ACS ENVIRONMENTAL Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Open Access
This article is licensed under CC-BY-NC-ND 4.0

pubs.acs.org/environau Review

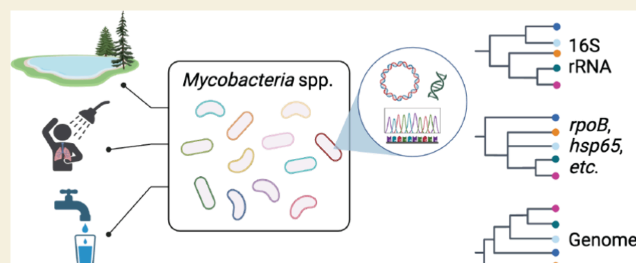
Toward Characterizing Environmental Sources of Non-tuberculous Mycobacteria (NTM) at the Species Level: A Tutorial Review of NTM Phylogeny and Phylogenetic Classification

Lin Zhang,[§] Tzu-Yu Lin,[§] Wen-Tso Liu, and Fangqiong Ling*

Cite This: *ACS Environ. Au* 2024, 4, 127−141 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Nontuberculous mycobacteria (NTM) are any mycobacteria that do not cause tuberculosis or leprosy. While the majority of NTM are harmless and some of them are considered probiotic, a growing number of people are being diagnosed with NTM infections. Therefore, their detection in the environment is of interest to clinicians, environmental microbiologists, and water quality researchers alike. This review provides a tutorial on the foundational approaches for taxonomic classifications, with a focus on the phylogenetic relationships among NTM revealed by the 16S rRNA gene, *rpoB* gene, and *hsp65* gene, and by genome-based approaches. Recent updates on the *Mycobacterium* genus taxonomy are also provided. A synthesis on the habitats of 189 mycobacterial species in a genome-based taxonomy framework was performed, with attention paid to environmental sources (e.g., drinking water, aquatic environments, and soil). The 16S rRNA gene-based classification accuracy for various regions was evaluated (V3, V3−V4, V3−V5, V4, V4−V5, and V1−V9), revealing overall excellent genus-level classification (up to 100% accuracy) yet only modest performance (up to 63.5% accuracy) at the species level. Future research quantifying NTM species in water systems, determining the effects of water treatment and plumbing conditions on their variations, developing high throughput species-level characterization tools for use in the environment, and incorporating the characterization of functions in a phylogenetic framework will likely fill critical knowledge gaps. We believe this tutorial will be useful for researchers new to the field of molecular or genome-based taxonomic profiling of environmental microbiomes. Experts may also find this review useful in terms of the selected key findings of the past 30 years, recent updates on phylogenomic analyses, as well as a synthesis of the ecology of NTM in a phylogenetic framework.



KEYWORDS: *Non-tuberculous Mycobacteria, Drinking Water, Phylogeny, Taxonomy, Ecology, 16S rRNA, rpoB, hsp65, Genome*

1. INTRODUCTION

Nontuberculous mycobacteria (NTM) are a diverse group of mycobacterial species that are ubiquitous in the environment. NTM can refer to any mycobacteria that do not cause tuberculosis or leprosy. While the majority of NTM are harmless, with some even considered probiotic (e.g., *Mycobacterium vaccae*[1,2]), many NTM species can lead to infections in populations with pre-existing conditions. Unlike the decline of *Mycobacterium tuberculosis* infections, the NTM infection rate is rising globally.[3] Different NTM species can have different clinical manifestations, such as NTM pulmonary diseases (NTM-PD), skin infections, cardiac infections, bone and joint infections, and lymphadenitis.[4] NTM-PD account for the majority of the NTM diseases, and the *M. avium* complex (MAC), the *M. abscessus* (MAB) complex, and *M. kansasii*, *M. fortuitum*, *M. gordonae*, *M. xenopi*, *M. chelonae*, and *M. malmoense* are the most common causes of NTM-PD.[4] This broad variety of NTM species can also differ in their susceptibility to antimicrobials.[5] Thus, correct species identification of NTM is clinically important.[5]

A better understanding of the environmental loading, dispersal, and selection of NTM can greatly contribute to curbing NTM diseases. Unlike *M. tuberculosis*, NTM infections are not typically transmitted person to person. Instead, NTM infections are usually associated with environmental exposure. NTM have been found in soil, aerosols, and water, including municipal water sources, such as those in households and healthcare facilities.[6,7] Activities such as farming, swimming, and public bath use can prolong exposure to NTM and increase the risk of infection.[5] Thus, characterizing NTM in the environment, in particular, at the species level, is critical to determining the presence of clinically relevant NTM species, as

Received: December 1, 2023
Revised: January 27, 2024
Accepted: January 29, 2024
Published: February 23, 2024

ACS Publications © 2024 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acsenvironau.3c00074
ACS Environ. Au 2024, 4, 127−141

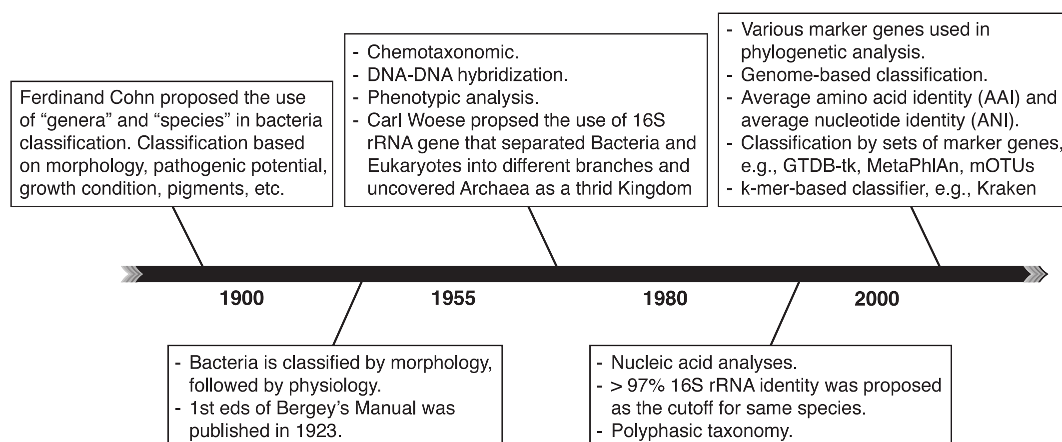


Figure 1. Development of prokaryotic taxonomic classification over time.

well as to prioritizing the targets in further dose−response relationship studies. Equally important, several NTM species have exhibited different sensitivities to disinfection[8] and varied potentials to aerosolize.[9] Geographical variations in species causing disease have been reported, though not fully understood.[10] Thus, improved understanding of the dispersal paths and selective forces that lead to the proliferations of NTM species will play an important role in controlling the spread of NTM in the built environment.

Using up-to-date taxonomy is critical in accurately reporting the presence and abundance of NTM species in the environment. Various culture-dependent and culture-independent methods have been applied to characterize NTM in the environment. Often, the choice of a method assumes certain rules to classify the microorganisms. Even within culture-independent methods, such as those involving the analyses of marker genes or metagenomes, analyses of the same environmental samples using different tools can lead to varied and sometimes conflicting taxonomic classifications. While some of the discrepancies are associated with the choices of underlying algorithms or parameters, other changes can arise from the way the taxonomy is done. There are continuous updates to the phylogeny of mycobacteria, and the picture of NTM taxonomy is constantly evolving, which can broadly influence the nomenclature used in scientific communications within and across disciplines. Such a dynamic situation motivates this tutorial review of the fundamental methodology and milestones in phylogenetic analyses as well as of recent updates in NTM taxonomy.

Readers should be aware of recent debates on the genus name of *Mycobacterium*, and also appreciate that taxonomy as a branch of science is an ongoing process that can integrate new findings and various opinions.[11−14] This review has four specific aims: (1) to introduce approaches and findings of the phylogeny of mycobacteria; (2) to highlight discoveries made by recent phylogenomic analyses; (3) to provide an analysis of the association between the ecology and the phylogenetic relationships among nontuberculous mycobacteria; and (4) to provide a reflection on the ability of the 16S rRNA gene and other markers to detect mycobacteria in the environment. Because numerous insightful reviews have discussed prokaryotic taxonomy,[15,16] the NTM ecology,[17] and NTM's positive and negative implications for public health,[18] we do our best here to provide only the necessary background. By referring to further sources when appropriate, we hope that interested readers can pursue these additional materials. We believe this tutorial will be useful for researchers new to the field of molecular or genome-based taxonomic profiling of environmental microbiomes. Experts may also find this review useful in terms of the selected key findings of the past 30 years, recent updates on phylogenomic analyses, and our synthesis of the ecology of NTM in a phylogenetic framework.

2. THE CLASSIFICATION OF MICROORGANISMS AS AN EVOLVING DISCUSSION

Unlike the naming of chemical elements, a relatively mature process, the classification and naming of microorganisms are evolving efforts. Taxonomy in biology is commonly defined as a branch of science that names and classifies organisms based on their shared properties. The fundamental tasks of taxonomy involve: (i) classification, (ii) nomenclature, and (iii) identification of new organisms. While there are many approaches to classifying organisms, it is generally agreed that taxonomy should be based on evolutionary relationships as the most natural way of arranging organisms.[20] Kämpfer and Glaeser asserted that "in microbiology, the ultimate ambition would be to establish a system that mirrors the taxonomic relationships as an "order in nature", which is now most often associated with "evolutionary order" back to the origin of life".[21] The journey of microbe classifications has progressed through phenotype-based classification, phylogenetic classification, polyphasic approaches, and ongoing developments of genome-based taxonomy (Figure 1).

2.1. Phenotype-Based Classification

Back in the 1870s, new concepts in the study of microorganisms opened the era of modern microbiology.[22] During this time, influenced by Darwin's theory, Ferdinand Cohn introduced the concept of species and genera in bacteria. Further, Cohn classified bacteria based on morphology, inspiring microbial taxonomy based on phenotypic classification, the arrangement of microorganisms by phenotypes, i.e., the observed properties of the organism that result from the expression of genes. For example, the first edition of Bergey's manual (1923) hierarchically classified microorganisms in different ranks, and the keys relied heavily on morphology, culturing conditions, and pathogenic characteristics.[15,23] Phenotypic classification remained commonly accepted, until it was acknowledged that its resolution of the evolutionary relationships of microorganisms was limited. This limitation

https://doi.org/10.1021/acsenvironau.3c00074
ACS Environ. Au 2024, 4, 127−141

even led to arguments in the 1940s-1960s that developing an evolution-based system to classify bacteria would be futile.[15,24]

It should be noted that phenotypic traits also include characteristics of the structural components of prokaryotic cells, based on which chemotaxonomic methods were developed, keying on such as DNA G+C contents, lipids, isoprenoid quinones, cytochromes, cell walls, and peptidoglycan.[21] While these methods rely on expressed genes and thus can be considered as "phenotypic", they provide information that is more consistent and robust, and they are widely used in modern taxonomic classification. For example, DNA−DNA hybridization is widely used in species classification. In terms of their broad application in taxonomy, however, these methods, like other phenotypic methods, can be applied only to cultured organisms, not to the vast diversity of yet-uncultured organisms.[16] In addition, like other phenotypic methods, relationships between organisms are inferred based on similarity (phenetic relationships), whereas truly interrogating evolutionary relationships requires methods that can be applied to inferring common ancestors (cladistic relationships).[16]

2.2. Phylogenetic Classification

Phylogeny refers to the evolutionary relationships between organisms. Young contended that the term "phylogenetic classification" was rarely defined precisely and asserted that "the central outcome of phylogenetic classification is that taxa be monophyletic."[25] Analysis of sequence data in the framework of evolutionary models makes it possible to make inferences about common ancestry. It is worth noting that making use of sequence data in classification does not mean the inferred relationships are cladistic because the method used for inference also matters.[19] Phenetic relationships are measured by similarities or dissimilarities, and cladistic relationships are measured by time or, in the absence of time, by topological units.[16,19]

The monumental work by Woese and Fox was the first to use small subunit rRNA as a useful molecular clock for inferring evolutionary relationships among prokaryotes.[26] Through this new lens, Woese and colleagues discovered *Archaea* as an entire Domain of Life that had previously been overlooked. Pace and colleagues further developed the clone library method, in which the 16S rRNA gene is first amplified and then sequenced, that enabled the discovery of numerous microbial diversities in environments previously overlooked by culture-based methods.[27,28] Studies utilizing 16S rRNA clone libraries, and later the amplicon sequencing technologies, provided an outline of the Tree of Life.[29] Comparative analysis of the 16S rRNA gene enabled the development of the first practical hierarchical classification system based on one molecular marker.[16] By the 1990s, Bergey's Manual had transitioned from a phenotypic-based classification to a 16S rRNA-based phylogenetic framework.[30]

