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 one another 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.

IMPORTANCE

Pseudomonas aeruginosa is a leading cause of chronic lung infections in cystic fibrosis (CF), and are correlated with lung function declines. Significant clinical efforts are, therefore, aimed at detecting infections and tracking them for phenotypic changes, such as mucoidy and antibiotic resistance. Both the detection and tracking of lung infections relies on sputum cultures, but due to improvements in CF therapies, sputum production is declining though risks for lung infections

persist. Therefore, we are working toward the development of breath-based diagnostics for CF lung infections. In this study we characterized of the volatile metabolomes of 81 *P. aeruginosa* clinical isolates collected from 17 CF patients over a duration of at least five years of a chronic lung infection. We found that the volatilome of *P. aeruginosa* adapts over time, and correlates to infection phenotype changes, suggesting it may be possible to track chronic CF lung infections with a breath test.

KEYWORDS

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

RUNNING TITLE

P. aeruginosa cystic fibrosis volatilome

67 **INTRODUCTION**

68 Cystic fibrosis (CF) is an autosomal recessive disease caused by a mutation in the
69 CFTR protein that regulates ion transport across epithelia. In the lungs, reduced or lost CFTR
70 function leads to defective mucociliary transport, facilitating infection and colonization by a
71 plethora of microorganisms (1). *Pseudomonas aeruginosa* is one of the most prevalent lung
72 pathogens in CF – especially after adolescence (2) – and is able to establish chronic infections
73 that can last for years to decades (3, 4). *P. aeruginosa* lung infection is associated with more
74 rapid lung function decline, increased risk of hospitalization, and increased risk of death (5, 6).
75 Significant clinical efforts are, therefore, aimed at diagnosing and treating new infections to
76 delay the onset of chronic infection (7).

77 As *P. aeruginosa* transitions from an acute infection to chronicity, it undergoes a variety
78 of genotypic, phenotypic, and metabolic changes (8, 9). It has been well-established that
79 several phenotypes are correlated with chronic infection, including reduced motility and
80 increased mucoidy, antibiotic resistance, and biofilm formation (9-11), and some of these
81 phenotypes are also correlated with poorer patient outcomes (6, 12, 13). In light of this, *P.*
82 *aeruginosa* phenotypes could be used to evaluate infection stage, CF disease progression, and
83 patient morbidity risk (10, 12, 14, 15). Poor access to infection sites in the lower airways,
84 however, delays the detection of new *P. aeruginosa* infections and reduces the feasibility and
85 accuracy of disease state tracking via bacterial phenotypes, whether by culture-dependent or
86 culture-independent methods (16).

87 We are developing breath-based diagnostics to detect new *P. aeruginosa* infections and
88 track chronic infection phenotypes in the CF lung. Breath consists of thousands of volatile
89 organic compounds containing a wealth of information about human health and disease (17,
90 18), and is being leveraged for the development of diagnostics for wide-ranging conditions (19,
91 20). In the context of chronic lung infections, breath-based diagnostics possess significant
92 advantages over other diagnostic modalities. First, breath provides a non-invasive way of

93 sampling the entire ventilated lung, a limitation of both sputum and bronchoalveolar lavage fluid
94 that leads to delayed diagnosis of new infections, or the detection of infection phenotype
95 changes (16, 21-24). Second, breath sampling captures metabolic information about the
96 pathogens *in situ*, eliminating the need for microbial enrichment steps and speeding the time-to-
97 diagnosis by days to weeks (25, 26). Additionally, *in situ* metabolic measurements provide
98 information on the physiology of the pathogens in the context of disease versus the context of
99 the lab. We propose that a breath-based diagnostic can be created for detecting and tracking
100 CF *P. aeruginosa* lung infections by correlating the microbial genotypic and phenotypic changes
101 that occur during chronic infections to changes in the infection volatile metabolome, or
102 "volatilome".

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

118 the *P. aeruginosa* volatilome will shrink during chronic infection, with a concomitant decrease in
119 chemical diversity and a reduction in inter-patient dissimilarity.
120 The primary goal of this study is to build a foundation for using volatilomes to diagnose
121 and track *P. aeruginosa* CF lung infections by exploring how the *in vitro* volatilome of *P.*
122 *aeruginosa* CF isolates change over the course of chronic infection. We used comprehensive
123 two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC×GC-
124 TOFMS) to characterize the volatilomes of 81 *P. aeruginosa* isolates from early and late chronic
125 lung infections from 17 persons with CF. Using an untargeted metabolomics approach, we
126 characterized the size and chemical composition of the CF *P. aeruginosa* volatilome, and
127 identified core volatiles that would be a primary source of breath biomarkers for diagnosing
128 infections. We also investigated how the *P. aeruginosa* volatilome is shaped by the CF lung
129 environment, characterizing changes in volatilome sizes and compositions over time, and how
130 these changes relate to intra-patient and inter-patient volatilome dissimilarities. A secondary
131 goal of this study is to provide further data on the *P. aeruginosa* volatilome via the largest single
132 analysis of *P. aeruginosa* headspace volatiles, to-date.
133
134 **METHODS**
135 **Bacterial isolates**
136 Eighty one *P. aeruginosa* isolates from 17 individuals with CF were acquired from the
137 Cystic Fibrosis Isolate Core at Seattle Children's Center for Global Infectious Disease
138 Research. From the majority of patients we obtained three isolates: one early isolate, defined as
139 the first cultured *P. aeruginosa* isolate, and two late isolates collected a minimum of five years
140 after the first isolate. For one patient (Patient 23), 32 additional isolates were collected at
141 intervals over 7.5 years between the collections of the early and two late isolates. For one
142 patient, only one isolate was included in this study. The full strain names of the isolates are

143 provided in Table S1, and all isolates are available upon request to the CF Isolate Core
144 (https://www.seattlechildrens.org/research/resources/cystic-fibrosis-isolate/).
145 We quantified five phenotypes that are correlated with chronic infections: proteases,
146 pyocyanin, rhamnolipids, twitching motility, and mucoidy. *P. aeruginosa* was cultured from
147 glycerol stocks on LB agar for 24 h, then a single colony was cultured aerobically to stationary
148 phase in LB Lennox broth (10 g tryptone, 5 g yeast extract, 5 g NaCl per liter) at 37 °C with
149 shaking at 200 RPM, unless otherwise noted. Results of phenotype assays are reported in
150 Table S1 as means of biological triplicates, except for mucoidy for which five replicates were
151 measured. Pyocyanin production was evaluated by methods adapted from (43). Cell-free *P.*
152 *aeruginosa* culture supernatant (7.5 mL) was extracted with 4.5 mL of chloroform, inverted for 2
153 min and centrifuged at 4,122 × *g* for 15 min. Three milliliters of the organic phase was removed
154 and extracted with 750 µL of 0.2 *N* hydrochloric acid and vortexed for 2 min. The aqueous
155 phase was aliquoted into a 96-well plate and absorbance was measured at 520 nm. Protease
156 production was evaluated by methods adapted from (15). A sterile wooden inoculation stick was
157 dipped into a culture of *P. aeruginosa* and then gently touched to the surface of a brain-heart
158 infusion-skim milk agar (1.5%) plate. Plates were incubated upright at 37 °C and zones of
159 clearance were measured at 48 h. Rhamnolipids production was evaluated by methods adapted
160 from (44). Proteose peptone-glucose-ammonium salts (PPGAS) medium was inoculated from
161 an LB pre-culture of *P. aeruginosa* and cultured at 37 °C to stationary phase with shaking at 200
162 RPM. Cultures were centrifuged at 21,694 × *g* for 1 min to pellet cells. The supernatant was
163 serially diluted two-fold in PPGAS and 20 µL of each dilution was spotted onto a microtiter plate
164 lid. Spots were classified as drops or collapsed drops and assigned numeric scores
165 corresponding to the number of dilutions needed to obtain a drop. Twitching motility was
166 evaluated by methods adapted from (45, 46). The pointed end of a sterile toothpick was touched
167 to the edge of a single *P. aeruginosa* colony and then stabbed to the bottom of a twitching

168 motility agar plate (per 1 L: 10 g tryptone, 5 g NaCl, 5 g yeast extract, 10 g agar). Plates were
169 incubated at 37 °C for 24 h and the radius of the interstitial biofilm was measured. For mucoidy,
170 frozen glycerol stocks of *P. aeruginosa* were streaked onto LB Lennox agar (1.5%) plates and
171 grown for 48 hrs. Colonies were visually inspected and the degree of mucoid morphology was
172 scored from 0 to 2, in 0.5 increments (0 = highly mucoid, 2 = non-mucoid). Twenty pairs of
173 isolates (one Early and one Late) from ten patients were selected for additional statistical
174 analyses, described below, based on changes to their phenotypes during chronic infections.
175

176 **Sample preparation**

177 Isolates were cultured as previously described (47). Briefly, isolates were cultured
178 aerobically for 16 h at 37 °C in 5 mL of lysogeny broth-Lennox (LB), and then diluted 1000-fold
179 into 25 mL of fresh LB and grown for 24 h under the same conditions. For metabolomics
180 analyses, cells were removed via centrifugation through a 0.2 *μ*m filter and 2 mL of each filtrate
181 was transferred to a 10 mL GC headspace vial with screw caps. Samples were prepared in
182 biological triplicate, with LB media controls prepared in parallel, and stored at –20 °C prior to
183 analysis.
184

185 **GC×GC analysis and data processing**

186 Culture filtrates were thawed and maintained at 4 °C until analyzed, as previously
187 described (47). Headspace volatiles were characterized using a Pegasus® 4D GC×GC-TOFMS
188 (LECO® Corporation, St. Joseph, MI) equipped with a MPS Pro® rail autosampler (Gerstel®
189 Inc, Linthicum Heights, MD). Column set configuration, and GC×GC, MS, and data processing
190 method parameters are previously reported (47), and summarized in Miscellaneous Table 1
191 (https://doi.org/10.1101/2020.06.13.126698).

192 Data processing steps are outlined in Miscellaneous Figure 1. Chromatographic artifacts
193 and suspected contaminants based on peak names and comparisons to blanks were removed
194 (Miscellaneous Table 2), as well as poorly modulated peaks eluting prior to 358 s (acetone
195 retention time) in the first dimension. Missing values were handled as follows: if detected in one
196 of three replicates, the measured value was permuted to 0; if detected in two of three replicates,
197 the missing value was imputed as half of the minimum detected value for that compound across
198 all samples.

199 Probabilistic quotient normalization (PQN) was applied to account for differences in peak
200 abundance due to variations in culture cell density, followed by a $\log_{10}$ transformation (48).
201 Peaks were further filtered out based on the following criteria: (1) within-triplicate intraclass
202 correlation coefficient (ICC) < 0.75 (definition: absolute agreement; model: two-way mixed
203 effects; type: mean of *k* measurements) (49); and (2) only detected in sterile media, or not
204 significantly greater in abundance compared to sterile media using the Wilcoxon rank-sum test
205 with Benjamini-Hochberg adjustment (significance threshold of .05) (50).

