EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original research article

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

Andries J van Tonder, Huw C Ellis, Colin p Churchward, Kartik Kumar, Newara Ramadan, Susan Benson, Julian Parkhill, Miriam F Moffatt, Michael R Loebinger, William OC Cookson

Please cite this article as: van Tonder AJ, Ellis HC, Churchward Cp, *et al*. *Mycobacterium avium* complex (MAC) genomics and transmission in a London hospital. *Eur Respir J* 2022; in press (https://doi.org/10.1183/13993003.01237-2022).

# *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 author: jp369@cam.ac.uk

¶ Contributed equally

## Article summary

Transmission of *Mycobacterium avium* complex (MAC) species is indirect and may involve environmental intermediates or asymptomatic carriage in the wider population.

## Abstract

### Background

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

### Methods

To investigate the potential for transmission of the *Mycobacterium avium* complex (MAC) we sequenced 996 isolates from 354 CF and non-CF patients at the Royal Brompton Hospital (RBH), London (collected 2013-2016), and analysed them in a global context. Epidemiological links were identified from patient records. Previously published genomes were used to characterise global population structures.

### Results

We identified putative transmission clusters in three MAC species, 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.

### Conclusions

CF and non-CF patients shared transmission chains, although the lack of epidemiological links suggested that most transmission is indirect and may involve environmental intermediates or asymptomatic carriage in the wider population.

## 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]. Globally, disease due to NTM infections is increasing in prevalence; the estimated prevalence of NTM disease in the United States of America (USA) rose from 2.4/100,000 in the early 1980s to 15.2/100,000 in 2013 [4]. NTM infections may be progressive and treatment requires prolonged multi-drug therapy [5]. Treatment 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 [6–8]. 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 LivaNova heater-cooler units (HCUs) contaminated during their manufacture [9–11].

Until recently the prevailing hypothesis was that NTM infections were due to independent acquisitions from environmental sources such as soil, contaminated drinking water distribution systems and household plumbing [12,13]. 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 [14–18]. An observational study of *M. abscessus,* across England showed that these dominant clones are also found in patients with other chronic respiratory diseases, but was unable to identify epidemiological links for most closely-related isolates, suggesting environmental acquisition [19]. However, a wider analysis demonstrated that transmission networks involve both people with CF and those without, and that these networks are global. It is therefore likely that transmission is complex, involving multiple patient cadres as well as environmental intermediates [20]. In the case of *M. chimaera*, a global genetic analysis of 250 isolates from patients, HCUs and the factory of origin suggested a point source contamination during manufacture causing global distribution followed by localised transmission [11].

Whilst transmission of MAC between CF patients in the USA has been investigated [21], little work has been done to examine whether transmission of MAC occurs 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.

## Methods

*Bacterial isolates and study period*

Isolates were collected from 363 patients attending the respiratory inpatient and outpatient clinics of the RBH between January 2013 and April 2016. DNA extractions were performed on confirmed MAC cultures and the Illumina Hiseq X10 platform was used to generate 2 x

150bp paired-end reads (Supplementary Text S1). Sequencing reads were deposited in the European Nucleotide Archive under project PRJEB21813 (Supplementary File 2).

*Whole genome sequencing and data analysis*

Following sample QC, whole species maximum likelihood phylogenetic trees were built with IQ-tree version 1.6.5 using a previously described pipeline [22] (Supplementary Text S1), including published sequences for context (Supplementary File 3). Pairwise SNP distances for within-patient longitudinal isolates for each species in the RBH datasets were used to calculate thresholds for defining transmission clusters (Supplementary Text S1).

**Results**

*Patient demographics*

A total of 354 patients were included in the study and the clinical and demographic characteristics of the study population are shown in Table 1.

*Species distribution*

996 isolates from 354 patients from 9 MAC species were sequenced (Table 2). Three species accounted for 926/996 (93.0%) of the MAC isolates; *M. avium (M. avium* subsp. *avium* and *M. avium* subsp. *hominissuis)*, *M. chimaera* and *M. intracellulare.* Most patients were infected with only a single species during the collection period. However, 45 of the 354 patients (12.7%) were infected with two or more species (figure 1a). In this group, most of the isolates collected were typically from a single species, with other species observed more infrequently (figure 1b). Subsequent analyses in this study will focus on the three predominant species in the dataset: *M. intracellulare, M. avium (M. avium* subsp. *avium* and *M. avium* subsp. *hominissuis)*, and *M. chimaera*.

*M. intracellulare*

162 genomes from 37 patients were identified as *M. intracellulare* (figure 2a). Eleven patients had CF, seventeen had BX, and seven had other lung conditions (COPD n=3; ILD n=3; asthma n=1; congenital pulmonary airway malformation [CPAM] n=1). Disease metadata were missing for two patients. Genomic clustering with fastBAPs identified nine lineages with three of these having more than ten genomes (figure 2a). Following remapping to local references for the three largest lineages, three putative transmission clusters were identified with the largest, Mi_FB3_1, composed of 16 patients (figure 2b; supplementary table 2). Of these 16 patients, eight had BX, seven CF and one ILD. During the sampling period, four patients were treated with antibiotics, with successful outcomes in three patients. 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 [23] and formed part of a clade containing isolates from the Mi_FB3_1 cluster (figure 2c).

*M. avium*

405 sequenced isolates collected from 176 patients were identified as *M. avium*. Sequence data for all *M. avium* isolates were mapped to a single reference (*M. avium* subsp. *avium* 104; NC008595.1) and a phylogenetic tree constructed with *M. avium* subsp. *paratuberculosis* (DRR263663) as an outgroup (supplementary figure 2). The structure of this phylogeny showed that there were two major clades which corresponded to the two subspecies *M. avium* subsp. *avium* (n=207; MAA) and *M. avium* subsp. *hominissuis* (n=198; MAH). Each of these subspecies were analysed separately.

*M. avium* subsp. *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 3a). Clustering of the 207 MAA genomes identified 18 fastBAPS lineages (figure 3a). Remapping of the four largest lineages 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).

Twenty one fastBAPS lineages were defined in the global phylogeny (figure 3c). Examination of the distribution of pairwise SNP distances for each lineage containing at least ten genomes revealed that there were two lineages containing RBH isolates with lower median pairwise SNP distances (MAA_FB10 and MAA_FB15; figure 3c; supplementary figure 3b). The 23 genomes in FB10 were collected in the UK between 2013 and 2016, with most coming from the National Mycobacterial Reference Service which characterises mycobacterial cultures from across the Midlands and North of England (figure 5a) [23]. 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.* [23]. The majority (35/46) of the isolates forming MAA_FB15 were collected in the USA with most from a study investigating *M. avium* in the community and household water in Philadelphia (figure 5b) [24]. The pairwise distances between two of the RBH isolates and other UK isolates and isolates from the USA in MAA_FB15 were under our threshold, suggesting potential cross-Atlantic transmission (supplementary figure 4a).

*M. avium* subsp. *hominissuis* (MAH)

198 genomes from 106 patients were characterised as MAH (figure 4a). 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 lineages and seven putative transmission clusters containing between two and 16 patients (supplementary table 4). The largest cluster of 16 isolates is shown in figure 4b 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 lineages in the global collection, MAH_FB6 and MAH_FB11 had lower median pairwise SNP distances and contained RBH isolates (figure 4c; supplementary figure 3c). Both lineages were composed completely or near-completely of isolates collected in the UK (figure 5c-d). 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 have formed a transmission cluster that included both isolates from this study and isolates from other UK studies (supplementary figure 4b) [23,25]. Two small transmission clusters, containing RBH and other UK isolates, were also identified in FB11.

*M. chimaera*

359 sequenced isolates collected from 155 patients were 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 lineage, Mc_FB3, corresponding to the previously characterised Group 1 accounted for 332/359 (92.5%) of the sequenced isolates with most of the remaining isolates (25/359; 7.0%) forming a second main lineage Mc_FB4 corresponding to Group 2 (figure 6a). Thirteen putative transmission clusters containing two to 106 patients were identified (supplementary table 5). Twelve potential epidemiological links between 15 different patients were found in transmission cluster Mc_FB3_1, the largest transmission cluster of 106 patients (figure 6b; supplementary table 5; supplementary table 6). 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.

