A High-Throughput Approach for Identification of Nontuberculous Mycobacteria in Drinking Water Reveals Relationship between Water Age and *Mycobacterium avium*

Sarah-Jane Haig,[a] Nadine Kotlarz,[a*] John J. LiPuma,[b] Lutgarde Raskin[a]

[a]Department of Civil & Environmental Engineering, University of Michigan, Ann Arbor, Michigan, USA
[b]Department of Pediatrics and Communicable Diseases, University of Michigan Medical School, Ann Arbor, Michigan, USA

ABSTRACT Nontuberculous mycobacteria (NTM) frequently detected in drinking water (DW) include species associated with human infections, as well as species rarely linked to disease. Methods for improved the recovery of NTM DNA and high-throughput identification of NTM are needed for risk assessment of NTM infection through DW exposure. In this study, different methods of recovering bacterial DNA from DW were compared, revealing that a phenol-chloroform DNA extraction method yielded two to four times as much total DNA and eight times as much NTM DNA as two commercial DNA extraction kits. This method, combined with high-throughput, single-molecule real-time sequencing of NTM *rpoB* genes, allowed the identification of NTM to the species, subspecies, and (in some cases) strain levels. This approach was applied to DW samples collected from 15 households serviced by a chloraminated distribution system, with homes located in areas representing short (<24 h) and long (>24 h) distribution system residence times. Multivariate statistical analysis revealed that greater water age (i.e., combined distribution system residence time and home plumbing stagnation time) was associated with a greater relative abundance of *Mycobacterium avium* subsp. *avium*, one of the most prevalent NTM causing infections in humans. DW from homes closer to the treatment plant (with a shorter water age) contained more diverse NTM species, including *Mycobacterium abscessus* and *Mycobacterium chelonae*. Overall, our approach allows NTM identification to the species and subspecies levels and can be used in future studies to assess the risk of waterborne infection by providing insight into the similarity between environmental and infection-associated NTM.

IMPORTANCE An extraction method for improved recovery of DNA from nontuberculous mycobacteria (NTM), combined with single-molecule real-time sequencing (PacBio) of NTM *rpoB* genes, was used for high-throughput characterization of NTM species and in some cases strains in drinking water (DW). The extraction procedure recovered, on average, eight times as much NTM DNA and three times as much total DNA from DW as two widely used commercial DNA extraction kits. The combined DNA extraction and sequencing approach allowed high-throughput screening of DW samples to identify NTM, revealing that the relative abundance of *Mycobacterium avium* subsp. *avium* increased with water age. Furthermore, the two-step barcoding approach developed as part of the PacBio sequencing method makes this procedure highly adaptable, allowing it to be used for other target genes and species.

KEYWORDS DNA extraction, drinking water, *Mycobacterium avium*, NTM, PacBio, premise plumbing

Management of the microbial quality of drinking water (DW) is aimed primarily at minimizing illness caused by waterborne pathogens. The increasing prevalence of infections due to nontuberculous mycobacteria (NTM) (1–4), combined with growing evidence linking such infections to DW (5–10), has highlighted NTM exposure through DW as an emerging public health challenge.

Only a few of the >150 NTM species that have been described (11) cause infections in humans (10, 12, 13). Among these are NTM frequently detected in DW and DW distribution systems (DS), including species commonly associated with human infections (e.g., *Mycobacterium avium* and *Mycobacterium abscessus*) (7, 10, 13), as well as species that rarely cause infections (e.g., *Mycobacterium frederiksbergense* and *Mycobacterium aurum*) (14, 15). Molecular surveys targeting short regions of the bacterial 16S rRNA gene are widely used to profile bacterial communities in DW; however, this target does not provide sufficient resolution to distinguish NTM species from each other (13). Improved methods for NTM identification in DW are thus critical for risk assessment of NTM infection through DW exposure.

NTM detection and differentiation in environmental samples such as DW is challenging in that NTM cells are hydrophobic and relatively impermeable, rendering them difficult to lyse (16). In addition, NTM may infect and replicate within protozoa (17–19), further limiting access to their nucleic acids. Finally, sequence analysis of relatively large (>750-bp) fragments of phylogenetically informative genes, such as *rpoB* and *hsp65*, is required to differentiate NTM species, as even full-length 16S rRNA gene sequences do not provide the resolution needed for discrimination to the species and subspecies levels (13, 20, 21). While studies to date have relied on Sanger sequencing of clone libraries of such genes (22–24), this approach is labor intensive and not well suited to high-throughput screening to detect NTM in environmental samples such as DW.

A major obstacle for high-throughput differentiation of NTM species and strains is the difficulty in sequencing long DNA fragments (>750 bp) with low error rates. The high guanine-cytosine content of NTM genomes (25) provides an additional challenge by increasing the potential for positional errors in several sequencing platforms (26). van der Wielen et al. (14) and Gomez-Smith et al. (15) applied high-throughput sequencing approaches targeting *hsp65* and using the 454 and MiSeq platforms, respectively. These approaches, however, failed to differentiate between some members of the *M. avium* and *M. abscessus* complexes, were limited in differentiating NTM at the strain level, and were prone to primer cross hybridization to other bacterial species.

In this study, we sought to determine if single-molecule real-time sequencing with the PacBio platform (Pacific Biosciences, Menlo Park, CA) could overcome these obstacles. This platform achieves long read lengths (>15,000 bp) with low error rates when used in conjunction with the circular consensus sequencing (CCS) approach, making it possible to distinguish single nucleotide polymorphisms (SNPs) and allowing, in some cases, strain level differentiation (27–29). Combining an improved method of DNA recovery from NTM with PacBio sequencing, we investigated the NTM composition of DW samples from 15 households serviced by the same DW treatment plant. In doing so, we assessed the impact of water age (i.e., combined residence time in the DS and stagnation time in premise plumbing) on the relative abundance of NTM species.

RESULTS

Evaluation of DNA extraction methods. Two phenol-chloroform methods (designated PC1 and PC2) and two commercially available DNA extraction spin kits (FastDNA and Maxwell) (Table 1) were compared with respect to their total and NTM DNA yields from DW biomass. The phenol-chloroform-based methods yielded two to four times as much total DNA as the kits and significantly more NTM DNA ($P < 0.001$), as determined by measuring NTM abundances by quantitative PCR (qPCR) targeting the *atpE* gene (Fig. 1). However, the *Pseudomonas* DNA yields of the four extraction methods, as measured by a qPCR assay targeting the 16S rRNA gene, were comparable (Fig. 1B). PC2 employed a higher sodium dodecyl sulfate (SDS) concentration than PC1 and, when

TABLE 1 Summary of the DNA extraction methods used in this study

Method	Abbreviation	Procedure
FastDNA spin kit for soil[a]	FastDNA	Mechanical (bead beating) and chemical lysis; genomic DNA purified via solutions in kit with spin filter columns
Maxwell LEV Blood DNA kit[b]	Maxwell	As described by Webster et al. (53); mechanical (bead beating) and chemical lysis; genomic DNA purified via solutions and magnetic beads in kit
Phenol-chloroform DNA extraction	PC1	Chemical lysis with modified UNEX buffer[c] with 0.0002% SDS and mechanical lysis (bead beating) for 45 s or 2 or 5 min
Phenol-chloroform DNA extraction	PC2	Chemical lysis with modified UNEX buffer with 0.09% SDS and mechanical lysis (bead beating) for 45 s or 2 or 5 min

[a]MP Biomedicals, Solon, OH.
[b]Promega, Madison, WI.
[c]Modified UNEX buffer composition (final concentrations): 2.22 M guanidinium thiocyanate (MP Biomedicals, Solon, OH), 0.0042% Tween 20, 0.0691 M sodium acetate, 0.1185 M sodium chloride, 0.1975% sodium sulfite, 0.0988% dithioerythritol, and 816 milliAnson U (1 milliAnson U is defined as the amount of enzyme that liberates Folin-positive amino acids and peptides corresponding to 1 μmol of tyrosine under assay conditions in 1 min with hemoglobin as the substrate) of proteinase K (Ambion, Waltham, MA).

used with 5 min of bead beating, resulted in significantly greater total DNA yields ($P = 0.043$) and NTM abundances ($P = 0.045$) (Fig. 1B) than the other conditions for PC2 and PC1. There were no significant differences in the total DNA yields obtained with increased bead-beating times (from 45 s to 2 min to 5 min) by either PC1 or PC2 (Fig. 1A) ($P = 0.10$, 0.19, and 0.66, respectively). Likewise, there was no significant difference in the total bacterial, NTM, and *Pseudomonas* abundances obtained with increased bead-beating times (from 45 s to 2 min to 5 min) by either PC1 or PC2 (Fig. 1B).

