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

Andries J van Tonder¹, Huw C Ellis^{2,3}, Colin P Churchward³, Kartik Kumar^{2,3}, Newara Ramadan⁴, Susan Benson⁴, Julian Parkhill^{1*}, Miriam F Moffatt^{¶3}, Michael R Loebinger^{¶2,3}, William OC Cookson^{¶2,3*}

¹Department 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

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

⁴Department 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.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

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.

Added value of this study

Here we present 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.

Implications of all the available evidence

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 *M. 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 (13–16). 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 (17). A 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 (18). 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; identify potential transmission chains involving patients with CF and other non-CF lung conditions; and 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).

Data collection

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

Sample collection

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

Culturing, DNA extraction and sequencing

NTM cultures were grown from bead stock cultures in BBL MGIT media (BD) in a Bactec MGIT 960 (BD) until the system indicated growth. In the absence of growth, DNA was extracted from the bead stock. DNA extractions were performed as previously described (<https://dx.doi.org/10.17504/protocols.io.bf28jqhw>). A total of 1189 DNA extracts were sequenced by the core pipeline teams at the Wellcome

Sanger Institute. Illumina libraries were created using the Nextera protocol and the Illumina Hiseq X10 platform was used to generate 2 x 150 bp paired-end reads. Raw sequencing reads were deposited at the European Nucleotide Archive under project PRJEB21813. All accessions used in this project are listed in Supplementary File 1.

Sequence QC, mapping and phylogenetics

Basic quality control metrics for the raw sequence data were generated using FastQC v0.11.9 (19). Sequence reads with similarity to *Mycobacterium* species were identified using Kraken v0.10.6 (20) and Bracken v1.0 (21). Samples with < 70% reads mapping to a *Mycobacterium* species were excluded from further analyses (n = 116). Three isolates identified as *Mycobacterium abscessus* were removed from the dataset. Sequence reads for each species were trimmed using Trimmomatic v0.33 (22) and mapped to appropriate references (Supplementary Table 1) using BWA mem v0.7.17 (minimum and maximum insert sizes of 50 and 1000 respectively) (23). Single nucleotide polymorphisms (SNPs) were called using SAMtools v1.2 mpileup and BCFtools v1.2 (minimum base call quality of 50 and minimum root squared mapping quality of 30) as previously described (24). Samples with reads that mapped to < 80% of the reference were excluded (n = 70). Variant sites were extracted from the resulting alignments using snp-sites v2.5.1 (25). Whole species maximum likelihood phylogenetic trees were built using IQ-tree v1.6.5 accounting for constant sites (-fconst; determined using snp-sites -C) with the built-in model testing (-m MFP) to determine the best phylogenetic model and 1000 ultrafast bootstraps (-bb 1000) (26).

Recombinant regions were identified and removed from alignments using GUBBINS (27) and new phylogenetic trees were constructed as described above. Pairwise SNP distances were calculated for all pairs of isolates using pairsnp (28).

Global collections

To provide context for each the isolates sequenced for each species in this study, datasets consisting of published sequenced isolates were assembled (Supplementary File 2) (12,29–48). Sequence data were downloaded from the European Nucleotide Archive (ENA) and trimmed; Sample QC, mapping and phylogenetic tree construction were performed as detailed above.

Genome assemblies

A previously published pipeline was used to produce annotated assemblies (49). Briefly, sequence reads were assembled with spades v 3.10.10 (50) and assemblies were improved by first scaffolding the assembled contigs using SSPACE v2.0 (51) and filling the sequence gaps with GapFiller v1.11 (52).

Transmission and epidemiological linkage

Genomic clusters were identified using fastBAPS (53) and new alignments were created for clusters $\geq$ ten isolates by aligning sequence reads for included isolates against the assembly that had the smallest number of contigs (using the method described above). In order to calculate a pairwise SNP threshold to determine putative transmission clusters within each genomic cluster, pairwise SNP distances for all isolates for each species in the RBH datasets were calculated. Using a

previously described method (54), the transmission threshold for each species, regardless of cluster, was calculated by taking the 95th percentile of the maximum within-patient isolate pairwise SNP distances for all patients and adding twice the number of mutations expected to occur in a six month period. To account for excess within-patient diversity observed in the *M. chimaera* FB1 and *M. a. avium* FB14 clusters (Supplementary Figure 1), pairwise SNP distances greater than 25 and 50 (assumed to result from infection with multiple lineages) were removed respectively before the above calculations were performed. Based on these results, the R library iGRAPH (55,56) and pairwise SNP thresholds of 16 (*M. intracellulare* and *M. a. hominissuis*), 30 (*M. chimaera*) and 58 SNPs (*M. a. avium*) were used to calculate putative transmission clusters in each genomic cluster. Finally, in order to identify possible epidemiological links between patients infected with the same transmission clusters, hospital stay records were examined for epidemiological contacts. The latter were defined as patients attending the same ward on the same day up to one year prior to the collection of the first sequenced isolate.

