Translating genomics into practice for real-time surveillance and response to carbapenemase-producing Enterobacteriaceae: evidence from a complex multi-institutional KPC outbreak

Jason C Kwong [1, 2, 3] , **Courtney Lane** [4, 5] , **Finn Romanes** [5] , **Anders Gonçalves da Silva** [1, 4] , **Marion Easton** [5] , **Katie Cronin** [6] , **Mary Jo Waters** [6] , **Takehiro Tomita** [4] , **Kerrie Stevens** [4] , **Mark B Schultz** [1, 4] , **Sarah L Baines** [1, 2] , **Norelle L Sherry** [1, 2, 3] , **Glen Carter** [1, 2] , **Andre Mu** [1, 2] , **Michelle Sait** [4] , **Susan A Ballard** [1, 4] , **Torsten Seemann** [1, 7] , **Timothy P Stinear** [1, 2] , **Benjamin P Howden** Corresp. [1, 2, 3, 4]

1 Doherty Applied Microbial Genomics, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

2 Department of Microbiology and Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

3 Department of Infectious Diseases, Austin Health, Heidelberg, Australia

4 Microbiological Diagnostic Unit Public Health Laboratory, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

5 Health Protection Branch, Department of Health and Human Services, Victoria State Government, Melbourne, Australia

6 Department of Microbiology, St Vincent's Hospital Melbourne, Fitzroy, Australia

7 Melbourne Bioinformatics, The University of Melbourne, Carlton, Australia

Corresponding Author: Benjamin P Howden
Email address: bhowden@unimelb.edu.au

Background. Until recently, KPC-producing Enterobacteriaceae were rarely identified in Australia. Following an increase in the number of incident cases across the state of Victoria, we undertook a real-time combined genomic and epidemiological investigation. The scope of this study included identifying risk factors and routes of transmission, and investigating the utility of genomics to enhance traditional field epidemiology for informing management of established widespread outbreaks.

Methods. All KPC-producing Enterobacteriaceae isolates referred to the state reference laboratory from 2012 onwards were included. Whole-genome sequencing (WGS) was performed in parallel with a detailed descriptive epidemiological investigation of each case, using Illumina sequencing on each isolate. This was complemented with PacBio long-read sequencing on selected isolates to establish high-quality reference sequences and interrogate characteristics of KPC-encoding plasmids.

Results. Initial investigations indicated the outbreak was widespread, with 86 KPC-producing Enterobacteriaceae isolates (*K. pneumoniae* 92%) identified from 35 different locations across metropolitan and rural Victoria between 2012-2015. Initial combined analyses of the epidemiological and genomic data resolved the outbreak into distinct nosocomial transmission networks, and identified healthcare facilities at the epicentre of KPC transmission. New cases were assigned to transmission networks in real-time, allowing focussed infection control efforts. PacBio sequencing confirmed a secondary transmission network arising from inter-species plasmid transmission. Insights from Bayesian transmission inference and analyses of within-host diversity informed the development of state-wide public health and infection control guidelines, including interventions such as an intensive approach to screening contacts following new case detection to minimise unrecognised colonisation.

Conclusions. A real-time combined epidemiological and genomic investigation proved critical to

identifying and defining multiple transmission networks of KPC Enterobacteriaceae, while data from either investigation alone were inconclusive. The investigation was fundamental to informing infection control measures in real-time and the development of state-wide public health guidelines on carbapenemase producing Enterobacteriaceae surveillance and management.

1 **Full Title: Translating genomics into practice for real-time surveillance and response to**

2 **carbapenemase-producing Enterobacteriaceae: evidence from a complex multi-institutional**

3 **KPC outbreak**

5 **Authors:** Jason C Kwong [1,2,3], Courtney Lane [4,5], Finn Romanes [5], Anders Gonçalves da Silva [1,4],

6 Marion Easton [5], Katie Cronin [6], Mary Jo Waters [6], Takehiro Tomita [4], Kerrie Stevens [4], Mark B

7 Schultz [1,4], Sarah L Baines [1,2], Norelle L Sherry [1,2,3], Glen Carter [1,2], Andre Mu [1,2], Michelle Sait

8 [4], Susan A Ballard [1,4], Torsten Seemann [1,7], Timothy P Stinear [1,2], Benjamin P Howden [1,2,3,4]

10 [1] Doherty Applied Microbial Genomics, and [2] Department of Microbiology and Immunology, The

11 University of Melbourne at The Peter Doherty Institute for Infection & Immunity, Melbourne,

12 Australia

13 [3] Department of Infectious Diseases, Austin Health, Heidelberg, Australia

14 [4] Microbiological Diagnostic Unit Public Health Laboratory, The University of Melbourne at The

15 Peter Doherty Institute for Infection & Immunity, Melbourne, Australia

16 [5] Health Protection Branch, Department of Health and Human Services, Victoria State

17 Government, Melbourne, Australia

18 [6] Department of Microbiology, St Vincent's Hospital Melbourne, Fitzroy, Australia

19 [7] Melbourne Bioinformatics, The University of Melbourne, Carlton, Australia

21 **Correspondence:**

22 Prof Benjamin P Howden

23 Microbiological Diagnostic Unit Public Health Laboratory

24   Doherty Institute for Infection & Immunity

25   792 Elizabeth Street

26   Melbourne, Victoria

27   Australia 3000

28   Email: bhowden@unimelb.edu.au

29   Phone: +61 3 8344 5701

30 **ABSTRACT**

31 **Background.** Until recently, KPC-producing Enterobacteriaceae were rarely identified in

32 Australia. Following an increase in the number of incident cases across the state of Victoria, we

33 undertook a real-time combined genomic and epidemiological investigation. The scope of this

34 study included identifying risk factors and routes of transmission, and investigating the utility of

35 genomics to enhance traditional field epidemiology for informing management of established

36 widespread outbreaks.

37 **Methods.** All KPC-producing Enterobacteriaceae isolates referred to the state reference laboratory

38 from 2012 onwards were included. Whole-genome sequencing (WGS) was performed in parallel

39 with a detailed descriptive epidemiological investigation of each case, using Illumina sequencing

40 on each isolate. This was complemented with PacBio long-read sequencing on selected isolates to

41 establish high-quality reference sequences and interrogate characteristics of KPC-encoding

42 plasmids.

43 **Results.** Initial investigations indicated the outbreak was widespread, with 86 KPC-producing

44 Enterobacteriaceae isolates (*K. pneumoniae* 92%) identified from 35 different locations across

45 metropolitan and rural Victoria between 2012-2015. Initial combined analyses of the

46 epidemiological and genomic data resolved the outbreak into distinct nosocomial transmission

47 networks, and identified healthcare facilities at the epicentre of KPC transmission. New cases were

48 assigned to transmission networks in real-time, allowing focused infection control efforts. PacBio

49 sequencing confirmed a secondary transmission network arising from inter-species plasmid

50 transmission. Insights from Bayesian transmission inference and analyses of within-host diversity

51 informed the development of state-wide public health and infection control guidelines, including

52   interventions such as an intensive approach to screening contacts following new case detection to

53   minimise unrecognised colonisation.

54   **Conclusions.** A real-time combined epidemiological and genomic investigation proved critical to

55   identifying and defining multiple transmission networks of KPC Enterobacteriaceae, while data

56   from either investigation alone were inconclusive. The investigation was fundamental to informing

57   infection control measures in real-time and the development of state-wide public health guidelines

58   on carbapenemase producing Enterobacteriaceae surveillance and management.

59 **INTRODUCTION**

60 Carbapenemase-producing Enterobacteriaceae (CPE) are among the most urgent antimicrobial

61 resistance threats worldwide (Centers for Disease Control and Prevention 2013). In addition to

62 producing the carbapenemase, capable of inactivating almost all beta-lactam antibiotics, including

63 penicillins, cephalosporins and carbapenems, these organisms frequently harbour multiple other

64 antibiotic resistance genes and mutations (Queenan & Bush 2007). Few therapeutic options exist,

65 and the available options are often limited by tolerability and efficacy (Falagas et al. 2014;

66 Tzouvelekis et al. 2012).

68 In Australia, carbapenemases have been rarely identified in Enterobacteriaceae, apart from the

69 weakly carbapenem-hydrolysing $bla_{\text{IMP-4}}$ which has established low level endemicity (Australian

70 Commission on Safety and Quality in Health Care 2013; Australian Group on Antimicrobial

71 Resistance 2014). Although small outbreaks and limited patient-to-patient transmission of CPE in

72 Australia have been described (Chang et al. 2015; Espedido et al. 2013; Kotsanas et al. 2013; Peleg

73 et al. 2005; Tai et al. 2015), the majority of patients identified with CPE are thought to have

74 acquired these organisms during international travel to endemic countries (Chua et al. 2014;

75 Fernando et al. 2010; Poirel et al. 2010; Sidjabat et al. 2011; Sidjabat et al. 2013). Limited

76 transmission in the community has also been reported (Blyth et al. 2014).

77

78 *Klebsiella pneumoniae* carbapenemases (KPCs) were first identified in the United States of

79 America in 1996, but have since spread worldwide to be endemic in several countries including

80 the USA, Israel, Italy, Greece, Brazil and China (Munoz-Price et al. 2013). Several gene variants

81 have been reported from the KPC family, though the most commonly reported alleles are KPC-2

82 and KPC-3, that differ by a single amino acid substitution (Tzouvelekis et al. 2012). The first

83 reported KPC-producing organism (KPC-2) in Australia was isolated in 2010 from a patient who

84 had returned to Sydney, New South Wales, after being hospitalised in Greece (Coatsworth et al.

85 2012). In 2012, KPC-2 was first identified in the state of Victoria in a patient who had been

86 repatriated to a metropolitan hospital also after a prolonged admission in a Greek hospital (Chua

87 et al. 2014).

88

89 Whole-genome sequencing (WGS) has emerged as a powerful tool for bacterial strain typing and

90 outbreak investigation (Kwong et al. 2015), and has been used in public health to assess

91 transmission of *Listeria monocytogenes* and other foodborne pathogens at jurisdictional, national

92 and international levels (Kwong et al. 2016). It has also been used at a single institution level to

93 investigate small-to-medium sized outbreaks of KPC-producing Enterobacteriaceae (Jiang et al.

94 2015; Kanamori et al. 2017; Marsh et al. 2015; Mathers et al. 2015; Ruppe et al. 2017; Snitkin et

95 al. 2012; Weterings et al. 2015). While the details of who infected whom can be traced among a

96 small group of individuals in conjunction with a detailed epidemiological investigation, the

97 transmission dynamics of larger established outbreaks across multiple institutions are more

98 difficult to resolve (Gilchrist et al. 2015).

99

100 Due to the increasing incidence of KPC-producing Enterobacteriaceae during 2012–2014 from

101 multiple sources across the state of Victoria (approximate population 6 million), we undertook an

102 outbreak investigation employing WGS in parallel with traditional outbreak epidemiology. At the

103 time this combined genomics and epidemiological investigation was initiated in June 2014, half

104 of the 30 cases reported were from two metropolitan hospitals that each reported prior internal

105   outbreaks of KPC-producing *K. pneumoniae* – one in 2012 and the other in early 2014. However,

106   no clear epidemiological links had been identified between the two hospitals. Additionally, the

107   remaining cases not identified through those two hospitals comprised patients in seven other

108   healthcare facilities and in the community. Although all but one of the isolates were ST258 by

109   multilocus sequence typing, if and where transmission of *KPC-producing K. pneumoniae* was

110   occurring was unknown. Here we report the utility of advanced genomic approaches including

111   sampling for within-host clonal diversity, Bayesian transmission modelling, and plasmid genome

112   reconstruction to assess the transmission dynamics of the outbreak to address these questions and

113   inform infection control management in real-time during the course of the outbreak.

114   **MATERIALS AND METHODS**

115   An overview of the workflow and methods used is shown in Figure 1.

116

117   **Laboratory and genomic methods**

118   ***Isolate selection and antimicrobial susceptibility assays.*** From 2012, following the first KPC

119   isolate in the state, all Victorian diagnostic microbiology laboratories were asked to refer suspected

120   carbapenemase-producing isolates to the Microbiological Diagnostic Unit Public Health

121   Laboratory (MDU PHL) for further testing. Suspected isolates were defined as Enterobacteriaceae

122   with a meropenem minimum inhibitory concentration (MIC) ≥0.5 mg/L or zone diameter (ZD)

123   ≤23 mm, or an imipenem MIC ≥2 mg/L or ZD ≤23 mm. Isolates with a positive colorimetric (e.g.

