Application Note

Exposomics, Genomics, Bioinformatics

Agilent

Trusted Answers

Association of *Nosema ceranae* Infection with Honey Bee (*Apis meillfera*) Exposome Profiles and Affected Biological Pathways

Authors

Anthony Macherone[1,2], Robert L. Broadrup[3], Christopher Mayack[4], and Helen K. White[3]

[1] Life Science and Chemical Analysis Group, Agilent Technologies, Santa Clara, CA 95051

[2] Department of Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD USA

[3] Department of Chemistry, Haverford College, Haverford, PA USA

[4] Department of Biology, Swarthmore College, Swarthmore, PA USA

Abstract

Foraging honey bee samples representative of 30 individual hives were collected simultaneously at the end of the 2015 foraging season. Polymerase chain reaction was used to determine if the hives were infected with *Nosema ceranae* – a common honey bee parasite. A second sample of foraging bees was collected from each hive at the same time for discovery-based exposomics using the Agilent 7890B GC combined with an Agilent 7200B GC/Q-TOF system. Significance testing and biological pathway analyses were conducted using Agilent MassProfiler Professional and Pathway Architect software. Significance testing determined that exposome profiles were associated with *N. ceranae* infection. This study illustrates how the integration of exposomics with genetic disease profiling yields a powerful multi-omics platform for discovery and generation of new hypotheses for targeted follow-up studies.

Introduction

Over the past decade, the western honey bee (*Apis mellifera*) population has succumbed to a host of environmental stressors including exposure to parasites, persistent organic pollutants, poor nutrition, and loss of habitat. These stressors represent environmental exposures that adversely affect the health of bees at the hive level. Current honey bee research offers insight into the reasons for the global health decline of honey bees, but this complex combination of stressors has made it difficult to pinpoint key features of disease causality.

The exposome represents cumulative exposures over individuals' lifetime. It includes, but is not limited to, environmental pollutants, diet, drugs, and the consequent alterations of internal biochemistry. The exposome paradigm associates exposures with biological response pathways, and offers a mechanism to investigate the nongenetic causative factors of chronic disease. To address the highly complex problem of honey bee population decline, we hypothesized that integrating exposome chemical profiles with semiquantitative PCR (sq-PCR), and analyzing the resulting multi-omics dataset with bioinformatics tools may identify chemical profiles and dysregulation of biological pathways statistically associated ($p < 0.05$) with disease and eventual collapse of a bee colony.

Experimental

Sample Collection

Biological samples were collected from 30 noncommercial apiary hives in seven geographical locations located in southeastern Pennsylvania. Sixty to one hundred bees were collected from each hive in 50 mL disposable FalconTM tubes, and immediately frozen in dry ice. All samples were stored in a –80 °C freezer prior to analysis.

Sample extraction

Samples for mass spectrometry were extracted by pulverizing 3 g of bees in 27 mL of water/acetonitrile/acetic acid (44 %:55 %:1 %). A modified QuEChERS extraction was performed as follows: to each sample was added 250 mg carbon (p/n 5982-4482), 500 primary secondary amine (PSA, p/n 5982-8382), 6 g of magnesium sulfate (p/n 5982-8082), and 1.5 g sodium acetate (p/n 5982-5751). The tubes were sealed, shaken and centrifuged at 3,000 rpm for 5 minutes. A 2-mL aliquot of the supernatant from each sample was applied to a conditioned SPE cartridge containing 250 mg carbon (p/n 12113041). Analytes were eluted from the SPE cartridge with 4 mL acetone/toluene (70:30 v/v) and reduced in volume using a gentle stream of nitrogen. Approximately 2 mL of each concentrated sample was transferred to 2-mL autosampler vials prior to GC-TOF analysis.

DNA Extraction and Semiquantitative PCR

Samples for PCR screening were prepared by adding 6 mL of RNase-free water to 30 bees. The bees were macerated with a tissue grinder to thoroughly homogenize the contents. To semiquantify *Nosema ceranae*, a PCR reaction with a RpS5 reference gene was performed based on methods adapted from the HBRC method1. Briefly, a 150-µL aliquot of the homogenate was added to 300 µL of a 1:1 mixture of phenol/chloroform. The solution was centrifuged at 13,000 rpm for 5 minutes, and the supernatant was added to another 300 µL of the 1:1 phenol/chloroform solution in a 1.5-mL microcentrifuge tube. The supernatant was drawn once again and transferred to a 1.5 mL microcentrifuge tube. This was followed by the addition of 30 µL sodium acetate and 600 µL of 95 % ethanol to precipitate the DNA overnight at –20 °C. A PCR reaction combined 1 µL of a 10 mM solution of each primer, 1.5 µL of 10x PCR buffer, 0.5 µL of 10 mM deoxynucleotide triphosphate (dNTP), 0.2 µL of a 25 mM magnesium chloride solution, 0.2 µL of 5 U/µL Taq DNA polymerase (New England BioLabs, Ipswich, MA), and 2 µL of template DNA. The PCR thermocycler program was 94 °C for 2.5 minutes, followed by 10 cycles of: 15 seconds at 94 °C, 30 seconds at 61.8 °C, and 45 seconds at 72 °C. Then, 20 cycles of: 15 seconds at 94 °C, 30 seconds at 61.8 °C, and 50 seconds at 72 °C, followed by an extension step at 72 °C for 7 minutes, and a final hold step of 4 °C. PCR products were identified by a 3 % agarose gel using electrophoresis, and confirmed using the Agilent 2100 Bioanalyzer. Open-source ImageJ was used to semiquantify *N. ceranae* based on product band intensity from the gel electrophoresis[2].

