



## Article

# Whole-Genome Sequencing Analysis to Identify Infection with Multiple Species of Nontuberculous Mycobacteria

Visal Khieu <sup>1,2</sup>, Pimjai Ananta <sup>1,3</sup>, Orawee Kaewprasert <sup>1,2</sup>, Marut Laohaviroj <sup>1,2</sup>, Wisen Namwat <sup>1,2</sup> and Kiaticchai Faksri <sup>1,2,\*</sup>

<sup>1</sup> Department of Microbiology, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand; visal.cn@gmail.com (V.K.); pimjais@kku.ac.th (P.A.); k\_orawee@kkumail.com (O.K.); marut@kku.ac.th (M.L.); wisenam@kku.ac.th (W.N.)

<sup>2</sup> Research and Diagnostic Center for Emerging Infectious Diseases (RCEID), Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand

<sup>3</sup> Clinical Laboratory Unit, Srinagarind Hospital, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand

\* Correspondence: kiaticchai@kku.ac.th

**Abstract:** Mixed infection with multiple species of nontuberculous mycobacteria (NTM) is difficult to identify and to treat. Current conventional molecular-based methods for identifying mixed infections are limited due to low specificity. Here, we evaluated the utility of whole-genome sequencing (WGS) analysis to detect and identify mixed NTM infections. Analytical tools used included PubMLST, MetaPhlan3, Kraken2, Mykrobe-Predictor and analysis of heterozygous SNP frequencies. The ability of each to identify mixed infections of NTM species was compared. Sensitivity was tested using 101 samples (sequence sets) including 100 *in-silico* simulated mixed samples with various proportions of known NTM species and one sample of known mixed NTM species from a public database. Single-species NTM control samples (155 WGS samples from public databases and 15 samples from simulated reads) were tested for specificity. Kraken2 exhibited 100% sensitivity and 98.23% specificity for detection and identification of mixed NTM species with accurate estimation of relative abundance of each species in the mixture. PubMLST (99% and 96.47%) and MetaPhlan3 (95.04% and 83.52%) had slightly lower sensitivity and specificity. Mykrobe-Predictor had the lowest sensitivity (57.42%). Analysis of read frequencies supporting single nucleotide polymorphisms (SNPs) could not detect mixed NTM samples. Clinical NTM samples ( $n = 16$ ), suspected on the basis of a 16S–23S rRNA gene sequence-based line-probe assay (LPA) to contain more than one NTM species, were investigated using WGS-analysis tools. This identified only a small proportion (37.5%, 6/16 samples) of the samples as mixed infections and exhibited only partial agreement with LPA results. LPAs seem to be inadequate for detecting mixed NTM species infection. This study demonstrated that WGS-analysis tools can be used for diagnosis of mixed infections with different species of NTM.

**Keywords:** mixed species infection; nontuberculous mycobacteria; NTM; whole-genome sequencing



**Citation:** 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. *Pathogens* **2021**, *10*, 879. <https://doi.org/10.3390/pathogens10070879>

Academic Editor: Leonardo A. Sechi

Received: 26 April 2021

Accepted: 8 July 2021

Published: 11 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Nontuberculous mycobacteria (NTM) are environmental mycobacteria that have been associated with human diseases since the 1930s [1]. NTMs are mostly opportunistic pathogens, becoming true pathogens in immunocompromised patients such as HIV patients [2]. NTMs are difficult to treat and diagnostically easily confused with the more common disease, tuberculosis. Moreover, mixed infections with different strains or species of NTM can occur [3,4]. Disease status [5] and treatment outcomes in people infected with multiple NTMs can differ from those with individual infections [4,6]. Misdiagnosis and inappropriate treatment is a common critical problem in multiple-NTM infections [7]. Cases of multiple-species NTM infections have been reported in many countries [3,4,8,9]. Such in-

fections may be pulmonary [4], cutaneous [8] or disseminated [10]. Accurate identification of mixed species infection will facilitate appropriate antibiotic treatment.

Conventional methods for NTM species identification include biochemical tests and use of phenotypic characteristics. These methods are time-consuming and sometimes cannot provide accurate identification [11]. Molecular methods based on the line-probe assay (LPA) and real-time PCR can be used for more accurate NTM species identification. In addition, also useful is sequencing of specific target genes including 16S rRNA [12], *rpoB* [13] and the *hsp65* gene [14]. The inability of a single gene to distinguish between genetically similar mycobacterial species has resulted in the use of multigene sequencing methods [15]. In addition, there is no gold standard for detection and identification of multiple species infection of NTM. LPA and multiplex real-time PCR/melting curve analysis have been applied to detect mixed infections with different NTM species based on the 16S-23S rRNA intergenic transcribed spacer (ITS) region and achieved good performance using material directly from clinical samples [16–18]. However, the resolution of these techniques is still low because not all NTM species, especially newly discovered species, have not been included in assay development. Therefore, accurate diagnosis of mixed infection requires use of another high-resolution method for confirmation.

The advent of whole-genome sequencing (WGS) allows the sequencing of the entire genome of microorganisms and has ability to sequence multiple species in samples from one individual [19]. Free WGS-analysis tools for species identification and diagnosis of many pathogens [20], including mycobacteria, are increasingly available. WGS analysis has been successfully used to identify infection caused by multiple strains of *M. tuberculosis* [21,22]. Mixed infection with different species of NTMs have generally been detected only incidentally [23,24]. No study has systematically evaluated and applied bioinformatics analysis approaches for the identification of infections due to multiple species of NTMs.

Here, we aim to evaluate the performance of WGS-analysis approaches for identification of mixed infections of NTMs. We use *in-silico* simulated mixtures with various proportions of known NTM species, WGS analysis of single and mixed NTM species datasets, and datasets from clinical samples of mixed NTM species (as detected by LPA) to evaluate the performance of several WGS-analysis approaches.

## 2. Results

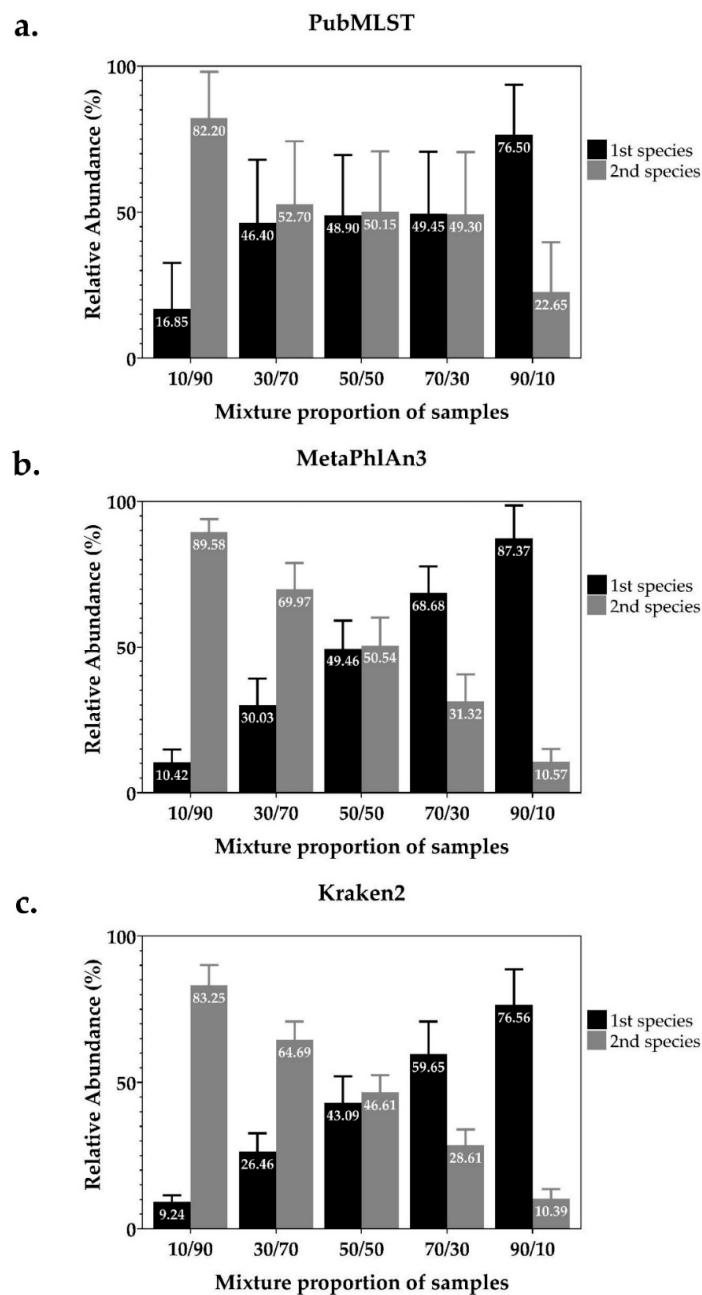
### 2.1. Detection and Identification of Mixed Infections from In-Silico Simulated Mixed-Species NTM Dataset