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 (Mc_FB6 [Group 1] figure 6c). Isolates from the two clusters with more than ten representatives (Mc_FB5 [Group 2] and Mc_FB6) were remapped to local references and new phylogenetic trees constructed (figure 7a [Mc_FB5] and figure 7b [Mc_FB6]). Most isolates in Mc_FB5 were collected from patients with respiratory conditions or 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 7b). Within Mc_FB6 a more deeply rooting (ancestral) diverse clade contained mainly human isolates and the larger shallower clade rooting within this contained most of the HCU isolates interspersed with human isolates (figure 7b-c). Due to the large number of isolates assigned to Mc_FB6 (n = 765), further genomic lineage assignment was performed using fastBAPS to identify five sub-lineages within Mc_FB6 (figure 7b-c). New mapping was performed for the three lineages with more than ten isolates (figure 8a-c). Two of these lineages, Mc_FB6_FB2 and Mc_FB6_FB4, were completely or nearly completely composed of patient isolates (figure 8b-c). Most of the isolates collected from HCUs or water supplies made up the bulk of Mc_FB6_FB1, though human isolates were distributed throughout the tree (figure 8a). Putative transmission clusters were calculated for all five lineages using 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 Mc_FB6_FB1, comprising 258 HCU isolates, 230 patient isolates and a single isolate from a water supply (supplementary table 7).

**Discussion**

The aims of this study were to characterise the population structure of MAC isolates collected in a London hospital and identify potential transmission between CF and non-CF patients. During the sampling period we identified 45 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 at different times. This has been observed previously in other longitudinal samples of patients with MAC lung disease [26], and suggests that single colonies from sputum swabs may be underrepresenting the underlying diversity of MAC species in patients. Therefore, to better understand the ecology and

transmission of MAC species in the lung, a deep sequencing approach in which plate sweeps from swabs or MGIT cultures are utilised should be considered for future studies.

Previous studies have used thresholds of 15-25 SNPs to identify putative transmission clusters in NTM species [14,19,24]. Rather than using a previously applied or arbitrary threshold, we used a method developed to define SNP thresholds in *Staphylococcus aureus* [27]. This allowed us to calculate a threshold for each species using the longitudinal isolates collected in the RBH. The thresholds calculated for *M. intracellulare* and MAH of 16 SNPs were similar to those previously applied [24]. 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 in patients with both MAA and *M. chimaera*.

Using these SNP thresholds, we were able to identify putative transmission clusters amongst all four of the MAC species/subspecies investigated. 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 introductions into the human population followed by onward transmission. Our analysis of *M. chimaera*, however, identified a single large transmission cluster which linked 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. The tight infection control surrounding patients with CF, which prioritises preventing cross-infection through hygiene and segregation, means that they are unlikely to have interacted with patients with other lung conditions whilst in hospital. Outpatient clinics are organised so that patients are seen individually and, during inpatient care, CF patients are given individual rooms with *en-suite* facilities. 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 there has been considerable debate whether NTM transmission between patients is occurring, even when transmission is strongly supported by genome-based analysis [14,19,28]. The limited evidence 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 [19]. Whilst there is certainly strong evidence for the same lineages being isolated from water supplies and patients in the same location [24], the absence of a single national water supply in even 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 equally worthy of consideration. This would explain the presence of local and long-distance 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

collecting isolates from the environment as well as from the general population, perhaps focussing on smokers where there is some evidence that they may provide a reservoir for NTM species and other opportunistic pathogens [20].

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* lineages that were known to be associated with HCUs, we were unable to identify large globally circulating lineages 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 lineage containing RBH isolates amongst the two *M. avium* subspecies that included isolates from outside the UK. This cluster comprised most of the isolates in MAA_FB15 (figure 5b). Given that most *M. avium* genomes have been collected in the UK or USA, a more comprehensive global sampling strategy could potentially reveal the existence of additional globally circulating clones.

When we calculated global transmission clusters for *M. chimaera* using a pairwise SNP threshold of 30 SNPs (Supplementary table 7), we found that most clusters either consisted of isolates from across Europe or included isolates from other parts of the world such as Australia. Notably, we identified a single large transmission cluster containing 489 isolates from 12 countries that included most of the HCU-associated isolates. Previous work investigating the global population structure of *M. chimaera* suggested that the majority of *M. chimaera* transmission globally was associated with point source contamination of HCUs during manufacturing which then led to direct infection of patients undergoing cardiac surgery [11]. Our study included isolates collected from HCUs and cardiac patients as well

as patients with lung conditions such as CF and BX. This enabled us to investigate the global population structure of *M. chimaera* in more detail than has previously been possible [11,29]. Strikingly, we discovered clear evidence for distinct lineages within the global phylogeny that were almost exclusively associated with patient isolates, indicating considerable onward transmission. In addition, 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, and that a similar derivation has happened on at least one other occasion.

Potential limitations of this study are the lack of environmental samples as well as the distinct bias with respect to country of isolation in the contextual dataset. Given the lack of direct patient-to-patient contact and epidemiological links, inclusion of samples collected from wards and hospital water supplies could potentially have revealed additional vectors of transmission. Assembling useful contextual datasets is reliant on what is available in the public domain and the geographical bias towards the USA and UK in these datasets is not confined to this study.

Our results showed that polymicrobial infections by different MAC species are not uncommon and that transmission is occurring between CF and non-CF patients even in the presence of strict infection controls. This suggests that all of the species and subspecies within the MAC, like *M. abscessus,* are capable of generating lineages that can sustain transmission within the human population, albeit to different extents. We also provided evidence that the lineage responsible for the HCU-vectored *M. chimaera* outbreak was

derived from an existing lineage which was already circulating amongst patients with pre-existing lung conditions, indicating that this is also true of *M. chimaera*, even after excluding the extensive HCU-vectored outbreak event. To better understand the transmission dynamics and to prevent continuing circulation of MAC and other NTM species, future studies should include sampling of environmental reservoirs and potential asymptomatic carriers.

**Acknowledgements**

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. Kartik Kumar's Clinical PhD was supported by a Lee Family Endowment.

**Conflicts of interest**

Julian Parkhill received consulting fees from and holds stock in Next Gen Diagnostics Llc. Michael Loebinger received consulting fees from Savara, Astra Zeneca, Insmed, Grifols, Zambon, Armarta and 30T. Additionally, he received honoraria from Insmed and Grifols. No other authors have potential competing interests.

**Tables**

Table 1. Patient clinical and demographic data

| | All patients (n = 354) |
|---|---|
| Gender, male | 174 (49.2%) |
| Median age at first culture, years (range) | 56 (5-93) |

| | |
|---|---|
| Median BMI at first culture (range) | 22.5 (13.4-43.4) |
| Current or former smokers | 28 (10.7%) |
| Underlying pulmonary disease | |
| None | 12 (3.4%) |
| Non-CF bronchiectasis (BX) | 147 (41.5%) |
| Cystic fibrosis (CF) | 87 (24.6%) |
| Chronic obstructive pulmonary disease (COPD) | 53 (15.0%) |
| Asthma | 32 (9.0%) |
| Allergic bronchopulmonary aspergillosis (ABPA) | 19 (5.4%) |
| Interstitial lung disease (ILD) | 17 (4.8%) |
| Other (pleural thickening, sarcoidosis) | 7 (2.0%) |
| | |
| Antibiotic treatment during study period | 55 (15.5%) |
| | |
| Clinical data unavailable | 20 (5.6%) |

Table 2. *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 |
| Total | 996 |

**References**


1. Griffith DE, Aksamit T, Brown-Elliott BA *et al*. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. Am J Respir Crit Care Med 2007; 175(4): 367–416.

2. Chan ED, Bai X, Kartalija M *et al*. Host immune response to rapidly growing mycobacteria, an emerging cause of chronic lung disease. Am J Respir Cell Mol Biol 2010; 43(4): 387–93.

3. Iseman MD, Marras TK. The importance of nontuberculous mycobacterial lung disease. Am J Respir Crit Care Med 2008; 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; 13(12): 2143–50.

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

6. 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; 15(7):817–26.

7. Hoefsloot W, van Ingen J, Andrejak C *et a*l. The geographic diversity of nontuberculous mycobacteria isolated from pulmonary samples: an NTM-NET collaborative study. Eur Respir J 2013; 42(6): 1604–13.

8. Olivier KN, Weber DJ, Wallace RJ *et al*. Nontuberculous mycobacteria. I: multicenter prevalence study in cystic fibrosis. Am J Respir Crit Care Med 2003; 167(6): 828–34.

9. Achermann Y, Rössle M, Hoffmann M *et al*. Prosthetic valve endocarditis and bloodstream infection due to *Mycobacterium chimaera*. J Clin Microbiol 2013; 51(6): 1769–73.

10. Sax H, Bloemberg G, Hasse B *et al*. Prolonged Outbreak of *Mycobacterium chimaera* Infection After Open-Chest Heart Surgery. Clin Infect Dis Off Publ Infect Dis Soc Am 2015; 61(1): 67–75.

