

# Phylogeographic Analysis of *Mycobacterium kansasii* Isolates from Patients with *M. kansasii* Lung Disease in Industrialized City, Taiwan

Patrick George Tobias Cudahy, Po-Chen Liu, Joshua L. Warren, Benjamin Sobkowiak, Chongguang Yang, Thomas R. Ioerger, Chieh-Yin Wu, Po-Liang Lu, Jann-Yuan Wang, Hsiao-Han Chang, Hung-Ling Huang, Ted Cohen, Hsien-Ho Lin

Little is known about environmental transmission of *Mycobacterium kansasii*. We retrospectively investigated potential environmental acquisition, primarily water sources, of *M. kansasii* among 216 patients with pulmonary disease from an industrial city in Taiwan during 2015–2017. We analyzed sputum mycobacterial cultures using whole-genome sequencing and used hierarchical Bayesian spatial network methods to evaluate risk factors for genetic relatedness of *M. kansasii* strains. The mean age of participants was 67 years; 24.1% had previously had tuberculosis. We found that persons from districts served by 2 water purification plants were at higher risk of being infected with genetically related *M. kansasii* isolates. The adjusted odds ratios were 1.81 (1.25–2.60) for the Weng Park plant and 1.39 (1.12–1.71) for the Fongshan plant. Those findings unveiled the association between water purification plants and *M. kansasii* pulmonary disease, highlighting the need for further environmental investigations to evaluate the risk for *M. kansasii* transmission.

*Mycobacterium kansasii*, a slow-growing nontuberculous mycobacteria (NTM), can cause destructive pulmonary diseases in humans that cause similar clinical manifestations to those of *M. tuberculosis* pulmonary disease (1). In recent years, *M. kansasii* has become one of the most frequently reported NTMs in the world (2), but whether increasing notifications in

some countries (3) reflects true increases in the incidence of *M. kansasii* disease or improvements in laboratory identification is not well understood.

Infection with *M. kansasii* has been predominantly reported in urban settings, in high-density and low-income communities, and among gold miners (4). Despite recent reports of potential human-to-human transmission for other NTMs (5), *M. kansasii* lung disease has been generally assumed to be acquired from environmental sources (6). Nonetheless, the precise route of transmission is yet to be characterized. *M. kansasii* is ubiquitous in the environment; tap water is reported as a major reservoir (6). The waterborne acquisition of *M. kansasii* is enabled by its intrinsic resistance to disinfectants, acid, and heat (7); its ability to survive in oligotrophic water; and its ability to form pipe surface biofilms (8). In previous studies, *M. kansasii* has been isolated from the water distribution system in the same communities in which cases with *M. kansasii* disease arise (9). However, evidence on environmental acquisition on the basis of genotyping of clinical or environmental isolates has been limited.

We investigated the environmental acquisition of *M. kansasii* in the industrial city of Kaohsiung, located in southern Taiwan. Our earlier multicenter study revealed that the number of patients with *M. kansasii*

Author affiliations: Yale School of Medicine, New Haven, Connecticut, USA (P.G.T. Cudahy); National Taiwan University, Taipei, Taiwan (P.-C. Liu, C.-Y. Wu, J.-Y. Wang, H.-H. Lin); Yale School of Public Health, New Haven (J.L. Warren, B. Sobkowiak, T. Cohen); Sun Yat-Sen University School of Public Health, Guangzhou, China (C. Yang); Texas A&M University, College Station, Texas, USA (T.R. Ioerger); Kaohsiung Medical

University, Kaohsiung, Taiwan (P.-L. Lu, H.-L. Huang); Kaohsiung Medical University Hospital, Kaohsiung (P.-L. Lu, H.-L. Huang); National Tsing Hua University, Hsinchu, Taiwan (H.-H. Chang); Kaohsiung Municipal Ta-Tung Hospital, Kaohsiung, (H.-L. Huang)

DOI: <https://doi.org/10.3201/eid3008.240021>

pulmonary infection was nearly 5-fold higher in Kaohsiung than in Taipei in northern Taiwan (10). Higher humidity levels, warmer temperatures, and industrial areas with greater air pollution in Kaohsiung might explain this difference. We previously identified 2 spatial hotspots of high risk for *M. kansasii* infection in Kaohsiung by using clinical data collected from patients in all tertiary medical centers of the city (11). We hypothesized that specific water supplies or heavy industrial areas might be associated with the risk for *M. kansasii* infection. In this study, we extended the previous spatial analysis to include whole-genome sequencing (WGS) data of clinical *M. kansasii* isolates along with geographic information from patients, area-level industrialization, and water supply systems to comprehensively assess the risk and drivers of *M. kansasii* in Kaohsiung.

## Methods

### Study Participants

We included 302 patients >20 years of age who had newly diagnosed *M. kansasii* lung disease, consistent with American Thoracic Society/Infectious Diseases Society of America guidelines (12) from 1 tertiary medical center and its affiliated regional hospitals in Kaohsiung during 2015–2017. We conducted chart reviews to collect clinical and demographic data from eligible persons.

### Mycobacterial WGS

Of the 302 patients, we performed WGS on mycobacterial isolates from 243 patients with high-quality culture samples from pretreatment sputum or bronchoalveolar lavage fluid. We performed sequencing in the Laboratory of Genomics and Bioinformatics Service at Texas A&M AgriLife (College Station, TX, USA) and constructed libraries by using the NEXT-FLEX Rapid XP DNA-Seq Kit (Revvity, <https://www.revvity.com>). We performed paired-end sequencing using the NovaSeq 6000 Sequencing System (Illumina, <https://www.illumina.com>) with a read length of 150 bp and anticipated minimum mean depth of coverage of 100×. We removed 22 samples with most reads from a species other than *M. kansasii* from analysis. One additional sample had a large degree of nonmycobacterial reads, and 4 samples were from participants missing clinical or residential data. We excluded those from analysis, leaving 216 samples for our primary analysis.

### Environmental Exposure Data

Our primary environmental exposure of interest was the water sources. We collected information on the

service areas of water purification plants in Kaohsiung through the official website of Taiwan Water Corporation (13). In total, 5 water purification plants provide service to most households in Kaohsiung. Among those, Pingding is the largest by volume (44% of the total water supply from the 5 plants), followed by Chengcing (29%), Kaotan (12%), Fongshan (10%), and Weng Park (4%) (13). The water supply networks of those purification plants overlap; a single district often receives service from >1 water plant (Figure 1). We identified areas of heavy industrial zoning in Kaohsiung and used a probabilistic approach to identify participants who had a high probability of working in 1 of those zones (Appendix, <https://wwwnc.cdc.gov/EID/article/30/8/24-0021-App1.pdf>).

### Data Analysis

We classified and filtered raw fastq reads using kraken2 (14) with a custom database, bracken (15), and ntm-profiler (16) (Appendix). We used Shovill (17) for de novo assembly using SPAdes (18). We aligned assemblies using skat (S.R. Harris, unpub. data, <https://www.biorxiv.org/content/10.1101/453142v1>) to the *M. kansasii* strain ATCC 12478, and all had >90% genomic coverage of the reference. We identified and masked regions of horizontal sequence transfer using Gubbins (19) and constructed a recombination-masked pairwise single-nucleotide polymorphism (SNP) distance matrix with snp-dists (20). We used this structure to create a phylogenetic tree with RAXML-NG (21) using the *M. kansasii* strain FDA-ARGOS 1615 as an outgroup, a generalized time reversible with gamma model of rate heterogeneity, and 50 starting trees (25 random and 25 parsimony) and calculated the transfer bootstrap expectation metric. To assess the association between genetic clade and water purification plant, we used Pearson  $\chi^2$  tests.

