SEQUENCE-BASED TYPING TO IDENTIFY NOVEL POLYMORPHISMS AMONG *LEGIONELLA PNEUMOPHILA* ISOLATES IN WESTERN PENNSYLVANIA

by

Brooke R. Bixler

BS, Pennsylvania State University, 2014

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Public Health

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Brooke R. Bixler

It was defended on

November 4, 2015

and approved by

Thesis Advisor: Jeremy J. Martinson, DPhil, Assistant Professor, Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh

Committee Member: Lee H. Harrison, MD, Professor, Epidemiology, Graduate School of Public Health, University of Pittsburgh

Committee Member: Nancy W. Glynn, PhD, Assistant Professor, Epidemiology, Graduate School of Public Health, University of Pittsburgh

Jeremy J. Martinson, DPhil

SEQUENCE-BASED TYPING TO IDENTIFY NOVEL POLYMORHPISMS AMONG *LEGIONELLA PNEUMOPHILA* ISOLATES IN WESTERN PENNSLYVANIA

Brooke R. Bixler, MPH

University of Pittsburgh, 2015

ABSTRACT

Legionella pneumophila is a gram-negative bacterium that can cause two types of human illness, Legionnaires' disease and Pontiac fever. The goal of this study was to determine the sequence types (STs) for several *L. pneumophila* isolates collected in western Pennsylvania and compare them to STs from other geographic areas. Investigating ST of *L. pneumophila* can help to establish control measures and determine sources of outbreaks, both of which are of great public health significance. Environmental and patient samples were collected in Pittsburgh and Erie, Pennsylvania from October 2013 through December 2014. Sequence-based typing (SBT) was conducted to determine the sequence type of *L. pneumophila* present in the samples. Out of the nine STs that were identified in western Pennsylvania, five were novel. The known STs found in this study were ST 8, 986, 154, and 1941. When compared to similar SBT studies done in Portugal, Canada, England, and Spain, the results of the Pittsburgh study proved to be unique due to the identification of the five novel STs. The known STs found in the study were also not commonly found in investigations in other geographic areas. Overall, the results from the Pittsburgh investigation indicate environmental and patient isolates from western Pennsylvania have a unique ST compared to other isolates of *L. pneumophila* from around the world.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank Jessica Schlackman for training and assisting me in the lab as well as making the *filC* primer. I would also like to thank Dr. Lee Harrison and Dr. Jane Marsh for providing me with an opportunity to perform this investigation. Lastly, I would like to thank the hospital's Infection Prevention team for supporting this investigation.

1.0 INTRODUCTION

Legionella pneumophila is a non-sporulating, gram-negative bacillus that can be found in the environment or in artificial reservoirs, specifically in water sources (1). *L. pneumophila* may be found in hot tubs, cooling towers, hot water tanks, showerheads, air conditioning systems, humidifiers, and large plumbing systems (2,3,4). *L. pneumophila* grows best in temperatures ranging from 20º- 60 ºC, and in the presence of nutrients such as sediment, scale, sludge, and biofilm (1). For these reasons, *L. pneumophila* tends to grow more efficiently in artificial environments as opposed to natural environments.

L. pneumophila can cause two forms of infection in humans, Pontiac fever and Legionnaires' disease. Legionnaires' disease is a severe form of pneumonia caused by the bacterium. A milder infection can lead to a non-pneumonic flu-like illness known as Pontiac fever (5).

Legionnaires' disease was named after a point-source outbreak in Philadelphia at a convention hosted by the American Legion in 1976 (6). *L. pneumophila* can be transmitted via inhalation of aerosolized contaminated water droplets, however infection is not transmissible from person to person (7,8). The average incubation period for Legionnaires' disease ranges from 2-10 days (7,9). Most healthy individuals will not become infected with *L. pneumophila* after being exposed. However, risk factors associated with infection have been identified including smoking or a history of smoking, chronic heart of lung disease, cancer or hematologic

malignancy, a weakened immune system, and use of steroids or immunosuppressive therapy (3). In addition, males and the elderly are at higher risk of becoming infected after being exposed (3).

It is estimated that Legionnaires' disease is responsible for up to 30% of community-acquired pneumonia infections that require admission to an intensive care unit worldwide (10). The proportion of community–acquired pneumonias that results in a severe pneumonia is higher for Legionnaires' disease than for other causes. Therefore, there is a higher mortality rate associated with pneumonia from Legionnaires' disease, making it of great public health significance. Continued research into Legionnaires' disease and *L. pneumophila* infections may allow public health professionals to reduce the risk of infection and lower mortality in patients with this disease.

Studies have shown that pathogenesis and ecology of *Legionella* are intrinsically related. Consequently, different areas of the world have case-fatality rates. The United States and Australia have similar case-fatality rates. There is a case-fatality rate of 14% for nosocomial infections and 5-10% for community-acquired infections. When comparing this to Europe, it can be seen that the overall case-fatality rate is around 12% (10).

