Mycobacterium avium complex (MAC) genomics and transmission in a London hospital

Andries J van Tonder[1], Huw C Ellis[2,3], Colin P Churchward[3], Kartik Kumar[2,3], Newara Ramadan[4], Susan Benson[4], Julian Parkhill[1*], Miriam F Moffatt¶[3], Michael R Loebinger¶[2,3], William OC Cookson¶[2,3*]

[1]Department of Veterinary Medicine, University of Cambridge, Cambridge

[2] Host Defence Unit, Department of Respiratory Medicine, Royal Brompton Hospital, Guy's and St Thomas' NHS Foundation Trust, London

[3]National Heart and Lung Institute, Imperial College London, London

[4]Department of Microbiology, Royal Brompton Hospital, Guy's and St Thomas' NHS Foundation Trust, London

* Corresponding authors: jp369@cam.ac.uk, w.cookson@imperial.ac.uk

¶ Contributed equally

Abstract

Background

Non-tuberculous mycobacteria (NTM) are ubiquitous environmental microorganisms and opportunistic pathogens in individuals with pre-existing lung conditions such as cystic fibrosis (CF) and non-CF bronchiectasis (BX). Whilst recent studies of *Mycobacterium abscessus* have identified transmission within single CF centres as well as nationally and globally, transmission of other NTM species is less well studied.

Methods

We sequenced 1000 Mycobacterium avium complex (MAC) isolates from CF and non-CF patients at the Royal Brompton Hospital (RBH), London. Epidemiological links were identified from patient records. Previously published genomes were used to characterise global population structures.

Findings

Analysis of the three most predominant MAC species identified putative transmission clusters that contained patients with CF, BX and other lung conditions, although few epidemiological links could be identified. For *M. avium*, lineages were largely limited to single countries, whilst for *M. chimaera*, global transmission clusters previously associated with heater cooler units (HCUs) were found. However, the origin of the major HCU-associated outbreak was a lineage already circulating in patients with pre-existing lung conditions.

Interpretation

CF and non-CF patients share transmission chains even in the presence of CF patient-focussed hospital control measures, although the lack of epidemiological links suggests that most transmission is indirect and may be due to environmental foci or else asymptomatic carriage in the wider population. The major HCU-associated *M. chimaera* lineage being derived from an already circulating lineage, suggests that HCUs are not the sole vector nor the ultimate source of this lineage. Future studies should include sampling of environmental reservoirs and potential asymptomatic carriers.

Funding

This project was supported by the Asmarley Trust, the Wellcome Trust and the NIHR Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust, Imperial College London. Huw Ellis' CRF was funded by a grant from the Welton Foundation.

Research in context

Evidence before this study

Whilst recent studies in *Mycobacterium abscessus* have identified transmission within single CF centres as well as nationally and globally, the transmission dynamics between CF and non-CF patients has not yet been comprehensively examined in the *Mycobacterium avium* complex (MAC). We searched PubMed and bioRxiv with the following search terms: "*Mycobacterium avium*" or "*Mycobacterium intracellulare*" or "*Mycobacterium chimaera*" and "whole genome" or "genomics" or "WGS" or "transmission", for articles published in English between January 1st 2015 and December 31st 2020. The searches returned 492 articles of which 20 were relevant. The relevant publications described the collection and sequencing of MAC isolates and provided genomes for global contextual analysis. Two papers explicitly examined transmission of the three main MAC species considered in this study: one investigated transmission of MAC species in CF centres in the United States of America whilst the other looked at the similarity of MAC isolates from community and household water in suburban Philadelphia. Three studies investigated the diversity of *M. chimaera* isolates associated with Heater Cooler Units (HCUs) although none of these expressly considered transmission using SNP thresholds.

73 **Added value of this study**

74 Here we present the first study to use a well-sampled longitudinal isolate dataset,
75 that includes both CF and non-CF patients from a single hospital setting, to
76 investigate transmission of MAC species. We identified transmission clusters in the
77 three predominant MAC species circulating in the hospital and showed that these
78 included both CF and non-CF patients. We then incorporated isolates from previous
79 studies to examine the global population structure of MAC species and showed that
80 for *M. avium* there were UK-specific lineages circulating amongst patients, whilst for
81 *M. chimaera* we could identify global lineages associated with HCUs. For the first
82 time, we also show that the predominant HCU-associate lineage is likely derived from
83 already circulating lineages associated with patients with respiratory diseases.

84 **Implications of all the available evidence**

85 Our study shows the value of integrating whole genome sequencing with
86 epidemiological data to perform high-resolution molecular analyses to characterise
87 MAC populations and identify transmission clusters. Knowledge of putative
88 transmission networks can improve responses to outbreaks and inform targeted
89 infection control and clinical practice.

90

91 **Introduction**

92 Non-tuberculous mycobacteria (NTM) are ubiquitous environmental microorganisms
93 found in soil and water and are considered opportunistic pathogens in humans.
94 Individuals with pre-existing genetic or acquired lung diseases such as cystic fibrosis
95 (CF), non-CF bronchiectasis (BX) and chronic obstructive pulmonary disease (COPD)
96 are more prone to NTM disease although individuals with no known immune

97 dysfunction can also present with NTM infections (1–3). Symptoms of NTM
98 pulmonary disease are variable but most patients will develop a chronic cough and
99 other symptoms may include fatigue, sputum production, chest pain, breathlessness,
100 fever and weight loss (1). Globally, disease due to NTM infections is increasing in
101 prevalence. For example, the estimated prevalence of NTM disease in the United
102 States of America (USA) rose from 2.4 cases/100,000 in the early 1980s to 15.2
103 cases/100,000 in 2013 (4), whilst in the United Kingdom (UK) the prevalence rose
104 from 0.9 cases/100,000 to 2.9 cases/100,000 between 1995 and 2006 (5). NTM
105 infections may be progressive and treatment requires prolonged multi-drug therapy
106 (6) and is often unsuccessful due to an absence of antimicrobial agents with low
107 toxicity and effective *in vivo* activity against NTM species (1).
108
109 A number of NTM species including *Mycobacterium abscessus* and members of the
110 *M. avium* Complex (MAC), notably *M. avium* and *M. intracellulare*, have emerged as
111 major respiratory pathogens in the past three decades (7–9). Another member of the
112 MAC, *M. chimaera*, has also been implicated in numerous global infections
113 associated with cardiothoracic surgery with the source of infections linked to heater-
114 cooler units (HCUs) contaminated during their manufacture (10–12).
115
116 Until recently the prevailing hypothesis was that infections caused by NTM were due
117 to independent acquisitions from environmental sources such as soil, contaminated
118 drinking water distribution systems and household plumbing. Recent studies of *M.*
119 *abscessus* in CF patients have however identified indirect patient-patient
120 transmission within a single CF centre as well as the presence of globally circulating

121 clones of *M. abscessus* amongst CF patients worldwide (13–16). An observational
122 study of *M. abscessus,* across England, revealed that these dominant clones are also
123 found in patients with other chronic respiratory diseases. The study however was
124 unable to identify epidemiological links for most closely-related isolates, suggesting
125 environmental acquisition may be involved (17). A recent study of *M. abscessus* has
126 demonstrated that transmission networks may involve both people with CF and those
127 without, and that these transmission networks are global. It is therefore likely that
128 transmission is complex, involving multiple patient cadres as well as environmental
129 intermediates (18). In the special case of *M. chimaera*, the high level of genetic
130 similarity between sequenced *M. chimaera* isolates collected from patients, HCUs
131 and the factory of origin suggested a point source contamination during manufacture
132 causing global distribution followed by localised transmission (12).
133
134 To date, little work has been done to examine whether similar patterns of transmission
135 in the MAC are occurring between patients with CF, BX or other chronic respiratory
136 diseases. Using a large collection of longitudinal isolates collected from patients
137 attending the Royal Brompton Hospital (RBH) in London, the aims of this study were
138 to characterise the population structure of MAC; identify potential transmission
139 chains involving patients with CF and other non-CF lung conditions; and place the
140 RBH isolates in a global context using previously published genomes.
141
142 **Methods**
143 *Consent/ethics*

144 The study (access to patients' clinical data) was approved by the NHS Health
145 Research Authority (HRA) and Health and Care Research Wales (HCRW) (REC
146 reference 21/HRA/2554).
147
148 *Data collection*
149 Clinical data pertaining to patients from whom NTM cultures were isolated were
150 collected from electronic health records at the RBH. Data included patients' sex, age
151 at the time of first positive NTM culture, height, weight, lung function test results,
152 comorbidities, medication history and date of death (where applicable).
153 Anonymization was undertaken by removing personal data, including patients'
154 hospital numbers, prior to analysis.
155
156 *Sample collection*
157 Isolates were collected from patients attending the respiratory inpatient and
158 outpatient clinics of the RBH between January 2013 and April 2016. The RBH
159 routinely archive all mycobacterial isolates cultured from their patients and this
160 archive was used without selection as the basis for the study.
161
162 *Culturing, DNA extraction and sequencing*
163 NTM cultures were grown from bead stock cultures in BBL MGIT media (BD) in a
164 Bactec MGIT 960 (BD) until the system indicated growth. In the absence of growth,
165 DNA was extracted from the bead stock. DNA extractions were performed as
166 previously described (https://dx.doi.org/10.17504/protocols.io.bf28jqhw). A total of
167 1189 DNA extracts were sequenced by the core pipeline teams at the Wellcome

168 Sanger Institute. Illumina libraries were created using the Nextera protocol and the
169 Illumina Hiseq X10 platform was used to generate 2 x 150 bp paired-end reads. Raw
170 sequencing reads were deposited at the European Nucleotide Archive under project
171 PRJEB21813. All accessions used in this project are listed in Supplementary File 1.
172
173 *Sequence QC, mapping and phylogenetics*
174 Basic quality control metrics for the raw sequence data were generated using FastQC
175 v0.11.9 (19). Sequence reads with similarity to *Mycobacterium* species were
176 identified using Kraken v0.10.6 (20) and Bracken v1.0 (21). Samples with < 70%
177 reads mapping to a *Mycobacterium* species were excluded from further analyses (n
178 = 116). Three isolates identified as *Mycobacterium abscessus* were removed from
179 the dataset. Sequence reads for each species were trimmed using Trimmomatic
180 v0.33 (22) and mapped to appropriate references (Supplementary Table 1) using BWA
181 mem v0.7.17 (minimum and maximum insert sizes of 50 and 1000 respectively) (23).
182 Single nucleotide polymorphisms (SNPs) were called using SAMtools v1.2 mpileup
183 and BCFtools v1.2 (minimum base call quality of 50 and minimum root squared
184 mapping quality of 30) as previously described (24). Samples with reads that mapped
185 to < 80% of the reference were excluded (n = 70). Variant sites were extracted from
186 the resulting alignments using snp-sites v2.5.1 (25). Whole species maximum
187 likelihood phylogenetic trees were built using IQ-tree v1.6.5 accounting for constant
188 sites (-fconst; determined using snp-sites -C) with the built-in model testing (-m MFP)
189 to determine the best phylogenetic model and 1000 ultrafast bootstraps (-bb 1000)
190 (26).
191

192 Recombinant regions were identified and removed from alignments using GUBBINS
193 (27) and new phylogenetic trees were constructed as described above. Pairwise SNP
194 distances were calculated for all pairs of isolates using pairsnp (28).
195
196 *Global collections*
197 To provide context for each the isolates sequenced for each species in this study,
198 datasets consisting of published sequenced isolates were assembled
199 (Supplementary File 2) (12,29–48). Sequence data were downloaded from the
200 European Nucleotide Archive (ENA) and trimmed; Sample QC, mapping and
201 phylogenetic tree construction were performed as detailed above.
202
203 *Genome assemblies*
204 A previously published pipeline was used to produce annotated assemblies (49).
205 Briefly, sequence reads were assembled with spades v 3.10.10 (50) and assemblies
206 were improved by first scaffolding the assembled contigs using SSPACE v2.0 (51)
207 and filling the sequence gaps with GapFiller v1.11 (52).
208
209 *Transmission and epidemiological linkage*
210 Genomic clusters were identified using fastBAPS (53) and new alignments were
211 created for clusters ≥ ten isolates by aligning sequence reads for included isolates
212 against the assembly that had the smallest number of contigs (using the method
213 described above). In order to calculate a pairwise SNP threshold to determine
214 putative transmission clusters within each genomic cluster, pairwise SNP distances
215 for all isolates for each species in the RBH datasets were calculated. Using a