124   Rapidec® CarbaNP (Nordmann et al. 2012)) or molecular (e.g. Cepheid GeneXpert® (Tenover et

125   al. 2013)) test were also requested. Duplicate KPC-producing Enterobacteriaceae isolates of the

126   same species from a patient during a single hospital admission were usually excluded by referring

127   laboratories. However, for some patients, multiple isolates were referred to MDU PHL if they

128   originated from different laboratories, and were subsequently included in the analyses.

129

130   Species identification was confirmed using matrix-assisted laser desorption/ionisation time-of-

131   flight (MALDI-TOF) mass spectrometry (VITEK® MS, bioMérieux, France) with routine

132   susceptibilities tested on selected isolates using VITEK® 2 Compact (AST-N246 cards,

133   bioMérieux, France). All suspected CPE isolates meeting the above criteria underwent testing for

134   carbapenemase genes using polymerase chain reaction (PCR) (Supplementary Table S1). Isolates

135   confirmed as KPC-producing organisms underwent WGS and were included in the genomic

136   analyses, with the respective patients included in the epidemiological investigation. Isolates

137   obtained prior to 2014 were sequenced retrospectively, while WGS was performed prospectively

138   during the outbreak on isolates collected from 2014 onwards.

139

140   ***Multiple colony sampling.*** To understand the genomic diversity present within a single individual

141   and assess the influence on reconstructing transmission networks, primary faecal specimens were

142   obtained from one patient. After overnight culture on Brilliance™ CRE (Thermo Fisher, Waltham,

143   Massachusetts, US) selective media, 10-15 colonies, including any that differed in colony

144   morphology, were selected from the plates for DNA extraction and WGS. Multiple colony

145   sampling and sequencing for other patient samples was not performed due to cost limitations.

146

147   ***DNA extraction and whole genome sequencing***. Bacterial cultures were purified for DNA

148   extraction by two successive single colony selections after streaking onto horse blood agar

149   incubated overnight at 37°C. DNA was extracted from a liquid suspension of the purified cultures

150   using the QIAmp DNA Mini Kit (Qiagen, Hilden, Germany) or the JANUS Chemagic Workstation

151   with the Chemagic Viral DNA/RNA kit (CMG-1033; PerkinElmer, Waltham, Massachusetts, US).

152

153   WGS was performed on the Illumina MiSeq or NextSeq platforms using Nextera XT libraries and

154   protocols (Illumina, San Diego, California, US) with a minimum average quality score of 30 and

155   a target sequencing depth of ≥50x as previously described (Carter et al. 2016). Isolates not meeting

156   these metrics were resequenced. Raw sequence data has been uploaded to the Sequence Read

157   Archive under BioProject PRJNA397262.

158

159   Single molecule real-time (SMRT) sequencing of one *Klebsiella pneumoniae* and one *Citrobacter*

160   *farmeri* isolate was also performed on the PacBio RS II platform (Pacific Biosciences, Menlo Park,

161   California, US) using previously reported methods (Carter et al. 2016). In brief, genomic DNA

162   was extracted using a GenElute™ Bacterial Genomic DNA kit (Sigma-Aldrich, St. Louis,

163   Missouri, USA) and SMRTbell libraries were prepared using manufacturer protocols (Pacific

164   Biosciences, Menlo Park, California). Libraries underwent an additional size-selection step using

165   a 20-kb template BluePippin size-selection protocol (Sage Science, Beverly, Massachusetts,

166   USA).

167

168   ***Bioinformatic analyses***. Illumina raw sequencing reads were trimmed to clip Nextera adapters and

169   low-quality sequence (Phred score <10) using *Trimmomatic* v0.36 (Bolger et al. 2014). The

170   trimmed reads were assembled *de novo* with the *SPAdes* v3.7.1 assembler (Bankevich et al. 2012)

171   and auto-annotated with *Prokka* v1.12-beta (Seemann 2014). The SMRT analysis portal

172   v2.3.0.140936 (Pacific Biosciences) was used for isolates sequenced on the PacBio RS II, with

173   raw sequence reads assembled *de novo* using the HGAP3 protocol, and error correction and

174   polishing with *Quiver* v1. Further error correction was performed by mapping Illumina short reads

175   from the same isolate to the PacBio assembly with *Pilon* v1.21 (Walker et al. 2014) and *Snippy*

176   v3.2 (https://github.com/tseemann/snippy). The completed genome assemblies have been

177   uploaded to GenBank under BioProject PRJNA397262.

178

179   From the annotated assemblies, the multi-locus sequence type (MLST) was determined *in silico*

180   as was the presence or absence of antimicrobial resistance ("resistome") and other genes (pan-

181   genome)   using   BLAST-based   tools   (https://github.com/tseemann/mlst)

182 (https://github.com/tseemann/abricate) (https://github.com/sanger-pathogens/Roary) (Page et al.

183 2015). The pan-genome data was imported into *FriPan* (https://github.com/drpowell/FriPan) for

184 web-browser visualisation using a Python script (https://github.com/kwongj/roary2fripan), and

185 annotated genomes were visualised in *Geneious* v7.1.5 (http://www.geneious.com/).

186

187 Sequencing reads were also aligned to a reference genome to produce a reference-based whole-

188 genome alignment including single nucleotide polymorphism (SNP), invariant and missing sites

189 (https://github.com/tseemann/snippy). This alignment was then trimmed to exclude plasmid sites

190 (https://github.com/kwongj/trim-aln). Putative regions of recombination were predicted using

191 *ClonalFrameML* v1.0 (Didelot & Wilson 2015), and masked in the alignment

192 (https://github.com/kwongj/cfml-maskrc). Core genome SNP sites were extracted from the

193 recombination-filtered alignment (Page et al. 2016) and a maximum likelihood phylogenetic tree

194 inferred from the resulting SNP alignment in *RAxML* v8.2.4 (Stamatakis 2014), using a General

195 Time Reversible model of nucleotide substitution with a Γ model of rate heterogeneity and four

196 rate categories, with 1000 bootstrap replicates to determine branch support. Hierarchical Bayesian

197 Analysis of Population Structure (*hierBAPS*) (Cheng et al. 2013) was used to provide further

198 support for identifying phylogenetic clades, with clustering performed using 8 levels in the

199 hierarchy (L) and the prior maximum number of clusters (maxK) set at 10. Publicly available

200 sequencing data was retrieved from the National Center for Biotechnology Information (NCBI)

201 Sequence Read Archive and GenBank for comparison with local genomes (Supplementary Tables

202 S2 and S3). Three reference genomes were used. The initial genomic analysis was performed using

203 KPNIH24 (GenBank accession CP008797), with PacBio sequences of internal isolates (*K.*

204 *pneumoniae* AUSMDU00008079, GenBank accession SAMN07452764; and *C. farmeri*

205   AUSMDU00008141, GenBank accession SAMN07452765) used for subsequent analyses to

206   minimise inaccuracies and biases from using distant genomes and draft genome assemblies as

207   references (Supplementary Figure S1).

208

209   ***Genomic context of the bla*$_{\text{KPC}}$ *gene.*** The genomic context of the *bla*$_{\text{KPC}}$ gene was investigated

210   using a custom Python script (https://github.com/kwongj/contig-puller) to extract and align *de*

211   *novo* assembled contigs carrying the gene. The flanking *Tn*4401 transposon regions were

212   compared to previously described isoforms (Chen et al. 2012; Kitchel et al. 2009; Naas et al. 2012;

213   Naas et al. 2008), using *BLAST+* (Camacho et al. 2009). Plasmid replicons were identified in

214   genome assemblies using *BLAST+* against the *PlasmidFinder* database (Carattoli et al. 2014).

215   Plasmids were presumed to carry KPC if *bla*$_{\text{KPC}}$ was identified on the same contiguous sequence

216   of DNA as the plasmid replicon.

217

218   ***Bayesian evolutionary analyses and transmission modelling.*** Bayesian evolutionary analyses

219   were also conducted to ascertain if the genomic signal could estimate the date of emergence of any

220   outbreak clusters, and if any additional data could help inform where and when transmission events

221   were taking place. A recombination-filtered chromosomal alignment was obtained as described

222   above. The subsequent alignment was used as input into *BEAST2* v2.4.3 (Bouckaert et al. 2014)

223   with collection dates entered for each isolate. The relationship between root-to-tip distance and

224   date of isolation was assessed in *TempEst* v1.5 (Rambaut et al. 2016). We fitted a model with a

225   relaxed log-normal clock (Drummond et al. 2006) to the alignment to account for inter-clade

226   variation, assuming a HKY model for nucleotide substitution with Γ distributed among site rate

227   variation (Hasegawa et al. 1985), and used a constant population size coalescent prior on the

228   genealogy. Nodes were selected for logging the likely time of the most recent common ancestor

229   (MRCA) for major clades identified in the maximum likelihood tree. We used eight Markov chain

230   Monte Carlo (MCMC) runs of 100 million states, with sampling every 5000 states, and a burn-in

231   of 50 percent. The posterior samples from each chain were checked in *Tracer* v1.6

232   (http://tree.bio.ed.ac.uk/software/tracer/) for convergence, and then grouped into a single chain.

233   The posterior samples for the dates of the nodes of interest were annotated on a maximum clade

234   credibility tree and exported.

235

236   The maximum clade credibility tree was used as input into *TransPhylo* v1.0 (Didelot et al. 2017)

237   to reconstruct the transmission chain. We based our prior for generation time (the time from

238   primary infection of an individual to any subsequent secondary infection) on a previous study on

239   the duration of carriage of KPC-producing Enterobacteriaceae (Zimmerman et al. 2013), assuming

240   detection of KPC as an indicator of likelihood for onwards transmission. We therefore used a

241   Gamma distribution with shape parameter 1.2 and scale 1.0 for the generation time prior, with a

242   distribution mean of 1.2 years (438 days), standard deviation 1.096 years (400 days), and mode

243   0.2 years (73 days). The MCMC was run for 100000 iterations, with transmissions inferred from

244   a consensus transmission tree.

245

246 ***Outbreak genomic investigation.*** An initial retrospective analysis using the bioinformatic methods

247 described above was performed in June 2014 on 41 KPC-producing *K. pneumoniae* isolates that

248 had undergone WGS, including *de novo* genome assembly, *in silico* multi-locus sequence typing

249 (MLST), resistance gene detection and phylogenetic analysis. Subsequently, a customised, in-

250 house pipeline was developed to streamline and automate handling of sequence data for ongoing

251 assessment and analysis (https://github.com/tseemann/nullarbor). A summary report was

252 generated for each set of isolates run through Nullarbor including quality control metrics, MLST,

253 resistome and pan-genome comparison, and a maximum likelihood phylogeny inferred from the

254 core SNP alignment (Price et al. 2010).

255

256 From this initial report, clonal complex (CC) 258 isolates underwent further analysis using a

257 PacBio-assembled reference genome of a local isolate for a higher resolution core genome SNP

258 alignment and phylogenetic comparison. As additional KPC-positive isolates were identified over

259 time, the *Nullarbor* pipeline was repeated in an iterative process to establish the genomic

260 relationship of new isolates to existing isolates, and where new CC258 isolates were identified,

261 the CC258-specific analysis was also repeated with the new isolates. These analyses were

262 interpreted together with epidemiological data in real-time as new cases emerged (see below –

263 "Outbreak reporting and oversight"). Additional detailed analyses, including recombination

264 filtering and Bayesian temporal analyses, were performed *ad hoc* to gain an overall understanding

265 of the outbreak, and although were not used to directly inform infection control of individual cases,

266 were used to inform overall outbreak management and guideline development.

267

268   **Epidemiological methods**

269   Following the results of the initial genomic analysis and concern regarding local transmission, the

270   ongoing genomic investigation was accompanied by collection of detailed epidemiological data,

271   retrospectively for cases detected prior to 2014 and in real-time from 2014.

