Children with cystic fibrosis are infected with multiple subpopulations of *Mycobacterium abscessus* with different antimicrobial resistance profiles

Running title: *M. abscessus* subclones in CF patients

Authors: Liam P. Shaw[1,2], Ronan M. Doyle[3,4], Ema Kavaliunaite[5], Helen Spencer[5], Francois Balloux[1], Garth Dixon[3,4], Kathryn A. Harris[3,4]

1: UCL Genetics Institute, UCL, London, UK.

2: Nuffield Department of Medicine, John Radcliffe Hospital, Oxford, UK.

3: Department of Microbiology, Virology and Infection Control, Great Ormond Street Hospital NHS Foundation Trust, London, UK

4: National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and UCL.

5: Paediatric Respiratory Medicine and Lung Transplantation, Great Ormond Street Hospital NHS Foundation Trust, London, United Kingdom

Corresponding author: Kathryn Harris (kathryn.harris@gosh.nhs.uk)

Summary:

Children with cystic fibrosis undergoing lung transplant harbour multiple subpopulations of *M. abscessus.* Subpopulations can have different antimicrobial resistance genotypes. Sputum isolates do not reflect the genetic diversity within a patient.

Abstract

Background: Children with cystic fibrosis (CF) can develop life-threatening infections of *Mycobacterium abscessus*. These present a significant clinical challenge, particularly when the strains involved are resistant to antibiotics. Recent evidence of within-patient subclones of *M. abscessus* in adults with CF suggests the possibility that within-patient diversity may be relevant for the treatment of pediatric CF patients.

Methods: We performed whole genome sequencing (WGS) on 32 isolates of *M. abscessus* from multiple body sites for two patients with CF undergoing treatment at Great Ormond Street Hospital, UK, in 2015.

Results: We found evidence of extensive diversity within patients over time. Clustering analysis of single nucleotide variants (SNVs) revealed that each patient harboured multiple subpopulations, which were differentially abundant between sputum, lung samples, chest wounds, and pleural fluid. Sputum isolates did not reflect overall within-patient diversity, including failing to detect subclones with mutations previously associated with macrolide resistance (*rrl* 2058/2059). Some variants were present at intermediate frequencies before lung transplant. The time of transplant coincided with extensive variation, suggesting that this event is particularly disruptive for the microbial community, but transplant did not clear the *M. abscessus* infection and both patients died as a result of this infection.

Conclusions: Isolates of *M. abscessus* from sputum do not always reflect the entire diversity present within the patient, which can include subclones with differing antimicrobial resistance profiles. Awareness of this phenotypic variability, with sampling of multiple body sites in conjunction with WGS, may be necessary to ensure the best treatment for this vulnerable patient group.

Keywords

Lung transplant; whole-genome sequencing; within-patient diversity; macrolides; physiological niches; ferric uptake.

Introduction

Mycobacterium abscessus is a nontuberculous mycobacteria (NTM) which has recently emerged as a major pathogen in cystic fibrosis (CF) patients[1]. Infection with *M. abscessus* is associated with poor clinical outcomes, particularly in conjunction with lung transplantation[2]. Treatment is challenging due to the intrinsic resistance of *M. abscessus* to many classes of antibiotics[3], along with certain genotypes drastically altering the efficacy of antibiotics[4]. The antimicrobial resistance (AMR) profile of isolates is highly relevant for treatment, but current diagnostic work mainly uses isolates from sputum, which may not reflect the full range of genetic diversity within the patient and therefore fail to recover the true AMR profile.

Minority variants from WGS have been used to infer the presence of multiple subpopulations (subclones) in longitudinal sputum isolates of *M. abscessus*[5]. However, it remains an open question whether patients harbour further unsampled genetic diversity. The lung is known to be capable of harbouring considerable pathogen diversity in chronic infections. For example, *M. tuberculosis* infections exist as multiple subpopulations with different AMR profiles[6,7]. The potential relevance of this genetic diversity for treatment is not yet known.

For this reason, we investigated longitudinal isolates from two patients infected with *M. abscessus subsp. abscessus* undergoing lung transplant at Great Ormond Street Hospital (Figure 1), and identified variable genomic positions within samples (single nucleotide variants, SNVs). By including isolates from sputum samples but also biologically important compartments such as pleural fluid, lung tissue, and swabs from chest wounds, we aimed to establish the extent and significance of within-patient variation in *M. abscessus* for this vulnerable group.

Methods

Patient cohort and sample collection

Great Ormond Street Hospital (GOSH) is a large regional centre for paediatric CF patients and the largest paediatric lung transplant centre in the UK. The two patients in this study (Patient 1 and Patient 2) were from other CF centres and were seen at GOSH for a transplant assessment, during the lung transplant procedure, and post-transplant (Table 1). Both patients in this study underwent regular respiratory microbiological diagnostic investigations, including specific stain and culture for mycobacteria on sputum pre- and post-transplant, and explanted lung tissue, bronchoalveolar lavage, pleural fluid and clamshell incision wound swabs post-transplant. All further microbiological investigations were carried out on 'sweeps' from pure culture plates. All *M. abscessus* isolates cultured in our laboratory are identified to sub-species level by PCR and sequencing of *hsp65* and *rpoB* genes and inducible macrolide resistance predicted by PCR and sequencing of the *erm(41)* gene as previously described[8]. VNTR profiles were obtained for selected isolates as previously described[9]. Phenotypic sensitivity data was obtained from the mycobacterial reference laboratory (Table 3).