As PC2 with 0.09% SDS and 5 min of bead beating yielded the greatest total DNA and NTM abundance, it was selected as the extraction method for the remainder of this study. The extraction efficiency of this procedure was determined by spiking 1 liter of phosphate-buffered saline (PBS) with *Escherichia coli* or *M. abscessus* cells. By using
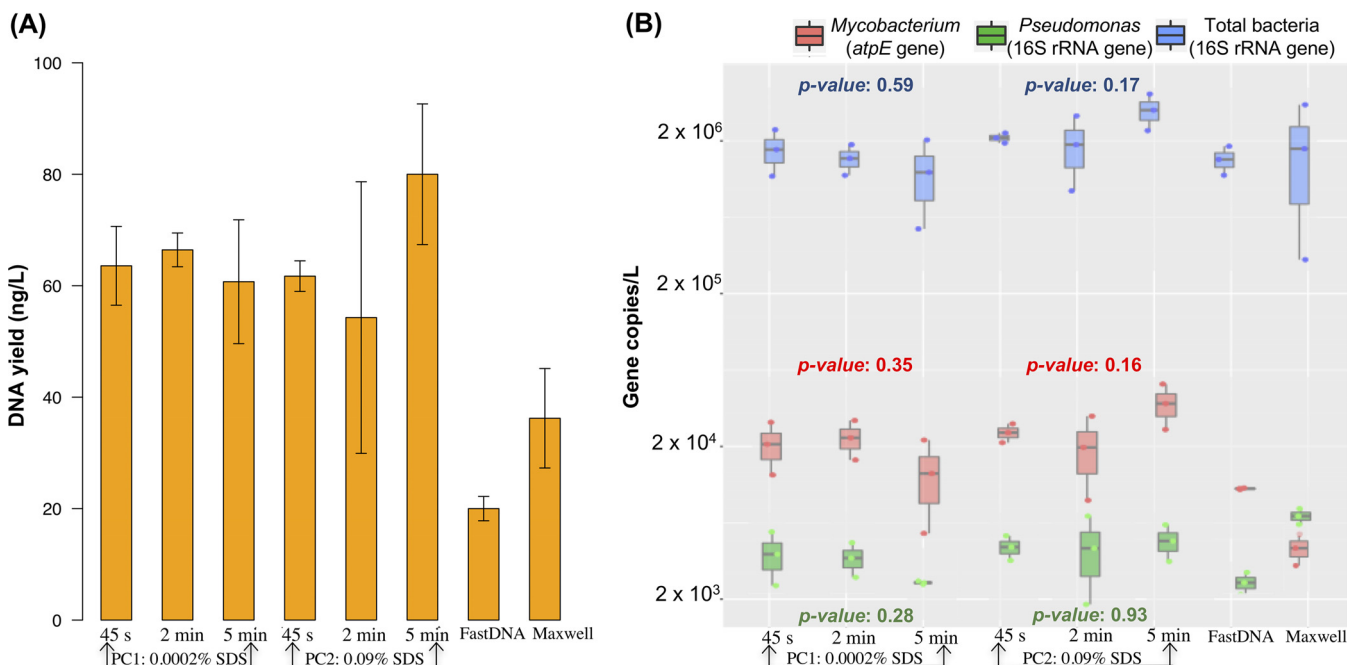

FIG 1 DNA extraction yields (A) and concentrations of NTM, *Pseudomonas*, and total bacteria as determined by qPCR targeting the NTM *atpE*, *Pseudomonas* 16S rRNA, and bacterial 16S rRNA genes, respectively (B), for four different DNA extraction methods on biomass filtered from 1 liter of DW. Bars and points represent the average values of triplicate experiments with the SD represented by whiskers. PC1 and PC2 are phenol-chloroform methods using the modified UNEX buffer (Table 1) with bead-beating times of 45 s, 2 min, and 5 min. *P* values represent the statistical significance of differences between bead-beating times within each method for each microbial group. FastDNA is the MP Biomedicals FastSpin kit for soil, and Maxwell is the Promega Maxwell LEV Blood DNA kit.
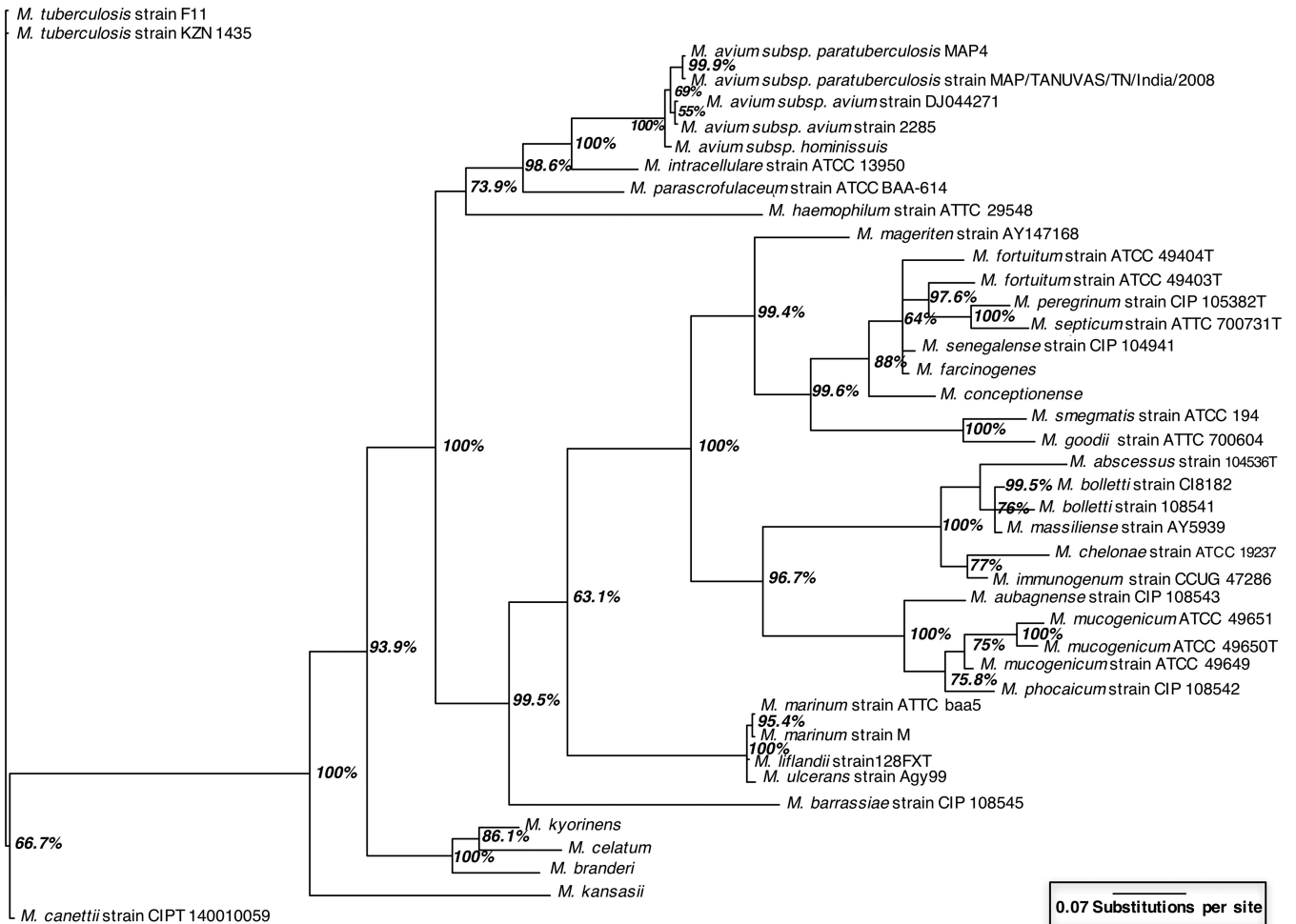
FIG 2 Phylogenetic tree created from sequence analysis of the 942- to 957-bp targeted section of the *rpoB* gene from 41 NTM for which full-length *rpoB* sequences are available in the GenBank database. The percentage on each branch indicates the Bayesian posterior probability value, and branch length indicates the relative inferred evolutionary distance between isolates. An *rpoB* gene sequence from *Mycobacterium tuberculosis* was used as the outgroup.

extraction-independent fluorescence microscopy to enumerate *E. coli* and *M. abscessus* cells and assuming a theoretical DNA content per cell, the theoretical DNA yield was estimated and compared with the measured DNA yields, revealing extraction efficiencies of 78.6% ± 9.5% and 26.7% ± 6.3% for *E. coli* and *M. abscessus* cells, respectively (see Text S1 in the supplemental material).

