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

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. We sequenced 996 Mycobacterium avium complex (MAC) isolates from CF and non-

CF patients at the Royal Brompton Hospital (RBH), London. Genomic analysis was used to analyse local transmission. Epidemiological links were identified from patient records. These and previously published genomes were used to characterise global population structures. Analysis of the three 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 immediate ancestor of the lineage causing the major HCU-associated outbreak was a lineage already circulating in patients with pre-existing lung conditions. CF and non-CF patients shared transmission chains even in the presence of CF patient-focussed hospital control measures, although the lack of epidemiological links suggested that most transmission is indirect and may involve environmental intermediates 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, while being responsible for a major global transmission event, are not the sole vector nor the ultimate source of this wider patient-infecting lineage. Future studies should include sampling of environmental reservoirs and potential asymptomatic carriers.

Author summary

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 believe this is the first study to use a well-sampled longitudinal isolate dataset, that includes both CF and non-CF patients from a single hospital setting, to investigate transmission of MAC species. We identified transmission clusters in the three predominant MAC species circulating in the hospital and showed that these included both CF and non-CF patients. We then incorporated isolates from previous studies to examine the global population structure of MAC species and showed that for *M. avium* there were UK-specific lineages circulating amongst patients, whilst for *M. chimaera* we could identify global lineages associated with HCUs. For the first time, we also show that the predominant HCU-associate lineage is likely derived from already circulating lineages associated with patients with respiratory diseases. Our study shows the value of integrating whole genome sequencing with epidemiological data to perform high-resolution molecular analyses to characterise MAC populations and identify transmission clusters. Knowledge of putative transmission networks can improve responses to outbreaks and inform targeted infection control and clinical practice.

Introduction

Non-tuberculous mycobacteria (NTM) are ubiquitous environmental microorganisms found in soil and water and are considered opportunistic pathogens in humans. Individuals with pre-existing genetic or acquired lung diseases such as cystic fibrosis (CF), non-CF bronchiectasis (BX) and chronic obstructive pulmonary disease (COPD) are more prone to NTM disease although individuals with no known immune dysfunction can also present with NTM infections (1–3). Symptoms of NTM pulmonary disease are variable but most patients will develop a chronic cough and

other symptoms may include fatigue, sputum production, chest pain, breathlessness, fever and weight loss (1). Globally, disease due to NTM infections is increasing in prevalence. For example, the estimated prevalence of NTM disease in the United States of America (USA) rose from 2.4 cases/100,000 in the early 1980s to 15.2 cases/100,000 in 2013 (4), whilst in the United Kingdom (UK) the prevalence rose from 0.9 cases/100,000 to 2.9 cases/100,000 between 1995 and 2006 (5). NTM infections may be progressive and treatment requires prolonged multi-drug therapy (6) and is often unsuccessful due to an absence of antimicrobial agents with low toxicity and effective *in vivo* activity against NTM species (1).

A number of NTM species including *Mycobacterium abscessus* and members of the *Mycobacterium avium* complex (MAC), notably *M. avium* and *M. intracellulare*, have emerged as major respiratory pathogens in the past three decades (7–9). Another member of the MAC, *M. chimaera*, has also been implicated in numerous global infections associated with cardiothoracic surgery with the source of infections linked to heater-cooler units (HCUs) contaminated during their manufacture (10–12).

Until recently the prevailing hypothesis was that infections caused by NTM were due to independent acquisitions from environmental sources such as soil, contaminated drinking water distribution systems and household plumbing. Recent studies of *M. abscessus* in CF patients have however identified indirect patient-patient transmission within a single CF centre as well as the presence of globally circulating clones of *M. abscessus* amongst CF patients worldwide (13–17). An observational study of *M. abscessus,* across England revealed that these dominant clones are also

found in patients with other chronic respiratory diseases. The study however was unable to identify epidemiological links for most closely-related isolates, suggesting environmental acquisition may be involved (18). Another recent study of *M. abscessus* has demonstrated that transmission networks may involve both people with CF and those without, and that these transmission networks are global. It is therefore likely that transmission is complex, involving multiple patient cadres as well as environmental intermediates (19). In the special case of *M. chimaera*, the high level of genetic similarity between sequenced *M. chimaera* isolates collected from patients, HCUs and the factory of origin suggested a point source contamination during manufacture causing global distribution followed by localised transmission (12).

To date, little work has been done to examine whether similar patterns of transmission in the MAC are occurring between patients with CF, BX or other chronic respiratory diseases. Using a large collection of longitudinal isolates collected from patients attending the Royal Brompton Hospital (RBH) in London, the aims of this study were to characterise the population structure of MAC; to identify potential transmission chains involving patients with CF and other non-CF lung conditions; and to place the RBH isolates in a global context using previously published genomes.

Methods

Consent/ethics

The study (access to patients' clinical data) was approved by the NHS Health Research Authority (HRA) and Health and Care Research Wales (HCRW) (REC reference 21/HRA/2554).

121

122 *Data collection*

123 Clinical data pertaining to patients from whom NTM cultures were isolated were
124 collected from electronic health records at the RBH. Data included patients' sex, age
125 at the time of first positive NTM culture, height, weight, lung function test results,
126 comorbidities, medication history and date of death (where applicable).
127 Anonymization was undertaken by removing personal data, including patients'
128 hospital numbers, prior to analysis.

129

130 *Sample collection*

131 Isolates were collected from patients attending the respiratory inpatient and
132 outpatient clinics of the RBH between January 2013 and April 2016. The RBH
133 routinely archive all mycobacterial isolates cultured from their patients and this
134 archive was used without selection as the basis for the study.

135

136 *Culturing, DNA extraction and sequencing*

137 Sputum samples were grown in BBL MGIT media (BD) in a Bactec MGIT 960 (BD)
138 until the system indicated Mycobacterial growth. Species confirmation was
139 performed using the HAIN GenoType System. Confirmed MAC cultures were
140 regrown from bead stock cultures in BBL MGIT media (BD) in a Bactec MGIT 960
141 (BD) until the system indicated growth. In the absence of growth, DNA was extracted
142 from the bead stock. DNA extractions were performed as previously described
143 (https://dx.doi.org/10.17504/protocols.io.bf28jqhw). A total of 1189 DNA extracts
144 were sequenced by the core pipeline teams at the Wellcome Sanger Institute. The

145 Illumina Hiseq X10 platform was used to generate 2 x 150 bp paired-end reads. Raw
146 sequencing reads were deposited at the European Nucleotide Archive under project
147 PRJEB21813. All accessions used in this project are listed in S1 Table.
148
149 *Sequence QC, mapping and phylogenetics*
150 Basic quality control metrics for the raw sequence data were generated using FastQC
151 v0.11.9 (20). Sequence reads with similarity to *Mycobacterium* species were
152 identified using Kraken v0.10.6 (21) and Bracken v1.0 (22). Samples with < 70%
153 reads mapping to a *Mycobacterium* species were excluded from further analyses (n
154 = 116). Seven isolates not belonging to the MAC (*M. abscessus, M. chelonae, M.*
155 *simiae)* were removed from the dataset. Sequence reads for each species were
156 trimmed using Trimmomatic v0.33 (23) and mapped to appropriate references (Table
157 A in S1 Data) using BWA mem v0.7.17 (minimum and maximum insert sizes of 50 bp
158 and 1000 bp respectively) (24). Single nucleotide polymorphisms (SNPs) were called
159 using SAMtools v1.2 mpileup and BCFtools v1.2 (minimum base call quality of 50 and
160 minimum root squared mapping quality of 30) as previously described (25). Samples
161 with reads that mapped to < 80% of the reference were excluded (n = 70). Variant
162 sites were extracted from the resulting alignments using snp-sites v2.5.1 (26). Whole
163 species maximum likelihood phylogenetic trees were built using IQ-tree v1.6.5
164 accounting for constant sites (-fconst; determined using snp-sites -C) with the built-
165 in model testing (-m MFP) to determine the best phylogenetic model and 1000
166 ultrafast bootstraps (-bb 1000) (27).
167

168 For higher-resolution phylogenies within fastBAPS lineages, recombinant regions
169 were identified and removed from alignments using GUBBINS (28) and new
170 phylogenetic trees were constructed as described above. Pairwise SNP distances
171 were calculated for all pairs of isolates using pairsnp (29).
172
173 *Global collections*
174 To provide context for each the isolates sequenced for each species in this study,
175 datasets consisting of published sequenced isolates were assembled (S2 Table)
176 (12,30–49). Sequence data were downloaded from the European Nucleotide Archive
177 (ENA) and trimmed; Sample QC, mapping and phylogenetic tree construction were
178 performed as detailed above. Only the first isolate from each patient was included
179 from the RBH isolates for each species/subspecies.
180
181 *Genome assemblies*
182 A previously published pipeline was used to produce annotated assemblies (50).
183 Briefly, sequence reads were assembled with spades v 3.10.10 (51) and assemblies
184 were improved by first scaffolding the assembled contigs using SSPACE v2.0 (52)
185 and filling the sequence gaps with GapFiller v1.11 (53).
186
187 *Transmission and epidemiological linkage*
188 Genomic lineages were identified using fastBAPS (54) and new alignments were
189 created for lineages ≥ ten isolates by aligning sequence reads for included isolates
190 against the assembly that had the smallest number of contigs (using the method
191 described above). In order to calculate a pairwise SNP threshold to determine

192 putative transmission clusters within each genomic lineage, pairwise SNP distances
193 for all isolates for each species in the RBH datasets were calculated. Using a
194 previously described method (55), the transmission threshold for each species,
195 regardless of lineage, was calculated by taking the 95th percentile of the maximum
196 within-patient isolate pairwise SNP distances for all patients and adding twice the
197 number of mutations expected to occur in a six month period. To account for excess
198 within-patient diversity observed in the *M. chimaera* FB1 and *M. avium* subsp. *avium*
199 FB14 lineages (Fig A in S1 Data), pairwise SNP distances greater than 25 and 50
200 (assumed to result from infection with multiple lineages) were removed respectively
201 before the above calculations were performed. Based on these results, the R library
202 iGRAPH (56,57) and pairwise SNP thresholds of 16 (*M. intracellulare* and *M. avium*
203 subsp. *hominissuis*), 30 (*M. chimaera*) and 58 SNPs (*M. avium* subsp. *avium*) were
204 used to calculate putative transmission clusters in each genomic lineage. Finally, in
205 order to identify possible epidemiological links between patients infected within the
206 same transmission clusters, hospital stay records were examined for epidemiological
207 contacts. The latter were defined as patients attending the same ward on the same
208 day up to one year prior to the collection of the first sequenced isolate.
209