We used the simulated mixed NTM species dataset (20 mixed samples, each with five different ratios of reads of the two species,  $n = 100$ ) (Table S1), to test the ability of four WGS-analysis tools to recognize mixed infections and identify the species present (Table 1). Kraken2 achieved a perfect score. MetaPhlan3 can detect species present in all 100 samples, but 5 (5%) samples were incorrectly reported as a mixture between *M. tuberculosis* and *M. canettii* that belong to the same species complex. PubMLST correctly identified species represented in 99 samples but failed to identify one sample which consisted of a 10/90 ratio of reads from *M. fortuitum* and *M. peregrinum*. All samples containing the *M. abscessus* complex were correctly identified into subspecies using MAB-MLST (additional tool of PubMLST). Mykrobe-Predictor correctly identified species in only 58 samples (58%), incorrectly claimed that a single species was present in 17 samples (17%) and incorrectly identified the species present in 25 mixed samples (25%), but mostly suggested species within the same species complex. Apart from Mykrobe-Predictor, all tools reported the expected proportion of each species: MetaPhlan3 and Kraken2 were the most accurate (Figure 1).

**Table 1.** Ability of different WGS-analysis approaches to detect NTM species represented in simulated datasets containing reads from two different NTM species in various ratios ( $n = 100$ )<sup>a</sup>.

Species Identification Types	PubMLST ( $n = 100$ )	MetaPhlan3 ( $n = 100$ )	Kraken2 ( $n = 100$ )	Mykrobe-Predictor ( $n = 100$ )
Correct mixed species	99 (99%)	95 (95%)	100 (100%)	58 (58%)
Incorrect mixed species	0	5 (5%) <sup>b</sup>	0	25 (25%)
Single species only	1 (1%)	0	0	17 (17%)

<sup>a</sup> Each combination of species at different proportions was considered as an individual sample in the total of 100 samples. <sup>b</sup> This tool detected *M. tuberculosis* and *M. canettii*, both belonging to the *M. tuberculosis* complex.



**Figure 1.** Average relative abundance (%) of each NTM species presented in simulated mixed samples were estimated by (a) PubMLST, (b) MetaPhlan3 and (c) Kraken2. (a) One sample was identified as a single species. (b) One simulated mixed NTM species sample levels and five samples were excluded.

The performance of the WGS-analysis tools for detection and identification of NTMs was also assessed using WGS data for 15 single-species datasets (processed through the ART simulator) from GenBank (Table S2). Both Kraken2 and PubMLST provided 100% specificity ( $n = 15$ ). Subspecies identification of the *M. abscessus* complex were correctly reported by MAB-MLST in all samples. MetaPhlAn3 and Mykrobe-Predictor correctly identified the single species represented in 14 (93.33%) datasets. The former method claimed that one dataset contained a mixture of *M. canettii* and *M. tuberculosis* and the latter misidentified the species present as *M. intracellulare*.

Besides the four WGS-analysis tools, we tried using the analysis of read frequencies supporting SNPs to identify species in simulated mixed-sequence samples of *M. intracellulare* and *M. abscessus*. Various proportions of the two species were represented in these simulated samples. However, there was no clear pattern of allele frequencies at SNP sites that made it possible to identify the number and proportions of each species represented (Figure S1). This approach was therefore not included in later comparisons.

### 2.2. Detection and Identification of NTM Species in Mixed Infections: Dataset from GenBank

We retrieved publicly available WGS data for one known mixed-species NTM sample (*M. hassiacum* and *M. peregrinum*), obtained from a bird. All WGS-analysis tools correctly reported both species to be present, except that Mykrobe-Predictor identified only one species (*M. hassiacum*) (Table 2). From metagenomic assembly analysis, of the three specific target genes, only the *rpoB* gene was found. This was represented by two sequence types, one with 100% similarity to *M. hassiacum* and one with 98.34% similarity to *M. peregrinum* (Table 2 and Table S3).

**Table 2.** Comparison of WGS-analysis tools for species-level identification of NTMs in a single mixed-species dataset from a previous study (RA%: relative abundance of reads in the dataset).

Accession Number	WGS-Analysis Tools							Metagenomic Assembly Analysis
	PubMLST		MetaPhlAn3		Kraken2		Mykrobe-Predictor	BLAST ( <i>rpoB</i> Gene)
	Species	RA (%)	Species	RA (%)	Species	RA (%)	Species	Species
SRR5043021 [25]	<i>M. hassiacum</i> , <i>M. peregrinum</i>	49 39	<i>M. hassiacum</i> , <i>M. peregrinum</i>	83.67 16.33	<i>M. hassiacum</i> , <i>M. peregrinum</i>	73.41 24.29	<i>M. hassiacum</i> *	<i>M. hassiacum</i> , <i>M. peregrinum</i>

\* Mykrobe-Predictor has no feature for proportion identification.

WGS datasets ( $n = 155$ ), each of a known single mycobacterial species, from a public database were used as single-species controls (Table S4). Kraken2 provided the highest specificity by correctly identifying the species in 152/155 (98.06%) datasets, followed by PubMLST (149/155, 96.12%), Mykrobe-Predictor (149/155, 96.12%) and MetaPhlAn3 (128/155, 82.58%) (Table 3). All four analysis tools correctly identified the same 120/155 (77.41%) samples. Notably, a sample of *M. intermedium* was correctly identified by Kraken2, but the other tools failed to do so. MAB-MLST correctly differentiated subspecies within the *M. abscessus* complex in 10/16 (62.5%) cases as *M. abscessus* and in 9/25 (36%) cases as *M. abscessus* subsp. *massiliense*.

**Table 3.** Ability of WGS-analysis approaches to identify NTM species present (single-species control) in datasets from a public database.

Species Identification Types	PubMLST ( $n = 155$ )	MetaPhlAn3 ( $n = 155$ )	Kraken2 ( $n = 155$ )	Mykrobe-Predictor ( $n = 155$ )
Correct single species	149 (96.12%)	128 (82.58%)	152 (98.06%)	149(96.12%)
Mixed species	4 (2.58%)	26 (16.77%)	3 (1.93%)	0
Incorrect species	1 (0.64%)	0	0	5 * (3.22%)
No result	1 (0.64%)	1 (0.64%)	0	1(0.64%)

\* One sample was only identified to the genus level (*Mycobacterium* sp.).

### 2.3. Comparison of the Four WGS-Analysis Tools for Identifying Mixed Infections with Different NTM Species

The sensitivity and specificity of the four WGS-analysis tools for detection and identification of species present in the simulated mixed NTM datasets and samples from a public database are shown (Table 4). Kraken2 had the highest overall sensitivity and specificity (100%, 98.23%). PubMLST exhibited the second-highest sensitivity and specificity (99%, 96.47%), followed by MetaPhlAn3. The sensitivity of Mykrobe-Predictor was very much lower (Table 4).

