# *PSEUDOMONAS AERUGINOSA* VOLATILOME CHARACTERISTICS AND ADAPTATIONS IN CHRONIC CYSTIC FIBROSIS LUNG INFECTIONS

Trenton J. Davis[a,b], Ava V. Karanjia[a,c], Charity N. Bhebhe[a], Sarah B. West[a], Matthew Richardson[d,e], and Heather D. Bean[a,b,#]

[a]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

[b]Center for Fundamental and Applied Microbiomics, The Biodesign Institute, Tempe, AZ, USA

[c]School of Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, AZ, USA

[d]Department of Respiratory Sciences, College of Life Sciences, University of Leicester, Leicester, UK

[e]NIHR Biomedical Research Centre (Respiratory theme), Institute for Lung Health, Leicester, UK

[#]Corresponding author: Heather D. Bean
Heather.D.Bean@asu.edu

**ABSTRACT**

*Pseudomonas aeruginosa* chronic lung infections in individuals with cystic fibrosis (CF) significantly reduce quality of life and increase morbidity and mortality. Tracking these infections is critical for monitoring patient health and informing treatments. We are working toward the development of novel breath-based biomarkers to track chronic *P. aeruginosa* lung infections *in situ*. Using comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry (GC×GC-TOFMS), we characterized the *in vitro* volatile metabolomes (or volatilomes) of 81 *P. aeruginosa* isolates collected from 17 CF patients over at least a five-year period of their chronic lung infections. We detected 539 volatiles produced by the *P. aeruginosa* isolates, 69 of which were core volatiles that were highly conserved. We found that each early infection isolate has a unique volatilome, and as infection progresses, the volatilomes of isolates from the same patient become increasingly dissimilar, to the point that these intra-patient isolates are no more similar to each other than to isolates from other patients. We observed that the size and chemical diversity of *P. aeruginosa* volatilomes do not change over the course of chronic infections; however, the relative abundances of core hydrocarbons, alcohols, and aldehydes do change, and are correlated to changes in phenotypes associated with chronic infections. This study indicates that it may be feasible to track *P. aeruginosa* chronic lung infections by measuring changes to the infection volatilome, and lays the groundwork for exploring the translatability of this approach to direct measurement using patient breath.

**KEYWORDS**

*Pseudomonas aeruginosa*; cystic fibrosis; metabolomics; volatile organic compounds; chronic infection; adaptation

**RUNNING TITLE**

*P. aeruginosa* cystic fibrosis volatilome

53 **INTRODUCTION**

54 Cystic fibrosis (CF) is an autosomal recessive disease caused by a mutation in the
55 CFTR protein that regulates ion transport across epithelia. In the lungs, reduced or lost CFTR
56 function leads to defective mucociliary transport, which facilitates infection and colonization by a
57 plethora of microorganisms (1). *Pseudomonas aeruginosa* is one of the most prevalent lung
58 pathogens in CF – especially after adolescence (2) – and is able to establish chronic infections
59 that can last for years to decades (3, 4). *P. aeruginosa* lung infection is associated with more
60 rapid lung function decline, increased risk of hospitalization, and increased risk of death (5, 6).
61 Significant clinical efforts are therefore aimed at diagnosing and treating new infections to delay
62 the onset of chronic infection (7). As *P. aeruginosa* transitions from an acute infection to
63 chronicity, it undergoes a variety of genotypic, phenotypic, and metabolic changes (8, 9). It has
64 been well-established that several phenotypes are correlated with chronic infection, including
65 reduced motility and increased mucoidy, antibiotic resistance, and biofilm formation (9-11), and
66 some of these phenotypes are also correlated with poorer patient outcomes (6, 12, 13). In light
67 of this, *P. aeruginosa* phenotypes could be used to evaluate infection stage, CF disease
68 progression, and patient morbidity risk (10, 12, 14, 15). Poor access to infection sites in the
69 lower airways, however, delays the detection of new *P. aeruginosa* infections and reduces the
70 feasibility and accuracy of disease state tracking via bacterial phenotypes, whether by culture-
71 dependent or culture-independent methods (16).

72 We are developing breath-based diagnostics to detect new *P. aeruginosa* infections and
73 track chronic infection phenotypes in the CF lung. Breath consists of thousands of volatile
74 organic compounds containing a wealth of information about human health and disease (17,
75 18), and is being leveraged for the development of diagnostics for wide-ranging conditions (19,
76 20). In the context of chronic lung infections, breath-based diagnostics possess significant
77 advantages over other diagnostic modalities. First, breath provides a non-invasive way of
78 sampling the entire ventilated lung, a limitation of both sputum and bronchoalveolar lavage fluid

79 that leads to delayed diagnosis of new infections, or the detection of infection phenotype
80 changes (16, 21-24). Second, breath sampling captures metabolic information about the
81 pathogens *in situ*, eliminating the need for microbial enrichment steps and speeding the time-to-
82 diagnosis by days to weeks (25, 26). Additionally, *in situ* metabolic measurements provide
83 information on the physiology of the pathogens in the context of disease versus the context of
84 the lab. We propose that a breath-based diagnostic can be created for detecting and tracking
85 CF *P. aeruginosa* lung infections by correlating the microbial genotypic and phenotypic changes
86 that occur during chronic infections to changes in the infection volatile metabolome, or
87 "volatilome".

88 In order to develop a set of biomarkers for diagnosing and tracking *P. aeruginosa* lung
89 infections in CF, we must consider the genomic, phenomic, and volatilomic diversity of the
90 species, and how that diversity may be altered via adaptation to the CF lung environment. A
91 recent effort to sequence more than 1300 *P. aeruginosa* isolates demonstrates that the
92 genomes are highly flexible, with only 1% of the pan-genome conserved across all isolates (27).
93 An untargeted analysis of the volatile metabolites produced *in vitro* by 24 *P. aeruginosa* clinical
94 isolates shows that the volatilomes are also diverse, with 18% of the pan-volatilome conserved
95 in this comparatively small study (28). In the context of CF, most early *P. aeruginosa* infections
96 in young patients are caused by unique strains that come from the patients' environments (29-
97 32). We therefore expect the volatilomes early in infection to have high inter-patient dissimilarity,
98 reflecting the genomic diversity of *P. aeruginosa* CF isolates (3, 33-35), and to be chemically
99 diverse (*i.e.*, volatiles from a wide range of chemical classes). Once chronic infections are
100 established, however, the "founder" *P. aeruginosa* strain evolves into a population of clonal sub-
101 strains (3, 8, 36-40), many of which harbor CF-typical loss-of-function mutations in regulatory
102 genes and sigma factors (*e.g.*, *lasR* and *mucA*) (3, 41, 42). From this, we posit that the size of
103 the *P. aeruginosa* volatilome will shrink during chronic infection, with a concomitant decrease in
104 chemical diversity and a reduction in inter-patient dissimilarity.

105 To build a foundation for using volatilomes to diagnose and track *P. aeruginosa* CF lung
106 infections, we explored how the *in vitro* volatilome of *P. aeruginosa* CF isolates change over the
107 course of chronic infection. We used comprehensive two-dimensional gas chromatography
108 coupled with time-of-flight mass spectrometry (GC×GC-TOFMS) to characterize the volatilomes
109 of 81 *P. aeruginosa* isolates from early and late chronic CF lung infections from 17 persons with
110 CF. Using an untargeted metabolomics approach, we characterized the size and chemical
111 composition of the CF *P. aeruginosa* volatilome, and identified core volatiles that would be a
112 primary source of breath biomarkers for diagnosing infections. We also investigated how the *P.*
113 *aeruginosa* volatilome is shaped by the CF lung environment, characterizing changes in
114 volatilome sizes and compositions over time, and how these changes relate to intra-patient and
115 inter-patient volatilome dissimilarities.
116
117 **METHODS**
118 **Bacterial isolates**
119 Eighty one *P. aeruginosa* isolates from 17 individuals with CF were acquired from the Cystic
120 Fibrosis Isolate Core at Seattle Children's Hospital. From the majority of patients we obtained
121 three isolates: one early isolate, defined as the first cultured *P. aeruginosa* isolate, and two late
122 isolates collected a minimum of five years after the first isolate. For one patient (Patient 23), 32
123 additional isolates were collected at intervals over 7.5 years between the collections of the early
124 and two late isolates. For one patient, only one isolate was included in this study. We quantified
125 five phenotypes that are correlated with chronic infections: proteases, pyocyanin, rhamnolipids,
126 twitching motility, and mucoidy, using published methods (see Supplementary Information for
127 phenotype characterization methods). Twenty pairs of isolates (one Early and one Late) from
128 ten patients were selected for additional statistical analyses, described below, based on
129 changes to their phenotypes during chronic infections.
130

131 **Sample preparation**

132 Isolates were cultured as previously described (43). Briefly, isolates were cultured aerobically
133 for 16 h at 37 °C in 5 mL of LB Lennox (LB) broth, and then diluted 1000-fold into 25 mL of fresh
134 LB and grown for 24 h under the same conditions. For metabolomics analyses, cells were
135 removed via centrifugation through a 0.2 *μ*m filter and 2 mL of each filtrate was transferred to a
136 10 mL GC headspace vial with screw caps. Samples were prepared in biological triplicate, with
137 LB media controls prepared in parallel, and stored at -20 °C prior to analysis.

138

139 **GC×GC analysis and data processing**

140 Culture filtrates were thawed and maintained at 4 °C until analyzed, as previously described
141 (43). Headspace volatiles were characterized using a Pegasus® 4D GC×GC-TOFMS (Leco®
142 Corporation, St. Joseph, MI) equipped with a MPS Pro® rail autosampler (Gerstel® Inc,
143 Linthicum Heights, MD). Column set configuration, and GC×GC, MS, and data processing
144 method parameters are previously reported (43), and summarized in Supplementary Table 1a-
145 d.

146 Data processing steps are outlined in Fig. S1. Chromatographic artifacts and suspected
147 contaminants based on peak names were removed (Supplementary Table 2), as well as peaks
148 eluting prior to 358 s (acetone retention time) in the first dimension. Missing values were
149 handled as follows: if detected in one of three replicates, the measured value was permuted to
150 0; if detected in two of three replicates, the missing value was imputed as half of the minimum
151 detected value for that compound across all samples.