Demographic and clinical data were extracted from the Patient Administration System (PIMS) and microbiological data from the Laboratory Information Management system (OMNI-client ISS) using SQL databases and Excel spreadsheets. Additional sources of information included CF and transplantation databases. Details of antimicrobial therapy administered prior to transplantation was provided by the referring CF centres. All investigations were performed in accordance with the Hospitals Research governance policies and procedures.

Whole genome sequencing

16 isolates from Patient 1 and 16 isolates from Patient 2 (Figure 1, Table 2) underwent DNA extraction as previously described[9] with the addition of a bead beating step. Total DNA concentration was determined using the Qubit high-sensitivity (HS) assay kit (Thermofisher) and a sequencing library was prepared from 50 ng of DNA using the Nextera Library Preparation kit (Illumina). Post-PCR clean-up was carried out using Ampure XP beads (Beckman). Library size was validated using the Agilent 2200 TapeStation with Agilent D1000 ScreenTape System and 150bp paired-end reads were sequenced on Illumina NextSeq 550 system. Raw sequencing reads have been deposited on ENA (Study Accession PRJEB28875) as well as two assemblies used as '*de novo* references' (see below).

Sequence data analysis

Initial variable nucleotide tandem repeat (VNTR) typing carried out as described previously[9,10] suggested the possibility of mixed infections, based on results intermediate between VNTR I and the closely related VNTR I* profile (differing at one locus). A preliminary mapping of all isolates to the standard *M. abscessus* strain ATCC 19977 chromosome (NCBI Accession: CU458896.1) showed that the mean coverage at 10X was ~91%, contrasted with >99% for a representative set of VNTR II isolates from another patient sequenced with the same protocol (not shown). In order to ensure we captured as much genetic diversity as possible, we therefore adopted a hybrid *de novo* and mapping approach. We selected the first isolate (temporally) for each patient and performed *de novo* assembly with Spades v3.10.0 with the --careful switch and otherwise default parameters[11]. After removing contigs with <10,000 bases to exclude small mobile genetic elements, this first *de novo* assembly was used as a new reference to map raw reads from other isolates using bwa mem v0.7.12 with default parameters[12]. This produced '*de novo* references' for Patient 1 and Patient 2 containing 5.17Mb and 5.28Mb respectively (Table

1). Contigs were reordered against the *M. abscessus* ATCC 19977 chromosome using Mauve v2.4.0 (2015-02-25)[13].

Variant identification and clustering analysis

In brief, the mapping file was sorted and indexed using picard v1.130, then GATK v3.30 was used to create a combined variant call format (VCF) file for each patient. Each position required a mapping depth > 30 in all samples from a patient to be included in downstream analysis. We manually inspected the 'self-mapping' of the reads from the first temporal sample to its own *de novo* assembly using IGV v2.4.10[14] to identify small regions where mapping was problematic. We removed SNVs within isolated regions where the self-mapping had unexpected peaks in coverage (Patient 1: contigs 12 (12,575-13,519bp) and 19 (96,916-97,032bp); Patient 2: contig 17 (12,718-12,770bp) as well as SNVs where the reference allele fraction from the self-mapping reads was < 5%.

As noted by Bryant *et al.*[5], patterns of linkage of variants in for *M. abscessus* can be suggestive of the existence of subpopulations. We aimed to establish a conservative lower bound for the number of clonal subpopulations within a patient, inferring their existence from the linkage patterns of variant frequencies across all samples. Patterns of linkage disequilibrium can also occur due to recombination, so we therefore attempted to remove local recombination in our analysis. Using the SNVs obtained via the mapping and filtering methods described above, we hierarchically clustered SNVs using Ward's minimum variance criterion[15] applied to Euclidean distances between allele frequencies with a dissimilarity threshold of 1 to define clusters. We removed clusters containing <4 SNVs. We also removed putative local recombination regions by removing clusters where the SNVs were distributed within a total range < 100,000bp (~2% of the *M. abscessus* genome).

In general, the inference of haplotype frequencies using variant frequencies from short sequencing reads for a microbial population undergoing recombination is a complex problem[16]. However, as we are not attempting to comment on abundances of subpopulations but only their presence, we did not need to infer haplotype frequencies. Observing *n* distinct clusters of variants within a patient over time (i.e. allele frequencies that co-vary in step with each other) means that there must be at least *n* bacterial haplotypes within the population producing these patterns. This fact holds even when recombination is present. Therefore, observing distinct clusters of linked variants tells us that distinct subpopulations of *M. abscessus* exist within individual samples.

Results

Individuals harbour extensive variation

We observed multiple positions in the *M. abscessus* genome which varied between different isolates within a patient over time (total variable positions used for clustering across all isolates, Patient 1: 54 positions, Patient 2: 64 positions), although isolates remained highly similar on average and were clearly the same infecting strain (mean inter-isolate SNV distances Patient 1: 2.07 +/- 0.92 SNVs, Patient 2: 1.96 +/- 1.81 SNVs). Subsets of these SNVs showed patterns of linkage across the *M. abscessus* genome (Figure 2a). Similar clustered patterns of linked SNVs could also arise due to recombination, but clustered SNVs were widely spread across the genome suggesting the presence of multiple subclones. Even if recombination were present, leading to mixtures of clusters (i.e. different haplotypes), then the observed abundance patterns still require multiple subclones. A neighbour-joining tree produced from distances between isolates based on these allele frequencies also suggested multiple subclones (Figure 2b).