Results

Patient demographics

The median age and BMI of the 354 patients included in the study was 56 years (range 5 - 93) and 22.5 (range 13.4 – 43.4) years respectively. One hundred and seventy-four patients (49.1%) were male and 38/354 (10.7%) were smokers. There were 147/354 (41.5%) patients with BX, 87/354 (24.6%) with CF, 53/354 (15.0%) with COPD, 32/354 (9.0%) with asthma, 19/354 (5.4%) with allergic bronchopulmonary aspergillosis (ABPA), 17/354 (4.8%) with interstitial lung disease (ILD) and 7/354

(2.0%) with other underlying respiratory conditions such as pleural thickening or sarcoidosis. Patients with no other respiratory disease or predisposition to NTM infections accounted for 3.4% (12/354) of the cohort. During the study period, 55/354 (15.5%) patients were on antibiotic treatment regimes. For 20 patients (5.6%) clinical data were unavailable.

Species distribution

A total of 1,000 isolates from 354 patients were successfully sequenced and the 11 MAC species identified are detailed in Table 1. The three predominant species amongst the sequenced isolates were *M. avium* (*M. a. avium* and *M. a. hominissuis*), *M. chimaera* and *M. intracellulare*. Together these accounted for 926/1000 (92.6%) of the MAC isolates sequenced. Most patients were infected with only a single species during the collection period. However, 46 of the 354 patients (13.0%) were infected with two or more species (Supplementary Figure 2). In this group of patients, the majority of isolates collected were typically from a single species, with other species observed more infrequently (Figure 1). Subsequent analyses in this study will focus on the three predominant species in the dataset: *M. intracellulare*, *M. avium* (*M. a. avium* and *M. a. hominissuis*), and *M. chimaera*.

M. intracellulare

A total of 162 genomes from 37 patients were characterised as *M. intracellulare* (Figure 2A). Eleven of the patients had CF, seventeen had BX, with the remaining seven having other lung conditions (COPD n = 3; ILD n = 3; asthma n = 1; congenital pulmonary airway malformation [CPAM] n = 1) and disease metadata were missing

for two patients. Genomic clustering with fastBAPs identified nine clusters with three of these having more than ten genomes. Following remapping to local references for the three largest fastBAPS clusters, three putative transmission clusters were identified with the largest, Mi_FB3_1, composed of 16 patients (Figure 3A; Supplementary Table 2). Of these 16 patients, eight had BX, seven CF and one ILD. The range of the number of isolates collected per patient was between one and 32. Four of the sixteen patients were also infected with other species or lineages during the sampling period with Mi_FB3_1 only being detected in 1/22 isolates collected for patient 218 (Figure 3A). During the time period that the sequenced isolates were collected, four of the patients (85, 95, 97 and 99), were being treated with antibiotics and three of the four patients were treated successfully. No epidemiological links were identified between patients in the year prior to isolate collection.

Comparison with 77 previously published *M. intracellulare* isolates showed that the most closely related contextual isolate was collected in the UK in 2015/2016 [32] and formed part of a clade containing isolates from the Mi_FB3_1 cluster (Figure 4A). The remainder of the RBH isolates were evenly distributed through the global phylogenetic tree.

M. avium

A total of 405 sequenced isolates collected from 176 patients were identified as *M. avium*. Of the patients infected with *M. avium*, 44 had CF and 81 BX. Sequence data for all *M. avium* isolates was mapped to a single reference (*Mycobacterium avium* 104; NC008595.1) and a phylogenetic tree constructed with a previously published

M. avium paratuberculosis isolate (DRR263663) chosen as an outgroup (Supplementary Figure 3). The structure of this phylogeny showed that there were two major clades which corresponded to the two subspecies *M. avium avium* (n = 207; MAA) and *M. avium hominissuis* (n = 198; MAH). Isolates for each of the two subspecies were next analysed separately. A small number of patients were infected with both subspecies (n = 6). No examples of the other two *M. avium* subspecies, *paratuberculosis* and *silvaticum*, were identified in this study.