152 Probabilistic quotient normalization (PQN) was applied to account for differences in peak
153 abundance due to variations in culture cell density, followed by a $\log_{10}$ transformation (44).
154 Peaks were further filtered out based on the following criteria: (1) within-triplicate intraclass
155 correlation coefficient (ICC) < 0.75 (definition: absolute agreement; model: two-way mixed

156 effects; type: mean of *k* measurements) (45); and (2) only detected in sterile media, or not
157 significantly greater in abundance compared to sterile media using the Wilcoxon rank-sum test
158 with Benjamini-Hochberg adjustment (significance threshold of 0.05) (46).

159 For statistical analyses beyond the reporting of detected peaks, peaks that showed
160 significant correlations to run order were identified using Kendall's tau with Benjamini-Hochberg
161 adjustment ($-0.6 \leq \tau \geq 0.6$, significance threshold of 0.05) and removed. Principal component
162 analysis (PCA) revealed an apparent batch effect that was attributed to, at least in part, a non-
163 biological phenomenon and was described previously (43). This batch effect was corrected
164 using an empirical Bayes approach (47). The geometric means of triplicates of the batch-
165 corrected data were used for further statistical analyses. For analysis of the ten paired early and
166 late chronic infection isolates, batch-corrected data was not used as all of these isolates were
167 contained within the same batch.

168 Peaks were putatively identified using published reporting standards (48). Level 2
169 identifications were determined based on the following criteria: (1) ≥ 80% mass spectral forward
170 match using the NIST 2011 library; (2a) first dimension retention times that possess a strong
171 linear fit with carbon number in homologous series ($r^2 \geq 0.995$, for the identification of select 2-,
172 3-, and 4-ketones), or (2b) experimentally determined linear retention indices (LRI) consistent
173 with published LRIs, as determined by the following acceptance criteria based on the
174 differences between experimental and published non-polar and polar column RIs for Grob's test
175 mix:

176

177 $$\left(\frac{\mathrm{RI}_{\mathrm{experimental}} - \mathrm{RI}_{\mathrm{non-polar}}}{\mathrm{RI}_{\mathrm{polar}} - \mathrm{RI}_{\mathrm{non-polar}}}\right) \times 100 = 0 - 43\%$$

178

179 Level 3 identifications were based on at least an 80% mass spectral forward match score. Level
180 4 identifications are those that meet none of the above-mentioned criteria.

181

182 **Statistical analyses**

183 Volatile compounds that were detected in at least 95% of all samples were classified as "core"
184 volatiles. Significant differences between inter-isolate phenotypes were tested using the
185 Wilcoxon signed-rank test (significance threshold of 0.05). The relatedness of isolates based on
186 their volatile metabolomes were assessed using agglomerative hierarchical clustering analysis
187 (HCA), non-metric multidimensional scaling (NMDS), and permutational multivariate analysis of
188 variances (PERMANOVA) on the Euclidean distances between isolates. Linear regression and
189 Pearson's correlation of Euclidean distances of Patient 23 isolates were used to assess intra-
190 patient isolate dissimilarity over time. All statistical analyses were performed using R version
191 3.3.2 (The R Foundation for Statistical Computing); information regarding the R packages used
192 are in Supplementary Table 3.

193

194 **Data availability**

195 Metabolomic data (chemical feature peak areas and retention time information) included in this
196 study will be made available at the NIH Common Fund's National Metabolomics Data
197 Repository (NMDR) website, the Metabolomics Workbench, at
198 www.metabolomicsworkbench.org.

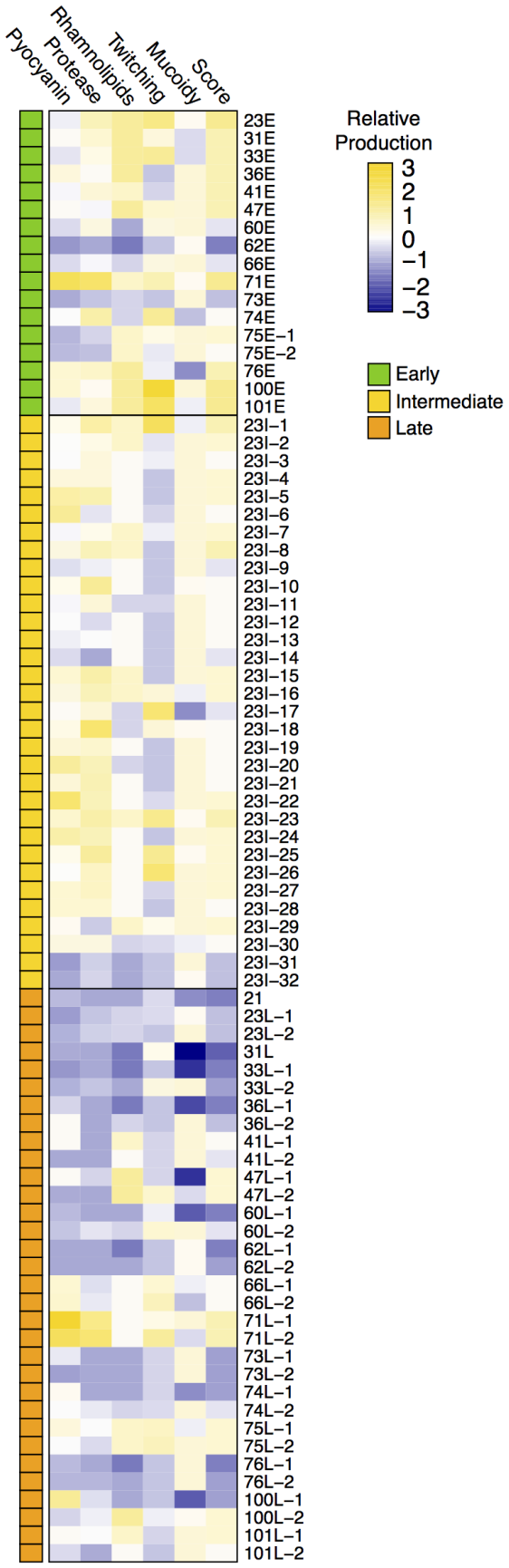
199

200 **RESULTS**

201 **Characteristics of *P. aeruginosa* CF isolates**

202 We obtained from a biorepository 81 *P. aeruginosa* chronic infection isolates, which had been
203 collected from 17 individuals with CF. From 16 patients, three *P. aeruginosa* isolates were
204 obtained: one isolate that was the first cultured *P. aeruginosa*, and two isolates that were
205 collected at least five years after the first. We refer to these isolates as Early and Late,
206 respectively. For one patient (Patient 75), four isolates were collected: two Early isolates

207 collected one month apart and two Late isolates, collected ten and 16 years after the first. For
208 one patient (Patient 23), 32 additional isolates were collected over the course of a 7.5 y infection
209 period, which we have termed Intermediate isolates. For one patient, only one isolate was
210 collected. All of these isolates were genetically characterized in a study by Smith *et al*., who
211 determined that the intra-patient replicate isolates are all clonally related, with the exception of
212 the four isolates from patient 75, which are actually two clonally-related Early/Late pairs
213 (isolates 75E-1 and 75L-2 are clonal, and isolates 75E-2 and 75L-1 are clonal) (3).
214 For all isolates, we characterized five clinically relevant phenotypes *in vitro*: mucoidy,
215 pyocyanin production, rhamnolipids production, protease production, and twitching motility (Fig.
216 1; Supplementary Table 4). These phenotypes are commonly altered during the course of
217 chronic CF lung infections (15). There were wide ranges in the expression of these phenotypes
218 across all isolates, as expected. Several isolate sets exhibited phenotypes consistent with age
219 of collection in that Early isolates possessed higher degrees of motility, higher quantities of
220 pyocyanin, rhamnolipids, and proteases (indicating intact quorum regulation), and lower
221 mucoidy compared to their cognate Late isolates (*e.g*., Patients 23, 31, 33, 36, and 76). The
222 associations between relative patient age at isolate collection and *P. aeruginosa* phenotypes
223 were not perfect, however. For example, Early isolate 62E had no detectable pyocyanin,
224 proteases, rhamnolipids, or twitching motility, and many of the Late isolates retained more early-
225 like phenotypes (e.g., both Late isolates of 41, 66, 71, 75, and 101, and 60L-1, 74L-2, and 100L-
226 2). Despite observable changes in isolate phenotype within patient, no phenotypes were
227 significantly different when comparing Late isolate(s) to the Early isolate. Collectively, the 81
228 isolates were highly varied, and represented the array of phenotypes we expect to observe in
229 the span of *P. aeruginosa* CF lung infections from initial infection to long-established chronic
230 infections.


Pyocyanin
Protease
Rhamnolipids
Twitching
Mucoidy
Score
23E
31E
33E
36E
41E
47E
60E
62E
66E
71E
73E
74E
75E−1
75E−2
76E
100E
101E
23I−1
23I−2
23I−3
23I−4
23I−5
23I−6
23I−7
23I−8
23I−9
23I−10
23I−11
23I−12
23I−13
23I−14
23I−15
23I−16
23I−17
23I−18
23I−19
23I−20
23I−21
23I−22
23I−23
23I−24
23I−25
23I−26
23I−27
23I−28
23I−29
23I−30
23I−31
23I−32
21
23L−1
23L−2
31L
33L−1
33L−2
36L−1
36L−2
41L−1
41L−2
47L−1
47L−2
60L−1
60L−2
62L−1
62L−2
66L−1
66L−2
71L−1
71L−2
73L−1
73L−2
74L−1
74L−2
75L−1
75L−2
76L−1
76L−2
100L−1
100L−2
101L−1
101L−2
Relative Production
3
2
1
0
−1
−2
−3
Early
Intermediate
Late

