

# Prevalence and Risk Factors of Nontuberculous Mycobacterial Infection in a Coastal Area, Southern China

**Juan He**

Shenzhen Nanshan Center for Chronic Disease Control

**Xujun Guo**

Shenzhen Nanshan Center for Chronic Disease Control

**Chunfang Lv**

Shenzhen Nanshan Center for Chronic Disease Control

**Yuzheng Fan**

Shenzhen Nanshan Center for Chronic Disease Control

**Yao Huang**

Shenzhen Nanshan Center for Chronic Disease Control

**Liuzhu Lu**

Shenzhen Nanshan Center for Chronic Disease Control

**Amel Kevin Alame Emane**

Shenzhen Nanshan Center for Chronic Disease Control

**Howard Eugene Takiff**

Shenzhen Nanshan Center for Chronic Disease Control

**Shengyuan Liu** (✉ [liushenglb@126.com](mailto:liushenglb@126.com))

Shenzhen Nanshan Center for Chronic Disease Control

---

## Research Article

**Keywords:** Nontuberculous mycobacteria, Prevalence, Risk factor, Whole-genome sequencing

**Posted Date:** February 2nd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1308258/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Nontuberculous mycobacterial (NTM) infections are becoming more common, but the epidemiologic characteristics of NTM in much of the world remain largely unknown. The purpose of this study was to evaluate the prevalence and risk factors for NTM infection in a southern coastal area of China.

**Results:** 1759 individuals suspected of having tuberculosis were included in the analysis, of whom 140 (7.96%) had NTM isolated from their sputa. The NTM species identified by the Kraken 2/Bracken and Hain methods were highly consistent. *M. abscessus* complex bacilli were the most prevalent species isolated (n=58, 41.43%), followed by the *Mycobacterium avium* complex (MAC) (n=41, 29.29%). Residency in Shenzhen for more than 2 years (OR 4.205; 95% CI 1.851-9.552;  $P = 0.001$ ) and a history of prior TB (OR 4.263; 95% CI 1.871-9.714;  $P = 0.001$ ) increased the risk for NTM infections. Whole-genome sequencing data from many of the MAC isolates showed high genomic diversity and heterogeneity.

**Conclusion:** *M. abscessus* was the most common causative NTM species found in symptomatic patients in this region. Living in Shenzhen for more than 2 years and having a previous history of TB were associated with an increased risk of NTM infection. MAC genomes are often heterogenous.

## Background

Nontuberculous mycobacteria (NTM) are environmental species that occasionally cause opportunistic infections in humans and animals that can be similar to infections with the *Mycobacterium tuberculosis* complex (MTBC) bacteria. Aided by the development of molecular methods to analyze NTM, more than 250 species and subspecies are now included in the List of Prokaryotic names with Standing in Nomenclature (LPSN) [1]. The incidence and prevalence of NTM infections have been increasing over the past several decades [2], and studies in four U.S. states showed that the rates of NTM positive cultures increased from 8.2 cases/100,000 in 1994 to 16 cases/100,000 in 2014 [3]. Data from Korea showed a similar trend of increasing NTM incidence but with regional differences in the rates [4]. Studies have documented higher healthcare costs for individuals with NTM disease, including one from Canada showing that during 2001–2012, the cumulative mean attributable 1-year costs associated with NTM disease were CAD \$14,953 (USD \$11,541) [5].

It is not easy to distinguish NTM from MTBC in clinical specimens because both are acid-fast bacilli that will be read as positive by conventional microscopy, and traditional methods to identify the NTM species are complicated [6]. However, molecular diagnostic techniques, such as the GenoType system (Hain Lifescience, Germany), have simplified NTM species identification [7]. Whole-genome sequencing (WGS) is also used to identify and describe NTM species more precisely, and it has the ability to detect multiple species present in samples from a single individual [8]. The taxonomic classification of WGS data and estimation of relative abundances from NTM isolates can be performed using Kraken / Bracken [9]. Kraken can classify WGS reads to bacterial species in the Kraken database, and Kraken 2, an updated

version of Kraken 1, can make the classifications faster [10]. The use of Bracken may improve on the Kraken results and add the capability of estimating bacterial abundance [11].

The principal NTM species causing pulmonary disease are members of the *Mycobacterium avium* complex (MAC), the most common of which include *M. avium*, *M. intracellulare*, and *M. chimaera* [12], but there is a great diversity within the MAC [13]. Newly identified members of the complex, such as *M. paraintracellulare* [14] have also been found in pulmonary infections. The identification of the causative species can provide important epidemiologic and clinical information to aid in the choice of appropriate therapy [15].

Shenzhen is a new city in the south of China. In the Nanshan district of Shenzhen, the Shenzhen Nanshan Center for Chronic Disease Control (CCDC) is the designated local TB facility, providing diagnosis of suspected TB or NTM infections. In the study reported here, we conducted a retrospective study of patients with NTM infections presenting to the Nanshan CCDC. Using both the Hain technique as well as WGS with Kraken 2, we identified the causative mycobacterial species and then analyzed the patient demographic characteristics to identify risk factors associated with NTM infections.

## Results

### NTM species

From a total of 2145 positive sputum cultures, 1759 isolates, from an equal number of patients, were included in the final analysis (Fig. 1). Of these, 1619 (92.04%) were identified as MTBC and 140 (7.96%) as NTM. A total of 12 different NTM species or species complexes were detected. Isolates belonging to the *M. abscessus* complex were the most common (n=58, 41.43%), followed by *M. sp. VKM Ac-1817D* (n=17, 12.14%), *M. kansasii* (n=16, 11.43%), *M. avium* (n=13, 9.29%), *M. intracellulare* (n=13, 9.29%), *M. fortuitum* (n=6, 4.29%), *M. colombiense* (n=5, 3.57%), *M. marseillense* (n=5, 3.57%), *M. chimaera* (n=4, 2.86%), *M. paraintracellulare* (n=1, 0.71%), *M. sinensis* (n=1, 0.71%) and *M. sp. YC-RL4* (n=1, 0.71%). Except for *M. abscessus*, *M. fortuitum*, *M. sinensis*, *M. sp. VKM Ac-1817D* and *M. sp. YC-RL4*, these isolates were slowly growing mycobacterial species (SGM) [16] (Table 1). Isolates belonging to the *M. abscessus* complex were not subdivided into species. *M. sinensis*, synonymous with *Mycobacterium sinense*, is a multidrug-resistant mycobacteria phylogenetically placed between rapidly growing nontuberculous mycobacteria (RGM) and SGM [17]. *M. sp. VKM Ac-1817D* [18] and *M. sp. YC-RL4* [19] are considered nonpathogenic environmental bacteria and may represent contaminants rather than pathogenic species associated with disease.