To assess the sensitivity of our phylogenetic analysis, we repeated the analysis with the addition of genomic sequences obtained from public sources and identified as obtained from outside of Taiwan. We used a list of available isolates (9) to download short-read sequences of 22 specimens from the National Center for Biotechnology Information Sequence Read Archive database. We speciated and filtered the sequences with the same analysis pipeline as our samples from Taiwan and repeated phylogenetic analyses using a tree constructed from the combined sample sets.

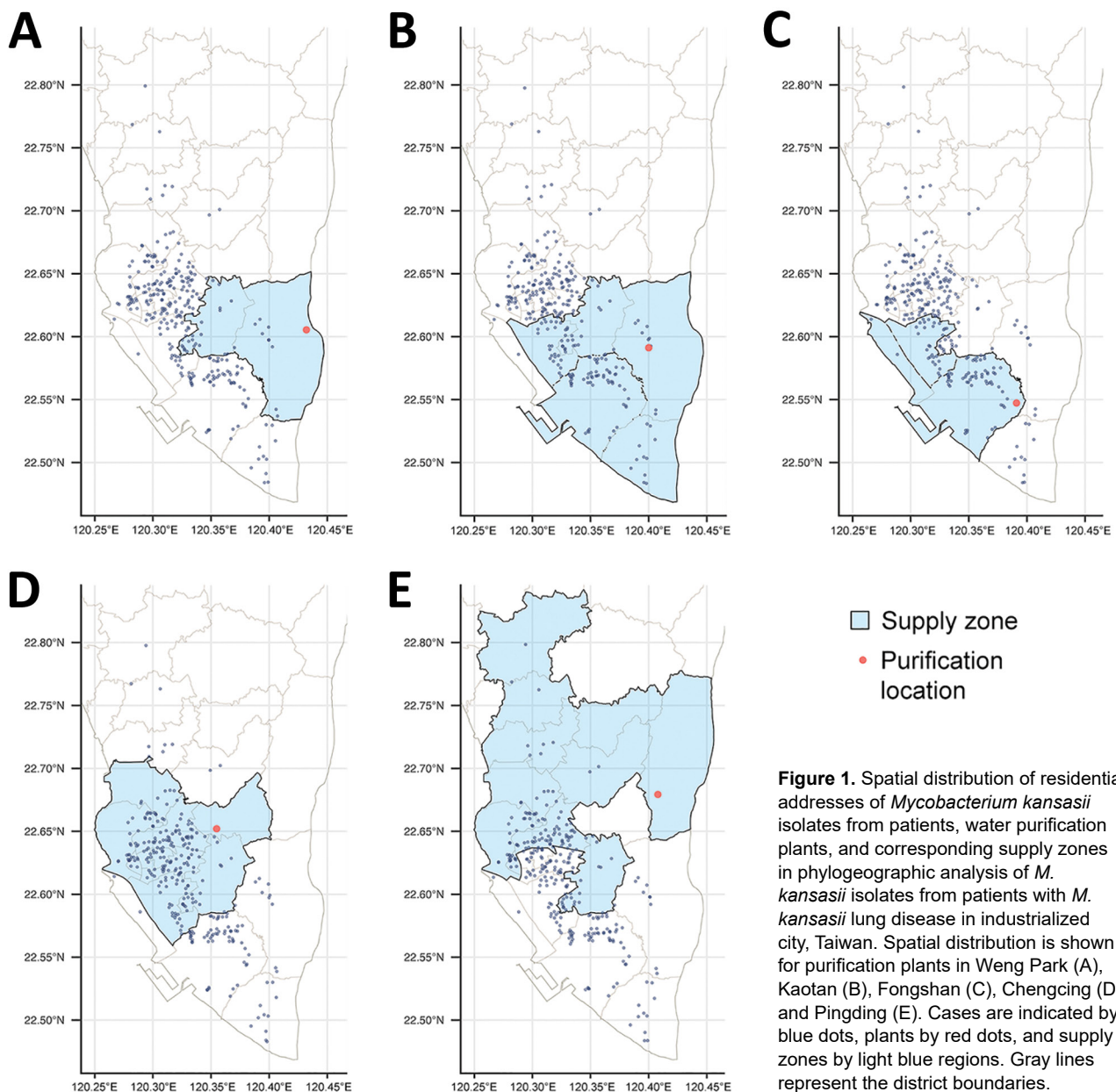
To analyze factors affecting the genetic relatedness of *M. kansasii* strains between pairs of persons from the study sample, we used hierarchical Bayesian spatial network methods implemented in the R package

GenePair (22). As opposed to standard regression modeling, GenePair methods are designed to account for correlation observed in dyadic data given that the same person is represented across multiple pairs (i.e., network dependence), spatial correlation between paired outcomes because of unmeasured transmission dynamics, and specific distributional features of genetic relatedness outcomes (23).

We modeled 2 outcomes of genetic relatedness: a binary outcome in which we classified pairs of strains as clustered or not based on a threshold SNP distance and a continuous measure of SNP distance between pairs of strains. For the binary outcome, the bimod-

al distribution of observed SNP distances between strains (Appendix Figure 1) suggested a threshold of  $\leq 45$  SNPs to classify pairs of strains as being within a cluster. In a sensitivity analysis, we also investigated a more conservative cutoff of  $\leq 32$  SNPs. We fit all models using Markov chain Monte Carlo sampling techniques and assessed convergence using trace plots and Geweke's diagnostic for all relevant model parameters. When making statistical inference, we reported posterior means and 95% equal-tailed quantile-based credible intervals.

Independent variables included the time between the date of each participant's sputum collection in



days, the combined age of both participants in years, the age difference between pairs of participants in years, whether both were male or female, whether either person in the pair had cavitory disease on chest radiograph, geographic distance between participant households in kilometers, whether participant households were both supplied by the same water purification plants, whether both participants resided in villages with a high degree of linkage to the 4 heavy industrial zones, and whether the patients had received their diagnosis and been treated in the same hospital. We chose demographic and clinical variables included in the models a priori on the basis of clinical relevance.

The study was approved by the institutional ethics committees of the participating hospitals (KMUHIRB-F(I)-20210173, KMUHIRB-E(I)-20210380). The study funders had no role in the design and conduct of the study, the analysis and interpretation of data, or in the preparation, review, or approval of the manuscript.

## Results

### Demographics

Of the 216 cases with complete information, the mean patient age was 67 years (SD 17.6); 146 (67.6%) were men and 70 (32.4%) women (Table). Mean body mass index was 21 (SD 4.1). Of the 216 case-patients, 46 (21.3%) were current smokers, 45 (20.8%) were ex-smokers, and 125 (57.9%) had never smoked. On chest radiograph, 42 (20.1%) persons had consolidation, 37 (17.1%) had fibrocavitory disease, 81 (38.8%) had nodular bronchiectasis, 3 (1.4%) had nodules, 1 (0.5%) had fibrosis, and 52 (24.1%) had mixed image patterns. Tuberculosis had been diagnosed previously in 52 (24.1%) persons.