231

232 **FIG 1** Relative phenotype expression of 81 *P. aeruginosa* CF chronic lung infection isolates.
233 Score represents the sum of pyocyanin, proteases, rhamnolipids, twitching, and mucoidy. Each
234 phenotype was independently standardized.
235
236 **The *in vitro* volatile metabolome of *P. aeruginosa* CF isolates**
237 Taking an untargeted metabolomics approach, we cultured the clinical *P. aeruginosa* isolates *in*
238 *vitro* and characterized their volatile metabolomes using GC×GC-TOFMS. Following extensive
239 data processing, including removal of analytically biased chromatographic features, we
240 conservatively attribute 539 non-redundant volatile compounds to the growth and metabolism of
241 the *P. aeruginosa* isolates (Supplementary Table 5). Among these, 69 compounds are shared
242 by at least 95% ($n \geq 77$) of the isolates, representing core volatiles. Using minimum
243 metabolomics reporting standards, we assigned compound identification levels 1-4 (1 high, with
244 2 being the highest in this study) based on a combination of mass spectral and chromatographic
245 characteristics. We were able to assign putative names to 14 core and 33 non-core compounds
246 with an ID level 2, many of which have been previously reported as *P. aeruginosa* volatiles
247 (Table 1). While we detected 2-aminoacetophonone (the volatile compound responsible for the
248 grape-like odor of *P. aeruginosa*), it possessed an intraclass correlation slightly lower than the
249 applied threshold of 0.75 and was ultimately filtered out of our peak list. We also identified
250 several compounds of a variety of chemical classes that have not been previously reported for
251 *P. aeruginosa*, including: alcohols (3-methyl-1-buten-1-ol, 1-butoxy-2-propanol, 2-butyl-1-
252 octanol), esters (butanoic acids, methyl isovalerate), aromatics (tetrahydrofuran,
253 tetramethylpyrazine, 2-butylfuran, 3-cyano-2,5-dimethylpyrazine, 2,4-(1,1-dimethyl)-phenol),
254 unsaturated ketones (acetoin, 4-undecanone, 1-octen-3-one, 3-penten-2-one, 6-methyl-2-
255 heptanone, 1-(4-ethylphenyl)-ethanone), and an aldehyde (decanal).
256
257

258 **Table 1** Putatively identified volatile compounds metabolized by *P. aeruginosa* CF isolates.

| Compound Name† | Formula | Chemical Class | Samples (*n* = 81) |
|---|---|---|---|
| ***Core volatile compounds*** | | | |
| **3-methylbutanal** | $C_5H_{10}O$ | ALD | 79 |
| thiocyanic acid, methyl ester | $C_2H_3NS$ | EST | 77 |
| 3-methyl-1-butanol | $C_5H_{12}O$ | ALC | 77 |
| **pyridine** | $C_5H_5N$ | ARO | 77 |
| **3-penten-2-one** | $C_5H_8O$ | KET | 77 |
| **3-methyl-1-buten-1-ol** | $C_5H_{10}O$ | ALC | 78 |
| 4-methyl-3-penten-2-one | $C_6H_{10}O$ | KET | 81 |
| hexanal | $C_6H_{12}O$ | ALD | 81 |
| **heptanal** | $C_7H_{14}O$ | ALD | 80 |
| 2,5-dimethylpyrazine | $C_6H_8N_2$ | ARO | 81 |
| 2-nonanone | $C_9H_{18}O$ | KET | 81 |
| **4-ethyl-1,2-dimethylbenzene** | $C_{10}H_{14}$ | ARO | 77 |
| **1,3,3-trimethyl-(bicyclo)-heptan-2-ol** | $C_{10}H_{18}O$ | ALC | 78 |
| **1-(4-ethylphenyl)-ethanone** | $C_{10}H_{12}O$ | KET | 79 |
| ***Non-core volatile compounds*** | | | |
| dimethyl sulfide | $C_2H_6S$ | OTH | 29 |
| **2-methyl-3-buten-2-ol** | $C_5H_{10}O$ | ALC | 35 |
| 2-butanone | $C_4H_8O$ | KET | 1 |
| **tetrahydrofuran** | $C_4H_8O$ | ARO | 1 |
| acetic acid | $C_2H_4O_2$ | CA | 55 |
| methyl thiolacetate | $C_3H_6OS$ | THI | 3 |
| 2,4-dimethylfuran | $C_6H_8O$ | ARO | 66 |
| **butanoic acid, methyl ester** | $C_5H_{10}O_2$ | EST | 3 |
| **3-hydroxybutan-2-one** | $C_4H_8O_2$ | KET | 18 |
| dimethyl disulfide | $C_2H_2S_2$ | OTH | 1 |
| **2-methyl-butanoic acid, methyl ester** | $C_6H_{12}O_2$ | EST | 3 |
| **Compound Name†** | **Formula** | **Chemical Class** | **Samples (*n* = 81)** |
| **3-methyl-butanoic acid, methyl ester** | $C_6H_{12}O_2$ | EST | 3 |
| furfural | $C_5H_4O_2$ | ARO | 67 |
| **2-butylfuran** | $C_8H_{12}O$ | ARO | 6 |
| 2-hexen-1-ol | $C_6H_{12}O$ | ALC | 51 |
| **1-butoxy-2-propanol** | $C_7H_{16}O_2$ | ALC | 61 |

| | | | |
|---|---|---|---|
| **6-methyl-2-heptanone** | $C_8H_{16}O$ | KET | 33 |
| **1-methylethylbenzene** | $C_9H_{12}$ | ARO | 57 |
| 3-octanone | $C_8H_{16}O$ | KET | 56 |
| **4,6-dimethyl-2-heptanone** | $C_9H_{18}O$ | KET | 47 |
| **2-ethyl-5-methylpyrazine** | $C_7H_{10}N_2$ | ARO | 58 |
| 4-nonanone | $C_9H_{18}O$ | KET | 49 |
| **tetramethylpyrazine** | $C_8H_{12}N_2$ | ARO | 5 |
| 2-octenal | $C_8H_{14}O$ | ALD | 48 |
| octanenitrile | $C_8H_{15}N$ | OTH | 73 |
| 2,4-octadienal | $C_8H_{12}O$ | ALD | 9 |
| 3-decanone | $C_{10}H_{20}O$ | KET | 41 |
| **decanal** | $C_{10}H_{20}O$ | ALD | 65 |
| 2-decanone | $C_{10}H_{20}O$ | KET | 13 |
| **4-undecanone** | $C_{11}H_{22}O$ | KET | 30 |
| 2-dodecanone | $C_{12}H_{24}O$ | KET | 18 |
| **2-butyl-1-octanol** | $C_{12}H_{26}O$ | ALC | 48 |
| **2,4-bis(1,1-dimethyl)-phenol** | $C_{14}H_{22}O$ | ARO | 26 |

[†]Bolded compounds have not been previously reported as *P. aeruginosa* volatile compounds. ALC = alcohol, ALD = aldehyde, ARO = aromatic, CA = carboxylic acid, EST = ester, HC = hydrocarbon, KET = ketone, THI = thiol, OTH = other.

Chemical classifications were assigned to ID level 2 and 3 compounds (187 compounds) using mass spectral match and retention time characteristics (Fig. 2; Fig. S2). Among Level 2 and 3 core volatiles, the majority of classified compounds were hydrocarbons (41%), followed by ketones (16%). Other oxidized compounds, including alcohols, aromatics, acids, and thiols, accounted for an additional 43%. The non-core volatiles had similar chemical compositions, which are consistent with previous studies of the *P. aeruginosa* volatilome (25, 49-52). Of the entire volatilome, 65% of the volatiles ($n$ = 352) possessed less than an 80% mass spectral match to the 2011 NIST MS library, and as such were classified as unknowns (Fig. S2).

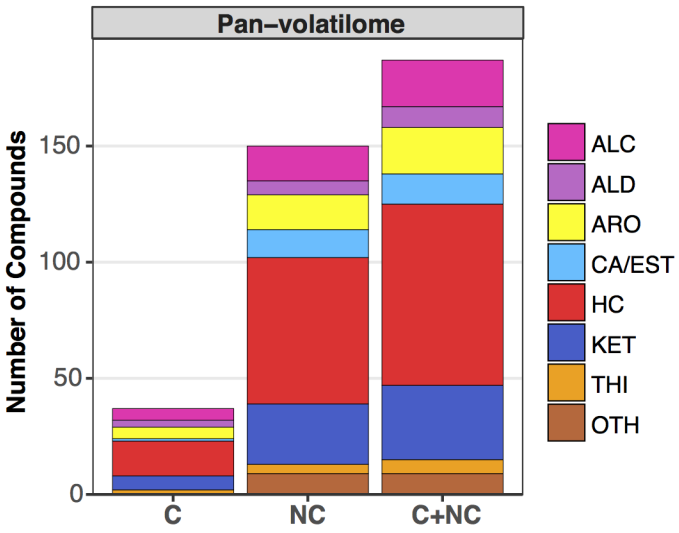
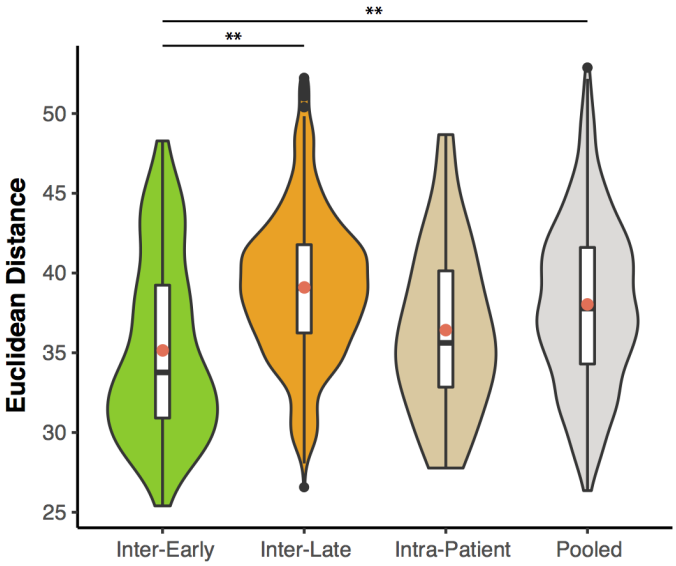



**FIG 2** Size and chemical composition of the volatilome produced by 81 *P. aeruginosa* chronic infection CF isolates, with unknown compounds excluded (Supplementary Figure 2 includes unknowns). C = Core compounds, detected in at least 95% of samples; NC = Non-core compounds; C+NC = sum of Core and Non-core compounds. ALC = alcohols, ALD = aldehydes, ARO = aromatics, CA/EST = carboxylic acids and esters, HC = hydrocarbons, KET = ketones, THI = thiols, OTH = other compounds.