**Table 4.** Performance of WGS-analysis tools for identifying multiple species in datasets of reads from NTMs.

Sample Groups	Number of Samples (Correct Species Identification/Total Samples)			
	PubMLST	MetaPhlAn3	Kraken2	Mykrobe-Predictor
Simulated mixed species <sup>a</sup>	99/100	95/100	100/100	58/100
Mixed species from public database	1/1	1/1	1/1	0/1
<b>Sensitivity of tool</b>	<b>99%</b> <b>(100/101)</b>	<b>95.04%</b> <b>(96/101)</b>	<b>100%</b> <b>(101/101)</b>	<b>57.42%</b> <b>(58/101)</b>
Simulated single species control	15/15	14/15	15/15	14/15
Single species control from public database	149/155	128/155	152/155	149/155
<b>Specificity of tool</b>	<b>96.47%</b> <b>(164/170)</b>	<b>83.52%</b> <b>(142/170)</b>	<b>98.23%</b> <b>(167/170)</b>	<b>95.88%</b> <b>(163/170)</b>

<sup>a</sup> Each combination of species at different proportions was considered as individual sample in the total of 100 samples.

### 2.4. Concordance between WGS-Analysis Tools and LPA for Detection and Identification of Mixed NTM Species in Clinical Samples

Sixteen clinical samples of mixed infection with different species of NTM (identified using LPA as the standard) were used to evaluate the performance of WGS-analysis tools (Table 5). MetaPhlAn3 was the most successful at detecting mixed infections (6/16, 37.5%). In three of these cases, MetaPhlAn3 correctly identified one of the species present and in the remaining cases none of the species reported was in agreement with the result from LPA. A further feature of the MetaPhlAn3 result was that, in three datasets, at least one species was represented by fewer than 0.2% of the reads (Table 6). The remaining tools inferred the presence of mixed infections in only a single sample each (6.25%), but without agreement among them. Even when disagreeing with the results of the LPA, the WGS-analysis tools often agreed with each other as to the identity of at least one species in a sample (Table 6).

**Table 5.** List of clinical samples of mixed infection with different species of mycobacteria detected by line-probe assay (LPA).

Sample ID	NTM Species (According to LPA)	Date of Collection	Sex	Age	Specimen Types
MIX80105	<i>M. abscessus</i> subsp. <i>massiliense</i> , <i>M. abscessus</i>	10 January 2014	M	43	Sputum
MIX80487	<i>M. intracellulare</i> , <i>M. scrofulaceum</i>	16 February 2015	M	49	Nasal Cavity (swab)
MIX80628	<i>M. goodii</i> , <i>M. simiae</i>	19 February 2014	M	34	Bone marrow
MIX80885	<i>M. avium</i> , <i>M. intracellulare</i>	12 March 2014	M	63	Sputum
MIX81256	<i>M. kansasii</i> , <i>M. intracellulare</i>	8 April 2014	F	50	Bronchial wash
MIX81523	<i>M. kansasii</i> , <i>M. malmeonse</i>	8 August 2016	F	56	Sputum
MIX81666	<i>M. tuberculosis</i> complex, <i>M. scrofulaceum</i>	8 July 2016	M	69	Sputum
MIX82390	<i>M. intracellulare</i> , <i>M. avium</i>	15 July 2014	M	34	Bone marrow
S12260	<i>M. fortuitum</i> , <i>M. peregrinum</i>	22 August 2016	F	75	Blood
S80510	<i>M. goodii</i> , <i>M. fortuitum</i>	17 February 2015	M	57	Sputum
S81158	<i>M. kansasii</i> , <i>M. scrofulaceum</i>	29 April 2013	M	28	Sputum
S81463	<i>M. simiae</i> , <i>M. fortuitum</i>	29 May 2013	F	61	Sputum
S81801	<i>M. fortuitum</i> , <i>M. peregrinum</i>	27 July 2016	F	47	Pus
S82945	<i>M. intracellulare</i> , <i>M. abscessus</i>	2 October 2013	F	81	Sputum
S83359	<i>M. intracellulare</i> , <i>M. fortuitum</i>	10 October 2014	F	33	Sputum
S83411	<i>M. kansasii</i> , <i>M. avium</i>	5 November 2013	M	75	Sputum

**Table 6.** Comparison between WGS analysis and LPA for detection and identification of mixed NTM species infection in Thai clinical samples.

Sample ID	LPA	WGS-Analysis Tools			
		PubMLST (MAB-MLST) (%RA)	MetaPhlan3 (%RA)	Kraken2 (%RA)	Mykrobe-Predictor
MIX80105	<i>M. abscessus</i> subsp. <i>massiliense</i> , <i>M. abscessus</i>	<i>M. abscessus</i> ( <i>M. abscessus</i> subsp. <i>massiliense</i> )	<i>M. abscessus</i>	<i>M. abscessus</i>	<i>M. abscessus</i>
MIX80487	<i>M. intracellulare</i> , <i>M. scrofulaceum</i>	<i>M. malmoense</i>	<i>M. malmoense</i>	<i>M. malmoense</i>	<i>M. parascrofulaceum</i>
MIX80628	<i>M. gordonae</i> , <i>M. simiae</i>	<i>M. sherrisii</i> , <i>M. asiaticum</i> (96, 3)	<i>M. sherrisii</i> , <i>M. asiaticum</i> , <i>M. simiae</i> (99.83, 0.16, 0.01)	<i>M. sherrisii</i>	<i>M. sherrisii</i>
MIX80885	<i>M. avium</i> , <i>M. intracellulare</i>	<i>M. asiaticum</i>	<i>M. sp. 1165178.9</i> , <i>M. lepraemurium</i> , <i>M. colombiense</i> (99.91, 0.06, 0.03)	<i>M. sp. 1165178.9</i>	<i>M. intracellulare</i>
MIX81256	<i>M. kansasii</i> , <i>M. intracellulare</i>	<i>M. intracellulare</i>	<i>M. intracellulare</i>	<i>M. intracellulare</i>	<i>M. intracellulare</i>
MIX81523	<i>M. kansasii</i> , <i>M. malmoense</i>	<i>M. attenuatum</i>	<i>M. kansasii</i>	<i>M. sp. MK136</i> <sup>a</sup>	<i>M. kansasii</i>
MIX81666	<i>M. tuberculosis</i> complex, <i>M. scrofulaceum</i>	<i>M. tuberculosis</i>	<i>M. canettii</i> <sup>b</sup> , <i>M. sp. E3198</i> , <i>M. sp. 852002-50816</i> , <i>SCH5313054-b</i> (96.70, 1.85, 1.45)	<i>M. tuberculosis</i>	<i>M. tuberculosis</i> , <i>M. intracellulare</i>
MIX82390	<i>M. intracellulare</i> , <i>M. avium</i>	<i>M. scrofulaceum</i>	<i>M. lepraemurium</i> , <i>M. scrofulaceum</i> (75.94, 24.06)	<i>M. sp. ACS4054</i> , <i>M. scrofulaceum</i> (9.23, 7.13)	<i>M. intracellulare</i>
S12260	<i>M. fortuitum</i> , <i>M. peregrinum</i>	<i>M. fortuitum</i>	<i>M. fortuitum</i>	<i>M. fortuitum</i>	<i>M. fortuitum</i>
S80510	<i>M. gordonae</i> , <i>M. fortuitum</i>	<i>M. abscessus</i> (UD)	<i>M. gordonae</i> , <i>M. abscessus</i> (81.26, 18.74)	<i>M. gordonae</i>	<i>M. gordonae</i>
S81158	<i>M. kansasii</i> , <i>M. scrofulaceum</i>	<i>M. attenuatum</i>	<i>M. kansasii</i>	<i>M. sp. MK136</i> <sup>a</sup>	<i>M. kansasii</i>
S81463	<i>M. simiae</i> , <i>M. fortuitum</i>	No result	<i>M. rhodesiae</i>	<i>M. sp. M26</i> <sup>c</sup>	<i>M. farcinogenes</i>
S81801	<i>M. fortuitum</i> , <i>M. peregrinum</i>	<i>M. fortuitum</i>	<i>M. fortuitum</i> , <i>M. abscessus</i> (99.99, 0.01)	<i>M. fortuitum</i>	<i>M. fortuitum</i>
S82945	<i>M. intracellulare</i> , <i>M. abscessus</i>	<i>M. abscessus</i> ( <i>M. abscessus</i> subsp. <i>massiliense</i> )	<i>M. abscessus</i>	<i>M. abscessus</i>	<i>M. abscessus</i>
S83359	<i>M. intracellulare</i> , <i>M. fortuitum</i>	<i>M. fortuitum</i>	<i>M. fortuitum</i>	<i>M. fortuitum</i>	<i>M. fortuitum</i>
S83411	<i>M. kansasii</i> , <i>M. avium</i>	<i>M. persicum</i>	<i>M. persicum</i>	<i>M. persicum</i>	<i>M. kansasii</i>
Single species <sup>d</sup>	0	14 (87.5%)	10 (62.5%)	15 (93.75%)	15 (93.75%)
Mixed species <sup>e</sup>	16(100%)	1 (6.25%)	6 (37.5%)	1 (6.25%)	1 (6.25%)
No result	0	1 (6.25%)	0	0	0
<b>Total</b>	<b>16</b>	<b>16</b>	<b>16</b>	<b>16</b>	<b>16</b>