The exact prevalence of infection with *L. pneumophila* and the prevalence of the bacterium in the environment worldwide is not known (10). This is due in part to inconsistencies in both detection and reporting of *L. pneumophila* findings between countries. There are different methods used to determine infection, and some countries do not perform extensive investigations when outbreaks occur. Another reason prevalence is unknown is due to the ability for some serotypes of *L. pneumophila* to go undetected in soil.

The European Working Group for Legionella Infections (EWGLI) developed a standardized sequence-based typing (SBT) scheme for characterizing *L. pneumophila* (11).

Although SBT is not the most discriminatory epidemiologic method, it has been shown to produce rapid results with excellent inter-laboratory reproducibility at a lower cost compared to whole genome sequencing and it is also an easier method to perform (12, 13, 14). SBT requires the use of the polymerase chain reaction (PCR). PCR amplifies genomic DNA by using two short DNA molecules, called forward and reverse primers, that adhere to a target sequence of DNA. A DNA polymerase enzyme binds to the sequence-specific primer: template complexes and uses DNA nucleotides to synthesize new strands of DNA. Multiple rounds of amplification result in generation of many copies of the targeted genomic DNA. The DNA sequence of these amplified molecules can then be determined by other enzymatic reactions. SBT for *L. pneumophila* requires amplification and sequencing of 7 loci (15). Two of the loci are housekeeping genes (*asd* and *neuA*), and the other five loci are genes associated with virulence (*filC, pilE, mip, proA, and mompS*) (15). Once the DNA sequences of these PCR-amplified genes have been determined, they can be entered into an online database (http://bioinformatics.phe.org.uk/legionella/legionella_sbt/php/sbt_homepage.php), which will assign an allele number. The allelic profile obtained from these seven genes defines a sequence type (ST) (16). STs are used to track the spread of *L. pneumophila* strains during an outbreak. STs will differ if there is even a single nucleotide polymorphism (SNP) between isolates. SNPs are the most common type of genetic variation that occur when one DNA base pair has been substituted for another (17).

Bacteria have a few different mechanisms in which they can exchange genetic material. Conjugation is when one bacterium connects to another to pass on genetic information (18). This process starts when the donor cell contacts a specific receptor. The receptor is on a cell that does not have the conjugative plasmid, which will be transferred. The genes that will be transferred

are on what is known as the tra operon. The tra operon encodes proteins that will construct a pilus. The pilus is a protein tube that will extend outside of the donor cell. The pilus will form a channel between the cells to share genetic information from the donor cell to the recipient cell. After conjugation is completed, the recipient cell will become a new donor cell.

Transduction is another way bacteria can share genetic information. In this process, genetic information is transferred by means of a virus, also known as a phage (18). The phage can insert its DNA into a recipient cell. This will then modify that cell's genome to incorporate the new material.

The last mechanism for genetic exchange that will be discussed is transformation. This is when bacteria get the new DNA from the surrounding environment (18). DNA from the donor cell is released into the environment. It is then absorbed and incorporated into the recipient cell's genome. Collectively, these mechanisms of horizontal and lateral gene transfer have the potential to transfer large segments of bacterial genetic material at one time, which would be detected as substantial ST differences if they involved any of the genes used for SBT.

The aim of this study was to determine the STs of 56 *L. pneumophila* isolates collected in Western Pennsylvania, and to compare these results to STs of *L. pneumophila* from other similar investigations done worldwide. For this investigation, it was hypothesized that *L. pneumophila* isolated from certain environmental and patient samples from the Pittsburgh area would be genetically distinct compared to other isolates of *L. pneumophila* from around the world.

2.0 METHODS

The current study focused on identification of *L. pneumophila* in both environmental and patient samples taken from hospitals in the Pittsburgh and Erie area, and compared the results to other samples of *L. pneumophila* worldwide. SBT was used to determine the genetic relationship among isolates collected in western Pennsylvania and to compare them to STs from other geographic regions to determine if any unique STs were identified.

L. pneumophila isolates were collected by the hospital system's clinical microbiology laboratory from October 2013 through December 2014 from six member health care institutions in Western PA. For the purpose of this study, the institutions have been named Hospital A, B, C, D, E, and F. Hospitals A, B, D, E, and F lie within a ten mile radius of one another. Hospital C is located approximately 130 miles from the hospitals in Pittsburgh. Isolates were cultured from 51 environmental and five patient sources. A majority of environmental isolates were collected from showers, faucets, sinks, hot water supply lines, and ice machines. Of the patient isolates, three isolates came from bronchioalveolar lavage (BAL) and two isolates were collected from sputum (SPU) samples. More information on isolates can be found in Table 1.

The study had a sample size of 56 isolates. Of these isolates, two were found not to be *L. pneumophila*, five were patient isolates, and 49 were environmental isolates. The two samples that were identified as non-*Legionella* were found to be contaminants of the cultures done at the beginning of the study. The study sample included 45 isolates from Hospital F, one isolate from

Hospital D, four isolates from Hospital C, one isolate from Hospital A, one isolate from Hospital B, and four isolates from Hospital E.

