated metabolites from the TP309-Xa21—PXO99 class (immunity) with their METLIN search results.

Retention time (min)	Mass (u)	Fold change	Empirical formula	Number of formulas	METLIN search (number of hits)
32.64	771.4705	2.4	C27H68N10O13P	93	0
1.11	296.9389	2.8	C7HNO8F2S	13	0
32.76	710.4604	4.2	C38H60N7O6	71	0
41.36	167.0575	6.1	C6H7N4O2	3	3
43.74	401.3279	9.7	C24H41N4O	9	0
40.60	849.5386	10.7	C29H72N17O10P	100	0
31.24	295.2517	11.5	C18H33NO2	2	0
25.80	329.2925	11.9	C19H39NO3	2	0
27.20	453.2855	12.4	C9H31N19O3	20	0
32.66	739.4514	12.9	C32H65N7O10S	85	0
47.38	934.5473	18.6	C53H79N2O10P	100	0
46.41	660.5333	20.3	C25H62N19O2	32	0
50.91	565.8811	20.7	C4HN6O16F6PS	32	0
49.41	948.5989	23.2	C58H76N8O4	100	0
2.38	221.0538	24.8	C6H7N5O3	4	0
45.17	817.5082	25.0	C34H63N19O3S	100	0
37.68	608.2646	27.8	C25H45N4O9PS	75	Harderoporphyrin
38.53	861.5044	28.0	C47H70N6O7P	100	erythromycin ethylsuccinate
2.07	129.0414	35.5	C5H7NO3	1	6
2.10	122.0383	36.2	C7H6O2	2	benzoic acid
37.25	624.2587	41.8	C30H33N12O2P	85	0
53.60	945.6066	64.5	C60H79N7OS	100	0
53.31	93.0454	97.1	C5H5N2	2	0
2.06	385.1011	98.6	C17H18N6OPS	36	0
52.22	157.9583	112.7	C3HN3OPS	1	0
51.76	580.4290	151.1	C24H57N10O4P	30	Vitamin K2
51.35	225.9444	203.8	C4HF6PS	4	0

Table 6. Down-regulated metabolites from the TP309-Xa21—PXO99 class (immunity) with their METLIN search results.

Retention time (min)	Mass (u)	Fold change	Empirical formula	Number of formulas	METLIN search (number of hits)
1.09	213.9057	3.2	none	0	0
45.66	452.3297	3.3	C23H44N6OS	17	0
32.52	600.4134	3.4	C22H57N12O3PS	40	Violaxanthin, Neoxanthin
1.09	73.0268	4.4	CHNF2P	1	0
36.95	281.6063	4.9	none	0	0
1.09	109.1268	10.0	none	0	0
47.21	524.3837	21.3	C30H48N6O2	24	0
1.28	103.0648	21.6	none	0	0
46.57	652.4474	29.2	C29H59N13PS	48	0
1.43	95.9816	48.3	none	0	0
47.20	540.3579	101.8	C29H47N7OP	32	0

was selected for further investigation. The sample was rerun on the Q-TOF LC/MS system. Separation and ion source conditions used were the same as those for the original TOF analyses. MS/MS spectra were acquired from each of the metabolites on the list of target masses.

Examination of the MS/MS spectrum from the selected metabolite (Figure 13) showed a base peak representing the loss of a carboxyl group (formic acid – CH_2O_2) from the precursor ion. A second significant peak represented the subsequent loss of CO.

By evaluating the MS/MS spectrum against the molecular structures included in the METLIN database search results (Figure 14), it was determined that only two of the six possible metabolites—pyroglutamic acid and pyrrolidonecarboxylic acid—could logically have produced the losses seen. These two compounds are enantiomers, and as such indistinguishable by mass spectrometry. If obtaining the precise identity of the metabolite had been critical, it could have been determined through reanalysis using standards and a chiral LC column.

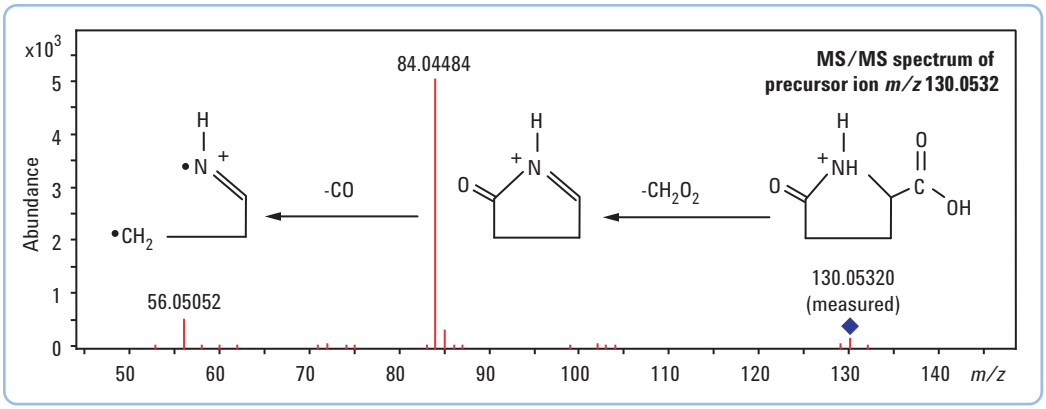


Figure 13. MS/MS spectrum of the precursor ion at *m/z* 130.0532 shows a base peak representing the loss of formic acid (CH_2O_2) and a peak representing a subsequent loss of CO. Evaluation of this information against the structures of the six possible identities generated by a search of the METLIN metabolite database reduced the list of possible identities to two.

