

1 **Title: *Stenotrophomonas maltophilia* natural history and evolution in the airways of adults**
2 **with cystic fibrosis**

3
4 Conrad Izydorczyk^a, Barbara J. Waddell^a, Christina S. Thornton^{a,b}, John M. Conly^{a,b,c,d}, Harvey
5 R. Rabin^{a,b}, Ranjani Somayaji^{a,b,c}, Michael G. Surette^{a,e}, Deirdre L. Church^{b,c,d}, Michael D.
6 Parkins^{a,b,c,#}

7 ^a Department of Microbiology, Immunology and Infectious Diseases, Cumming School of
8 Medicine, University of Calgary, Calgary, Alberta, Canada

9 ^b Department of Medicine, Cumming School of Medicine, University of Calgary and Alberta
10 Health Services, Calgary, Alberta, Canada

11 ^c Snyder Institute for Chronic Diseases, Cumming School of Medicine, University of Calgary
12 and Alberta Health Services, Calgary, Alberta, Canada

13 ^d Department of Pathology and Laboratory Medicine, Cumming School of Medicine, University
14 of Calgary and Alberta Health Services, Calgary, Alberta, Canada

15 ^e Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton,
16 Ontario, Canada

17

18

19

20

21 **Running Title: *S. maltophilia* infection in cystic fibrosis**

22

23

24

25

26 Please address correspondence to:

27 Dr. Michael D. Parkins,

28 Professor, Department of Medicine and Microbiology, Immunology & Infectious Diseases,
29 Cumming School of Medicine, University of Calgary

30 3330 Hospital Drive NW, Calgary, AB CANADA T2N 4N1

31 mdparkin@ucalgary.ca

32

33

34

35

36
37 Abstract Word Count: 250 words
38
39 Importance Word Count: 116
40
41 Article Word Count: 4938 (3831 excluding Materials and Methods)

42 **ABSTRACT**

43 *Stenotrophomonas maltophilia* is an opportunistic pathogen infecting person with cystic fibrosis
44 (pwCF) and portends a worse prognosis. Studies of *S. maltophilia* infection dynamics have been
45 limited by cohort size and follow-up. We investigated the natural history, transmission potential,
46 and evolution of *S. maltophilia* in a large Canadian cohort of pwCF over a thirty-seven year
47 period. *S. maltophilia* was recovered at least once in 25.5% of the cohort. Yearly isolates from 74
48 pwCF (23%) were typed by pulsed-field gel electrophoresis, and shared pulsotypes underwent
49 whole-genome sequencing. Most pwCF were infected by unique strains, but serial infections
50 with different strains, and strains shared between patients, were observed. In chronic carriage,
51 longer time periods between positive collection dates increased the likelihood that subsequent
52 isolates were unrelated. Isolates from individual pwCF were largely clonal, with genetic diversity
53 driven by gene content differences. Disproportionate progression of CF lung disease was not
54 observed amongst those infected with multiple strains over time (versus a single) or amongst
55 those with shared clones (versus strains only infecting one patient). We did not observe evidence
56 of patient-to-patient transmission despite relatedness between isolates. Instead, genomic analyses
57 suggested common, indirect sources as their origins. Sixteen multi-mutated genes were identified
58 as having a potential role in adaptation of *S. maltophilia* to CF, including in a regulator of an
59 efflux pump and in an iron acquisition gene cluster. The information derived from a genomics-
60 based understanding of the natural history of *S. maltophilia* infection within CF provides unique
61 insight into its potential for in-host evolution.

62

63 **IMPORTANCE**

64 In this largest and longest single center study of *S. maltophilia* causing infections in persons with
65 cystic fibrosis, we concluded that patient-to-patient infection transmission had not occurred. We
66 determined that infection by a new *S. maltophilia* strain was more likely the longer the time
67 between its recovery in sputum, suggesting infection by individual strains is generally short-
68 lived. Amongst bacterial isolates belonging to the same clonal complex, isolates could be better
69 differentiated by their gene content than mutations, suggesting gene gain/loss may contribute
70 more to the genetic diversity of these strains than mutation. Infection by multiple strains, or a
71 shared strain found in at least one other person, was not associated with progression to end-stage
72 lung disease.

73 **Introduction**

74 *Stenotrophomonas maltophilia* is an opportunistic Gram-negative pathogen increasingly
75 recognized for its potential to cause a variety of human infections(1), particularly among
76 immunocompromised individuals such as those with cystic fibrosis (CF). The prevalence of *S.*
77 *maltophilia* in persons with CF (pwCF) has increased in recent decades(2, 3), with chronic
78 infections associated with adverse clinical outcomes, including increased pulmonary
79 exacerbation frequency, hospitalization, requirements for intravenous antibiotic treatments (4, 5),
80 poorer baseline health(4, 6), variably accelerated lung function decline(4, 7–9), and a higher risk
81 of progression to end-stage lung disease(10).

82 Several studies have investigated the natural history of *S. maltophilia* infection in CF(7, 11–15).
83 In general, *S. maltophilia* infection appears to be short-lived in many individuals(16). However,
84 in those with chronic infection (defined as ≥ 2 positive cultures within a year after initial
85 acquisition), infection by multiple genotypes over time is common(11–14). In contrast, co-
86 infection by multiple sub-strains and cross-infection of multiple patients by the same strain are
87 infrequent(11, 16–18). Moreover, an initial infecting *S. maltophilia* strain may adapt to the CF
88 lung environment, diversifying under selective pressures into highly successful, antibiotic
89 resistant sub-lineages(13).

90 Studies of *S. maltophilia* natural history in CF are limited, however, in their inclusion of
91 relatively small numbers of pwCF, with a focus on those with chronic infection, short duration of
92 follow-up, and use of low-resolution molecular typing methods(11, 12, 15); only a single study
93 used whole-genome sequencing for strain assessment(14). Further, none of these studies
94 investigated the potential for *S. maltophilia* to spread between pwCF.

95 Herein, we performed a retrospective investigation of the natural history and potential for clinic-
96 associated, patient-to-patient transmission of *S. maltophilia* at a greater resolution and across a
97 large cohort. We drew on the Calgary Adult CF Biobank, which includes every isolate from
98 every clinical encounter from the entire CF cohort attending the clinic. We hypothesized that
99 strain replacement events would be common and that shared strains would be observed only in a
100 minority of patients, but that clinic-associated transmission was not the source of new acquisition
101 and infections.

102 RESULTS

103 Study and Sample Population

104 Between 1979-2016, 321 individuals with CF lung disease were followed by the Southern
105 Alberta CF Clinic, representing 2640.64 person-years of observation. Characteristics of the
106 cohort are summarized in **Error! Reference source not found.**. During the study, eighty-two of
107 the 321 pwCF (25.5%) had ≥ 1 *S. maltophilia* positive sputum cultures collected. The prevalence
108 of pwCF with ≥ 1 *S. maltophilia* positive sputum cultures in any given 5- and 1-year windows
109 during the study period was 16.2% (IQR 12.9%-17.6%) and 8.74% (IQR 4.48%-10.3%),
110 respectively. Of these individuals, 23 (28%) had only one isolate collected, while 14 pwCF
111 (17.1%) had two positive cultures, and 45 had ≥ 3 (54.9%). The median number of positive
112 cultures in each pwCF of the cohort was three (IQR 1-5, range 1-65). Within the biobank, a total
113 of 424 cultures and 447 individual *S. maltophilia* isolates of different morphological appearance
114 (median 1, mean 1.05 isolates/culture date, range 1-2) were present. The natural history of
115 patients with ≥ 1 *S. maltophilia* isolates isolated by routine clinical microbiology testing over the
116 duration of their attendance at the clinic is displayed in **Error! Reference source not found.**

117 For each pwCF, we collected their first and last *S. maltophilia* isolates, as well as intermediate
118 isolates collected 1-3 years apart, and performed typing by PFGE. A total of 162 isolates were
119 typed from 74/82 pwCF (90.2%) (median 1 isolate/pwCF, range 1-15). Isolates from the
120 remaining eight pwCF were either not recoverable by culture or missing from the biobank. These
121 pwCF did not differ by age, sex, dF508 homozygosity, pancreatic insufficiency status, or
122 ppFEV1 at incident isolate(s) from those with typed isolates but were more likely to have only
123 one *S. maltophilia* positive sputum culture (7/8 vs. 16/74 patients with one isolate, Fisher's exact
124 test $p=0.0004$). Of those with typed isolates, thirty-six pwCF (48.6%) had ≥ 2 isolates typed,
125 spanning a collective 397.5 person-years of observation.