210 **Results**

211 *Patient demographics*

212 The median age and BMI of the 354 patients included in the study was 56 years
213 (range 5 - 93) and 22.5 (range 13.4 – 43.4) respectively. One hundred and seventy-
214 four patients (49.2%) were male and 38/354 (10.7%) were smokers. There were
215 147/354 (41.5%) patients with BX, 87/354 (24.6%) with CF, 53/354 (15.0%) with

216 COPD, 32/354 (9.0%) with asthma, 19/354 (5.4%) with allergic bronchopulmonary
217 aspergillosis (ABPA), 17/354 (4.8%) with interstitial lung disease (ILD) and 7/354
218 (2.0%) with other underlying respiratory conditions such as pleural thickening or
219 sarcoidosis. Patients with no other respiratory disease or predisposition to NTM
220 infections accounted for 3.4% (12/354) of the cohort. During the study period, 55/354
221 (15.5%) patients were on antibiotic treatment regimes. For 20 patients (5.6%) clinical
222 data were unavailable.
223
224 *Species distribution*
225 A total of 996 isolates from 354 patients were successfully sequenced and the nine
226 MAC species identified are detailed in Table 1. The three predominant species
227 amongst the sequenced isolates were *M. avium (M. avium* subsp. *avium* and *M. avium*
228 subsp. *hominissuis)*, *M. chimaera* and *M. intracellulare.* Together these accounted
229 for 926/996 (93.0%) of the MAC isolates sequenced. Most patients were infected
230 with only a single species during the collection period. However, 45 of the 354
231 patients (12.7%) were infected with two or more species (Fig 1A). In this group of
232 patients, most of the isolates collected were typically from a single species, with other
233 species observed more infrequently (Fig 1B). Subsequent analyses in this study will
234 focus on the three predominant species in the dataset: *M. intracellulare, M. avium (M.*
235 *avium* subsp. *avium* and *M. avium* subsp. *hominissuis)*, and *M. chimaera*.
236
237 *M. intracellulare*
238 A total of 162 genomes from 37 patients were identified as *M. intracellulare* (Fig 2A).
239 Eleven of the patients had CF, seventeen had BX, with the remaining seven having

240 other lung conditions (COPD n = 3; ILD n = 3; asthma n = 1; congenital pulmonary
241 airway malformation [CPAM] n = 1) and disease metadata were missing for two
242 patients. Genomic clustering with fastBAPs identified nine lineages with three of
243 these having more than ten genomes (Fig 2A). Following remapping to local
244 references for the three largest fastBAPS lineages, three putative transmission
245 clusters were identified with the largest, Mi_FB3_1, composed of 16 patients (Fig 2B;
246 Table B in S1 Data). Of these 16 patients, eight had BX, seven CF and one ILD. The
247 range of the number of isolates collected per patient was between one and 32. Four
248 of the sixteen patients were also infected with other species or lineages during the
249 sampling period with Mi_FB3_1 only being detected in 1/22 isolates collected for
250 patient 218 (Fig 2B). During the time period that the sequenced isolates were
251 collected, four of the patients (85, 95, 97 and 99), were being treated with antibiotics
252 and three of the four patients were treated successfully. No epidemiological links
253 were identified between patients in the year prior to isolate collection. Comparison
254 with 77 previously published *M. intracellulare* isolates showed that the most closely
255 related contextual isolate was collected in the UK in 2015/2016 [32] and formed part
256 of a clade containing isolates from the Mi_FB3_1 cluster (Fig 2C). The remainder of
257 the RBH isolates were evenly distributed through the global phylogenetic tree.
258
259 *M. avium*
260 A total of 405 sequenced isolates collected from 176 patients were identified as *M.*
261 *avium*. Of the patients infected with *M. avium*, 44 had CF and 81 BX. Sequence data
262 for all *M. avium* isolates was mapped to a single reference (*M. avium* subsp. *avium*
263 104; NC008595.1) and a phylogenetic tree constructed with a previously published

264 *M. avium* subsp. *paratuberculosis* isolate (DRR263663) chosen as an outgroup (Fig B
265 in S1 Data). The structure of this phylogeny showed that there were two major clades
266 which corresponded to the two subspecies *M. avium* subsp. *avium* (n = 207; MAA)
267 and *M. avium* subsp. *hominissuis* (n = 198; MAH). Isolates for each of the two
268 subspecies were next analysed separately. A small number of patients were infected
269 with both subspecies (n = 6). No examples of the other two *M. avium* subspecies,
270 *paratuberculosis* and *silvaticum*, were identified in this study.
271
272 *M. avium avium (MAA)*
273 Of the 76 patients infected with MAA, 15 had CF, 39 had BX and 16 had other lung
274 conditions (COPD n = 10; ILD n = 3; ABPA n = 3; asthma n = 6) whilst three patients,
275 including one smoker, had no pre-existing respiratory disease (metadata were
276 unavailable for three patients; Fig 3A). Clustering of the 207 MAA genomes identified
277 17 fastBAPS lineages (Fig 3A). Remapping of the four largest fastBAPS lineages
278 allowed the identification of seven putative transmission clusters comprising between
279 two and ten patients (Table C in S1 Data) with the largest cluster of ten isolates
280 comprised of patients with CF (n = 2), BX (n = 6) and asthma or COPD (n =2; Fig 3B).
281
282 A total of 21 fastBAPS lineages were defined in the global phylogeny (Fig 3C).
283 Examination of the distribution of pairwise SNP distances for each fastBAPS lineage
284 containing at least ten genomes revealed that there were two lineages containing
285 RBH isolates with lower median pairwise SNP distances (FB10 and FB15; Fig 3C and
286 Fig CB in S1 Data). The 23 genomes, including two RBH genomes, in FB10 were
287 collected in the UK between 2013 and 2016, with most coming from the National

288 Mycobacterial Reference Service. The latter characterises mycobacterial cultures
289 from across the Midlands and North of England (Fig 5A) (37). If the SNP threshold
290 used to calculate putative transmission clusters circulating in the RBH was applied,
291 then one isolate from the RBH (40-1) formed part of a cluster comprised of isolates
292 from Quan *et al.* (37). The majority (35/46) of the isolates forming FB15 were collected
293 in the USA with most from a study investigating *M. avium* in the community and
294 household water in Philadelphia (Fig 5B) (33). The pairwise distances between two
295 of the RBH isolates and other UK isolates and isolates from the USA in FB15 were
296 within the threshold used to define transmission, suggesting potential cross-Atlantic
297 transmission (Fig DA in S1 Data).
298
299 *M. avium hominissuis (MAH)*
300 A total of 198 genomes from 106 patients were characterised as MAH (Fig 4A). Thirty-
301 three patients had CF, 43 BX and 21 had other lung conditions (COPD n = 14; ILD n
302 = 1; ABPA n = 5; sarcoidosis n = 1; asthma n = 11). Four patients had no pre-existing
303 respiratory disease. Disease metadata were unavailable for the remaining five
304 patients. Genomic clustering identified 17 lineages and seven putative transmission
305 clusters containing between two and 16 patients (Table D in S1 Data). The largest
306 cluster of 16 isolates is shown in Figure 4B and comprised patients with CF (n = 4),
307 BX (n = 7), COPD (n = 3) and a single patient with no pre-existing lung condition
308 (metadata were unavailable for one patient). Two pairs of patients in each of the
309 clusters, MAH_FB8_1 and MAH_FB14_9, were found to have epidemiological links.
310

311 Of the 20 defined fastBAPS lineages in the global collection, FB6 and FB11 had lower
312 median pairwise SNP distances and contained RBH isolates (Fig 4C and Fig CC in
313 S1 Data). Both lineages were composed completely or near-completely of isolates
314 collected in the UK (Fig 5C and Fig 5D). The star-like structure of the FB6 phylogeny
315 is suggestive of a point source outbreak and applying a transmission SNP threshold
316 of 16 SNPs showed that 41/45 of the isolates in FB6 would have formed a
317 transmission cluster that included both isolates from this study and isolates from
318 other UK studies (Fig DB in S1 Data) (35,37). Two small transmission clusters,
319 containing RBH and other UK isolates, were also identified in FB11.
320
321 *M. chimaera*
322 There were 359 sequenced isolates collected from 155 patients identified as *M.*
323 *chimaera*; 37 of the patients had CF, 60 had BX whilst 50 patients had other lung
324 conditions (COPD n = 24; ILD n =10; ABPA n = 11; pleural thickening with enfolded
325 lung n = 1; lung non-mucinous adenocarcinoma n = 1; Wegener's granulomatosis n
326 = 1; *M. tuberculosis* in lymph nodes during same endoscopy procedure n = 1; primary
327 ciliary dyskinesia n = 1; asthma n = 17). There were five patients with no pre-existing
328 respiratory disease and metadata were unavailable for a further nine patients. A single
329 predominant lineage, Mc_FB1, corresponding to the previously characterised Group
330 1 accounted for 332/359 (92.5%) of the sequenced isolates with the majority of the
331 remaining isolates (25/359; 7.0%) forming a second main lineage Mc_FB2
332 corresponding to Group 2 (Fig 6A). Thirteen putative transmission clusters containing
333 two to 106 patients were identified (Table E in S1 Data). Twelve potential
334 epidemiological links between 15 different patients were found in transmission cluster