Evaluation of NTM identification method. Newly developed *rpoB* primers (see Text S2) produced amplicons of various lengths (942 to 957 bp) and enabled the differentiation of the 41 NTM that currently have full-length *rpoB* sequences in the GenBank database, including species and subspecies within the difficult-to-distinguish *M. avium* complex (*M. avium*, *M. avium* subsp. *paratuberculosis*, and *Mycobacterium intracellulare*) and subspecies of *Mycobacterium fortuitum* and *Mycobacterium mucogenicum* (Fig. 2). Over the *rpoB* amplicon, there were 3 to 125 variable regions distinguishing NTM, representing SNPs and insertion sites.

In multiplex sequencing, reads are sorted into sample libraries on the basis of the detection of a unique, sample-specific barcode added to the DNA to be sequenced. We implemented a two-step PCR approach in which *rpoB*-targeted primers containing an M13 motif amplify the template in the first step. A dilution of the amplicons produced by this step served as the template for successive amplification with bar-coded primers containing complements of the M13 motifs (Fig. 3). To evaluate whether our *rpoB* primers and barcoding strategy could be used for NTM differentiation in DW, *rpoB*
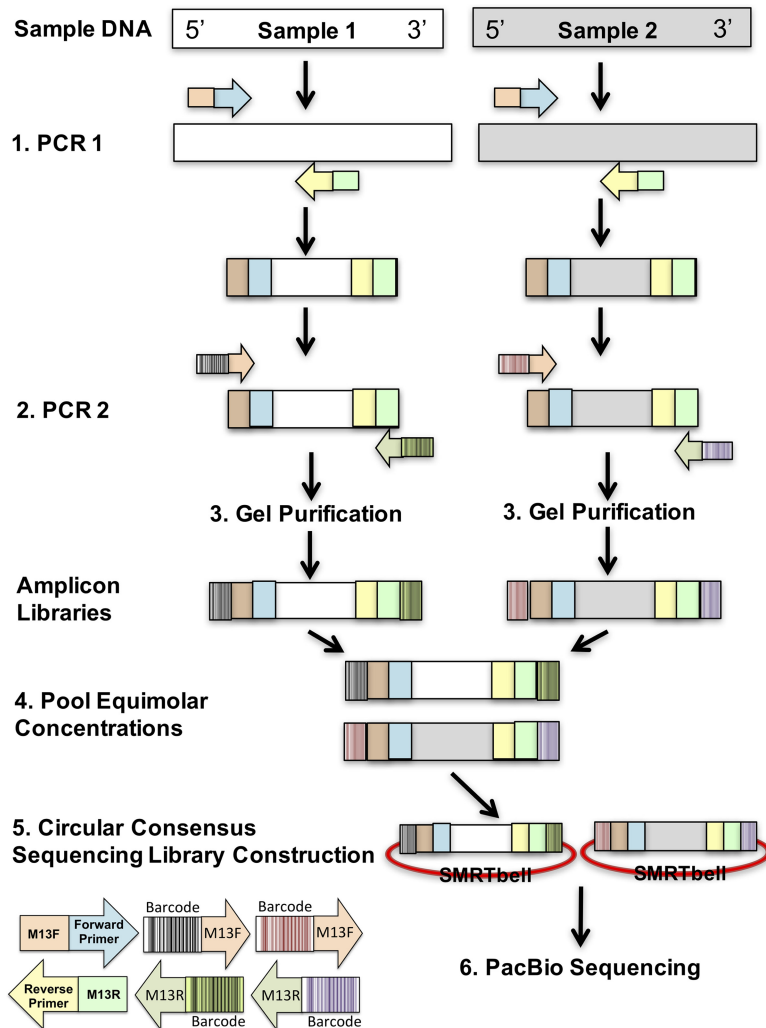
FIG 3 Schematic overview of the PacBio barcoding approach. For simplicity, this schematic shows how amplicon libraries are created from two separate samples. The approach includes two PCRs, the first specific to the target with M13 motifs on the 5′ and 3′ ends of the forward and reverse primers, respectively. The second PCR uses the product from step 1 and involves the addition of unique forward and reverse barcodes to the amplicons. Overall, the amplicon library created for each sample possesses a unique combination of barcodes, allowing both sample libraries to be pooled and distinguished postsequencing. Upon pooling, equal concentrations of the amplicon libraries (step 4) are submitted for PacBio sequencing by the CCS approach entailing circularization of the amplicons (SMRTbell).

amplicons from 16 DW samples were pooled. Fifteen of the samples consisted of biomass from 1 liter of DW collected from homes after at least 6 h of stagnation in premise plumbing, and the 16th sample was a positive control. After quality filtering, 49,728 sequences remained and >99% of them were taxonomically classified as *Mycobacterium* at a sequence identity of ≥85% (30). NTM species and strains were assigned at a sequence identity of ≥99.8%. The *rpoB* amplicons from the positive control, a mixture of two NTM operational taxonomic units (OTUs), were correctly identified as *M. avium* subsp. *avium* and *M. chelonae* (Fig. 4). The experimentally determined abundances of these two OTUs were very similar (97.5 and 2.5%) to those expected (95 and 5%, respectively). The taxonomic and abundance matches obtained for this positive control suggested a high level of confidence for NTM identification in DW samples by this sequencing method.