126 Bacterial Strain typing

127 A variety of pulsotypes were identified amongst recovered isolates (**Error! Reference source**
128 **not found.** and 2). Most pwCF (53/74, 71.6%) were infected by only a single pulsotype, but
129 infection over time by different pulsotypes over time was also observed. The median number of
130 pulsotypes/patient was 1 (IQR 1-2), but fifteen (20.3%), four (5.4%), and two (2.7%) pwCF had
131 infection by two, three, or four pulsotypes, respectively. Most pulsotypes were detected in one or
132 a few sputum cultures and were transient or quickly replaced. Indeed, in most pwCF with a
133 history of multiple pulsotypes, sputum cultures negative for *S. maltophilia* growth separated
134 cultures with different pulsotypes, suggesting clearance of the initial strain and subsequent
135 infection by a new pulsotype, rather than competitive replacement of a prior strain by a more fit
136 new strain. Four (5.4%) instances of prolonged (>1 year), continuous infection by a single or
137 multiple pulsotypes were also observed: pwCF A055's second pulsotype (1.4 years), A130's
138 single pulsotype (1.9 years), A176's single pulsotype (2.8 years), A362's single pulsotype (3.6
139 years), and A357's two pulsotypes (2.9 years). However, recovery of a given pulsotype over a
140 prolonged period of time was also observed with negative cultures in between the positive

141 cultures (e.g., patients A006 and A052), suggesting clearance and re-infection by a common
142 strain rather than continuous infection.

143 Recovery of a new pulsotype after detection of a prior pulsotype was significantly associated
144 with time between typed cultures (**Error! Reference source not found.**). When considering
145 intervals of greater vs. less than one, two, and five years, the odds of a new pulsotype relative to
146 a prior pulsotype increased with increasing time between typed cultures. Similarly, the risk of a
147 new pulsotype was always greater in longer than shorter intervals, and the relative risk increased
148 when comparing intervals >1 to >2 years. A slight decrease in the relative risk was observed
149 between the >2 years to >5 years comparisons (3.38 vs. 3.14, respectively), but this was likely
150 due to most pwCF having both new/prior pulsotypes identified sooner than by five years.

151 Recovery of a prior pulsotype after detection of another was rare and observed in only four
152 (5.4%) pwCF (**Error! Reference source not found.** pwCF A057, A090, A145, and A357). In all
153 cases, re-recovery of the prior pulsotype occurred <1 year after detection of the new pulsotype.
154 However, in pwCF A057, A090, and A145, at least one *S. maltophilia* negative culture separated
155 their second pulsotype from the re-recovery of their first pulsotype, whereas there were no
156 additional cultures (either positive or negative) between the second pulsotype and re-recovery of
157 the first in pwCF A357.

158 Most pwCF (63/74) were infected with pulsotypes unique to themselves, but five pulsotypes
159 were shared among ≥ 2 pwCF, corresponding to five shared STs that were identified (**Error!
160 Reference source not found.**). These five STs corresponded to a collective thirty-three isolates
161 from ten pwCF and underwent WGS, along with seventeen isolates from five non-shared STs to
162 serve as non-shared controls (**Error! Reference source not found.**). An additional isolate from
163 a unique pulsotype was sequenced as it belonged to a patient (A055) with many isolates in
164 shared ST-199; it was subsequently found to also be ST-199. Three further sequenced isolates
165 (SM055, SM111, and SM113) initially identified as belonging to shared pulsotypes did not
166 match any shared STs, for a total of 54 isolates sequenced.

167 Sequenced STs and isolates appeared to be a random sample from the global pool of *S.*
168 *maltophilia* diversity and were scattered among the twenty-three previously identified
169 monophyletic lineages of the species complex(19) (Figure 1). However, most STs and isolates
170 belonged to the Sm6 lineage – a lineage previously reported to harbor most of the strains of *S.*
171 *maltophilia* recovered from humans(19). Within their respective lineages, isolates clustered by
172 ST and multiple STs were observed to comprise some lineages. The number of SNPs separating
173 STs ranged from 10^4 - 10^5 , dependent on intra- vs. inter-lineage comparisons. In contrast, the
174 number of SNPs separating intra-ST isolates was up to three orders of magnitude lower (10^2 - 10^3
175 SNPs). A similar trend was observed with wgMLST allele distances, with inter-ST distances (on
176 the order of 10^3 alleles) being an order of magnitude greater than intra-ST distances.

177 **Intra-Patient Genetic Diversity and Clonal Relatedness**

178 Isolates from the same individual and of the same ST were mostly clonal in SNP and wgMLST
179 allele phylogenies (**Error! Reference source not found., Error! Reference source not found.,**
180 and Supplementary Figure 3), forming monophyletic clades with relatively small intra-clade SNP

181 and wgMLST allele distances. Evidence of a molecular clock was also observed among some
182 STs (e.g., Supplementary Figure 3 panels d and e) with isolates collected later in time more
183 distantly related to their common ancestor than earlier isolates. Only a single instance of non-
184 monophyletic clustering of a patient's isolates was noted. In this case, one of pwCF A055's ST-
185 199 isolates (SM003, pulsotype F) was localized in a separate clade with three of A090's isolates
186 and was hundreds of SNPs/alleles distant from A055's other isolates (all pulsotype C). SM003,
187 however, had a different pulsotype and was collected two years prior to the earliest of A055's
188 other isolates, and was no more or less distant to isolates from other pwCF than it was to A055's,
189 further suggesting it was a separate lineage, independently acquired and not an ancestor of
190 A055's later isolates. A second instance of a possible non-clonal relatedness was observed with
191 patient A344's SM137 (ST-365) isolate, which differed by more than 50 SNPs/alleles to their
192 other ST-365 isolates despite all isolates having been collected within one year. Excluding the
193 former case, intra-patient pairwise SNP and wgMLST allelic distances ranged from 0-58 SNPs
194 (median 10, IQR 7-22) and 2-54 alleles (median 16, IQR 9-23), respectively. Inter-pwCF SNP
195 and allele distances ranged from 54-478 SNPs (median 253, IQR 200-258.5) and 56-461 alleles
196 (median 260, IQR 206.5-269.5), respectively.

197 A median of 4472.5 coding sequences (IQR 4311.75-4472.5) were annotated per isolate genome.
198 Intra-pwCF isolates exhibited greater variability in relatedness with respect to differences in gene
199 content than SNPs/wgMLST alleles and could be as different from one another as to isolates
200 from another pwCF. The pairwise number of genes present/absent among intra-pwCF isolates
201 ranged from 6-608 genes (median 86, IQR 33.5-289), whereas the inter-pwCF equivalent was
202 311-1195 genes (IQR 4311.75-4472.5). Clustering patterns based on gene presence/absence also
203 recovered the clonal relationships observed in SNP/wgMLST allele phylogenies but with longer
204 branches between even closely related isolates (**Error! Reference source not found.**).

205 **Transmission**

206 The potential for patient-to-patient transmission of *S. maltophilia* among patients within shared
207 STs was simultaneously investigated using four complementary analyses: SNP/wgMLST
208 distances, phylogenetics, gene content analysis, and six-month carriage overlap. Collectively,
209 sixteen pairs of patients were identified among the five shared STs (median one patient pair per
210 shared ST, range 1-10), for eleven of which ≥ 2 isolates were available for at least one pwCF in
211 the pair (**Error! Reference source not found.** and **Error! Reference source not found.**). Of the
212 sixteen pairs of patients analyzed, nine pairs had no analyses supporting transmission, and seven
213 pairs had one analysis supporting potential for transmission. In no pairs of patients was support
214 for transmission provided by ≥ 2 analyses.

215 ST-specific SNP and wgMLST allele distances were smaller among intra-pwCF isolate pairs
216 than inter-pwCF isolate pairs in all but one case. Specifically, patient A057's single isolate (ST-
217 199) was similarly distant to A055's isolates as the observed SNP distances between some of
218 patient A344's isolates (ST-365). However, these latter distances likely represent a separate,
219 distinctly acquired sub-strain in this patient or a hypermutating strain with SNP distances greater
220 than typically expected between non-hypermutating isolates, and this overlap of distances was
221 not observed among wgMLST allele distances. Phylogenetic support for transmission was not

222 observed for any isolate pairs (**Error! Reference source not found.**). In all cases where ≥ 2
223 isolates were available for at least one patient in a given pair, isolates clustered by patient with
224 shorter branches to other isolates from the same patient than to isolate(s) from other patients.
225 Similarly, gene content support was not observed, as hierarchical clustering recovered the same
226 clonal relationships as phylogenetic analysis, albeit with longer branches between some intra-
227 patient isolates (**Error! Reference source not found.**). Carriage support was the most common
228 and observed for six pairs of isolates.

229 **Multi-mutated Genes**

230 The accumulation of multiple independent mutations in a given gene may be an indicator of
231 adaptive pressures from CF-associated infections acting on that gene, as may a higher ratio of
232 non-synonymous to synonymous mutations(20, 21).

