
Population Genomics of *Legionella longbeachae* and Hidden Complexities of Infection Source Attribution

Rodrigo Bacigalupe, Diane Lindsay, Giles Edwards, J. Ross Fitzgerald

Legionella longbeachae is the primary cause of legionellosis in Australasia and Southeast Asia and an emerging pathogen in Europe and the United States; however, our understanding of the population diversity of *L. longbeachae* from patient and environmental sources is limited. We analyzed the genomes of 64 *L. longbeachae* isolates, of which 29 were from a cluster of legionellosis cases linked to commercial growing media in Scotland in 2013 and 35 were non-outbreak-associated isolates from Scotland and other countries. We identified extensive genetic diversity across the *L. longbeachae* species, associated with intraspecies and interspecies gene flow, and a wide geographic distribution of closely related genotypes. Of note, we observed a highly diverse pool of *L. longbeachae* genotypes within compost samples that precluded the genetic establishment of an infection source. These data represent a view of the genomic diversity of *L. longbeachae* that will inform strategies for investigating future outbreaks.

Legionellosis presents as 2 clinically distinct forms: an influenza-like illness called Pontiac fever and a severe pneumonia known as Legionnaires' disease (1). In Europe and the United States, most legionellosis cases are caused by *Legionella pneumophila* serogroup 1 (1,2); <5% of cases are caused by nonpneumophila *Legionella* spp. (3,4). In Australasia, New Zealand, and some countries in Asia, infections caused by *L. longbeachae* occur at comparable levels to infections caused by *L. pneumophila* (5–7). Unlike *L. pneumophila* infections, which are typically linked to artificial water systems, *L. longbeachae* infections are associated with exposure to soil, compost, and potting mixes (8).

The number of legionellosis cases caused by *L. longbeachae* is increasing worldwide (7), with a notable rise reported across Europe (9–11). Within the United Kingdom, most *L. longbeachae* infections have been identified in

Scotland, where 6 cases were diagnosed during 2008–2012 (12) and another 6 were diagnosed in the summer of 2013 and represented a singular increased incidence or cluster with all patients requiring intensive care hospitalization (11). Epidemiologic investigation revealed that most patients from the 2013 cluster were avid gardeners, and *L. longbeachae* was isolated from respiratory secretions and from samples of the growing media they had used for gardening before becoming ill (11,12). However, an investigation into the provenance of the growing media did not reveal a single commercial or manufacturing source that would suggest a common origin for the *L. longbeachae* associated with the outbreak (11).

Molecular typing methods used to discriminate between *L. longbeachae* and other *Legionella* spp. and between the 2 *L. longbeachae* serogroups have limited efficacy, and although considerable evidence supports growing media as a source for *L. longbeachae* infections (13,14), there is still a lack of genetic evidence for an epidemiologic link. Furthermore, a population genomic study involving large numbers of *L. pneumophila* isolates has been conducted (15,16), but the same has not been done for *L. longbeachae*, so the diversity of environmental and pathogenic genotypes and the relationship between them remains unknown for *L. longbeachae*. To examine the etiology of the 2013 cluster of legionellosis cases in Scotland in the context of *L. longbeachae* species diversity, we analyzed the genomes of 70 *Legionella* spp. isolates from 4 countries over 18 years.

Materials and Methods

Bacterial Isolates

We sequenced 65 isolates that had previously been identified as *L. longbeachae*. These isolates were obtained during 1996–2014 from several patients, growing media samples (including compost and soil), and a hot water supply. Of these isolates, 55 were from Scotland (29 from the 2013 cluster of infections and 26 from other clinical and environmental samples) and 10 were from patients and

Author affiliations: The Roslin Institute, University of Edinburgh, Midlothian, Scotland, UK (R. Bacigalupe, J.R. Fitzgerald); Glasgow Royal Infirmary, Glasgow, Scotland, UK (D. Lindsay, G. Edwards)

DOI: <http://dx.doi.org/10.3201/eid2305.161165>

environmental compost samples in New Zealand (online Technical Appendix Table, <https://wwwnc.cdc.gov/EID/article/23/5/16-1165-Techapp1.pdf>).

In our analysis, we also included all publicly available genome sequences for *L. longbeachae*: *L. longbeachae* NSW150 (serogroup 1) and *L. longbeachae* C-4E7 (serogroup 2) isolated from patients in Australia; and *L. longbeachae* D-4968 (serogroup 1), *L. longbeachae* ATCC39642 (serogroup 1), and *L. longbeachae* 98072 (serogroup 2) isolated from patients in the United States (17–19). We sequenced multiple isolates ($n = 2$ to 5) for each of 3 patients and their linked growing media samples from the 2013 outbreak in Scotland and for 2 additional compost samples. The species of all isolates had been determined by serotyping or macrophage infectivity potentiator (mip) gene sequencing (20,21).

Bacterial Culture, Genomic DNA Isolation, and Whole-Genome Sequencing

We cultured *Legionella* spp. isolates in a microaerophilic and humid environment at 37°C on BCYE (buffered charcoal yeast extract) agar plates for 48 h. We then picked individual colonies from the plates and grew them in ACES-buffered yeast extract broth containing *Legionella* BCYE Growth Supplement (Oxoid Ltd., Basingstoke, UK) with shaking at 37°C for 24–48 h. We extracted genomic DNA from fresh cultures by using the QIAGEN DNeasy Blood and Tissue Kit (QIAGEN Benelux B.V., Venlo, the Netherlands).

We prepared sequencing libraries by using the Nextera XT kit for MiSeq or HiSeq (all from Illumina, San Diego, CA, USA) sequencing at Edinburgh Genomics, University of Edinburgh (Edinburgh, Scotland, UK). For each isolate, one 2×250 -bp or two 2×200 -bp paired-end sequencing runs were carried out using the MiSeq and HiSeq technologies, respectively. Raw reads were quality checked using FastQC v0.10.1 (22), and primers were trimmed by using Cutadapt (23). We used wgsim software (24) to simulate sequence reads for publicly available, complete whole-genome sequences.

Bioinformatic Analysis and Data Deposition

A detailed description of the bioinformatic analyses is available in the online Technical Appendix. The sequence data for the 65 genomes of *Legionella* spp. sequenced in this study were deposited in the SRA database (accession no. PRJEB14754).

Results

Limitations of Current Typing Approaches for *Legionella* spp. Identification

We sequenced 65 isolates obtained from several patients and environmental samples over 18 years in different

countries and previously identified as *L. longbeachae*. To confirm the species identity of the *Legionella* isolates, we constructed a phylogenetic tree that included all *Legionella* type strains for which cultures are available, based on the 16S rRNA gene sequence (25). We also built phylogenetic trees based on the whole-genome content and core-genome diversity. For each approach, 64 of the 70 isolates examined co-segregated within the *L. longbeachae*-specific clade, 4 isolates clustered with *Legionella anisa*, and 2 belonged to a separate clade that was distinct from all known *Legionella* spp. (Figure 1; online Technical Appendix Figures 1, 2). The species identities were further supported by determination of the average nucleotide identity values (online Technical Appendix Figure 3), a widely used method for bacterial species delineation based on genomic relatedness (26). Of note, *L. anisa* is the most common nonpneumophila *Legionella* spp. in Europe (27–29). In addition, *L. longbeachae* isolates 13.8642 (from a compost sample from Scotland) and 13.8295 (from a patient in New Zealand) belong to a putative novel *Legionella* spp. Overall, the data indicate that current serotyping methods and mip gene sequencing are limited in their capacity to identify *L. longbeachae* to the species level.

To investigate the genetic relatedness of *L. longbeachae* strains associated with the 2013 outbreak to temporally and geographically distinct isolates, we constructed a core genome-based neighbor-joining tree of the 64 confirmed *L. longbeachae* isolates obtained from 4 countries over 18 years (online Technical Appendix Figure 4). This phylogenetic tree presents a comet-like pattern, with 2 distinct clades separated by 9,911 single-nucleotide polymorphisms, representing the major serogroups (serogroups 1 and 2) previously identified for *L. longbeachae* (20), each containing isolates from patient and environmental samples from different years. In contrast with findings from a previous analysis of 2 isolates of *L. longbeachae* serogroup 1 (20), we observed a higher diversity among the 56 isolates within serogroup 1 (online Technical Appendix Figures 1, 4); this finding is not unexpected, given the difference in the number of genomes examined. Nevertheless, compared with isolates from the same serogroup in other *Legionella* spp., such as *L. pneumophila* serogroup 1 (2% polymorphism) (20), *L. longbeachae* serogroup 1 exhibits a lower diversity (<0.1% polymorphism). Although serogroup 1 and 2 clades contained isolates from Scotland, Australasia, and the United States, 96% of the isolates from Scotland (including all of the 2013 outbreak isolates) belonged to serogroup 1, suggesting that serogroup 1 may be more clinically relevant in Scotland than in some other countries where *L. longbeachae* is a more established cause of legionellosis. However, analysis of more isolates from different countries would be required to investigate this observation further.