206 For statistical analyses beyond the reporting of detected peaks, peaks that showed
207 significant correlations to run order were identified using Kendall's tau with Benjamini-Hochberg
208 adjustment ($-0.6 \leq \tau \geq 0.6$, significance threshold of .05) and removed. Principal component
209 analysis (PCA) revealed an apparent batch effect that was attributed to, at least in part, a non-
210 biological phenomenon and was described previously (47). This batch effect was corrected
211 using an empirical Bayes approach (51). The geometric means of triplicates of the batch-
212 corrected data were used for further statistical analyses. For analysis of the ten paired early and
213 late chronic infection isolates, batch-corrected data was not used as all of these isolates were
214 contained within the same batch.

215 Peaks were putatively identified using published reporting standards (52). Level 2
216 identifications were determined based on the following criteria: (1) ≥ 80% mass spectral forward
217 match using the NIST 2011 library; (2a) first dimension retention times that possess a strong

linear fit with carbon number in homologous series ($r^2 \geq 0.995$, for the identification of select 2-, 3-, and 4-ketones), or (2b) experimentally determined linear retention indices (LRI) consistent with published LRIs, as determined by the following acceptance criteria based on the differences between experimental and published non-polar and polar column RIs for Grob's test mix:

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

Level 3 identifications were based on at least an 80% mass spectral forward match score. Chemical classifications (alcohols, aldehydes, aromatics, carboxylic acids, esters, hydrocarbons, ketones, thiols, or other) were assigned based primarily on their mass spectral identity, and secondarily by their chromatographic characteristics. Using polar second dimension columns, compounds of differing functional classes elute at easily discernible second-dimension retention times in a stratified manner, typically (from smallest ${}^2t_R$ to largest): hydrocarbons < ketones < aromatics (53-57). In this study, compounds identified as hydrocarbon, ketones, and aromatics eluted at approximately 0.71 s ± 0.05 s, 0.88 s ± 0.22 s, and 1.12 s ± 0.26 s, respectively. Level 4 identifications are those that meet none of the above-mentioned criteria.

Statistical analyses

A Phenotype Score was determined, for each replicate, by scaling the data for each phenotype to a range of 0 to 1, where 1 is the maximum value and 0 is the minimum value (excluding outliers). Scaled phenotype data were then summed to yield the Score, and averaged across replicates. Patient age was not included in calculation of Scores. Two-tailed Wilcoxon signed-rank tests (with continuity correction, where appropriate; significant threshold of .05) were used to test for significant differences between Late and Early isolate Scores.

243 Volatile compounds that were detected in at least 95% of all samples were classified as
244 "core" volatiles. Significant differences between inter-isolate phenotypes were tested using the
245 Wilcoxon signed-rank test (significance threshold of .05). Significant differences between intra-
246 and inter-patient volatilomes were tested using one-way analysis of variance (ANOVA)
247 (significance threshold of .05), and Tukey's honestly significant differences (HSD) multiple
248 comparison procedure (significance threshold of .05). The relatedness of isolates based on their
249 volatilomes were assessed using agglomerative hierarchical clustering analysis (HCA), non-
250 metric multidimensional scaling (NMDS), and permutational multivariate analysis of variances
251 (PERMANOVA) on the Euclidean distances between isolates. Linear regression and Pearson's
252 correlation of Euclidean distances of Patient 23 isolates were used to assess intra-patient
253 isolate dissimilarity over time. All statistical analyses were performed using R version 3.3.2 (The
254 R Foundation for Statistical Computing) with the following packages (version): *ICC* (2.3.0),
255 *pairwiseAdonis* (0.0.1), *pheatmap* (1.0.8), *stats* (3.3.2), and *sva* (3.22.0).
256
257 **Data availability**
258 Metabolomic data (chemical feature peak areas and retention time information) included
259 in this study are available at the NIH Common Fund's National Metabolomics Data Repository
260 (NMDR) website, the Metabolomics Workbench, at www.metabolomicsworkbench.org, where it
261 has been assigned Project ID PR000970, Study ID ST001414 (DOI:
262 http://dx.doi.org/10.21228/M89Q4F). Additional Miscellaneous Information mentioned can be
263 found alongside Supplementary Information at bioRxiv (DOI:
264 https://doi.org/10.1101/2020.06.13.126698)
265
266 **RESULTS**
267 **Characteristics of *P. aeruginosa* CF isolates**

268 We obtained from a biorepository 81 *P. aeruginosa* chronic infection isolates, which had
269 been collected from 17 individuals with CF. From 14 patients, two or three *P. aeruginosa*
270 isolates were obtained: one isolate that was the first cultured *P. aeruginosa*, and one or two
271 isolates that were collected at least five years after the first. We refer to these isolates as Early
272 and Late, respectively. For one patient (Patient 75), four isolates were collected: two Early
273 isolates collected one month apart and two Late isolates, collected ten and 16 years after the
274 first. For one patient (Patient 23), 32 additional isolates were collected over the course of a 7.5 y
275 infection period, which we have termed Intermediate isolates. For one patient, only one isolate
276 was collected. These isolates were genetically characterized in a study by Smith *et al*., who
277 determined that the intra-patient replicate isolates are all clonally related, with the exception of
278 the four isolates from patient 75, which are actually two clonally-related Early/Late pairs
279 (isolates 75E-1 and 75L-2 are clonal, and isolates 75E-2 and 75L-1 are clonal) (3).
280 For all isolates, we characterized five clinically relevant phenotypes *in vitro*: mucoidy,
281 pyocyanin production, rhamnolipids production, protease production, and twitching motility (Fig.
282 1; Table S1). These phenotypes are commonly altered during the course of chronic CF lung
283 infections (15). There were wide ranges in the expression of these phenotypes across all
284 isolates, as expected. Several isolate sets exhibited phenotypes consistent with age of
285 collection in that Early isolates possessed higher degrees of motility, higher quantities of
286 pyocyanin, rhamnolipids, and proteases (indicating intact quorum regulation), and lower
287 mucoidy compared to their cognate Late isolates (*e.g.*, Patients 23, 31, 33, 36, and 76). The
288 associations between relative patient age at isolate collection and *P. aeruginosa* phenotypes
289 were not perfect, however. For example, Early isolate 62E had no detectable pyocyanin,
290 proteases, rhamnolipids, or twitching motility, and many of the Late isolates retained more early-
291 like phenotypes (e.g., both Late isolates of 41, 66, 71, 75, and 101, and 60L-1, 74L-2, and 100L-
292 2). Despite observable changes in isolate phenotype within patient, no phenotypes were
293 significantly different when comparing Late isolate(s) to the Early isolate. Collectively, the 81

294 isolates were highly varied, and represented the array of phenotypes we expect to observe in
295 the span of *P. aeruginosa* CF lung infections from initial infection to long-established chronic
296 infections.
297
298 **The *in vitro* volatilome of *P. aeruginosa* CF isolates**
299 Taking an untargeted metabolomics approach, we cultured the clinical *P. aeruginosa*
300 isolates *in vitro* and characterized their volatilomes using GC×GC-TOFMS. Following extensive
301 data processing, including removal of analytically biased chromatographic features, we
302 conservatively attribute 539 non-redundant volatile compounds to the growth and metabolism of
303 the *P. aeruginosa* isolates (Table S2). Among these, 69 compounds are shared by at least 95%
304 ($n \geq 77$) of the isolates, representing core volatiles. Using minimum metabolomics reporting
305 standards, we assigned compound identification levels 1-4 (1 high, with 2 being the highest in
306 this study) based on a combination of mass spectral and chromatographic characteristics. We
307 were able to assign putative names to 14 core and 33 non-core compounds with an ID level 2
308 (Table 1), many of which have been previously reported as *P. aeruginosa* volatiles (Table 1).
309 While we detected 2-aminoacetophenone (the volatile compound responsible for the grape-like
310 odor of *P. aeruginosa*), it possessed an intraclass correlation slightly lower than the applied
311 threshold of 0.75 and was ultimately filtered out of our peak list. We also identified several
312 compounds of a variety of chemical classes that have not been previously reported for *P.*
313 *aeruginosa*, including: alcohols (3-methyl-1-buten-1-ol, 1-butoxy-2-propanol, 2-butyl-1-octanol),
314 esters (butanoic acids, methyl isovalerate), aromatics (tetrahydrofuran, tetramethylpyrazine, 2-
315 butylfuran, 3-cyano-2,5-dimethylpyrazine, 2,4-(1,1-dimethyl)-phenol), unsaturated ketones (3-
316 hydroxy-butan-2-one, 4-undecanone, 1-octen-3-one, 3-penten-2-one, 6-methyl-2-heptanone, 1-
317 (4-ethylphenyl)-ethanone), and an aldehyde (decanal).
318 Chemical classifications were assigned to ID level 2 and 3 compounds (187 compounds)
319 using mass spectral match and retention time characteristics (Fig. 2). Among Level 2 and 3 core

320 volatiles, the majority of classified compounds were hydrocarbons (41%), followed by ketones
321 (16%). Other oxidized compounds, including alcohols, aromatics, acids, and thiols, accounted
322 for an additional 43%. The non-core volatiles had similar chemical compositions, which are
323 consistent with previous studies of the *P. aeruginosa* volatilome (25, 28, 53, 58, 59). Of the
324 entire volatilome, 65% of the volatiles (*n* = 352) possessed less than an 80% mass spectral
325 match to the 2011 NIST MS library, and as such were classified as unknowns.
326

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

Compound Name[a]	Formula	Chemical Class	Samples (*n* = 81)	Refs.[b]
Core volatile compounds				
3-methylbutanal	$C_5H_{10}O$	Aldehyde	79	
thiocyanic acid, methyl ester	C_2H_3NS	Ester	77	(53)
3-methyl-1-butanol	$C_5H_{12}O$	Alcohol	77	(58, 60, 61)
pyridine	C_5H_5N	Aromatic	77	
3-penten-2-one	C_5H_8O	Ketone	77	
3-methyl-1-buten-1-ol	$C_5H_{10}O$	Alcohol	78	
4-methyl-3-penten-2-one	$C_6H_{10}O$	Ketone	81	(28, 62)
hexanal	$C_6H_{12}O$	Aldehyde	81	(58, 60, 62)
heptanal	$C_7H_{14}O$	Aldehyde	80	
2,5-dimethylpyrazine	$C_6H_8N_2$	Aromatic	81	(53)
2-nonanone	$C_9H_{18}O$	Ketone	81	(28, 53, 58, 59, 61-67)
4-ethyl-1,2-dimethylbenzene	$C_{10}H_{14}$	Aromatic	77	
1,3,3-trimethyl-(bicyclo)-heptan-2-ol	$C_{10}H_{18}O$	Alcohol	78	
1-(4-ethylphenyl)-ethanone	$C_{10}H_{12}O$	Ketone	79	
Non-core volatile compounds				
dimethyl sulfide	C_2H_6S	Other	29	(28, 53, 58, 60, 61, 64, 68, 69)
2-methyl-3-buten-2-ol	$C_5H_{10}O$	Alcohol	35	
2-butanone	C_4H_8O	Ketone	1	(28, 53, 58, 59, 61, 63, 66, 70)
tetrahydrofuran	C_4H_8O	Aromatic	1	
acetic acid	$C_2H_4O_2$	Carboxylic acid	55	
methyl thioacetate	C_3H_6OS	Thiol	3	(58, 60, 62)

