Supplementary Materials for

Population-level genomics identifies the emergence and global spread of a human transmissible multidrug-resistant nontuberculous mycobacterium.

Josephine M Bryant[1,2*], Dorothy M Grogono[2,3*], Daniela Rodriguez-Rincon[2], Isobel Everall[1], Karen P Brown[2,3], Pablo Moreno[4], Deepshikha Verma[5], Emily Hill[5], Judith Drijkoningen[2], Peter Gilligan[6], Charles R Esther[6], Peadar G Noone[6], Olivia Giddings[6], Scott C. Bell[7,8,9], Rachel Thomson[10], Claire E. Wainwright[8,11], Chris Coulter[12], Sushil Pandey[12], Michelle E Wood[7,8,9], Rebecca E Stockwell[7,8], Kay A Ramsay[7,8], Laura J Sherrard[7], Timothy J Kidd[14,16], Nassib Jabbour[15,17], Graham R Johnson[17], Luke D Knibbs[18], Lidia Morawska[17], Peter D Sly[13], Andrew Jones[19], Diana Bilton[19], Ian Laurenson[20], Michael Ruddy[21], Stephen Bourke[22], Ian CJW Bowler[23], Stephen J Chapman[23], Andrew Clayton[24], Mairi Cullen[25], Thomas Daniels[25], Owen Dempsey[26], Miles Denton[27], Maya Desai[28], Richard J Drew[29], Frank Edenborough[30], Jason Evans[21], Jonathan Folb[31], Helen Humphrey[32], Barbara Isalska[25], Søren Jensen-Fangel[33], Bodil Jönsson[34], Andrew M. Jones[25], Terese L Katzenstein[35], Troels Lillebaek[36], Gordon MacGregor[37], Sarah Mayell[29], Michael Millar[38], Deborah Modha[39], Edward F Nash[40], Christopher O'Brien[22], Deirdre O'Brien[41], Chandra Ohri[39], Caroline S Pao[38], Daniel Peckham[28], Felicity Perrin[42], Audrey Perry[22], Tania Pressler[35], Laura Prtak[31], Tavs Qvist[35], Ali Robb[22], Helen Rodgers[43], Kirsten Schaffer[41], Nadia Shafi[3], Jakko van Ingen[44], Martin Walshaw[45], Danie Watson[38], Noreen West[46], Joanna Whitehouse[40], Charles S Haworth[3], Simon R Harris[1], Diane Ordway[5], Julian Parkhill[1**] & R. Andres Floto[2,3**].

correspondence to: Andres Floto: arf27@cam.ac.uk or
Julian Parkhill: parkhill@sanger.ac.uk

This PDF file includes:

Materials and Methods

Sample collection.

All major Cystic Fibrosis centres in the UK, and the five mainland mycobacterial reference laboratories (National Mycobacterial Reference Laboratory (NMRL); Regional Centre for Mycobacteriology, Birmingham; Regional Centre for Mycobacteriology, Newcastle; Scottish Mycobacteria Reference Laboratory and Wales Centre for Mycobacteriology); Cystic Fibrosis centres from St Vincent's Hospital Dublin (Republic of Ireland), University of North Carolina Chapel Hill (USA), Gotheburg (Sweden); Copenhagen and Skejby (Denmark), Queensland (Australia), and Nijmegen (Netherlands) were asked to submit clinical isolates of *Mycobacterium abscessus* and (because they are frequently mis-speciated in clinical laboratories) *M. chelonae* from stored collections or prospectively for the period of the study. Where possible, bacterial samples taken directly from the original Mycobacterial growth indicator tube (MGIT) cultures were preferred, but clonal cultures and DNA-only samples were also accepted. Ethical approval for the study was obtained nationally for centres in England and Wales from the National Research Ethics Service (NRES; REC reference: 12/EE/0158) and the National Information Governance Board (NIGB; ECC 3-03 (f)/2012) and locally for other centres. All sequence data associated with this study is deposited in the European Nucleotide Archive under project accession ERP001039.

DNA extraction and sequencing.

All bacterial samples were sub-cultured onto solid media. Biomass from solid cultures (taken from sweeps of colonies) was mixed with 425-600 µm glass beads, pulse-vortexed for 5 minutes, incubated at 80 ^{o}C for 10 minutes and then centrifuged. The supernatant was then transferred to a fresh tube and DNA extracted using a Qiagen QIAamp DNA mini kit as previously described [17]. The above, combined with DNA samples sent directly by collaborating centres, resulted in a DNA collection of 1,173 isolates which were then subjected to multiplexed paired-end sequencing using the Illumina Hiseq platform. Reads were mapped using SMALT [38] to three different reference sequences (ATCC19977 [39] and *M. a. massiliense* and *M. a. bolletii de novo* assemblies described previously [17]), representing the three different subspecies. The reference with the greatest number of successfully mapped reads was chosen for downstream analysis. Samples that did not map with at least 70% identity to any of the three reference genomes were *de novo* assembled and the resultant contigs were analysed using BLAST. All were identified as non-*M. abscessus* NTMs so were excluded from any downstream analysis. We thus obtained complete whole genome sequences for 1080 isolates of *M. abscessus* from 503 individuals for further analysis.

Variant calling

Consensus variants: Consensus variants were called using SAMtools and bcftools [40] using filters designed to keep false positives to a minimum [41], which include a minimum base quality of 50, a minimum mapping quality of 30, support from at least four reads (two forward and two reverse) and an absence of heterozygosity.

Detection of minority variants: In a study of longitudinal *M. abscessus* sampling in 11 chronically infected CF individuals, whole genome sequencing was carried out on DNA extracted from sweeps of *M. abscessus,* maintaining a sample of the within-patient diversity which could be captured using a high depth of sequencing (x100). In order to detect minority variants, where all reads do not agree on a consensus sequence,

stringent filters were applied. As a first step, stringent mapping was used where in addition to the default SMALT [38] parameters a minimum nucleotide identity of 0.98 was applied, which avoided the mapping of reads with more than one mis-match which could be considered poor quality. Possible high quality minority variants were then extracted from the mapping output by parsing the VCF file [43]. To distinguish sequencing errors from true variants, a minority variant had to be supported by at least four reads, where at least two reads were mapped to each strand, and a strand bias P value cutoff of 0.05 was applied. To avoid heterozygous positions which may arise due to mis-mapping, the base positions had to have a depth of coverage within a normal range (+/- 50% of the average). In addition, any positions within 200bp of another putative minority variant were also removed, as these are more likely to represent mapping error than a true *de novo* generated variant.

Inferring the presence of subclones from minority variants: In patient two (Papworth outbreak [17]), 17 minority variants were found to occur multiple times in different samples, and 14 of these co-occurred with one or more other variant at the same frequency, indicating linkage on the same genetic background. This suggests the presence of several subclones within the patient. Using the observed linkage between these variants, hypothetical lineages could be reconstructed. The evolutionary succession of these lineages and the variants associated with them could be inferred if variants sometimes occurred at a lower frequency (but never higher) or at a higher frequency (but never lower) than another linked variant. One of these subclones was found to be present in Patient 28, for which there is strong evidence of transmission from patient two [17]. Using the frequency of the "terminal" variants, the frequency of the sub-lineages could be tracked over time. Minority variant analysis to determine subclone evolution was performed for isolates from all 11 patients identified within the transmission clusters at Papworth Hospital *[17]*.

Phylogenetic analyses

Species-wide phylogeny

Sequences from this study, along with 29 publicly available isolates (see Supplementary Table 1), were mapped to the type strain *M. abscessus* ATCC19977 using SMALT version 0.7.4 [38]. An alignment of the 356,526 variable positions in the core genome of the isolates was used to construct a maximum likelihood tree using RAxML, with 100 bootstrap replicates performed [42]. The bootstrap support for the major nodes is shown in Figure S1. The percentage average nucleotide identity (ANI) was calculated by applying the ANIm tool in the Jspecies package [43] to *de novo* assembled genomes [44].

Phylogenetic clustering

Maximum likelihood trees were built for the three individual subspecies datasets using RAxML v. 7.0.4 [40] using only the first available isolate per patient to avoid patient-specific clustering. Two different methods were used to identify clusters in the phylogeny. The first is BAPs [45], a Bayesian based method which hierarchically clusters the data. However this method assigns all samples to a cluster so is unable to distinguish dense clusters from more sparsely distributed clusters. In order to distinguish clustered from non-clustered isolates we used a second, novel, method to search for significantly dense nodes in the phylogenies. The identification of highly dense nodes allows us to identify monophyletic clades with minimal genetic distance between isolates, thus indicating a cluster. TreeGubbins [46] achieves this by calculating the density of each node (mean descendent branch length) and comparing it to the expected density (mean branch length of remaining tree) using a one-dimensional scanning

statistic likelihood function. Starting with the node with the highest likelihood, statistical significance is assessed through permuting the branch lengths across the tree and similarly maximising the likelihood function on each permuted tree. This is repeated iteratively until deemed no longer significant (P>0.05). All of the leaf-descendants (isolates) of each significant node were considered a cluster and any isolates which did not form part of a significant cluster were considered non-clustered. There was good agreement between the major clusters identified via both methods.

Only clusters from more than one site (to exclude possible point source outbreaks) and greater than five isolates (to ensure enough comparisons for calculations of summary statistics) were considered dominant circulating clones (DCCs) and included in downstream analysis.

When calculating pairwise differences between isolates from the same site, different sites and different countries, variants were called using a reference-based mapping approach (as above) to a separate *de novo* assembled reference (using Velvet v1.2.03 [44]) for each BAPs cluster. To avoid multiple comparisons, for each isolate, the isolate with the closest genetic distance (measured as pairwise SNPs) was identified and classed as being from the same site, a different site but same country or a different site from a different country.