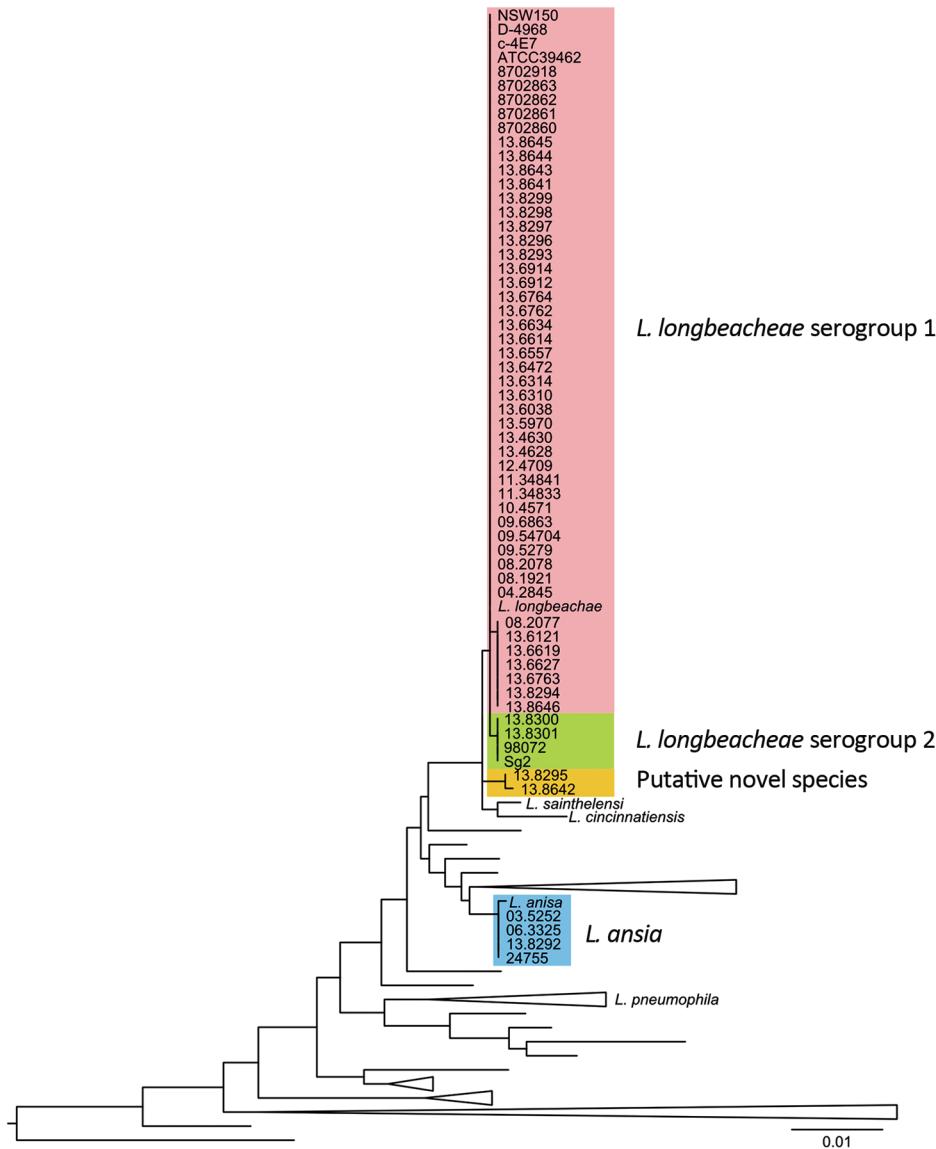


Figure 1. 16S rRNA gene-based phylogenetic tree of the sequenced genomes and all the cultured and type *Legionella* spp. strains available in the ribosomal database project (<http://rdp.cme.msu.edu/>), as accessed in May 2015. Scale bar indicates the mean number of nucleotide substitutions per site.

Effect of Recombination on *L. longbeachae* Serogroup 1 Population Structure

It is established that recombination has played a key role in shaping the evolutionary history of *L. pneumophila*, but its effect on *L. longbeachae* population structure is unknown (22,30). This knowledge is critical because for highly recombinant bacteria, recombination networks may represent evolutionary relationships more explicitly than traditional phylogenetic trees. Therefore, we constructed a recombination network of all serogroup 1 isolates by using the neighbor-net algorithm of SplitsTree4 (31). The resultant network displayed a reticulate topology with an extensive reticulated background from which clusters of isolates emerge, supporting an evolutionary history involving recombination ($p < 0.01$ by ϕ test) (32), followed by clonal expansion and subsequent additional recombination

events among some lineages (online Technical Appendix Figure 5). Using BratNextGen (33), we identified a total of 94 predicted recombination events affecting more than half of the core genome (1.74 Mb of 3.36 Mb) and representing recent and ancient recombination events of different sizes (range 1,350 bp–350 Kbp) distributed across the phylogeny (online Technical Appendix Figure 6). Given the reported limitation in sensitivity of BratNextGen for the identification of all recombination events (34), we also used ClonalFrameML (35), an algorithm that uses maximum likelihood inference to simultaneously detect recombination in bacterial genomes and account for it in phylogenetic reconstruction. The estimated average length of the recombined fragments was 8,047 bp, and the ratio of recombination to mutation was 1.42, indicating a greater role for recombination over mutation in the diversification

of *L. longbeachae*. This estimate is in accordance with early estimates for *L. pneumophila* based on multiple gene sequence data (36), but it is low compared with recent estimates based on whole-genome sequence data [recombination to mutation ratios of 16.8 (30) or 47.93 (37)]. Differences in the clonal diversity of *Legionella* spp. sequence datasets used to determine recombination rates could affect the estimates. Reconstruction of the phylogeny after removal of all predicted recombinant sequences resulted in a tree with largely similar clusters of isolates but with reduced branch lengths and variation in the position of nodes deep in the phylogeny (Figure 2).

Accessory Genome Analysis Indicates Extensive Interspecies and Intraspecies Gene Flow

The extent to which horizontal gene transfer occurs among *L. longbeachae* isolates and between *L. longbeachae*

and other *Legionella* spp. is unknown. In our study, the pangenome of *L. longbeachae* represented by the 56 serogroup 1 isolates was 6,890 genes, including a core genome of 2,574 genes; the average gene content was 3,558 genes per strain. The accessory genome, which included only strain-dependent genes varied from 809 to 1,155 genes, depending on the strain. A parsimony clustering analysis based on the presence or absence of all genes classified the isolates in a manner distinct from that in a core genome-based maximum-likelihood tree, suggesting extensive horizontal gene transfer among *L. longbeachae* isolates (online Technical Appendix Figures 1, 2). BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) analysis of all assembled contigs was used to filter for plasmid-related homologous sequences, revealed 2 major plasmids: pLLO, described previously in *L. longbeachae* NSW150 (20), and pLELO, originally identified

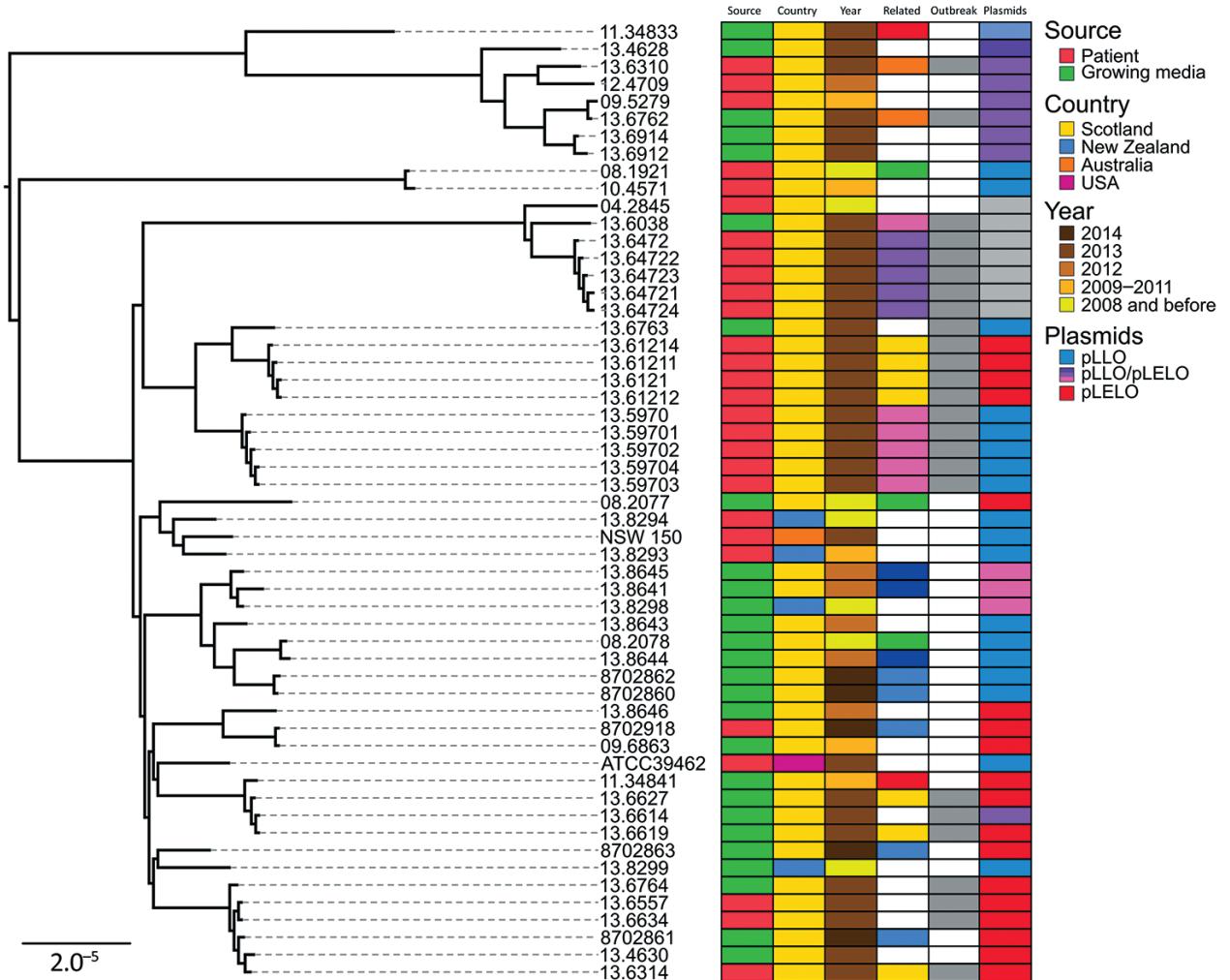


Figure 2. Core genome-based maximum-likelihood phylogeny of *Legionella longbeachae* serogroup 1 isolates corrected for recombination; source, country, year of isolation, relatedness and plasmid carriage are indicated. Related isolates are shown in the same color; those from the 2013 outbreak are indicated by gray. Isolates from the same patient are clustered together but do not cosegregate with cognate compost samples. Scale bar indicates the mean number of nucleotide substitutions per site.

