SUPPLEMENTARY MATERIAL

BREATH METABOLOME OF MICE INFECTED WITH *PSEUDOMONAS AERUGINOSA*

Giorgia Purcaro[1,2], Mavra Nasir[3], Flavio A. Franchina[1,4], Christiaan A. Rees[3], Minara Aliyeva[5], Nirav Daphtary [5], Matthew J. Wargo[5], Lennart K. A. Lundblad[6,7], Jane E. Hill[1,3,*]

[1]Thayer School of Engineering, 14 Engineering Drive, Dartmouth College, Hanover, NH, 03755, United States.

[2] Now at: Gembloux Agro-Bio Tech, University of Liège, Gembloux, 5030, Belgium

[3]Geisel School of Medicine, 1 Rope Ferry Road, Dartmouth College, Hanover, NH, 03755, United States.

[4] Now at: Department of Chemistry, University of Liège, Liège (Sart-Tilman), 4000, Belgium

[5] Larner College of Medicine, 149 Beaumont Avenue, University of Vermont, Burlington, VT, 05405, United States.

[6]THORASYS Thoracic Medical Equipment Inc. 6560 de l'Esplanade, Suite 103, Montreal, Quebec H2V 4L5, Canada

[7]Meakins-Christie Laboratories, McGill University, 1001 Boulevard Décarie, Montréal, QC H4A 3J1, Canada

[*]Corresponding author. Tel: 1 (603) 646-8656; Fax: 1 (603) 646-8778

E-mail address: Jane.E.Hill@dartmouth.edu

Materials and Methods

Analytical instrumentation and statistical analysis

Thermal desorption tubes were analyzed in a Pegasus 4D (LECO Corporation, St. Joseph, MI) GC×GC-TOFMSinstrument with an Agilent 7890 GC equipped with a Thermal Desorption Unit (TDU), Cooled Inlet System (CIS), and a Multi-Purpose Sampler (MPS) autosampler (Gerstel, Linthicum Heights, MD). Data acquisition and analysis were performed using ChromaTOF software, version 4.50 (LECO Corp.).

Chromatographic data were processed and aligned using ChromaTOF. For peak identification, a signal-to-noise (S/N) cutoff was set at 100:1, and resulting peaks were identified by a forward search of the NIST 2011 library. For tentative peak identification, a forward match score of 750 (of 1000) was required. For the alignment of peaks across chromatograms, maximum first and second-dimension retention time deviations were set at 6s and 0.2s, respectively, and the inter-chromatogram spectral match threshold was set at 600. Artifacts such as siloxane and phthalates were identified and deleted with the support of the script tool available in ChromaTOF®. The scripts used to identify linear siloxanes (defined as “siloxane” in the script) and phthalates were adopted from Weggler et al. [26]. Additional scripts were also developed to define cyclosiloxanes and silanols [27]. The flagged peaks using the aforementioned scripts were then manually checked and deleted from the peak table and not considered for further data analysis.

All statistical analyses were performed using R (version 3.3.0). Data were log-transformed and normalized using Probabilistic Quotient Normalization [28]. The ‘limma’ package was used to assess and address batch effect in our dataset by the month in which the experiment was conducted. The function removeBatchEffect() was used to fit a linear model to the data to adjust for batch related coefficients.

Random forest additional information RF creates many de-correlated decision trees from a randomly selected subset of volatile compounds and predicts the sample class assignment for the remaining samples not used in the building of decision trees obtaining a more accurate and stable prediction. Two-thirds of the samples in the training set are randomly selected with replacement for each decision tree and the remaining one-third are used to calculate the performance (accuracy) of the RF classification model. RF is a non-parametric approach that can deal with highly collinear data and it is resistant to different type of outliers.

Table S1. Summary of GC×GC-TOFMS parameters.