METLIN Metabolite Database

Home | About | Metabolites | MS/MS | FTMS | LC/MS | Help

Metabolites (Metabolites 1-6 of 6) Change Query

MID	Mass	Name	Formula	CAS	KEGG	Structure
284	129.0426	1-Pyrroline-5-carboxylic acid, 3-hydroxy-	$C_5H_7NO_3$	22573-88-2		
3251	129.0426	Pyroglutamic acid	$C_5H_7NO_3$	98-79-3	C01879	
5769	129.0426	Pyrrolidonecarboxylic acid	$C_5H_7NO_3$			
6196	129.0426	Pyrroline hydroxycarboxylic acid	$C_5H_7NO_3$			
6343	129.0426	N-Acryloylglycine	$C_5H_7NO_3$			
6564	129.0426	1-Pyrroline-4-hydroxy-2-carboxylate	$C_5H_7NO_3$			

Figure 14. A search of the METLIN Personal metabolite database generated a list of six possible identities for a metabolite with a molecular weight of approximately 129.0414 u.

Conclusions

In order to better understand the mechanisms of infection and immunity between rice and bacterial leaf blight (BLB), a study was undertaken to identify metabolites related to infection and resistance. A two-step LC/MS approach was employed. Rapid differential expression analysis of samples using time-of-flight (TOF) mass spectrometry (MS) was followed by targeted identification of differentially expressed metabolites using quadrupole time-of-flight (Q-TOF) MS/MS.

In total, seven different classes were compared. Clear differences in the metabolites of the different rice/bacteria classes were detected. Based on relatively few metabolites, the rice lines and state of infection were clearly distinguishable.

A significant amount was also learned about instrumentation and software requirements for this type of study. For example, the number of replicates required is determined by natural (sample) variability and technical (instrumentation) variability. In this experiment, the natural variations in both the rice and bacteria dictated multiple biological replicates. However, because of the outstanding reproducibility of the Agilent 6210 Time-of-Flight LC/MS system used for profiling, additional technical replicates were not required.

Data analysis plays an essential role in large-scale metabolomics studies. This study convincingly demonstrated that having a range of statistical-analysis tools is essential to obtaining the best results. Analysis of the profiling data by principal component analysis (PCA) alone could barely distinguish between the two rice lines. The combination of PCA with one- or two-way analysis of variance (ANOVA) clearly distinguished not only the rice lines, but infection state regardless of rice line. Further, the ability to apply fold-change filtering and visualize the results made it easier to compare differences in abundance and select targets for the second phase of the study: metabolite identification. Without the wide range of statistical-analysis tools in the Agilent GeneSpring MS software used for data analysis, it would have been impossible to obtain as much information as was obtained from this study.

Proceeding from the profiling phase of the study to the metabolite identification phase, was facilitated by two factors. The extremely good mass accuracy of the TOF profiling data and the powerful METLIN Personal metabolite database together made it possible to narrow the list of possible metabolite identities to a managable number. The MS/MS spectra obtained by reanalysis using an Agilent 6510 Q-TOF LC/MS identified the metabolite as one of two enantiomers. This also highlighted the challenges of metabolomics. Even with the largest metabolite database commercially available and accurate-mass MS/MS spectra, conclusive identification of some compounds will require further analysis. In this case, it would have required reanalysis using standards and a chiral LC column.

References

1. Song et. al. (1995) *Science* 270: 1804–1806
2. Sang-Wong Lee et. al. PNAS, 103(49) 18395–18400, 2006
3. Weckworth et. al. *Proteomics* 2004, 4, 78–83

Acknowledgments

The authors would like to express special thanks to Dr. Pamela Ronald and Dr. Oliver Fiehn of the Genome Center at the University of California, Davis.

Dr. Ronald is Chair of the Plant Genomics Program and Professor in the Department of Plant Pathology at U.C Davis, and a leading researcher in the fundamental processes of grasses such as rice and switchgrass.

Dr. Fiehn is an Associate Professor in the Department of Molecular and Cellular Biology & Genome Center at U.C. Davis. He is a leader in the emerging field of metabolomics.

Drs. Ronald and Fiehn provided the samples for the research outlined in the application note as well as invaluable background information and advice.

About Agilent Technologies

Agilent Technologies is a leading supplier of life science research systems that enable scientists to understand complex biological processes, determine disease mechanisms, and speed drug discovery. Engineered for sensitivity, reproducibility, and workflow productivity, Agilent's life science solutions include instrumentation, microfluidics, software, microarrays, consumables, and services for genomics, proteomics, and metabolomics applications.

Learn more:
www.agilent.com/chem/metabolomics

Buy online:
www.agilent.com/chem/store

Find an Agilent customer center in your country:
www.agilent.com/chem/contactus

U.S. and Canada
1-800-227-9770

agilent_inquiries@agilent.com

Europe
info_agilent@agilent.com

Asia Pacific
adinquiry_aplsca@agilent.com

Printed in the U.S.A. February 14, 2007

5989-6234EN

Agilent Technologies