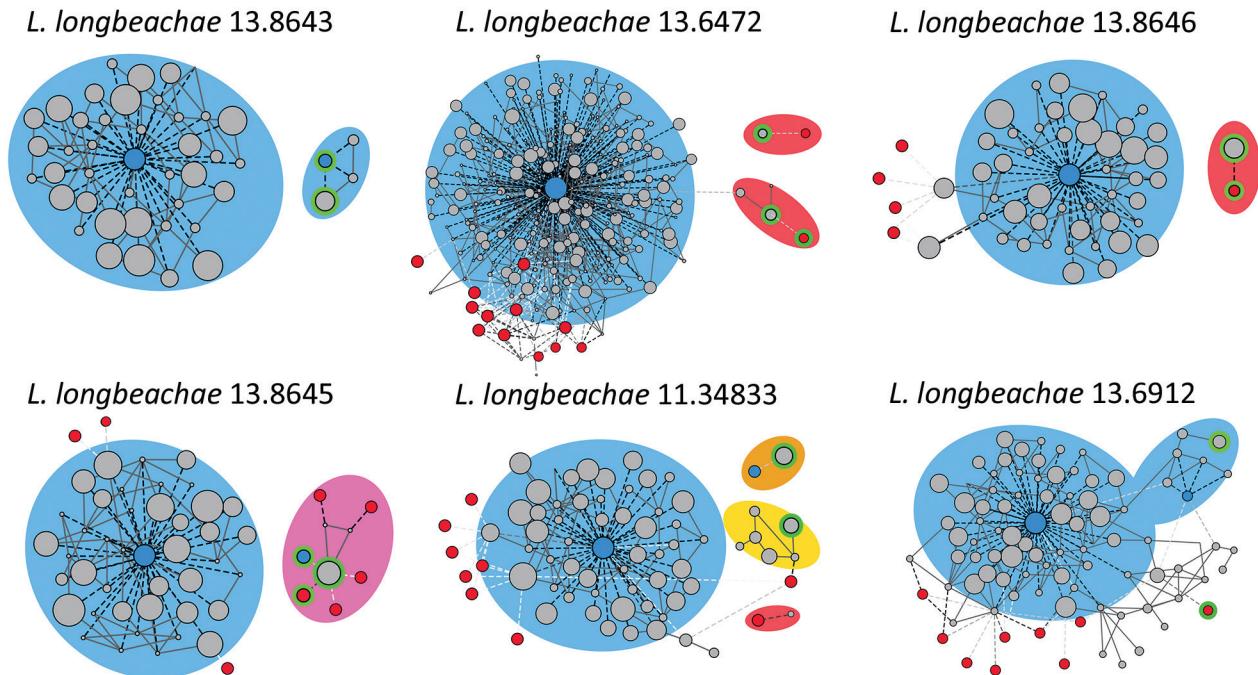


Figure 3. *Legionella longbeachae* plasmid analysis: contigs networks reconstructions for 6 representative *L. longbeachae* types of plasmid content. The networks of the contigs representing the main chromosome and plasmids comprising the genome obtained by using PLACNET (38), a program enabling reconstruction of plasmids from whole-genome sequence datasets. The sizes of the contig nodes (in gray) are proportional to their lengths; continuous lines correspond to scaffold links. Dashed lines represent BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) hits to the *L. longbeachae* (blue) or *L. pneumophila* (red) strains; intensity of the line is proportional to the hit (white indicates low, black indicates high). Green lines correspond to plasmid contigs. Background colors indicate species relatedness for the main chromosome and plasmids (blue for *L. longbeachae*, red for *L. pneumophila*, pink for a combination of both, and yellow for previously unidentified genomic content).

in *L. pneumophila subsp. pneumophila* (22). Of the 55 serogroup 1 isolates, 36 contained sequences for the pLLO and pLELO plasmids. Of note, the distribution of these plasmids among the *L. longbeachae* isolates correlated with the gene content–based clustering, whereas the distribution of plasmids in the core genome–based tree was independent of the phylogeny (Figure 2). In addition, 11 isolates appeared to contain plasmids with sequences homologous to those for pLLO and pLELO, which is indicative of recombinant forms of the plasmid. Further examination of plasmid diversity using a modified version of PLACNET (38), a program enabling reconstruction of plasmids from whole-genome sequence datasets, confirmed that some plasmids consisted of a mosaic of recombinant fragments homologous to pLELO, pLLO, or other unknown plasmids (Figure 3). Overall, these data indicate the high prevalence of specific plasmids among *L. longbeachae* isolates and reveal extensive recombination and horizontal gene transfer among different *Legionella* spp (39). The high prevalence of plasmids in *L. longbeachae* is notable, considering these elements may be less common in *L. pneumophila* (30).

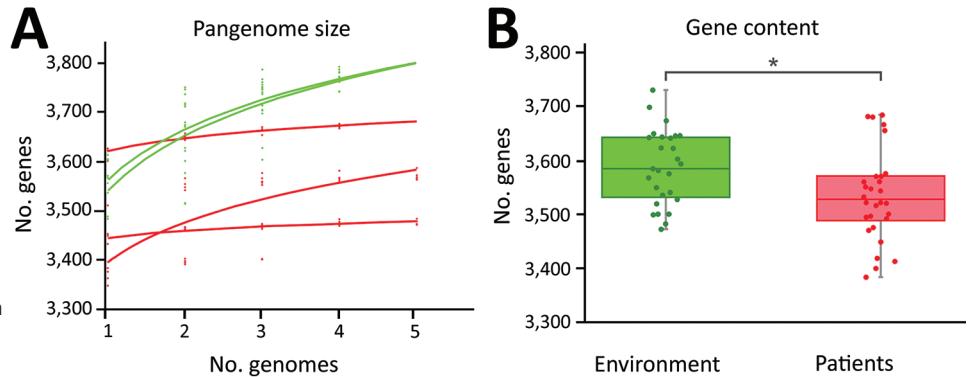
To examine the possibility that clinical and environmental isolates of *L. longbeachae* contained genomic

differences reflecting their distinct origins, we compared their accessory genome content. For isolates obtained from a single patient sample, the accessory genome was highly conserved compared with those for environmental isolates from a single compost sample or closely related environmental isolates from distinct compost samples (Figure 4, panel A). In addition, considering the average gene content of all sequenced isolates (28 clinical and 27 environmental), the gene content for *L. longbeachae* from growing media samples (3,586 genes) was significantly higher than that for isolates from patients (3,533 genes; 2-sample *t*-test, $t = 2.5213$; d.f. = 53; $p = 0.01474$) (Figure 4, panel B). The data imply that gene loss occurs during human infection or that *L. longbeachae* strains with reduced gene content have enhanced human infectivity. However, we did not identify a specific enriched gene or functional category in clinical or environmental samples (data not shown).

Source Attribution Confounded by Complex Serogroup 1 Populations within Environmental Samples

Having accounted for the influence of recombination on the phylogeny of *L. longbeachae*, we investigated the diversity of isolates associated with 5 patients and their linked compost samples obtained during 2008–2014, including

Figure 4. Variation in gene content between environmental and patient *Legionella longbeachae* samples. A) Increase in pangenome size with every addition of a *L. longbeachae* genome. Environmental isolates pangenomes (green) are larger and continue increasing after the addition of 5 genomes, consistent with an open pangenome, but the within-patient pangenome plateaus quickly, consistent with a more closed pangenome. B) Average gene content of environmental isolates is significantly higher than that of clinical isolates ($p = 0.01474$).



3 patients from the 2013 outbreak in Scotland. Of note, isolates from the 2013 outbreak were distributed across several subclades of the tree, indicating that the infections were caused by different strains (Figure 2). However, all isolates from a single patient clustered together, consistent with a monoclonal etiology of each infection. Of note, for all 5 patients, clinical isolates were not closely allied to the environmental isolates obtained from linked compost samples, and therefore a genetic link between patient and compost samples could not be established. Most subclades included isolates of diverse geographic origin, consistent with a wide distribution for *L. longbeachae* strains; however, 3 *L. longbeachae* isolates originating from Australasia (strains 13.8294, 13.8293, and NSW150) belonged to their own region-specific cluster (Figure 2).

We hypothesized that the lack of genetic relatedness between *L. longbeachae* isolates from patients and linked compost samples could be explained by a highly diverse population of *L. longbeachae* in growing media samples compounded by a sampling strategy consisting of a single sequenced isolate. All 5 compost samples for which we had >1 isolate contained isolates distributed across multiple clades in the phylogenetic tree. In particular, 5 isolates from the same growing media sample linked to a patient infected in Edinburgh in 2014 were distributed across 4 distinct clades, demonstrating that within a single environmental sample, considerable species diversity may be represented (Figure 2). Taken together, these data suggest that for future outbreak investigations, extensive sampling of environmental samples may be required to identify genotypes responsible for episodes of legionellosis infection, if indeed they are present.