Within-patient variation includes antimicrobial resistance mutations

Macrolide resistance in *M. abscessus* is driven by mutations at established positions in the *rrl* gene. Both patients received macrolides almost continuously throughout the six months prior to transplant (Figure 1). While initial isolates taken earlier in treatment were susceptible, we observed that resistance alleles at these positions (C/G) increased in abundance over time (Figure 3). Notably, all sputum isolates from Patient 1 showed a susceptible allele at position 2059 whereas isolates from pleural fluid and clamshell incision wound swabs carried a resistance allele (A2059C, Figure 3a). There was also substantial variation within sets of isolates taken on the same day. For example, three isolates from different lymph node samples taken on the day of transplant for Patient 2 showed completely different macrolide resistance profiles,

9

most strikingly at position 2058 (Figure 3b). These positions were not among the SNVs clustered into subpopulation structure clusters in Patient 2, but *rrl* 2059 was part of cluster 1.D in Patient 1, demonstrating that these resistance alleles can arise spontaneously but also persist linked to genetic background.

We also observed variable positions in the *rrs* gene in both patients (Figure 4). The *rrs* gene codes for 16S rRNA and is often a site of the emergence of aminoglycoside resistance (e.g. to amikacin), particularly the last few 100bp of the gene where the secondary structure of the rRNA can be affected by multiple mutations[17]. The *de novo* reference assembly for the gene in both patients was identical to the previously characterised sequence from amikacin-resistant *M. abscessus*[17,18], which was consistent with the measured AMR phenotype of the first sputum sample in Patient 2 but not Patient 1 (Table 3). Subsequently, an isolate from lung tissue on the day of transplant had a different allele at position 1174 in patient 1 (C→T; Figure 4), and was partially resistant when phenotyped (Table 3), suggesting this mutation may have been involved in this resistance. In Patient 2, we observed both A and G at position 1374, corresponding to the A1400G mutation which confers high-level amikacin resistance in *M. tuberculosis*[19]. The G allele was dominant by the end of treatment, suggesting that it may have conferred higher resistance and/or been an important compensatory mutation.

Other sites also showed high levels of variation in both patients. After sorting SNVs by the standard deviation of the reference allele fraction across isolates (Supplementary Dataset 1), the most highly variable position in Patient 1 was in the putative ferric uptake regulator FurB (MAB_1678c). This variant was present at high abundance in samples 49 days after transplant, with the reference allele fraction only present at <2% in one pleural fluid sample (Supplementary Dataset 1). Ferric uptake regulation has been associated with virulence of pathogenic mycobacteria; in mycobacterial infections the host response deprives the bacteria of

10

iron to prevent replication[20]. Iron is important for growth and virulence in *M. abscessus*[21] with gallium used as a treatment because of its ability to inhibit iron-dependent enzymes[22]. The most variable position in Patient 2 was within a putative linoleoyl-CoA desaturase (MAB_2148), and the second most highly variable position in patient 2 was within the cell division control protein 48 CDC48 (MAB_0347). Population heterogeneity via asymmetric cell division has been suggested as a factor facilitating the survival of *M. tuberculosis* across host physiological niches[23], and the control of cell division is probably similarly important in the survival of *M. abscessus* across body sites. Both patients had a variable position within *erm(41)* (MAB_2297) although at different positions, which confers inducible resistance to macrolides[24].

Sputum samples do not reflect overall within-patient diversity

The first sample from both patients was from sputum. Frequencies of the reference alleles were significantly associated with body site in Patient 2 (Kruskal-Wallis rank-sum test, $p<0.001$) but not Patient 1 (Kruskal-Wallis rank-sum test, $p=0.13$). Reference allele frequencies were significantly higher in subsequent sputum isolates compared to non-sputum isolates for the majority of SNVs in both patients (Patient 1: 51/54 SNVs with $p<0.05$ after Benjamini-Hochberg correction, Patient 2: 61/64 SNVs with $p<0.05$ after Benjamini-Hochberg correction), suggesting that sputum isolates tended to be more similar to the initial sputum isolate used as a reference, even for Patient 1 where 3/3 subsequent sputum isolates were post-transplant (immunosuppressed). This also suggests that non-sputum isolates harboured additional diversity that was not well-sampled using sputum.

Discussion

In this retrospective study we sought to establish the extent of within-patient variability of *M. abscessus* in two patients who developed severe complications following lung transplant as part

of treatment for CF. We used WGS to characterise this variability in isolates from longitudinal clinical samples, and were able to reveal patterns of linkage of SNVs consistent with the presence of multiple subpopulations within patients.