2,4-dimethylfuran	C_6H_8O	Aromatic	66	(28)
butanoic acid, methyl ester	$C_5H_{10}O_2$	Ester	3	
3-hydroxybutan-2-one	$C_4H_8O_2$	Ketone	18	
dimethyl disulfide	$C_2H_2S_2$	Other	1	(28, 58-62, 66, 68-73)
2-methyl-butanoic acid, methyl ester	$C_6H_{12}O_2$	Ester	3	
3-methyl-butanoic acid, methyl ester	$C_6H_{12}O_2$	Ester	3	
furfural	$C_5H_4O_2$	Aromatic	67	(60)
2-butylfuran	$C_8H_{12}O$	Aromatic	6	
2-hexen-1-ol	$C_6H_{12}O$	Alcohol	51	(28)
1-butoxy-2-propanol	$C_7H_{16}O_2$	Alcohol	61	
6-methyl-2-heptanone	$C_8H_{16}O$	Ketone	33	
1-methylethylbenzene	C_9H_{12}	Aromatic	57	
3-octanone	$C_8H_{16}O$	Ketone	56	(28, 61)
4,6-dimethyl-2-heptanone	$C_9H_{18}O$	Ketone	47	
2-ethyl-5-methylpyrazine	$C_7H_{10}N_2$	Aromatic	58	
4-nonanone	$C_9H_{18}O$	Ketone	49	(28)
tetramethylpyrazine	$C_8H_{12}N_2$	Aromatic	5	
2-octenal	$C_8H_{14}O$	Aldehyde	48	(53)
octanenitrile	$C_8H_{15}N$	Other	73	(28)
2,4-octadienal	$C_8H_{12}O$	Aldehyde	9	(28)
3-decanone	$C_{10}H_{20}O$	Ketone	41	(53)
decanal	$C_{10}H_{20}O$	Aldehyde	65	
2-decanone	$C_{10}H_{20}O$	Ketone	13	(28)
4-undecanone	$C_{11}H_{22}O$	Ketone	30	
2-dodecanone	$C_{12}H_{24}O$	Ketone	18	(28, 53)
2-butyl-1-octanol	$C_{12}H_{26}O$	Alcohol	48	
2,4-bis(1,1-dimethyl)-phenol	$C_{14}H_{22}O$	Aromatic	26	

[a]Bolded compounds have not been previously reported as *P. aeruginosa* volatile compounds.
[b]Previous references for these volatiles in *P. aeruginosa* cultures are provided.

327

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

329 Because every patient is infected by a genetically-unique strain, we hypothesized that
330 the dissimilarities of intra-patient volatilomes would be lower than inter-patient. Additionally, due
331 to the greater metabolic potential of early-infection isolates with intact regulatory networks, we
332 hypothesized that the volatilomes of inter-patient Early isolates would be more dissimilar than

333 those of inter-patient Late isolates. We calculated the pairwise Euclidean distances as a
334 measure of dissimilarity between all Early isolate volatilomes ("inter-Early dissimilarity"), all Late
335 isolate volatilomes ("inter-Late dissimilarity"), and between Early and Late isolate volatilomes
336 from the same patient ("intra-patient dissimilarity"). Significant differences between means were
337 tested using one-way ANOVA with Tukey's HSD multiple comparison procedure. Interestingly,
338 the inter-Early isolate dissimilarity was the lowest (mean 35.2, median 33.8) and the inter-Late
339 isolate dissimilarity was the highest (mean 39.1, median 39.1), and these differences were
340 significant (95% CI: 3.05 – 4.86; $P = 0+$) (Fig. 3). Similarly, mean inter-Late volatilome
341 dissimilarity was significantly greater than that of intra-Patient volatilomes (mean 36.4, median
342 35.6) (95% CI: 0.41 – 4.97; $P = .013$). Even more intriguing was that the inter-Early dissimilarity
343 was lower than even the intra-patient dissimilarity, though this difference was not significant
344 (95% CI: -1.11 – 3.66; $P = .515$). For comparison, we also calculated the pairwise Euclidean
345 distances of all isolates in this study (Pooled; mean 38.0, median 37.9); the mean inter-Early
346 and mean inter-Late dissimilarities were both significantly lower than the mean dissimilarity of
347 the Pooled isolates (95% CIs: 1.99 – 3.77, -1.65 – 0.51; $P = 0+$).
348
349 Visualizing the Euclidean distances between isolate volatilomes via non-metric
350 multidimensional scaling (NMDS) and ordered dissimilarity images (ODI), we looked for
351 clustering of clonal isolates, indicating intra-patient similarities. The NMDS and ODI of the
352 Euclidean distances of all 81 isolate volatilomes show a primary cluster comprised mostly of
353 intermediate isolates of Patient 23 (Fig. S1). To reduce the influence of the over-represented
354 Patient 23 isolates in data visualization, we truncated our data set to 48 isolates by removing
355 the Intermediate isolates of Patient 23, and the lone Patient 21 isolate. The ODI plot and NMDS
356 of the truncated set revealed three clusters of similarity, none of which include all isolates from a
357 single patient (Fig. 4; Fig S2a,b). This refutes our hypothesis that isolates from the same patient
358 maintain a similar volatilome over time. To determine if any inter-patient or intra-patient isolate

359 pairs are significantly different, we performed permutational multivariate analysis of variance
360 (PERMANOVA) on the Euclidean distances for all possible pairs. Following Benjamini-Hochberg
361 adjustment of *P* values, no pairs of isolates were identified as having significantly different
362 volatilomes ($q \geq .1$ for all isolate pairs). However, PERMANOVA conducted on all isolates
363 indicated strong significant differences between the 81 volatilomes (*pseudo-*$F_{80,161} = 28.0$, P
364 $< .01$), as well as between the volatilomes of the truncated set of 48 isolates (*pseudo-*$F_{47,96}$ =
365 29.2, $P = 0+$), reinforcing the inference from the ODI and NMDS that the CF *P. aeruginosa*
366 volatilome is heterogeneous.

367 Though we do not find that patient isolates maintain similar volatilomes over the duration
368 of chronic infection, it is noteworthy that the majority of Patient 23's isolates have strong
369 similarities (Fig. S1). Taking a closer look at the 35 isolates from Patient 23 (Fig. 5; Fig. S2c),
370 the ODI and NMDS revealed four distinct similarity neighborhoods: one neighborhood consisting
371 of only the Early isolate 23E, which stands alone; a second consisting of the four latest-collected
372 isolates 23I-32, 23I-33, 23L-1, 23L-2; a third consisting of isolates 23I-7, 23I-9, 23I-12, 23I-14,
373 23I-29; and a fourth encompassing the remainder of the isolates. The four latest-collected
374 isolates are less defined by their similarity to each other than their dissimilarity to all other
375 isolates collected during the infection. A linear regression of the Euclidean distances between
376 the Early isolate and the Intermediate and Late isolates as a function of patient age showed that
377 isolates became increasingly dissimilar to the Early isolate over time (Pearson's r = 0.70, P =
378 0+) (Fig. 6). These observations are underpinned by the phenotypic differences of the four
379 latest isolates, and the number and types of known mutations that were accumulated in the 23L
380 versus 23E isolates (3). As described by Smith *et al*., the *P. aeruginosa* infection sampled from
381 Patient 23 (referred to as Patient 1 in the referenced publication) diversified from patient ages
382 1.5 to 3 y into a population of isolates, with mutations in virulence, motility, quorum sensing, iron
383 transport, efflux, and transcription and translation genes. The volatilome dissimilarities we
384 measured reflect the genotypic diversification of the infection as a function of time (Fig. 6).

385

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

387 We compared the chemical characteristics of the Early and Late volatilomes, positing
388 that the size and chemical diversity of the volatilomes would decrease from Early to Late
389 infection due to loss-of-function mutations. To perform the comparison, we needed to balance
390 the size of the isolate groups, and did so by selecting ten pairs of *P. aeruginosa* isolates that
391 had large changes in phenotypes over the duration of infection, thereby enhancing the
392 differences we may find in the Early vs. Late volatilomes (Fig. 1; Table S1). Visual inspection of
393 the ten pairs of GC×GC chromatograms suggests reductions in the number and variety of
394 volatile compounds produced by Late isolates (Fig. 7, Miscellaneous Figure 2). Contrary to the
395 appearance of the chromatograms, however, the overall size and chemical compositions of the
396 Early and Late volatilomes were similar to each other (Fig. S3a-d), with hydrocarbons
397 representing approximately 50% and alcohols and ketones together representing approximately
398 30% of the 410 Early and 441 Late volatiles. We quantified the chemical richness and diversity
399 of the pooled Early and pooled Late isolates using the Shannon-Wiener diversity index (Table
400 S3) and found the volatilomes of Early and Late isolates were similar, whether using all volatiles
401 or only the non-core volatiles.

402 The uniformity in the Early versus Late volatilomes is not a function of data aggregation,
403 as the volatilomes of the individual isolates are highly similar in number and chemical
404 composition, with the exception of one Late isolate (76L) that has a slow-growth phenotype (Fig.
405 S3e-h). Furthermore, when looking at individual isolate pairs, for the vast majority there were no
406 differences in the richness and diversity of their volatilomes (Table S3). Together, these results
407 suggest that there are not major differences between the number and diversity of chemical
408 compounds when comparing Early and Late infection isolates. Hierarchical clustering analysis
409 (HCA) of the *P. aeruginosa* isolates by their volatilomes underscores this observation; we

410 observed no discernible clustering of the Early versus Late isolates based on the presence and
411 absence of volatiles (Fig. S4a).
412
413 **Abundance of volatile compounds in Early and Late isolates**
414 Though we did not observe any differences in the Early and Late isolate volatilomes
415 based on presence and absence of metabolites, the chromatograms suggest a reduced
416 volatilome in late infection. We therefore posited that while the numbers of volatiles do not
417 significantly change from Early to Late infection isolates, the relative concentrations of volatiles
418 do. Using volatile abundances, we performed HCA on the 20 selected Early and Late isolates
419 and found clustering by the time of isolate collection and phenotype score (Fig. S4b). We were
420 interested in whether or not clustering still occurs in the larger, less curated set of 48 isolates
421 (*i.e.*, the truncated set), which has more discordance between phenotype and time of collection
422 (Fig. 1). HCA of the truncated set of isolates shows that Early isolates are generally more similar
423 to one another than Late isolates are (Fig. 8), reinforcing observations made by calculating
424 Euclidean distances; however, significant proportions of the Early and the Late isolates are mis-
425 clustered. Interestingly, the majority of the Late isolates clustering with the Early isolates in the
426 upper clade have more early-like phenotypes, and several Early isolates that have more late-
427 like phenotypes (*e.g.*, 62E, 73E, 70E) cluster in the lower clade with Late isolates, indicating a
428 relationship between the isolate phenomes and volatilomes. We also observed clustering by
429 phenotype when the volatilomes of all 81 isolates were analyzed (Fig. S5). Similar phenomena
430 are observed when clustering the truncated set or all 81 isolates by the core volatilome (Fig.
431 S6), and even when using only the 23 core alcohols, aldehydes, and hydrocarbons (Fig. 9; Fig.
432 S7), suggesting that a set of conserved volatiles could be identified and used as biomarkers for
433 detecting phenotypic changes in chronic infections.
434
435

436 **DISCUSSION**

437 *P. aeruginosa* isolates from cystic fibrosis lung infections have a large and diverse
438 volatilome. We conservatively attribute 539 detected volatiles to *P. aeruginosa* growth and
439 metabolism, only 13% of which were core volatiles. For volatiles that we could assign to a
440 chemical class, hydrocarbons represented the highest proportion, followed by ketones, which
441 we observed both in the pan-volatilome and for individual patient isolates. Approximately 40% of
442 core volatiles have been identified in other *P. aeruginosa* metabolomic studies, including the
443 most frequently detected *in vitro* volatiles 2-butanone, 2-nonanone, dimethyl sulfide, and
444 dimethyl disulfide (see Table 1 for references). The majority of the core volatiles we detected,
445 however, are reported here for the first time. This opens up speculation that there may be a
446 unique core volatile profile for strains colonizing the CF lung, though due to the inter-strain
447 variation in the *P. aeruginosa* volatilome, many more CF and non-CF isolates will need to be
448 characterized via untargeted GC×GC analyses to determine if this may be the case. To date, a
449 number of studies have attempted to identify collections of volatiles to differentiate *P.*
450 *aeruginosa*-positive or -negative subjects from breath (58, 74-81) and lavage (25) with
451 promising results. Compared to this study, however, the referenced analyses were either
452 targeted at single compounds (e.g., hydrogen cyanide or 2-aminoacetophenone), did not
453 characterize the volatilome of *P. aeruginosa* in detail, and/or did not draw inferences to infection
454 stage or changes across time. This study represents the first untargeted comparative analysis of
455 *P. aeruginosa* volatilomes over long-term chronic infections, expanding the body of knowledge
456 on *P. aeruginosa* metabolism and broadening the potential applications for breath-based
457 diagnostics.