233 One-hundred ninety-eight protein coding genes across all analyzed STs (see methods) were
234 identified with mutations arising during infection in CF (“CF genes”); 1042 genes had mutations
235 acquired outside of CF infection (“non-adaptive genes”). Mutations in CF genes were more
236 likely to be both non-synonymous (Fisher’s exact test adjusted $p=0.00648$) and stop mutations
237 (adjusted $p=0.00309$) than synonymous, compared to mutations in non-adaptive genes.
238 Enrichment analysis did not identify any gene ontology (GO) categories significantly associated
239 with CF vs. non-adaptive genes, however.

240 Sixteen CF genes were multi-mutated across STs, two were multi-mutated across and within
241 STs, and six were multi-mutated only within STs (**Error! Reference source not found.**).
242 Neither multi-mutated CF nor non-adaptive genes were more likely to have non-synonymous or
243 stop mutations than synonymous mutations compared to their non-multi-mutated counterparts,
244 respectively (Fisher’s exact test unadjusted and adjusted $p>0.05$). Similarly, mutations in CF
245 multi-mutated genes were not more likely to be non-synonymous (unadjusted and adjusted
246 $p>0.05$) nor stop mutations, but a trend was observed among stop mutations (unadjusted
247 $p=0.004$, adjusted $p=0.059$). When CF multi-mutated genes were separated into across-ST and
248 within-ST subcategories and compared, a trend favoring non-synonymous (unadjusted $p=0.0039$,
249 adjusted $p=0.055$) but not stop (unadjusted and adjusted $p>0.05$) mutations was observed.
250 Neither across-ST nor within-ST multi-mutated CF genes were more likely to have non-
251 synonymous or stop mutations than synonymous mutations compared to non-multi-mutated CF
252 genes, although trends were observed among non-synonymous mutations in the across-ST
253 (unadjusted $p=0.0650$, adjusted $p>0.05$) and within-ST (unadjusted $p=0.0183$, adjusted $p>0.05$)
254 groups, and among stop mutations in the across-ST group (unadjusted $p=0.0771$, adjusted
255 $p>0.05$).

256 Multi-mutated regions included a variety of genes and intergenic regions (Supplementary Table
257 2). Notably, two multi-mutated intergenic regions were clustered around the same set of genes
258 involved in iron acquisition (an outer membrane hemin receptor and hemin uptake protein
259 HemP/HmuP), both of which were also (singly) mutated in CF. Multi-mutated genes included
260 genes associated with efflux transporters, basic metabolism, protein transport, virulence, and
261 hypothetical proteins. For example, the *smeT* gene had acquired two independent non-

262 synonymous mutations, including a Leu166Gln mutation found in nine isolates from a single
263 patient previously identified and associated with overexpression of the *smeDEF* efflux
264 pump(22).

265 **Clinical Outcomes**

266 The 82 pwCF with at least one *S. maltophilia* positive sputum culture were followed for a
267 median of 10.1 years (IQR 6.2 to 17.6 years). During this time, 54 pwCF (65.9%) progressed to
268 end-stage lung disease (defined as ppFEV₁ <40). Amongst those with advanced lung disease, 23
269 (28.0%) required lung transplantation. In total, 39 (47.6%) died during the study period. PwCF
270 who succumbed to end-stage lung disease or received transplants were not more likely to have
271 been infected with multiple strain types as compared to those who had stable lung function (14
272 [25.9%] vs. 6 [21.4%], p=0.79). Patients who were infected with a shared clone (≥ 2 patients)
273 were not more likely progress to end stage lung disease as compared to those with stable lung
274 function (6 [11.1%]vs 4 [14.3%], p=0.73). In particular ST5, infecting 5 individuals did not
275 portend a worse prognosis (p=0.83).

276

277 DISCUSSION

278 We retrospectively analyzed a large, comprehensive collection of *S. maltophilia* isolates from all
279 pwCF attending the Southern Alberta Adult CF Clinic collected over thirty-seven years in order
280 to understand the natural history of infection and potential for pwCF-pwCF infection
281 transmission. Approximately a quarter of pwCF attending the clinic had ≥ 1 *S. maltophilia*
282 positive sputum culture over the study duration, while the average percentages of pwCF with
283 positive cultures within any given 5- and 1-year windows were 16.2% and 8.74%, respectively.
284 This total proportion of pwCF is greater than that reported in some(23) but not other(24) studies.
285 While most patients had infection with only a single strain, detection of multiple *S. maltophilia*
286 strains over time was common, as previously reported(11–14), but co-infection was not. And
287 while others have reported infection with *S. maltophilia* portends a worsened prognosis relative
288 to those uninfected(4–10), we did not observe differences in our cohort based on whether a share
289 was shared or unique to a single individual, or whether pwCF carried multiple strain types over
290 time versus were only ever infected with a single strain type.

291 Individual strains were mostly clonal, with SNP and wgMLST allele distances consistent with
292 close relatedness. However, even clonal isolates could often be differentiated by their gene
293 content, suggesting that the gain/loss of genes may contribute more to the genetic diversity of
294 these strains than mutation. Consistent with our hypothesis, most patients carried unique strains,
295 and while shared, genetically closely related strains were observed in some pwCF, patient-to-
296 patient associated transmission, and infection within the healthcare system, was considered
297 unlikely due to a lack of supporting evidence.

298 A limitation of current studies of *S. maltophilia* in CF is their inclusion of relatively small
299 numbers of patients (typically only those chronically infected) and short study periods (with
300 infrequent studies extending up to ten years(14)). Thus, comprehensive longitudinal clinic-wide
301 assessments of *S. maltophilia* infection in CF are lacking. Further, most studies have used
302 traditional molecular strain typing methods (rep-PCR and pulsed-field gel electrophoresis
303 (PFGE)) for strain assessment(11, 12, 15), with only a single study using whole-genome
304 sequencing (WGS) on multiple chronically infected patients(14). This latter point is particularly
305 relevant, as many studies have identified a significant proportion of patients with shared strains
306 (as defined by molecular methods). While shared strains as determined through molecular
307 methods may indicate the potential for infection transmission(25, 26), it is not sufficient to
308 identify a transmission event(27–30). This is key, as independent acquisition of the same strain
309 without a CF intermediary is well known to occur with other CF pathogens(26, 28, 29),
310 confounding our ability to understand infection transmission. To date, no studies of *S.*
311 *maltophilia* in CF have investigated its potential to spread between patients.

312 By utilizing the Calgary Adult CF Clinic Biobank – a unique, one-of-a-kind resource – we were
313 able to provide a broad picture of *S. maltophilia* infection dynamics, genetic diversity, and
314 potential for clinic-associated patient-to-patient infection transmission across an entire CF clinic
315 over a period of 37 years. While previous studies of *S. maltophilia* in CF focused on detailed
316 analyses of many isolates from individual patients(12, 13), utilized molecular methods as a
317 baseline for strain typing(11, 15), or focused on relatively small numbers of patients over short

318 timeframes(14), we demonstrated the pertinence of their findings to the entire clinic level. At the
319 same time, we were able to achieve a finer resolution in the patterns and relationships of
320 infecting strains compared to previous works(14) by analyzing sequenced isolates in an ST-
321 specific manner. Indeed, it is now well recognized that the choice of reference genome in SNP
322 calling-based studies is critical and that single-reference analyses are inadequate(31).

323 Recently, *S. maltophilia* has been suggested to exist as a species complex consisting of 23
324 “species-like lineages”(19). This may partially explain the high level of strain diversity and rapid
325 changes in infecting strain type observed in this work, since a very diverse pool of potentially
326 infectious strains exists under the same species classification. Our results are in agreement with
327 the previous finding that detection of multiple *S. maltophilia* strains over time is common(11–14)
328 and extend previous studies by demonstrating that this pattern may persist for several decades.
329 This pattern of rapid strain acquisition is also consistent with a hypothesis of independent
330 environmental acquisition as the source of new infections in pwCF in CF cohorts with adequate
331 infection control protocols, as has been suggested for other CF pathogens(26, 29, 32). This is
332 further supported by a lack of epidemiological evidence for infection transmission, and the clonal
333 nature of infecting strains. The proportion of pwCF with shared strains here is consistent with
334 previous studies(14). The clonal nature of infecting strains is also in line with observations of
335 other CF pathogens(20). No evidence of the circulation of any epidemic strains was observed,
336 unlike what has been observed in some strains of *P. aeruginosa*, *Burkholderia cenocepacia*, and
337 *Mycobacterium abscessus massiliense*(27, 33).