**Note:** MAB-MLST refers to *M. abscessus* complex-multilocus sequence typing, UD refers to un-differentiable,<sup>a</sup> *M. sp. MK136* reported by the tool refers to *M. attenuatum*,<sup>b</sup> *M. canettii* is a member of the *M. tuberculosis* complex, <sup>c</sup> *M. sp. M26* reported from the tool refers to *M. massiliopolynesiensis*. %RA refers to relative abundance of different species within mixed samples. Red text highlights samples in which the WGS-analysis tool identified multiple species of NTM, and the relative proportions of each estimated by the tool. <sup>d</sup> Number of samples in which the tool claimed only a single species was present. <sup>e</sup> Number of samples in which the tool identified two species of NTM.

The metagenomic assembly analysis could detect only two genes; 16S rRNA and *rpoB*. In the positive control (known mixed NTM dataset), the two sequence types of *rpoB* were found, indicating mixed NTM species. Only one clinical sample (1/16, 6.25%) of suspected mixed-NTM according to LPA yielded the concordant mixed sequence types (*M. tuberculosis* with 100% similarity and *Mycobacterium sp. MOTT-01* or *M. parascrofulaceum* with 94.20% similarity) (Tables S3 and S5).

### 3. Discussion

WGS analysis is increasingly being used for clinical laboratory diagnosis of bacterial infections [25,26] including those due to NTMs [23]. However, no study has previously evaluated WGS-analysis approaches for detection and classification of mixed-species NTM infections. Here, we have demonstrated that particular WGS-analysis tools can be used for this purpose.

In clinical samples, it is difficult to detect mixed infections of NTM species and to identify the species involved. The current approach, using molecular probes specific for a single gene target, has limitations, especially when it comes to species identification [16,17]. The presence of more than one species of NTM in cultured specimens might not be apparent from colonial morphologies. In this study, we analyzed WGS data from cultures identified as including more than one species of NTM, as diagnosed by routine laboratory analysis using LPA.

We assessed preexisting WGS-analysis tools including the metagenomics approaches (MetaPhlan3 and Kraken2), web-based species identification relying on ribosomal multilocus sequence typing (PubMLST) and a drug-resistance prediction tool for *M. tuberculosis* (Mykrobe-Predictor). An analysis pipeline based on the read frequencies supporting SNPs, which has been used to detect mixed strain infection [22], was also included.

There is as yet no gold standard for detection and classification of mixed NTM infections. To ensure that we had accurate identification of the species involved, we produced simulated WGS datasets ( $n = 100$ ) that included reads from various pairs of mycobacterial species ( $n = 15$ ) frequently found in the clinical setting and various ratios of reads from the two species. WGS data ( $n = 155$ ) of single species of common pathogenic NTMs, from a public database, were included for comparison. To ensure the purity of single-species control data, the simulated sequence reads were generated from reference genome data for each species ( $n = 15$ ). A WGS dataset of known mixed NTMs isolated from a bird was also included. Finally, the WGS-analysis tools were assessed using data from clinical samples ( $n = 16$ ) suspected (on the basis of LPA analysis) to contain more than one NTM species.

First, we tried the approach of analyzing SNP allele frequencies to identify the mixed NTM species represented in the simulated dataset. This approach has been successfully used to identify mixed-strain infections of *M. tuberculosis* [22] using WGS data. We could not find appropriate clustering patterns of alleles (Figure S1). We tried to map the sequence reads to reference genomes of different species, but could not identify mixed NTM species. This could be due to the higher diversity among full species compared with subspecies analysis, or to the greater complexity of the genome sequences of NTM compared to *M. tuberculosis* [27].

Expecting that they might perform better, we then evaluated well-known metagenomics analysis tools including Kraken2 and MetaPhlan3. Kraken2 identifies pathogens based on exact *k-mers* alignment [28] and has been successfully applied to detect food-borne and vector-borne pathogens in clinical samples [29,30]. A previous study used Kraken2 to classify NTM isolates into correct species and the results were found to be concordant with conventional PCR and direct sequencing [31]. Here, we found Kraken2 to have high sensitivity (100%) and specificity for detecting and identifying NTMs in mixed infections. Furthermore, Kraken2 correctly identified the proportions of the minor species in the mixture. The specificity of Kraken2 based on single-species controls was 98.23%: misidentification of three species might have been due to limitations of the database associated with the tool or to the presence of closely related species. Both the database and the algorithm of each analysis tool plays a role in its performance for species identification [30,32,33]. The higher performance of Kraken2 might be due to its use of an exact *k-mer* alignment to the *k-mer* of the lowest-common ancestor (LCA) of the taxa and higher specificity due to the default option of  $k = 35$ . However, differentiating subspecies of *M. abscessus* that are associated with different drug-resistance patterns [34,35] remains a challenge for Kraken2. One sample of *M. intermedium* was correctly identified by Kraken2, whereas other tools failed to

do this. Overall, Kraken2 provided the most reliable results for identifying single-species controls.

MetaPhlan3 is another metagenomics tool that can identify and estimate the proportions of members of a microbial community using unique clade-specific marker genes [36]. Previously, this tool has been applied to identify the composition of bacterial species in the intestine and skin [37,38]. However, had no previous study has assessed the performance of this tool for identifying mixed-species NTM infections. Here, we found that MetaPhlan3 had high sensitivity (95.04%). Its main failing was in confusing *M. tuberculosis* with *M. canettii*, species that belong to the same complex. This tool could also reliably estimate the relative abundance of each species represented in the dataset. The specificity of MetaPhlan3 was 83.52%, due in part to its failure to distinguish between closely related species in the same species complex [39]. Similarly, subspecies of *M. abscessus* could not be identified by this tool.

Previously, PubMLST has been used for NTM detection and identified species correctly ( $n = 29$ ) when compared to other methods including MLSTverse [40]. However, this previous study investigated only samples with single NTM species. Here, we demonstrated that PubMLST can be used to identify species in mixed NTM infections with high sensitivity (99%) and specificity (96.47%). In one simulated mixed sample consisting of reads from *M. fortuitum* (10%) and *M. peregrinum* (90%), only the latter species was identified. Six samples of single-species controls were misidentified as incorrect species or as mixed species. In addition, PubMLST could not identify *M. abscessus* at the subspecies level. An additional MLST-based tool is available for this purpose (MAB-MLST [41]). This extended web-based tool provided only moderate performance for identifying the members of the *M. abscessus* complex due to limitations of the specific gene profiles or database.

Mykrobe-Predictor [42] was developed to predict drug-resistance in *M. tuberculosis* and *Staphylococcus aureus*. This tool showed high concordance (96%) with LPA for detecting clinically significant NTM species and incidentally found mixed species in NTM samples [23]. Here, we showed that this tool has moderate performance for identifying multiple NTM species in samples. However, this tool had the worst performance among those we assessed. For example, it misidentified mixed NTM due to *M. intracellulare* and *M. scrofulaceum* (that both belong to the *M. avium* complex) and *M. fortuitum* and *M. peregrinum* (that both belong to the *M. fortuitum* complex) as due to one species in each case. It also gave misidentifications when the proportion of the minor species was as low as 10%. In addition, this tool does not have the ability to detect the proportion of each species represented in the mixture. Thus, Mykrobe-Predictor, while reliable for diagnosis of mycobacteria [23], is not recommended for identification of mixed NTM species.