Coalescent analyses

The BEAST package (v1.7.5) [24], a program for Bayesian Markov chain Monte Carlo (MCMC) analysis of genetic sequences, was used to estimate the age of clusters. Input XML files were created using BEAUTi [24] from the whole genome SNP-alignment and the associated dates of isolation for each isolate. Again, only the first isolate from each patient was used and possible recombination was removed [47]. For all analyses, three independent MCMC chains of 100 million states were run using a GTR evolutionary model and a variety of combinations of clock (strict and log normal relaxed) and population (constant, skyline and exponential) size models. Tracer (v1.5) [48] was used to assess convergence (after an initial burn-in period of 10 %), agreement between the three runs and that all effective sample size (ESS) values were greater than 200. BEAST analyses of substitution rate and time rely on the assumption of the presence of a molecular clock. This was assessed by carrying out a root-tip distance randomization procedure as described previously [49]. Briefly this is carried out by plotting the root to tip distance of each isolate in the phylogeny against the date of their isolation. The resulting linear regression co-efficient is then compared to what would be expected by chance alone by producing 99 comparison datasets, in which the isolation dates were permuted on the phylogeny. This test yielded a significant result for *M. a. abscessus cluster* 1, but not *M. a. abscessus* cluster 2 or *M. a. massiliense* cluster 1 which is likely to be due to inadequate sampling. As *M. a. massiliense* cluster 1 was of particular interest, we used the substitution rate from *M. a. abscessus* cluster 1 (mean 3.5 $x10^{-7}$, 1.83-5.11$x10^{-7}$ 95% higher posterior density interval; strict clock, constant population growth; maximum clade credibility tree shown in Figure S4), which we demonstrated to have a highly similar within-patient substitution rate (**Figure S12**) to estimate the age of internal nodes in the phylogeny. Calculations were made excluding any SNPs associated with recombination events. Figure 3a shows a maximum likelihood phylogeny of *M. a. massiliense* cluster 1 (recombination events not removed) with the estimated ages of nodes of interest marked on for illustration purposes. The estimated age of *M. a abscessus* cluster 2 is shown in Figure S5.

Clinical data collection

Clinical data was collected for a subset of 345 patients with Cystic Fibrosis including information on whether or not patients fulfilled American Thoracic Society (ATS) criteria for *M. abscessus* NTM pulmonary disease (NTM-PD; [1]) namely: two or more sputum (or one or more bronchoscopic) culture positive for *M. abscessus;* radiological (HRCT) changes consistent with NTM infection; symptoms attributable to NTM-PD, and clinical exclusion of other diagnoses. Clinical outcome was recorded as one of the following: less than 6 months treatment; infection cleared (defined as either culture conversion or sustained clinical improvement where no further cultures were available); or chronic infection (persistently culture positive or clinical deterioration where no further cultures were available).

Aerosol analysis

To assess the potential for airborne spread of nontuberculous mycobacteria between individuals with CF, we set out to determine whether *M. abscessus* could be detected in cough aerosols within the respirable size range and, if so, to define the distance these potentially infectious aerosols could travel and the duration for which they remain airborne, using our validated detection system (as previous described *[35])*.

A clinically stable 17 year old male with CF who was chronically infected (>2 years) with *M. abscessus* consented to participate in this study (previously approved by The Children's Health Queensland HHS Human Research Ethics Committee HREC/14/QRCH/88 and TPCH Research Governance Office SSA/14/QPCH/202. On the day of the study, his FEV_1 was 3.9L and FVC was 4.8L (79% and 92% predicted respectively). The patient underwent several cough manoeuvers while placed in both the distance rig and duration rigs, as previously described *[35]* with minor modifications. In the distance rig, aerosol flight distances of 2 and 4 metres were used respectively with 5 minutes of voluntary cough for each study. In the duration rig, aerosols generated by 2 minutes of voluntary cough for each study. were allowed to remain suspended for 5, 15 and 45 minutes before extraction.

Collection of the cough aerosols were via six-stage Anderson chambers loaded with culture plates (Middlebrook 7H11 media supplemented with OADC (oleic acid-albumin-dextrose-catalase) enrichment (100 mL/L) and glycerol (5 mL/L)). Cough spray droplets were able to impact naturally onto exposed 'settle plates' placed on the floor of the distance rig 750 mm from the patient (to detect large droplets). Zeihl-Neelsen stain was used to presumptively identify acid fast bacilli (AFB) organisms. Positive isolates were then subcultured (to confirm purity) and genomic DNA extracted (as above) from these samples as well as previously collected sputum and bronchoscopic samples and subjected to multiplexed paired-end sequencing on an Illumina Miseq platform.

Cough aerosols were also collected in parallel by an Optical particle counter *[35].* The patient produced a mean aerosol concentration of 2.2×10^6 particles / m^3 of diluted breath when tested using the distance rig, and 6.9×10^6 particles / m^3 of diluted breath when tested on the duration rig. In our cohort study of CF patients with chronic *Pseudomonas aeruginosa* infection, the interquartile range of aerosol concentrations for the distance and duration rigs were 1.3×10^6 - 1.4×10^7 and 1.1×10^6 and 2.9×10^7 particles / m^3 respectively.

Environmental surface sampling within hospital

Surface sampling was performed using sterile sponges in neutralizing buffer (Technical Consultants, TS/15-B). In the inpatient setting samples were taken from the bedside

table, bedside cabinet, bed head, and around the sink area. Sponges were then stored at 2-7 °C before processing, placed into an adapted concentrator (Amicon Pro Affinity Concentrator, exchange device and filter removed), inside a falcon tube and centrifuged for 15-30 mins at 3000 g. Extracted fluid was vortexed for 15 seconds to homogenise and 100µl cultured onto a Middlebrook 7H11 selective agar plate supplemented with PANTA. Plates were incubated at 30 ^{o}C and monitored for growth. Colonies with mycobacterial morphology were picked and stained with Kinyoun carbol fuchsin to determine if they were acid-fast and if so, were processed for DNA extraction as described above.

In vitro phenotyping of clustered and unclustered clinical *M. abscessus* isolates

A panel of representative clinical isolates (27 clustered and 17 unclustered *M. a. abscessus*; 25 clustered and 13 unclustered *M. a. massiliense*) were streaked on Middlebrook 7H11 agar supplemented with OADC (Becton Dickinson, UK) and single colonies were selected and grown in liquid media (Middlebrook 7H9 supplemented with ADC) and subjected to the following phenotypic assays: *Colony morphotyping:* Isolates were grown on CBA plates at 37°C with 5% CO_2 for 5-10 days before colony morphotypes were ascribed as smooth, rough, or intermediate/mixed as previously [17]; *Drug susceptibility testing:* Antibiotic susceptibilities were evaluated for each isolate grown in cation-adjusted Mueller-Hinton broth with TES (Thermo Scientific, UK) using RAPMYCO Sensititre plates (Trek Diagnostic Systems, UK) according to Clinical and Laboratory Standards Institute (CLSI) protocols as previously [17]; *Biofilm formation:* Isolates were grown in Middlebrook 7H9 supplemented with ADC, placed into MBEC Biofilm Inoculators (Innovotech, Canada), and grown for 5 days at 37°C and 210 rpm. Biofilm production was quantified using crystal violet according to manufacturer's instructions. Briefly pegs were washed with water, fixed with 99% methanol for 15 min, air-dried and stained with 2% crystal violet for 20 min. Pegs were then washed gently with running tap water and bound crystal violet released using 33% acetic acid and quantified by measuring absorbance at 590 nm; *Cytokine release.* Primary monocyte-derived human macrophages were generated as previously [50] from peripheral blood samples from consenting healthy volunteers. Cells were infected (at MOI of 10:1) with *M. abscessus* isolates and levels of TNFα and IL-8 in supernatants were measured as described previously [50] using Luminex; *Macrophage phagocytosis and intracellular survival:* Differentiated THP-1 cells [50] were incubated with *M. abscessus* isolates (at MOI of 10:1) at 37 ^{o}C for 2 h, washed and either immediately lysed with sterile water, scraped and plated onto CBA plates to enumerate uptake of mycobacteria through counting colony forming units (CFU), or incubated for 24 or 48h before lysis and CFU counting to quantify intracellular survival of bacteria [51]. Data points averages of at least three independent replicates performed in at least triplicate.

Multifactorial analysis of phenotypes

Multifactorial analysis (MFA) was performed with the FactoMineR package for R [52]. Prior to MFA execution the data set was cleansed of missing values and variables were transformed as follows. Antibiotic MICs were transformed to a factor of S, I, and R using appropriate cutoffs. Morphotype, which could be rough, smooth or mixed, codified into two binary variables Morpho_Rough and Morph_Smooth, setting both to 1 when a mixture was found. The analysis was run separately for *M. a. abscessus* and *M. a. massiliense,* and the data was labelled with its clustered/un-clustered status, although this wasn't considered as a variable for the MFA. Phenotypic variables were grouped in the following manner: intracellular survival: 24hrs and 48hrs; biofilm formation (single

group); morphotype: rough or smooth; cytokine production: TNFα and IL-8; antibiotic MIC: amikacin and clarithromycin; phagocytosis (single group). Other variables (such as MICs for other antibiotics) were not included as they did not vary within the data set.

Experimental infections in mice

Specific-pathogen-free female *SCID* mice, from 6 to 8 weeks old, were purchased from the Jackson Laboratories, Bar Harbor, Maine. Mice were maintained in the Biosafety Level III animal laboratory at Colorado State University, and were given sterile water, mouse chow, bedding and enrichment for the duration of the experiments. The specific pathogen-free nature of the mouse colonies was demonstrated by testing sentinel animals. All experimental protocols were approved by the Animal Care and Usage Committee of Colorado State University. The CSU animal assurance welfare number is A3572-01.

Ex vivo infection: Bone marrow-derived macrophages from *SCID* mice were infected as previously described [53] with mycobacteria at a macrophage per bacterial ratio of 1:1, and 2 hours later, the monolayers were washed to remove extracellular bacilli. The numbers of intracellular mycobacteria were measured by plating. Briefly, monolayers were washed at each time point to remove extracellular bacilli and 1 ml double-distilled H_2O containing 0.05% Tween 80 was added to monolayers and incubated for 10 min to lyse macrophages. After passing through a 26-gauge needle five times, the lysates were serially diluted and plated onto Middlebrook 7H11 agar plates. Monolayers that weren't lysed were replenished with fresh medium. BMDM cell death was assayed by trypan blue exclusion and determined by flow cytometry using BD™ Viability Counting Beads, as described by the manufacturer (BD PharMingen, San Jose, CA USA 95131) cell viability-staining methods [54].