338 Since most intra-pwCF isolates were clonal with limited SNP and wgMLST allele diversity but
339 could differ from one another to the same degree as from isolates from different pwCF with
340 respect to gene content, our data suggests that gene gain/loss may be a stronger driver of *S.*
341 *maltophilia* evolution in CF. Indeed, it has been suggested that *S. maltophilia* as a species
342 evolves primarily via recombination and gene gain/loss(34), and our data supports this to be the
343 case in CF as well. However, our analysis of mutations arising during infection in CF found that
344 these were enriched in non-synonymous and stop codon-introducing mutations compared to
345 mutations separating strains prior to their introduction to the CF airways, suggesting that at least
346 some of these genes may be under adaptive pressure(21). While not statistically significant,
347 similar mutational spectral trends were observed for CF genes with multiple mutations across
348 STs and within STs as well. Indeed, multiple independent mutations at a given locus may be
349 indicative of adaptive pressure on the locus(35–37), which we observed in sixteen loci across
350 STs and eight loci within STs, further suggesting that selection acting on mutations is also
351 present within these strains.

352 We recognize several limitations of this work. Firstly, as a single-center retrospective analysis,
353 we were limited to previously sampled isolates at a single Canadian clinic, with varying numbers
354 sampled from different patients. Thus, some patients may have had denser sampling than others
355 based on frequency of healthcare encounters. Moreover, given the magnitude of the collection in
356 the Calgary Adult CF Clinic Biobank, only one *S. maltophilia* isolate per morphologically
357 distinct colony type is stored per sputum culture. As such, we were limited to a single
358 representative isolate and unable to measure strain diversity at any single point in time within a

359 given sputum culture. In some cases, this meant that only a single isolate was available for a
360 given pwCF, limiting the types of phylogenetic relationships that could be observed for inferring
361 transmission. The magnitude of the collection in the Calgary Adult CF Clinic Biobank also
362 meant that we had to employ strict selection criteria for isolate typing so that not every *S.*
363 *maltophilia* isolate was typed by PFGE. Likewise, sequencing of only isolates belonging to
364 shared strains means that any inherent evolutionary differences between shared and non-shared
365 strains were not analyzed. The draft nature of genome sequencing performed also means that the
366 gene content of sequenced isolates may not be perfectly known.

367 In conclusion, we have demonstrated that *S. maltophilia* infection in pwCF are a random draw
368 from the broader *S. maltophilia* species complex diversity. Infection within individual pwCF is
369 driven by unique strains that are likely of environmental origins, as observed with other CF
370 pathogens. While some patients may carry genetically related strains, these do not appear to be
371 associated with patient-to-patient transmission but more likely acquired from indirect
372 environmental sources. The infection process is largely clonal at the SNP level, but significant
373 diversity is present and driven by differences in gene content within strains.

374 **METHODS**

375 **Patient Population and Strains**

376 In this retrospective single-center cohort study, we analyzed longitudinally collected *S.*
377 *maltophilia* isolates from pwCF attending the Calgary Adult CF Clinic (CACFC) between 1979-
378 2016. All patients attending this clinic undergo routine sputum testing quarterly and as required
379 clinically. Each pathogen recovered from real-time clinical investigations is frozen at -80°C and
380 included in the CACFC Biobank. Each distinct colony morphotype of each pathogen is collected
381 and frozen, separately.

382 Inclusion criteria for pwCF in this study included a confirmed diagnosis of CF (38), aged ≥ 18
383 years, and ≥ 1 *S. maltophilia* positive sputum cultures collected. Patients entering the cohort who
384 had received a life-saving lung transplant were excluded. PwCF who entered the cohort prior to
385 the transplant were censored at the time of the transplant. This study received approval from the
386 University of Calgary's Conjoint Health Research Ethics Board (REB-15-2744).

387 **Bacterial Strain Typing**

388 To assess for strain diversity and relatedness given the magnitude of samples in the CACFC
389 Biobank, representative yearly *S. maltophilia* isolates from all pwCF with ≥ 1 *S. maltophilia*
390 positive sputum cultures (one morphotype per sputum culture) were typed by pulsed-field gel
391 electrophoresis (PFGE) using protocols adapted from Parkins et al.(39). For all pwCF with ≥ 2
392 positive sputum cultures, we included the first, last, and yearly intermediate isolates when viable.
393 Pulsotypes differing by ≤ 3 bands with $\geq 80\%$ similarity were considered to potentially represent
394 the same strain(40). Shared pulsotypes were defined as those representing the same strain and
395 found in ≥ 2 pwCF.

396 Two groups of isolates were selected for whole genome sequencing (WGS): *i*) isolates belonging
397 to all shared strains and *ii*) isolates belonging to a select number of non-shared strains (i.e.,
398 present in only a single patient). The former was sequenced to assess for potential transmission
399 between patients; the latter were selected as a comparison set to allow for the observation of
400 intra-patient genetic distances in the absence of infection transmission. In total, 34 isolates
401 belonging to shared pulsotypes and 17 isolates from five non-shared multi-locus sequence typing
402 (MLST) sequence types (STs) were sequenced. Genomic DNA was extracted using the Promega
403 Wizard® Genomic DNA Purification Kit. Genomic libraries were prepared using the Nextera
404 XT DNA Library Prep Kit and sequenced using either an Illumina HiSeq (2x250 bp reads) or
405 MiSeq (2x300 bp reads) instrument.

406 **Public Genomes (Lineages Analysis)**

407 Publicly available *S. maltophilia* genomes were used to supplement those of our clinic cohort to
408 better understand the placement of our genomes amongst the *S. maltophilia* species complex
409 (Supplementary Table 3). Ten genomes from each of the 23 *S. maltophilia* lineages identified by
410 Groschel et al.(19) were downloaded and processed.

411 **Bioinformatic Analyses**

412 The full details of bioinformatic analyses are described in the supplementary methods. In brief,
413 sequencing reads were trimmed using Trimmomatic(41) (v0.39) and *in silico* MLST performed
414 with stringMLST(42) (v0.6.3). Isolate genomes were assembled with Unicycler(43) (v0.4.8) and
415 annotated using RASTtk as implemented in the PATRIC Command Line Interface toolkit(44)
416 (v1.035). Pangenome analysis was performed using Panaroo(45) (v1.2.8). Core genome
417 phylogenies were generated using IQ-Tree(46), and gene presence/absence clustering was
418 performed using the Ape(47) (v5.3) R package.

419 Single-nucleotide polymorphism (SNP) calling was performed a) in an ST-specific manner and
420 b) for all isolates sequenced in this work against a single reference (*S. maltophilia* strain K279a,
421 GCF_000072485.1) using Snippy(48) (v4.6.0). Reference genomes for ST-specific SNP calling
422 are found in Supplementary Table 4. Phylogenies were generated using IQ-Tree(46) (v2.0.3) and
423 corrected for recombination using ClonalFrameML(49) (v1.12). Snp-dists(50) (v0.7.0) was used
424 to obtain pairwise SNP distance matrices.

425 **Transmission Analysis**

426 The potential for transmission to have occurred between pwCF infected with the same *S.*
427 *maltophilia* STs was simultaneously assessed with four complementary analyses, each offering a
428 different type of support for a hypothesis of transmission. These analyses included: *i*)
429 SNP/wgMLST allele distance support: inter-pwCF isolate pairs with SNP/wgMLST allele
430 distances overlapping with the distribution of intra-pwCF distances; this latter distribution was
431 compiled in an ST-specific manner but then combined across all sequenced (shared and non-
432 shared) STs, *ii*) phylogenetic support: mixed clustering/interspersion of isolates from ≥ 2 pwCF
433 within the same clade or encompassing the genetic diversity of one pwCF's isolates within the
434 diversity of another's isolates, both with strong Ultrafast bootstrap support ($\geq 95\%$), *iii*) gene
435 content support: mixed clustering/interspersion of isolates from ≥ 2 pwCF within the same clade or
436 encompassing the genetic diversity of one pwCF's isolates within the diversity of another's
437 isolates based on neighbor joining clustering, and *iv*) concurrent carriage support: detection of ≥ 1
438 *S. maltophilia* positive sputum cultures within six months in a given patient pair. The
439 combination of carriage support and at least two other analyses would warrant an individual case
440 review examining evidence that involved patients attended clinic/hospital or other healthcare
441 encounter within 48 hours of each other. The effect of cumulative support from all five analyses
442 would be required to support a hypothesis of transmission between a pair of patients. A lack of
443 support in any analysis was considered to exclude the possibility of transmission.

444 **Multi-mutated Genes Analysis**

445 Complete details of multi-mutated gene analysis are presented in the supplementary methods. In
446 brief, for each ST, we identified all bacterial genes with mutations that accumulated within a
447 pwCF during infection in CF (termed CF genes/mutations) as well as genes with mutations
448 arising prior to infection in CF (termed non-adaptive genes/mutations). Each of these genes was
449 then classified as multi-mutated if it had ≥ 2 mutations, and multi-mutated genes were further
450 subdivided into across-ST or within-ST, depending on which STs the contributing mutations
451 occurred in. The distributions of synonymous, non-synonymous, and stop mutations were then

452 compared between multi-mutated and non-multi-mutated genes using Fisher's exact tests in
453 GraphPad Prism (v9.4.1).