NTM identification in DW. DW samples were collected from 15 homes serviced by the same municipal DW treatment plant (Ann Arbor, MI). The DS water residence times of these homes varied from 11 to 113 h according to the city's hydraulic model. Homes
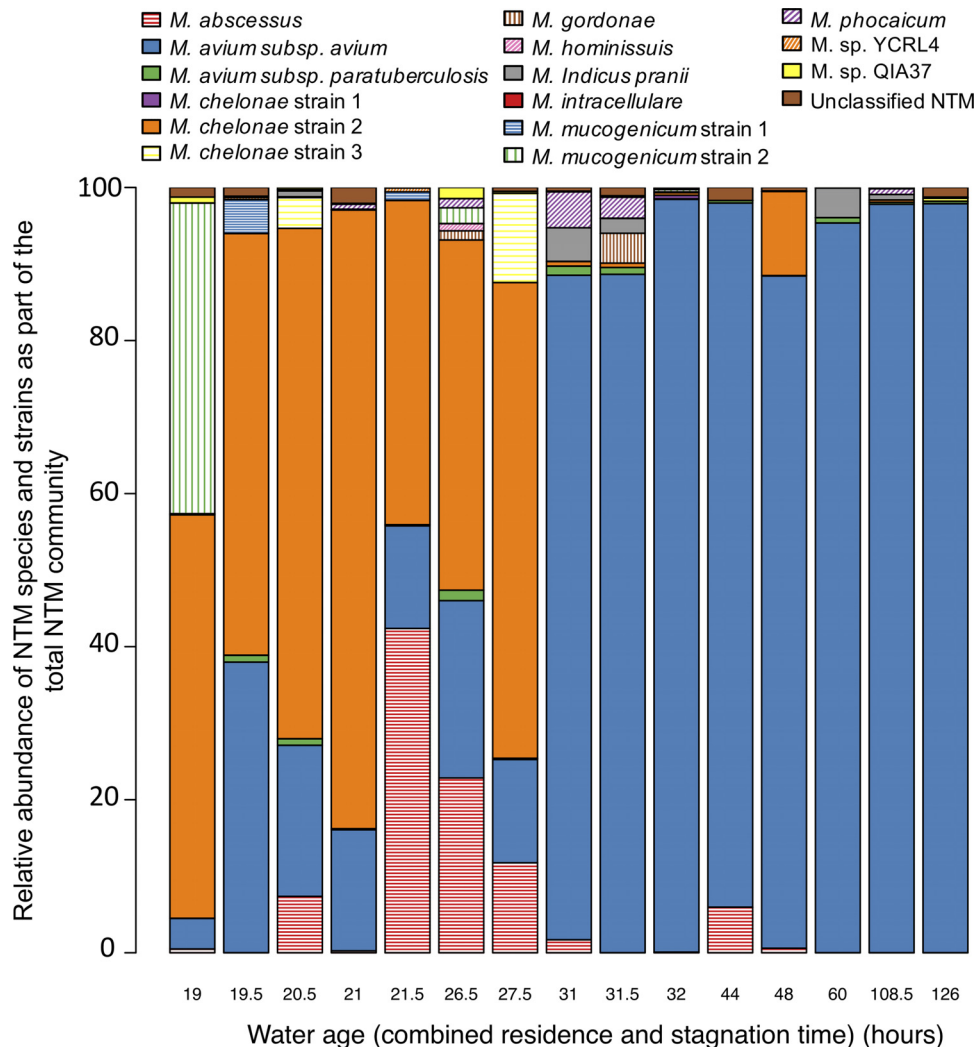
FIG 4 Relative abundances of NTM species and strains as part of the total NTM community in biomass from 15 DW samples collected from premise plumbing. Water age represents the total of water residence time in the distribution system and stagnation time in premise plumbing.

were divided into groups near (<24-h DS residence time) and far from (>24-h DS residence time) the DW treatment plant, corresponding to low and high water ages (i.e., DS residence time combined with premise plumbing stagnation time), respectively.

The sequencing effort resulted in 3,059 ± 1,023 (average ± standard deviation [SD]) NTM *rpoB* gene sequences per sample. The number of NTM OTUs present in DW samples ranged from 4 to 10 (see Text S3). Overall, 16 NTM OTUs were identified, including three unclassified and nine classified NTM species, representing both rapidly (e.g., *M. abscessus*) and slowly (e.g., *M. avium*) growing species and including three distinct strains of *M. chelonae* and two strains of *M. mucogenicum* (Fig. 4). There was no statistically significant difference in NTM diversity between the water age groups ($P = 0.057$), as measured by using the Shannon index (see Text S3). DW samples with a lower water age had a more even NTM community structure (average Pielou's evenness ± SD of 0.49 ± 0.11), and *M. avium* subsp. *avium*, *M. abscessus*, and several *M. chelonae* strains comprised the majority of the NTM community. In comparison, DW samples with a higher water age exhibited a more uneven NTM community structure (average Pielou's evenness of 0.17 ± 0.09) (Fig. 4). *M. avium* subsp. *avium* dominated samples with a higher water age.

Permutational multivariate analysis of variance revealed that the NTM community structures of the two water age groups were significantly different from each other ($P =$

0.001), with water age explaining 81% of the difference ($P = 0.03$). The residual monochloramine concentration (C) multiplied by water age (i.e., the combined DS residence time and premise plumbing stagnation time [t]), referred to collectively as Ct, was not a significant parameter in the multivariate model explaining NTM community structure. However, a greater Ct was significantly associated with a greater NTM qPCR abundance ($P < 0.001$), with the average concentration across all samples corresponding to $6.47 \times 10^6 \pm 1.03 \times 10^6$ NTM *atpE* gene copies/liter. Further, additional multivariate analysis using estimated *M. avium* subsp. *avium* concentrations, obtained by combining *M. avium* subsp. *avium* relative abundances determined by the PacBio sequencing method with NTM qPCR concentrations (see Text S4), revealed that neither the monochloramine concentration nor Ct was a significant factor explaining *M. avium* subsp. *avium* abundance ($P = 0.90$ and 0.08, respectively).

DISCUSSION

Improved NTM DNA recovery. DNA extraction methods impact the abundance and types of microorganisms recovered from complex environmental and biological samples owing to physiological differences among microorganisms and differing extraction efficiencies (31, 32). Given the health risks posed by certain NTM species that can be found in DW, high-throughput methods for NTM identification in DW are needed.

Previous studies assessing the abundance of NTM in DW (14, 22, 23, 33) have generally relied on commercially available DNA extraction kits. These kits typically incorporate only modest physical disruption of bacterial cells (e.g., bead-beating times of <2 min) and do not include reagents necessary for efficient nucleic acid extraction from a range of hard-to-lyse microorganisms (34–36). Our study revealed that two commonly used commercial DNA extraction kits performed poorly in comparison with a phenol-chloroform procedure (PC2) that recovered, on average, eight times as much NTM and three times as much total DNA from DW biomass. The better performance of the PC2 method is attributed to more aggressive physical, enzymatic, and chemical lysis of NTM cell walls, which consist of peptidoglycan, arabinogalactan, and mycolic acids that are covalently linked to form a hydrophobic, relatively impermeable complex (16, 37). Specifically, the longer bead-beating time and increased concentrations of proteinase K and surfactants (SDS and Tween 20) likely improved physical lysis and aided in solubilizing and denaturing proteins (38) in the NTM cell wall, respectively. However, although PC2 yielded significantly more total DNA than the commercial kits, using this method did not result in significant differences in the abundance of bacterial 16S rRNA genes. This apparent discrepancy may be explained by differences in the potential of the extraction methods used to obtain other fractions of DNA, including DNA of *Eukarya*, which have complex cell walls, making them difficult to lyse (35).

Despite the impact DNA extraction procedures have on the recovery of different bacterial groups, few studies report DNA extraction efficiencies, making comparisons between studies challenging. In this study, we found that PC2 resulted in an extraction efficiency of 26.7% for NTM, which was three to eight times as high as the extraction efficiencies achieved with the commercial kits tested. While this is significantly lower than the extraction efficiency of approximately 45% reported by Kaevska and Slana (39), the lack of an extraction-independent method to verify the concentrations of cells used to spike samples limits the interpretation of their results. In contrast, we computed our extraction efficiency by comparing the measured DNA yields (see Text S1) to a theoretical DNA yield determined by enumeration of cells by an extraction-independent method. Conversely, recovery of Gram-negative bacteria (*E. coli*) by the PC2 method was 78.6%, considerably higher than values reported in another study, in which the extraction efficiency did not exceed 43.3% (40). Collectively, our results indicate differences in extraction efficiency between commercially available kits and our phenol-chloroform method using a custom lysis buffer with respect to NTM in DW samples. Future studies focused on risk assessment or accurate quantification of NTM should include positive DNA extraction controls to assess method effectiveness. With-

out such controls, results cannot be put into context and artificially low extraction yields will result in underestimation of the levels of NTM in DW (31).

Water age and NTM distribution. The development of *rpoB* primers (see Text S2) combined with a high-throughput PacBio sequencing assay (Fig. 3) allowed us to determine the diversity of NTM in DW samples without the need for cultivation. Furthermore, the two-step PCR and barcoding approach reduced template DNA requirements, which is useful for samples that have low biomass concentrations, such as DW samples. In addition, the barcodes with M13 motifs can be used with other gene targets, thereby removing the need to purchase gene-specific, individually barcoded primers for different assays.

