Malays J Pathol 2026; 48(1): 63 – 68

ORIGINAL ARTICLE

Exploring the distribution of *Legionella pneumophila* sequence types in the water supply system of Klang Valley integrated transit system

Jun Leong WONG[1,2], Yun Fong NGEOW[1,2], Sook Fan YAP[1,2,*]

[1]Department of Pre-clinical Sciences, M. Kandiah Faculty of Medicine and Health Sciences, Universiti Tunku Abdul Rahman, Jalan Sungai Long, 43000 Kajang, Selangor, Malaysia; [2]Dr. Wu Lien-Teh Centre for Research in Communicable Diseases, M. Kandiah Faculty of Medicine and Health Sciences, Universiti Tunku Abdul Rahman, Jalan Sungai Long, 43000 Kajang, Selangor, Malaysia

Abstract

Introduction: Legionella pneumophila, a microorganism that thrives in both natural freshwater and man-made water systems, is a significant pathogen that causes Legionnaires' disease, a potentially fatal form of pneumonia. This study aimed to investigate the distribution of *L. pneumophila* sequence types (ST) within the water supply system of the Klang Valley Integrated Transit System (KVITS) in Malaysia. *Materials and Methods:* Sequence-Based Typing (SBT) was used to determine the sequence type of the *L. pneumophila* isolates by amplifying seven different loci (*flaA, pilE, asd, mip, mompS, proA*, and *neuA*), as per the European Working Group for Legionella Infections (protocol version 5.0). *Results: L. pneumophila* was isolated from five out of 80 samples (6.3%). These isolates comprised five distinct sequence types: ST1, ST22, ST2210, ST3017, and ST3029. Three isolates typed as ST1, ST22 and ST2210 belong to serogroup 1. Phylogenetic analysis suggested multiple sources of contamination. Conclusion: This study suggests the need for a comprehensive water management plan for KVITS, including routine testing and risk assessments, to reduce the risk of Legionnaires' disease outbreaks.

Keywords: Legionella pneumophila, Sequence-based typing, Klang Valley Integrated Transit System, Waterborne pathogens

INTRODUCTION

Legionella pneumophila (*L. pneumophila*) is a bacterium responsible for Legionnaires' disease, a form of pneumonia that can be fatal in vulnerable populations, such as the elderly and immunocompromised individuals.[1] *L. pneumophila* thrives in natural freshwater environments, including lakes and streams, and can also colonise man-made water systems, including cooling towers, hot tubs, and plumbing systems.[2] The bacteria can survive and multiply in water systems where temperatures are between 20 and 50 degrees Celsius especially in systems that provide nutrients such as algae and other organic matter.[3]

When water containing *L. pneumophila* is aerosolised, it poses a significant health risk as individuals can inhale contaminated water droplets containing the bacterium and subsequently acquire Legionnaires' disease.[4] Globally, *L. pneumophila* is recognised as a leading cause of community-acquired pneumonia, with outbreaks often traced to large water systems that serve public spaces, such as hotels, hospitals, and transportation hubs.[5] Despite global concerns, there remains limited research on the occurrence and distribution of *L. pneumophila* in public transit systems, which are high-traffic areas. In the local context, they serve over 800,000 commuters within the Klang Valley daily.[6] Therefore, studying the distribution of *L. pneumophila* sequence types in water supply systems would help to identify potential sources of contamination and the route of dissemination from contaminated sources. Given the potential risk of *Legionella* contamination in water systems, it is crucial for organisations responsible for public health and

**Address for correspondence:* Sook Fan YAP. M. Kandiah Faculty of Medicine and Health Sciences, Universiti Tunku Abdul Rahman, Jalan Sungai Long, 43000 Kajang, Selangor, Malaysia. Tel: +603-90860288; Email: *yapsf@utar.edu.my*

safety to take appropriate measures to prevent and control the spread of *L. pneumophila* in their water systems to protect public health.[1,4,7]

Previous studies on *L. pneumophila* have largely focused on cooling towers and healthcare settings, leaving a gap in the understanding of the risks associated with potable water systems in public transportation networks.[8] Identifying the specific sequence types (ST) of *L. pneumophila* in water systems is one way to trace sources of contamination and assess the effectiveness of water treatment protocols. Sequence-based typing (SBT) has emerged as a powerful tool for identifying *L. pneumophila* at the genetic level, allowing for a deeper understanding of the bacterium's distribution across different environments.[4,9]