A major criticism of taxonomic classification based on the 16S rRNA gene is its lack of resolution below the genus level. Stackebrandt and Goebel showed that DNA−DNA hybridization and the 16S rRNA gene correlated well below an rRNA sequence similarity of 97.5%, whereas the relationship above that value can vary largely.[31] Other criticisms involve the missing diversity caused by primer mismatches and PCR-produced chimera sequences which can corrupt tree topologies.[32−34] It was hence proposed that data in addition to 16S rRNA gene sequences, such as chemotaxonomic analysis or genome sequences, be used in species classification.[21,35]

2.3. Polyphasic Approach

A polyphasic approach integrates phenotypic and genotypic (genomic) data to perform classification. Young analyzed the literature on the polyphasic approach and articulated two different ideas behind the way multifaceted data is used in polyphasic classification.[25] On one hand, the definition by Colwell, which introduced the term "polyphasic taxonomy", refers to classifications based on a consensus of all available methods.[25,36] On the other hand, Vandamme and colleagues proposed polyphasic taxonomy based on a consensus of data gathered by all available methods 'that would be consistent with phylogenetic classification'.[25,37] Young further ventured that Vandamme and colleagues "may have interpreted polyphasic taxonomy as meaning that, at the higher levels of phylum, division, class, order and family, taxa are ordered in phylogenetic terms (based on sequence analysis), but at lower levels, at species and perhaps genus levels, taxa are established as phenetic groups based on a consensus of phenotypic data."[25]

2.4. Genome-Based Classification

With progress in sequencing technologies, important advances have been made toward genome-based taxonomic classification. First and foremost, systematic sequencing of type strains has improved the availability of genome resources. For instance, the Genomic Encyclopedia of Bacteria and Archaea (GEBA) and GCM 2.0 focused on type strains, the taxonomic reference materials of prokaryotic species, and generated genome sequences for over 4,000 type strains.[38,39] These efforts, as well as genome sequences gained from new organism descriptions, have grown the number of sequenced type strains to over 18,000.[40] Second, genome-based species definitions, such as Average Nucleotide Identity (ANI), Average Amino Acid Identity (AAI), and digital DNA:DNA hybridization (dDDH), have been proved to be well-correlated with DNA:DNA hybridization, and have subsequently become widely used for species demarcation.[35,41,42] Multiple independent analyses of genomes, as well as environmental genomes—metagenome-assembled genomes (MAGs) or sequence-assembled genomes (SAGs)—further validated a discontinuity in bacterial genome similarity at ∼95% ANI, the previously proposed species boundary.[43] Phylogenetic trees constructed from sequences of single-copy, vertically inherited proteins provide higher resolutions than those built from single phylogenetic marker genes, and have been proposed as a reference for taxonomy.[44,45] Parks and colleagues proposed the Genome Taxonomy database (GTDB) taxonomy, based on a phylogeny inferred from the concatenation of 120 ubiquitous single-copy marker proteins.[44] The GTDB framework addressed polyphyletic groups in previous sequenced-based taxonomy.[44] In addition, they performed rank normalization based on Relative Evolutionary Distances (RED), derived from the reference tree.[46] Initially covering 94,759 bacterial genomes, the GTDB has grown to span 317,542 genomes organized into 65,703 species clusters.[47] By using GTDB-tk, a toolkit to compare newly acquired genomes to GTDB references, the genome-based taxonomy framework can be applied to yet-uncultured organisms based on the MAGs or SAGs recovered from the environment.[46] These tools allow automated classifications of newly acquired genomes in a consistent taxonomic framework.

https://doi.org/10.1021/acsenvironau.3c00074
ACS Environ. Au 2024, 4, 127−141

2.5. Benefits of Embracing Taxonomy Updates for Environmental Science and Engineering (ESE) Researchers

The National Academies encourages a convergence approach, which is built on the belief that many societal challenges can be addressed only in a truly transdisciplinary manner.[48] Taxonomy provides a common language among scientists across disciplines. New microorganisms are being discovered every year, and the names of previously discovered organisms can also be revised. Embracing an updated taxonomy of microorganisms can be beneficial in several ways. First, it will help ESE researchers to better utilize resources developed in evolutionary biology and clinical microbiology such as sequence-based or metagenomic taxonomic classification databases. Second, having a common language will be critical for the development of novel molecular and metagenomic tools for the environmental monitoring of organisms of concern. For the results of environmental monitoring to be used across disciplinary boundaries, they must be reported in a common language.[49] Third, the evolving taxonomy of opportunistic pathogens, such as the discovery of new virulent species or subspecies, will likely affect the applicability of existing quantitative microbial risk assessment (QMRA) frameworks, thus create demands for continued research.[50] Last, but not the least, accurate species detection is increasingly required by the Clinical and Laboratory Standards Institute's guidance for appropriate antimicrobial susceptibility testing (AST). Thus, the environmental surveillance of antimicrobial resistance can also benefit from an up-to-date taxonomy of microorganisms.[49]

Multiple resources are available for researchers to stay up to date with the current taxonomy of microorganisms. The International Code of Nomenclature of Bacteria stipulates that valid novel and revised taxonomy be published in the International Journal of Systematic and Evolutionary Microbiology (IJSEM). This can occur either as a primary publication, or as acceptance on an IJSEM validation list for taxa previously and effectively published in a non-IJSEM journal.[51] In addition, the Journal of Clinical Microbiology (JCM) publishes biennial updates on newly described genera and species, as well as on taxonomic revisions of medically important microorganisms.[52] Upon JCM publication, manufacturers based in the United States usually incorporate the revised names in their databases.[49]

3. PHENOTYPIC CLASSIFICATIONS OF MYCOBACTERIA

Bergey's manual chapter on *Mycobacterium* describes the genus as aerobic to microaerophilic, as slightly curves or straight rods, and as acid-alcohol fast in certain growth stages.[30] The acid-fastness is due to the high mycolic acid content in their cell walls. Colonies are white to cream colored, and some may produce pigments with or without light stimulation.[30] Traditional classification of nontuberculous mycobacteria was based on growth rates, the production of pigments, and colony morphologies.[53] While there were 126 validated species under the *Mycobacterium* genus at the time of Bergey's 2012 manual, the described diversity under this genus has expanded since then. At the time of our literature analysis for this review, there were 189 species with valid published names (Figure 1). As molecular and culturing techniques grow, clinical and environmental studies will likely detect new species.

Based on their growth rates, mycobacteria are often split first into rapid and slow growers, i.e., rapid growing mycobacteria (RGM) or slow-growing mycobacteria (SGM). RGM show visible growth from dilute inocula within 7 days, whereas SGM require more than 7 days to show visible growth.[54] Notably, several groups of NTM are considered drinking water associated pathogens that can cause infections in immunocompromised or otherwise susceptible individuals,[55] such as *M. avium*, *M abscessus*, and the *M. chelonae* complexes, which are slow growers (Figure 1). Regarding the phenotypic classification of NTM, Runyon proposed classification based on growth rates as well as pigment production, whether yellow pigment is produced, and whether light is required to produce pigments.[53] Under Runyon classification, NTM were divided into four groups: group I, slow-growing organisms producing a yellow-orange pigment when exposed to light (photochromogens); group II, slow-growing organisms producing a yellow-orange pigment regardless of whether they are grown in the dark or the light (scotochromogens); group III, slow-growing organisms that never produce pigments, regardless of culture conditions (nonphotochromogens); and group IV, rapid-growing nonpigmented strains.[53] The split into rapid and slow growers and then complexes or groups is often detected in phylogenetic and phylogenomic data, although the split within slow-growing NTM was not supported by phylogenetic analyses.[56]

3.1. Clinical Species Identification Methods Based on Phenotypic Classification

Phenotypic classification remains a useful tool for species identification of NTM in clinical settings. Traditionally, mycobacteria are identified based on their growth rates, the presence or absence of pigmentation, and biochemical tests on isolates recovered from clinical specimens.[57] Newer chromatographic/chemotaxonomic methods have been developed utilizing the lipid-rich composition of the cell wall of mycobacteria.[57] High-performance liquid chromatography (HPLC) of cell wall mycolic acids has gained popularity in reference laboratories.[58] However, the fact that some species can share common HPLC profiles makes this method lack sufficient discriminative power. Modern methods such as matrix-assisted-laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) are equipped with rich databases, widely used for the species identification and subspecies typing of clinically relevant organisms.[16] For example, MALDI-TOF MS has become a reliable tool for NTM identification. It produces unique spectral fingerprints based on extracted proteins.[59] The latest released version (Mycobacteria Library v6.0) of the mycobacterial spectra database represents 178 species with 807 strains. Using Mycobacteria Library v6.0, MALDI-TOF MS achieved correct identification to the species level for 95 out of 101 NTM isolates cultivated from clinical samples.[60]

4. *MYCOBACTERIUM* PHYLOGENY BASED ON 16S AND OTHER MARKER GENE APPROACHES

4.1. Phylogeny Inferred from 16S rRNA and the 16S rRNA Gene

Stahl and Urbance built the first 16S rRNA phylogeny for the *Mycobacterium* genus.[61] They sequenced the 16S rRNA genes of 20 *Mycobacterium* strains as well as outgroups, and they found that 16S rRNA phylogeny reflected the grouping of RGM and SGM. Further, they found that RGM were

polyphyletic, whereas SGM were monophyletic and RGM were more basally placed. They hypothesized that RGM were the ancestors of SGM, which branched out during evolution.[61] Overt pathogens (*M. bovis*, *M. kansasii*, the *M. avium*−*M. intracellulare*−*M. scrofulaceum* complex, and *M. paratuberculosis*) were closely related and clustered within the slow-growing clade. Their analysis also suggested that some strains that used to be considered as mycobacteria (e.g., *Mycobacterium chitae*) were quite distant from the rest of the strains and thus could have been misclassified to the genus.

In the same year, Rogall and colleagues utilized PCR-amplified 16S rRNA genes to build the phylogeny of mycobacteria.[56] Their study included 19 species of mycobacteria as well as outgroups. Rogall and colleagues confirmed a division of RGM and SGM.[56] In addition, an intermediate position was proposed for *M. simiae*.[56] While other SGM had a more extended helix between positions 370 and 450, *M. simiae* had a short helix that was characteristic of a known RGM, consistent with their intermediate positioning. These findings supported RGM being a more ancestral state, with SGM appearing later. Interestingly, within slow growing mycobacteria, phylogenetic relationships inferred from the 16S rRNA gene sequences did not reflect the Runyon classification.[56] In both studies (Stahl and Urbance, and Rogall and colleagues), there were strains that were indistinguishable based on the RNA analysis or DNA sequence analysis. For example, *M. gastri* was indistinguishable from *M. kansaii* in the 16S rRNA gene analysis.[56]

Tortoli refined the phylogeny of NTM by expanding the analysis to 88 strains.[62] An overall division between the RGM and SGM was confirmed. Well-defined clusters were detected around *M. simiae* (SGM), *M. terrae* (SGM), *M. avium* (SGM), *M. fortuitum* (RGM) and *M. chelonae* (RGM). As expected, RGM encompassed species such as the *M. fortuitum* group, *M. chelonae* group, and *M. smegmatis* group; SGM encompassed species such as the *M. Avium* group, *M. simiae* group, and *M. scrofulaceum* group.[62] However, three of the 88 analyzed strains showed unexpected positioning (*M. tusciae*, *M. trivale*, and *M. intermedium*) in terms of their growth rates.

