Infection, Genetics and Evolution 65 (2018) 200–209

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Insights into the long-term persistence of *Legionella* in facilities from whole-genome sequencing

Megan Wells[a,1], Erica Lasek-Nesselquist[b,*,1], Dianna Schoonmaker-Bopp[b], Deborah Baker[b], Lisa Thompson[b], Danielle Wroblewski[b], Elizabeth Nazarian[b], Pascal Lapierre[b], Kimberlee A. Musser[b]

[a] Union College, 807 Union St., Schenectady, NY 12308, USA
[b] Wadsworth Center, New York State Department of Health, 120 New Scotland Ave., Albany, NY 12208, USA

ARTICLE INFO

Keywords:
Legionella pneumophila
Legionella micdadei
Long-term persistence
Whole-genome sequencing analysis (WGS)
Substitution rate
Evolution

ABSTRACT

We investigated the value of whole-genome sequencing (WGS) and single nucleotide polymorphism (SNP) analyses in determining the relationships among and evolutionary rates of *Legionella* species with long-term persistence in three healthcare facilities. We examined retrospective clinical and environmental isolates of *Legionella micdadei* and *Legionella pneumophila* serogroup 1 isolates with identical PFGE DNA fingerprints sampled over the course of up to 18 years. WGS analyses demonstrated that heterogeneous populations of *Legionella* were present within each facility despite displaying the same PFGE profiles. Additionally, clustering of some clinical isolates with those from a separate but related institution exposed a source of infection not previously detected, underscoring the importance of considering phylogenetic relationships when assessing epidemiological links. The data supported an average substitution rate of 0.80 SNPs per genome per year for L. *micdadei* but a reliable estimate for L. *pneumophila* serogroup 1 could not be obtained due to complicating factors such as non-chronological links among isolates and inadequate sampling depths. While the substitution rate for L. *micdadei* is consistent with previous estimates for L. *pneumophila*, the lack of a temporal signal in our sequence data for L. *pneuomphila* serogroup 1 isolates suggests either insufficient change to provide an estimate or variable evolutionary rates, which could reflect the presence of both actively dividing and viable but non-culturable *Legionella* spp. in the built environment. This study highlights the increased discriminatory power of WGS SNP analysis as compared to PFGE, emphasizes the need for extended sampling, and provides insight into the evolution of *Legionella* from longitudinal investigations.

1. Introduction

Legionella are gram-negative, rod shaped, fastidious bacteria that cause Legionellosis (Nazarian et al., 2015). Legionellosis describes two diseases: Pontiac fever (Nazarian et al., 2015; Bartram et al., 2007) and Legionnaires' disease (LD) - a serious, potentially fatal pneumonia-like disease caused by pathogenic *Legionella* species. There are 62 known species of *Legionella* (Whiley et al., 2014) but *Legionella pneumophila* serogroup 1 is the most common pathogenic serogroup, related to 80–90% of LD cases (Yu et al., 2002; Fields et al., 2002). *Legionella micdadei* is the next most frequent cause of LD in the United States (Reingold et al., 1984).

Legionella live ubiquitously in fresh and salt water and have been found consistently in water sources within the built environment, which includes non-potable cooling towers and potable water in hospitals, hotels, and nursing homes (Prussin et al., 2017). *Legionella* are environmentally-derived pathogens that become infectious when contaminated water is aerosolized and inhaled into the lungs (Nazarian et al., 2015; Whiley et al., 2014). A challenge to water maintenance is that *Legionella* can survive in biofilms that provide nutrients and protection for extended periods of time (Berjeaud et al., 2016; Rogers et al., 1994). *Legionella* can also exist as parasites of amoebae, two ciliates, and a slime mold, which serve as defense shields from antimicrobial agents, disinfection via chlorination, and unfavorable environmental conditions (Lau and Ashbolt, 2009; Rowbotham, 1980). A combination of these factors allows *Legionella* to persist in the built environment (Berjeaud et al., 2016) with the co-evolution of *Legionella* and protist hosts likely also facilitating the ability of this microbe to infect humans