216 previously described method (54), the transmission threshold for each species,
217 regardless of cluster, was calculated by taking the 95th percentile of the maximum
218 within-patient isolate pairwise SNP distances for all patients and adding twice the
219 number of mutations expected to occur in a six month period. To account for excess
220 within-patient diversity observed in the *M. chimaera* FB1 and *M. a. avium* FB14
221 clusters (Supplementary Figure 1), pairwise SNP distances greater than 25 and 50
222 (assumed to result from infection with multiple lineages) were removed respectively
223 before the above calculations were performed. Based on these results, the R library
224 iGRAPH (55,56) and pairwise SNP thresholds of 16 (*M. intracellulare* and *M. a.*
225 *hominissuis*), 30 (*M. chimaera*) and 58 SNPs (*M. a. avium*) were used to calculate
226 putative transmission clusters in each genomic cluster. Finally, in order to identify
227 possible epidemiological links between patients infected with the same transmission
228 clusters, hospital stay records were examined for epidemiological contacts. The latter
229 were defined as patients attending the same ward on the same day up to one year
230 prior to the collection of the first sequenced isolate.
231
232 **Results**
233 *Patient demographics*
234 The median age and BMI of the 354 patients included in the study was 56 years
235 (range 5 - 93) and 22.5 (range 13.4 – 43.4) years respectively. One hundred and
236 seventy-four patients (49.1%) were male and 38/354 (10.7%) were smokers. There
237 were 147/354 (41.5%) patients with BX, 87/354 (24.6%) with CF, 53/354 (15.0%) with
238 COPD, 32/354 (9.0%) with asthma, 19/354 (5.4%) with allergic bronchopulmonary
239 aspergillosis (ABPA), 17/354 (4.8%) with interstitial lung disease (ILD) and 7/354

240 (2.0%) with other underlying respiratory conditions such as pleural thickening or
241 sarcoidosis. Patients with no other respiratory disease or predisposition to NTM
242 infections accounted for 3.4% (12/354) of the cohort. During the study period, 55/354
243 (15.5%) patients were on antibiotic treatment regimes. For 20 patients (5.6%) clinical
244 data were unavailable.
245
246 *Species distribution*
247 A total of 1,000 isolates from 354 patients were successfully sequenced and the 11
248 MAC species identified are detailed in Table 1. The three predominant species
249 amongst the sequenced isolates were *M. avium (M. a. avium* and *M. a. hominissuis)*,
250 *M. chimaera* and *M. intracellulare.* Together these accounted for 926/1000 (92.6%)
251 of the MAC isolates sequenced. Most patients were infected with only a single
252 species during the collection period. However, 46 of the 354 patients (13.0%) were
253 infected with two or more species (Supplementary Figure 2). In this group of patients,
254 the majority of isolates collected were typically from a single species, with other
255 species observed more infrequently (Figure 1). Subsequent analyses in this study will
256 focus on the three predominant species in the dataset: *M. intracellulare, M. avium (M.*
257 *a. avium* and *M. a. hominissuis)*, and *M. chimaera*.
258
259 *M. intracellulare*
260 A total of 162 genomes from 37 patients were characterised as *M. intracellulare*
261 (Figure 2A). Eleven of the patients had CF, seventeen had BX, with the remaining
262 seven having other lung conditions (COPD n = 3; ILD n = 3; asthma n = 1; congenital
263 pulmonary airway malformation [CPAM] n = 1) and disease metadata were missing

264 for two patients. Genomic clustering with fastBAPs identified nine clusters with three
265 of these having more than ten genomes. Following remapping to local references for
266 the three largest fastBAPS clusters, three putative transmission clusters were
267 identified with the largest, Mi_FB3_1, composed of 16 patients (Figure 3A;
268 Supplementary Table 2). Of these 16 patients, eight had BX, seven CF and one ILD.
269 The range of the number of isolates collected per patient was between one and 32.
270 Four of the sixteen patients were also infected with other species or lineages during
271 the sampling period with Mi_FB3_1 only being detected in 1/22 isolates collected for
272 patient 218 (Figure 3A). During the time period that the sequenced isolates were
273 collected, four of the patients (85, 95, 97 and 99), were being treated with antibiotics
274 and three of the four patients were treated successfully. No epidemiological links
275 were identified between patients in the year prior to isolate collection.
276
277 Comparison with 77 previously published *M. intracellulare* isolates showed that the
278 most closely related contextual isolate was collected in the UK in 2015/2016 [32] and
279 formed part of a clade containing isolates from the Mi_FB3_1 cluster (Figure 4A). The
280 remainder of the RBH isolates were evenly distributed through the global
281 phylogenetic tree.
282
283 *M. avium*
284 A total of 405 sequenced isolates collected from 176 patients were identified as *M.*
285 *avium*. Of the patients infected with *M. avium*, 44 had CF and 81 BX. Sequence data
286 for all *M. avium* isolates was mapped to a single reference (*Mycobacterium avium*
287 104; NC008595.1) and a phylogenetic tree constructed with a previously published

288 *M. avium paratuberculosis* isolate (DRR263663) chosen as an outgroup
289 (Supplementary Figure 3). The structure of this phylogeny showed that there were
290 two major clades which corresponded to the two subspecies *M. avium avium* (n =
291 207; MAA) and *M. avium hominissuis* (n = 198; MAH). Isolates for each of the two
292 subspecies were next analysed separately. A small number of patients were infected
293 with both subspecies (n = 6). No examples of the other two *M. avium* subspecies,
294 *paratuberculosis* and *silvaticum*, were identified in this study.
295
296 *M. avium avium (MAA)*
297 Of the 76 patients infected with MAA, 15 had CF, 39 had BX and 16 had other lung
298 conditions (COPD n = 10; ILD n = 3; ABPA n = 3; asthma n = 6) whilst three patients,
299 including one smoker, had no pre-existing respiratory disease (metadata were
300 unavailable for three patients; Figure 2B). Clustering of the 207 MAA genomes
301 identified 17 fastBAPS clusters (Figure 2B). Remapping of the four largest fastBAPS
302 clusters allowed the identification of seven putative transmission clusters comprising
303 between two and ten patients (Supplementary Table 3) with the largest cluster of ten
304 isolates comprised of patients with CF (n = 2), BX (n = 6) and asthma or COPD (n =2;
305 Figure 3B).
306
307 A total of 21 fastBAPS clusters were defined in the global phylogeny. Examination of
308 the distribution of pairwise SNP distances for each fastBAPS cluster containing at
309 least ten genomes revealed that there were two clusters containing RBH isolates with
310 lower median pairwise SNP distances (FB10 and FB15; Figure 4B and Supplementary
311 Figure 5B). The 23 genomes, including two RBH genomes, in FB10 were collected

312 in the UK between 2013 and 2016, with most coming from the National Mycobacterial
313 Reference Service. The latter characterises mycobacterial cultures from across the
314 Midlands and North of England (Figure 5A) [32]. If the SNP threshold used to
315 calculate putative transmission clusters circulating in the RBH was applied, then one
316 isolate from the RBH (40-1) formed part of a cluster comprised of isolates from Quan
317 *et al.* [32]. The majority (35/46) of the isolates forming FB15 were collected in the USA
318 with most from a study investigating *M. avium* in the community and household water
319 in Philadelphia (Figure 5B) [28]. The pairwise distances between two of the RBH
320 isolates and other UK isolates and isolates from the USA were within the threshold
321 used to define transmission, with the above suggesting potential cross-Atlantic
322 transmission (Supplementary Figure 6A).
323
324 *M. avium hominissuis (MAH)*
325 A total of 198 genomes from 106 patients were characterised as MAH (Figure 2C).
326 Thirty-three patients had CF, 43 BX and 21 had other lung conditions (COPD n = 14;
327 ILD n = 1; ABPA n = 5; sarcoidosis n = 1; asthma n = 11). Four patients had no pre-
328 existing respiratory disease. Disease metadata were unavailable for the remaining five
329 patients. Genomic clustering identified 17 clusters and seven putative transmission
330 clusters containing between two and 16 patients (Supplementary Table 4). The
331 largest cluster of 16 isolates is shown in Figure 3C and comprised patients with CF
332 (n = 4), BX (n = 7), COPD (n = 3) and a single patient with no pre-existing lung condition
333 (metadata were unavailable for one patient). Two pairs of patients in each of the
334 clusters, MAH_FB8_1 and MAH_FB14_9, were found to have epidemiological links.
335

336 Of the 20 defined fastBAPS clusters in the global collection, FB6 and FB11 had lower
337 median pairwise SNP distances and contained RBH isolates (Figure 4C and
338 Supplementary Figure 5C). Both clusters were composed completely or near-
339 completely of isolates collected in the UK (Figure 5C and Figure 5D). The star-like
340 structure of the FB6 phylogeny is suggestive of a point source outbreak and applying
341 a transmission SNP threshold of 16 SNPs showed that 41/45 of the isolates in FB6
342 would form a transmission cluster that included both isolates from this study and
343 isolates from other UK studies (Supplementary Figure 6B). Two small transmission
344 clusters, containing RBH and other UK isolates, were also identified in FB11.
345
346 *M. chimaera*
347 There were 359 sequenced isolates collected from 155 patients identified as *M.*
348 *chimaera*; 37 of the patients had CF, 60 had BX whilst 50 patients had other lung
349 conditions (COPD n = 24; ILD n =10; ABPA n = 11; pleural thickening with enfolded
350 lung n = 1; lung non-mucinous adenocarcinoma n = 1; Wegener's granulomatosis n
351 = 1; *M. tuberculosis* in lymph nodes during same endoscopy procedure n = 1; primary
352 ciliary dyskinesia n = 1; asthma n = 17). There were five patients with no pre-existing
353 respiratory disease and metadata were unavailable for a further nine patients. A single
354 predominant lineage, Mc_FB1, corresponding to the previously characterised Group
355 1 accounted for 332/359 (92.5%) of the sequenced isolates with the majority of the
356 remaining isolates (25/359; 7.0%) forming a second main lineage Mc_FB2
357 corresponding to Group 2 (Figure 2D). Thirteen putative transmission clusters
358 containing two to 106 patients were identified (Supplementary Table 5). Twelve
359 potential epidemiological links (Supplementary Table 6) were found in transmission

360 cluster Mc_RB_FB3_1, the largest transmission cluster of 106 patients. Of the
361 patients in this cluster 43 had BX, 24 CF, 13 COPD, 14 asthma, seven ILD, two ABPA,
362 one pleural thickening with enfolded lung, one lung non-mucinous adenocarcinoma
363 and one primary ciliary dyskinesia. Three patients had no pre-existing lung condition
364 and disease status was missing for five patients. Fifteen patients underwent antibiotic
365 treatment during the study period (Supplementary Figure 4).
366
367 The global phylogenetic tree containing 155 isolates from RBH and 671 previously
368 published isolates was topologically similar to the tree built using only the isolates
369 from RBH with a single major clade with low genetic diversity present (FB6; Figure
370 6A). Isolates from the two clusters with more than ten representatives (FB5 and FB6)
371 were remapped to local references. Recombination was removed and new
372 phylogenetic trees constructed (Figure 6B [FB6] and Figure 6C [FB5]). Most isolates
373 in FB5 were collected from patients with respiratory conditions or else patients with
374 infections following cardiac surgeries. There was, however, a distinct sub-clade
375 rooting within this diversity containing only isolates from HCUs or hospital water
376 supplies (Figure 6C). Within FB6 a more deeply rooting diverse clade contained
377 mainly human isolates and the larger shallower clade rooting within this contained
378 most of the HCU isolates interspersed with human isolates. Due to the large number
379 of isolates assigned to FB6 (n = 765), further genomic clustering was performed using
380 fastBAPS to identify six sub-clusters within FB6. New mapping was performed for
381 the four clusters with more than ten isolates (FB6_FB3 to FB6_FB6; Figure 7A-D;).
382 Three of these clusters, FB6_FB3, FB6_FB4 and FB6_FB5, were completely or nearly
383 completely composed of patient isolates (Figure 7B-D). Most of the isolates collected

384 from HCUs or water supplies made up the bulk of FB6_FB6, though human isolates
385 were distributed throughout the tree (Figure 7A). Putative transmission clusters were
386 calculated for all five fastBAPS lineages described above used a pairwise SNP
387 threshold of 30 SNPs (Supplementary Table 7). Transmission clusters containing
388 RBH isolates were found in each of the lineages with the largest, found in FB6_FB6,
389 comprising 255 HCU isolates, 228 patient isolates and a single isolate from a water
390 supply.
391
392 **Discussion**
393 The aims of this study were to characterise the population structure of MAC and
394 identify potential transmission between CF and non-CF patients. We used a dataset
395 of 1,000 sequenced MAC isolates collected in a London hospital. This is, to date, the
396 largest sequenced MAC isolate collection analysed from a single location. We
397 showed that polymicrobial infection with MAC is not uncommon. The large,
398 longitudinal dataset enabled us to identify putative transmission clusters for all four
399 species/sub-species we considered. Most of these transmission clusters contained
400 both patients with CF and non-CF patients, showing that transmission of MAC is
401 occurring despite the presence of hospital-based control measures focussing on
402 patients with CF. The lack of clear epidemiological links suggests that the majority
403 of transmission is not due to direct contact but rather through an alternative route
404 such as an environmental vector or asymptomatic carriage in the wider population.
405 Comparison of the isolates collected as part of this study with published isolates
406 identified a small number of lineages likely circulating in the UK. With the exception
407 of the major HCU-associated lineage in *M. chimaera*, the comparison did not identify