Discussion

Our findings reveal the population genomic structure for *L. longbeachae*, an emerging pathogen in Europe and

the United States, and includes a genome-scale investigation into an outbreak of *L. longbeachae* legionellosis. We provide evidence for extensive recombination and lateral gene transfer among *L. longbeachae*, including the presence of widely distributed mosaic plasmids that have likely recombined with plasmids from other *Legionella* spp., suggesting an ecologic overlap or shared habitat. Our analysis highlights the need to account for recombination events when determining the genetic relatedness of *L. longbeachae* isolates.

Our application of whole-genome sequencing for diagnostic purposes revealed the misidentification, using current serotyping methods, of several *L. anisa* isolates as *L. longbeachae* and led to the identification of a putative novel *Legionella* sp. linked to legionellosis. These findings highlight the limitations of current typing methods for differentiation of *Legionella* spp. and accurate identification of legionellosis etiology.

We used whole-genome sequencing to attempt to establish a genetic link between legionellosis infections and associated compost samples. Our inability to establish a link probably reflects the traditional strategy of single isolate sampling, which when applied to a highly diverse pool of *L. longbeachae* genotypes fails to detect the infecting genotype. We suggest that the approach to investigating the source of future legionellosis cases linked to growing media will require a radical revision of sampling protocols to maximize the chances of isolating the infecting strain, if present. Taken together, our findings provide a view of the population structure of *L. longbeachae* and highlight the complexities of tracing the origin of legionellosis associated with growing media. Overall, our findings demonstrate the resolution afforded by whole-genome sequencing for understanding the biology underpinning legionellosis and provide information that should be considered for future epidemiologic investigations.

Acknowledgment

We are grateful to Carmen Buchrieser for providing the original sequence reads for *L. longbeachae* strains ATCC39642, 98072, and C-4E7. We thank David Harte for supplying the New Zealand strains.

Funding was provided by the Chief Scientist's Office Scotland (grant ETM/421 to J.R.F.) and by the Biotechnology and Biological Sciences Research Council (ISP3 grant BB/J004227/1 to J.R.F.).

Mr. Bacigalupe is a PhD candidate at the Roslin Institute, University of Edinburgh. His primary research focuses on the evolution, adaptation, and outbreak dynamics of bacterial pathogens.

References

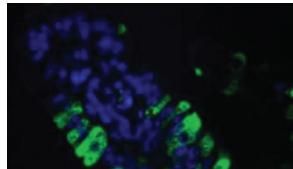
- Fields BS, Benson RF, Besser RE. *Legionella* and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev*. 2002; 15:506–26. <http://dx.doi.org/10.1128/CMR.15.3.506-526.2002>
- European Centre for Disease Prevention and Control. Surveillance report. Legionnaires' disease in Europe, 2010. 2012 [cited 2016 Jul 9]. <http://ecdc.europa.eu/en/publications/publications/sur-legionnaires-disease-surveillance-2010.pdf>
- Joseph CA, Ricketts KD; European Working Group for Legionella Infections. Legionnaires disease in Europe 2007-2008. *Euro Surveill*. 2010;15:19493.
- Marston BJ, Lipman HB, Breiman RF. Surveillance for Legionnaires' disease. Risk factors for morbidity and mortality. *Arch Intern Med*. 1994;154:2417–22. <http://dx.doi.org/10.1001/archinte.1994.00420210049006>
- Li JS, O'Brien ED, Guest C. A review of national legionellosis surveillance in Australia, 1991 to 2000. *Commun Dis Intell Q Rep*. 2002;26:461–8.
- Cramp GJ, Harte D, Douglas NM, Graham F, Schousboe M, Sykes K. An outbreak of Pontiac fever due to *Legionella longbeachae* serogroup 2 found in potting mix in a horticultural nursery in New Zealand. *Epidemiol Infect*. 2010;138:15–20. <http://dx.doi.org/10.1017/S0950268809990835>
- Whiley H, Bentham R. *Legionella longbeachae* and legionellosis. *Emerg Infect Dis*. 2011;17:579–83. <http://dx.doi.org/10.3201/eid1704.100446>
- Yu VL, Plouffe JF, Pastoris MC, Stout JE, Schousboe M, Widmer A, et al. Distribution of *Legionella* species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey. *J Infect Dis*. 2002;186:127–8. <http://dx.doi.org/10.1086/341087>
- García C, Ugalde E, Campo AB, Miñambres E, Kovács N. Fatal case of community-acquired pneumonia caused by *Legionella longbeachae* in a patient with systemic lupus erythematosus. *Eur J Clin Microbiol Infect Dis*. 2004;23:116–8. <http://dx.doi.org/10.1007/s10096-003-1071-7>
- den Boer JW, Yzerman EPF, Jansen R, Bruin JP, Verhoef LPB, Neve G, et al. Legionnaires' disease and gardening. *Clin Microbiol Infect*. 2007;13:88–91. <http://dx.doi.org/10.1111/j.1469-0691.2006.01562.x>
- Potts A, Donaghy M, Marley M, Othieno R, Stevenson J, Hyland J, et al. Cluster of Legionnaires disease cases caused by *Legionella longbeachae* serogroup 1, Scotland, August to September 2013. *Euro Surveill*. 2013;18:20656. <http://dx.doi.org/10.2807/1560-7917.ES2013.18.50.20656>
- Lindsay DSJ, Brown AW, Brown DJ, Pravinkumar SJ, Anderson E, Edwards GF. *Legionella longbeachae* serogroup 1 infections linked to potting compost. *J Med Microbiol*. 2012;61:218–22. <http://dx.doi.org/10.1099/jmm.0.035857-0>
- Steele TW, Lanser J, Sangster N. Isolation of *Legionella longbeachae* serogroup 1 from potting mixes. *Appl Environ Microbiol*. 1990;56:49–53.
- Koide M, Arakaki N, Saito A. Distribution of *Legionella longbeachae* and other legionellae in Japanese potting soils. *J Infect Chemother*. 2001;7:224–7. <http://dx.doi.org/10.1007/s101560170017>
- Reuter S, Harrison TG, Köser CU, Ellington MJ, Smith GP, Parkhill J, et al. A pilot study of rapid whole-genome sequencing for the investigation of a *Legionella* outbreak. *BMJ Open*. 2013;3:e002175. <http://dx.doi.org/10.1136/bmjopen-2012-002175>
- Rao C, Benhabib H, Ensminger AW. Phylogenetic reconstruction of the *Legionella pneumophila* Philadelphia-1 laboratory strains through comparative genomics. *PLoS One*. 2013;8:e64129. <http://dx.doi.org/10.1371/journal.pone.0064129>
- Cazalet C, Gomez-Valero L, Rusniok C, Lomma M, Dervins-Ravault D, Newton HJ, et al. Analysis of the *Legionella longbeachae* genome and transcriptome uncovers unique strategies to cause Legionnaires' disease. *PLoS Genet*. 2010;6:e1000851. <http://dx.doi.org/10.1371/journal.pgen.1000851>
- Kozak NA, Buss M, Lucas CE, Frace M, Govil D, Travis T, et al. Virulence factors encoded by *Legionella longbeachae* identified on the basis of the genome sequence analysis of clinical isolate D-4968. *J Bacteriol*. 2010;192:1030–44. <http://dx.doi.org/10.1128/JB.01272-09>
- Gomez-Valero L, Rusniok C, Jarraud S, Vacherie B, Rouy Z, Barbe V, et al. Extensive recombination events and horizontal gene transfer shaped the *Legionella pneumophila* genomes. *BMC Genomics*. 2011;12:536. <http://dx.doi.org/10.1186/1471-2164-12-536>
- Ratcliff RM, Lanser JA, Manning PA, Heuzenroeder MW. Sequence-based classification scheme for the *Legionella* targeting the mip gene. *J Clin Microbiol*. 1998;36:1560–7.
- Fallon RJ, Abraham WH. Experience with heat-killed antigens of *L. longbeachae* serogroups 1 and 2, and *L. jordanis* in the indirect fluorescence antibody test. *Zentralbl Bakteriol Mikrobiol Hyg A*. 1983;255:8–14.
- Babraham Bioinformatics. FastQC. 2010 [cited 2016 Jul 9]. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011;17:10–12.
- GitHub, Inc. wgsim Read Simulator [cited 2016 Jul 9]. <https://github.com/lh3/wgsim>
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 2014;42(D1):D633–42. <http://dx.doi.org/10.1093/nar/gkt1244>
- Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol*. 2014;64:346–51. <http://dx.doi.org/10.1099/ijs.0.059774-0>
- Health Protection Scotland. Surveillance report: legionellosis in Scotland 2013–2014. 2015 Sep 1 [cited 2015 Aug 15]. <http://www.hps.scot.nhs.uk/resp/wrdetail.aspx?id=65135&wrttype=6>
- van der Mee-Marquet N, Domelier AS, Arnault L, Bloc D, Laudat P, Hartemann P, et al. *Legionella anisa*, a possible indicator of water contamination by *Legionella pneumophila*. *J Clin Microbiol*. 2006;44:56–9. <http://dx.doi.org/10.1128/JCM.44.1.56-59.2006>
- Svarrer CW, Uldum SA. The occurrence of *Legionella* species other than *Legionella pneumophila* in clinical and environmental samples in Denmark identified by mip gene sequencing and matrix-assisted laser desorption ionization time-of-flight mass

