Comparative analysis of the volatile metabolomes of *Pseudomonas aeruginosa* clinical isolates

Heather D Bean$^{1,4}$, Christiaan A Rees$^{2,4}$ and Jane E Hill$^{1,2,4}$

1 School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA
2 Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA
3 Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA
4 These authors made equal contributions to this work.

E-mail: jane.e.hill@dartmouth.edu

Keywords: *Pseudomonas aeruginosa*, GC × GC-TOFMS, bacterial infections, biomarkers, comparative metabolomics

Abstract

*Pseudomonas aeruginosa* is a nearly ubiquitous Gram-negative organism, well known to occupy a multitude of environmental niches and cause human infections at a variety of bodily sites, due to its metabolic flexibility, secondary to extensive genetic heterogeneity at the species level. Because of its dynamic metabolism and clinical importance, we sought to perform a comparative analysis on the volatile metabolome (the ‘volatilome’) produced by *P. aeruginosa* clinical isolates. In this study, we analyzed the headspace volatile molecules of 24 *P. aeruginosa* clinical isolates grown in vitro, using 2D gas chromatography time-of-flight mass spectrometry (GC × GC-TOFMS). We identified 391 non-redundant compounds that we associate with the growth and metabolism of *P. aeruginosa* (the ‘pan-volatilome’). Of these, 70 were produced by all 24 isolates (the ‘core volatilome’). Sixty-five of the detected compounds could be assigned putative compound identifications, of which 43 had not previously been associated with *P. aeruginosa*. Using the accessory volatile molecules, we determined the inter-strain variation in the metabolomes of these isolates, clustering strains by their metabotypes. Assessing the extent of metabolic diversity in *P. aeruginosa* through an analysis of the volatile molecules that it produces is a critical next step in the identification of novel diagnostic or prognostic biomarkers.

Introduction

*Pseudomonas aeruginosa* is a bacterium that is well known for its metabolic flexibility [1, 2], enabling it to thrive in numerous environmental niches. Additionally, *P. aeruginosa* is a broad-host range opportunistic pathogen, infecting plants, invertebrates, and animals (including humans), and is implicated in a wide range of clinical infections including otitis media, wound infections, and acute and chronic pneumonias [3, 4]. Seminal genomics analyses over the past decade attribute *P. aeruginosa*’s capabilities to a relatively large genome (mean size of 6.6 Mbp [5]), containing an average of 6175 genes per isolate [6], as well as its disproportionate large number of regulatory elements, comprising approximately 10% of an isolate’s genes [5]. More recently, several medium- and large-scale *P. aeruginosa* genomics studies (<20 and >100 genomes, respectively) [6–11] have provided a more nuanced view of the genetic capacity of the species. In one of the largest studies to-date, Mosquera-Rendon et al analyzed 181 genomes, and identified 16 820 non-redundant genes, constituting the predicted *P. aeruginosa* pan-genome. Only 15% of the genes (n = 2305) were identified as core genes, defined as being present in >95% of sequenced isolates [6]. More than half were accessory genes, defined as being present in 5–95% of *P. aeruginosa* genomes, and the remainder (31%) consisted of rare genes, harbored by less than 5% of the strains. Therefore, genetic diversity across strains likely contributes to *P. aeruginosa*’s metabolic diversity and niche ubiquity [6, 7].

Complicating the mission to understand *P. aeruginosa*’s flexibility is our limited ability to translate genomic data into confident predictions on niche occupancy, phenotypes, and metabolism. Several sticking points obstruct progress. First, the vast majority of *P. aeruginosa* genes that have been sequenced lack any
known or predicted function. For example, the majority of rare genes are without functional annotations in the KEGG database [6], and Ozer and colleagues reported that nearly 40% of P. aeruginosa core genes are ‘poorly characterized’, without any known or predicted function. Second, as little as 17% of an individual isolate’s genome consists of core genes [7], and the remainder is highly variable from one isolate to another [6, 8, 9]. Transcriptional analyses of the accessory genes from 150 P. aeruginosa clinical isolates revealed that sub-groups of strains share only small sets of accessory genes, and no large common gene clusters were identifiable [8]. Therefore, bottom-up approaches to P. aeruginosa metabolic network discovery (i.e., using genomes to predict metabolomes) generates limited information. More directly, in spite of previous studies, we know only a fraction of the P. aeruginosa metabolome [2, 12–14], particularly secondary metabolites, hindering top-down approaches (using metabolites to predict genetic pathways) to identify unknown metabolism-associated genes and their interactions.

A more thorough understanding of the P. aeruginosa metabolome could have significant positive impact on the clinical management of this pathogen [15, 16]. In particular, identifying the complete metabolic repertoire of clinical strains—or the clinical pan-metabolome—could shed light on mechanisms by which this organism establishes disease or evades therapy. Characterizing the core metabolome could lead to new therapeutic targets as well as the identification of biomarkers for highly sensitive diagnostics of any P. aeruginosa infection, regardless of strain or infection site. Conversely, the accessory metabolome could yield biomarkers that are specific to clinically-relevant phenotypes, such as virulence, antibiotic resistance, or mucoidy. Rare or unique metabolites could be used as sentinels or tracers of epidemic strains.

We and others have been studying the P. aeruginosa volatile metabolome, or ‘volatilome’, yielding approximately 125 unique volatile metabolites for this species via analysis of roughly 100 unique strains [17–37]. We hypothesize that, akin to the characteristics of the pan-genome, there is a core volatilome for P. aeruginosa clinical strains that constitutes a small fraction of the full volatilome of this species. We analyzed the volatile molecules from 24 isolates, cultured in vitro, using comprehensive 2D gas chromatography—time-of-flight mass spectrometry (GC × GC-TOFMS). We detected nearly 400 volatile compounds that contribute to the P. aeruginosa pan-volatilome, and identified 70 metabolites that are core to these clinical strains under the conditions tested here, 43 of which have not been previously identified as part of the P. aeruginosa volatile metabolome. In addition, we identified a total of 321 accessory and rare volatile molecules and used these data to assess the inter-strain variability in their volatile profiles, and to estimate the size of the pan-volatilome and core volatilome under these conditions.