408 globally circulating lineages like those observed in *M. abscessus*. Surprisingly, given
409 the previous description of the *M. chimaera* HCU clade as representing a point source
410 outbreak, we provide strong evidence that HCU-associated *M. chimaera* isolates are
411 derived from existing lineages circulating amongst patients with pre-existing lung
412 conditions, and that this has happened on at least two separate occasions. This
413 indicates that, while infection and distribution of HCUs has contributed to the
414 transmission of these lineages, they were not the primary source of these lineages,
415 which were already circulating in the human population.
416
417 During the isolate sampling period we identified 46 patients infected with more than
418 one species or lineage. This suggests that many of the patients had polymicrobial
419 infections and that certain lineages were preferentially sampled and cultured at
420 different times. This has been observed previously in other longitudinal samples of
421 patients with MAC lung disease (37). These observations suggest that single colony
422 picks during culturing potentially may be underrepresenting the underlying diversity
423 of MAC species in patients with MAC lung disease. Therefore, to better understand
424 the ecology and transmission of MAC species in the lung, a deep sequencing
425 approach in which colony sweeps rather than colony picks are utilised should be
426 considered for future studies.
427
428 Previous studies have used SNP thresholds of 15-25 SNPs to identify putative
429 transmission clusters in NTM species (13,17,32). Rather than using a previously
430 applied or arbitrary threshold, we used a method that had previously been used to
431 define SNP thresholds in *Staphylococcus aureus* (54). This allowed us to calculate a

432 SNP threshold for each species using the longitudinal isolates collected by the RBH.
433 The thresholds calculated for *M. intracellulare* and MAH of 16 SNPs were similar to
434 those previously applied. Whilst we obtained higher thresholds of 30 SNPs and 58
435 SNPs for *M. chimaera* and MAA respectively, this did not result in an inflated number
436 of transmission clusters, as applying a threshold of 16 SNPs to the MAA dataset
437 would have resulted in only two fewer clusters. These larger values were
438 predominantly due to high levels of within-host diversity observed amongst patient
439 isolates in single clusters in both MAA and *M. chimaera*. For instance, in MAA_FB14
440 isolates from a single patient collected a day apart were as much as 101 SNPs
441 different. This highlighted the likelihood of polymicrobial infections being common
442 amongst the patients we included in the study.
443
444 Using the SNP thresholds calculated for each species, we were able to identify
445 putative transmission clusters amongst all four of the MAC species/subspecies we
446 considered. Broadly, two different patterns of infection were observed: amongst *M.*
447 *intracellulare* and the two *M. avium* subspecies, our results suggested that there have
448 been multiple independent acquisitions followed by limited transmission. Our analysis
449 of *M. chimaera*, however, identified a single large transmission cluster was
450 responsible for infecting 106/155 (68.4%) of our patients infected with this species.
451 Most of these transmission clusters contained both CF and non-CF patients and a
452 small number also contained patients with no pre-existing lung conditions.
453
454 Using our criteria, we were unable to identify potential epidemiological links for most
455 patients included in the transmission clusters. The exception to this was the largest

456 *M. chimaera* transmission cluster where we found 12 potential epidemiological links
457 up to a year before sampling began. Interestingly, four patients were part of more
458 than one link, with patient 218 overlapping with other patients in this transmission
459 cluster on the same wards six times over a nine-month period. The tight infection
460 control surrounding patients with CF means that they are unlikely to have interacted
461 with patients with other lung conditions whilst in hospital. This implies that
462 transmission is likely occurring *via* other pathways, such as through environmental
463 intermediates or through a reservoir of healthy, asymptomatic carriers in the wider
464 population.
465
466 This observation is not unique and for some time, there has been considerable debate
467 whether NTM transmission between patients is occurring, even when transmission is
468 strongly supported by genome-based analysis (13,17,57). The near complete
469 absence of epidemiological links between patients in transmission clusters as well as
470 the ubiquitous presence of NTM species in the environment, especially water
471 supplies, has led some researchers to suggest that the dissemination of NTM
472 lineages at a national level is associated with exposure to contaminated water
473 supplies (17). Whilst there is certainly strong evidence for the same lineages being
474 isolated from water supplies and patients in the same location [28], the absence of a
475 single national water supply in a comparatively small country like the UK would
476 suggest that closely related isolates collected from different geographical locations
477 are unlikely to be due to a single contaminated water supply. It is however possible
478 that transmission networks may include local water supplies as intermediates. The
479 alternative hypothesis that NTM populations are being maintained in apparently

480 healthy individuals with no symptoms of NTM lung disease is a compelling one. This
481 would explain the presence of local transmission clusters in the absence of direct
482 epidemiological links as well as the presence of phylogeographical structure in NTM
483 trees. To better address how NTM species are being transmitted, future studies
484 should focus on attempting to collect isolates from the general population, perhaps
485 focussing on smokers where there is some evidence that they may provide a better
486 reservoir for NTM species and other opportunistic pathogens (18).
487
488 The inclusion of published genomes alongside the RBH isolates allowed us to use
489 our defined pairwise SNP thresholds to identify putative MAC lineages circulating in
490 the UK and globally. Apart from the *M. chimaera* clusters that we found associated
491 with HCUs, we were unable to identify large globally circulating clusters like those
492 observed in *M. abscessus*. This potentially suggests different evolutionary dynamics
493 amongst the MAC species or, in the case of *M. intracellulare*, may simply reflect a
494 lack of comprehensive sampling. We were only able to identify a single cluster
495 containing RBH isolates amongst the two *M. avium* subspecies that included isolates
496 from outside the UK. This cluster comprised the majority of isolates in FB15 (Figure
497 5B). Given that most available *M. avium* genomes have been collected in the UK or
498 USA, a more comprehensive global sampling strategy could potentially reveal the
499 existence of globally circulating clones.
500
501 A different picture emerged when we calculated global transmission clusters for *M.*
502 *chimaera* using a pairwise SNP threshold of 30 SNPs (Supplementary Table 7). Here
503 we found most clusters consisted of isolates from across Europe or else included

504 isolates from across the globe. Notably, we identified a single large transmission
505 cluster containing 484 isolates from 12 countries that included most of the HCU-
506 associated isolates.
507
508 Previous work investigating the global population structure of *M. chimaera* has
509 suggested that the majority of *M. chimaera* transmission globally was associated with
510 point source contamination of HCUs during manufacturing which then led to onward
511 transmission to patients undergoing cardiac surgery (12). In our study we have
512 assembled the largest *M. chimaera* dataset analysed to date, including isolates
513 collected from HCUs and cardiac patients as well as from patients with lung
514 conditions such as CF and BX. This has enabled us to investigate the global
515 population structure of *M. chimaera* in more detail than has previously been possible.
516 Strikingly, we discovered clear evidence for distinct lineages within the global
517 phylogeny that were almost exclusively associated with patient isolates. In fact most
518 of the HCU-associated isolates were clustered together in a single lineage that was
519 derived from the lineages containing only human isolates (Figure 6B). This was also
520 true of a second, smaller, lineage that contained only HCU and environmental
521 isolates. The results suggest that the predominant HCU-associated lineage is
522 descended from a *M. chimaera* population already circulating amongst patients with
523 respiratory diseases such as CF, and that a similar derivation has happened at least
524 once more. Unfortunately, due to the absence of even a moderate temporal signal
525 (data not shown), we were unable to place any estimates of when these lineages first
526 emerged and whether these dates occurred before the infected HCUs were first
527 manufactured.

528

529 In conclusion, using a large longitudinal sample of MAC isolates collected within a
530 single hospital as well as previously published global contextual isolates, we were
531 able to identify polymicrobial infections by different MAC species and lineages in
532 multiple patients as well as identify transmission clusters circulating within the
533 hospital, at a national level and, less commonly, globally. We show that the origin of
534 the major HCU-associated *M. chimaera* lineage is likely a previously circulating
535 lineage associated with patients with respiratory diseases. To better understand the
536 transmission dynamics and to prevent continuing circulation of MAC and other NTM
537 species, future studies should include sampling of environmental reservoirs and
538 potential asymptomatic carriers.

539

540 **Author contributions**

541 **Conceptualisation**

542 MM, WC, JP

543 **Data curation**

544 HE, KK, AvT, CC

545 **Formal analysis**

546 AvT, HE, KK

547 **Funding acquisition**

548 WC, MM, ML, JP

549 **Investigation**

550 AvT, HE, KK, CC

551 **Methodology**

AvT, CC

Project administration

AvT, CC, MM, WC

Resources

CC, NR, SB

Software

NA

Supervision

JP, WC, MM, ML

Validation

AvT

Visualisation

AvT

Writing – original draft

AvT, JP

Writing – review & editing

AvT, HE, CC, KK, MM, ML, JP, WC

Conflicts of interest

JP is a paid consultant to NextGen Diagnostics LLc.

Data sharing

574 Raw sequencing reads were deposited at the European Nucleotide Archive under
575 project PRJEB21813; all accessions used in this project are listed in Supplementary
576 File 1.
577

578 **References**

579 1. Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, et al.
580 An official ATS/IDSA statement: diagnosis, treatment, and prevention of
581 nontuberculous mycobacterial diseases. Am J Respir Crit Care Med. 2007 Feb
582 15;175(4):367–416.

583 2. Chan ED, Bai X, Kartalija M, Orme IM, Ordway DJ. Host immune response to
584 rapidly growing mycobacteria, an emerging cause of chronic lung disease. Am J
585 Respir Cell Mol Biol. 2010 Oct;43(4):387–93.

586 3. Iseman MD, Marras TK. The importance of nontuberculous mycobacterial lung
587 disease. Am J Respir Crit Care Med. 2008 Nov 15;178(10):999–1000.

588 4. Donohue MJ, Wymer L. Increasing Prevalence Rate of Nontuberculous
589 Mycobacteria Infections in Five States, 2008-2013. Ann Am Thorac Soc. 2016
590 Dec;13(12):2143–50.

591 5. Moore JE, Kruijshaar ME, Ormerod LP, Drobniewski F, Abubakar I. Increasing
592 reports of non-tuberculous mycobacteria in England, Wales and Northern
593 Ireland, 1995-2006. BMC Public Health. 2010 Oct 15;10:612.

594 6. Philley JV, Griffith DE. Treatment of slowly growing mycobacteria. Clin Chest
595 Med. 2015 Mar;36(1):79–90.

596 7. Adjemian J, Olivier KN, Prevots DR. Epidemiology of Pulmonary Nontuberculous
597 Mycobacterial Sputum Positivity in Patients with Cystic Fibrosis in the United
598 States, 2010-2014. Ann Am Thorac Soc. 2018 Jul;15(7):817–26.

599 8. Hoefsloot W, van Ingen J, Andrejak C, Angeby K, Bauriaud R, Bemer P, et al.
600 The geographic diversity of nontuberculous mycobacteria isolated from
601 pulmonary samples: an NTM-NET collaborative study. Eur Respir J. 2013
602 Dec;42(6):1604–13.

603 9. Olivier KN, Weber DJ, Wallace RJ, Faiz AR, Lee J-H, Zhang Y, et al.
604 Nontuberculous mycobacteria. I: multicenter prevalence study in cystic fibrosis.
605 Am J Respir Crit Care Med. 2003 Mar 15;167(6):828–34.

606 10. Achermann Y, Rössle M, Hoffmann M, Deggim V, Kuster S, Zimmermann DR,
607 et al. Prosthetic valve endocarditis and bloodstream infection due to
608 Mycobacterium chimaera. J Clin Microbiol. 2013 Jun;51(6):1769–73.

609 11. Sax H, Bloemberg G, Hasse B, Sommerstein R, Kohler P, Achermann Y, et
610 al. Prolonged Outbreak of *Mycobacterium chimaera* Infection After Open-Chest
611 Heart Surgery. Clin Infect Dis Off Publ Infect Dis Soc Am. 2015 Jul 1;61(1):67–
612 75.

613 12. van Ingen J, Kohl TA, Kranzer K, Hasse B, Keller PM, Katarzyna Szafrańska
614 A, et al. Global outbreak of severe *Mycobacterium chimaera* disease after
615 cardiac surgery: a molecular epidemiological study. Lancet Infect Dis. 2017
616 Oct;17(10):1033–41.

617 13. Bryant JM, Grogono DM, Greaves D, Foweraker J, Roddick I, Inns T, et al.
618 Whole-genome sequencing to identify transmission of *Mycobacterium*
619 *abscessus* between patients with cystic fibrosis: a retrospective cohort study.
620 Lancet Lond Engl. 2013 May 4;381(9877):1551–60.

621 14. Bryant JM, Grogono DM, Rodriguez-Rincon D, Everall I, Brown KP, Moreno
622 P, et al. Emergence and spread of a human-transmissible multidrug-resistant
623 nontuberculous mycobacterium. Science. 2016 Nov 11;354(6313):751–7.

624 15. Everall I, Nogueira CL, Bryant JM, Sánchez-Busó L, Chimara E, Duarte R da
625 S, et al. Genomic epidemiology of a national outbreak of post-surgical
626 *Mycobacterium abscessus* wound infections in Brazil. Microb Genomics. 2017
627 May;3(5):e000111.