335 Mc_FB3_1, the largest transmission cluster of 106 patients (Fig 6B; Table F in S1
336 Data). Of the patients in this cluster 43 had BX, 24 CF, 13 COPD, 14 asthma, seven
337 ILD, two ABPA, one pleural thickening with enfolded lung, one lung non-mucinous
338 adenocarcinoma and one primary ciliary dyskinesia. Three patients had no pre-
339 existing lung condition and disease status was missing for five patients. Fifteen
340 patients underwent antibiotic treatment during the study period.
341
342 The global phylogenetic tree containing 155 isolates from RBH and 671 previously
343 published isolates was topologically similar to the tree built using only the isolates
344 from RBH with a single major clade with low genetic diversity present (FB6; Fig 6C).
345 Isolates from the two clusters with more than ten representatives (FB5 and FB6) were
346 remapped to local references. Recombination was removed and new phylogenetic
347 trees constructed (Fig 7A [FB5] and Fig 7B [FB6]). Most isolates in FB5 were collected
348 from patients with respiratory conditions or else patients with infections following
349 cardiac surgeries. There was, however, a distinct sub-clade rooting within this
350 diversity containing only isolates from HCUs or hospital water supplies (Fig 7B).
351 Within FB6 a more deeply rooting diverse clade contained mainly human isolates and
352 the larger shallower clade rooting within this contained most of the HCU isolates
353 interspersed with human isolates (Fig 7B and Fig 7C). Due to the large number of
354 isolates assigned to FB6 (n = 765), further genomic lineage assignment was
355 performed using fastBAPS to identify five sub-lineages within FB6 (Fig 7B and Fig
356 7C). New mapping was performed for the three lineages with more than ten isolates
357 (FB6_FB1, FB6_FB2 and FB6_FB4; Fig 8A-8C;). Two of these lineages, FB6_FB2
358 and FB6_FB4, were completely or nearly completely composed of patient isolates

359 (Fig 8B and Fig 8C). Most of the isolates collected from HCUs or water supplies
360 made up the bulk of FB6_FB1, though human isolates were distributed throughout
361 the tree (Fig 8A). Putative transmission clusters were calculated for all five fastBAPS
362 lineages described above used a pairwise SNP threshold of 30 SNPs (Table G in S1
363 Data). Transmission clusters containing RBH isolates were found in each of the
364 lineages with the largest, found in FB6_FB1, comprising 258 HCU isolates, 230
365 patient isolates and a single isolate from a water supply (Table G in S1 Data).
366

367 **Discussion**

368 The aims of this study were to characterise the population structure of MAC and
369 identify potential transmission between CF and non-CF patients. We used a dataset
370 of 996 sequenced MAC isolates collected in a London hospital. This is, to date, the
371 largest sequenced MAC isolate collection analysed from a single location. We
372 showed that polymicrobial infection with MAC is not uncommon. The large,
373 longitudinal dataset enabled us to identify putative transmission clusters for all four
374 species/sub-species we considered. Most of these transmission clusters contained
375 both patients with CF and non-CF patients, showing that transmission of MAC is
376 occurring despite the presence of hospital-based control measures focussing on
377 patients with CF. The lack of clear epidemiological links suggests that the majority
378 of transmission is not due to direct patient-to-patient transmission but rather through
379 intermediates such as a local environmental vector or asymptomatic carriage in the
380 wider population. Comparison of the isolates collected as part of this study with
381 published isolates identified a small number of lineages likely circulating in the UK.
382 With the exception of the major HCU-associated lineage in *M. chimaera*, the

383 comparison did not identify globally circulating lineages like those observed in *M.*
384 *abscessus*. Surprisingly, given the previous description of the *M. chimaera* HCU
385 clade as representing an environmental point source outbreak (12), we provide
386 evidence that the lineage which was responsible for the HCU-vectored *M. chimaera*
387 outbreak was derived from an existing lineage which was already circulating amongst
388 patients with pre-existing lung conditions, and that this introduction into HCUs
389 happened on at least two separate occasions (only one of which led to the global
390 outbreak). This indicates that, while infection and distribution of HCUs was clearly
391 responsible for the global transmission of one sub-lineage, they were not the primary
392 source of this lineage, which was derived from a lineage that was already circulating
393 in the human population prior to the HCU outbreak. Further, there is evidence of
394 onward transmission into non-cardiac patients of the HCU-sub-lineage, indicating
395 that local establishment and transmission can still occur independent of direct
396 infection by HCUs.
397
398 During the isolate sampling period we identified 45 patients infected with more than
399 one species or lineage. This suggests that many of the patients had polymicrobial
400 infections and that certain lineages were preferentially sampled and cultured at
401 different times. This has been observed previously in other longitudinal samples of
402 patients with MAC lung disease (38). These observations suggest that single colony
403 picks during culturing potentially may be underrepresenting the underlying diversity
404 of MAC species in patients with MAC lung disease. Therefore, to better understand
405 the ecology and transmission of MAC species in the lung, a deep sequencing

406 approach in which colony sweeps rather than colony picks are utilised should be
407 considered for future studies.
408
409 Previous studies have used SNP thresholds of 15-25 SNPs to identify putative
410 transmission clusters in NTM species (13,18,33). Rather than using a previously
411 applied or arbitrary threshold, we used a method that had previously been used to
412 define SNP thresholds in *Staphylococcus aureus* (55). This allowed us to calculate a
413 SNP threshold for each species using the longitudinal isolates collected by the RBH.
414 The thresholds calculated for *M. intracellulare* and MAH of 16 SNPs were similar to
415 those previously applied. Whilst we obtained higher thresholds of 30 SNPs and 58
416 SNPs for *M. chimaera* and MAA respectively, this did not result in an inflated number
417 of transmission clusters, as applying a threshold of 16 SNPs to the MAA dataset
418 would have resulted in only two fewer clusters. These larger values were
419 predominantly due to high levels of within-host diversity observed amongst patient
420 isolates in single clusters in both MAA and *M. chimaera*. For instance, in MAA_FB14
421 isolates from a single patient collected a day apart were as much as 101 SNPs
422 different. This highlighted the likelihood of polymicrobial infections being common
423 amongst the patients we included in the study.
424
425 Using the SNP thresholds calculated for each species, we were able to identify
426 putative transmission clusters amongst all four of the MAC species/subspecies we
427 considered. Broadly, two different patterns of infection were observed: amongst *M.*
428 *intracellulare* and the two *M. avium* subspecies, our results suggested that there have
429 been multiple independent introductions into the human population followed by

430 limited transmission. Our analysis of *M. chimaera*, however, identified a single large
431 transmission cluster was responsible for infecting 106/155 (68.4%) of our patients
432 infected with this species. Most of these transmission clusters contained both CF
433 and non-CF patients and a small number also contained patients with no pre-existing
434 lung conditions.
435
436 Using our criteria, we were unable to identify potential epidemiological links for most
437 patients included in the transmission clusters. The exception to this was the largest
438 *M. chimaera* transmission cluster where we found 12 potential epidemiological links
439 up to a year before sampling began. Interestingly, four patients were part of more
440 than one link, with patient 218 overlapping with other patients in this transmission
441 cluster on the same wards six times over a nine-month period. The tight infection
442 control surrounding patients with CF, which prioritises preventing cross-infection
443 through hygiene and segregation, means that they are unlikely to have interacted with
444 patients with other lung conditions whilst in hospital. Outpatient clinics are organised
445 so that patients are seen individually and, during inpatient care, CF patients are given
446 individual rooms with en-suite facilities. This implies that transmission is likely
447 occurring *via* other pathways, such as through environmental intermediates or
448 through a reservoir of healthy, asymptomatic carriers in the wider population. This
449 observation is not unique and for some time, there has been considerable debate
450 whether NTM transmission between patients is occurring, even when transmission is
451 strongly supported by genome-based analysis (13,18,58). The near complete
452 absence of epidemiological links between patients in transmission clusters as well as
453 the ubiquitous presence of NTM species in the environment, especially water

454 supplies, has led some researchers to suggest that the dissemination of NTM
455 lineages at a national level is associated with exposure to contaminated water
456 supplies (18). Whilst there is certainly strong evidence for the same lineages being
457 isolated from water supplies and patients in the same location [28], the absence of a
458 single national water supply even in a comparatively small country like the UK would
459 suggest that closely related isolates collected from different geographical locations
460 are unlikely to be due to a single contaminated water supply. It is however possible
461 that transmission networks may include local water supplies as intermediates. The
462 alternative hypothesis that NTM populations are being maintained in apparently
463 healthy individuals with no symptoms of NTM lung disease is equally worthy of
464 consideration. This would explain the presence of local transmission clusters in the
465 absence of direct epidemiological links as well as the presence of phylogeographical
466 structure in NTM trees. To better address how NTM species are being transmitted,
467 future studies should focus on attempting to collect isolates from the general
468 population, perhaps focussing on smokers where there is some evidence that they
469 may provide a better reservoir for NTM species and other opportunistic pathogens
470 (19).
471
472 The inclusion of published genomes alongside the RBH isolates allowed us to use
473 our defined pairwise SNP thresholds to identify putative MAC lineages circulating in
474 the UK and globally. Apart from the *M. chimaera* lineages that were known to be
475 associated with HCUs, we were unable to identify large globally circulating lineages
476 like those observed in *M. abscessus*. This potentially suggests different evolutionary
477 dynamics amongst the MAC species or, in the case of *M. intracellulare*, may simply

478 reflect a lack of comprehensive sampling. We were only able to identify a single
479 lineage containing RBH isolates amongst the two *M. avium* subspecies that included
480 isolates from outside the UK. This cluster comprised the majority of isolates in FB15
481 (Figure 5B). Given that most available *M. avium* genomes have been collected in the
482 UK or USA, a more comprehensive global sampling strategy could potentially reveal
483 the existence of additional globally circulating clones.
484
485 A different picture emerged when we calculated global transmission clusters for *M.*
486 *chimaera* using a pairwise SNP threshold of 30 SNPs (Supplementary Table 7). Here
487 we found that most clusters either consisted of isolates from across Europe or else
488 also included isolates from other parts of the world such as Australia. Notably, we
489 identified a single large transmission cluster containing 489 isolates from 12 countries
490 that included most of the HCU-associated isolates. Previous work investigating the
491 global population structure of *M. chimaera* has suggested that the majority of *M.*
492 *chimaera* transmission globally was associated with point source contamination of
493 HCUs during manufacturing which then led to onward transmission to patients
494 undergoing cardiac surgery (12). In our study we have assembled the largest *M.*
495 *chimaera* dataset analysed to date, including isolates collected from HCUs and
496 cardiac patients as well as from patients with lung conditions such as CF and BX.
497 This has enabled us to investigate the global population structure of *M. chimaera* in
498 more detail than has previously been possible (12,59). Strikingly, we discovered clear
499 evidence for distinct lineages within the global phylogeny that were almost exclusively
500 associated with patient isolates. In addition, most of the HCU-associated isolates
501 were clustered together in a single lineage that was derived from the lineages