11. van Ingen J, Kohl TA, Kranzer K *et al*. Global outbreak of severe *Mycobacterium chimaera* disease after cardiac surgery: a molecular epidemiological study. Lancet Infect Dis 2017; 17(10): 1033–41.

12. Falkinham JO. Nontuberculous mycobacteria in the environment. Clin Chest Med 2002; 23(3): 529–51.

13. Falkinham JO. Nontuberculous mycobacteria from household plumbing of patients with nontuberculous mycobacteria disease. Emerg Infect Dis 2011; 17(3): 419–24.

14. Bryant JM, Grogono DM, Greaves D *et al*. Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. Lancet 2013; 381(9877): 1551–60.

15. Bryant JM, Grogono DM, Rodriguez-Rincon D *et al*. Emergence and spread of a human-transmissible multidrug-resistant nontuberculous mycobacterium. Science 2016; 354(6313): 751–7.

16. Everall I, Nogueira CL, Bryant JM *et al*. Genomic epidemiology of a national outbreak of post-surgical *Mycobacterium abscessus* wound infections in Brazil. Microb Genomics 2017; 3(5): e000111.

17. Aitken ML, Limaye A, Pottinger P *et al*. Respiratory outbreak of *Mycobacterium abscessus* subspecies massiliense in a lung transplant and cystic fibrosis center. Am J Respir Crit Care Med 2012; 185(2): 231–2.

18. Davidson RM, Hasan NA, Epperson LE *et al*. Population Genomics of *Mycobacterium abscessus* from U.S. Cystic Fibrosis Care Centers. Ann Am Thorac Soc 2021; 18(12): 1960–9.

19. Lipworth S, Hough N, Weston N *et al*. Epidemiology of *Mycobacterium abscessus* in England: an observational study. Lancet Microbe; 2(10): e498–507.

20. Ruis C, Bryant JM, Bell SC *et al*. Dissemination of *Mycobacterium abscessus* via global transmission networks. Nat Microbiol 2021; 6(10): 1279–88.

21. Hasan NA, Davidson RM, Epperson LE *et al*. Population Genomics and Inference of *Mycobacterium avium* Complex Clusters in Cystic Fibrosis Care Centers, United States. Emerg Infect Dis 2021; 27(11): 2836–46.

22. Harris SR, Feil EJ, Holden MTG *et al*. Evolution of MRSA during hospital transmission and intercontinental spread. Science 2010; 27(5964): 469–74.

23. Quan TP, Bawa Z, Foster D *et al*. Evaluation of Whole-Genome Sequencing for Mycobacterial Species Identification and Drug Susceptibility Testing in a Clinical Setting: a Large-Scale Prospective Assessment of Performance against Line Probe Assays and Phenotyping. J Clin Microbiol 2018; 56(2): e01480-17.

24. Lande L, Alexander DC, Wallace RJ *et al*. *Mycobacterium avium* in Community and Household Water, Suburban Philadelphia, Pennsylvania, USA, 2010-2012. Emerg Infect Dis 2019; 25(3): 473–81.

25. Votintseva AA, Pankhurst LJ, Anson LW *et al*. Mycobacterial DNA extraction for whole-genome sequencing from early positive liquid (MGIT) cultures. J Clin Microbiol 2015; 53(4): 1137–43.

26. Operario DJ, Pholwat S, Koeppel AF *et al*. *Mycobacterium avium* Complex Diversity within Lung Disease, as Revealed by Whole-Genome Sequencing. Am J Respir Crit Care Med 2019; 200(3): 393–6.

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

28. Bronson RA, Gupta C, Manson AL *et al*. Global phylogenomic analyses of *Mycobacterium abscessus* provide context for non cystic fibrosis infections and the evolution of antibiotic resistance. Nat Commun 2021; 12(1): 5145.

29. Schreiber PW, Kohl TA, Kuster SP *et al*. 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; 27(11): 1613–20.

**Figure Legends**

**Figure 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.

**Figure 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 salmon boxes. Scale bars are shown in SNPs per site.

**Figure 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 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 avium* subsp. *avium* isolates (n = 344). 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 salmon boxes. Scale bars are shown in SNPs per site.

**Figure 4. Population structure and transmission of *Mycobacterium avium* subsp. *hominissuis*.** A) Midpoint-rooted maximum likelihood phylogenetic tree for *Mycobacterium avium* subsp. *hominissuis* isolates collected at the Royal Brompton Hospital (n = 198). 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. *hominissuis* MAH_FB8_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 avium* subsp. *hominissuis* isolates (n = 236). 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 salmon boxes. 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* 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

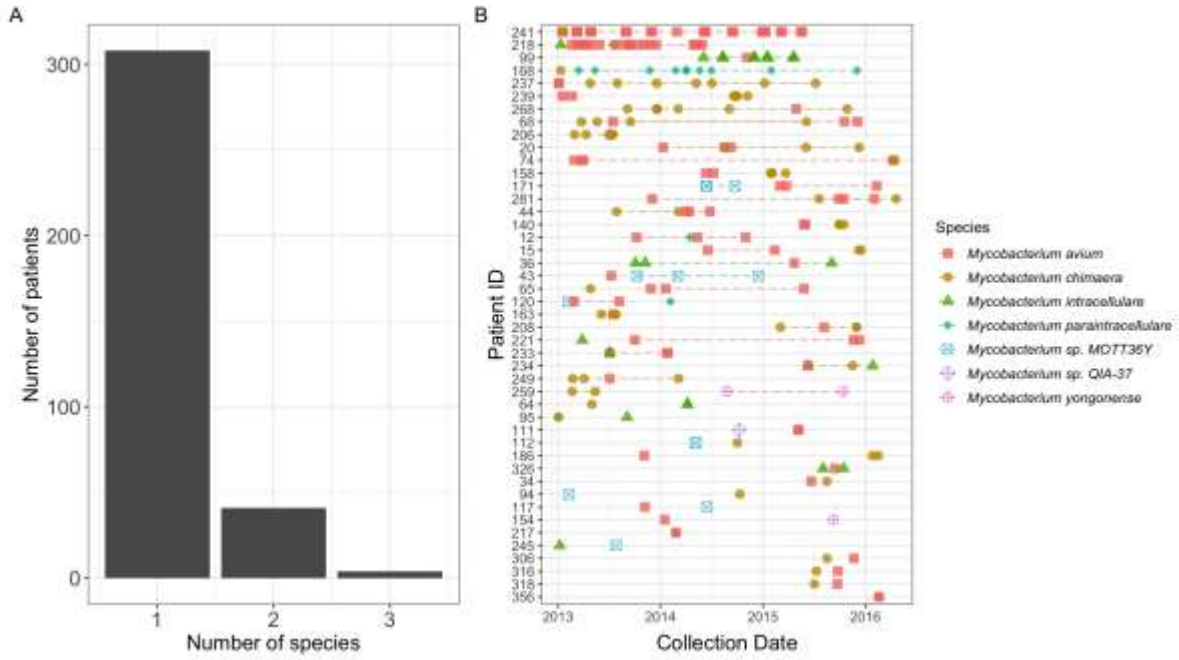
per site. The source of isolate, country of collection and study are shown as datastrips to the right of the phylogenies. Isolates from the Royal Brompton Hospital are highlighted in red. Branch lengths are shown in SNPs per site.

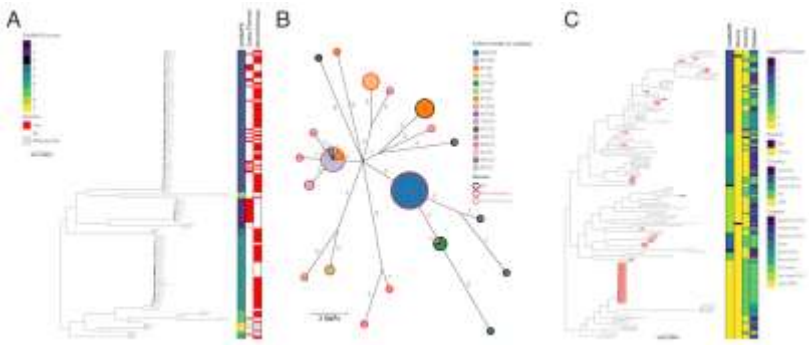
**Figure 6. Population structure and transmission of *Mycobacterium chimaera.*** A) Midpoint-rooted maximum likelihood phylogenetic tree for *Mycobacterium chimaera* isolates collected at the Royal Brompton Hospital (n = 359). The taxa are clustered according to their sequence similarities. The lengths 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 chimaera* Mc_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 according to patient disease status and labelled by patient number. The edges represent the pairwise SNP distance between the isolate(s); C) Rooted maximum likelihood phylogenetic tree for global *Mycobacterium chimaera* isolates (n = 826). 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 salmon boxes. Scale bars are shown in SNPs per site.