M. avium avium (MAA)

Of the 76 patients infected with MAA, 15 had CF, 39 had BX and 16 had other lung conditions (COPD n = 10; ILD n = 3; ABPA n = 3; asthma n = 6) whilst three patients, including one smoker, had no pre-existing respiratory disease (metadata were unavailable for three patients; Figure 2B). Clustering of the 207 MAA genomes identified 17 fastBAPS clusters (Figure 2B). Remapping of the four largest fastBAPS clusters allowed the identification of seven putative transmission clusters comprising between two and ten patients (Supplementary Table 3) with the largest cluster of ten isolates comprised of patients with CF (n = 2), BX (n = 6) and asthma or COPD (n = 2; Figure 3B).

A total of 21 fastBAPS clusters were defined in the global phylogeny. Examination of the distribution of pairwise SNP distances for each fastBAPS cluster containing at least ten genomes revealed that there were two clusters containing RBH isolates with lower median pairwise SNP distances (FB10 and FB15; Figure 4B and Supplementary Figure 5B). The 23 genomes, including two RBH genomes, in FB10 were collected

in the UK between 2013 and 2016, with most coming from the National Mycobacterial Reference Service. The latter characterises mycobacterial cultures from across the Midlands and North of England (Figure 5A) [32]. If the SNP threshold used to calculate putative transmission clusters circulating in the RBH was applied, then one isolate from the RBH (40-1) formed part of a cluster comprised of isolates from Quan *et al.* [32]. The majority (35/46) of the isolates forming FB15 were collected in the USA with most from a study investigating *M. avium* in the community and household water in Philadelphia (Figure 5B) [28]. The pairwise distances between two of the RBH isolates and other UK isolates and isolates from the USA were within the threshold used to define transmission, with the above suggesting potential cross-Atlantic transmission (Supplementary Figure 6A).

M. avium hominissuis (MAH)

A total of 198 genomes from 106 patients were characterised as MAH (Figure 2C). Thirty-three patients had CF, 43 BX and 21 had other lung conditions (COPD n = 14; ILD n = 1; ABPA n = 5; sarcoidosis n = 1; asthma n = 11). Four patients had no pre-existing respiratory disease. Disease metadata were unavailable for the remaining five patients. Genomic clustering identified 17 clusters and seven putative transmission clusters containing between two and 16 patients (Supplementary Table 4). The largest cluster of 16 isolates is shown in Figure 3C and comprised patients with CF (n = 4), BX (n = 7), COPD (n = 3) and a single patient with no pre-existing lung condition (metadata were unavailable for one patient). Two pairs of patients in each of the clusters, MAH_FB8_1 and MAH_FB14_9, were found to have epidemiological links.

Of the 20 defined fastBAPS clusters in the global collection, FB6 and FB11 had lower median pairwise SNP distances and contained RBH isolates (Figure 4C and Supplementary Figure 5C). Both clusters were composed completely or near-completely of isolates collected in the UK (Figure 5C and Figure 5D). The star-like structure of the FB6 phylogeny is suggestive of a point source outbreak and applying a transmission SNP threshold of 16 SNPs showed that 41/45 of the isolates in FB6 would form a transmission cluster that included both isolates from this study and isolates from other UK studies (Supplementary Figure 6B). Two small transmission clusters, containing RBH and other UK isolates, were also identified in FB11.

M. chimaera

There were 359 sequenced isolates collected from 155 patients identified as *M. chimaera*; 37 of the patients had CF, 60 had BX whilst 50 patients had other lung conditions (COPD n = 24; ILD n = 10; ABPA n = 11; pleural thickening with enfolded lung n = 1; lung non-mucinous adenocarcinoma n = 1; Wegener's granulomatosis n = 1; *M. tuberculosis* in lymph nodes during same endoscopy procedure n = 1; primary ciliary dyskinesia n = 1; asthma n = 17). There were five patients with no pre-existing respiratory disease and metadata were unavailable for a further nine patients. A single predominant lineage, Mc_FB1, corresponding to the previously characterised Group 1 accounted for 332/359 (92.5%) of the sequenced isolates with the majority of the remaining isolates (25/359; 7.0%) forming a second main lineage Mc_FB2 corresponding to Group 2 (Figure 2D). Thirteen putative transmission clusters containing two to 106 patients were identified (Supplementary Table 5). Twelve potential epidemiological links (Supplementary Table 6) were found in transmission

cluster Mc_RB_FB3_1, the largest transmission cluster of 106 patients. Of the patients in this cluster 43 had BX, 24 CF, 13 COPD, 14 asthma, seven ILD, two ABPA, one pleural thickening with enfolded lung, one lung non-mucinous adenocarcinoma and one primary ciliary dyskinesia. Three patients had no pre-existing lung condition and disease status was missing for five patients. Fifteen patients underwent antibiotic treatment during the study period (Supplementary Figure 4).