Isolates from the same patient were similar — for example, all within-patient inter-isolate distances were within the threshold of 25 SNVs previously suggested for inferring potential transmission events[25], and while mixed populations could theoretically lead to different transmission inferences, here between-patient variation was significantly larger — but this does not mean that the variation is not clinically significant. We have demonstrated that within-patient variation can contain biologically and clinically important variation. Notably, we observed that variation at *rrl* 2058/2059, associated with macrolide resistance, developed over the course of treatment. We observed extensive variation in isolates from Patient 2 prior to and on the day of transplant (i.e. before the patient was immunosuppressed). Isolates from Patient 1 prior to and on the day of transplant displayed no variation at these positions, so presumably the phenotypic macrolide resistance reported in these samples at this time was due entirely to inducible resistance conferred by *erm(41)*. Nonetheless, *rrl* 2058/2059 variants were present in later isolates, suggesting that macrolide use still has a therapeutic impact even in the presence of a functional *erm(41)* gene and drives the selection of high-level macrolide resistance. We also observed variation in the *rrs* gene, another source of resistance to antibiotics targeting ribosomal function (e.g. amikacin), at both previously recorded and novel positions, including an allele that rose in dominance over the course of treatment in Patient 2 (Figure 4).

When resistance to antibiotics is driven by point mutations at single positions, natural mutation rates will lead to the repeated presence of naturally occurring resistant cells. Conservatively taking values from the more slow-growing and non-recombining *M. tuberculosis,* a mutation rate of $\sim 8 \times 10^{-9}$ mutations per site per month[26] and a typical extracellular population of $\sim 10^9$

cells[27] clearly means that during treatment a typical within-patient population will repeatedly give rise to cells with the *rrl* 2058 mutation (for example). Typically such mutations have fitness costs, so remain at low frequency, but in the presence of antibiotics they rapidly achieve dominance. In *M. tuberculosis,* combination therapy is specifically designed to combat this selection of resistance. The strong resistance selection effect we observe for macrolides in *M. abscessus* highlights the weakness of current treatment regimens, particularly the lack of good companion medications.

Based on this data we would question how useful phenotypic testing of isolates recovered from a limited number of sputum isolates is for guiding antimicrobial therapy, as this strategy is unlikely to capture the diversity present in the full sample. Similarly, even though WGS is clearly a valuable tool, if restricted to the analysis of sputum isolates it may also fail to capture an accurate AMR profile. Previous studies on *M. tuberculosis* have shown that mycobacterial culture reduces the diversity recovered from sputum samples[28,29]. It is therefore possible that the diversity found across different sample sites in this study may have been present in sputum samples, but lost in the culture step. For *M. tuberculosis*, direct sequencing from sputum samples using capture-based enrichment methods has been shown to recover sample diversity not present in liquid culture[30]. Extending this to this case, it is possible that WGS at high depth applied directly to sputum samples could identify the variants detected between sample sites.

We observed highly variable positions in genes with direct relevance for the survival of *M. abscessus* across different physiological niches e.g. the regulation of ferric uptake (MAB_1678c) and the control of cell division (MAB_0347). In a chronic infection, mycobacteria must cope with considerable host stresses, including Fe starvation, which enables the persistence of *M. tuberculosis* in granulomas[31]. Phenotypic diversity due to expression can also occurs: colony morphotype ('smooth' or 'rough') has previously been linked to phenotype, although both

13

morphotypes appear to be capable of aggregation and intracellular survival[32]. Further diversity may come from the subpopulations harboured at different locations within the patient's lung. Regional selective pressures within the lung have been shown to drive the diversification of *Pseudomonas aeruginosa,* another chronic CF pathogen, and we would expect similar dynamics for *M. abscessus.* In particular, different body regions may have different antibiotic antibiotic concentrations. It has been recently shown for *M. tuberculosis* that antibiotic concentrations vary across biopsy sites in cavities, and that this variation is associated with different MICs and resistance-associated variants[7]. It seems highly plausible that similar effects exist in *M. abscessus* infections, and this is an important area for further research.

Our findings suggest that the wider diversity present within patients chronically infected with *M. abscessus* is not well-sampled with sputum, and that body site influences subpopulation structure. More widespread sampling of multiple body sites would provide a more accurate picture of the AMR profile of *M. abscessus* infecting a patient, and may be necessary to guide targeted antimicrobial therapy prior to transplant. However, in practical terms this would mean taking biopsies, which carries a significant clinical risk and would probably be unfeasible in patients awaiting transplant. An alternative solution to improve patient management before transplant might rely on deep-sequencing of multiple sputum samples. Such a strategy might capture a sufficient fraction of the total within-patient diversity to provide accurate information about the presence of minor variants, in particular those conferring AMR.

Acknowledgements

References

1   Bryant JM, Grogono DM, Rodriguez-Rincon D, *et al.* Emergence and spread of a human-transmissible multidrug-resistant nontuberculous mycobacterium. *Science (80-)* 2016; **354**. DOI:10.1126/science.aaf8156.

2   Watkins RR, Lemonovich TL. Evaluation of infections in the lung transplant patient. *Curr Opin Infect Dis* 2012; **25**: 193–8.

3   Nessar R, Cambau E, Reyrat JM, Murray A, Gicquel B. Mycobacterium abscessus: a new antibiotic nightmare. *J Antimicrob Chemother* 2012; **67**: 810–8.

4   Koh W-J, Jeong B-H, Kim S-Y, *et al.* Mycobacterial Characteristics and Treatment Outcomes in *Mycobacterium abscessus* Lung Disease. *Clin Infect Dis* 2017; **64**: 309–16.

5   Bryant JM, Grogono DM, Rodriguez-Rincon D, *et al.* Emergence and spread of a human-transmissible multidrug-resistant nontuberculous mycobacterium. *Science (80-)* 2016; **354**: 751–7.

6   Kaplan G, Post FA, Moreira AL, *et al.* Mycobacterium tuberculosis growth at the cavity surface: a microenvironment with failed immunity. *Infect Immun* 2003; **71**: 7099–108.