502 containing only human isolates (Figure 6B). This was also true of a second, smaller,
503 lineage that contained only HCU and environmental isolates. The results suggest that
504 the predominant HCU-associated lineage is descended from a *M. chimaera*
505 population already circulating amongst patients with respiratory diseases such as CF,
506 and that a similar derivation has happened at least once more. Unfortunately, due to
507 the absence of even a moderate temporal signal (data not shown), we were unable
508 to place any estimates of when these lineages first emerged.
509
510 Potential limitations of this study are the lack of environmental samples as well as the
511 distinct bias with respect to country of isolation in the contextual dataset. Given, the
512 lack of direct patient to patient contact and epidemiological links, the inclusion of
513 samples collected from the wards admitting patients and hospital water supplies
514 would potentially have revealed additional vectors of transmission. Assembling
515 useful contextual datasets is reliant on what is available in the public domain and the
516 geographical bias towards the USA and UK in the contextual datasets we used is not
517 confined to this study. For many pathogens, most studies have focussed on resource
518 richer settings. However, given the increasing availability of genome sequencing, it
519 is likely that NTM datasets from other settings will be available to include in future
520 analyses.
521
522 In conclusion, using a large longitudinal sample of MAC isolates collected within a
523 single hospital as well as previously published global contextual isolates, we were
524 able to identify polymicrobial infections by different MAC species and lineages in
525 multiple patients as well as identify transmission clusters representing lineages

526 circulating within the hospital, at a national level and, less commonly, globally. We
527 show that the origin of the major HCU-associated *M. chimaera* lineage is likely a
528 previously circulating lineage associated with patients with respiratory diseases. To
529 better understand the transmission dynamics and to prevent continuing circulation of
530 MAC and other NTM species, future studies should include sampling of
531 environmental reservoirs and potential asymptomatic carriers.
532

533 **Data availability**

534 Raw sequencing reads were deposited at the European Nucleotide Archive under
535 project PRJEB21813; all accessions used in this project are listed in Supplementary
536 File 1.
537

538 **Code availability**

539 The R code used to perform data analyses in this study is available in GitHub
540 (https://github.com/avantonder/MAC).
541

542 **Funding**

543 This project was supported by the Asmarley Trust, the Wellcome Trust and the NIHR
544 Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield
545 NHS Foundation Trust, Imperial College London. Huw Ellis' CRF was funded by a
546 grant from the Welton Foundation.
547

548 **Conflicts of interest**

549 JP is a paid consultant to NextGen Diagnostics LLc.

550

551 **References**

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

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

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

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

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

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

567 7. Adjemian J, Olivier KN, Prevots DR. Epidemiology of Pulmonary Nontuberculous
568 Mycobacterial Sputum Positivity in Patients with Cystic Fibrosis in the United States,
569 2010-2014. Ann Am Thorac Soc. 2018 Jul;15(7):817–26.

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

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

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

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

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

585 13. Bryant JM, Grogono DM, Greaves D, Foweraker J, Roddick I, Inns T, et al. Whole-genome
586 sequencing to identify transmission of Mycobacterium abscessus between patients with
587 cystic fibrosis: a retrospective cohort study. Lancet Lond Engl. 2013 May
588 4;381(9877):1551–60.

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

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

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

598 17. Davidson RM, Hasan NA, Epperson LE, Benoit JB, Kammlade SM, Levin AR, et al.
599 Population Genomics of Mycobacterium abscessus from U.S. Cystic Fibrosis Care
600 Centers. Ann Am Thorac Soc. 2021 Dec;18(12):1960–9.

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

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

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

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

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

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

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

618 25. Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, et al. Evolution of
619 MRSA during hospital transmission and intercontinental spread. Science. 2010 Jan
620 22;327(5964):469–74.

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

624 27. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
625 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol.
626 2015 Jan;32(1):268–74.

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

630 29. Tonkin-Hill G. pairsnp [Internet]. 2021 [cited 2021 Dec 14]. Available from:
631 https://github.com/gtonkinhill/pairsnp

632 30. Caverly LJ, Spilker T, LiPuma JJ. Complete Genome Sequence of Mycobacterium
633 abscessus subsp. bolletii. Genome Announc. 2016 Jun 9;4(3):e00543-16.

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

638 32. Bryant JM, Thibault VC, Smith DGE, McLuckie J, Heron I, Sevilla IA, et al. Phylogenomic
639 exploration of the relationships between strains of Mycobacterium avium subspecies
640 paratuberculosis. BMC Genomics. 2016 Jan 26;17:79.

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

644 34. Yoon J-K, Kim TS, Kim J-I, Yim J-J. Whole genome sequencing of Nontuberculous
645 Mycobacterium (NTM) isolates from sputum specimens of co-habiting patients with
646 NTM pulmonary disease and NTM isolates from their environment. BMC Genomics.
647 2020 Apr 23;21(1):322.

648 35. Votintseva AA, Pankhurst LJ, Anson LW, Morgan MR, Gascoyne-Binzi D, Walker TM, et
649 al. Mycobacterial DNA extraction for whole-genome sequencing from early positive
650 liquid (MGIT) cultures. J Clin Microbiol. 2015 Apr;53(4):1137–43.

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

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

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

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

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

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

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

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

676 44. Ghodousi A, Borroni E, Peracchi M, Palù G, Fallico L, Rassu M, et al. Genomic analysis of
677 cardiac surgery-associated Mycobacterium chimaera infections in Italy. PloS One.
678 2020;15(9):e0239273.

679 45. Mac Aogáin M, Roycroft E, Raftery P, Mok S, Fitzgibbon M, Rogers TR. Draft Genome
680 Sequences of Three Mycobacterium chimaera Respiratory Isolates. Genome Announc.
681 2015 Dec 3;3(6):e01409-15.

682 46. Perkins KM, Lawsin A, Hasan NA, Strong M, Halpin AL, Rodger RR, et al. Notes from the
683 Field: Mycobacterium chimaera Contamination of Heater-Cooler Devices Used in
684 Cardiac Surgery - United States. MMWR Morb Mortal Wkly Rep. 2016 Oct
685 14;65(40):1117–8.

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

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

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

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

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

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

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

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

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

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

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

58. Bronson RA, Gupta C, Manson AL, Nguyen JA, Bahadirli-Talbott A, Parrish NM, et al. Global phylogenomic analyses of Mycobacterium abscessus provide context for non cystic fibrosis infections and the evolution of antibiotic resistance. Nat Commun. 2021 Aug 26;12(1):5145.

59. Schreiber PW, Kohl TA, Kuster SP, Niemann S, Sax H. The global outbreak of Mycobacterium chimaera infections in cardiac surgery-a systematic review of whole-genome sequencing studies and joint analysis. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. 2021 Nov;27(11):1613–20.

Tables and Figures

Table 1. *Mycobacterium avium* complex species identified in the Royal Brompton Hospital Collection.

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. colombiense	1
Mycobacterium sp. QIA-37	1

726

727

728

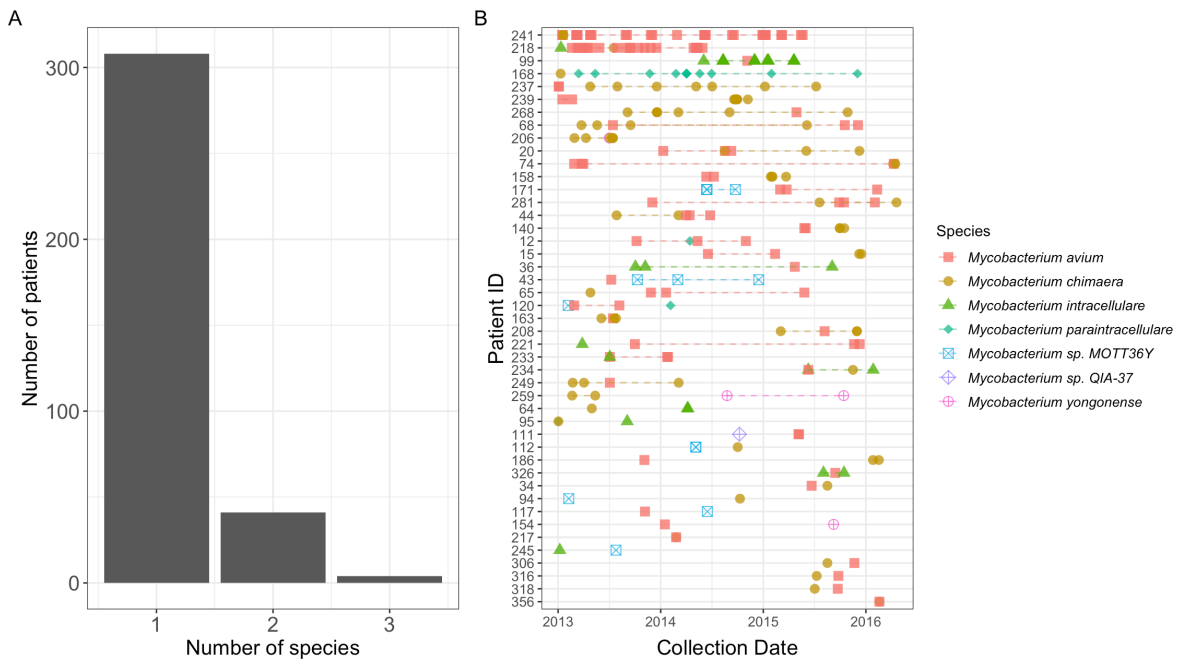


Fig 1. Patients infected by more than one *Mycobacterium avium* Complex (MAC) species. A) Histogram showing the number of *Mycobacterium avium* complex species identified in each patient; B) Timeline of patients infected by more than one *Mycobacterium avium* Complex (MAC) species. Each point represents 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.