272

273   *Data Collection.* All patients from whom KPC-producing Enterobacteriaceae were identified from

274   the first isolate in 2012 to 31 December 2015 were included in the epidemiological investigation.

275   A case report form was developed to collect detailed epidemiological data, especially regarding

276   patient hospitalisation details. Patients, or their next of kin, were interviewed by phone to ascertain

277   demographic information on age, sex, country of birth, and risk factors for CPE such as

278   hospitalisation and medical procedures, overseas travel, and comorbidities. Where a patient was

279   in a healthcare facility at the time of specimen collection, or reported hospitalisation in the 12

280   months prior to initial CPE identification, medical records from each hospitalisation were

281   examined to obtain data on specimen collection, clinical details of infection or colonisation,

282   procedures and patient movements both before and after CPE identification, where available.

283   Treating doctors, general practitioners and infection control personnel were contacted to obtain

284   additional and/or missing information. Patient hospital admission data were collated and used to

285   identify putative transmission networks of patients linked by proximity in time and space.

286

287   Once identified with KPC, patients were assumed to be colonised indefinitely, thus subsequent

288   KPC isolates were deemed to constitute the same patient episode and epidemiological data were

289   not re-collected. KPC-producing Enterobacteriaceae isolated from normally sterile sample sites

290   *e.g.* blood, cerebrospinal fluid, pleural fluid, were considered to represent "infections". Isolates

291   obtained from non-sterile sites (*e.g.* urine, wound swab, aspirate from intra-abdominal collection)

292   where clinical evidence of infection was present and the patient was treated with antibiotics with

293   activity against KPC-producing organisms, or where the treating clinician identified infection but

294   elected to palliate the patient, were also considered to represent "infections". Other isolates were

295   considered to represent "colonisations", unless no epidemiological data were available

296   ("unknown").

297

Infection control investigation

299   For patients where local transmission of KPC-producing Enterobacteriaceae was suspected,

300   interviews were conducted with infection control practitioners (ICPs) at the facility with putative

301   transmission. Data were collected on environmental and contact screening activities, patient follow

302   up and patient management alerts, isolation and collection of subsequent screening specimens.

303

Outbreak reporting and oversight

305   Data from both genomic and epidemiological investigations were analysed together in real-time,

306   and reported back to the Victorian Department of Health and Human Services. Where transmission

307   events were recognised, the healthcare institutions involved were also informed. An incident

308   management team was established through the Department of Health and Human Services to

309   oversee the investigation and management of the outbreak.

310

Ethics approval and consent to participate

312   Data were collected as part of an outbreak investigation through the Victorian Department of

313   Health and Human Services under the Public Health and Wellbeing Act 2008

314 (https://www2.health.vic.gov.au/about/legislation/public-health-and-wellbeing-act). No consent

315 was required for the project.

316

317 **Availability of data and materials**

318 The raw Illumina short-read sequencing data and completed PacBio genome assemblies supporting

319 the conclusions of this article are available in the National Center for Biotechnology Information

320 database under BioProject PRJNA397262. In-house scripts used in the bioinformatics analyses are

321 available at https://github.com/tseemann and https://github.com/kwongj.

322 **RESULTS**

323 A total of 86 KPC-producing Enterobacteriaceae were referred to MDU PHL between 2012-2015

324 from 69 patients in 26 different healthcare facilities, six general practices and three aged care

325 facilities across both metropolitan and rural Victoria (Figure 2). Specimen source was reported for

326 85 of the isolates, with most specimens being urine (n = 42; 49%), faeces or rectal swabs (n = 20;

327 23%), or blood cultures (n = 10; 12%) (Supplementary Table S4). Almost all isolates were KPC-

328 2 producing Enterobacteriaceae, apart from two KPC-3 producing *K. pneumoniae* isolates from

329 one patient. Overall, 79 of the isolates were *K. pneumoniae*, five were *Citrobacter farmeri*, with

330 single isolates of *Klebsiella oxytoca* and *Citrobacter freundii*.

331

332 **Initial genomic analysis**

333 In the initial retrospective genomic analysis of isolates collected prior to June 2014, 40/41 (98%)

334 *K. pneumoniae* isolates collected from 29 patients were clonal complex (CC) 258 by *in silico*

335 MLST, with all of these belonging to clade 1 of sequence type (ST) 258 *K. pneumoniae* described

336 by Deleo *et al*. (Deleo et al. 2014), based on analysis of the capsular polysaccharide gene island.

337 Phylogenetically, these isolates clustered together in comparison to other previously reported

338 international ST258 *K. pneumoniae* isolates (Supplementary Figures S2 & S3). Analysis of the

339 local ST258 isolates revealed three distinct phylogenetic clades involving patients in 20 different

340 healthcare locations, supported by pairwise SNP distributions and Bayesian Analysis of the

341 Population Structure (Figure 3 & Supplementary Figure S4). These results formed the basis for

342 ongoing prospective genomic analyses.

343

344 **Epidemiological analysis**

345 A total of 57 patients were found to harbour isolates belonging to CC258 from 2012-2015. Most

346 patients presented to the Emergency Department (n=7) or were admitted to hospital (n=44) at the

347 time of initial specimen collection. Thirty-three (58%) were male, and the median age of the

348 affected patients was 74 (range 20 – 94; IQR 62-83) (Supplementary Table S5). Clinical infection

349 was suspected in 61% of the cases, with the most common presentation being urinary tract

350 infection (n=24; 42%). Sepsis was reported in 63% of those with infection due to KPC-producing

351 *K. pneumoniae*. KPC-producing isolates were obtained from patients with clinical infection more

352 commonly during 2012-2014 than 2015, where >50% of the KPC-producing isolates identified

353 were colonising or screening isolates (Supplementary Figure S5).

354

355 Of patients from whom a complete travel history was able to be obtained (n=46), 22 (48%) reported

356 overseas travel since 1996, and only eight (14%) reported travel in the 12 months prior to

357 hospitalisation, strongly reinforcing the suspicion of local transmission of KPC. All 57 patients

358 were found to have been hospitalised in Australian healthcare facilities in the 12 months prior to

359 initial positive specimen collection, with a median length of stay of 61 inpatient days (range 0-181

360 days; IQR 34-101). Four patients had fewer than ten inpatient days in an Australian healthcare

361 facility during this period, three of whom had recent overseas hospitalisation, for which complete

362 hospitalisation data could not be collected. Additional epidemiological data is detailed in the

363 Supplementary Appendix.

364

365 A detailed analysis of patient admission data was undertaken in attempt to identify location or

366 source of acquisition. Forty-one patients (72%) had previously attended one facility (Facility F) in

367   the 12 months prior to initial identification of a KPC-producing isolate. However, the remaining

368   16 patients reported no hospitalisation in this facility, and 18 additional facilities were identified

369   in which two or more patients had been admitted. Three putative transmission networks were

370   identified based on overlapping admissions to the same hospital ward at the same time among

371   patients detected with KPC. It became apparent that due to the sheer number of hospitalisations

372   and the complexity of patient movements, drawing firm conclusions on who transmitted to whom

373   and where transmission had occurred solely from epidemiological data would be difficult (Figure

374   4 & Supplementary Figure S6).

375

376   **Combined analysis of CC258 KPC-2-producing *K. pneumoniae* isolates**

377   Despite identification of three distinct phylogenetic clades in the initial genomic analysis (Figure

378   3), we were unable to exactly define where transmission was occurring as the samples were

379   distributed across multiple locations. Of the three clades (designated as genomic cluster A, B and

380   C), cluster C had the best correlation with an epidemiologically-defined transmission network,

381   comprising isolates from patients in the same healthcare facility at the same time (Figure 4).

382   However, other correlations were not immediately apparent. Cluster A included isolates from three

383   patients in three different facilities that were within the same healthcare network and/or were

384   geographically located in the same region within a 20 km radius, though were 30 km away from

385   the fourth patient in a different healthcare network in the cluster. Cluster B included isolates from

386   patients in 12 different locations across the state of Victoria, separated by up to 500 km. Seven

387   isolates were obtained from patients attending local general practice (primary care) clinics.

388

389 Using the genomic data to enhance and allow flexibility in the epidemiological data (*e.g.* including

390 admissions to adjacent wards or admissions separated by a few days as "overlapping"), it became

391 apparent that each of the genomic clusters corresponded to a separate transmission network (Figure

392 5). The largest of these (cluster B) was located at a single institution (Facility F), and as further

393 isolates were identified, two subclusters (B1, B2) were defined. Cluster B2 included isolates

394 arising from suspected transmission in mid-2014 in an inpatient aged care ward, while cluster

395 included isolates from a number of patients who were admitted to Facility F in 2012, though to

396 several different wards. A second transmission network in cluster B1 was identified in another

397 aged care ward in 2015. The predicted transmission network corresponding with cluster C was

398 also confirmed, with the isolate genomes differing by <5 SNPs, and the corresponding

399 epidemiological data for the respective patients showing overlapping hospital admissions to a

400 single hospital ward in late 2013 (Figure 6).

401

402 Combining epidemiological and genomic data also revealed secondary transmission events, where

403 patients with undetected colonisation at one hospital were transferred to a second healthcare

404 service, with subsequent diagnosis and onwards transmission. Patients whose isolates were

405 phylogenetically clustered with other isolates from a known transmission network, but who had

406 not previously been admitted to the hospital where transmission was occurring, were flagged as

407 potentially having acquired KPC through a secondary transmission event. Through this, two

408 putative secondary transmission events were later identified from these main transmission

409 networks (see example case study in Figure 6).

410

Bayesian evolutionary and transmission analysis

Although from the combined genomic and epidemiological analyses, it appeared that local transmission within separate hospitals was driving the outbreak, local isolates from each cluster were closer to other local isolates (median pairwise SNP distance = 16; IQR 9-38) than to international ST258 isolates from GenBank (median pairwise SNP distance = 243; IQR 80-258). To investigate the possibility of initial inter-hospital transmission leading to subsequent transmission networks within each hospital, we explored whether Bayesian evolutionary and transmission modelling could be used to provide additional insight.

From these analyses, genomic clusters A and C each appeared to be derived from separate overseas importations of KPC-producing *K. pneumoniae* with subsequent spread within a single hospital or healthcare network rather than inter-hospital transmission, with a most recent common ancestor (MRCA) predicted to have occurred prior to 2010 (median 2007.78; 95% HPD 2003.30–2011.09) when the first KPC-producing organism was reported in Australia (Coatsworth et al. 2012), and 2012, when the first KPC-producing organism was reported in Victoria (Chua et al. 2014) (Figure 7). Similarly, the unclustered isolates represented separate individual importations in these analyses, supporting the known history of recent overseas travel for these patients (Supplementary Table S5).

Cluster B likely arose from another overseas importation (MRCA cluster A and B: median 2002.52; 95% HPD 1991.72–2008.79), though the timed phylogeny (Figure 7) suggested strains of KPC-producing *K. pneumoniae* from cluster B may have been circulating in Victoria since 2011 (MRCA cluster B1 and B2: median 2010.59; 95% HPD 2008.67–2011.86). A number of internal

434   nodes within cluster B1 were dated to 2012 (Supplementary Figure S8), corresponding with

435   overlapping patient admissions at Facility F at that time, which affirmed the putative transmission

436   network in Facility F in 2012. In both clusters B1 and B2, many isolates were linked by

437   transmission events that were predicted to have occurred during 2013, despite a decline in new

438   cases detected during this period. Notably, many of the isolates derived from these transmission

439   events were not detected until later, in 2014 or 2015. This supported the epidemiological

440   hypothesis that detection of previously unrecognised colonisation was driving a large proportion

441   of the new cases in 2014 and 2015, rather than new transmission events (Figure 8). For each of the

442   suspected transmission networks identified in the combined outbreak investigation, posterior

443   probability distributions for the MRCA were generally consistent with the epidemiological data

444   (Supplementary Figure S8).

445

446   From the *TransPhylo* analysis, the mean reproductive number (R_0) across the outbreak was

447   calculated to be 1.63 (SD = 0.13), accounting for an average estimated sampling proportion of 0.27

448   (SD = 0.05).