454 CF genes were also analyzed for enrichment of GO categories using OmicsBox (v2.1.14). Gene
455 sequences were obtained from annotated isolate assemblies and Blast run via CloudBlast as
456 implemented in OmicsBox. Association testing was performed with Fisher's exact tests via the
457 Enrichment Analysis tool, with correction for multiple testing performed by the Benjamini-
458 Hochberg procedure and the false discovery rate set to 0.05.

459 **Statistical Analyses**

460 Characteristics of the pwCF cohort were descriptively summarized. Associations between
461 clinical/demographic factors and patients with included/excluded isolates and time between
462 PFGE-typed sputum cultures and detection of new/prior pulsotypes were performed using
463 Fisher's exact tests in SPSS (v28.0.1.0). Statistical analyses for associations between clinical
464 outcomes and carriage of multiple/shared strains were performed using R v.4.1.1 (27). All *P*-
465 values were adjusted for multiple comparisons using the Holm-Bonferroni method (T-test or
466 ANOVA). Categorical variables were presented as numbers and frequencies. Continuous
467 variables were presented as mean±standard deviation (SD) or median (interquartile range), as
468 appropriate. End-stage lung disease was defined as percent predicted forced expiratory volume in
469 1 second (ppFEV₁) as less than or equal to 40.

470 **ACKNOWLEDGEMENTS**

471 The authors gratefully acknowledge the contributions of the staff of the Calgary Adult CF Clinic
472 and Alberta Precision Laboratories for their continued efforts in collecting, stocking, and
473 maintaining of samples in the CACFC Biobank. They also thank the Cystic Fibrosis Foundation
474 (CFF) for its financial support.

475

476 **DATA AVAILABILITY**

477 The sequencing data for the 54 isolates that underwent whole-genome sequencing in this study is
478 available from the National Center for Biotechnology Information (NCBI) Sequence Read
479 Archive (SRA) repository under BioProject PRJNA943478 (REVIEWER LINK
480 <https://dataview.ncbi.nlm.nih.gov/object/PRJNA943478?reviewer=1oc0g1ib56d5das4hi33v3he8g>
481 [g](https://dataview.ncbi.nlm.nih.gov/object/PRJNA943478?reviewer=1oc0g1ib56d5das4hi33v3he8g)). Publicly available datasets from previous studies used to support the conclusions of this
482 article are also available from NCBI – see Supplementary Table 3 for sample accessions and
483 associated BioProject IDs. High quality images of all Figures and Supplementary Figures are
484 available from a figshare digital repository: <https://doi.org/10.6084/m9.figshare.c.6465634.v1>.

485

486 **References**

- 487 1. Brooke JS. 2021. Advances in the Microbiology of *Stenotrophomonas maltophilia*. Clinical
488 Microbiology Reviews 34:e00030-19.
- 489 2. Salsgiver EL, Fink AK, Knapp EA, LiPuma JJ, Olivier KN, Marshall BC, Saiman L. 2016.
490 Changing Epidemiology of the Respiratory Bacteriology of Patients With Cystic Fibrosis.
491 Chest 149:390–400.
- 492 3. Hatziagorou E, Orenti A, Drevinek P, Kashirskaya N, Mei-Zahav M, De Boeck K, Pflieger
493 A, Sciensano MT, Lammertyn E, Macek M, Olesen HV, Farge A, Naehrlich L, Ujhelyi R,
494 Fletcher G, Padoan R, Timpare Z, Malakauskas K, Fustik S, Gulmans V, Turcu O, Pereira
495 L, Mosescu S, Rodic M, Kayserova H, Krivec U, Vazquez-Cordero C, de Monestrol I,
496 Lindblad A, Jung A, Makukh H, Carr SB, Cosgriff R, Zolin A. 2020. Changing
497 epidemiology of the respiratory bacteriology of patients with cystic fibrosis—data from the
498 European cystic fibrosis society patient registry. Journal of Cystic Fibrosis 19:376–383.
- 499 4. Berdah L, Taytard J, Leyronnas S, Clement A, Boelle P-Y, Corvol H. 2018.
500 *Stenotrophomonas maltophilia*: A marker of lung disease severity. Pediatric Pulmonology
501 53:426–430.
- 502 5. Waters V, Yau Y, Prasad S, Lu A, Atenafu E, Crandall I, Tom S, Tullis E, Ratjen F. 2011.
503 *Stenotrophomonas maltophilia* in Cystic Fibrosis Serologic Response and Effect on Lung
504 Disease. Am J Respir Crit Care Med 183:635–640.

- 505 6. Com G, Carroll JL, Castro MM, Tang X, Jambhekar S, Berlinski A. 2014. Predictors and
506 Outcome of Low Initial Forced Expiratory Volume in 1 Second Measurement in Children
507 with Cystic Fibrosis. *The Journal of Pediatrics* 164:832–838.
- 508 7. Barsky EE, Williams KA, Priebe GP, Sawicki GS. 2017. Incident *Stenotrophomonas*
509 *maltophilia* infection and lung function decline in cystic fibrosis. *Pediatric Pulmonology*
510 52:1276–1282.
- 511 8. Waters V, Atenafu EG, Salazar JG, Lu A, Yau Y, Matukas L, Tullis E, Ratjen F. 2012.
512 Chronic *Stenotrophomonas maltophilia* infection and exacerbation outcomes in cystic
513 fibrosis. *Journal of Cystic Fibrosis* 11:8–13.
- 514 9. Cogen J, Emerson J, Sanders DB, Ren C, Schechter MS, Gibson RL, Morgan W, Rosenfeld
515 M. 2015. Risk Factors for Lung Function Decline in a Large Cohort of Young Cystic
516 Fibrosis Patients. *Pediatr Pulmonol* 50:763–770.
- 517 10. Waters V, Atenafu EG, Lu A, Yau Y, Tullis E, Ratjen F. 2013. Chronic *Stenotrophomonas*
518 *maltophilia* infection and mortality or lung transplantation in cystic fibrosis patients.
519 *Journal of Cystic Fibrosis* 12:482–486.
- 520 11. Vidigal PG, Dittmer S, Steinmann E, Buer J, Rath P-M, Steinmann J. 2014. Adaptation of
521 *Stenotrophomonas maltophilia* in cystic fibrosis: Molecular diversity, mutation frequency
522 and antibiotic resistance. *International Journal of Medical Microbiology* 304:613–619.
- 523 12. Pompilio A, Crocetta V, Ghosh D, Chakrabarti M, Gherardi G, Vitali LA, Fiscarelli E, Di
524 Bonaventura G. 2016. *Stenotrophomonas maltophilia* Phenotypic and Genotypic Diversity

- 525 during a 10-year Colonization in the Lungs of a Cystic Fibrosis Patient. *Front Microbiol*
526 7:1551.
- 527 13. Chung H, Lieberman TD, Vargas SO, Flett KB, McAdam AJ, Priebe GP, Kishony R. 2017.
528 Global and local selection acting on the pathogen *Stenotrophomonas maltophilia* in the
529 human lung. *Nat Commun* 8:14078.
- 530 14. Esposito A, Pompilio A, Bettua C, Crocetta V, Giacobazzi E, Fiscarelli E, Jousson O, Di
531 Bonaventura G. 2017. Evolution of *Stenotrophomonas maltophilia* in Cystic Fibrosis Lung
532 over Chronic Infection: A Genomic and Phenotypic Population Study. *Front Microbiol* 8.
- 533 15. Pompilio A, Savini V, Fiscarelli E, Gherardi G, Di Bonaventura G. 2020. Clonal Diversity,
534 Biofilm Formation, and Antimicrobial Resistance among *Stenotrophomonas maltophilia*
535 Strains from Cystic Fibrosis and Non-Cystic Fibrosis Patients. *Antibiotics (Basel)* 9:E15.
- 536 16. Krzewinski JW, Nguyen CD, Foster JM, Burns JL. 2001. Use of Random Amplified
537 Polymorphic DNA PCR To Examine Epidemiology of *Stenotrophomonas maltophilia* and
538 *Achromobacter (Alcaligenes) xylosoxidans* from Patients with Cystic Fibrosis. *J Clin*
539 *Microbiol* 39:3597–3602.
- 540 17. Demko CA, Stern RC, Doershuk CF. 1998. *Stenotrophomonas maltophilia* in cystic
541 fibrosis: incidence and prevalence. *Pediatr Pulmonol* 25:304–308.
- 542 18. Denton M, Todd NJ, Kerr KG, Hawkey PM, Littlewood JM. 1998. Molecular
543 Epidemiology of *Stenotrophomonas maltophilia* Isolated from Clinical Specimens from
544 Patients with Cystic Fibrosis and Associated Environmental Samples. *J Clin Microbiol*
545 36:1953–1958.