To extend the analysis to a real-world situation, we used WGS data from 16 clinical samples identified by LPA as likely representing mixed infections of NTM species [43,44]. In our study, MetaPhlan3 detected the highest number (6/16 samples) of mixed NTM infections. However, the species identifications were only partially concordant with the LPA results, or disagreed completely. In addition, in three samples, the proportion of reads from the minor species was assessed as being lower than 0.2% that might not be the actual mixed NTM or clinically insignificant. Notably, the majority (75%) of the suspected mixed-NTM samples detected by LPA were reported to contain only single species by at least three out of the four WGS-analysis tools. Previously, Mykrobe-Predictor concordantly reported a small proportion of mixed species samples (2/25, 8%) compared to LPA [23]. In the majority of our suspected mixed samples (13/16, 81.25%), Mykrobe-Predictor identified only one species that was concordant with the LPA.

We further analyzed the samples using a metagenomic assembly approach. The one known mixed sample obtained from the public database contained two sequence types of *rpoB*, indicating a mixture of NTM species. However, only one sequence type of each gene (16S rRNA and *rpoB* gene) was obtained from most of clinical samples. Only one of our clinical sample was identified as mixed NTM species based on the *rpoB* sequence types found. Our failure to detect the *hsp65* gene from metagenomic assemblies might be due to



the low quality of, or errors in the draft assemblies [45]. In some metagenomic assemblies, we could not detect the 16S rRNA gene. Gene prediction tools typically focus on complete rRNAs, but the 16S rRNA gene is relatively short (>1200 bp) and is commonly fragmented in sequencing reads [46]. Therefore, we concluded that the *rpoB* gene provides the best performance for identification of NTM species using the metagenomic assemblies analysis approach.

Notably, the metagenomics analysis approach was typically in agreement with at least two WGS-analysis tools but only in partial agreement with LPA. Due to the different analysis algorithms, the WGS-analysis approach seems to have higher sensitivity than the metagenomic analysis approach. The results from the latter suggested that there might be only single species of NTM present in suspected mixed-species samples according to LPA. Most of the WGS-analysis tools as well as the metagenomic analysis approach correctly identified the species present in the one known mixed NTM-species sample, which had been confirmed by conventional laboratory methods. In addition, the results from simulated mixed-species datasets indicate that WGS-analysis tools are reliable for detection and identification of mixed NTM species. Taken together, it is likely that LPA has low specificity to detect mixed NTM species in clinical samples. The discordance between LPA and WGS analysis might be due to nature of the genetic targets and their resolution. However, lack of a gold standard hampered our ability to reach a clear conclusion.

Limitations of our study should be noted. A relatively small number ( $n = 16$ ) of clinical samples suspected to be cases of mixed NTM infection was used. Use of Sanger sequencing of the specific target gene(s) to confirm the identities of species in multiple colonies grown from a single clinical sample might be a gold standard to confirm the presence of mixed NTM infections when these are suggested by LPA. However, validation of the LPA results was not included in our study and this is considered a major limitation. Although we used LPA as the routine laboratory method for comparison, it can generate false-positive and false-negative results due to unsuccessful hybridization caused by heterogeneity within the probe-binding site [44]. It also has varying sensitivity for identification of multiple species [47]. Therefore, we used positive and negative controls derived from WGS datasets of known NTM species. A further limitation of LPA is the cross reactivity of some probes for the *M. avium-intracellulare-scrofulaceum* group, *M. fortuitum* complex and *M. intracellulare* Type 2, thus reducing specificity [48]. Although we have illustrated that WGS-analysis tools achieved excellent performance for diagnosis of mixed NTM infections in simulated datasets, we could not clearly illustrate their performance in real clinical isolates. Only a degree of concordance with LPA was demonstrated. Additional investigations should use a higher number of clinical mixed infections with a greater range of NTM species. Identification of species should be confirmed using various methods such as biochemical tests. WGS analysis should also be carried out on samples from colonies spiked with known species and on sequences derived from DNA extracted directly from clinical sample material.

## 4. Materials and Methods

### 4.1. Study Population: Clinical Samples of Mixed NTM Species

Twenty-three clinical culture samples of mixed infection with different species of NTMs from the biobank of Srinagarind Hospital, Khon Kaen Province, Thailand, collected during 2012–2016, were used in this study. The NTM species in these samples were identified by line-probe assay using INNO-LiPA MYCOBACTERIA v2 (INNOGENETICS GmbH, Heiden Germany) according to the manufacturer's protocol [49].

### 4.2. Sample Preparation and WGS of Clinical Samples

Genomic DNA of the 23 samples was extracted from multiple loops of colonies using the cetyl-trimethyl-ammonium bromide-sodium chloride (CTAB) method [50]. All clinical samples of mixed NTM species infection were sent for sequencing by NovogeneAIT, Hong Kong, using the HiSeq (Illumina) platform generating 150-bp paired-end reads.

Unfortunately, sequencing failed for 7 of the 23 samples because there was insufficient material. The characteristics of each sample ( $n = 16$ ) are shown in Table 5.

#### 4.3. In-Silico Simulated Samples Containing Various Proportions of Reads from Different NTM Species

Positive control datasets of mixed species of NTM were simulated from WGS data of known single NTM species. WGS data (FASTA format) from 14 known NTM species and one *M. tuberculosis* strain (HN-506) were retrieved from NCBI Genbank and the European Nucleotide Archive (ENA). These were used to provide the simulated mixed samples ( $2 \times 150$  bp based on Illumina HiSeq) using the ART simulator [51] with various mean depths of coverage ( $10\times$ ,  $30\times$ ,  $50\times$ ,  $70\times$ ,  $90\times$  and  $100\times$ ) across the genome. Then, 100 simulated datasets of mixed NTM species were produced by mixing of sequence reads from two different NTM species, with various percentages of reads from the first and second species (10/90, 30/70, 50/50, 70/30, and 90/10). Simulated reads of individual species ( $n = 15$ ) with a mean coverage of  $100\times$  from the ART simulator were used as non-mixed (single species) samples of mycobacteria. WGS data in FASTA format and simulated mixed species of NTM samples are listed in Table S6.

#### 4.4. WGS Data Samples from a Public Database

The WGS data of 14 species of NTM plus *M. tuberculosis* (in total 155 datasets), sequenced using the Illumina platform, were randomly selected from the Sequence Read Archive (SRA) database. One mixed sample from a bird, which was reported as a mixed infection with different species of NTM [24], was also included. The reads of each dataset were extracted using Fastq-dump from SRA Toolkit version 2.9.1 (<http://ncbi.github.io/sra-tools/> accessed on 15 April 2021) [52]. The list of sequence samples is shown in Table S4.

#### 4.5. Bioinformatics Analysis

##### 4.5.1. QC Check and Data Preparation of Sequence Reads

The quality of sequence reads obtained from clinical samples and the public database were checked by FastQC version 0.11.5 [53]. All sequence reads greater than 75 bp were retained. Reads shorter than 75 bp and potentially contaminating adapter sequences were removed by Trimmomatic version 0.36 [54] using the options LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:75. The filtered reads were then used in the downstream applications.

##### 4.5.2. Detection of Mixed Species of NTM Using Analysis of Read Frequencies Supporting SNP Alleles

This approach used the read frequencies supporting SNPs to distinguish the mixture of species in samples. Because of the common prevalence of mixed NTM infections [4] due to *M. intracellulare* and *M. abscessus*, simulated datasets based on different percentages of reads from these two species (10/90, 30/70, 50/50, 70/30, and 90/10) were selected for analysis using this method. All these datasets were mapped to both *M. abscessus* UC22 (GenBank Accession number: CP012044) and *M. intracellulare* FLAC0181 reference genomes (GenBank Accession number: CP023149.1) using BWA-MEM version 0.7.17 [55]. SAMtools version 0.1.19 [56] was used to convert and sort mapped sequences to the SAM-BAM format. Re-alignment of the mapped reads was done using GATK version 3.4.0 [57]. Variant calling and filtering were then performed to generate the intersection variant set between SAMtools and GATK. SAMtools pileup was used to generate the combined nucleotide frequency for each positional SNP. Outputs of this step were extracted to construct the graph of SNP allele frequencies using GraphPad Prism 8 (GraphPad, San Diego, CA, USA). Mixed infection was visually determined based on the pattern of heterozygous SNP allele frequencies in the samples across genome depending on the number and proportion of species/strains present.