The sequence-based typing (SBT) protocol for molecular epidemiological typing of *L. pneumophila* (Version 5.0) was followed (16). Genomic DNA was extracted using a NucliSENs easyMAG system following manufacturer's instructions using off-board lysis (bioMérieux, Durham, NC). PCR for SBT was performed as described by the EWGLI SBT (SBT) Database for *Legionella* using AmpliTaq Gold DNA Polymerase (Applied Biosystems, Grand Island, NY).

PCR primers targeting seven gene loci (*flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA*, and *neuA*) were used for amplification (Table 2). There were some cases of failed amplification of the neuA locus, in which the neuAh protocol for this known phenomenon was followed (16, 19). A second set of primers was designed for the *flaA* loci. This primer was named *filC* and used in addition to the forward primer for *flaA*. New *filC* primers were designed using PrimerSelect software (DNAstar, Madison, Wisconsin) using GenBank reference genome NC_002942.5. The *filC* gene is located at positions 1478004-1479758 in this reference genome sequence. The total product length of the *filC* PCR product obtained using the new primers is 1578 base pairs. The *filC* primer sequences were *filC* forward CCATCAACCGAGGTATAAAGAAAT and *filC* reverse TCGGGTACAGTAAATTCACAACAA (Table 2).

PCR products were purified by polyethylene glycol (PEG) precipitation, and pre-processed for DNA sequencing by incubation with the enzymes provided in the Exo-Sap It kit (Affymetrix, Santa Clara, CA). These enzymes remove any unincorporated PCR primers and dNTPs remaining after the PCR itself, as these would impede the subsequent sequencing reaction. Sanger DNA sequencing was performed with the BigDye Terminator v3.1 Cycle

Sequencing Kit according to manufacturer's instructions, using the same PCR primers described on the EWGLI website (16). Sequencing products were detected by capillary electrophoresis using an Applied Biosystems 3730xl Genome Analyzer and the data were analyzed using DNAstar Lasergene SeqMan Pro software (v.12; DNAstar, Madison, WI). Sequence alleles and sequence types (ST) were determined by querying the EWGLI SBT database.

A minimum spanning tree (MST) (Figure 1) was created to visualize the SBT data and show genetic linages among isolates. MST analysis was performed using BioNumerics software version 6.6 (Applied Maths, Austin, Texas), using the predefined MST for categorical data template (Figure 1). A clonal complex (CC) was defined by having ≤ 2 locus variants (20).

Use of the previously published reverse primer for one of the loci, *flaA*, resulted in many failed amplifications. Therefore, a new reverse primer, *filC* was created and used with the forward primer for *flaA* to perform successful amplification and sequencing using *filC* along with *flaA*. These primers differ in positions of the primers, fragment size, and primer sequence.

3.0 RESULTS

A total of nine STs of *L. pneumophila* were identified in this investigation, of which five were new STs. These new STs were named A, B, C, E, and F. ST 1941 was identified as ST D at the beginning of the current study, but was declared to be ST 1941 by the EWGLI before the end of the investigation. The known *L. pneumophila* STs found in this study were ST 8, 986, 154, and 1941. Out of a total study sample size of n=56, 2 isolates were in fact contaminants of the cultures and subsequently identified as non- *Legionella* (LEG 6 and 14), 49 isolates were environmental isolates and 5 isolates were from patients. All patient isolates were collected from respiratory samples. Of the isolates collected from patients, sample LEG 11 and sample LEG 12 are genetically related in that the sequences are the same at all seven loci and were grouped into ST F. However, isolates 11 and 12 are different at all seven loci compared to the other three patient isolates. These isolates, LEG 13, 17, and 25 were grouped into ST C.

Figure 1 shows the MST based on ST. There were three predominant clusters. ST F was the most prevalent of the nine ST identified. Isolates from ST F consisted of 14 environmental isolates collected from Hospital F, one environmental isolate from Hospital B, one environmental isolate from Hospital A, and two patient isolates collected from BAL samples at Hospital F. Eleven environmental isolates collected at Hospital E and Hospital F comprise the ST154 CC (Figure 1, green outline). The ST8 CC is comprised of 13 environmental isolates and three patient isolates (Figure 1, pink outline). Most of the isolates in this cluster fall under ST 8,

but it is important to note ST C because this is where the patient isolates can be found. There is only a difference of a single SNP between ST 8 and ST C. This SNP is found in the *asd* locus at allele 3. There is a difference of a thymine (T) nucleotide base when there is expected to be a cytosine (C) nucleotide base at the base pair position 226.

Environmental isolates from Hospital C belong to ST E. This ST was not closely related to any other ST found in the study. A single environmental isolate from Hospital F belonged to ST 1941 (LEG15). ST 986 also had a single isolate.

4.0 DISCUSSION

In this study, there were five new STs, ST A, B, C, E, and F, that had not been seen before, according to the EWGLI website. The results of the new STs have been submitted to EWGLI with assignments pending.