7   Dheda K, Lenders L, Magombedze G, *et al.* Drug-Penetration Gradients Associated with Acquired Drug Resistance in Patients with Tuberculosis. *Am J Respir Crit Care Med* 2018; **198**: 1208–19.

8   Blauwendraat C, Dixon GLJ, Hartley JC, Foweraker J, Harris KA. The use of a two-gene sequencing approach to accurately distinguish between the species within the Mycobacterium abscessus complex and Mycobacterium chelonae. *Eur J Clin Microbiol Infect Dis* 2012; **31**: 1847–53.

9   Harris KA, Kenna DTD, Blauwendraat C, *et al.* Molecular fingerprinting of *Mycobacterium abscessus* strains in a cohort of pediatric cystic fibrosis patients. *J Clin Microbiol* 2012. DOI:10.1128/JCM.00155-12.

10   Harris KA, Underwood A, Kenna DTD, *et al.* Whole-genome sequencing and epidemiological analysis do not provide evidence for cross-transmission of mycobacterium abscessus in a cohort of pediatric cystic fibrosis patients. *Clin Infect Dis* 2015; **60**: 1007–16.

11   Bankevich A, Nurk S, Antipov D, *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 2012; **19**: 455–77.

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

13   Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS One* 2010; **5**: e11147.

14   Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013; **14**: 178–92.

16

15   Ward JH. Hierarchical Grouping to Optimize an Objective Function. *J Am Stat Assoc* 1963; **58**: 236–44.

16   Hong LZ, Hong S, Wong H, *et al.* BAsE-Seq: a method for obtaining long viral haplotypes from short sequence reads. *Genome Biol* 2014; **15**: 517.

17   Nessar R, Reyrat JM, Murray A, Gicquel B. Genetic analysis of new 16S rRNA mutations conferring aminoglycoside resistance in Mycobacterium abscessus. *J Antimicrob Chemother* 2011; **66**: 1719–24.

18   Prammananan T, Sander P, Brown BA, *et al.* A single 16S ribosomal RNA substitution is responsible for resistance to amikacin and other 2-deoxystreptamine aminoglycosides in Mycobacterium abscessus and Mycobacterium chelonae. *J Infect Dis* 1998; **177**: 1573–81.

19   Alangaden GJ, Kreiswirth BN, Aouad A, *et al.* Mechanism of resistance to amikacin and kanamycin in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 1998; **42**: 1295–7.

20   Neyrolles O, Wolschendorf F, Mitra A, Niederweis M. Mycobacteria, metals, and the macrophage. *Immunol Rev* 2015; **264**: 249–63.

21   Abdalla MY, Ahmad IM, Switzer B, Britigan BE. Induction of heme oxygenase-1 contributes to survival of Mycobacterium abscessus in human macrophages-like THP-1 cells. *Redox Biol* 2015; **4**: 328–39.

22   Olakanmi O, Kesavalu B, Pasula R, Abdalla MY, Schlesinger LS, Britigan BE. Gallium Nitrate Is Efficacious in Murine Models of Tuberculosis and Inhibits Key Bacterial Fe-Dependent Enzymes. *Antimicrob Agents Chemother* 2013; **57**: 6074–80.

23   Kieser KJ, Rubin EJ. How sisters grow apart: mycobacterial growth and division. *Nat Rev Microbiol* 2014; **12**: 550–62.

24   Maurer FP, Castelberg C, Quiblier C, Bottger EC, Somoskovi A. Erm(41)-dependent inducible resistance to azithromycin and clarithromycin in clinical isolates of Mycobacterium abscessus. *J Antimicrob Chemother* 2014; **69**: 1559–63.

25   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**: 1551–60.

26   Eldholm V, Balloux F. Antimicrobial Resistance in Mycobacterium tuberculosis : The Odd One Out. *Trends Microbiol* 2016; **24**: 637–48.

27   Lalande L, Bourguignon L, Maire P, Goutelle S. Mathematical modeling and systems pharmacology of tuberculosis: Isoniazid as a case study. *J Theor Biol* 2016; **399**: 43–52.

28   Martín A, Herranz M, Ruiz Serrano MJ, Bouza E, García de Viedma D. The clonal composition of Mycobacterium tuberculosis in clinical specimens could be modified by culture. *Tuberculosis* 2010; **90**: 201–7.

29   Hanekom M, Streicher EM, Van de Berg D, *et al.* Population Structure of Mixed Mycobacterium tuberculosis Infection Is Strain Genotype and Culture Medium Dependent. *PLoS One* 2013; **8**: e70178.

30   Doyle RM, Burgess C, Williams R, *et al.* Direct Whole-Genome Sequencing of Sputum Accurately Identifies Drug-Resistant *Mycobacterium tuberculosis* Faster than MGIT Culture Sequencing. *J Clin Microbiol* 2018; **56**. DOI:10.1128/JCM.00666-18.

31   Kurthkoti K, Amin H, Marakalala MJ, *et al.* The Capacity of *Mycobacterium tuberculosis* To Survive Iron Starvation Might Enable It To Persist in Iron-Deprived Microenvironments of Human Granulomas. *MBio* 2017; **8**. DOI:10.1128/mBio.01092-17.

