Comparison of Whole-Genome Sequences of *Legionella pneumophila* in Tap Water and in Clinical Strains, Flint, Michigan, USA, 2016

Emily Garner, Connor L. Brown, David Otto Schwake, William J. Rhoads, Gustavo Arango-Argoty, Liqing Zhang, Guillaume Jospin, David A. Coil, Jonathan A. Eisen, Marc A. Edwards, Amy Pruden

During the water crisis in Flint, Michigan, USA (2014–2015), 2 outbreaks of Legionnaires' disease occurred in Genesee County, Michigan. We compared whole-genome sequences of 10 clinical *Legionella pneumophila* isolates submitted to a laboratory in Genesee County during the second outbreak with 103 water isolates collected the following year. We documented a genetically diverse range of *L. pneumophila* strains across clinical and water isolates. Isolates belonging to 1 clade (3 clinical isolates, 3 water isolates from a Flint hospital, 1 water isolate from a Flint residence, and the reference Paris strain) had a high degree of similarity (2–1,062 single-nucleotide polymorphisms), all *L. pneumophila* sequence type 1, serogroup 1. Serogroup 6 isolates belonging to sequence type 2518 were widespread in Flint hospital water samples but bore no resemblance to available clinical isolates. *L. pneumophila* strains in Flint tap water after the outbreaks were diverse and similar to some disease-causing strains.

Legionnaires' disease is a severe form of pneumonia caused by inhalation of virulent species of aerosolized *Legionella* bacteria. In January 2016, the Michigan Department of Health and Human Services (MDHHS) and the Genesee County Health Department publicly announced 2 Legionnaires' disease outbreaks in Genesee County, Michigan, USA (*1*,*2*). The first outbreak occurred from June 2014 through March 2015 and the second from May 2015 through October 2015; a total of 90 cases and 12 deaths were documented (*1*–*3*). From April 2014 through October 2015, the city of Flint, in Genesee County, switched its drinking water source from Detroit Water and Sewer Department (DWSD), which used corrosion control, to the corrosive Flint River, without implementing federally mandated corrosion control; this new water source led to elevated lead in tap water over a prolonged period, now called the Flint water crisis (*4*). This disruption in water quality likely also stimulated the growth of *L. pneumophila*, the species most frequently identified as the causative agent of Legionnaires' disease (*5*,*6*), in Flint's distribution and plumbing systems (*7*).

Our prior work associated the Legionnaires' disease outbreaks with factors known to be conducive to *Legionella* growth: elevated iron (a consequence of corroded iron water mains), reduced free chlorine disinfectant residuals, and elevated water temperatures (*7*,*8*). Later, Zahran and colleagues reported that the odds of Flint residents being referred for Legionnaires' disease treatment while the Flint River was the source of tap water increased 6.3-fold and confirmed our report of associations with low chlorine residuals (*9*), but the odds analysis, which was based on the use of referral date rather than symptom onset date, excluded many healthcare-associated cases (*10*). Furthermore, during the second outbreak, *Legionella* spp. and *L. pneumophila* genes were found to be higher in the tap water of large buildings in Flint than in other water systems in US areas not experiencing outbreaks (*8*). Conversely, levels of the *mip* gene, which is specific to *L. pneumophila*, were largely below detection in Flint single-family residences, at least during the later stages of the water crisis when they were measured (2015–2016) (*8*). Large buildings with extensive plumbing networks, such as hospitals, are generally more susceptible to *Legionella* growth than are simpler plumbing systems characteristic of single-family homes (*11*); however, residences

Author affiliations: West Virginia University, Morgantown, West Virginia, USA (E. Garner); Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA (E. Garner, C.L. Brown, W.J. Rhoads, G. Arango-Argoty, L. Zhang, M.A. Edwards, A. Pruden); Methodist University, Fayetteville, North Carolina, USA (D.O. Schwake); University of California at Davis, Davis, California, USA (G. Jospin, D.A. Coil, J.A. Eisen)

DOI: https://doi.org/10.3201/eid2511.181032

are also of interest for *Legionella* growth, given concerns about the high rate of sporadic Legionnaires' disease (*12*) and potential for exposure in the home.

Our study objective was to use next-generation DNA sequencing to compare *L. pneumophila* isolated from Flint tap water after the second Legionnaires' disease outbreak with tap water isolates from neighboring drinking water systems outside of Flint that were never served by Flint River water and clinical strains received during the second outbreak at a regional reference laboratory in Genesee County. Within Flint, *Legionella* isolates were obtained from the tap water of a hospital, a large public building, and single-family residences several months after the water source was switched back to DWSD. In addition to serogroup testing, we used whole-genome sequencing to compare isolates in terms of sequence type (ST), average nucleotide identity, and single-nucleotide polymorphisms (SNPs).

Materials and Methods

Water Sample Collection and *Legionella* Isolation

After Flint resumed purchasing water with corrosion control from the original supplier, DWSD, water sampling campaigns were conducted 5 months (March 7–9, 2016), 8 months (June 21–27, 2016), and 10 months (August 15–16, 2016) later. Samples were collected from residences, small businesses, a large public building, and a hospital in Flint; as controls, samples were collected from buildings located outside of Flint that used DWSD or well water (Table 1). The March 2016 campaign targeted sampling of residences, small businesses, a large public building, and a hospital; samples were collected from hot (flushed for 30 seconds) and cold (stagnant) taps at each location. Samples were collected from the kitchen sink in homes and from restrooms in public buildings. The June 2016 campaign extensively sampled homes as part of a water heater cleaning campaign; the following samples were collected before and after a cleaning protocol: hot and cold stagnant kitchen tap samples, a stagnant shower sample of blended hot and cold water, a hot flushed kitchen tap sample, the water heater drain valve, and a flushed cold water sample from the outside hose bib or nearest tap to the service entry point. The August 2016 campaign targeted sampling from hot (flushed 30 seconds) and cold (stagnant) water taps from homes and small businesses. *Legionella* was cultured according to standard methods (*13*), and colonies were streaked to isolation.

Clinical Isolates

MDHHS provided 11 clinical isolates from de-identified Legionnaires' patients who received a diagnosis in 2015; however, 1 isolate could not be cultured and was deemed nonviable. When we initiated this study, we assumed that all 11 isolates originated from patients with some history of exposure in Flint or Genesee County during the Flint water crisis. However, we later learned that the commonality among clinical isolates was that they had been submitted to a Genesee County laboratory for analysis during the second outbreak and that 3 of the 11 isolates originated from patients who resided and received treatment outside of Genesee County (J. McFadden, MDHHS, pers. comm., 2017 Feb 1). Because the clinical isolates in this study were de-identified, comparison with the water isolates is described in terms of "*L. pneumophila* known to be capable of causing LD." We also included publicly available DNA sequence information from clinical reference strains in the analysis (Appendix 1 Table 2, https://wwwnc.cdc.gov/EID/article/25/11/18-1032-App1.pdf).

Whole-Genome Analysis of *L. pneumophila* Isolates

Whole-genome sequencing was conducted by MicrobesNG (https://microbesng.uk) on an Illumina MiSeq (https://www.illumina.com) with 2 × 250-bp paired-end reads and Nextera library preparation (Illumina). Sequencing was performed for a representative subset of each building type and water source, including 103 water isolates and the 10 available clinical *L. pneumophila* isolates (Appendix 1 Table 1). To verify DNA integrity, DNA extracts were quantified via a Qubit 2.0 fluorometer (https://www.thermofisher.com) and analyzed via gel electrophoresis. Positive (*L. pneumophila* strain 130b) and negative (*Stenotrophomonas maltophilia*) control strains were also sequenced, and 3 clinical strains were sequenced in duplicate on 2 MiSeq runs to evaluate run-to-run variation (Appendix 1 Figures 1–3). On average, 806,825 reads were obtained per isolate (range 280,380–2,031,828 reads). Reads were trimmed by using Trimmomatic (*14*), and de novo assemblies were generated by using SPAdes (*15*).