Table 1  
NTM species identified by Kraken 2 / Bracken (n=140)

| Classification  | Species                      | Number              | Percentage (%) |
|---|------------------------------|---------------------|----------------|
| SGM   | <i>M. kansasii</i>           | 16                  | 11.43          |
|   | <i>M. avium</i>              | 13                  | 9.29           |
|   | <i>M. intracellulare</i>     | 13                  | 9.29           |
|   | <i>M. colombiense</i>        | 5                   | 3.57           |
|   | <i>M. marseillense</i>       | 5                   | 3.57           |
|   | <i>M. chimaera</i>           | 4                   | 2.86           |
|   | <i>M. paraintracellulare</i> | 1                   | 0.71           |
|   | RGM                          | <i>M. abscessus</i> | 58             |
| <i>M. fortuitum</i>   |                              | 6                   | 4.29           |
| Others  | <i>M. sp. VKM Ac-1817D</i>   | 17                  | 12.14          |
|   | <i>M. sp. YC-RL4</i>         | 1                   | 0.71           |
|   | <i>M. sinensis</i>           | 1                   | 0.71           |
| Total   |                              | 140                 | 100            |
| SGM: slowly growing nontuberculous mycobacteria; RGM: rapidly growing nontuberculous mycobacteria |                              |                     |                |

### Demographic characteristics and risk factors of NTM infection

Of the 140 individuals with NTM infection, 83 (59.3%) were male and 53 (37.8%) were over 50 years of age (Table 2). 97.7% (1719/1759) of the entire study population were internal migrants to Shenzhen from other regions of China. Univariable logistic regression suggested a significant association of NTM infection with gender, older age, longer period of residence in Shenzhen, retired or unemployed status and prior history of TB. However, after multivariable regression, only longer residence in Shenzhen and prior TB history remained associated with NTM infection. Compared to patients who had lived in Shenzhen for less than 1 year, those who had resided in Shenzhen for more than 2 years had a significantly higher prevalence of NTM than MTBC infections [odds ratio (OR) 4.205; 95% confidence interval (95%CI) 1.851-9.552;  $P = 0.001$ ]. There was no significant difference in the percentage of NTM vs. MTBC infections in patients who had lived in Shenzhen for 1-2 years ( $P = 0.065$ ). We also found that the prevalence of NTM infections was significantly higher in patients with a prior history of TB disease than in those with no TB history (OR 4.263; 95% CI 1.871-9.714;  $P = 0.001$ ). The results of species identification on isolates from 26 patients with 2-3 sputum positive specimens taken at different times showed that in 2 patients the

isolates from the initial sputa were identified as MTBC but *M. abscessus* was found in their follow-up sputa (Supplementary Table S1). In addition, NTM was isolated less frequently than MTBC in patients presenting with cough (OR 0.548; 95% CI 0.310-0.970;  $P = 0.039$ ). No significant interaction of patients' age and residence years was observed ( $P = 0.120$ ).

Table 2  
Demographic and clinical characteristics of the 1759 cases with mycobacteria infections

| Characteristics       | No. of mycobacteria infections, n (%) |                |              | Univariate regression  |         | Multivariable regression             |              |
|-----------------------|---------------------------------------|----------------|--------------|------------------------|---------|--------------------------------------|--------------|
|                       | All                                   | MTBC           | NTM          | OR (95%CI)             | P value | OR (95%CI)                           | P value      |
| Gender                |                                       |                |              |                        |         |                                      |              |
| Female                | 558<br>(31.7)                         | 501<br>(30.9)  | 57<br>(40.7) | Rf                     |         |                                      |              |
| Male                  | 1201<br>(68.3)                        | 1118<br>(69.1) | 83<br>(59.3) | 0.653<br>(0.458-0.929) | 0.018   |                                      |              |
| Age group, years      |                                       |                |              |                        |         |                                      |              |
| <30                   | 873<br>(49.6)                         | 822<br>(50.8)  | 51<br>(36.5) | Rf                     |         |                                      |              |
| 30~49                 | 536<br>(30.5)                         | 500<br>(30.9)  | 36<br>(25.7) | 1.160<br>(0.747-1.803) | 0.508   |                                      |              |
| 50~69                 | 290<br>(16.5)                         | 247<br>(15.2)  | 43<br>(30.7) | 2.806<br>(1.825-4.313) | <0.001  |                                      |              |
| ≥70                   | 60<br>(3.4)                           | 50<br>(3.1)    | 10<br>(7.1)  | 3.224<br>(1.545-6.726) | 0.002   |                                      |              |
| Residence years in SZ |                                       |                |              |                        |         |                                      |              |
| (0,1]                 | 513<br>(32.5)                         | 488<br>(33.5)  | 25<br>(20.9) | Rf                     |         |                                      |              |
| (1,2]                 | 177<br>(11.2)                         | 164<br>(11.2)  | 13<br>(10.8) | 1.547<br>(0.774-3.095) | 0.217   | 2.785<br>(0.937-8.280)               | 0.065        |
| (2,+∞)                | 889<br>(56.3)                         | 807<br>(55.3)  | 82<br>(68.3) | 1.983<br>(1.250-3.147) | 0.004   | <b>4.205</b><br><b>(1.851-9.552)</b> | <b>0.001</b> |
| Occupation            |                                       |                |              |                        |         |                                      |              |
| Active                | 1048<br>(59.6)                        | 978<br>(60.4)  | 70<br>(50.0) | Rf                     |         |                                      |              |

MTBC: *Mycobacterium tuberculosis* complex; NTM: Nontuberculous mycobacteria; SZ: Shenzhen; OR: Odds Ratio; CI: Confidence Interval.