### Sources of Residential Water Purification

Participant households were served by a median of 2 (maximum 3) of 8 different water purification plants. Chengcinghu supplied 128 (59.3%) of participant households, Pingding supplied 84 (38.9%), Kaotan supplied 99 (45.8%), Fongshan supplied 70 (32.4%), Weng Park supplied 22 (10.2%), Lingkou supplied 4 (1.9%), and Baolai and Lujhu each supplied 1 (0.5%) (Table; Figure 1).

### Factors Associated with Genetic Relatedness between Pairs of Participants

Before analyzing genetic relatedness of *M. kansasii* strains, we masked areas of horizontal sequence transfer as previously described. This process reduced the median number of SNPs per strain from 23

(interquartile range [IQR] 537) to 13 (IQR 14.25). We constructed hierarchical Bayesian models to evaluate factors associated with being part of a genetic cluster (i.e., *M. kansasii* SNP distance between isolates of <45) as well as factors associated with continuous SNP distance. The model results (Figure 2) indicated that for each pair of participants, if both participant residences were supplied by the Weng Park water purification plant, their *M. kansasii* isolates had increased likelihood of being genetically clustered (odds ratio [OR] 1.81, 95% credible interval [CrI] 1.25–2.60), after adjusting for the other relevant factors. This finding was mirrored in the SNP model, where the SNP distances between the *M. kansasii* isolates of those pairs of persons were ≈12% smaller on average (risk ratio [RR] 0.88, 95% CrI 0.85–0.92). With a clustered model and a more conservative cutoff of 32 SNPs, the magnitude of the association was similar but not statistically significant (OR 1.54, 95% CrI 0.96–2.45) (Appendix Figure 2). We found a lower magnitude of association for the Fongshan water purification plant with a clustering OR of 1.39 (95% CrI 1.12–1.71) and a RR of SNP distance of 0.96 (0.94–0.98). We observed an inverse but statistically insignificant association between linear spatial distance and odds of clustering; adjusted OR was 0.77 (95% CrI 0.20–2.88) for every 1 km increase in spatial distance. Having a linkage to the same heavy industrial zone and sharing the same healthcare facility were not significantly associated with genetic relatedness, nor did they significantly alter the effect sizes for other variables in the model. Therefore, we removed those 2 variables from the final multivariable models for parsimony.

The statistical models also provided an estimate of a spatially referenced random effect parameter for each person that described the participants' residual (after adjustment for other factors) risk of being infected with an *M. kansasii* isolate that was genetically similar to other participants in the study. We mapped posterior mean estimates for those parameters (Appendix Figure 3). From visual inspection, we noted no obvious areas of increased residual risk.

### Tree Structure of *M. kansasii* Population

We created a maximum-likelihood phylogram that showed 3 main clades (Figure 3). Participants with residential water supplied by the Weng Park water purification plant were more likely to be in clade A. Of the 73 persons in clade A, 14 (19.2%) were supplied by Weng Park. Of 58 participants in clade B, 3 (5.2%) had water purified by Weng Park; 5 (6.1%) of 82 participants in clade C had water purified by Weng Park ( $p = 0.008$  for the association between genetic clade

and Weng Park). Participants supplied by the Fongshan water purification plant were more likely to be in clade B. Of the 58 participants in clade B, 27 (46.6%) were supplied by Fongshan. Of 73 participants in clade A, 15 (20.5%) were supplied by Fongshan; 27 (32.9%) of 82 in clade C were supplied by Fongshan ( $p = 0.006$  for the association between genetic clade and Fongshan). We conducted a sensitivity analysis that included an additional 22 publicly available *M. kansasii* isolates from outside Taiwan in the phylogenetic analysis. The resulting phylogeny revealed the 3 major clades, which contained the same Taiwan samples as in the main analysis (Appendix Figure 4). Similarly, when we constructed a phylogeny of our samples from Taiwan without masking recombinant regions, the same clades were identified with identical members in clades A and B. Clade C in the phylogeny constructed without recombination masking contained all of the samples seen in the primary analysis but also overlapped with 59 samples from clade B.

**Table.** Demographics of included population in study of phylogeographic analysis of *Mycobacterium kansasii* isolates from patients with *M. kansasii* lung disease in industrialized city, Taiwan\*

Characteristic	Value
Total no. participants	216
Mean age, y (SD)	67.29 (17.59)
Sex	
F	70 (32.4)
M	146 (67.6)
Mean body mass index (SD)	20.98 (4.14)
Smoking status	
Current smoker	46 (21.3)
Ex-smoker	45 (20.8)
Never smoked	125 (57.9)
Chest radiograph findings	
Consolidation	42 (20.1)
Fibrocavitary disease	37 (17.1)
Fibrosis	1 (0.5)
Nodular bronchiectasis	81 (38.8)
Nodule	3 (1.4)
Other†	52 (24.1)
Prior tuberculosis	52 (24.1)
Chronic obstructive disease	51 (23.6)
Pulmonary disease	
Bronchiectasis	35 (16.2)
Asthma	19 (8.8)
Pneumoconiosis	4 (1.9)
Lung cancer	9 (4.2)
Inhaled corticosteroid use	19 (8.8)
Residential water purification plant‡	
Baolai	1 (0.5)
Weng Park	22 (10.2)
Lingkou	4 (1.9)
Lujhu	1 (0.5)
Kaotan	99 (45.8)
Fongshan	70 (32.4)
Chengcinghu	128 (59.3)
Pingding	84 (38.9)

\*Values are no. (%) except as indicated.

†Other patterns included mixtures of the aforementioned patterns.

‡Participant residences may be supplied by >1 water purification plant.