Thermal desorption tubes and desorption process	
Sorbent material	Carbopack Y, Carbopack X, Carboxen 1000
Solvent venting time (TDU temperature; flow rate)	10 min (30 °C; 60 mL/min)
Thermal desorption time (TDU temperature; flow rate)	5 min (330 °C; 60 mL/min)
Cryofocusing time (CIS temperature)	5 min (-120 °C)
Sample desorption time, CIS temperature, injection mode	180 s, 330 °C, splitless
GC×GC	
D1 column (L × ID × d_f)	Rxi-624Sil (60 m × 0.25 mm × 1.4 µm)
D2 column (L × ID × d_f)	Stabilwax (1 m × 0.25 mm × 0.5 µm)
D1 oven temperature	35 to 235°C at 3.5 °C/min
D2 oven and modulator temperature	+ 5°C and + 25 °C relative to D1 oven temp
Modulation period (hot/cold jet)	2.5 s total (alternating 0.85 s/0.4 s hot/cold)
Carrier gas (flow rate)	Helium (2 mL/min)
Transfer line temperature	260 °C
MS	
Acquisition mass range	30 - 400 *m/z*
Acquisition frequency	200 Hz
Ion source temperature	200 °C

TDU: Thermal desorption unit, CIS: L: length; ID: internal diameter; d_f: phase thickness

Results

Murine infection

The establishment of lung infections were verified by bacterial counts (CFU/lung) in lung homogenates and by evaluation of leukocyte differentiation in BAL fluid (**Supplementary Figure 1**).

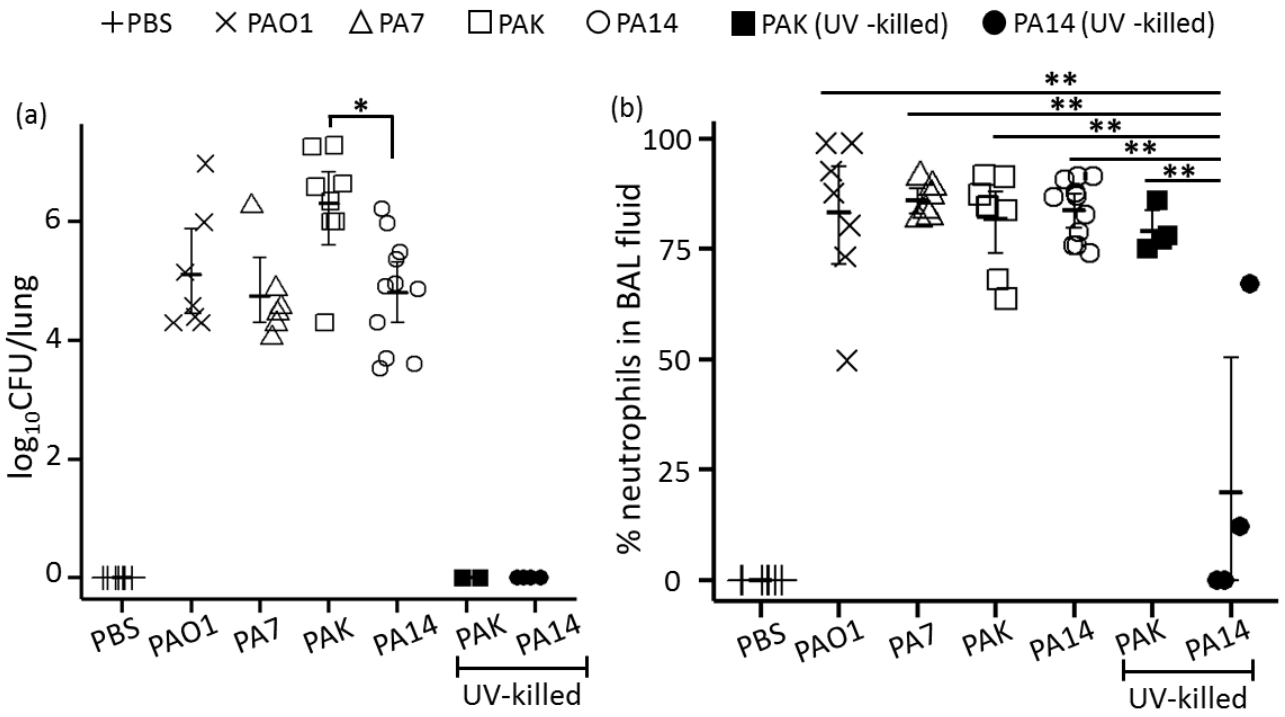

Figure S1 (a) *Bacterial CFU recovered from murine lung homogenates, and* ***(b)*** *neutrophils in bronchoalveolar lavage fluid, 24 hours post-inoculation. Each symbol represents results for an individual mouse and the bars are the mean for each group. Error bars represent the 95% confidence intervals. Significance was assessed using one-way ANOVA with Tukey's post-hoc test: * $p < 0.05$, ** $p < 0.001$. The number of mice in each experimental group were as follows: PBS = 7, PAO1 = 7, PA7 = 6, PAK = 8, PA14 = 11, UV-killed PAK = 4, and UV-killed PA14 = 4.*