458 Despite the trend toward reduced or loss-of-phenotype expression in late infection
459 isolates – which we hypothesized would correspond to a reduced metabolome size – we
460 observed that Late isolates produce the same number and variety of volatile compounds as
461 Early isolates. However, the relative abundance of volatiles produced by the Late isolates is

462 significantly reduced. Our data suggest that mutations that arise in *P. aeruginosa* chronic
463 infection isolates do not result in complete inhibition of metabolic pathways, but may instead
464 reduce metabolic flux. This observation has implications for diagnostics, signifying that presence
465 and absence of *P. aeruginosa* volatile metabolites would provide little diagnostic value for
466 tracking chronic lung infections. Rather, a breath-based diagnostic for monitoring *P. aeruginosa*
467 adaptations in CF will likely need to measure the relative abundance of metabolites. Importantly,
468 a subset of core volatiles can be used to cluster isolates based on commonalities in
469 phenotypes, indicating that a diagnostic based upon the relative abundances of conserved
470 hydrocarbons, alcohols, and aldehydes (or a few selected volatiles from these classes) may be
471 sufficient to track *P. aeruginosa* infection in most patients. It should be noted that core *P.*
472 *aeruginosa* volatiles are not necessarily unique to *P. aeruginosa*. In fact, a review of the mVOC
473 2.0 microbial volatile organic compound database (82) shows nearly half of the core volatiles
474 (and a third of named compounds overall) have been reported as being produced by other CF
475 lung bacterial pathogens, including *Staphylococcus aureus*, *Klebsiella pneumoniae*,
476 *Haemophilus influenzae*, *Stenotrophomonas maltophilia*, and *Burkholderia spp*. This further
477 highlights the necessity of developing a breath-based diagnostic based on the relative
478 abundances of sets of volatiles, which has been shown to accurately discriminate between
479 human pathogens (distantly or closely related to *P. aeruginosa*) in numerous *in vitro*, *ex vivo*,
480 and *in vivo* analyses (25, 62, 83-85).

481 Based on the genomic heterogeneity of *P. aeruginosa* (3, 33-35), we hypothesized that
482 every patient would have a unique *P. aeruginosa* volatilome, and that intra-patient isolates
483 would have more similar volatilomes than inter-patient isolates. To the contrary, we observed
484 that the volatilomes of Early and Late isolates from a single patient are no more similar to one
485 another than they are to any other isolate (intra-patient vs. pooled data, Fig. 3). Interestingly, the
486 volatilomes of inter-patient Early isolates are even more similar than intra-patient isolates, while
487 the volatilomes of inter-patient Late isolates have the highest dissimilarity. These data indicate

488 that the genomic similarity within patient is less influential on the volatilome than the post-
489 transcriptional diversity that evolves within a chronic *P. aeruginosa* infection population. The
490 data from Patient 23, however, suggest that intra-patient isolates are more likely to have similar
491 volatilomes when there are fewer mutational and phenotypic differences between them. From
492 these results, we propose that a breath-based diagnostic to track *P. aeruginosa* chronic lung
493 infections would be personalized, with each patient's breathprint at initial infection serving as a
494 baseline to which subsequent breathprints are compared. Sustained, significant deviations from
495 the patient's baseline breathprint would indicate that the *P. aeruginosa* population phenotype
496 has changed, and therefore the patient may be at higher risk for associated poor clinical
497 outcomes (*e.g.*, higher risk of exacerbation) (15) and at higher risk of infection eradication failure
498 after antibiotic treatment (12). In a longitudinal case study of a single CF patient, we have
499 shown that changes in the dominant species colonizing the lung correlate to changes in the
500 volatile metabolites detected in sputum (86), and therefore monitoring for breathprint deviations
501 may also work for *P. aeruginosa* infections, as these data suggest. Despite this, the issue of
502 specificity becomes a concern unless biomarkers for individual *P. aeruginosa* phenotypes are
503 identified and used for tracking; this is the subject of ongoing work.
504 We highlight some limitations in this study. Although we make reference to Early isolates
505 and those phenotypes associated with early infection, it is important to recognize that these
506 isolates were merely the first culturable *P. aeruginosa* from the patients included in this study. It
507 is likely that some patients had *P. aeruginosa* in the lung for some years in spite of persistent
508 negative cultures into late childhood or adolescence (15), exemplified by isolates 62E and 73E
509 exhibiting late-like phenotypes. Phenotypes discordant with collection time or apparent infection
510 stage underscores a major drawback of culture-based methods for infection staging (*e.g.*, *in*
511 *vitro* phenotyping) in that the respiratory microbiome is prone to under-sampling, especially
512 during initial colonization, due to the complex structure of the lung. The increased use of new
513 therapeutics (*e.g.*, CFTR modulators) also impacts culture-based diagnostics as they reduce

514 sputum production, limiting the opportunity for clinical cultures. Utilizing breath-based
515 biomarkers, on the other hand, can overcome these challenges as breath represents a sample
516 of the entire ventilated lung environment.

517 *P. aeruginosa* chronic infection phenotypes are positively correlated with advanced
518 patient age (15, 87-89), and we observe a moderate, but significant correlation (Kendall's tau –
519 0.41, $P < .001$) between subject age and late-infection phenotype scores in this study. The
520 variability of isolate volatilomes we observed might, therefore, be explained by the age of their
521 corresponding patients. Testing this would require analyzing isolate volatilomes from a larger
522 number of subjects that were diagnosed with their first *P. aeruginosa* infections at younger and
523 older ages. In either case, the issue of the potential covariance between age and volatilome on
524 *P. aeruginosa* diagnostics could be overcome by taking a multifactorial approach as opposed to
525 volatile biomarkers only. We may find that using a combination of several clinical predictors
526 such as age, *P. aeruginosa* phenotype, and volatilome together provides more diagnostic power
527 than any individual metric. It will be important to explore this more closely in future work.
528 Additionally, while we only presented volatilome trends in relation to five clinically-relevant
529 phenotypes in this study (production of pyocyanin, rhamnolipids, and proteases, twitching
530 mobility, and mucoidy), there are additional clinically-relevant phenotypes, including antibiotic
531 resistance, that should be included in future analyses.

532 In summary, current methods for tracking *P. aeruginosa* lung infection progression are
533 predominantly culture-based (16, 21), and our results suggest a potential role for *in vitro* volatile
534 metabolomics in staging chronic infections in conjunction with established clinical laboratory
535 methods. The clinically-correlated *P. aeruginosa* phenotypes we measured in this study –
536 mucoidy, motility, and quorum-regulated traits – are sensitive to environment (90-93); as such,
537 the metabolomes we characterized from *in vitro* cultures are unlikely to directly reflect the
538 metabolomes these isolates produced in the CF lung. For the development of a breath test for
539 tracking infections *in situ*, it will be necessary to utilize culture models that are more reflective of

540 the lung (94-97) and to collect *ex vivo* and *in vivo* infection volatilomes (25, 75, 98). The utility of
541 breath for tracking chronic lung infections is being demonstrated in tuberculosis lung infection
542 models (26), and therefore we anticipate that the development of a breath-based diagnostic for
543 tracking *P. aeruginosa* chronic lung infections is feasible. It will also be important to ensure that
544 the *P. aeruginosa* volatiles we have identified are detectable in breath. This is the primary goal
545 of the observational clinical study IMproving *P. Aeruginosa* deteCTion using Breath (IMPACT-
546 Breath) that is underway. Despite the limitations of *in vitro* analyses, this study presents the first
547 comprehensive analysis of the *P. aeruginosa* volatilome from chronic CF lung infections, and
548 lays the groundwork for the application of volatile metabolomics in tracking CF lung disease.
549

550 **ACKNOWLEDGEMENTS**

551 Funding to support this work was provided by the Cystic Fibrosis Foundation (Hill17P0,
552 Hill18A0-CI), the National Institutes of Health (R56HL139846), and the ASU School of Life
553 Sciences Undergraduate Research (SOLUR) program. Isolates were obtained from the Cystic
554 Fibrosis Isolate Core at Seattle Children's Research Institute, funded by the National Institutes
555 of Health (P30 DK089507) and the Cystic Fibrosis Foundation (HOFFMA20Y2-OUT).
556

557 **AUTHOR CONTRIBUTIONS**

558 HB conceived of the study; HB, TD designed the experiments; AK, CB, SW, and TD collected
559 phenotype data; TD collected metabolomics data; TD and MR processed and analyzed the
560 metabolomics data; HB and TD wrote and edited the manuscript; all authors approved the final
561 version.
562

563 **CONFLICTS OF INTEREST**

564 The authors declare that the research was conducted in the absence of any commercial or
565 financial relationships that could be construed as a potential conflict of interest.

566

567 **REFERENCES**

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

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

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

577 4. Li ZH, Kosorok MR, Farrell PM, Laxova A, West SEH, Green CG, Collins J, Rock MJ,
578 Splaingard ML. 2005. Longitudinal development of mucoid *Pseudomonas aeruginosa*
579 infection and lung disease progression in children with cystic fibrosis. JAMA 293:581-
580 588.

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

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

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

589 8. Winstanley C, O'Brien S, Brockhurst MA. 2016. *Pseudomonas aeruginosa* evolutionary
590 adaptation and diversification in cystic fibrosis chronic lung infections. Trends Microbiol
591 24:327-337.

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

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

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

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

604 13. Courtney JM, Bradley J, McCaughan J, O'Connor TM, Shortt C, Bredin CP, Bradbury I,
605 Elborn JS. 2007. Predictors of mortality in adults with cystic fibrosis. Pediatr Pulmonol
606 42:525-532.

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

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

612 16. Héry-Arnaud G, Nowak E, Caillon J, David V, Dirou A, Revert K, Munck MR, Frachon I,
613 Haloun A, Horeau-Langlard D, Le Bihan J, Danner-Boucher I, Ramel S, Pelletier MP,
614 Rosec S, Gouriou S, Poulhazan E, Payan C, Ferec C, Rault G, Le Gal G, Le Berre R.
615 2017. Evaluation of quantitative PCR for early diagnosis of *Pseudomonas aeruginosa*
616 infection in cystic fibrosis: A prospective cohort study. Clin Microbiol Infect 23:203-207.