With the expansion of the 16S rRNA gene sequences from NTM, it was clear that the 16S rRNA gene polymorphism within the *Mycobacterium* genus was limited and further identification of species based on the 16S rRNA gene was difficult. Some species have very similar or identical 16S rRNA gene sequences, such as *M. kansasii* and *M. gastri*, *M. senegalense* and *M. farcinogenes*, *M. marinum* and *M. ulcerans*, *M. malmoense* and *M. szulgai*, *M. abscessus* and *M. chelonae*, and members of the *M. tuberculosis* complex.[63,64] Multiple studies showed that phylogenetic trees constructed from the 16S rRNA gene sequences showed poor robustness when examined using a bootstrap analysis.[62,64]

4.2. Phylogeny Inferred from *hsp65*, *rpoB*, and Other Markers

To better resolve the phylogenetic relationships among mycobacteria, researchers have explored other markers, either as independent resources or for use in conjunction with the 16S rRNA gene. In particular, the heat shock protein (*hsp65*) and RNA polymerase gene (*rpoB*) have been extensively investigated. Tortoli et al. showed that the phylogenetic tree constructed from the *rpoB* gene reflected the SGM/RGM separation, whereas the tree constructed from *hsp65* was less clear.[62] However, *hsp65* was useful in the classification of certain *Mycobacterium* species.[62,65−69] Other species pairs that were previously indistinguishable using 16S rRNA, such as *M. chelonae*/*M. abscessus*, and *M. gastri*/*M. kansasii*, can be well differentiated by *hsp65*.[65,66] Nonetheless, *hsp65* was unable to identify some of the *Mycobacterium* species, especially for *M. simiae* and *M. fortuitum*.[65] The *rpoB* gene was reported to be effective for the differentiation of RGM species, as well as some SGM such as *M. gastri* and *M. kansasii*,[65,68,70] but not effective in identifying *M. terrae*.[65] Many other markers have been explored to build the phylogeny of NTM. 16S-23S rRNA internal transcribed spacer (ITS) sequences were used for classification of SGM.[71] 16S-23S rRNA ITS was able to differentiate *M. gastri* and *M. kansasii* but not *M. marinum* and *M. ulcerans*.[71] In addition, phylogeny or species classification based on 23S rRNA, *smpB*, *sodA*, *tuf*, *tmRNA*, *gyrA*, *dnaK*, *secA1*, and *ssrA* have been investigated.[68,69,69,72−75]

Incorporating information from multiple markers has been shown to provide more robust phylogeny about mycobacteria, such as improved robustness in the nodes, in particular, the deeper nodes. Devulder and colleagues used concatenations of the 16S rRNA gene and the *hsp65*, *rpoB*, and *sod* genes to construct the evolutionary relationships among mycobacteria.[64] Their study included 97 strains, including all types of strains known at that time. Concatenation of multiple genes largely improved the tree robustness.[64] In the tree built from the concatenated genes, the percentage of nodes with bootstrap values >50% was 60.49%, a large improvement over the 35.05% in the 16S rRNA gene tree.[64] The SGM/RGM separation was robustly detected except for *M. doricum*, a slow grower that was placed among rapid growers.[64] Using the concatenated genes to compute the distance between species, most species were significantly different, with the exception of three species pairs: *M. farcinogenes*−*M. senegalense*, *M. marinum*−*M. ulcerans*, and *M. murale*−*M. tokaiense*.[64] Mignard and Flandrois constructed evolutionary relationships among two sets of *Mycobacterium* strains from seven individual genes (*hsp65*, *rpoB*, 16S rRNA, *smpB*, *sodA*, tmRNA, and *tuf*), using both a concatenation method and a superdistance-matrix method.[69,76] The concatenation method resulted in a robust phylogeny, where 74% of the nodes were well supported (aLRT statistic > 75%). The superdistance-matrix method, as a kind of supertree method that combines multiple trees to infer phylogeny, was more computationally efficient, yet appeared to be more sensitive to addition of new data.[69] Both methods showed SGM/RGM separation. Using an evolution model, the study computed evolution rates from various genes, and showed that 16S rRNA exhibited the slowest evolution, and the *smpB* gene exhibited the most rapid evolution.[69] The *M. terrae* complex was placed furthest away from the rest of the slow growers.[69]

4.3. Clinical Species Identification Methods Based on Marker Genes

Phylogenetic classification based on marker genes has facilitated the development of various species identification tools used in clinical microbiology. For instance, PCR restriction-enzyme analysis (PRA) based on the *rpoB* gene and *hsp65* gene can rapidly identify mycobacterial species by comparing the patterns of fragments produced by selected restriction enzymes.[77] Yet misidentifications can arise due to shared and ambiguous patterns between species.[63] Since their development more than 30 years ago, commercial DNA probes based on various marker genes have been applied in many

https://doi.org/10.1021/acsenvironau.3c00074
ACS Environ. Au 2024, 4, 127−141

clinical mycobacterial laboratories.[78,79] For example, INNO-LiPA MYCOBACTERIA, which targets the 16S-23S rRNA spacer region, was able to identify 16 *Mycobacterium* species with an overall sensitivity of 100% and a specificity of 94.4%.[80] Two versions of the GenoType Mycobacterium system targeting the 23S rRNA gene region were reported to have accuracies of 92.6% and 89.9% in identifying 23 and 14 species, respectively.[81] The performance of commercial DNA probe systems, however, can depend on the choices of mycobacterial species in the validation experiments. With many new species revealed by taxonomic studies in the past decade, it has been shown that commercial DNA probes have limited ability to identify less frequently encountered species.[82]

5. *MYCOBACTERIUM* PHYLOGENY BASED ON GENOME SEQUENCES

As genome sequencing technology becomes more affordable and accessible, more genomes of *Mycobacterium* species have become available in public databases. These genome data allow for the construction of more robust and comprehensive phylogenetic trees with greater resolution. A number of studies have conducted phylogenetic analyses based on subsets of core proteins/genes from *Mycobacterium* genomes.[11,83−86]

5.1. Update on the Phylogeny of NTM Based on Genomic Data

Fedrizzi and colleagues performed the first comprehensive characterization of the genome diversity of NTM.[85] Their analysis utilized a set of genomes representing 99 isolates. This analysis used a concatenated core genome approach, where 243 fully conserved genes within the genus were analyzed using a maximum-likelihood approach. The phylogenetic tree built from concatenated core genomes showed high concordance with the full-length 16S rRNA gene tree. A separation between rapid and slow growers was apparent, and the *M. terrae* complex occupied an intermediate position.[56,85] As in the previous phylogeny constructed from 16S rRNA gene sequences, *M. doricum*, a slow grower, was placed among the rapid growers.[64,85] The evolutionary relationships indicated that rapid growers appear more ancestral, and there was a common ancestor for the entire genus, most closely related to the present *M. abscessus* complex.[85] The authors noted that *M. triviale* branched basally to the *M. terrae* complex, supporting the hypothesis that this rapid grower might represent the link in the evolutionary path toward the slow growing species.[85]

As for groups within the *Mycobacterium* genus, the analysis revealed nine potential complexes/groups.[85] Among rapid growers, monophyletic groups associated with *M. abscessus*, *M. fortuitum*, and *M. smegmatis* complexes formed. In terms of slow growers, monophyletic groups formed around *M. celatum*, *M. kansasii*, *M. marinum*, *M. tuberculosis*, and *M. avium*. The *M. simiae* species appeared paraphyletic, which is consistent with observations of their diversified phenotypes. This extensive phylogenetic analysis suggested new group/complex assignments for several species such as *M. kubicae*, *M. shimoidei*, *M. bohemicum*, and *M. nebraskense*.

Tortoli and colleagues analyzed the ANI of 144 mycobacterial strains, covering type strains of described species, among others.[87] The ANI showed similar patterns to the phylogeny constructed by Fedrizzi and colleagues.[85] Overall, there was a separation of SGM/RGM, and the *M. terrae* group was intermediately placed.[87] The authors proposed eight clusters from the ANI analysis, namely the *M. abscessus-chelonae* complex, *M. fortuitum-smegmatis* group, *M. terrae* complex, *M. celatum* group, *M. xenopi* group, a "pathogens group" (around the *M. tuberculosis* complex and *M. leprae* species, but also including NTM), the *M. simiae* complex, and the *M. avium* complex.[85] It is worth noting that some species pairs showed an ANI above 95%, thus the authors recommended the fusion of some species.[85]

Gupta and colleagues constructed phylogenetic trees for 150 *Mycobacterium* species.[11] They constructed three phylogenetic trees, respectively, based on 1941 core proteins for the genus *Mycobacterium*, 136 core proteins for the phylum *Actinobacteria*, and 8 conserved housekeeping proteins. In all three trees, members were consistently grouped into five clades.[11] Based on the clustering of "*Tuberculosis-Simiae*" (emended *Mycobacterium*), "*Fortuitum-Vaccae*" (*Mycolicibacterium*), "*Terrae*" (*Mycolicibacter*), "*Triviale*" (*Mycolicibacillus*), and "*Abscessus-Chelonae*" (*Mycobacteroides*), the authors proposed a five-genera split in the genus taxonomy.[11] Among these clusters, "*Abscessus-Chelonae*" and "*Fortuitum-Vaccae*" were rapid growers and the other three were slow growers. "*Tuberculosis-Simiae*" contained most of the clinically important *Mycobacterium* species, including *M. avium*, *M. tuberculosis*, and *M. leprae*, and "*Fortuitum-Vaccae*" was primarily comprised of environmental species.[11]

5.2. Debates on the Mycobacterial Genus Taxonomy

Based on phylogenetic and comparative genomic analyses, Gupta and colleagues proposed that the genus *Mycobacterium* should be divided into five genera, namely *Mycolibacterium*, *Mycolicibacter*, *Mycolibacillus*, *Mycobacteroides*, and the emended *Mycobacterium*.[11] They calculated pairwise amino acid identities (AAI) and found that members within the same clade showed AAI higher than those of members from different clades. In addition, they identified specific molecular markers, conserved signature indels (CSIs) and conserved signature proteins (CSPs), for each clade. Those unique molecular markers most likely came from the initial introduction into the common ancestors and then were vertically inherited,[88] which provided further evidence for the monophyletic nature of the observed clades.

The five-genera split was controversial. A paper authored by twenty-three researchers from multiple countries expressed concerns about the potential disruptions that could be caused by splitting and renaming *Mycobacterium*, because it is a medically important genus.[12] The authors contended that "substantial cost and effort will be needed to update procedures and laboratory information management systems in hospitals and laboratories which support physicians and care-givers in their daily work".[12] This article also clarified that, based on the International Code of Nomenclature of Prokaryotes (ICNP), the previous names were still valid, and anyone is free to use them.[12]

Meehan and colleagues analyzed the within- and across-clade genome-relatedness for the original as well as the Gupta-proposed mycobacterial genus, and then compared these results to widely accepted genus boundaries.[14] They found that the intragenus and intergenus scores calculated from the original *Mycobacterium* genus were consistent with the widely reported genus cutoffs, i.e., 94.5% for the 16S rRNA gene similarity, 65% for Amino Acid Identity (AAI), and 50% for Percentage of Conserved Proteins (POCP), whereas those computed from the Gupta genera were less clear.[14] In fact, the intergenus comparisons based on the above metrics did not