The global phylogenetic tree containing 155 isolates from RBH and 671 previously published isolates was topologically similar to the tree built using only the isolates from RBH with a single major clade with low genetic diversity present (FB6; Figure 6A). Isolates from the two clusters with more than ten representatives (FB5 and FB6) were remapped to local references. Recombination was removed and new phylogenetic trees constructed (Figure 6B [FB6] and Figure 6C [FB5]). Most isolates in FB5 were collected from patients with respiratory conditions or else patients with infections following cardiac surgeries. There was, however, a distinct sub-clade rooting within this diversity containing only isolates from HCUs or hospital water supplies (Figure 6C). Within FB6 a more deeply rooting diverse clade contained mainly human isolates and the larger shallower clade rooting within this contained most of the HCU isolates interspersed with human isolates. Due to the large number of isolates assigned to FB6 ($n = 765$), further genomic clustering was performed using fastBAPS to identify six sub-clusters within FB6. New mapping was performed for the four clusters with more than ten isolates (FB6_FB3 to FB6_FB6; Figure 7A-D). Three of these clusters, FB6_FB3, FB6_FB4 and FB6_FB5, were completely or nearly completely composed of patient isolates (Figure 7B-D). Most of the isolates collected

from HCUs or water supplies made up the bulk of FB6_FB6, though human isolates were distributed throughout the tree (Figure 7A). Putative transmission clusters were calculated for all five fastBAPS lineages described above used a pairwise SNP threshold of 30 SNPs (Supplementary Table 7). Transmission clusters containing RBH isolates were found in each of the lineages with the largest, found in FB6_FB6, comprising 255 HCU isolates, 228 patient isolates and a single isolate from a water supply.

Discussion

The aims of this study were to characterise the population structure of MAC and identify potential transmission between CF and non-CF patients. We used a dataset of 1,000 sequenced MAC isolates collected in a London hospital. This is, to date, the largest sequenced MAC isolate collection analysed from a single location. We showed that polymicrobial infection with MAC is not uncommon. The large, longitudinal dataset enabled us to identify putative transmission clusters for all four species/sub-species we considered. Most of these transmission clusters contained both patients with CF and non-CF patients, showing that transmission of MAC is occurring despite the presence of hospital-based control measures focussing on patients with CF. The lack of clear epidemiological links suggests that the majority of transmission is not due to direct contact but rather through an alternative route such as an environmental vector or asymptomatic carriage in the wider population. Comparison of the isolates collected as part of this study with published isolates identified a small number of lineages likely circulating in the UK. With the exception of the major HCU-associated lineage in *M. chimaera*, the comparison did not identify

globally circulating lineages like those observed in *M. abscessus*. Surprisingly, given the previous description of the *M. chimaera* HCU clade as representing a point source outbreak, we provide strong evidence that HCU-associated *M. chimaera* isolates are derived from existing lineages circulating amongst patients with pre-existing lung conditions, and that this has happened on at least two separate occasions. This indicates that, while infection and distribution of HCUs has contributed to the transmission of these lineages, they were not the primary source of these lineages, which were already circulating in the human population.

During the isolate sampling period we identified 46 patients infected with more than one species or lineage. This suggests that many of the patients had polymicrobial infections and that certain lineages were preferentially sampled and cultured at different times. This has been observed previously in other longitudinal samples of patients with MAC lung disease (37). These observations suggest that single colony picks during culturing potentially may be underrepresenting the underlying diversity of MAC species in patients with MAC lung disease. Therefore, to better understand the ecology and transmission of MAC species in the lung, a deep sequencing approach in which colony sweeps rather than colony picks are utilised should be considered for future studies.