Genome sequences are available in GenBank under BioProject PRJNA453403. *Legionella* species assignments were determined via blastn (https://blast.ncbi.nlm.nih.gov) for isolate 16S rRNA gene sequences. Average nucleotide identity was calculated as previously described (*16*), and SNPs were identified by using kSNP3.0 (*17*). We also included 9 *L. pneumophila* reference strains (Appendix 1 Table 2). We performed sequence-based typing targeting the *flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA*, and *neuA* alleles (*18*) by using the mompS tool (*19*).

Serogroup Analysis

We identified *L. pneumophila* isolates belonging to serogroup 1 via detection of the *wzm* gene (*20*) in whole genome sequences. We verified DNA sequence-based classifications and determined unknown serogroups by using direct fluorescent antibody staining with fluorescein isothiocyanate–conjugated antibodies (m-TECH, http://www.4m-tech.com).

Table 1. Total number of buildings sampled, number of samples collected, and number of isolates analyzed for *Legionella*, Flint, Michigan, USA*

Water sample source	March 2016			June 2016			August 2016		
	No. buildings or samples	No. (%) positive	No. isolates analyzed	No. buildings or samples	No. (%) positive	No. isolates analyzed	No. buildings or samples	No. (%) positive	No. isolates analyzed
Flint residences	**5**	**0**		**32**†	**2 (6)**		**10**‡	**2 (20)**	
Hot (flushed)	5	0	0	62	2 (3)	3	14	1 (7)	1
Hot (stagnant)	NS			62	2 (3)	4	NS		
Cold (flushed)	NS			61	1 (2)	2	NS		
Cold (stagnant)	5	0	0	61	1 (2)	4	11	4 (36)	1
Water heater drain valve	NS			62	1 (2)	5	NS		
Shower (hot and cold)	NS			62	1 (2)	2	3	1 (33)	1
Hospitals	**1**	**1 (100)**		**NS**			**NS**		
Hot (flushed)	19	16 (84)	56	NS			NS		
Cold (stagnant)	19	6 (32)	14	NS			NS		
Buildings receiving DWSD water	**4**	**0**		**NS**			**8**	**0**	
Hot (flushed)	4	0	0	NS			8	0	0
Cold (stagnant)	4	0	0	NS			8	0	0
Flint large buildings	**2**	**1 (50)**		**NS**			**NS**		
Hot (flushed)	5	0	0	NS			NS		
Cold (stagnant)	5	1 (20)	1	NS			NS		
Buildings receiving well water	**1**	**1 (100)**		**NS**			**NS**		
Hot (flushed)	4	4 (100)	5§	NS			NS		
Cold (stagnant)	3	2 (67)	4¶	NS			NS		
Flint small businesses	**6**	**0**		**NS**			**8**	**0**	
Hot (flushed)	6	0	0	NS			8	0	0
Cold (stagnant)	6	0	0	NS			8	0	0

*Positive samples indicate presumptive *L. pneumophila* identified by performing culture according to the method described in (*12*). Unless otherwise noted, identification as *L. pneumophila* was confirmed by using whole-genome sequencing. Boldface indicates total buildings sampled. Blank cells indicate that data were not reported when applicable samples were not collected. NS, no samples of this type were collected.
†1 of the 32 homes was also sampled in March 2016.
‡5 of the 10 homes were also sampled in March 2016; 1 of the 10 was sampled in June 2016 (but not in March 2016; samples from this house were positive on both dates).
§4 of 5 isolates were a non–*L. pneumophila* species, according to whole-genome sequencing.
¶4 of 4 isolates were a non–*L. pneumophila* species, according to whole-genome sequencing.

Results

Legionella Isolate Characterization

Of the 515 total water samples collected and from which *L. pneumophila* isolation was attempted (Table 1), 43 samples (8%) were positive for *Legionella*. Of these, 22 (58%) of 38 hospital samples from March 2016, eight (2%) of 370 residence samples from June 2016 (positives originating from 2 separate residences), and 6 (21%) of 28 residence samples from August 2016 (positives originating from different taps in a single residence) were positive for culturable *L. pneumophila*. No isolates were obtained from businesses receiving DWSD water, but 6 (86%) of 7 taps at the school serviced by well water were identified as positive (although 5 of these were later determined to be *Legionella* species other than *L. pneumophila*).

16S rRNA genes mined from whole-genome sequences indicated that all clinical and water isolates, except for 8 of the 9 well water isolates, were *L. pneumophila*. The positive control strain was correctly identified as *L. pneumophila*; SNP analysis further classified it according to its known provenance (130b), and the negative control strain was also confirmed to be *S. maltophilia* (i.e., not *Legionella*). Serogrouping via presence of the *wzm* gene for serogroup 1 and direct fluorescent antibody staining for other serogroups indicated that all *L. pneumophila* isolates belonged to serogroups 1 and 6 (Table 2).

L. pneumophila isolates from clinical and water samples belonged to several STs (Table 2). Of serogroup 1 isolates, all belonged to STs 1, 44, 159, 192, 211, 213, or 222 or to a previously uncharacterized ST that we submitted to the European Working Group for *Legionella* Infections database (http://www.ewgli.org) and that has now been designated as ST2513. Serogroup 6 isolates all belonged to a previously uncharacterized ST that has now been designated as ST2518. Most hospital isolates belonged to ST2518, and isolates originating from residential tap water belonged primarily to ST192. Only ST1 was represented by both clinical and water isolates, specifically, 3 clinical isolates, 3 isolates from hospital tap water, and 1 isolate from residential tap water.

When classified according to SNP similarity, isolates formed distinct clades that were generally consistent with the ST classification (Figure). The ST1 clade varied by 2–1,062

Table 2. Summary of *Legionalla pneumophila* isolates from Flint, Michigan, USA, 2016

ST	Serogroup	Isolate origin*
1	1	3 hospital water (HH17, HH25, HH56), 1 residence water (RH08), 3 clinical (C2, C3, C7)
44	1	1 clinical (C6)
159	1	1 clinical (C1)
192	1	19 residence water (RC01, RC02, RC03, RC04, RC06, RC07, RD01, RD02, RD03, RD04, RD05, RH02, RH03, RH04, RH05, RH07 RH07, RS01, RS02)
211	1	1 clinical (C8)
213	1	2 clinical (C4, C5)
222	1	1 clinical (C9)
2513†	1	1 clinical (C10)
2518†	6	66 hospital water (HC01, HC02, HC03, HC04, HC05, HC06, HC07, HC08, HC09, HC10, HC11, HC12, HC13, HC14, HH01, HH02, HH03, HH04, HH05, HH06, HH07, HH08, HH09, HH10, HH11, HH12, HH13, HH14, HH15, HH16, HH18, HH19, HH20, HH21, HH22, HH23, HH24, HH26, HH27, HH29, HH30, HH31, HH32, HH33, HH34, HH35, HH36, HH37, HH38, HH39, HH40, HH41, HH42, HH43, HH44, HH45, HH46, HH47, HH48, HH49, HH50, HH51, HH52, HH53, HH54, HH55), 1 public building (PC01), 1 well water (WH03)
ND	ND	HH28, RC05, RH01, RS03, WC01, WC02, WC03, WC04, WH01, WH02, WH04, WH05

*First letter of isolate name indicates building type/location: H, hospital; R, residence; W, school using well water; P, large public building. Second letter indicates sample collection location; H, hot water tap; C, cold water tap; D, water heater drain valve; S, shower. Numerals 1–10 indicate clinical strains. ND, not determined because of lack of *L. pneumophila*–specific alleles or insufficient genome coverage; ST, sequence type. †New sequence types from this study submitted to European Working Group for Legionella Infections database (http://www.ewgli.org).

SNPs, and isolates varied from the reference Paris strain by 371–505 SNPs. In particular, clinical isolate C3 shared the highest degree of similarity with Flint tap water isolates (38–46 SNPs). Some degree of variation is expected to be associated with variability in sequencing accuracy because the 3 clinical isolates that were sequenced in duplicate on multiple MiSeq lanes differed from their replicate by 0–10 SNPs. Several other distinct clades emerged in which water isolates were grouped primarily by building type. A large clade of ST2518 isolates included most samples from the hospital, 1 sample from well water, and 1 sample from a large public building. Another clade contained only isolates originating from Flint residence water samples belonging to ST192. The SNP results were confirmed by phylogenetic analysis and average nucleotide identity comparison (Appendix 1 Figures 1–3; Appendix 2, https://wwwnc.cdc.gov/EID/article/25/11/18-1032-App2.xlsx).

