

SHORT REPORT

Real-time investigation of a *Legionella pneumophila* outbreak using whole genome sequencing

R. M. A. GRAHAM*, C. J. DOYLE AND A. V. JENNISON

Public Health Microbiology, Communicable Disease, Department of Health, Forensic and Scientific Services, Brisbane, QLD, Australia

Received 15 August 2013; Final revision 8 January 2014; Accepted 3 February 2014;
first published online 27 February 2014

SUMMARY

Legionella pneumophila is the main pathogen responsible for outbreaks of Legionnaires' disease, which can be related to contaminated water supplies such as cooling towers or water pipes. We combined conventional molecular methods and whole genome sequence (WGS) analysis to investigate an outbreak of *L. pneumophila* in a large Australian hospital. Typing of these isolates using sequence-based typing and virulence gene profiling, was unable to discriminate between outbreak and non-outbreak isolates. WGS analysis was performed on isolates during the outbreak, as well as on unlinked isolates from the Public Health Microbiology reference collection. The more powerful resolution provided by analysis of whole genome sequences allowed outbreak isolates to be distinguished from isolates that were temporally and spatially unassociated with the outbreak, demonstrating that this technology can be used in real-time to investigate *L. pneumophila* outbreaks.

Key words: Bacterial infections, *Legionella*, Legionnaire's disease, molecular biology, outbreaks.

Legionella pneumophila is a Gram-negative facultative intracellular pathogen, and is a causative agent of legionellosis, a form of pneumonia that is potentially fatal [1]. *L. pneumophila* exists naturally in aquatic environments but can be found as a contaminant in air-conditioning cooling towers and hot-water systems where it can be transmitted by aerosol to cause infection [1]. As a species, *L. pneumophila* shows a high degree of genomic plasticity [2]; however, the presence of a globally distributed clone, responsible for outbreaks and sporadic cases in several countries has

been described [3]. A study of *L. pneumophila* serogroup 1 in Australia demonstrated the presence of a predominant AFLP genotype in clinical isolates [4] although it is not clear if this genotype corresponds to that of the globally distributed clone mentioned above. The predominance of a particular *L. pneumophila* genotype can make epidemiological typing in outbreak situations difficult, since not all typing methods have the necessary resolving power to discriminate outbreak isolates from non-outbreak isolates. Whole genome sequencing provides a high level of resolution and has been used successfully to characterize disease outbreaks and to elucidate transmission sources and links between patients [5]. A retrospective pilot study into the use of whole genome sequence (WGS) analysis for the investigation of a *L. pneumophila* outbreak suggested that this approach could be used in

* Author for correspondence: Dr R. M. A. Graham, Public Health Microbiology, Communicable Disease, PO Box 594, Department of Health, Forensic and Scientific Services, Brisbane, QLD 4108, Australia.
(Email: rikki_graham@health.qld.gov.au)

conjunction with other typing techniques to identify links between isolates and environmental sources [6]. However, to the best of our knowledge, until now there have been no reports of WGS technology being used to investigate an outbreak in real-time in order to provide timely advice to public health units and aid in epidemiological investigation.

In May 2013, an outbreak of legionellosis caused by *L. pneumophila* serogroup 1 occurred in a large Australian hospital. Two patients were positive for *L. pneumophila* by culture, and one was positive by urinary antigen test; however, *L. pneumophila* was not isolated from this patient. Initial epidemiological investigations strongly suggested that the hot-water supply of the hospital was linked with these cases. Sequence-based typing (SBT) and virulence gene profiling were initially utilized as typing methods as part of the outbreak response but these were not able to discriminate outbreak isolates from non-outbreak isolates, probably due to the clonal nature of *L. pneumophila* present in the region. Methods such as spoligotyping and variable genetic element typing have been shown to have the potential to discriminate between isolates belonging to the same SBT sequence type [7, 8]; however, these methods can be labour intensive and may still have limited resolution compared to that possible by comparison of the whole genome. Whole genome sequencing of bacteria is becoming increasingly rapid and inexpensive due to advances in next-generation sequencing technology. It is now possible to sequence and analyse the entire genomes of several isolates within days and the information generated has the potential to provide a very high level of discrimination. For this reason, WGS analysis was used during this outbreak to compare isolates at the genomic level and provide further insight into possible links between the clinical and environmental isolates.