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

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

621 19. Oakley-Girvan I, Davis SW. 2018. Breath based volatile organic compounds in the
622 detection of breast, lung, and colorectal cancers: A systematic review. Cancer Biomark
623 21:29-39.

624 20. Amann A, Miekisch W, Schubert J, Buszewski B, Ligor T, Jezierski T, Pleil J, Risby T.
625 2014. Analysis of exhaled breath for disease detection. Annu Rev Anal Chem, Vol 7
626 7:455-482.

627 21. Smith WD, Bardin E, Cameron L, Edmondson CL, Farrant KV, Martin I, Murphy RA,
628 Soren O, Turnbull AR, Wierre-Gore N, Alton EW, Bundy JG, Bush A, Connett GJ, Faust
629 SN, Filloux A, Freemont PS, Jones AL, Takats Z, Webb JS, Williams HD, Davies JC.
630 2017. Current and future therapies for *Pseudomonas aeruginosa* infection in patients
631 with cystic fibrosis. FEMS Microbiol Lett 364.

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

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

638 24. West SEH, Zeng L, Lee BL, Kosorok MR, Laxova A, Rock MJ, Splaingard MJ, Farrell
639 PM. 2002. Respiratory infections with *Pseudomonas aeruginosa* in children with cystic
640 fibrosis - early detection by serology and assessment of risk factors. JAMA 287:2958-
641 2967.

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

645 26. Beccaria M, Bobak C, Maitshotlo B, Mellors TR, Purcaro G, Franchina FA, Rees CA,
646 Nasir M, Black A, Hill JE. 2019. Exhaled human breath analysis in active pulmonary
647 tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and
648 chemometric techniques. J Breath Res 13.

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

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

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

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

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

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

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

669 34. Cramer N, Klockgether J, Wrasman K, Schmidt M, Davenport CF, Tummler B. 2011.
670 Microevolution of the major common *Pseudomonas aeruginosa* clones C and PA14 in
671 cystic fibrosis lungs. Environ Microbiol 13:1690-704.

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

675 36. Bragonzi A, Paroni M, Nonis A, Cramer N, Montanari S, Rejman J, Di Serio C, Doring G,
676 Tummler B. 2009. *Pseudomonas aeruginosa* microevolution during cystic fibrosis lung

677 infection establishes clones with adapted virulence. Am J Respir Crit Care Med 180:138-
678 45.

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

682 38. Clark ST, Guttman DS, Hwang DM. 2018. Diversification of *Pseudomonas aeruginosa*
683 within the cystic fibrosis lung and its effects on antibiotic resistance. FEMS Microbiol Lett
684 365.

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

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

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

693 42. Ciofu O, Mandsberg LF, Bjarnsholt T, Wassermann T, Hoiby N. 2010. Genetic
694 adaptation of *Pseudomonas aeruginosa* during chronic lung infection of patients with
695 cystic fibrosis: Strong and weak mutators with heterogeneous genetic backgrounds
696 emerge in *mucA* and/or *lasR* mutants. Microbiology 156:1108-1119.

697 43. Essar DW, Eberly L, Hadero A, Crawford IP. 1990. Identification and characterization of
698 genes for a second anthranilate aynthase in *Pseudomonas aeruginosa*:

699 Interchangeability of the two anthranilate synthases and evolutionary implications. J
700 Bacteriol 172:884-900.

701 44. Caiazza NC, Shanks RMQ, O'Toole GA. 2005. Rhamnolipids modulate swarming
702 motility patterns of *Pseudomonas aeruginosa*. J Bacteriol 187:7351-7361.

703 45. Turnbull L, Whitchurch CB. 2014. Motility assay: Twitching motility. Methods Mol Biol
704 1149:73-86.

705 46. Glessner A, Smith RS, Iglewski BH, Robinson JB. 1999. Roles of *Pseudomonas*
706 *aeruginosa* las and rhl quorum-sensing systems in control of twitching motility. J
707 Bacteriol 181:1623-9.

708 47. Higgins Keppler EA, Jenkins CL, Davis TJ, Bean HD. 2018. Advances in the application
709 of comprehensive two-dimensional gas chromatography in metabolomics. Trends Anal
710 Chem 109:275-286.

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

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

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

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

720 52. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O,
721 Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN,
722 Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. 2007. Proposed
723 minimum reporting standards for chemical analysis Chemical Analysis Working Group
724 (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3:211-221.

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

728 54. Vogt L, Gröger T, Zimmermann R. 2007. Automated compound classification for ambient
729 aerosol sample separations using comprehensive two-dimensional gas
730 chromatography–time-of-flight mass spectrometry. J Chromatogr A 1150:2-12.

731 55. Seeley JV, Kramp F, Hicks CJ. 2000. Comprehensive two-dimensional gas
732 chromatography via differential flow modulation. Anal Chem 72:4346-4352.

733 56. Adahchour M, Beens J, Vreuls RJJ, Brinkman UAT. 2006. Recent developments in
734 comprehensive two-dimensional gas chromatography (GC x GC): IV. Further
735 applications, conclusions and perspectives. Trends Anal Chem 25:821-840.

736 57. Baptista I, Santos M, Rudnitskaya A, Saraiva JA, Almeida A. 2019. A comprehensive
737 look into the volatile exometabolome of enteroxic and non-enteroxic *Staphylococcus*
738 *aureus* strains. Int J Biochem Cell Biol 108:40-50.

739 58. Shestivska V, Španêl P, Dryahina K, Sovová K. 2012. Variability in the concentrations of
740 volatile metabolites emitted by genotypically different strains of *Pseudomonas*
741 *aeruginosa*. Journal of Applied Microbiology.

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

745 60. Lawal O, Knobel H, Weda H, Nijsen TME, Goodacre R, Fowler SJ, BreathDx c. 2018.
746 TD/GC-MS analysis of volatile markers emitted from mono- and co-cultures of
747 *Enterobacter cloacae* and *Pseudomonas aeruginosa* in artificial sputum. Metabolomics
748 14:66.

749 61. Filipiak W, Sponring A, Baur MM, Filipiak A, Ager C, Wiesenhofer H, Nagl M, Troppmair
750 J, Amann A. 2012. Molecular analysis of volatile metabolites released specifically by
751 *Staphylococcus aureus* and *Pseudomonas aeruginosa*. BMC Microbiol 12:113.

752 62. Neerincx AH, Geurts BP, Habets MFJ, Booij JA, van Loon J, Jansen JJ, Buydens LMC,
753 van Ingen J, Mouton JW, Harren FJM, Wevers RA, Merkus P, Cristescu SM, Kluijtmans
754 LAJ. 2016. Identification of *Pseudomonas aeruginosa* and *Aspergillus fumigatus* mono-
755 and co-cultures based on volatile biomarker combinations. J Breath Res 10:9.

756 63. Boots AW, Smolinska A, van Berkel J, Fijten RRR, Stobberingh EE, Boumans MLL,
757 Moonen EJ, Wouters EFM, Dallinga JW, Van Schooten FJ. 2014. Identification of
758 microorganisms based on headspace analysis of volatile organic compounds by gas
759 chromatography-mass spectrometry. J Breath Res 8:12.

760 64. Zscheppank C, Wiegand HL, Lenzen C, Wingender J, Telgheder U. 2014. Investigation
761 of volatile metabolites during growth of *Escherichia coli* and *Pseudomonas aeruginosa*
762 by needle trap-GC-MS. Anal Bioanal Chem 406:6617-28.

763 65. Senecal AG, Magnone J, Yeomans W, Powers EM. 2002. Rapid detection of pathogenic
764 bacteria by volatile organic compound (VOC) analysis. Proc SPIE 4575:610-613.

765 66. Zechman JM, Aldinger S, Labows JN, Jr. 1986. Characterization of pathogenic bacteria
766 by automated headspace concentration-gas chromatography. J Chromatogr 377:49-57.

767 67. Briard B, Heddergott C, Latge JP. 2016. Volatile compounds emitted by *Pseudomonas*
768 *aeruginosa* stimulate growth of the fungal pathogen *Aspergillus fumigatus*. mBio
769 7:e00219.

770 68. Preti G, Thaler E, Hanson CW, Troy M, Eades J, Gelperin A. 2009. Volatile compounds
771 characteristic of sinus-related bacteria and infected sinus mucus: Analysis by solid-
772 phase microextraction and gas chromatography-mass spectrometry. J Chromatogr B
773 Analyt Technol Biomed Life Sci 877:2011-8.

774 69. Allardyce RA, Langford VS, Hill AL, Murdoch DR. 2006. Detection of volatile metabolites
775 produced by bacterial growth in blood culture media by selected ion flow tube mass
776 spectrometry (SIFT-MS). J Microbiol Methods 65:361-5.

777 70. Rudzinski CM, Herzig-Marx R, Lin J, Szpiro A, Johnson B. 2004. Pathogen detection
778 using headspace analysis abstr Scientific Conference on Chemical and Biological
779 Defense Research, Hunt Valley, MD, MIT Lincoln Laboratory,

780 71. Carroll W, Lenney W, Wang T, Spanel P, Alcock A, Smith D. 2005. Detection of volatile
781 compounds emitted by *Pseudomonas aeruginosa* using selected ion flow tube mass
782 spectrometry. Pediatr Pulmonol 39:452-6.

783 72. Thorn RM, Reynolds DM, Greenman J. 2011. Multivariate analysis of bacterial volatile
784 compound profiles for discrimination between selected species and strains *in vitro*. J
785 Microbiol Methods 84:258-64.

786 73. Schöller C, Molin S, Wilkins K. 1997. Volatile metabolites from some gram-negative
787 bacteria. Chemosphere 35:1487-95.

788 74. Lammers M, Kos R, De Vries R, Brinkman P, Dagelet JWF, Rutjes NW, Majoor CJ,
789 Terheggen-Lagro SWJ, Weersink EJM, Bos LDJ, Sterk PJ, Maitland-Van Der Zee A-H,
790 Neerincx AH, Group TACR. 2018. Detection of *Pseudomonas aeruginosa* infections in
791 cystic fibrosis patients by enose technology. 52:PA3412.

792 75. Neerincx AH, Brinkman P, Van der Gouwe P, Paff T, Fens N, Gerritsen MG, Haarman E,
793 Van Aalderen WMC, Majoor CJ, Weersink EJM, Bos LD, Sterk PJ. 2018. Detection of
794 *Pseudomonas aeruginosa* in exhaled breath of cystic fibrosis patients. Eur Respir J 52.

795 76. Pabary R, Huang J, Kumar S, Alton EW, Bush A, Hanna GB, Davies JC. 2016. Does
796 mass spectrometric breath analysis detect *Pseudomonas aeruginosa* in cystic fibrosis?
797 Eur Respir J 47:994-7.

798 77. Gilchrist FJ, Belcher J, Jones AM, Smith D, Smyth AR, Southern KW, Spanel P, Webb
799 AK, Lenney W. 2015. Exhaled breath hydrogen cyanide as a marker of early
800 *Pseudomonas aeruginosa* infection in children with cystic fibrosis. ERJ Open Res 1.