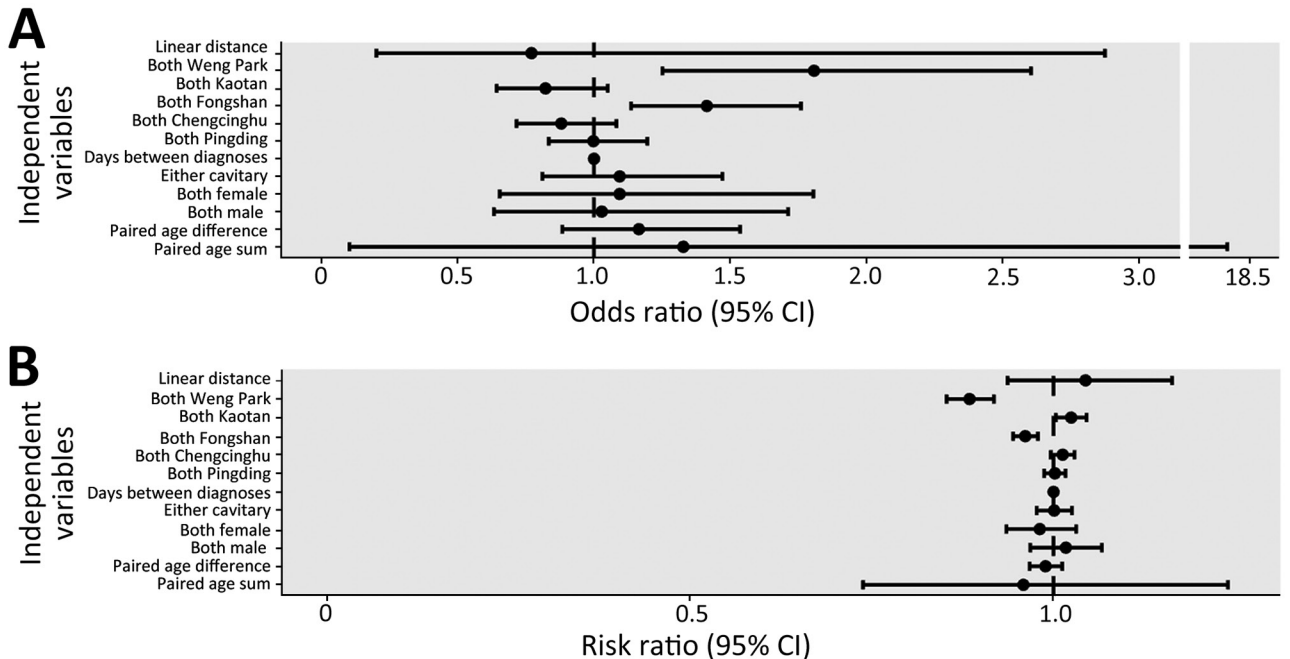
## Discussion

Using densely sampled cases of *M. kansasii* infection from a tropical metropolitan city in southern Taiwan, we investigated environmental factors associated with genetic relatedness using WGS data. We found that pairs of persons with *M. kansasii* lung disease living in a district served by specific water purification plants (Weng Park and Fongshan) were at greater risk of being infected with genetically similar *M. kansasii* isolates. The association of smaller spatial distances between pairs of persons had a statistically significant trend toward more genetically similar *M. kansasii* strains.

We previously conducted a spatial analysis of 537 *M. kansasii* cases from 4 major tertiary hospitals in Kaohsiung and identified 2 suspicious spatial clusters (11). In this analysis, we found that the Weng Park and Fongshan water purification plants, both near 1 of the previously identified spatial hotspots, were associated with increased genetic relatedness among case pairs.

Weng Park, the water purification plant with the strongest signal of genetic relatedness (adjusted OR 1.81, 95% CrI 1.25–2.60) among *M. kansasii* cases, accounted for only 4% of the total water supply in the study area but was associated with 22 (10%) of the total 216 cases. We also observed that Weng Park was significantly associated with 1 major *M. kansasii* clade in our phylogenetic analysis. Because water from different water plants mixed in underground water pipes, determining the source of supply for a particular household was difficult. Given the relatively low number of households supplied by Weng Park, the cases and households labeled as Weng Park might have also received their supply from other water plants, such as Kaotan and Pingding (Figure 1). The misclassification of water purification plants at the household level would lead us to underestimate the association between Weng Park and genetic relatedness of *M. kansasii* among case pairs.

This study does not provide a mechanism for the association between Weng Park and Fongshan water plants and the genetic clustering of *M. kansasii*-infected cases. Each water plant applied different methods of water purification, resulting in differences in pH value or organic matter content that might affect the risk for contamination and growth of *M. kansasii* (6). Different sources of raw water might also potentially influence the microbiological ecology; a 2003 analysis found that water from Weng Park had substantially higher general hardness than other water plants, suggesting a higher contribution from underground water as the source (13). The presence of more sediment accumulation in pipelines from underground water than from surface water might accelerate biofilm development (24). As



**Figure 2.** Associations of environmental and clinical risk factors with genetic relatedness based on pair-level data using hierarchical Bayesian regression models in phylogeographic analysis of *Mycobacterium kansasii* isolates from patients with *M. kansasii* lung disease in industrialized city, Taiwan. A) Odds ratios for pairs of *M. kansasii* isolates to be in a genetic cluster (using the single-nucleotide polymorphism [SNP] cutoff of 45). An odds ratio of >1 suggests that the risk factor was associated with genetic clustering. B) Risk ratios for increase in SNP distance between pairs of isolates. A risk ratio of <1 suggests that the risk factor was associated with a shorter pairwise SNP distance. The 3 smaller water purification plants (Lingkou, Baolai, and Lujhu) were not considered in the analysis as they together only provided service to 6 participants.

an example of the effect of the water treatment and distribution system on NTM abundance, a previous study in the United States reported low NTM relative abundances in Mississippi River water, a source for the drinking water system, but high relative abundances in the distribution system and tap water (25). Further environmental samplings should be conducted to examine the distribution of *M. kansasii* isolates in different water sources and plants, particularly Weng Park and Fongshan water plants.

Recent analyses of *M. abscessus*, another pathogenic NTM, have suggested that transmission might occur from person-to-person, especially in cystic fibrosis patients who might attend the same clinic (5). Our study did not find a potential link between shared clinic visits and the genetic clustering of *M. kansasii* in the investigated cases by analyzing pairs of persons' strains treated at the same hospital.

Although WGS has been widely applied to understand the transmission dynamics of *M. tuberculosis*, its application for studying NTM transmission including *M. kansasii* is still limited (9,26,27). The population structure and genomic diversity of *M. kansasii* on a global scale have been previously reported (28), but the genomic diversity over a well-defined geo-

graphy has not been characterized. We used the cutoff of 45 and 32 SNPs to define genetic relatedness on the basis of the empirical SNP distribution in our population, but the relationship between those measures of genetic distance and transmission remains uncertain. A major difference between *M. tuberculosis* and *M. kansasii* is that genomic recombination occurs frequently in *M. kansasii* through distributive conjugal transfer (28). A recent study using 60 samples of *M. kansasii* from different provinces of China revealed that the pairwise SNP distance (after masking recombinant regions) of those isolates were all within 20 SNPs and suggested a threshold of 4 SNPs to define clustering (29). Both that study and ours excluded recombinant regions using the same methods, further complicating the interpretation of our analysis in terms of transmission inference. In addition, selecting an effective threshold to cluster cases with a potential shared exposure site might be dependent on the local epidemiology and genetic diversity of the tested population. Along with this study, 2 other large analyses of clinical *M. kansasii* whole-genome sequences (9,29) showed that, after masking recombinant regions, the resulting sequences appeared to belong to homogeneous clusters with maximum SNP distances on the

order of 100 SNPs. The largest analysis to date created a phylogeny that estimated the most recent common ancestor of a global collection of clinical isolates to be timed to the early 1900s (9).