In general, the STs of the Pittsburgh study are unique when compared to other studies done worldwide (Table 3). A Portuguese study of the genetic diversity of clinical isolates of *L. pneumophila* resulted in identification of 25 different ST profiles (21). An outbreak investigation done in West Midlands, England in 2012 resulted in a single ST, which was found to be ST1268 (22). In a similar outbreak investigation conducted in Barrow-in Furness, Cumbria, England there was also just a single ST found, which was identified as ST78 (23). In a separate outbreak investigation performed in Comunidad Valenciana, Spain, there were distinct STs identified from both clinical and environmental isolates (24). There were eight known STs found in the Spanish study, as well as a few unique, previously unidentified STs. A 2012 outbreak investigation of *L. pneumophila* in Quebec City, Canada resulted in identification of five different STs (25). When comparing the results of these previous studies to the current study conducted in Pittsburgh, there was no overlap between the previously identified STs in other studies and STs found in the Pittsburgh study. This shows that the prevalent STs found in Western PA are not common among other areas where *L. pneumophila* has been studied using SBT. While the exact prevalence of various STs are unknown, the results of this study show that

a distinctive set of STs is present in Western PA, and that these STs do not appear to be found anywhere else.

When compared to other investigations of *L. pneumophila*, ST 1 is the most prevalent ST identified from both clinical and environmental isolates worldwide (26). Most of the studies done outside of Western PA identified isolates that were ST 1. The study done in Western PA did not contain ST 1 or any other STs that were similar to other investigations done recently worldwide. There were also five previously unidentified STs found from the study conducted in Pittsburgh.

In addition to having a unique set of STs, in most cases, there were no ST trends between the environmental sources such as the isolates from ice machines being in one sequence type or the isolates from sinks being in one sequence type. The exception to this was the isolates collected from Hospital C, which were all collected from sinks or faucets (Table 1).

The isolates from Hospital C are unique compared to the other isolates in this study, which could be due to the water source for the Hospital C being on a different water system compared to the other hospitals in the study. Hospital C is located in Erie, PA, which has a water supply from Lake Erie (27). This water source is different than the rest of the hospitals in the study, which are located in Allegheny County and get water from the Allegheny River (28). Water sources that are this far apart would be expected to have some different types of *L. pneumophila*, given that some STs are localized and others, such as ST 1, are seen worldwide (26).

Isolates LEG 13, 17, and 25, were patient samples that were found to only have one SNP difference compared to the environmental isolates from ST 8. Though it is impossible to determine the exact environmental source that caused the patient infections, with a difference of

only one SNP, there is a strong association between isolates LEG 13, 17, and 25 and the environmental sources in ST 8 from Hospital F.