The STs of 8 isolates derived from well water could not be determined because *L. pneumophila*–specific alleles were absent, suggesting that the isolates were mistakenly phenotypically characterized as *L. pneumophila* on the basis of colony morphology. Average nucleotide identity values comparing these isolates with the positive control *L. pneumophila* strain (130b) were 62.645%–62.969%, whereas average nucleotide identity values of a single species are generally >95% (*21*). These 8 isolates seem to be most closely related to *L. taurinensis*, *L. rubrilucens*, or *L. erythra*, because the 16S rRNA genes extracted from these genomes shared >99% nt similarity to all 3 species.

Discussion

When considered per capita, the Legionnaires' outbreaks in Genesee County are among the largest in US history. However, to our knowledge, few clinical sputum isolates were collected or preserved from these outbreaks; for most cases, only urine-antigen testing was conducted. A common problem in the United States is reliance on urine-antigen testing and lack of collection of clinical *Legionella* isolates; these practices unfortunately limit the ability to track sources of infection, learn from past outbreaks, and prevent future outbreaks (*22*,*23*). Among the clinical sputum isolates that were sent to Genesee County laboratories during the outbreaks, none were from patients residing in homes serviced by Flint water (S. Lyon-Callo, MDHHS, pers. comm., 2018 Apr 5); thus, direct examination of potential residential exposure is not possible from this study. Given that 68% of patients' residences were confirmed to not have been serviced by Flint water (*3*), the potential exists that a portion of the remaining 32% had some residential exposure in Flint.

Another challenge of tracking sources of Legionnaires' disease is limited availability of water isolates. Given that the outbreaks were not publicly announced until 3 months after the conclusion of the second outbreak (January 2016), few environmental specimens were collected or preserved when the outbreaks were occurring. Analysis of any other water isolates that might exist from the time of the crisis would be valuable for learning more about this outbreak. MDHHS reported that 106 environmental *Legionella* specimens were retained at a Flint hospital but were not submitted to the State Health Department as had been requested (Sarah Lyon-Callo, MDHHS, pers. comm., 2019 Apr 30). Thus, a more definitive study of environmental sources of the outbreaks is not possible without a wider collection of clinical and environmental isolates.

Our study provides a survey of the landscape of genetic diversity among *Legionella* isolates collected from tap water from a range of building types served by the Flint drinking water distribution system over the 1-year period after the switch back to DWSD water. We compared these isolates with clinical isolates and with isolates from tap water of neighboring water systems never served by the Flint River or DWSD. Although it was not possible to collect water isolates during the actual outbreaks, previous studies have demonstrated that a single strain of

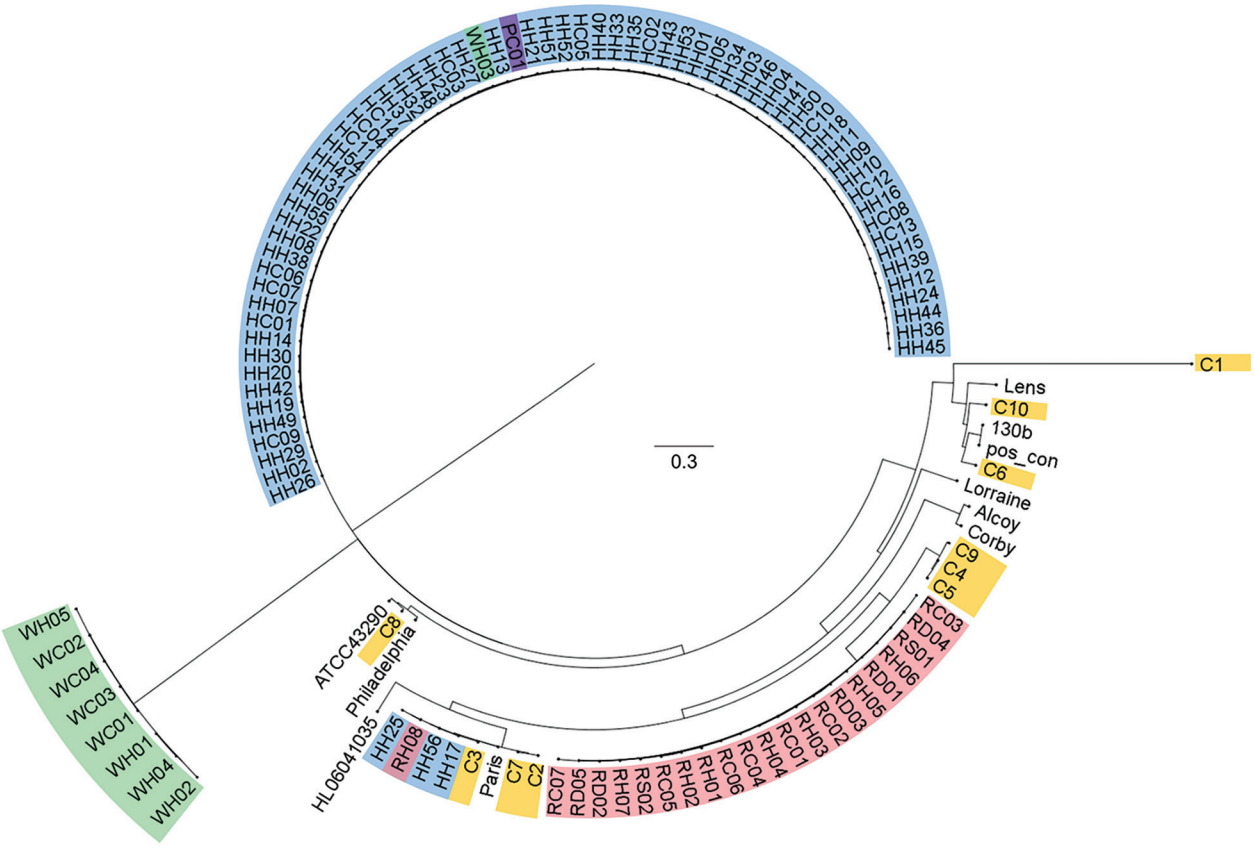


Figure. Single-nucleotide polymorphism (SNP) analysis of isolates from study of *Legionella pneumophila* in tap water, Flint, Michigan, USA. Analysis was conducted in kSNP3.0 (https://sourceforge.net/projects/ksnp/) and visualized by using FigTree 14.3 (https://github.com/rambaut/figtree/releases/tag/v1.4.3). Isolate sources: yellow, clinical samples; blue, hospital water; red, residence water; purple, public building water; green, buildings supplied by well water. With the exception of buildings supplied by well water, all buildings were serviced by Flint municipal water. Reference strains are detailed in Appendix 1 Table 2 (https://wwwnc.cdc.gov/EID/article/25/11/18-1032-App1.pdf). Scale bar indicates nucleotide substitutions per site.

L. pneumophila can colonize buildings and persist over multiple years (*24–26*). Thus, it is reasonable to assume that water isolates collected in 2016 were probably representative of strains colonizing building water systems over the previous months or even years.