628 16. Aitken ML, Limaye A, Pottinger P, Whimbey E, Goss CH, Tonelli MR, et al.
629 Respiratory outbreak of *Mycobacterium abscessus* subspecies massiliense in a
630 lung transplant and cystic fibrosis center. Am J Respir Crit Care Med. 2012 Jan
631 15;185(2):231–2.

632 17. Lipworth S, Hough N, Weston N, Muller-Pebody B, Phin N, Myers R, et al.
633 Epidemiology of *Mycobacterium abscessus* in England: an observational study.
634 Lancet Microbe. 2021 Oct;2(10):e498–507.

635 18. Ruis C, Bryant JM, Bell SC, Thomson R, Davidson RM, Hasan NA, et al.
636 Dissemination of *Mycobacterium abscessus* via global transmission networks.
637 Nat Microbiol. 2021 Oct;6(10):1279–88.

638 19. Babraham Bioinformatics - FastQC A Quality Control tool for High
639 Throughput Sequence Data [Internet]. [cited 2021 Dec 14]. Available from:
640 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

641 20. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence
642 classification using exact alignments. Genome Biol. 2014 Mar 3;15(3):R46.

643 21. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species
644 abundance in metagenomics data. PeerJ Comput Sci. 2017 Jan 2;3:e104.

645 22. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
646 sequence data. Bioinforma Oxf Engl. 2014 Aug 1;30(15):2114–20.

647 23. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
648 transform. Bioinforma Oxf Engl. 2009 Jul 15;25(14):1754–60.

649 24. Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, et al.
650 Evolution of MRSA during hospital transmission and intercontinental spread.
651 Science. 2010 Jan 22;327(5964):469–74.

652 25. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-
653 sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb
654 Genomics. 2016 Apr;2(4):e000056.

655 26. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and
656 effective stochastic algorithm for estimating maximum-likelihood phylogenies.
657 Mol Biol Evol. 2015 Jan;32(1):268–74.

658 27. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al.
659 Rapid phylogenetic analysis of large samples of recombinant bacterial whole
660 genome sequences using Gubbins. Nucleic Acids Res. 2015 Feb 18;43(3):e15.

661 28. Tonkin-Hill G. pairsnp [Internet]. 2021 [cited 2021 Dec 14]. Available from:
662 https://github.com/gtonkinhill/pairsnp

663 29. Caverly LJ, Spilker T, LiPuma JJ. Complete Genome Sequence of
664 *Mycobacterium abscessus* subsp. bolletii. Genome Announc. 2016 Jun
665 9;4(3):e00543-16.

666 30. Hasan NA, Davidson RM, Epperson LE, Kammlade SM, Rodger RR, Levin
667 AR, et al. Population Genomics of Nontuberculous Mycobacteria Recovered
668 from United States Cystic Fibrosis Patients [Internet]. 2019 Jun [cited 2021 Dec
669 14] p. 663559. Available from:
670 https://www.biorxiv.org/content/10.1101/663559v1

671 31. Bryant JM, Thibault VC, Smith DGE, McLuckie J, Heron I, Sevilla IA, et al.
672 Phylogenomic exploration of the relationships between strains of
673 *Mycobacterium avium* subspecies paratuberculosis. BMC Genomics. 2016 Jan
674 26;17:79.

675 32. Lande L, Alexander DC, Wallace RJ, Kwait R, Iakhiaeva E, Williams M, et al.
676 *Mycobacterium avium* in Community and Household Water, Suburban
677 Philadelphia, Pennsylvania, USA, 2010-2012. Emerg Infect Dis. 2019
678 Mar;25(3):473–81.

679 33. Yoon J-K, Kim TS, Kim J-I, Yim J-J. Whole genome sequencing of
680 Nontuberculous Mycobacterium (NTM) isolates from sputum specimens of co-
681 habiting patients with NTM pulmonary disease and NTM isolates from their
682 environment. BMC Genomics. 2020 Apr 23;21(1):322.

683 34. Votintseva AA, Pankhurst LJ, Anson LW, Morgan MR, Gascoyne-Binzi D,
684 Walker TM, et al. Mycobacterial DNA extraction for whole-genome sequencing

685 from early positive liquid (MGIT) cultures. J Clin Microbiol. 2015 Apr;53(4):1137–
686 43.

687 35. Pfeiffer W, Braun J, Burchell J, Witte CL, Rideout BA. Whole-genome
688 analysis of mycobacteria from birds at the San Diego Zoo. PloS One.
689 2017;12(3):e0173464.

690 36. Quan TP, Bawa Z, Foster D, Walker T, Del Ojo Elias C, Rathod P, et al.
691 Evaluation of Whole-Genome Sequencing for Mycobacterial Species
692 Identification and Drug Susceptibility Testing in a Clinical Setting: a Large-Scale
693 Prospective Assessment of Performance against Line Probe Assays and
694 Phenotyping. J Clin Microbiol. 2018 Feb;56(2):e01480-17.

695 37. Operario DJ, Pholwat S, Koeppel AF, Prorock A, Bao Y, Sol-Church K, et al.
696 *Mycobacterium avium* Complex Diversity within Lung Disease, as Revealed by
697 Whole-Genome Sequencing. Am J Respir Crit Care Med. 2019 Aug
698 1;200(3):393–6.

699 38. Rivas AE, Hollinger C, Oehler DA, Robbe-Austerman S, Paré JA. DIAGNOSIS
700 AND MANAGEMENT OF MYCOBACTERIOSIS IN A COLONY OF LITTLE
701 PENGUINS (EUDYPTULA MINOR). J Zoo Wildl Med Off Publ Am Assoc Zoo Vet.
702 2019 Jun 13;50(2):427–36.

703 39. Chand M, Lamagni T, Kranzer K, Hedge J, Moore G, Parks S, et al. Insidious
704 Risk of Severe *Mycobacterium chimaera* Infection in Cardiac Surgery Patients.
705 Clin Infect Dis Off Publ Infect Dis Soc Am. 2017 Feb 1;64(3):335–42.

706 40. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, et
707 al. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-
708 genome sequencing: a prospective study. Lancet Respir Med. 2016 Jan;4(1):49–
709 58.

710 41. Epperson LE, Timke M, Hasan NA, Godo P, Durbin D, Helstrom NK, et al.
711 Evaluation of a Novel MALDI Biotyper Algorithm to Distinguish *Mycobacterium*
712 *intracellulare* From *Mycobacterium chimaera*. Front Microbiol. 2018;9:3140.

713 42. Acosta F, Pérez-Lago L, Ruiz Serrano MJ, Marín M, Kohl TA, Lozano N, et al.
714 Fast update of undetected *Mycobacterium chimaera* infections to reveal
715 unsuspected cases. J Hosp Infect. 2018 Dec;100(4):451–5.

716 43. Ghodousi A, Borroni E, Peracchi M, Palù G, Fallico L, Rassu M, et al.
717 Genomic analysis of cardiac surgery-associated *Mycobacterium chimaera*
718 infections in Italy. PloS One. 2020;15(9):e0239273.

719 44. Mac Aogáin M, Roycroft E, Raftery P, Mok S, Fitzgibbon M, Rogers TR. Draft
720 Genome Sequences of Three *Mycobacterium chimaera* Respiratory Isolates.
721 Genome Announc. 2015 Dec 3;3(6):e01409-15.

722 45. Perkins KM, Lawsin A, Hasan NA, Strong M, Halpin AL, Rodger RR, et al.
723 Notes from the Field: *Mycobacterium chimaera* Contamination of Heater-Cooler

724 Devices Used in Cardiac Surgery - United States. MMWR Morb Mortal Wkly
725 Rep. 2016 Oct 14;65(40):1117–8.

726 46. Svensson E, Jensen ET, Rasmussen EM, Folkvardsen DB, Norman A,
727 Lillebaek T. *Mycobacterium chimaera* in Heater-Cooler Units in Denmark Related
728 to Isolates from the United States and United Kingdom. Emerg Infect Dis. 2017
729 Mar;23(3):507–9.

730 47. Williamson D, Howden B, Stinear T. *Mycobacterium chimaera* Spread from
731 Heating and Cooling Units in Heart Surgery. N Engl J Med. 2017 Feb
732 9;376(6):600–2.

733 48. Zhang X, Lin J, Feng Y, Wang X, McNally A, Zong Z. Identification of
734 *Mycobacterium chimaera* in heater-cooler units in China. Sci Rep. 2018 May
735 18;8(1):7843.

736 49. Page AJ, De Silva N, Hunt M, Quail MA, Parkhill J, Harris SR, et al. Robust
737 high-throughput prokaryote de novo assembly and improvement pipeline for
738 Illumina data. Microb Genomics. 2016 Aug;2(8):e000083.

739 50. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
740 SPAdes: a new genome assembly algorithm and its applications to single-cell
741 sequencing. J Comput Biol J Comput Mol Cell Biol. 2012 May;19(5):455–77.

742 51. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-
743 assembled contigs using SSPACE. Bioinforma Oxf Engl. 2011 Feb 15;27(4):578–
744 9.

745 52. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller.
746 Genome Biol. 2012 Jun 25;13(6):R56.

747 53. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical
748 Bayesian analysis of population structure. Nucleic Acids Res. 2019 Jun
749 20;47(11):5539–49.

750 54. Coll F, Raven KE, Knight GM, Blane B, Harrison EM, Leek D, et al. Definition
751 of a genetic relatedness cutoff to exclude recent transmission of meticillin-
752 resistant *Staphylococcus aureus*: a genomic epidemiology analysis. Lancet
753 Microbe. 2020 Dec;1(8):e328–35.

754 55. R Core Team (2020). — European Environment Agency [Internet]. [cited 2021
755 Dec 14]. Available from: https://www.eea.europa.eu/data-and-
756 maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-
757 team-2006

758 56. Csardi G, Nepusz T. The Igraph Software Package for Complex Network
759 Research. InterJournal. 2005 Nov 30;Complex Systems:1695.

760 57. Bronson RA, Gupta C, Manson AL, Nguyen JA, Bahadirli-Talbott A, Parrish
761 NM, et al. Global phylogenomic analyses of *Mycobacterium abscessus* provide

762 context for non cystic fibrosis infections and the evolution of antibiotic
763 resistance. Nat Commun. 2021 Aug 26;12(1):5145.

764

765 **Tables and Figures**

766 Table 1. *Mycobacterium avium* Complex species identified in Royal Brompton

767 Hospital Collection.

768

Species	Number of genomes
M. avium	405
M. chimaera	359
M. intracellulare	162
Mycobacterium sp. MOTT36Y	39
M. paraintracellulare	15
M. yongonense	10
M. marseillense	4
M. simiae	3
M. colombiense	1
M. chelonae	1
Mycobacterium sp. QIA-37	1

769

770

771

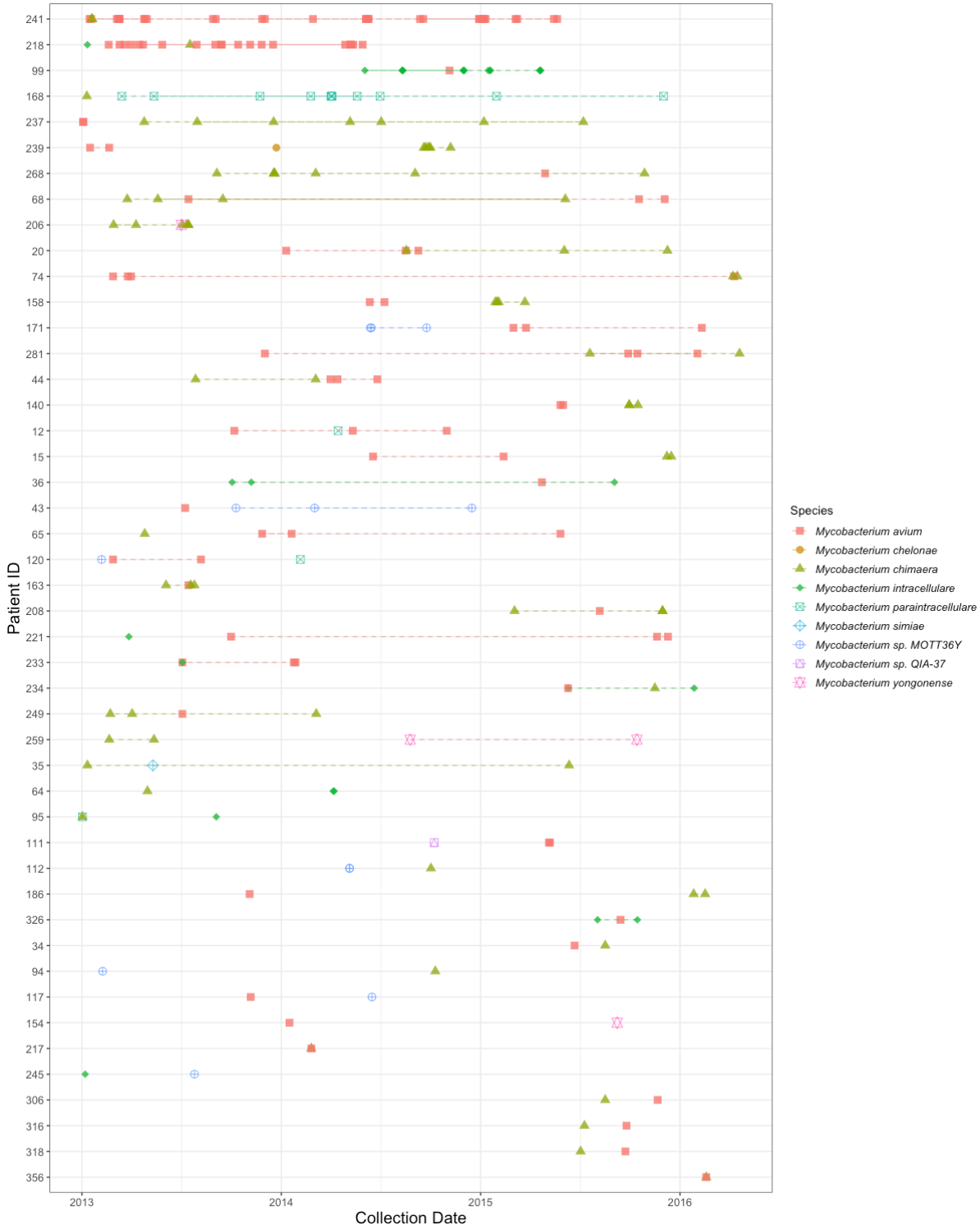


Figure 1. Timeline of patients infected by more than one *Mycobacterium avium* Complex (MAC) species. Each point represents an isolate and is coloured according to the species identified. Isolates from the species in a patient are linked

together. Patients with one isolate or multiple isolates from a single species were excluded.