- spectrometry. *Clin Microbiol Infect.* 2012;18:1004–9. <http://dx.doi.org/10.1111/j.1469-0691.2011.03698.x>
30. Underwood AP, Jones G, Mentasti M, Fry NK, Harrison TG. Comparison of the *Legionella pneumophila* population structure as determined by sequence-based typing and whole genome sequencing. *BMC Microbiol.* 2013;13:302. <http://dx.doi.org/10.1186/1471-2180-13-302>
 31. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 2006;23:254–67. <http://dx.doi.org/10.1093/molbev/msj030>
 32. Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. *Genetics.* 2006;172:2665–81. <http://dx.doi.org/10.1534/genetics.105.048975>
 33. Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, et al. Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Res.* 2012;40:e6. <http://dx.doi.org/10.1093/nar/gkr928>
 34. de Been M, van Schaik W, Cheng L, Corander J, Willems RJ. Recent recombination events in the core genome are associated with adaptive evolution in *Enterococcus faecium*. *Genome Biol Evol.* 2013;5:1524–35. <http://dx.doi.org/10.1093/gbe/evt111>
 35. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLOS Comput Biol.* 2015;11:e1004041. <http://dx.doi.org/10.1371/journal.pcbi.1004041>
 36. Coscollá M, Comas I, González-Candelas F. Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. *Mol Biol Evol.* 2011;28:985–1001. <http://dx.doi.org/10.1093/molbev/msq278>
 37. Sánchez-Busó L, Comas I, Jorques G, González-Candelas F. Recombination drives genome evolution in outbreak-related *Legionella pneumophila* isolates. *Nat Genet.* 2014;46:1205–11. <http://dx.doi.org/10.1038/ng.3114>
 38. Lanza VF, de Toro M, Garcillán-Barcia MP, Mora A, Blanco J, Coque TM, et al. Plasmid flux in *Escherichia coli* ST131 sublineages, analyzed by plasmid constellation network (PLACNET), a new method for plasmid reconstruction from whole genome sequences. *PLoS Genet.* 2014;10:e1004766. <http://dx.doi.org/10.1371/journal.pgen.1004766>
 39. Cazalet C, Rusniok C, Brüggemann H, Zidane N, Magnier A, Ma L, et al. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat Genet.* 2004;36:1165–73. <http://dx.doi.org/10.1038/ng1447>

Address for correspondence: J. Ross Fitzgerald, The Roslin Institute, University of Edinburgh, Midlothian, EH259RG Scotland, UK; email: Ross.Fitzgerald@ed.ac.uk

April 2015: Emerging Viruses

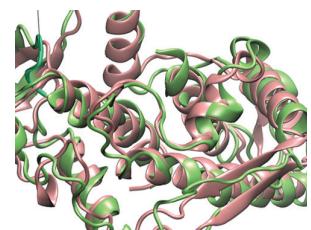
- Reappearance of Chikungunya, Formerly Called Dengue, in the Americas
- Hantavirus Pulmonary Syndrome, Southern Chile, 1995–2012
- Animal-Associated Exposure to Rabies Virus among Travelers, 1997–2012
- Evolution of Ebola Virus Disease from Exotic Infection to Global Health Priority, Liberia, Mid-2014
- Population Structure and Antimicrobial Resistance of Invasive Serotype IV Group B Streptococcus, Toronto, Ontario, Canada
- Sequence Variability and Geographic Distribution of Lassa Virus, Sierra Leone



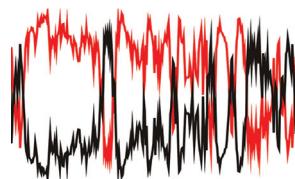
- Norovirus Genotype Profiles Associated with Foodborne Transmission, 1999–2012
- Deaths Associated with Respiratory Syncytial and Influenza Viruses among Persons >5 Years of Age in HIV-Prevalent Area, South Africa, 1998–2009
- Influenza A(H7N9) Virus Transmission between Finches and Poultry
- Highly Pathogenic Avian Influenza A(H5N1) Virus Infection among Workers at Live Bird Markets, Bangladesh, 2009–2010
- Increased Risk for Group B Streptococcus Sepsis in Young Infants Exposed to HIV, Soweto, South Africa, 2004–2008

- La Crosse Virus in *Aedes japonicus japonicus* Mosquitoes in the Appalachian Region, United States
- Multidrug-Resistant *Salmonella enterica* Serotype Typhi, Gulf of Guinea Region, Africa
- Reassortant Avian Influenza A(H9N2) Viruses in Chickens in Retail Poultry Shops, Pakistan, 2009–2010
- Candidate New Rotavirus Species in Sheltered Dogs, Hungary
- Severity of Influenza A(H1N1) Illness and Emergence of D225G Variant, 2013–14 Influenza Season, Florida, USA

- Close Relationship of Ruminant Pestiviruses and Classical Swine Fever Virus
- Peste des Petits Ruminants Virus in Heilongjiang Province, China, 2014
- Enterovirus 71 Subgenotype B5, France, 2013



- West Nile Virus Infection Incidence Based on Donated Blood Samples and Neuroinvasive Disease Reports, Northern Texas, USA, 2012
- Influenza A(H10N7) Virus in Dead Harbor Seals, Denmark



**EMERGING
INFECTIOUS DISEASES®**

<http://wwwnc.cdc.gov/eid/articles/issue/21/4/table-of-contents>

Hidden Complexities of Source Attribution for *Legionella longbeachae* Infections Revealed by Population Genomics

Technical Appendix

Materials and Methods

Genome Assemblies and Variant Calling

De novo assemblies of the *Legionella* isolates were produced using SPAdes 2.5.1. (1) (k values of 21, 33, 55, 77, 99 and 127), generating a median of 106 contigs per genome (range, 38–402 contigs), with an average of 4.16 Mb in length (3.98–4.52 Mb) and an average N50 of 130 Kb (29 Kb–291 Kb).

The error-corrected reads produced by SPAdes were mapped against the *Legionella longbeachae* reference genome of strain NSW 150 (GenBank accession number NC_013861) using BWA 0.5.9 (2) with default parameters. SNPs were called using Samtools 1.18 (3) and those absent in at least 30% of the reads, with quality below 30 and depth below 3 were filtered out. The output from this filtering was used to construct consensus genomes of all the isolates for further phylogenetic analyses.

Analysis of Genome Content

The contigs were annotated using Prokka v1.10 (4) and orthologous genes were clustered using the algorithm OrthoMCL (5) integrated in the software Get_homologs (6). We selected the options `-f 50` (filters by 50% length difference within clusters) and `-t 0` (for reporting all the clusters), resulting in 1801 core genome clusters. This program was also run using the Sg1 isolates only as input, specifying the options minimum percentage coverage (`-C 80`) and percentage identity (`-S 85`), which generated a core genome of 2574 gene clusters.

We also used JSpecies (7) to compute the average nucleotide identity values (BLAST; ANIb) between several pairs of isolates. These ANIb results were represented on a plot where isolates were clustered according to the 16S rRNA phylogenetic tree. In addition, a pangenomic

tree of the OMCL binary matrix from `get_homologs.pl` using all the *Legionella* isolates was constructed using the `compare_clusters.pl` script.

Evolutionary and Phylogenetic Analysis

To confirm the identity of the isolates, a Neighbor-Joining tree based on the 16S rRNA gene of the sequenced genomes and all the cultured type *Legionella* strains available in the Ribosomal Database Project (8) (as of 01/06/2015) was constructed. The RNAmmer 1.2 server (9) was used to identify the 16S rRNA genes in the de novo assemblies, which were then aligned using MUSCLE with default parameters (10). The Neighbor-Joining tree was estimated using the Hasegawa–Kishino–Yano model and 1000 bootstrap resampling replicates using the program Geneious 5.4.6 (11).

To construct a phylogeny based on the whole genome sequence data, the 1801 orthologous open reading frames identified using OrthoMCL were aligned using MUSCLE 3.8.31 (10). Individual protein alignments were translated back to DNA alignments using `pal2nal` v14 (12) and the resultant alignments were concatenated using `catfasta2phym.pl` (13) into an 1110024 bp long super-alignment. A ML phylogenetic tree was estimated based on this alignment using RAxML. C-4E7 was excluded from the original clustering as the low quality of the assembly significantly reduced the size of the core genome.

The *L. longbeachae* phylogeny was reconstructed using a Neighbor-Joining approach in Splitstree4 (14). Phylogenies of *L. longbeachae* Sg1 isolates before and after removing recombination were reconstructed from the genome alignments using RAxML 7.2.6 (15).

Detection of Recombination

Recombination was examined using the SplitsTree4 program (version 4.13.1) (14). A phylogenetic network was computed on the *L. longbeachae* Sg1 multiple genome alignment using the Neighbor-Net method implemented in this software. The statistical significance of the tree was confirmed using a Phi test (16). Recombination was detected on the core genome alignment of the Sg1 isolates using BratNextGen (17). After drawing a PSA tree, we selected a cutoff of 0.042, which split the tree into 8 clusters. We used 20 iterations for the recombination learning algorithm and after performing 100 replicate runs in a single processor we selected a threshold of 5% for estimating the significance of recombination. Finally, the *L. longbeachae* Sg1 ML tree and the whole genome alignment were used as input for ClonalFrameML (18) to generate a phylogeny with branch lengths corrected for recombination. 100 pseudo-bootstrap

replicates were used to estimate the uncertainty in the EM model and the option - ignore_user_sites with the list of non-core coordinates was parsed.