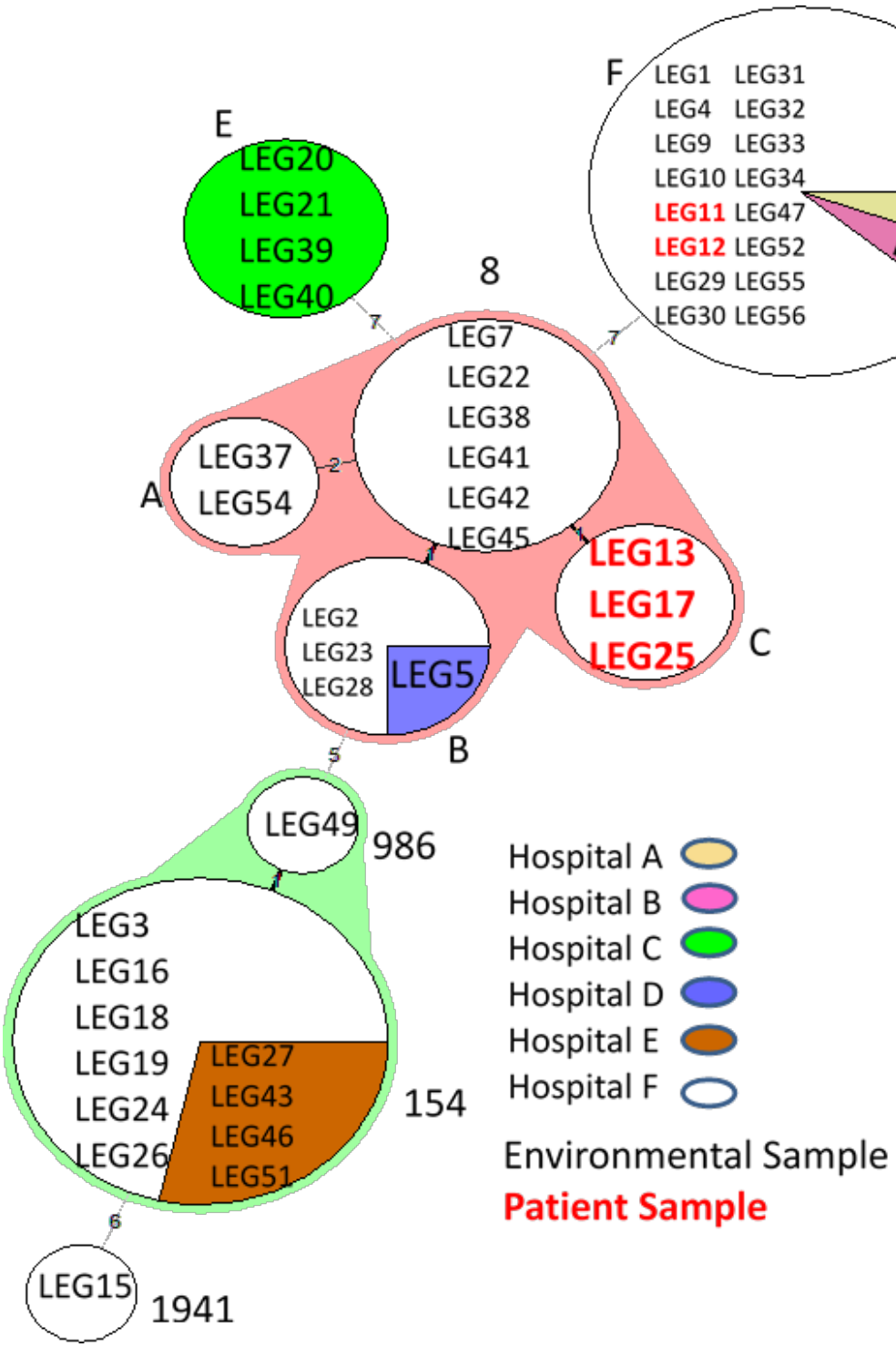
While SBT is currently used as a way to investigate outbreaks of *Legionella*, it does limit the results of studies that use it in that SBT describes genetic linages only as opposed to whole genome sequencing that can determine the entire genome of organisms. Even though SBT is relatively cheap and produces fast results, it is not the most discriminatory method when compared to whole genome sequencing. Whole genome sequencing would be able to differentiate STs among different hospitals. For example in ST F, isolates from Hospitals A, B, and F are all clustered together. While they might have similar STs, a difference would be expected among them since samples were obtained from three different institutions. Some of the isolates from the study have been sent out for whole genome sequencing, which will be completed at a later time in 2015.

5.0 CONCLUSIONS

For the environmental isolates, we can conclude Hospital C has a unique ST that could be due to the location of the hospital. ST F included the largest number of isolates, and had isolates from three different hospitals, all located in Allegheny County. Isolates in ST 8, ST A, ST B, and ST C are all related by two or less locus variations. Isolates in ST 986 and ST 154 are also related by a single locus variation. The prevalence of environmental isolates belonging to ST 8 and ST 154 CC from multiple locations in Pittsburgh suggests that these *L. pneumophila* genetic lineages are present in the Allegheny County water system.

While ST 8, ST 986, ST 154, and ST 1941 have been previously identified, they appear to not be highly prevalent in other recent outbreak investigations. To the best of our knowledge, ST A, ST B, ST C, ST E, and ST F are entirely unique to the Western Pennsylvania area. This demonstrates that environmental and patient isolates from the Western Pennsylvania area have a unique ST compared to other isolates of *L. pneumophila* worldwide. With this finding, more is known about the molecular epidemiology of *L. pneumophila*, which can be used in the future to better understand how the bacterium is spread worldwide. It is of great public health significance to know and understand as much as possible about *L. pneumophila* because it causes both community-acquired pneumonias and nosocomial pneumonia.

APPENDIX: TABLES AND FIGURE



Circles represent a unique genotype. The sizes of circles are proportional to the number of isolates. The lines connecting the sequence types show the number of locus variants. Colored clusters represent clonal complexes.

Figure 1. Minimum Spanning Tree for *L. pneumophila* SBT.

Table 1. Isolate information including allelic profile, culture date, source, submitter, and additional comments.