Previous studies have used SNP thresholds of 15-25 SNPs to identify putative transmission clusters in NTM species (13,17,32). Rather than using a previously applied or arbitrary threshold, we used a method that had previously been used to define SNP thresholds in *Staphylococcus aureus* (54). This allowed us to calculate a

SNP threshold for each species using the longitudinal isolates collected by the RBH. The thresholds calculated for *M. intracellulare* and MAH of 16 SNPs were similar to those previously applied. Whilst we obtained higher thresholds of 30 SNPs and 58 SNPs for *M. chimaera* and MAA respectively, this did not result in an inflated number of transmission clusters, as applying a threshold of 16 SNPs to the MAA dataset would have resulted in only two fewer clusters. These larger values were predominantly due to high levels of within-host diversity observed amongst patient isolates in single clusters in both MAA and *M. chimaera*. For instance, in MAA_FB14 isolates from a single patient collected a day apart were as much as 101 SNPs different. This highlighted the likelihood of polymicrobial infections being common amongst the patients we included in the study.

Using the SNP thresholds calculated for each species, we were able to identify putative transmission clusters amongst all four of the MAC species/subspecies we considered. Broadly, two different patterns of infection were observed: amongst *M. intracellulare* and the two *M. avium* subspecies, our results suggested that there have been multiple independent acquisitions followed by limited transmission. Our analysis of *M. chimaera*, however, identified a single large transmission cluster was responsible for infecting 106/155 (68.4%) of our patients infected with this species. Most of these transmission clusters contained both CF and non-CF patients and a small number also contained patients with no pre-existing lung conditions.

Using our criteria, we were unable to identify potential epidemiological links for most patients included in the transmission clusters. The exception to this was the largest

M. chimaera transmission cluster where we found 12 potential epidemiological links up to a year before sampling began. Interestingly, four patients were part of more than one link, with patient 218 overlapping with other patients in this transmission cluster on the same wards six times over a nine-month period. The tight infection control surrounding patients with CF means that they are unlikely to have interacted with patients with other lung conditions whilst in hospital. This implies that transmission is likely occurring *via* other pathways, such as through environmental intermediates or through a reservoir of healthy, asymptomatic carriers in the wider population.

This observation is not unique and for some time, there has been considerable debate whether NTM transmission between patients is occurring, even when transmission is strongly supported by genome-based analysis (13,17,57). The near complete absence of epidemiological links between patients in transmission clusters as well as the ubiquitous presence of NTM species in the environment, especially water supplies, has led some researchers to suggest that the dissemination of NTM lineages at a national level is associated with exposure to contaminated water supplies (17). Whilst there is certainly strong evidence for the same lineages being isolated from water supplies and patients in the same location [28], the absence of a single national water supply in a comparatively small country like the UK would suggest that closely related isolates collected from different geographical locations are unlikely to be due to a single contaminated water supply. It is however possible that transmission networks may include local water supplies as intermediates. The alternative hypothesis that NTM populations are being maintained in apparently

healthy individuals with no symptoms of NTM lung disease is a compelling one. This would explain the presence of local transmission clusters in the absence of direct epidemiological links as well as the presence of phylogeographical structure in NTM trees. To better address how NTM species are being transmitted, future studies should focus on attempting to collect isolates from the general population, perhaps focussing on smokers where there is some evidence that they may provide a better reservoir for NTM species and other opportunistic pathogens (18).

The inclusion of published genomes alongside the RBH isolates allowed us to use our defined pairwise SNP thresholds to identify putative MAC lineages circulating in the UK and globally. Apart from the *M. chimaera* clusters that we found associated with HCUs, we were unable to identify large globally circulating clusters like those observed in *M. abscessus*. This potentially suggests different evolutionary dynamics amongst the MAC species or, in the case of *M. intracellulare*, may simply reflect a lack of comprehensive sampling. We were only able to identify a single cluster containing RBH isolates amongst the two *M. avium* subspecies that included isolates from outside the UK. This cluster comprised the majority of isolates in FB15 (Figure 5B). Given that most available *M. avium* genomes have been collected in the UK or USA, a more comprehensive global sampling strategy could potentially reveal the existence of globally circulating clones.

A different picture emerged when we calculated global transmission clusters for *M. chimaera* using a pairwise SNP threshold of 30 SNPs (Supplementary Table 7). Here we found most clusters consisted of isolates from across Europe or else included

isolates from across the globe. Notably, we identified a single large transmission cluster containing 484 isolates from 12 countries that included most of the HCU-associated isolates.