| Characteristics              | No. of mycobacteria infections, n (%) |               |              | Univariate regression  |         | Multivariable regression       |              |
|------------------------------|---------------------------------------|---------------|--------------|------------------------|---------|--------------------------------|--------------|
|                              | All                                   | MTBC          | NTM          | OR (95%CI)             | P value | OR (95%CI)                     | P value      |
| Students                     | 88<br>(5.0)                           | 79<br>(4.9)   | 9<br>(6.4)   | 1.592<br>(0.766-3.306) | 0.213   |                                |              |
| Retired or Unemployed        | 623<br>(35.4)                         | 562<br>(34.7) | 61<br>(43.6) | 1.516<br>(1.059-2.171) | 0.023   |                                |              |
| Annual income, yuan          |                                       |               |              |                        |         |                                |              |
| <30000                       | 309<br>(35.0)                         | 288<br>(35.5) | 21<br>(29.2) | Rf                     |         |                                |              |
| 30000~50000                  | 212<br>(24.0)                         | 196<br>(24.2) | 16<br>(22.2) | 1.120<br>(0.570-2.199) | 0.743   |                                |              |
| 50000~80000                  | 163<br>(18.5)                         | 144<br>(17.7) | 19<br>(26.4) | 1.810<br>(0.943-3.473) | 0.075   |                                |              |
| >80000                       | 199<br>(22.5)                         | 183<br>(22.6) | 16<br>(22.2) | 1.199<br>(0.610-2.358) | 0.599   |                                |              |
| Clinical factors (yes v no)  |                                       |               |              |                        |         |                                |              |
| TB history                   | 91<br>(5.2)                           | 72<br>(4.4)   | 19<br>(13.6) | 3.374<br>(1.969-5.780) | <0.001  | <b>4.263<br/>(1.871-9.714)</b> | <b>0.001</b> |
| Hypertension                 | 24<br>(1.4)                           | 22<br>(1.4)   | 2<br>(1.4)   | 1.052<br>(0.245-4.521) | 0.946   |                                |              |
| Diabetes mellitus            | 62<br>(3.5)                           | 57<br>(3.5)   | 5<br>(3.6)   | 1.015<br>(0.400-2.575) | 0.975   |                                |              |
| Pulmonary cavity             | 427<br>(24.3)                         | 399<br>(24.6) | 28<br>(20.0) | 0.764<br>(0.498-1.174) | 0.220   |                                |              |
| Clinical symptoms (yes v no) |                                       |               |              |                        |         |                                |              |

MTBC: *Mycobacterium tuberculosis* complex; NTM: Nontuberculous mycobacteria; SZ: Shenzhen; OR: Odds Ratio; CI: Confidence Interval.

| Characteristics | No. of mycobacteria infections, n (%) |               |              | Univariate regression      |         | Multivariable regression                  |              |
|-----------------|---------------------------------------|---------------|--------------|----------------------------|---------|---|--------------|
|                 | All                                   | MTBC          | NTM          | OR (95%CI)                 | P value | OR (95%CI)                                | P value      |
| Cough           | 791<br>(45.0)                         | 738<br>(45.6) | 53<br>(37.9) | 0.727<br>(0.510-<br>1.037) | 0.079   | <b>0.548</b><br><b>(0.310-<br/>0.970)</b> | <b>0.039</b> |
| Hemoptysis      | 169<br>(9.6)                          | 159<br>(9.8)  | 10<br>(7.1)  | 0.706<br>(0.364-<br>1.372) | 0.305   |   |              |
| Chest Pain      | 113<br>(6.4)                          | 106<br>(6.5)  | 7<br>(5.0)   | 0.751<br>(0.343-<br>1.647) | 0.475   |   |              |
| Fever           | 72<br>(4.1)                           | 65<br>(4.0)   | 7<br>(5.0)   | 1.258<br>(0.566-<br>2.799) | 0.573   |   |              |
| Others          | 46<br>(2.6)                           | 44<br>(2.7)   | 2<br>(1.4)   | 0.519<br>(0.124-<br>2.163) | 0.368   |   |              |

MTBC: *Mycobacterium tuberculosis* complex; NTM: Nontuberculous mycobacteria; SZ: Shenzhen; OR: Odds Ratio; CI: Confidence Interval.

### Consistency of the Kraken 2 / Bracken and Hain method

Of the 1759 total, 1032 were also analyzed by the Hain test for mycobacterial species. 1030 of 1032 isolates were identified by Hain as NTM or MTBC, while 2/1032 were found to be co-infections with NTM and MTBC. The agreement between the Kraken 2 / Bracken method and the Hain test (MTBC (n=931), NTM (n=74)) was 97.57% (1005/1030) with a kappa value of 0.842 ( $P < 0.001$ ). Among the 74 isolates identified as NTM by both methods (Table 3), the Kraken 2 / Bracken method identified 33 as *M. abscessus*, of which 29/33 (87.88%) were also identified as *M. abscessus* using Hain, 2/33 were identified as *M. gordonae*, 1/33 as a mixed infection of *M. gordonae* and *M. abscessus* and 1/33 as an unidentified species. For the 5 strains designated as *M. fortuitum* by Kraken 2 / Bracken, 3/5 (60%) were identified as *M. fortuitum* by Hain, and 2/5 were unidentified species. The species designations by Kraken 2 / Bracken and Hain were identical for *M. avium* (8/8), *M. kansasii* (8/8) and *M. intracellulare* (7/7) isolates.