**Relationships in the *P. aeruginosa* CF volatilome within and between patients**

Because every patient is infected by a genetically-unique strain, we hypothesized that the dissimilarities of intra-patient volatilomes would be lower than inter-patient. Additionally, due to the greater metabolic potential of early-infection isolates with intact regulatory networks, we hypothesized that the volatilomes of inter-patient Early isolates would be more dissimilar than those of inter-patient Late isolates. We calculated the pairwise Euclidean distances as a measure of dissimilarity between all early isolate volatilomes (“inter-Early dissimilarity”), all late isolate volatilomes (“inter-Late dissimilarity”), and between early and late isolate volatilomes from the same patient (“intra-patient dissimilarity”). Interestingly, the inter-Early isolate dissimilarity was the lowest (mean 35.2, median 33.8) and the inter-Late isolate dissimilarity was

289 the highest (mean 39.1, median 39.1), and these differences were significant ($P < 0.001$) (Fig.
290 3). Even more intriguing was that the inter-Early dissimilarity was lower than even the intra-
291 patient dissimilarity, though this difference was not significant ($P = 0.13$). For comparison, we
292 also calculated the pairwise Euclidean distances of all isolates in this study (Pooled; mean 38.0,
293 median 37.9); the mean inter-Early dissimilarity was also significantly lower than the mean
294 dissimilarity of the Pooled isolates ($P < 0.001$).



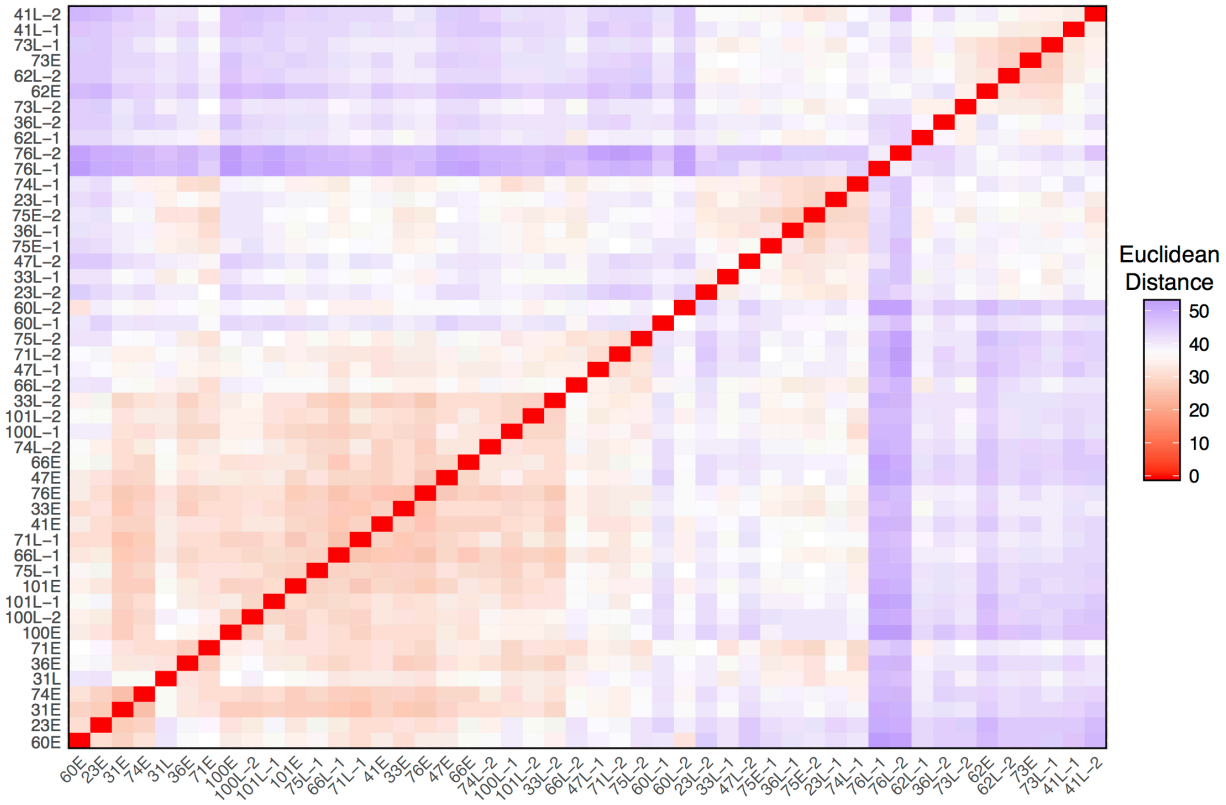

295

296 **FIG 3** Violin plots and box plots of pairwise dissimilarities (defined as Euclidean distance) of *P.*
297 *aeruginosa* chronic infection isolate volatilomes. Inter-Early are inter-patient Early-Early isolate
298 dissimilarities. Inter-Late are inter-patient Late-Late isolate dissimilarities. Intra-Patient are intra-
299 patient Early-Late isolate dissimilarities. Pooled represents all pairwise volatilome dissimilarities.
300 Red circles indicate the mean.

301

302 Visualizing the Euclidean distances between isolate volatilomes via non-metric
303 multidimensional scaling (NMDS) and ordered dissimilarity images (ODI), we looked for
304 clustering of clonal isolates, indicating intra-patient similarities. The NMDS and ODI of the
305 Euclidean distances of all 81 isolate volatilomes show a primary cluster comprised mostly of
306 intermediate isolates of Patient 23 (Fig. S3, Fig. S4). To reduce the influence of the over-

307 represented Patient 23 isolates in data visualization, we truncated our data set to 48 isolates by
308 removing the Intermediate isolates of Patient 23. The ODI plot and NMDS of the truncated set
309 revealed three clusters of similarity, none of which include all isolates from a single patient (Fig.
310 4; Fig. S5a). This refutes our hypothesis that isolates from the same patient maintain a similar
311 volatilome over time. To determine if any inter-patient or intra-patient isolate pairs are
312 significantly different, we performed permutational multivariate analysis of variance
313 (PERMANOVA) on the Euclidean distances for all possible pairs. Following Benjamini-Hochberg
314 adjustment of *P* values, no pairs of isolates were identified as having significantly different
315 volatilomes ($P_{adj} \geq 0.1$ for all isolate pairs). However, PERMANOVA conducted on all isolates
316 indicated strong significant differences between the 81 volatilomes ($pseudo\text{-}F_{80,161} = 28.0$, $P <$
317 $0.01$), as well as between the volatilomes of the truncated set of 49 isolates ($pseudo\text{-}F_{47,96} =$
318 $29.2$, $P < 0.001$), reinforcing the inference from the ODI and NMDS that the CF *P. aeruginosa*
319 volatilome is heterogeneous.




320

321 **FIG 4** Ordered dissimilarity image (ODI) of the truncated set of 48 *P. aeruginosa* clinical isolates
322 depicting volatilome dissimilarity defined by Euclidean distance.
323
324 Though we do not find that patient isolates maintain similar volatilomes over the duration
325 of chronic infection, it is noteworthy that the majority of Patient 23's isolates have strong
326 similarities (Fig. S3, Fig. S4). Taking a closer look at the 35 isolates from Patient 23 (Fig. 5; Fig.
327 S6), the ODI and NMDS revealed four distinct similarity neighborhoods: one neighborhood
328 consisting of only the Early isolate 23E, which stands alone; a second consisting of the four
329 latest-collected isolates 23I-32, 23I-33, 23L-1, 23L-2; a third consisting of isolates 23I-7, 23I-9,
330 23I-12, 23I-14, 23I-29; and a fourth encompassing the remainder of the isolates. The four latest-
331 collected isolates are less defined by their similarity to each other than their dissimilarity to all
332 other isolates collected during the infection. A linear regression of the Euclidean distances
333 between the Early isolate and the Intermediate and Late isolates as a function of patient age
334 showed that isolates became increasingly dissimilar to the Early isolate over time (Pearson's r =
335 0.70, $P < 0.001$) (Fig. 6). These observations are underpinned by the phenotypic differences of
336 the four latest isolates, and the number and types of known mutations that were accumulated in
337 the 23L versus 23E isolates (3). As described by Smith *et al*., the *P. aeruginosa* infection
338 sampled from Patient 23 (referred to as Patient 1 in the referenced publication) diversified from
339 patient ages 1.5 to 3 y into a population of isolates, with gene mutations in virulence, motility,
340 quorum sensing, iron transport, efflux, and transcription and translation genes. The volatilome
341 dissimilarities we measured reflect the genotypic diversification of the infection as a function of
342 time (Fig. 6).

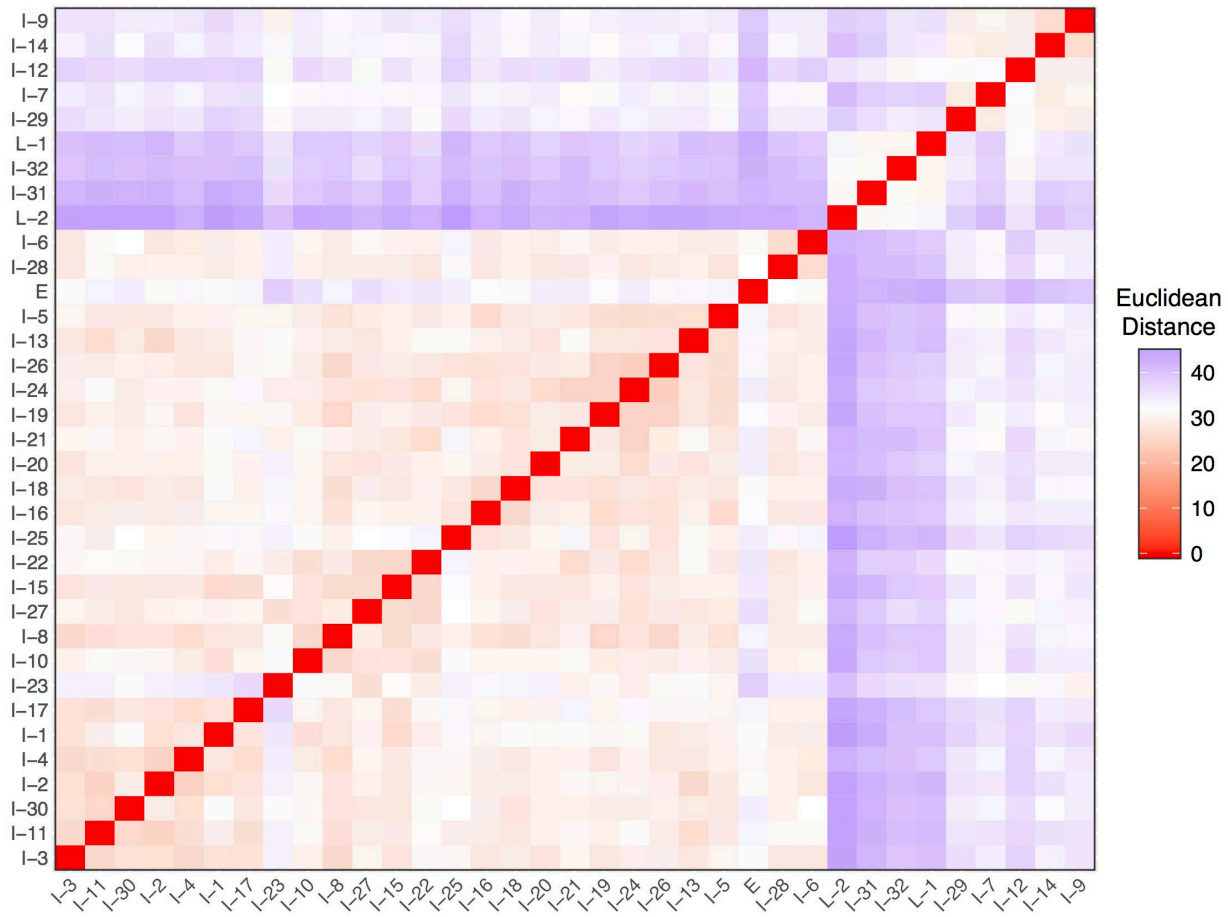



**FIG 5** Ordered dissimilarity image (ODI) of the 35 *P. aeruginosa* clinical isolates from Patient 23 depicting volatilome dissimilarity defined by Euclidean distance. For visual clarity, the isolate labels do not include patient number (*e.g.*, 23)