- 546 19. Gröschel MI, Meehan CJ, Barilar I, Diricks M, Gonzaga A, Steglich M, Conchillo-Solé O,
547 Scherer I-C, Mamat U, Luz CF, De Bruyne K, Utpatel C, Yero D, Gibert I, Daura X,
548 Kampmeier S, Rahman NA, Kresken M, van der Werf TS, Alio I, Streit WR, Zhou K,
549 Schwartz T, Rossen JWA, Farhat MR, Schaible UE, Nübel U, Rupp J, Steinmann J,
550 Niemann S, Kohl TA. 2020. The phylogenetic landscape and nosocomial spread of the
551 multidrug-resistant opportunist *Stenotrophomonas maltophilia*. Nat Commun 11:2044.
- 552 20. Caballero JD, Clark ST, Coburn B, Zhang Y, Wang PW, Donaldson SL, Tullis DE, Yau
553 YCW, Waters VJ, Hwang DM, Guttman DS. 2015. Selective Sweeps and Parallel
554 Pathoadaptation Drive *Pseudomonas aeruginosa* Evolution in the Cystic Fibrosis Lung.
555 mBio 6:e00981-15.
- 556 21. Diaz Caballero J, Clark ST, Wang PW, Donaldson SL, Coburn B, Tullis DE, Yau YCW,
557 Waters VJ, Hwang DM, Guttman DS. 2018. A genome-wide association analysis reveals a
558 potential role for recombination in the evolution of antimicrobial resistance in *Burkholderia*
559 *multivorans*. PLoS Pathog 14.
- 560 22. Sánchez P, Alonso A, Martinez JL. 2002. Cloning and Characterization of SmeT, a
561 Repressor of the *Stenotrophomonas maltophilia* Multidrug Efflux Pump SmeDEF.
562 Antimicrobial Agents and Chemotherapy 46:3386–3393.
- 563 23. Goss CH, Mayer-Hamblett N, Aitken ML, Rubenfeld GD, Ramsey BW. 2004. Association
564 between *Stenotrophomonas maltophilia* and lung function in cystic fibrosis. Thorax
565 59:955–959.

- 566 24. Capaldo C, Beauruelle C, Saliou P, Rault G, Ramel S, Héry-Arnaud G. 2020. Investigation
567 of *Stenotrophomonas maltophilia* epidemiology in a French cystic fibrosis center. *Respir*
568 *Med Res* 78:100757.
- 569 25. Gabrielaite M, Bartell JA, Nørskov-Lauritsen N, Pressler T, Nielsen FC, Johansen HK,
570 Marvig RL. 2021. Transmission and Antibiotic Resistance of *Achromobacter* in Cystic
571 Fibrosis. *J Clin Microbiol* 59:e02911-20.
- 572 26. Stapleton PJ, Izydorczyk C, Clark S, Blanchard A, Wang PW, Yau Y, Waters V, Guttman
573 DS. 2020. *Pseudomonas aeruginosa* strain sharing in early infection among children with
574 cystic fibrosis. *Clin Infect Dis* <https://doi.org/10.1093/cid/ciaa788>.
- 575 27. Parkins MD, Somayaji R, Waters VJ. 2018. Epidemiology, Biology, and Impact of Clonal
576 *Pseudomonas aeruginosa* Infections in Cystic Fibrosis. *Clinical Microbiology Reviews*
577 31:e00019-18.
- 578 28. Izydorczyk C, Waddell B, Edwards BD, Greysen-Wong J, Surette MG, Somayaji R, Rabin
579 HR, Conly JM, Church DL, Parkins MD. 2020. Epidemiology of *E. coli* in Cystic Fibrosis
580 Airways Demonstrates the Capacity for Persistent Infection but Not Patient-Patient
581 Transmission. *Front Microbiol* 11.
- 582 29. Doyle RM, Rubio M, Dixon G, Hartley J, Klein N, Coll P, Harris KA. 2020. Cross-
583 transmission Is Not the Source of New *Mycobacterium abscessus* Infections in a
584 Multicenter Cohort of Cystic Fibrosis Patients. *Clinical Infectious Diseases* 70:1855–1864.

- 585 30. Izydorczyk C, Waddell BJ, Weyant RB, Surette MG, Somayaji R, Rabin HR, Conly JM,
586 Church DL, Parkins MD. 2022. The natural history and genetic diversity of *Haemophilus*
587 *influenzae* infecting the airways of adults with cystic fibrosis. 1. Sci Rep 12:15765.
- 588 31. Valiente-Mullor C, Beamud B, Ansari I, Francés-Cuesta C, García-González N, Mejía L,
589 Ruiz-Hueso P, González-Candelas F. 2021. One is not enough: On the effects of reference
590 genome for the mapping and subsequent analyses of short-reads. PLOS Computational
591 Biology 17:e1008678.
- 592 32. Yan J, Kevat A, Martinez E, Teese N, Johnson K, Ranganathan S, Harrison J, Massie J,
593 Daley A. 2019. Investigating transmission of *Mycobacterium abscessus* amongst children in
594 an Australian cystic fibrosis centre. J Cyst Fibros <https://doi.org/10.1016/j.jcf.2019.02.011>.
- 595 33. Ledson MJ, Gallagher MJ, Corkill JE, Hart CA, Walshaw MJ. 1998. Cross infection
596 between cystic fibrosis patients colonised with *Burkholderia cepacia*. Thorax 53:432–436.
- 597 34. Yu D, Yin Z, Li B, Jin Y, Ren H, Zhou J, Zhou W, Liang L, Yue J. 2016. Gene flow,
598 recombination, and positive selection in *Stenotrophomonas maltophilia*: mechanisms
599 underlying the diversity of the widespread opportunistic pathogen. Genome 59:1063–1075.
- 600 35. Wood TE, Burke JM, Rieseberg LH. 2005. Parallel genotypic adaptation: when evolution
601 repeats itself. Genetica 123:157–170.
- 602 36. Bailey SF, Rodrigue N, Kassen R. 2015. The effect of selection environment on the
603 probability of parallel evolution. Mol Biol Evol 32:1436–1448.

- 604 37. Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we
605 learned about the genetics of adaptation? *Trends Ecol Evol* 23:26–32.
- 606 38. Farrell PM, Rosenstein BJ, White TB, Accurso FJ, Castellani C, Cutting GR, Durie PR,
607 LeGrys VA, Massie J, Parad RB, Rock MJ, Campbell PW. 2008. Guidelines for Diagnosis
608 of Cystic Fibrosis in Newborns through Older Adults: Cystic Fibrosis Foundation
609 Consensus Report. *The Journal of Pediatrics* 153:S4–S14.
- 610 39. Parkins MD, Glezerson BA, Sibley CD, Sibley KA, Duong J, Purighalla S, Mody CH,
611 Workentine ML, Storey DG, Surette MG, Rabin HR. 2014. Twenty-Five-Year Outbreak of
612 *Pseudomonas aeruginosa* Infecting Individuals with Cystic Fibrosis: Identification of the
613 Prairie Epidemic Strain. *J Clin Microbiol* 52:1127–1135.
- 614 40. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH,
615 Swaminathan B. 1995. Interpreting chromosomal DNA restriction patterns produced by
616 pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol*
617 33:2233–2239.
- 618 41. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
619 sequence data. *Bioinformatics* 30:2114–2120.
- 620 42. Gupta A, Jordan IK, Rishishwar L. 2017. stringMLST: a fast k-mer based tool for
621 multilocus sequence typing. *Bioinformatics* 33:119–121.
- 622 43. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome
623 assemblies from short and long sequencing reads. *PLOS Computational Biology*
624 13:e1005595.

- 625 44. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N,
626 Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao
627 C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC,
628 Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V,
629 Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2020. The PATRIC Bioinformatics Resource
630 Center: expanding data and analysis capabilities. *Nucleic Acids Research* 48:D606–D612.
- 631 45. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA,
632 Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J. 2020.
633 Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology*
634 21:180.
- 635 46. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A,
636 Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
637 Inference in the Genomic Era. *Molecular Biology and Evolution* 37:1530–1534.
- 638 47. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R
639 language. *Bioinformatics* 20:289–290.
- 640 48. Seemann T. Snippy.
- 641 49. Didelot X, Wilson DJ. 2015. ClonalFrameML: Efficient Inference of Recombination in
642 Whole Bacterial Genomes. *PLoS Comput Biol* 11.
- 643 50. Seemann T. snp-dists.
- 644

Tables

Table 1. Summary characteristics of pwCF with at least one *S. maltophilia* positive sputum culture between 1979-2016.