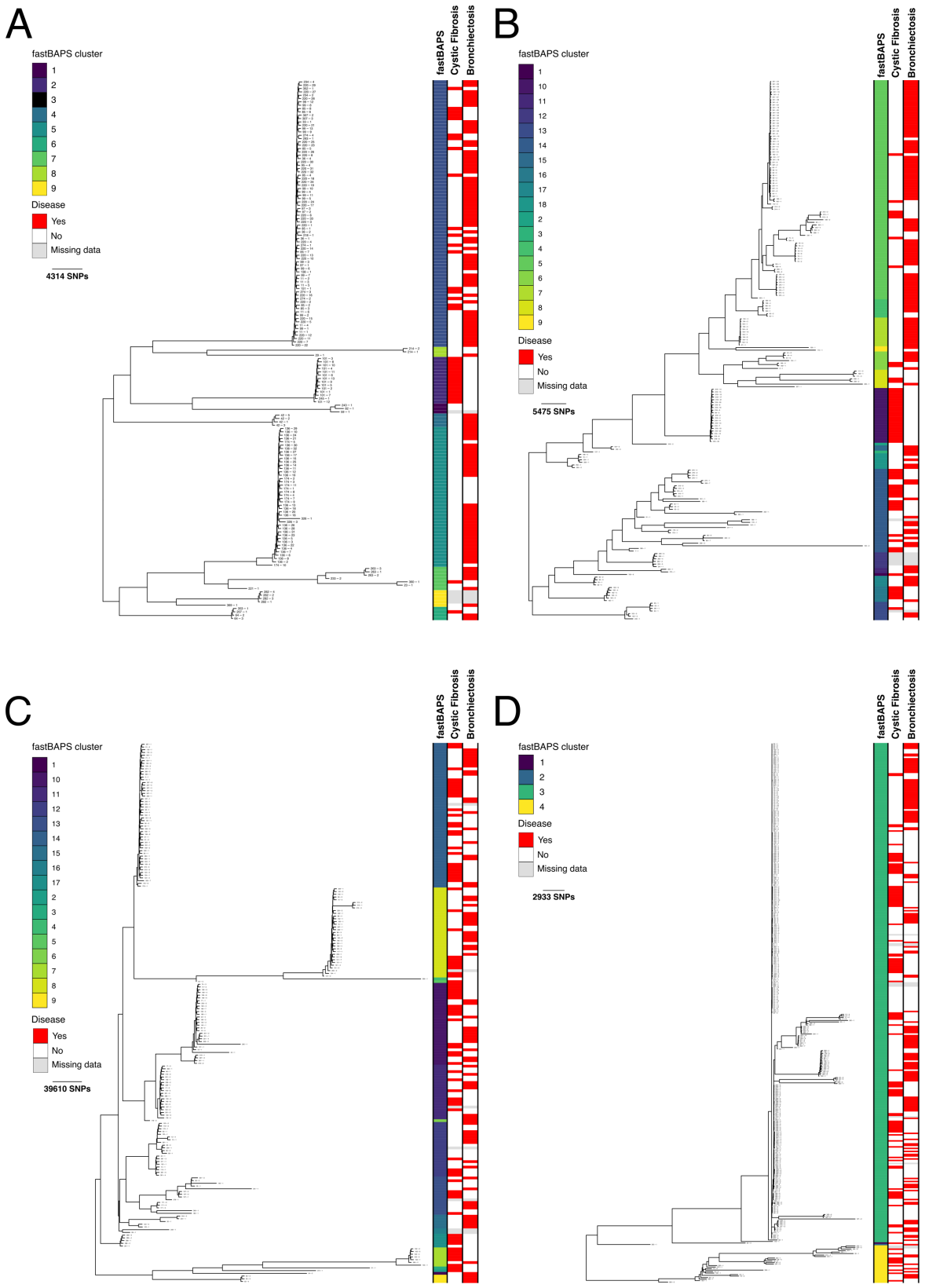
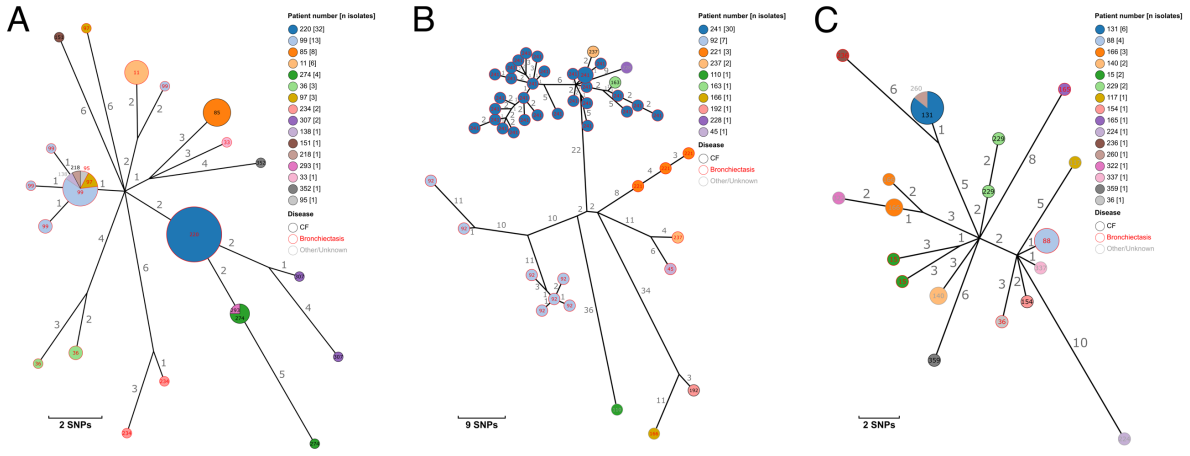


Figure 2: Population structure of *Mycobacterium avium* Complex species collected at the Royal Brompton Hospital. Midpoint-rooted Maximum likelihood

782 phylogenetic trees for A) *Mycobacterium intracellulare* (n = 162); B) *Mycobacterium*
783 *avium avium* (n = 207); C) *Mycobacterium avium hominosuis* (n = 198) and D)
784 *Mycobacterium chimaera* (n = 359). fastBAPS clusters are shown as a coloured
785 datastrip in each plot. The taxa are clustered according to their sequence similarities.
786 The lengths of the branches are scaled in nucleotide substitutions per site. The
787 disease status of the patients (that isolates were collected from) for cystic fibrosis
788 and non-cystic fibrosis bronchiectasis are represented by red bars for Yes and white
789 bars for No with missing patient data shown by gray bars. Scale bars are shown in
790 SNPs per site.

791



792

793

794

795 **Figure 3. Largest transmission clusters amongst isolates collected at the Royal**
796 **Brompton Hospital.** A) *Mycobacterium intracellulare* (Mi_FB3_1); B) *Mycobacterium*
797 *avium avium* (MAA_FB5_1) and C) *Mycobacterium avium hominosuis* (MAH_FB8_1).
798 Each node represents an isolate or isolates identical (0 SNPs) to each other and the
799 size of the node is proportional to the number of identical isolates. Nodes are
800 coloured and labelled by patient number. The outer ring of each node is coloured

according to the disease status of the patient (cystic fibrosis = black, non-cystic fibrosis bronchiectasis, grey = other lung condition). The edges represent the pairwise SNP distance between the isolate(s).

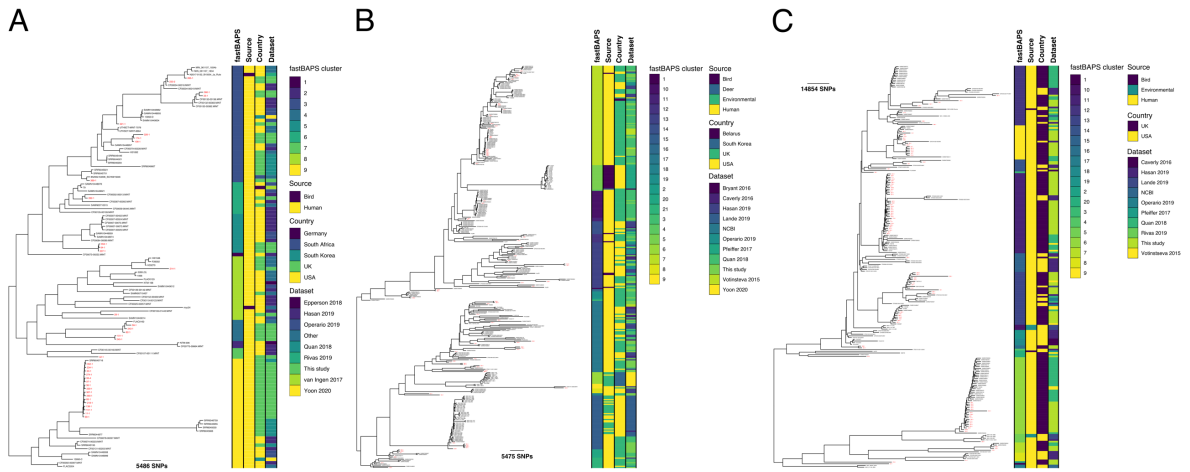


Figure 4. Global population structure of *Mycobacterium avium* Complex species. Midpoint-rooted maximum likelihood phylogenetic trees for A) *Mycobacterium intracellulare* (n = 114); B) *Mycobacterium avium avium* (n = 344) and C) *Mycobacterium avium hominosuis* (n = 236). The taxa are clustered according to their sequence similarities. The lengths of the branches are scaled in nucleotide substitutions per site. fastBAPS cluster, source of isolate, country of collection and study are shown as datastrips to the right of the phylogenies. Isolates from the Royal Brompton Hospital are highlighted in red. Scale bars are shown in SNPs per site.