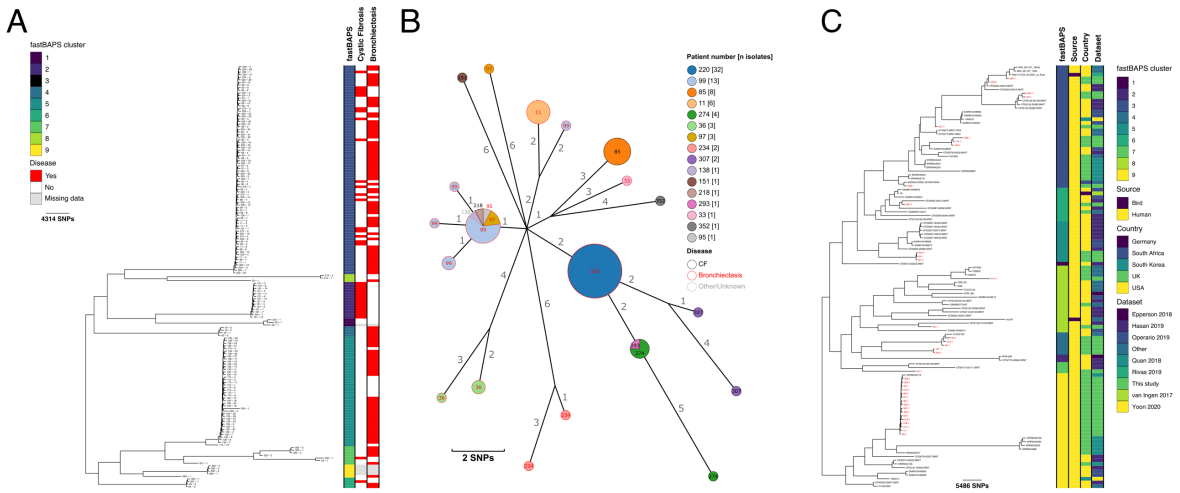


Fig 2: Population structure and transmission of *Mycobacterium intracellulare*. A) Midpoint-rooted maximum likelihood phylogenetic tree for *Mycobacterium intracellulare* isolates collected at the Royal Brompton Hospital (n = 162). The taxa are clustered according to their sequence similarities. The lengths of the branches are scaled in nucleotide substitutions per site. The disease status of the patients (that isolates were collected from) for cystic fibrosis and non-cystic fibrosis bronchiectasis are represented by red bars for Yes and white bars for No with missing patient data shown by gray bars. Scale bars are shown in SNPs per site.; B) *Mycobacterium intracellulare* Mi_FB3_1 transmission cluster. 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 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 = red, grey = other lung condition). The edges represent the pairwise SNP distance between the isolate(s); C) Midpoint-rooted maximum likelihood phylogenetic tree for global *Mycobacterium intracellulare* isolates (n = 114). The taxa are clustered according to their sequence similarities. The lengths of the branches are scaled in nucleotide substitutions per

site. fastBAPS lineage, 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.

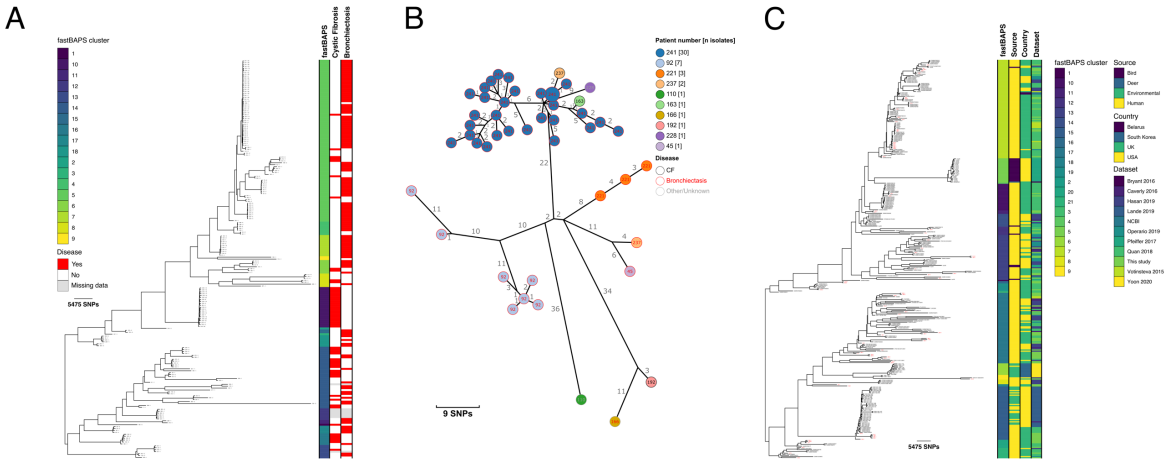
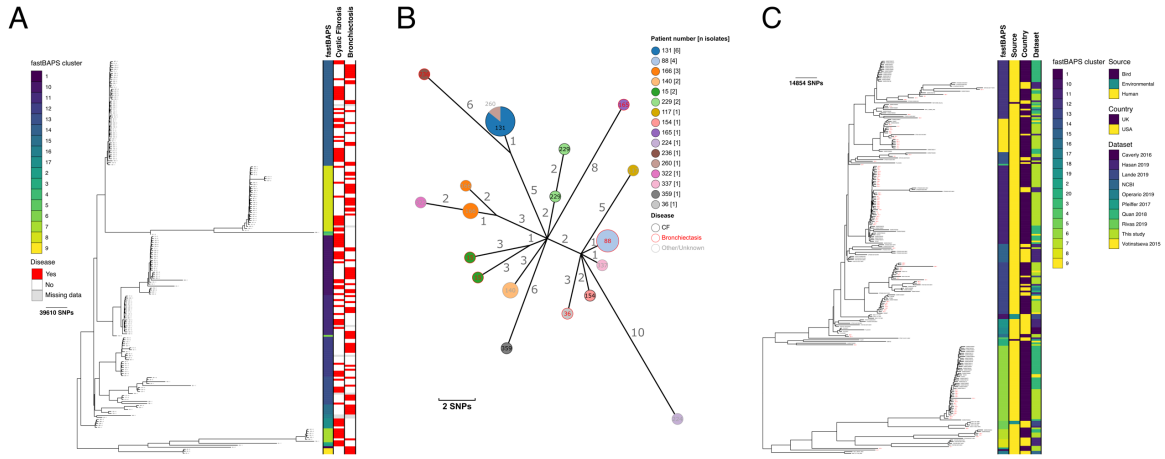


Fig 3. Population structure and transmission of *Mycobacterium avium* subsp. *avium*. A) Midpoint-rooted maximum likelihood phylogenetic tree for *Mycobacterium avium* subsp. *avium* isolates collected at the Royal Brompton Hospital (n = 207). The taxa are clustered according to their sequence similarities. The lengths of the branches are scaled in nucleotide substitutions per site. The disease status of the patients (that isolates were collected from) for cystic fibrosis and non-cystic fibrosis bronchiectasis are represented by red bars for Yes and white bars for No with missing patient data shown by gray bars. Scale bars are shown in SNPs per site; B) *Mycobacterium avium* subsp. *avium* MAA_FB5_1 transmission cluster. 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 and labelled by patient number. The outer ring of each node is coloured according to the

775 disease status of the patient (cystic fibrosis = black, non-cystic fibrosis
776 bronchiectasis = red, grey = other lung condition). The edges represent the pairwise
777 SNP distance between the isolate(s); C) Midpoint-rooted maximum likelihood
778 phylogenetic tree for global *Mycobacterium avium* subsp. *avium* isolates (n = 344).
779 The taxa are clustered according to their sequence similarities. The lengths of the
780 branches are scaled in nucleotide substitutions per site. fastBAPS lineage, source of
781 isolate, country of collection and study are shown as datastrips to the right of the
782 phylogenies. Isolates from the Royal Brompton Hospital are highlighted in red. Scale
783 bars are shown in SNPs per site.

784

785



786

787

788 **Fig 4. Population structure and transmission of *Mycobacterium avium* subsp.**
789 ***hominissuis*.** A) Midpoint-rooted maximum likelihood phylogenetic tree for
790 *Mycobacterium avium* subsp. *Hominissuis* isolates collected at the Royal Brompton
791 Hospital (n = 198). The taxa are clustered according to their sequence similarities.
792 The lengths of the branches are scaled in nucleotide substitutions per site. The

793 disease status of the patients (that isolates were collected from) for cystic fibrosis
794 and non-cystic fibrosis bronchiectasis are represented by red bars for Yes and white
795 bars for No with missing patient data shown by gray bars. Scale bars are shown in
796 SNPs per site; B) *Mycobacterium avium* subsp. *hominissuis* MAH_FB8_1
797 transmission cluster. Each node represents an isolate or isolates identical (0 SNPs)
798 to each other and the size of the node is proportional to the number of identical
799 isolates. Nodes are coloured and labelled by patient number. The outer ring of each
800 node is coloured according to the disease status of the patient (cystic fibrosis = black,
801 non-cystic fibrosis bronchiectasis = red, grey = other lung condition). The edges
802 represent the pairwise SNP distance between the isolate(s); C) Midpoint-rooted
803 maximum likelihood phylogenetic tree for global *Mycobacterium avium* subsp.
804 *hominissuis* isolates (n = 236). The taxa are clustered according to their sequence
805 similarities. The lengths of the branches are scaled in nucleotide substitutions per
806 site. fastBAPS lineage, source of isolate, country of collection and study are shown
807 as datastrips to the right of the phylogenies. Isolates from the Royal Brompton
808 Hospital are highlighted in red. Scale bars are shown in SNPs per site.
809