Demographics¹		
	Age at first episode (median, IQR) (years)	24.9 (20.3-32.3)
	Sex (% female) (n/N)	58.5 (48/82)
CF Characteristics and Comorbidities¹		
	F508del Homozygous (%) (n/N)	66.2 (45/68)
	F508del Heterozygous (%) (n/N)	23.5 (16/68)
	Pancreatic insufficient (%) (n/N)	90.0 (72/80)
	CFRD (%) (n/N)	22.5 (18/80)
	CFLD (%) (n/N)	18.8 (15/80)
	Osteopenia (%) (n/N)	43.0 (34/79)
	Osteoporosis (%) (n/N)	7.59 (6/79)
Determinants of Disease State¹		
	ppFEV1 (median, IQR)	59 (38-76.5)
	% with Mild CF Lung Disease (ppFEV1 predicted ≥ 70) (n/N)	32.4 (24/74)
	% with Moderate CF Lung Disease ($40 < \text{ppFEV1} < 70$) (n/N)	40.5 (30/74)
	% with Severe CF Lung Disease (ppFEV1 ≤ 40) (n/N)	27.0 (20/74)
	ppFVC (median, IQR)	81.5 (58-93.3)
	Use of supplemental oxygen (%) (n/N)	22.5 (18/80)
	Use of enteral nutrition (%) (n/N)	12.5 (10/80)
Co-Infections¹		
	% with PA co-infection (n/N)	63.51 (47/74)
	% with MSSA co-infection (n/N)	44.6 (33/74)
	% with MRSA co-infection (n/N)	4.05 (3/74)
	% with HI co-infection (n/N)	6.76 (5/74)
Disease Modifying CF Medications¹		
	Inhaled Tobramycin (%) (n/N)	30.0 (24/80)
	Inhaled Colistin (%) (n/N)	1.25 (1/80)
	Inhaled Aztreonam (%) (n/N)	2.50 (2/80)
	Azithromycin (%) (n/N)	12.5 (10/80)
	Inhaled DNase (%) (n/N)	50.0 (40/80)
	Inhaled Hypertonic Saline (%) (n/N)	15.0 (12/80)

¹ All values were based on demographic and clinical factors at first *S. maltophilia* infection episode (first positive sputum culture).

F508del = deletion of phenylalanine at position 508 in protein polypeptide

CFRD = cystic fibrosis related diabetes

CFLD = cystic fibrosis liver disease

ppFEV1 = percent predicted forced expiratory volume in one second

ppFVC = percent predicted forced vital capacity

PA = *Pseudomonas aeruginosa*

MSSA = methicillin susceptible *Staphylococcus aureus*

MRSA = methicillin resistant *Staphylococcus aureus*

HI = *Haemophilus influenzae*

Table 2. Odds and relative risks of increasing time between typed *S. maltophilia* sputum cultures and likelihood of detection of a new vs. prior pulsotype.

Time Between Typed Cultures (X years)	Risk of New Pulsotype (%)	Risk of Prior Pulsotype (%)	Odds New vs. Prior Pulsotype	Risk of New Pulsotype in $\leq X$ Years (%)	Relative Risk [new pulsotype $>X/\leq X$] (95% CI)	Unadjusted P¹
>1	21/37 (56.8)	16/37 (43.2)	1.31	11/50 (22.0)	2.58 (1.43-4.67)	0.00147
>2	15/18 (83.3)	3/18 (16.7)	5.00	17/69 (24.6)	3.38 (2.13-5.37)	0.00008
>5	9/11 (90.9)	1/11 (9.1)	9.00	22/76 (28.9)	3.14 (2.11-4.68)	0.000131

1 Raw p values not corrected for multiple testing.

Table 3. Shared *S. maltophilia* sequence types identified among patients attending the Calgary Adult CF Clinic.

Shared Pulsotypes/Sequence Types				
Sequence Type	Corresponding Pulsotype(s)	Number of Isolates²	Number of Patients³	Patients
5	A	7	5	A057, A061, A275, A372, A376
39	B	2	2	A061, A200
199	C, F ¹	19	3	A055, A057, A090
220	D	3	2	A090, A371
224	E	3	2	A013, A200
Non-shared (Control) Pulsotypes/Sequence Types⁴				
Sequence Type	Corresponding Pulsotype(s)	Number of Isolates	Number of Patients	Patients
23	G	4	1	A130
91	H	5	1	A374
246	I	2	1	A057
365	J	3	1	A344
Novel 2	K	3	1	A090

1 A single isolate (SM003) belonging to patient A055 belonged to a unique pulsotype (F) but was included due to sharing an ST with pulsotype C isolates. All pulsotype C isolates were more closely related to each other than to the pulsotype F isolate.

2 Total number of isolates is 54; 3 isolates that did not correspond to any shared/non-shared STs are not included here.

3 Sum of number of patients is greater than total number of patients with shared pulsotypes due to some patients having isolates belonging to ≥ 2 shared pulsotypes.

4 Select non-shared sequence types were additionally included as controls for intra-patient genetic diversity.

Table 4. Assessment for infection transmission risk within patient pairs among all identified shared STs.

ST	Patient Pair	SNP/wgMLST Distance Support	Phylogenetic Support	Gene Content Support	Carriage Support	Epidemiological Support
5	A057/A061	No	No	No	No	No
	A057/A275	No	No	No	Yes	No
	A057/A372	No	No	No	No	No
	A057/A376	No	No	No	No	No
	A061/A275	No	No	No	No	No
	A061/A372	No	No	No	Yes	No
	A061/A376	No	No	No	Yes	No
	A275/A372	No	No	No	No	No
	A275/A376	No	No	No	No	No
39	A372/A376	No	No	No	Yes	No
	A061/A200	No	-- ²	No	No	No
199	A055/A057	Partial¹	No	No	No	No
	A055/A090	No	No	No	Yes	No
	A057/A090	No	No	No	No	No
220	A090/A371	No	No	No	Yes	No
224	A013/A200	No	No	No	No	No

¹ Distance support here is among SNP but not wgMLST distances, and only among 12/14 isolate pairs between these two pwCF.

² Only one isolate/patient was available for this pair, so phylogenetic support was unable to be assessed.

Figures

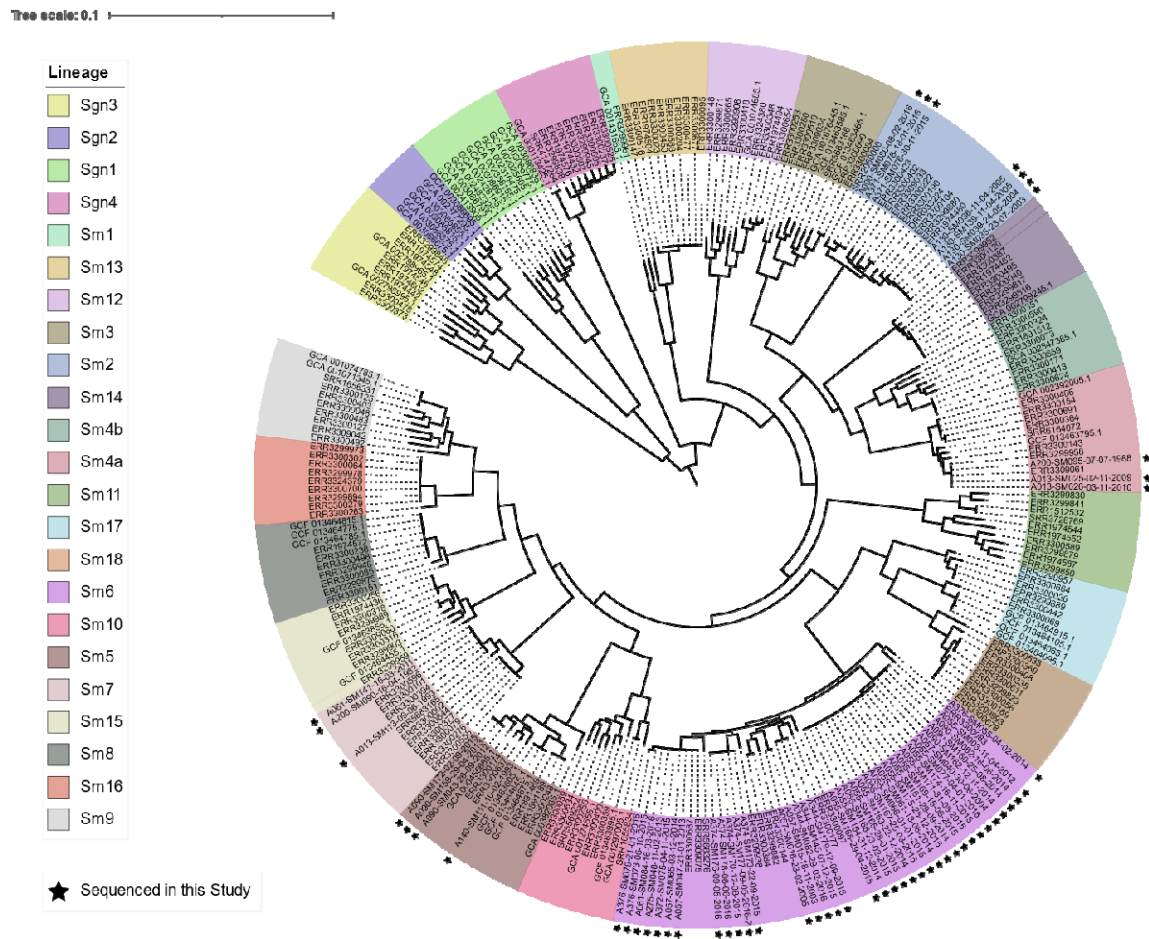


Figure 1. Core genome phylogeny (midpoint rooted) of 23 phylogenetic lineages comprising the *S. maltophilia* species complex. Lineage colors are presented in the same order in the legend as in the phylogeny (clockwise, starting with lineage Sgn3). Isolates sequenced in this study are marked by black stars. The phylogeny was constructed from an alignment of 1947 core genes.