Unlike previous studies employing high-throughput screening methods for NTM in environmental samples (14, 15), this study achieved species, subspecies, and in some cases strain level resolution because of the superior differential power of the *rpoB* gene (Fig. 2). Optimization of the primer set resulted in the amplification of significantly fewer non-NTM sequences (0.06% in this study compared to 68% in the study of van der Wielen et al. [14], which used the *hsp65* gene). Hence, fewer sequences needed to be discarded, making this new methodology more economical and effective. Application of the sequencing approach to DW samples revealed 13 NTM species, 5 of which are considered clinically relevant (10, 12, 13), including *M. abscessus*, *M. avium* subsp. *avium*, *M. chelonae*, *M. intracellulare*, and *M. mucogenicum*, and have been identified by previous culture-based DW studies (7, 8, 41, 42). The striking difference between the NTM present in DW reported in this study, including several clinically relevant NTM species, and the NTM present in Dutch DW (14), consisting mainly of uncultured NTM with unknown clinical relevance, calls for future research. It is possible that this difference is due to the differing disinfection regimes employed: chloramination with an average concentration of 2.04 mg/liter as Cl_2 in this study versus the absence of a residual disinfectant in Dutch DW DS. NTM are resistant to chlorine and chloramines (42); thus, chloramine likely exerted a selective pressure in the DS in this study. Conversely, the diversity of NTM species observed in DW samples in this study was similar to that found in homes in the Netherlands (14). However, the NTM diversity was 50% lower than that found in DS (14) and biofilm samples collected from water mains in St. Paul, MN, which also contained chloramine as the residual disinfectant, albeit at a higher concentration (an average of 3.5 mg/liter as Cl_2) than in this study (15). Such differences in NTM diversity are plausible because of the differing sample types (planktonic biomass versus biofilm), source water quality across geographic locations (Michigan versus the Netherlands versus Minnesota), disinfection practices (residual disinfection with chloramine versus no residual disinfection), and DNA extraction methods used (phenol-chloroform versus the MP Biomedical FastSpin kit for soil).

Previously, an increase in NTM abundance with water age was observed in a simulated DW DS (33); however, differences in NTM species and strain distribution for different water ages have not been described. Such differences in NTM community evenness and membership for different water ages may be due to variable growth rates and susceptibility of NTM to disinfectants and their by-products (43, 44), although more work is needed to identify the specific selective pressures that impact NTM profiles in DW. In addition, it was surprising that the NTM community compositions at relatively similar water ages of 27.5 and 31 h were dramatically different (Fig. 4). This may be due to inaccurate estimates of DS residence times, which were calculated by using the city's hydraulic water distribution model calibrated by using average water demand values across 2010 (winter 2015-2016 water demand values, corresponding to the sampling period, were not available). More accurate DS residence times could be obtained by improving the hydraulic water distribution model and calibration, but that was beyond the scope of this study.

Combining relative abundances of NTM determined by the PacBio sequencing method with NTM qPCR concentrations (see Text S3 and S4) suggests that the concentrations of potentially clinically relevant NTM in DW samples (e.g., $3.53 \times 10^4 \pm$

3.32 × 10⁴ *M. avium* subsp. *paratuberculosis atpE* gene copies/liter) may have been sufficient to result in NTM infections, as previously demonstrated through dose-response models (N50 [the dose at which 50% of the population is expected to be affected], 2.70 × 10² CFU) (45). Since these results were obtained with DNA-based methods that include DNA associated with nonviable NTM, it will be important in future work to validate the results obtained by methods that separate viable NTM (e.g., culture-independent methods that allow separation of viable cells [46]) prior to sequencing efforts to determine the risk of NTM infection posed by DW exposure. Furthermore, before infection mitigation strategies can be developed, future research must focus on determining the virulence of the potentially clinically relevant DW strains. This is of particular importance, as previous work has shown that environmentally isolated opportunistic pathogens can be as virulent as disease-causing, clinically isolated strains (47).

Conclusion. Methods are needed that yield high NTM DNA levels during extraction from low-biomass samples and recover relatively large, phylogenetically informative segments of NTM DNA to advance our understanding of the potential health risks posed by NTM in DW systems. Our use of a DNA extraction method that employs enhanced physical disruption and chemical lysis of NTM cells resulted in considerably greater yields of NTM DNA from DW samples than those obtained with commonly used commercial DNA extraction kits. The development of a two-step barcoding procedure with PacBio sequencing provided a high-throughput method to detect and differentiate NTM present in DW. Applying these methods to DW samples collected from homes receiving municipally treated DW suggests that water age impacts NTM distribution in DW. Most notably, our findings indicate that DW from homes with DS residence times of >24 h had higher concentrations of *M. avium* subsp. *avium* than DW from homes closer to the treatment plant. If supported by future, larger-scale studies, the factors contributing to this intriguing finding can be elucidated. This will, in turn, allow for better estimates of the health risks posed by NTM in DW and will provide a foundation for studies to mitigate this risk.

MATERIALS AND METHODS

Water sample collection and monochloramine analysis. From October 2015 to February 2016, 1-liter cold water samples were collected from kitchen faucets after at least 6 h of stagnation from 15 homes serviced by the Ann Arbor, MI, DW treatment plant. This first 1-liter sample represents the cold water in each home's plumbing; the minimum premise plumbing volume for the homes sampled was 3.5 liters. The DW treatment plant obtains raw water from the Huron River (80 to 85%) and from groundwater wells (15 to 20%) and provides lime softening, coagulation, flocculation, sedimentation, ozonation, filtration, and chloramination (48). The homes were located in two broadly defined areas, (i) close to the treatment plant, receiving water with shorter residence times in the DS (<24 h), and (ii) farther from the plant, receiving water with longer residence times (>24 h).

The temperature, pH, and total and free chlorine concentrations of all samples were measured on site. The average (±SD) water temperature and pH of the 15 samples were 20.4°C ± 1.9°C and 8.6 ± 0.6, respectively. Total and free chlorine concentrations were measured with a DR900 spectrophotometer (Hach, Loveland, CO). The combined chlorine (monochloramine) concentration was estimated by subtracting the free chlorine concentration from the total chlorine concentration. DS residence times for each home sampled were estimated by performing 168 h (7 days) of simulations in EPANET (49) for each home by using the City of Ann Arbor's hydraulic water distribution model and average daily water demands across 2010 for each home. The monochloramine concentration (C) multiplied by water age (i.e., the combined DS residence time and premise plumbing stagnation time [t]), referred to collectively as Ct, was calculated for each sample.

Biomass was filtered from each sample with a 0.2-μm polycarbonate filter (Isopore Membrane Filters; EMD Millipore, Billerica, MA) and stored at −80°C.

DNA extraction methods. Four DNA extraction methods were tested in triplicate with 1-liter samples of DW from a composite sample of 30 liters of cold water collected from the Environmental and Water Resources Engineering Building at the University of Michigan (Ann Arbor). The four methods tested included the commercially available FastSpin kit for soil (MP Biomedicals, Solon, OH) and Maxwell LEV Blood DNA kit (Promega, Madison, WI) and two variations of a standard phenol-chloroform method (PC1 and PC2) with a modified version of the universal nucleic acid extraction buffer (35) (Table 1). The two phenol-chloroform-based methods (PC1 and PC2) differed in surfactant concentration. Details of the procedure are provided in Text S5. DNA extractions with the FastSpin and Maxwell kits were performed as outlined by Haig et al. (50) and Webster et al. (51), respectively. The performance of the different extraction procedures was evaluated by qPCRs targeting NTM, *Pseudomonas*, and total bacteria as

described below. Subsequently, DNA was extracted from 15 DW samples collected from homes and one mock community by PC2, the phenol-chloroform-based extraction method with 0.09% SDS and 5 min of bead beating.