Table 3  
Comparison between Kraken 2 / Bracken and Hain for 74 strains isolated as NTM

| Kraken 2 / Bracken                   | Hain                                 | Consistency   |
|--------------------------------------|--------------------------------------|---------------|
| <b>33/74</b> <i>M. abscessus</i>     | <b>29/33</b> <i>M. abscessus</i>     | 29/33, 87.88% |
|                                      | 1/33 <i>M. gordonae</i>              |               |
|                                      | 1/33 <i>M. gordonae/M. abscessus</i> |               |
|                                      | 1/33 <i>M. chelonae</i>              |               |
|                                      | 1/33 Unidentified species            |               |
| <b>8/74</b> <i>M. avium</i>          | <b>8/8</b> <i>M. avium</i>           | 8/8, 100%     |
| <b>8/74</b> <i>M. kansasii</i>       | <b>8/8</b> <i>M. kansasii</i>        | 8/8, 100%     |
| 8/74 <i>M. sp. VKM Ac-1817D</i>      | 8/8 <i>M. fortuitum</i>              | 0%            |
| <b>7/74</b> <i>M. intracellulare</i> | <b>7/7</b> <i>M. intracellulare</i>  | 7/7, 100%     |
| <b>5/74</b> <i>M. fortuitum</i>      | <b>3/5</b> <i>M. fortuitum</i>       | 3/5, 60%      |
|                                      | 2/5 Unidentified species             |               |
| 2/74 <i>M. colombiense</i>           | 2/2 <i>M. intracellulare</i>         | 0%            |
| 2/74 <i>M. marseillense</i>          | 2/2 <i>M. intracellulare</i>         | 0%            |
| 1/74 <i>M. chimaera</i>              | 1/1 <i>M. malmoense/M. avium</i>     | 0%            |
| Total (n=74)                         | Total (n=74)                         | 55/74, 74.32% |
| NTM: Nontuberculous mycobacteria.    |                                      |               |

### Discrepancies between Hain and Kraken 2 / Bracken

The Hain test on the sputum cultures of eight patients found NTMs (2 *M. fortuitum*, 1 *M. kansasii*, 2 *M. abscessus* and 3 unidentified species), while Kraken 2 / Bracken analysis of the WGS of these 8 cultures found MTBC (Supplementary Table S2). Although the sputum smears were negative for seven of these eight patients, all were treated for tuberculosis based on clinical judgement. The WGS analysis was performed retrospectively, much later and did not influence therapy. Seven completed their treatment successfully and the eighth, after 4 months of the standard TB regimen, was transferred for NTM treatment. For this eighth patient, *M. abscessus* was diagnosed in their sputum by the Hain test and only 59% of WGS reads from their sputum culture mapped to MTBC.

In another seventeen patients, Hain found MTBC in the sputum cultures while Kraken 2 / Bracken found NTM in their WGS: 6 *M. abscessus* (> 95% reads); 3 *M. kansasii* (85%, 98% and 99% reads); 1 *M. colombiense* (96% reads); 1 *M. intracellulare* (58% reads); 1 *M. fortuitum* (47% reads); 1 *M. chimaera* (41%

reads); 3 *M. sp. VKM Ac-1817D* (38%, 48% and 50% reads); and 1 *M. sp. YC-RL4* (21% reads). Of these, 12/17 completed treatment for TB and 3 were lost to follow-up, perhaps because they went to their hometowns to be treated. 2 patients were transferred for MDR-TB treatment after a month or two on the standard drug sensitive TB regimen. In the sputa of two patients, Hain found mixed infections of *M. abscessus* and MTBC, and *M. gordonae* and MTBC, while Kraken 2 / Bracken found *M. abscessus* (99% of reads) and MTBC (67% of reads), respectively. The patient with *M. abscessus* and MTBC, who had a cough > 2 weeks, a 2+ positive sputum and a prior history of TB, received 8 months of TB treatment (2HRZES3/6HRE) and was cured. The patient with *M. gordonae* and MTBC, who had no symptoms, a negative sputum and no prior history of TB, completed standard TB therapy (2HRZE/4HR).

## MAC diversity revealed by WGS

Taxonomic classification of WGS data from the MAC isolates (*M. avium*, *M. intracellulare*, *M. colombiense*, *M. marseillense*, *M. chimaera* and *M. paraintracellulare*) was performed using Kraken 2 followed by Bayesian re-estimation of relative abundances using Bracken (Fig. 2). The 41 samples belonging to the MAC were clustered by their Bray-Curtis dissimilarity distances into two major groups: one predominantly *M. avium*; and a second, mixed group mainly comprised of *M. intracellulare*, *M. colombiense* and *M. marseillense*. The isolates in the *M. avium* cluster had reads that mapped almost exclusively to *M. avium* (96%), whereas the reads from isolates within the mixed group mapped to different MAC species: in the *M. intracellulare* cluster, the percentage of reads classified as *M. intracellulare* ranged from 31–84%; in the *M. marseillense* cluster, 34%-96% of reads were classified as *M. marseillense*; in the *M. chimaera* cluster, 24%-88% were classified as *M. chimaera*; and in the *M. colombiense* cluster 78%-96% of reads were classified as *M. colombiense* (Supplementary Table S3).

## Discussion

The increasing prevalence of pulmonary NTM disease is often neglected as a public health concern in China [20], but we isolated NTM from sputum cultures of nearly 8% (7.96%, 140/1759) of suspected TB patients, similar to a nationwide survey that found NTM infections in 7.7% of patients suspected of having tuberculosis in southern China [21]. A systematic review and meta-analysis showed that the prevalence of NTM infections and the mycobacterial species isolated vary by geographic distribution, with higher rates of NTM found in southern, high-humidity coastal areas than in inland regions [22].

Molecular diagnostics provide a more rapid and precise identification of MTBC and NTM than conventional biochemical methods [23], thus facilitating early and appropriate treatment. Using Kraken 2 / Bracken to distinguish between MTBC and NTM and to estimate the abundance of MAC, we found that the species designations were highly consistent with species identification with the Hain test, suggesting that either method can be used for clinical diagnosis to guide treatment. Strains belonging to the *M. abscessus* complex ( $n=58$ , 41.43%) were the most frequently isolated NTM in the Nanshan District and were also the most common in the Guangzhou area of southern China [24]. In contrast, studies in other areas of China, such as Beijing (north) [25], Changchun (northeast) [26], Nanjing (southeast) [27] found

that *M. intracellulare* was the most common NTM species isolated from patient specimens. In our study, isolates belonging to the MAC ( $n=41$ , 29.29%), were the second most common group of NTM species identified, and included *M. avium* ( $n=13$ , 9.29%), *M. intracellulare* ( $n=13$ , 9.29%), *M. colombiense* ( $n=5$ , 3.27%), *M. marseillense* ( $n=5$ , 3.27%), *M. chimaera* ( $n=4$ , 2.86%) and *M. paraintracellulare* ( $n=1$ , 0.71%). The geographical variation in the prevalence of causative organisms of NTM infections likely reflects the diversity of environmental NTM present across China. However, while nearly all of the WGS reads from isolates identified as *M. avium* and *M. chimaera* mapped to these species, reads from the other isolates classified as belonging to the MAC showed genomic heterogeneity, with reads mapping to several MAC species. A similar heterogeneity was previously reported in a study that found interspecies and intraspecies MAC diversity that varied over time in follow-up isolates from patients with MAC lung disease [9]. It was suggested that this diversity was perhaps because infections of certain MAC species are polymicrobial or the reference genomes do not fully capture the genomic intricacies of the microorganisms in this complex. These findings may have clinical significance for the diagnosis and treatment of MAC lung disease.