In vivo infection: SCID mice were challenged with clustered and unclustered isolates of *M. a. abscessus* and *M. a. massiliense,* using an intravenous infection calibrated to deliver $1.0x10^7$ bacilli per animal. At days 1, 20 and 40 following infection, bacterial loads in the lungs, lung histology, and mononuclear and lymphocytic cellular expressions were determined as previously described [55]. Bacterial counts were determined by plating serial dilutions of homogenates of lungs on nutrient 7H11 agar and counting colony-forming units after 5-10 days incubation at 30^{o}C. A total of five animals were infected for each time point. The accessory lobe of the lung from each mouse was fixed with 10% formalin in phosphate buffered saline (PBS). Tissue sections were stained using haematoxylin and eosin and acid-fast stains as previously reported [56, 57]. Data are presented using the mean values from 5 mice per group and from values from replicate samples and duplicate or triplicate assays. The Student t-test test was used to assess statistical significance between groups of mice.

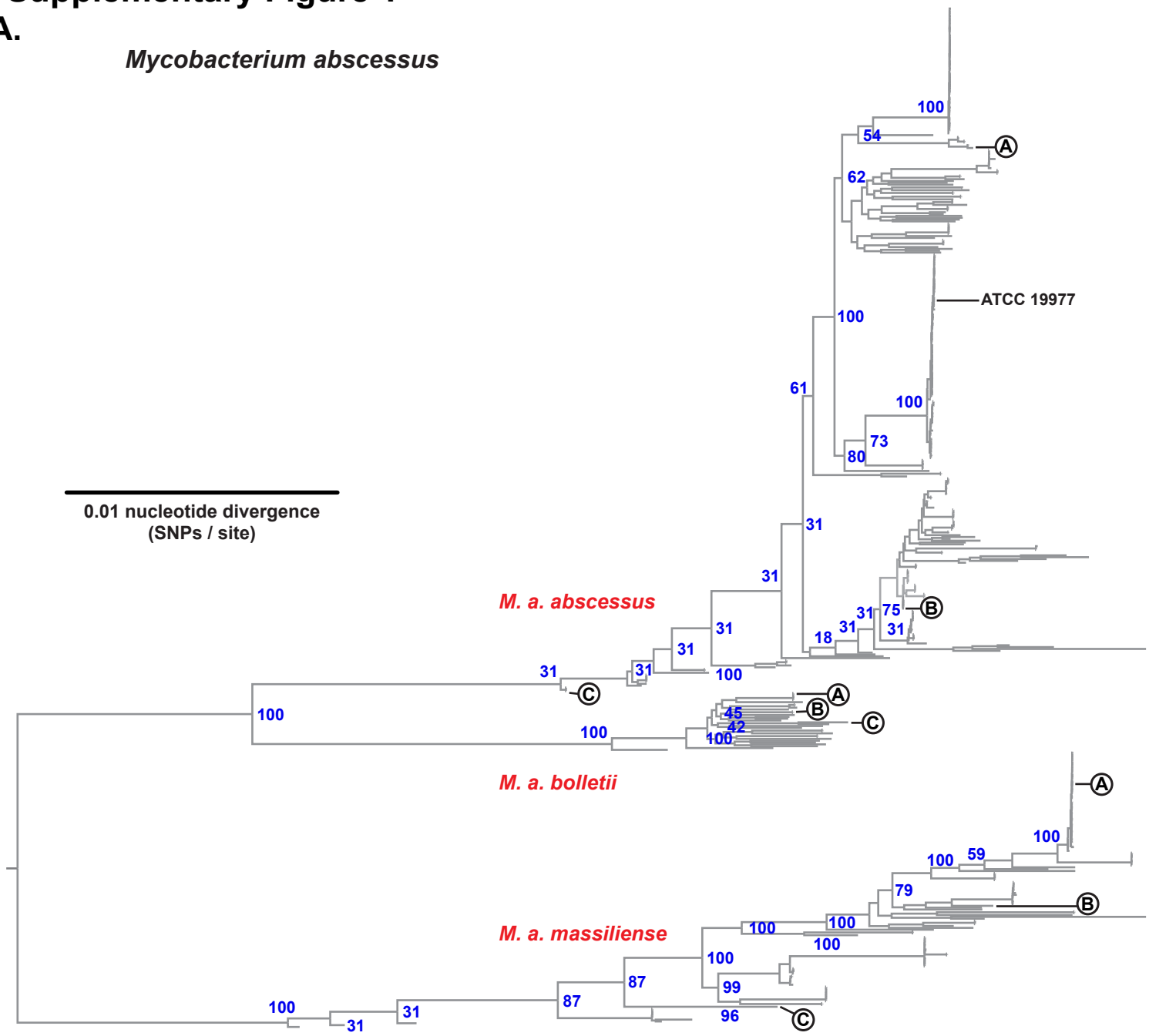
SUPPLEMENTARY REFERENCES

38. H. Ponstingl. SMALT, http://www.sanger.ac.uk/science/tools/smalt-0 (Wellcome Trust Sanger Institute, Hinxton), accessed. 2012.
39. F. Ripoll et al., Non mycobacterial virulence genes in the genome of the emerging pathogen Mycobacterium abscessus. *PloS one* **4**, e5660 (2009).
40. H. Li et al., The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079 (2009).
41. S. R. Harris et al., Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469-474 (2010).

42. A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).
43. M. Richter, R. Rosselló-Móra, Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* **106**, 19126-19131 (2009).
44. D. R. Zerbino, Using the Velvet de novo assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics* Chapter 11, Unit 11.15 (2010).
45. L. Cheng, T. R. Connor, J. Siren, D. M. Aanensen, J. Corander, Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol Biol Evol* **30**, 1224-1228 (2013).
46. S. Harris. (2016), https://github.com/simonrharris/tree_gubbins.
47. N. J. Croucher et al., Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **43**, e15 (2015).
48. D. A. J. Rambaut A, Tracer v1.5, Available from http://beast.bio.ed.ac.uk/Tracer. (2007).
49. C. Firth et al., Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol Biol Evol* **27**, 2038-2051 (2010).
50. L. Hepburn et al., Innate immunity. A Spaetzle-like role for nerve growth factor β in vertebrate immunity to Staphylococcus aureus. *Science* **346**, 641-646 (2014).
51. M. Schiebler et al., Functional drug screening reveals anticonvulsants as enhancers of mTOR-independent autophagic killing of Mycobacterium tuberculosis through inositol depletion. *EMBO Mol Med* **7**, 127-139 (2015).
52. S. Lê, J. Josse, F. Husson, FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software* **25**, 1-18 (2008).
53. S. Shang et al., Increased virulence of an epidemic strain of Mycobacterium massiliense in mice. *PLoS One* **6**, e24726 (2011).
54. D. Ordway et al., Animal model of Mycobacterium abscessus lung infection. *J Leukoc Biol* **83**, 1502-1511 (2008).
55. A. Obregón-Henao et al., Susceptibility of Mycobacterium abscessus to antimycobacterial drugs in preclinical models. *Antimicrob Agents Chemother* **59**, 6904-6912 (2015).
56. D. Ordway et al., Influence of Mycobacterium bovis BCG Vaccination on Cellular Immune Response of Guinea Pigs Challenged with Mycobacterium tuberculosis. *Cellul. Vaccine Immunol* **15**, 1248-1258 (2008).
57. D. Ordway et al., The hypervirulent Mycobacterium tuberculosis strain HN878 induces a potent TH1 response followed by rapid down-regulation. *J. Immunol*. **179.**, 522-531 (2007).
58. Y. Saito *et al.*, Characterization of endonuclease III (nth) and endonuclease VIII (nei) mutants of Escherichia coli K-12. *J Bacteriol* **179**, 3783-3785 (1997).

Supplementary Figure 1

A.



B.

	M. a. abscessus A	M. a. abscessus B	M. a. abscessus C	ATCC19977	M. a. bolletii A	M. a. bolletii B	M. a. bolletii C	M. a. massiliense A	M. a. massiliense B	M. a. massiliense C	M. chelonae
M. a. abscessus A	100	99.2	98.8	99.4	97.4	97.4	97.4	97.2	97.3	97.7	85.3
M. a. abscessus B	99.2	100	98.8	99.3	97.3	97.2	97.3	97.2	97.2	97.6	85.3
M. a. abscessus C	98.9	98.8	100	98.9	97.8	97.8	97.9	97.2	97.2	97.6	85.2
ATCC19977	99.3	99.3	98.8	100	97.3	97.3	97.3	97.3	97.2	97.7	85.3
M. a. bolletii A	97.4	97.2	97.8	97.3	100	99.4	99.4	96.8	96.9	97.1	85.2
M. a. bolletii B	97.4	97.2	97.8	97.3	99.4	100	99.5	96.7	96.9	97.2	85.2
M. a. bolletii C	97.4	97.3	97.9	97.4	99.4	99.5	100	96.9	96.9	97.2	85.2
M. a. massiliense A	97.2	97.1	97.1	97.3	96.8	96.7	96.9	100	99.3	98.5	85.2
M. a. massiliense B	97.2	97.2	97.2	97.2	96.8	96.9	96.9	99.4	100	98.7	85.2
M. a. massiliense C	97.7	97.6	97.6	97.7	97.1	97.2	97.2	98.5	98.7	100	85.3
M. chelonae	85.3	85.2	85.2	85.2	85.2	85.2	85.2	85.2	85.2	85.2	100

Fig. S1 Phylogeny and subspecies structure of global clinical isolates of *M. abscessus*
(A) Maximum likelihood phylogenetic tree of clinical isolates of M. abscessus collected from 517 patients (using one isolate per patient) in this study in addition to publically-available genomes (as shown in Figure 1), shown with natural branch lengths, Bootstrap values from 100 trees for major nodes (*blue*), and isolates used to calculate the average nucleotide identity in (B) (three representatives for each subspecies (A-C)and the ATCC reference isolate for *M. abscessus*). **(B)** Analysis of average nucleotide identity (ANI) between isolates from across the global phyolgenetic tree (A-C in (A)) and *M. chelonae* (ATCC 35752; accession NZ_CP010946.1) indicate that all *M. abscessus* isolates have >95% ANI (confirming that they belong to the same species). ANI values between subspecies (96-97%) and within subspecies (>98%) are consistent with the presence of three distinct subspecies which, given their different clinical trajectories (at least for non-CF individuals), is likely to be an important reclassification.