449

450   **Plasmid analysis**

451   In late 2015, other Enterobacteriaceae carrying bla_{KPC-2} were detected, including *Citrobacter*

452   *farmeri*, *Citrobacter freundii*, and *Klebsiella oxytoca* (Figure 2), which raised concern about KPC-

453   plasmid spread. Analysis of the *de novo* assembled draft genomes indicated the bla_{KPC-2} carrying

454   plasmids in these organisms matched a bla_{KPC-2} carrying plasmid found in the locally circulating

455   *K. pneumoniae* isolates, suggesting inter-species plasmid movement, and subsequent organism

456   transmission. All *K. pneumoniae* ST258 isolates were found to harbour bla_{KPC-2} within the "a"

457   isoform of the characteristic Tn*4401* transposon, and usually on a contig containing sequence

458   encoding the replication regulator RNAI from IncFII(K) plasmids. Analysis of the PacBio

459   reference *K. pneumoniae* genome assembly showed the presence of both IncFIB(pQil) and

460   IncFIB(K) plasmids. The ST258 draft genome assemblies also all contained IncFIB replicons,

461   though the assemblies were inadequate to confirm the plasmid types carrying Tn*4401*.

462   Chromosomal integration of the Tn*4401* within the putative diguanylate cyclase gene, *ycdT*, was

463   found in two epidemiologically linked ST258 *K. pneumoniae* isolates. One ST258 *K. pneumoniae*

464   isolate lacked bla_{KPC-2}, and subsequent repeat susceptibility testing confirmed the isolate to be

465   susceptible to carbapenems. Re-testing of 10 colonies subcultured from the stored glycerol stock

466   for that patient sample found only 7/10 displayed carbapenemase hydrolytic acitivity with the

467   CarbaNP test, indicating *in vivo* or *in vitro* loss of the bla_{KPC} gene.

468

469   Despite presence of the Tn*4401* and IncFII(K) RNAI in the *C. farmeri* draft assemblies, they

470   lacked the IncFIB replicons detected in the *K. pneumoniae* genomes. To determine if interspecies

471   plasmid transmission had occurred, one *C. farmeri* isolate underwent PacBio sequencing.

472   Sequence comparisons between the KPC-plasmid genome from this isolate and the KPC-plasmid

473   genome from the *K. pneumoniae* PacBio reference genome indicated almost identical plasmids.

474   The *C. farmeri* plasmid carried an IncR-type replicon and lacked a Type I restriction modification

475   system, but otherwise comprised an identical sequence to 85% of the *K. pneumoniae* KPC plasmid

476   following a recombination and inversion event (Figure 9). The *C. freundii* isolate was also found

477   to have an IncR-type plasmid, but lacked other gene content found in the *C. farmeri* IncR plasmids,

478   suggesting mobilisation of the Tn*4401* transposon. The exact location of bla_{KPC-2} and plasmid type

479   of the *C. freundii* and *K. oxytoca* isolates were unable to be determined due to the limitations of

low

480 the assembled short-read sequencing data. Phylogenetic analysis of the *C. farmeri* isolates inferred

481 from core genome (chromosomal) SNPs confirmed the isolates were closely related

482 (Supplementary Figure S9).

483

484 **Within-host diversity**

485 To assess the within-host genomic diversity of KPC-producing Enterobacteriaceae, the isolate

486 genomes from 17 patients who had multiple isolates obtained and sequenced were compared. Of

487 these, one patient had 32 isolates collected from seven samples over an 8-month period from

488 February 2016 to October 2016, including multiple colony sampling from two faecal samples. The

489 other 16 patients had multiple (median = 2; range 2-4) isolates referred to MDU PHL, obtained

490 through recurrent presentations at different healthcare facilities.

491

492 Genomic comparisons indicated isolates from the same patient differed by up to 21 SNPs in the

493 core genome. This included the 32 isolates obtained from patient 70, with several internal lineages

494 emerging from the same common ancestor within that individual (Figure 10). Pangenome analyses

495 of these isolates also demonstrated changes in the accessory genome correlating with the emerging

496 internal lineages in the core genome (chromosomal) phylogeny, and lineage-specific plasmid

497 carriage (Supplementary Figure S10). These changes included subtle differences in plasmid types,

498 acquisition/loss/switching of transposons, and loss/gain of heavy metal operons. There was a poor

499 molecular clock signal for these isolates (root-to-tip divergence R^2 = 0.026), and no further

500 temporal analyses were undertaken.

501

low

502   When compared to other outbreak isolates, multiple isolates from the same host clustered together

503   within a single subclade in the phylogenetic tree. This supported the concept of a "transmission

504   bottleneck", where despite an assumed diverse within-host population of KPC-producing

505   Enterobacteriaceae, transmission only involved a single isolate, with subsequent within-host

506   evolution following the initial transmission event. This also suggested multiple acquisitions of

507   different strains of KPC-producing *K. pneumoniae* by the same patient were uncommon among

508   the patient cohort. Of patients with multiple isolates spanning a six-month period, only patients 18

509   and 70 appeared to have isolates from differing subclades, though bootstrap branch support for

510   these subclades was marginal (0.70-0.75).

511

512   **Environmental diversity**

513   A limited number of KPC-producing environmental *K. pneumoniae* isolates obtained at a single

514   time-point from frequently touched surfaces in the room of patient 75 underwent WGS and were

515   compared to the other patient isolates (Figure 10). As expected, the environmental isolates

516   clustered with the clinical isolate from patient 75, as well as those from patient 72, who had

517   previously shared a (different) room with patient 75. Notably, the genomic diversity between these

518   clinical and environmental isolates was less than the within-host diversity seen within patient 70,

519   supporting the hypotheses that the direction of transmission was from patient 75 to the

520   environment, rather than environment to patient, and that transmission had occurred between

521   patients 72 and 75 when they previously shared a room.

522

Infection control investigation

523

524 Qualitative interviews were conducted with infection control practitioners at three facilities

525 regarding five transmission networks (four primary clusters of KPC-producing ST258 *K.*

526 *pneumoniae* including two secondary transmission events at different locations, in addition to one

527 cluster of KPC-producing *C. farmeri*) of suspected local transmission involving 17 patients. The

528 majority of transmission events were thought to have occurred due to unrecognised colonisation,

529 and prior to the implementation of contact precautions – single room isolation with *en-suite*

530 bathroom, with requirements for use of personal protective equipment (gloves and gown) by staff

531 and visitors entering the room. However, two instances of apparent transmission to and from

532 patients under contact precautions were reported. In each of these situations, after being admitted

533 directly into contact precautions due to known colonisation with other multidrug-resistant

534 organisms, a patient acquired a KPC-producing *K. pneumoniae* isolate that was highly related to

535 other isolates in a genomic cluster (<5 SNPs in the core genome).

536

537 Due to the lack of local guidelines at the time of the outbreak, the management of wards where

538 transmission was found to have occurred was at the discretion of the individual healthcare facilities

539 involved. A retrospective review of the methods used to interrupt transmission identified a number

540 of bundled approaches, with the common measures being contact screening, enhanced cleaning

541 (in contrast to standard cleaning), and isolation of identified cases under contact precautions

542 (Supplementary Table S6).

543

544 **Impact of whole-genome sequencing on the outbreak investigation and development of local**

545 **containment guidelines**

546   In response to the genomic and epidemiological evidence of an evolving outbreak with local

547   transmission, a state-wide management guideline was developed for the surveillance and

548   containment of carbapenemase-producing Enterobacteriaceae incorporating the routine use of

549   WGS to determine relatedness between isolates (Department of Health & Human Services 2015).

550

551   Several insights were gleaned through the use of genomics that were fundamental to the outbreak

552   investigation and informed development of the guidelines and their focus on hospitals. For

553   example, through the epidemiological investigation, the identification of KPC in previously

554   hospitalised patients suggested nosocomial rather than community transmission, though which

555   hospital transmission was occurring within was not immediately clear, given the number of

556   hospitals and prior overlapping hospitalisations involved. The resolution of the genomic data

557   refined epidemiological hypotheses to pinpoint transmission areas in individual hospitals to focus

558   infection control efforts, with Bayesian analyses of the genomic data also supporting the

559   epidemiological findings. By determining when and where transmission was occurring, it was

560   evident that many new cases were readmissions of patients with previously unrecognised

561   colonisation, rather than new transmission events, making it difficult to "control" the outbreak.

562   This highlighted the need to intensively screen at-risk patient contacts, including flagging those

563   that had been discharged from an affected ward prior to recognition of transmission for screening

564   and pre-emptive isolation upon subsequent hospital presentations. This real-time change-in-

565   practice was reflected by the increasing proportion of cases identified as "colonisations" rather

566   than "infections" due to intensified screening practices over time (Supplementary Figure S5).

567 **DISCUSSION**

568 The emergence of carbapenemase producing Enterobacteriaceae (CPE) is a major threat to human

569 health (Australian Commission on Safety and Quality in Health Care 2013; Centers for Disease

570 Control and Prevention 2013; World Health Organization 2014), with significant interventions

571 required at state and national levels to contain the spread once established (Schwaber & Carmeli

572 2014). Here, we describe the largest outbreak of CPE reported in Australia, and demonstrate how

573 a combined genomics and epidemiological investigation delineated the outbreak into five separate

574 nosocomial transmission networks (four clusters of *K. pneumoniae* and one cluster of *C. farmeri*)

575 across the state over four years, resulting in targeted interventions for each transmission area. The

576 use of WGS for outbreak analysis has been well established, including for transmission of

577 multidrug resistant hospital pathogens (Palmore & Henderson 2013; Snitkin et al. 2012), but these

578 studies have been predominantly small scale, single institution studies, and often retrospective.

579 Genomics has rarely been used prospectively and in real-time during a complex multi-institutional

580 outbreak requiring a coordinated state-wide public health response. Due to prolonged colonisation,

581 many CPE cases linked to hospitals with suspected transmission were identified through other

582 healthcare facilities, and some through general practice. With the number of potential transmission

583 opportunities in retrospective hospitalisation data from all facilities, the epidemiological

584 investigation would have been difficult to interpret and translate into focussed interventions

585 without the resolution offered by genomics. Similarly, accurate interpretation of the genomic data

586 would have been difficult without the supporting epidemiologic data. As others have pointed out,

587 the two must go hand-in-hand (Tong 2013). By integrating our detailed epidemiological

588 investigation with genomic analyses, we were able to refine our hypotheses, and coordinate an

589   effective public health response to target areas with ongoing transmission, emphasising the ability

590   of WGS to enhance surveillance systems and outbreak investigations.

591

592   Despite integrating genomic data with detailed epidemiological information, we still encountered

593   several challenges. Accurate inference of plasmid transmission can be challenging from short-read

594   Illumina sequencing data, with repetitive elements such as insertion sequences that accompany

595   mobile resistance elements frequently confounding short-read assemblers and read-aligners

596   (Conlan et al. 2014). As we demonstrated here and as reported by others (Conlan et al. 2014), long-

597   read sequencing technology such as Pacific Biosciences single-molecule sequencing can be highly

598   useful in resolving and tracking the diversity of plasmids carrying carbapenemases and other

599   resistance genes, although we could only sequence a limited representation of the outbreak isolates

600   due to cost limitations. The data gained from these completed genomes were invaluable. The use

601   of a local internal reference provided additional confidence in SNP calling among highly clonal

602   isolates. We were able to compare completely assembled plasmids between different species,

603   which otherwise would not have been possible. Many isolates in the outbreak were hypothesised

604   to carry bla_{KPC-2} on either an IncFIB(K) or IncFIB(pQil) plasmid, with long-read sequencing

605   showing it was possible for isolates to harbour multiple plasmids of the same incompatibility

606   group. As we found in two isolates and as others have recently reported (Mathers et al. 2017),

607   chromosomal integration of Tn*4401* also occasionally occurs. The movement of these resistance

608   elements adds complexity to understanding transmission dynamics using genomics, and highlights

609   the limitations of short-read sequencing data to fully characterise these elements.

610

611   It is increasingly evident that the subtly diverse populations of a single clonal type due to within-

612   host evolution (Golubchik et al. 2013), and more diverse populations from transmission of mixed

613   infections (Eyre et al. 2013; Hatherell et al. 2016), can impair accurate reconstruction of

614   transmission pathways from genomic data (Worby et al. 2014). In our study, we found some

615   individuals not only had different plasmid variants, but also had distinct within-host evolutionary

616   lineages, indicating a complex evolutionary history of transmission, within-host evolution, and

617   plasmid movement, mirroring recent reports in other patients with prolonged KPC colonisation

618   (Conlan et al. 2016). However, measuring this diversity is difficult in outbreak investigations.

