

1 **Genome sequencing links persistent outbreak of legionellosis in Sydney to an emerging**
2 **clone of *Legionella pneumophila* ST211**

3

4 VJ Timms¹, R Rockett¹, NL Bachmann², E Martinez¹, Q Wang³, SC-A Chen^{3,4}, N Jeffreys³,
5 PJ Howard³, A Smith⁵, S Adamson⁶, R Gilmour⁶, V Sheppard⁶, and V Sintchenko^{1,3,4}

6 ¹ Centre for Infectious Diseases and Microbiology- Public Health, Westmead Hospital

7 ² Centenary Institute, University of Sydney

8 ³ Centre for Infectious Diseases and Microbiology Laboratory Services, Pathology West,
9 Sydney, New South Wales, Australia

10 ⁴ Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney Medical School, The
11 University of Sydney, New South Wales, Australia

12 ⁵ Forensic and Analytical Science Services

13 ⁶ Health Protection, NSW Health, North Sydney, New South Wales, Australia

14

15

16 **Keywords:**

17 *Legionella pneumophila*; Legionnaires' disease; prevention and control; molecular
18 epidemiology; bacterial genomics

19

20

21

22

23

24

25 **Abstract**

26 The city of Sydney, Australia, experienced a persistent outbreak of *Legionella pneumophila*
27 serogroup 1 (Lp1) pneumonia in 2016. To elucidate the source and bring the outbreak to a
28 close we examined the genomes of clinical and environmental Lp1 isolates recovered over 7
29 weeks. A total of 48 isolates from patients and cooling towers were sequenced and compared
30 using SNP-based, core-genome MLST and pangenome approaches. All three methods
31 confirmed phylogenetic relatedness between isolates associated with outbreaks in the Central
32 Business District (March and May) and Suburb 1. These isolates were designated “Main
33 cluster” and consisted of isolates from two patients from the CBD March outbreak, one
34 patient and one tower isolate from Suburb 1 and isolates from two cooling towers and three
35 patients from the CDB May outbreak. All main cluster isolates were sequence type ST211
36 which has only ever been reported in Canada. Significantly, pangenome analysis identified
37 mobile genetic elements containing a unique T4ASS that was specific to the main cluster and
38 co-circulating clinical strains, suggesting a potential mechanism for increased fitness and
39 persistence of the outbreak clone. Genome sequencing was key in deciphering the
40 environmental sources of infection among the spatially and temporally coinciding cases of
41 legionellosis in this highly populated urban setting. Further, the discovery of a unique T4ASS
42 emphasises the potential contribution of genome recombination in the emergence of
43 successful Lp1 clones.

44

45

46

47

48 **Introduction**

49 Between February to May 2016, Sydney, Australia experienced several apparent outbreaks of
50 human legionellosis. The particular features of these outbreaks were that they spanned three
51 precincts within the Greater Sydney region, 18 km apart over seven and a half weeks and
52 included the Central Business District (CBD), an area visited by over 610,000 visitors per
53 day. The agent responsible was found to be *Legionella pneumophila* serogroup 1 (Lp1), a
54 ubiquitous organism with a natural habitat of water reservoirs and amoebae that also
55 sporadically infects humans and during this period it was responsible for the death of two
56 patients.

57 The genus *Legionella* contains over 60 species. All species are found in the
58 environment and about half of these can cause a life-threatening illness in humans. *L.*
59 *pneumophila* is the major pathogen of this group and is further subdivided into 16 serogroups
60 with the majority (80-84%) of legionellosis cases caused by Lp1 (Ambrose et al.,
61 2014)(Kozak-Muiznieks et al., 2014)(Raphael et al., 2016). Since the first description of
62 Legionnaires disease in 1976 when a large outbreak of pneumonia occurred at an American
63 Legion conference in Philadelphia (Fraser et al., 1977), this infection has been implicated in
64 multiple sporadic cases and community outbreaks globally, most commonly in males over 50,
65 people with co-morbidities and smokers (Frieden, Jaffe, Stephens, & Thacker, 2011).

66 The *Legionella* are highly recombinant bacteria with the latest
67 recombination/mutation rate reported at 47.93 per site (Sánchez-Busó, Comas, Jorques, &
68 González-Candelas, 2014)(Gomez-Valero et al., 2011) coupled with a suite of mobile genetic
69 elements, including plasmids and pathogenicity islands (Khodr et al., 2016; Wee, Woolfit,
70 Beatson, & Petty, 2013)(Mercante, Morrison, Desai, Raphael, & Winchell, 2016)(Chien et
71 al., 2004). These mobile genetic elements encode virulence and fitness factors, such as type

72 IV secretion systems, known to be imperative for intracellular replication and survival in both
73 amoebae and human macrophages (Gomez-Valero & Buchrieser, 2013)(Voth, Broederdorf,
74 & Graham, 2012). While gene sequencing based typing (SBT) methods have been used for
75 outbreak investigation (Gaia et al., 2005), the limited resolution of SBT have led to a wider
76 application of high resolution whole genome sequencing (WGS) in the investigation of
77 community outbreaks of legionellosis. WGS technological advancements now provide
78 increased discrimination of outbreak isolates especially for endemic clones such as ST1
79 (Bartley et al., 2016; Reuter et al., 2013)(Graham, Doyle, & Jennison, 2014). This study was
80 aimed to elucidate the source and probable increased persistence of the complex Sydney 2016
81 outbreak by examining the genomes of clinical and environmental *L. pneumophila* isolates
82 collected over that period.