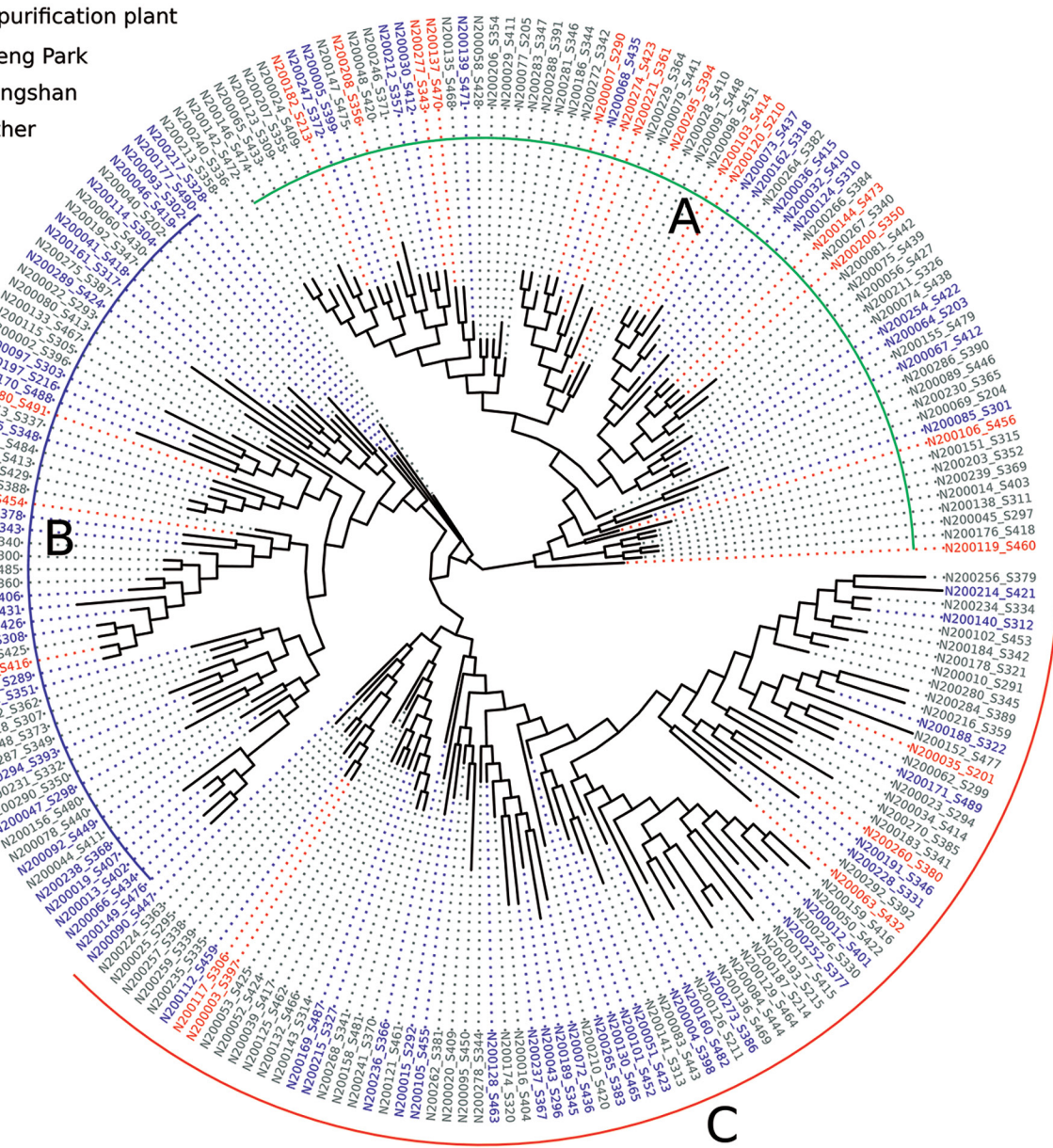
A major strength of this study is the combination of WGS and detailed spatial information on the environmental determinants of interest, including the water supply and heavy industrial zoning. The novel Bayesian hierarchical modeling approach correctly accounts for the correlation of pairwise spatial-genetic data and the simultaneous adjustment for potential

confounders. Previous studies on the transmission of *M. kansasii* have mostly applied conventional genotyping methodologies (e.g., restriction fragment length polymorphism or targeted PCR analysis) (30,31).

The first limitation of this study is that the *M. kansasii* isolates came from 1 major tertiary medical center and its affiliated hospitals (accounting for 56% of all *M. kansasii* isolates from all major medical centers in Kaohsiung during the same period) (11). *M. kansasii* infection is not a notifiable disease in Kaohsiung, and thus the population coverage of this

Water purification plant

- Weng Park
- Fongshan
- Other



**Figure 3.** Maximum-likelihood phylogram of clinical *Mycobacterium kansasii* isolates from patients with *M. kansasii* lung disease in industrialized city, Taiwan. Phylogeny with major clades are labeled as A, B, and C; colored text indicates source of water for patient households.

analysis is not comprehensive. The suboptimal population coverage posed a challenge in identifying environmental exposures. Second, our analysis revealed an association between certain water purification plants (Weng Park and Fongshan) and genetic relatedness, but the route of transmission cannot be confirmed without environmental sampling. Weng Park only accounted for 10% of total cases and Fongshan accounted for 32% of cases. Environmental determinants of most *M. kansasii* cases remain to be elucidated. Third, we only have crude (district-level) spatial coverage of the water purification plants, and misclassification of water supply at the household level might well occur. Last, we used the residential address as the proxy for exposure assessment, and we were not able to obtain definitive workplace exposures.

In conclusion, our novel spatial phylogenetic analysis in a densely sampled, well-defined geography revealed an independent association between certain water purification plants and the genetic relatedness of *M. kansasii* isolates. Our approach demonstrated the utility of combining WGS sequencing and readily available clinical and environmental information to obtain useful insights into the transmission of *M. kansasii* to trigger further environmental investigation and control measures.

### Acknowledgments

We thank laboratory members Mo-Hua Li and Yu-Wei Chen for helping with mycobacterial culture and DNA extraction of *M. kansasii* isolates.

Funding for this study was provided by Taiwan Ministry of Science and Technology (MOST 111-2628-B-002-045, MOST 112-2314-B-037 -073 -MY3), Kaohsiung Municipal Ta-Tung Hospital (KMTTH-111-039, KMTTH-112-R002), and National Health Research Institutes (NHRI-EX108-10805PI). P.G.T.C. was funded by a career development grant from the Fogarty International Center of the National Institutes of Health (5K01TW011194).

All genomes are shared through the US National Center for Biotechnology Information Sequence Read Archive.

P.G.T.C., P.C.L., H.L.H., T.C., and H.H.L. conceptualized the study. P.G.T.C., J.L.W., and T.C. did the genomic and statistical analyses. B.S., C.Y., and T.I. contributed to the genomic analyses. P.C.L. contributed to the spatial analysis. P.L.L. provided isolates for analyses. P.G.T.C., P.C.L., H.L.H., T.C., and H.H.L. wrote the initial draft of the manuscript. All authors edited and reviewed the final manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

### About the Author

Dr. Cudahy is an infectious disease specialist and assistant professor at Yale School of Medicine. His primary research interest is mycobacterial infection outcomes and epidemiology.