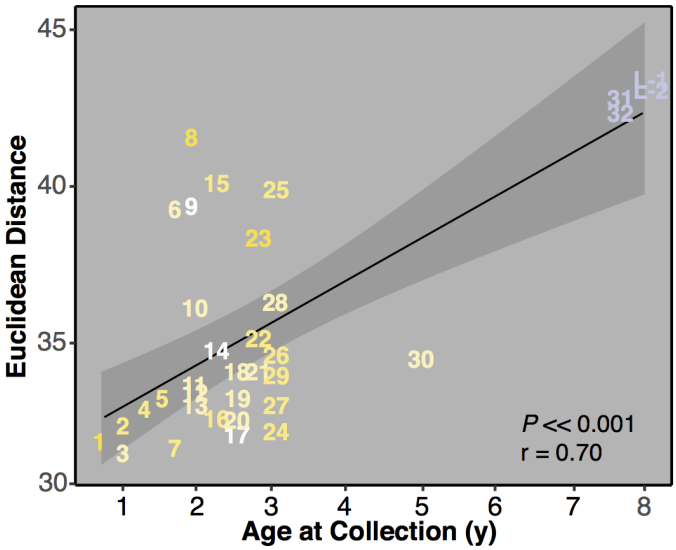



**FIG 6** Dissimilarity of Patient 23 volatilomes over time, defined as Euclidian distance from the Early isolate (not shown). For visual clarity, the isolate labels do not include patient number (*e.g.*, 23) and intermediate isolate labels have been truncated to only include the numeral that signifies collection order (*e.g.*, I-1 is represented as 1). The shade of the isolate labels depict the relative phenotype Score (Fig. 1). The black line represents the linear regression fit, and the shaded region represents the standard error of the regression line.

**Chemical characteristics of the volatilomes of Early and Late isolates**

We compared the chemical characteristics of the Early and Late volatile metabolomes, positing that the size and chemical diversity of the volatilomes would decrease from Early to Late infection due to loss-of-function mutations. To perform the comparison, we needed to balance the size of the isolate groups, and did so by selecting ten pairs of *P. aeruginosa* isolates that had large changes in phenotypes over the duration of infection, thereby enhancing the differences we may find in the Early vs. Late volatilomes (Fig. 1; Supplementary Table 4). Visual inspection of the ten pairs of GC×GC chromatograms suggests reductions in the number and variety of volatile compounds produced by Late isolates (Fig. 7, Fig. S7). Contrary to the appearance of the chromatograms, however, the overall size and chemical compositions of the

365 Early and Late volatilomes were similar to each other (Fig. S8), with hydrocarbons representing
366 approximately 50% and alcohols and ketones together representing approximately 30% of the
367 410 Early and 441 Late volatiles. We quantified the chemical richness and diversity of the
368 pooled Early and pooled Late isolates using the Shannon-Wiener diversity index
369 (Supplementary Table 6) and found the volatilomes of Early and Late isolates were similar,
370 whether using all volatiles or only the non-core volatiles.

371 The uniformity in the Early versus Late volatilomes is not a function of data aggregation,
372 as the volatilomes of the individual isolates are highly similar in number and chemical
373 composition, with the exception of one Late isolate (76L) that has a slow-growth phenotype (Fig.
374 S9). Furthermore, when looking at individual isolate pairs, for the vast majority there were no
375 differences in the richness and diversity of their volatilomes (Supplementary Table 6). Together,
376 these results suggest that there are not major differences between the number and diversity of
377 chemical compounds when comparing Early and Late infection isolates. Hierarchical clustering
378 analysis (HCA) of the *P. aeruginosa* isolates by their volatilomes underscores this observation;
379 we observed no discernible clustering of the Early versus Late isolates based on the presence
380 and absence of volatiles (Fig. S10).

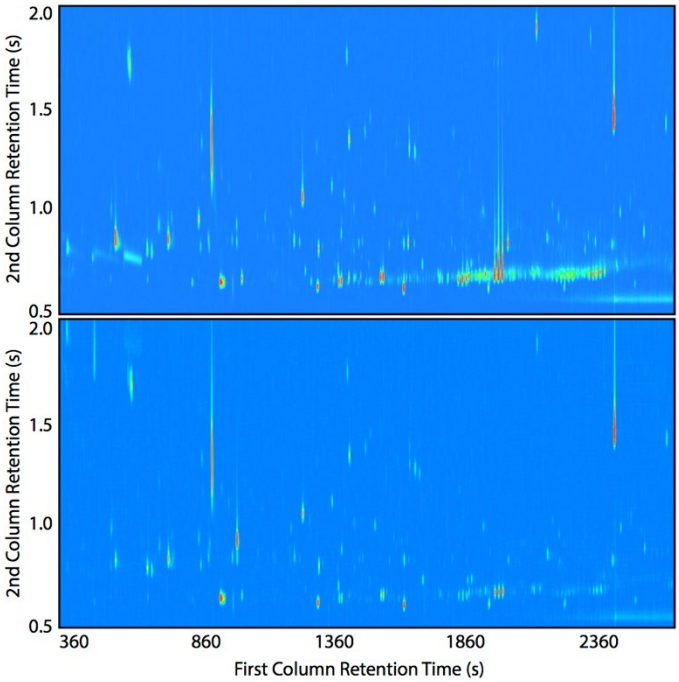



**FIG 7** Representative GC×GC chromatograms of Early (top) and Late (bottom) chronic infection isolates (Patient 41). Dark blue represents the baseline, and peak intensity is depicted using a color gradient from light blue (low) to dark red (high). Chromatographic regions of ${}^{1}t_R$ < 200 s and ${}^{2}t_R$ < 0.5 s were excluded for visual clarity. All chromatograms for the ten selected Early and Late isolate pairs are provided in Fig. S7.

**Abundance of volatile compounds in Early and Late isolates**

Though we did not observe any differences in the Early and Late isolate volatilomes based on presence and absence of metabolites, the chromatograms suggest a reduced volatilome in late infection. We therefore posited that while the numbers of volatiles do not significantly change from Early to Late infection isolates, the relative concentrations of volatiles do. Using volatile abundances, we performed HCA on the 20 selected Early and Late isolates and found modest clustering by the time of isolate collection (Fig. S11). We were interested in whether or not clustering still occurs in the larger, less curated set of 48 isolates (*i.e.*, the truncated set), which has more discordance between phenotype and time of collection (Fig. 1). HCA of the truncated

397 set of isolates shows that Early isolates are generally more similar to one another than Late
398 isolates are (Fig. 8), reinforcing observations made by calculating Euclidean distances;
399 however, significant proportions of the Early and the Late isolates are mis-clustered.
400 Interestingly, the majority of the Late isolates clustering with the Early isolates in the upper clade
401 have more early-like phenotypes, and several Early isolates that have more late-like phenotypes
402 (*e.g.*, 62E, 73E, 70E) cluster in the lower clade with Late isolates, indicating a relationship
403 between the isolate phenomes and volatilomes. We also observed clustering by phenotype
404 when the volatilomes of all 81 isolates were analyzed (Fig. S12), and this was seen in the set of
405 ten paired isolates as well. Similar phenomena are observed when clustering the truncated set
406 or all 81 isolates by the core volatilome (Fig. S13, Fig. S14), and even when using only the 23
407 core alcohols, aldehydes, and hydrocarbons (Fig. 9 and Fig. S15), suggesting that a set of
408 conserved volatiles could be identified and used as biomarkers for detecting phenotypic
409 changes in chronic infections.