83

84 **Methods**

85 **Bacterial isolates and serogrouping.** Between February and May 2016, cases were
86 identified by positive urinary Lp1 antigen results and/or positive respiratory tract cultures for
87 Lp1 and clinical findings (details available on request). Clinical and environmental Lp1
88 isolates from three geographically distinct locations, the Central Business District (CBD), 18
89 km south of the CBD (Suburb 1) and 12 km west of the CBD (Suburb 2), were studied (Table
90 1 and Figure 1). Isolates from CBD spanned February to March 2016 (CBD March outbreak)
91 and from April to May (CBD May outbreak). There was a period of six weeks where no cases
92 from the CBD were detected between March and April. Environmental isolates from cooling
93 towers were referred from the Forensic and Analytical Science Service, whilst clinical
94 isolates were cultured from sputum samples or bronchoalveolar lavage fluid referred to the
95 Centre for Infectious Diseases and Microbiology Laboratory Services, ICPMR-Pathology

96 West, Westmead Hospital, Sydney. Isolates were grown on buffered charcoal yeast extract
97 (BCYE) agar and incubated at 37°C for up to 7 days and identified by standard phenotypic
98 characteristics (Versalovic et al., 2011). Serogrouping of each isolate was performed using
99 the Legionella Latex Test (Oxoid, ThermoFisher Scientific, North Ryde, NSW, Australia).
100 The full co-ordinated public health response is detailed in the Public Health Investigation into
101 the Legionella outbreaks in Sydney CBD (Griffiths et al., 2016).

102 For whole genome analysis, the type strains *L. pneumophila* ATCC 33152
103 (Philadelphia AE017354) and *L. pneumophila* ATCC33215 (Chicago 2 933085) were
104 included for comparison of assembly and typing pipelines. In addition, two Lp1 clinical
105 isolates from previous unrelated outbreaks in Sydney (2013 Reference and 2015 Reference)
106 were included as unrelated outgroups.

107 **DNA extraction and WGS.** Genomic DNA was extracted from pure cultures using the
108 DNeasy Blood & Tissue Kit (QIAGEN, Chadstone, VIC, Australia) with a 3 hour Proteinase
109 K digestion at 56°C. The quality of each DNA sample was measured using a Nanodrop ND-
110 1000 spectrophotometer (Nanodrop Technologies, ThermoFisher Scientific) and the ratios
111 260/280 nm and 260/230 nm inspected for the presence of organic matter and solvent
112 residues, respectively. DNA purity was determined as satisfactory if the 260/280 ratio fell
113 between 1.6 and 2.2 and the 260/230 ratio fell between 1.8 and 2.2. Purity of each DNA
114 sample was also inspected on 1.5% agarose gel. Paired-end indexed libraries of 150 bp in
115 length were prepared from an input of 1 ng of purified DNA with the Nextera XT library
116 preparation kit (Illumina, Scoresby, VIC, Australia) as per manufacturer's instructions. DNA
117 libraries were then sequenced using the NextSeq 500 (Illumina).

118 **Genome assembly and analysis.** To identify SNPs from whole genome sequencing data,
119 FASTQ files were imported into Geneious (version 8.0.4) and mapped to a curated reference
120 of *L. pneumophila* Philadelphia (accession number NC_002942) using the bwa plugin

121 (version 0.7.10) with bacteriophage, insertion sequences and other repeat regions removed
122 according to a previous publication (Coil et al., 2008). Quality based variant detection was
123 performed using CLC Genomics Workbench v 7.0 (CLC bio Aarhus, Denmark). Variant
124 detection thresholds were set for a minimum coverage of 10 and minimum variant frequency
125 of 75%. SNPs were excluded if they were in regions with a minimum fold coverage of <10,
126 within 10-bp of another SNP or <15-bp from the end of a contig. Maximum likelihood
127 phylogenetic trees were constructed from SNP matrices using the GTR model with 100
128 bootstrap replications.

129 Sequencing reads were assembled with Spades (Bankevich et al., 2012) and annotated
130 with Prokka (Seemann, 2014). Sequence-based typing (SBT) was performed with seven loci
131 by uploading identified alleles to the database and obtaining loci number (Gaia et al., 2005).
132 In addition, core genome multi-locus sequence typing (cgMLST) was conducted using the
133 Ridom SeqSphere software (Qiagen) employing gene definitions determined from closed or
134 complete *L. pneumophila* genomes from NCBI GenBank (Moran-Gilad et al., 2015). Using
135 these criteria, the core genome was determined to be 1530 genes with an accessory genome
136 of 1370 genes.

137 Further pangenome assessment and visualisation was performed using the Roary
138 pipeline (Page et al., 2015) that includes alignment using MAFFT (Kato, Misawa, Kuma, &
139 Miyata, 2002) and tree building with FastTree (Price et al., 2010). Specific protein and
140 nucleotide comparisons were made (including the unique T4SS) to all available Legionella
141 genomes on NCBI using the relevant Blast database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

142 The genomic data have been deposited in the NCBI Sequence Read Archive (SRA)
143 (<http://www.ncbi.nlm.nih.gov/Traces/sra/>) under accession number (TBA).

144

145 **Results**

146 Four outbreaks of severe community-acquired pneumonia in adults with positive urinary
147 antigen for Lp1 were identified in metropolitan Sydney in the first half of 2016. Nine human
148 cases were linked to the CBD March outbreak and six cases in the CBD May outbreak, all
149 with common exposure links to the Sydney CBD. Detailed epidemiological information on
150 the CBD cases can be found in the Public Health Investigation Report (Griffiths et al., 2016).
151 Four cases were linked in April to Suburb 1. Also in May, five cases were linked to Suburb 2.
152 One case (Case 12) in May was linked to both Suburb 2 and the CBD (Figures 2 and 3). Lp1
153 was cultured from respiratory samples of twelve patients and from fourteen cooling towers.
154 All clinical and environmental isolates collected, together with two control strains and two
155 Lp1 isolates obtained from epidemiologically unrelated cases were subjected to genome
156 sequencing (Table 1). All 48 genomes were *de novo* assembled and their genome size ranged
157 between 3.2-3.4 Mb with a G+C of 39%. The core genome as determined by the Roary
158 pipeline, of this dataset consisted of 2181 genes. The accessory genome had 2273 genes out
159 of a total of 5919 genes (including 99 soft core genes and 1366 shell genes that were not
160 counted in the core and accessory respectively).