Materials and methods

Bacterial isolates and sample preparation

Twenty-four P. aeruginosa clinical isolates from infections of the eye (n = 6), ear (n = 1), throat (n = 1), respiratory tract (n = 7), abdomen (n = 3), urinary tract (n = 1), blood (n = 1), and skin (n = 4) were cultured aerobically at 37 °C in LB-Lennox (10 g tryptone, 5 g yeast extract, 5 g NaCl per liter). After 24 h of growth, the bacterial cultures were cooled on ice, cells pelleted via centrifugation, and the supernatants filtered through 0.22 µm PES membranes to remove all remaining cells. Five milliliters of supernatant and a stir bar were sealed into 20 ml GC headspace vials with PTFE/silicone caps. Three biological replicates were prepared for each isolate. Six LB-Lennox sterile media controls were prepared in parallel through all steps of sample processing. All samples were stored at −20 °C prior to analysis.

Volatile molecule collection and analysis

Culture supernatants were stirred and heated to 50 °C and the headspace was sampled for 30 min via solid phase microextraction (SPME) using a 2 cm triphase fiber (divinylbenzene/carboxen/polydimethylsiloxane, 50/30 µm; Supelco/Sigma-Aldrich, St. Louis, MO). The volatile molecules were desorbed for 180 s at 270 °C, and injected splitlessly onto a comprehensive 2D gas chromatograph coupled with a time-of-flight mass spectrometer (GC × GC-TOFMS; Pegasus 4D, LECO Corp., St. Joseph, MI), equipped with an autosampler (Multipurpose Sampler, Gerstel, Inc., Linthicum Heights, MD). Separation of volatile molecules was achieved using a 2D column set consisting of an Rxi-624Sil (60 m × 250 µm × 1.4 µm; length × internal diameter × film thickness; Restek, Bellefonte, PA) first column and a Stabilwax (1 m × 250 µm × 0.5 µm; Restek) second column, joined by a press-fit connection. The primary oven (containing the Rxi-624Sil column) was initiated at 35 °C (0.5 min hold) and ramped to 230 °C (5 min hold) at 5 °C min⁻¹. The secondary oven (containing the Stabilwax column) and the quad-jet modulator were heated in step with the primary oven, with a +5 °C and +30 °C offset, respectively, relative to the primary oven temperature. A 2 s modulation period was used, with 0.5 s alternating hot and cold pulses. The helium carrier gas flow rate was 2 ml min⁻¹. The mass spectrometer transfer line was heated to 250 °C. Mass spectra were acquired over the range of 30–500 m/z at a rate of 200 Hz. Retention indices (RI) for the sample volatile molecules were calculated using external alkane standards (C₆–C₃₇). Headspace volatile molecules of a pure retention index mixture (Sigma-Aldrich, St. Louis, MO) were sampled using the 2 cm triphase SPME fiber for 10 min at 50 °C and desorbed at a 30:1 split. RIs for compounds eluting prior to hexane (C₆) or after pentadecane (C₁₅) were extrapolated.
Data processing and analysis

Data collection, processing, and alignment were performed using ChromaTOF and the Statistical Compare software package, version 4.50 (Leco Corp.). The baseline was drawn through the middle of the noise and the signal-to-noise (S/N) cutoff for initial peak finding was set to 100 for a minimum of two apexing masses. Subpeaks were combined when the second dimension retention time shift was ≤0.1 s early for subsequent modulation periods and the mass spectral match score was ≥600. For aligning peaks across chromatograms, maximum retention time shifts of 2 s (one modulation period) in the first dimension and 0.2 s in the second dimension were allowed, and minimum mass spectral similarity matches of 600 were required. A second round of peak discovery was performed on the aligned chromatograms, adding masses. Subpeaks were combined when the second dimension retention time shift was 0.1 s early for one chromatogram and the signal-to-noise (S/N) cutoff for initial peak finding was set to 100 for a minimum of two apexing masses.

Peaks were defined as absent, detected, or present for each clinical isolate of P. aeruginosa if they were observed in 0, 1–2, or all 3 biological replicates, respectively. Using these definitions, volatile molecules were classified as core, rare, or accessory metabolites according to the following criteria:

- core: present in all 24 clinical isolates;
- rare: present in only one clinical isolate and detected in ≤ 6 of the remaining 23 isolates;
- accessory: present in 2–23 clinical isolates or present in one isolate and detected in > 6 of the remaining 23 isolates.

Level 3 compounds were identified on ≥85% mass spectral match to the NIST 2011 library. Level 4 compounds have mass spectral matches <85%, but can still be differentiated based upon mass spectral data. Chemical classifications assigned to level 4 compounds are based upon characteristic mass fragments or patterns of neutral loss.

Statistical analyses

All statistical analyses were performed using R version 3.2.2. Peaks eluting before 358 s were removed from the peaktables, and the relative abundances of the remaining compounds were normalized across chromatograms using Probabilistic Quotient Normalization [39]. Intraclass correlations (ICC) were calculated for the 24 P. aeruginosa strains, and peaks with: (1) ICC > 0.4, and (2) greater mean intensities in samples versus sterile media controls (including zeros for blanks, but excluding zeros for samples) were retained for further analysis. For peaks that were detected in both the samples and the sterile media controls, the statistical significance of the differences between sample and control peak intensities were determined using the one-tailed non-parametric Mann–Whitney U-test [40] with Benjamini–Hochberg correction [41], with a significance threshold of p < 0.05 selected. For peaks that had high strain-to-strain variability within the P. aeruginosa samples (yielding large U-test p-values), an ICC > 0.8 was used as an alternative measure of statistical significance. Only peaks that were either (1) significantly more abundant in samples relative to sterile media, or (2) not detected in any sterile media samples are reported, and these peaks constitute the pan-volatilome described in this study.