Isolate	flaA	pilE	asd	mip	mompS	proA	neuA	ST	CultDate	Source	Hospital	Source of Sample
1	19	25	25	33	49	28	225	F	7/14/14	ENV	F	Faucet
2	1	4	16	1	1	1	9	B	1/30/14	ENV	F	Ice Machine
3	11	14	16	16	15	13	2	154	5/29/14	ENV	F	Shower
4	19	25	25	33	49	28	225	F	11/8/14	ENV	F	Ice Machine
5	1	4	16	1	1	1	9	B	5/30/14	ENV	D	Pitt Shower
6									7/14/14	ENV	F	Ice Machine
7	1	4	3	1	1	1	9	8	8/14/14	ENV	F	Sink
8	19	25	25	33	49	28	225	F	7/11/14	ENV	B	Ice Machine
9	19	25	25	33	49	28	225	F	7/25/14	ENV	F	Sink
10	19	25	25	33	49	28	225	F	8/15/14	ENV	F	Exam Room
11	19	25	25	33	49	28	225	F	10/24/13	BAL	F	
12	19	25	25	33	49	28	225	F	12/5/14	BAL	F	
13	1	4	~3 (C22 6T)	1	1	1	9	C	6/13/14	SPUT	F	subcultured from agar slant; copy of LEG25
14									2/13/14		F	
15	12	8	11	15	20	12	2	1941	7/31/14		F	
16	11	14	16	16	15	13	2	154	2/19/14		F	
17	1	4	~3 (C22 6T)	1	1	1	9	C	6/14/14	BAL	F	
18	11	14	16	16	15	13	2	154	7/7/14	ENV	F	Sink Faucet
19	11	14	16	16	15	13	2	154	7/7/14	ENV	F	Sink Faucet
20	23	12	31	6	14	31	220	E	7/1/14	ENV	C	Sink Faucet
21	23	12	31	6	14	31	220	E	7/1/14	ENV	C	Faucet
22	1	4	3	1	1	1	9	8	12/24/13	ENV	F	Ice Machine Reservoir
23	1	4	16	1	1	1	9	B	1/22/14	ENV	F	Ice Machine
24	11	14	16	16	15	13	2	154			F	
25	1	4	~3 (C22 6T)	1	1	1	9	C	6/13/14	SPUT	F	subcultured from agar plate; copy of LEG13
26	11	14	16	16	15	13	2	154	7/7/14	ENV	F	Sink Faucet
27	11	14	16	16	15	13	2	154	7/2/14	ENV	E	Sink Faucet
28	1	4	16	1	1	1	9	B	8/18/14	ENV	F	Water Line
29	19	25	25	33	49	28	225	F	8/18/14	ENV	F	Sink
30	19	25	25	33	49	28	225	F	8/26/14	ENV	F	Water Line
31	19	25	25	33	49	28	225	F	9/30/14	ENV	F	Shower
32	19	25	25	33	49	28	225	F	8/29/14	ENV	F	Water Line

Table 1 Continued

33	19	25	25	33	49	28	225	F	8/29/14	ENV	F	Sink Water
34	19	25	25	33	49	28	225	F	8/29/14	ENV	F	Sink Faucet
35	19	25	25	33	49	28	225	F	8/29/14	ENV	A	Ice Machine
36	11	14	16	16	15	13	2	154	10/1/14	ENV	F	Sink Faucet
37	1	4	3	10	1	1	203	A	9/26/14	ENV	F	Shower
38	1	4	3	1	1	1	9	8	9/11/14	ENV	F	Sink Faucet
39	23	12	31	6	14	31	220	E		ENV	C	Sink
40	23	12	31	6	14	31	220	E		ENV	C	Sink
41	1	4	3	1	1	1	9	8		ENV	F	Shower
42	1	4	3	1	1	1	9	8		ENV	F	Sink
43	11	14	16	16	15	13	2	154		ENV	E	Sink
44	11	14	16	16	15	13	2	154		ENV	F	Shower
45	1	4	3	1	1	1	9	8		ENV	F	Water Line
46	11	14	16	16	15	13	2	154		ENV	E	
47	19	25	25	33	49	28	225	F		ENV	F	
48	1	4	3	1	1	1	9	8		ENV	F	
49	1	14	16	16	15	13	2	986		ENV	F	Sink Water
50	11	14	16	16	15	13	2	154		ENV	F	Sink Water
51	11	14	16	16	15	13	2	154		ENV	E	
52	19	25	25	33	49	28	225	F		ENV	F	Sink Water
53	11	14	16	16	15	13	2	154			F	
54	1	4	3	10	1	1	203	A	12/23/14		F	
55	19	25	25	33	49	28	225	F	12/23/14	ENV	F	
56	19	25	25	33	49	28	225	F	8/18/14	ENV	F	Water

ENV=Environmental
SPUT= Sputum
BAL= Bronchioalveolar lavage

Table 2. This table represents the genes used in this study as well as the primer sequences used for DNA amplification and sequencing.

Gene	Primer Sequence (5'-3')
flaA F	GCG TAT TGC TCA AAA TAC TG
flaA R	CCA TTA ATC GTT AAG TTG TAG
pilE F	CAC AAT CGG ATG GAA CAC AAA CTA
pilE R	GCT GGC GCA CTC GGT ATC T
asd F	CCC TAA TTG CTC TAC CAT TCA GAT G
asd R	CGA ATG TTA TCT GCG ACT ATC CAC
mip F	GCT GCA ACC GAT GCC AC
mip R	CAT ATG CAA GAC CTG AGG GAA C
mompS F	TTG ACC ATG AGT GGG ATT GG
mompS R	TGG ATA AAT TAT CCA GCC GGA CTT C
proA F	GAT CGC CAA TGC AAT TAG
proA R	ACC ATA ACA TCA AAA GCC
neuA F	CCG TTC AAT ATG GGG CTT CAG
neuA R	CGA TGT CGA TGG ATT CAC TAA TAC
filC F	CCATCAACCGAGGTATAAAGAAAT
filC R	TCGGGTACAGTAAATTCACAACAA

The primers shown in this table were previously established and routinely used for SBT of *L. pneumophila*, with the exception of *filC*, which was used in addition to *flaA* primers(15).