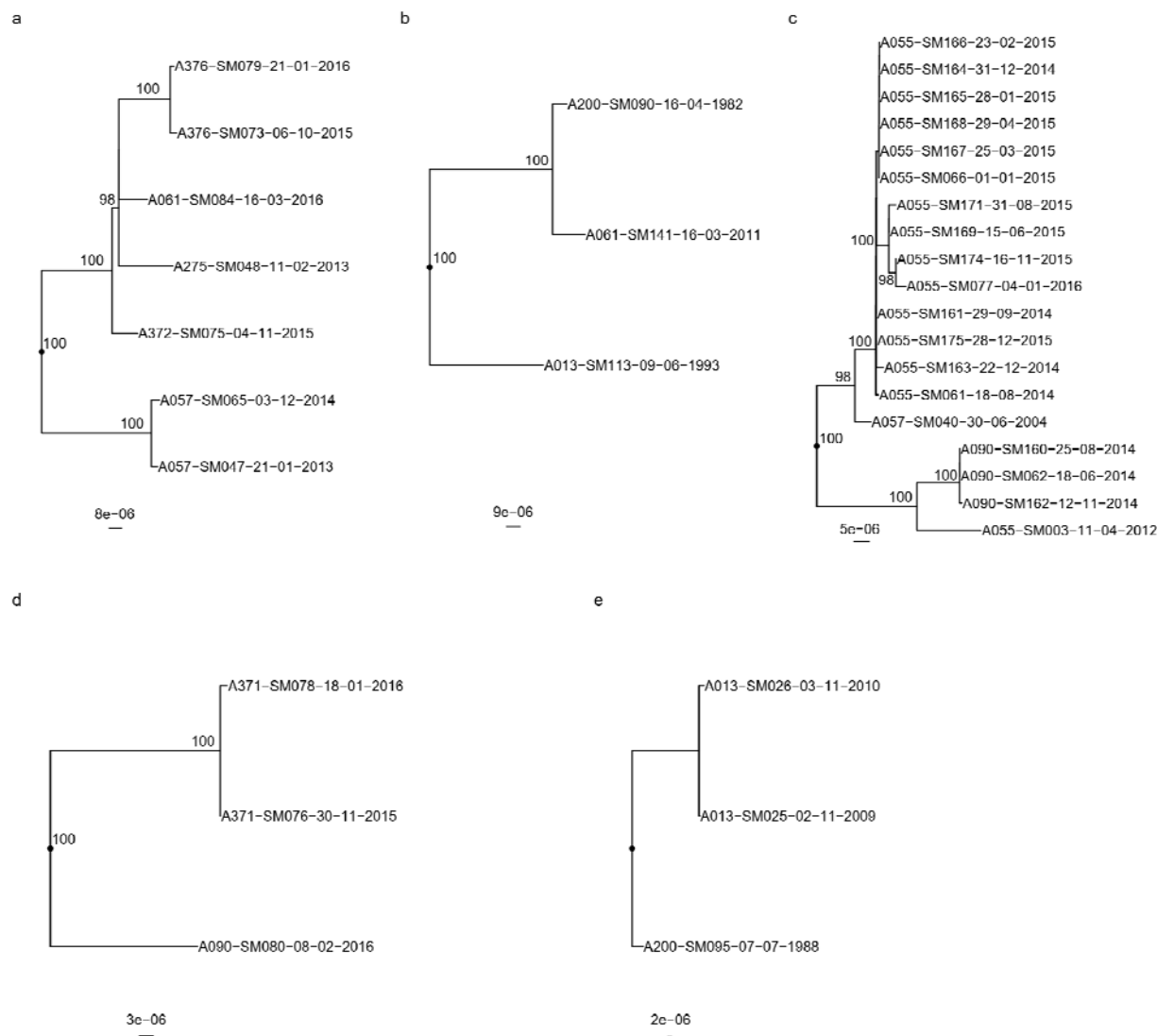


Figure 2. Recombination-corrected maximum likelihood phylogenies of isolates belonging to shared STs: (a) ST-5, (b) ST-39, (c) ST-199, (d) ST-220, (e) ST-224. Each phylogeny is rooted at the midpoint of the branch where outgroups attach. Isolate A013-SM113-09-06-1993 in (b) is a novel single-locus variant of ST-39 and is included as an outgroup. UltraFast bootstrap support is indicated only in clades with $\geq 95\%$ support. Scale bars are in units of SNPs/site. Isolate names are presented in the format “Patient_Identification_Number-Isolate_Identification_Number-dd-mm-yyyy”.

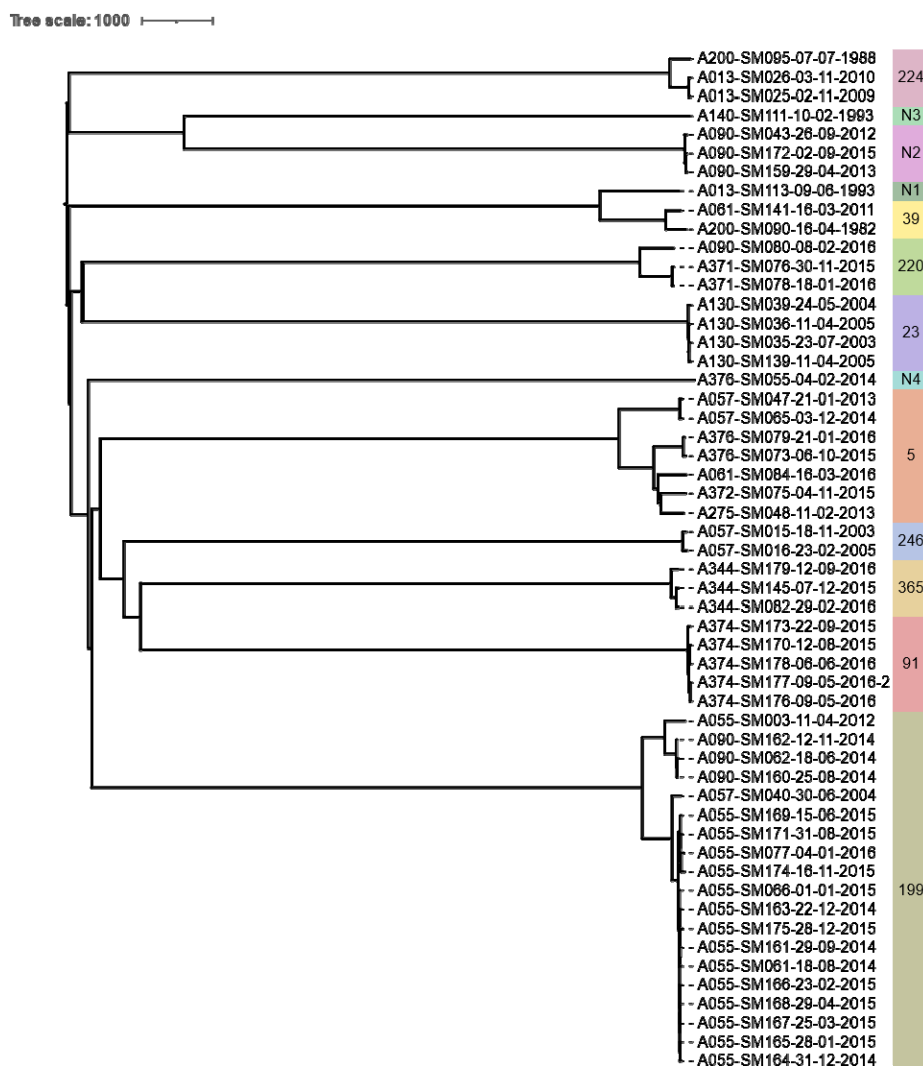


Figure 3. Neighbor-joining phylogeny constructed from wgMLST allele data from all isolates sequenced in this study. STs are indicated by text and colored bands on the right. Isolate names are presented in the format “Patient_Identification_Number-Isolate_Identification_Number-dd-mm-yyyy”.