#### 4.5.3. Detection of Mixed Species of NTM Using PubMLST

Before analysis with PubMLST [58], an assembled sequence file is required. De-novo assembly of the filtered reads for all samples was performed using SPAdes version 3.11.1 [59] (<http://cab.spbu.ru/software/spades> accessed on 15 April 2021). The quality of scaffold files was checked using QUAST version 5.0.2 [60] and then the files were used as the input for submission to the PubMLST web interface (<https://pubmlst.org/> accessed on 15 April 2021). For samples apparently containing members of the *M. abscessus* complex, subspecies present were identified using MAB-MLST (<https://pubmlst.org/mabscessus/> accessed on 15 April 2021), which is an additional tool offered by the PubMLST webpage.

#### 4.5.4. Detection of Mixed Species of NTM Using MetaPhlan3

MetaPhlan3 [36] (<https://github.com/biobakery/MetaPhlan/wiki/MetaPhlan-3.0> accessed on 15 April 2021) was used. The paired-end filtered reads of samples were used as the input for the MetaPhlan3 classification analysis with default parameters. The *mpa\_v30\_CHOCOPHlan\_201901* database was used for the analysis.

#### 4.5.5. Detection of Mixed Species of NTM Using Kraken2

Detection of mixed NTM species from samples was done using Kraken2 [28] with the Maxikraken2 database and *k-mers* approach. NTM species identification of each paired-end read was done using the options `-use-names` and `-report`, which provided the taxonomic names associated with each classified sequence and standard ranks for each taxon. Next, Bracken tool [61] was used to estimate genus- and species-level abundance using output from Kraken2 classification results as the input.

#### 4.5.6. Detection of Mixed Species of NTM Using Mykrobe-Predictor

Mykrobe-Predictor [42] (<https://github.com/Mykrobe-tools/mykrobe> accessed on 15 April 2021) was used for mycobacteria species identification. The sequence reads were used to identify mycobacteria species by default parameters.

#### 4.5.7. Detection of Mixed Species of NTM Using Metagenomic Assembly Analysis Based on 16S rRNA, *rpoB* and *hsp65* Genes

The metagenomic assemblies of the samples were called using Prokka 1.13.7 [62]. The sequences of 16S rRNA, *rpoB* and *hsp65* genes used for NTM species identification [14] were extracted and compared to those in the GenBank database using the BLAST algorithm (<http://blast.ncbi.nlm.nih.gov/Blast.cgi> accessed on 15 April 2021). Multiple-sequence alignment (MSA) was performed using Seaview version 5 [63]. The presence of distinctly different sequence types from a single sample in the MSA was taken as evidence of mixed NTM species.

### 4.6. Statistical Analysis

Since there is no gold-standard method for detecting and identifying species in mixed infections of NTM species, simulated sequence reads from each of the two NTM species retrieved from reference and publicly available genome sequences were used as positive controls to calculate sensitivity. Sequence reads based on genomes of single NTM species and sequence reads from public databases were used as negative control to calculate specificity. The sensitivity and specificity of the various WGS-analysis tools for identification of mixed infection with different NTM species were calculated. Sensitivity = true positive/(true positive + false negative). Specificity = true negative/(true negative + false positive).

## 5. Conclusions

We evaluated diagnostic performance of WGS-analysis tools for identification of mixed NTM species infection. Kraken2 provided the highest sensitivity and specificity for this, together with accurate estimation of relative abundance of each species in the samples.

PubMLST and MetaPhlan3 had slightly lower performance but might be useful ancillary methods. LPA seems to have inadequate performance to detect mixed NTM species infection in our study. Accurate species identification will assist choice of appropriate treatment and reduce the mortality rates caused by NTM infection.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/pathogens10070879/s1>, Figure S1: Graph of heterozygous SNP frequencies for five different proportions of simulated mixed NTM species samples (*M. intracellulare* mixed with *M. abscessus*) mapped against each reference genome *M. intracellulare* and *M. abscessus*, respectively; Table S1: simulated datasets of sequence reads representing two NTM species in various proportions and the identifications made by four WGS-analysis tools; Table S2: WGS analysis tools for in silico simulated non-mixed (single species) NTM species identification; Table S3: two different sequence types of *rpoB* gene were identified by metagenomic assembly analysis of a known mixed NTM dataset (SRR5043021) and similar analysis of a Thai clinical sample; Table S4: summary results of WGS analysis for identifying NTM species based on data from a public database; Table S5: comparison between WGS-analysis tools, metagenomic assembly analysis (16S rRNA and *rpoB* gene) and LPA for detection and identification of mixed NTM species infection in known mixed-NTM samples from a public database and Thai clinical samples; Table S6: sources of WGS data and simulated mixed species of NTM samples.

**Author Contributions:** K.F. designed the study and grant management. K.F. supervised V.K. to conduct the experiment. P.A. provided the clinical samples and clinical data. O.K. facilitate the data analysis and result interpretation. M.L. and W.N. assisted the study design, the data analysis and result interpretation. V.K. retrieved the WGS data from public databases and performed data analysis. K.F. and V.K. interpreted the results. K.F. and V.K. wrote the manuscript text. K.F. edited the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** The project was funded by Invitation Research Grant, Faculty of Medicine (Grant No. IN64254) and Research and Diagnostic Center for Emerging Infectious Diseases (RCEID), Khon Kaen University, Thailand.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Khon Kaen University Ethics Committee in Human Research (Ethics Number HE591454).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The WGS data for 16 samples are available in the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra> accessed on 15 April 2021) with the Accession No. PRJNA716753.

**Acknowledgments:** V.K. was supported by the Postgraduate Scholarships for International Student (PSIS) of Faculty of Medicine, Khon Kaen University. We would like to acknowledge David Blair for editing the MS via Publication Clinic KKU, Thailand.

**Conflicts of Interest:** The authors declare that there are no competing interests.

## References

1. Wagner, D.; Young, L. Nontuberculous mycobacterial infections: A clinical review. *Infection* **2004**, *32*, 257–270. [[CrossRef](#)] [[PubMed](#)]
2. Henry, M.; Inamdar, L.; O’riordain, D.; Schweiger, M.; Watson, J. Nontuberculous mycobacteria in non-HIV patients: Epidemiology, treatment and response. *Eur. Respir. J.* **2004**, *23*, 741–746. [[CrossRef](#)]
3. Hirabayashi, R.; Nakagawa, A.; Takegawa, H.; Tomii, K. A case of pleural effusion caused by *Mycobacterium fortuitum* and *Mycobacterium mageritense* coinfection. *BMC Infect. Dis.* **2019**, *19*, 1–3. [[CrossRef](#)] [[PubMed](#)]
4. Shin, S.H.; Jhun, B.W.; Kim, S.-Y.; Choe, J.; Jeon, K.; Huh, H.J.; Ki, C.-S.; Lee, N.Y.; Shin, S.J.; Daley, C.L. Nontuberculous mycobacterial lung diseases caused by mixed infection with *Mycobacterium avium* complex and *Mycobacterium abscessus* complex. *Antimicrob. Agents Chemother.* **2018**, *62*, e01105–18. [[CrossRef](#)] [[PubMed](#)]
5. Lim, H.-J.; Park, C.M.; Park, Y.S.; Lee, J.; Lee, S.-M.; Yang, S.-C.; Yoo, C.-G.; Kim, Y.W.; Han, S.K.; Yim, J.-J. Isolation of multiple nontuberculous mycobacteria species in the same patients. *Int. J. Infect. Dis.* **2011**, *15*, e795–e798. [[CrossRef](#)]
6. Wallace, R.J., Jr.; Zhang, Y.; Brown, B.A.; Dawson, D.; Murphy, D.T.; Wilson, R.; Griffith, D.E. Polyclonal *Mycobacterium avium* complex infections in patients with nodular bronchiectasis. *Am. J. Respir. Crit. Care Med.* **1998**, *158*, 1235–1244. [[CrossRef](#)]