A total of nine *L. pneumophila* serogroup 1 isolates from the hospital where the outbreak occurred were referred to or isolated by the Public Health Microbiology laboratory as part of the outbreak response. Two of these were clinical isolates from the two legionellosis patients and seven isolates were from the hot-water system of the hospital. Culture, identification and serogrouping were confirmed by the *Legionella* reference laboratory using standard methods [9]. Additional isolates included as part of the WGS analysis were clinical and water isolates from the Public Health Microbiology reference collection, which are described in more detail below.

SBT performed using the seven loci scheme described by the European Legionnaires' Disease Surveillance Network [10] identified all nine isolates as belonging to sequence type (ST) 1, and virulence gene profiling using the method described by Huang *et al.* [11] demonstrated that all isolates were positive for the *lvh* and *rtxA* regions. However, the epidemiological conclusions that could be drawn based on this data alone were limited because the majority of *L. pneumophila* serogroup 1 isolates typed by the Public Health Microbiology laboratory have also been found to be ST1 and to possess the same virulence gene profile.

In order to expand upon the results obtained by SBT and virulence gene profiling, and to provide an improved level of discrimination between isolates from different sources, the isolates from the two legionellosis patients (P1 and P2) and an isolate from the hot-water supply of the hospital where the outbreak occurred (W1) were subjected to WGS analysis, as were three isolates from our reference collection that were considered to be temporally and/or spatially unassociated with this outbreak. These consisted of two clinical isolates, C1, isolated from a 2011 patient at the same hospital, and C2, isolated in 2008 from a patient at another hospital located ~5 km from the hospital where the current outbreak occurred, and one environmental isolate, W2, isolated in 2012 from the hot-water system of a building located ~30 km from the hospital where the current outbreak occurred.

DNA was extracted from the isolates using the MasterPure DNA extraction kit (Epicentre, USA) according to the manufacturer's instructions. Fragment libraries of the genomic DNA were generated using the Ion Plus Fragment library kit and were sequenced on an Ion Torrent PGM (Life Technologies, USA) according to the manufacturer's instructions. Analysis of the genomic data generated was performed using CLC Genomics Workbench v. 4.9 (CLC Bio, Denmark). Sequencing reads were mapped to a reference genome, *L. pneumophila* strain Paris (Genbank accession no. CR628336) [12].

Comparison of the genome sequences from the isolates indicated that they were highly similar to the *L. pneumophila* Paris strain, which is also ST1. However, in all of the isolates sequenced, the sequences for the resistance-related genomic island R1 described in the Paris strain by D'Auria *et al.* [13] were mostly absent, with only the first four genes present. P1, P2, W1 and C1 also lacked sequences for

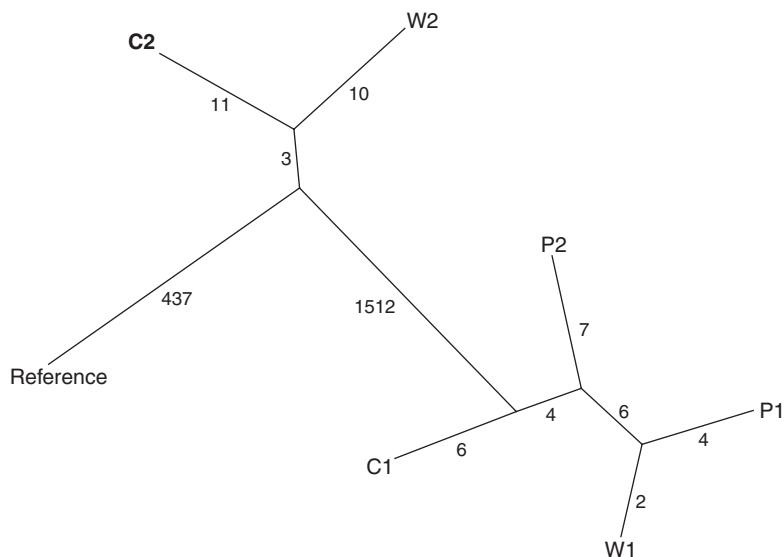


Fig. 1. Maximum-likelihood tree of *L. pneumophila* isolate single nucleotide polymorphisms (SNPs). Branch numbers indicate the number of SNPs.

72 of the 142 coding regions found on the Paris strain plasmid (pLPP, Genbank accession no. NC_006365) including the genes for the F-type IV secretion system (T4SSA). These isolates did, however, possess sequences for the T4SSA genes found on the Lorraine strain plasmid (pLELO, Genbank accession no. NC_018141) [2] as well as the sequences for several other coding regions from this plasmid. C2 and W2 differed from these isolates in that they possessed the sequences for the entire pLPP and showed no significant similarity to sequences from pLELO. Whole genome shotgun sequences for P1, W1, W2, P2, C1 and C2 have been submitted to Genbank with accession nos. AWQT00000000, AVAP00000000, AVNJ00000000, AWES00000000, AVOW00000000 and AVOV00000000, respectively.