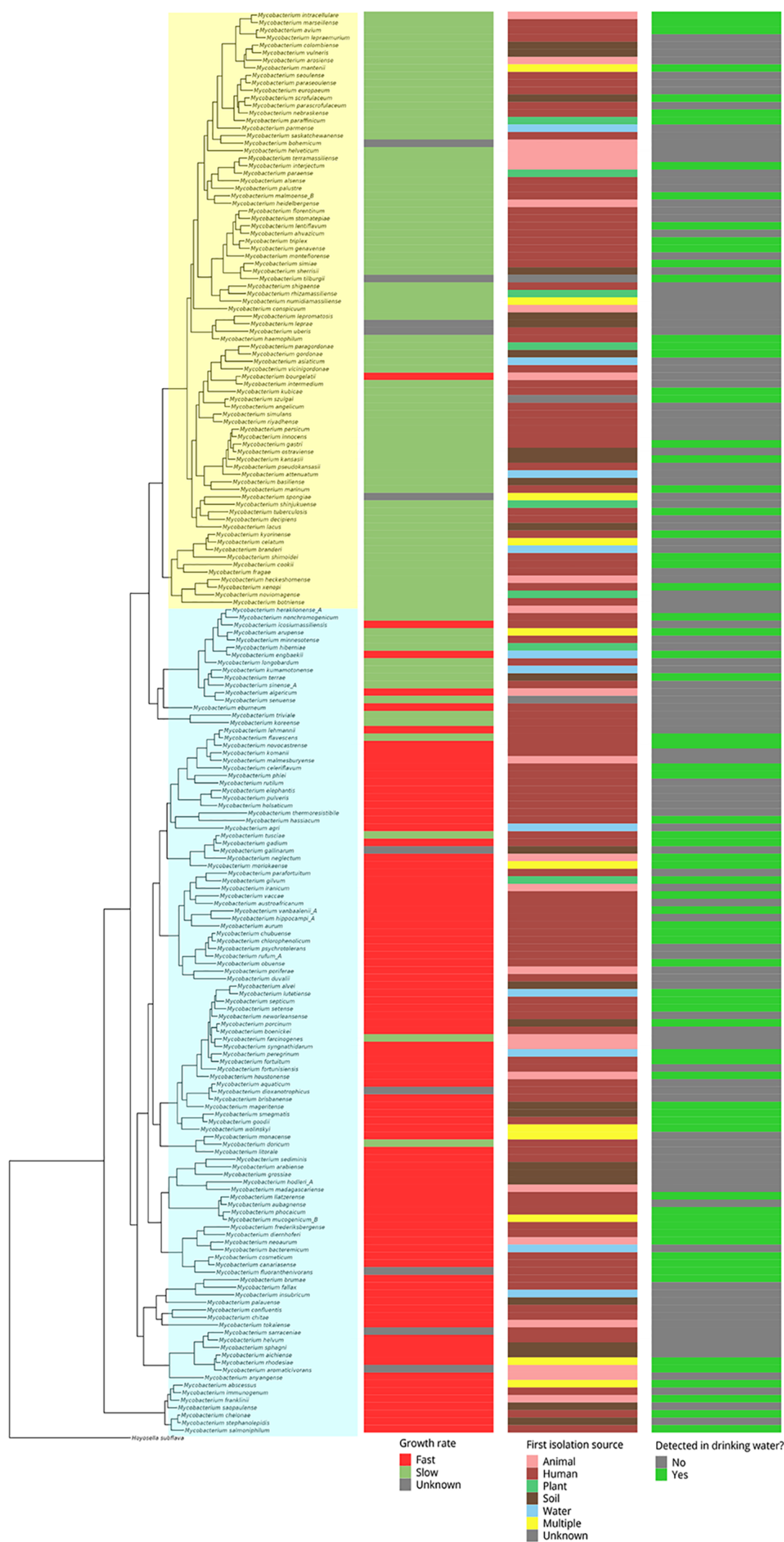


Figure 2. Phylogenetic tree for mycobacterial species, annotated with habitat and growth rate information. Shown are 189 mycobacterial species retrieved from the bacterial reference tree of the Genome Taxonomy Database (r207). The lineages affected by the five-genera split are in blue, and those not affected are in yellow. Each species is annotated with growth rate groups (fast growing, slow growing, or unknown), first isolation source, and drinking water detection. Table S1 lists the studies supporting the annotations.

https://doi.org/10.1021/acsenvironau.3c00074
ACS Environ. Au 2024, 4, 127−141

separate any of the genera proposed by Gupta and colleagues.[14] When the mean alignment fraction (AF) and ANI were examined for the original mycobacterial genus and compared to previously reported genus demarcation boundaries (AF 0.33 and ANI 73.98%),[89] the original *Mycobacterium* fell within the 95% confidence interval of the ANI boundary but showed a considerably low AF.[14] Thus, the authors proposed that the original genus label should be reapplied to all of the species within this group.

The taxonomic labeling of potential pathogens can have wide-ranging implications in species identification and treatment. The ICNP states in its Rule 56a that *nomen periculosum*, 'a name whose application is likely to lead to accidents endangering health or life or both or of serious economic consequences", should be avoided. Based on phylogenomic analyses, and the potential disruptions a five-genus naming could produce in medical practice, Meehan and colleagues proposed that five-genera names be rejected.[14]

5.2.1. How Does the Mycobacterial Genus Taxonomy Debate Affect Environmental Science and Engineering Researchers? The five-genera split of the mycobacteria, which was later reconsituted,[14] can cause confusions in the environmental monitoring of NTM. For instance, the United States Centers for Disease Control listed 31 mycobacterial species as "opportunistic pathogens of premise plumbing", but only 19 would have been considered as falling in the "*Mycobacterium*" genus under the five-genera split taxonomy, which could easily create confusion when reporting NTM in the environment. In particular, *Mycobacterium abscessus*, *Mycobacterium chelonae*, and *Mycobacterium fortuitum* were renamed as "*Mycobacteroides abscessus*", "*Mycobacteroides chelonae*," and "*Mycolicibacterium fortuitum*", respectively. Those three species are often associated with infections, and all have been frequently detected in drinking water (Figure 2). After the reconstitution of the five-genera into one genus, the confusion was cleared up. Moving forward, we believe the following three aspects deserve particular attention from environmental science and engineering researchers:

(1) Since the mycobacterial genus nomenclature has been reconstituted, using the updated names (shown as "conventional" in Table 1) will be critical to fostering cross-disciplinary communication around NTM.

(2) Because the reconstitution took place relatively recently, reports using the legacy five-genera split taxonomy are still frequently encountered in the literature. Thus, systematic review and meta-analysis investigating NTM relevant to environmental sources, such as those steps in the development of dose−response models, will need to include the legacy species names in their searches (shown as "Nomenclature under five-genera split" in Table 1 and blue colored lineages in Figure 2). Literature searches based only on "*Mycobacterium*" will lead to results that underrepresent the NTM in the environment.

(3) In metagenomic species profiling, utilizing up-to-date databases can help prevent propagation of perilous names in the scientific literature. The current authoritative source of genome-based taxonomy, the GTDB reference (R207 and later), reflects the genus reconstitution. Results generated from earlier versions of the GTDB could still contain five-genera taxonomy and should be manually edited.

Table 1. NTM in Premise Plumbing Included in a Watch List from the Water Management Program of the CDC,[129] Using Both Conventional Nomenclatures and Names under the Five-Genera Split[a]

Conventional	Nomenclature under five-genera split
Mycobacterium abscessus	*Mycobacteroides abscessus*
Mycobacterium bolettii	*Mycobacteroides bolettii*
Mycobacterium massiliense	*Mycobacteroides massiliense*
Mycobacterium chelonae	*Mycobacteroides chelonae*
Mycobacterium mucogenicum	*Mycolicibacterium mucogenicum*
Mycobacterium phocaicum	*Mycolicibacterium phocaicum*
Mycobacterium fortuitum	*Mycolicibacterium fortuitum*
Mycobacterium cosmeticum	*Mycolicibacterium cosmeticum*
Mycobacterium mageritense	*Mycolicibacterium mageritense*
Mycobacterium porcinum	*Mycolicibacterium porcinum*
Mycobacterium septicum	*Mycolicibacterium septicum*
Mycobacterium immunogenum	*Mycobacteroides immunogenum*
Mycobacterium goodii	*Mycolicibacterium goodii*
Mycobacterium wolinskyi	*Mycolicibacterium wolinskyi*
Mycobacterium Aurum	*Mycolicibacterium Aurum*
Mycobacterium simiae	*Mycobacterium simiae*
Mycobacterium avium	*Mycobacterium avium*
Mycobacterium intracellulare	*Mycobacterium intracellulare*
Mycobacterium chimaera	*Mycobacterium chimaera*
Mycobacterium avium subsp. hominissuis	*Mycobacterium avium subsp. hominissuis*
Mycobacterium colombiense	*Mycobacterium colombiense*
Mycobacterium scrofulaceum	*Mycobacterium scrofulaceum*
Mycobacterium parascrofulaceum	*Mycobacterium parascrofulaceum*
Mycobacterium xenopi	*Mycobacterium xenopi*
Mycobacterium arupense	*Mycolicibacter arupensis*
Mycobacterium kansasii	*Mycobacterium kansasii*
Mycobacterium hemophilum	*Mycobacterium hemophilum*
Mycobacterium nonchromogenicum	*Mycolicibacter nonchromogenicus*
Mycobacterium triviale	*Mycolicibacillus trivializ*
Mycobacterium terrae	*Mycolicibacter terrae*
Mycobacterium gordonae	*Mycobacterium gordonae*

[a]A proposal has been published on IJSEM to reject the five-genera split (*nomina rejicienda*) as perilous names (*nomen periculosum*).[14]

5.3. Clinical Species Identification Methods Based on Whole Genome Sequencing

Advances in whole genome sequencing (WGS), which generates the complete genomes of bacterial isolates, have the potential to improve the species-level characterization of mycobacteria in clinical settings. Quan and colleagues compared the performance of WGS and standard laboratory diagnostic workflows (the GenoType testing kits) in NTM species identification.[90] They performed WGS on 1902 isolates, using the Illumina Miseq sequencing platform. The species were identified using Mykrobe, a bioinformatics tool tailored for mycobacteria.[91] The very high agreement (96%) between WGS and GenoType tests in NTM species identification supported the use of WGS as a routine mycobacterial species identification method. For complex reasons, four isolates had persistently discordant results between WGS and GenoType tests: *M. fortuitum*, *M. peregrinum*, *M. intracellulare*, and *M. chelonae*. Long read sequencing using PacBio or Oxford Nanopore Technologies (ONT) is a useful tool to improve genome assemblies. Matsumoto and colleagues developed an NTM identification method, mlstverse, incorporating hybrid assembly using

https://doi.org/10.1021/acsenvironau.3c00074
ACS Environ. Au 2024, 4, 127−141

sequence reads from Illumina and ONT. They demonstrated that hybrid assembly not only helped achieve very high accuracy in species classification, but was also much more successful in identifying subspecies than traditional methods (mass spectrometry and multilocus sequencing typing).[92]

Besides species identification, WGS can be a powerful tool in many other clinical applications. For example, Mykrobe can also predict drug resistance for mycobacteria using WGS data.[90,91] In addition, WGS has been used in multiple countries to characterize clinical isolates of foodborne pathogens and to support epidemiological investigations.[93] Ongoing research is applying WGS to track the sources of NTM infections.[94] Genome sequences of isolates from patients and their household environments can be used to prove or rule out a source of infection. A recent study by Lande and colleagues performed WGS on *M. avium* isolates from household plumbing biofilms and those from patient respiratory specimens.[95] Single nucleotide variant (SNV)-based phylogeny analyses showed distances of 4−51 SNVs between the isolates from respiratory specimens and the isolates of the associated household plumbing biofilms. Intriguingly, five biofilm-patient pairs were separated by less than 15 SNVs, suggesting close relatedness and a possible plumbing-related source.[95] However, the environmental source of the NTM remains undetermined, pending the establishment of definite thresholds of the SNV distance. This task remains challenging because host and environmental factors can affect mutation rates; additionally, there are usually time lags between acquiring a strain and making a disease diagnosis; thus, it can be difficult to characterize the time for mutations to accumulate in the environmental or the clinical isolates. Future studies to understand the genetic diversity of NTM in the environment and their evolution under environmentally relevant conditions will likely complement studies of host conditions and contribute to a more holistic understanding.

6. REFLECTIONS AND RECOMMENDATIONS

6.1. Species-Level Characterization of NTM in Drinking Water: A Critical Research Gap

NTM infections are rising globally, and inhalation of aerosolized water is increasingly recognized as an important transmission route. Nevertheless, NTM are not yet routinely monitored in municipal water supplies or premise plumbing, and knowledge of their transmission and potential selections in engineered water systems remains limited. The mechanisms driving species-level variations of NTM in drinking water distribution systems and premise plumbing are also poorly understood. Several notable studies have provided interesting examples of species- and clade-level variations. Haig and colleagues found that greater water age was associated with a less diverse community of NTM and a greater relative abundance of *M. avium*.[96] Gebert and colleagues compared NTM species in homes receiving municipal water and those receiving well water. Higher relative abundances of potentially pathogenic NTM, such as species from the *M. mucogenicum/M. phocaicum* clade were found in homes receiving municipal water, whereas groups rarely considered pathogenic, including *M. nebraskense* and *M. gordonae* clades, were more abundant in homes receiving well water.[97] Warmer locations with higher shower water chlorine concentrations correlated with higher relative abundances of the *M. mucogenicum/M. phocaicum* clade.[97] Ghosh and colleagues reported that the operating conditions of simulated reclaimed water distribution systems can select for different NTM species.[98] They found that chlorinated water was dominated by the rapidly growing *Fortuitum-Vaccae* clade, while chloraminated water contained a more diverse community, including the *Abscessus-Chelonae*, *Terrae*, and *Fortuitum-Vaccae* clades. Field and experimental research investigating the effects of geography, water sources (surface or groundwater), as well as plumbing conditions (such as water age, pipe material, temperature, and residual disinfectant levels) on NTM species diversity and composition will be crucial to elucidating the environmental transmission mechanisms of NTM.