32   Clary G, Sasindran SJ, Nesbitt N, *et al. Mycobacterium abscessus* Smooth and Rough Morphotypes Form Antimicrobial-Tolerant Biofilm Phenotypes but Are Killed by Acetic Acid. *Antimicrob Agents Chemother* 2018; **62**. DOI:10.1128/AAC.01782-17.

Tables

	Patient 1	**Patient 2**
Total bases	5,172,759	5,275,491
Mean depth of coverage	25.3X	74.1X
Number of contigs	28	46

Table 1. Summary statistics for the de novo reference genomes of the two patients in this study. References were assembled for the first temporal sample from each patient (patient_1_S1 and patient_2_S1; see Methods).

Patient	*M. abscessus* sub-species	Sex	CF genotype	Date of last smear-positive sample	Date of last *M. abscessus* isolate
Patient 1	ABS	Female	F508del/F508del	-166 days	+56 days
Patient 2	ABS	Male	F508del/W1282X	-81 days	+59 days

Table 2: patient clinical data. Dates are relative to the day of transplant.

Sample	Body site	Date isolated	Colony morphotype	VNTR profile	Phenotypic susceptibility
patient_1_S1	Sputum	-166 days	Rough	I	Amikacin-S; Cipro-R; Clarithromycin-R; Doxycycline-R; Augmentin-R
patient_1_S2	Lung tissue	0 days	Rough	I	Amikacin-PR; Cipro-R; Clarithromycin-R; Doxycycline-R; Linezolid-R; Co-trem-R; Cefotaxime-R
patient_1_S3	Sputum	10 days	Smooth	I	n/a
patient_1_S4	Sputum	15 days	Rough	I	n/a
patient_1_S5	Pleural Fluid	21 days	Smooth	I	n/a
patient_1_S6	Pleural Fluid	38 days	Rough	I*	n/a
patient_1_S7	Pleural Fluid	38 days	Rough	I*	n/a
patient_1_S8	Pleural Fluid	40 days	Smooth	I*	n/a
patient_1_S9	BAL	42 days	Smooth	I*	n/a
patient_1_S10	Pleural Fluid	49 days	Smooth	I*	n/a
patient_1_S11	Pleural Fluid	49 days	Smooth	I*	n/a
patient_1_S12	Pleural Fluid	49 days	Smooth	I*	n/a
patient_1_S13	Chest wound swab	49 days	Smooth	I / I*	n/a
patient_1_S14	Chest wound swab	49 days	Smooth	I	n/a
patient_1_S15	Sputum	51 days	Smooth	I / I *	n/a
patient_1_S16	Chest wound swab	56 days	Rough	I	n/a
patient_2_S1	Sputum	-121 days	Rough	n/a	Amikacin-R; Cipro-R; Clarithromycin-R; Doxycycline-R; Linezolid-R; Co-trem-R; Cefotaxime-R
patient_2_S2	Sputum	-81 days	Smooth	I	n/a
patient_2_S3	Lymph node tissue	0 days	Rough	I	n/a
patient_2_S4	Lymph node tissue	0 days	Smooth	n/a	n/a

patient_2_S5	Lymph node tissue	0 days	Rough	I	n/a
patient_2_S6	Lung tissue	0 days	n/a	I	n/a
patient_2_S7	Sputum	14 days	Smooth	n/a	n/a
patient_2_S8	BAL	27 days	Smooth	I	n/a
patient_2_S9	Pleural Fluid	33 days	Smooth	I	n/a
patient_2_S10	BAL	41 days	Smooth	I / I*	n/a
patient_2_S11	Pleural Fluid	42 days	Smooth	I	n/a
patient_2_S12	Pleural Fluid	42 days	Smooth	I	Amikacin-R; Ciprofloxacin-R; Clarithromycin-R; Doxycycline-R; Linezolid-R; Co-trimoxazole-R; Cefoxitin-R; Tobramycin-R; Moxyfloxacin-R
patient_2_S13	Drain site swab	49 days	Smooth	n/a	n/a
patient_2_S14	Pleural Fluid	50 days	Smooth	n/a	n/a
patient_2_S15	Pleural Fluid	58 days	Smooth	I	n/a
patient_2_S16	Pleural Fluid	59 days	Smooth	n/a	n/a

Table 3: Routine microbiology data about isolates in this study. Dates are relative to the date of transplant for each patient. An asterisk (*) indicates that the VNTR profile differed at one locus. n/a indicates data not available. Phenotypic susceptibility was available for a minority of isolates (S=susceptible, R=resistant, PR=partially resistant).

Figure legends

Figure 1. Overview of the sampling scheme and antibiotic treatment. Treatment regime including intravenous antibiotics (boxed coloured lines) and oral antibiotics (faint coloured lines) for both patients during the six-month period prior to lung transplant (red vertical line). The sampling scheme is represented at the bottom of each panel (coloured points).

Figure 2. Linkage patterns of SNVs across samples suggest the presence of closely-related subpopulations within patients. (a, b) SNVs were grouped into clusters (colours) using an unsupervised clustering technique, showing clear patterns of abundance across samples (see Methods). Genome position was inferred by ordering *de novo* contigs against the *M. abscessus* ATCC 19977 reference genome. **(c, d)** Midpoint-rooted neighbour-joining tree based on Euclidean distances between samples using these clustered SNVs show this variation within patients over time (numbers) and body site (colours).