161 Genome comparisons based on SNP-based mapping, cgMLST and pangenome
162 produced congruent results and confirmed phylogenetic relatedness between isolates
163 associated with outbreaks in the CBD (March and May) and Suburb 1 (Figures 2 and 3 and
164 Supplementary Figure 1) and these isolates were designated “Main cluster”. The CBD March
165 outbreak showed two patients, Cases 1 and 3 clustered together (three isolates were recovered
166 from one patient) (Figures 2 and 3). For the next outbreak in Suburb 1, one patient clustered
167 with one tower and in the CDB May outbreak, isolates from two cooling towers and three
168 patients were of the same genetic cluster. All isolates from the main cluster were also typed
169 as ST211. Isolates from all patients linked to Suburb 2 (Cases 9, 10, 11 and 12) and from one

170 patient linked to each of the CBD (Case 2) and Suburb 1 outbreaks (Cases 4) appeared to be
171 genomically distinct and not related to the main cluster (Figure 2). The maximum SNP
172 difference between outbreak isolates was 87 between CBD tower isolates and Suburb 1
173 clinical isolates. There were 44 SNP differences between Suburb 1 tower isolates and Suburb
174 1 clinical isolates. Amongst all clinical isolates from the outbreak cluster, there was an
175 average difference of 66 SNPs and isolates from the same case had less than 45 SNPs. The
176 two towers harbouring ST211 strains were approximately 800m apart on different buildings.

177 Pangenome analysis further revealed three regions of interest related to the outbreak
178 cluster (Figure 3). The first (Region I) was a large region of 178 genes that was present in the
179 bottom clade only and included the main cluster. Region I was also found in all reference
180 strains, the two past outbreak isolates, closely related environmental strains BE3 and BE4 and
181 an outlier GC4. Region I included the genes for Dot/Icm, *luxR*, Ankyrin repeats and beta-
182 lactamase among others indicating that this was a Type IV B secretion system (T4BSS). The
183 second region (Region II) consisted of 58 genes that were unique to the outbreak cluster, the
184 outlier patient isolates BC9 and BC10 co-circulating at the same time and two closely related
185 environmental isolates, BE3 and BE4 (Figure 3). This was found to carry another T4SS, an A
186 F-type secretion system (T4ASS) (Supplementary Table 1) and this region had a higher G+C
187 content of 42% compared to the rest of the genome. Using BLAST this region was found to
188 have between 99-100% homology to similar conjugative elements in only four other *L.*
189 *pneumophila* Lp1 strains, two closed genomes, LPE509 (NC_020521.1) and C1-S
190 (CP05932.1) and two further clinical isolates, one ST37 isolated in 2003 in the U.K.
191 (LT632617.1) and the other an ST42 clinical isolate from Germany (LT632616.1). This
192 unique region is 73-kb with an insertion site adjacent to three tRNA genes (tRNA^{Arg},
193 tRNA^{Lys}, tRNA^{Lys}). Two more small regions were unique to the outbreak cluster and closely

194 related isolates BE3 and BE4. These regions denoted as Region III contained mainly
195 hypothetical proteins, acytransferase genes and a calcium transporting ATPase.

196

197 **Discussion**

198 Our findings indicated that cases amongst four temporally and spatially separate outbreaks of
199 legionellosis that occurred in quick succession in metropolitan Sydney in 2016 were caused
200 by a common clone of Lp1 ST211. The geographical distance between sites of potential
201 exposure was much larger than previously recognised (Knox et al, 2016) and this coupled
202 with the temporal differences between these outbreaks led to the initial assumption that they
203 were related to separate breaches of environmental health controls. However, phylogenetic
204 analysis on both clinical and environmental isolates from these outbreaks by three
205 independent genomic approaches confirmed that a common Lp1 clone was responsible across
206 three of the four outbreaks. The CBD March outbreak was consonant with the Suburb 1
207 outbreak 18 km away and both of these were genetically related to the CBD May outbreak
208 that occurred six weeks later. Factors such as cooling tower maintenance and other
209 environmental considerations may have contributed to this wave of Lp1 disease in Sydney.

210 This report describes the direct comparison of the resolution power of SNP-based and
211 core genome combined with a pangenome based analyses in the investigation of Lp1
212 outbreaks (David, Mentasti, et al., 2016). Our experience suggested that these three
213 approaches have comparable discrimination power and should be supplemented by the
214 analysis of mobile genetic elements given the high recombination rates in Lp1 (Wang et al.,
215 2015)(Sánchez-Busó et al., 2014). Importantly, the outbreaks in Sydney demonstrated the
216 presence of Lp1 ST211 in the Southern Hemisphere. This ST has not been reported in the
217 USA or Europe and was thought to reside exclusively in the colder climate of Canada where

218 it was first identified in 1989 (Tijet et al., 2010). In Ontario, ST211 was initially thought to be
219 contributing to sporadic cases which peaked in 1999, however, further investigation found
220 that it was responsible for 12.5% of isolates, including eight times in the same hospital.
221 ST211 has since become one of the most persistent and predominant STs in Ontario and
222 along with other STs, has almost completely replaced the previously dominant Lp1 ST1.
223 There is also a suggestion that ST1 may be the predominant ST in Australia (Graham et al.,
224 2014) and further investigations will determine if ST211 is replacing ST1 as the dominant
225 clone in Sydney.

226 In addition to the assessment of potential relationships between clinical cases and
227 environmental sources, WGS of *Legionella* isolates enables the examination of diversity and
228 genomic structures that may have augmented their persistence in the environment (Sánchez-
229 Busó et al., 2016). Our pangenome analysis revealed a unique conjugative Tra (F-type)
230 element carrying a T4ASS and this was present in all outbreak isolates but more significantly
231 it was also found in closely related environmental isolates and isolates that infected Case 2 at
232 the same time as the CBD March outbreak. Interestingly, this element had 99% nucleotide
233 homology to four previously reported clinical strains from the U.S.A. and Europe and an
234 environmental isolate from Japan, although none of these other isolates are ST211 but are
235 STs in the top five of outbreak causing clones (David, Rusniok, et al., 2016).