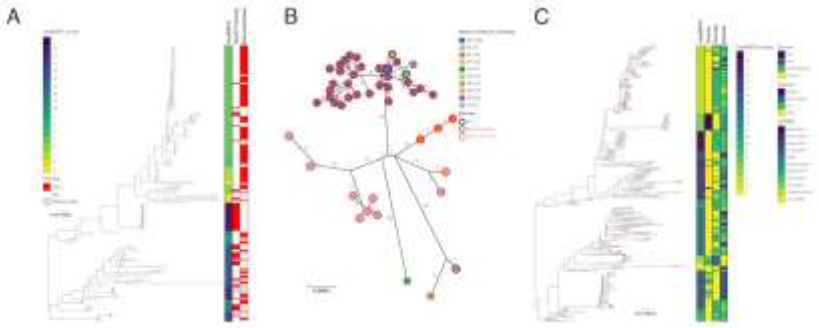
**Figure 7. *Mycobacterium chimaera* global fastBAPS cluster phylogenies.** Rooted maximum likelihood phylogenetic trees for A) *Mycobacterium chimaera* FB5 (n = 45); B) *Mycobacterium chimaera* FB6 (n = 765) and C) *Mycobacterium chimaera* FB6 with major

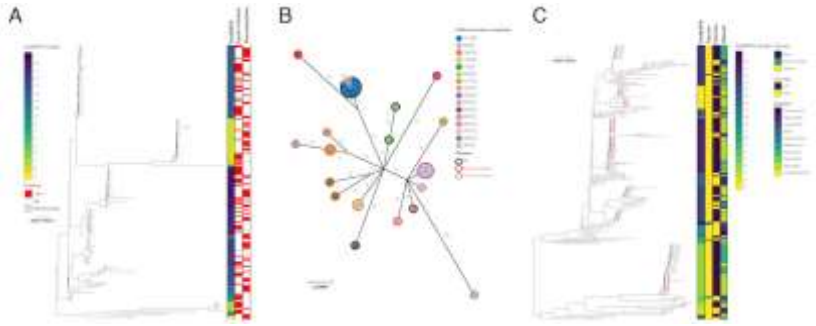
clades collapsed; pie charts are shown for each major clade showing the distribution of sources for that clade. The taxa are clustered according to their sequence similarities. The lengths of the branches are scaled in nucleotide substitutions per site. The source of isolate, country of collection and 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 for *Mycobacterium chimaera* FB5.

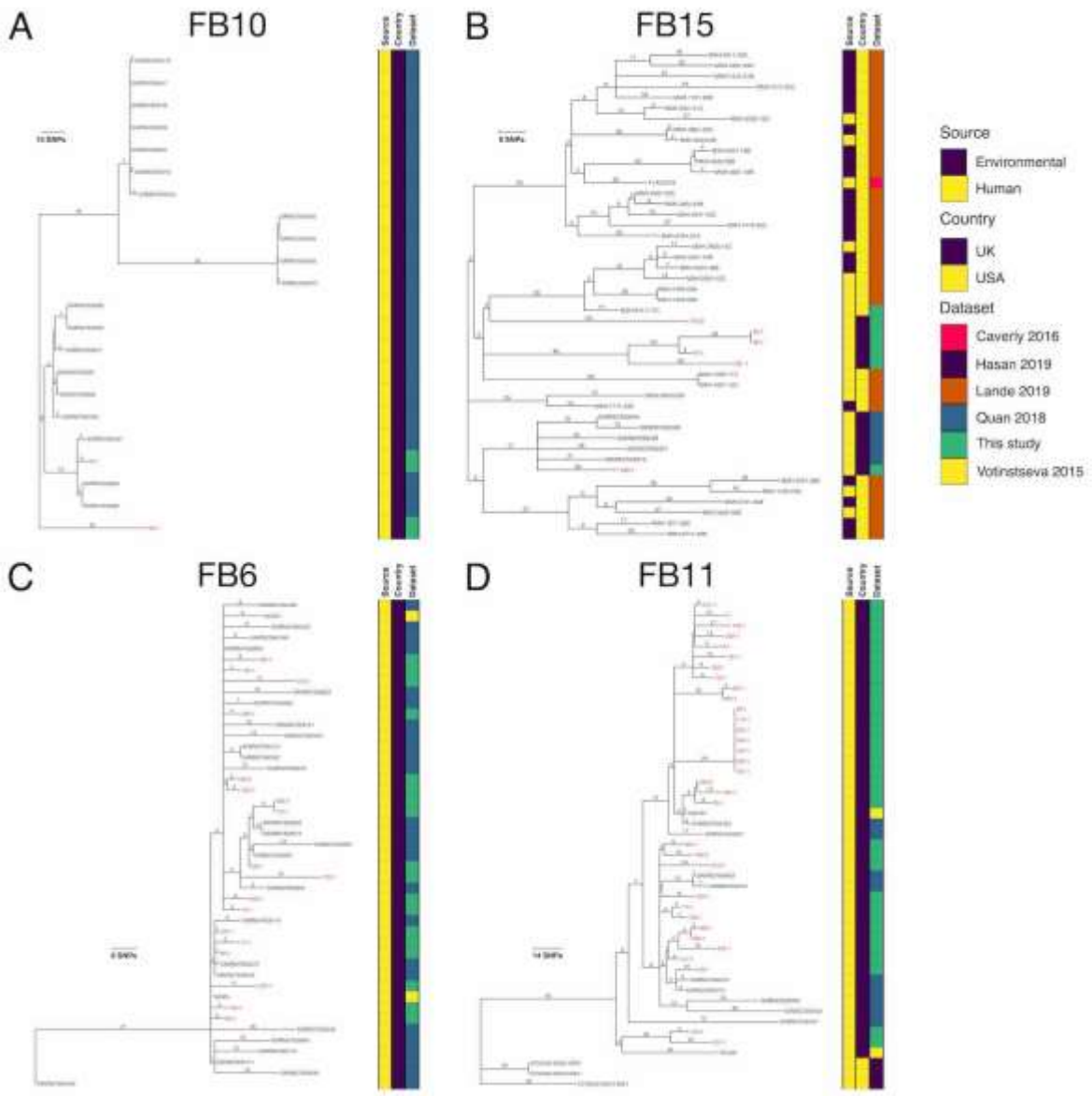
**Figure 8: *Mycobacterium chimaera* fastBAPS cluster FB6 phylogenies.** Maximum likelihood phylogenetic trees for A) FB6_FB1 (n = 540); B) FB6_FB2 (n = 186); and C) FB6_FB4 (n = 34). The taxa are clustered according to their sequence similarities. The lengths of the branches are scaled in nucleotide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

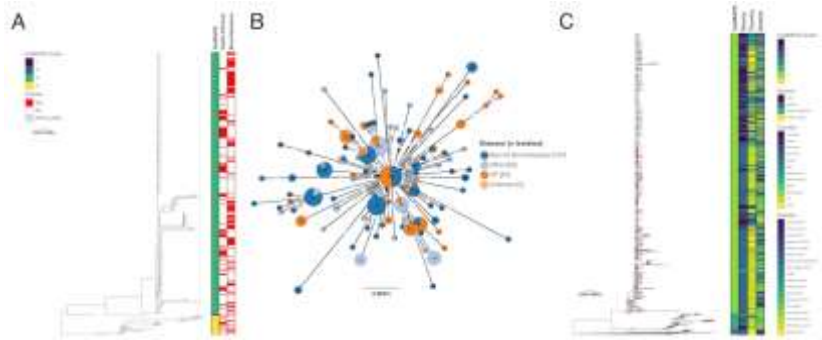

A
300
200
100
0
Number of patients
1
2
3
Number of species
B
Patient ID
2013
2014
2015
2016
Collection Date
Species
Mycobacterium avium
Mycobacterium chimaera
Mycobacterium intracellulare
Mycobacterium paraintracellulare
Mycobacterium sp. MDTT35Y
Mycobacterium sp. QIA-37
Mycobacterium yongonense