Extraction efficiency. The extraction efficiency of method PC2 was determined by spiking 1 liter of PBS with *E. coli* ATCC 15597 at $1.6 \times 10^7 \pm 7.1 \times 10^6$ CFU/liter or *M. abscessus* ATCC 19977 at $1.4 \times 10^7 \pm 1.1 \times 10^7$ CFU/liter. The cultures were grown in accordance with ATCC standard procedures. Resulting DNA yields were measured with the Qubit Fluorometer (Life Technologies, Inc., Waltham, MA). Final extraction efficiencies were calculated by comparing the measured results to the extraction-independent total cell counts determined by staining with 4′,6-diamidino-2-phenylindole (DAPI; Thermo, Fisher Scientific, Waltham, MA) and visualization with a Zeiss Axio Observer D1 fluorescence microscope (Zeiss, Peabody, MA). Theoretical DNA yields from *E. coli* and *M. abscessus* cells were calculated as outlined by Dolezel et al. (52) (see Text S3).

qPCR. qPCR assays targeting genes for NTM, *Pseudomonas*, and total bacteria were performed by previously published methods (34, 53, 54) as detailed in Text S6. All DNA samples were processed in triplicate with negative controls and standards. Assays were conducted with 96-well polypropylene plates on a CFX96 real-time quantitative thermocycler (Bio-Rad, Hercules, CA). Each 10-μl reaction mixture contained 5 μl of Fast EvaGreen qPCR master mix (Biotium, Fremont, CA), 0.8 μl of each primer (0.4 μM; IDT, Coralville, IA), 2.4 μl of water, and 1 μl of template DNA (~0.2 ng $\cdot$ μl^{-1}). PCR conditions for NTM and total 16S rRNA gene assays consisted of 40 cycles, whereas the *Pseudomonas* assay was performed with 35 cycles. Melting curve analysis of the PCR products was conducted following each assay to confirm that the fluorescence signal originated from specific PCR products and not from primer dimers or other artifacts. For all qPCR assays, a linear relationship between the log of the standards' DNA copy number and the calculated threshold cycle value across the specified concentration range was confirmed (R^2 value of >0.99 in all cases). Amplification efficiencies, calculated by the method described by Pfaffl (55), varied from 1.8 to 2.0 across all assays, and these values are consistent with those reported in other studies (54, 56).

***rpoB* primer design.** The 41 currently available full-length NTM *rpoB* DNA sequences in the GenBank database (57) were aligned with Kalign (58) and included several strains of *M. avium* subsp. *avium*, *Mycobacterium bolletii*, *M. chelonae*, *Mycobacterium marinum*, and *M. mucogenicum*. The primers developed by Macheras et al. (59) were modified on the basis of this alignment, resulting in amplicons of 942 to 957 bp (see Text S2). Subsequent *in-silico* analysis suggested that certain versions of the degenerate primers would hybridize with non-NTM species. Therefore, equimolar concentrations of individual, nondegenerate versions of the primers that primarily amplified NTM were combined to improve primer specificity. The nondegenerate forward and reverse primer stocks comprised six forward and four reverse primers, respectively (see Text S2).

MrBayes version 3.2 (60) was used to create a Bayesian inference phylogenetic tree with alignment performed with ClustalW. Specifically, the 942- to 957-bp amplicons were fitted to a general time-reversible gamma-distributed rate variation model by using a Markov chain Monte Carlo analysis performed for 10^6 generations with sampling every 10^3 generations. The average SD of split frequencies was <0.007 after 10^6 generations, indicating convergence. Posterior probabilities were averaged over the final 75% of trees (25% burn-in). A phylogenetic tree was plotted with FigTree v1.4.2.

PacBio RS II sequencing. Identification of NTM to the species level was performed with 15 DW samples collected in Ann Arbor, MI, and a control sample consisting of 95% *M. avium* subsp. *avium* (DJO-44271) and 5% *M. chelonae* (19237) by a two-step PCR procedure similar to that described in reference 61 (Fig. 3). The control sample was generated by combining the aforementioned two NTM species at the ratio stated after individually amplifying pure cultures of them. In the first PCR, the NTM *rpoB* genes were amplified with the newly designed *rpoB* primers (see Text S2) with either M13F or M13R motifs amended to the 5′ ends of the primers (Fig. 3). In the second PCR, products from the first PCR were amplified with primers that consisted of M13F or M13R motifs at the 3′ end and unique 16-bp PacBio barcodes at the 5′ end (see Text S2). The resulting amplicons were composed of asymmetrical barcodes (Fig. 3). Each 10-μl reaction mixture contained 1$\times$ Phusion Flash High-Fidelity PCR master mix (Thermo, Fisher Scientific, Waltham, MA), 0.4 μM each primer (IDT, Coralville, IA), 0.625 mg/ml bovine serum albumin (Life Technologies, Inc., Waltham, MA), 0.5 M betaine, 5% dimethyl sulfoxide (Sigma-Aldrich, St. Louis, MO), and 1 μl of template. The thermal cycling conditions used for each PCR are shown in Text S2. To reduce chimera formation and PCR biases, cycle numbers were kept as low as possible. The second PCR was performed 9 to 18 separate times with products purified with the Qiagen Gel Cleanup kit (Qiagen, Germantown, MD) and precipitated with PEG 6000 (50), and the resulting yields were measured with the Qubit Fluorometer (Life Technologies, Inc., Waltham, MA). Equimolar concentrations of each sample (15 ng/μl) were pooled. Amplicons were processed without shearing, and SMRTbell libraries were constructed in accordance with the PacBio 1-kb Template Preparation protocol for CCS, allowing a single insert to be sequenced multiple times (29). The polymerase-template complexes were immobilized at the bottom of the zero-mode waveguides with megabead loading and sequenced on a PacBio RS II system with C2 sequencing chemistry and one SMRT cell (University of Michigan, Ann Arbor, MI). A 1-kb collection protocol was utilized with a 180-min movie window for signal detection.

PacBio data processing and analysis. PacBio data processing was performed by the Bioinformatics Core at the University of Michigan and consisted of demultiplexing CCS by the RS_ReadsOfInsert protocol on the PacBio SMRTanalysis Portal with the following parameters: a full-pass number of >10, a predictive accuracy of >90%, and the most flexible barcode stringency.

Phylogeny of OTUs in the noncontrol samples was established on the basis of parameters obtained by testing a UPARSE analysis workflow (62) with the control sample. The resulting workflow consisted of

(i) employing CUTADAPT (63) with forward and reverse M13 sequences to delineate full amplicon sequences, (ii) using USEARCH (64) to discard reads with >1.0 total expected errors, (iii) removing equipment- and reagent-derived DNA (background contamination) from dereplicated reads by using the neutral-model-based approach developed by Bassis et al. (65), (iv) clustering into OTUs with a minimum size of two reads, and (v) employing the UTAX algorithm to assign lineage information to those OTUs on the basis of a custom-made *rpoB* database. For OTUs that could not be assigned to a species/strain or were assigned with a low score (confidence value of <0.8), we employed BLAST (66) to identify the closest relatives.

The custom-made *rpoB* database was generated by downloading all 70,507 currently available bacterial *rpoB* sequences from GenBank (57), identifying the relevant amplicon sequences with USEARCH and the primer pair sequences (allowing an edit distance of 4 for each primer), and retrieving lineage information from the NCBI taxonomy database (67). Five yeast RPB2 sequences were used as decoys to train the UTAX algorithm on the bacterial data set. The resulting database consisted of 11,572 *rpoB* sequences. By using the aforementioned workflow, classifications were assigned to OTUs from the experimental samples by using global alignments with USEARCH and an identity threshold of >99.8%.

Statistical analyses. Significant differences between the total DNA yields and *Mycobacterium*, *Pseudomonas*, and total bacterial gene copy numbers obtained by different extraction methods were identified by Wilcoxon tests. Significant differences in the NTM community structure were determined by nonparametric multiple analysis of variance. All statistical analyses were performed with the R statistical software (68), with statistical significance determined by P values of <0.05.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio .02354-17.

TEXT S1, PDF file, 0.04 MB.
TEXT S2, PDF file, 0.1 MB.
TEXT S3, PDF file, 0.05 MB.
TEXT S4, PDF file, 0.7 MB.
TEXT S5, PDF file, 0.1 MB.
TEXT S6, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Morimoto K, Iwai K, Uchimura K, Okumura M, Yoshiyama T, Yoshimori K, Ogata H, Kurashima A, Gemma A, Kudoh S. 2014. A steady increase in nontuberculous mycobacteriosis mortality and estimated prevalence in Japan. Ann Am Thorac Soc 11:1–8. https://doi.org/10.1513/AnnalsATS .201303-067OC.

2. Bar-On O, Mussaffi H, Mei-Zahav M, Prais D, Steuer G, Stafler P, Hananya S, Blau H. 2015. Increasing nontuberculous mycobacteria infection in cystic fibrosis. J Cyst Fibros 14:53–62. https://doi.org/10.1016/j.jcf.2014 .05.008.