Embracing the evolving and up-to-date taxonomy of NTM in environmental characterization research will facilitate the communication of these results across disciplinary boundaries. The latest developments in metagenomic taxonomic classification tools that incorporate the genome-based taxonomy framework will be indispensable in this endeavor. Alignment and k-mer based tools, such as mOTUs2,[99] MetaPhlAn,[100] Kraken2,[101] and Bracken[102] can provide taxonomic profiling from metagenomic reads. Those approaches, while tremendously useful, can be limited by the availability of curated taxonomic trees.[103] Deep learning approaches can circumvent this bottleneck. For example, DeepMicrobes, a deep learning-based computational framework for taxonomic classification, was shown to outperform other state-of-the-art tools in taxonomic classification of human gut metagenomes.[103] Nevertheless, deep learning tools will need to be trained when new species are added. Their performance in taxonomic classifications for environmental metagenomes, e.g., drinking water metagenomes, where new species are anticipated, requires validation.[104]

6.2. What Does the Lack of Complex-Level Association between Mycobacteria and Environments Mean for Water Treatment and Risk Assessments?

Genomic characterizations have revealed five complexes in the *Mycobacterium* genus: *Tuberculosis-Simiae*", "*Fortuitum-Vaccae*", "*Terrae*", "*Triviale*", and "*Abscessus-Chelonae*".[11,14,85] Here, we specifically ask whether there is any association between these complexes and their environmental habitats, in particular, the drinking water environment. If such associations are present, they will help prioritize targets for water treatment and distribution system designs, as well as risk assessments based on their environmental prevalence. If such relationships are not present, then there are strong justifications for (1) systematically examining the selection of water treatment and plumbing conditions on all complexes and (2) developing the QMRA frameworks for diverse mycobacterial complexes.

To pursue these questions, we annotated the mycobacterial species represented by the Genome Taxonomy Database (GTDB) entry 207 with their environmental sources. The origins of three species (1.6%) have not been reported. The names of the species and studies about their taxonomic classification are included in Table S1. As expected, almost all of the slow-growing and fast-growing species are classified into two well-defined clusters. The slow-growing group contains the human pathogens *M. tuberculosis* complex and opportunistic pathogens such as the *M. avium* complex. Most slow-growing species were first isolated from a human or an animal host. However, there was no clear clustering pattern between the phylogenetic relationships and whether a clade is detected in

https://doi.org/10.1021/acsenvironau.3c00074
ACS Environ. Au 2024, 4, 127−141

an environmental source, such as the drinking water environment (Figure 2).

Ecological forces shaping microbial communities can be conceptualized as dispersal, selection, and stochasticity.[105] In the phylogenetic tree, none of the complexes show a strong association with drinking water. We speculate that there is not a very strong selection for NTM at the species level in the aquatic environment and that their distribution is mainly driven by dispersal limitation. This lack of association has important implications in extending treatment and distribution system design and QMRA research to diverse complexes. In particular, the quantitative risk assessment framework for NTM is currently available only for the *M. avium* complex.[50] However, from Figure 2, none of the NTM clades can be precluded from drinking water. Therefore, extending the NTM risk assessment framework to other complexes and species is crucially important to understanding and managing the risks of NTM infections through water systems.

6.3. Methodological Considerations: Mycobacterial Taxonomic Profiling Based on Short and Long Reads of Marker Genes

While various techniques can perform species or subspecies identification for NTM isolates or molecular diagnostics in clinical samples, high throughput detection of NTM in the environment remains an ongoing topic of research.[17] Given the pervasive use of 16S rRNA gene amplicon sequences for microbiome taxonomic composition studies, we asked how well the amplicon sequencing approach could reveal the taxonomic compositions related to NTM. We retrieved the full-length 16S rRNA gene sequences from the SILVA v138 alignment, and trimmed the sequences with five sets of 16S universal primers commonly used for sequence extraction and classifier training, including those amplifying the V3 region (336*F*/518R), V3−V4 region (336*F*/806R), V3−V5 region (336*F*/909R), V4 region (515*F*/806R), and theV4-V5 region (515*F*/909R).[106,107] We then examined the accuracy of the taxonomic classification at the genus and species levels, respectively (Table 2).

As expected, the full length 16S rRNA gene had the highest classification power. At the genus level, 100% of the sequences were correctly classified to the genus *Mycobacterium*, whereas at the species level, only 63.5% of the sequences were correctly classified. Among the partial 16S rRNA sequences, regions V3−V4, V3−V5, and V4−V5 were able to correctly classify 100% of the sequences to *Mycobacterium*. At the species level, the classification accuracy was poor. In even the best performance (the V3 region), only 22.8% of the sequences were classified to the correct species. We also observed unresolved clusters in the phylogeny trees of all partial length 16S rRNA sequences. Thus, results from NTM detection based on 16S rRNA gene analyses should be interpreted with caution.

Table 2. Taxonomic Classification Based on 16S rRNA Gene Partial Sequences

Region	Primer sets	Percentage of sequences classified to a genus label (%)	Percentage of sequences classified to a *correct* genus label (%)	Percentage of sequences classified to a species label (%)	Percentage of sequences classified to a *correct* species label (%)
V3	336*F*/518R	96.4	93.9	29.4	22.8
V3−V4	336*F*/806R	100	100	20.8	16.2
V3−V5	336*F*/909R	100	100	15.2	11.7
V4	515*F*/806R	100	93.4	6.6	4.1
V4−V5	515*F*/909R	100	100	5.1	3.6
V1−V9	Full length	100	100	73.1	63.5

Developments in long-read sequencing can shed light on new strategies for the species level characterization of NTM in the environment. In terms of long read sequencing, Haig and colleagues developed a high throughput assay utilizing the *rpoB* sequence to differentiate environmental NTM species.[108] They used the assay to amplify a region of the *rpoB* gene that is over 900 bp long and then generated high-quality long-read sequences using the PacBio technology. When applied to drinking water samples, the new assay revealed nine known species, including *M. abscessus*, *M. avium*, *M. chelonae*, and *M. mucogenicum*, and it even provided subspecies resolutions in some cases. Despite the potential of long-read sequencing, the choice of the marker genes is still highly important. For example, our analyses showed that classification using the full-length 16S rRNA gene was still less than satisfactory for *Mycobacterium* (Table 2, 63.5% accuracy). Thus, even long-read sequencing of the 16S rRNA genes might not be a good strategy for species-level characterization of NTM.

6.4. Utilizing NTM Genome Data to Characterize Functional Potentials in a Phylogenetic Framework

The increased availability of NTM genome data and the updated genome-based phylogeny opened new opportunities to incorporate analyses of functions and functional genes into the phylogenetic framework. The Mykrobe platform specifically incorporated genomes from diverse mycobacteria and predicted antimicrobial susceptibility from genome data.[91] Generic tools for functional gene annotations can also be utilized to develop custom pipelines. Functional gene annotation approaches used in recent studies have typically involved two steps: (1) annotating genes using gene prediction software, such as Prokka[109] and Prodigal,[110] and (2) identifying functional genes by comparing them to databases, such as EggNOG,[111] KEGG,[112] COG,[113] RefSeq,[114] Pfam,[115] as well as customized databases, using tools such as hidden Markov models (HMM),[116] BLAST,[117] and Diamond.[118] Still, further study is required to construct specific databases tailored for functional gene annotations in NTM.

Detailed insights have been drawn from characterizing the genomes of NTM. For instance, the functional gene differences between rapid and slow growers were found to occur mainly in cell wall lipids synthesis genes and transporter genes.[85,119] Genes for biosynthesizing dimycocerosate esters (DIM) were found only in some slower growing pathogens, such as the *M. tuberculosis* complex, *M. leprae*, *M. kansasii*, *M. marinum*, *M. ulcerans*, and *M. hemophilum*.[120] Transporter genes, including the livFGMH and ABC operons, were enriched in rapid growers, while these genes were lost in slow growing species.[119] In terms of shared and distinct genes of NTM in various environments, between 51 and 731 genes out of 4000−7000 open reading frames were present in specific clades, with the functions of the majority of these genes yet to be characterized.[85] However, the presence of these distinctive genes in mycobacterial species sheds light on gene targets for novel rapid detection in the environment.

https://doi.org/10.1021/acsenvironau.3c00074
ACS Environ. Au 2024, 4, 127−141

In another recent update enabled by genomic characterization, it has been found that the necessity of certain virulence genes in pathogenic and nonpathogenic *Mycobacterium* may differ, posing the possibility of differentiating the two groups. For example, the protein ESX-3 is reported to be conserved in all *Mycobacterium* species,[85,121] and the absence of ESX-3 can disrupt the growth of *M. bovis* BCG by interfering with iron acquisition, but the absence was tolerable by nonpathogenic *M. smegmatis*.[121]

Despite the current understanding of functions in NTM, drinking water remains an under-characterized environment, with few reference genomes.[122] Because the drinking water environment is unique in its oligotrophism, disinfectants, and highly variable temperature, we speculate that mycobacterial isolates from the drinking water environment will reveal new insights. Future research expanding NTM reference genomes from the drinking water environment will be a crucial next step.

7. METHODS

7.1. Visualization of the Environmental Habitats for NTM

The whole genome tree was retrieved from GTDB r207,[123] and trimmed using ape (5.6−2)[124] and ggtree (3.2.1)[125] in R version 4.1.2.[126] GTDB r207 was built from the concatenation of 120 ubiquitous single-copy bacterial marker genes. The tree was first trimmed to keep tips belonging to *Mycobacterium*. This process retained 189 *Mycobacterium* species with valid published names. The tree was further trimmed to retain only one tip for one species.

Then, we annotated the *Mycobacterium* genome tree with growth rate and habitat information on each species (Figure 2). A literature review was conducted to investigate the first isolation source of each species. Among the 189 species investigated, 106 (56.1%) were isolated from human specimens, 25 (13.2%) were isolated from soil/sediment samples, 25 (13.2%) were isolated from animal samples, 11 (5.8%) were isolated from water samples, and eight (4.2%) were isolated from plant samples. In addition, 11 species had strains of multiple sources in their first taxonomic characterization, and they are denoted as "multiple sources" in Figure 2.

We further examined the presence of each *Mycobacterium* species in drinking water systems. Keyword searches using each species name and "drinking water" were performed in PubMed and Google Scholar. A manual curation step was performed after the search to check the relevance of the retrieved papers. In total, 74 out of the 189 *Mycobacterium* species (39.2%) have been reported in drinking water ecosystems, using culture-dependent or culture-independent methods (Table S1).

7.2. Retrieval and Analysis of 16S rRNA Gene Sequences Related to the Mycobacteria

We retrieved and filtered out 197 nonredundant full length 16S rRNA sequences of 197 *Mycobacterium* species from Silva v138.[127] The sequences were imported into Qiime2 v2022.2 for processing.[128] Phylogenetic trees of each biomarker were generated by using maximum likelihood with the fasttree plugin in Qiime2. *Streptococcus aglactiae* and *Clostridioides difficile* were used as the outgroup.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsenvironau.3c00074.

List of *Mycobacterium* species with validly published names from the Genome Taxonomy Database. Growth rate, first isolation source, and drinking water detection of each species was retrieved after literature review (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Fangqiong Ling − *Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;* orcid.org/0000-0003-1546-5647; Email: fangqiong@wustl.edu

Authors

Lin Zhang − *Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;* orcid.org/0000-0001-9115-6585

Tzu-Yu Lin − *Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;* orcid.org/0000-0003-0548-6495

Wen-Tso Liu − *Department of Civil and Environmental Engineering, University of Illinois, Urbana−Champaign, Urbana, Illinois 61801, United States;* orcid.org/0000-0002-8700-9803

Complete contact information is available at: https://pubs.acs.org/10.1021/acsenvironau.3c00074

Author Contributions

[§]L.Z. and T.-Y.L. contributed equally to this paper.

Funding

F.L., L.Z., and T.-Y.L.'s participation is partially supported by the National Science Foundation Division of Chemical, Bioenergy, Environmental and Transport Systems (CBET) 2047470.

Notes

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the sponsoring organizations, agencies, or the United States government.

The authors declare no competing financial interest.

■ REFERENCES

(1) Hunt, J. R. F.; Martinelli, R.; Adams, V. C.; Rook, G. a. W.; Brunet, L. R. Intragastric Administration of Mycobacterium Vaccae Inhibits Severe Pulmonary Allergic Inflammation in a Mouse Model. *Clinical & Experimental Allergy* **2005**, *35* (5), 685−690.

(2) Matthews, D. M.; Jenks, S. M. Ingestion of Mycobacterium Vaccae Decreases Anxiety-Related Behavior and Improves Learning in Mice. *Behavioural Processes* **2013**, *96*, 27−35.

(3) Brode, S. K.; Daley, C. L.; Marras, T. K. The Epidemiologic Relationship between Tuberculosis and Non-Tuberculous Mycobacterial Disease: A Systematic Review. *Int. J. Tuberc Lung Dis* **2014**, *18* (11), 1370−1377.