236 The T4SS of *Legionella* are part of the dynamic accessory genome, known to
237 contribute to fitness and virulence and play a crucial role in intracellular replication and
238 survival (Khodr et al., 2016)(Rolando & Buchrieser, 2014; Schroeder et al., 2010; Voth et al.,
239 2012). Unique T4ASS have been found in previous outbreak isolates (Graham et al., 2014)
240 including from a recent outbreak in Western Canada, unique in its dry, cold conditions that
241 were initially thought to be too harsh for survival of *L. pneumophila* (Knox et al., 2016). The
242 element described in this study contained genes homologous to the Lvh region of other *L.*

243 *pneumophila* strains and genes from this region are thought to assist in intracellular
244 replication (Bandyopadhyay, Liu, Gabbai, Venitelli, & Steinman, 2007)(Bandyopadhyay,
245 Lang, Rasaputra, & Steinman, 2013). In addition, it also contained *csrA*, a carbon storage
246 regulator known to control the switch from replicative to transmissive phase in *Legionella*
247 which is crucial for efficient replication both *in vitro* and intracellularly (Fettes, Forsbach-
248 Birk, Lynch, & Marre, 2001)(Forsbach-Birk, Mcnealy, Shi, Lynch, & Marre, 2004; Molofsky
249 & Swanson, 2003). It is possible that these genomic features reflect high recombination
250 potential and fitness of Lp1 ST211 and help to explain its emergence as an outbreak clone.
251 Earlier reports attempted to define outbreak strains from non-outbreak strains and genome
252 sequencing has identified recombination as a major contributor to Lp1 variability which
253 occurs across all strains, although some STs appear to have different recombination rates
254 (Sánchez-Busó et al., 2014). The genomes described in this study are the only publicly
255 available ST211 genomes that can be used for further outbreak comparisons given the
256 isolates from Canada were typed using the SBT scheme only and not with WGS.

257 The identification of elements such as the T4SSA described here can alert
258 investigators to the presence of a *L. pneumophila* clone with possible superior virulence and
259 persistence abilities. In addition, given the dependency of *L. pneumophila* on T4SS for
260 intracellular survival and replication, these systems offer an potential target for antibacterial
261 agents and vaccines (Voth et al., 2012).

262 In conclusion, SNP-based, core genome and pan genome based analyses of Lp1
263 isolates can assist in deciphering and confirming transmission pathways during the
264 investigation of complex outbreaks of legionellosis. Comparative genomics of clinical and
265 environmental isolates of Lp1 suggested that the emerging ST211 clone was responsible for
266 three sequential outbreaks of legionellosis in metropolitan Sydney separated geographically
267 and temporally over seven weeks. Significantly, pangenome analysis identified mobile

268 genetic elements that included a T4ASS in the outbreak strain and co-circulating strains that
269 may have augmented the fitness and persistence of the outbreak clone in the environment.

270

271

272 **Acknowledgements**

273 The authors thank public health professionals from NSW Health and the City of Sydney, and
274 laboratory scientists from NSW Health Pathology for providing clinical and environmental
275 isolates and their assistance in collecting epidemiological and environmental information.

276

277 **References**

278 Ambrose, J., Hampton, L. M., Fleming-Dutra, K. E., Marten, C., McClusky, C., Perry, C., ...
279 Cersovsky, S. B. (2014). Large outbreak of Legionnaires' disease and Pontiac fever at a
280 military base. *Epidemiol Infect*, *142*(11), 2336–2346.

281 <https://doi.org/10.1017/s0950268813003440>

282 Bandyopadhyay, P., Lang, E. A. S., Rasaputra, K. S., & Steinman, H. M. (2013). Implication
283 of the VirD4 coupling protein of the Lvh type 4 secretion system in virulence
284 phenotypes of *Legionella pneumophila*. *Journal of Bacteriology*, *195*(15), 3468–75.

285 <https://doi.org/10.1128/JB.00430-13>

286 Bandyopadhyay, P., Liu, S., Gabbai, C. B., Venitelli, Z., & Steinman, H. M. (2007).
287 Environmental Mimics and the Lvh Type IVA Secretion System Contribute to
288 Virulence-Related Phenotypes of *Legionella pneumophila*. *INFECTION AND*

289 *IMMUNITY*, *75*(2), 723–735. <https://doi.org/10.1128/IAI.00956-06>

290 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ...

- 291 Pevzner, P. a. (2012). SPAdes: A New Genome Assembly Algorithm and Its
292 Applications to Single-Cell Sequencing. *Journal of Computational Biology*, 19(5), 455–
293 477. <https://doi.org/10.1089/cmb.2012.0021>
- 294 Bartley, P. B., Ben Zakour, N. L., Stanton-Cook, M., Muguli, R., Prado, L., Garnys, V., ...
295 Beatson, S. A. (2016). Hospital-wide eradication of a nosocomial *Legionella*
296 *pneumophila* serogroup 1 outbreak. *Clinical Infectious Diseases*.
297 <https://doi.org/10.1093/cid/civ870>
- 298 Chien, M., Morozova, I., Shi, S., Sheng, H., Chen, J., Gomez, S. M., ... Russo, J. J. (2004).
299 The Genomic Sequence of the Accidental Pathogen *Legionella pneumophila*. *Science*,
300 305(5692).
- 301 Coil, D. a, Vandersmissen, L., Ginevra, C., Jarraud, S., Lammertyn, E., & Anné, J. (2008).
302 Intragenic tandem repeat variation between *Legionella pneumophila* strains. *BMC*
303 *Microbiology*, 8, 218. <https://doi.org/10.1186/1471-2180-8-218>
- 304 David, S., Mentasti, M., Tewolde, R., Aslett, M., Harris, S. R., Afshar, B., ... Harrison, T. G.
305 (2016). Evaluation of an optimal epidemiologic typing scheme for *Legionella*
306 *pneumophila* with whole genome sequence data using validation guidelines. *Journal of*
307 *Clinical Microbiology*, 54(June), JCM.00432-16. <https://doi.org/10.1128/JCM.00432-16>
- 308 David, S., Rusniok, C., Mentasti, M., Gomez-Valero, L., Harris, S. R., Lechat, P., ...
309 Buchrieser, C. (2016). Multiple major disease-associated clones of *Legionella*
310 *pneumophila* have emerged recently and independently. *Genome Research*, 26(11),
311 1555–1564. <https://doi.org/10.1101/gr.209536.116>
- 312 Fettes, P. S., Forsbach-Birk, V., Lynch, D., & Marre, R. (2001). Overexpression of a
313 *Legionella pneumophila* homologue of the *E. coli* regulator *csrA* affects cell size,
314 flagellation, and pigmentation. *International Journal of Medical Microbiology* □: *IJMM*,