Our study provides reasonable evidence that plumbing served by the Flint drinking water system was colonized by strains of *L. pneumophila* capable of causing Legionnaires' disease, particularly serogroup 1 and ST1. Although no epidemiologic links have been made between clinical cases and cooling tower exposures in these outbreaks, direct or indirect use of tap water (e.g., via feed to cooling towers) is possible. High degrees of similarity (2–1,062 SNPs) were noted between the ST1 isolates of clinical and water origin, a finding that was consistent with phylogenetic and average nucleotide identity analysis (Appendix 1 Figures 1–3; Appendix 2). The highest degree of similarity between clinical and water isolates was between C3 and RH08 (38 SNPs), HH25 (40 SNPs), HH17 (45 SNPs), and HH56 (46 SNPs). C2 differed from water isolates by 1,053–1,062 SNPs, and C7 differed from water isolates by 1,041–1,049 SNPs. With the exception of 1 isolate (WH03), isolates obtained from tap water from buildings never served by the Flint River were markedly distinct from those originating from residences or hospitals in Flint as well as the clinical isolates. The low number of SNPs between replicate genomes sequenced in this study (0–10) suggests that the extent to which technical variation in whole-genome sequencing contributed to observed sequence variation is low. Previous studies have documented that although some Legionnaires' disease outbreaks are characterized by *L. pneumophila* clinical strains that differ by as few as <5 SNPs, other outbreaks may differ by as many as 418 core SNPs (*27*). Thus, the SNP variability between water and clinical strains of ST1 in this study, particularly C3, is comparable to the documented range of variation in other outbreaks. In contrast, clinical strain C2 varied from the Paris reference strain by only 505, C3 by 371, and C7 by 491 SNPs. Therefore, ascertaining what level of SNP divergence between strains is demonstrative of a common source or virulent strain is

challenging. Given the well-established pathogenicity of the Paris strain, the results are also suggestive of genomic similarity among virulent strains of *Legionella*. Regardless, the similarity between C3 and strains isolated from Flint tap water samples (38–46 SNPs) is notable.

ST1 water isolates were collected from taps of a hospital and a residence, indicating that this ST seems to have been somewhat widespread in the water distribution system, spanning multiple Flint buildings. However, the presence of several distinct phylogenetic clades of *L. pneumophila* isolated from Flint water systems further demonstrates that a single strain of *L. pneumophila* did not dominate the system citywide. We hypothesize that this finding is likely the result of conditions favorable to *Legionella* growth, which we previously documented in the Flint system (*7*), facilitating the proliferation of multiple strains of *L. pneumophila* in different buildings and parts of the system. Similarly, the broad distribution of clinical isolates across 7 STs supports the hypothesis that any waterborne exposures that resulted in Legionnaires' disease could hypothetically have originated from a diverse array of *L. pneumophila* strains and exposure sources.

All clinical isolates characterized in this study belonged to *L. pneumophila* serogroup 1, which is identified as the cause of ≥57% of reported Legionnaires' cases in the United States (*6*). ST1 (belonging to serogroup 1) has been widely implicated in Legionnaires' outbreaks worldwide, including outbreaks in France (*28*), China (*29*), Germany (*30*), Canada (*31*), and the United States (*32*). In the United States, ST1 is thought to be both the most common cause of sporadic Legionnaires' disease cases and the most common waterborne ST found in potable and nonpotable water (*32*). ST1 isolates are highly conserved at the nucleotide level (*33*), making it challenging to link clinical cases with environmental sources because of the prevalence of ST1 and lack of genetic variability.

Water isolates belonging to serogroup 6, all classified as ST2518, were widespread in samples collected from a Flint hospital in March 2016. A study of *L. pneumophila* isolates collected from Flint tap water in September and October 2016 also found that serogroup 6 isolates were widespread in residential premise plumbing water samples, although these isolates all belonged to STs 367 and 461 (*34*). Byrne and colleagues found that serogroup 6 strains were at least as infectious for macrophages as a known virulent laboratory strain, emphasizing the potential for Legionnaires' disease to be caused by strains other than serogroup 1 (*34*), although more research is needed to confirm the relevance of serogroup 6 strains for human infectivity. In our study, none of the clinical strains available for analysis were serogroup 6.

It is noteworthy that 19% of hot water and 12% of cold water taps were positive for culturable *L. pneumophila*. Although *L. pneumophila* typically multiplies at 25°C–37°C (*35*) and prospers in hot water plumbing systems (*36*), it has also been widely documented in cold water taps; one molecular analysis–based study found that as many as 47% of surveyed cold water taps were positive for genes specific to *L. pneumophila* serogroup 1 (*37*).

When MDHHS recently conducted an epidemiologic characterization of the Genesee County Legionnaires' disease cases recorded in 2014 and 2015, although a lack of clinical isolates hampered a comprehensive investigation, they found that exposures that occurred at 1 Flint hospital potentially explained most cases (*1–3*). Our study provides complementary whole-genome sequencing–based characterization of clinical isolates and tap water *L. pneumophila* isolates collected after the Flint outbreaks. Notably, we found a high degree of similarity between 4 water isolates originating from Flint tap water and 3 regional clinical strains known to cause Legionnaires' disease. Our study also established that a variety of *L. pneumophila* strains were culturable from Flint tap water and that they tended to cluster genetically by residence versus hospital origin. Likewise, we found notable diversity among clinical strains, spanning 7 STs. Thus, multiple *L. pneumophila* strains were associated with the Flint 2014–2015 Legionnaires' outbreaks, potentially resulting from multiple sources of exposure, although further epidemiologic investigation is needed to identify whether multiple sources were involved and whether there were any common sources of exposure. Although we did not intend for this study to provide an epidemiologic analysis of precise sources of *Legionella* exposure for Legionnaires' patients, our publicly available data could support such studies in the future.

Acknowledgments

We thank the members of the Flint Water Study Team at Virginia Polytechnic Institute and State University who volunteered their time to collect samples and the Flint citizens and businesses that supported this study. We thank Joan Rose for providing access to the laboratory at Michigan State University to support the fieldwork.

This study was supported by the US National Science Foundation (RAPID award no. 1556258), a Graduate Research Fellowship Program Grant (DGE 0822220), and supplementary funding (associated with award no. 1336650). Additional support was provided by the Alfred P. Sloan Foundation Microbiology of the Built Environment program, the Michigan Department of Environmental Quality for a study of effects of flushing residential hot water heaters (June 2016), the American Water Works Association Abel Wolman Fellowship, and the Institute for Critical Technology and Applied Science and Engineering of the Exposome Center at Virginia Polytechnic Institute and State University.

In addition to the funding sources noted, author M.A.E. contributed discretionary research funding to support this study. No funds were received from the State of Michigan or MDHHS for this study. M.A.E. was subpoenaed to testify by the defense in the felony case against MDHHS employees Eden Wells and Nick Lyon in March 2018 and did so without compensation. In addition, coauthors M.A.E. and A.P. were principal investigators on a $263,000 subcontract to the National Sanitation Foundation on a project funded by the State of Michigan to study point-of-use filters, but this project was canceled by the National Sanitation Foundation midway through the study without any samples having been collected or processed.

About the Author

Dr. Garner is an assistant professor of civil and environmental engineering at West Virginia University. Her research focuses on the dissemination of opportunistic pathogens and antibiotic-resistant bacteria in water systems, with a particular emphasis on water sustainability and water reuse.