Supplementary Figure 2

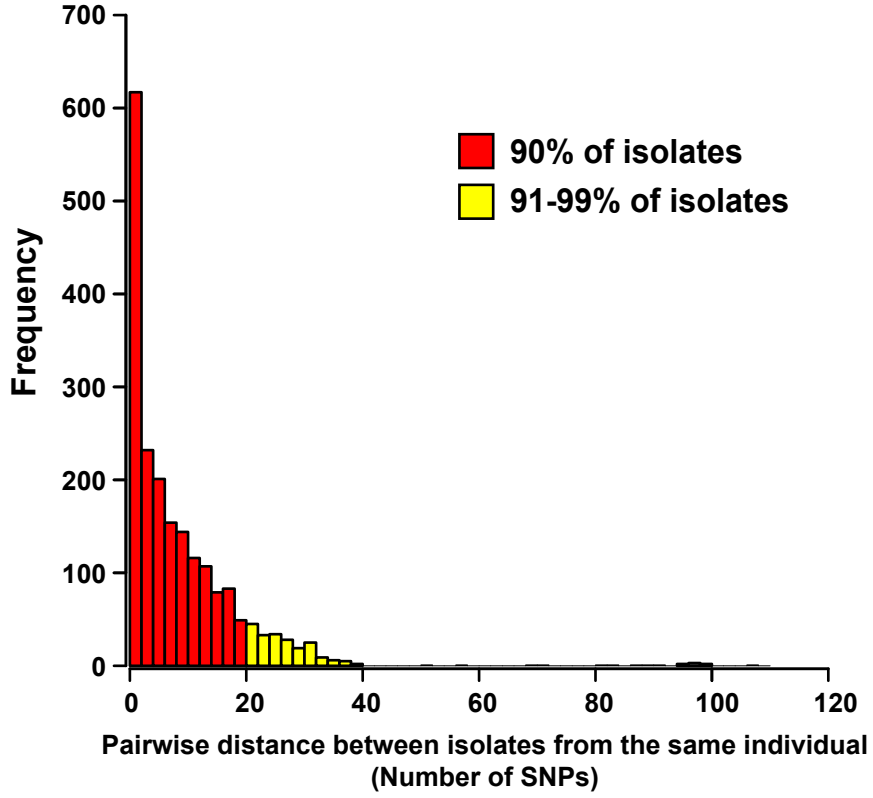


Fig. S2
Genetic variation within individuals infected with *M. abscessus*.
Differences in whole genome sequences, calculated as pairwise distances (in numbers of single nucleotide polymorphisms; SNPs) of isolates of *M. abscessus* from the same individual. 90% of isolates had less that 20 SNPs difference (*red*) while 99% of isolates had less that 38 SNPs (*yellow*). Analysis based on whole genome sequencing of 864 *M. abscessus* isolates from 202 infected individuals.

Supplementary Figure 3

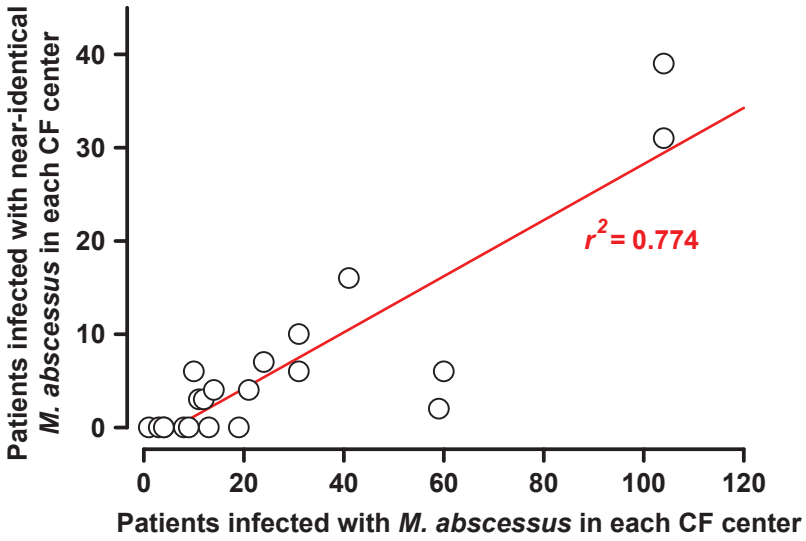


Fig. S3.
Relationship between the numbers of patients with near-identical (<20 SNPs) *M. abscessus* isolates in each CF center and the total numbers of individuals infected with *M. abscessus*. Linear regression (*red*).

Supplementary Figure 4

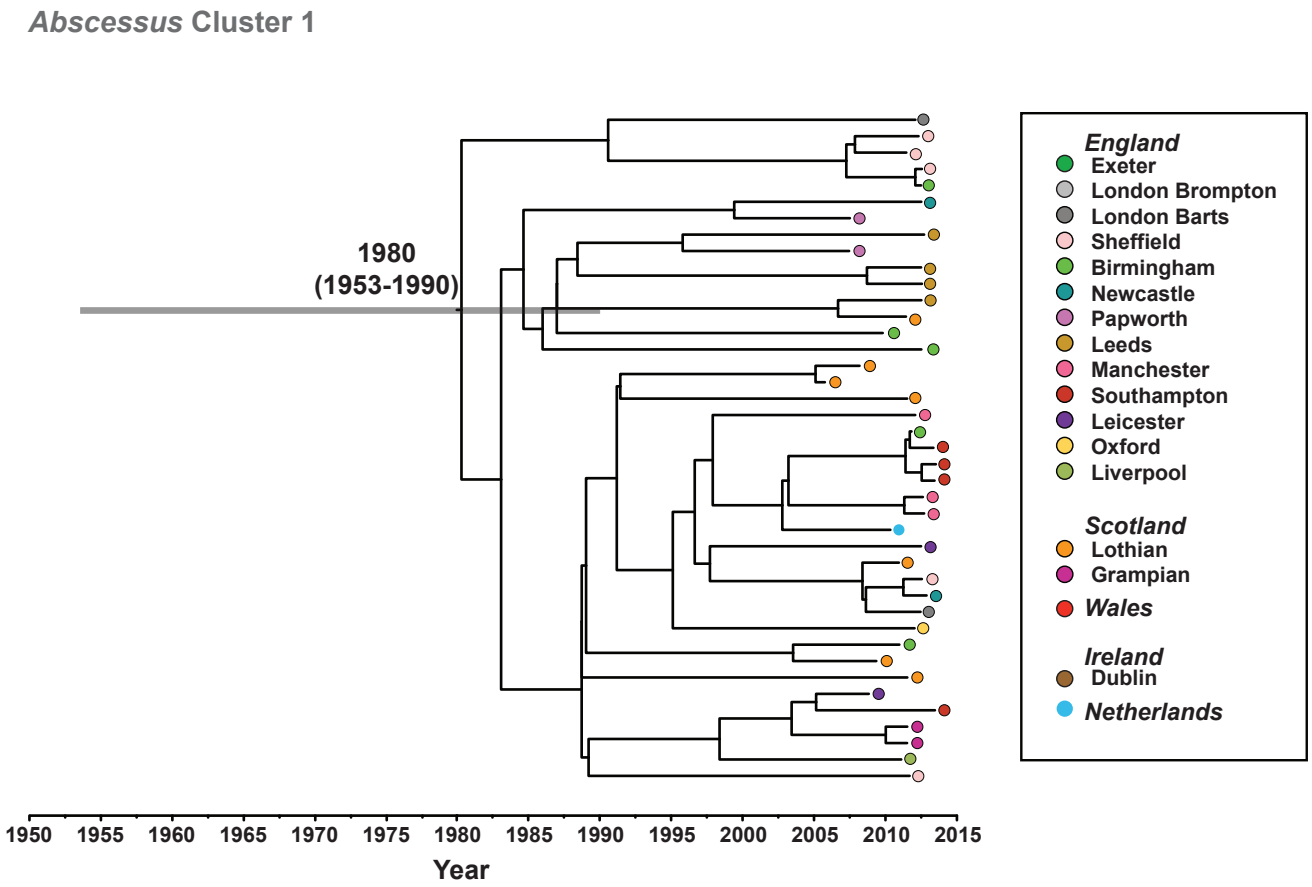


Fig. S4 Maximum clade credibility tree of *M. a. abscessus* cluster 1, restricted to one isolate per patient where isolation date was available. The tree was produced by TreeAnnotator in the BEAST package using 9,000 trees (excluding 10% burn-in) on a single BEAST run using the strict clock and constant population models. The mean age and 95% higher probability densities of the most recent common ancestor are labelled.