GC-TOF Mass Spectrometry

An Agilent 7890B GC combined with an Agilent 7200B GC/Q-TOF system was used for discovery-based chemical profiling of bee exposomes. The GC was configured with a 40 m × 0.25 mm, 0.25 µm DB5-MS DuraGuard column (Agilent J&W 122-5532G) operated at 1.2 mL/minute helium in constant flow mode. A 0.2 µL pulsed splitless injection was made into a 250 °C isothermal split/splitless inlet. The oven program was 80 °C (1 minute), then 10 °C/min to 310 °C (6 minutes). The transfer line temperature was 300 °C. The mass spectrometer was operated in electron ionization, high resolution TOF mode. The source and quadrupole (RF only) temperatures were 275 °C and 150 °C, respectively. High resolution, accurate mass (HRAM) spectral information was collected at 5 Hz over a mass range of 50 Da to 800 Da.

Data Analysis

To extract chemical features from the raw mass spectra, chromatographic deconvolution was performed using Agilent MassHunter Unknowns Analysis B.08.00 software package. Briefly, chemical features were identified if the calculated signal-to-noise of a chromatographic peak was >3:1, and at least three extracted ion chromatograms could be aligned based on peak shape and retention time. Each identified feature was tested against a known commercial spectral library (NIST11)[3] for mass, number of ions, and ion ratio similarities. If a match score >0.7 was determined, the feature was annotated with chemical names, CAS number, and other pertinent information. Nonannotated chemical features were excluded from statistical analyses.

Instrument Conditions

Parameter	Value
Column	Agilent J&W DB5-MS DuraGuard, 40 m × 0.25 mm, 0.25 µm (122-5532G)
Gas flow	Helium, 1.2 mL/min, constant flow mode
Injection	0.2 µL pulsed splitless
Inlet temperature	250 °C, isothermal split/splitless inlet
Oven program	80 °C (1 minute), then 10 °C/min to 310 °C (6 minutes)
Transfer line temperature	300 °C
MS Mode	EI, high res TOF
Source temperature	275 °C
Quadrupole temperature	150 °C, RF only

Statistical Significance Testing and Pathway Analyses

The multi-omic dataset was statistically analyzed using Agilent MassProfiler Professional (MPP) and Pathway Architect software. The variables included *N. ceranae* Load (ranging from 0.00 to 40.02), and Phenotype (healthy, unhealthy, collapsed, unknown) determined by organoleptic observation at the time of sample collection. The data were aligned in MPP, allowing for a retention time variation of 0.15 minutes and a mass extraction window of ±20 ppm. All raw data were Log_2 transformed. The baseline of the data was determined using the baseline to the median of all log-transformed response values across all samples. Significance testing was performed using built-in features of MPP, first Filter by Flags, wherein chemical entities were retained if they were present in at least two samples, followed by Filter by Frequency to retain chemical entities that appear in 100 % of samples. One-way ANOVA of the reduced dataset yielded 20 chemical entities with a p-value <0.05 and a fold-change (distance from median value) >2. These were screened against known *Apis mellifera* biological pathways downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[4,5] using Agilent Pathway Architect tools. Figure 1 illustrates how the *N. ceranae* data are connected to deconvoluted mass spectra and analyzed in MPP to identify chemical entities, associations, and the biological pathways affected in the disease state.