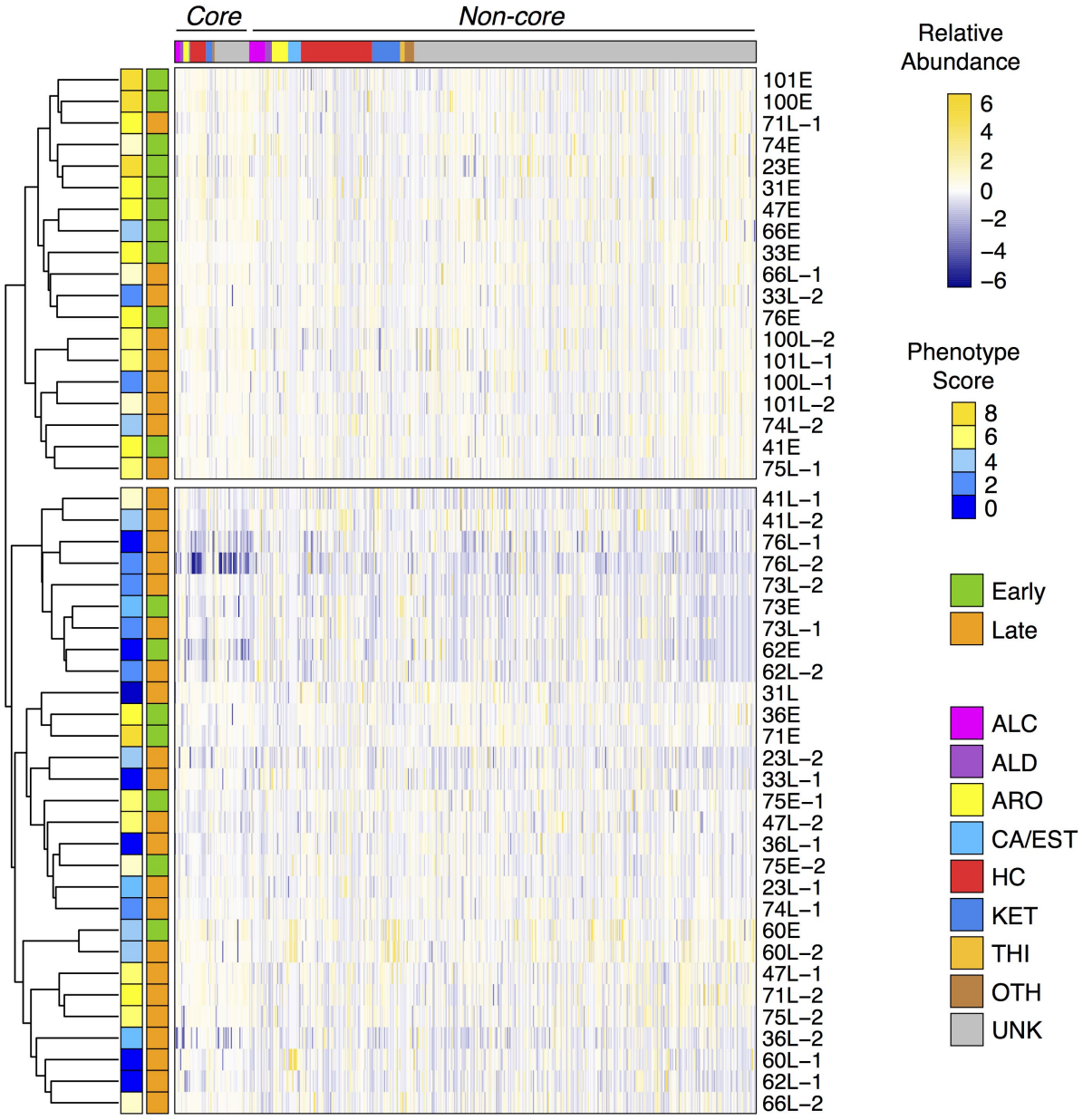



**FIG 8** Hierarchical clustering analysis (HCA) of the truncated set of 48 *P. aeruginosa* clinical CF isolates, based on the relative abundance of 539 volatile compounds. Volatiles are in columns (standardized relative abundance). Clustering is based on rows (isolates), which are color-coded by their phenotype score (left color block) and relative time of collection (right color block). ALC = alcohols, ALD = aldehydes, ARO = aromatics, CA/EST = carboxylic acids and esters, HC = hydrocarbons, KET = ketones, THI = thiols, OTH = other compounds, UNK = unknown identity.

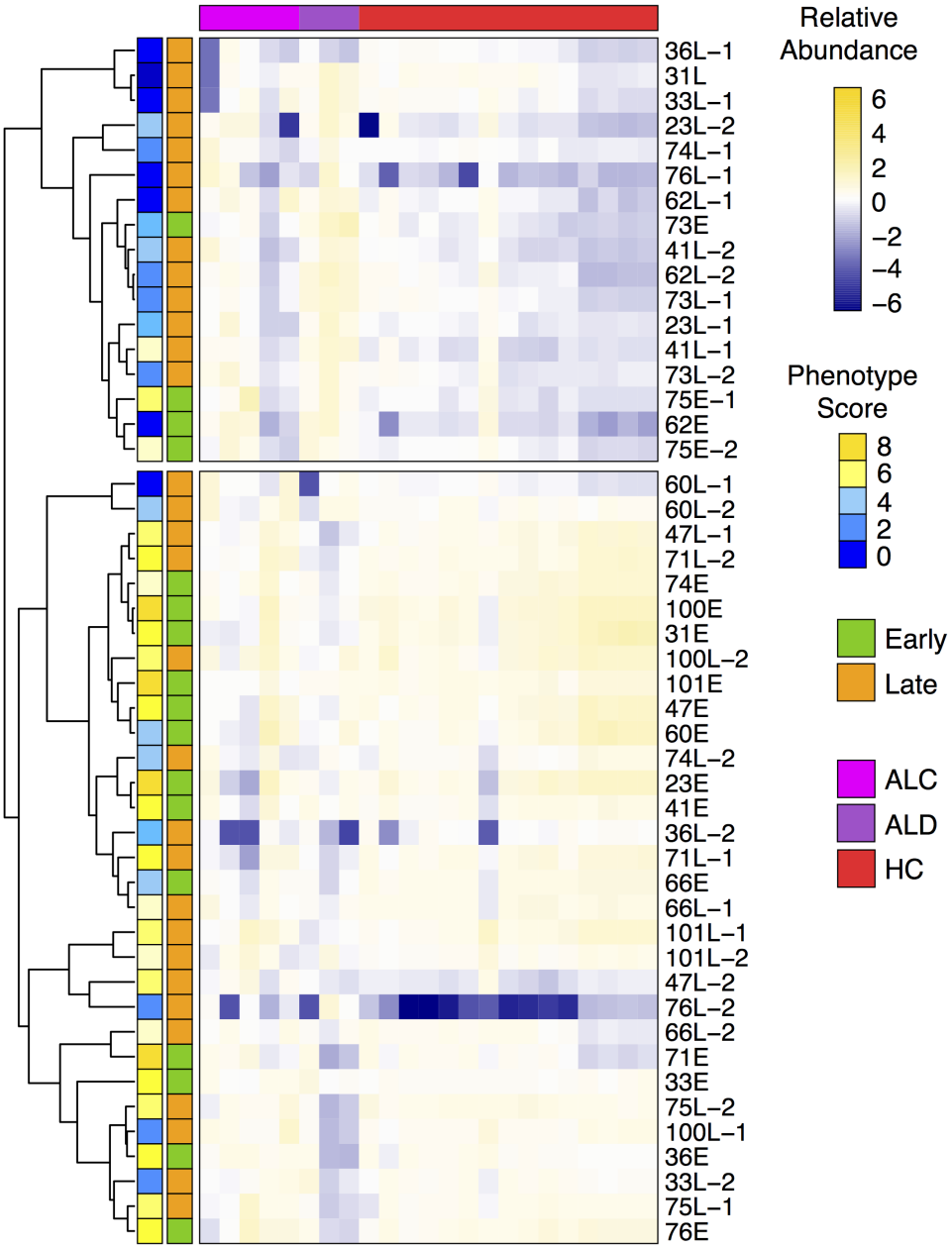



**FIG 9** Hierarchical clustering analysis (HCA) of the truncated set of 48 *P. aeruginosa* clinical CF isolates, based on the relative abundance of 23 core alcohols, aldehydes, and hydrocarbons. Volatiles are in columns (standardized relative abundance). Clustering is based on rows (isolates), which are color-coded by their phenotype score (left color block) and relative time of collection (right color block). ALC = alcohols, ALD = aldehydes, HC = hydrocarbons.

425 **DISCUSSION**

426 *P. aeruginosa* isolates from cystic fibrosis lung infections have a large and diverse
427 volatilome. We conservatively attribute 539 detected volatiles to *P. aeruginosa* growth and
428 metabolism, only 13% of which were core volatiles. For volatiles that we could assign to a
429 chemical class, hydrocarbons represented the highest proportion, followed by ketones, which
430 we observed both over the pan-volatilome and for individual patient isolates. The chemical
431 compositions are similar to previous volatilome studies of *P. aeruginosa* (49-52), and many of
432 the volatiles we detected have been frequently reported in other studies (*e.g.*, 2-butanone, 2-
433 nonanone, dimethyl sulfide, dimethyl disulfide). Only just over 40% of the core volatiles have
434 been identified in other *P. aeruginosa* metabolomic studies, suggesting that a CF-specific *P.*
435 *aeruginosa* volatilome is unlikely. However, due to the inter-strain variation in the *P. aeruginosa*
436 volatilome, many more CF and non-CF isolates will need to be characterized to test this
437 hypothesis.

438 Despite the trend toward reduced or loss-of-phenotype expression in late infection
439 isolates, which we hypothesized would correspond to a reduced metabolome size, we observed
440 that Late isolates produce the same number and variety of volatile compounds as Early isolates.
441 However, the relative abundance of volatiles produced by the Late isolates is significantly
442 reduced. Our data suggest that mutations that arise in *P. aeruginosa* chronic infection isolates
443 do not result in complete inhibition of metabolic pathways, but may instead reduce metabolic
444 flux. This observation has implications for diagnostics, signifying that presence and absence of
445 *P. aeruginosa* volatile metabolites would provide little diagnostic value for tracking chronic lung
446 infections. Rather, a breath-based diagnostic for monitoring *P. aeruginosa* adaptations in CF will
447 likely need to measure the relative abundance of metabolites, particularly in relation to a
448 baseline. Importantly, a subset of core volatiles can be used to cluster isolates based on
449 commonalities in phenotypes, indicating that a diagnostic based upon the relative abundances

450 of conserved hydrocarbons, alcohols, and aldehydes (or a few selected volatiles from these
451 classes) may be sufficient to track *P. aeruginosa* infection in most patients.
452 Based on the genomic heterogeneity of *P. aeruginosa* (3, 33-35), we hypothesized that
453 every patient would have a unique *P. aeruginosa* volatilome, and that intra-patient isolates
454 would have more similar volatilomes than inter-patient isolates. To the contrary, we observed
455 that the volatilomes of isolates from a single patient are no more similar to one another than
456 they are to any other isolate (intra-patient vs. pooled data, Fig. 3). Interestingly, the volatilomes
457 of inter-patient Early isolates are even more similar than intra-patient isolates, while the
458 volatilomes of inter-patient Late isolates have the highest dissimilarity. These data indicate that
459 the genomic similarity within patient is less influential on the volatilome than the post-
460 transcriptional diversity that evolves within a chronic *P. aeruginosa* infection population. The
461 data from Patient 23, however, suggest that intra-patient isolates are more likely to have similar
462 volatilomes when there are fewer mutational and phenotypic differences between them. From
463 these results, we propose that a breath-based diagnostic to track *P. aeruginosa* chronic lung
464 infections would be personalized, with each patient's breathprint at initial infection serving as a
465 baseline to which subsequent breathprints are compared. Sustained, significant deviations from
466 the baseline breathprint would indicate that the *P. aeruginosa* population phenotype has
467 changed, and therefore the patient may be at higher risk for associated poor clinical outcomes
468 (*e.g.*, higher risk of exacerbation) (15) and at higher risk of infection eradication failure after
469 antibiotic treatment (12). In a longitudinal case study of a single CF patient, we have shown that
470 changes in the dominant species colonizing the lung correlate to changes in the volatile
471 metabolites detected in sputum (53), and therefore monitoring for breathprint deviations may
472 also work for *P. aeruginosa* infections. Despite this, the issue of specificity becomes a concern
473 unless biomarkers for individual *P. aeruginosa* phenotypes are identified and used for tracking;
474 this is the subject of ongoing work.

