Whole Genome Sequencing of Nontuberculous Mycobacterium (NTM) Isolates from Sputum and Environmental Specimens in Co-Habiting Patients with NTM Pulmonary Disease

CURRENT STATUS: UNDER REVIEW

BMC Genomics

Jung-Ki Yoon
Seoul National University College of Medicine

Taek Soo Kim
Seoul National University Hospital

Jong-II Kim
Seoul National University Graduate School

Jae-Joon Yim
Seoul National University College of Medicine

yimjj@snu.ac.kr

Corresponding Author

DOI:
10.21203/rs.2.17722/v1

SUBJECT AREAS
Epigenetics & Genomics
Pulmonology

KEYWORDS
Non-tuberculosis mycobacterium, whole genome sequencing, transmission, non-tuberculous mycobacterial pulmonary disease, phylogenomics
Abstract
Background: Nontuberculous mycobacterium (NTM) species are ubiquitous microorganisms. NTM pulmonary disease (NTM-PD) is caused not by human-to-human transmission but by independent environmental acquisition. However, recent studies using next-generation sequencing (NGS) have reported trans-continental spread of Mycobacterium abscessus among patients with cystic fibrosis.

Results: We investigated NTM genomes through NGS to examine transmission patterns in three pairs of co-habiting NTM-PD patients who were suspected of patient-to-patient transmission. Three pairs of patients with NTM-PD co-habiting for at least 15 years were enrolled: a mother and a daughter with M. avium PD, a couple with M. intracellulare PD, and a second couple, one of whom was infected with M. intracellulare PD and the other of whom was infected with M. abscessus subsp. massiliense PD. Whole genome sequencing was performed using NTM colonies isolated from patients and environmental specimens. Genetic distances were estimated based on single nucleotide polymorphisms (SNPs) in the NTM genomes. Comparing SNPs in the consensus regions, the minimum pairwise SNP distances of NTM isolates derived from the two pairs of patients infected with the same NTM species were over 10,000. In phylogenetic analysis, the NTM isolates from patients with M. avium PD clustered with isolates from different environmental sources.

Conclusions: In conclusion, considering the genetic distances between NTM strains, the likelihood of patient-to-patient transmission in pairs of co-habiting NTM-PD patients without overt immune deficiency is minimal.

Background
The prevalence of nontuberculous mycobacterial pulmonary disease (NTM-PD) is increasing in developed countries (1–4). Several explanations for this epidemiological change have been proposed, including awareness and improved detection of NTM-PD, increased populations with risk factors such as bronchiectasis or use of immunosuppressants, and disinfection of drinking water in urban areas resulting in selective advantages for NTM (1).

The development of next-generation sequencing (NGS) technology enables the identification of massive numbers of single nucleotide polymorphisms (SNPs) by whole genome sequencing (WGS).
Using SNPs as genetic fingerprints and comparing them among multiple samples, phylogenetic analysis has been able to identify the source of infection for pathogens such as *Vibrio cholera* (5), *Staphylococcus aureus* (6), *Pseudomonas aeruginosa* (7), and NTM (8–11).

Since NTM is a ubiquitous microorganism, it is generally assumed that all patients with NTM-PD acquire NTM from their environment, not from other infected individuals. However, recent NGS studies showed that this might not be the case, at least for patients with cystic fibrosis (8, 9). Bryant and colleagues collected *Mycobacterium abscessus* isolates from patients with cystic fibrosis, performed WGS, and analysed phylogenetic relationships among these isolates. Surprisingly, they found strong evidence supporting human-to-human transmission of *M. abscessus* among patients with cystic fibrosis, and identified some *M. abscessus* isolates that were widespread globally (9). Although this finding remains controversial, these reports have raised concerns that NTM might be transmitted not only among immunosuppressed individuals such as cystic fibrosis patients but also among immunocompetent individuals (10). This would be especially important for hospital infection control, since isolation practices for NTM-PD patients without cystic fibrosis are not as strict as those for pulmonary tuberculosis patients generally.

Recently, we diagnosed three pairs of NTM-PD patients with no immunodeficiency who had been co-habiting for at least 15 years. We investigated the genomes of NTM isolates derived from the patients and environmental samples in their houses to understand the source of infection using WGS.