801 78. Gilchrist FJ, Bright-Thomas RJ, Jones AM, Smith D, Spanel P, Webb AK, Lenney W.
802 2013. Hydrogen cyanide concentrations in the breath of adult cystic fibrosis patients with
803 and without *Pseudomonas aeruginosa* infection. J Breath Res 7:026010.

804 79. Robroeks C, van Berkel J, Dallinga JW, Jobsis Q, Zimmermann LJI, Hendriks HJE,
805 Wouters MFM, van der Grinten CPM, van de Kant KDG, van Schooten FJ, Dompeling E.
806 2010. Metabolomics of volatile organic compounds in cystic fibrosis patients and
807 controls. Pediatric Research 68:75-80.

808 80. Scott-Thomas AJ, Syhre M, Pattemore PK, Epton M, Laing R, Pearson J, Chambers ST.
809 2010. 2-aminoacetophenone as a potential breath biomarker for *Pseudomonas*
810 *aeruginosa* in the cystic fibrosis lung. BMC Pulmonary Medicine 10:10.

811 81. Enderby B, Smith D, Carroll W, Lenney W. 2008. Hydrogen cyanide as a biomarker for
812 *Pseudomonas aeruginosa* in the breath of children with cystic fibrosis. Pediatr Pulmonol
813 44:142-147.

814 82. Lemfack MC, Gohlke BO, Toguem SMT, Preissner S, Piechulla B, Preissner R. 2018.
815 mVOC 2.0: A database of microbial volatiles. Nucleic Acids Res 46:D1261-D1265.

816 83. Dryahina K, Sovová K, Nemec A, Španěl P. 2016. Differentiation of pulmonary bacterial
817 pathogens in cystic fibrosis by volatile metabolites emitted by their *in vitro* cultures:
818 *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Stenotrophomonas maltophilia* and
819 the *Burkholderia cepacia* complex. J Breath Res 10:037102.

820 84. Zhu J, Bean HD, Jiménez-Díaz J, Hill JE. 2013. Secondary electrospray ionization-mass
821 spectrometry (SESI-MS) breathprinting of multiple bacterial lung pathogens, a mouse
822 model study. J Appl Physiol 114:1544-1549.

823 85. Zhu J, Jiménez-Díaz J, Bean HD. 2013. Robust detection of *P. aeruginosa* and *S.*
824 *aureus* acute lung infections by secondary electrospray ionization-mass spectrometry
825 (SESI-MS) breathprinting: From initial infection to clearance. J Breath Res 7.

826 86. Hahn A, Whiteson K, Davis TJ, Phan J, Sami I, Koumbourlis AC, Freishtat RJ, Crandall
827 KA, Bean HD. 2020. Longitudinal associations of the cystic fibrosis airway microbiome
828 and volatile metabolites: A case study. Front Cell Infect Microbiol 10.

829 87. Hoffman LR, Kulasekara HD, Emerson J, Houston LS, Burns JL, Ramsey BW, Miller SI.
830 2009. *Pseudomonas aeruginosa lasR* mutants are associated with cystic fibrosis lung
831 disease progression. J Cyst Fibros 8:66-70.

832 88. Ren CL, Konstan MW, Yegin A, Rasouliyan L, Trzaskoma B, Morgan WJ, Regelmann
833 W, Scientific Advisory Group I, Coordinators of the Epidemiologic Study of Cystic F.
834 2012. Multiple antibiotic-resistant *Pseudomonas aeruginosa* and lung function decline in
835 patients with cystic fibrosis. J Cyst Fibros 11:293-9.

836 89. Hunter RC, Klepac-Ceraj V, Lorenzi MM, Grotzinger H, Martin TR, Newman DK. 2012.
837 Phenazine content in the cystic fibrosis respiratory tract negatively correlates with lung
838 function and microbial complexity. Am J Respir Cell Mol Biol 47:738-745.

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

843 91. Bragonzi A, Worlitzsch D, Pier GB, Timpert P, Ulrich M, Hentzer M, Andersen JB,
844 Givskov M, Conese M, Doring G. 2005. Nonmucoid *Pseudomonas aeruginosa*
845 expresses alginate in the lungs of patients with cystic fibrosis and in a mouse model. J
846 Infect Dis 192:410-419.

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

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

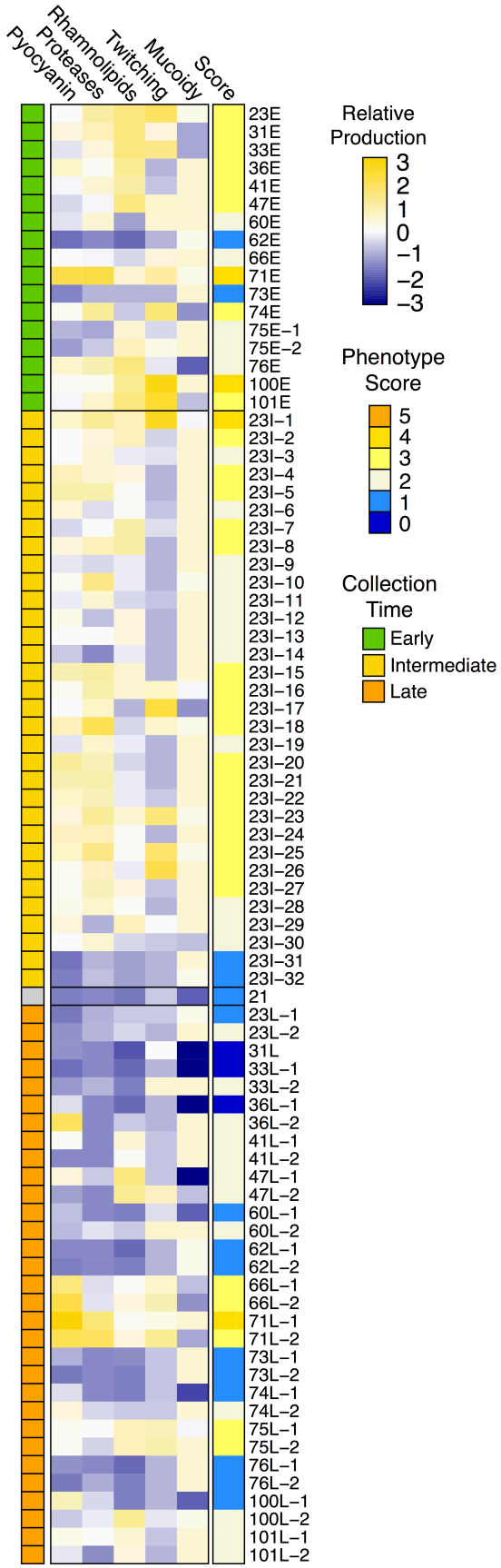
854 94. Harrison F, Muruli A, Higgins S, Diggle SP. 2014. Development of an *ex vivo* porcine
855 lung model for studying growth, virulence, and signaling of *Pseudomonas aeruginosa*.
856 Infect Immun 82:3312-3323.

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

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

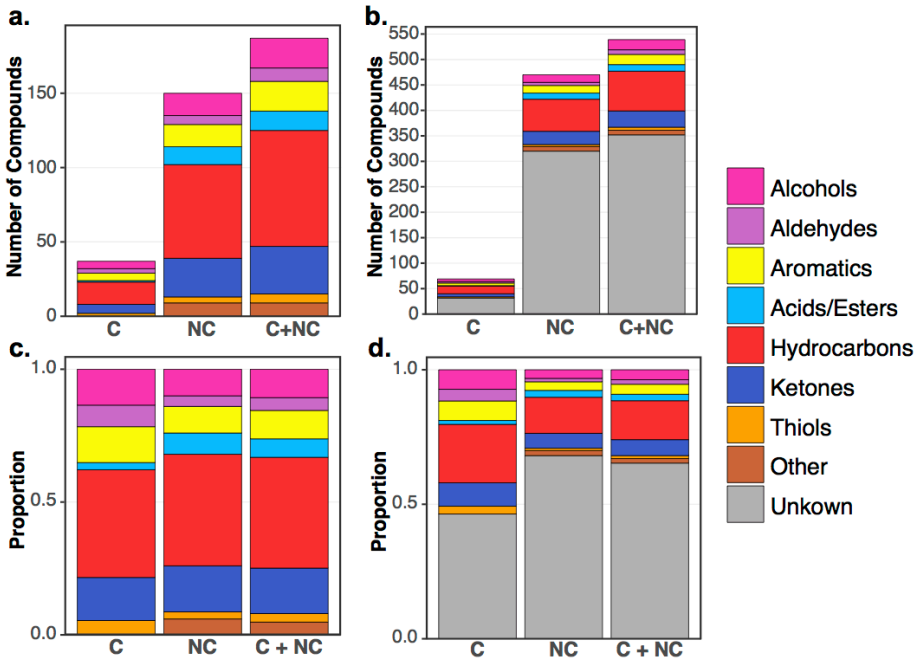
862 97. Sønderholm M, Kragh KN, Koren K, Jakobsen TH, Darch SE, Alhede M, Jensen PO,
863 Whiteley M, Kühl M, Bjarnsholt T. 2017. *Pseudomonas aeruginosa* aggregate formation
864 in an alginate bead model system exhibits *in vivo*-like characteristics. Appl Environ
865 Microbiol 83.

866 98. Bean HD, Jiménez-Díaz J, Zhu J, Hill JE. 2015. Breathprints of model murine bacterial
867 lung infections are linked with immune response. Eur Respir J 45:21-24.
868


Pyocyanin
Proteases
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
Phenotype Score
5
4
3
2
1
0
Collection Time
Early
Intermediate
Late

869

870 **FIG 1** Relative phenotype production of 81 *P. aeruginosa* CF chronic lung infection isolates.
871 Each phenotype was independently scaled to a range of 0 to 1, with 0 representing the
872 minimum data point and 1 representing the maximum data point (excluding outliers), and the
873 relative productions are depicted in the heat map as standardized values (mean-centered and
874 scaled to unit variance) across isolates. Phenotype Score ("Score") is the sum of the scaled,
875 non-standardized phenotype values, rounded to the nearest integer.



876

877 **FIG 2** Size and chemical composition of volatile compounds metabolized by 81 *P. aeruginosa*
878 CF chronic infection isolates. **(a)** Identified compounds; **(b)** Identified and unknown compounds;
879 **(c)** Identified compounds, scaled to 100%; **(d)** Identified and unknown compounds, scaled to
880 100%. C = Core compounds, defined as detected in 95% or more of samples; NC = Non-core
881 compounds; C+NC = sum of Core and Non-core compounds.