Supplementary Figure 5

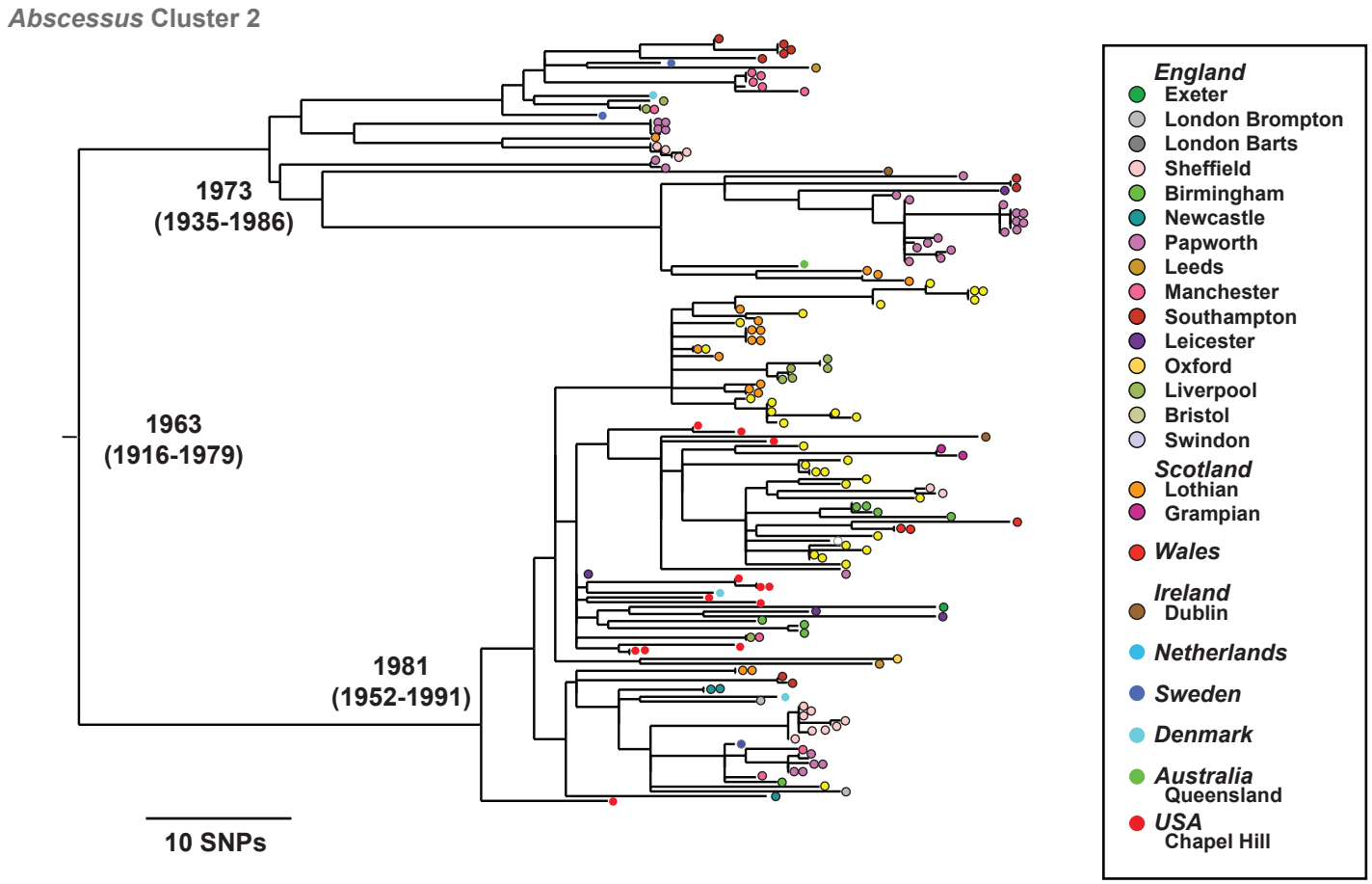


Fig. S5 Dating of *M. a. abscessus* Cluster 2

Maximum likelihood phylogeny of *M. a. abscessus* Cluster 2. Recombination was removed and total tree depth calculated. The age and 95% error of the most recent common ancestor (MRCA) were estimated using the substitution rate inferred from a Bayesian analysis of *M. a. abscessus* cluster 1 (Figure S4).

Supplementary Figure 6

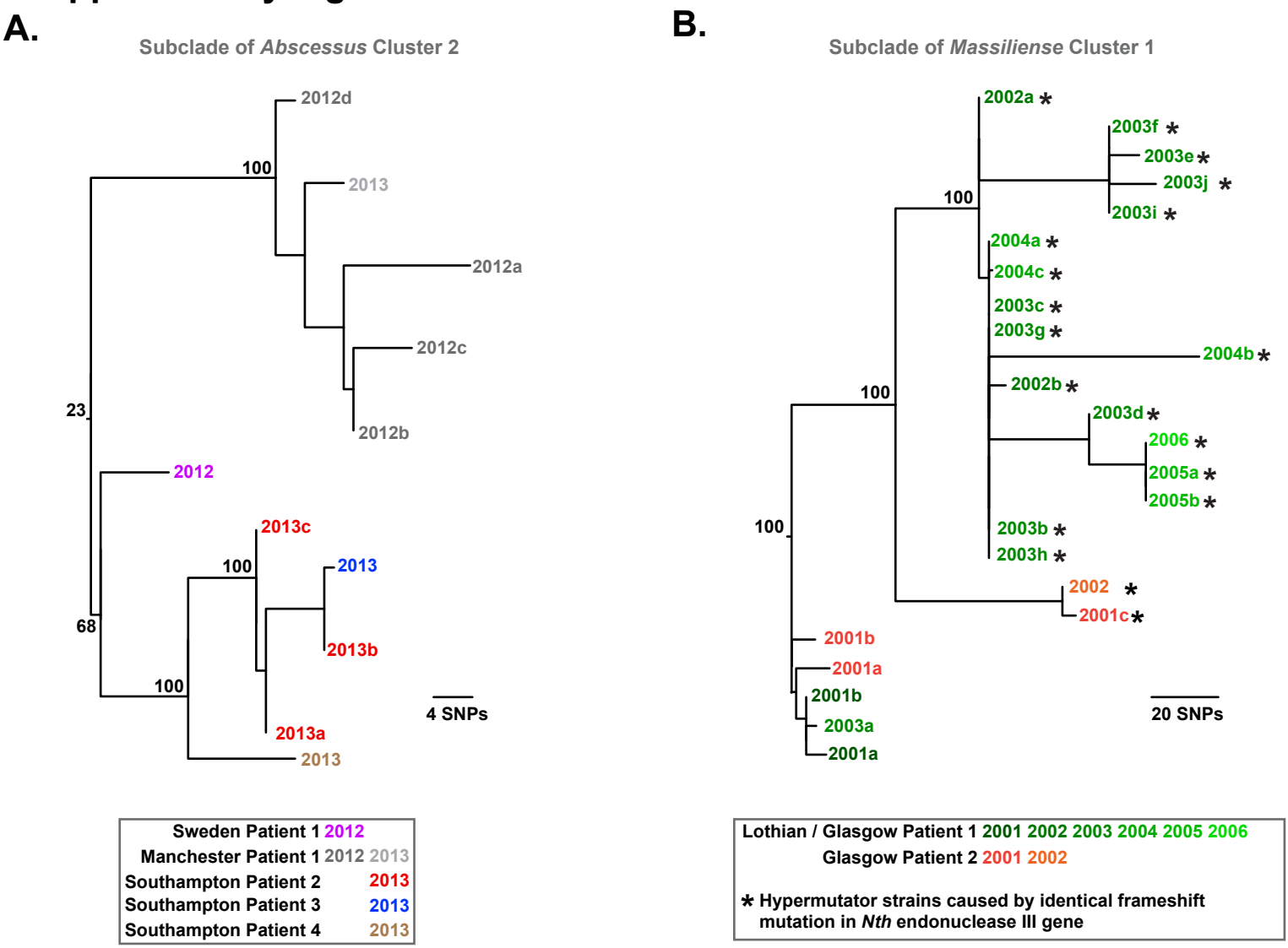


Fig. S6 Examples of paraphyletic relationships of *M. abscessus* isolates from different patients
Whole genome sequencing of longitudinal samples collected from over 50 individuals with CF revealed multiple examples (two of which are shown here) of paraphyletic relationships (nesting of one patient's sampled diversity within another) between isolates from different patients; a pattern strongly indicative of patient-to-patient transmission *[25]*. The maximum likelihood trees were constructed with 100 bootstrap replicates (support for the major nodes is labelled). (**A**) Paraphyly observed for Southampton Patients 2 and 3. Closely related isolates are shown for context, and demonstrate that we observe similar levels of diversity within a single patient (Manchester Patient 1) as we do between patients. (**B**) Paraphyly seen in two patients from Scotland. Greater diversity was observed between isolates, probably due to acquisition of a frameshift mutation in the *Nth* endonuclease III gene, known to cause a hypermutator phenotype *[58]*. Both hypermutator and non-hypermutator strains were found in Patients 1 and 2, suggesting that either (i) multiple transmission events had occurred or (ii) that a single infecting inoculum was sufficiently large to transmit this genetic heterogeneity.

Supplementary Figure 7

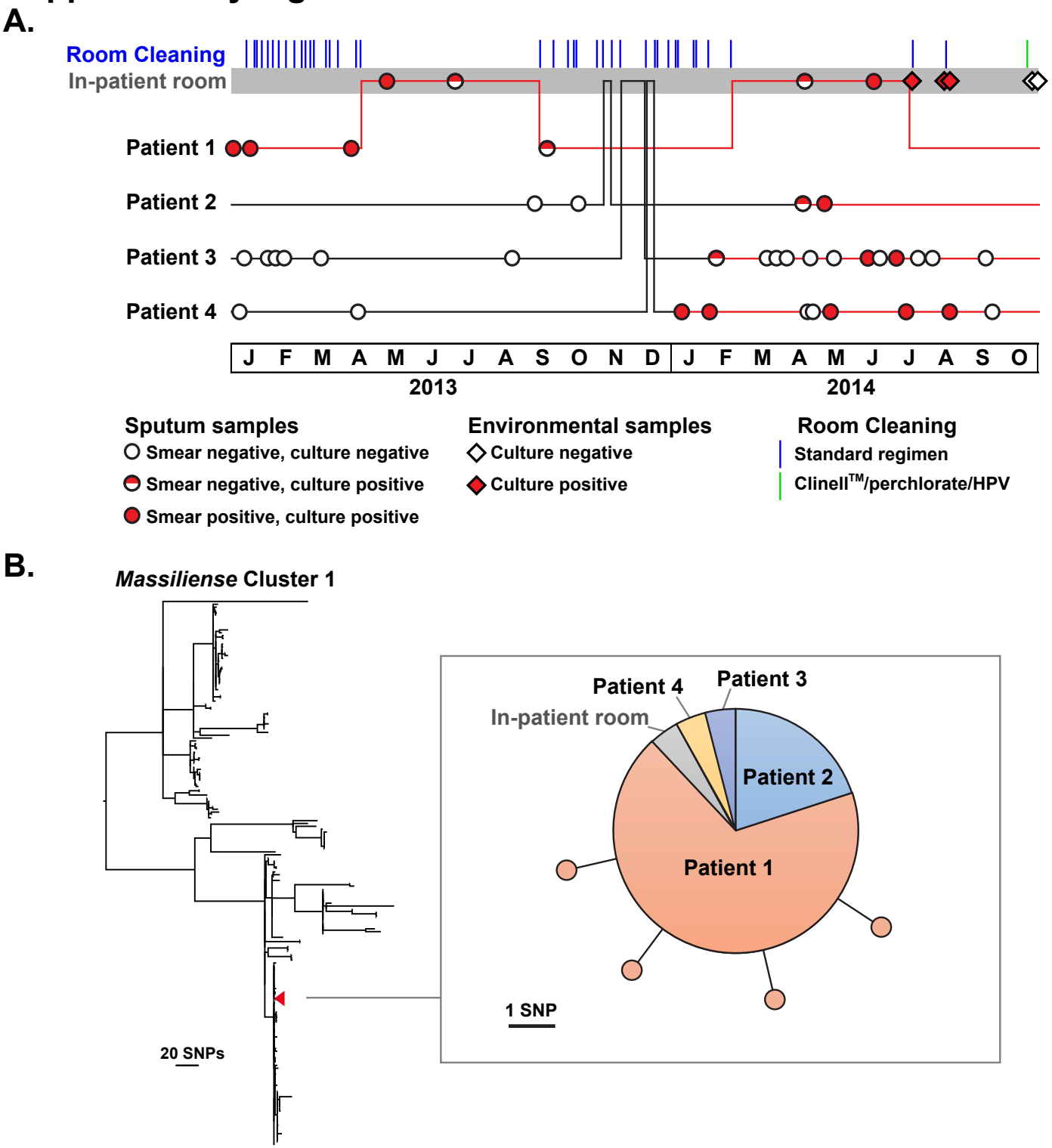
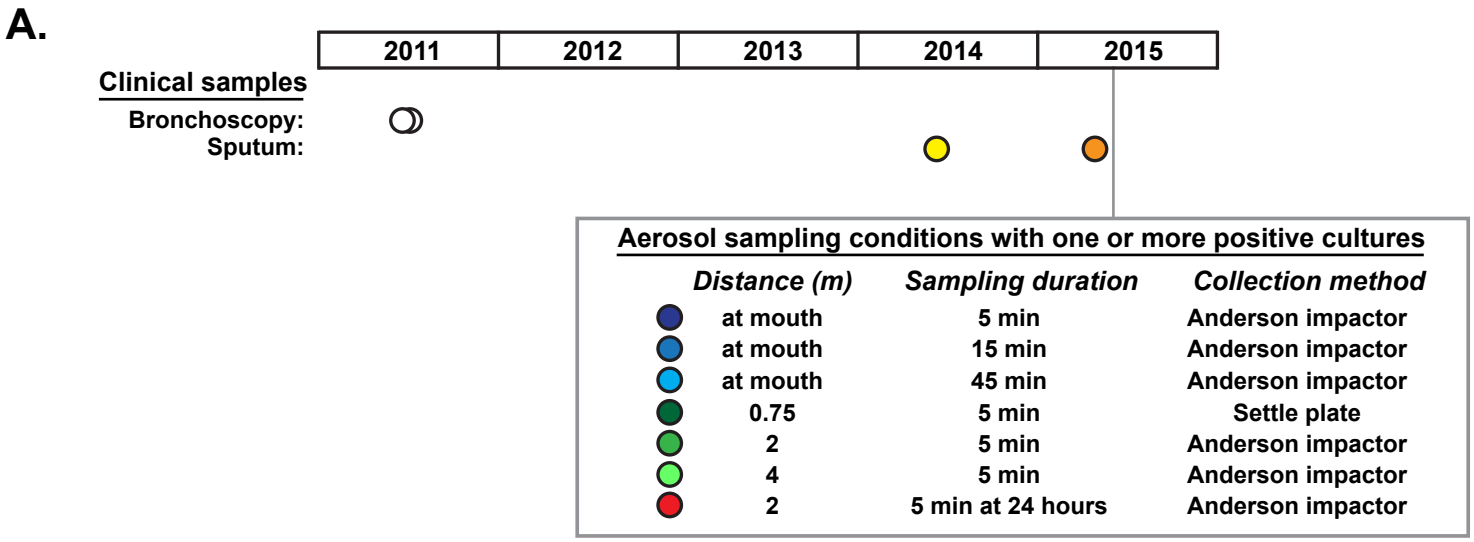