Single nucleotide polymorphisms (SNPs) in the genomes of the isolates tested were identified using CLC Genomics Workbench v. 4.9, filtering for coding region SNPs that were present in 100% of mapped reads at regions with a minimum coverage of 15 reads. SNPs were analysed by generating a maximum-likelihood tree using BioNumerics v. 6.5 (Applied Maths, Belgium). This analysis revealed that the isolates formed two distinct groups separated by 1512 SNPs. The patient and environmental isolates P1, P2, and W1 clustered together into one group which was highly related genetically, differing by a maximum of 17 SNPs (Fig. 1). The other group consisted of C2, W2 and the *L. pneumophila* Paris reference

strain. The C1 isolate clustered with the outbreak isolates and was found to differ from P1, P2 and W1 by only 20, 17 and 18 SNPs, respectively. This isolate was isolated in 2011 from a patient at the same hospital, meaning that although it was not temporally associated, it was spatially associated with the isolates from the recent outbreak and may indicate a persistent presence of this strain at this location.

Phylogenetic analysis of the isolates was performed by multilocus sequence analysis (MLSA) using 28 genes that have previously been shown to be powerful for predicting the relatedness of *Legionella* and other bacterial genomes [2]. The sequences for three of the genes used in the original MLSA study (*lepA*, *metG*, *thdF*) were not complete in one or more of the genomes of the isolates tested and were therefore excluded from the analysis. The sequences of the remaining 28 genes were concatenated into a single sequence 41 250 bp long and the concatenated sequences were aligned using CLC Genomics Workbench v. 4.9. A pairwise comparison was performed using CLC Genomics Workbench v. 4.9 and a maximum-likelihood tree was built using Geneious v. 6.1 (Biomatters, New Zealand). The phylogenetic tree constructed using the concatenated sequences separated the isolates into two distinct groups (Fig. 2), with one group, group 1, consisting of P1, P2, C1 and W1. The isolates in this group had identical nucleotide sequences, with the exception of P2, which differed by one nucleotide. Group 2 consisted

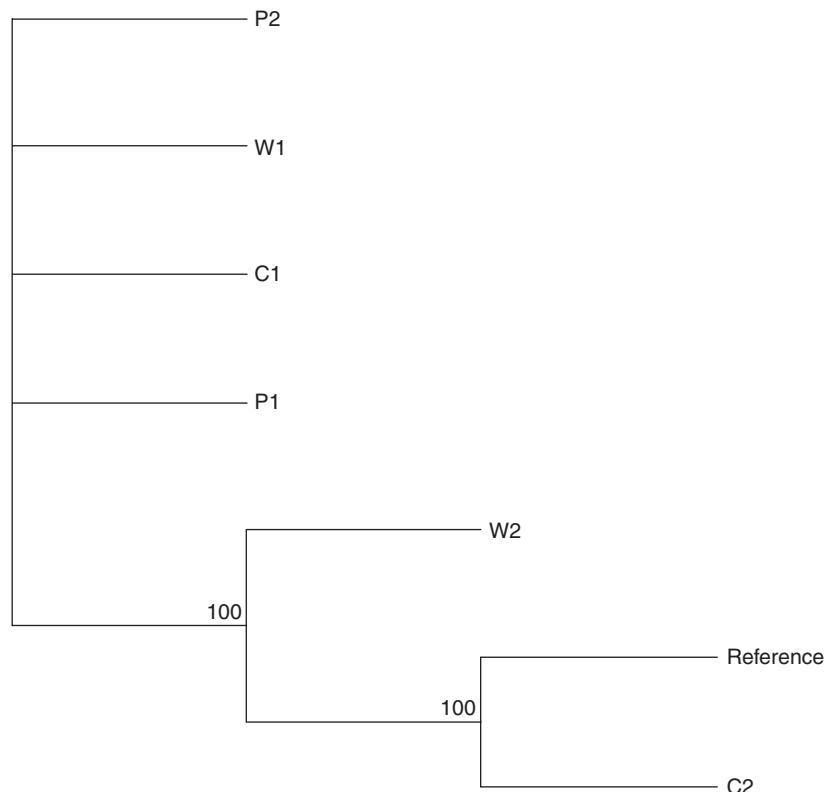


Fig. 2. Phylogenetic tree built using 28 housekeeping gene sequences concatenated into one sequence 41 205 bp long. The numbers next to the branches represent the percentages of the support for the groups.

of C2, W2 and the reference strain. This group differed from group 1 by 71 nucleotides and within this group the sequences differed by 2 bp. These groupings correlate with the clustering produced by the SNP analysis.