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

552 AvT, CC

553 **Project administration**

554 AvT, CC, MM, WC

555 **Resources**

556 CC, NR, SB

557 **Software**

558 NA

559 **Supervision**

560 JP, WC, MM, ML

561 **Validation**

562 AvT

563 **Visualisation**

564 AvT

565 **Writing – original draft**

566 AvT, JP

567 **Writing – review & editing**

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

569

570 **Conflicts of interest**

571 JP is a paid consultant to NextGen Diagnostics LLC.

572

573 **Data sharing**

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

References

1. Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, et al. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am J Respir Crit Care Med*. 2007 Feb 15;175(4):367–416.
2. Chan ED, Bai X, Kartalija M, Orme IM, Ordway DJ. Host immune response to rapidly growing mycobacteria, an emerging cause of chronic lung disease. *Am J Respir Cell Mol Biol*. 2010 Oct;43(4):387–93.
3. Iseman MD, Marras TK. The importance of nontuberculous mycobacterial lung disease. *Am J Respir Crit Care Med*. 2008 Nov 15;178(10):999–1000.
4. Donohue MJ, Wymer L. Increasing Prevalence Rate of Nontuberculous Mycobacteria Infections in Five States, 2008–2013. *Ann Am Thorac Soc*. 2016 Dec;13(12):2143–50.
5. Moore JE, Kruijshaar ME, Ormerod LP, Drobniewski F, Abubakar I. Increasing reports of non-tuberculous mycobacteria in England, Wales and Northern Ireland, 1995–2006. *BMC Public Health*. 2010 Oct 15;10:612.
6. Philley JV, Griffith DE. Treatment of slowly growing mycobacteria. *Clin Chest Med*. 2015 Mar;36(1):79–90.
7. Adjemian J, Olivier KN, Prevots DR. Epidemiology of Pulmonary Nontuberculous Mycobacterial Sputum Positivity in Patients with Cystic Fibrosis in the United States, 2010–2014. *Ann Am Thorac Soc*. 2018 Jul;15(7):817–26.
8. Hoefsloot W, van Ingen J, Andrejak C, Angeby K, Bauriaud R, Bemer P, et al. The geographic diversity of nontuberculous mycobacteria isolated from pulmonary samples: an NTM-NET collaborative study. *Eur Respir J*. 2013 Dec;42(6):1604–13.
9. Olivier KN, Weber DJ, Wallace RJ, Faiz AR, Lee J-H, Zhang Y, et al. Nontuberculous mycobacteria. I: multicenter prevalence study in cystic fibrosis. *Am J Respir Crit Care Med*. 2003 Mar 15;167(6):828–34.
10. Achermann Y, Rössle M, Hoffmann M, Deggim V, Kuster S, Zimmermann DR, et al. Prosthetic valve endocarditis and bloodstream infection due to *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 Szafrńska
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, Koeppl 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-](https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-team-2006)
756   [maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-](https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-team-2006)
757   [team-2006](https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-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

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

Tables and Figures

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

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

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

776 together. Patients with one isolate or multiple isolates from a single species were
777 excluded.
778

779

780 **Figure 2: Population structure of *Mycobacterium avium* Complex species**

781 **collected at the Royal Brompton Hospital. Midpoint-rooted Maximum likelihood**

phylogenetic trees for A) *Mycobacterium intracellulare* (n = 162); B) *Mycobacterium avium avium* (n = 207); C) *Mycobacterium avium hominosis* (n = 198) and D) *Mycobacterium chimaera* (n = 359). fastBAPS clusters are shown as a coloured datastrip in each plot. 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.

Figure 3. Largest transmission clusters amongst isolates collected at the Royal Brompton Hospital. A) *Mycobacterium intracellulare* (Mi_FB3_1); B) *Mycobacterium avium avium* (MAA_FB5_1) and C) *Mycobacterium avium hominosis* (MAH_FB8_1). 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, 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 hominosis* (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* FB10; B) *Mycobacterium avium* FB15; C) *Mycobacterium avium hominosis* FB6 and D) *Mycobacterium avium hominosis* 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

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
<i>M. intracellulare</i>	ATCC 13950	5.4	South Korea	2012
<i>M. avium avium</i>	104	5.5	USA	1983
<i>M. avium hominosis</i>	TH135	5.0	Japan	2004-2008
<i>M. chimaera</i>	DSM 44623	5.9	Italy	1999-2003