P. aeruginosa **specific volatile molecules in murine breath**

In total, 3000 peaks were detected across the murine breath and room air samples analyzed in our study. After removal of known contaminants and artifacts (e.g., siloxanes, phthalates, atmospheric gases), 1802 peaks were selected for further analysis. PCA scores plot showed that samples clustered by month of analysis (**Figure S2a**). The contribution of batch effect in untargeted metabolomics is an important variable to consider prior to any statistical analysis, since it can decrease the power to detect biological relationships and introduce

spurious correlations. Therefore, a batch correction algorithm was applied removing the apparent trend. The batch corrected matrix was used for further analysis (**Figure S2b).**

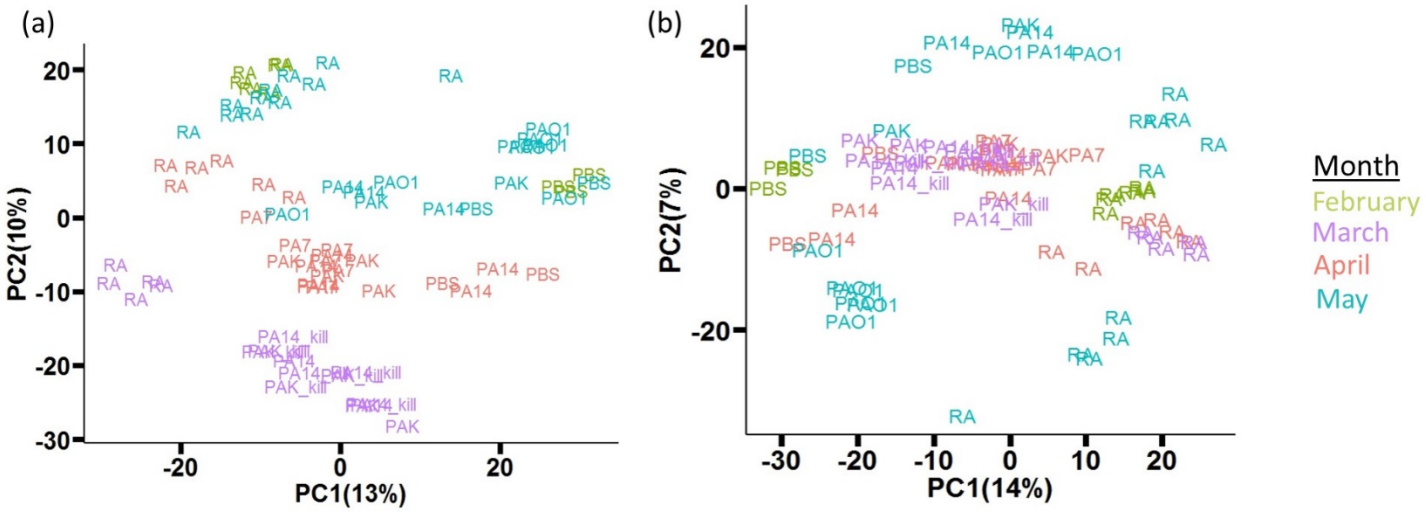

Figure S2 *Principle component analysis scores plot of 76 samples in the study (**a**) before, and (**b**) after batch correction. RA = room air.*

The number of peaks was then further reduced by the removal of features not statistically different from room air samples ($p > 0.05$), resulting in a total of 527 peaks. However, it was not possible visually discriminate the Live-group from controls (PBS and UV-killed breath samples) by performing a hierarchical cluster analysis of these 527 VOCs (**Figure S3**).