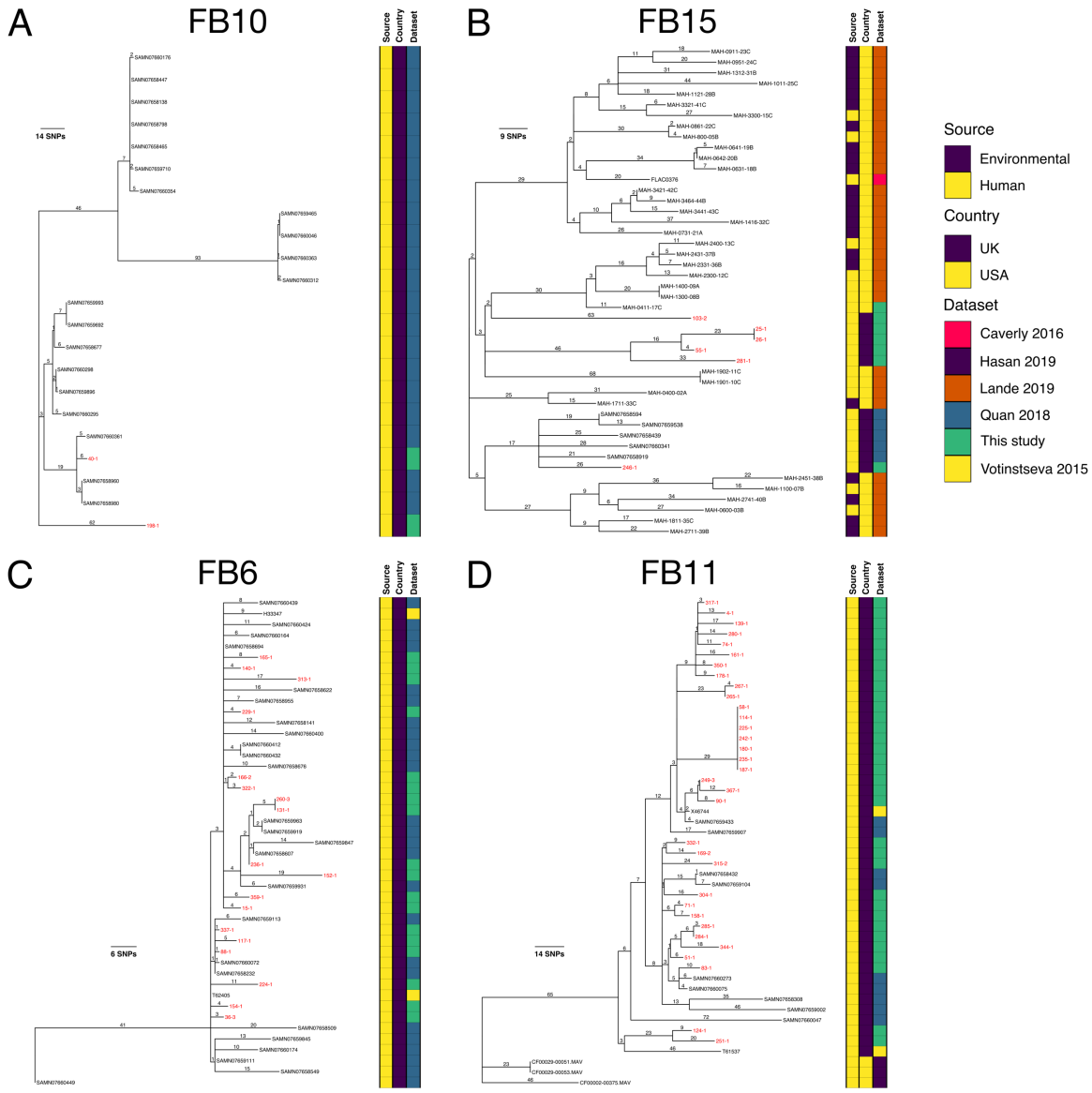
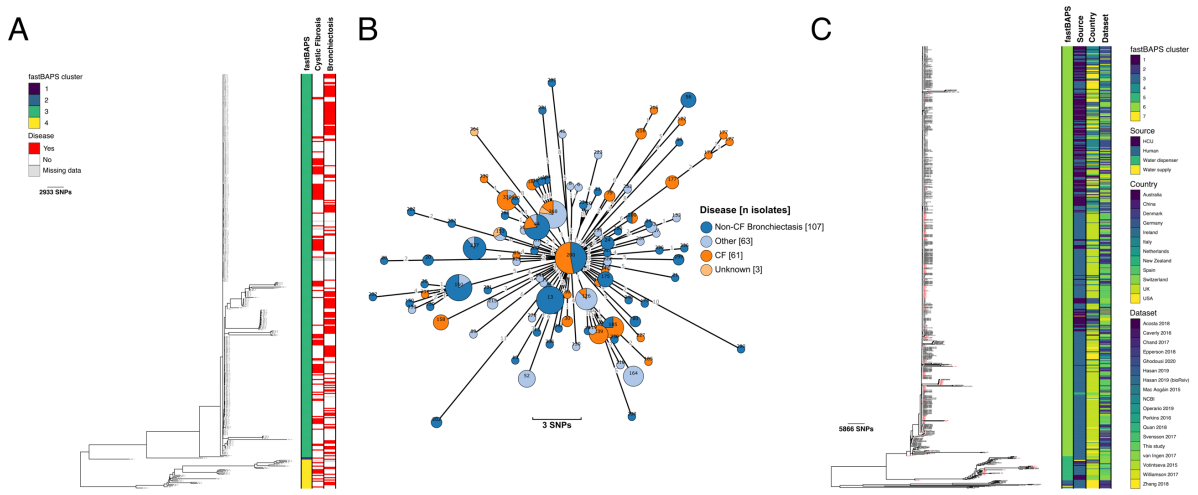


Fig 5: *Mycobacterium avium* global fastBAPS phylogenies. Midpoint-rooted maximum likelihood phylogenetic trees for A) *Mycobacterium avium* subsp. *avium* FB10; B) *Mycobacterium avium* subsp. *avium* FB15; C) *Mycobacterium avium* subsp. *hominissuis* FB6 and D) *Mycobacterium avium* subsp. *hominissuis* 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 as datastrips to the right of the phylogenies. Isolates from the

819 Royal Brompton Hospital are highlighted in red. Branch lengths are shown in SNPs

820 per site.

821



822

823

824 **Fig 6. Population structure and transmission of *Mycobacterium chimaera.*** A)

825 Midpoint-rooted maximum likelihood phylogenetic tree for *Mycobacterium chimaera*

826 isolates collected at the Royal Brompton Hospital (n = 359). The taxa are clustered

827 according to their sequence similarities. The lengths of the branches are scaled in

828 nucleotide substitutions per site. The disease status of the patients (that isolates were

829 collected from) for cystic fibrosis and non-cystic fibrosis bronchiectasis are

830 represented by red bars for Yes and white bars for No with missing patient data

831 shown by gray bars. Scale bars are shown in SNPs per site; B) *Mycobacterium*

832 *chimaera* Mc_FB3_1 transmission cluster. Each node represents an isolate or isolates

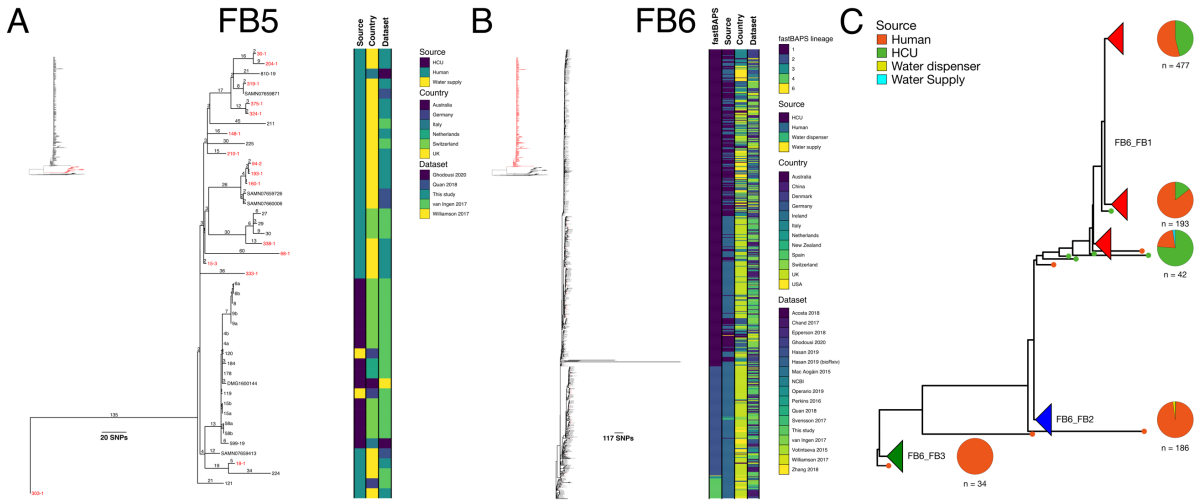
833 identical (0 SNPs) to each other and the size of the node is proportional to the number

834 of identical isolates. Nodes are coloured according to patient disease status and

835 labelled by patient number. The edges represent the pairwise SNP distance between

836 the isolate(s); C) Rooted maximum likelihood phylogenetic tree for global

837 *Mycobacterium chimaera* isolates (n = 826). The taxa are clustered according to their
838 sequence similarities. The lengths of the branches are scaled in nucleotide
839 substitutions per site. fastBAPS lineage, source of isolate, country of collection and
840 study are shown as datastrips to the right of the phylogenies. Isolates from the Royal
841 Brompton Hospital are highlighted in red. Scale bars are shown in SNPs per site.
842

843 

844 **Fig 7. *Mycobacterium chimaera* global fastBAPS cluster phylogenies.** Rooted
845 maximum likelihood phylogenetic trees for A) *Mycobacterium chimaera* FB5 (n = 45);
846 B) *Mycobacterium chimaera* FB6 (n = 765) and C) *Mycobacterium chimaera* FB6 with
847 major clades collapsed; pie charts are shown for each major clade showing the
848 distribution of sources for that clade. The taxa are clustered according to their
849 sequence similarities. The lengths of the branches are scaled in nucleotide
850 substitutions per site. The source of isolate, country of collection and study are shown
851 as datastrips to the right of the phylogenies. Isolates from the Royal Brompton
852 Hospital are highlighted in red. Branch lengths are shown in SNPs per site for
853 *Mycobacterium chimaera* FB5.
854