475 Current methods for tracking *P. aeruginosa* lung infection progression are predominantly
476 culture-based (16, 21), and our results suggest a potential role for *in vitro* volatile metabolomics
477 in staging chronic infections in conjunction with established clinical laboratory methods. The
478 clinically-correlated *P. aeruginosa* phenotypes we measured in this study – mucoidy, motility,
479 and quorum-regulated traits – are sensitive to environment (54-57); as such, the metabolomes
480 we characterized from *in vitro* cultures are unlikely to directly reflect the metabolomes these
481 isolates produced in the CF lung. For the development of a breath test for tracking infections *in*
482 *situ*, it will be necessary to utilize culture models that are more reflective of the lung (58-61) and
483 to collect *ex vivo* and *in vivo* infection volatilomes (25, 62, 63). The utility of breath for tracking
484 chronic lung infections is being demonstrated in tuberculosis lung infection models (26), and
485 therefore we anticipate that the development of a breath-based diagnostic for tracking *P.*
486 *aeruginosa* chronic lung infections is feasible. Despite the limitations of *in vitro* analyses, this
487 study presents the first comprehensive analysis of the *P. aeruginosa* volatilome from chronic CF
488 lung infections, and lays the groundwork for the application of volatile metabolomics in tracking
489 CF lung disease.
490

491 **ACKNOWLEDGEMENTS**

492 Funding to support this work was provided by the Cystic Fibrosis Foundation (Hill17P0,
493 Hill18A0-CI), the National Institutes of Health (R56HL139846), and the ASU School of Life
494 Sciences Undergraduate Research (SOLUR) program. Isolates were obtained from the Cystic
495 Fibrosis Isolate Core at Seattle Children's Research Institute, funded by the National Institutes
496 of Health (P30 DK089507) and the Cystic Fibrosis Foundation (HOFFMA20Y2-OUT).
497

498 **AUTHOR CONTRIBUTIONS**

499 HB conceived of the study; HB, TD designed the experiments; AK, CB, SW, and TD collected
500 phenotype data; TD collected metabolomics data; TD and MR processed and analyzed the

501 metabolomics data; HB and TD wrote and edited the manuscript; all authors approved the final
502 version.
503
504 **CONFLICTS OF INTEREST**
505 The authors declare that the research was conducted in the absence of any commercial or
506 financial relationships that could be construed as a potential conflict of interest.
507
508 **REFERENCES**


509 1. Gibson RL, Burns JL, Ramsey BW. 2003. Pathophysiology and management of
510 pulmonary infections in cystic fibrosis. Am J Respir Crit Care Med 168:918-951.

511 2. Marshall B, Faro A, Fink A, Loeffler D, Elbert A, O'Neil T, Rush T, Rizvi S. 2018. Cystic
512 Fibrosis Foundation patient registry: 2017 annual data report. Cystic Fibrosis
513 Foundation, Foundation CF, Bethesda, MD.

514 3. Smith EE, Buckley DG, Wu ZN, Saenphimmachak C, Hoffman LR, D'Argenio DA, Miller
515 SI, Ramsey BW, Speert DP, Moskowitz SM, Burns JL, Kaul R, Olson MV. 2006. Genetic
516 adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. Proc
517 Natl Acad Sci U S A 103:8487-8492.

518 4. Li ZH, Kosorok MR, Farrell PM, Laxova A, West SEH, Green CG, Collins J, Rock MJ,
519 Splaingard ML. 2005. Longitudinal development of mucoid *Pseudomonas aeruginosa*
520 infection and lung disease progression in children with cystic fibrosis. JAMA 293:581-
521 588.

522 5. Emerson J, Rosenfeld M, McNamara S, Ramsey B, Gibson RL. 2002. *Pseudomonas*
523 *aeruginosa* and other predictors of mortality and morbidity in young children with cystic
524 fibrosis. Pediatr Pulmonol 34:91-100.

525 6. Kosorok MR, Zeng L, West SEH, Rock MJ, Splaingard ML, Laxova A, Green CG, Collins
526 J, Farrell PM. 2001. Acceleration of lung disease in children with cystic fibrosis after
527 *Pseudomonas aeruginosa* acquisition. Pediatr Pulmonol 32:277-287.

528 7. Davidson AG, Chilvers MA, Lillquist YP. 2012. Effects of a *Pseudomonas aeruginosa*
529 eradication policy in a cystic fibrosis clinic. Curr Opin Pulm Med 18:615-21.

530 8. Winstanley C, O'Brien S, Brockhurst MA. 2016. *Pseudomonas aeruginosa* evolutionary
531 adaptation and diversification in cystic fibrosis chronic lung infections. Trends Microbiol
532 24:327-337.

533 9. Hogardt M, Heesemann J. 2013. Microevolution of *Pseudomonas aeruginosa* to a
534 chronic pathogen of the cystic fibrosis lung, p 91-118. *In* Dobrindt U, Hacker JH,
535 Svanborg C (ed), Between pathogenicity and commensalism, vol 358. Springer-Verlag
536 Berlin, Berlin.

537 10. Sanders DB. 2014. Phenotypes that matter: *Pseudomonas aeruginosa* and progression
538 of cystic fibrosis lung disease. Am J Respir Crit Care Med 190:245-246.

539 11. Pritt B, O'Brien L, Winn W. 2007. Mucoid *Pseudomonas* in cystic fibrosis. Am J Clin
540 Pathol 128:32-34.

541 12. Mayer-Hamblett N, Ramsey BW, Kulasekara HD, Wolter DJ, Houston LS, Pope CE,
542 Kulasekara BR, Armbruster CR, Burns JL, Retsch-Bogart G, Rosenfeld M, Gibson RL,
543 Miller SI, Khan U, Hoffman LR. 2014. *Pseudomonas aeruginosa* phenotypes associated
544 with eradication failure in children with cystic fibrosis. Clin Infect Dis 59:624-631.

545 13. Courtney JM, Bradley J, McCaughan J, O'Connor TM, Shortt C, Bredin CP, Bradbury I,
546 Elborn JS. 2007. Predictors of mortality in adults with cystic fibrosis. Pediatr Pulmonol
547 42:525-532.

548 14. Hauser AR, Jain M, Bar-Meir M, McColley SA. 2011. Clinical significance of microbial
549 infection and adaptation in cystic fibrosis. Clin Microbiol Rev 24:29-70.

550 15. Mayer-Hamblett N, Rosenfeld M, Gibson RL. 2014. *Pseudomonas aeruginosa in vitro*
551 phenotypes distinguish cystic fibrosis infection stages and outcomes. Am J Respir Crit
552 Care Med 190:289-297.

553 16. Héry-Arnaud G, Nowak E, Caillon J, David V, Dirou A, Revert K, Munck MR, Frachon I,
554 Haloun A, Horeau-Langlard D, Le Bihan J, Danner-Boucher I, Ramel S, Pelletier MP,
555 Rosec S, Gouriou S, Poulhazan E, Payan C, Ferec C, Rault G, Le Gal G, Le Berre R.
556 2017. Evaluation of quantitative PCR for early diagnosis of *Pseudomonas aeruginosa*
557 infection in cystic fibrosis: A prospective cohort study. Clin Microbiol Infect 23:203-207.

558 17. Kim K-H, Jahan SA, Kabir E. 2012. A review of breath analysis for diagnosis of human
559 health. Trends Anal Chem 33:1-8.

560 18. Ahmed WM, Lawal O, Nilsen TM, Goodacre R, Fowler SJ. 2017. Exhaled volatile
561 organic compounds of infection: A systematic review. ACS Infect Dis 3:695-710.

562 19. Oakley-Girvan I, Davis SW. 2018. Breath based volatile organic compounds in the
563 detection of breast, lung, and colorectal cancers: A systematic review. Cancer Biomark
564 21:29-39.

565 20. Amann A, Miekisch W, Schubert J, Buszewski B, Ligor T, Jezierski T, Pleil J, Risby T.
566 2014. Analysis of exhaled breath for disease detection. Annu Rev Anal Chem, Vol 7
567 7:455-482.

568 21. Smith WD, Bardin E, Cameron L, Edmondson CL, Farrant KV, Martin I, Murphy RA,
569 Soren O, Turnbull AR, Wierre-Gore N, Alton EW, Bundy JG, Bush A, Connett GJ, Faust
570 SN, Filloux A, Freemont PS, Jones AL, Takats Z, Webb JS, Williams HD, Davies JC.
571 2017. Current and future therapies for *Pseudomonas aeruginosa* infection in patients
572 with cystic fibrosis. FEMS Microbiol Lett 364.

573 22. Burns JL, Gibson RL, McNamara S, Yim D, Emerson J, Rosenfeld M, Hiatt P, McCcoy
574 K, Castile R, Smith AL, Ramsey BW. 2001. Longitudinal assessment of *Pseudomonas*
575 *aeruginosa* in young children with cystic fibrosis. J Infect Dis 183:444-452.

576 23. Hoppe JE, Towler EE, Wagner BD, Accurso FJ, Sagel SD, Zemanick ET. 2015. Sputum
577 induction improves detection of pathogens in children with cystic fibrosis. Pediatr
578 Pulmonol 50:638-646.

579 24. West SEH, Zeng L, Lee BL, Kosorok MR, Laxova A, Rock MJ, Splaingard MJ, Farrell
580 PM. 2002. Respiratory infections with *Pseudomonas aeruginosa* in children with cystic
581 fibrosis - early detection by serology and assessment of risk factors. JAMA 287:2958-
582 2967.

583 25. Nasir M, Bean HD, Smolinska A, Rees CA, Zemanick ET, Hill JE. 2018. Volatile
584 molecules from bronchoalveolar lavage fluid can 'rule-in' *Pseudomonas aeruginosa* and
585 'rule-out' *Staphylococcus aureus* infections in cystic fibrosis patients. Sci Rep 8.