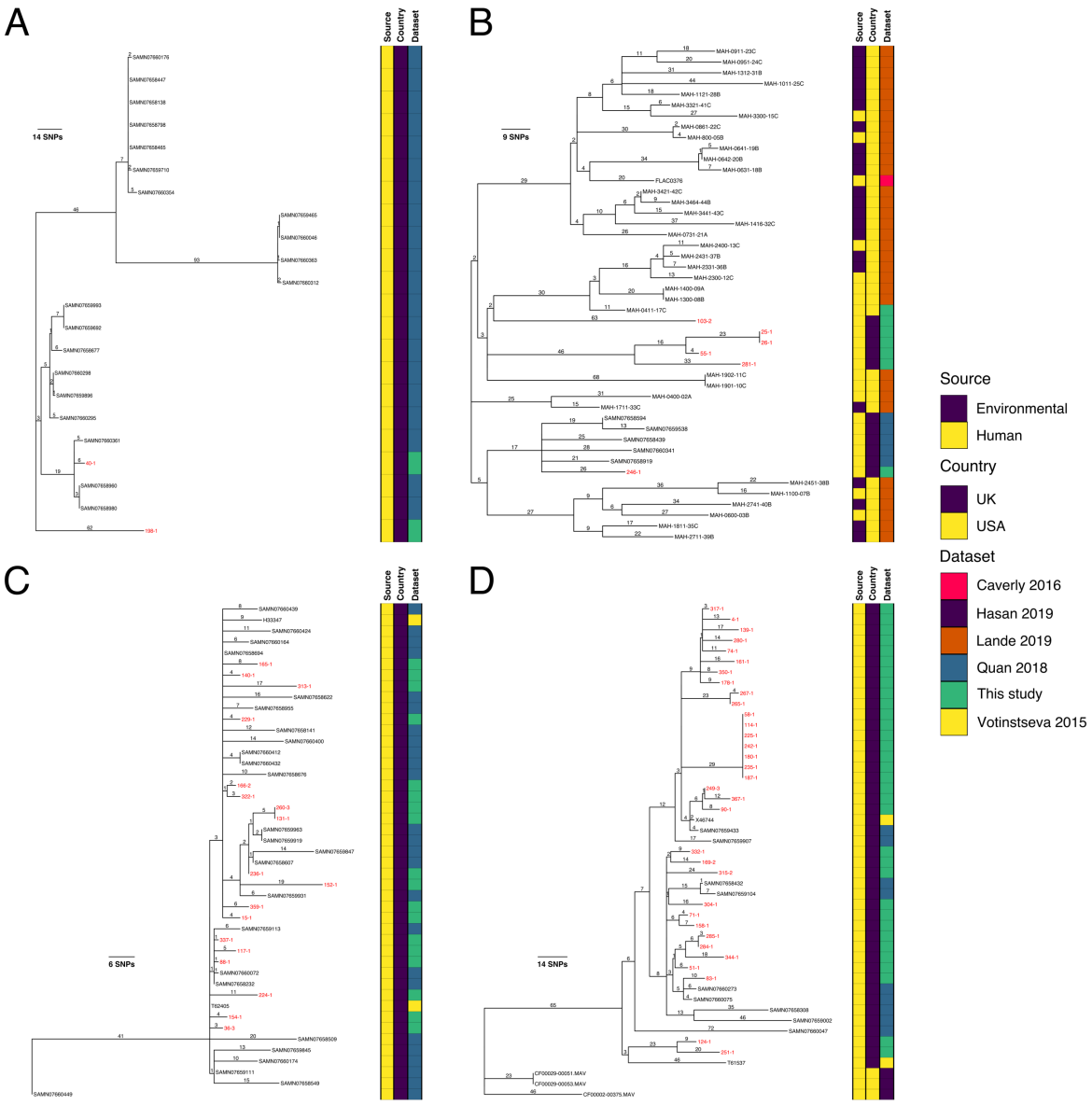
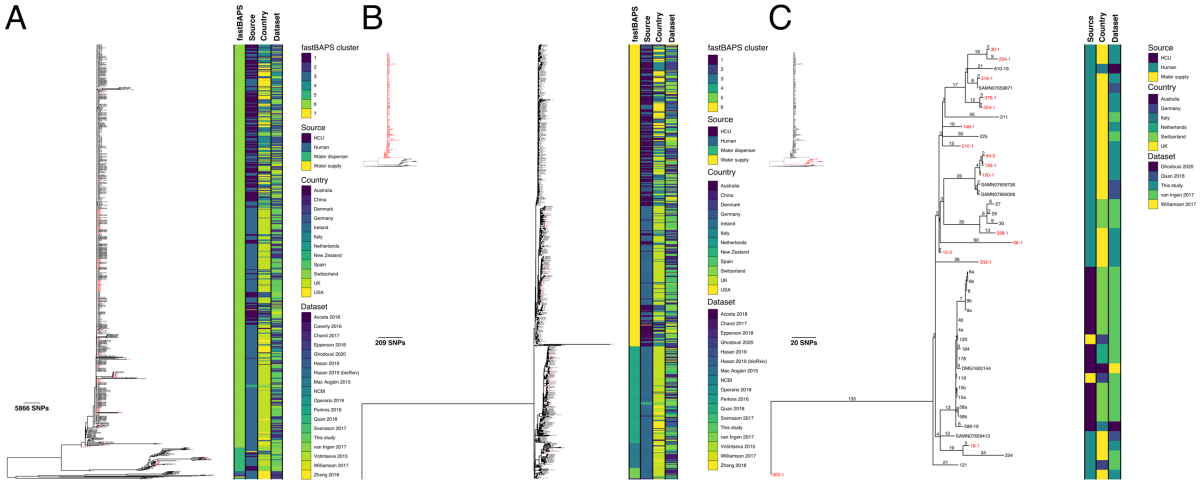


Figure 5: *Mycobacterium avium* global fastBAPS phylogenies. Midpoint-rooted maximum likelihood phylogenetic trees for A) *Mycobacterium avium avium* FB10; B) *Mycobacterium avium avium* FB15; C) *Mycobacterium avium hominosuis* FB6 and D) *Mycobacterium avium hominosuis* FB11. The taxa are clustered according to their sequence similarities. The lengths of the branches are scaled in nucleotide substitutions per site. The source of isolate, country of collection and study are shown

823 as datastrips to the right of the phylogenies. Isolates from the Royal Brompton

824 Hospital are highlighted in red. Branch lengths are shown in SNPs per site.

825

826 

827

828 **Figure 6. Global population structure of *Mycobacterium chimaera.*** Midpoint-

829 rooted maximum likelihood phylogenetic trees for A) all *Mycobacterium chimaera*

830 isolates B) *Mycobacterium chimaera* FB6 C) *Mycobacterium chimaera* FB5. The taxa

831 are clustered according to their sequence similarities. The lengths of the branches

832 are scaled in nucleotide substitutions per site. The source of isolate, country of

833 collection and study are shown as datastrips to the right of the phylogenies. Isolates

834 from the Royal Brompton Hospital are highlighted in red. Branch lengths are shown

835 in SNPs per site for *Mycobacterium chimaera* FB5.

836

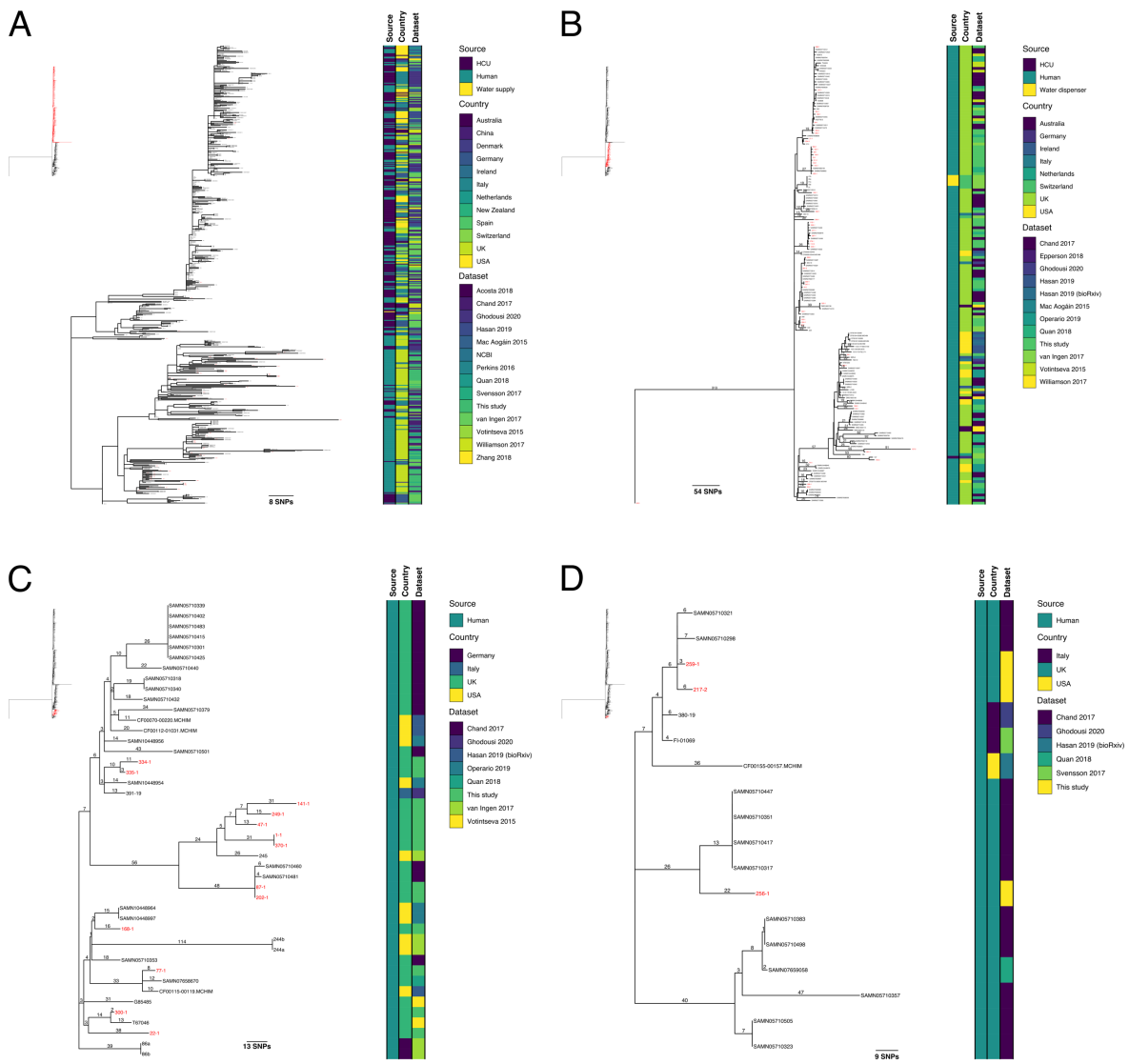


Figure 7: *Mycobacterium chimaera* fastBAPS cluster FB6 phylogenies. Maximum likelihood phylogenetic trees for A) FB6_FB6; B) FB6_FB4; C) FB6_FB3 and D) FB6_FB5. The taxa are clustered according to their sequence similarities. The lengths of the branches are scaled in nucleotide substitutions per site. The source of isolate, country of collection and study are shown as datastrips to the right of the phylogenies. Isolates from the Royal Brompton Hospital are highlighted in red. Branch lengths are shown in SNPs per site.

847 **Supplementary Material**

848 Supplementary Table 1. Reference genomes used in this study

Species/ subspecies	Reference	Length (Mb)	Country of isolation	Year of isolation
M. intracellulare	ATCC 13950	5.4	South Korea	2012
M. avium avium	104	5.5	USA	1983
M. avium hominosuis	TH135	5.0	Japan	2004-2008
M. chimaera	DSM 44623	5.9	Italy	1999-2003

849

850 Supplementary Table 2. *Mycobacterium intracellulare* transmission clusters

fastBAPS cluster	Transmission cluster	Patients (n)	Bronchiectosis (n)	Cystic Fibrosis (n)	Other lung conditions (n)	No pre-existing lung condition (n)	Patients undergoing treatment (n)
Mi_FB2	Mi_FB2_1	2	0	2	0	0	1
Mi_FB3	Mi_FB3_1	16	8	7	1	0	4
Mi_FB5	Mi_FB5_1	3	2	0	1	0	2

851

852

853 Supplementary Table 3. *Mycobacterium avium avium* transmission clusters

fastBAPS cluster	Transmission cluster	Patients (n)	Bronchiectosis (n)	Cystic Fibrosis (n)	Other lung conditions (n)	No pre-existing lung condition (n)	Patients undergoing treatment (n)
MAA_FB5	MAA_FB5_1	10	6	2	2	0	4
MAA_FB5	MAA_FB5_4	2	1	1	0	0	0
MAA_FB5	MAA_FB5_5	2	1	0	1	0	1
MAA_FB5	MAA_FB5_12	2	1	1	0	0	1
MAA_FB7	MAA_FB7_1	2	2	0	0	0	1
MAA_FB10	MAA_FB10_1	2	0	2	0	0	0
MAA_FB14	MAA_FB14_1	2	2	0	0	0	1

854

855

856 Supplementary Table 4. *Mycobacterium avium hominosuis* transmission clusters

fastBAPS cluster	Transmission cluster	Patients (n)	Bronchiectosis (n)	Cystic Fibrosis (n)	Other lung conditions (n)	No pre-existing lung condition (n)	Patients undergoing treatment (n)
MAH_FB8	MAH_FB8_1	16	7	4	3	1	0
MAH_FB10	MAH_FB10_1	4	2	2	0	0	1
MAH_FB11	MAH_FB11_1	6	3	1	2	0	0
MAH_FB12	MAH_FB12_2	4	1	2	1	0	2
MAH_FB14	MAH_FB14_3	4	2	2	0	0	0
MAH_FB14	MAH_FB14_7	2	0	2	0	0	0
MAH_FB14	MAH_FB14_9*	7	3	2	1	0	1

857

858 Supplementary Table 5. *Mycobacterium chimaera* transmission clusters

859

fastBAPS cluster	Transmission cluster	Patients (n)	Bronchiectasis (n)	Cystic Fibrosis (n)	Other lung conditions (n)	No pre-existing lung condition (n)	Patients undergoing treatment (n)
Mc_FB3	Mc_FB3_1	106	43	24	25	3	15
Mc_FB3	Mc_FB3_3	2	1	1	0	0	0
Mc_FB3	Mc_FB3_4	2	2	0	0	0	0
Mc_FB3	Mc_FB3_6	2	1	0	1	0	0
Mc_FB3	Mc_FB3_13	4	1	1	1	1	1
Mc_FB3	Mc_FB3_16	2	0	0	1	1	1
Mc_FB3	Mc_FB3_17	2	1	1	0	0	0
Mc_FB3	Mc_FB3_19	2	0	1	0	1	0
Mc_FB3	Mc_FB3_25	2	1	1	0	0	1