Globally, there is limited research on the occurrence and distribution of *L. pneumophila* in public transit systems, which are high-traffic areas. The Klang Valley Integrated Transit System (KVITS) is a large-scale public transportation network in Malaysia that serves millions of people each year through a comprehensive system that includes mass rapid transit (MRT), light rail transit (LRT), monorail, bus rapid transit (BRT) and commuter rail services. Additionally, the system includes the KLIA Transit, providing seamless connectivity between different modes of transportation and promoting sustainable and accessible public transportation options for residents and visitors of the Klang Valley. KVITS water supply exemplifies a large, complex water system that could serve as a reservoir for *L. pneumophila.* Given the increasing number of public transit users, the potential public health implications of *L. pneumophila* contamination in such a water system is significant and cannot be overstated.

This study aimed to examine the distribution of *L. pneumophila* sequence types in the water supply of the KVITS to evaluate potential public health risks and inform future water management practices. The findings of this study could have significant implications not only for the KVITS but also for other large-scale public transportation networks worldwide, emphasizing the need for improved monitoring and control strategies to prevent the spread of *L. pneumophila.*[4]

MATERIALS AND METHODS

Water samples were collected from faucets in toilet facilities at 80 stations within the KVITS in April 2019. Out of the total 112 stations in the network, only 80 stations were equipped with toilet facilities. From each of these 80 stations, one sample was collected randomly, either from male or female toilets. Samples were collected using 125 mL sterile polypropylene bottles and transported to the laboratory within 12 hours to ensure sample integrity.

Upon arrival, each sample was filtered using a sterile filter membrane with a pore size of 0.45µm. The filter membrane was then placed in a 50ml Falcon tube containing 20mL of distilled water and heated in a water bath at 48°C for 30 minutes to inhibit and reduce contamination by other microorganisms. After heating, the tube was vortexed for 1 minute to dislodge any trapped *L. pneumophila* cells from the membrane. A 100 µL aliquot was plated onto each quadrant of a Buffered Charcoal Yeast Extract (BCYE) agar plate (Isolab, Malaysia) and inoculated with a sterile inoculation loop by streaking in 4 quadrants. The plate was then incubated at 35°C with 3% CO_2 for up to 10 days. Colonies suspected to be *L. pneumophila* growing on BCYE were sub-cultured on another BCYE and blood agar. Colonies that grew on BCYE but not on blood agar were examined by Gram staining. Colonies that exhibited Gram-negative bacilli morphology were subsequently identified as *L. pneumophila* using a latex agglutination test kit with serogroup 1-15 antisera (Microgen, UK).

Genomic DNA was extracted from each *L. pneumophila* isolate using a commercial DNA extraction kit (Thermo Scientific GeneJET Genomic DNA Purification Kit #K0721, USA). Briefly, 1 mL of *L. pneumophila* culture was centrifuged at 12,000 g for 5 minutes. The pellet was then resuspended in 180 µL of lysis buffer containing proteinase K and incubated at 56°C for 10 minutes. After incubation, 200 µL of ethanol was added to the lysate, and the mixture was applied to a GeneJET purification column. The column was washed with buffer after which the purified DNA was eluted in 50 µL of elution buffer. The concentration and purity of the extracted DNA were measured using a NanoDrop spectrophotometer.

The SBT method developed by the European Working Group for *Legionella* Infections (SBT protocol version 5.0 and *neuAh* protocol version 1.0) was used to determine the sequence type of the *L. pneumophila* isolates. Briefly, genomic DNA was extracted from each *L. pneumophila* isolate using a commercial DNA extraction kit. The SBT method involved amplifying seven different loci (*flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA*,

Table 1: Reference sequences and regions used for allele assignment

Target	Amplicon size	GenBank accession no. of reference sequence	Target region included for allele assignment	Size of target region (nt)
flaA	414	X83232	653-834	182
pilE	460	AF048690	103-435	333
asd	576	AF034213	538-1010	473
mip	559	AJ496269	117-518	402
mompS	711	AF078136	523-874	352
proA	481	M32884	1134-1538	405
neuA	459	AJ007311	229-582	354
neuAh	791	FR750546	229-579	351

and *neuA*) using PCR, followed by sequencing of the amplicons (Table 1). The sequences were compared to a reference database of known sequence types to assign the sequence type of each isolate. Due to the inaccessibility of the EWGLI SBT database (www.hpa-bioinformatics.org.uk/legionella/legionella_sbt/php/sbt_homepage.php) at the time of this study, the sequence data for the seven loci were manually submitted to Public Health England (legionella-sbt@phe.gov.uk), who provided the corresponding ST numbers. Additionally, the sequenced amplicons for each locus of each isolate were submitted to GenBank.