Table 3. This table shows the sequence types (STs) from the studies done using SBT worldwide.

Place of Investigation	Sequence Type (ST)
Western PA	ST 8, ST 154, ST 986, ST 1941, ST A, ST B, ST C, ST E, ST F
Portugal	ST 1, ST 16, ST 20, ST 22, ST 23, ST 37, ST 42, ST 44, ST 62, ST 75, ST 94, ST 98, ST 99, ST 100, ST 101, ST 102, ST 103, ST 146, ST 172, ST 173, ST 174, ST 436, ST 785, ST 1009, ST 1010
West Midlands, England	ST1268
Barrow-in Furness, Cumbria	ST78
Comunidad Valenciana, Spain	ST 1, ST 48, ST 301, ST 1146, ST 1393, ST 1394, ST 1396, ST1397, others (unidentified)
Quebec City, Canada	ST 1, ST 62, ST 150, ST 213, ST 284

BIBLIOGRAPHY

1. Pancer, K. (2013). Sequence-based typing of Legionella pneumophila strains isolated from hospital water distribution systems as complementary element of risk assessment of legionellosis in Poland. *Annals of Agricultural and Environmental Medicine 2013, 20*(3), 436-440. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pubmed/24069845
2. Pancer, K., Matuszewska, R., Bartosik, M., Kacperski, K., & Krogulska, B. (2013). Persistent colonization of 2 hospital water supplies by L. pneumophila strains through 7 years- Sequence-based typing and serotyping as useful tools for complex risk analysis. *Annals of Agricultural and Environmental Medicine 2013, 20*(4), 487-494. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pubmed/24364435
3. Cassier, P., Campese, C., Le Strat, Y., Che, D., Ginevra, C., Etienne, J., & Jarraud, S. (2015). Epidemiologic characteristics associated with ST23 clones compared to ST1 and ST47 clones of Legionnaires disease cases in France. *New Microbes and New Infections*, *3*, 29–33. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4337934/
4. Gilmour, M. W., Bernard, K., Tracz, D. M., Olson, A. B., Corbett, C. R., Burdz, T., … Berry, J. D. (2007). Molecular typing of a Legionella pneumophila outbreak in Ontario, Canada. Journal of Medical Microbiology, 56(Pt 3), 336–341. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2884934/
5. Qin, T., Zhou, H., Ren, H., Guan, H., Li, M., Zhu, B., & Shao, Z. (2014). Distribution of Sequence-Based Types of Legionella pneumophila Serogroup 1 Strains Isolated from Cooling Towers, Hot Springs, and Potable Water Systems in China. *Applied and Environmental Microbiology*, *80*(7), 2150–2157. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3993137/
6. Euser, S. M., Boogmans, B., Brandsema, P., Wouters, M., & Den Boer, J. W. (2014). Legionnaires' disease after using an industrial pressure test pump: a case report. *Journal of Medical Case Reports*, *8*, 31. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3917411/
7. Demirjian, A., Lucas, C., Garrison, L., Kozak-Muiznieks, N., States, S., Brown, E., . . . Hicks, L. (2015). The Importance of Clinical Surveillance in Detecting Legionnaires' Disease Outbreaks: A Large Outbreak in a Hospital With a Legionella Disinfection System--Pennsylvania, 2011-2012. *Clinical Infectious Diseases, 60*(11), 1596-1602. Retrieved July 31, 2015, from http://cid.oxfordjournals.org/