- 315 291(5), 353–60. <https://doi.org/10.1078/1438-4221-00141>
- 316 Forsbach-Birk, V., Mcnealy, T., Shi, C., Lynch, D., & Marre, R. (2004). Reduced expression
317 of the global regulator protein CsrA in *Legionella pneumophila* affects virulence-
318 associated regulators and growth in *Acanthamoeba castellanii*. *International Journal of*
319 *Medical Microbiology*, 294, 15–25. <https://doi.org/10.1016/j.ijmm.2003.12.003>
- 320 Fraser, D. W., Tsai, T. R., Orenstein, W., Parkin, W. E., Beecham, H. J., Sharrar, R. G., ...
321 Team*, the F. I. (1977). Legionnaires' Disease. *New England Journal of Medicine*,
322 297(22), 1189–1197. <https://doi.org/10.1056/NEJM197712012972201>
- 323 Frieden, T. R., Jaffe, H. W., Stephens, J. W., & Thacker, S. B. (2011). Centers for Disease
324 Control and Prevention MMWR Editorial and Production Staff. *MMWR*, 60.
- 325 Gaia, V., Fry, N. K., Afshar, B., Lück, P. C., Meugnier, H., Etienne, J., ... Harrison, T. G.
326 (2005). Consensus sequence-based scheme for epidemiological typing of clinical and
327 environmental isolates of *Legionella pneumophila*. *Journal of Clinical Microbiology*,
328 43(5), 2047–52. <https://doi.org/10.1128/JCM.43.5.2047-2052.2005>
- 329 Gomez-Valero, L., & Buchrieser, C. (2013). Genome dynamics in Legionella: The basis of
330 versatility and adaptation to intracellular replication. *Cold Spring Harbor Perspectives*
331 *in Medicine*, 3(6). <https://doi.org/10.1101/cshperspect.a009993>
- 332 Gomez-Valero, L., Rusniok, C., Jarraud, S., Vacherie, B., Rouy, Z., Barbe, V., ... Gibbs, A.
333 (2011). Extensive recombination events and horizontal gene transfer shaped the
334 *Legionella pneumophila* genomes. *BMC Genomics*, 12(1), 536.
335 <https://doi.org/10.1186/1471-2164-12-536>
- 336 Graham, R. M., Doyle, C. J., & Jennison, A. V. (2014). Real-time investigation of a
337 *Legionella pneumophila* outbreak using whole genome sequencing. *Epidemiology and*
338 *Infection*, 142(11), 2347–51. <https://doi.org/10.1017/S0950268814000375>

- 339 Griffiths, E., Gilmour, R., Carlin, M., Adamson, S., Timms, V. J., Chen, S., ... Mcanulty, J.
340 (2016). *Public Health Investigation into the Legionella Outbreaks in Sydney CBD*.
341 Sydney.
- 342 Katoh, K., Misawa, K., Kuma, K.-I., & Miyata, T. (2002). MAFFT: a novel method for rapid
343 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*,
344 30(14), 3059–3066.
- 345 Khodr, A., Kay, E., Gomez-Valero, L., Ginevra, C., Doublet, P., Buchrieser, C., & Jarraud, S.
346 (2016). Molecular epidemiology, phylogeny and evolution of Legionella. *Infection*,
347 *Genetics and Evolution*, 43, 108–122. <https://doi.org/10.1016/j.meegid.2016.04.033>
- 348 Knox, N. C., Weedmark, K. A., Conly, J., Ensminger, A. W., Hosein, F. S., & Drews, S. J.
349 (2016). Unusual Legionnaires' outbreak in cool, dry Western Canada: an investigation
350 using genomic epidemiology. *Epidemiology and Infection*, 145(July 1976), 1–12.
351 <https://doi.org/10.1017/S0950268816001965>
- 352 Kozak-Muiznieks, N. A., Lucas, C. E., Brown, E., Pondo, T., Taylor, T. H., Frace, M., ...
353 Winchell, J. M. (2014). Prevalence of Sequence Types among Clinical and
354 Environmental Isolates of *Legionella pneumophila* Serogroup 1 in the United States
355 from 1982 to 2012. *Journal of Clinical Microbiology*, 52(1), 201–211.
356 <https://doi.org/10.1128/JCM.01973-13>
- 357 Mercante, J. W., Morrison, S. S., Desai, H. P., Raphael, B. H., & Winchell, J. M. (2016).
358 Genomic Analysis Reveals Novel Diversity among the 1976 Philadelphia Legionnaires'
359 Disease Outbreak Isolates and Additional ST36 Strains. *PLoS ONE*, 11(9).
360 <https://doi.org/10.1371/journal.pone.0164074>
- 361 Molofsky, A. B., & Swanson, M. S. (2003). *Legionella pneumophila* CsrA is a pivotal
362 repressor of transmission traits and activator of replication. *Molecular Microbiology*,