As NTM are thought to be acquired from the environment, a substantial number of patients have no known risk factors, but we identified two patient characteristics that were statistically associated with NTM infections: residency in Shenzhen longer than 2 years and prior history of TB. The role of residency duration has been reported previously in Canada [28] and may indicate that recent immigrants are less likely to acquire infections with the local NTM. Our preliminary speculation is that the acquisition of NTM infections may be related to the drinking water or trace metals in surface water. It has been reported that concentrations of the trace metals molybdenum and vanadium in surface water sources were associated with NTM infection in Oregon [29]. Water age (i.e., combined distribution system residence time and home plumbing stagnation time) has also been reported to be associated with a greater relative abundance of *M. avium* [30]. The migrants from other areas of China to Shenzhen would be exposed to the local drinking and surface water that may gradually alter their immune status to increase the possibility of NTM infections.

Patients previously diagnosed with TB also have a greater risk of NTM infections. The reasons for this phenomenon remain unclear, but it could be associated with a weakened immune system or damaged lung architecture resulting from the previous TB disease [31]. To illustrate this possibility, we performed species identification on isolates from 26 patients with 2-3 sputum positive specimens taken at different times and found that in 2 patients the isolates from the initial sputa were identified as MTBC but *M. abscessus* was found in their follow-up sputa. Although we found that cough is less common in NTM than in pulmonary TB patients and was identified as a “protective” factor, NTM infection should nevertheless be added to the differential diagnosis in patients with chronic coughs unresponsive to routine antibiotic therapy [32]. Contrary to previous reports [33] older age, female gender and retired or unemployed status were not identified as risk factors when analyzed by multivariable regression.

Our study has several limitations. First, because Kraken requires abundant computing power, we used a reduced-sensitivity MiniKraken database that has been shown to be nearly equivalent to the full database

[34], but is perhaps slightly less precise. Second, underlying health conditions, such as chronic obstructive pulmonary disease (COPD), bronchiectasis and HIV infection pose a significant risk for NTM infections [35], but we did not record these comorbidities or symptoms in the clinical data and therefore couldn't determine whether they were associated with NTM infections. NTM infections can be notoriously difficult to treat due to intrinsic resistance to many common antibiotics, but we were unable to collect information on treatments or outcomes because patients found to have NTM do not receive their treatment in Nanshan, but rather at the central Shenzhen City CCDC. These aspects, which are missing in the current analysis, should be addressed in future studies.

## Conclusions

In conclusion, we found that strains belonging to the *M. abscessus* complex were the most common NTM isolated from sputa of patients suspected to have TB, and that residency in Shenzhen for more than 2 years and patients with TB history had an increased risk for NTM infections. We also found high genomic diversity and heterogeneity in the MAC isolates. Future studies should try to correlate treatment and outcomes with the NTMs identified in patient sputa.

## Methods

### Study design and population

The Nanshan District is located in the southwest coastal area of Shenzhen City and is one of the most prosperous regions in China. It has a population of approximately 1.8 million and in 2020 had a Gross Domestic Product (GDP) of about 100\$ billion. All suspected tuberculosis (TB) patients presenting to the Shenzhen Nanshan CCDC between January 2015 to December 2019 were eligible for inclusion in this study if they had positive sputum cultures. Aliquots of all isolates were maintained at -80 C and re-grown on Löwenstein-Jensen (L-J) medium or in a mycobacteria growth indicator tube (MGIT) tube, from which genomic DNA was isolated for sequencing. For patients with more than one positive sputum culture, the isolate grown from the earliest sputum specimen was selected. A total of 1896 unique isolates were successfully genome sequenced, and 1759 isolates were included in analysis, of which 140 (7.96%, 140/1759) were found to be NTM.

The species were also identified by the Hain genotyping method (1032/1759). Sputum specimens were cultured on Löwenstein-Jensen (L-J) medium and the positive cultures were sent to the central Shenzhen CCDC for species identification by Hain test, where 82 isolates were identified as NTM. Of those 82 isolates, 74 isolates were also identified as NTM by Kraken 2 / Bracken (Supplementary Figure S1). We compared the results of Kraken 2 / Bracken and Hain methods for species identification of the 74 isolates.

## WGS

Genomic DNA of the 1896 strains were extracted using the cetyltrimethylammonium bromide (CTAB) method [36]. For each strain, a paired-end library was constructed and sequenced on an Illumina platform (Illumina, San Diego, CA, USA). The raw data was analyzed using in-house Python scripts to obtain the coverage and depth. The taxonomic classification was based on Kraken 2 / Bracken [37]. The MiniKraken database was downloaded and used with a publicly available script [38]. Isolates with a sequence depth less than 20X or genome coverage less than 95% were excluded from the analysis [39].

## Statistical analysis

We extracted the basic demographic and clinical data for the enrolled patients from the routine TB surveillance system of the Shenzhen Nanshan CCDC. Information on social factors was obtained from a structured questionnaire routinely administered to all patients with positive sputum cultures. The risk factors for NTM were analyzed using univariable and multivariable logistic regressions, with estimates of ORs and 95% CIs. Factors with a *P* value less than 0.2 in the univariable regression were included in the multivariable analysis and considered to be significantly associated with NTM if their *P* values were less than 0.05 in the final multivariable model. The consistency of the Kraken 2 / Bracken and Hain method was evaluated using the kappa statistic. All statistical analysis was performed in SPSS v22.0 (IBM, USA). Missing data was not included in the analyses. Visualization of the relative species abundance of MAC was performed in R v4.1.1.