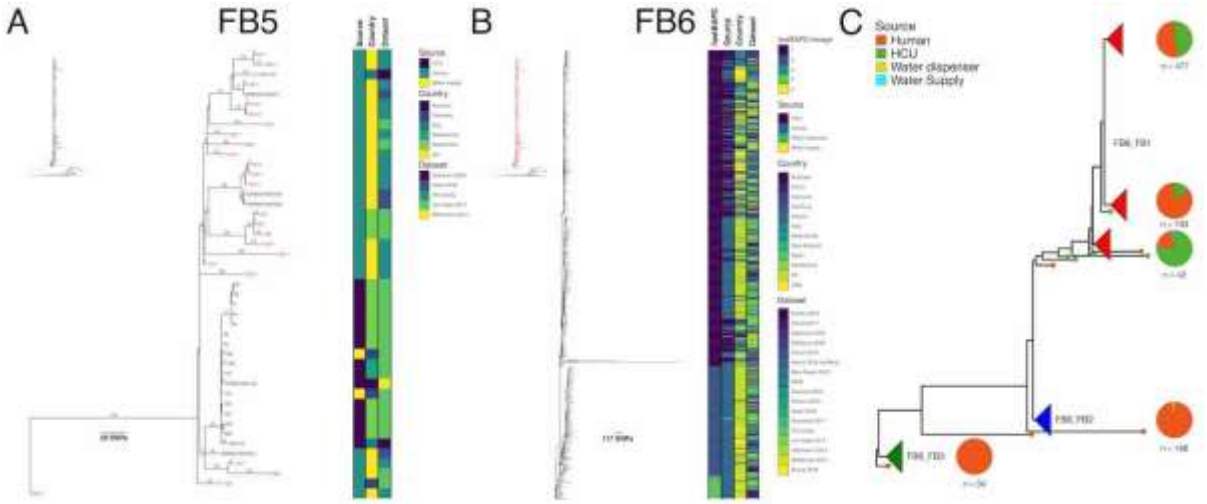

A
B
C

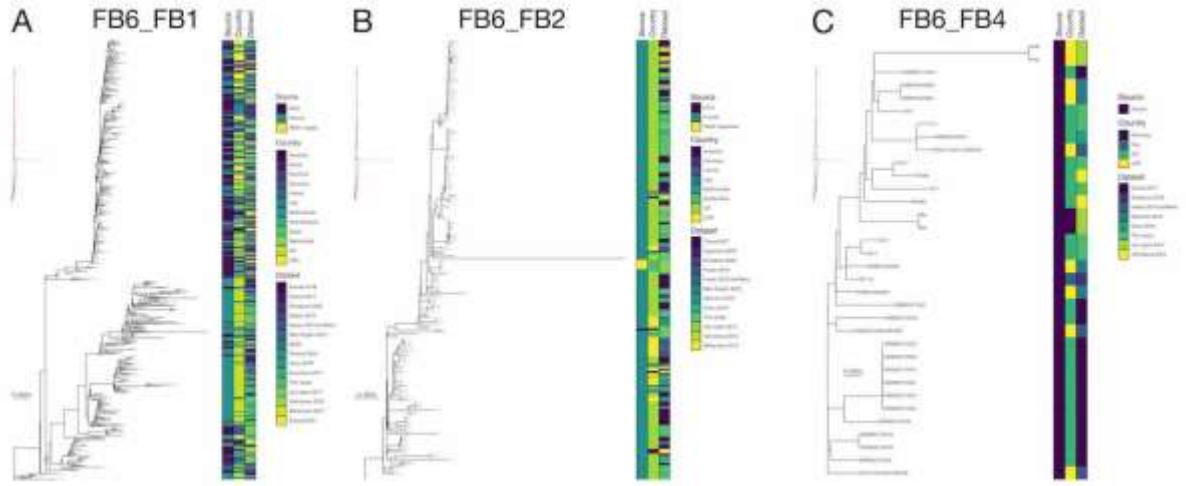

A
B
C


A
B
C


A
FB10
B
FB15
C
FB6
D
FB11
Source
Country
Dataset
Source
Environmental
Human
Country
UK
USA
Dataset
Caverly 2016
Hasan 2019
Lande 2019
Quan 2018
This study
Votinstseva 2015


A
B
C


A
FB5
B
FB6
C
Source
Human
HCU
Water dispenser
Water Supply


A
FB6_FB1
B
FB6_FB2
C
FB6_FB4

**Supplementary Text S1 - Detailed 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).

*Culturing, DNA extraction and sequencing*

Sputum samples were grown in BBL MGIT media (Becton Dickinson; BD) in a BACTEC MGIT 960 (BD) culture system until the system indicated likely Mycobacterial growth. NTM species confirmation was performed using the GenoTypeMycobacterium CM VER 2.0 (Hain Lifescience) System. Confirmed MAC cultures were regrown from bead stock cultures in BBL MGIT media (BD) in a BACTEC MGIT 960 (BD) culture system 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.

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.

*Sequence QC, mapping and phylogenetics*

Basic quality control metrics for the raw sequence data were generated using FastQC v0.11.9 [1]. Sequence reads with similarity to *Mycobacterium* species were identified using Kraken v0.10.6 [2] and Bracken v1.0 [3]. Samples with < 70% reads mapping to a *Mycobacterium* species were excluded from further analyses (n = 116). Seven isolates not belonging to the MAC (*M. abscessus, M. chelonae, M. simiae)* were removed from the dataset. Sequence reads for each species were trimmed using Trimmomatic v0.33 [4] and mapped to appropriate references (supplementary table 1) using BWA mem v0.7.17 (minimum and maximum insert sizes of 50 bp and 1000 bp respectively) [5]. 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 [6]. 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 [7]. 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) [8].

For higher-resolution phylogenies within fastBAPS lineages, recombinant regions were identified and removed from alignments using GUBBINS [9] and new phylogenetic trees were constructed as described above. Pairwise SNP distances were calculated for all pairs of isolates using pairsnp [10].

*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 3) [11–31]. Sequence data were downloaded from the European Nucleotide Archive (ENA) and trimmed;

Sample QC, mapping and phylogenetic tree construction were performed as detailed above. Only the first isolate from each patient was included from the RBH isolates for each species/subspecies.

*Genome assemblies*

A previously published pipeline was used to produce annotated assemblies [32]. Briefly, sequence reads were assembled with spades v 3.10.10 [33] and assemblies were improved by first scaffolding the assembled contigs using SSPACE v2.0 [34] and filling the sequence gaps with GapFiller v1.11 [35].

*Transmission and epidemiological linkage*

Genomic lineages were identified using fastBAPS [36] and new alignments were created for lineages ≥ 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 lineage, pairwise SNP distances for all isolates for each species in the RBH datasets were calculated. Using a previously described method [37], the transmission threshold for each species, regardless of lineage, was calculated by taking the $95^{th}$ 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. avium* subsp. *avium* FB14 lineages (Fig A in S1 Data), 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 [38,39] and pairwise SNP thresholds of 16 (*M. intracellulare* and *M. avium* subsp. *hominissuis*), 30 (*M. chimaera*) and

58 SNPs (*M. avium* subsp. *avium*) were used to calculate putative transmission clusters in each genomic lineage. Finally, in order to identify possible epidemiological links between patients infected within 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.

**References**


1. Babraham Bioinformatics. FastQC A Quality Control tool for High Throughput Sequence Data. www.bioinformatics.babraham.ac.uk/projects/fastqc/ Date last updated: December 14 2021. Date last accessed: December 14 2021.

2. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 2014; 15: R46.

3. Lu J, Breitwieser FP, Thielen P *et al*. Bracken: estimating species abundance in metagenomics data. PeerJ Comput Sci 2017; 3: e104.

4. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30: 2114–20.

5. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25: 1754–60.

6. Harris SR, Feil EJ, Holden MTG *et al*. Evolution of MRSA during hospital transmission and intercontinental spread. Science 2010; 327: 469–74.

7. Page AJ, Taylor B, Delaney AJ *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genomics 2016; 2: e000056.

8. Nguyen L-T, Schmidt HA, von Haeseler A, *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 2015; 32: 268–74.

9. Croucher NJ, Page AJ, Connor TR *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res 2015; 43: e15.

10. Tonkin-Hill G. pairsnp. https://github.com/gtonkinhill/pairsnp. Date last updated: December 2021. Date last accessed: December 14 2021 .

11. van Ingen J, Kohl TA, Kranzer K *et al.* Global outbreak of severe *Mycobacterium chimaera* disease after cardiac surgery: a molecular epidemiological study. Lancet Infect Dis 2017; 17: 1033–41.

12. Caverly LJ, Spilker T, LiPuma JJ. Complete Genome Sequence of *Mycobacterium abscessus* subsp. *bolletii*. Genome Announc 2016; 4: e00543-16.

13. Hasan NA, Davidson RM, Epperson LE *et al.* Population Genomics of Nontuberculous Mycobacteria Recovered from United States Cystic Fibrosis Patients. bioRxiv 2019; 663559.