(4) 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.; Iseman, M.; Olivier, K.; Ruoss, S.; von Reyn, C. F.; Wallace, R. J.; Winthrop, K. An Official ATS/IDSA Statement: Diagnosis, Treatment, and Prevention of Nontuberculous Mycobacterial Diseases. *Am. J. Respir Crit Care Med.* **2007**, *175* (4), 367−416.

(5) Hamed, K. A.; Tillotson, G. A Narrative Review of Nontuberculous Mycobacterial Pulmonary Disease: Microbiology, Epidemiology, Diagnosis, and Management Challenges. *Expert Review of Respiratory Medicine* **2023**, *17*, 973.

(6) Tzou, C. L.; Dirac, M. A.; Becker, A. L.; Beck, N. K.; Weigel, K. M.; Meschke, J. S.; Cangelosi, G. A. Association between

Mycobacterium Avium Complex Pulmonary Disease and Mycobacteria in Home Water and Soil. *Annals ATS* **2020**, *17* (1), 57−62.

(7) Thomson, R.; Tolson, C.; Carter, R.; Coulter, C.; Huygens, F.; Hargreaves, M. Isolation of Nontuberculous Mycobacteria (NTM) from Household Water and Shower Aerosols in Patients with Pulmonary Disease Caused by NTM. *Journal of Clinical Microbiology* **2013**, *51* (9), 3006−3011.

(8) Falkinham, J. O. Factors Influencing the Chlorine Susceptibility of Mycobacterium Avium, Mycobacterium Intracellulare, and Mycobacterium Scrofulaceum. *Appl. Environ. Microbiol.* **2003**, *69* (9), 5685−5689.

(9) Parker, B. C.; Ford, M. A.; Gruft, H.; Falkinham, J. O. Epidemiology of Infection by Nontuberculous Mycobacteria. IV. Preferential Aerosolization of Mycobacterium Intracellulare from Natural Waters. *Am. Rev. Respir. Dis.* **1983**, *128* (4), 652−656.

(10) Chu, H.; Zhao, L.; Xiao, H.; Zhang, Z.; Zhang, J.; Gui, T.; Gong, S.; Xu, L.; Sun, X. Systematic Review/Meta-Analysis Prevalence of Nontuberculous Mycobacteria in Patients with Bronchiectasis: A Meta-Analysis. *Arch Med. Sci.* **2014**, *4*, 661−668.

(11) Gupta, R. S.; Lo, B.; Son, J. Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus Mycobacterium into an Emended Genus Mycobacterium and Four Novel Genera. *Front Microbiol* **2018**, *9*, 67.

(12) Tortoli, E.; Brown-Elliott, B. A.; Chalmers, J. D.; Cirillo, D. M.; Daley, C. L.; Emler, S.; Floto, R. A.; Garcia, M. J.; Hoefsloot, W.; Koh, W.-J.; Lange, C.; Loebinger, M.; Maurer, F. P.; Morimoto, K.; Niemann, S.; Richter, E.; Turenne, C. Y.; Vasireddy, R.; Vasireddy, S.; Wagner, D.; Wallace, R. J.; Wengenack, N.; van Ingen, J. Same Meat, Different Gravy: Ignore the New Names of Mycobacteria. *Eur. Respir. J.* **2019**, *54* (1), 1900795.

(13) Oren, A.; Trujillo, M. E. On the Valid Publication of Names of Mycobacteria. *Eur. Respir. J.* **2019**, *54* (4), 1901483.

(14) Meehan, C. J.; Barco, R. A.; Loh, Y.-H. E.; Cogneau, S.; Rigouts, L. Reconstituting the Genus Mycobacterium. *Int. J. Syst. Evol Microbiol* **2021**, *71* (9), No. 004922.

(15) Hugenholtz, P.; Chuvochina, M.; Oren, A.; Parks, D. H.; Soo, R. M. Prokaryotic Taxonomy and Nomenclature in the Age of Big Sequence Data. *ISME J.* **2021**, *15* (7), 1879−1892.

(16) Kämpfer, P.; Glaeser, S. P. Prokaryotic Taxonomy in the Sequencing Era − the Polyphasic Approach Revisited. *Environmental Microbiology* **2012**, *14* (2), 291−317.

(17) Dowdell, K.; Haig, S.-J.; Caverly, L. J.; Shen, Y.; LiPuma, J. J.; Raskin, L. Nontuberculous Mycobacteria in Drinking Water Systems - the Challenges of Characterization and Risk Mitigation. *Curr. Opin Biotechnol* **2019**, *57*, 127−136.

(18) Claeys, T. A.; Robinson, R. T. The Many Lives of Nontuberculous Mycobacteria. *J. Bacteriol.* **2018**, *200* (11), e00739−17.

(19) Sneath, P. H. A. Analysis and Interpretation of Sequence Data for Bacterial Systematics: The View of a Numerical Taxonomist. *Systematic and Applied Microbiology* **1989**, *12* (1), 15−31.

(20) Mayr, E. Biological Classification: Toward a Synthesis of Opposing Methodologies. *Science* **1981**, *214* (4520), 510−516.

(21) Kämpfer, P.; Glaeser, S. P. Prokaryotic Taxonomy in the Sequencing Era–the Polyphasic Approach Revisited. *Environ. Microbiol.* **2012**, *14* (2), 291−317.

(22) Drews, G. The Roots of Microbiology and the Influence of Ferdinand Cohn on Microbiology of the 19th Century. *FEMS Microbiol Rev.* **2000**, *24* (3), 225−249.

(23) Bergey, D. H.; Breed, R. S.; Murray, E. G. D.; Hitchens, A. P. *Manual of Determinative Bacteriology*, 5th ed.; Bailliere, Tindall & Cox: London, 1939.

(24) Stanier, R. Y.; Van Niel, C. B. The Main Outlines of Bacterial Classification. *J. Bacteriol.* **1941**, *42* (4), 437−466.

(25) Young, J. M. Implications of Alternative Classifications and Horizontal Gene Transfer for Bacterial Taxonomy. *Int. J. Syst. Evol Microbiol* **2001**, *51* (3), 945−953.

(26) Woese, C. R.; Fox, G. E. Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74* (11), 5088−5090.

(27) Lane, D. J.; Pace, B.; Olsen, G. J.; Stahl, D. A.; Sogin, M. L.; Pace, N. R. Rapid Determination of 16S Ribosomal RNA Sequences for Phylogenetic Analyses. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82* (20), 6955−6959.

(28) Hugenholtz, P.; Pitulle, C.; Hershberger, K. L.; Pace, N. R. Novel Division Level Bacterial Diversity in a Yellowstone Hot Spring. *J. Bacteriol.* **1998**, *180* (2), 366−376.

(29) Pace, N. R. Mapping the Tree of Life: Progress and Prospects. *Microbiol Mol. Biol. Rev.* **2009**, *73* (4), 565−576.

(30) Bergey, D. H. *Bergey's Manual® of Systematic Bacteriology*; Springer Science & Business Media, 2001.

(31) Stackebrandt, E.; Goebel, B. M. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology* **1994**, *44* (4), 846−849.

(32) Janda, J. M.; Abbott, S. L. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *J. Clin Microbiol* **2007**, *45* (9), 2761−2764.

(33) Schulz, F.; Eloe-Fadrosh, E. A.; Bowers, R. M.; Jarett, J.; Nielsen, T.; Ivanova, N. N.; Kyrpides, N. C.; Woyke, T. Towards a Balanced View of the Bacterial Tree of Life. *Microbiome* **2017**, *5* (1), 140.

(34) DeSantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G. L. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72* (7), 5069−5072.

(35) Konstantinidis, K. T.; Tiedje, J. M. Towards a Genome-Based Taxonomy for Prokaryotes. *J. Bacteriol.* **2005**, *187* (18), 6258−6264.

(36) Colwell, R. R. Polyphasic Taxonomy of the Genus Vibrio: Numerical Taxonomy of Vibrio Cholerae, Vibrio Parahaemolyticus, and Related Vibrio Species. *J. Bacteriol.* **1970**, *104* (1), 410−433.

(37) Vandamme, P.; Pot, B.; Gillis, M.; de Vos, P.; Kersters, K.; Swings, J. Polyphasic Taxonomy, a Consensus Approach to Bacterial Systematics. *Microbiol Rev.* **1996**, *60* (2), 407−438.

(38) Wu, D.; Hugenholtz, P.; Mavromatis, K.; Pukall, R.; Dalin, E.; Ivanova, N. N.; Kunin, V.; Goodwin, L.; Wu, M.; Tindall, B. J.; Hooper, S. D.; Pati, A.; Lykidis, A.; Spring, S.; Anderson, I. J.; D'haeseleer, P.; Zemla, A.; Singer, M.; Lapidus, A.; Nolan, M.; Copeland, A.; Han, C.; Chen, F.; Cheng, J.-F.; Lucas, S.; Kerfeld, C.; Lang, E.; Gronow, S.; Chain, P.; Bruce, D.; Rubin, E. M.; Kyrpides, N. C.; Klenk, H.-P.; Eisen, J. A. A Phylogeny-Driven Genomic Encyclopaedia of Bacteria and Archaea. *Nature* **2009**, *462* (7276), 1056.

(39) Wu, L.; McCluskey, K.; Desmeth, P.; Liu, S.; Hideaki, S.; Yin, Y.; Moriya, O.; Itoh, T.; Kim, C. Y.; Lee, J.-S.; Zhou, Y.; Kawasaki, H.; Hazbón, M. H.; Robert, V.; Boekhout, T.; Lima, N.; Evtushenko, L.; Boundy-Mills, K.; Bunk, B.; Moore, E. R. B.; Eurwilaichitr, L.; Ingsriswang, S.; Shah, H.; Yao, S.; Jin, T.; Huang, J.; Shi, W.; Sun, Q.; Fan, G.; Li, W.; Li, X.; Kurtböke, I.; Ma, J. The Global Catalogue of Microorganisms 10K Type Strain Sequencing Project: Closing the Genomic Gaps for the Validly Published Prokaryotic and Fungi Species. *Gigascience* **2018**, *7* (5), No. giy026.

(40) Whitman, W. B.; Chuvochina, M.; Hedlund, B. P.; Hugenholtz, P.; Konstantinidis, K. T.; Murray, A. E.; Palmer, M.; Parks, D. H.; Probst, A. J.; Reysenbach, A.-L.; Rodriguez-R, L. M.; Rossello-Mora, R.; Sutcliffe, I.; Venter, S. N. Development of the SeqCode: A Proposed Nomenclatural Code for Uncultivated Prokaryotes with DNA Sequences as Type. *Systematic and Applied Microbiology* **2022**, *45* (5), No. 126305.

(41) Jain, C.; Rodriguez-R, L. M.; Phillippy, A. M.; Konstantinidis, K. T.; Aluru, S. High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. *Nat. Commun.* **2018**, *9* (1), 5114.

(42) Goris, J.; Konstantinidis, K. T.; Klappenbach, J. A.; Coenye, T.; Vandamme, P.; Tiedje, J. M. DNA−DNA Hybridization Values and

Their Relationship to Whole-Genome Sequence Similarities. *International Journal of Systematic and Evolutionary Microbiology* **2007**, *57* (1), 81−91.

(43) Olm, M. R.; Crits-Christoph, A.; Diamond, S.; Lavy, A.; Matheus Carnevali, P. B.; Banfield, J. F. Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. *mSystems* **2020**, *5* (1), No. e00731-19.

(44) Parks, D. H.; Chuvochina, M.; Waite, D. W.; Rinke, C.; Skarshewski, A.; Chaumeil, P.-A.; Hugenholtz, P. A Standardized Bacterial Taxonomy Based on Genome Phylogeny Substantially Revises the Tree of Life. *Nat. Biotechnol.* **2018**, *36* (10), 996−1004.

(45) Wu, D.; Jospin, G.; Eisen, J. A. Systematic Identification of Gene Families for Use as "Markers" for Phylogenetic and Phylogeny-Driven Ecological Studies of Bacteria and Archaea and Their Major Subgroups. *PLoS One* **2013**, *8* (10), e77033.

(46) Chaumeil, P.-A.; Mussig, A. J.; Hugenholtz, P.; Parks, D. H. GTDB-Tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database. *Bioinformatics* **2020**, *36*, 1925.

(47) Parks, D. H.; Chuvochina, M.; Rinke, C.; Mussig, A. J.; Chaumeil, P.-A.; Hugenholtz, P. GTDB: An Ongoing Census of Bacterial and Archaeal Diversity through a Phylogenetically Consistent, Rank Normalized and Complete Genome-Based Taxonomy. *Nucleic Acids Res.* **2022**, *50* (D1), D785−D794.