619   Although single-colony sampling and sequencing arguably provides the most informative data, it

620   may be subject to colony selection bias, so some uncertainty remains. This can be offset by

621   sampling more colonies, but the time and cost of sequencing each colony can quickly become

622   prohibitive in an outbreak investigation and real-world public health environment. Although some

623   have attempted to detect the presence of mixed infections in short-read datasets through analysis

624   of short-read mapping from deep sequencing (Eyre et al. 2013), this relies on WGS performed on

625   a sweep of a primary culture plate, introducing the potential for exogenous DNA in the dataset,

626   and becomes considerably more complicated with fluxes in accessory genome content. As we

627   found, bla_{KPC}-plasmids can be lost even through laboratory passage of stored isolates. Even if the

628   within-host genomic diversity can be captured, incorporating these data into models can be

629   difficult. Recent attempts to account for elements such as within-host diversity and unsampled data

630   in reconstructing transmission trees have proved successful in some situations (Didelot et al. 2017;

631   Jombart et al. 2014; Klinkenberg et al. 2017; Ypma et al. 2013). However, each of these make

632   several assumptions, such as the presence of a complete transmission bottleneck that does not

633   allow for repeated acquisition, a constant reproduction number across the outbreak, few or no

634   unsampled cases, and that genetic variation is accumulated in pathogens in a clock-like fashion.

635   The within-host diversity and horizontal transmission of mobile genetic elements in outbreak

636   investigations complicates analyses further, and consequently, accurate reconstruction of exact

637   transmission routes remains difficult for large and moderate-sized CPE outbreaks.

638

639   Our study has some limitations and points of note. Firstly, although most patient isolates had a

640   clear epidemiological link to other isolates within the defined genomic clusters, not every isolate

641   was able to be linked to another patient with KPC or a history of recent overseas travel, indicating

642   a larger pool of unrecognised colonisation, though influenced in part by differing stringency in

643   contact screening and data collection. For example, no patient in cluster A reported overseas travel

644   since 1996, when KPC was first detected. Furthermore, although these patients were located at

645   healthcare facilities in the same geographical region of Victoria and part of the same healthcare

646   service, they did not have any overlapping inpatient admissions. Thus, it is likely that there are

647   other individuals or environmental sources that serve as intermediary reservoirs of isolates not

648   captured in our sampling, that facilitate ongoing transmission. Although environmental and contact

649   screening was initially performed using methods published by others (Centers for Disease Control

650   and Prevention 2015), based on our data, the optimal approaches are still not well understood.

651   Secondly, although the outbreak has slowed, cases are still emerging despite the extensive

652   investigation and interventions implemented by the individual healthcare services. However, as

653   shown in our analyses, many of these may be unrecognised colonisation following previous

654   exposure rather than new transmission events. As a result of this knowledge, through the

655   development of our guidelines, we have redirected infection control resources into intensive

656   screening to identify individuals with asymptomatic colonisation, focussing efforts on healthcare

657 facilities with ongoing transmission. Thirdly, as discussed above, capturing within-host population

658 genomic diversity and reconstructing entire plasmid genomes from each clinical sample through

659 sequencing would further enhance future outbreak investigations, but are currently limited by the

660 cost of performing such analyses routinely. Sequencing cost and accuracy limitations also prohibit

661 the routine use of more rapid sequencing technologies, such as PacBio SMRT and Oxford

662 Nanopore sequencing, which could potentially provide more rapid results than Illumina

663 sequencing to inform infection control decisions for prospectively sequenced isolates, though in

664 the context of an outbreak spanning several years, turn-around times for Illumina sequencing can

665 still be considered "real-time". Finally, although we conducted transmission inference analyses,

666 these methods require further validation for routine implementation in outbreak investigations of

667 multidrug-resistant pathogens. For example, we used *TransPhylo* in preference to other

668 transmission modelling software as it did not require a high sampling proportion, and it provided

669 the ability to account for and estimate the proportion of missing cases. However, it is difficult to

670 know the true accuracy of these estimates, as even transmission chains arising from densely

671 sampled wards were predicted to have unsampled cases by *TransPhylo*. Several possibilities to

672 explain these unsampled cases include the limited sensitivity of the screening methods used, and

673 the role of staff, visitors and the hospital environment in reconstructing transmission networks.

674 Additionally, although transmission chains were predicted by *TransPhylo*, the resolution of the

675 epidemiological data was insufficient to confirm the precise sequence of transmission events in a

676 complex outbreak, with multiple potential pathways. Robust validation of these transmission

677 models is required before being applied in outbreak investigations to determine who infected

678 whom – results that may subsequently have considerable ethical or legal implications.

679

680 **CONCLUSIONS**

681 Our study reports a real-world prospective utilisation of WGS, including the difficulties

682 encountered, to enhance a complex epidemiological investigation in real-time for an important

683 pathogen. In contrast to most reported outbreaks of CPE within a single institution, our outbreak

684 demanded a coordinated public health response. Previously the domain and responsibility of

685 individual hospitals and healthcare facilities, the spread of CPE has become a public health issue.

686 Through our experience, WGS has since been incorporated into our state guidelines for the

687 management of CPE (Department of Health & Human Services 2015). Developed during the

688 outbreak investigation, this system entails routine WGS of all suspected CPE isolates with

689 concurrent epidemiological investigation to allow prospective, centralised comparison of isolates

690 and epidemiological data from multiple health services. In turn, this enables identification of

691 potential transmission events between patients geographically and temporally dispersed at

692 identification, translating into focussed interventions at the designated transmission locations. For

693 such extensively drug-resistant organisms with limited treatment options, all feasible interventions

694 towards reducing the early burden of disease are warranted.

695 **ACKNOWLEDGEMENTS**

696 We wish to acknowledge and thank the referring clinical microbiology laboratories for facilitating

697 transfer of isolates to MDU PHL, and the hospital infection control and infectious diseases staff

698 who assisted and facilitated epidemiological data collection from the corresponding patients. We

699 also thank Dr Xavier Didelot, Imperial College London for his advice and assistance with

700 *TransPhylo*, and Dr Sally Partridge, University of Sydney for her expertise in Gram-negative

701 plasmids.

702

703

704   **REFERENCES**

705

706   Australian Commission on Safety and Quality in Health Care. 2013. Recommendations for the control of

707   multi-drug resistant Gram-negatives: carbapenem resistant Enterobacteriaceae. Sydney: Australian

708   Commission   on   Safety   and   Quality   in   Health   Care.   Available   from:

709   https://www.safetyandquality.gov.au/publications/mrgn-guidelines-enterobacteriacea/.

710   Australian Group on Antimicrobial Resistance. 2014. The evolution of carbapenemases in

711   Enterobacteriaceae in Australia. Available from: http://www.agargroup.org/surveys.

712   Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI,

713   Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, and

714   Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell

715   sequencing. *J Comput Biol* 19:455-477.

716   Blyth CC, Pereira L, and Goire N. 2014. New Delhi metallo-beta-lactamase-producing enterobacteriaceae

717   in an Australian child who had not travelled overseas. *Med J Aust* 200:386.

718   Bolger AM, Lohse M, and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.

719   *Bioinformatics* 30:2114-2120.

720   Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, and Drummond

721   AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*

722   10:e1003537.

723   Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, and Madden TL. 2009.

724   BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

725   Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup F, and

726   Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid

727   multilocus sequence typing. *Antimicrob Agents Chemother* 58:3895-3903.

728   Carter GP, Buultjens AH, Ballard SA, Baines SL, Tomita T, Strachan J, Johnson PD, Ferguson JK,

729   Seemann T, Stinear TP, and Howden BP. 2016. Emergence of endemic MLST non-typeable

730   vancomycin-resistant Enterococcus faecium. *J Antimicrob Chemother* 71:3367-3371.

731   Centers for Disease Control and Prevention, US Department of Health and Human Services. 2013.

732   Antibiotic resistance threats in the United States, 2013. Available from:

733   https://www.cdc.gov/drugresistance/threat-report-2013/index.html.

734   Centers for Disease Control and Prevention, US Department of Health and Human Services. 2015.

735   Facility Guidance for Control of Carbapenem-resistant Enterobacteriaceae (CRE). Available

736   from: https://www.cdc.gov/hai/organisms/cre/cre-toolkit/index.html.

737   Chang LW, Buising KL, Jeremiah CJ, Cronin K, Poy Lorenzo YS, Howden BP, Kwong J, Cocks J, Blood

738   A, Greenough J, and Waters MJ. 2015. Managing a nosocomial outbreak of carbapenem-resistant

739   Klebsiella pneumoniae: an early Australian hospital experience. *Intern Med J* 45:1037-1043.

740   Chen L, Chavda KD, Mediavilla JR, Jacobs MR, Levi MH, Bonomo RA, and Kreiswirth BN. 2012.

741   Partial excision of blaKPC from Tn4401 in carbapenem-resistant Klebsiella pneumoniae.

742   *Antimicrob Agents Chemother* 56:1635-1638.

743   Cheng L, Connor TR, Siren J, Aanensen DM, and Corander J. 2013. Hierarchical and spatially explicit

744   clustering of DNA sequences with BAPS software. *Mol Biol Evol* 30:1224-1228.

745   Chua KY, Grayson ML, Burgess AN, Lee JY, and Howden BP. 2014. The growing burden of multidrug-

746   resistant infections among returned Australian travellers. *Med J Aust* 200:116-118.

747   Coatsworth NR, Huntington PG, Hardiman RP, Hudson BJ, and Fernandes CJ. 2012. A case of

748   carbapenemase-producing Klebsiella pneumoniae in Australia. *Pathology* 44:42-44.

749   Conlan S, Park M, Deming C, Thomas PJ, Young AC, Coleman H, Sison C, Program NCS, Weingarten

750   RA, Lau AF, Dekker JP, Palmore TN, Frank KM, and Segre JA. 2016. Plasmid Dynamics in

751   KPC-Positive Klebsiella pneumoniae during Long-Term Patient Colonization. *MBio* 7.

752   Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, Snitkin ES, Clark TA, Luong K, Song Y,

753   Tsai YC, Boitano M, Dayal J, Brooks SY, Schmidt B, Young AC, Thomas JW, Bouffard GG,

754   Blakesley RW, Program NCS, Mullikin JC, Korlach J, Henderson DK, Frank KM, Palmore TN,

755   and Segre JA. 2014. Single-molecule sequencing to track plasmid diversity of hospital-associated

756   carbapenemase-producing Enterobacteriaceae. *Sci Transl Med* 6:254ra126.

757   Deleo FR, Chen L, Porcella SF, Martens CA, Kobayashi SD, Porter AR, Chavda KD, Jacobs MR,

758   Mathema B, Olsen RJ, Bonomo RA, Musser JM, and Kreiswirth BN. 2014. Molecular dissection

759   of the evolution of carbapenem-resistant multilocus sequence type 258 Klebsiella pneumoniae.

760   *Proc Natl Acad Sci U S A* 111:4988-4993.

761   Department of Health & Human Services, Victoria State Government. 2015. Victorian guidelines on

762   carbapenemase-producing Enterobacteriaceae. Melbourne: Victoria State Government. Available

763   from: https://www2.health.vic.gov.au/about/publications/policiesandguidelines/carbapenemase-

764   producing-enterobacteriaceae-guidelines.

765   Didelot X, Fraser C, Gardy J, and Colijn C. 2017. Genomic Infectious Disease Epidemiology in Partially

766   Sampled and Ongoing Outbreaks. *Mol Biol Evol* 34:997-1007.

767   Didelot X, and Wilson DJ. 2015. ClonalFrameML: efficient inference of recombination in whole bacterial

768   genomes. *PLoS Comput Biol* 11:e1004041.

769   Drummond AJ, Ho SY, Phillips MJ, and Rambaut A. 2006. Relaxed phylogenetics and dating with

770   confidence. *PLoS Biol* 4:e88.