**Results**

**Patient Characteristics**

Three pairs of patients with NTM-PD who had been co-habiting for at least 15 years were enrolled (Table 1). A mother and a daughter (Patients A and B) with *M. avium* PD had lived in an apartment in an urban area (HOME-1) for 15 years. A couple (Patients C and D) with *M. intracellulare* PD had lived in a house in a rural area (HOME-2) for 30 years. A second couple (Patients E and F) with *M. intracellulare* PD and *M. abscessus* subsp. *massiliense* PD had lived in an apartment in an urban area (HOME-3) for 30 years. No patients were suspected of any immunodeficiency disorders. They were HIV-negative, were not taking any immunosuppressants and had no history of recurrent infection of any organs.

**NTM Isolation and Sequencing**
Among 12 environmental specimens from HOME-1, seven specimens from either the kitchen or the bathroom were culture-positive. Subsequently, 18 morphologically distinct isolates were purified (Supplementary Table 1, Additional File 1). However, none of the 15 environmental specimens from HOME-2 yielded any NTM isolates and only one of 15 specimens from HOME-3 yielded NTM isolates.

| Patient | Age/Relationship | NTM Species          | Habitat                                      | Radiologic findings                                      |
|---------|------------------|----------------------|----------------------------------------------|----------------------------------------------------------|
| Patient A | 81/Mother       | M. avium            | HOME-1 Apartment (urban area) for 15 years. | Bronchiectasis and centrilobular nodules in RML/LUL lingular segments |
| Patient B | 51/Daughter     | M. avium            | HOME-2 House (rural area) for 30 years with high soil environment | Centrilobular nodules with branching opacity in RUL/RML/LUL lingular segments |
| Patient C | 77/Husband       | M. intracellularare | HOME-2 House (rural area) for 30 years with high soil environment | Lung nodule in RUL |
| Patient D | 71/Wife          | M. intracellularare | HOME-3 Apartment (urban area) for 30 years. | Multiple branching opacity and centrilobular nodules in both |
| Patient E | 62/Husband       | M. intracellularare | HOME-3 Apartment (urban area) for 30 years. | Bronchiectasis with peribronchial infiltration in LLL |
| Patient F | 61/Wife          | M. abscessus subsp. massiliense | HOME-3 Apartment (urban area) for 30 years. | Patchy opacity and nodules in RUL |

RUL: right upper lobe, RML: right middle lobe, LUL: left upper lobe, LLL: left lower lobe

On average, 19.8 million sequencing reads were obtained for each isolate (See Supplementary Table 2, Additional File 1). According to the identification protocol using nine reference genomes, 12 isolates from environmental specimens in HOME-1, one isolate from Patient A, and one isolate from Patient B were identified as *M. avium* subsp. *avium*. Isolates from Patients C, D, and E were identified as *M. intracellularare*, and one isolate from Patient F was identified as *M. abscessus*. Two isolates from environmental specimens in HOME-1 were identified as *M. fortuitum*, while the remaining four isolates from HOME-1 and three isolates from HOME-3 were unidentifiable.

The mean sequencing depth of identified isolates was 313× (210–487×). The core region of *M. avium* subsp. *avium* consisted of 4.53 Mbp of 5.17 Mbp (88%), and that of *M. intracellularare* consisted of 5.20 Mbp of 5.40 Mbp (96%). The core region for *M. abscessus* was not defined because only one isolate was identified. Among 52,377 (44,670–55,621) SNPs and indels for each isolate, 38,042 (35,059–40,335) high-confidence SNPs in *M. avium* subsp. *avium* were detected (Supplementary Table 2, Additional File 1).
Pairwise SNP Distances and Phylogenetic Analysis

A total of 42,982 genomic positions with high-confidence SNPs were identified in *M. avium* subsp. *avium* and used for further analysis. Pairwise SNP distances between every pair of isolates at those positions were calculated and clusters on the histogram of SNP distances were observed (Figure 1). The SNP distance between the isolates from Patient A and B was 15,513. By contrast, the SNP distances among all replicates (between the isolates from patient A and the kitchen in HOME-1, and between the isolates from patient B and the bathroom) were less than 150 and these isolates clustered together closely in the phylogenetic tree (Figure 1). Using high-confidence SNPs from 14 *M. avium* subsp. *avium* isolates, phylogenetic analysis was performed (Figure 2). The isolate from Patient A and its replicates clustered with the specimens from the kitchen (scale on surface of kitchen faucet), while the isolate from Patient B clustered with the isolates from the bathroom (hot water from bathroom faucet, hot water from showerhead) and the kitchen (cold water from kitchen faucet).

In HOME-2, the SNP distance between the *M. intracellulare* isolates from Patients C and D was 25,136, even larger than the SNP distances from Patient E in HOME-3. The SNP distance between Patients C and E was 16,916 and the SNP distance between Patients D and E was 15,150. In HOME-3, different species of NTM were isolated (*M. intracellulare* from Patient E and *M. abscessus* subsp. *massiliense* from patient F), and the SNP distance was not calculated.