### References

- Huang HL, Lu PL, Lee CH, Chong IW. Treatment of pulmonary disease caused by *Mycobacterium kansasii*. *J Formos Med Assoc.* 2020;119(Suppl 1):S51-7. <https://doi.org/10.1016/j.jfma.2020.05.018>
- Martin-Casabona N, Bahrmand AR, Bennedsen J, Thomsen VO, Curcio M, Fauville-Dufaux M, et al.; Spanish Group for Non-Tuberculosis Mycobacteria. Non-tuberculous mycobacteria: patterns of isolation. A multi-country retrospective survey. *Int J Tuberc Lung Dis.* 2004;8:1186-93.
- Marras TK, Chedore P, Ying AM, Jamieson F. Isolation prevalence of pulmonary non-tuberculous mycobacteria in Ontario, 1997-2003. *Thorax.* 2007;62:661-6. <https://doi.org/10.1136/thx.2006.070797>
- Corbett EL, Hay M, Churchyard GJ, Herselman P, Clayton T, Williams BG, et al. *Mycobacterium kansasii* and *M. scrofulaceum* isolates from HIV-negative South African gold miners: incidence, clinical significance and radiology. *Int J Tuberc Lung Dis.* 1999;3:501-7.
- Bolden N, Mell JC, Logan JB, Planet PJ. Phylogenomics of nontuberculous mycobacteria respiratory infections in people with cystic fibrosis. *Paediatr Respir Rev.* 2023;46:63-70.
- Vaerewijck MJM, Huys G, Palomino JC, Swings J, Portaels F. Mycobacteria in drinking water distribution systems: ecology and significance for human health. *FEMS Microbiol Rev.* 2005;29:911-34. <https://doi.org/10.1016/j.femsre.2005.02.001>
- Blanc SM, Robinson D, Fahrenfeld NL. Potential for nontuberculous mycobacteria proliferation in natural and engineered water systems due to climate change: a literature review. *City Environ Interact.* 2021;11:100070. <https://doi.org/10.1016/j.cacint.2021.100070>
- Gan Y, Rahmatika I, Kurisu F, et al. The fate and risk of nontuberculous mycobacteria in the water supply system: a review. *H2Open J.* 2022;5:180-197.
- Luo T, Xu P, Zhang Y, Porter JL, Ghanem M, Liu Q, et al. Population genomics provides insights into the evolution and adaptation to humans of the waterborne pathogen *Mycobacterium kansasii*. *Nat Commun.* 2021;12:2491. <https://doi.org/10.1038/s41467-021-22760-6>
- Huang HL, Cheng MH, Lu PL, Shu CC, Wang JY, Wang JT, et al. Epidemiology and predictors of NTM pulmonary infection in Taiwan—a retrospective, five-year multicenter study. *Sci Rep.* 2017;7:16300. <https://doi.org/10.1038/s41598-017-16559-z>
- Liu BC, Huang HL, Chan TC, Lee SJ, Lin JN, Lee CH, et al. Spatial cluster analysis of *Mycobacterium kansasii* infection in Kaohsiung, Taiwan. *Taiwan J Public Health.* 2021;40:713-23.
- Daley CL, Iaccarino JM, Lange C, Cambau E, Wallace RJ Jr, Andrejak C, et al. Treatment of nontuberculous mycobacterial pulmonary disease: an official ATS/ERS/ESCMID/IDSA clinical practice guideline. *Clin Infect Dis.* 2020;71:e1-36. <https://doi.org/10.1093/cid/ciaa241>
- Kuo J, Chen Y. Statistical analysis and exploration of drinking water quality in Kaohsiung City. Kaohsiung City Government; 2007 [cited 2022 Sep 16]. <https://ksep.kcg.gov.tw/FileDownload/FileUpload/20191027122705617901.odt>



14. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20:257. <https://doi.org/10.1186/s13059-019-1891-0>
15. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci.* 2017;3:e104. <https://doi.org/10.7717/peerj-cs.104>
16. Phelan J. NTM-profiler [cited 2022 Sep 14]. <https://zenodo.org/records/10851060>
17. Seemann T. Shovill [cited 2022 Sep 14]. <https://github.com/tseemann/shovill>
18. Pribelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. *Curr Protoc Bioinformatics.* 2020;70:e102. <https://doi.org/10.1002/cpbi.102>
19. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 2015;43:e15–15. <https://doi.org/10.1093/nar/gku1196>
20. Seemann T. snp-dists [cited 2022 Sep 14]. <https://github.com/tseemann/snp-dists>
21. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics.* 2019;35:4453–5. <https://doi.org/10.1093/bioinformatics/btz305>
22. Warren JL, Chitwood MH, Sobkowiak B, Colijn C, Cohen T. Spatial modeling of *Mycobacterium tuberculosis* transmission with dyadic genetic relatedness data. *Biometrics.* 2023;79:3650–63. <https://doi.org/10.1111/biom.13836>
23. Hoff PD. Random effect models for network data. In: *Dynamic social network modeling and analysis: workshop summary and papers.* Washington: The National Academies Press; 2003. p. 303–12.
24. Learbuch KLG, Smidt H, van der Wielen PWJJ. Water and biofilm in drinking water distribution systems in the Netherlands. *Sci Total Environ.* 2022;831:154940. <https://doi.org/10.1016/j.scitotenv.2022.154940>
25. Hull NM, Holinger EP, Ross KA, Robertson CE, Harris JK, Stevens MJ, et al. Longitudinal and source-to-top New Orleans, LA, U.S.A. drinking water microbiology. *Environ Sci Technol.* 2017;51:4220–9. <https://doi.org/10.1021/acs.est.6b06064>
26. Bryant JM, Grogono DM, Rodriguez-Rincon D, Everall I, Brown KP, Moreno P, et al. Emergence and spread of a human-transmissible multidrug-resistant nontuberculous mycobacterium. *Science.* 2016;354:751–7. <https://doi.org/10.1126/science.aaf8156>
27. Nikolayevskyy V, Niemann S, Anthony R, van Soolingen D, Tagliani E, Ködmön C, et al. Role and value of whole genome sequencing in studying tuberculosis transmission. *Clin Microbiol Infect.* 2019;25:1377–82. <https://doi.org/10.1016/j.cmi.2019.03.022>
28. Tagimi F, Pillonel T, Bertelli C, Jatón K, Greub G. Pathogenic determinants of the *Mycobacterium kansasii* complex: an unsuspected role for distributive conjugal transfer. *Microorganisms.* 2021;9:348. <https://doi.org/10.3390/microorganisms9020348>
29. Guo Y, Cao Y, Liu H, Yang J, Wang W, Wang B, et al. Clinical and microbiological characteristics of *Mycobacterium kansasii* pulmonary infections in China. *Microbiol Spectr.* 2022;10:e0147521. <https://doi.org/10.1128/spectrum.01475-21>
30. Picardeau M, Prod'Hom G, Raskine L, LePennec MP, Vincent V. Genotypic characterization of five subspecies of *Mycobacterium kansasii*. *J Clin Microbiol.* 1997;35:25–32. <https://doi.org/10.1128/jcm.35.1.25-32.1997>
31. Kwenda G, Churchyard GJ, Thorrold C, Heron I, Stevenson K, Duse AG, et al. Molecular characterisation of clinical and environmental isolates of *Mycobacterium kansasii* isolates from South African gold mines. *J Water Health.* 2015;13:190–202. <https://doi.org/10.2166/wh.2014.161>

---

Address for correspondence: Hsien-Ho Lin, Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei 100, Taiwan; email: hsienho@ntu.edu.tw; Hung-Ling Huang, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung 807, Taiwan; email: 990325kmuh@gmail.com

Article DOI: <https://doi.org/10.3201/eid3008.240021>

*EID cannot ensure accessibility for supplementary materials supplied by authors. Readers who have difficulty accessing supplementary content should contact the authors for assistance.*

# Phylogeographic Analysis of *Mycobacterium kansasii* Isolates from Patients with *M. kansasii* Lung Disease in Industrialized City, Taiwan

## Appendix

### Supplementary methods

#### Mycobacterial isolation and species identification

We collected pre-treatment respiratory samples from bronchoalveolar lavage fluid or sputum before *M. kansasii* lung disease was diagnosed. We inoculated all samples on both Mycobacteria Growth Indicator Tube (MGIT, BD Biosciences, Sparks, MD, USA) and Löwenstein–Jensen(L-J) agar. Species identification was performed with Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) spectrometry analyses from colonies isolated on L-J solid culture medium using the Bruker Biotyper system with Mycobacteria Library v2.0 (Bruker Daltonics GmbH & Co. KG., Bremen, Germany). We extracted genomic DNA from each collected isolate following the cetyltrimethylammonium bromide (CTAB) protocol (1). The DNA samples with sufficient quantity and an OD260/OD280 ratio of 1.5 or higher were stored at  $-20^{\circ}\text{C}$  until sequencing.