3. Qvist T, Gilljam M, Jönsson B, Taylor-Robinson D, Jensen-Fangel S, Wang M, Svahn A, Kötz K, Hansson L, Hollsing A, Hansen CR, Finstad PL, Pressler T, Høiby N, Katzenstein TL, Scandinavian Cystic Fibrosis Study Consortium (SCFSC). 2015. Epidemiology of nontuberculous mycobacteria among patients with cystic fibrosis in Scandinavia. J Cyst Fibros 14:46–52. https://doi.org/10.1016/j.jcf.2014.08.002.

4. Ringshausen FC, Wagner D, de Roux A, Diel R, Hohmann D, Hickstein L,

Welte T, Rademacher J. 2016. Prevalence of nontuberculous mycobacterial pulmonary disease Germany 2009–2014. Emerg Infect Dis 22: 1102–1105. https://doi.org/10.3201/eid2206.151642.

5. Carson LA, Bland LA, Cusick LB, Favero MS, Bolan GA, Reingold AL, Good RC. 1988. Prevalence of nontuberculous mycobacteria in water supplies of hemodialysis centers. Appl Environ Microbiol 54:3122–3125.

6. Nishiuchi Y, Maekura R, Kitada S, Tamaru A, Taguri T, Kira Y, Hiraga T, Hirotani A, Yoshimura K, Miki M, Ito M. 2007. The recovery of *Mycobacterium avium-intracellulare* complex (MAC) from the residential bathrooms of patients with pulmonary MAC. Clin Infect Dis 45:347–351. https://doi.org/10.1086/519383.

7. Falkinham JO, III, Iseman MD, de Haas P, van Soolingen D. 2008. *Mycobacterium avium* in a shower linked to pulmonary disease. J Water Health 6:209–213. https://doi.org/10.2166/wh.2008.032.

8. Pierce ES. 2009. Possible transmission of *Mycobacterium avium* subspecies *paratuberculosis* through potable water: lessons from an urban

cluster of Crohn's disease. Gut Pathog 1:17. https://doi.org/10.1186/1757 -4749-1-17.

9. Falkinham JO, III. 2011. Nontuberculous mycobacteria from household plumbing of patients with nontuberculous mycobacteria disease. Emerg Infect Dis 17:419–424. https://doi.org/10.3201/eid1703.101510.

10. Thomson R, Tolson C, Carter R, Coulter C, Huygens F, Hargreaves M. 2013. Isolation of nontuberculous mycobacteria (NTM) from household water and shower aerosols in patients with pulmonary disease caused by NTM. J Clin Microbiol 51:3006–3011. https://doi.org/10.1128/JCM .00899-13.

11. Magee JG, Ward AC. 2015. Mycobacterium, p 1–84. In Whitman WB (ed), Bergey's manual of systematics of archaea and bacteria. Wiley, New York, NY.

12. Lipuma JJ. 2010. The changing microbial epidemiology in cystic fibrosis. Clin Microbiol Rev 23:299–323. https://doi.org/10.1128/CMR.00068-09.

13. Halstrom S, Price P, Thomson R. 2015. Review. Environmental mycobacteria as a cause of human infection. Int J Mycobacteriol 4:81–91.

14. van der Wielen PW, Heijnen L, van der Kooij D. 2013. Pyrosequence analysis of the hsp65 genes of nontuberculous Mycobacterium communities in unchlorinated drinking water in the Netherlands. Appl Environ Microbiol 79:6160–6166. https://doi.org/10.1128/AEM.01591-13.

15. Gomez-Smith CK, LaPara TM, Hozalski RM. 2015. Sulfate reducing bacteria and mycobacteria dominate the biofilm communities in a chloraminated drinking water distribution system. Environ Sci Technol 49: 8432–8440. https://doi.org/10.1021/acs.est.5b00555.

16. de los Reyes FL, Ritter W, Raskin L. 1997. Group-specific small-subunit rRNA hybridization probes to characterize filamentous foaming in activated sludge systems. Appl Environ Microbiol 63:1107–1117.

17. Steinert M, Birkness K, White E, Fields B, Quinn F. 1998. Mycobacterium avium bacilli grow saprozoically in coculture with Acanthamoeba polyphaga and survive within cyst walls. Appl Environ Microbiol 64: 2256–2261.

18. Berry D, Horn M, Xi C, Raskin L. 2010. Mycobacterium avium infections of Acanthamoeba strains: host strain variability grazing-acquired infections and altered dynamics of inactivation with monochloramine. Appl Environ Microbiol 76:6685–6688. https://doi.org/10.1128/AEM.00644-10.

19. Kennedy GM, Morisaki JH, Champion PAD. 2012. Conserved mechanisms of Mycobacterium marinum pathogenesis within the environmental amoeba Acanthamoeba castellanii. Appl Environ Microbiol 78: 2049–2052. https://doi.org/10.1128/AEM.06965-11.

20. Tortoli E. 2012. Phylogeny of the genus Mycobacterium: many doubts few certainties. Infect Genet Evol 12:827–831. https://doi.org/10.1016/j .meegid.2011.05.025.

21. de Zwaan R, van Ingen J, van Soolingen D. 2014. Utility of rpoB gene sequencing for identification of nontuberculous mycobacteria in the Netherlands. J Clin Microbiol 52:2544–2551. https://doi.org/10.1128/ JCM.00233-14.

22. Lee H, Park HJ, Cho SN, Bai GH, Kim SJ. 2000. Species identification of mycobacteria by PCR-restriction fragment length polymorphism of the rpoB gene. J Clin Microbiol 38:2966–2971.

23. Adékambi T, Drancourt M, Raoult D. 2009. The rpoB gene as a tool for clinical microbiologists. Trends Microbiol 17:37–45. https://doi.org/10 .1016/j.tim.2008.09.008.

24. Moghim S, Sarikhani E, Nasr Esfahani BN, Faghri J. 2012. Identification of nontuberculous mycobacteria species isolated from water samples using phenotypic and molecular methods and determination of their antibiotic resistance patterns by E-test method in Isfahan Iran. Iran J Basic Med Sci 15:1076–1082.

25. Wayne LG, Gross WM. 1968. Base composition of deoxyribonucleic acid isolated from mycobacteria. J Bacteriol 96:1915–1919.

26. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341. https://doi.org/10.1186/1471 -2164-13-341.

27. Guo X, Lehner K, O'Connell K, Zhang J, Dave SS, Jinks-Robertson S. 2015. SMRT sequencing for parallel analysis of multiple targets and accurate SNP phasing. G3 (Bethesda) 5:2801–2808. https://doi.org/10.1534/g3 .115.023317.

28. Betz-Stablein BD, Töpfer A, Littlejohn M, Yuen L, Colledge D, Sozzi V, Angus P, Thompson A, Revill P, Beerenwinkel N, Warner N, Luciani F. 2016. Single-molecule sequencing reveals complex genome variation of hepatitis B virus during 15 years of chronic infection following liver transplantation. J Virol 90:7171–7183. https://doi.org/10.1128/JVI .00243-16.

29. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 38:e159. https://doi.org/10.1093/nar/ gkq543.

30. Adékambi T, Shinnick TM, Raoult D, Drancourt M. 2008. Complete rpoB gene sequencing as a suitable supplement to DNA-DNA hybridization for bacterial species and genus delineation. Int J Syst Evol Microbiol 58:1807–1814. https://doi.org/10.1099/ijs.0.65440-0.

31. Hwang C, Ling F, Andersen GL, LeChevallier MW, Liu WT. 2012. Evaluation of methods for the extraction of DNA from drinking water distribution system biofilms. Microbes Environ 27:9–18. https://doi.org/10 .1264/jsme2.ME11132.

32. Knudsen BE, Bergmark L, Munk P, Lukjancenko O, Priemé A, Aarestrup FM, Pamp SJ. 2016. Impact of sample type and DNA isolation procedure on genomic inference of microbiome composition. mSystems 1:e00095 -16. https://doi.org/10.1128/mSystems.00095-16.