Fig. S7 Example of within-hospital Person-to-Person transmission of *M. abscessus* potentially through fomite spread. (A) Timelines (from January 2013) for four CF patients exposed at different times to the same inpatient room (*grey bar*). The index case (Patient 1), who was chronically smear- and culture-positive for *M. abscessus,* had a prolonged admission to the room (2013). The same room was subsequently occupied by Patients 2, 3, and 4 and represents the only environment shared with Patient 1 (though never contemporaneously). Each patient subsequently became sputum culture-positive for *M. abscessus*. The room was subjected to multiple standard cleaning regimens (indicated by short blue vertical bars on the time line) involving changing curtains and cleaning walls, floors and all surfaces with a combined detergent and chlorine-based product (Actichlor Plus™). Environmental sampling (at the end of Patient 1's second prolonged admission in 2014, *diamonds*) using surface swabs before and after standard cleaning, detected *M abscessus* from sinks and surface rails. Subsequent environmental samples after cleaning the room with Clinell™ wipes, concentrated perchlorate solution (for sink areas), and hydrogen peroxide vapour (HPV; *green bar*) were all negative for *M. abscessus*. Patient timelines change from black to red once *M. abscessus* infection detected. **(B)** All Patient and environmental isolates were identical and part of a global dominant circulating clone (*Massiliense* Cluster 1) and therefore did not originate from the hospital environment. Phylogenetic position of patient and environmental isolates in (A) within *Massiliense* Cluster 1 indicated by red arrowhead.This single monophyletic clade is represented as a minimum spanning tree (*insert*), where the size of each circle represents the number of identical isolates, and all had mutational amikacin (16S) and macrolide (23S) resistance, further supporting a patient rather than environmental source for the outbreak.

Supplementary Figure 8



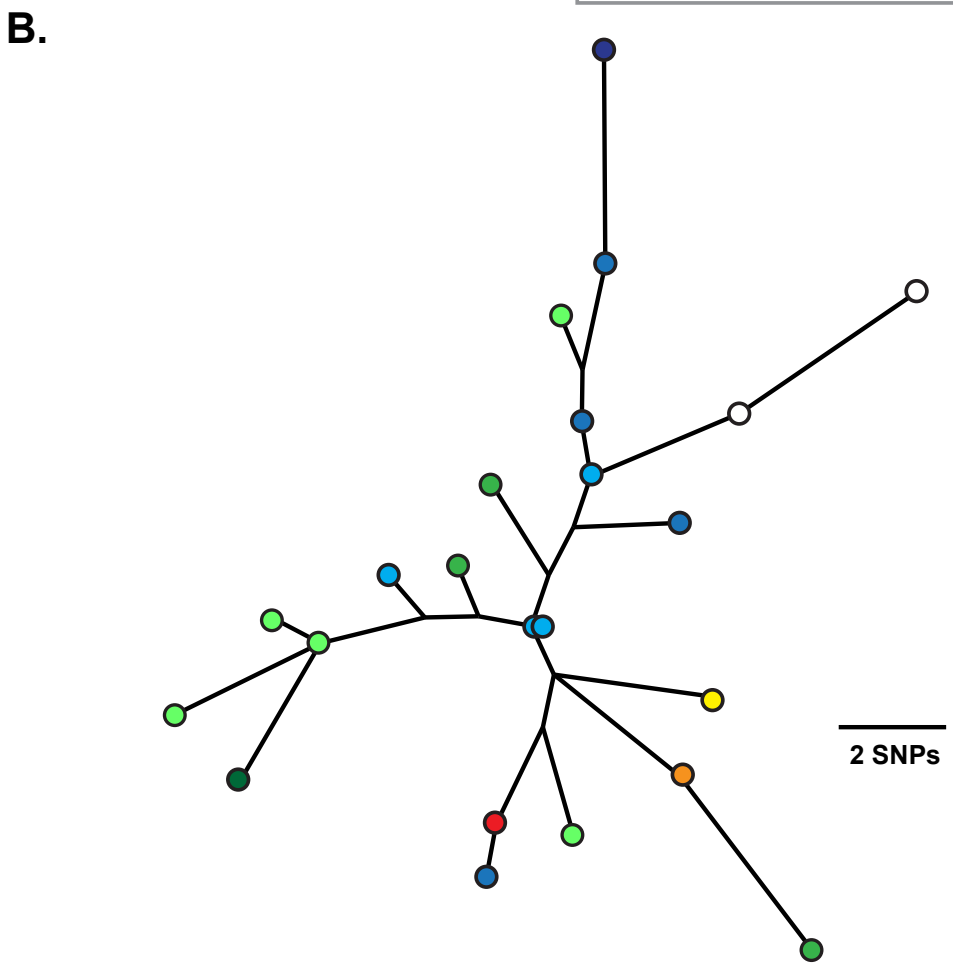


Fig. S8

Detection of viable *M. abscessus* in cough bio-aerosols and droplets generated by an individual with CF. To examine whether long-lived infectious aerosols might represent a possible mode by which *M. abscessus* could transmit between individuals with CF, we analysed the aerosols and droplets produced during coughing by a consented CF individual who was chronically infected (smear-positive) with *M. abscessus*. Both distance travelled and viability over time were evaluated using distance and duration rigs incorporating Andersen cascade impactors (detecting airborne droplet nuclei) and settle plates (detecting droplet spray) as previously described *[35]* (see *Supplementary Methods* for experimental details). **(A)** Longitudinal samples of *M. a. massiliense* isolated by bronchoscopy (*white*), from sputum (*yellow, orange*), and conditions during cough analysis yielding at least 1 colony (*grey box*). The majority of mycobacterial aerosol particles detected were less than 4.7 μm diameter and therefore respiratable. The positive sample detected after 24h is likely to have come from a fomite source within the rig created during the initial cough testing that became airborne, since the equipment was thoroughly flushed with filtered air before testing. **(B)** Maximum likelihood phylogeny of whole genome sequences of isolates recovered from the aerosol experiment described in (A). The tree was constructed using variants called by mapping against the *M. a. massiliense* reference sequence *[17]*. Isolates were tightly clustered (each colony varying by 4 or less SNPs from any other) and genetically distinct (at least 14,161 SNPs) from all other sequenced patient or environment isolates, implying a unique patient source. There was no observable relationship of samples within the cluster with respect to the distance and time collected.