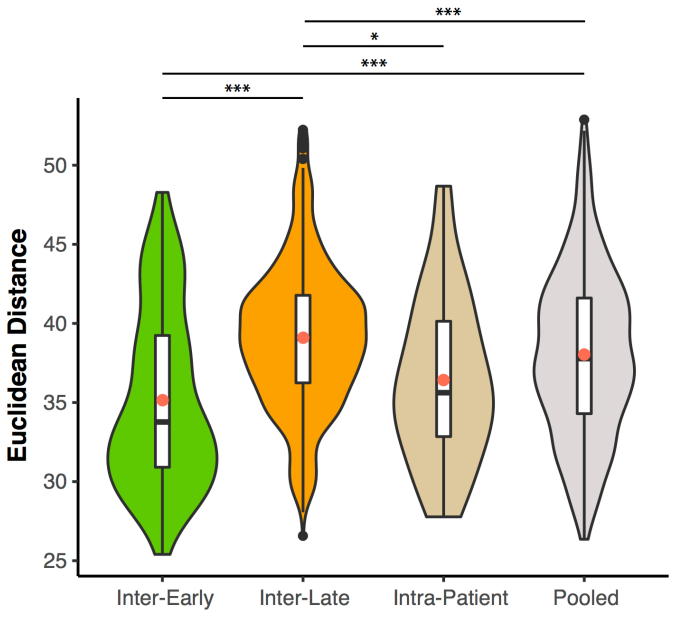


FIG 3 Violin plots and box plots of pairwise dissimilarities (defined as Euclidean distance) of *P. aeruginosa* chronic infection isolate volatilomes. Inter-Early are between-patient Early-Early isolate dissimilarities. Inter-Late are between-patient Late-Late isolate dissimilarities. Intra-Patient are within-patient Early-Late isolate dissimilarities. Pooled represents all pairwise volatilome dissimilarities. Red circles indicate the mean. Significant differences between means were identified using one-way ANOVA with Tukey's HSD. *$P < .05$, **$P < .01$, ***$P < .001$.

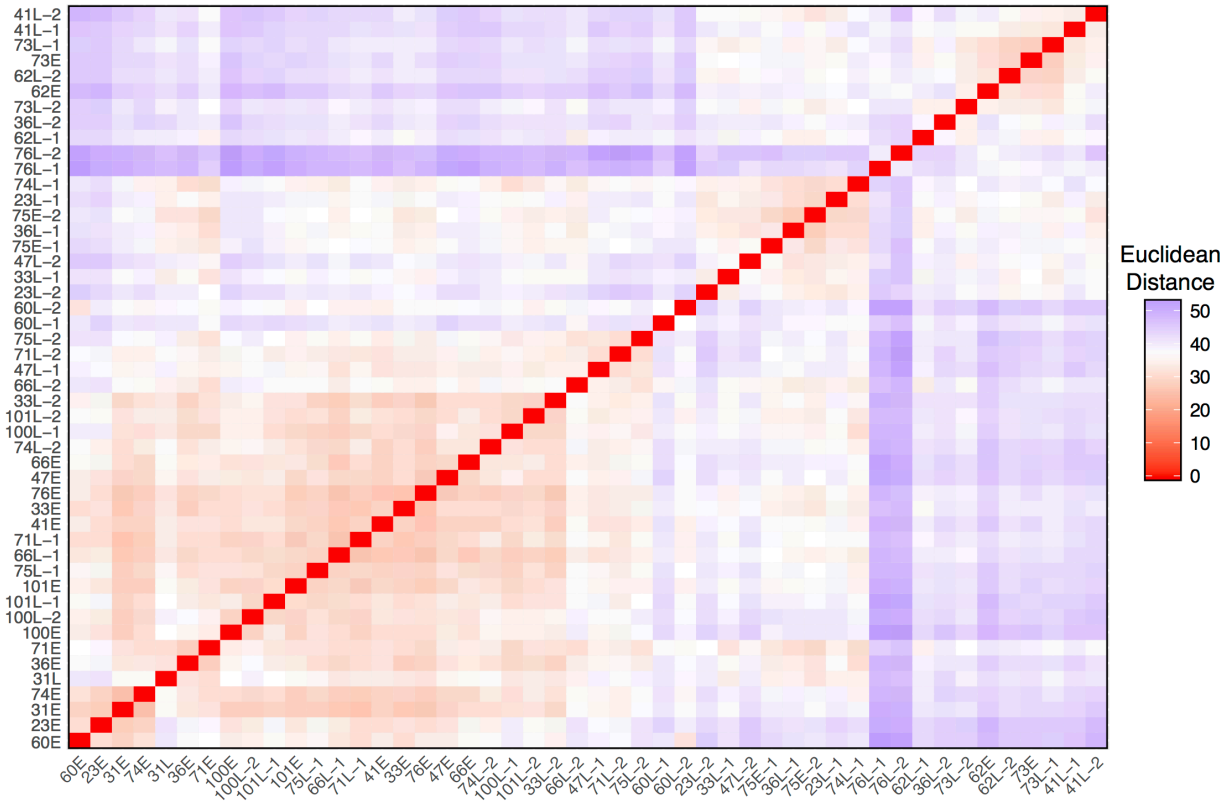


FIG 4 Ordered dissimilarity image (ODI) of the truncated set of 48 *P. aeruginosa* clinical isolates depicting volatilome dissimilarity defined by Euclidean distance.

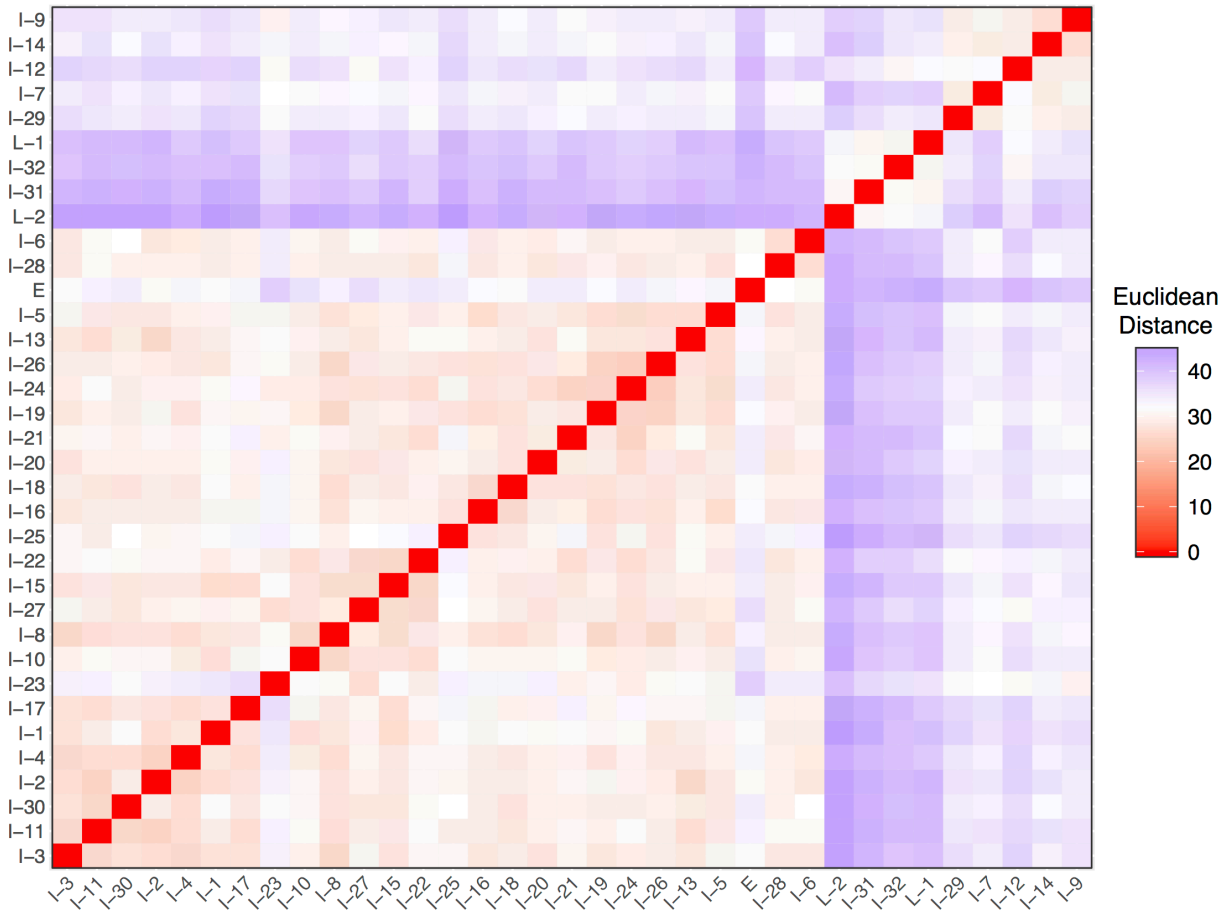


FIG 5 Ordered dissimilarity image (ODI) of the 35 *P. aeruginosa* clinical isolates from Patient 23 depicting volatilome dissimilarity defined by Euclidean distance. For visual clarity, the isolate labels do not include patient number (*e.g.*, 23).