Peaks were defined as absent, detected, or present for each clinical isolate of P. aeruginosa if they were observed in 0, 1–2, or all 3 biological replicates, respectively. Using these definitions, volatile molecules were classified as core, rare, or accessory metabolites according to the following criteria:

- core: present in all 24 clinical isolates;
- rare: present in only one clinical isolate and detected in ≤ 6 of the remaining 23 isolates;
- accessory: present in 2–23 clinical isolates or present in one isolate and detected in > 6 of the remaining 23 isolates.

For inter-strain metabolomics comparisons, the average peak intensity of each compound in a given isolate was calculated as the arithmetic mean of the three biological replicates, and an arbitrarily small value of 1 was assigned for compounds that were not detected. Average peak intensities were (1) log-transformed, (2) mean-centered, and (3) unit-scaled. Mean-centering and unit-scaling were performed for each compound individually. Hierarchical clustering analysis (HCA) was employed to assess the metabolic relatedness of isolates using all volatile molecules belonging to the accessory suite, with a Euclidean distance metric.

Accumulation and rarefaction curves were created using a methodology previously described by Humbert and colleagues for the analysis of genomic data [42]. A binary scale was generated to describe the occurrence of volatile metabolites in each isolate (0 if absent or detected, 1 if present). A total of 500 iterations were performed to generate both the accumulation and rarefaction curves, using a subset of data selected at random and without replacement.

Results and discussion

The volatile metabolome of P. aeruginosa clinical isolates grown aerobically in rich medium

From the collection, processing, and alignment of 78 GC × GC chromatograms—72 P. aeruginosa samples and 6 sterile medium controls—we identified 1752 non-redundant peaks that were observed in at least 3 chromatographic analyses. Through statistical analysis...
Comparisons between the samples and controls (see data in supplementary table S1, available at: stacks.iop.org/JBR/10/047102/mmedia), we conservatively attribute 391 peaks to the P. aeruginosa pan-volatilome (table S1) and identified 70 volatile molecules that were present in all three biological replicates of all 24 isolates, which we defined as the core volatilome (figure 1 and table S1). Two hundred and sixty-nine volatile molecules were present in a subset of isolates, deemed the accessory volatilome, and 52 were produced by only single clinical isolates in the study, which we call the rare volatilome (table S1). Based on mass spectral fragmentation patterns and 2nd dimension retention times ($t_R$) [43], chemical class assignments could be determined for a subset of the volatile molecules within the core ($n = 36$), accessory ($n = 54$), and rare ($n = 5$) volatilomes (figure 2).

Ketones were the most highly-represented class (38 volatile molecules of 95, 40% of the total), which recapitulates several previous studies that identified ketones as key constituents of the P. aeruginosa volatilome [18, 26, 32]. Hydrocarbons were next highest, constituting 28% of all compounds for which we could determine chemical class, while the remaining five classes (alcohols, aldehydes, esters, aromatics, and others), accounted for only 29% combined. Identifying chemical classes by electron impact mass spectrometry and GC × GC retention times is easier for some classes than others, which introduced bias into our analysis. For example, ketones and hydrocarbons are readily identifiable by the combination of their base ions of $m/z = 43, 57,$ and 71 and their second dimension retention times ($t_R = 0.80 \pm 0.12$ s and $0.65 \pm 0.03$ s, respectively; table S1) making these compound classes easy to detect via extracted ion chromatograms (XICs) and GC × GC retention time patterns. Alcohols, in contrast, are characterized by the neutral loss of water (i.e. ions in the mass spectra that differ by $m/z = 18$), which cannot be visualized using XICs, and these compounds do not have narrowly-defined $t_R$, like the hydrocarbons (table S1). This bias, in combination with the incompleteness of mass spectral and retention index libraries for many metabolites, makes it impossible to draw conclusions about differences in the variety of volatile compounds observed between the core, accessory, and rare metabolomes. However, it is interesting to note that ketones are well represented across the three volatilome categories (figure 2), suggesting a variety of roles for these compounds in P. aeruginosa diagnostics, with core ketones contributing to a suite of species-level biomarkers, as proposed by Shestivska et al [32], and combinations of accessory or rare ketones serving as unique strain identifiers.

Volatile molecules associated with the in vitro growth of P. aeruginosa clinical isolates

To provide putative identifications for the P. aeruginosa volatile molecules that we observed, we compared both mass spectral and retention time data to published
Table 1. Putatively identified volatile metabolites of *P. aeruginosa* clinical isolates.