## Abbreviations

NTM: Nontuberculous mycobacterial; MAC: *Mycobacterium avium* complex; WGS: Whole-genome sequencing; MTBC: *Mycobacterium tuberculosis* complex; LPSN: List of Prokaryotic names with Standing in Nomenclature; CCDC: Center for Chronic Disease Control; SGM: slowly growing mycobacterial species; RGM: rapidly growing nontuberculous mycobacteria; COPD: chronic obstructive pulmonary disease; GDP: Gross Domestic Product; TB: tuberculosis; L-J: Löwenstein-Jensen; OR: Odds Ratio; CI: Confidence Interval; MGIT: mycobacteria growth indicator tube; CTAB: cetyltrimethylammonium bromide.

## Declarations

### Ethics approval and consent to participate

The study was approved by the Ethics Review Committee of the Shenzhen Nanshan Center for Chronic Disease Control. Strains were collected as part of the routine clinical management of patients, according to the national guidelines in China. Therefore, informed consent was not sought.

### Consent for publication

Not applicable.

### Availability of data and materials

All data generated or analysed during this study are included in this published article.

## Competing interests

The authors declare that they have no competing interests.

## Funding

This work was supported by the Sanming project of Medicine in Shenzhen (SZSM201603029) and the Key Disciplines of Medicine in Nanshan District.

## Authors' contributions

JH and XG participated in the design of the study, the statistical analysis and drafted the manuscript. HT and SL participated in the design of the study and revised the manuscript carefully. AKAE assisted in the species identification. YF and YH collected the data. CL and LL cultured strains and did the sample preparation. All authors read and approved the final manuscript.

## Acknowledgments

The authors thank Guiyuan Han and Zijin Jiao for their contributions in part of the collation of the data, both of whom mentioned above are from Shenzhen Nanshan CCDC.

## References

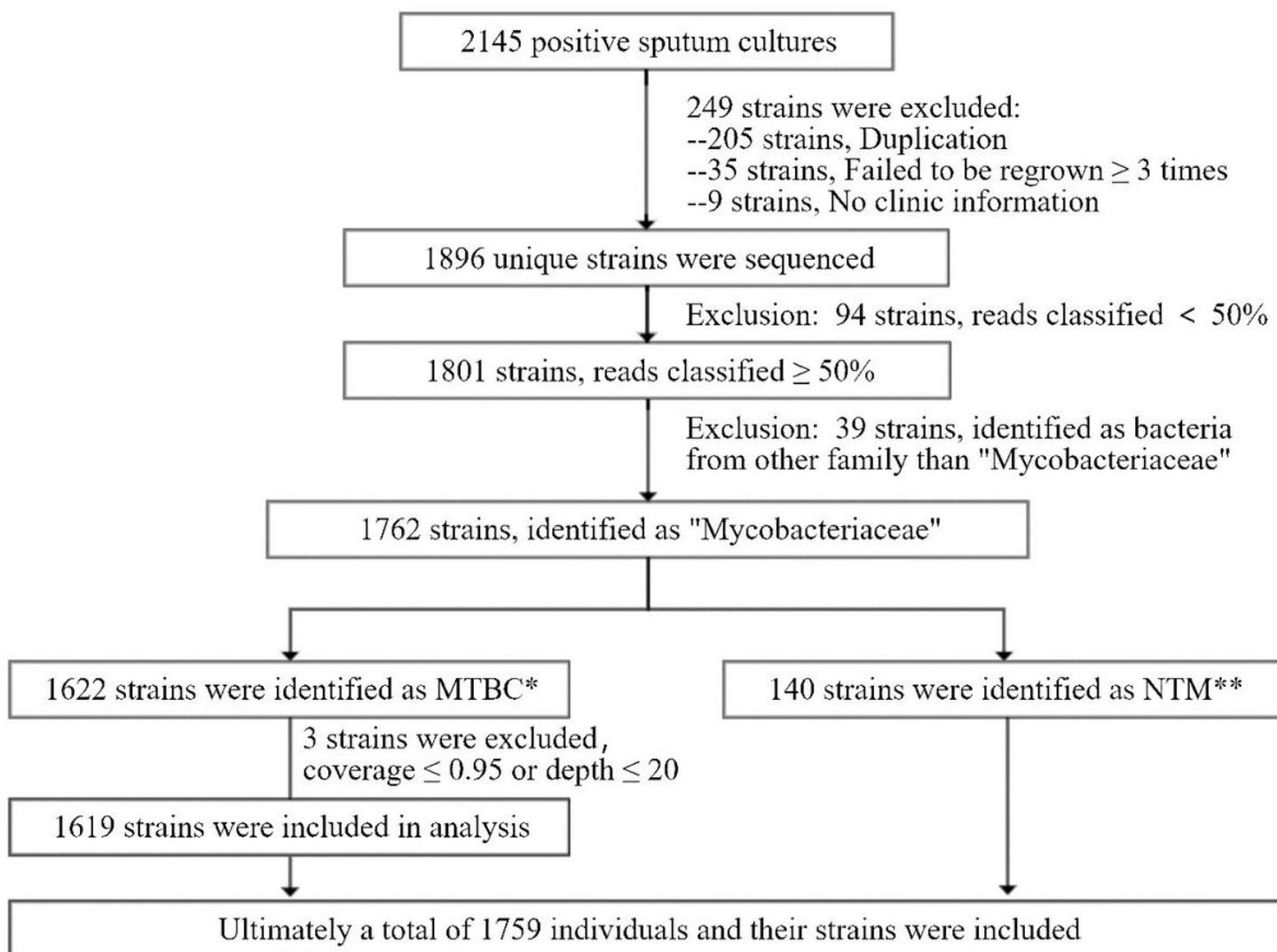
1. Parte AC. LPSN - List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *Int J Syst Evol Microbiol.* 2018;68:1825–9.
2. Ratnatunga CN, Lutzky VP, Kupz A, Doolan DL, Reid DW, Field M, et al. The Rise of Non-Tuberculosis Mycobacterial Lung Disease. *Front Immunol.* 2020;11:303.
3. Donohue MJ. Increasing nontuberculous mycobacteria reporting rates and species diversity identified in clinical laboratory reports. *BMC Infect Dis.* 2018;18:163.
4. Park SC, Kang MJ, Han CH, Lee SM, Kim CJ, Lee JM, et al. Prevalence, incidence, and mortality of nontuberculous mycobacterial infection in Korea: a nationwide population-based study. *BMC Pulm Med.* 2019;19:140.
5. Ramsay LC, Shing E, Wang J, Marras TK, Kwong JC, Brode SK, et al. Costs Associated with Nontuberculous Mycobacteria Infection, Ontario, Canada, 2001-2012. *Emerg Infect Dis.* 2020;26:2097–107.
6. Gopaldaswamy R, Shanmugam S, Mondal R, Subbian S. Of tuberculosis and non-tuberculous mycobacterial infections - A comparative analysis of epidemiology, diagnosis and treatment. *J Biomed Sci.* 2020;27:1–17.
7. Safianowska A, Walkiewicz R, Nejman-Gryz P, Chazan R, Grubek-Jaworska H. [Diagnostic utility of the molecular assay GenoType MTBC (HAIN Lifesciences, Germany) for identification of tuberculous