Overall, comparison of these *L. pneumophila* isolates at the genomic level was consistent with the reported clonal nature of *L. pneumophila* serogroup 1 ST1. Comparison of genomic features, SNP analysis and MLSA demonstrated that the level of genomic diversity that exists in *L. pneumophila* isolated from this geographical region is low, which is reflected by the inability of some typing methods to discriminate between spatially and temporally unassociated isolates. The WGS analysis determined that there was far less genetic diversity between the isolates from the outbreak-related patients and those from the hospital hot-water system than there was to other spatially and/or temporally related isolates. This information was provided to the state public health units during the outbreak investigation and used as part of the response to the outbreak [14]. Comparison of genomic features, SNP analysis and MLSA also demonstrated that isolate C1 was highly similar to, and likely to be

the same as, the outbreak strain, producing molecular evidence that the case in the same hospital 2 years previously was related to the current outbreak. The use of whole genome sequencing during this outbreak demonstrates how this technology can be used to identify links between environmental and patient isolates and to inform the public health response in real-time outbreak situations of *L. pneumophila*, even in regions showing endemicity of a clonal strain.

ACKNOWLEDGEMENTS

The authors acknowledge the laboratories that referred specimens and isolates that were included in this study, and thank the staff of the Public Health Microbiology laboratory for technical assistance. This work was funded in part by the Queensland Health Forensic and Scientific Services Research and Development Fund.

DECLARATION OF INTEREST

None.

REFERENCES

1. Fields BS, Benson RF, Besser RE. Legionella and Legionnaires' disease: 25 years of investigation. *Clinical Microbiology Reviews* 2002; **15**: 506–526.
2. Gomez-Valero L, et al. Extensive recombination events and horizontal gene transfer shaped the *Legionella pneumophila* genomes. *BMC Genomics* 2011; **12**: 536.
3. Yu VL, et al. Distribution of *Legionella* species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey. *Journal of Infectious Diseases* 2002; **186**: 127–128.
4. Huang B, et al. A predominant and virulent *Legionella pneumophila* serogroup 1 strain detected in isolates from patients and water in Queensland, Australia, by an amplified fragment length polymorphism protocol and virulence gene-based PCR assay. *Journal of Clinical Microbiology* 2004; **42**: 4164–4168.
5. Robinson E, Walker T, Pallen M. Genomics and outbreak investigation: from sequence to consequence. *Genome Medicine* 2013; **5**: 36.
6. Reuter S, et al. A pilot study of rapid whole-genome sequencing for the investigation of a *Legionella* outbreak. *BMJ Open* 2013; **3**(1).
7. Ginevra C, et al. *Legionella pneumophila* sequence type 1/Paris pulsotype subtyping by spoligotyping. *Journal of Clinical Microbiology* 2012; **50**: 696–701.
8. Pannier K, Heuner K, Lück C. Variable genetic element typing: a quick method for epidemiological subtyping of *Legionella pneumophila*. *European Journal of Clinical Microbiology and Infectious Diseases* 2010; **29**: 481–487.
9. Winn WCJ (ed.). *Legionella*, 7th edn. Washington, DC: ASM Press, 1999, pp. 572–585.
10. Edwards MT, Fry NK, Harrison TG. Clonal population structure of *Legionella pneumophila* inferred from allelic profiling. *Microbiology* 2008; **154**: 852–864.
11. Huang B, et al. Distribution of 19 major virulence genes in *Legionella pneumophila* serogroup 1 isolates from patients and water in Queensland, Australia. *Journal of Medical Microbiology* 2006; **55**: 993–997.
12. Cazalet C, et al. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nature Genetics* 2004; **36**: 1165–1173.
13. D'Auria G, et al. *Legionella pneumophila* pangenome reveals strain-specific virulence factors. *BMC Genomics* 2010; **11**: 181.
14. Queensland Department of Health. Review of the prevention and control of *Legionella pneumophila* infection in Queensland (<http://www.health.qld.gov.au/legionnaires/docs/cho-legionella-report.pdf>). Queensland Government, 2013.