Plasmid Analysis

We used PLACNET, a software that constructs a network of contigs interactions, for the identification and visualization of plasmids (19). Bowtie2 v2.0.6 (20) was first used to find all possible scaffold links of the contigs by mapping the reads to them. Length and insert sizes of the reads mapped were calculated using Picard-Tools v1.90 (21). These files and metrics were parsed as input for placnet.pl, which produced a plasmid network from which we extracted the scaffolds. We then performed a BLAST search of the contigs assembly files to a database containing all the bacteria and plasmids genomes available in the NCBI ftp site (22) (created in March 2015) and results were filtered as follows: contigs longer than 200 bp, with a bitscore below $1e-26$ and that had at least a blast hit over 5% of the contig size. The results were further analyzed to classify the nodes into one of these categories: “hit completely to a single reference genome,” “split nodes that hit to a single reference genome” and “nodes that hit several genomes.” The hits and the scaffolds were combined into a network that was uploaded into Cytoscape (23). Recommendations given in the PLACNET manual were followed to visualize the chromosome and plasmids networks. BLAST was finally used to search for *Legionella* spp. plasmid related sequences in the contigs.

References

1. 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.
2. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2010;26:589–95.
3. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
4. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30:2068–9.
5. Li L, Stoeckert CJJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13:2178–89.

6. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol.* 2013;79:7696–701.
7. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A.* 2009;106:19126–31.
8. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 2014;42(D1):D633–42.
9. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007;35:3100–8.
10. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
11. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28:1647–9.
12. Suyama M, Torrents D, Bork P, Delbru M. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 2006;34:609–12.
13. Nylander JA. Nylander/catfasta2phym. <https://github.com/nylander/catfasta2phym>
14. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 2006;23:254–67.
15. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006;22:2688–90.
16. Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. *Genetics.* 2006;172:2665–81.
17. Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, et al. Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Res.* 2012;40:e6.
18. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLOS Comput Biol.* 2015;11:e1004041.
19. Lanza VF, de Toro M, Garcillan-Barcia MP, Mora A, Blanco J, De Cruz F, et al. Plasmid flux in *Escherichia coli* ST131 sublineages, analyzed by Plasmid Constellation Network (PLACNET), a new method for plasmid reconstruction from whole genome sequences. *PLoS Genet.* 2014;10:e1004766.

20. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
21. Picard-tools 1.9. <https://sourceforge.net/projects/picard/files/picard-tools/>
22. NCBI ftp site. <ftp://ftp.ncbi.nlm.nih.gov/>
23. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011;27:431–2.

Technical Appendix Table. Isolates examined in the current study

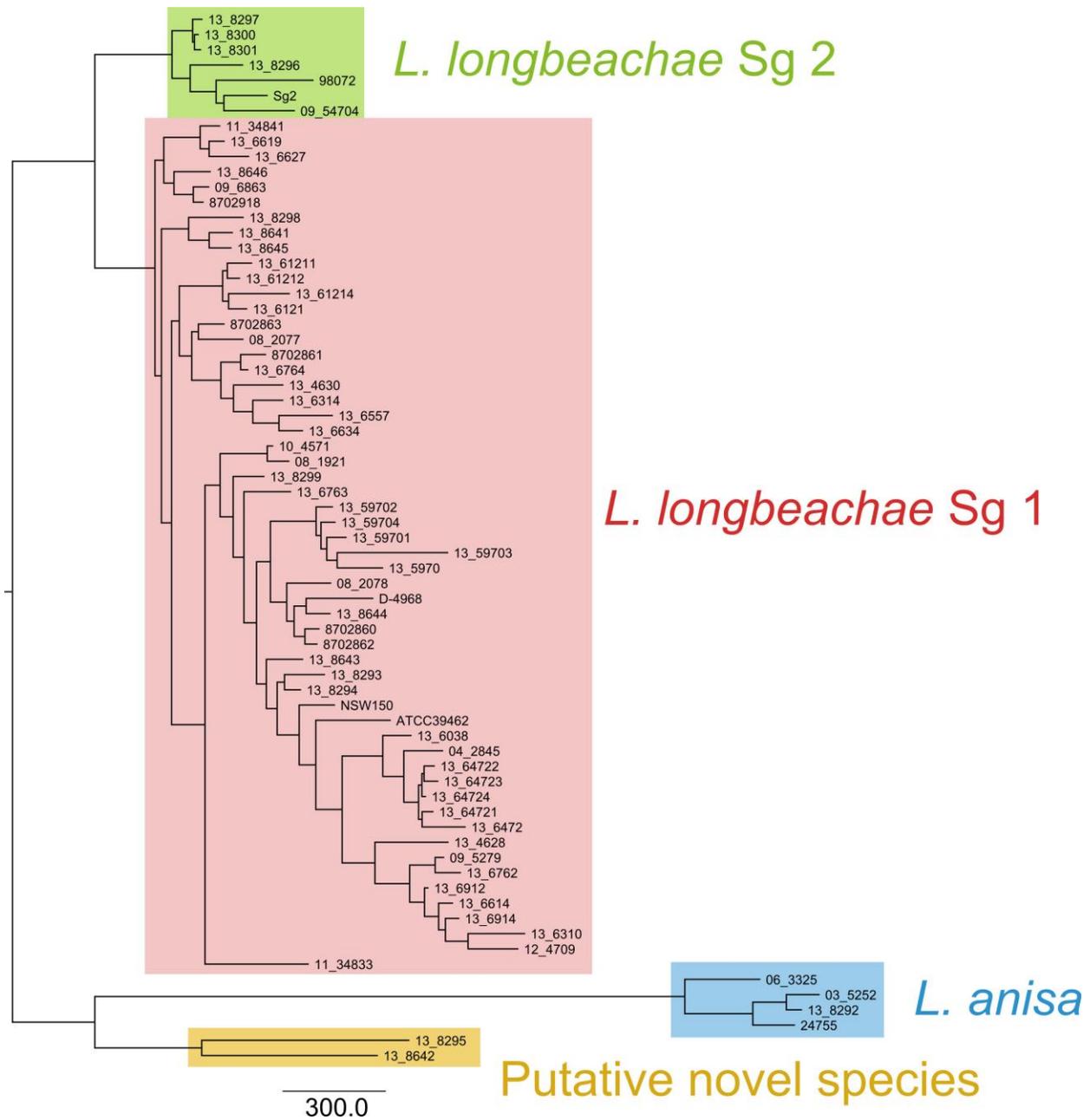
Identifier	Species/serogroup	Date*	Source	Country	Linked to†
02.4755	<i>L. anisa</i>	Sep 24, 2002	Hot water supply	Scotland	-
03.5252	<i>L. anisa</i>	Nov 12, 2003	Patient	Scotland	-
04.2845	<i>L. longbeachae</i> Sg1	Jun 7, 2004	Patient	Scotland	-
06.3325	<i>L. anisa</i>	Jul 18, 2006	Patient	Scotland	-
08.1921	<i>L. longbeachae</i> Sg1	Apr 1, 2008	Patient	Scotland	08.2077,08.2078
08.2077	<i>L. longbeachae</i> Sg1	Apr 17, 2008	Compost-IMS	Scotland	08.1921
08.2078	<i>L. longbeachae</i> Sg1	Jul 31, 2008	Compost Direct	Scotland	08.1921
09.5279	<i>L. longbeachae</i> Sg1	May 20, 2009	Patient	Scotland	09.5470–4
09.5470–4	<i>L. longbeachae</i> Sg2	Jun 2, 2009	Compost-Direct	Scotland	09.5279
09.6863	<i>L. longbeachae</i> Sg1	Nov 19, 2009	Compost	Scotland	Patient negative
10.4571	<i>L. longbeachae</i> Sg1	Mar 19, 2010	Patient	Scotland	Compost negative
11.3483(3)	<i>L. longbeachae</i> Sg1	May 13, 2011	Compost	Scotland	11.3484(1)
11.3484(1)	<i>L. longbeachae</i> Sg1	May 13, 2011	Compost	Scotland	11.3483(3)
12.4709	<i>L. longbeachae</i> Sg1	Jun 18, 2012	Patient	Scotland	No compost
13.8641	<i>L. longbeachae</i> Sg1	Jun 18, 2012	Compost	Scotland	13.8644/45
13.8642	New species	Jun 18, 2012	Compost	Scotland	13.8643
13.8643	<i>L. longbeachae</i> Sg1	Jun 18, 2012	Compost	Scotland	13.8642
13.8644	<i>L. longbeachae</i> Sg1	Jun 18, 2012	Compost	Scotland	13.8641/45
13.8645	<i>L. longbeachae</i> Sg1	Jun 18, 2012	Compost	Scotland	13.8641/44
13.8646	<i>L. longbeachae</i> Sg1	Jun 26, 2013	Compost	Scotland	-
13.4628	<i>L. longbeachae</i> Sg1	Jun 26, 2013	Compost	Scotland	-
13.4630	<i>L. longbeachae</i> Sg1	Aug 23, 2013	Compost	Scotland	-
13.5970‡	<i>L. longbeachae</i> Sg1	Aug 23, 2013	Patient	Scotland	13.6038
13.59701‡	<i>L. longbeachae</i> Sg1	Aug 23, 2013	Patient	Scotland	13.6038
13.59702‡	<i>L. longbeachae</i> Sg1	Aug 23, 2013	Patient	Scotland	13.6038
13.59703‡	<i>L. longbeachae</i> Sg1	Aug 23, 2013	Patient	Scotland	13.6038
13.59704‡	<i>L. longbeachae</i> Sg1	Aug 27, 2013	Patient	Scotland	13.6038
13.6038‡	<i>L. longbeachae</i> Sg1	Aug 30, 2013	Top soil	Scotland	13.5970
13.6121‡	<i>L. longbeachae</i> Sg1	Aug 30, 2013	Patient	Scotland	13.6619/27
13.61211‡	<i>L. longbeachae</i> Sg1	Aug 30, 2013	Patient	Scotland	13.6619/27
13.61212‡	<i>L. longbeachae</i> Sg1	Aug 30, 2013	Patient	Scotland	13.6619/27
13.61214‡	<i>L. longbeachae</i> Sg1	Sep 6, 2013	Patient	Scotland	13.6619/27
13.6310‡	<i>L. longbeachae</i> Sg1	Sep 6, 2013	Patient	Scotland	13.6762
13.6314‡	<i>L. longbeachae</i> Sg1	Sep 13, 2013	Patient	Scotland	13.6619
13.6472‡	<i>L. longbeachae</i> Sg1	Sep 13, 2013	Patient	Scotland	No compost
13.64721‡	<i>L. longbeachae</i> Sg1	Sep 13, 2013	Patient	Scotland	No compost
13.64722‡	<i>L. longbeachae</i> Sg1	Sep 13, 2013	Patient	Scotland	No compost
13.64723‡	<i>L. longbeachae</i> Sg1	Sep 17, 2013	Patient	Scotland	No compost
13.64724‡	<i>L. longbeachae</i> Sg1	Sep 19, 2013	Patient	Scotland	No compost
13.6557‡	<i>L. longbeachae</i> Sg1	Sep 19, 2013	Patient	Scotland	No compost
13.6614‡	<i>L. longbeachae</i> Sg1	Sep 19, 2013	Compost	Scotland	-
13.6619‡	<i>L. longbeachae</i> Sg1	Sep 20, 2013	Compost	Scotland	13.6121
13.6627‡	<i>L. longbeachae</i> Sg1	Sep 25, 2013	Compost	Scotland	13.6121
13.6634‡	<i>L. longbeachae</i> Sg1	Sep 25, 2013	Patient	Scotland	No compost
13.6762‡	<i>L. longbeachae</i> Sg1	Sep 25, 2013	Compost	Scotland	13.6310
13.6763‡	<i>L. longbeachae</i> Sg1	Oct 1, 2013	Compost	Scotland	-
13.6764‡	<i>L. longbeachae</i> Sg1	Oct 1, 2013	Compost	Scotland	-
13.6912	<i>L. longbeachae</i> Sg1	May 29, 2014	Compost	Scotland	-
13.6914	<i>L. longbeachae</i> Sg1	May 7, 2014	Compost	Scotland	-
8702918	<i>L. longbeachae</i> Sg1	May 8, 2014	Patient	Scotland	870286x
8702860	<i>L. longbeachae</i> Sg1	May 9, 2014	Soil	Scotland	8702918
8702861	<i>L. longbeachae</i> Sg1	May 10, 2014	Soil	Scotland	8702918
8702862	<i>L. longbeachae</i> Sg1	Apr 1, 2008	Soil	Scotland	8702918
8702863	<i>L. longbeachae</i> Sg1	Apr 17, 2008	Soil	Scotland	8702918
13.8292	<i>L. anisa</i>	2010	Compost	New Zealand	13.8293