586 26. Beccaria M, Bobak C, Maitshotlo B, Mellors TR, Purcaro G, Franchina FA, Rees CA,
587 Nasir M, Black A, Hill JE. 2019. Exhaled human breath analysis in active pulmonary

588 tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and
589 chemometric techniques. J Breath Res 13.

590 27. Freschi L, Vincent AT, Jeukens J, Emond-Rheault JG, Kukavica-Ibrulj I, Dupont MJ,
591 Charette SJ, Boyle B, Levesque RC. 2019. The *Pseudomonas aeruginosa* pan-genome
592 provides new insights on its population structure, horizontal gene transfer, and
593 pathogenicity. Genome Biol Evol 11:109-120.

594 28. Bean HD, Rees CA, Hill JE. 2016. Comparative analysis of the volatile metabolomes of
595 *Pseudomonas aeruginosa* clinical isolates. J Breath Res 10:047102.

596 29. Kelly NM, Fitzgerald MX, Tempany E, O'Boyle C, Falkiner FR, Keane CT. 1982. Does
597 *Pseudomonas* cross-infection occur between cystic-fibrosis patients. Lancet 2:688-90.

598 30. Grothues D, Koopmann U, von der Hardt H, Tummler B. 1988. Genome fingerprinting of
599 *Pseudomonas aeruginosa* indicates colonization of cystic fibrosis siblings with closely
600 related strains. J Clin Microbiol 26:1973-7.

601 31. Thomassen MJ, Demko CA, Doershuk CF, Root JM. 1985. *Pseudomonas aeruginosa*
602 isolates: Comparisons of isolates from campers and from sibling pairs with cystic
603 fibrosis. Pediatr Pulmonol 1:40-5.

604 32. Wolz C, Kiosz G, Ogle JW, Vasil ML, Schaad U, Botzenhart K, Doring G. 1989.
605 *Pseudomonas aeruginosa* cross-colonization and persistence in patients with cystic
606 fibrosis. Use of a DNA probe. Epidemiol Infect 102:205-14.

607 33. Ashish A, Paterson S, Mowat E, Fothergill JL, Walshaw MJ, Winstanley C. 2013.
608 Extensive diversification is a common feature of *Pseudomonas aeruginosa* populations
609 during respiratory infections in cystic fibrosis. J Cyst Fibros 12:790-3.

610 34. Cramer N, Klockgether J, Wrasman K, Schmidt M, Davenport CF, Tummler B. 2011.
611 Microevolution of the major common *Pseudomonas aeruginosa* clones c and pa14 in
612 cystic fibrosis lungs. Environ Microbiol 13:1690-704.

613 35. Mowat E, Paterson S, Fothergill JL, Wright EA, Ledson MJ, Walshaw MJ, Brockhurst
614 MA, Winstanley C. 2011. *Pseudomonas aeruginosa* population diversity and turnover in
615 cystic fibrosis chronic infections. Am J Respir Crit Care Med 183:1674-1679.

616 36. Bragonzi A, Paroni M, Nonis A, Cramer N, Montanari S, Rejman J, Di Serio C, Doring G,
617 Tummler B. 2009. *Pseudomonas aeruginosa* microevolution during cystic fibrosis lung
618 infection establishes clones with adapted virulence. Am J Respir Crit Care Med 180:138-
619 45.

620 37. Kordes A, Preusse M, Willger SD, Braubach P, Jonigk D, Haverich A, Warnecke G,
621 Haussler S. 2019. Genetically diverse *Pseudomonas aeruginosa* populations display
622 similar transcriptomic profiles in a cystic fibrosis explanted lung. Nat Commun 10:3397.

623 38. Clark ST, Guttman DS, Hwang DM. 2018. Diversification of *Pseudomonas aeruginosa*
624 within the cystic fibrosis lung and its effects on antibiotic resistance. FEMS Microbiol Lett
625 365.

626 39. Schick A, Kassen R. 2018. Rapid diversification of *Pseudomonas aeruginosa* in cystic
627 fibrosis lung-like conditions. Proc Natl Acad Sci U S A 115:10714-10719.

628 40. Wilder CN, Allada G, Schuster M. 2009. Instantaneous within-patient diversity of
629 *Pseudomonas aeruginosa* quorum-sensing populations from cystic fibrosis lung
630 infections. Infect Immun 77:5631-9.

631 41. Smania AM, Segura I, Pezza RJ, Becerra C, Albesa I, Argarana CE. 2004. Emergence
632 of phenotypic variants upon mismatch repair disruption in *Pseudomonas aeruginosa*.
633 Microbiology 150:1327-1338.

634 42. Ciofu O, Mandsberg LF, Bjarnsholt T, Wassermann T, Hoiby N. 2010. Genetic
635 adaptation of *Pseudomonas aeruginosa* during chronic lung infection of patients with
636 cystic fibrosis: Strong and weak mutators with heterogeneous genetic backgrounds
637 emerge in muca and/or lasr mutants. Microbiology 156:1108-1119.

638 43. Higgins Keppler EA, Jenkins CL, Davis TJ, Bean HD. 2018. Advances in the application
639 of comprehensive two-dimensional gas chromatography in metabolomics. Trends Anal
640 Chem 109:275-286.

641 44. Dieterle F, Ross A, Schlotterbeck G, Senn H. 2006. Probabilistic quotient normalization
642 as robust method to account for dilution of complex biological mixtures. Application in $^{1}$H
643 NMR metabonomics. Anal Chem 78:4281-4290.

644 45. Koo TK, Li MY. 2016. A guideline of selecting and reporting intraclass correlation
645 coefficients for reliability research. J Chiropr Med 15:155-163.

646 46. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and
647 powerful approach to multiple testing. J R Stat Soc Series B Methodol 57:289-300.

648 47. Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression
649 data using empirical bayes methods. Biostatistics 8:118-27.

650 48. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O,
651 Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN,
652 Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. 2007. Proposed

653 minimum reporting standards for chemical analysis Chemical Analysis Working Group
654 (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3:211-221.

655 49. Bean HD, Rees CA, Hill JE. 2016. Comparative analysis of the volatile metabolomes of
656 *Pseudomonas aeruginosa* clinical isolates. J Breath Res 10.

657 50. Bean HD, Dimandja J-MD, Hill JE. 2012. Bacterial volatile discovery using solid-phase
658 microextraction and comprehensive two-dimensional gas chromatography–time-of-flight
659 mass spectrometry. J Chromatogr B 901:41-46.

660 51. Shestivska V, Španêl P, Dryahina K, Sovová K. 2012. Variability in the concentrations of
661 volatile metabolites emitted by genotypically different strains of *Pseudomonas*
662 *aeruginosa*. Journal of Applied Microbiology.

663 52. Labows JN, McGinley KJ, Webster GF, Leyden JJ. 1980. Headspace analysis of volatile
664 metabolites of *Pseudomonas aeruginosa* and related species by gas chromatography-
665 mass spectrometry. J Clin Microbiol 12:521-6.

666 53. Hahn A, Whiteson, K., Davis, T. J., Phan, J., Sami, I., Koumbourlis, A. C., Freishtat, R.
667 J., Crandall, K. A., Bean, H. D. 2020. Longitudinal associations of the cystic fibrosis
668 airway microbiome and volatile metabolites: A case study. Front Cell Infect Microbiol 10.

669 54. Grosso-Becerra MV, Santos-Medellín C, González-Valdez A, Méndez JL, Delgado G,
670 Morales-Espinosa R, Servín-González L, Alcaraz LD, Soberón-Chávez G. 2014.
671 *Pseudomonas aeruginosa* clinical and environmental isolates constitute a single
672 population with high phenotypic diversity. BMC Genomics 15.

673 55. Bragonzi A, Worlitzsch D, Pier GB, Timpert P, Ulrich M, Hentzer M, Andersen JB,
674 Givskov M, Conese M, Doring G. 2005. Nonmucoid *Pseudomonas aeruginosa*

675 expresses alginate in the lungs of patients with cystic fibrosis and in a mouse model. J
676 Infect Dis 192:410-419.

677 56. Moskowitz SM, Emerson JC, McNamara S, Shell RD, Orenstein DM, Rosenbluth D,
678 Katz MF, Ahrens R, Hornick D, Joseph PM, Gibson RL, Aitken ML, Benton WW, Burns
679 JL. 2011. Randomized trial of biofilm testing to select antibiotics for cystic fibrosis airway
680 infection. Pediatr Pulmonol 46:184-192.

681 57. Cornforth DM, Dees JL, Ibberson CB, Huse HK, Mathiesen IH, Kirketerp-Moller K,
682 Wolcott RD, Rumbaugh KP, Bjarnsholt T, Whiteley M. 2018. *Pseudomonas aeruginosa*
683 transcriptome during human infection. Proc Natl Acad Sci U S A 115:E5125-E5134.

684 58. Harrison F, Muruli A, Higgins S, Diggle SP. 2014. Development of an *ex vivo* porcine
685 lung model for studying growth, virulence, and signaling of *Pseudomonas aeruginosa*.
686 Infect Immun 82:3312-3323.

687 59. Bayes HK, Ritchie N, Irvine S, Evans TJ. 2016. A murine model of early *Pseudomonas*
688 *aeruginosa* lung disease with transition to chronic infection. Sci Rep 6.

689 60. Crabbé A, Ledesma MA, Nickerson CA. 2014. Mimicking the host and its
690 microenvironment *in vitro* for studying mucosal infections by *Pseudomonas aeruginosa*.
691 Pathog Dis 71:1-19.

692 61. Sønderholm M, Kragh KN, Koren K, Jakobsen TH, Darch SE, Alhede M, Jensen PO,
693 Whiteley M, Kühl M, Bjarnsholt T. 2017. *Pseudomonas aeruginosa* aggregate formation
694 in an alginate bead model system exhibits *in vivo*-like characteristics. Appl Environ
695 Microbiol 83.

696 62. Bean HD, Jiménez-Díaz J, Zhu J, Hill JE. 2015. Breathprints of model murine bacterial
697 lung infections are linked with immune response. Eur Respir J 45:21-24.

698 63. Neerincx AH, Brinkman P, Van der Gouwe P, Paff T, Fens N, Gerritsen MG, Haarman E,
699 Van Aalderen WMC, Majoor CJ, Weersink EJM, Bos LD, Sterk PJ, Anke-Hilse Maitland-
700 Van der Z. 2018. Detection of *Pseudomonas aeruginosa* in exhaled breath of cystic
701 fibrosis patients. Eur Respir J 52.
702