Supplementary Figure 9

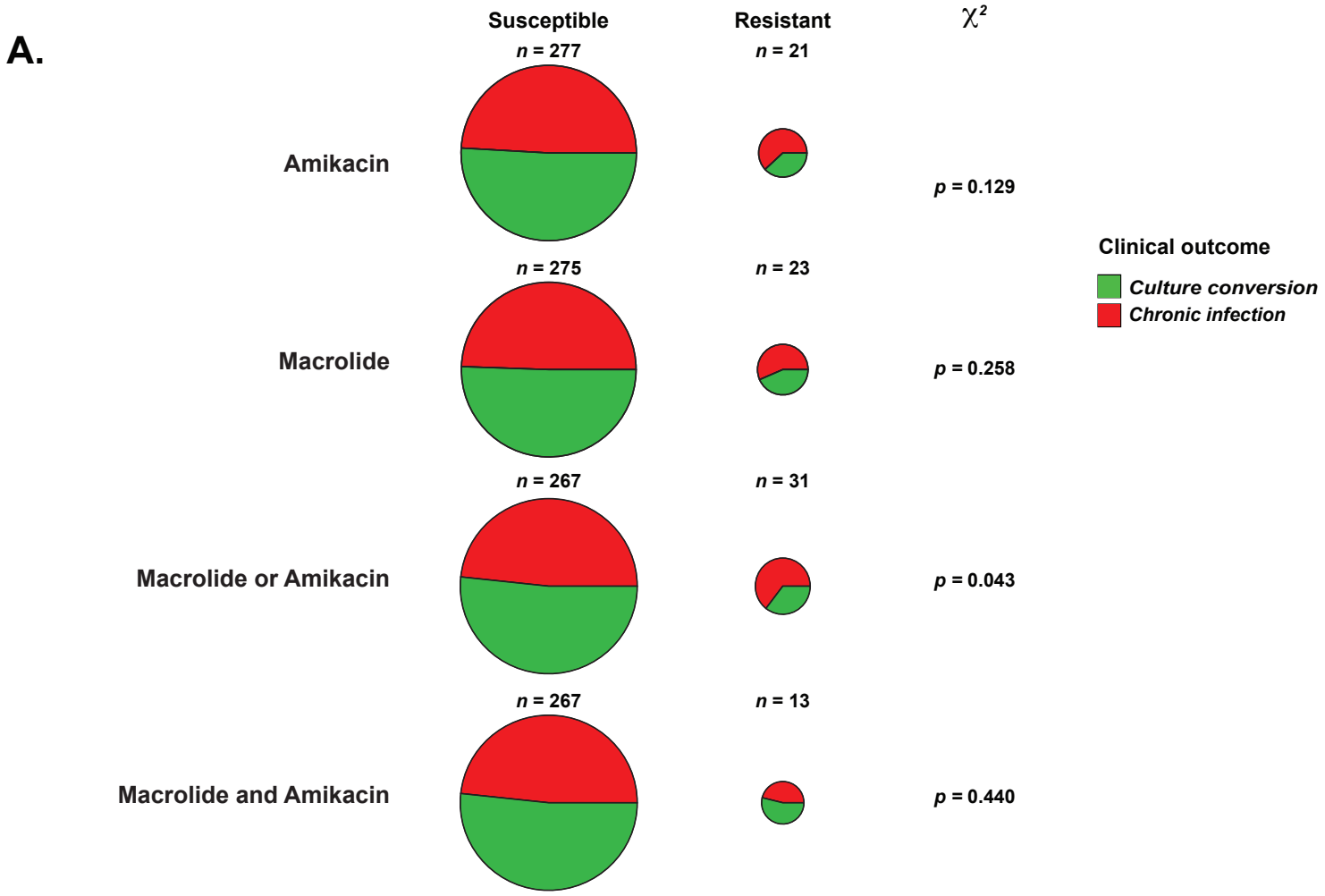


Fig. S9 Comparison of clinical outcomes for CF individuals infected with *M. abscessus* isolates based on genetic clustering or drug resistance of isolates.

A,B. Clinical outcomes (culture-conversion, *green*; chronic infection, *red*) of individuals infected with isolates that were (**A**) amikacin-resistant (with 16S rRNA mutations; *top*) or macrolide-resistant (with 23S rRNA mutations; *middle*), or either amikacin or macrolide resistance (*bottom*) irrespective of whether they were genetically clustered or unclustered or (**B**) unclustered or clustered isolates of *M. abscessus* classified as macrolide and amikacin susceptible (no 23S or 16S mutation, *top*) or macrolide susceptible (no 23S mutation, *bottom*). All comparisons are shown since many patients infected with amikacin-resistant isolates were never treated with amikacin. Uncorrected Chi square test (one tailed) used for comparisons.

Supplementary Figure 10

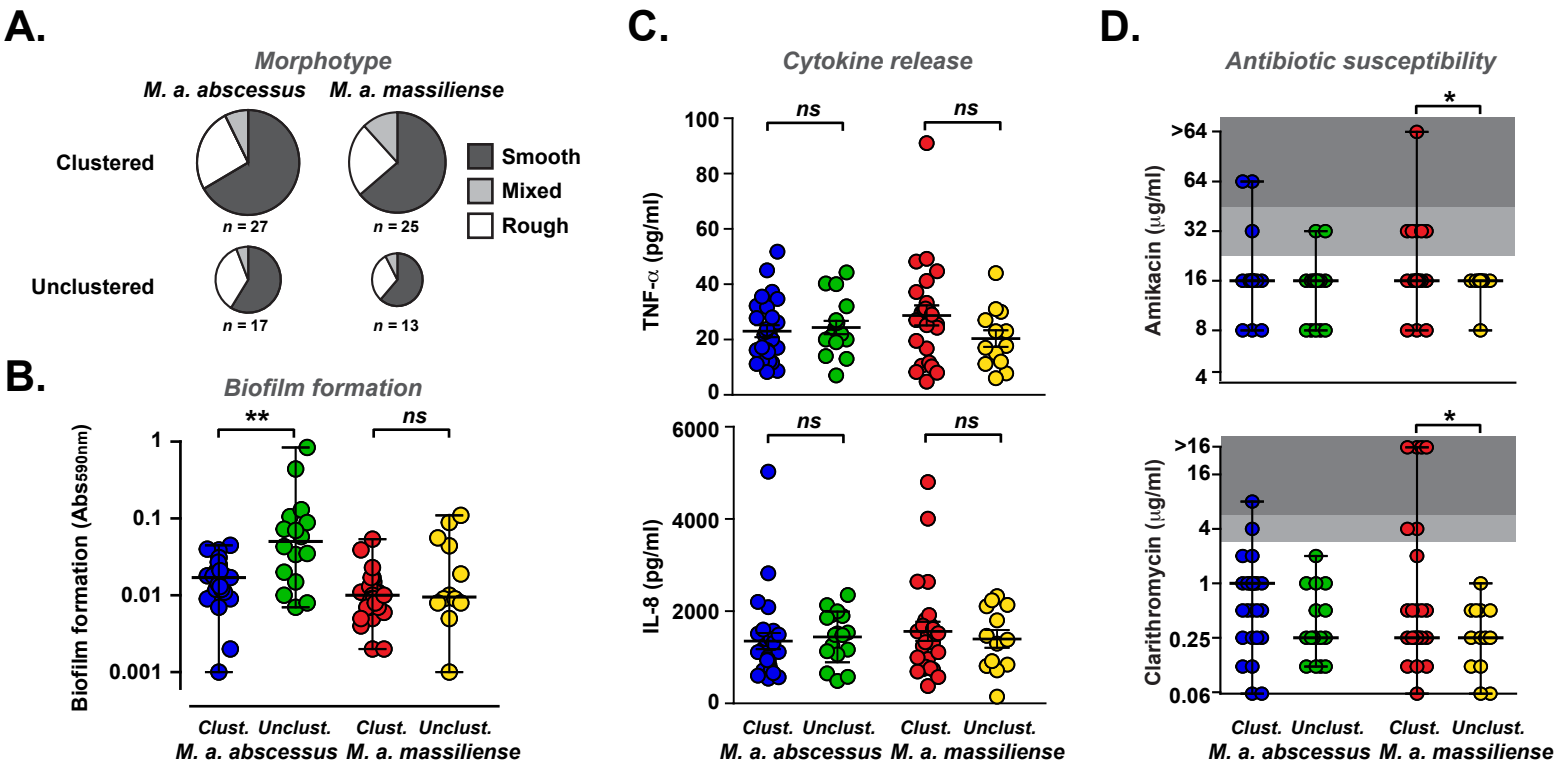


Fig. S10.
Phenotypic analysis of representative isolates of clustered and unclustered *M. abscessus* isolates. *In vitro* evaluation of representative clustered and unclustered isolates (27 clustered (*blue*) and 17 unclustered (*green*) *M. a. abscessus*; 25 clustered (*red*) and 13 unclustered (*yellow*) *M. a. massiliense*) for (A) morphotype, (B) Biofilm formation, (C) TNF-α and IL-8 release from infected primary human macrophages, and (D) antibiotic susceptibility (measured by broth microdilution) to amikacin and clarithromycin. Data points averages of at least three independent replicates. * $p < 0.05$; ** $p < 0.005$.

Supplementary Figure 11

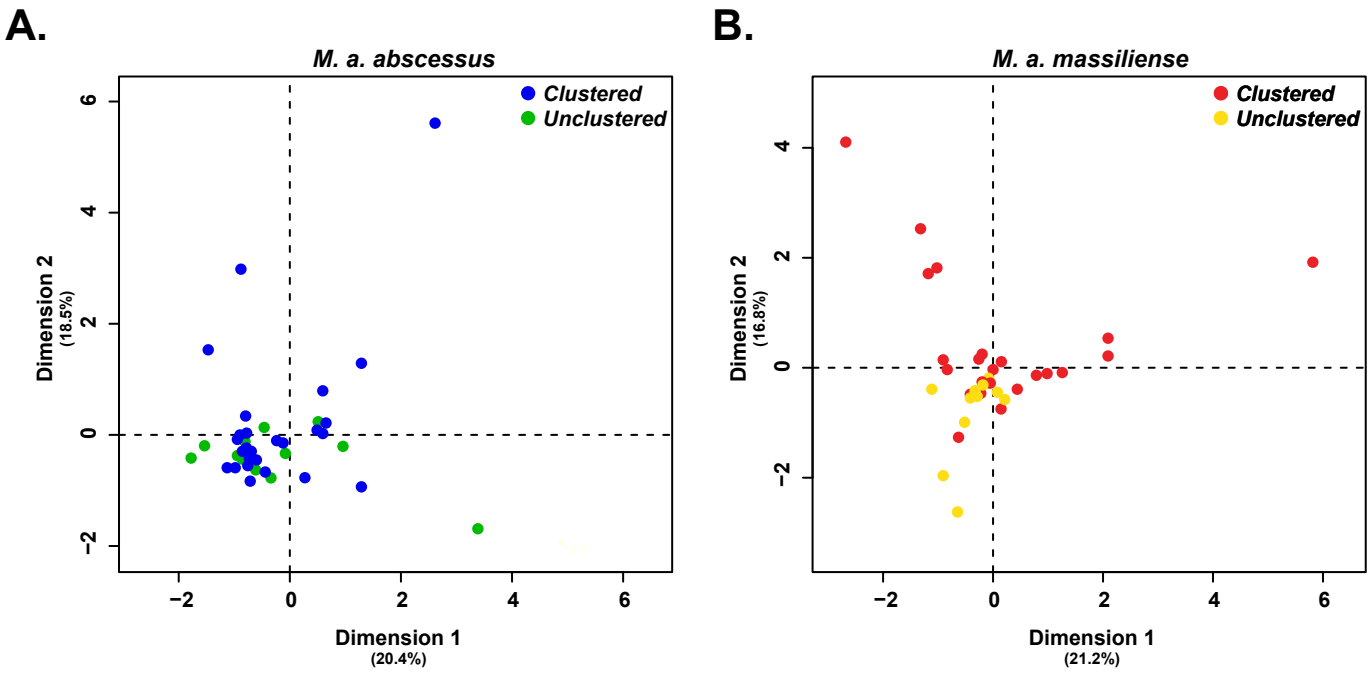


Fig. S11
Multifactorial analysis (incorporating morphotype, biofilm production, antibiotic resistance, phagocytic potential, intracellular survival, cytokine stimulation data for each isolate) of clustered (*blue*) and unclustered (*green*) *M. a. abscessus* (*left*), and clustered (*red*) and unclustered (*yellow*) *M. a. massiliense* (*right*) revealed no obvious segregation between groups.