Discussion

Recent NGS studies suggested the patient-to-patient transmission of *M. abscessus* among cystic fibrosis patients. (8, 9) In response to these findings, the NTM guidelines have been updated. The British Thoracic Society (12), US Cystic Fibrosis Foundation and European Cystic Fibrosis Society (13) recommended that infection control policies minimize risks of patient-to-patient transmission for patients with both cystic fibrosis and *M. abscessus*-PD. However, these guidelines were only designed for patients with underlying factors and contain no statements for patients without immunodeficiencies. In this context, it is clinically important to understand if there is any evidence of patient-to-patient transmission among patients without predisposing factors.

Among the three pairs of co-habiting NTM-PD patients without overt immunodeficiency examined here, no patient pairs had near-identical NTM isolates (SNP distance <150). A pair of patients were infected with *M. avium*-PD isolates that were genetically distinct but similar to separate isolates from environmental specimens in their homes. Although we could not identify environmental sources for
the remaining two pairs of patients, one patient pair was also infected with genetically distinct *M. intracellulare* isolates (SNP distance > $10^4$), and the last pair of patients was infected by different species of NTM (*M. intracellulare* and *M. abscessus* subsp. *massiliense*). Considering that the estimated mutation rate of NTM is $\sim$0.5–1.8 mutations per year (8), there was no evidence of patient-to-patient transmission in our study.

Household water sources are considered a major reservoir for NTM, especially for *M. avium*. One report showed that samples from 17 (46%) of 37 households yielded the same species of NTM found in the patient. In seven of those households, the patient isolate and a plumbing isolate exhibited the same repetitive sequence-based PCR DNA fingerprint. Another study reported that seven of 20 (35%) patients with NTM-PD had NTM isolates with identical genotypes to isolates from household water or shower aerosols (14). Recently, variable number tandem repeat genotyping and WGS revealed that 11 of 21 (52%) patients with *M. avium*-PD had genetically matched isolates to household isolates (11).

In our study, a phylogenetic tree based on WGS was consistent with these previous reports and demonstrated the environmental sources of two *M. avium*-PD patients living together (HOME–1). Patient A seemed to have acquired NTM from the kitchen, while Patient B acquired NTM from the bathroom or kitchen.

Although we thoroughly collected environmental specimens from water and biofilms on the faucets and showerhead, we could not culture any NTMs from HOME–2, and identified only one NTM from HOME–3. A previous study reported that NTM species could be isolated from 22 of 37 (59%) households of NTM-PD patients. In addition, a positive correlation was observed between the number of samples collected per house and the number of NTM-positive samples (15). We collected 15 water samples from each of the two houses; this number is equal to or larger than that of most studies, suggesting that there may be sources of NTM-PD other than water. The patients living in HOME–2 (Patient C and D) were farmers and their home was a high soil environment that included a yard and hay (Table 1). Given that previous studies reported that a majority of *M. intracellulare* was isolated from soil samples, especially in high soil environments (14, 16, 17), the source of *M. intracellulare* for these patients could be soil instead of water.
In HOME–3, three morphologically distinct colonies were cultured from one environmental specimen and sequenced separately. However, we failed to identify the species because no isolates matched to NTM reference genomes. Considering that the sequencing reads of NTM isolates from patients E and F aligned to NTM reference genomes, the environmental specimen isolates might not be closely related to NTM in HOME–3. A previous study showed that, among 19 patients with *M. avium*-PD or *M. intracellularare*-PD living in low soil environments, no patients had genetically identical isolates compared with soil sample isolates from their houses (17). Our results suggested that a patient with *M. intracellularare*-PD (Patient E) in HOME–3 (low soil) may have been exposed to the NTM outside the house. The lag time between the acquisition of NTM from the environment and diagnosis of NTM-PD also makes it more difficult to identify the source of NTM (14).

**Conclusions**

NTM isolates derived from three pairs of co-habiting NTM-PD patients were not genetically identical based on WGS data. The sources of *M. avium* in patients in one pair was identified: one acquired infection from the kitchen and the other from the bathroom or kitchen. The likelihood of patient-to-patient transmission appeared minimal in these three pairs of NTM-PD patients without overt immune deficiency.