(48) *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond*; National Academies Press: Washington, D.C., 2014. DOI: 10.17226/18722.

(49) Carroll, K. C.; Munson, E.; Butler-Wu, S. M.; Patrick, S. Point-Counterpoint: What's in a Name? Clinical Microbiology Laboratories Should Use Nomenclature Based on Current Taxonomy. *Journal of Clinical Microbiology* **2023**, *61* (1), No. e01732-22.

(50) Hamilton, K. A.; Weir, M. H.; Haas, C. N. Dose Response Models and a Quantitative Microbial Risk Assessment Framework for the Mycobacterium Avium Complex That Account for Recent Developments in Molecular Biology, Taxonomy, and Epidemiology. *Water Res.* **2017**, *109*, 310−326.

(51) Oren, A.; Göker, M. Validation List No. 213. Valid Publication of New Names and New Combinations Effectively Published Outside the IJSEM. *Int. J. Syst. Evol Microbiol* **2023**, *73* (9), 1 DOI: 10.1099/ijsem.0.005997.

(52) *Journal of Clinical Microbiology*. https://journals.asm.org/journal/jcm/resources-clinical-microbiologists?doi=10.1128%2Feissn.1098-660X&publicationCode=jcm (accessed 2023-11-28).

(53) Runyon, E. H. Anonymous Mycobacteria in Pulmonary Disease. *Med. Clin North Am.* **1959**, *43* (1), 273−290.

(54) Kim, C.-J.; Kim, N.-H.; Song, K.-H.; Choe, P. G.; Kim, E. S.; Park, S. W.; Kim, H.-B.; Kim, N.-J.; Kim, E.-C.; Park, W. B.; Oh, M. Differentiating Rapid- and Slow-Growing Mycobacteria by Difference in Time to Growth Detection in Liquid Media. *Diagn Microbiol Infect Dis* **2013**, *75* (1), 73−76.

(55) Proctor, C.; Garner, E.; Hamilton, K. A.; Ashbolt, N. J.; Caverly, L. J.; Falkinham, J. O.; Haas, C. N.; Prevost, M.; Prevots, D. R.; Pruden, A.; Raskin, L.; Stout, J.; Haig, S.-J. Tenets of a Holistic Approach to Drinking Water-Associated Pathogen Research, Management, and Communication. *Water Res.* **2022**, *211*, No. 117997.

(56) Rogall, T.; Wolters, J.; Flohr, T.; Böttger, E. C. Towards a Phylogeny and Definition of Species at the Molecular Level within the Genus Mycobacterium. *Int. J. Syst. Bacteriol* **1990**, *40* (4), 323−330.

(57) Kent, P. T. *Public Health Mycobacteriology: A Guide for the Level III Laboratory*; U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, 1985.

(58) Butler, W. R.; Guthertz, L. S. Mycolic Acid Analysis by High-Performance Liquid Chromatography for Identification of Mycobacterium Species. *Clin Microbiol Rev.* **2001**, *14* (4), 704−726. Table of contents.

(59) Buckwalter, S. P.; Olson, S. L.; Connelly, B. J.; Lucas, B. C.; Rodning, A. A.; Walchak, R. C.; Deml, S. M.; Wohlfiel, S. L.; Wengenack, N. L. Evaluation of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry for Identification of Mycobacterium Species, Nocardia Species, and Other Aerobic Actinomycetes. *J. Clin Microbiol* **2016**, *54* (2), 376−384.

(60) Markanović, M.; Makek, M. J.; Glodić, G.; Kuliš, T.; Mareković, I. Evaluation and Clinical Impact of MALDI Biotyper Mycobacteria Library v6.0 for Identification of Nontuberculous Mycobacteria by MALDI-TOF Mass Spectrometry. *J. Mass Spectrom* **2023**, *58* (4), No. e4915.

(61) Stahl, D. A.; Urbance, J. W. The Division between Fast- and Slow-Growing Species Corresponds to Natural Relationships among the Mycobacteria. *J. Bacteriol.* **1990**, *172* (1), 116−124.

(62) Tortoli, E. Phylogeny of the Genus Mycobacterium: Many Doubts, Few Certainties. *Infect Genet Evol* **2012**, *12* (4), 827−831.

(63) Tortoli, E. Microbiological Features and Clinical Relevance of New Species of the Genus Mycobacterium. *Clin Microbiol Rev.* **2014**, *27* (4), 727−752.

(64) Devulder, G.; de Montclos, M. P.; Flandrois, J. P. A Multigene Approach to Phylogenetic Analysis Using the Genus Mycobacterium as a Model. *Int. J. Syst. Evol Microbiol* **2005**, *55* (1), 293−302.

(65) Kim, S. H.; Shin, J. H. Identification of Nontuberculous Mycobacteria Using Multilocous Sequence Analysis of 16S rRNA, Hsp65, and rpoB. *J. Clin Lab Anal* **2018**, *32* (1), No. e22184.

(66) Ringuet, H.; Akoua-Koffi, C.; Honore, S.; Varnerot, A.; Vincent, V.; Berche, P.; Gaillard, J. L.; Pierre-Audigier, C. Hsp65 Sequencing for Identification of Rapidly Growing Mycobacteria. *J. Clin Microbiol* **1999**, *37* (3), 852−857.

(67) Kim, H.; Kim, S.-H.; Shim, T.-S.; Kim, M.-N.; Bai, G.-H.; Park, Y.-G.; Lee, S.-H.; Chae, G.-T.; Cha, C.-Y.; Kook, Y.-H.; Kim, B.-J. Differentiation of Mycobacterium Species by Analysis of the Heat-Shock Protein 65 Gene (Hsp65). *Int. J. Syst. Evol Microbiol* **2005**, *55* (4), 1649−1656.

(68) Adékambi, T.; Drancourt, M. Dissection of Phylogenetic Relationships among 19 Rapidly Growing Mycobacterium Species by 16S rRNA, Hsp65, sodA, recA and rpoB Gene Sequencing. *Int. J. Syst. Evol Microbiol* **2004**, *54* (6), 2095−2105.

(69) Mignard, S.; Flandrois, J.-P. A Seven-Gene, Multilocus, Genus-Wide Approach to the Phylogeny of Mycobacteria Using Supertrees. *Int. J. Syst. Evol Microbiol* **2008**, *58* (6), 1432−1441.

(70) Blackwood, K. S.; He, C.; Gunton, J.; Turenne, C. Y.; Wolfe, J.; Kabani, A. M. Evaluation of recA Sequences for Identification of Mycobacterium Species. *J. Clin Microbiol* **2000**, *38* (8), 2846−2852.

(71) Roth, A.; Fischer, M.; Hamid, M. E.; Michalke, S.; Ludwig, W.; Mauch, H. Differentiation of Phylogenetically Related Slowly Growing Mycobacteria Based on 16S-23S rRNA Gene Internal Transcribed Spacer Sequences. *J. Clin Microbiol* **1998**, *36* (1), 139−147.

(72) Guillemin, I.; Cambau, E.; Jarlier, V. Sequences of Conserved Region in the A Subunit of DNA Gyrase from Nine Species of the Genus Mycobacterium: Phylogenetic Analysis and Implication for Intrinsic Susceptibility to Quinolones. *Antimicrob. Agents Chemother.* **1995**, *39* (9), 2145−2149.

(73) Stone, B. B.; Nietupski, R. M.; Breton, G. L.; Weisburg, W. G. Comparison of Mycobacterium 23S rRNA Sequences by High-Temperature Reverse Transcription and PCR. *Int. J. Syst. Bacteriol* **1995**, *45* (4), 811−819.

(74) 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* (1), 144−153.

(75) Zelazny, A. M.; Calhoun, L. B.; Li, L.; Shea, Y. R.; Fischer, S. H. Identification of Mycobacterium Species by secA1 Sequences. *J. Clin Microbiol* **2005**, *43* (3), 1051−1058.

(76) Criscuolo, A.; Berry, V.; Douzery, E. J. P.; Gascuel, O. SDM: A Fast Distance-Based Approach for (Super)Tree Building in Phylogenomics. *Systematic Biology* **2006**, *55* (5), 740−755.

(77) Huang, C.-C.; Chen, J.-H.; Hu, S.-T.; Chiou, C.-S.; Huang, W.-C.; Hsu, J.-Y.; Lu, J.-J.; Shen, G.-H. Combined Rpo B Duplex PCR and Hsp65 PCR Restriction Fragment Length Polymorphism with Capillary Electrophoresis as an Effective Algorithm for Identification