| #  | Compound name | Formula | Weight | CAS   | Chemical class | ID level | $t_R$ (s) | $t_R$ (s) | RI | Samples (72) |
|----|---------------|---------|--------|-------|----------------|----------|----------|----------|----|--------------|
| 1  | acetone       | C₅H₆O  | 58     | 67-64-1 | KET           | 2        | 358      | 0.76     | 550| 72           |
| 2  | dimethyl sulfide | C₅H₁₀S | 62     | 75-18-3 | OTH           | 2        | 366      | 0.69     | 554| 72           |
| 3  | methyl vinyl ketone | C₅H₈O | 70     | 78-94-4 | KET           | 2        | 518      | 0.85     | 628| 72           |
| 4  | 2-butanone    | C₅H₁₀O | 72     | 78-93-3 | KET           | 2        | 534      | 0.78     | 636| 72           |
| 5  | 3-methyl-2-butanone | C₆H₁₄O | 86     | 563-80-4 | KET           | 2        | 666      | 0.75     | 700| 72           |
| 6  | 3-methyl-3-buten-2-one | C₆H₁₀O | 84     | 814-78-8 | KET           | 2        | 690      | 0.81     | 710| 72           |
| 7  | 2-pentanone   | C₅H₁₀O | 86     | 107-87-9 | KET           | 2        | 729      | 0.78     | 728| 72           |
| 8  | 2,4-dimethylfuran | C₆H₁₀O | 96     | 3710-43-8 | OTH       | 2        | 740      | 0.75     | 732| 72           |
| 10 | 2,3-pentanedione | C₆H₁₀O₂ | 100    | 600-14-6 | KET          | 2        | 743      | 0.86     | 734| 72           |
| 11 | dimethyl disulfide | C₂H₆S₂ | 94     | 624-92-0 | OTH           | 2        | 843      | 0.84     | 777| 72           |
| 12 | methyl isobutyl ketone | C₄H₁₀O | 100    | 108-10-1 | KET          | 2        | 850      | 0.76     | 780| 72           |
| 13 | 3-methyl-2-pentanone | C₅H₁₀O | 100    | 563-61-7 | KET          | 2        | 882      | 0.76     | 794| 72           |
| 14 | 2,3-hexanediene | C₆H₁₀O₂ | 114    | 3848-24-6 | KET      | 2        | 947      | 0.83     | 822| 72           |
| 16 | pyrrole       | C₅H₇N  | 67     | 109-97-7 | OTH          | 2        | 973      | 1.88     | 833| 72           |
| 17 | 2,4-dimethyl-1-heptene | C₇H₁₄O | 126    | 19549-87-2 | CH       | 2        | 1004     | 0.65     | 847| 72           |
| 19 | 3-methyl-3-penten-2-one | C₅H₁₀O | 98     | 565-62-8 | KET          | 2        | 1092     | 0.85     | 884| 72           |
| 21 | 3-heptanone   | C₅H₁₀O | 114    | 106-35-4 | KET          | 2        | 1184     | 0.76     | 925| 72           |
| 22 | 2-heptanone   | C₅H₁₀O | 114    | 110-43-0 | KET          | 2        | 1200     | 0.78     | 933| 72           |
| 24 | decane        | C₅H₁₀ | 142    | 124-18-5 | CH           | 2        | 1346     | 0.64     | 999| 72           |
| 29 | dimethyl trisulfide | C₂H₆S₃ | 126    | 3658-80-8 | OTH     | 2        | 1386     | 0.90     | 1018|72            |
| 30 | 3-octanone    | C₆H₁₂O | 126    | 106-68-3 | KET          | 2        | 1404     | 0.75     | 1027|72            |
| 31 | 2,4,6-trimethylpyridine | C₈H₁₁N | 121    | 108-75-8 | ARO         | 2        | 1405     | 0.85     | 1027|72            |
| 41 | 2-nonanone    | C₉H₁₈O | 142    | 821-55-6 | KET          | 2        | 1627     | 0.76     | 1137|72            |
| 42 | acetophenone  | C₈H₁₂O | 120    | 98-86-2  | KET          | 2        | 1633     | 1.06     | 1140|72            |

(Continued)
or experimental values, when available (see materials and methods for details). Amongst the 391 peaks that we associate with the growth and metabolism of 
P. aeruginosa clinical isolates, we were able to assign putative identifications to 65 (17%) using a stringent threshold for mass spectral matching ($\geq 850/1000$ relative to the NIST 2011 library), with corroboration via retention time data for a subset of compounds (table 1). Of these 65 compounds, 25 belong to the core volatilome, and are included in figure 1, 36 are in the accessory volatilome, and 4 in the rare volatilome. We identified several oft-reported 
P. aeruginosa-associated volatile compounds in the core volatilome, such as the dimethyl sulfuric compounds (dimethyl sulfide, disulfide, and trisulfide; table 1) [17, 18, 25–34, 36]. 
P. aeruginosa’s highly characteristic odor compound, 2-aminoacetophenone [18, 20, 26, 29, 32], was present in 23 of the isolates in this study and absent from one (Urinary-1) making it part of the accessory metabolome. We identified eight 2-ketones in the core and accessory volatilomes, four each of the 3- and 4-ketones for the first time. Of the volatile compounds we identified, only 22 have previously been associated with the growth and metabolism of 
P. aeruginosa (supplementary table S1) [17, 18, 20–22, 25–36], and therefore we have added 43 novel compounds to the volatilome of this species. Many of the newly-reported volatile molecules were observed as part of the core but had not been observed in the headspace of approximately 100 previously-studied 
P. aeruginosa strains [17–37], likely because of low relative abundance (e.g., 3-hexanone), co-elutions with other compounds in 1D gas chromatography (e.g., 4-nonanone), or both (e.g., 3-methylheptane). Our high yield of 
P. aeruginosa volatile molecules in this study demonstrates the power and utility of GC × GC-TOFMS in discovering metabolites via untargeted analyses and this, in turn, reflects the diversity of compounds produced by this organism.