- 363 50(2), 445–61.
- 364 Moran-Gilad, J., Prior, K., Yakunin, E., Harrison, T. G., Underwood, A., Lazarovitch, T., ...
365 Harmsen, D. (2015). Design and application of a core genome multilocus sequence
366 typing scheme for investigation of Legionnaires disease incidents. *Eurosurveillance*.
367 <https://doi.org/10.2807/1560-7917.ES2015.20.28.21186>
- 368 Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., ...
369 Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis.
370 *Bioinformatics (Oxford, England)*, 31(22), 3691–3.
371 <https://doi.org/10.1093/bioinformatics/btv421>
- 372 Price, M. N., Dehal, P. S., Arkin, A. P., Nawrocki, E., Kolbe, D., Eddy, S., ... Meyer, F.
373 (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments.
374 *PLoS ONE*, 5(3), e9490. <https://doi.org/10.1371/journal.pone.0009490>
- 375 Raphael, B. H., Baker, D. J., Nazarian, E., Lapierre, P., Bopp, D., Kozak-Muiznieks, N. A.,
376 ... Winchell, J. M. (2016). Genomic Resolution of Outbreak-Associated *Legionella*
377 *pneumophila* Serogroup 1 Isolates from New York State. *Applied Environmental*
378 *Microbiology*, 82(12), 3582–3590. <https://doi.org/10.1128/AEM.00362-16>
- 379 Reuter, S., Harrison, T. G., Köser, C. U., Ellington, M. J., Smith, G. P., Parkhill, J., ... Török,
380 M. E. (2013). A pilot study of rapid whole-genome sequencing for the investigation of a
381 *Legionella* outbreak. *BMJ Open*, 3(1), e002175. [https://doi.org/10.1136/bmjopen-2012-](https://doi.org/10.1136/bmjopen-2012-002175)
382 [002175](https://doi.org/10.1136/bmjopen-2012-002175)
- 383 Rolando, M., & Buchrieser, C. (2014). *Legionella pneumophila* type IV effectors hijack the
384 transcription and translation machinery of the host cell. *Trends in Cell Biology*, 24(12),
385 771–778. <https://doi.org/10.1016/j.tcb.2014.06.002>
- 386 Sánchez-Busó, L., Comas, I., Jorques, G., & González-Candelas, F. (2014). Recombination

- 387 drives genome evolution in outbreak-related *Legionella pneumophila* isolates. *Nature*
388 *Genetics*, 46(11), 1205–1211. <https://doi.org/10.1038/ng.3114>
- 389 Sánchez-Busó, L., Guiral, S., Crespi, S., Moya, V., Camaró, M. L., Olmos, M. P., ...
390 González-Candelas, F. (2016). Genomic investigation of a legionellosis outbreak in a
391 persistently colonized hotel. *Frontiers in Microbiology*, 6(JAN).
392 <https://doi.org/10.3389/fmicb.2015.01556>
- 393 Schroeder, G. N., Petty, N. K., Mousnier, A., Harding, C. R., Vogrin, A. J., Wee, B., ...
394 Frankel, G. (2010). *Legionella pneumophila* strain 130b possesses a unique combination
395 of type IV secretion systems and novel Dot/Icm secretion system effector proteins.
396 *Journal of Bacteriology*, 192(22), 6001–16. <https://doi.org/10.1128/JB.00778-10>
- 397 Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14),
398 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- 399 Tijet, N., Tang, P., Romilowych, M., Duncan, C., Ng, V., Fisman, D. N., ... Guyard, C.
400 (2010). New endemic *Legionella pneumophila* serogroup I clones, Ontario, Canada.
401 *Emerging Infectious Diseases*, 16(3), 447–454. <https://doi.org/10.3201/eid1603.081689>
- 402 Versalovic, J., Jorgensen, J. H., Funke, G., Warnock, D. W., Landry, M. L., & Carroll, K. C.
403 (Eds.). (2011). *Manual of Clinical Microbiology, 10th Edition*. American Society of
404 Microbiology. <https://doi.org/10.1128/9781555816728>
- 405 Voth, D. E., Broederdorf, L. J., & Graham, J. G. (2012). Bacterial Type IV Secretion
406 Systems: Versatile Virulence Machines. *Future Microbiology*, 7(2), 241–57.
407 <https://doi.org/10.2217/fmb.11.150>
- 408 Wang, Q., Holmes, N., Martinez, E., Howard, P., Hill-Cawthorne, G., & Sintchenko, V.
409 (2015). It is not all about single nucleotide polymorphisms: Comparison of mobile
410 genetic elements and deletions in *Listeria monocytogenes* genomes links cases of

411 hospital-acquired listeriosis to the environmental source. *Journal of Clinical*
412 *Microbiology*, 53(11), 3492–3500. <https://doi.org/10.1128/JCM.00202-15>

413 Wee, B. A., Woolfit, M., Beatson, S. A., & Petty, N. K. (2013). A Distinct and Divergent
414 Lineage of Genomic Island-Associated Type IV Secretion Systems in *Legionella*.
415 <https://doi.org/10.1371/journal.pone.0082221>

416

417 **Table 1.** *Legionella pneumophila* serogroup 1 isolates included in the study.

Isolate identification number	Origin	Case number in outbreak	Association with location	SBT type
BC7	Clinical	Case 1	CBD March*	211
BC10	Clinical	Case 2	CBD March	42
BC5	Clinical	Case 3	CBD March	211
Reference control 1	Clinical	ATCC33152		
Reference control 2	Clinical	ATCC33215		
BC2	Clinical	Case 1	CBD March	211
BC1	Clinical	Case 1	CBD March	211
BC9	Clinical	Case 2	CBD March	42
Reference 2013	Clinical	Reference case 1	Not related	762
Reference 2015	Clinical	Reference case 2	Not related	ND
BE3	Environmental	Tower 1	CBD March	ND
BE11	Environmental	Tower 2	CBD March	1
BE12	Environmental	Tower 3	CBD March	1
BE13	Environmental	Tower 4	CBD March	1
BE7	Environmental	Tower 5	CBD March	1
BE5	Environmental	Tower 6	CBD March	1
BE17	Environmental	Tower 6	CBD March	1
BE20	Environmental	Tower 7	CBD March	284
BE19	Environmental	Tower 8	CBD March	1
BE14	Environmental	Tower 9	CBD March	1
BE10	Environmental	Tower 9	CBD March	1
BE6	Environmental	Tower 9	CBD March	1
BE8	Environmental	Tower 10	CBD March	1
RC3	Clinical	Case 4	Suburb 1	ND
RC1	Clinical	Case 5	Suburb 1	211
RC2	Clinical	Case 5	Suburb 1	211
RE6	Environmental	Tower 11	Suburb 1	211
RE4	Environmental	Tower 11	Suburb 1	211