Figure 3. Variants in the *rrl* gene (23S rRNA) arise during treatment and are present in isolates from concurrent samples. Relative allele fractions at these positions show that although the initial sputum isolate was susceptible for both patients, resistance appeared to develop during treatment. Samples are ordered by time, with boxes indicating samples on the same day. Samples taken on the day of transplant are shown in bold text. N.B. Here following the usual convention for the *rrl* gene we use *E. coli* numbering. Positions 2058 and 2059 in *E. coli* correspond to 2269 and 2270 in *M. abscessus.* (S = sputum, BAL = bronchalveolar lavage, CW = chest wound, DSW = drain site swab, LT = lung tissue, LN = lymph node, PF = pleural fluid).

Figure 4. Variants in the *rrs* (16S rRNA) gene over the course of treatment. Patient 1: position 1174. Patient 2: position 1374, previously associated with amikacin resistance (see text). Numbering relative to ATCC 19977 reference. (S = sputum, BAL = bronchalveolar lavage, CW = chest wound, DSW = drain site swab, LT = lung tissue, LN = lymph node, PF = pleural fluid).

Supplementary data

Supplementary Dataset 1. SNVs used for clustering in both patients. A list of the variant positions used for clustering to identify subpopulations, sorted in decreasing order of standard deviation across samples. In addition to the position of the variant in the *de novo* assembly reference, its position (if present) in the ATCC 19977 is given, along with the relevant UniProt ID. Samples are sorted in time order.
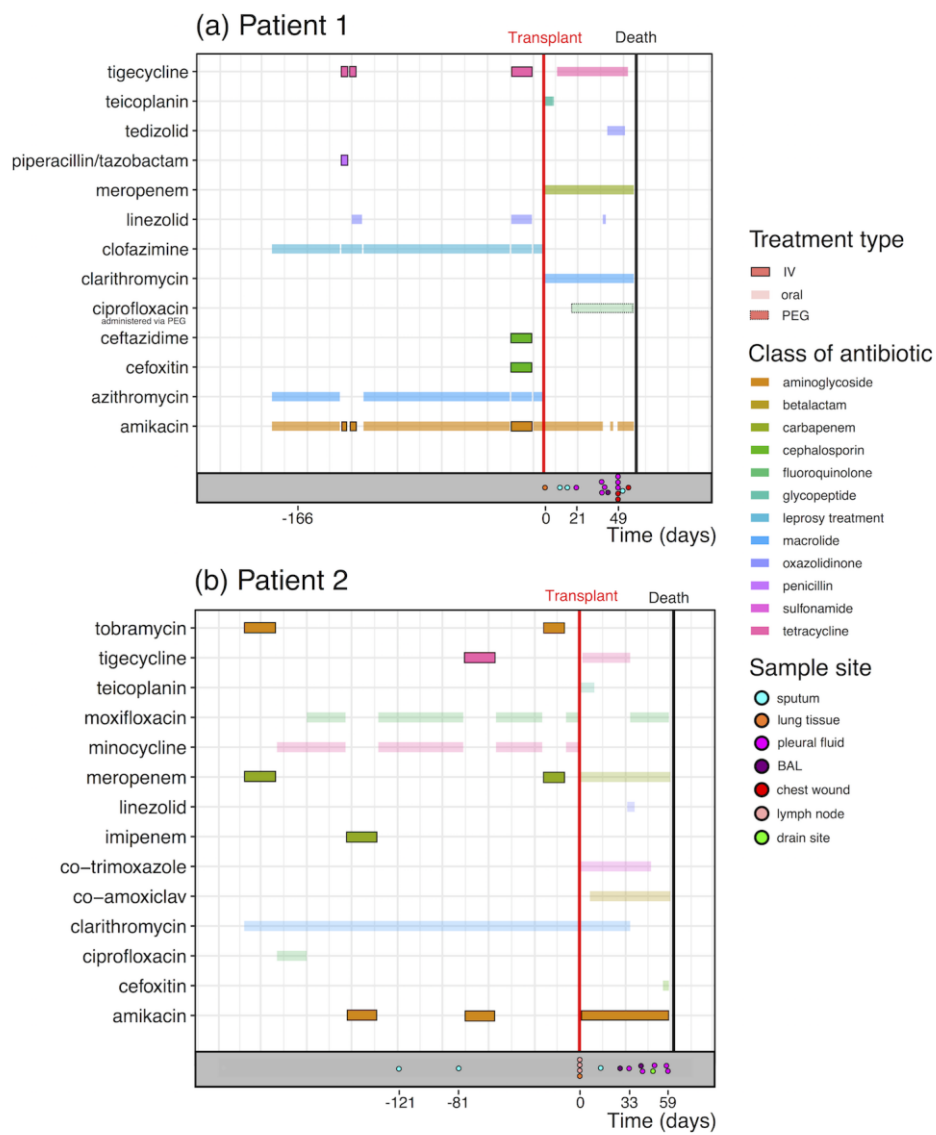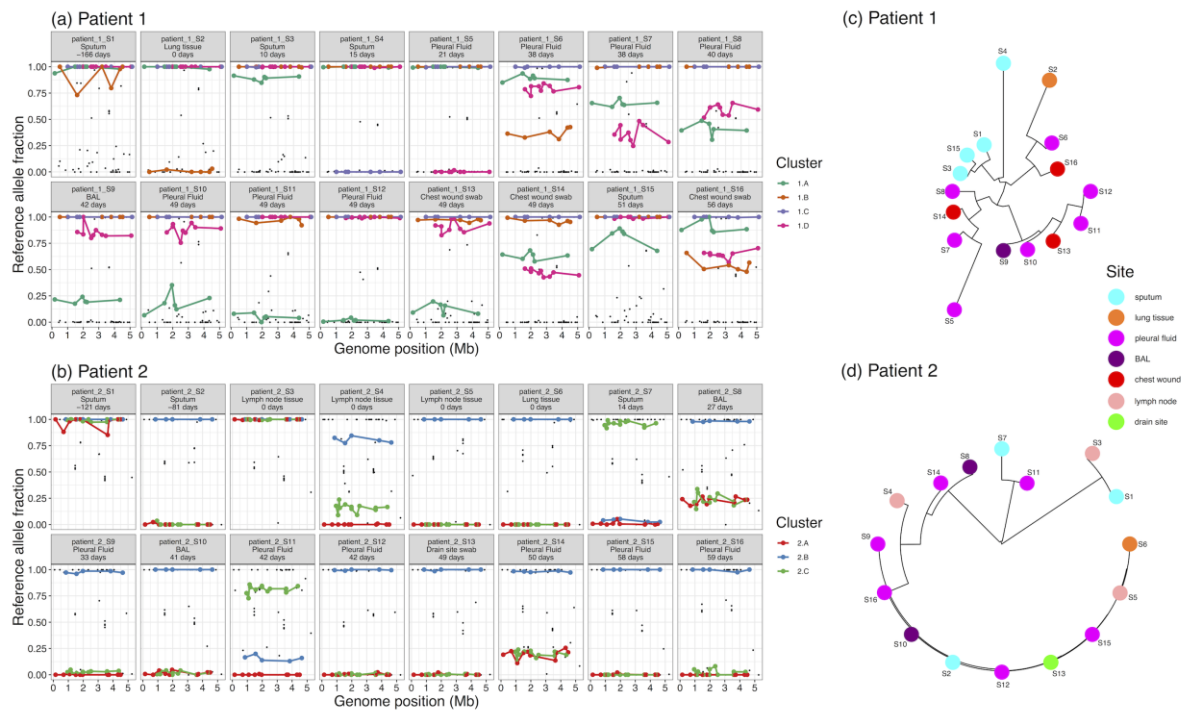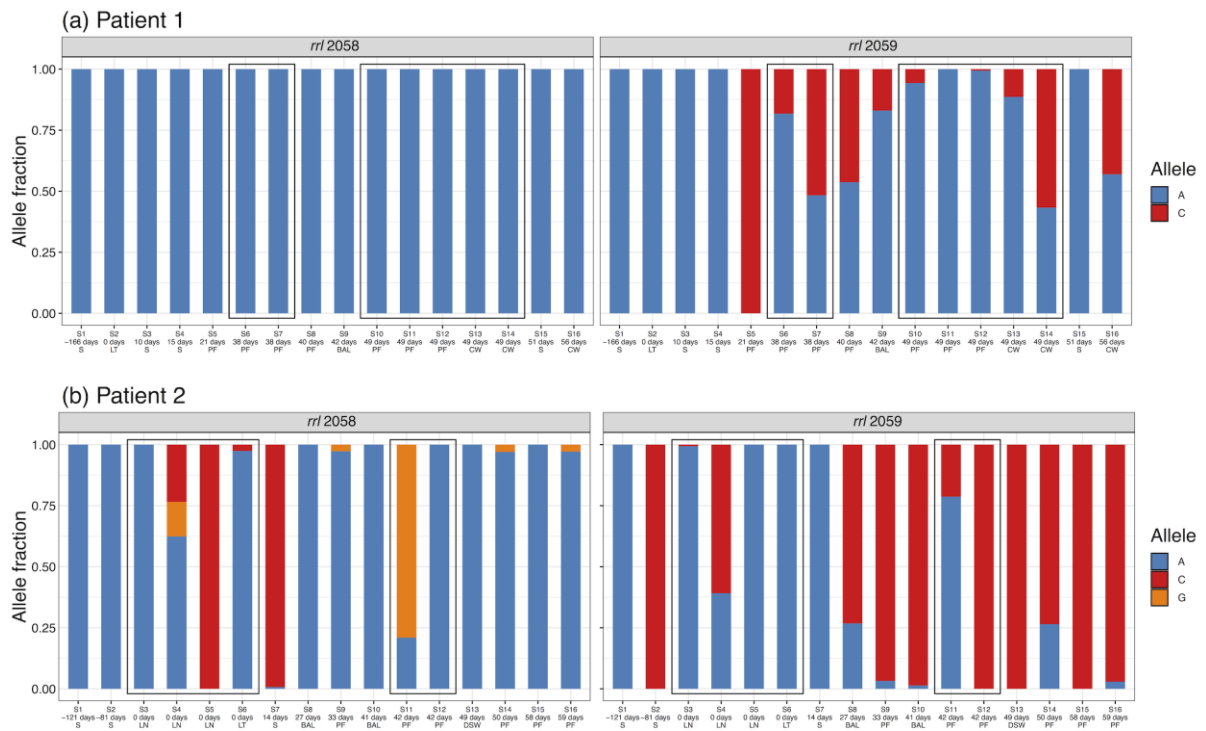
Figure 1

Figure 2

Figure 3

(a) Patient 1

(b) Patient 2

Figure 4