849

850 **Supplementary Table 2. *Mycobacterium intracellulare* transmission clusters**

fastBA PS cluster	Transmiss ion cluster	Patien ts (n)	Bronchieoct osis (n)	Cysti c Fibro sis (n)	Other lung conditio ns (n)	No pre- existin g lung conditi on (n)	Patients undergoi ng treatmen t (n)
Mi_FB 2	Mi_FB2_1	2	0	2	0	0	1
Mi_FB 3	Mi_FB3_1	16	8	7	1	0	4
Mi_FB 5	Mi_FB5_1	3	2	0	1	0	2

851

852

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

fastBAP S cluster	Transmiss ion cluster	Patie nts (n)	Bronchieoct osis (n)	Cysti c Fibro sis (n)	Other lung conditi ons (n)	No pre- existin g lung conditi on (n)	Patients undergo ing treatme nt (n)
MAA_F B5	MAA_FB5 _1	10	6	2	2	0	4
MAA_F B5	MAA_FB5 _4	2	1	1	0	0	0
MAA_F B5	MAA_FB5 _5	2	1	0	1	0	1
MAA_F B5	MAA_FB5 _12	2	1	1	0	0	1
MAA_F B7	MAA_FB7 _1	2	2	0	0	0	1
MAA_F B10	MAA_FB1 0_1	2	0	2	0	0	0
MAA_F B14	MAA_FB1 4_1	2	2	0	0	0	1

854

855

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

fastBAP S cluster	Transmissi on cluster	Patie nts (n)	Bronchieoct osis (n)	Cysti c Fibro sis (n)	Other lung conditi ons (n)	No pre- existin g lung conditi on (n)	Patients undergo ing treatme nt (n)
MAH_F B8	MAH_FB8 _1	16	7	4	3	1	0
MAH_F B10	MAH_FB1 0_1	4	2	2	0	0	1
MAH_F B11	MAH_FB1 1_1	6	3	1	2	0	0
MAH_F B12	MAH_FB1 2_2	4	1	2	1	0	2
MAH_F B14	MAH_FB1 4_3	4	2	2	0	0	0
MAH_F B14	MAH_FB1 4_7	2	0	2	0	0	0
MAH_F B14	MAH_FB1 4_9*	7	3	2	1	0	1

857

858 Supplementary Table 5. *Mycobacterium chimaera* transmission clusters

859

fastBA PS cluster	Transmiss ion cluster	Patien ts (n)	Bronchieoct osis (n)	Cysti c Fibro sis (n)	Other lung conditio ns (n)	No pre- existin g lung conditi on (n)	Patients undergoi ng treatmen t (n)
Mc_FB 3	Mc_FB3_ 1	106	43	24	25	3	15
Mc_FB 3	Mc_FB3_ 3	2	1	1	0	0	0
Mc_FB 3	Mc_FB3_ 4	2	2	0	0	0	0
Mc_FB 3	Mc_FB3_ 6	2	1	0	1	0	0
Mc_FB 3	Mc_FB3_ 13	4	1	1	1	1	1
Mc_FB 3	Mc_FB3_ 16	2	0	0	1	1	1
Mc_FB 3	Mc_FB3_ 17	2	1	1	0	0	0
Mc_FB 3	Mc_FB3_ 19	2	0	1	0	1	0
Mc_FB 3	Mc_FB3_ 25	2	1	1	0	0	1

Mc_FB 3	Mc_FB3_ 30*	4	0	1	0	0	0
Mc_FB 4	Mc_FB4_ 1	2	1	1	0	0	1
Mc_FB 4	Mc_FB4_ 2	9	3	3	3	0	2
Mc_FB 4	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

864

865 Supplementary Table 7. *Mycobacterium chimaera* global transmission clusters

866 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

867

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

874

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

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

Supplementary Figure 3. Population structure *Mycobacterium avium* at the Royal Brompton Hospital. Maximum likelihood phylogenetic tree of 406 *Mycobacterium avium* isolates rooted with a *M. avium 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.

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).

Supplementary Figure 5. Global fastBAPS cluster pairwise SNP distance distributions. Boxplots showing isolate pairwise SNP distances for A) *Mycobacterium intracellulare*; B) *Mycobacterium avium avium* and C) *Mycobacterium avium hominosis*

897

898

899 **Supplementary Figure 6: Distribution of pairwise SNP distances between**
900 **isolates from the Royal Brompton Hospital and other studies.** Histograms
901 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