The *neuAh* homolog, as described by Farhat *et al*.,[10] has been identified in *L. pneumophila* non-serogroup 1 strains, particularly when the *neuA* gene fails to amplify using the standard *neuA* primers in the SBT protocol. Therefore, when *neuA* amplification was unsuccessful, the *neuAh* allele result was reported, following the established SBT allele order (*flaA, pilE, asd, mip, mompS, proA,* and *neuAh*) in accordance with the ESGLI guidelines (version 1.0)

Statistical Analysis

Sequences were aligned using MUSCLE with default parameters prior to tree construction. Phylogenetic analysis was conducted using the Maximum Likelihood (ML) method in MEGA version 11 (Molecular Evolutionary Genetics Analysis, Pennsylvania State University, State College, PA, USA). The evolutionary history was inferred using the Tamura-Nei model. The tree with the highest log likelihood (-4681.06) is presented. Bootstrap analysis with 1000 replicates was performed to assess branch reliability, with bootstrap values displayed next to the branches. Initial tree(s) for the heuristic search were automatically generated using the Neighbor-Joining (NJ) and BioNJ algorithms from a pairwise distance matrix estimated using the Tamura-Nei model. The topology with the best log likelihood value was selected. A discrete Gamma distribution (G = 0.0500) was used to model evolutionary rate differences among sites with five rate categories. The final dataset comprised five nucleotide sequences with a total alignment length of 2502 positions.

RESULTS

A total of 80 water samples were collected from various faucet stations within the KVITS in April 2019. Following laboratory processing, *L. pneumophila* was isolated from five of the samples, yielding a 6.3% isolation rate. These isolates were subjected to SBT to determine their serogroups and ST, with the SBT analysis focusing on seven genetic loci: *flaA, pilE, asd, mip, mompS, proA, neuA or neuAh.*

The sequence typing identified five distinct *L. pneumophila* isolates, which were classified into different sequence types and serogroups. The detailed typing results are summarised in Table 2 below. The nucleotide sequences for each locus of these isolates were deposited in GenBank, with accession numbers listed in Table 3.

A phylogenetic tree was constructed using the ML method to show the evolutionary relationships among the identified sequence types. Since ST1 (cc055) isolated in this study is identical to the NCBI reference genome *L. pneumophila* ST1 (accession no. NC_006368), it was used as a reference to compare the genetic relationships among the other identified sequence types. The tree (Figure 1) shows *ST1* and *ST2210* isolates clustering closely together, while *ST22* is more distantly related. The sequence types

Table 2: Sequence-based typing data for *L. pneumophila* samples

Sample ID	Serogroup	ST	*flaA*	*pilE*	*asd*	*mip*	*mompS*	*proA*	*neuA/neuAh*
cc055	1	1	1	4	3	1	1	1	1
cc056	1	2210	1	4	3	1	93	30	1
cc087	1	22	2	3	6	10	2	1	6
cc041	2-15	3017	16	21	33	37	41	1	215
cc084	2-15	3029	1	12	3	6	1	1	220

Abbreviations: Sequence Type (ST), Flagellin A (flaA), Pilin E (pilE), Aspartate-semialdehyde dehydrogenase (asd), Macrophage infectivity potentiator (mip), Major outer membrane protein S (mompS), Proline iminopeptidase (proA), N-Acetylneuraminic acid hydrolase (neuA). neuAh is a homologue of neuA.

belonging to serogroups 2-15 (*ST3017* and *ST3029*) form distinct clusters, indicating their genetic divergence from the serogroup 1 isolates.

The presence of multiple sequence types (ST1, ST22, ST2210, ST3017 and ST3029) within the KVITS water system highlights the genetic diversity of *L. pneumophila* in this environment. The genetic divergence observed among these sequence types indicates varied sources of contamination or different environmental pressures within the water system influencing their persistence and distribution. These findings underscore the need for comprehensive monitoring and targeted water management strategies to mitigate the risks posed by genetically diverse *L. pneumophila* strains.