RE7	Environmental	Tower 11	Suburb 1	211
RE5	Environmental	Tower 11	Suburb 1	211
RE18	Environmental	Tower 12	Suburb 1	1
RE2	Environmental	Tower 12	Suburb 1	1
RE16	Environmental	Tower 12	Suburb 1	1
RE15	Environmental	Tower 12	Suburb 1	1
RE3	Environmental	Tower 12	Suburb 1	1
RE1	Environmental	Tower 10	Suburb 1	284
BC6	Clinical	Case 6	CBD May	211
BC4	Clinical	Case 7	CBD May	211
BC3	Clinical	Case 7	CBD May	211
BC8	Clinical	Case 8	CBD May	211
BE4	Environmental	Tower 13	CBD May	ND
BE1	Environmental	Tower 14	CBD May	211
BE9	Environmental	Tower 9	CBD May	1
BE2	Environmental	Tower 14	CBD May	211
GC3	Clinical	Case 9	Suburb 2	1983
GC2	Clinical	Case 10	Suburb 2	ND
GC1	Clinical	Case 11	Suburb 2	84
GC4	Clinical	Case 12	Suburb 2/CBD May	ND

418 *Abbreviations: CBD March: Isolates from the Central Business District during February to

419 March 2016. CBD May, isolates from the Central Business District between April to May

420 2016.

421 **ND – Not Determined due to one or more alleles being new or combination of alleles being

422 new.

423

424 **Figure 1:** Map of metropolitan Sydney indicating locations of the outbreak.

425

426 **Figure 2:** SNP-based mapping phylogeny of all outbreak isolates between February-May
427 2016 in Sydney.

428

429 **Figure 3:** The pan genome of the complete Sydney 2016 dataset (n=48). **A.** Maximum
430 likelihood tree showing outbreak cluster as clade 3. **B.** Pangenome sorted from core genes on
431 the left to accessory genes to the right. **C.** Heatmap showing presence (blue) and absence
432 (white) of genomic regions. The T4SS unique to the outbreak cluster (Region II), the outlier
433 (BC9 and BC10 in blue) and related environmental strains (BE3 and BE4 in blue) are
434 indicated by red circles. Region I demonstrates genes that were unique to clade 3 including
435 reference isolates and the outbreak cluster. Region III indicates genes that were unique to the
436 outbreak cluster and related environmental strains BE3 and BE4 only.

437

438 **Supplementary Table 1.** Genes present in T4SS found exclusively in outbreak cluster, BC9,
439 BC10 (Case 2), BE3 and BE4.

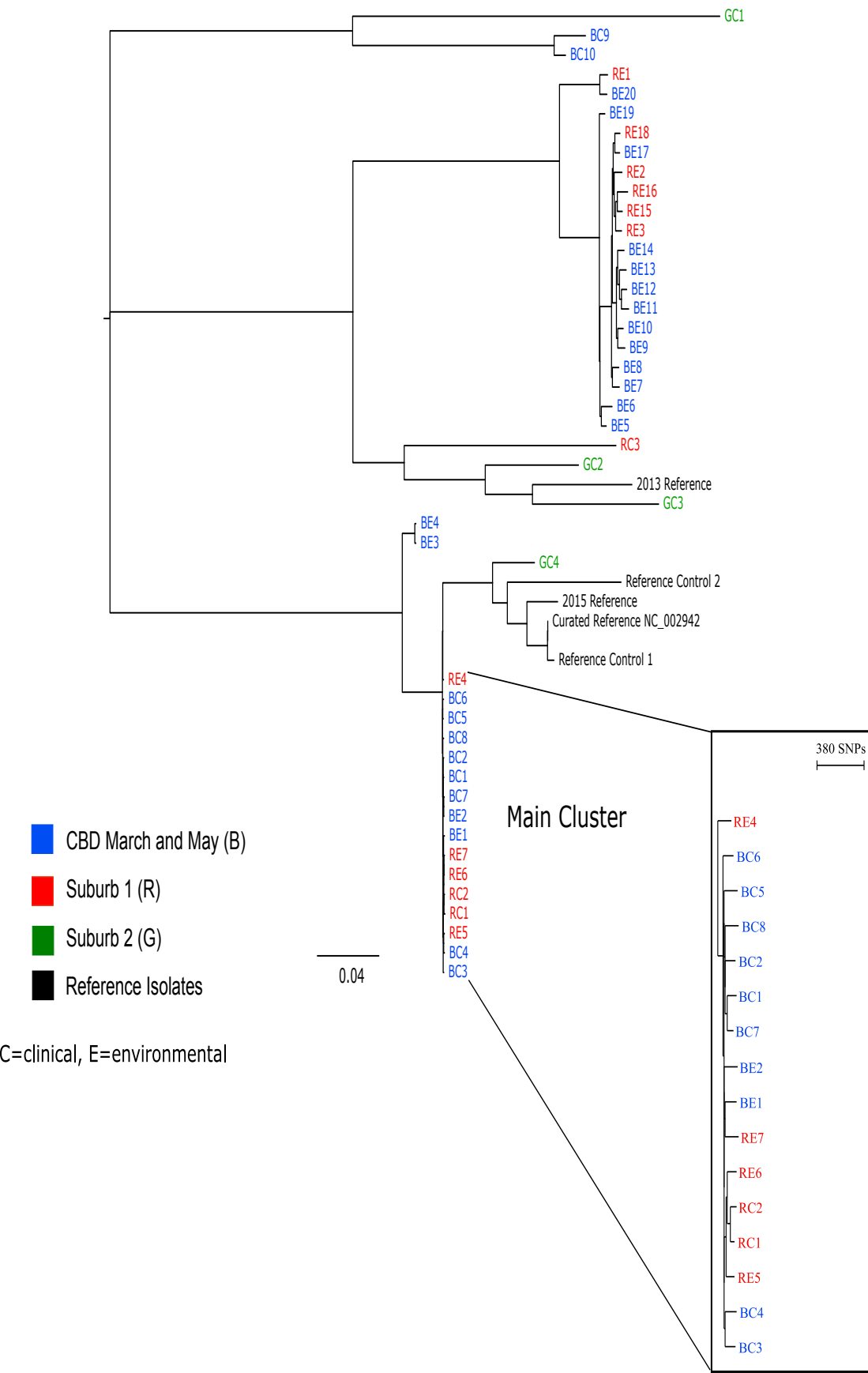
440 **Supplementary Figure 1:** cgMLST phylogeny of all outbreak isolates between February-
441 May 2016 in Sydney.