### Characterizing 
P. aeruginosa by its volatilome

Employing definitions analogous to those used for comparative genomics, we designate the pan-volatilome as including any volatile molecule produced by this set of clinical isolates, whereas the core volatilome includes the set that is shared across all isolates. Using the individual metabolome data for each strain, we estimated the sizes of the pan-volatilome and core volatilome for these 24 
P. aeruginosa clinical isolates via accumulation and rarefaction curves, respectively (figure 3). Both curves approach asymptotes (figure 3), indicating that the rate of discovery of new volatiles, and the rate of restraining the core volatilome is not unbounded in this set of experiments [44]. That is, a suitable number of isolates have been included in this study to capture the variation in the 
P. aeruginosa volatilome under these growth conditions and analytical methodology, and the addition of more strains is unlikely to significantly change the estimated core and pan-volatilome sizes, which stand at 70 and 391 compounds, respectively. If either curve was unbounded, that would indicate that the data we have

### Table 1. (Continued)

| # | Compound name | Formulas | Weight | CAS | Chemical class. | ID level | $t_R$ (s) | $t_R$ (s) | RI | Samples (72) |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 3-methylbutylacetate | $C_9H_{12}O_2$ | 172 | 72218-58-7 | EST | 3 | 1700 | 0.72 | 1175 | 69 |
| 4 | 2,4-octadienal | $C_6H_{10}O$ | 124 | 30361-28-5 | ALD | 2 | 1713 | 0.90 | 1182 | 16 |
| 5 | 3,6-dimethylpyrazine-2-carbonitrile | $C_7H_{11}N$ | 133 | 2435-47-4 | ARO | 3 | 1800 | 1.11 | 1229 | 61 |
| 6 | isopropyl phenyl ketone | $C_9H_{12}O$ | 148 | 611-70-1 | KET | 2 | 1974 | 0.93 | 1328 | 37 |
| 7 | 2-undecanone | $C_{11}H_{22}O$ | 170 | 112-12-9 | KET | 2 | 2000 | 0.75 | 1344 | 71 |
| 8 | decanenitrile | $C_9H_{16}O$ | 140 | 32064-72-5 | KET | 2 | 1701 | 0.79 | 1175 | 9 |
| 9 | 2,6,10-trimethyl-dodecane | $C_{12}H_{24}O$ | 184 | 6137-26-4 | KET | 2 | 2128 | 0.72 | 1423 | 50 |
| 10 | 2-aminoacetophenone | $C_8H_9NO$ | 135 | 551-93-9 | KET | 2 | 2112 | 1.42 | 1413 | 69 |
| 11 | 4-dodecanone | $C_{12}H_{24}O$ | 184 | 6175-49-1 | KET | 2 | 2169 | 0.74 | 1447 | 67 |
| 12 | 2-dodecanone | $C_{12}H_{24}O$ | 184 | 6175-49-1 | KET | 2 | 2169 | 0.74 | 1447 | 67 |
| 13 | 4-tridecanone | $C_{13}H_{26}O$ | 198 | 26215-90-7 | KET | 2 | 2289 | 0.73 | 1519 | 45 |
| 14 | 1-(2-nitrophenyl)-ethanone | $C_8H_7NO_3$ | 165 | 53475-15-3 | OTH | 3 | 1231 | 0.90 | 947 | 9 |
| 15 | 7-methyl-4-octanone | $C_9H_{16}O$ | 140 | 32064-72-5 | KET | 2 | 1701 | 0.79 | 1175 | 9 |
| 16 | 3,6-dimethylpyrazine-2-carbonitrile | $C_7H_7N_3$ | 133 | 2435-47-4 | ARO | 3 | 1800 | 1.11 | 1229 | 61 |
| 17 | 8-methyl-4-nonanone | $C_{10}H_{20}O$ | 156 | 6137-29-7 | KET | 3 | 1704 | 0.72 | 1177 | 4 |
| 18 | 2,6,10-trimethyl-dodecane | $C_{12}H_{24}O$ | 184 | 6137-26-4 | KET | 2 | 2128 | 0.72 | 1423 | 50 |
| 19 | 2,6,10-trimethyl-dodecane | $C_{12}H_{24}O$ | 184 | 6137-26-4 | KET | 2 | 2128 | 0.72 | 1423 | 50 |
| 20 | 2-dodecanone | $C_{12}H_{24}O$ | 184 | 6175-49-1 | KET | 2 | 2169 | 0.74 | 1447 | 67 |
| 21 | 4-tridecanone | $C_{13}H_{26}O$ | 198 | 26215-90-7 | KET | 2 | 2289 | 0.73 | 1519 | 45 |
| 22 | 1-(2-nitrophenyl)-ethanone | $C_8H_7NO_3$ | 165 | 53475-15-3 | OTH | 3 | 1231 | 0.90 | 947 | 9 |

**Rare volatile molecules**

| # | Compound name | Formulas | Weight | CAS | Chemical class. | ID level | $t_R$ (s) | $t_R$ (s) | RI | Samples (72) |
|---|---|---|---|---|---|---|---|---|---|---|
| 23 | 3-(methylthio)-2-butanol | $C_5H_{10}OS$ | 118 | 53475-15-3 | OTH | 3 | 1231 | 0.90 | 947 | 9 |
| 24 | 7-methyl-4-octanone | $C_9H_{16}O$ | 142 | 20809-46-5 | KET | 3 | 1512 | 0.73 | 1079 | 3 |
| 25 | 2-nonanone-4-one | $C_9H_{16}O$ | 140 | 32064-72-5 | KET | 2 | 1701 | 0.79 | 1175 | 9 |
| 26 | 8-methyl-4-nonanone | $C_{10}H_{20}O$ | 156 | 6137-29-7 | KET | 3 | 1704 | 0.72 | 1177 | 4 |

* Compounds in **bold** typeface have not been previously reported in 
P. aeruginosa volatilome studies. # = core analyte number used in figure 1; Chem. Class. = chemical classifications, ALC = alcohol, ALD = aldehyde, ARO = aromatic, CH = hydrocarbon, EST = ester, KET = ketone, OTH = other; ID level = confidence of the chemical identification, on a scale of 1–4 (see materials and methods), with 2 being the highest in this study [38]; $t_R$ = 1st and 2nd dimension retention times, respectively; RI = retention index.
collected is not sufficient for size estimations, and that more strains would be required to characterize the volatilome in this study [44]. However, we note that rare volatile molecules are underrepresented in this data set. Similar to the impact of rare genes on pan-genome estimates [6, 44], these compounds are expected to drive the upper-bounds of the accumulation curve, which plateau early in our analysis, while not necessarily strongly affecting the mean. We took a very conservative approach in the identification of rare volatile molecules, including only the rare peaks that were (1) present in all three biological replicates and (2) had low intra-strain versus inter-strain variability (ICC > 0.8). This approach disadvantaged the identification of compounds that were detected near the signal-to-noise threshold, which are susceptible to being reported as undetected in some replicates.