References

1. Michigan Department of Health and Human Services. Legionellosis outbreak—Genesee County, June, 2014–March, 2015 full analysis. 2016 [cited 2018 Jun 22]. https://www.michigan.gov/documents/mdhhs/6-14_to_3-15_Legionellosis_Report_Full_Analysis_Results_511708_7.pdf
2. Michigan Department of Health and Human Services. Legionellosis outbreak—Genesee County, May 2015–October 2015 full analysis. 2016 [cited 2018 Jun 22]. https://gchd.us/wp-content/uploads/2016/08/Updated-LD-Report-4-11-16.pdf
3. Michigan Department of Health and Human Services. Epidemiology of Legionnaires' disease in Genesee County, Michigan, 2014–2017 chart book. 2018 [cited 2018 Jun 22]. https://content.govdelivery.com/attachments/MIDHHS/2018/05/29/file_attachments/1014713/GC%2BLegionella%2BChartBook%2BFINAL.pdf
4. Pieper KJ, Martin R, Tang M, Walters L, Parks J, Roy S, et al. Evaluating lead in water levels during the Flint Water Crisis and associated remediation interventions. Environ Sci Technol. 2017;52:8124–32. https://doi.org/10.1021/acs.est.8b00791
5. Dooling KL, Toews K, Hicks LA, Garrison LE, Bachaus B, Zansky S. Active Bacterial Core Surveillance for Legionellosis—United States, 2011–2013. MMWR Morb Mortal Wkly Rep. 2015;64:2011–3.
6. Marston BJ, Lipman HB, Breiman RF. Surveillance for Legionnaires' disease. Risk factors for morbidity and mortality. Arch Intern Med. 1994;154:2417–22. https://doi.org/10.1001/archinte.1994.00420210049006
7. Rhoads WJ, Garner E, Ji P, Zhu N, Parks J, Schwake DO, et al. Distribution system operational deficiencies coincide with reported Legionnaires' disease clusters in Flint, Michigan. Environ Sci Technol. 2017;51:11986–95. https://doi.org/10.1021/acs.est.7b01589
8. Schwake DO, Garner E, Strom OR, Pruden A, Edwards MA. *Legionella* DNA markers in tap water coincident with a spike in Legionnaires' disease in Flint, MI. Environ Sci Technol Lett. 2016;3:311–5. https://doi.org/10.1021/acs.estlett.6b00192
9. Zahran S, McElmurry SP, Kilgore PE, Mushinski D, Press J, Love NG, et al. Assessment of the Legionnaires' disease outbreak in Flint, Michigan. Proc Natl Acad Sci U S A. 2018;115:E1730–9. https://doi.org/10.1073/pnas.1718679115
10. Michigan Department of Health and Human Services. MDHHS response to Flint Area Community Health and Environment Partnership, Proceedings of the National Academy of Sciences Article. February 5, 2018 [cited 2018 Jun 22]. https://www.michigan.gov/documents/mdhhs/MDHHS_Response_to_FACHEP_Proceedings_of_the_National_Academy_of_Sciences_Article_FINAL_613088_7.pdf
11. Sabria M, Yu VL. Hospital-acquired legionellosis: solutions for a preventable infection. Lancet Infect Dis. 2002;2:368–73. https://doi.org/10.1016/S1473-3099(02)00291-8
12. Hicks LA, Garrison LE, Nelson GE, Hampton LM; Centers for Disease Control and Prevention (CDC). Legionellosis—United States, 2000–2009. MMWR Morb Mortal Wkly Rep. 2011;60:1083–6.
13. Centers for Disease Control and Prevention, Procedures for the recovery of *Legionella* from the environment [cited 2018 Jun 22]. https://www.epa.gov/sites/production/files/2018-11/documents/cdc_legionella_method.pdf
14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170
15. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77. https://doi.org/10.1089/cmb.2012.0021
16. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. 2007; 57:81–91.
17. Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics. 2015;31:2877–8. https://doi.org/10.1093/bioinformatics/btv271
18. Gaia V, Fry NK, Afshar B, Lu PC, Etienne J, Peduzzi R, et al. Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumophila*. J Clin Microbiol. 2005;43:2047–52.
19. Gordon M, Yakunin E, Valinsky L, Chalifa-Caspi V, Moran-Gilad J; ESCMID Study Group for Legionella Infections. A bioinformatics tool for ensuring the backwards compatibility of *Legionella pneumophila* typing in the genomic era. Clin Microbiol Infect. 2017;23:306–10. https://doi.org/10.1016/j.cmi.2017.01.002
20. Mérault N, Rusniok C, Jarraud S, Gomez-Valero L, Cazalet C, Marin M, et al.; DELPH-I Study Group. Specific real-time PCR for simultaneous detection and identification of *Legionella pneumophila* serogroup 1 in water and clinical samples. Appl Environ Microbiol. 2011;77:1708–17. https://doi.org/10.1128/AEM.02261-10
21. Rodriguez-R LM, Konstantinidis KT. Bypassing cultivation to identify bacterial species: culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species [cited 2019 Sep 4]. https://www.asmscience.org/content/journal/microbe/10.1128/microbe.9.111.1
22. Berkelman RL, Pruden A. Prevention of Legionnaires' disease in the 21st century by advancing science and public health practice. Emerg Infect Dis. 2017;23:1905–7. https://doi.org/10.3201/eid2311.171429
23. Phin N, Parry-Ford F, Harrison T, Stagg HR, Zhang N, Kumar K, et al. Epidemiology and clinical management of Legionnaires' disease. Lancet Infect Dis. 2014;14:1011–21. https://doi.org/10.1016/S1473-3099(14)70713-3
24. Perola O, Kauppinen J, Kusnetsov J, Kärkkäinen UM, Lück PC, Katila ML. Persistent *Legionella pneumophila* colonization of a

hospital water supply: efficacy of control methods and a molecular epidemiological analysis. APMIS. 2005;113:45–53. https://doi.org/10.1111/j.1600-0463.2005.apm1130107.x
25. Rangel-Frausto MS, Rhomberg P, Hollis RJ, Pfaller MA, Wenzel RP, Helms CM, et al. Persistence of *Legionella pneumophila* in a hospital's water system: a 13-year survey. Infect Control Hosp Epidemiol. 1999;20:793–7. https://doi.org/10.1086/501586
26. Scaturro M, Fontana S, Dell'eva I, Helfer F, Marchio M, Stefanetti MV, et al. A multicenter study of viable PCR using propidium monoazide to detect *Legionella* in water samples. Diagn Microbiol Infect Dis. 2016;85:283–8. https://doi.org/10.1016/j.diagmicrobio.2016.04.009
27. Raphael BH, Baker DJ, Nazarian E, Lapierre P, Bopp D, Kozak-muiznieks NA, et al. Genomic resolution of outbreak-associated *Legionella pneumophila* serogroup 1 isolates from New York State. Appl Environ Microbiol. 2016;82:3582–90.
28. Ginevra C, Jacotin N, Diancourt L, Guigon G, Arquilliere R, Meugnier H, et al. *Legionella pneumophila* sequence type 1/Paris pulsotype subtyping by spoligotyping. J Clin Microbiol. 2012;50:696–701. https://doi.org/10.1128/JCM.06180-11
29. Qin T, Zhang W, Liu W, Zhou H, Ren H, Shao Z, et al. Population structure and minimum core genome typing of *Legionella pneumophila.* Sci Rep. 2016;6:21356. https://doi.org/10.1038/srep21356
30. Borchardt J, Helbig JH, Lück PC. Occurrence and distribution of sequence types among *Legionella pneumophila* strains isolated from patients in Germany: common features and differences to other regions of the world. Eur J Clin Microbiol Infect Dis. 2008;27:29–36. https://doi.org/10.1007/s10096-007-0392-3
31. Reimer AR, Au S, Schindle S, Bernard KA. *Legionella pneumophila* monoclonal antibody subgroups and DNA sequence types isolated in Canada between 1981 and 2009: Laboratory Component of National Surveillance. Eur J Clin Microbiol Infect Dis. 2010;29:191–205. https://doi.org/10.1007/s10096-009-0840-3
32. Kozak-Muiznieks NA, Morrison SS, Mercante JW, Ishaq MK, Johnson T, Caravas J, et al. Comparative genome analysis reveals a complex population structure of *Legionella pneumophila* subspecies. Infect Genet Evol. 2018;59:172–85. https://doi.org/10.1016/j.meegid.2018.02.008
33. Mercante JW, Caravas JA, Ishaq MK, Kozak-Muznieks NA, Raphael BH, Winchell JM. Genomic heterogeneity differentiates clinical and environmental subgroups of Legionella pneumophila sequence type 1. PLoS One. 2018;13:e0206110
34. Byrne BG, McColm S, McElmurry SP, Kilgore PE, Sobeck J, Sadler R, et al. Prevalence of infection-competent serogroup 6 *Legionella pneumophila* within premise plumbing in southeast Michigan. MBio. 2018;9:e00016–8. https://doi.org/10.1128/mBio.00016-18
35. Wadowsky RM, Wolford R, McNamara AM, Yee RB. Effect of temperature, pH, and oxygen level on the multiplication of naturally occurring *Legionella pneumophila* in potable water. Appl Environ Microbiol. 1985;49:1197–205.
36. Rhoads WJ, Ji P, Pruden A, Edwards MA. Water heater temperature set point and water use patterns influence *Legionella pneumophila* and associated microorganisms at the tap. Microbiome. 2015;3:1–13.
37. Donohue MJ, O'Connell K, Vesper SJ, Mistry JH, King D, Kostich M, et al. Widespread molecular detection of *Legionella pneumophila* serogroup 1 in cold water taps across the United States. Environ Sci Technol. 2014;48:3145–52. https://doi.org/10.1021/es4055115

Address for correspondence: Amy Pruden, Virginia Polytechnic Institute and State University, Civil and Environmental Engineering, 418 Durham Hall, 1145 Perry St, Blacksburg, VA 24061-0131, USA; email: apruden@vt.edu

Article DOI: https://doi.org/10.3201/eid2511.181032

Comparison of Whole-Genome Sequences of *Legionella pneumophila* in Tap Water and in Clinical Strains, Flint, Michigan, USA, 2016

Appendix 1

Materials and Methods

Sample Collection and Preservation

Samples were collected from all taps into sterile polypropylene bottles (Nalgene, Rochester, NY) with 24 mg of sodium thiosulfate per liter added as a chlorine quenching agent. All samples were transported to the lab at room temperature and processed within ≈30 hours of sampling.