of Mycobacterial Species from Clinical Isolates. *BMC Microbiology* **2012**, *12* (1), No. 137.
(78) Lee, M.-R.; Cheng, A.; Huang, Y.-T.; Liu, C.-Y.; Chung, K.-P.; Wang, H.-C.; Liang, S.-K.; Liao, C.-H.; Yu, C.-J.; Hsueh, P.-R. Performance Assessment of the DR. TBDR/NTM IVD Kit for Direct Detection of Mycobacterium Tuberculosis Isolates, Including Rifampin-Resistant Isolates, and Nontuberculous Mycobacteria. *J. Clin Microbiol* **2012**, *50* (10), 3398−3401.
(79) Drake, T. A.; Hindler, J. A.; Berlin, O. G.; Bruckner, D. A. Rapid Identification of Mycobacterium Avium Complex in Culture Using DNA Probes. *J. Clin Microbiol* **1987**, *25* (8), 1442−1445.
(80) 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* (9), 4418−4420.
(81) Richter, E.; Rüsch-Gerdes, S.; Hillemann, D. Evaluation of the GenoType Mycobacterium Assay for Identification of Mycobacterial Species from Cultures. *J. Clin Microbiol* **2006**, *44* (5), 1769−1775.
(82) Tortoli, E.; Pecorari, M.; Fabio, G.; Messinò, M.; Fabio, A. Commercial DNA Probes for Mycobacteria Incorrectly Identify a Number of Less Frequently Encountered Species. *J. Clin Microbiol* **2010**, *48* (1), 307−310.
(83) Prasanna, A. N.; Mehra, S. Comparative Phylogenomics of Pathogenic and Non-Pathogenic Mycobacterium. *PLoS One* **2013**, *8* (8), No. e71248.
(84) Wee, W. Y.; Dutta, A.; Choo, S. W. Comparative Genome Analyses of Mycobacteria Give Better Insights into Their Evolution. *PLoS One* **2017**, *12* (3), No. e0172831.
(85) Fedrizzi, T.; Meehan, C. J.; Grottola, A.; Giacobazzi, E.; Fregni Serpini, G.; Tagliazucchi, S.; Fabio, A.; Bettua, C.; Bertorelli, R.; De Sanctis, V.; Rumpianesi, F.; Pecorari, M.; Jousson, O.; Tortoli, E.; Segata, N. Genomic Characterization of Nontuberculous Mycobacteria. *Sci. Rep* **2017**, *7*, 45258.
(86) Gómez-Valero, L.; Rocha, E. P. C.; Latorre, A.; Silva, F. J. Reconstructing the Ancestor of Mycobacterium Leprae: The Dynamics of Gene Loss and Genome Reduction. *Genome Res.* **2007**, *17* (8), 1178−1185.
(87) Tortoli, E.; Fedrizzi, T.; Meehan, C. J.; Trovato, A.; Grottola, A.; Giacobazzi, E.; Serpini, G. F.; Tagliazucchi, S.; Fabio, A.; Bettua, C.; Bertorelli, R.; Frascaro, F.; De Sanctis, V.; Pecorari, M.; Jousson, O.; Segata, N.; Cirillo, D. M. The New Phylogeny of the Genus Mycobacterium: The Old and the News. *Infect Genet Evol* **2017**, *56*, 19−25.
(88) Gao, B.; Gupta, R. S. Phylogenetic Framework and Molecular Signatures for the Main Clades of the Phylum Actinobacteria. *Microbiol Mol. Biol. Rev.* **2012**, *76* (1), 66−112.
(89) Barco, R. A.; Garrity, G. M.; Scott, J. J.; Amend, J. P.; Nealson, K. H.; Emerson, D. A Genus Definition for Bacteria and Archaea Based on a Standard Genome Relatedness Index. *mBio* **2020**, *11* (1), e02475-19.
(90) Quan, T. P.; Bawa, Z.; Foster, D.; Walker, T.; del Ojo Elias, C.; Rathod, P.; Iqbal, Z.; Bradley, P.; Mowbray, J.; Walker, A. S.; Crook, D. W.; Wyllie, D. H.; Peto, T. E. A.; Smith, E. G. 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* (2), No. e01480-17.
(91) Hunt, M.; Bradley, P.; Lapierre, S. G.; Heys, S.; Thomsit, M.; Hall, M. B.; Malone, K. M.; Wintringer, P.; Walker, T. M.; Cirillo, D. M.; Comas, I.; Farhat, M. R.; Fowler, P.; Gardy, J.; Ismail, N.; Kohl, T. A.; Mathys, V.; Merker, M.; Niemann, S.; Omar, S. V.; Sintchenko, V.; Smith, G.; van Soolingen, D.; Supply, P.; Tahseen, S.; Wilcox, M.; Arandjelovic, I.; Peto, T. E. A.; Crook, D. W.; Iqbal, Z. Antibiotic Resistance Prediction for Mycobacterium Tuberculosis from Genome Sequence Data with Mykrobe. *Wellcome Open Res.* **2019**, *4*, 191.
(92) 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. *Emerging Microbes & Infections* **2019**, *8* (1), 1043−1053.
(93) Nadon, C.; Walle, I. V.; Gerner-Smidt, P.; Campos, J.; Chinen, I.; Concepcion-Acevedo, J.; Gilpin, B.; Smith, A. M.; Kam, K. M.; Perez, E.; Trees, E.; Kubota, K.; Takkinen, J.; Nielsen, E. M.; Carleton, H. PulseNet International: Vision for the Implementation of Whole Genome Sequencing (WGS) for Global Food-Borne Disease Surveillance. *Eurosurveillance* **2017**, *22* (23), 30544.
(94) 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), 322.
(95) Lande, L.; Alexander, D. C.; Wallace, R. J.; Kwait, R.; Iakhiaeva, E.; Williams, M.; Cameron, A. D. S.; Olshefsky, S.; Devon, R.; Vasireddy, R.; Peterson, D. D.; Falkinham, J. O. Mycobacterium Avium in Community and Household Water, Suburban Philadelphia, Pennsylvania, USA, 2010−2012. *Emerg Infect Dis* **2019**, *25* (3), 473−481.
(96) Haig, S.-J.; Kotlarz, N.; LiPuma, J. J.; 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* (1), No. e02354-17.
(97) Gebert, M. J.; Delgado-Baquerizo, M.; Oliverio, A. M.; Webster, T. M.; Nichols, L. M.; Honda, J. R.; Chan, E. D.; Adjemian, J.; Dunn, R. R.; Fierer, N. Ecological Analyses of Mycobacteria in Showerhead Biofilms and Their Relevance to Human Health. *mBio* **2018**, *9* (5), 01614-18.
(98) Ghosh, S.; Strom, L.; Zhu, N. J.; Donaldson, K. E.; Kurowski, A.; Majeed, H.; Falkinham, J. O. I.; Edwards, M. A.; Pruden, A. Metagenomic Profiling of Pathogens in Simulated Reclaimed Water Distribution Systems Operated at Elevated Temperature Reveals Distinct Response of Mycobacteria to Filtration and Disinfection Conditions. *ACS EST Water* **2023**, *3* (3), 720−732.
(99) Milanese, A.; Mende, D. R.; Paoli, L.; Salazar, G.; Ruscheweyh, H.-J.; Cuenca, M.; Hingamp, P.; Alves, R.; Costea, P. I.; Coelho, L. P.; Schmidt, T. S. B.; Almeida, A.; Mitchell, A. L.; Finn, R. D.; Huerta-Cepas, J.; Bork, P.; Zeller, G.; Sunagawa, S. Microbial Abundance, Activity and Population Genomic Profiling with mOTUs2. *Nat. Commun.* **2019**, *10* (1), 1014.
(100) Truong, D. T.; Franzosa, E. A.; Tickle, T. L.; Scholz, M.; Weingart, G.; Pasolli, E.; Tett, A.; Huttenhower, C.; Segata, N. MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling. *Nat. Methods* **2015**, *12* (10), 902−903.
(101) Wood, D. E.; Lu, J.; Langmead, B. Improved Metagenomic Analysis with Kraken 2. *Genome Biol.* **2019**, *20* (1), 257.
(102) Lu, J.; Breitwieser, F. P.; Thielen, P.; Salzberg, S. L. Bracken: Estimating Species Abundance in Metagenomics Data. *PeerJ. Comput. Sci.* **2017**, *3*, No. e104.
(103) Liang, Q.; Bible, P. W.; Liu, Y.; Zou, B.; Wei, L. DeepMicrobes: Taxonomic Classification for Metagenomics with Deep Learning. *NAR Genomics and Bioinformatics* **2020**, *2* (1), No. lqaa009.
(104) Park, H.; Lim, S. J.; Cosme, J.; O'Connell, K.; Sandeep, J.; Gayanilo, F.; Cutter Jr, G. R.; Montes, E.; Nitikitpaiboon, C.; Fisher, S.; Moustahfid, H.; Thompson, L. R. Investigation of Machine Learning Algorithms for Taxonomic Classification of Marine Metagenomes. *Microbiology Spectrum* **2023**, *11* (5), No. e05237-22.
(105) Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*; Princeton University Press, 2011. DOI: 10.1515/9781400837526.
(106) Wang, F.; Men, X.; Zhang, G.; Liang, K.; Xin, Y.; Wang, J.; Li, A.; Zhang, H.; Liu, H.; Wu, L. Assessment of 16S rRNA Gene Primers for Studying Bacterial Community Structure and Function of Aging Flue-Cured Tobaccos. *AMB Express* **2018**, *8* (1), No. 182.
(107) Youssef, N. H.; Couger, M. B.; McCully, A. L.; Criado, A. E. G.; Elshahed, M. S. Assessing the Global Phylum Level Diversity within the Bacterial Domain: A Review. *J. Adv. Res.* **2015**, *6* (3), 269−282.

(108) Haig, S.-J.; Kotlarz, N.; LiPuma, J. J.; 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* (1), No. e02354-17.
(109) Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30* (14), 2068−2069.
(110) Hyatt, D.; Chen, G.-L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinformatics* **2010**, *11* (1), 119.
(111) Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S. K.; Cook, H.; Mende, D. R.; Letunic, I.; Rattei, T.; Jensen, L. J.; von Mering, C.; Bork, P. eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. *Nucleic Acids Res.* **2019**, *47* (D1), D309−D314.
(112) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27−30.
(113) Tatusov, R. L.; Galperin, M. Y.; Natale, D. A.; Koonin, E. V. The COG Database: A Tool for Genome-Scale Analysis of Protein Functions and Evolution. *Nucleic Acids Res.* **2000**, *28* (1), 33−36.
(114) O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C. M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina, W.; Joardar, V. S.; Kodali, V. K.; Li, W.; Maglott, D.; Masterson, P.; McGarvey, K. M.; Murphy, M. R.; O'Neill, K.; Pujar, S.; Rangwala, S. H.; Rausch, D.; Riddick, L. D.; Schoch, C.; Shkeda, A.; Storz, S. S.; Sun, H.; Thibaud-Nissen, F.; Tolstoy, I.; Tully, R. E.; Vatsan, A. R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M. J.; Kimchi, A.; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T. D.; Pruitt, K. D. Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* **2016**, *44* (D1), D733−D745.
(115) Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G. A.; Sonnhammer, E. L. L.; Tosatto, S. C. E.; Paladin, L.; Raj, S.; Richardson, L. J.; Finn, R. D.; Bateman, A. Pfam: The Protein Families Database in 2021. *Nucleic Acids Res.* **2021**, *49* (D1), D412−D419.
(116) Eddy, S. R. Hidden Markov Models. *Curr. Opin. Struct. Biol.* **1996**, *6* (3), 361−365.
(117) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403−410.
(118) Buchfink, B.; Reuter, K.; Drost, H.-G. Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND. *Nat. Methods* **2021**, *18* (4), 366−368.
(119) Bachmann, N. L.; Salamzade, R.; Manson, A. L.; Whittington, R.; Sintchenko, V.; Earl, A. M.; Marais, B. J. Key Transitions in the Evolution of Rapid and Slow Growing Mycobacteria Identified by Comparative Genomics. *Frontiers in Microbiology* **2020**, *10*, 3019.
(120) Onwueme, K. C.; Vos, C. J.; Zurita, J.; Ferreras, J. A.; Quadri, L. E. N. The Dimycocerosate Ester Polyketide Virulence Factors of Mycobacteria. *Prog. Lipid Res.* **2005**, *44* (5), 259−302.
(121) Siegrist, M. S.; Unnikrishnan, M.; McConnell, M. J.; Borowsky, M.; Cheng, T.-Y.; Siddiqi, N.; Fortune, S. M.; Moody, D. B.; Rubin, E. J. Mycobacterial Esx-3 Is Required for Mycobactin-Mediated Iron Acquisition. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (44), 18792−18797.
(122) Hull, N. M.; Ling, F.; Pinto, A. J.; Albertsen, M.; Jang, H. G.; Hong, P.-Y.; Konstantinidis, K. T.; LeChevallier, M.; Colwell, R. R.; Liu, W.-T. Drinking Water Microbiome Project: Is It Time? *Trends Microbiol.* **2019**, *27* (8), 670−677.
(123) *GTDB Release 207 statistics*. https://gtdb.ecogenomic.org/stats/r207 (accessed 2022-10-24).
(124) Paradis, E.; Schliep, K. Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R. *Bioinformatics* **2019**, *35* (3), 526−528.
(125) Yu, G. Using Ggtree to Visualize Data on Tree-Like Structures. *Curr. Protoc Bioinformatics* **2020**, *69* (1), No. e96.
(126) R Core Team. *R: A Language and Environment for Statistical Computing*.; R Foundation for Statistical Computing: Vienna, Austria, 2013.
(127) Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F. O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res.* **2012**, *41*, D590−596.
(128) Bolyen, E.; Rideout, J. R.; Dillon, M. R.; Bokulich, N. A.; Abnet, C. C.; Al-Ghalith, G. A.; Alexander, H.; Alm, E. J.; Arumugam, M.; Asnicar, F.; Bai, Y.; Bisanz, J. E.; Bittinger, K.; Brejnrod, A.; Brislawn, C. J.; Brown, C. T.; Callahan, B. J.; Caraballo-Rodríguez, A. M.; Chase, J.; Cope, E. K.; Da Silva, R.; Diener, C.; Dorrestein, P. C.; Douglas, G. M.; Durall, D. M.; Duvallet, C.; Edwardson, C. F.; Ernst, M.; Estaki, M.; Fouquier, J.; Gauglitz, J. M.; Gibbons, S. M.; Gibson, D. L.; Gonzalez, A.; Gorlick, K.; Guo, J.; Hillmann, B.; Holmes, S.; Holste, H.; Huttenhower, C.; Huttley, G. A.; Janssen, S.; Jarmusch, A. K.; Jiang, L.; Kaehler, B. D.; Kang, K. B.; Keefe, C. R.; Keim, P.; Kelley, S. T.; Knights, D.; Koester, I.; Kosciolek, T.; Kreps, J.; Langille, M. G. I.; Lee, J.; Ley, R.; Liu, Y.-X.; Loftfield, E.; Lozupone, C.; Maher, M.; Marotz, C.; Martin, B. D.; McDonald, D.; McIver, L. J.; Melnik, A. V.; Metcalf, J. L.; Morgan, S. C.; Morton, J. T.; Naimey, A. T.; Navas-Molina, J. A.; Nothias, L. F.; Orchanian, S. B.; Pearson, T.; Peoples, S. L.; Petras, D.; Preuss, M. L.; Pruesse, E.; Rasmussen, L. B.; Rivers, A.; Robeson, M. S.; Rosenthal, P.; Segata, N.; Shaffer, M.; Shiffer, A.; Sinha, R.; Song, S. J.; Spear, J. R.; Swafford, A. D.; Thompson, L. R.; Torres, P. J.; Trinh, P.; Tripathi, A.; Turnbaugh, P. J.; Ul-Hasan, S.; vander Hooft, J. J. J.; Vargas, F.; Vázquez-Baeza, Y.; Vogtmann, E.; von Hippel, M.; Walters, W.; Wan, Y.; Wang, M.; Warren, J.; Weber, K. C.; Williamson, C. H. D.; Willis, A. D.; Xu, Z. Z.; Zaneveld, J. R.; Zhang, Y.; Zhu, Q.; Knight, R.; Caporaso, J. G. Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2. *Nat. Biotechnol.* **2019**, *37* (8), 852−857.
(129) Reduce Risk from Water. HAIs. *CDC*. https://www.cdc.gov/hai/prevent/environment/water.html (accessed 2023-11-29).