7. Prevots, D.R.; Marras, T.K. Epidemiology of human pulmonary infection with nontuberculous mycobacteria: A review. *Clin. Chest Med.* **2015**, *36*, 13–34. [[CrossRef](#)]
8. Singh, A.K.; Marak, R.S.; Maurya, A.K.; Das, M.; Nag, V.L.; Dhole, T.N. Mixed cutaneous infection caused by *Mycobacterium szulgai* and *Mycobacterium intermedium* in a healthy adult female: A rare case report. *Case Rep. Dermatol. Med.* **2015**, *2015*, 607519. [[CrossRef](#)]
9. Bekou, V.; Büchau, A.; Flaig, M.J.; Ruzicka, T.; Hogardt, M. Cutaneous infection by *Mycobacterium haemophilum* and *kansasii* in an IgA-deficient man. *BMC Dermatol.* **2011**, *11*, 1–5. [[CrossRef](#)]
10. Lévy-Frébault, V.; Pangon, B.; Buré, A.; Katlama, C.; Marche, C.; David, H. *Mycobacterium simiae* and *Mycobacterium avium-M. intracellulare* mixed infection in acquired immune deficiency syndrome. *J. Clin. Microbiol.* **1987**, *25*, 154–157. [[CrossRef](#)] [[PubMed](#)]
11. Häfner, B.; Haag, H.; Geiss, H.-K.; Nolte, O. Different molecular methods for the identification of rarely isolated non-tuberculous mycobacteria and description of new *hsp65* restriction fragment length polymorphism patterns. *Mol. Cell. Probes* **2004**, *18*, 59–65. [[CrossRef](#)]
12. Kirschner, P.; Bottger, E.C. Species identification of mycobacteria using rDNA sequencing. In *Mycobacteria Protocols*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 349–361.
13. Kim, B.-J.; Lee, S.-H.; Lyu, M.-A.; Kim, S.-J.; Bai, G.-H.; Kim, S.-J.; Chae, G.-T.; Kim, E.-C.; Cha, C.-Y.; Kook, Y.-H. Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). *J. Clin. Microbiol.* **1999**, *37*, 1714. [[CrossRef](#)] [[PubMed](#)]
14. Ringuet, H.; Akoua-Koffi, C.; Honore, S.; Varnerot, A.; Vincent, V.; Berche, P.; Gaillard, J.; Pierre-Audigier, C. *hsp65* sequencing for identification of rapidly growing mycobacteria. *J. Clin. Microbiol.* **1999**, *37*, 852. [[CrossRef](#)]
15. Dai, J.; Chen, Y.; Dean, S.; Morris, J.G.; Salfinger, M.; Johnson, J.A. Multiple-genome comparison reveals new loci for *Mycobacterium* species identification. *J. Clin. Microbiol.* **2011**, *49*, 144. [[CrossRef](#)]
16. Hwang, S.M.; Lim, M.S.; Hong, Y.J.; Kim, T.S.; Park, K.U.; Song, J.; Lee, J.H.; Kim, E.C. Simultaneous detection of *Mycobacterium tuberculosis* complex and nontuberculous mycobacteria in respiratory specimens. *Tuberculosis* **2013**, *93*, 642–646. [[CrossRef](#)]
17. Xu, Y.; Liang, B.; Du, C.; Tian, X.; Cai, X.; Hou, Y.; Li, H.; Zheng, R.; Li, J.; Liu, Y. Rapid identification of clinically relevant *Mycobacterium* species by multicolor melting curve analysis. *J. Clin. Microbiol.* **2019**, *57*, e01096-18. [[CrossRef](#)] [[PubMed](#)]
18. Kim, S.H.; Shin, J.H. Identification of Nontuberculous Mycobacteria from Clinical Isolates and Specimens using AdvanSure Mycobacteria GenoBlot Assay. *Jpn. J. Infect. Dis.* **2020**, *73*, 278–281. [[CrossRef](#)] [[PubMed](#)]
19. Deurenberg, R.H.; Bathoorn, E.; Chlebowicz, M.A.; Couto, N.; Ferdous, M.; García-Cobos, S.; Kooistra-Smid, A.M.; Raangs, E.C.; Rosema, S.; Veloo, A.C. Application of next generation sequencing in clinical microbiology and infection prevention. *J. Biotechnol.* **2017**, *243*, 16–24. [[CrossRef](#)]
20. Lindgreen, S.; Adair, K.L.; Gardner, P.P. An evaluation of the accuracy and speed of metagenome analysis tools. *Sci. Rep.* **2016**, *6*, 1–14. [[CrossRef](#)]
21. Gan, M.; Liu, Q.; Yang, C.; Gao, Q.; Luo, T. Deep whole-genome sequencing to detect mixed infection of *Mycobacterium tuberculosis*. *PLoS ONE* **2016**, *11*, e0159029. [[CrossRef](#)]
22. Sobkowiak, B.; Glynn, J.R.; Houben, R.M.; Mallard, K.; Phelan, J.E.; Guerra-Assunção, J.A.; Banda, L.; Mzembe, T.; Viveiros, M.; McNerney, R. Identifying mixed *Mycobacterium tuberculosis* infections from whole genome sequence data. *BMC Genom.* **2018**, *19*, 1–15. [[CrossRef](#)]
23. Quan, T.P.; Bawa, Z.; Foster, D.; Walker, T.; Del Ojo Elias, C.; Rathod, P.; Iqbal, Z.; Bradley, P.; Mowbray, J.; Walker, A.S. Evaluation of whole-genome sequencing for mycobacterial species identification and drug susceptibility testing in a clinical setting: A large-scale prospective assessment of performance against line probe assays and phenotyping. *J. Clin. Microbiol.* **2018**, *56*, e01480-17. [[CrossRef](#)]
24. Pfeiffer, W.; Braun, J.; Burchell, J.; Witte, C.L.; Rideout, B.A. Whole-genome analysis of mycobacteria from birds at the San Diego Zoo. *PLoS ONE* **2017**, *12*, e0173464. [[CrossRef](#)]
25. Hasman, H.; Saputra, D.; Sicheritz-Ponten, T.; Lund, O.; Svendsen, C.A.; Frimodt-Møller, N.; Aarestrup, F.M. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *J. Clin. Microbiol.* **2014**, *52*, 139–146. [[CrossRef](#)]
26. Pankhurst, L.J.; Del Ojo Elias, C.; Votintseva, A.A.; Walker, T.M.; Cole, K.; Davies, J.; Fermont, J.M.; Gascoyne-Binzi, D.M.; Kohl, T.A.; Kong, C. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: A prospective study. *Lancet Respir. Med.* **2016**, *4*, 49–58. [[CrossRef](#)]
27. Fedrizzi, T.; Meehan, C.J.; Grottola, A.; Giacobazzi, E.; Serpini, G.F.; Tagliazucchi, S.; Fabio, A.; Bettua, C.; Bertorelli, R.; De Sanctis, V. Genomic characterization of nontuberculous mycobacteria. *Sci. Rep.* **2017**, *7*, 1–14. [[CrossRef](#)] [[PubMed](#)]
28. Wood, D.E.; Lu, J.; Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 1–13. [[CrossRef](#)] [[PubMed](#)]
29. Doster, E.; Rovira, P.; Noyes, N.R.; Burgess, B.A.; Yang, X.; Weinroth, M.D.; Linke, L.; Magnuson, R.; Boucher, C.; Belk, K.E. A cautionary report for pathogen identification using shotgun metagenomics; a comparison to aerobic culture and polymerase chain reaction for *Salmonella enterica* identification. *Front. Microbiol.* **2019**, *10*, 2499. [[CrossRef](#)] [[PubMed](#)]
30. Vijayvargiya, P.; Jeraldo, P.R.; Thoendel, M.J.; Greenwood-Quaintance, K.E.; Esquer Garrigos, Z.; Sohail, M.R.; Chia, N.; Pritt, B.S.; Patel, R. Application of metagenomic shotgun sequencing to detect vector-borne pathogens in clinical blood samples. *PLoS ONE* **2019**, *14*, e0222915. [[CrossRef](#)] [[PubMed](#)]