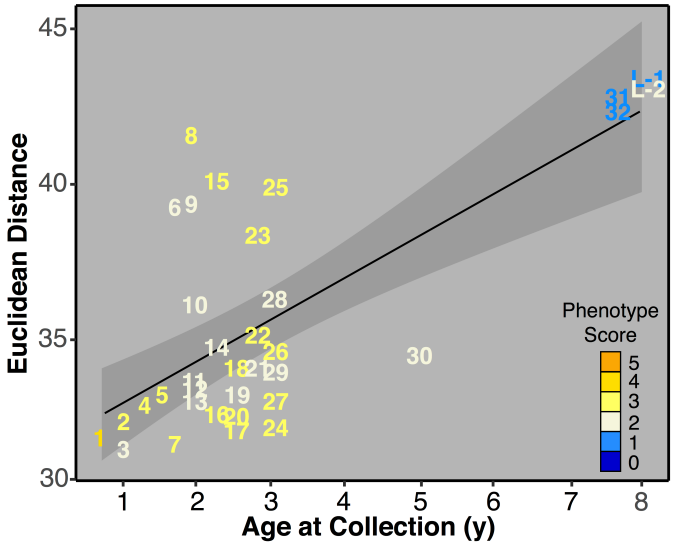
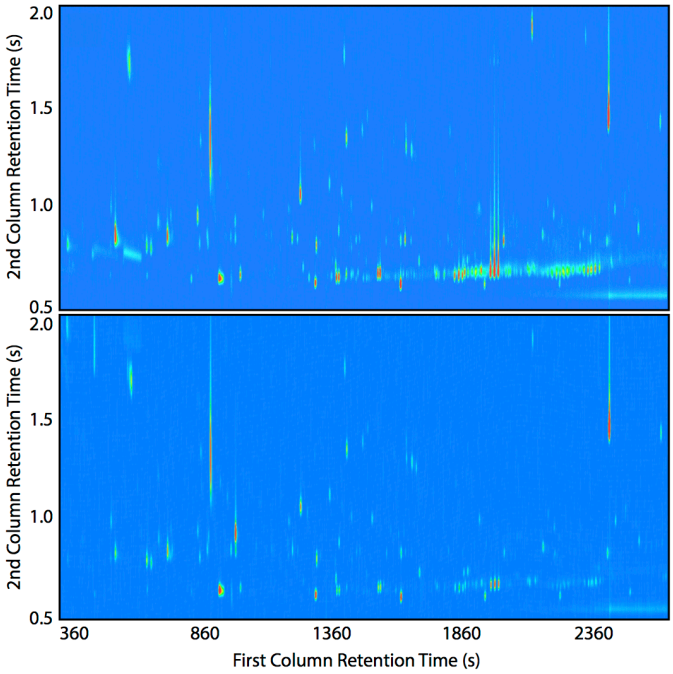
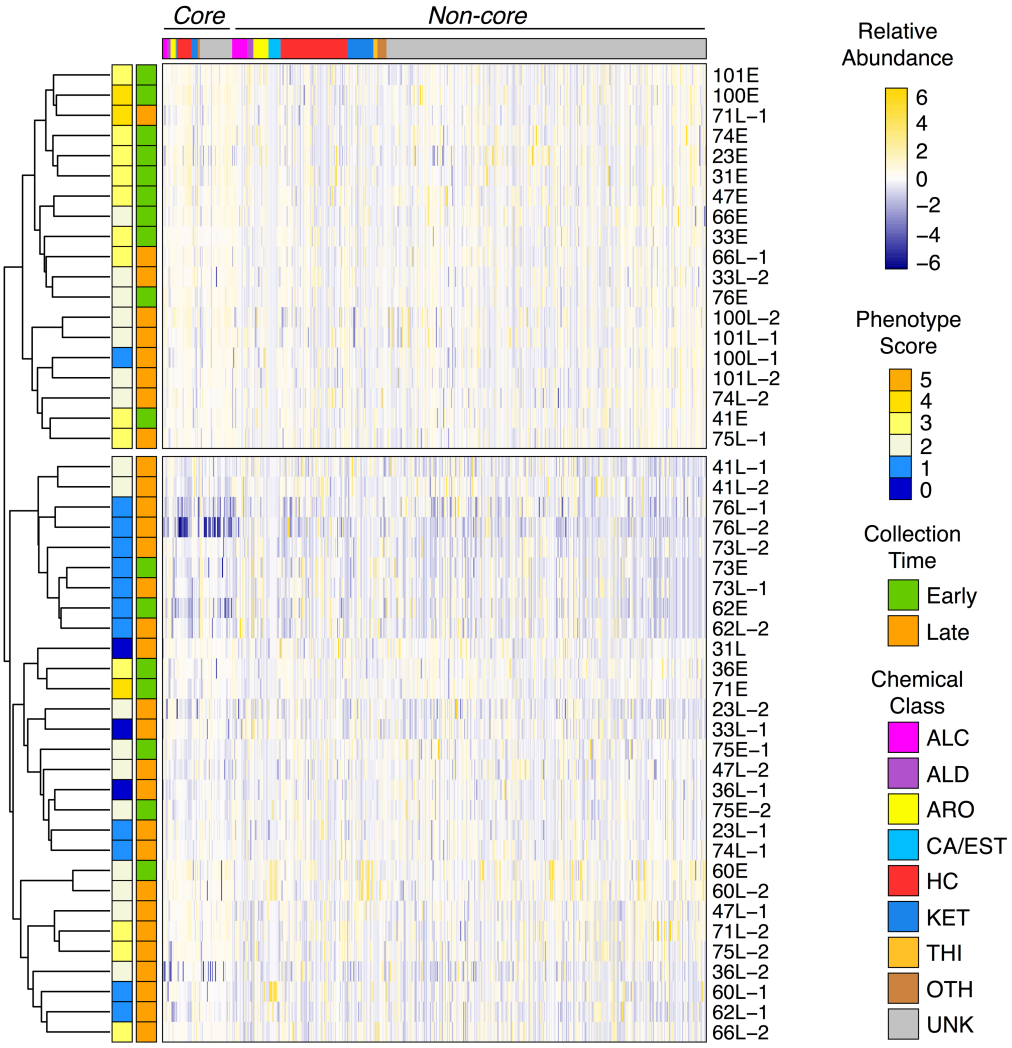


FIG 6 Dissimilarity of Patient 23 volatilomes over time, defined as Euclidean distance from the Early isolate (not shown). For visual clarity, the isolate labels do not include patient number (*e.g.*, 23) and intermediate isolate labels have been truncated to only include the numeral that signifies collection order (*e.g.*, I-1 is represented as 1). The black line represents the linear regression fit (r = 0.70, *P* = 0+), and the shaded region represents the standard error of the regression line.



906 **FIG 7** Representative GC×GC chromatograms of Early (top) and Late (bottom) chronic infection
907 isolates (41E and 41L-2 are depicted). Dark blue represents the baseline, and peak intensity is
908 depicted using a color gradient from light blue (low) to dark red (high). Chromatographic regions
909 of $^{1}t_{R}$ < 200 s and $^{2}t_{R}$ < 0.5 s were excluded for visual clarity. All chromatograms for the ten
910 selected Early and Late isolate pairs are provided in Supplementary Fig. 7.
911



912

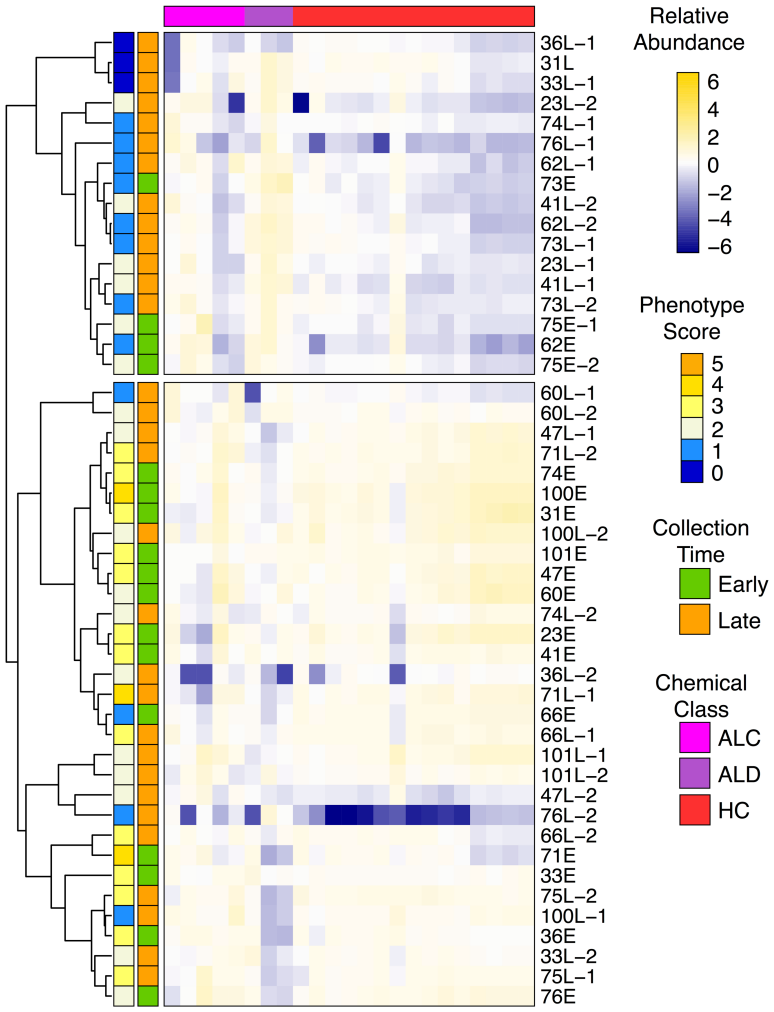
913 **FIG 8** Hierarchical clustering analysis (HCA) of the truncated set of 48 *P. aeruginosa* clinical CF
914 isolates, based on the relative abundance of 539 volatile compounds. Volatiles are in columns
915 (standardized relative abundance). Clustering is based on rows (isolates), which are color-
916 coded by their phenotype score (left color block) and relative time of collection (right color

917 block). ALC = alcohols, ALD = aldehydes, ARO = aromatics, CA/EST = carboxylic acids and

918 esters, HC = hydrocarbons, KET = ketones, THI = thiols, OTH = other compounds, UNK =

919 unknown identity.

920



921

922 **FIG 9** Hierarchical clustering analysis (HCA) of the truncated set of 48 *P. aeruginosa* clinical CF

923 isolates, based on the relative abundance of 23 core alcohols, aldehydes, and hydrocarbons.

924 Volatiles are in columns (standardized relative abundance). Clustering is based on rows

925 (isolates), which are color-coded by their phenotype score (left color block) and relative time of

926 collection (right color block). ALC = alcohols, ALD = aldehydes, HC = hydrocarbons.

927

928 **FIG S1**. Non-metric multidimensional scaling (NMDS) plots of 81 *P. aeruginosa* clinical
929 isolates depicting volatilome dissimilarity defined by Euclidean distance **(a)** colored by
930 collection time and **(b)** colored by patient; **(c)** Ordered dissimilarity image (ODI) of 81 *P.*
931 *aeruginosa* clinical isolates depicting volatilome dissimilarity defined by Euclidean
932 distance.
933
934 **FIG S2**. Non-metric multidimensional scaling (NMDS) plots of the truncated set of 48 *P.*
935 *aeruginosa* clinical isolates depicting volatilome dissimilarity defined by Euclidean
936 distance **(a)** colored by collection time and **(b)** colored by patient; **(c)** NMDS plot of the
937 35 *P. aeruginosa* clinical isolate volatilomes from Patient 23 depicting volatilome
938 dissimilarity defined by Euclidean distance.
939
940 **FIG S3**. Size and chemical composition of volatile compounds metabolized by the
941 selected ten Early-Late pairs of *P. aeruginosa* chronic infection CF isolates. **(a)**
942 Identified compounds; **(b)** Identified and unknown compounds; **(c)** Identified
943 compounds, scaled to 100%; **(d)** Identified and unknown compounds, scaled to 100%.
944 Size and chemical composition of volatile compounds metabolized by each of the ten
945 Early-Late pairs of *P. aeruginosa* chronic infection CF isolates. **(e)** Identified
946 compounds; **(f)** Identified and unknown compounds; **(g)** Identified compounds, scaled to
947 100%; **(h)** Identified and unknown compounds, scaled to 100% C = Core compounds,
948 defined as detected in 95% or more of samples; NC = Non-core compounds; C+NC =
949 sum of Core and Non-core compounds.
950

951 **FIG S4**. Hierarchical clustering analysis (HCA) of the ten Early-Late pairs of *P.*
952 *aeruginosa* clinical CF isolates, based on **(a)** presence and absence and **(b)** relative
953 abundance of 472 volatile compounds. Core compounds are present in at least 95% of
954 isolates; Non-core compounds are present in less than 95%. Volatiles are in columns.
955 Clustering is based on rows (isolates), which are color-coded by their phenotype score
956 (left color block) and relative time of collection (right color block).
957
958 **FIG S5**. Hierarchical clustering analysis (HCA) of the 81 *P. aeruginosa* clinical CF
959 isolates, based on the relative abundance of 539 volatile compounds. Core compounds
960 are present in at least 95% of isolates; Non-core compounds are present in less than
961 95%. Volatiles are in columns (standardized relative abundance). Clustering is based on
962 rows (isolates), which are color-coded by their phenotype score (left color block) and
963 relative time of collection (right color block).
964
965 **FIG S6**. Hierarchical clustering analysis (HCA) of **(a)** the truncated set of 48 and **(b)** full
966 set of 81 *P. aeruginosa* clinical CF isolates, based on the relative abundance of the 69
967 core volatile compounds. Volatiles are in columns (standardized relative abundance).
968 Clustering is based on rows (isolates), which are color-coded by their phenotype score
969 (left color block) and relative time of collection (right color block).
970
971 **FIG S7**. Hierarchical clustering analysis (HCA) of the 81 *P. aeruginosa* clinical CF
972 isolates, based on the relative abundance of 23 core alcohols, aldehydes, and
973 hydrocarbons. Volatiles are in columns (standardized relative abundance). Clustering is

974 based on rows (isolates), which are color-coded by their phenotype score (left color

975 block) and relative time of collection (right color block).

976