771   Espedido BA, Steen JA, Ziochos H, Grimmond SM, Cooper MA, Gosbell IB, van Hal SJ, and Jensen SO.

772   2013. Whole genome sequence analysis of the first Australian OXA-48-producing outbreak-

773   associated Klebsiella pneumoniae isolates: the resistome and in vivo evolution. *PLoS One*

774   8:e59920.

775   Eyre DW, Cule ML, Griffiths D, Crook DW, Peto TE, Walker AS, and Wilson DJ. 2013. Detection of

776   mixed infection from bacterial whole genome sequence data allows assessment of its role in

777   Clostridium difficile transmission. *PLoS Comput Biol* 9:e1003059.

778 Falagas ME, Lourida P, Poulikakos P, Rafailidis PI, and Tansarli GS. 2014. Antibiotic treatment of

779   infections due to carbapenem-resistant Enterobacteriaceae: systematic evaluation of the available

780   evidence. *Antimicrob Agents Chemother* 58:654-663.

781 Fernando GA, Collignon PJ, and Bell JM. 2010. A risk for returned travellers: the "post-antibiotic era".

782   *Med J Aust* 193:59.

783 Gilchrist CA, Turner SD, Riley MF, Petri WA, Jr., and Hewlett EL. 2015. Whole-genome sequencing in

784   outbreak analysis. *Clin Microbiol Rev* 28:541-563.

785 Golubchik T, Batty EM, Miller RR, Farr H, Young BC, Larner-Svensson H, Fung R, Godwin H, Knox K,

786   Votintseva A, Everitt RG, Street T, Cule M, Ip CL, Didelot X, Peto TE, Harding RM, Wilson DJ,

787   Crook DW, and Bowden R. 2013. Within-host evolution of Staphylococcus aureus during

788   asymptomatic carriage. *PLoS One* 8:e61319.

789 Hasegawa M, Kishino H, and Yano T. 1985. Dating of the human-ape splitting by a molecular clock of

790   mitochondrial DNA. *J Mol Evol* 22:160-174.

791 Hatherell HA, Colijn C, Stagg HR, Jackson C, Winter JR, and Abubakar I. 2016. Interpreting whole

792   genome sequencing for investigating tuberculosis transmission: a systematic review. *BMC Med*

793   14:21.

794 Jiang Y, Wei Z, Wang Y, Hua X, Feng Y, and Yu Y. 2015. Tracking a hospital outbreak of KPC-

795   producing ST11 Klebsiella pneumoniae with whole genome sequencing. *Clin Microbiol Infect*

796   21:1001-1007.

797 Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, and Ferguson N. 2014. Bayesian reconstruction of

798   disease outbreaks by combining epidemiologic and genomic data. *PLoS Comput Biol*

799   10:e1003457.

800 Kanamori H, Parobek CM, Juliano JJ, van Duin D, Cairns BA, Weber DJ, and Rutala WA. 2017. A

801   Prolonged Outbreak of KPC-3-Producing Enterobacter cloacae and Klebsiella pneumoniae

802   Driven by Multiple Mechanisms of Resistance Transmission at a Large Academic Burn Center.

803   *Antimicrob Agents Chemother* 61.

804 Kitchel B, Rasheed JK, Patel JB, Srinivasan A, Navon-Venezia S, Carmeli Y, Brolund A, and Giske CG.

805   2009. Molecular epidemiology of KPC-producing Klebsiella pneumoniae isolates in the United

806   States: clonal expansion of multilocus sequence type 258. *Antimicrob Agents Chemother*

807   53:3365-3370.

808 Klinkenberg D, Backer JA, Didelot X, Colijn C, and Wallinga J. 2017. Simultaneous inference of

809   phylogenetic and transmission trees in infectious disease outbreaks. *PLoS Comput Biol*

810   13:e1005495.

811 Kotsanas D, Wijesooriya WR, Korman TM, Gillespie EE, Wright L, Snook K, Williams N, Bell JM, Li

812   HY, and Stuart RL. 2013. "Down the drain": carbapenem-resistant bacteria in intensive care unit

813   patients and handwashing sinks. *Med J Aust* 198:267-269.

814 Kwong JC, McCallum N, Sintchenko V, and Howden BP. 2015. Whole genome sequencing in clinical

815   and public health microbiology. *Pathology* 47:199-210.

816 Kwong JC, Mercoulia K, Tomita T, Easton M, Li HY, Bulach DM, Stinear TP, Seemann T, and Howden

817   BP. 2016. Prospective Whole-Genome Sequencing Enhances National Surveillance of Listeria

818   monocytogenes. *J Clin Microbiol* 54:333-342.

819 Marsh JW, Krauland MG, Nelson JS, Schlackman JL, Brooks AM, Pasculle AW, Shutt KA, Doi Y,

820   Querry AM, Muto CA, and Harrison LH. 2015. Genomic Epidemiology of an Endoscope-

821   Associated Outbreak of Klebsiella pneumoniae Carbapenemase (KPC)-Producing K.

822   pneumoniae. *PLoS One* 10:e0144310.

823 Mathers AJ, Stoesser N, Chai W, Carroll J, Barry K, Cherunvanky A, Sebra R, Kasarskis A, Peto TE,

824   Walker AS, Sifri CD, Crook DW, and Sheppard AE. 2017. Chromosomal Integration of the

825   Klebsiella pneumoniae Carbapenemase Gene, blaKPC, in Klebsiella Species Is Elusive but Not

826   Rare. *Antimicrob Agents Chemother* 61.

827 Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Giess A, Yeh AJ, Didelot X, Turner SD, Sebra R,

828   Kasarskis A, Peto T, Crook D, and Sifri CD. 2015. Klebsiella pneumoniae carbapenemase

829   (KPC)-producing K. pneumoniae at a single institution: insights into endemicity from whole-

830   genome sequencing. *Antimicrob Agents Chemother* 59:1656-1663.

831 Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican M, Cornaglia G, Garau J,

832   Gniadkowski M, Hayden MK, Kumarasamy K, Livermore DM, Maya JJ, Nordmann P, Patel JB,

833   Paterson DL, Pitout J, Villegas MV, Wang H, Woodford N, and Quinn JP. 2013. Clinical

834   epidemiology of the global expansion of Klebsiella pneumoniae carbapenemases. *Lancet Infect*

835   *Dis* 13:785-796.

836 Naas T, Cuzon G, Truong HV, and Nordmann P. 2012. Role of ISKpn7 and deletions in blaKPC gene

837   expression. *Antimicrob Agents Chemother* 56:4753-4759.

838 Naas T, Cuzon G, Villegas MV, Lartigue MF, Quinn JP, and Nordmann P. 2008. Genetic structures at the

839   origin of acquisition of the beta-lactamase bla KPC gene. *Antimicrob Agents Chemother* 52:1257-

840   1263.

841 Nordmann P, Poirel L, and Dortet L. 2012. Rapid detection of carbapenemase-producing

842   Enterobacteriaceae. *Emerg Infect Dis* 18:1503-1507.

843 Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, and

844   Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*

845   31:3691-3693.

846 Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, and Harris SR. 2016. SNP-sites: rapid

847   efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* 2:e000056.

848 Palmore TN, and Henderson DK. 2013. Managing transmission of carbapenem-resistant

849   enterobacteriaceae in healthcare settings: a view from the trenches. *Clin Infect Dis* 57:1593-1599.

850 Peleg AY, Franklin C, Bell JM, and Spelman DW. 2005. Dissemination of the metallo-beta-lactamase

851   gene blaIMP-4 among gram-negative pathogens in a clinical setting in Australia. *Clin Infect Dis*

852   41:1549-1556.

853   Poirel L, Lagrutta E, Taylor P, Pham J, and Nordmann P. 2010. Emergence of metallo-beta-lactamase
854   NDM-1-producing multidrug-resistant Escherichia coli in Australia. *Antimicrob Agents*
855   *Chemother* 54:4914-4916.

856   Price MN, Dehal PS, and Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large
857   alignments. *PLoS One* 5:e9490.

858   Queenan AM, and Bush K. 2007. Carbapenemases: the versatile beta-lactamases. *Clin Microbiol Rev*
859   20:440-458, table of contents.

860   Rambaut A, Lam TT, Max Carvalho L, and Pybus OG. 2016. Exploring the temporal structure of
861   heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2:vew007.

862   Ruppe E, Olearo F, Pires D, Baud D, Renzi G, Cherkaoui A, Goldenberger D, Huttner A, Francois P,
863   Harbarth S, and Schrenzel J. 2017. Clonal or not clonal? Investigating hospital outbreaks of KPC-
864   producing Klebsiella pneumoniae with whole-genome sequencing. *Clin Microbiol Infect*.

865   Schwaber MJ, and Carmeli Y. 2014. An ongoing national intervention to contain the spread of
866   carbapenem-resistant enterobacteriaceae. *Clin Infect Dis* 58:697-703.

867   Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068-2069.

868   Sidjabat H, Nimmo GR, Walsh TR, Binotto E, Htin A, Hayashi Y, Li J, Nation RL, George N, and
869   Paterson DL. 2011. Carbapenem resistance in Klebsiella pneumoniae due to the New Delhi
870   Metallo-beta-lactamase. *Clin Infect Dis* 52:481-484.

871   Sidjabat HE, Kennedy K, Silvey A, Collignon P, and Paterson DL. 2013. Emergence of bla(OXA-181)-
872   carrying ColE plasmid in Klebsiella pneumoniae in Australia. *Int J Antimicrob Agents* 41:294-
873   296.

874   Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Group NCSP, Henderson DK, Palmore TN, and Segre JA.
875   2012. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-
876   genome sequencing. *Sci Transl Med* 4:148ra116.

877   Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
878   phylogenies. *Bioinformatics* 30:1312-1313.

879   Tai AY, Stuart RL, Sidjabat HE, Lemoh CN, Rogers BA, Graham M, Paterson DL, and Korman TM.

880   2015. Local acquisition and nosocomial transmission of Klebsiella pneumoniae harbouring the

881   blaNDM-1 gene in Australia. *Med J Aust* 202:270-272.

882   Tenover FC, Canton R, Kop J, Chan R, Ryan J, Weir F, Ruiz-Garbajosa P, LaBombardi V, and Persing

883   DH. 2013. Detection of colonization by carbapenemase-producing Gram-negative Bacilli in

884   patients by use of the Xpert MDRO assay. *J Clin Microbiol* 51:3780-3787.

885   Tong SY. 2013. Genomic polish for shoe-leather epidemiology. *Nat Rev Microbiol* 11:8.

886   Tzouvelekis LS, Markogiannakis A, Psichogiou M, Tassios PT, and Daikos GL. 2012. Carbapenemases

887   in Klebsiella pneumoniae and other Enterobacteriaceae: an evolving crisis of global dimensions.

888   *Clin Microbiol Rev* 25:682-707.

889   Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J,

890   Young SK, and Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant

891   detection and genome assembly improvement. *PLoS One* 9:e112963.

892   Weterings V, Zhou K, Rossen JW, van Stenis D, Thewessen E, Kluytmans J, and Veenemans J. 2015. An

893   outbreak of colistin-resistant Klebsiella pneumoniae carbapenemase-producing Klebsiella

894   pneumoniae in the Netherlands (July to December 2013), with inter-institutional spread. *Eur J

895   Clin Microbiol Infect Dis* 34:1647-1655.

896   Worby CJ, Lipsitch M, and Hanage WP. 2014. Within-host bacterial diversity hinders accurate

897   reconstruction of transmission networks from genomic distance data. *PLoS Comput Biol*

898   10:e1003549.

899   World Health Organization. 2014. Antimicrobial resistance: global report on surveillance. Available

900   from: http://www.who.int/drugresistance/documents/surveillancereport/en/.

901   Ypma RJ, van Ballegooijen WM, and Wallinga J. 2013. Relating phylogenetic trees to transmission trees

902   of infectious disease outbreaks. *Genetics* 195:1055-1062.

903   Zimmerman FS, Assous MV, Bdolah-Abram T, Lachish T, Yinnon AM, and Wiener-Well Y. 2013.

904   Duration of carriage of carbapenem-resistant Enterobacteriaceae following hospital discharge. *Am*

905   *J Infect Control* 41:190-194.