Identifier	Species/serogroup	Date*	Source	Country	Linked to†
13.8293	<i>L. longbeachae</i> Sg1	2010	Patient	New Zealand	13.8292
13.8294	<i>L. longbeachae</i> Sg1	2004	Patient	New Zealand	-
13.8295	New species	2013	Patient	New Zealand	-
13.8296	<i>L. longbeachae</i> Sg2	2012	Sump drain	New Zealand	-
13.8297	<i>L. longbeachae</i> Sg2	2007	Compost	New Zealand	13.8301
13.8298	<i>L. longbeachae</i> Sg1	1996	Compost	New Zealand	-
13.8299	<i>L. longbeachae</i> Sg1	2003	Compost	New Zealand	-
13.8300	<i>L. longbeachae</i> Sg2	2011	Compost	New Zealand	-
13.8301	<i>L. longbeachae</i> Sg2	2007	Patient	New Zealand	13.8297
Sg2	<i>L. longbeachae</i> Sg2	-	-	-	-
NSW 150	<i>L. longbeachae</i> Sg1	-	Patient	Australia	-
C-4E7	<i>L. longbeachae</i> Sg2	-	Patient	Australia	-
D-4968	<i>L. longbeachae</i> Sg1	-	Patient	USA	-
ATCC39642	<i>L. longbeachae</i> Sg1	-	Patient	USA	-
98072	<i>L. longbeachae</i> Sg2	-	Patient	USA	-

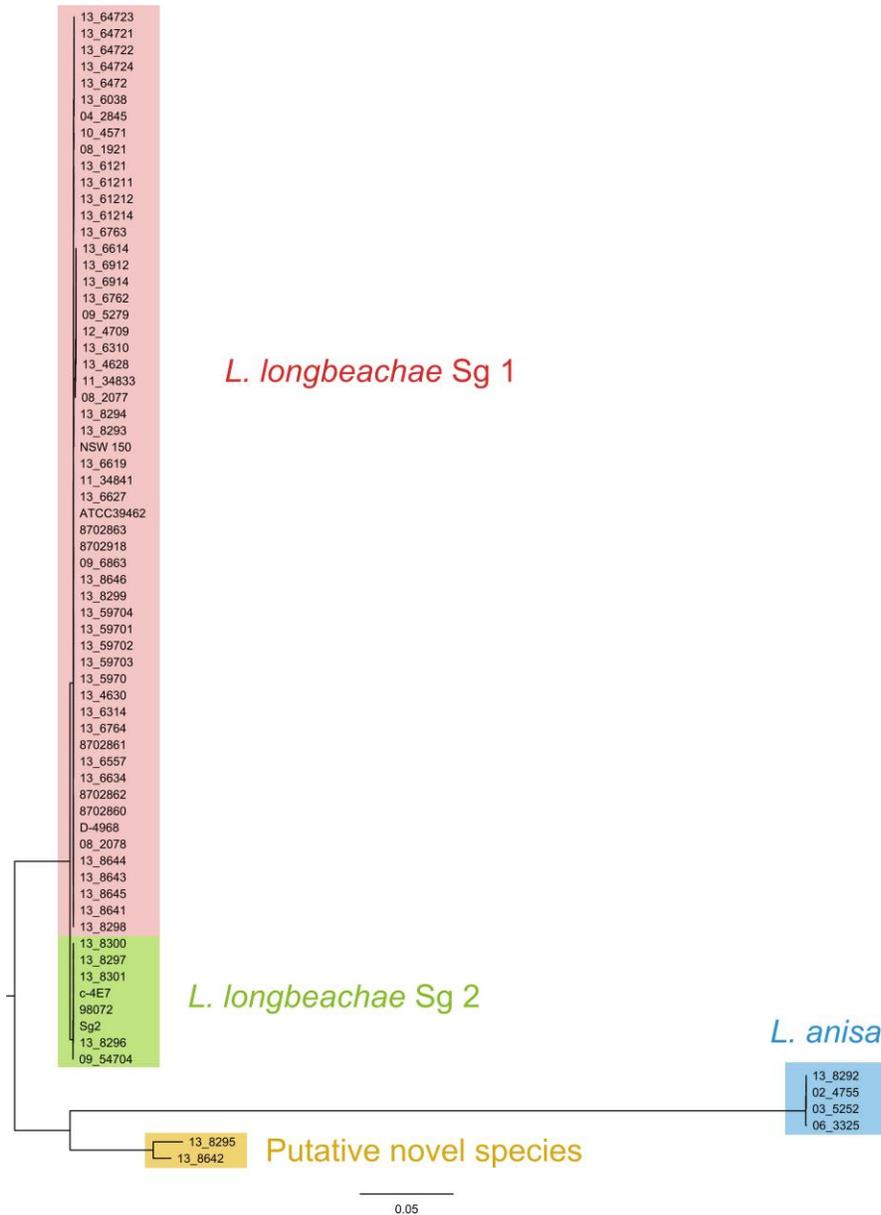
*Date received in the reference laboratory.

†Isolates that are linked to each other, as patient and cognate compost samples. Multiple isolates from each of the three patients or environmental samples share common identifiers (13.5970, 13.6121, 13.6472, 870286).

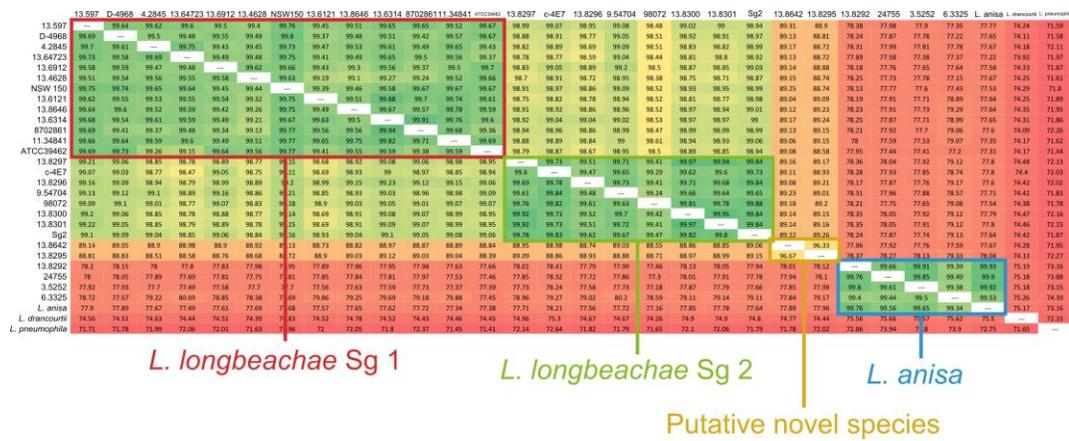
‡Isolates from the 2013 summer cluster of diseases in Scotland.