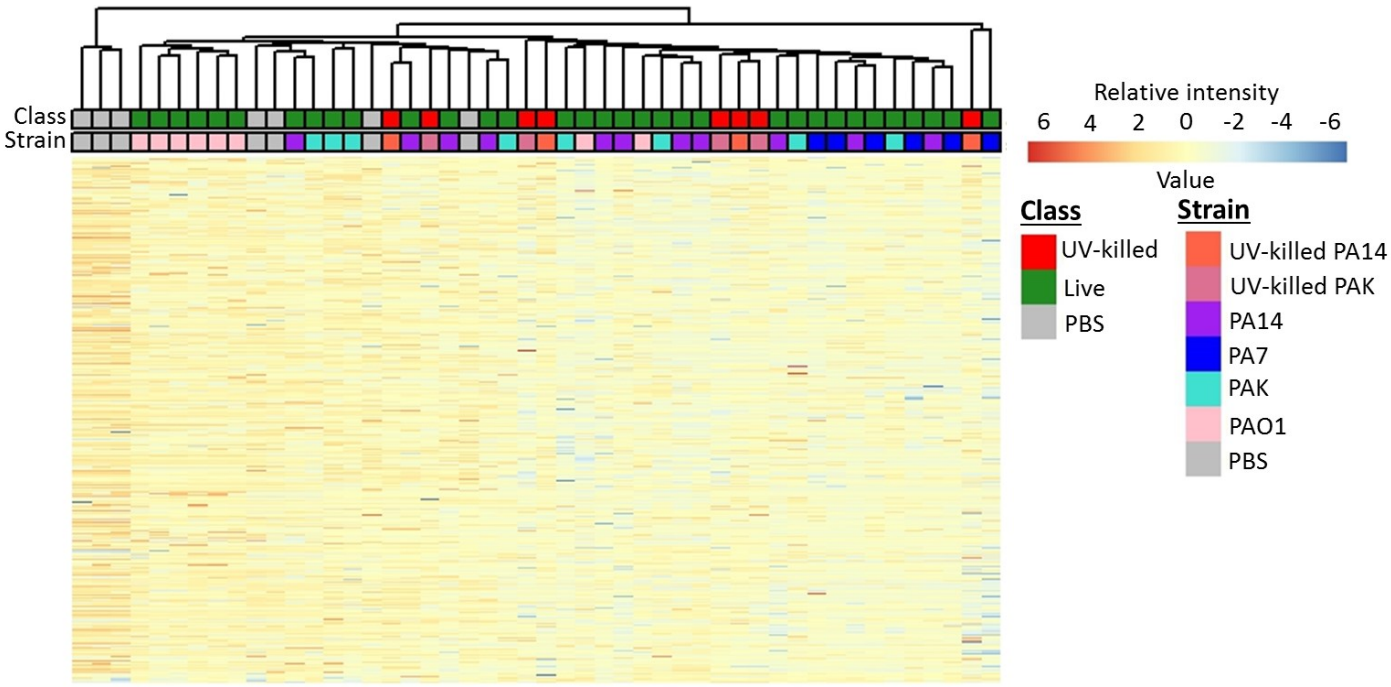

Figure S3. *Heatmap and hierarchical cluster analysis of breath profiles using 527 features that were statistically different from room air samples ($p < 0.05$, Mann-Whitney U test with Benjamini-Hochberg correction).*

Strain level evaluation in mice with P. aeruginosa lung infection

To assess whether volatile molecules in breath can identify strain level relationships we first performed a multiclass statistical test and removed the features not statistically significant for at least one strain ($p > 0.05$). The data matrix was thus reduced to 155 features. Two main clusters were depicted using hierarchical cluster analysis (**Figure S4**). The breath samples of mice inoculated with PAO1 and PA7 clustered well and separately from each other, while PA14 and PAK were interspersed almost equally in the two clusters (**Figure S4**).