#### WGS classification and filtering

We removed 22 samples with most reads from a species other than *M. kansasii* from analysis. One additional sample had a large degree of non-mycobacterial reads, and 4 samples were from participants missing clinical or residential data. We excluded these from analysis, leaving 216 samples for our primary analysis. For *M. kansasii* samples with more than 5% of

reads identified as non-*M. kansasii* (as determined by bracken), we included a filtering step with Kraken Tools to remove non-mycobacterial reads. These filtered samples had a range of non-mycobacterial reads of 5 to 38% before filtering.

### **Analysis of genetic relatedness**

The unit of analysis was the pair of individuals from the study sample. The goal was to assess the associations between pair-level genetic relatedness and its potential determinants including environmental exposure, spatial distance, demographics, and clinical features (investigated at the individual and pair level). For the binary outcome of clustering, we used a logistic regression framework to estimate the associations between individual-level and pair-wise variables, and genetic clustering. For the count valued outcome of SNP distance, we used negative binomial regression. Our Markov chain Monte Carlo sampling used standard weakly informative prior distributions specified for the model parameters. In total we collected 110,000 MCMC samples from the algorithms, with a burn-in of 10,000 states, and thinned by a factor of 10 to reduce posterior autocorrelation.

### **Analysis on heavy industry exposure and genetic relatedness of *M. kansasii***

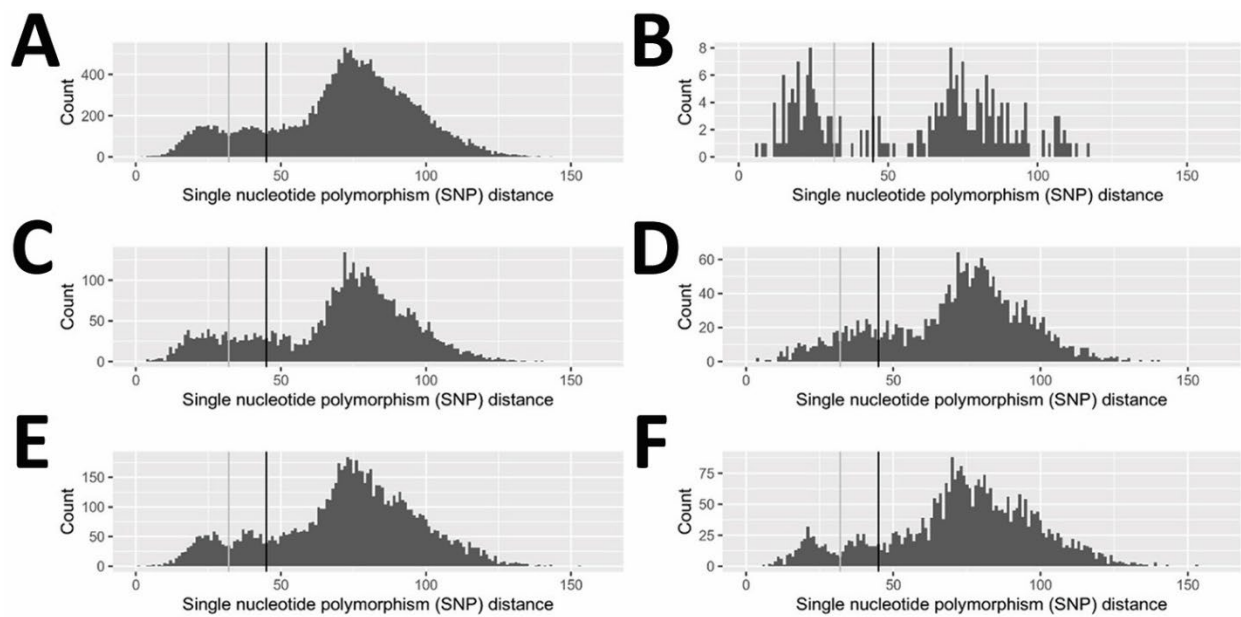
As described in the introduction section, we hypothesized that specific water supplies or heavy industrial zones might be associated with the risk of *M. kansasii* infection. To explore the potential association between heavy industry exposure and genetic relatedness of *M. kansasii*, we collected geographic information on heavy industrial zoning in Kaohsiung from the Open Data Platform of Kaohsiung City Government. Since people only work in the industrial zones during the day and we did not have information on specific work addresses, we used a probabilistic approach to identify study participants with a high probability of working in one of the four (Linhai, Dafa, Fongshan, and Linyuan) heavy industrial zones in Kaohsiung. In brief, we used the aggregated mobility information inferred from mobile phone calling data to identify villages with the highest day-time connectivity to each of the four heavy industrial zones. We determined whether each of the study participants are connected to each of the four industrial zones or not based on their residential village.

Following our approach in analyzing the association between potential risk factors and genetic relatedness, we conducted our analysis for two outcomes of genetic relatedness: 1) a binary outcome where we classified pairs of strains as in the same cluster based on a threshold

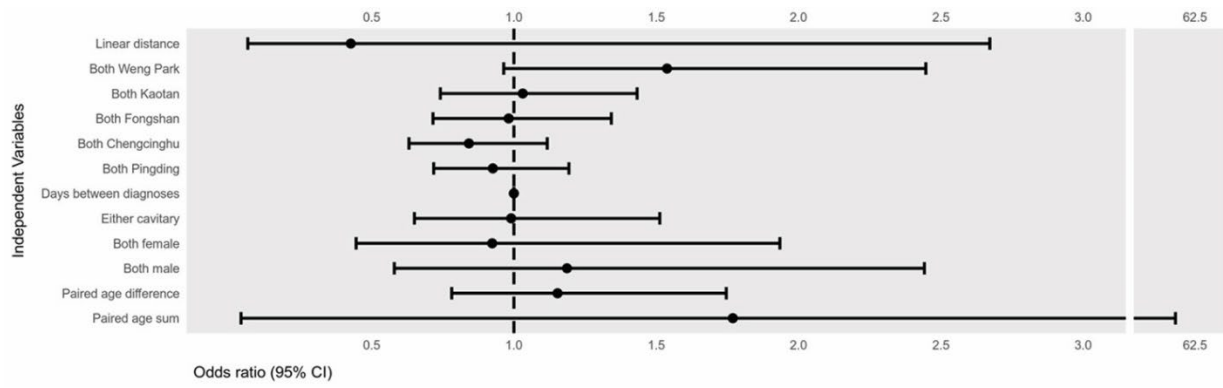
SNP distance and 2) a continuous measure of SNP distance between pairs of strains. We applied the hierarchical Bayesian spatial network methods to estimate the association between heavy industrial zoning and genetic relatedness (see the methods section for details). In the hierarchical Bayesian regression analysis based on paired data, there was no independent significant association between heavy industrial zoning (based on residential address among districts in the top 5% linkage to each industrial area) and risk of genetic relatedness after adjusting for other covariates for both binary and continuous genetic measures.