- mycobacteria]. *Pneumonol Alergol Pol.* 2009;77:517–20.
8. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, et al. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol.* 2017;243:16–24.
  9. Operario DJ, Pholwat S, Koeppel AF, Prorock A, Bao Y, Sol-Church K, et al. Mycobacterium avium Complex Diversity within Lung Disease, as Revealed by Whole-Genome Sequencing. *Am J Respir Crit Care Med.* 2019;200:393–6.
  10. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20:257.
  11. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: Estimating species abundance in metagenomics data. *PeerJ Comput Sci.* 2017;2017:1–14.
  12. Daley CL. Mycobacterium avium Complex Disease. *Microbiol Spectr.* 2017;5.
  13. Tortoli E. Microbiological features and clinical relevance of new species of the genus Mycobacterium. *Clin Microbiol Rev.* 2014;27:727–52.
  14. Lee S-Y, Kim B-J, Kim H, Won Y-S, Jeon CO, Jeong J, et al. Mycobacterium paraintracellulare sp. nov., for the genotype INT-1 of Mycobacterium intracellulare. *Int J Syst Evol Microbiol.* 2016;66:3132–41.
  15. Daley CL, Iaccarino JM, Lange C, Cambau E, Wallace RJ, Andrejak C, et al. Treatment of Nontuberculous Mycobacterial Pulmonary Disease: An Official ATS/ERS/ESCMID/IDSA Clinical Practice Guideline. *Clin Infect Dis an Off Publ Infect Dis Soc Am.* 2020;71:905–13.
  16. Shinnick TM, Good RC. Mycobacterial taxonomy. *Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol.* 1994;13:884–901.
  17. Zhang Z-Y, Sun Z-Q, Wang Z-L, Hu H-R, Wen Z-L, Song Y-Z, et al. Identification and pathogenicity analysis of a novel non-tuberculous mycobacterium clinical isolate with nine-antibiotic resistance. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis.* 2013;19:91–6.
  18. Bragin EY, Shtratnikova VY, Schelkunov MI, Dovbnaya D V, Donova M V. Genome-wide response on phytosterol in 9-hydroxyandrostenedione-producing strain of Mycobacterium sp. VKM Ac-1817D. *BMC Biotechnol.* 2019;19:39.
  19. Ren L, Jia Y, Ruth N, Qiao C, Wang J, Zhao B, et al. Biodegradation of phthalic acid esters by a newly isolated Mycobacterium sp. YC-RL4 and the bioprocess with environmental samples. *Environ Sci Pollut Res Int.* 2016;23:16609–19.
  20. Tan Y, Deng Y, Yan X, Liu F, Tan Y, Wang Q, et al. Nontuberculous mycobacterial pulmonary disease and associated risk factors in China: A prospective surveillance study. *J Infect.* 2021;83:46–53.
  21. Liu C-F, Song Y-M, He W-C, Liu D-X, He P, Bao J-J, et al. Nontuberculous mycobacteria in China: incidence and antimicrobial resistance spectrum from a nationwide survey. *Infect Dis Poverty.* 2021;10:59.
  22. Zhou L, Xu D, Liu H, Wan K, Wang R, Yang Z. Trends in the Prevalence and Antibiotic Resistance of Non-tuberculous Mycobacteria in Mainland China, 2000-2019: Systematic Review and Meta-

- Analysis. *Front public Heal.* 2020;8:295.
23. Huh HJ, Kim S-Y, Jhun BW, Shin SJ, Koh W-J. Recent advances in molecular diagnostics and understanding mechanisms of drug resistance in nontuberculous mycobacterial diseases. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis.* 2019;72:169–82.
  24. Pang Y, Tan Y, Chen J, Li Y, Zheng H, Song Y, et al. Diversity of nontuberculous mycobacteria in eastern and southern China: a cross-sectional study. *The European respiratory journal.* 2017;49.
  25. Huang J-J, Li Y-X, Zhao Y, Yang W-H, Xiao M, Kudinha T, et al. Prevalence of nontuberculous mycobacteria in a tertiary hospital in Beijing, China, January 2013 to December 2018. *BMC Microbiol.* 2020;20:158.
  26. Chunfang W, Jihong R, Yu W, Yunhong Z, Xuejuan S, Xiuyun J, et al. Prevalence of Nontuberculous Mycobacterial Disease in the Changchun District of China. *Curr Microbiol.* 2021;78:1643–7. doi:10.1007/s00284-021-02422-y.
  27. Hu C, Huang L, Cai M, Wang W, Shi X, Chen W. Characterization of non-tuberculous mycobacterial pulmonary disease in Nanjing district of China. *BMC Infect Dis.* 2019;19:764.
  28. Hernández-Garduño E, Elwood RK. Demographic risk factors of pulmonary colonization by non-tuberculous mycobacteria. *Int J Tuberc lung Dis Off J Int Union against Tuberc Lung Dis.* 2010;14:106–12.
  29. Lipner EM, French JP, Falkinham JO 3rd, Crooks JL, Mercaldo RA, Henkle E, et al. NTM Infection Risk and Trace Metals in Surface Water: A Population-Based Ecologic Epidemiologic Study in Oregon. *Ann Am Thorac Soc.* 2021.
  30. Haig S-J, Kotlarz N, LiPuma JJ, Raskin L. A High-Throughput Approach for Identification of Nontuberculous Mycobacteria in Drinking Water Reveals Relationship between Water Age and *Mycobacterium avium*. *MBio.* 2018;9.
  31. Xu J, Li P, Zheng S, Shu W, Pang Y. Prevalence and risk factors of pulmonary nontuberculous mycobacterial infections in the Zhejiang Province of China. *Epidemiol Infect.* 2019;147.
  32. Finegold I. Nontuberculous mycobacterial infection in the differential diagnosis of chronic cough. *Allergy asthma Proc.* 2008;29:343–4.
  33. Chan ED, Iseman MD. Slender, older women appear to be more susceptible to nontuberculous mycobacterial lung disease. *Gend Med.* 2010;7:5–18.
  34. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014;15:R46.
  35. Martínez González S, Cano Cortés A, Sota Yoldi LA, García García JM, Alba Álvarez LM, Palacios Gutiérrez JJ. Non-Tuberculous Mycobacteria. An Emerging Threat? *Arch Bronconeumol.* 2017;53:554–60.
  36. Schiebelhut LM, Abboud SS, Gómez Daglio LE, Swift HF, Dawson MN. A comparison of DNA extraction methods for high-throughput DNA analyses. *Mol Ecol Resour.* 2017;17:721–9.