14. Bryant JM, Thibault VC, Smith DGE *et al*. Phylogenomic exploration of the relationships between strains of *Mycobacterium avium* subspecies *paratuberculosis*. BMC Genomics 2016; 17: 79.

15. Lande L, Alexander DC, Wallace RJ *et al*. *Mycobacterium avium* in Community and Household Water, Suburban Philadelphia, Pennsylvania, USA, 2010-2012. Emerg Infect Dis 2019; 25: 473–81.

16. Yoon J-K, Kim TS, Kim J-I *et al*. Whole genome sequencing of Nontuberculous Mycobacterium (NTM) isolates from sputum specimens of co-habiting patients with NTM pulmonary disease and NTM isolates from their environment. BMC Genomics 2020; 21: 322.

17. Votintseva AA, Pankhurst LJ, Anson LW *et al.* Mycobacterial DNA extraction for whole-genome sequencing from early positive liquid (MGIT) cultures. J Clin Microbiol 2015; 53: 1137–43.

18. Pfeiffer W, Braun J, Burchell J *et al*. Whole-genome analysis of mycobacteria from birds at the San Diego Zoo. PloS One 2017; 12: e0173464.

19. Quan TP, Bawa Z, Foster D *et al*. Evaluation of Whole-Genome Sequencing for Mycobacterial Species Identification and Drug Susceptibility Testing in a Clinical Setting: a Large-Scale Prospective Assessment of Performance against Line Probe Assays and Phenotyping. J Clin Microbiol 2018; 56: e01480-17.

20. Operario DJ, Pholwat S, Koeppel AF *et al*. *Mycobacterium avium* Complex Diversity within Lung Disease, as Revealed by Whole-Genome Sequencing. Am J Respir Crit Care Med 2019; 200: 393–6.

21. Rivas AE, Hollinger C, Oehler DA *et al*. DIAGNOSIS AND MANAGEMENT OF MYCOBACTERIOSIS IN A COLONY OF LITTLE PENGUINS (EUDYPTULA MINOR). J Zoo Wildl Med Off Publ Am Assoc Zoo Vet 2019; 50: 427–36.

22. Chand M, Lamagni T, Kranzer K *et al*. Insidious Risk of Severe *Mycobacterium chimaera* Infection in Cardiac Surgery Patients. Clin Infect Dis Off Publ Infect Dis Soc Am 2017; 64: 335–42.

23. Pankhurst LJ, Del Ojo Elias C, Votintseva AA *et al.* Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med 2016; 4: 49–58.

24. Epperson LE, Timke M, Hasan NA *et al*. Evaluation of a Novel MALDI Biotyper Algorithm to Distinguish *Mycobacterium intracellulare* From *Mycobacterium chimaera*. Front Microbiol 2018; 9: 3140.

25. Acosta F, Pérez-Lago L, Ruiz Serrano MJ *et al*. Fast update of undetected *Mycobacterium chimaera* infections to reveal unsuspected cases. J Hosp Infect 2018; 100: 451–5.

26. Ghodousi A, Borroni E, Peracchi M *et al*. Genomic analysis of cardiac surgery-associated *Mycobacterium chimaera* infections in Italy. PloS One 2020; 15: e0239273.

27. Mac Aogáin M, Roycroft E, Raftery P *et al*. Draft Genome Sequences of Three *Mycobacterium chimaera* Respiratory Isolates. Genome Announc 2015; 3: e01409-15.

28. Perkins KM, Lawsin A, Hasan NA *et al*. Notes from the Field: *Mycobacterium chimaera* Contamination of Heater-Cooler Devices Used in Cardiac Surgery - United States. MMWR Morb Mortal Wkly Rep 2016; 65: 1117–8.

29. Svensson E, Jensen ET, Rasmussen EM *et al*. *Mycobacterium chimaera* in Heater-Cooler Units in Denmark Related to Isolates from the United States and United Kingdom. Emerg Infect Dis 2017; 23: 507–9.

30. Williamson D, Howden B, Stinear T. *Mycobacterium chimaera* Spread from Heating and Cooling Units in Heart Surgery. N Engl J Med 2017; 376: 600–2.

31. Zhang X, Lin J, Feng Y *et al*. Identification of *Mycobacterium chimaera* in heater-cooler units in China. Sci Rep 2018; 8: 7843.

32. Page AJ, De Silva N, Hunt M *et al.* Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. Microb Genomics 2016; 2: e000083.

33. Bankevich A, Nurk S, Antipov D *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol J Comput Mol Cell Biol 2012; 19:455–77.

34. Boetzer M, Henkel CV, Jansen HJ *et al*. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 2011; 27: 578–9.

35. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol 2012; 13: R56.

36. Tonkin-Hill G, Lees JA, Bentley SD *et al*. Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Res 2019; 47: 5539–49.

37. Coll F, Raven KE, Knight GM *et al*. Definition of a genetic relatedness cutoff to exclude recent transmission of meticillin-resistant *Staphylococcus aureus*: a genomic epidemiology analysis. Lancet Microbe 2020; 1: e328–35.

38. R Core Team. R: A language and environment for statistical computing. 2013.

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

Supplementary Table 1. 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 |

Supplementary Table 2. *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 |

Supplementary Table 3. *M. avium* subsp. *avium* transmission clusters

| fastBAPS cluster | Transmission cluster | Patients (n) | Bronchiectasis (n) | Cystic Fibrosis (n) | Other lung conditio | No pre-existing lung | Patients undergoing |
|---|---|---|---|---|---|---|---|

| | | | | | ns (n) | conditio n (n) | treatment (n) |
|---|---|---|---|---|---|---|---|
| MAA_FB 5 | MAA_FB5_ 1 | 10 | 6 | 2 | 2 | 0 | 4 |
| MAA_FB 5 | MAA_FB5_ 4 | 2 | 1 | 1 | 0 | 0 | 0 |
| MAA_FB 5 | MAA_FB5_ 5 | 2 | 1 | 0 | 1 | 0 | 1 |
| MAA_FB 5 | MAA_FB5_ 12 | 2 | 1 | 1 | 0 | 0 | 1 |
| MAA_FB 7 | MAA_FB7_ 1 | 2 | 2 | 0 | 0 | 0 | 1 |
| MAA_FB 10 | MAA_FB10 _1 | 2 | 0 | 2 | 0 | 0 | 0 |
| MAA_FB 14 | MAA_FB14 _1 | 2 | 2 | 0 | 0 | 0 | 1 |

Supplementary Table 4. *M. avium* subsp. *hominissuis* transmission clusters

| fastBAPS cluster | Transmissio n cluster | Patien ts (n) | Bronchiecta sis (n) | Cystic Fibros is (n) | Other lung conditio ns (n) | No pre-existing lung conditio n (n) | Patients undergoi ng treatment (n) |
|---|---|---|---|---|---|---|---|
| MAH_FB | MAH_FB8_ | 16 | 7 | 4 | 3 | 1 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 8 | 1 | | | | | | |
| MAH_FB 10 | MAH_FB10 _1 | 4 | 2 | 2 | 0 | 0 | 1 |
| MAH_FB 11 | MAH_FB11 _1 | 6 | 3 | 1 | 2 | 0 | 0 |
| MAH_FB 12 | MAH_FB12 _2 | 4 | 1 | 2 | 1 | 0 | 2 |
| MAH_FB 14 | MAH_FB14 _3 | 4 | 2 | 2 | 0 | 0 | 0 |
| MAH_FB 14 | MAH_FB14 _7 | 2 | 0 | 2 | 0 | 0 | 0 |
| MAH_FB 14 | MAH_FB14 _9 | 7 | 3 | 2 | 1 | 0 | 1 |

Supplementary Table 5. *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_1 | 4 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 3 | | | | | | |
| Mc_FB3 | Mc_FB3_16 | 2 | 0 | 0 | 1 | 1 | 1 |
| Mc_FB3 | Mc_FB3_17 | 2 | 1 | 1 | 0 | 0 | 0 |
| Mc_FB3 | Mc_FB3_19 | 2 | 0 | 1 | 0 | 1 | 0 |
| Mc_FB3 | Mc_FB3_25 | 2 | 1 | 1 | 0 | 0 | 1 |
| Mc_FB3 | Mc_FB3_30 | 4 | 0 | 1 | 0 | 0 | 0 |
| Mc_FB4 | Mc_FB4_1 | 2 | 1 | 1 | 0 | 0 | 1 |
| Mc_FB4 | Mc_FB4_2 | 9 | 3 | 3 | 3 | 0 | 2 |
| Mc_FB4 | Mc_FB4_3 | 6 | 1 | 2 | 2 | 0 | 1 |

Supplementary Table 6. *Mycobacterium chimaera* epidemiological links (n =15 patients)