31. Yoon, J.-K.; Kim, T.S.; Kim, J.-I.; Yim, J.-J. Whole genome sequencing of Nontuberculous Mycobacterium (NTM) isolates from sputum specimens of co-habiting patients with NTM pulmonary disease and NTM isolates from their environment. *BMC Genomics* **2020**, *21*, 1–7. [[CrossRef](#)]
32. Breitwieser, F.P.; Lu, J.; Salzberg, S.L. A review of methods and databases for metagenomic classification and assembly. *Brief. Bioinforma.* **2019**, *20*, 1125–1136. [[CrossRef](#)]
33. Miossec, M.J.; Valenzuela, S.L.; Pérez-Losada, M.; Johnson, W.E.; Crandall, K.A.; Castro-Nallar, E. Evaluation of computational methods for human microbiome analysis using simulated data. *PeerJ* **2020**, *8*, e9688. [[CrossRef](#)]
34. Griffith, D.E.; Aksamit, T.; Brown-Elliott, B.A.; Catanzaro, A.; Daley, C.; Gordin, F.; Holland, S.M.; Horsburgh, R.; Huitt, G.; Iademarco, M.F. An official ATS/IDSA statement: Diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am. J. Respir. Crit. Care Med.* **2007**, *175*, 367–416. [[CrossRef](#)]
35. Haworth, C.S.; Banks, J.; Capstick, T.; Fisher, A.J.; Gorsuch, T.; Laurenson, I.F.; Leitch, A.; Loebinger, M.R.; Milburn, H.J.; Nightingale, M. British Thoracic Society guidelines for the management of non-tuberculous mycobacterial pulmonary disease (NTM-PD). *Thorax* **2017**, *72*, ii1–ii64. [[CrossRef](#)]
36. Beghini, F.; McIver, L.J.; Blanco-Míguez, A.; Dubois, L.; Asnicar, F.; Maharjan, S.; Mailyan, A.; Thomas, A.M.; Manghi, P.; Valles-Colomer, M. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *bioRxiv* **2020**, *10*, e65088.
37. Zinkernagel, M.S.; Zysset-Burri, D.C.; Keller, I.; Berger, L.E.; Leichtle, A.B.; Largiadèr, C.R.; Fiedler, G.M.; Wolf, S. Association of the intestinal microbiome with the development of neovascular age-related macular degeneration. *Sci. Rep.* **2017**, *7*, 1–9. [[CrossRef](#)]
38. Rebollar, E.A.; Gutiérrez-Preciado, A.; Noecker, C.; Eng, A.; Hughey, M.C.; Medina, D.; Walke, J.B.; Borenstein, E.; Jensen, R.V.; Belden, L.K. The skin microbiome of the neotropical frog *Craugastor fitzingeri*: Inferring potential bacterial-host-pathogen interactions from metagenomic data. *Front. Microbiol.* **2018**, *9*, 466. [[CrossRef](#)] [[PubMed](#)]
39. Benjak, A.; Avanzi, C.; Benito, Y.; Breyse, F.; Chartier, C.; Boschirolì, M.-L.; Fourichon, C.; Michelet, L.; Pin, D.; Flandrois, J.-P. Highly reduced genome of the new species *Mycobacterium uberis*, the causative agent of nodular thelitis and tuberculoid scrotoitis in livestock and a close relative of the leprosy bacilli. *Msphere* **2018**, *3*, e00405-18. [[CrossRef](#)] [[PubMed](#)]
40. Matsumoto, Y.; Kinjo, T.; Motooka, D.; Nabeya, D.; Jung, N.; Uechi, K.; Horii, T.; Iida, T.; Fujita, J.; Nakamura, S. Comprehensive subspecies identification of 175 nontuberculous mycobacteria species based on 7547 genomic profiles. *Emerg. Microbes Infect.* **2019**, *8*, 1043–1053. [[CrossRef](#)]
41. Wuzinski, M.; Bak, A.K.; Petkau, A.; Demczuk, W.H.B.; Soualhia, H.; Sharma, M.K. A multilocus sequence typing scheme for *Mycobacterium abscessus* complex (MAB-multilocus sequence typing) using whole-genome sequencing data. *Int. J. Mycobacteriol.* **2019**, *8*, 273.
42. Bradley, P.; Gordon, N.C.; Walker, T.M.; Dunn, L.; Heys, S.; Huang, B.; Earle, S.; Pankhurst, L.J.; Anson, L.; De Cesare, M. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat. Commun.* **2015**, *6*, 1–15. [[CrossRef](#)]
43. Mijs, W.; De Vreese, K.; Devos, A.; Pottel, H.; Valgaeren, A.; Evans, C.; Norton, J.; Parker, D.; Rigouts, L.; Portaels, F. Evaluation of a commercial line probe assay for identification of *Mycobacterium* species from liquid and solid culture. *Eur. J. Clin. Microbiol. Infect. Dis.* **2002**, *21*, 794–802. [[CrossRef](#)]
44. Sarkola, A.; Mäkinen, J.; Marjamäki, M.; Marttila, H.; Viljanen, M.; Soini, H. Prospective evaluation of the GenoType assay for routine identification of mycobacteria. *Eur. J. Clin. Microbiol. Infect. Dis.* **2004**, *23*, 642–645. [[CrossRef](#)]
45. Jung, H.; Ventura, T.; Chung, J.S.; Kim, W.-J.; Nam, B.-H.; Kong, H.J.; Kim, Y.-O.; Jeon, M.-S.; Eyun, S.-I. Twelve quick steps for genome assembly and annotation in the classroom. *PLoS Comput. Biol.* **2020**, *16*, e1008325. [[CrossRef](#)]
46. Huang, Y.; Gilna, P.; Li, W. Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **2009**, *25*, 1338–1340. [[CrossRef](#)]
47. Scarparo, C.; Piccoli, P.; Rigon, A.; Ruggiero, G.; Nista, D.; Piersimoni, C. Direct identification of mycobacteria from MB/BacT alert 3D bottles: Comparative evaluation of two commercial probe assays. *J. Clin. Microbiol.* **2001**, *39*, 3222–3227. [[CrossRef](#)] [[PubMed](#)]
48. Tortoli, E.; Mariottini, A.; Mazzarelli, G. Evaluation of INNO-LiPA MYCOBACTERIA v2: Improved reverse hybridization multiple DNA probe assay for mycobacterial identification. *J. Clin. Microbiol.* **2003**, *41*, 4418–4420. [[CrossRef](#)] [[PubMed](#)]
49. García-Agudo, L.; Jesús, I.; Rodríguez-Iglesias, M.; García-Martos, P. Evaluation of INNO-LiPA mycobacteria v2 assay for identification of rapidly growing mycobacteria. *Braz. J. Microbiol.* **2011**, *42*, 1220–1226. [[CrossRef](#)] [[PubMed](#)]
50. Larsen, M.H.; Biermann, K.; Tandberg, S.; Hsu, T.; Jacobs, W.R., Jr. Genetic manipulation of *Mycobacterium tuberculosis*. *Curr. Protoc. Microbiol.* **2007**, *6*, 10A-2. [[CrossRef](#)] [[PubMed](#)]
51. Huang, W.; Li, L.; Myers, J.R.; Marth, G.T. ART: A next-generation sequencing read simulator. *Bioinformatics* **2012**, *28*, 593–594. [[CrossRef](#)]
52. Team, S.T.D. Available online: <http://ncbi.github.io/sra-tools/> (accessed on 2 March 2021).
53. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*; Babraham Institute: Cambridge, UK, 2010.
54. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
55. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.

56. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
57. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytzky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)]
58. Jolley, K.A.; Maiden, M.C. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinform.* **2010**, *11*, 1–11. [[CrossRef](#)] [[PubMed](#)]
59. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)]
60. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUASt: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)]
61. Lu, J.; Breitwieser, F.P.; Thielen, P.; Salzberg, S.L. Bracken: Estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **2017**, *3*, e104. [[CrossRef](#)]
62. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)]
63. Gouy, M.; Tannier, E.; Comte, N.; Parsons, D.P. Seaview Version 5: A Multiplatform Software for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation. In *Multiple Sequence Alignment*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 241–260.