Mc_FB3	Mc_FB3_30*	4	0	1	0	0	0
Mc_FB4	Mc_FB4_1	2	1	1	0	0	1
Mc_FB4	Mc_FB4_2	9	3	3	3	0	2
Mc_FB4	Mc_FB4_3*	6	1	2	2	0	1

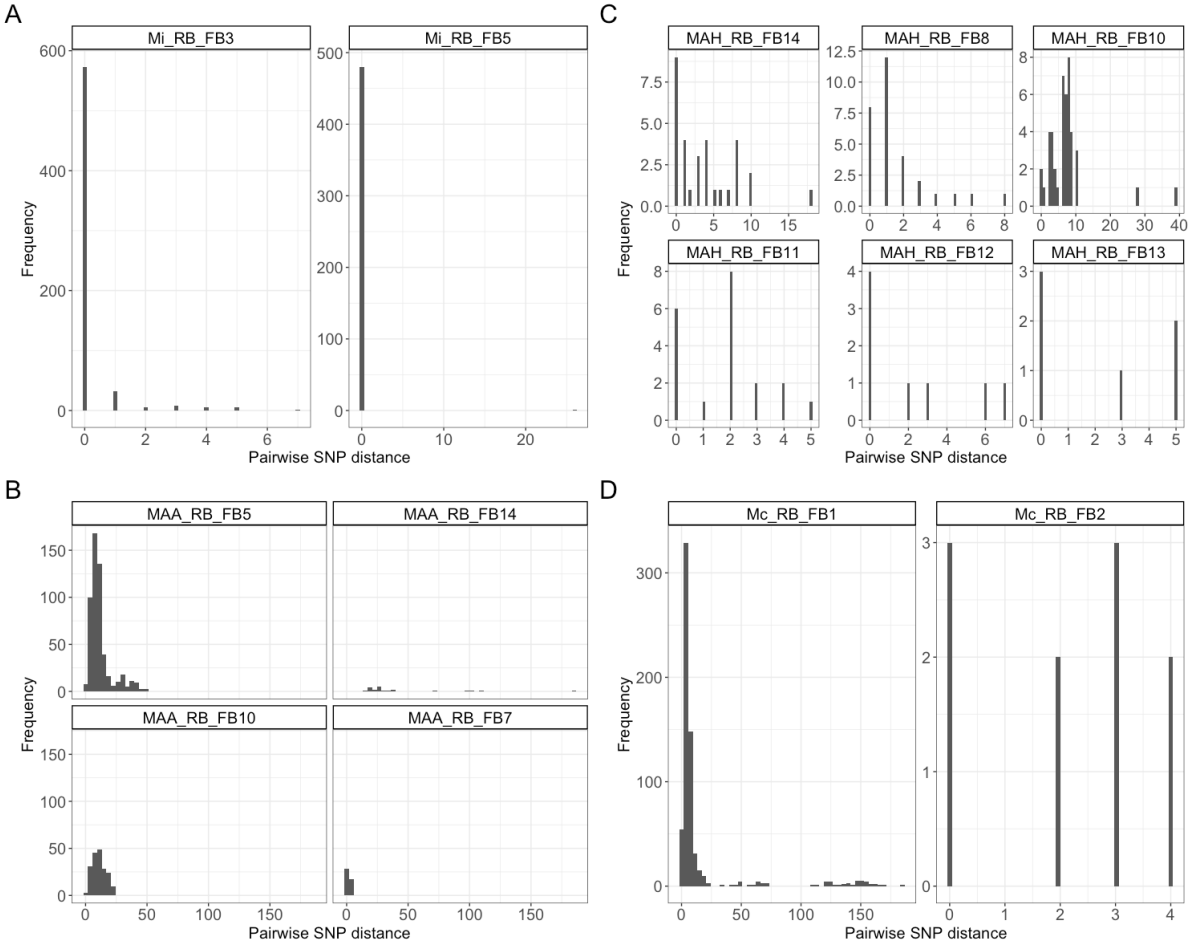
Supplementary Table 6. *Mycobacterium chimaera* epidemiological links

Patients with epidemiological link	Hospital ward
24/218	FOUL
163/218	FOUL
175/218	FOUL
130/218	LIND
76/218	FOUL
32/163	LIND
79/241	LIND
241/272	LIND
177/186	LIND
122/175/306	FOUL
30/122	FOUL

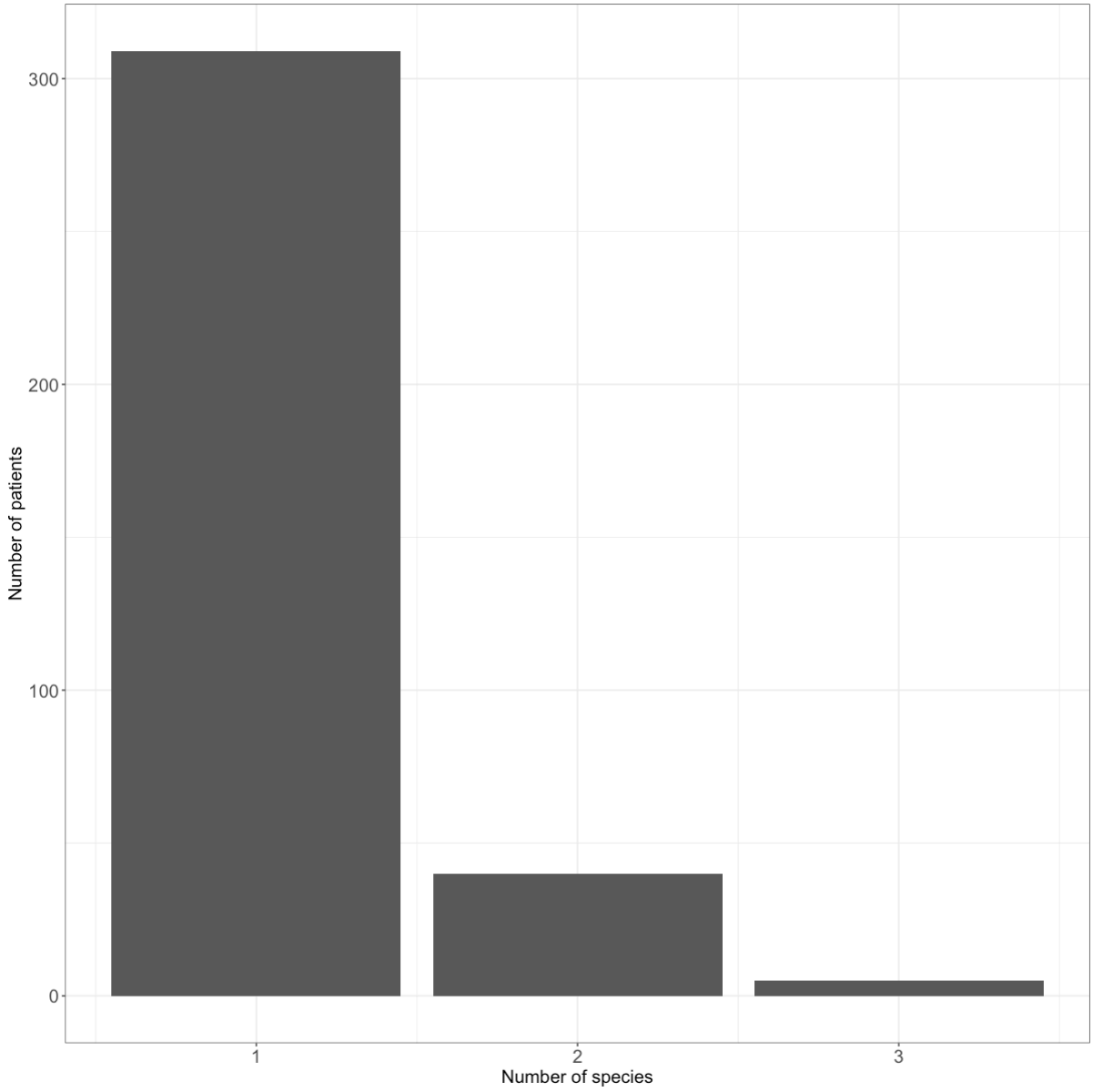
30/218	FOUL

Supplementary Table 7. *Mycobacterium chimaera* global transmission clusters containing RBH isolates

fastBAPS cluster	Transmission cluster	Total isolates (n)	Patients (n)	HCU (n)	Other (n)
FB5	FB5_1	21	4	15	2
FB5	FB5_2	7	7	0	0
FB5	FB5_3	5	5	0	0
FB5	FB5_7	4	4	0	0
FB6_FB3	FB6_FB3_7	4	4	0	0
FB6_FB3	FB6_FB3_9	3	3	0	0
FB6_FB3	FB6_FB3_12	3	3	0	0
FB6_FB3	FB6_FB3_21	2	2	0	0
FB6_FB4	FB6_FB4_1	21	21	0	0
FB6_FB4	FB6_FB4_2	30	30	0	0
FB6_FB4	FB6_FB4_13	10	10	0	0
FB6_FB5	FB6_FB5_1	6	6	0	0
FB6_FB5	FB6_FB5_1	484	228	255	1
FB6_FB5	FB6_FB5_3	3	3	0	0
FB6_FB5	FB6_FB5_5	3	3	0	0
FB6_FB5	FB6_FB5_7	4	4	0	0



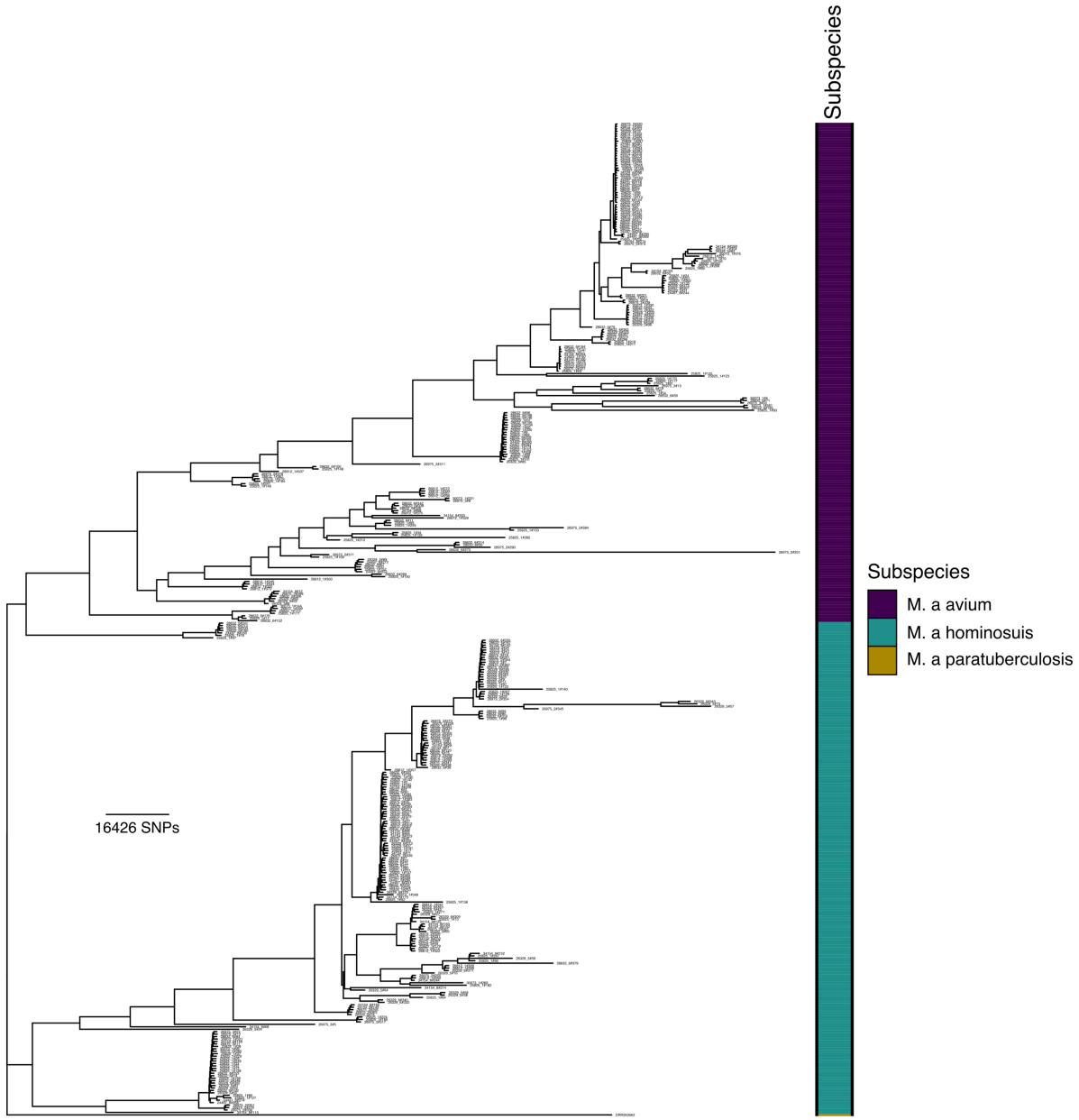
Supplementary Figure 1. Within-host isolate pairwise SNP diversity in fastBAPS clusters for A) *Mycobacterium intracellulare*; B) *Mycobacterium avium avium;* C) *Mycobacterium avium hominosuis* and D) *Mycobacterium chimaera*.