33. Wang H, Edwards M, Falkinham JO, III, Pruden A. 2012. Molecular survey of the occurrence of Legionella spp., Mycobacterium spp., Pseudomonas aeruginosa and amoeba hosts in two chloraminated drinking water distribution systems. Appl Environ Microbiol 78:6285–6294. https://doi .org/10.1128/AEM.01492-12.

34. Radomski N, Roguet A, Lucas FS, Veyrier FJ, Cambau E, Accrombessi H, Moilleron R, Behr MA, Moulin L. 2013. atpE gene as a new useful specific molecular target to quantify Mycobacterium in environmental samples. BMC Microbiol 13:277. https://doi.org/10.1186/1471-2180-13-277.

35. Hill VR, Narayanan J, Gallen RR, Ferdinand KL, Cromeans T, Vinjé J. 2015. Development of a nucleic acid extraction procedure for simultaneous recovery of DNA and RNA from diverse microbes in water. Pathogens 4:335–354. https://doi.org/10.3390/pathogens4020335.

36. Lever MA, Torti A, Eickenbusch P, Michaud AB, Šantl-Temkiv T, Jørgensen BB. 2015. A modular method for the extraction of DNA and RNA and the separation of DNA pools from diverse environmental sample types. Front Microbiol 6:476. https://doi.org/10.3389/fmicb.2015.00476.

37. Hett EC, Rubin EJ. 2008. Bacterial growth and cell division: a mycobacterial perspective. Microbiol Mol Biol Rev 72:126–156. https://doi.org/10 .1128/MMBR.00028-07.

38. Bhairi SM. 2001. Detergents: a guide to the properties and uses of detergents in biological systems. Calbiochem-Novabiochem Corporation, San Diego, CA. http://sevierlab.vet.cornell.edu/resources/CALBIOCHEM -DetergentsIV.pdf.

39. Kaevska M, Slana I. 2015. Comparison of filtering methods, filter processing and DNA extraction kits for detection of mycobacteria in water. Ann Agric Environ Med 22:429–432. https://doi.org/10.5604/12321966 .1167707.

40. Mumy KL, Findlay RH. 2004. Convenient determination of DNA extraction efficiency using an external DNA recovery standard and quantitative-competitive PCR. J Microbiol Methods 57:259–268. https:// doi.org/10.1016/j.mimet.2004.01.013.

41. September SM, Brözel VS, Venter SN. 2004. Diversity of nontuberculoid Mycobacterium species in biofilms of urban and semiurban drinking water distribution systems. Appl Environ Microbiol 70:7571–7573. https://doi.org/10.1128/AEM.70.12.7571-7573.2004.

42. Donohue MJ, Mistry JH, Donohue JM, O'Connell K, King D, Byran J, Covert T, Pfaller S. 2015. Increased frequency of nontuberculous mycobacteria detection at potable water taps within the United States. Environ Sci Technol 49:6127–6133. https://doi.org/10.1021/acs.est.5b00496.

43. Le Dantec C, Duguet JP, Montiel A, Dumoutier N, Dubrou S, Vincent V. 2002. Chlorine disinfection of atypical mycobacteria isolated from a water distribution system. Appl Environ Microbiol 68:1025–1032. https:// doi.org/10.1128/AEM.68.3.1025-1032.2002.

44. Pryor M, Springthorpe S, Riffard S, Brooks T, Huo Y, Davis G, Sattar SA. 2004. Investigation of opportunistic pathogens in municipal drinking water under different supply and treatment regimes. Water Sci Technol 50:83–90.

45. Breuninger KJ, Weir MH. 2015. Development of an interspecies nested dose-response model for Mycobacterium avium subspecies paratuberculosis. Risk Anal 35:1479–1487. https://doi.org/10.1111/risa.12380.

46. Hammes F, Berney M, Wang Y, Vital M, Köster O, Egli T. 2008. Flowcytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. Water Res 42: 269–277. https://doi.org/10.1016/j.watres.2007.07.009.

47. Struve C, Krogfelt KA. 2004. Pathogenic potential of environmental

Klebsiella pneumoniae isolates. Environ Microbiol 6:584–590. https://doi.org/10.1111/j.1462-2920.2004.00590.x.

48. Pinto AJ, Xi C, Raskin L. 2012. Bacterial community structure in the drinking water microbiome is governed by filtration processes. Environ Sci Technol 46:8851–8859. https://doi.org/10.1021/es302042t.

49. Rossman LA. 2000. EPANET 2 users manual. U.S. Environmental Protection Agency, Cincinnati, OH. http://www.innovyze.com/products/epanet/download/P1007WWU.pdf.

50. Haig SJ, Schirmer M, D'amore R, Gibbs J, Davies RL, Collins G, Quince C. 2015. Stable-isotope probing and metagenomics reveal predation by protozoa drives *E. coli* removal in slow sand filters. ISME J 9:797–808. https://doi.org/10.1038/ismej.2014.175.

51. Webster TM, Smith AL, Reddy RR, Pinto AJ, Hayes KF, Raskin L. 2016. Anaerobic microbial community response to methanogenic inhibitors 2-bromoethanesulfonate and propynoic acid. Microbiologyopen 5:537–550. https://doi.org/10.1002/mbo3.349.

52. Dolezel J, Bartos J, Voglmayr H, Greilhuber J. 2003. Nuclear DNA content and genome size of trout and human. Cytometry A 51:127–129. https://doi.org/10.1002/cyto.a.10013.

53. Spilker T, Coenye T, Vandamme P, LiPuma JJ. 2004. PCR-based assay for differentiation of *Pseudomonas aeruginosa* from other *Pseudomonas* species recovered from cystic fibrosis patients. J Clin Microbiol 42:2074–2079. https://doi.org/10.1128/JCM.42.5.2074-2079.2004.

54. Haig SJ, Quince C, Davies RL, Dorea CC, Collins G. 2014. Replicating the microbial community and water quality performance of full-scale slow sand filters in laboratory-scale filters. Water Res 61:141–151. https://doi.org/10.1016/j.watres.2014.05.008.

55. Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45–e45. https://doi.org/10.1093/nar/29.9.e45.

56. Fierer N, Jackson JA, Vilgalys R, Jackson RB. 2005. Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. Appl Environ Microbiol 71:4117–4120. https://doi.org/10.1128/AEM.71.7.4117-4120.2005.

57. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. Nucleic Acids Res 44:D67–D72. https://doi.org/10.1093/nar/gkv1276.

58. Lassmann T, Sonnhammer EL. 2005. Kalign—an accurate and fast mul-tiple sequence alignment algorithm. BMC Bioinformatics 6:298. https://doi.org/10.1186/1471-2105-6-298.

59. Macheras E, Roux AL, Bastian S, Leão SC, Palaci M, Sivadon-Tardy V, Gutierrez C, Richter E, Rüsch-Gerdes S, Pfyffer G, Bodmer T, Cambau E, Gaillard JL, Heym B. 2011. Multilocus sequence analysis and *rpoB* sequencing of *Mycobacterium abscessus* (*sensu lato*) strains. J Clin Microbiol 49:491–499. https://doi.org/10.1128/JCM.01274-10.

60. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029.

61. Berry D, Mahfoudh KB, Wagner M, Loy A. 2011. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. Appl Environ Microbiol 77:7846–7849. https://doi.org/10.1128/AEM.05220-11.

62. Edgar RC. 2013. Uparse: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996–998. https://doi.org/10.1038/nmeth.2604.

63. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.

64. Edgar RC, Flyvbjerg H. 2015. Error filtering pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31:3476–3482. https://doi.org/10.1093/bioinformatics/btv401.

65. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, Beck JM, Curtis JL, Huffnagle GB. 2015. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. mBio 6:e00037-15. https://doi.org/10.1128/mBio.00037-15.

66. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. Nucleic Acids Res 36:W5–W9. https://doi.org/10.1093/nar/gkn201.

67. NCBI Resource Coordinators. 2017. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 45(D1):D12–D17. https://doi.org/10.1093/nar/gkw1071.

68. R Development Core Team. 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://statlab.uni-heidelberg.de/projects/r/refman.pdf.