DISCUSSION

This study represents the first known report of *L. pneumophila* isolation from water obtained from toilet faucets in a rapid transit rail network in Malaysia. Despite a relatively low isolation rate of 6.3%, the colony counts for all five isolates exceeded 200 colony-forming units per milliliter (cfu/mL), highlighting the potential for significant contamination in certain areas of the KVITS. Current guidelines for *Legionella* testing primarily focus on cooling towers, with contamination thresholds above 100 cfu/mL typically necessitating decontamination efforts. However, there is absence of an established guideline for acceptable *Legionella* levels in potable water systems, such as those in public transportation networks.[1,4]

The presence of high colony counts in the KVITS water system suggests that certain parts of the network could serve as reservoirs of *L. pneumophila*, posing potential infection risks to passengers and staff. These findings underscore the need for targeted risk assessments to identify specific sources of contamination and periodic monitoring to detect trends in *Legionella* presence and concentration. Such assessments could guide the development of appropriate remediation strategies, including enhanced water treatment protocols and infrastructure upgrades.

The identification of distinct ST of *L. pneumophila*, such as ST1, ST22, ST2210, ST3017, and ST3029, provides useful insights into the genetic diversity of *Legionella* in the KVITS water supply. SBT enabled the precise identification of these ST, illustrating the genetic variations and potential virulence differences

Table 3: GenBank accession numbers for each locus of *Legionella pneumophila* isolates

Sample ID	*flaA*	*pilE*	*asd*	*mip*	*mompS*	*proA*	*neuA (h)*
cc055	PV200170	PV200200	PV200175	PV200185	PV200190	PV200180	PV200195
cc056	PV200171	PV200201	PV200176	PV200186	PV200191	PV200181	PV200196
cc087	PV200172	PV200202	PV200177	PV200187	PV200192	PV200182	PV200197
cc041	PV200173	PV200203	PV200178	PV200188	PV200193	PV200183	PV200198
cc084	PV200174	PV200204	PV200179	PV200189	PV200194	PV200184	PV200199

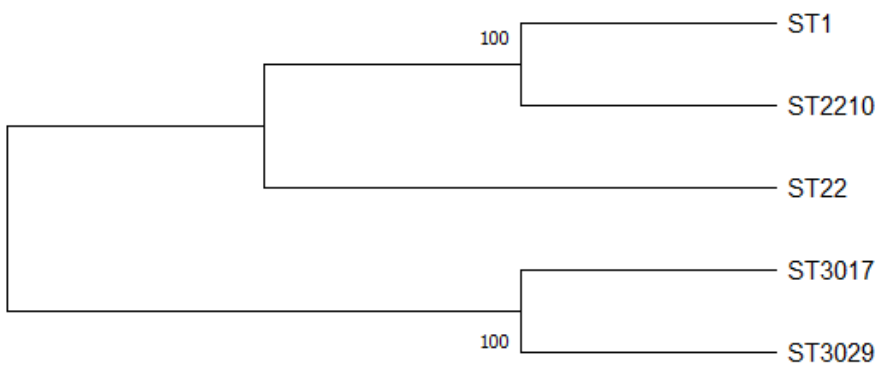

Figure 1. Phylogenetic tree illustrating the evolutionary relationships among *L. pneumophila* sequence types (ST1, ST22, ST2210, ST3017, and ST3029) isolated from the KVITS water system. Sequences were aligned using MUSCLE, and the tree was constructed using the Maximum Likelihood (ML) method with the Tamura-Nei model in MEGA11. Bootstrap values (1000 replicates) are displayed at branch points. The reference sequence used was *L. pneumophila* ST1 (GenBank accession: NC_006368).

among the isolates. Notably, the phylogenetic analysis revealed clustering patterns that suggest distinct evolutionary lineages, particularly among isolates belonging to serogroups 2-15. The use of ST1 as a reference sequence allowed for a clearer comparison of the genetic divergence observed among the other sequence types, supporting the conclusion that multiple sources of contamination may be present within the KVITS water system. This genetic differentiation may have implications for the pathogenicity and transmissibility of *L. pneumophila* strains in the KVITS water system.[11]

The presence of *L. pneumophila* in the KVITS water supply raises important public health concerns, especially given the increasing utilisation of public transportation and the potential for widespread exposure to contaminated water aerosols. The absence of stringent regulatory guidelines for *Legionella* in potable water systems within public transportation networks further complicates the mitigation of infection risks. To address these gaps, we recommend the implementation of a comprehensive water management plan by KVITS authorities. The water management plan should encompass regular monitoring, risk assessments, and prompt remedial actions to reduce *Legionella* concentrations to safe levels, thereby minimising the risk of Legionnaires' disease outbreaks.[4]

Additionally, the findings of this study underscore the necessity for broader regulatory oversight and the development of specific guidelines for *Legionella* control in public transportation water systems. Enhanced surveillance and stricter control measures could significantly improve the safety of water systems in high-traffic environments, ultimately protecting public health.