* Corresponding author.
E-mail address: Erica.lasek-nesselquist@health.ny.gov (E. Lasek-Nesselquist).
[1] Co-first authors.

https://doi.org/10.1016/j.meegid.2018.07.040
Received 23 May 2018; Received in revised form 27 July 2018; Accepted 30 July 2018
Available online 31 July 2018

Table 1
Legionella isolates examined in each facility. UNK, unknown; CLN, clinical isolate; ENV, environmental isolate; RM, room; ERM, exam room; PRM, patient room; Date, date of collection or data sample was received; SRA ID, SRA accession number. The PFGE profile is the NYS-specific PFGE pattern designation.

Facility	Sample ID	Source	Date	PFGE profile	SRA ID
1	1-A	UNK	8/17/99	LmiS13001	SAMN09744557
1	1-B	CLN – Bronchial	3/26/04	LmiS13001	SAMN09727221
1	1-C	ENV – Sink hot water	3/26/04	LmiS13001	SAMN09727222
1	1-D	ENV – Return line water	3/26/04	LmiS13001	SAMN09727223
1	1-E	ENV – Sink hot water	3/26/04	LmiS13001	SAMN09727224
1	1-F	ENV – Return line flush	3/26/04	LmiS13001	SAMN09727225
1	1-G	ENV – High zone heater	3/26/04	LmiS13001	SAMN09744558
1	1-H	CLN - Bronchoscopy	8/16/04	LmiS13001	SAMN09744559
1	1-I	CLN - Sputum	3/11/05	LmiS13001	SAMN09744560
1	1-J	CLN – Bronchoscopy	3/30/05	LmiS13001	SAMN09727226
1	1-K	CLN – Tracheal aspirate	3/13/05	LmiS13001	SAMN09727227
1	1-L	CLN – Sputum	3/31/05	LmiS13001	SAMN09727228
1	1-M	CLN – Sputum	4/25/05	LmiS13001	SAMN09727229
1	1-N	CLN – Sputum	10/28/06	LmiS13001	SAMN09727230
1	1-O	CLN – Bronchoscopy	10/30/06	LmiS13001	SAMN09744561
1	1-P	ENV – Water	11/15/06	LmiS13001	SAMN09727231
1	1-Q	ENV – Water	11/15/06	LmiS13001	SAMN09744562
1	1-R	ENV – Water	11/15/06	LmiS13001	SAMN09727232
1	1-S	ENV - Water	11/15/06	LmiS13001	SAMN09727233
1	1-T	CLN - Sputum	7/16/08	LmiS13001	SAMN09727234
2	2-A	CLN – Sputum	9/2/03	LpnS13041	SAMN09727235
2	2-B	CLN – Sputum	9/2/04	LpnS13041	SAMN09727236
2	2-C	CLN – Sputum	1/4/05	LpnS13041	SAMN09727237
2	2-D	CLN – Bronchial	1/6/05	LpnS13041	SAMN09727238
2	2-E	CLN – Sputum	1/30/06	LpnS13041	SAMN09727239
2	2-F	CLN – Sputum	1/30/06	LpnS13041	SAMN09727240
2	2-G	CLN – Sputum	2/7/06	LpnS13041	SAMN09727241
2	2-H	CLN – Sputum	2/17/06	LpnS13041	SAMN09727242
2	2-I	ENV – RM 6–3404 water	2/15/06	LpnS13041	SAMN09727243
2	2-J	ENV – ERM C4 water	2/15/06	LpnS13041	SAMN09727244
2	2-K	ENV – ERM 13 water	2/18/06	LpnS13041	SAMN09727245
2	2-L	CLN – Sputum	12/6/06	LpnS13041	SAMN09727246
2	2-M	CLN – Sputum	10/9/08	LpnS13041	SAMN09727247
2	2-N	CLN – Sputum	9/24/10	LpnS13041	SAMN09727248
3	3-A	CLN - Unknown	2/20/90	LpnS13066	SAMN09727249
3	3-B	ENV - Water	6/21/90	LpnS13066	SAMN09727250
3	3-C	ENV – PRM 218 water	6/21/90	LpnS13066	SAMN09727251
3	3-D	CLN – Sputum	6/25/92	LpnS13066	SAMN09727252
3	3-E	CLN – Sputum	5/11/94	LpnS13066	SAMN09727253
3	3-F	ENV – PRM 209 water	6/10/94	LpnS13066	SAMN09727254
3	3-G	ENV – PRM 432 water	8/18/95	LpnS13066	SAMN09727255
3	3-H	CLN – Unknown	8/20/95	LpnS13066	SAMN09727256
3	3-I	ENV – RM 504 water	12/12/06	LpnS13066	SAMN09727257
3	3-J	ENV – Hot water tank water	12/12/06	LpnS13066	SAMN09727258
3	3-K	CLN - Tissue	8/4/07	LpnS13066	SAMN09727259
3	3-L	ENV – RM 605 bathRM water	8/13/07	LpnS13066	SAMN09727260
3	3-M	ENV – RM 305 sink water	8/13/07	LpnS13066	SAMN09727261
3	3-N	ENV – Cooling tower water	8/17/07	LpnS13066	SAMN09727262
Nursing Home	NH-1	ENV – Cooling tower SUMP	8/3/04	LpnS13105	SAMN09727263
Nursing Home	NH-2	CLN - Sputum	8/3/04	LpnS13105	SAMN09727263
Nursing Home	NH-3	ENV – RM 512 sink swab	8/2/04	LpnS13106	SAMN09727265
Nursing Home	NH-4	CLN – West shower RM water	8/3/04	LpnS13107	SAMN09727266