37. Khieu V, Ananta P, Kaewprasert O, Laohaviroj M, Namwat W, Faksri K. Whole-Genome Sequencing Analysis to Identify Infection with Multiple Species of Nontuberculous Mycobacteria. *Pathog (Basel, Switzerland)*. 2021;10.
38. Schmeier S. 6.4.Kraken2. <https://genomics.sschmeier.com/ngs-taxonomic-investigation/index.html#kraken2>. Accessed 30 Aug 2021.
39. Yang T, Wang Y, Liu Q, Jiang Q, Hong C, Wu L, et al. A population-based genomic epidemiological study of the source of tuberculosis infections in an emerging city: Shenzhen, China. *Lancet Reg Heal West Pacific*. 2021;8:100106.

## Figures

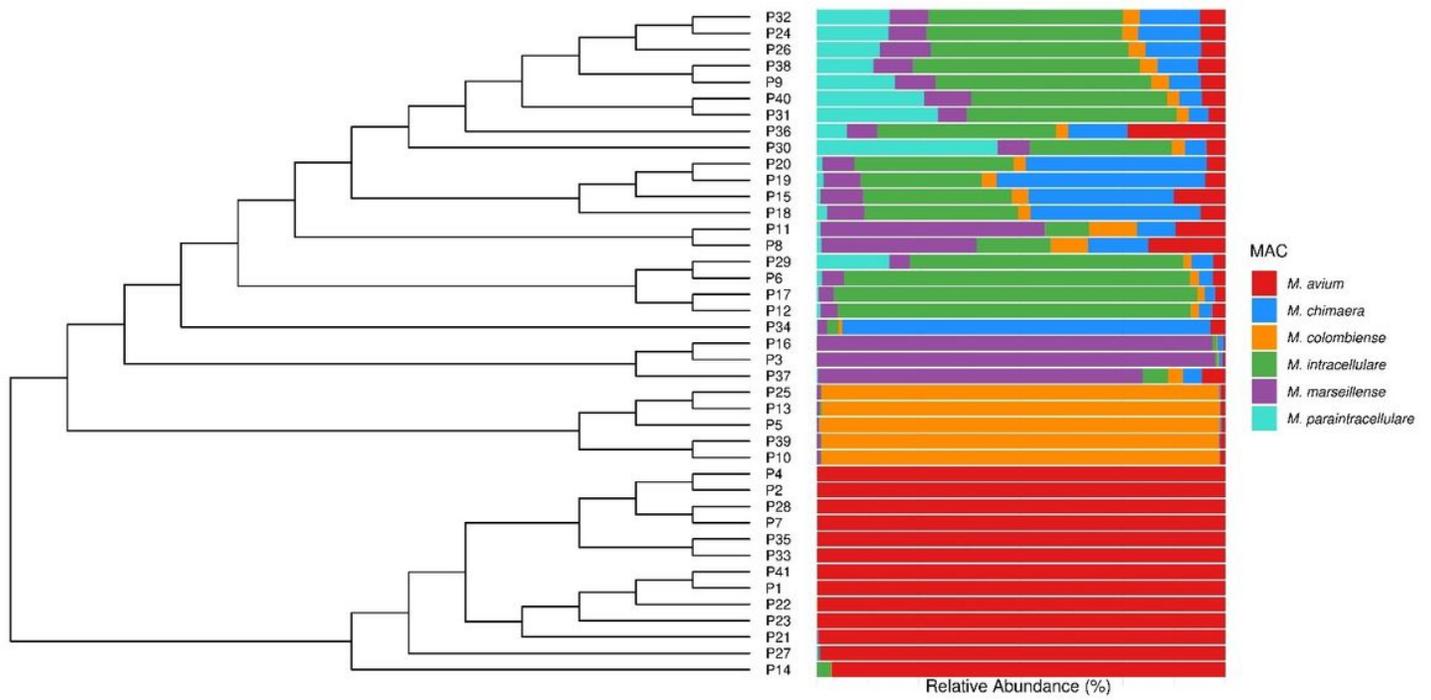


**Figure 1**

Flow map of the inclusion of 1759 cases. MTBC: *Mycobacterium tuberculosis* complex; NTM: Nontuberculous mycobacteria.

\*: The reads of 1622 strains were mainly assigned to the “*Mycobacterium tuberculosis complex*” subgenus by Kraken 2 / Bracken.

\*\* : The reads of 140 strains were mainly assigned to the nontuberculous mycobacteria species by Kraken 2 / Bracken.



**Figure 2**

Identification of *Mycobacterium avium* complex (MAC) isolates, as determined by taxonomic classification using Kraken 2 / Bracken. Each of the 41 isolates is organized vertically and clustered on their Bray-Curtis dissimilarity into two major groups: a predominantly *M. avium* and a mixed group mainly included *M. colombiense*, *M. intracellulare* and *M. marseillense*, as seen on the dendrogram to the left.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)