8. Yates, T. A., Bruin, J. P., Harrison, T. G., & Mannes, T. (2013). Molecular diagnostics and the public health management of legionellosis. *BMJ Case Reports*, *2013*, bcr2013008713. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3645249/
9. BEYRER, K., LAI, S., DREESMAN, J., LEE, J. V., JOSEPH, C., HARRISON, T., … WINDORFER, A. (2007). Legionnaires' disease outbreak associated with a cruise liner, August 2003: epidemiological and microbiological findings. *Epidemiology and Infection*, *135*(5), 802–810. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2870631/
10. Bartram, J., Chartier, Y., Lee, J., Pond, K., & Surman-Lee, S. (Eds.). (2007). *Legionella and the prevention of legionellosis* (276). World Health Organization.
11. Moran-Gilad, J., Mentasti, M., Lazarovitch, T., Huberman, Z., Stocki, T., Sadik, C., . . . Grotto, I. (2013). Molecular epidemiology of Legionnaires' disease in Israel. Clinical Microbiology and Infection, 20(7), 690-696. Retrieved July 31, 2015, from http://www.clinicalmicrobiologyandinfection.com/article/S1198-743X(14)61160-5/fulltext
12. Gaia, V., Fry, N. K., Afshar, B., Lück, P. C., Meugnier, H., Etienne, J., … Harrison, T. G. (2005). Consensus Sequence-Based Scheme for Epidemiological Typing of Clinical and Environmental Isolates of *Legionella pneumophila*. *Journal of Clinical Microbiology*, *43*(5), 2047–2052. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1153775/
13. Afshar, B., Fry, N. K., Bellamy, W., Underwood, A. P., Harrison, T. G., & and Members of the European Working Group for Legionella Infections . (2007). External Quality Assessment of a DNA Sequence-Based Scheme for Epidemiological Typing of *Legionella pneumophila* by an International Network of Laboratories . *Journal of Clinical Microbiology*, *45*(10), 3251–3256. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2045376/
14. Scaturro, M., Losardo, M., De Ponte, G., & Ricci, M. L. (2005). Comparison of Three Molecular Methods Used for Subtyping of *Legionella pneumophila* Strains Isolated during an Epidemic of Legionellosis in Rome. *Journal of Clinical Microbiology*, *43*(10), 5348–5350. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1248497/
15. Sánchez-Busó, L., Coscollá, M., Pinto-Carbó, M., Catalán, V., & González-Candelas, F. (2013). Genetic Characterization of *Legionella pneumophila* Isolated from a Common Watershed in Comunidad Valenciana, Spain. *PLoS ONE*, *8*(4), e61564. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3636276/
16. Legionella pneumophila Sequence Based Typing. (n.d.). Retrieved July 31, 2015, from http://bioinformatics.phe.org.uk/legionella/legionella_sbt/php/sbt_homepage.php
17. What are single nucleotide polymorphisms (SNPs)? (2015, November 9). Retrieved November 12, 2015, from http://ghr.nlm.nih.gov/handbook/genomicresearch/snp
18. Genetic Exchange. (2008, November 6). Retrieved November 12, 2015, from https://cspwproject.wordpress.com/genetic-exchange/
19. Farhat, C., Mentasti, M., Jacobs, E., Fry, N. K., & Lück, C. (2011). The *N*-Acylneuraminate Cytidyltransferase Gene, *neuA*, Is Heterogenous in Legionella pneumophila Strains but Can Be Used as a Marker for Epidemiological Typing in

the Consensus Sequence-Based Typing Scheme . *Journal of Clinical Microbiology*, *49*(12), 4052–4058. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3232988/

20. MLST - Home. (n.d.). Retrieved September 30, 2015, from http://eburst.mlst.net/3.asp
21. Chasqueira, M., Rodrigues, L., Nascimento, M., Ramos, M., & Marques, T. (2014). Genetic diversity and evolutionary relationships among Legionella pneumophila clinical isolates, Portugal, 1987 to 2012. *Eurosurveillance, 19*(46), 20965-20971. Retrieved July 31, 2015, from http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20965
22. Coetzee N, Duggal H, Hawker J, Ibbotson S, Harrison TG, Phin N, Laza-Stanca V, Johnston R, Iqbal Z, Rehman Y, Knapper E, Robinson S, Aigbogun N. An outbreak of Legionnaires' disease associated with a display spa pool in retail premises, Stoke-on-Trent, United Kingdom, July 2012 . Euro Surveill. 2012;17(37), 20271-20273. Retrieved July 31, 2015, from http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20271
23. Bennett, E., Ashton, M., Calvert, N., Chaloner, J., Cheesbrough, J., Egan, J., . . . Gent, R. (2013). Barrow-in-Furness: A large community legionellosis outbreak in the UK. *Epidemiol. Infect. Epidemiology and Infection, 142*(8), 1763-1777. Retrieved July 31, 2015, from http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=9284315&fileId=S0950268813002483
24. Coscollá, M., Fernández, C., Colomina, J., Sánchez-Busó, L., & González-Candelas, F. (2013). Mixed infection by Legionella pneumophila in outbreak patients. *International Journal of Medical Microbiology, 304*(3-4), 307-313. Retrieved July 31, 2015, from http://www.sciencedirect.com/science/article/pii/S1438422113001756
25. Lévesque, S., Plante, P.-L., Mendis, N., Cantin, P., Marchand, G., Charest, H., … Tremblay, C. (2014). Genomic Characterization of a Large Outbreak of *Legionella pneumophila* Serogroup 1 Strains in Quebec City, 2012. *PLoS ONE*, *9*(8), e103852. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4126679/
26. Kozak-Muiznieks, N. A., Lucas, C. E., Brown, E., Pondo, T., Taylor, T. H., Frace, M., … Winchell, J. M. (2014). Prevalence of Sequence Types among Clinical and Environmental Isolates of Legionella pneumophila Serogroup 1 in the United States from 1982 to 2012. *Journal of Clinical Microbiology*, *52*(1), 201–211. Retrieved July 31, 2015, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3911437/
27. Our Water. (n.d.). Retrieved September 30, 2015, from https://www.eriewater.org/our-water/
28. PWSA - Pittsburgh Water and Sewer Authority. (n.d.). Retrieved September 30, 2015, from http://www.pgh2o.com/