The size of the pan-volatilome in this study indicates that the ~125 volatile molecules that have been previously recorded for P. aeruginosa are but a fraction of the species’ volatile metabolome. And, our data on 24 isolates will necessarily underestimate the clinical pan-volatilome. The accumulation and rarefaction curves used to quantify a ‘pan-genome’ and ‘core genome’ assume a genetic diversity that is representative of the species as a whole, and the same assumptions are made when evaluating the core and pan-volatilome. Because the underlying genetic characteristics of these isolates is presently unknown, we cannot verify the isolates used here are representative of P. aeruginosa clinical isolates more broadly. Additionally, the pan-genome studies of P. aeruginosa indicate that many dozens of strains should be included for completeness [6, 7], and that these numbers can prove insufficient if there are dis-
The accumulation curve is largely shaped by the accessory volatiles in this study—i.e., the volatile molecules shared by subsets of the isolates—illustrating that these compounds capture the variation in the volatiles of the 24 clinical isolates. Using only the accessory volatile molecules for the analysis, we quantified the inter-strain variability and relatedness of the isolates. We hypothesized that isolates from related types of infection would be more metabolically-similar and would cluster together. The peak intensities for each accessory volatile were mean-centered and unscaled across all 24 clinical isolates, and visualized with a heat map (figure 4). Using these data, we calculated the metabolic relatedness amongst strains, depicted via a dendrogram with branch length inversely proportional to metabolic similarity (i.e., shorter branch lengths indicate more metabolically-similar isolates). Interestingly, four of the six eye infection isolates (Eye-1, -4, -5, -6) cluster together based on a few highly similar blocks of volatile metabolites. This hints at the ability to use the accessory volatome as a source of biomarkers for characterizing P. aeruginosa clinical strains linked to this specific site of infection, but the rest of the cluster analysis data indicate that type of information may not be generalizable to all infection sites. We do not, however, have sufficient genomic data on these isolates to determine the underlying source of the metabolic relatedness of the eye infection isolates under these growth conditions, to identify the reasons why most isolates from the same infection sites do not cluster, or to explore why the metabolome of respiratory-1 (RESP-1) exhibits greater similarity to sterile LB medium than to any of the other isolates studied. Additionally, we hypothesize that the growth conditions will dictate the infection site-specific biomarkers we could identify, and studies are underway to test this hypothesis.

A previous study by Shestivska et al explored the relationship between the volatile metabolomes and genomes of P. aeruginosa clinical isolates, and found that none existed [32]. However, that study was limited in both genomics and metabolomics data, utilizing the sequences of seven housekeeping genes and the relative abundances of six conserved metabolites to characterize the relatedness of 36 isolates. Based upon the sizes and variation of the pan-genome and pan-volatilome of P. aeruginosa, we hypothesize that correlations between genomes and metabolomes will be uncovered, but that much larger datasets and sample sizes will be required. We posit that metabolomic clustering of the eye infection isolates could also reflect shared phenotypes that are common amongst these types of infections, and the relationships between P. aeruginosa genotypes, phenotypes, and metabotypes will be topics of future studies.

Conclusions

To the best of our knowledge, we report here the most comprehensive study to-date to identify volatile metabolites produced by P. aeruginosa clinical isolates. Utilizing concepts from the field of comparative genomics, we performed a comparative analysis of the volatilomes of 24 P. aeruginosa clinical isolates grown aerobically in rich media, identifying 391 unique compounds in the pan-volatilome, of which 70 compounds are core to all of the clinical isolates in this study. As reported previously, P. aeruginosa produces a wide array of ketone compounds, and the use of GC × GC-TOFMS in this study allowed us to putatively identify 17 novel ketones in the P. aeruginosa volatilome repertoire, as well as 26 volatiles in other chemical classes. Two-thirds of the volatiles we identified were categorized as accessory volatiles, present in a subset of the P. aeruginosa isolates we characterized, which drives the variation in the volatilomes of these strains. We observed very little similarity in the volatile metabolomes of strains from related infection sites, but future investigations of P. aeruginosa clinical isolates will explore how genotype and phenotype relate to the metabotype, as well as the influence of in vitro growth conditions on the identification of volatile biomarkers for P. aeruginosa infections.

In the characterization of the P. aeruginosa volatilome, there is much more to be done. Metabolomes (as with transcriptomes, proteomes) are conditional, and therefore complete characterization of the pan-metabolome requires the exploration of many different strains and environmental conditions, including interactions with other species, which can modulate the P. aeruginosa metabolome [34, 45–49]. In light of these challenges, the pursuit of identifying and characterizing the P. aeruginosa pan-volatilome is a tall order, yet the data collected towards its completion will be valuable in the development of new diagnostic tools, therapeutic interventions, and basic insight into the metabolism of this pervasive opportunistic pathogen. This study represents a step in that direction.