Samples and Variables

Sample	Average *Nosema* load	Hive health
1	0.000	Unknown
2	2.144	Unknown
3	18.321	Healthy
4	0.000	Healthy
5	0.578	Unknown
6	0.000	Unknown
7	0.983	Healthy
8	0.000	Collapsed
9	6.683	Collapsed
10	0.000	Healthy
11	2.565	Unknown
12	40.020	Unknown
13	5.063	Unknown
14	3.566	Healthy
15	0.121	Unknown

Sample	Average *Nosema* load	Hive health
16	0.080	Unknown
17	0.000	Healthy
18	0.000	Healthy
19	0.606	Collapsed
20	0.000	Healthy
21	1.224	Healthy
22	0.000	Healthy
23	3.290	Unknown
24	29.821	Unknown
25	0.000	Collapsed
26	0.837	Unknown
27	0.000	Healthy
28	0.491	Collapsed
29	0.000	Unhealthy
30	2.246	Unknown

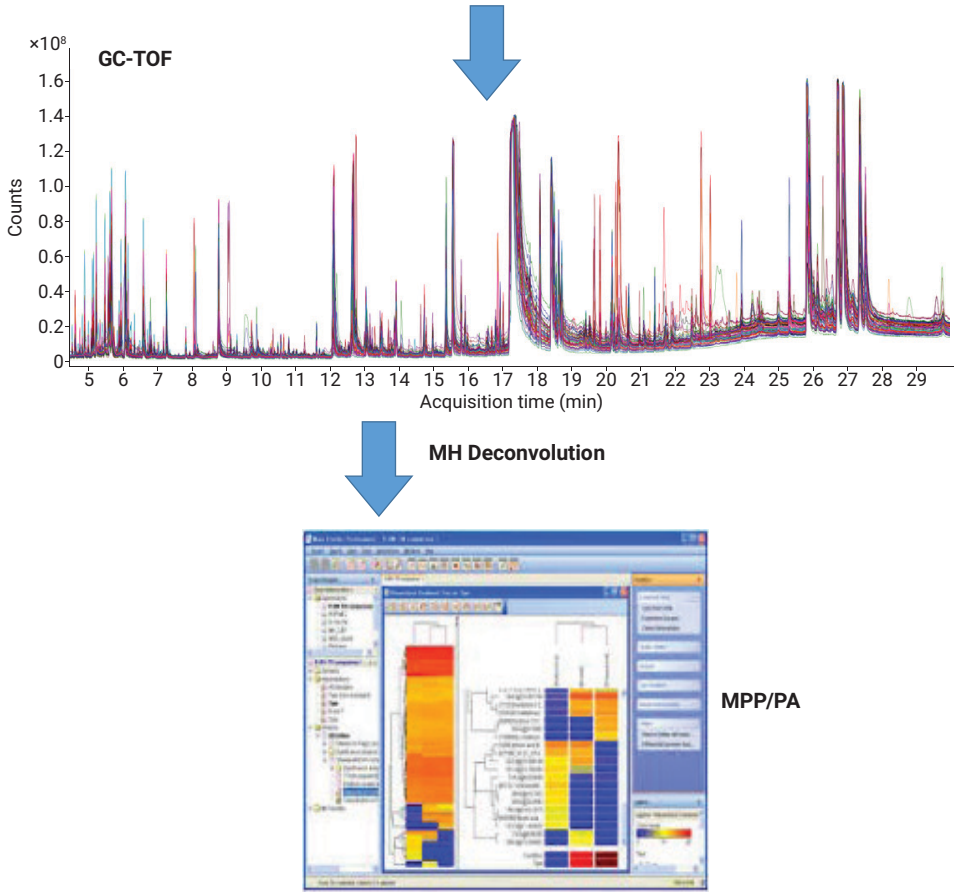

Figure 1. Connecting multi-omic data in Agilent MPP. *N. ceranae* load and hive health conditions for each sample are connected to the corresponding deconvoluted mass spectral data and chemical entity annotations in MPP. Quality and statistical testing is performed within the MPP environment to identify chemicals significantly associated with infection and hive health. Pathway Architect features are used to interrogate known honey bee biological pathways with the chemical entity data subset. Dysregulated chemical entities within each pathway are determined and saved for interpretation by the investigator.

Results and Discussion

The *N. ceranae* load data were approximately log-normal distributed. Significance testing identified multiple chemical biomarkers that may affect honey bee health. Many of the annotated chemicals were associated with known biological pathways of the western honey bee including phenylalanine metabolism, caffeine metabolism, fatty acid biosynthesis, and nicotinate and nicotinamide metabolism. For example, in the Ubiquinone and Other Terpenoid-Quinone Biosynthesis Pathway, a putative direct relationship between *N. ceranae* load and β-tocopherol and γ-tocopherol was identified. Therein, lower *N. ceranae* loads align with a down-regulation of these tocopherols, and higher loads align with increased tocopherol abundance (Table 1). In the Nicotinate and Nicotinamide Pathway, it was determined that nicotinamide levels in unhealthy hives was significantly upregulated.