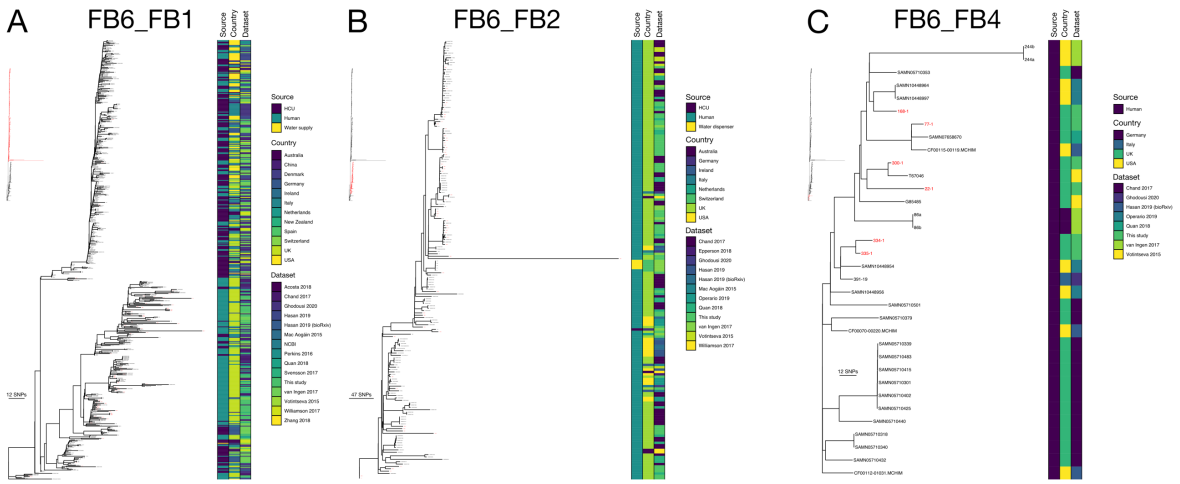


Fig 8: *Mycobacterium chimaera* fastBAPS cluster FB6 phylogenies. Maximum likelihood phylogenetic trees for A) FB6_FB1 (n = 540); B) FB6_FB2 (n = 186); and C) FB6_FB4 (n = 34). The taxa are clustered according to their sequence similarities. The lengths of the branches are scaled in nucleotide substitutions 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.

Supplementary Material

Table A. Reference genomes used in this study.

Species/ subspecies	Reference	Length (Mb)	Country of isolation	Year(s) of isolation
M. intracellulare	ATCC 13950	5.4	South Korea	2012
M. avium subsp. *avium*	104	5.5	USA	1983

M. avium subsp. *hominissuis*	TH135	5.0	Japan	2004-2008
M. chimaera	DSM 44623	5.9	Italy	1999-2003

866

867 Table B. *Mycobacterium intracellulare* transmission clusters

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)
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

868

869

870 Table C. *M. avium* subsp. *avium* transmission clusters

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)

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

Table D. *M. avium* subsp. *hominissuis* transmission clusters

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)

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

Table E. *Mycobacterium chimaera* transmission clusters

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

877
878
879 Table F. *Mycobacterium chimaera* epidemiological links
880

Patients with epidemiological link	Hospital ward	Dates of overlapping stay
24/218	FOUL	02/02/12-12/02/12
163/218	FOUL	13/02/12
175/218	FOUL	05/04/12-16/04/12
130/218	LIND	30/04/12
76/218	FOUL	06/06/12-11/06/12
32/163	LIND	29/06/12
79/241	LIND	05/07/12
241/272	LIND	06/07/12
177/186	LIND	06/08/12
122/175/306	FOUL	25/10/12-30/10/12
30/122	FOUL	05/11/12-06/11/12
30/218	FOUL	12/11/12-13/11/12

881

882 Table G. *Mycobacterium chimaera* global transmission clusters containing RBH
883 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_6	4	4	0	0
FB6_FB1	FB6_FB1_1	489	230	258	1
FB6_FB1	FB6_FB1_2	6	6	0	0
FB6_FB1	FB6_FB1_8	5	5	0	0
FB6_FB1	FB6_FB1_21	4	4	0	0
FB6_FB2	FB6_FB2_1	22	22	0	0
FB6_FB2	FB6_FB2_2	36	36	0	0
FB6_FB2	FB6_FB2_8	16	16	0	0
FB6_FB2	FB6_FB2_17	10	10	0	0
FB6_FB2	FB6_FB2_32	11	11	0	0
FB6_FB2	FB6_FB2_35	2	2	0	0
FB6_FB4	FB6_FB4_8	4	4	0	0
FB6_FB4	FB6_FB4_11	3	3	0	0
FB6_FB4	FB6_FB4_16	2	2	0	0

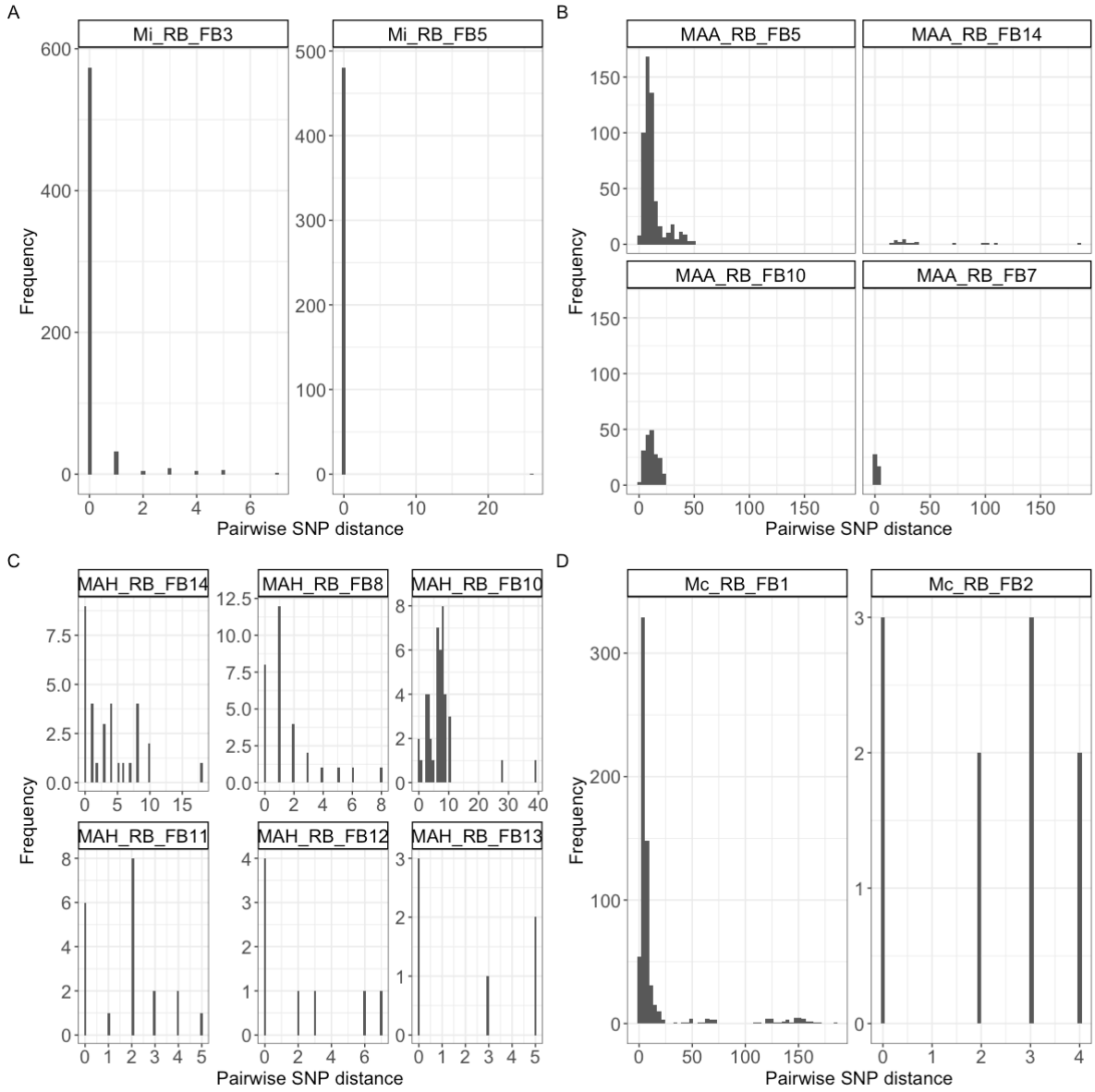


Fig A. Within-host isolate pairwise SNP diversity in fastBAPS lineages for A) *Mycobacterium intracellulare*; B) *Mycobacterium avium* subsp. *avium;* C) *Mycobacterium avium* subsp. *hominissuis* and D) *Mycobacterium chimaera*.