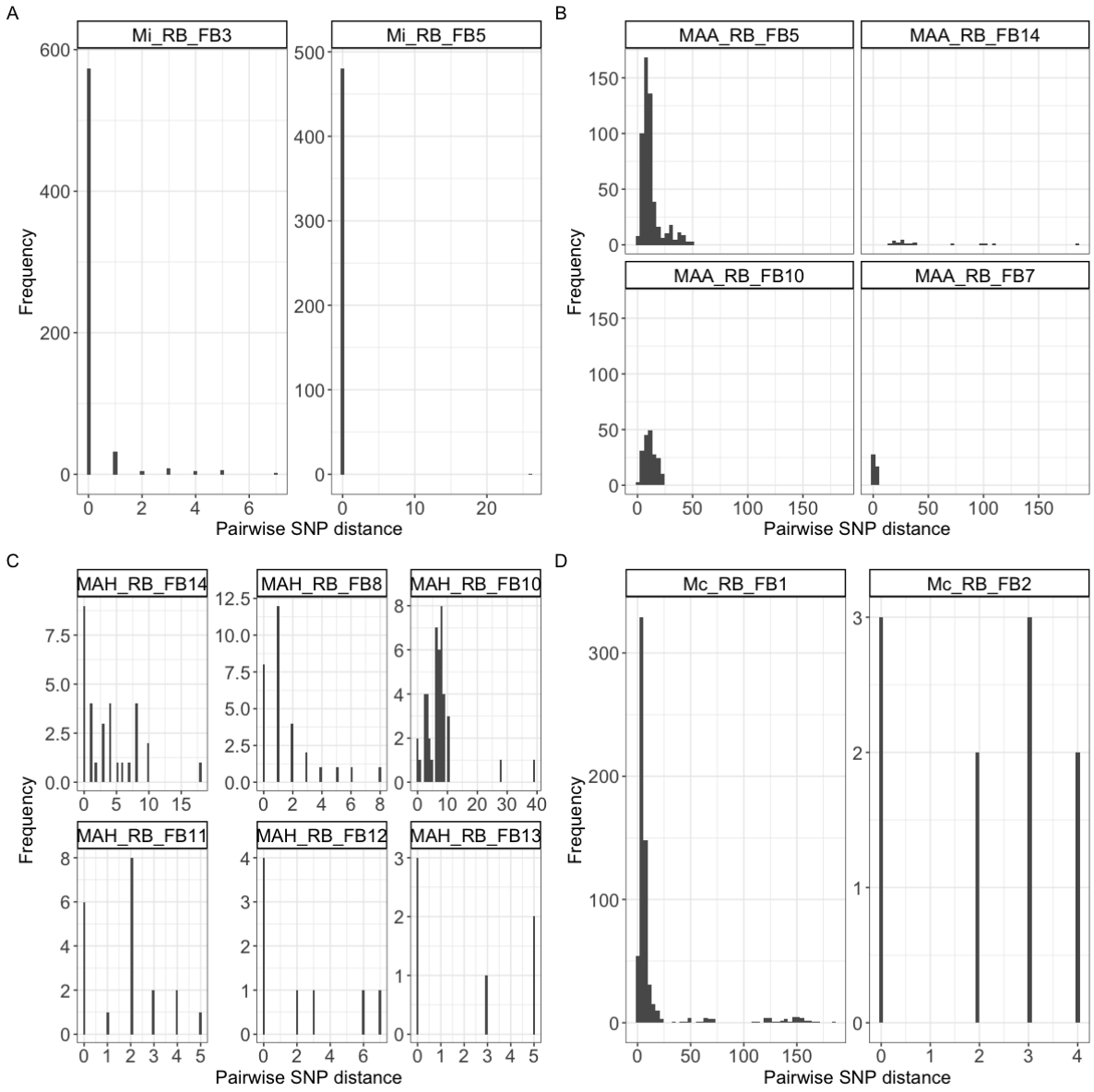
| Patients with epidemiological link | Hospital ward | Dates of overlapping stay |
|---|---|---|
| 24/218 | FOUL | 02/02/12-12/02/12 |
| 163/218 | FOUL | 13/02/12 |
| 175/218 | FOUL | 05/04/12-16/04/12 |
| 130/218 | LIND | 30/04/12 |
| 76/218 | FOUL | 06/06/12-11/06/12 |
| 32/163 | LIND | 29/06/12 |

| | | |
|---|---|---|
| 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 |

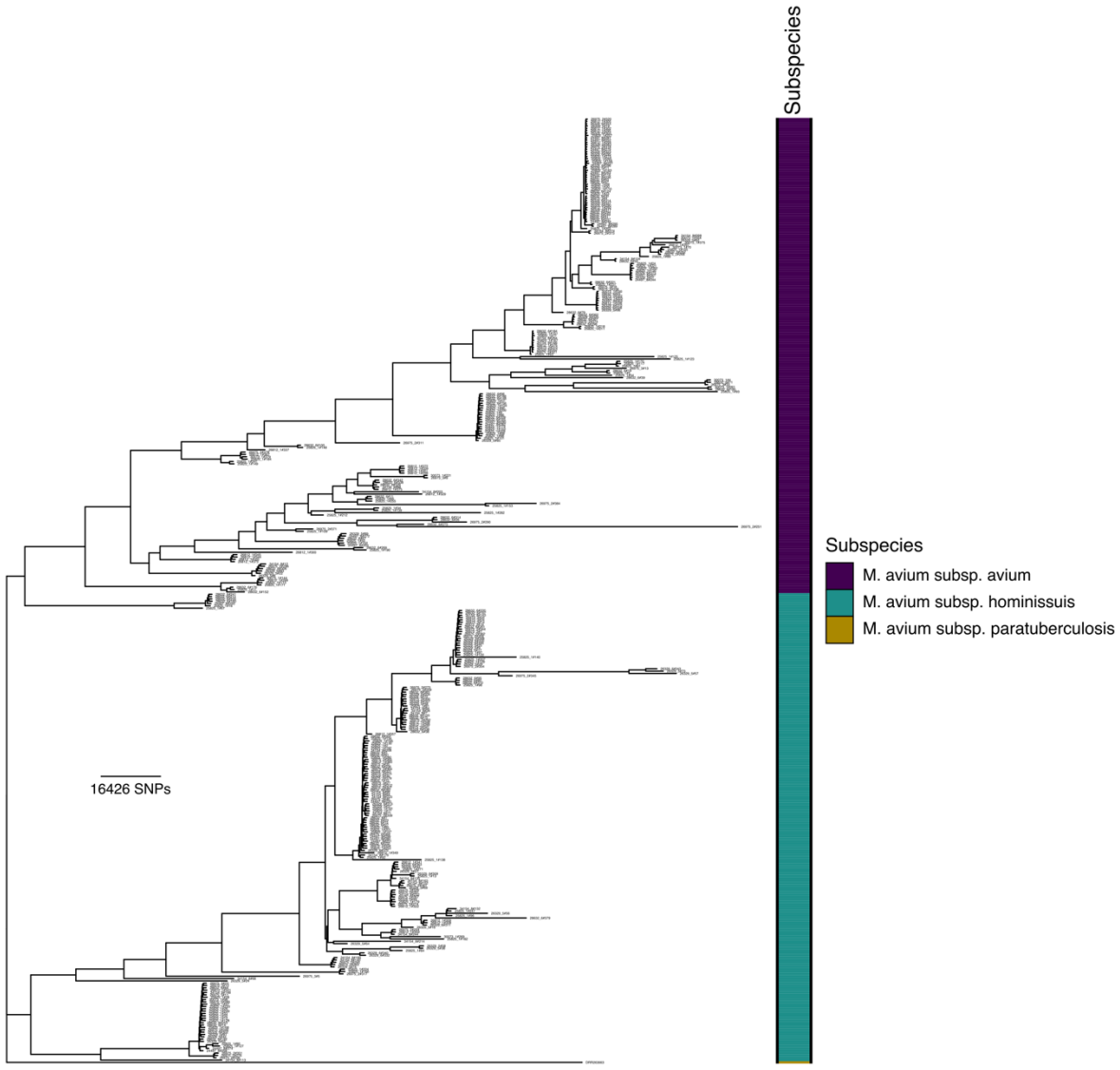
Supplementary Table 7. *Mycobacterium chimaera* global transmission clusters containing RBH isolates

| fastBAPS cluster | Transmission cluster | Total isolates (n) | Patients (n) | HCU (n) | Other (n) |
|---|---|---|---|---|---|
| FB5 | FB5_1 | 21 | 4 | 15 | 2 |
| FB5 | FB5_2 | 7 | 7 | 0 | 0 |
| FB5 | FB5_3 | 5 | 5 | 0 | 0 |
| FB5 | FB5_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 |