**Methods**

**Patient Enrolment**

Pairs of adult patients living together and diagnosed with NTM-PD at Seoul National University Hospital satisfying the following inclusion criteria were enrolled: age >18 years; typical respiratory symptoms such as chronic cough, sputum, or haemoptysis; findings suggestive of NTM-PD on computed tomography; identification of NTM in ≥2 sputum cultures or in ≥1 bronchoscopic specimen; living in the same household prior to diagnosis with NTM-PD; and consented to collection of environmental samples in their home. This study was approved by an institutional review board (IRB Number: 1804-064-936) and registered at ClinicalTrial.gov (NCT03532438).

**Sample Collection and NTM Isolation**
The most recently-isolated NTM from participant sputum or bronchial wash specimens from participants was collected. In addition, we visited patients’ homes to collect environmental specimens. One litre of water was aseptically collected into sterile containers directly from all faucets and showers in bathrooms and kitchens. Swabs were taken from the inside of faucets and showerheads in bathrooms and kitchens. As previously described (14, 18), each water specimen was passed through a 0.45 µm filter. Filters were rinsed with 10 mL of sterile distilled water and transferred to Middlebrook 7H11 plates and Mycobacterial Growth Indicator Tubes (MGITs). Swabs were washed in sterile distilled water and then processed in the same manner as the water samples. All culture-positive MGITs were transferred to new Middlebrook 7H11 plates. If culture-positive Middlebrook 7H11 plates had more than two morphologically distinct colonies, each distinct colony was separately transferred to a 3% Ogawa media plate, incubated, and purified as a single colony. If the colonies were homogenous and discrete on Middlebrook 7H11 plates, we collected the isolates without a further purification step.

**DNA Preparation and Sequencing**

All biomass from Middlebrook 7H11 plates or Ogawa media plates taken from sweeps of colonies was mixed with 425–600 µm glass beads, vortexed for 5 min, incubated at 80°C for 10 min and then centrifuged. DNA was extracted from the supernatant using a QIAamp DNA mini kit (Qiagen inc, Hilden, Germany) as previously described (8). Subsequently, DNA was sequenced using an Illumina HiSeq or NovaSeq instrument (Illumina, San Diego, USA). To validate the method, two patient isolates and an environmental specimen isolate were sequenced twice as replicates.

To identify species of NTM in each isolate, reads were separately mapped to nine mycobacterium reference genomes using BWA (19): *Mycobacterium avium* subsp. *avium* 2285, *M. avium* subsp. *paratuberculosis* strK10, *Mycobacterium intracellulare* ATCC 13950, *M. abscessus* ATCC 19977, *Mycobacterium fortuitum* CT6, *Mycobacterium kansasii* ATCC 12478, *Mycobacterium sinense* JDM601, *Mycobacterium smegmatis* MC2 155, and *Mycobacterium tuberculosis* H37Rv. Among the nine mycobacterium reference genomes, the genome with ≥70% reads aligned was considered the
species name of each isolate and was used for further analysis. To exclude error-prone regions, we defined a “core region” for each NTM species, a collection of 100-bp genomic region bins present in all of the same NTM isolates. Bins were excluded from the core region if the mean depth of any isolates in the bin were <100× or >600×.

**Variant Calling and Phylogenetic Analysis**

Variants were called using a combination of SAMtools and bcftools (20) with the following filters: minimum base quality of 50, minimum mapping quality of 30, support from at least four reads, and absence of heterozygosity. High-confidence SNPs were defined as SNPs in the core regions of each NTM species. SNP distance and the number of different high-confidence SNPs were calculated for each pair of isolates. Maximum likelihood trees were built using RAxML under a GTR model (21). All sequence data was deposited in the National Center for Biotechnology Information database as BioProject ID PRJNA577108.

**List Of Abbreviations**

MGIT: mycobacterial growth indicator tubes

NGS: next-generation sequencing

NTM: non-tuberculous mycobacterium

NTM-PD: non-tuberculous mycobacterial pulmonary disease

SNP: single nucleotide polymorphism

WGS: whole genome sequencing

**Declarations**

**Ethics approval and consent to participate**

All participants were informed about the study and written consent was obtained prior to the enrollment. Ethical approval of this study was approved by the institutional review board of Seoul National University Hospital (IRB Number: 1804-064-936).

**Consent for publication**
All the study participants provided written informed consent for publication of research results without their names disclosed.

**Availability of data and materials**

All sequence data supporting the conclusions of this article is available in the National Center for Biotechnology Information database as BioProject ID PRJNA577108.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

Not applicable.

**Authors Contributions**

J-JY designed the experiments. J-KY and J-JY participated in data collection. TSK isolated the bacteria. J-KY, TSK, J-IK and J-JY analyzed the data and wrote the paper. J-IK and J-JY supervised the study. All the authors read and approved the final manuscript.