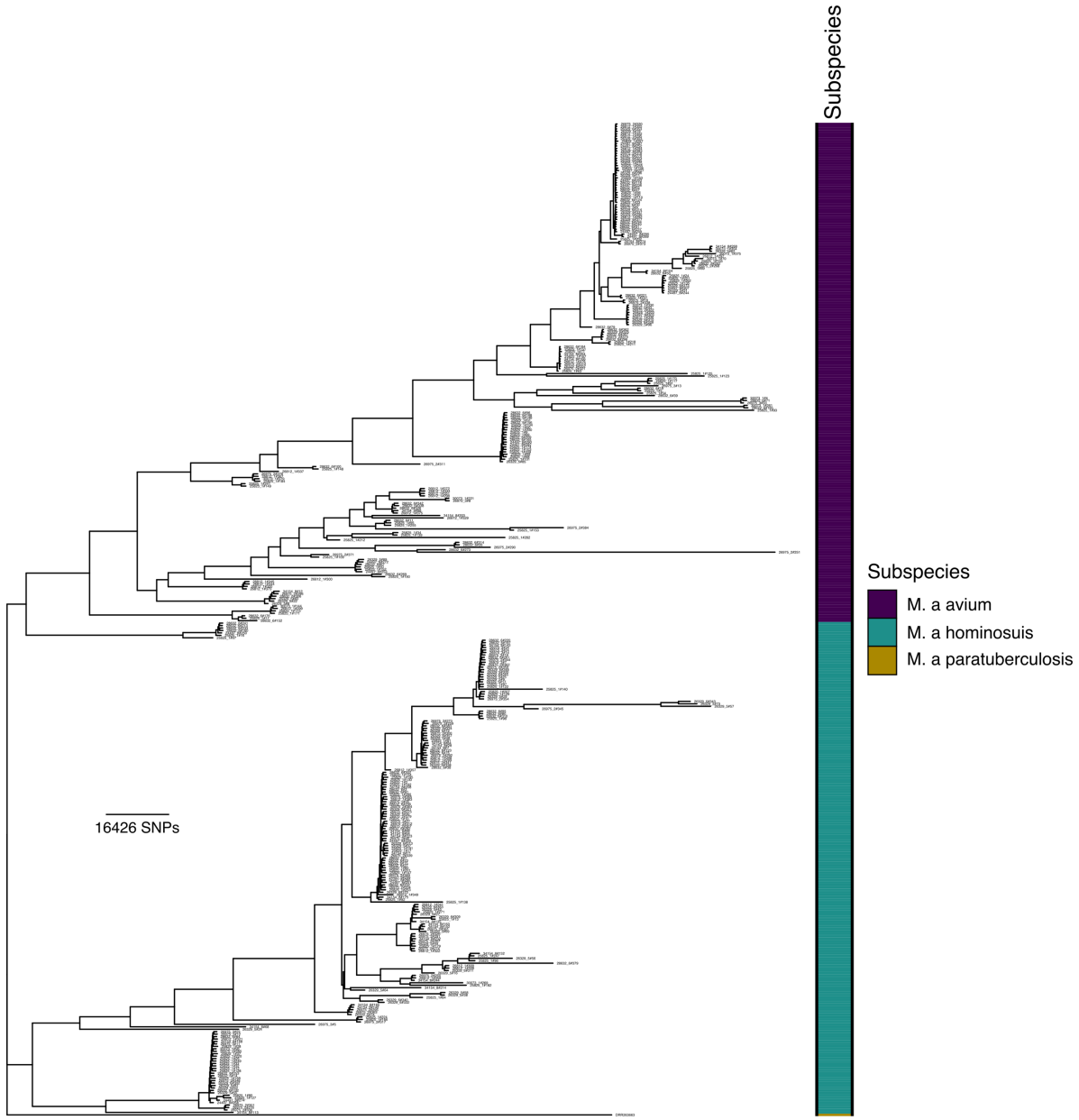


Fig B. Population structure of *Mycobacterium avium* at the Royal Brompton Hospital. Maximum likelihood phylogenetic tree of 406 *Mycobacterium avium* isolates rooted with a *Mycobacterium avium* subsp. *paratuberculosis* isolate (DRR263663). The subspecies of each isolate is shown in the datastrip to the right of the phylogeny. The scale bar is shown in SNPs per site.

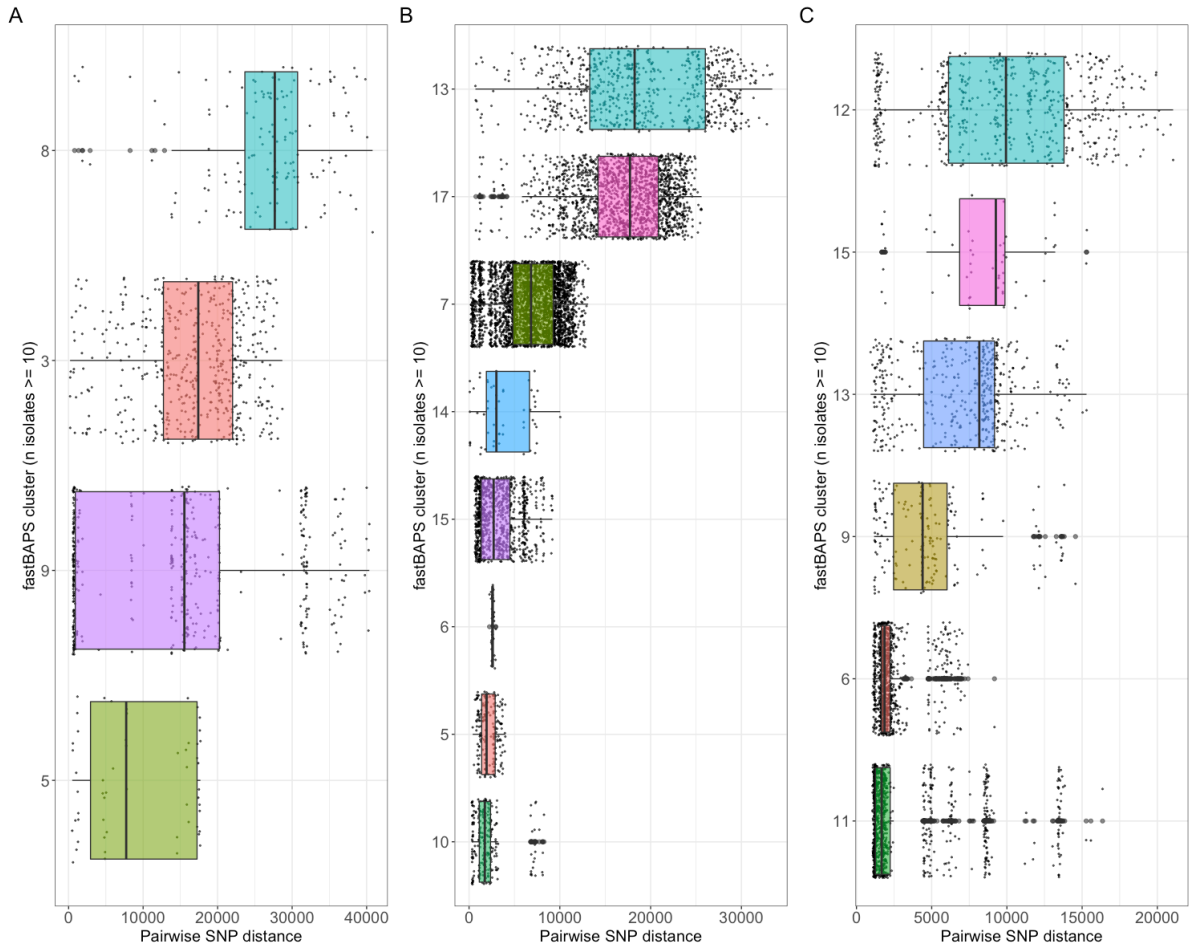


Fig C. Global fastBAPS cluster pairwise SNP distance distributions. Boxplots showing isolate pairwise SNP distances for A) *Mycobacterium intracellulare*; B) *Mycobacterium avium* subsp. *avium* and C) *Mycobacterium avium* subsp. *hominissuis*

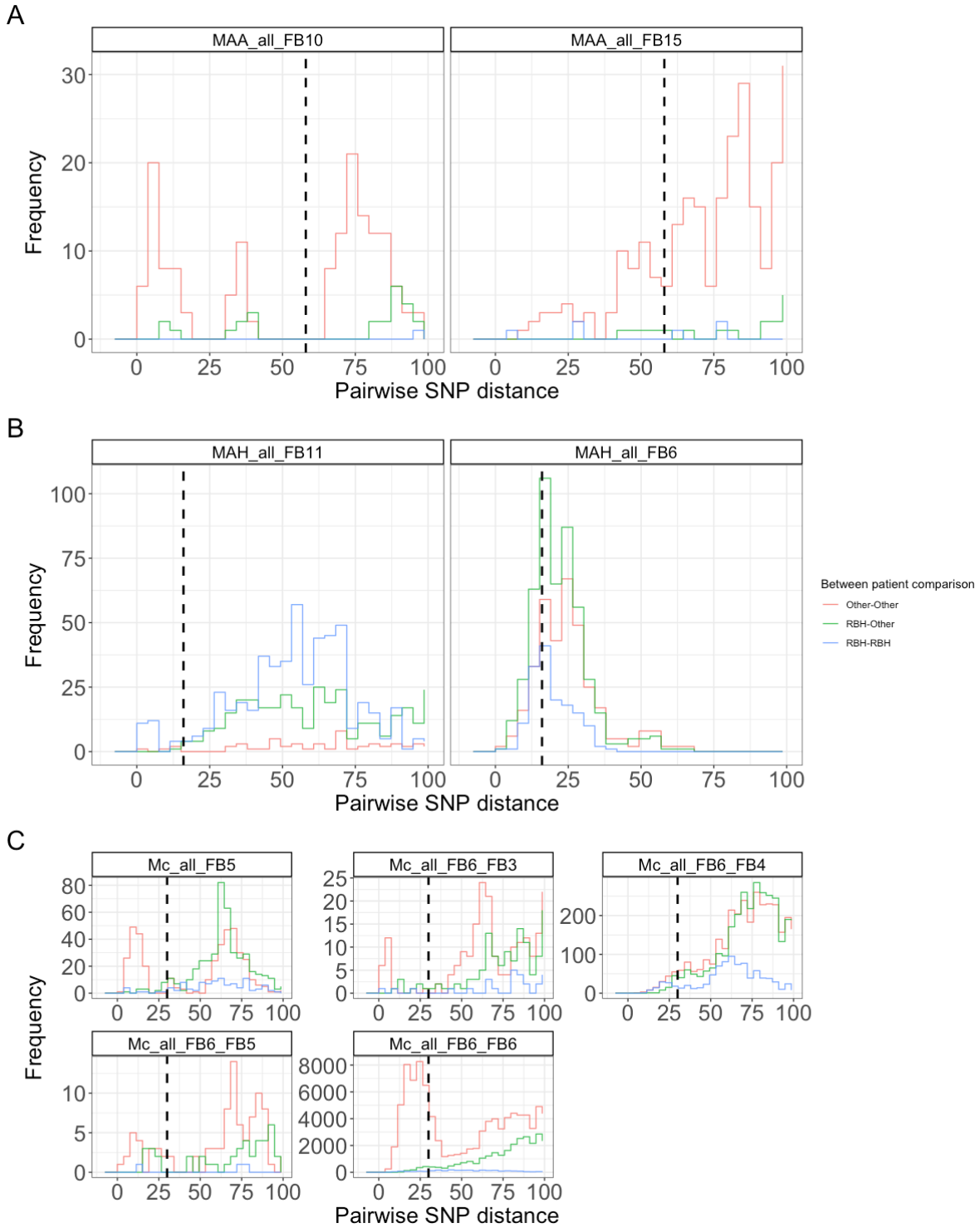


Fig D: Distribution of pairwise SNP distances between isolates from the Royal Brompton Hospital and other studies. Histograms showing distribution of pairwise

907 SNP distances for fastBAPS clusters for A) *Mycobacterium avium* subsp. *avium*; B)

908 *Mycobacterium avium* subsp. *hominissuis* and C) *Mycobacterium chimaera*.