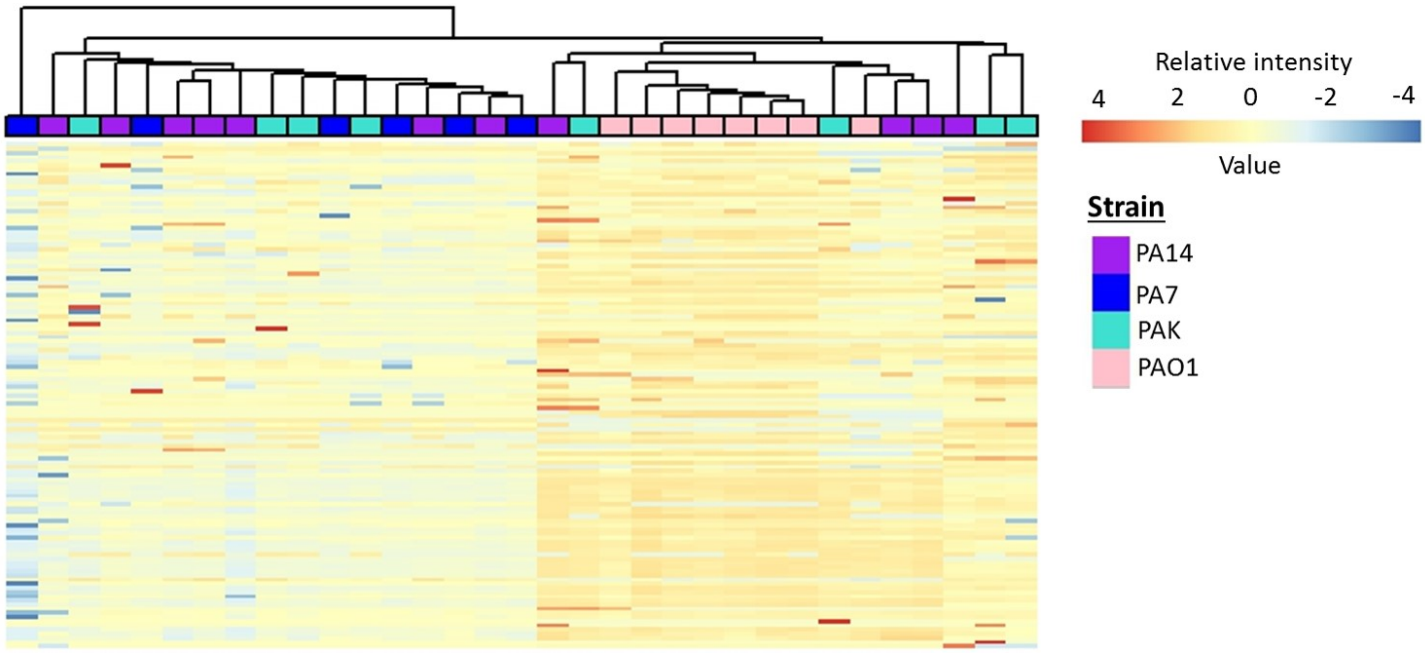

Figure S4. *Heatmap and hierarchical cluster analysis of breath profiles using 155 features that were statistically different between PA14, PA7, PAK and PAO1 ($p < 0.05$, Kruskal-Wallis test with Benjamini-Hochberg correction).*

References

1. Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R, Chun-Rong L, Guenthner D, Bovee D, Olson M V, Manoil C. Comprehensive transposon mutant library of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. U. S. A.* [Internet] 2003 [cited 2018 Jun 13]; 100: 14339–14344Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.2036282100.

2. Rahme LG, Stevens EJ, Wolfort SF, Shao J, Tompkins RG, Ausubel FM. Common virulence factors for bacterial pathogenicity in plants and animals. *Science (80-.).* [Internet] 1995 [cited 2018 Jun 13]; 268: 1899–1902Available from: http://www.ncbi.nlm.nih.gov/pubmed/7604262.

3. Minamishima Y, Takeya K, Ohnishi Y, Amako K. Physicochemical and biological properties of fibrous Pseudomonas bacteriophages. *J. Virol.* [Internet] 1968 [cited 2018 Jun 13]; 2: 208–213Available from: http://www.ncbi.nlm.nih.gov/pubmed/4986905.

4. Roy PH, Tetu SG, Larouche A, Elbourne L, Tremblay S, Ren Q, Dodson R, Harkins D, Shay R, Watkins K, Mahamoud Y, Paulsen IT. Complete genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7. Ahmed N, editor. *PLoS One* [Internet] 2010 [cited 2018 Jun 13]; 5: e8842Available from: http://dx.plos.org/10.1371/journal.pone.0008842.