**Acknowledgements**

Not applicable

**References**

1. Adjemian J, Olivier KN, Seitz AE, Holland SM, Prevots DR. Prevalence of nontuberculous mycobacterial lung disease in U.S. Medicare beneficiaries. American journal of respiratory and critical care medicine. 2012;185(8):881-6.

2. Park YS, Lee CH, Lee SM, Yang SC, Yoo CG, Kim YW, et al. Rapid increase of non-tuberculous mycobacterial lung diseases at a tertiary referral hospital in South Korea. Int J Tuberc Lung Dis. 2010;14(8):1069-71.
3. Izumi K, Morimoto K, Hasegawa N, Uchimura K, Kawatsu L, Ato M, et al. Epidemiology of Adults and Children Treated for Nontuberculous Mycobacterial Pulmonary Disease in Japan. Ann Am Thorac Soc. 2019;16(3):341–7.

4. Lee H, Myung W, Koh WJ, Moon SM, Jhun BW. Epidemiology of Nontuberculous Mycobacterial Infection, South Korea, 2007–2016. Emerg Infect Dis. 2019;25(3):569–72.

5. Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, et al. The origin of the Haitian cholera outbreak strain. N Engl J Med. 2011;364(1):33–42.

6. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science. 2010;327(5964):469–74.

7. Quick J, Cumley N, Wearn CM, Niebel M, Constantinidou C, Thomas CM, et al. Seeking the source of Pseudomonas aeruginosa infections in a recently opened hospital: an observational study using whole-genome sequencing. BMJ open. 2014;4(11):e006278.

8. Bryant JM, Grogono DM, Greaves D, Foweraker J, Roddick I, Inns T, 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.

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

10. Harris KA, Underwood A, Kenna DT, Brooks A, Kavaliunaite E, Kapatai G, 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(7):1007–16.

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

12. Haworth CS, Banks J, Capstick T, Fisher AJ, Gorsuch T, Laurenson IF, et al. British Thoracic Society guidelines for the management of non-tuberculous mycobacterial pulmonary disease (NTM-PD).
12.

Thorax. 2017;72(Suppl 2):ii1-ii64.

13.Floto RA, Olivier KN, Saiman L, Daley CL, Herrmann JL, Nick JA, et al. US Cystic Fibrosis Foundation and European Cystic Fibrosis Society consensus recommendations for the management of non-tuberculous mycobacteria in individuals with cystic fibrosis. Thorax. 2016;71 Suppl 1:i1–22.

14.Thomson R, Tolson C, Carter R, Coulter C, Huygens F, Hargreaves M. Isolation of nontuberculous mycobacteria (NTM) from household water and shower aerosols in patients with pulmonary disease caused by NTM. J Clin Microbiol. 2013;51(9):3006-11.

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

16.Feazel LM, Baumgartner LK, Peterson KL, Frank DN, Harris JK, Pace NR. Opportunistic pathogens enriched in showerhead biofilms. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(38):16393–9.

17.Fujita K, Ito Y, Hirai T, Maekawa K, Imai S, Tatsumi S, et al. Genetic relatedness of Mycobacterium avium-intracellulare complex isolates from patients with pulmonary MAC disease and their residential soils. Clin Microbiol Infect. 2013;19(6):537–41.

18.Thomson RM, Carter R, Tolson C, Coulter C, Huygens F, Hargreaves M. Factors associated with the isolation of Nontuberculous mycobacteria (NTM) from a large municipal water system in Brisbane, Australia. BMC Microbiol. 2013;13:89.

19.Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

20.Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987-93.

21.Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019.

Figures
Figure 1

Pairwise SNP distance distribution of M. avium isolates X-axis shows SNP distance and y-axis shows the frequency of distances between any two M. avium isolates in HOME-1. The inner box plot shows a magnified view from 0 to 1,500.
Phylogenetic analysis of NTM isolates from M. avium-PD patients and their home in urban area Patient A and B, a mother and daughter, lived for 15 years in HOME-1, an apartment in an urban area. Both were infected with M. avium-PD. A phylogenetic tree was constructed using the genome sequences of 14 M. avium subsp. avium isolates and three replicates (Patient Ar, 1-9M-1r, and 1-12M-2r). The SNP distances between the clusters around patients A and B were more than 15,000. The blue box indicates patient’s NTM isolates, the red box indicates isolates identified in the bathroom, and the green box indicates isolates identified in the kitchen.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
Additional file 1.xlsx