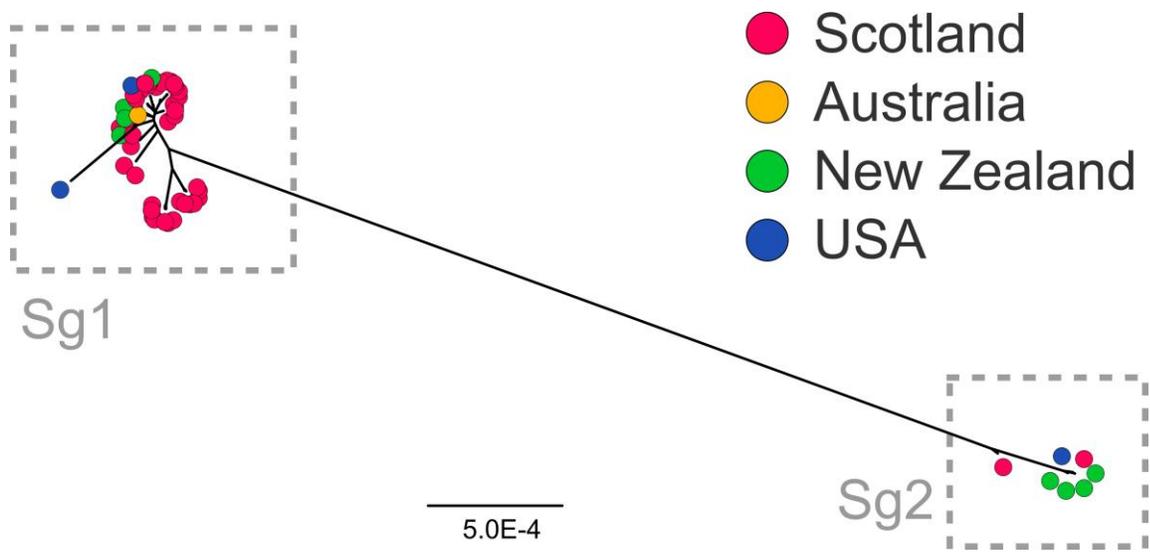
Technical Appendix Figure 1. Parsimony based tree of the OMCL pangenomic matrix obtained for all the sequenced genomes. Scale bar indicates the gene content differences.



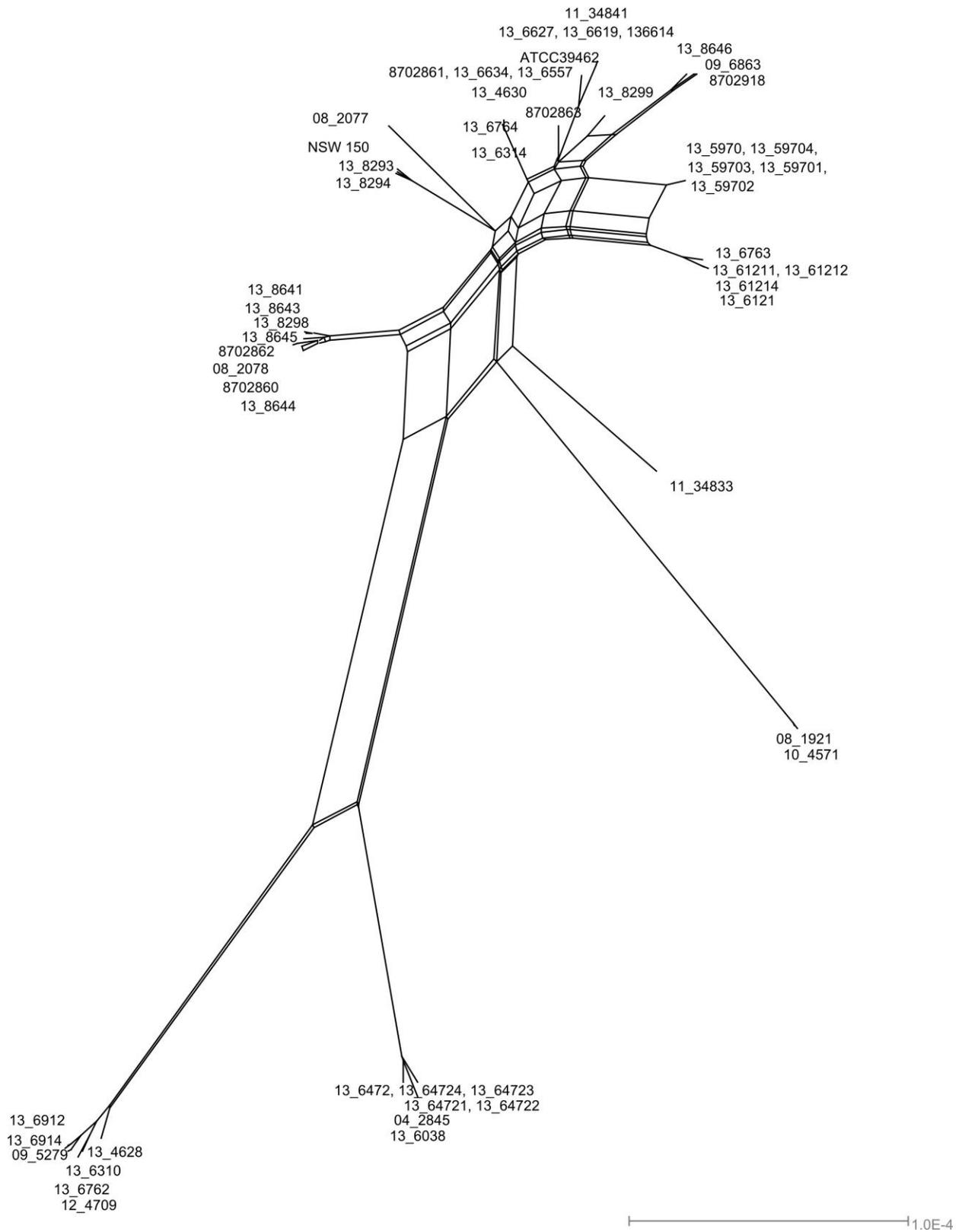
Technical Appendix Figure 2. Maximum likelihood tree of a core gene alignment of all the isolates included in the study. The tree shows the same clusters as the 16S rRNA gene based tree and the parsimony pangenome tree. Scale bar indicates the mean number of nucleotide substitutions per site.



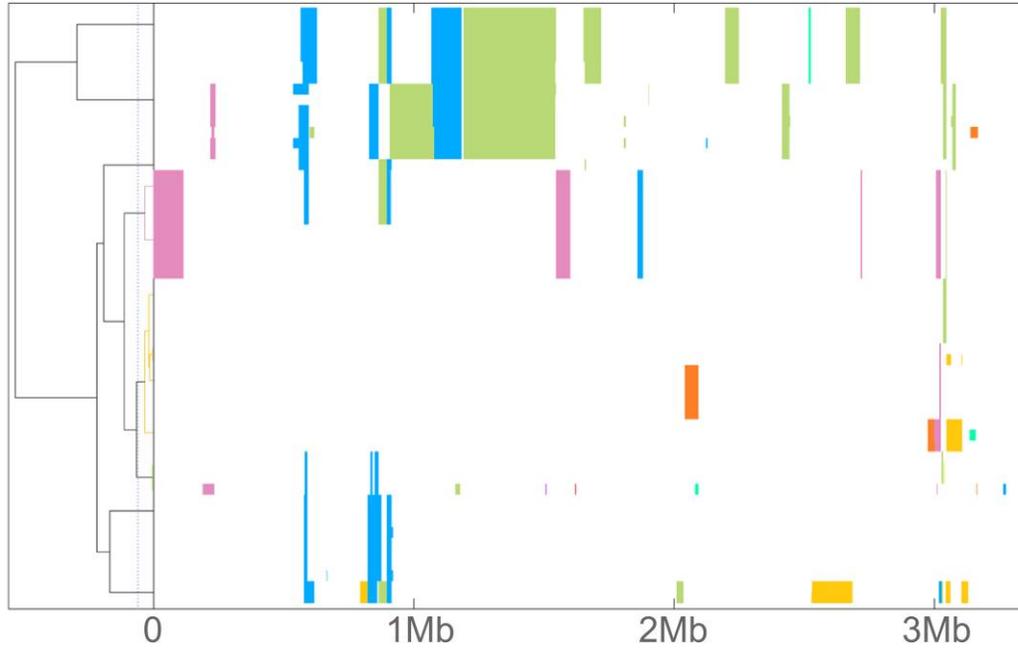
Technical Appendix Figure 3. Average Nucleotide Identity (ANI) comparison matrix of several genomes of the study sorted by similarity. Green, high ANI values; Red, low ANI values. The genomes of *Legionella drancourtii* and *Legionella pneumophila* were included in the matrix for comparative purposes, but none of the isolates showed an ANI higher than 76% with these two species.



Technical Appendix Figure 4. Neighbor-Joining phylogeny based on the core genome of *Legionella longbeachae* isolates. Isolates are colored by geographic source, and dashed boxes indicate the defined or predicted serogroups to which the isolates belong.



Technical Appendix Figure 5. Neighbor-Joining split network for the *Legionella longbeachae* Serogroup 1 isolates based on the consensus alignment obtained from mapping every isolate to the reference chromosome NSW 150. Scale bar indicates the mean number of nucleotide substitutions per site.



Technical Appendix Figure 6. Recombinant regions of the core genome alignment of 55 *L. longbeachae* Sg1 isolates as identified using BratNextGen. On the left, a clustering tree of the isolates with colored branches indicating cluster relationships. On the right, significant recombinant segments predicted, with similar color in a column representing recombinant regions for those isolates have the same origin.