874

875 **Supplementary Figure 2. Histogram showing the number of *Mycobacterium***

876 ***avium* Complex species identified in each patient.**



877

878 **Supplementary Figure 3. Population structure *Mycobacterium avium* at the**

879 **Royal Brompton Hospital.** Maximum likelihood phylogenetic tree of 406

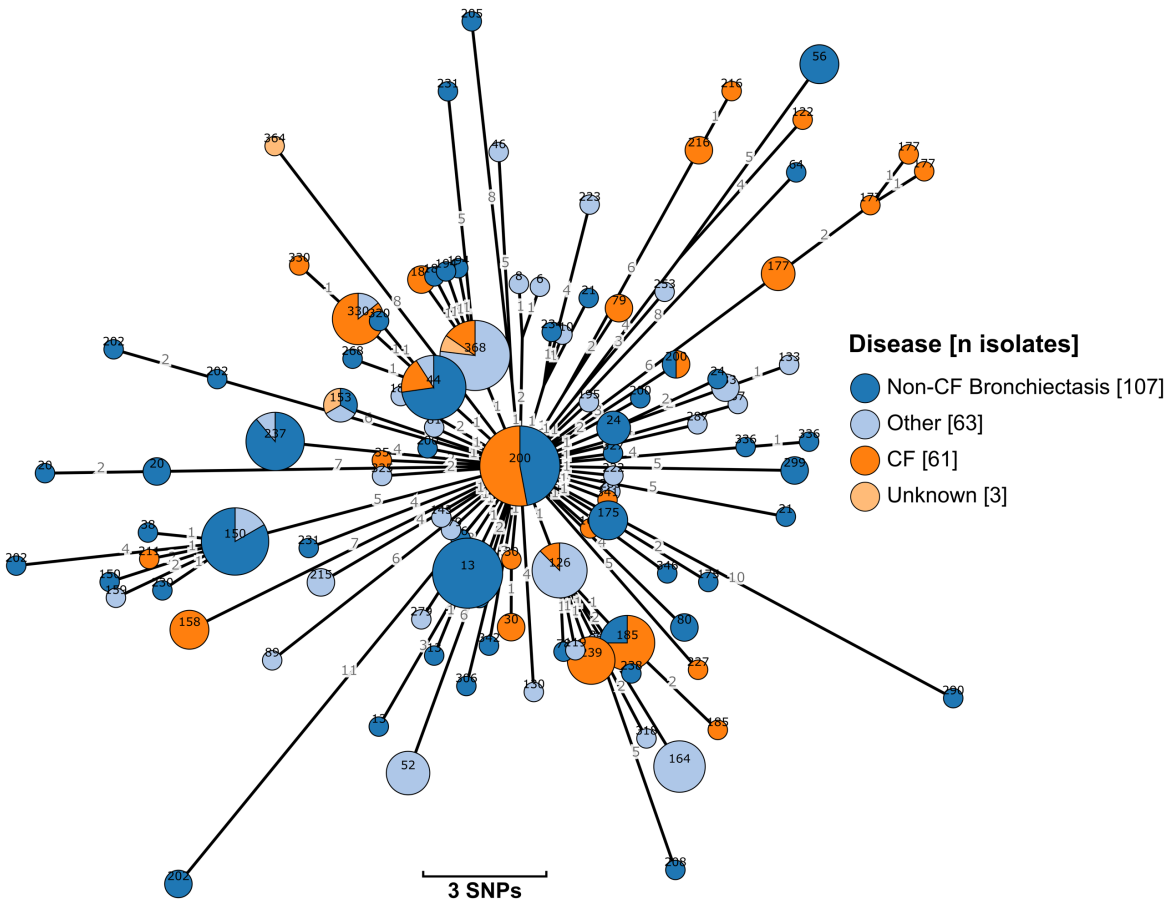
880 *Mycobacterium avium* isolates rooted with a *M. avium paratuberculosis* isolate

881 (DRR263663). The subspecies of each isolate is shown in the datastrip to the right

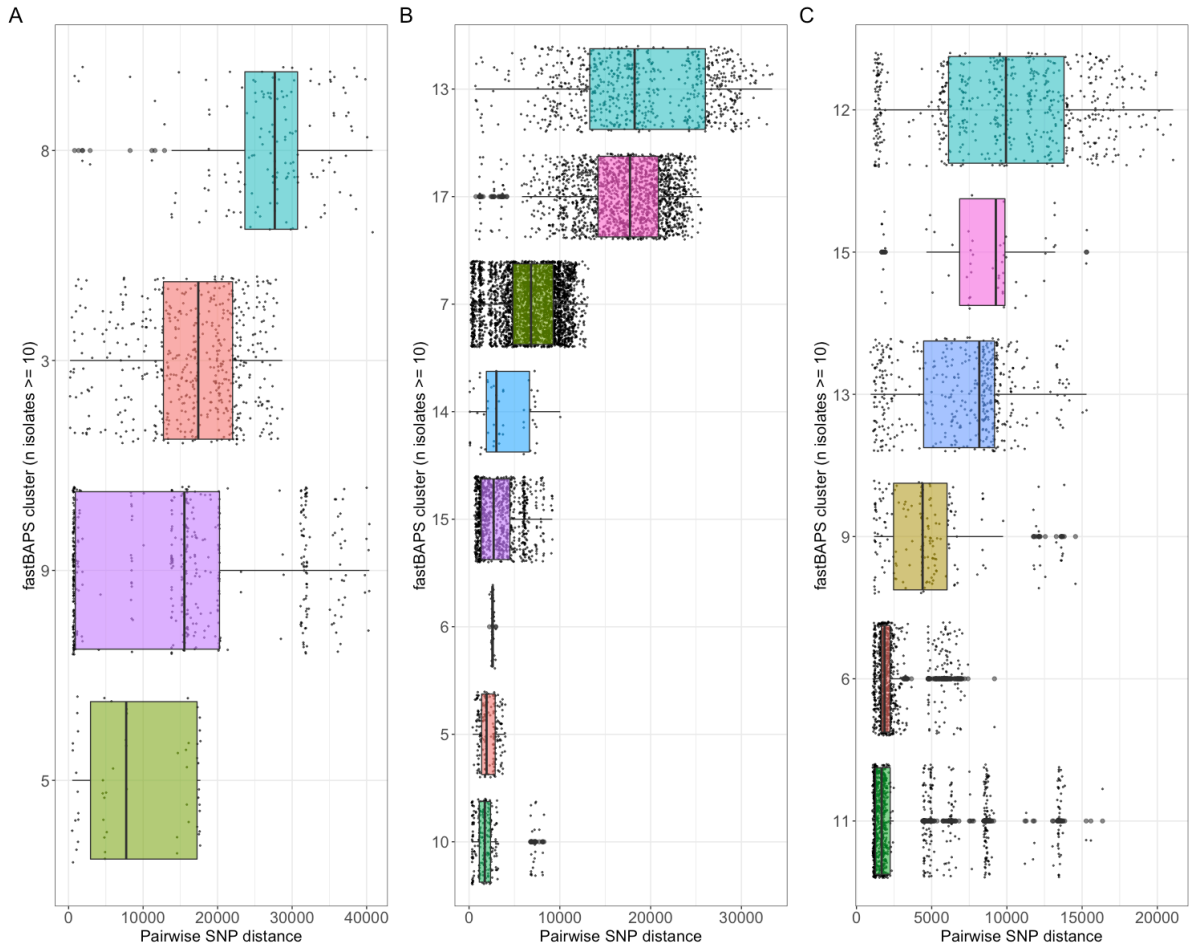
882 of the phylogeny. The scale bar is shown in SNPs per site.

883

884



Supplementary Figure 4. Largest transmission cluster of *Mycobacterium chimaera* isolates collected at the Royal Brompton Hospital. Each node represents an isolate or isolates identical (0 SNPs) to each other and the size of the node is proportional to the number of identical isolates. Nodes are coloured according to patient disease status and labelled by patient number. The edges represent the pairwise SNP distance between the isolate(s).



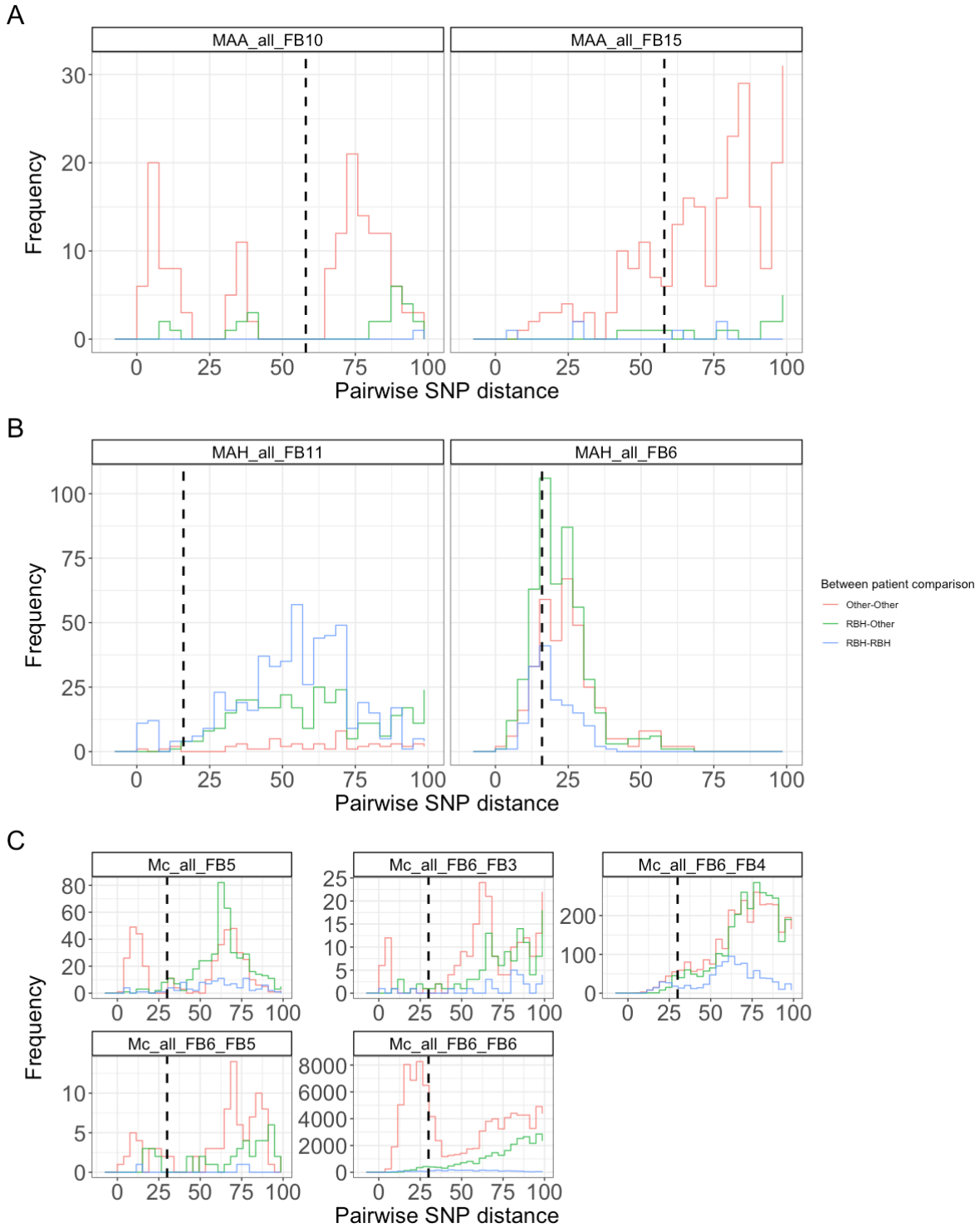
892

893 **Supplementary Figure 5. Global fastBAPS cluster pairwise SNP distance**

894 **distributions.** Boxplots showing isolate pairwise SNP distances for A)

895 *Mycobacterium intracellulare*; B) *Mycobacterium avium avium* and C) *Mycobacterium*

896 *avium hominosuis*



Supplementary Figure 6: Distribution of pairwise SNP distances between isolates from the Royal Brompton Hospital and other studies. Histograms showing distribution of pairwise SNP distances for fastBAPS clusters for A)

902 *Mycobacterium avium avium*; B) *Mycobacterium avium hominosuis* and C)
903 *Mycobacterium chimaera*.
904
905
906