Acknowledgments

We thank Dr Joseph Schwartzman and Dr Michael Zegans at Dartmouth-Hitchcock Medical Center, and Dr Deborah Hogan and Dr George O’Toole at the Geisel School of Medicine at Dartmouth College for providing P. aeruginosa strains used in this study. This work was funded through a Cystic Fibrosis Foundation Pilot and Feasibility Award (Hill16I0), a Dartmouth Neukom Institute CompX Faculty Grant, and a Dartmouth Clinical and Translational Science Institute Synergy Award. Dartmouth College holds an institutional program unifying population and laboratory based sciences award from the Burroughs Wellcome Fund, and CR was supported by this grant (Grant #: 1014106). HB was supported on a Cystic Fibrosis Foundation Postdoctoral Fellowship (Bean12F0). Research reported...
in this publication was supported by The Dartmouth Clinical and Translational Science Institute, under award number UL1TR001086 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH). The content is solely the responsibility of the author(s) and does not necessarily represent the official views of the NIH.

References

[1] Williams H D, Ziosnik J E A and Ryall B 2007 Oxygen, cyanide and energy generation in the cystic fibrosis pathogen Pseudomonas aeruginosa Advances in Microbial Physiology vol 52, ed R K Poole (London: Academic) pp 1–71
[2] Aral H 2011 Regulation and function of versatile aerobic and anaerobic respiratory metabolism in Pseudomonas aeruginosa Front. Microbiol. 2 103
[3] Bodey G P, Bolivar R, Fainstein V and Jadeja L 1983 Infections caused by Pseudomonas aeruginosa Rev. Infect. Dis. 5 279–313
[4] Lyczak J B, Cannon C L and Pier G B 2000 Establishment of Pseudomonas aeruginosa infection: lessons from a versatile opportunistic Microbes Infect. 2 1051–60
[5] Stover C K et al 2000 Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen Nature 406 959–64
[6] Mosquera-Rendon J, Rada-Bravo A M, Cardenas-Brito S, Chouvarine P, Tuemmler B and Neueseler S 2014 The extensive set of accessory Pseudomonas aeruginosa genomic components FEMS Microbiol. Lett. 356 235–41
[7] Frenzli L et al 2015 Clinical utilization of genomics data produced by the international Pseudomonas aeruginosa consortium Front. Microbiol. 6 1036
[8] Pohl S, Klockgether J, Eckowerider D, Khaledi A, Schniederjans M, Chouvarine P, Tuemmler B and Neueseler S 2014 The extensive set of accessory Pseudomonas aeruginosa genomic components FEMS Microbiol. Lett. 356 235–41
[9] Klockgether J, Cramer N, Wiehlmann L, Davenport C F and Tuemmler B 2011 Pseudomonas aeruginosa genomic structure and diversity Front. Microbiol. 2 150
[10] Hilkert R et al 2015 Intercellular gradient of virulence in the Pseudomonas aeruginosa pan-genome from disease and environment Environ. Microbiol. 17 29–46
[11] Ozer E A, Allen J P and Hauser A R 2014 Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools spine and AGEnt BMC Genomics 15 1–17
[12] Gross H and Loper J E 2009 Genomics of secondary metabolite production by Pseudomonas spp Nat. Prod. Rep. 26 1608–46
[13] Rojo F 2010 Carbon catabolite repression in Pseudomonas: optimizing metabolic versatility and interactions with the environment FEMS Microbiol. Rev. 34 658–84
[14] Sonnleitner E and Haas D 2011 Small RNAs as regulators of primary and secondary metabolism in Pseudomonas species Appl. Microbiol. Biotechnol. 91 63–79
[15] Croussilles A, Maunders E, Bartlett S, Fan C, Ukor F E, Abdelhamid Y, Baker Y, Floto A, Spring D R and Welch M 2015 Which microbial factors really are important in Pseudomonas aeruginosa infections? Future Microbiol. 10 1625–36
[16] Lamers R P and Burrows L L 2016 Pseudomonas aeruginosa: targeting cell-wall metabolism for new antibacterial discovery and development Future Med. Chem. 8 975–92
[17] Filipiak W, Sponring A, Bauer M, Filipiak A, Ager C, Wiesenhofer H, Nagl M, Tröppmair J and Amann A 2012 Molecular analysis of volatile metabolites released specifically by Staphylococcus aureus and Pseudomonas aeruginosa BMC Microbiol. 12 113
[18] Bean H D, Dimandja J-M D and Hill J E 2012 Bacterial volatile discovery using solid phase microextraction and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry J. Chromatogr. B 901 41–6
[19] Shestivska V, Nemec A, Dřevínek P, Sovová K, Dryahina K and Španěl P 2011 Quantification of methyl thioicynate in the headspace of Pseudomonas aeruginosa cultures and in the breath of cystic fibrosis patients by selected ion flow tube mass spectrometry Rapid Commun. Mass Spectrom. 25 2459–67
[20] Cox C D and Parker J 1979 Use of 2-aminoacetophenone production in identification of Pseudomonas aeruginosa J. Clin. Microbiol. 9 479–84 (PMID: 110829)
[21] Senecal A G, Magrone J, Yeomans W and Powers E M 2002 Rapid detection of pathogenic bacteria by volatile organic compound (VOC) analysis Proc. SPIE 4575 121–31
[22] Savelev S U, Perry J D, Bourke S J, Jary H, Taylor R, Fisher A J, Corris P A, Petrie M and De Souza A 2011 Volatile biomarkers of Pseudomonas aeruginosa in cystic fibrosis and noncystic fibrosis bronchiectasis Lett. Appl. Microbiol. 52 610–3
[23] Whiteson K L, Meinardi S, Lim Y W, Schmieder R, Maughan H, Quinn R, Blake D R, Conrad D and Rohwer F 2014 Breath gas metabolites and bacterial metagenomes from cystic fibrosis airways indicate active pH neutral 2,3-butanedione fermentation ISME J. 8 1247–58