Aliquots (250–500 mL) were filter-concentrated onto a sterile 0.22 μm pore size mixed-cellulose ester membrane (Millipore, Billerica, MA) and resuspended in 5 mL sterile tap water before culturing *Legionella* according to standard methods (*1*). *L. pneumophila* colonies were streaked to isolation to obtain pure cultures. DNA was extracted from *Legionella* cultures by resuspending colonies in 50 μl of molecular grade water, freezing at –20°C, and rapidly thawing at 90°C for 10 minutes. Extracts were centrifuged at 10,000×g for 5 minutes to remove cell debris. Quantities of *Legionella* spp. and *L. pneumophila* gene markers from these samples have been published previously (*2*,*3*).

Whole Genome Sequence Analysis

16S rRNA gene sequences were extracted from sequence data using the Rapid Annotations Using Subsystem Technology server (*4*) and *Legionella* species assignments were determined via BLASTn of the sequence against the NCBI nucleotide database via the web server. Phylogenetic trees were generated using FastTree (*5*) based on extracted 16S rRNA gene sequences and 37 single-copy housekeeping genes in nucleotide space and amino acid space

using PhyloSift (*6*). ANI was calculated as previously described (*7*) and SNPs were identified using kSNP3.0 (*8*) with maximum likelihood estimation. Nine known *L. pneumophila* genomes associated with previous LD outbreaks were included in the analysis as reference strains for comparison (Table S2).

Sequence-based typing was performed targeting the *flaA, pilE, asd, mip, mompS, proA,* and *neuA* alleles (*9*) using the mompS tool (*10*). This bioinformatics tool facilitates backward compatibility of whole genome sequence typing by ensuring that the *momp*S allele, of which *L. pneumophila* can carry multiple non-identical copies, is typed based on the copy of the gene that would be identified by the traditional PCR-based typing scheme. STs were identified from allele profiles using the European Working Group for *Legionella* Infections database for *L. pneumophila*.

Serogroup Analysis

L. pneumophila isolate genomes belonging to serogroup 1 were identified via detection of the *wzm* gene (*11*) in whole genomes using BLAST with a minimum nucleotide identity of 98% and e-value of 1e-5. DNA sequence-based classifications were verified and unknown serogroups were determined using direct fluorescent antibody staining with FITC-conjugated antibodies (m-TECH, Milton, GA). To address problems with non-specific binding when stained cells were prepared according to manufacturer instructions, the protocol was modified as follows: isolates grown in buffered yeast extract broth (per liter: 10 g yeast extract, 1 g α ketoglutaric acid, 10 g 2-(carbamoylmethylamino)ethanesulfonic acid, 0.4 g L-cystine monohydrochloride, 0.25 g ferric pyrophosphate) were centrifuged at 5,000xg and resuspended in 1X phosphate buffered saline (PBS). To separate 25 µl aliquots of cells suspended in PBS, 5 µl of each FITC-conjugated antibodies were added and the suspension was incubated at 20°C for 30 minutes. Cells were washed with 1X PBS three times, then viewed with an AxioSkop2 plus fluorescence microscope (Carl Zeiss Microscopy, Oberkocken, Germany).

References

1. United States Centers for Disease Control and Prevention. Procedures for the Recovery of *Legionella* from the Environment. 2005.

2. Rhoads WJ, Garner E, Ji P, Zhu N, Parks J, Schwake DO, et al. Distribution System Operational Deficiencies Coincide with Reported Legionnaires' Disease Clusters in Flint, Michigan. Environ Sci Technol. 2017;51:11986–95. PubMed https://doi.org/10.1021/acs.est.7b01589

3. Schwake DO, Garner E, Strom OR, Pruden A, Edwards MA. *Legionella* DNA Markers in Tap Water Coincident with a Spike in Legionnaires' Disease in Flint, MI. Environ Sci Technol Lett. 2016;3:311–5. https://doi.org/10.1021/acs.estlett.6b00192

4. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75. PubMed https://doi.org/10.1186/1471-2164-9-75

5. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:e9490. PubMed https://doi.org/10.1371/journal.pone.0009490

6. Darling AE, Jospin G, Lowe E, Matsen FA IV, Bik HM, Eisen JA. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ. 2014;2:e243. PubMed https://doi.org/10.7717/peerj.243

7. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA – DNA hybridization values and their relationship to whole-genome sequence similarities. 2017;57:81–91.

8. Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics. 2015;31:2877–8. PubMed https://doi.org/10.1093/bioinformatics/btv271

9. Gaia V, Fry NK, Afshar B, Lu PC, Etienne J, Peduzzi R, et al. Consensus Sequence-Based Scheme for Epidemiological Typing of Clinical and Environmental Isolates of *Legionella pneumophila*. 2005;43(5):2047–52.

10. Gordon M, Yakunin E, Valinsky L, Chalifa-Caspi V, Moran-Gilad J; ESCMID Study Group for Legionella Infections. A bioinformatics tool for ensuring the backwards compatibility of *Legionella pneumophila* typing in the genomic era. Clin Microbiol Infect. 2017;23:306–10. PubMed https://doi.org/10.1016/j.cmi.2017.01.002

11. Mérault N, Rusniok C, Jarraud S, Gomez-Valero L, Cazalet C, Marin M, et al.; DELPH-I Study Group. Specific real-time PCR for simultaneous detection and identification of *Legionella pneumophila* serogroup 1 in water and clinical samples. Appl Environ Microbiol. 2011;77:1708–17. PubMed https://doi.org/10.1128/AEM.02261-10

Appendix Table 1. Summary of environmental and clinical *L. pneumophila* isolates subject to whole genome sequencing. All clinical isolates were provided by the Michigan Department of Health and Human Services. All clinical isolates are from patients who resided outside of Flint.

SampleID[a]	Year Collected	Month Collected	Isolate Type	Building Type[b]	Sample Tap or Source	Flushed/ Stagnant	Water Source (April 2014-October 2015)	ST	SG
C1	2015		Clinical					159	1
C2*	2015		Clinical					1	1
C3*	2015		Clinical					1	1
C4	2015		Clinical					213	1
C5	2015		Clinical					213	1
C6	2015		Clinical					44	1
C7	2015		Clinical					1	1
C8	2015		Clinical					211	1
C9*	2015		Clinical					222	1
C10	2015		Clinical					2513	1
HC01	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC02	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC03	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC04	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC05	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC06	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC07	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC08	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC09	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC10	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC11	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC12	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC13	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HC14	2016	March	Water	Hospital	Cold	Stagnant	Flint	2518	6[d]
HH01	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[c]
HH02	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH03	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH04	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[c]
HH05	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH06	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH07	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH08	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH09	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH10	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH11	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH12	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH13	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[c]
HH14	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH15	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH16	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH17	2016	March	Water	Hospital	Hot	Stagnant	Flint	1	1
HH18	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH19	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH20	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH21	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH22	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH23	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH24	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH25	2016	March	Water	Hospital	Hot	Stagnant	Flint	1	1
HH26	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH27	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH28	2016	March	Water	Hospital	Hot	Stagnant	Flint	ND	ND
HH29	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH30	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH31	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH32	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH33	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH34	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH35	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH36	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH37	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH38	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH39	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH40	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]