442

Figure 1: Map of metropolitan Sydney indicating locations of the Lp1 outbreak



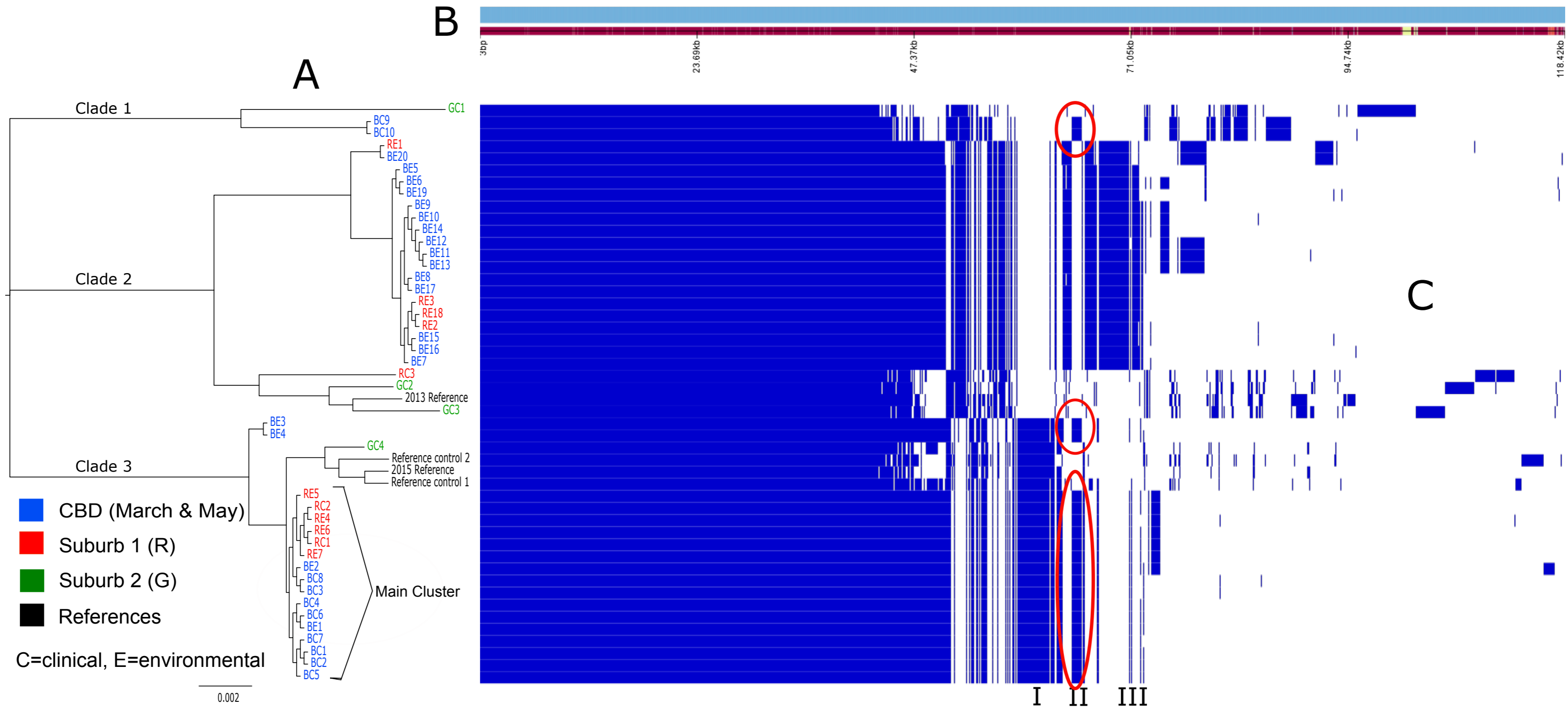
Figure 1: SNP-based mapping phylogeny of all 48 Lp1 outbreak isolates between February-May 2016 in Sydney



- CBD March and May (B)
- Suburb 1 (R)
- Suburb 2 (G)
- Reference Isolates

C=clinical, E=environmental

Figure 2: The pan genome of the complete Sydney 2016 Lp1 dataset (n=48)



A. Maximum likelihood tree showing outbreak cluster as clade 3. B. Pangenome sorted from core genes on the left to accessory genes to the right.

C. Heatmap showing presence (blue) and absence (white) of genomic regions. The T4SS unique to the outbreak cluster (Region II), the outlier (BC9 and BC10 in blue) and related environmental strains (BE3 and BE4 in blue) are indicated by red circles.

Region I demonstrates genes that were unique to clade 3 including reference isolates and the outbreak cluster. Region III indicates genes that were unique to the outbreak cluster and related environmental strains BE3 and BE4 only.