Many of the associated chemicals were found at relative concentration levels (represented by ion abundance) 10-fold higher than the median value over all samples (Figure 2). When comparing healthy to unhealthy hives, niacinamide, and semioxamazide (oxalic acid, a common treatment for diseased hives) were significantly associated ($p << 0.00001$) with *N. ceranae* infection. Other chemicals moderately associated with *N. ceranae* infection when comparing healthy to unhealthy hives include: 1-methyl-1H-imidazole ($p = 0.0025$), a histidine and histamine mimetic; 1 -tridecyn-4-ol ($p = 0.0040$), a 13-carbon alcohol potentially associated with the chemical communication of the bees; 5-hydroxy-4',7-dimethoxyflavinone ($p = 0.0069$), derived from flower pollen; and tris(2,4-di-*tert*-butylphenyl)phosphate ($p = 0.021$), a known plasticizer.

Table 1. Ubiquinone and other terpenoid-quinone biosynthesis. β-tocopherol and γ-tocopherol are downregulated (blue) in hives with very low to no *N. ceranae* infection as determined by sq-PCR, while hives with higher levels of infection show upregulated levels of these tocopherols (red).

Entity	Averaged *Nosema ceranae* Load (Log_2 transformed)							
β-tocopherol	0.244	-6.96	-4.69	-3.87	3.14	2.45	3.92	13.5
γ-tocopherol	-0.286	-4.70	-4.70	-5.11	4.70	4.70	4.70	15.65

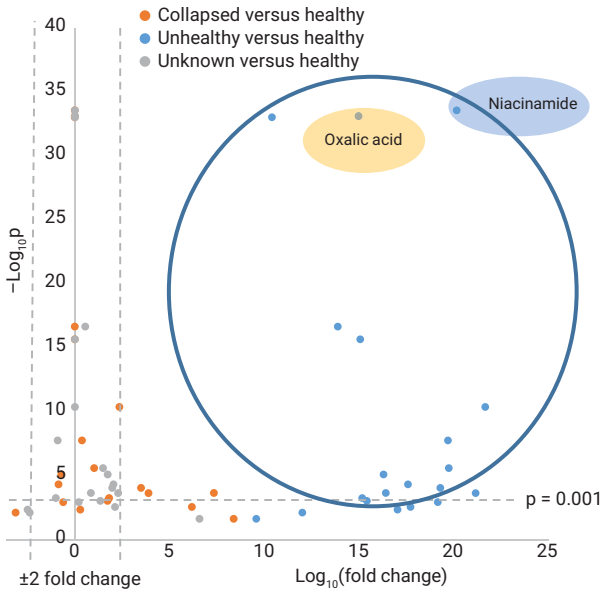

Figure 2. Pseudo-Manhattan plot. $-Log_2$(transformed p values) versus log(transformed fold-change) for collapsed versus healthy, unhealthy versus healthy, and unknown versus healthy hives. Vertical dotted lines illustrate ±2 log_2 normalized relative concentration fold-change. The horizontal line represents the log_2(normalized α) for $p = 0.001$ (0.1 % occurs by chance). Significance testing identified a chemical entity list of 20 chemicals significantly ($p<0.05$) associated with *N. ceranae* infection. When comparing this list in the collapsed, unhealthy, or unknown hive health against healthy hives, most of the chemical entities in the collapsed and unknown hives fall within the ±2-fold change window. However, comparing unhealthy hives health against healthy hives, most entities have >10-fold change, with statistical significance ($p <0.001$).

Conclusion

The initial results from this study tend to confirm the hypotheses that as parasite infestation increases, there is a measurable change in the exposome chemical profile. The identified chemicals can be associated with known biological pathways of the western honey bees. We acknowledge that the small sample size of the pilot study may not elicit the appropriate effect needed to detect more subtle associations with changes in honey bee health. We caution that these results must be confirmed with more extensive longitudinal sampling and follow-up targeted hypothesis-driven studies.

References

1. Hamiduzzaman, M. M.; Guzman-Novoa, E.; Goodwin, P. H. A multiplex PCR assay to diagnose and quantify Nosema infections in honey bees (*Apis mellifera*). *J. Invertebr. Pathol.* **2010**, *105*, 151-155.
2. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. *Nature Methods* **2012**, *9, 7*, 671-675.
3. National Institute of Standards and Technology, NIST Standard Reference Database 1A. Retrieved February 17, **2013**, from NIST.gov: http://www.nist. gov/srd/nist1a. cfm#2011.
4. Kanehisa, M.; *et al. Nucleic Acids Res*. **2016**, *44*, D457-D462.
5. Kanehisa, M.; Goto, S. *Nucleic Acids Res*. **2000**, *28*, 27-30.

www.agilent.com/chem

Printed in the USA, September 21, 2017
5991-8357EN

Agilent
Trusted Answers