SampleID[a]	Year Collected	Month Collected	Isolate Type	Building Type[b]	Sample Tap or Source	Flushed/ Stagnant	Water Source (April 2014-October 2015)	ST	SG
HH41	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH42	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH43	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH44	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH45	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH46	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH47	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH48	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH49	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH50	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH51	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[c]
HH52	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH53	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH54	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH55	2016	March	Water	Hospital	Hot	Stagnant	Flint	2518	6[d]
HH56	2016	March	Water	Hospital	Hot	Stagnant	Flint	1	1
PC01	2016	March	Water	Public building	Cold	Stagnant	Flint	2518	6[c]
RC01	2016	June	Water	Residence	Cold	Stagnant	Flint	192	1
RC02	2016	June	Water	Residence	Cold	Flushed	Flint	192	1
RC03	2016	August	Water	Residence	Cold	Stagnant	Flint	192	1
RC04	2016	June	Water	Residence	Cold	Stagnant	Flint	192	1
RC05	2016	June	Water	Residence	Cold	Stagnant	Flint	2514[c]	1
RC06	2016	June	Water	Residence	Cold	Stagnant	Flint	192	1
RC07	2016	June	Water	Residence	Cold	Flushed	Flint	192	1
RD01	2016	June	Water	Residence	HWHDV	NA	Flint	192	1
RD02	2016	June	Water	Residence	HWHDV	NA	Flint	192	1
RD03	2016	June	Water	Residence	HWHDV	NA	Flint	192	1
RD04	2016	June	Water	Residence	HWHDV	NA	Flint	192	1
RD05	2016	June	Water	Residence	HWHDV	NA	Flint	192	1
RH01	2016	June	Water	Residence	Hot	Flushed	Flint	ND	1
RH02	2016	June	Water	Residence	Hot	Stagnant	Flint	192	1
RH03	2016	June	Water	Residence	Hot	Flushed	Flint	192	1
RH04	2016	June	Water	Residence	Hot	Flushed	Flint	192	1
RH05	2016	June	Water	Residence	Hot	Stagnant	Flint	192	1
RH06	2016	June	Water	Residence	Hot	Stagnant	Flint	192	1
RH07	2016	June	Water	Residence	Hot	Stagnant	Flint	192	1
RH08	2016	August	Water	Residence	Hot	Stagnant	Flint	1	1
RS01	2016	June	Water	Residence	Shower	Stagnant	Flint	192	1
RS02	2016	August	Water	Residence	Shower	Stagnant	Flint	192	1
RS03	2016	June	Water	Residence	Shower	Stagnant	Flint	ND	ND
WC01	2016	March	Water	Well Water	Cold	Stagnant	well	NA	NA
WC02	2016	March	Water	Well Water	Cold	Stagnant	well	NA	NA
WC03	2016	March	Water	Well Water	Cold	Stagnant	well	NA	NA
WC04	2016	March	Water	Well Water	Cold	Stagnant	well	NA	NA
WH01	2016	March	Water	Well Water	Hot	Stagnant	well	NA	NA
WH02	2016	March	Water	Well Water	Hot	Stagnant	well	NA	NA
WH03	2016	March	Water	Well Water	Hot	Stagnant	well	2518	6[d]
WH04	2016	March	Water	Well Water	Hot	Stagnant	well	NA	NA
WH05	2016	March	Water	Well Water	Hot	Stagnant	well	NA	NA
pos_con			+ control					42	1
pos_con			+ control					42	1
neg_con			- control					NA	NA
neg_con			- control					NA	NA

[a]Isolates were named according to the following system: First letter indicates building type/location (H = hospital; R = residence; W = school using well water; p = large public building), second letter indicates sample collection location (hot water tap (H), cold water tap (C),water heater drain valve (D), shower (S)), followed by a unique numeric identifier. Clinical strains are denoted C1–10.

[b]Unless otherwise indicated, all buildings were serviced by Flint municipal water derived from the Flint River during the Flint Water Crisis

[c]Verified serogroup 6 using direct fluorescent antibody staining

[d]Presumed serogroup 6 based on direct fluorescent antibody staining of a phylogenetically diverse subset of isolates belonging to ST 2518

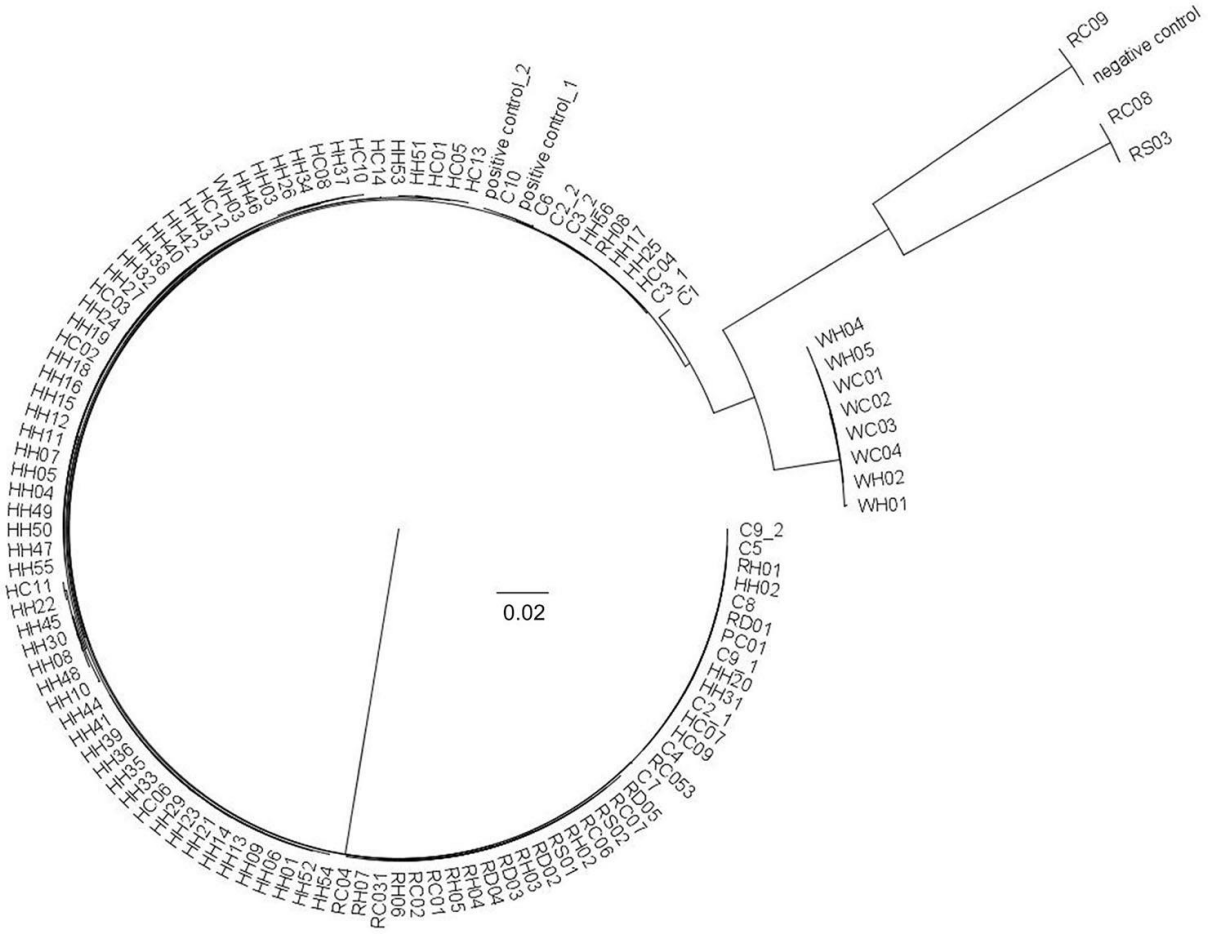
*indicates isolate was prepared and sequenced twice with consistent results as an additional control.