Supplementary Figure 12

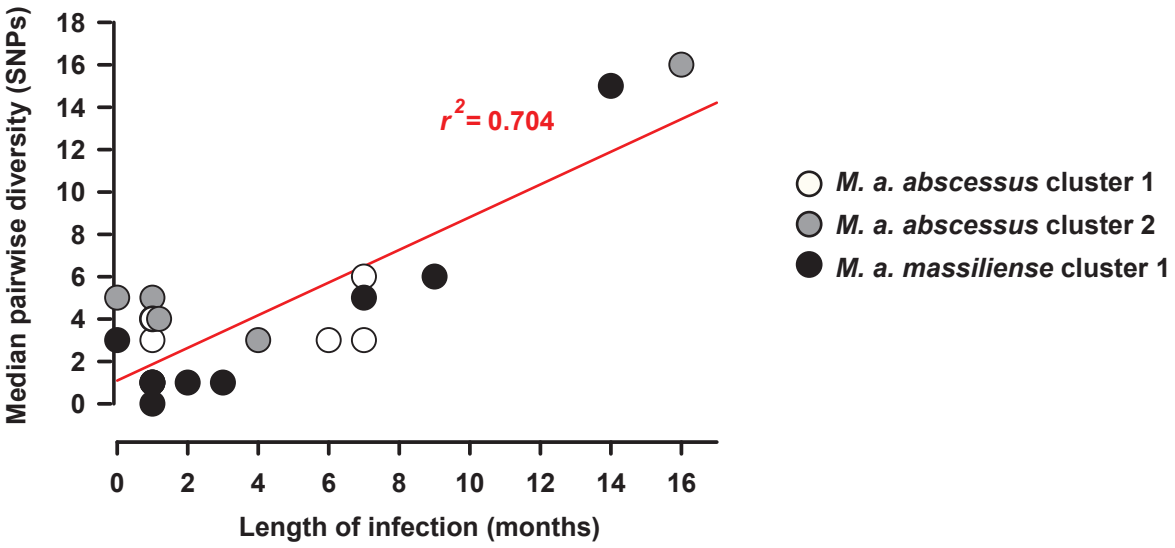


Fig. S12
Relationship of median pairwise diversity (measured in single nucleotide polymorphisms; SNPs) of the three main dominant circulating clones *M. a. abscessus* cluster 1 (*white*) and cluster 2 (*grey*) and *M. a. massiliense* (*black*) with the duration of infection of an individual patient (in months).

M. a. abscessus		
Reference	Citation	Figure 1
M93	Choo SW et al. **Genome sequence of the *Mycobacterium abscessus* strain M93.** *J Bacteriol* 2012, **194**:3278	
M94	Choo SW et al. **Analysis of the genome of *Mycobacterium abscessus* strain M94 reveals an uncommon cluster or tRNAs.** *J Bacteriol* 2012, **194**:5724	
M152	Ngeow YF et al. **Genome sequence of *Mycobacterium abscessus* strain M152.** *J Bacteriol* 2012, **194**:6662	
4S 0116 R	Davidson RM et al. **Genome sequencing of *Mycobacterium abscessus* isolates from patients in the united states and comparisons to globally**	
6G 0125 R	**diverse clinical strains.** *J Clin Microbiol* 2014, **52** (10):3573-82	
3A 0119 R		
9808		
ATCC 19977T	Ripoll F et al. **Non mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*.** *PLoS One* 2009	
V06705	Pang S et al. **Whole-genome sequencing of *Mycobacterium abscessus* clinical strain V06705.** *Genome Announc.* 2013, **1**(5)e0069-13	
CF	Pawlik A et al. **Identification and characterization of the genetic changes responsible for the characteristic smooth-to-rough morphotype**	
	alterations of clinically persistent *Mycobacterium abscessus*. *Mol Microbiol* 2013, **90**(3):612-29	

M. a. massiliense		
Reference	Citation	Figure 1
M18	Ngeow YF et al. **Genome sequence of *Mycobacterium massiliense* M18, isolated from a lymph node biopsy specimen.** *J Bacteriol* 2012, **194**(15):4125	
M115	Ngeow YF et al. **Genomic analysis of *Mycobacterium massiliense* strain M115, an isolate from human sputum.** *J Bacteriol* 2012, **194**:4786	
M139	Ngeow YF et al. Genomic analysis of Mycobacterium abscessus strain M139, which has an ambiguous subspecies taxonomic position. J Bacteriol 2012, 194:6002-6003	
M148	Choo SW et al. **Genomic reconnaissance of clinical isolates of emerging human pathogen *Mycobacterium abscessus* reveals high evolutionary**	
M154	**potential.** *Sci Rep* 2014, **4**: 4061	
	Choo SW et al. Annotated genome sequence of Mycobacterium massiliense strain M154, belonging to the recently created taxon Mycobacterium abscessus subsp. bolletii comb. nov. J Bacteriol 2012, 194:4778	
M156	Choo SW et al. **Whole-genome shotgun sequencing of *Mycobacterium abscessus* M156, an emerging clinical pathogen in Malaysia.** *Genome Announc*	
M172	2013, **1**(1) pii: e00063-12	
	Choo SW et al. **Genome analysis of *Mycobacterium massiliense* strain M172, which contains a putative mycobacteriophage.** *J Bacteriol* 2012, **194**(18):5128	
CCUG48898	Tettelin H et al. **Genomic insights into the emerging human pathogen *Mycobacterium massiliense*.** *J Bacteriol* 2012, **194**:5450	
47J26	Chan J et al. **Whole-genome sequencing of the emerging pathogen *Mycobacterium abscessus* strain 47J26.** *J Bacteriol* 2012 **194**:549	
MAB 091912 2446	Tettelin H et al. **High-level relatedness among *Mycobacterium abscessus* subsp. *massiliense* strains from widely separated outbreaks.** *Emerg Infect*	
MAB 082312 2258	*Dis* 2014, **20**(3):364-371	
MAB 2b 0107		
1S 151 930		
5S 0817		
Asan 50594	Kim BJ et al. **Complete genome sequence of *Mycobacterium massiliense* clinical strain Asan 50594, belonging to the Type II genotype.** *Genome*	
	Announc 2013, **1**(4) pii: e00429-13	
CRM0020	Davidson R et al. **Phylogenomics of Brazilian epidemic isolates of *Mycobacterium abscessus* subsp. *bolletii* reveals relationships of global**	
	outbreak strains. *Infect Genet Evol* 2013, **20**:292-7	
GO-06	Raiol T et al. **Complete genome sequence of *Mycobacterium massiliense*.** *J Bacteriol* 2012, **194**:5455	

M. a. bolletii		
Reference	Citation	Figure 1
CIP108541	Choi GE et al. **Draft genome sequence of *Mycobacterium abscessus* subsp. *bolletii* BD (T).** *J Bacteriol* 2012, **194**:2756-2757	
M24	Wong Y et al. **Draft genome sequence of *Mycobacterium bolletii* Strain M24, a rapidly growing Mycobacterium of contentious taxonomic status.** *J*	
	Bacteriol 2013, **194**(16): 4475	

Table S1.

Additional *M. abscessus* whole genome sequences (with identifier, citation and symbol) used to create maximum likelihood phylogenetic tree of global isolates in Figure 1.

Supplementary Table 2

Cluster	Number of patients	One CF center	One country	International
Absc. 1	119			■
Absc. 2	64			■
Absc. 3	11			■
Absc. 4	7	■		
Absc. 5	7	■		
Absc. 6	5		■	
Absc. 7	28			■
Absc. 8	6			■
Absc. 9	5		■	
Absc. 10	5		■	
Absc. 11	4		■	
Absc. 12	5			■
Absc. 13	8			■
Absc. 14	3	■		
Absc. 15	3			■
Absc. 16	2	■		
Absc. 17	12			■
Absc. 18	4			■
Absc. 19	2	■		
Absc. 20	2	■		
Absc. 21	7			■
Absc. 22	3			■
Absc. 23	3	■		
Absc. 24	6			■
Absc. 25	2	■		
Absc. 26	2			■
Absc. 27	2		■	
Absc. 28	2	■		
Boll. 1	5			■
Boll. 2	2		■	
Boll. 3	2	■		
Boll. 4	2	■		
Mass. 1	48			■
Mass. 2	9		■	
Mass. 3	7			■
Mass. 4	16			■
Mass. 5	9			■
Mass. 6	4			■
Mass. 7	4		■	
Mass. 8	6		■	
Mass. 9	4		■	
Mass. 10	2	■		
Mass. 11	3	■		
Mass. 12	3	■		
Mass. 13	2	■		
Mass. 14	4			■

Table S2. Numbers of CF individuals infected with *M. abscessus* clusters and their geographical distribution. Clusters of *M. abscessus* subspecies *M. a. abscessus* (Absc.), *M. a. massiliense* (Mass.), and *M. a. bolletii* (Boll.) were identified through heirarchical branch density analysis using TreeGubbins (see *Supplementary Methods; Figure 2*). The number of CF individuals infected by each cluster is shown together with their geographical distribution (black boxes indicate clusters that are limited to one CF center, one country, or have spread internationally).