906

907

908

909   **FIGURE LEGENDS**

910

911   **Figure 1: Workflow summary diagram of methods used in this study.**

912   Microbiology methods (orange), bioinformatic methods (blue), and epidemiological methods

913   (green) used to generate data for analysis and reporting to public health authorities (purple) are

914   shown. Results of combined prospective epidemiological and genomic analyses performed

915   iteratively during the outbreak were reported to the Department of Health and Human Services in

916   real-time. CPE = carbapenemase-producing Enterobacteriaceae; KPC = *Klebsiella pneumoniae*

917   carbapenemase; MALDI-TOF = matrix-assisted laser desorption/ionisation time-of-flight; PCR =

918   polymerase chain reaction; WGS = whole-genome sequencing; SNP = single nucleotide

919   polymorphism; AMR = antimicrobial resistance; MLST = multi-locus sequence typing.

920

921   **Figure 2: Incidence of new KPC-producing Enterobacteriaceae cases referred to MDU PHL,**

922   **2012-2015.**

923   Blocks are coloured by the species and KPC allele of the referred isolates. Repeated detections of

924   KPC-producing isolates from the same patient have been excluded.

925

926   **Figure 3: The initial maximum likelihood phylogenetic tree comprised three dominant**

927   **clades**.

928   The tree includes 29 ST258 *K. pneumoniae* isolates collected from 29 patients prior to June 2014,

929   with external nodes coloured according to healthcare facility at time of sample collection.

930   Recurrent isolates from each patient have been excluded. Multiple patients in Facility A (purple)

931   and Facility F (orange) were colonised/infected with KPC-producing *K. pneumoniae*,

932 corresponding to known previous outbreaks in those hospitals. Support values (%) from 1000

933 bootstrap replicates are shown for major branches. Major phylogenetic clades have been labelled

934 cluster A (green shading), B (orange shading), and C (purple shading) in the order that the clades

935 emerged, with the larger clade B comprising two subclades, B1 and B2. Corresponding clusters

936 identified through Bayesian Analysis of the Population Structure (BAPS) are also shown. The tree

937 was rooted using an outgroup isolate (*K. pneumoniae* NJST258_1, GenBank accession:

938 CP006923.1; not shown in the tree) from a different ST258 clade.

939

940 **Figure 4: Epidemiological links based on overlapping patient admissions were unable to**

941 **resolve where and when transmission was occurring for most isolates.**

942 Links in the above network connect patients based on overlapping patient admissions in the same

943 hospital ward at the same time (minimum of one day overlap). Links that occurred after detection

944 of bla_{KPC} in both patients have been excluded. Nodes are numbered by patient, and coloured by

945 healthcare facility (as in Figure 3) at the time of sample collection. The epidemiological network

946 connecting patients 22, 23, 24 and 33 correlates with the closely related genomes in Cluster C

947 (Figure 3).

948

949 **Figure 5: Combined analysis, with genomic relationships between isolates overlaid upon**

950 **epidemiological data, delineated multiple transmission networks.**

951 A maximum likelihood phylogenetic tree is shown on the left, labelled with major genomic clusters

952 and supporting branch bootstrap values (%) from 1000 replicates for major branches. Nodes of the

953 tree are coloured by healthcare facility at the time of sample collection as in Figures 3 and 4.

954 Coloured horizontal bars on the right indicate healthcare facility admissions over time (x axis),

955 with different colours representing different healthcare networks. Black diamonds (◑) indicate

956 first detection of KPC for each patient.

957

958 **Figure 6: Combined analysis of genomic and epidemiological data of the cluster C network**

959 **identified secondary transmission.**

960 Example case study: Patients 22 and 23 both reported overseas hospitalisation in the 12 months

961 prior to first detection of KPC – patient 22 in Vietnam for spinal surgery following a motor vehicle

962 accident, and patient 23 in Greece for stem cell therapy. Both patients had undergone rectal

963 screening on admission to Ward 1 in Facility A, with patient 22 being placed in intensive contact

964 precautions for the duration of his hospital admissions after isolating another multidrug-resistant

965 organism, though was required to use shared bathroom facilities with patients in the adjacent room.

966 Having required treatment with meropenem for both hospital-acquired pneumonia and a surgical

967 wound infection, patient 22 was later diagnosed with a KPC-producing *K. pneumoniae* indwelling

968 catheter-associated urinary infection in January 2014. Twelve days later, patient 23 was

969 subsequently found to have a polymicrobial sacral wound infection, with cultures including KPC-

970 producing *K. pneumoniae* from sacral tissue. In response to this, all patients on the ward who had

971 been admitted to the same room and/or shared bathroom facilities with patients 22 and 23 were

972 screened, with the subsequent identification of patient 24. Alerts were placed on the records of

973 patients meeting the criteria who had been previously discharged. Environmental screening of the

974 rooms and bathrooms was conducted, with no KPC-producing organisms identified, and extended

975 bleach cleaning with changes of curtains, chairs and other furnishings was conducted for the entire

976 ward.

977

978 Patient 33 was also admitted to Ward 1 in Facility A in February 2014, subsequent to the

979 identification of KPC-2 in patients 22, 23 and 24. This patient was not screened as he had not been

980 admitted to the same room, nor had he shared a bathroom with the identified cases. He also

981 reported no recent history of overseas travel. However, he was identified in July 2014 through

982 routine screening at Facility A following transfer from Ward 3, Facility U, located 25 km away. A

983 KPC-producing isolate from Patient 39 was identified in September 2014, and although the isolate

984 genomically clustered with isolates from Patients 22, 23, 24 and 33 identified at Facility A, she

985 had no previous presentations to that healthcare facility. However, immediately prior to

986 identification of KPC, she had also been in Facility U on Ward 3, though she was admitted there

987 13 days after Patient 33's discharge. Given this was the only plausible epidemiological link to the

988 other cluster C patients, secondary transmission was presumed to have occurred in Ward 3, Facility

989 U.

990

991 **Figure 7: Bayesian evolutionary analysis indicates each of the phylogenetic clades**

992 **corresponding to the genomic clusters emerged prior to the detection of KPC in Victoria.**

993 A maximum clade credibility timed phylogeny from Bayesian evolutionary analysis of local

994 CC258 *K. pneumoniae* isolate genomes is shown, with median node heights displayed. The thin

995 red bars indicate 95% highest posterior density (HPD) intervals for the most recent common

996 ancestor (MRCA) for major clades and defined genomic clusters (indicated on the right). The

997 shaded grey region indicates the recent period when KPC isolates were detected in Victoria.

998

999 **Figure 8: An inferred transmission tree shows that undetected colonisation was significant**

1000 **in propagating the outbreak.**

1001   Solid nodes represent the posterior mean time of KPC acquisition by individuals and are coloured

1002   by the corresponding genomic cluster, with empty circles representing inferred unsampled

1003   individuals contributing to the transmission tree. Braches are shaded by number of missing links

1004   in the transmission tree, with lighter branches representing increasing numbers of missing links

1005   implicated.

1006

1007   **Figure 9: KPC plasmids from the *C. farmeri* isolates were almost identical to KPC plasmids**

1008   **from *K. pneumoniae* outbreak isolates, despite differing replication proteins.**

1009   BLAST comparison between an IncFIB (pQIL-like) plasmid genome from a ST258 *Klebsiella*

1010   *pneumoniae* isolate, AUSMDU00008079 (above), and an IncR plasmid genome from a

1011   *Citrobacter farmeri* isolate, AUSMDU00008141 (below), from the outbreak. The grey shading

1012   indicates corresponding DNA regions of high nucleotide identity transcribed in opposing

1013   directions, with the Tn*4401* transposon harbouring bla_{KPC-2} highlighted in red. The plasmid

1014   genomes have been orientated to their respective replicons and downstream plasmid partitioning

1015   genes, *parA* and *parB*.

1016

1017   **Figure 10: Genomes of isolates from the same host group together in the phylogeny.**

1018   Maximum likelihood tree of the study isolates with additional isolates obtained up to 1 November

1019   2016 included. Multiple isolates from the same patient have been coloured by patient, with the

1020   accompanying graph indicating the collection dates for the corresponding isolates. Thirty-two

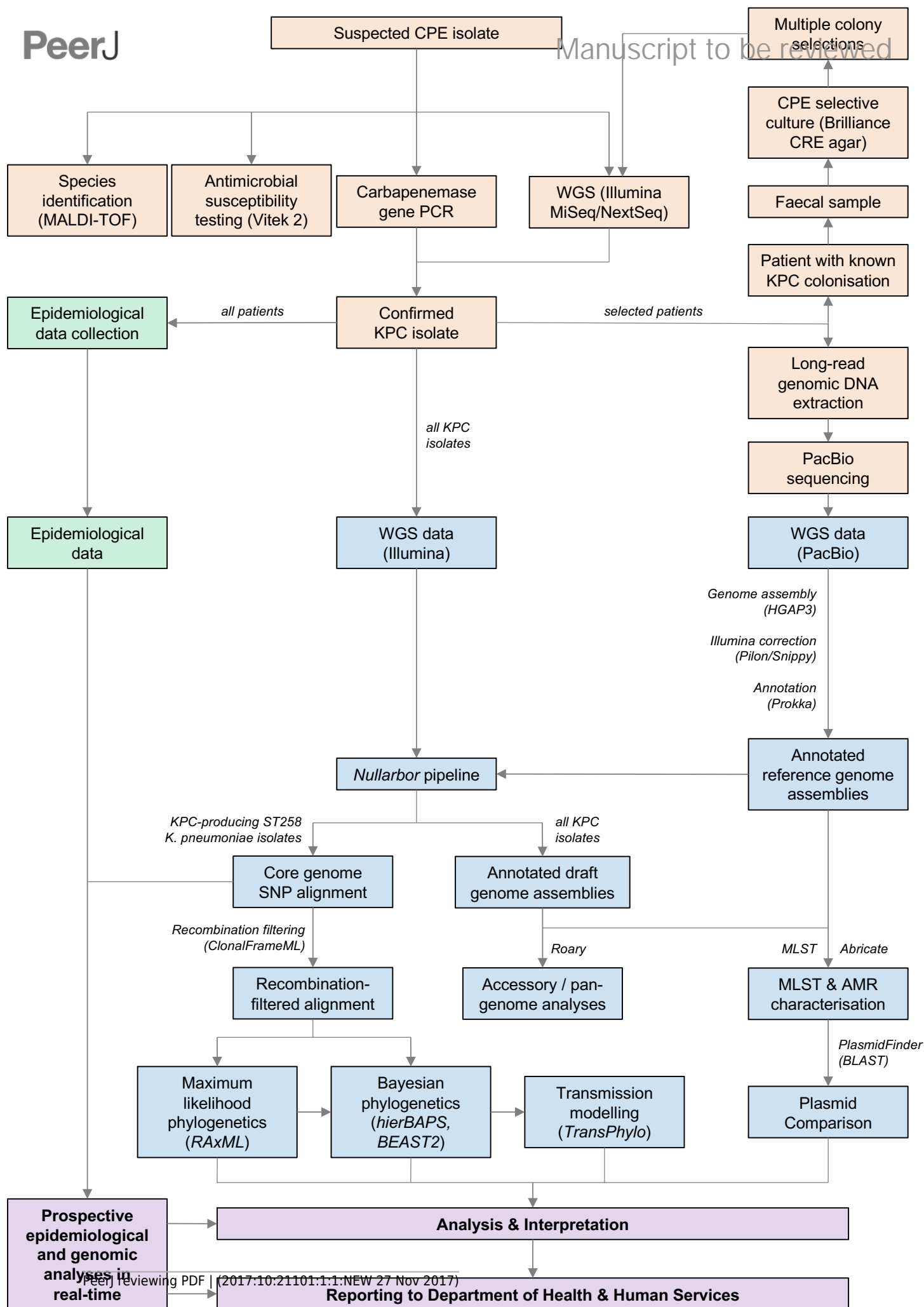
1021   isolates from six clinical samples obtained from patient 70 (blue) over 8 months have been