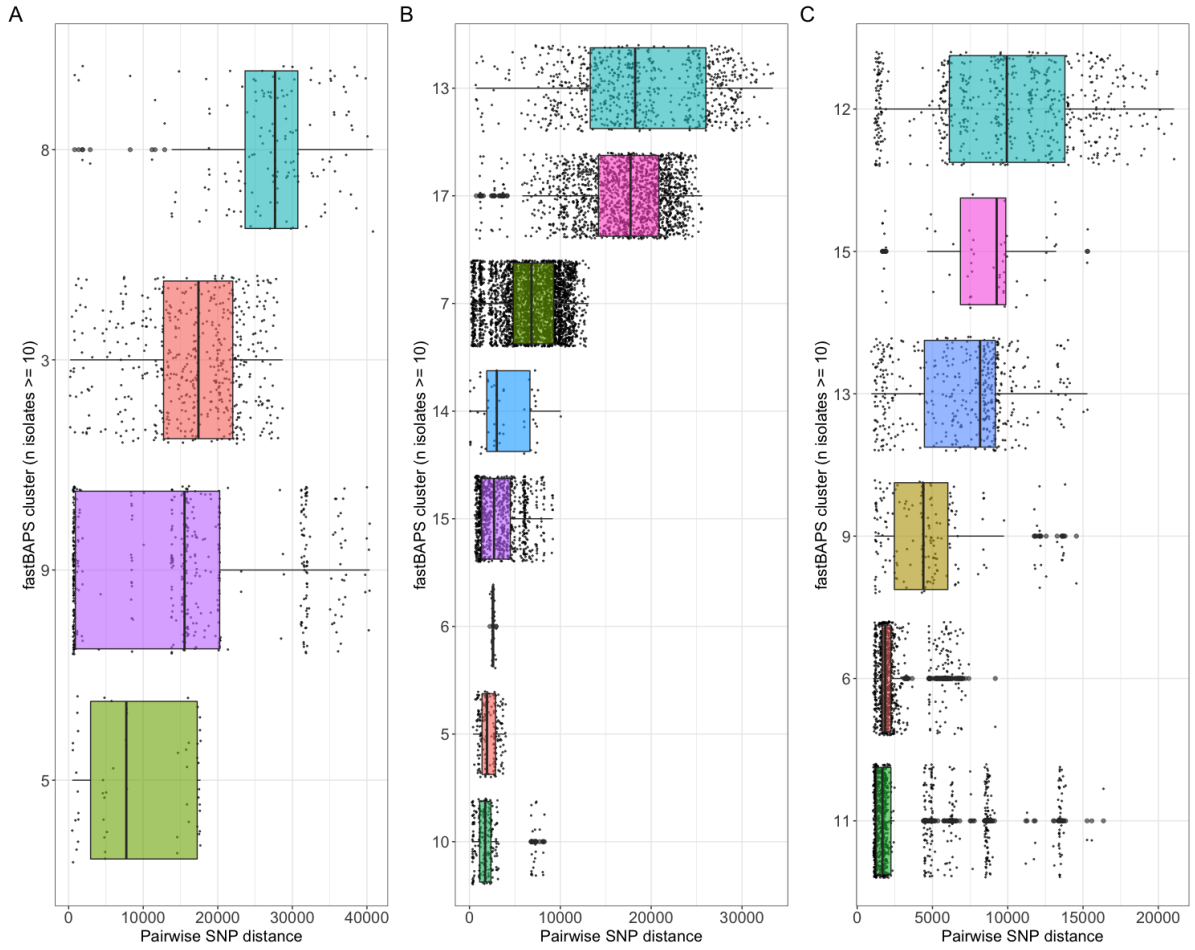


**Supplementary Figure 1. Within-host isolate pairwise SNP diversity in fastBAPS lineages** for A) *Mycobacterium intracellulare*; B) *Mycobacterium avium* subsp. *avium;* C) *Mycobacterium avium* subsp. *hominissuis* and D) *Mycobacterium chimaera*.

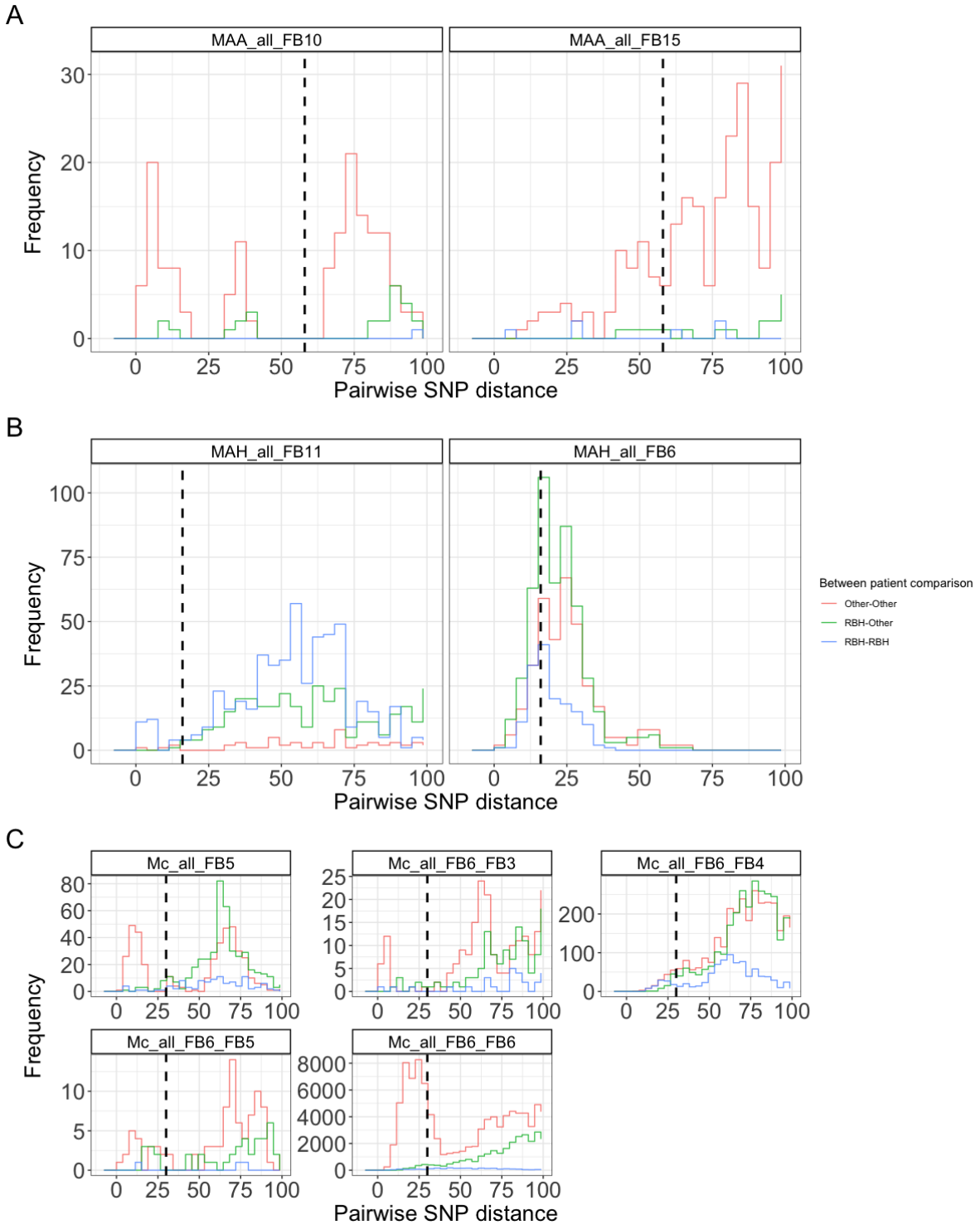



**Supplementary Figure 2. Population structure of *Mycobacterium avium* at the Royal Brompton Hospital.** Maximum likelihood phylogenetic tree of 406 *Mycobacterium avium* isolates rooted with a *Mycobacterium avium* subsp. *paratuberculosis* isolate (DRR263663).

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.




**Supplementary Figure 3. 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*




**Supplementary Figure 4: Distribution of pairwise SNP distances between isolates from the Royal Brompton Hospital and other studies**. Histograms showing distribution of

pairwise SNP distances for fastBAPS clusters for A) *Mycobacterium avium* subsp. *avium*; B) *Mycobacterium avium* subsp. *hominissuis* and C) *Mycobacterium chimaera*.