[24] Julak J, Stranska E, Rosova V, Geppert P, Spanel P and Smith D 2006 Bronchoalveolar lavage examined by solid phase microextraction, gas chromatography–mass spectrometry and selected ion flow tube mass spectrometry J. Microbiol. Methods 65 76–86
[25] Zechman J M, Aldinger S and Labows J N 1986 Characterization of pathogenic bacteria by automated headspace concentration gas–chromatography J. Chromatogr. 377 49–67
[26] Labows J N, McGinley K J, Webster G F and Leyden J J 1980 Headspace analysis of volatile metabolites of Pseudomonas aeruginosa and related species by gas chromatography–mass spectrometry J. Clin. Microbiol. 12 521–6 (PMID: 6775012)
[27] Carroll W, Lenney W, Wang T S, Spanel P, Alcock A and Smith D 2003 Detection of volatile compounds emitted by Pseudomonas aeruginosa using selected ion flow tube mass spectrometry Pediatr. Pulmonol. 39 152–6
[28] Allard G, Langford V S, Hill L A and Murdoch D R 2006 Detection of volatile metabolites produced by bacterial growth in blood culture media by selected ion flow tube mass spectrometry (SIFT–MS) J. Microbiol. Methods 65 361–5
[29] Preti G, Thaler E, Hanson C W, Troy M, Eades J and Gelperin A 2009 Volatile compounds characteristic of sinus–related bacteria and infected sinus mucus: analysis by solid-phase microextraction and gas chromatography–mass spectrometry J. Chromatogr. B 877 2011–8
[30] Scholler C, Molin S and Wållin K 1997 Volatile metabolites from some gram-negative bacteria Chemosphere 35 1487–95
[31] Rudzinski C M, Herzig-Mars R, Lin J, Sapio A and Johnson B (ed) 2004 Pathogen detection using headspace analysis Scientific Conf. on Chemical and Biological Defense Research
[32] Shestivska V, Španěl P, Dryahina K, Sovová K, Smith D, Musilek M and Nemec A 2012 Variability in the concentrations of volatile metabolites emitted by genotypically different strains of Pseudomonas aeruginosa J. Appl. Microbiol. 113 701–13
[33] Zacheppank C, Wiegand H L, Lenzen C, Wingender J and Telgheimer U 2014 Investigation of volatile metabolites during growth of Escherichia coli and Pseudomonas aeruginosa by needle trap–GC-MS Anal. Bioanal. Chem. 406 6617–28
[34] Neerinck A H et al 2016 Identification of Pseudomonas aeruginosa and Aspergillus fumigatus mono- and co-cultures based on volatile biomarker combinations J. Breath Res. 10 016002
[35] Boots A W et al 2014 Identification of microorganisms based on headspace analysis of volatile organic compounds by gas chromatography–mass spectrometry J. Breath Res. 8 027106
[36] Thorns R M S, Reynolds D M and Greenman J 2011 Multivariate analysis of bacterial volatile compound profiles for discrimination between selected species and strains in vitro J. Microbiol. Methods 84 238–44
[37] Nizito K D, Perrault K A, Troobnikoff A N, Ueland M, Shoma S, Iredell J R, Middleton P G and Forbes S L 2016 In vitro volatile organic compound profiling using GC × GC-TOFMS to differentiate bacteria associated with lung infections: a proof-of-concept study J. Breath Res. 10 026008
[38] Sumner L W et al 2007 Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI) Metabolomics 3 211–21
[39] Dieterle F, Ross A, Schlotterbeck G and Senn H 2006 Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabolomics Anal. Chem. 78 4281–90
[40] Mann H B and Whitney D R 1947 On a test of whether one of two random variables is stochastically larger than the other Ann. Math. Stat. 18 50–60
[41] Benjamini Y and Hochberg Y 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing J. R. Stat. Soc. B 57 289–300
[42] Humbert J-F et al 2013 A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium Microcystis aeruginosa PLoS One 8 e70747
[43] Blomberg J, Schoenmakers P J, Beens J and Tijssen R 1997 Comprehensive two-dimensional gas chromatography (GC × GC) and its applicability to the characterization of complex (petrochemical) mixtures J. High Resolut. Chromatogr. 20 539–44
[44] Tettelin H, Riley D, Cattuto C and Medini D 2008 Comparative genomics: the bacterial pan-genome Curr. Opin. Microbiol. 11 472–7
[45] Filkins I M, Graber J A, Olson D G, Dolben E L, Lynd L R, Bhuj S and O’Toole G A 2015 Coculture of Staphylococcus aureus with Pseudomonas aeruginosa drives S. aureus towards fermentative metabolism and reduced viability in a cystic fibrosis model J. Bacteriol. 197 2252–64
[46] Price K E, Orazi G, Ruoff K L, Hebert W P, O’Toole G A and Mastoridis P 2015 Mannitol does not enhance tobramycin killing of Pseudomonas aeruginosa in a cystic fibrosis model system of biofilm formation PLoS One 10 e0141192
[47] Price K E, Naimie A A, Griffin E F, Bay C and O’Toole G A 2016 Tobramycin-treated Pseudomonas aeruginosa PA14 enhances Streptococcus constellatus 7155 biofilm formation in a cystic fibrosis model system J. Bacteriol. 198 237–47
[48] Anderson G G, Moreau-Marquis S, Stanton B A and O’Toole G A 2008 In vitro analysis of tobramycin-treated Pseudomonas aeruginosa biofilms on cystic fibrosis-derived airway epithelial cells Infect. Immun. 76 1423–33
[49] Briard B, Heddergott C and Latge J P 2016 Volatile compounds emitted by Pseudomonas aeruginosa stimulate growth of the fungal pathogen Aspergillus fumigatus MBio 7 e00219