## References

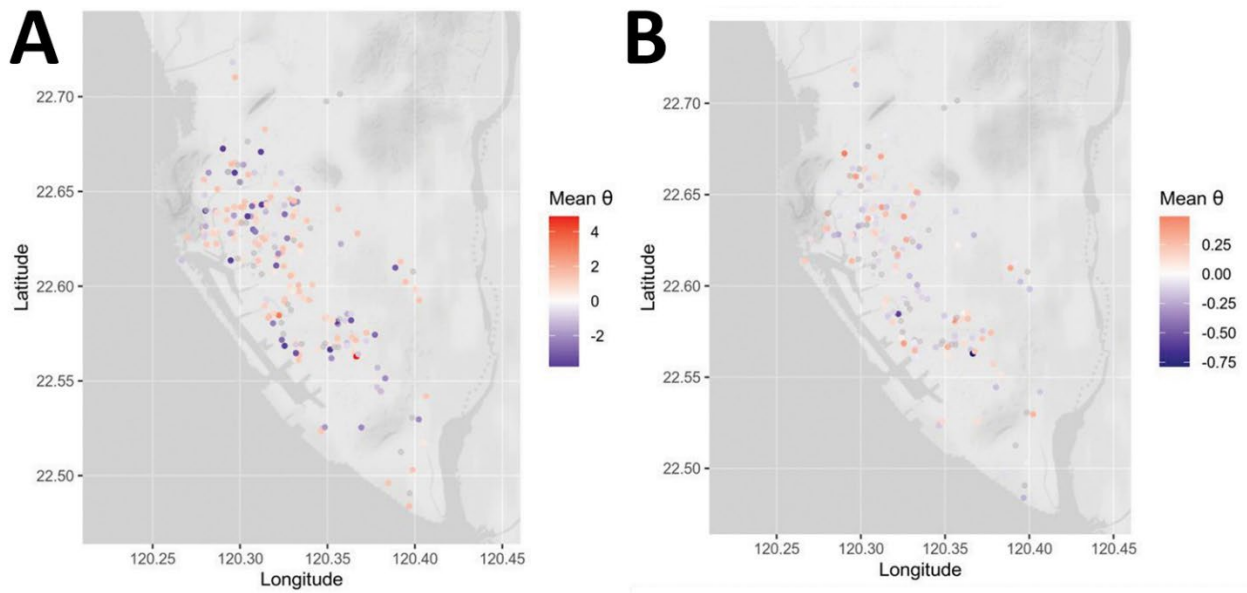
1. Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs WR Jr. Genetic manipulation of *Mycobacterium tuberculosis*. *Curr Protoc Microbiol*. 2007;10:10A.2. PMID 18770603



**Appendix Figure 1.** Single nucleotide polymorphism (SNP) distances between pairs of isolates, stratified by water supply to residential addresses of participants. Black line indicates cluster cutoff of 45 SNPs. Grey line indicates cluster cutoff of 32 SNPs. Count indicates the number of participants pairs with that SNP distance between their isolates.



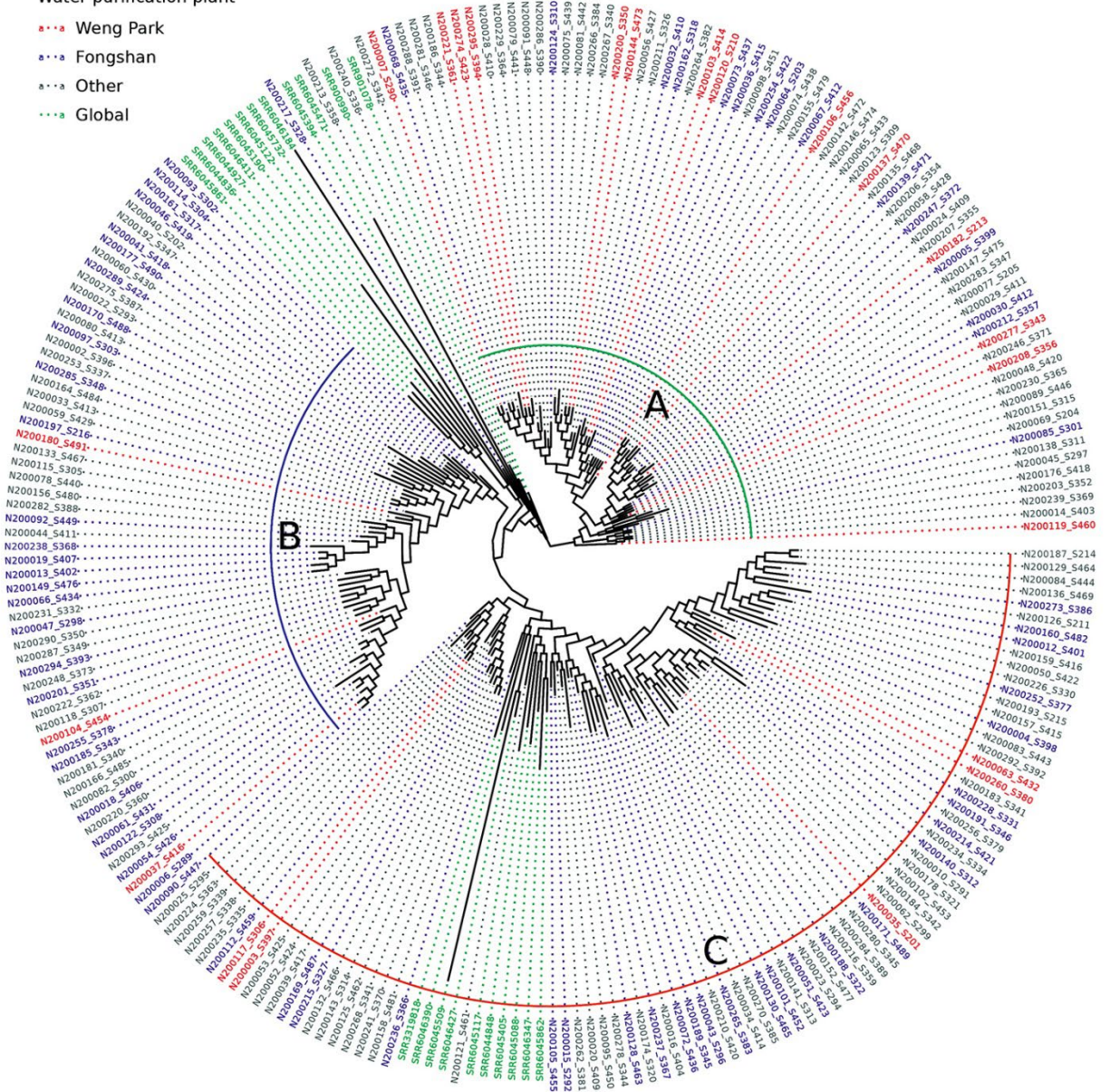
**Appendix Figure 2.** Cluster model (32 SNP cutoff) estimates.



**Appendix Figure 3.** Participant level random effects ( $\theta$ ). A) Cluster model estimates, B) SNP model estimates. Participant home locations (jittered) covered by individual level random effects value (if significant)

Water purification plant

- a Weng Park
- b Fongshan
- c Other
- d Global



Appendix Figure 4. Phylogeny of *M. kansasii* isolates including global samples.