This study has several limitations. Only one sample was collected randomly from either the male or female toilets at each station. While this sampling strategy was systematic, it may not fully capture the distribution of *L. pneumophila* across the entire water system of all KVITS stations. Additionally, owing to the unavailability of the EWGLI SBT database at the time of the study, sequence data had to be submitted manually to Public Health England for ST assignment. Although STs were successfully assigned, direct access to the database could have enabled more immediate and broader comparative analysis with global datasets during the research process.

CONCLUSION

Overall, this research highlights the critical need for ongoing surveillance, improved water management practices, and a greater understanding of the genetic diversity of *L. pneumophila* in large-scale water systems. The insights gained from this study could inform similar efforts globally, guiding the development of strategies to prevent the spread of Legionnaires' disease in public transportation networks.[4]

Acknowledgements: We would like to acknowledge Ms. Sargit Kaur A/P Chanan Singh for her valuable assistance in this study. We also gratefully acknowledge Dr. Baharak Afshar from Public Health England for his assistance in providing the Sequence Type assignments for the *L. pneumophila* isolates. This research was financially supported by the Research Funds (Grant numbers: IPSR/RMC/UTARRF/2019-C1/W02 and IPSR/RMC/UTARRF/2021-C2/W02) from Universiti Tunku Abdul Rahman, Malaysia.

Informed Consent Statement: Informed consent was waived as this study did not involve human participants, animals, or any living organisms. It focused solely on environmental water sampling.

Authors' contributions: JLW: performed all experimental assays and prepared the draft; SFY and YFN: supervised the study and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. Lupia T, Corcione S, Shbaklo N, *et al*. *Legionella pneumophila* infections during a 7-year retrospective analysis (2016–2022): epidemiological, clinical features and outcomes in patients with Legionnaires' disease. Microorganisms. 2023; 11(2): 498.
2. Stout JE, Yu VL, Muraca P, Joly J, Troup N, Tompkins LS. Potable water as a cause of sporadic cases of community-acquired legionnaires' disease. N Engl J Med. 1992; 326: 151-5. doi: 10.1056/NEJM199201163260302.
3. Lau HY, Ashbolt NJ. The role of biofilms and protozoa in *Legionella* pathogenesis: implications for drinking water. J Appl Microbiol. 2009; 107: 368-78.
4. Phin N, Parry-Ford F, Harrison TG, *et al*. Epidemiology and clinical management of Legionnaires' disease. Lancet Infect Dis. 2014; 14: 1011-21.
5. Cunha BA, Burillo A, Bouza E. Legionnaires' disease. Lancet. 2016; 387: 376-85.
6. Ministry of Transport Malaysia. Statistic of Rail Transport [Data set]. [cited 2024 Sep 22]. [2024, September 22]. Available from https://mot.gov.my/en/statistic-of-rail-transport
7. Underwood AP, Jones G, Mentasti M, Fry NK, Harrison TG. Comparison of the *Legionella pneumophila* population structure as determined by sequence-based typing and whole genome sequencing. BMC Microbiol. 2013; 13: 302.
8. Quaranta G, Vincenti S, Ferriero AM, *et al*. *Legionella* on board trains: effectiveness of environmental surveillance and decontamination. BMC Public Health. 2012; 12: 618.
9. Steinert M, Hentschel U, Hacker J. *Legionella pneumophila*: an aquatic microbe goes astray. FEMS Microbiol Rev. 2002; 26: 149-62.
10. Farhat C, Mentasti M, Jacobs E, Fry NK, Lück C. The N-acylneuraminate cytidyltransferase gene, neuA, is heterogeneous in *Legionella pneumophila* strains but can be used as a marker for epidemiological typing in the consensus sequence-based typing scheme. J Clin Microbiol. 2011; 49: 4052-8.
11. Zhan XY, Zhu QY. Molecular typing of *Legionella pneumophila* isolates from environmental water samples and clinical samples using a five-gene sequence typing and standard sequence-based typing. PLoS One. 2018; 13(2): e0190986.