(Rowbotham, 1980; Cazalet et al., 2004).

Environmental sampling of epidemiologically-linked sources in facilities allows the comparison of environmental isolates of *Legionella* to those from clinical cases to determine relatedness. The comparison of environmental and clinical *Legionella* isolates has historically been performed with pulsed-field gel electrophoresis (PFGE) (Schoonmaker et al., 1992), with samples showing matching band patterns considered to be identical strains (Tenover et al., 1995). However, because of the long-term persistence of *Legionella* spp. in built environments, it can be difficult to identify outbreak sources. Additionally, PFGE analysis is a subjective method that provides limited resolution as it considers only a small fraction of the genome. Other methods, such as sequence-based typing (SBT) or core genome multi-locus sequence typing (cgMLST), which rely on allelic characterization to determine relatedness offer improved resolution in comparison to PFGE. Whole genome sequencing (WGS) single nucleotide polymorphism (SNP) analysis is a highly discriminatory method capable of assessing over 99% of the genome to quantify genetic relatedness among isolates (Raphael et al., 2016). For example, the increased resolution of WGS SNP analysis has distinguished isolates from separate outbreaks despite displaying the same PFGE pattern (Raphael et al., 2016) and has confirmed nosocomial sources of LD acquired in hospitals (Rosendahl Madsen et al., 2017; Reuter et al., 2013; Graham et al., 2014; Bartley et al., 2016).

WGS has also revealed that *Legionella* spp. are genetically diverse and acquire variation through mutation, horizontal gene transfer (HGT), acquisition of mobile elements, and recombination (Cazalet et al., 2004; Schjørring et al., 2014; Gomez-Valero and Buchrieser, 2013). Recombination and HGT are particularly prevalent in shaping *Legionella* genomes (Gomez-Valero and Buchrieser, 2013; Sánchez-Busó et al., 2014a; Coscollá and González-Candelas, 2007; Gomez-Valero et al., 2011). Recombination contributed up to 98% of the SNPs among *Legionella pneumophila* isolates sampled over the course of 11 years and