HWHDV = hot water heater drain valve; ST = sequence type; SG = serogroup

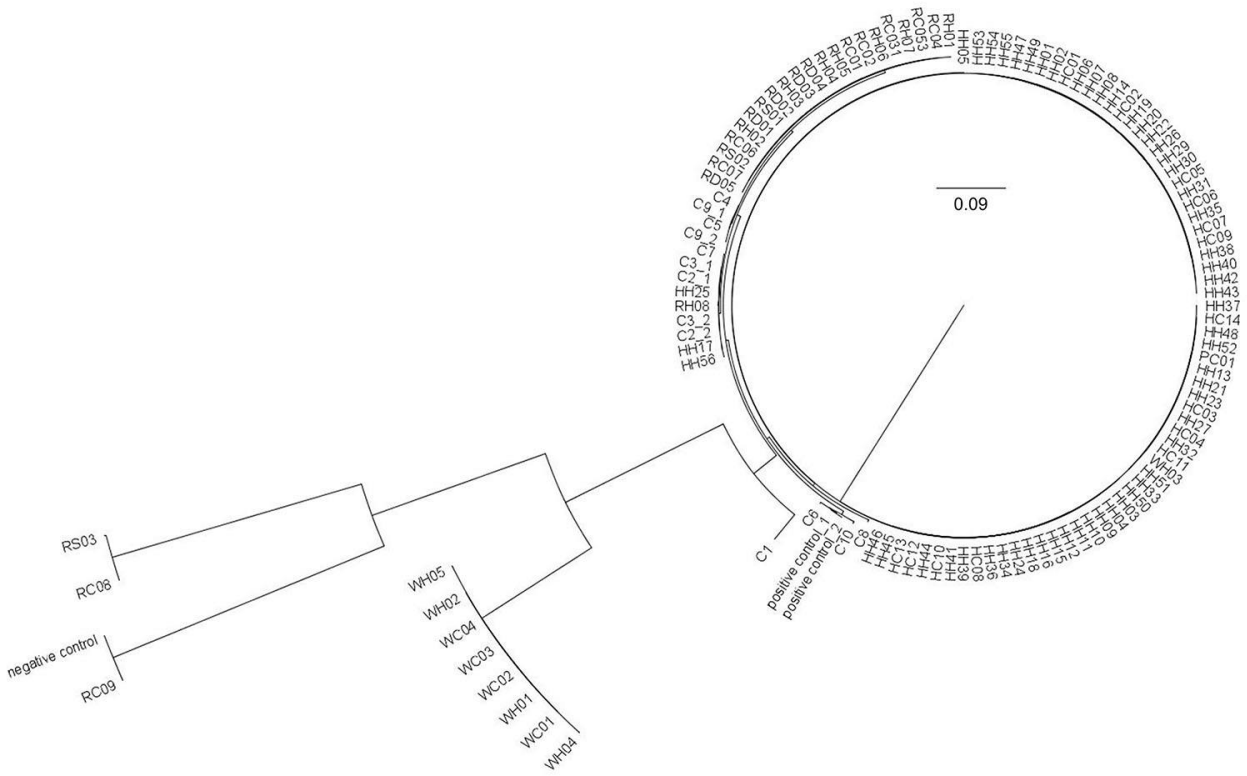
ND = could not be determined due to insufficient genome coverage; NA = not applicable

Appendix Table 2. Clinical reference strains selected for comparison to water isolates.

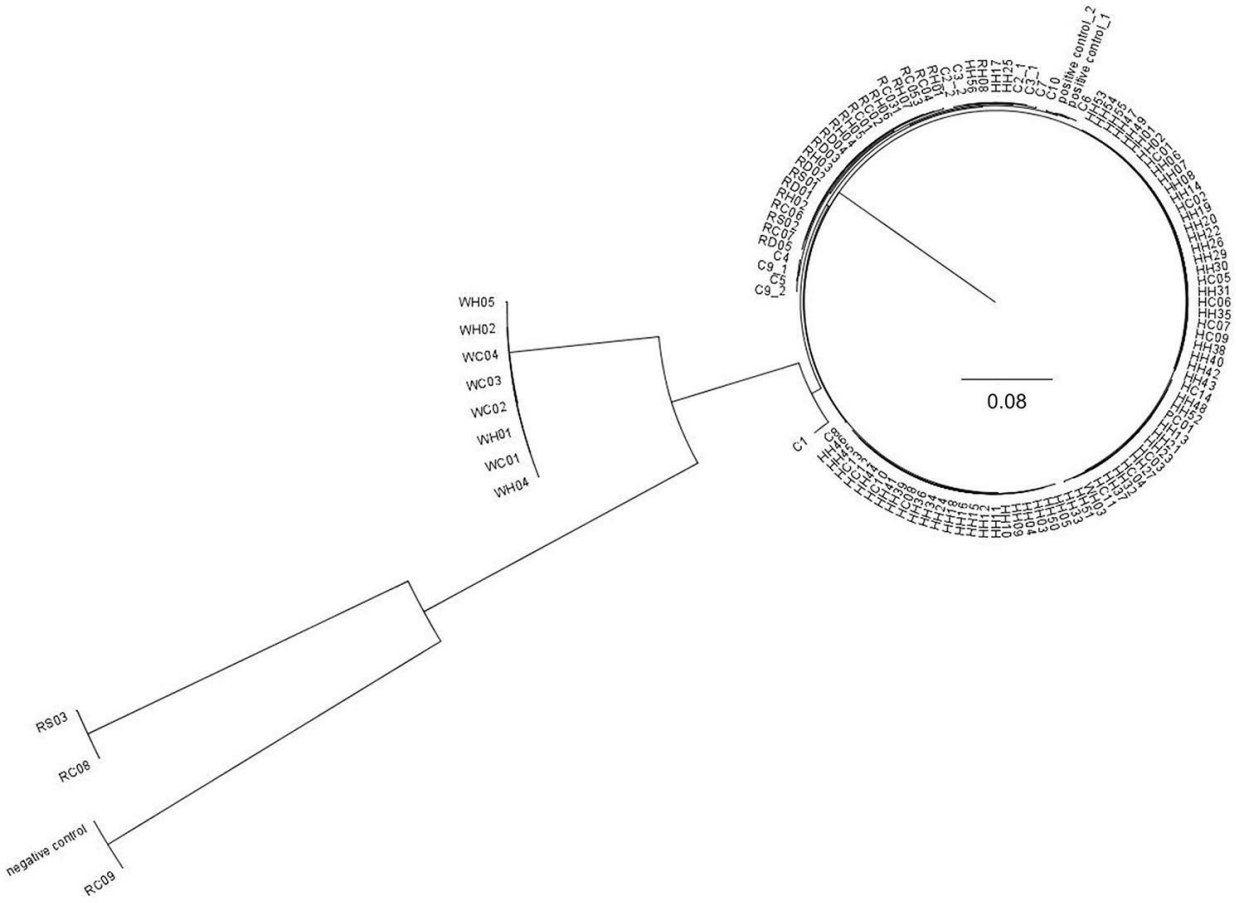
Sample ID	GenBank Accession Number	Origin	Serogroup	Sequence Type
LP Philadelphia	AE017354.1	USA	1	ST-136
LP ATCC 43290	CP003192.1	USA	12	ST-187
LP Alcoy	CP001828.1	Spain	1	ST-578
LP Corby	CP000675.2	UK	1	ST-51
LP Lens	CR628337.1	France	1	ST-15
LP 130b	FR687201.1	USA	1	ST-42
LP Paris	CR628336.1	France	1	ST-1
LP Lorraine	FQ958210.1	France	1	ST-47
LPHL06041035	FQ958211.1	France	1	ST-734



Appendix Figure 1. Phylogenetic tree generated using FastTree (Price et al. 2010) based on extracted 16S rRNA gene sequences using PhyloSift (Darling et al. 2014). Sample names appended with "_1" and "_2" represent isolates sequenced in duplicate on two different MiSeq runs.



Appendix Figure 2. Phylogenetic tree generated using FastTree (Price et al. 2010) based on 37 single-copy housekeeping genes in amino acid space using PhyloSift (Darling et al. 2014). Sample names appended with "_1" and "_2" represent isolates sequenced in duplicate on two different MiSeq runs.



Appendix Figure 3. Phylogenetic tree generated using FastTree (Price et al. 2010) based on 37 single-copy housekeeping genes in nucleotide space using PhyloSift (Darling et al. 2014). Sample names appended with "_1" and "_2" represent isolates sequenced in duplicate on two different MiSeq runs.