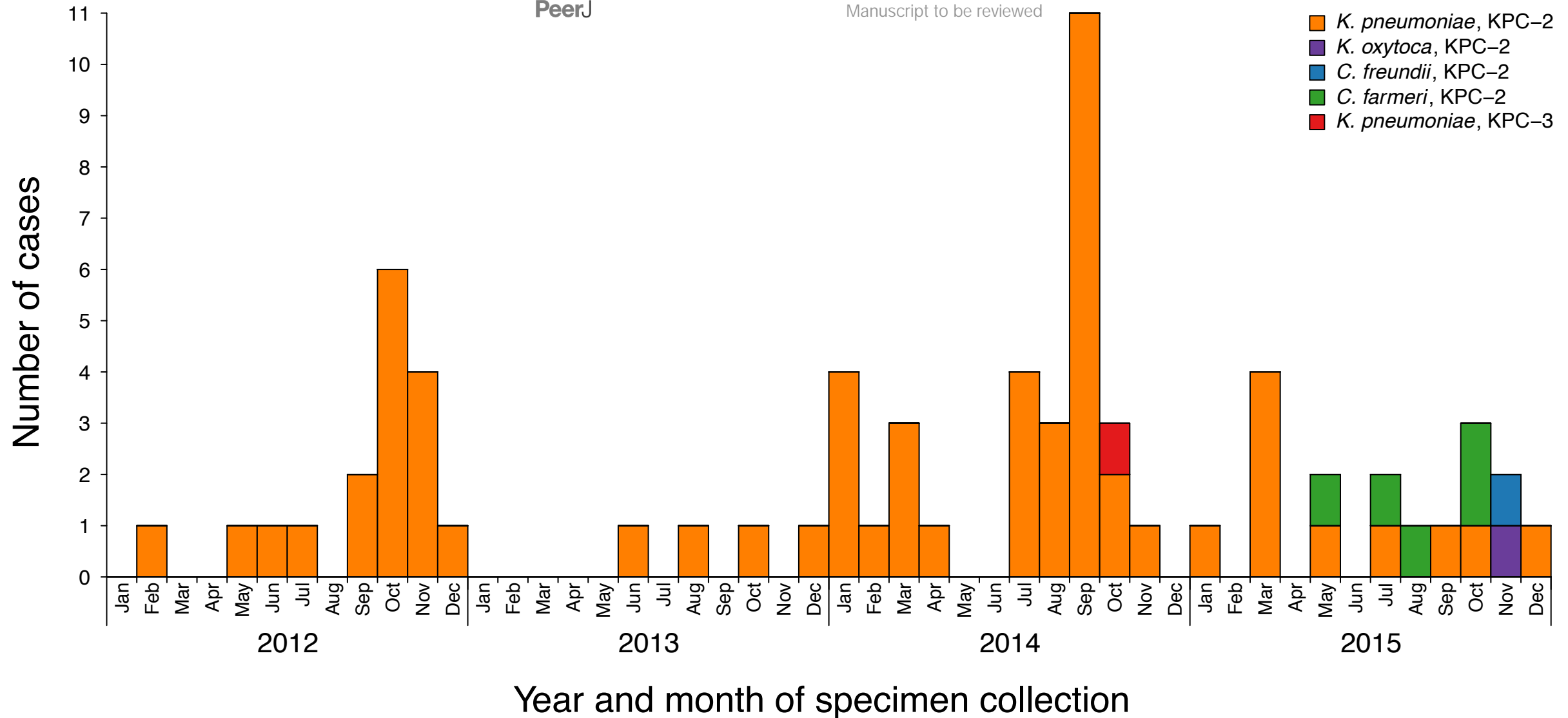
1022   highlighted to illustrate the within-host lineages emerging in this patient. Six environmental

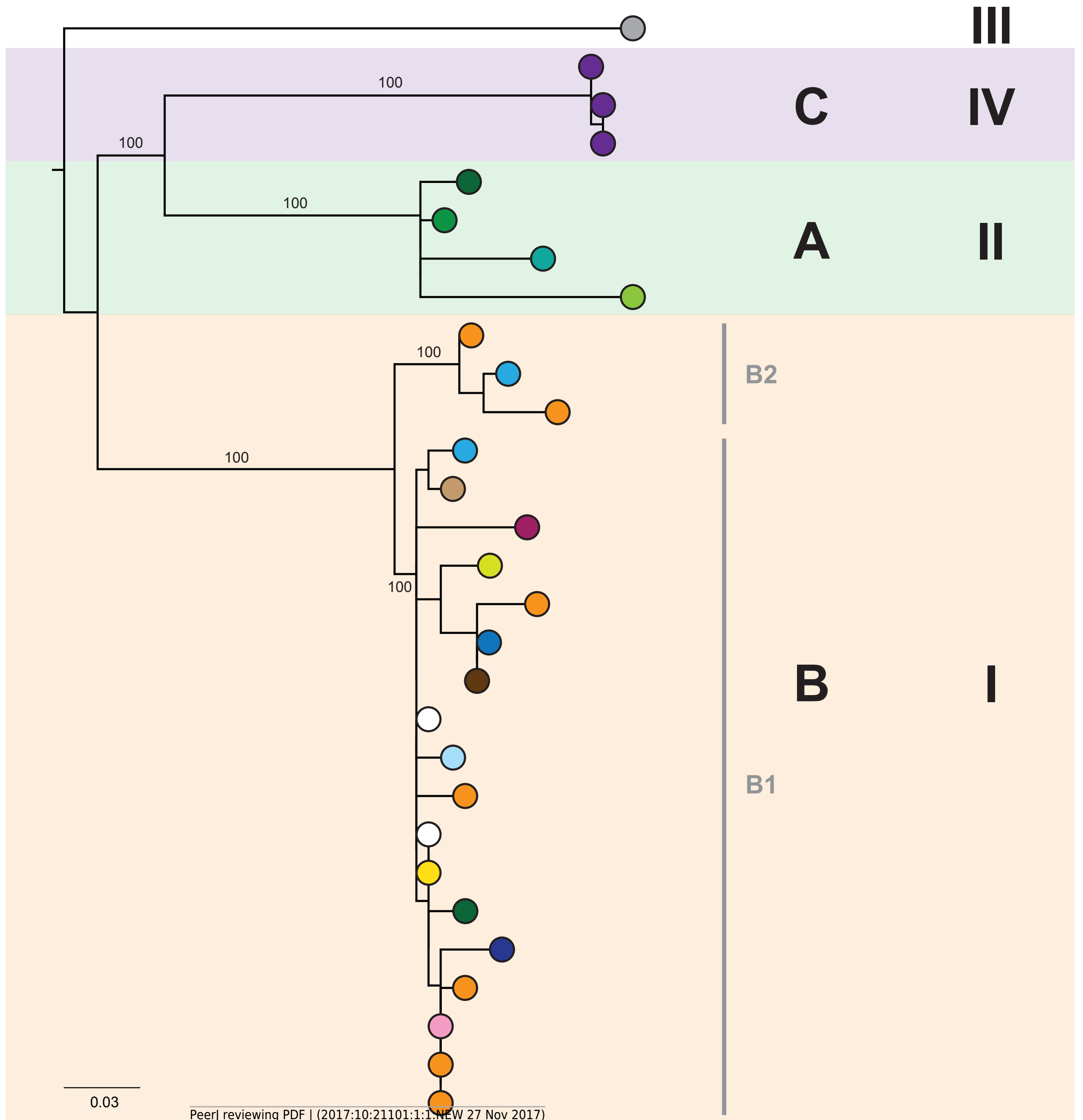
1023 isolates from the room of patient 75 are also shown (light pink). Bootstrap values (%) from 1000

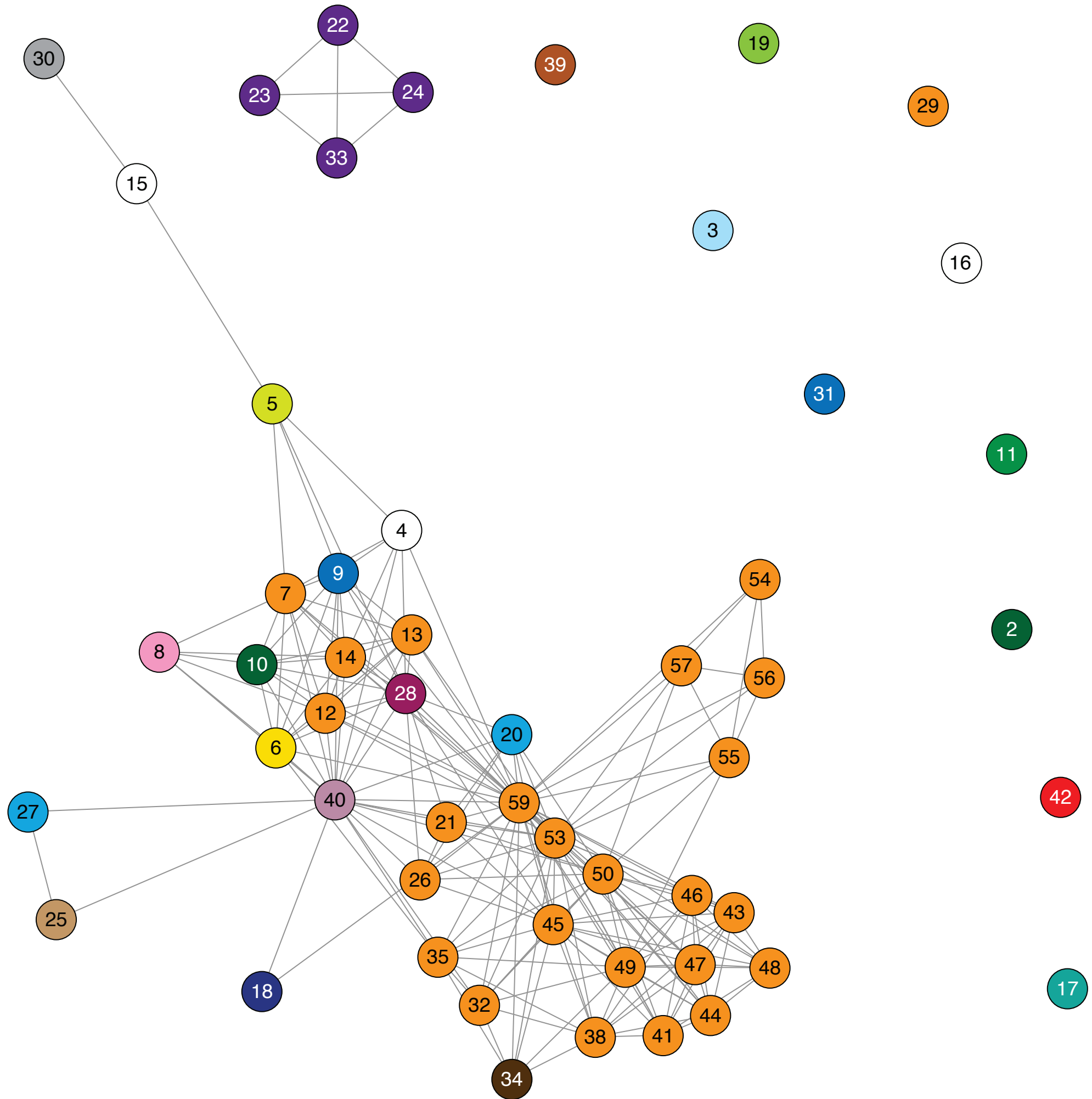
1024 replicates have been displayed for major branches in the tree.

Workflow summary diagram of methods used in this study.

Microbiology methods (orange), bioinformatic methods (blue), and epidemiological methods (green) used to generate data for analysis and reporting to public health authorities (purple) are shown. Results of combined prospective epidemiological and genomic analyses performed iteratively during the outbreak were reported to the Department of Health and Human Services in real-time. CPE = carbapenemase-producing Enterobacteriaceae; KPC = *Klebsiella pneumoniae* carbapenemase; MALDI-TOF = matrix-assisted laser desorption/ionisation time-of-flight; PCR = polymerase chain reaction; WGS = whole-genome sequencing; SNP = single nucleotide polymorphism; AMR = antimicrobial resistance; MLST = multi-locus sequence typing.

Incidence of new KPC-producing Enterobacteriaceae cases referred to MDU PHL, 2012-2015.

Blocks are coloured by the species and KPC allele of the referred isolates. Repeated detections of KPC-producing isolates from the same patient have been excluded.

The initial maximum likelihood phylogenetic tree comprised three dominant clades.

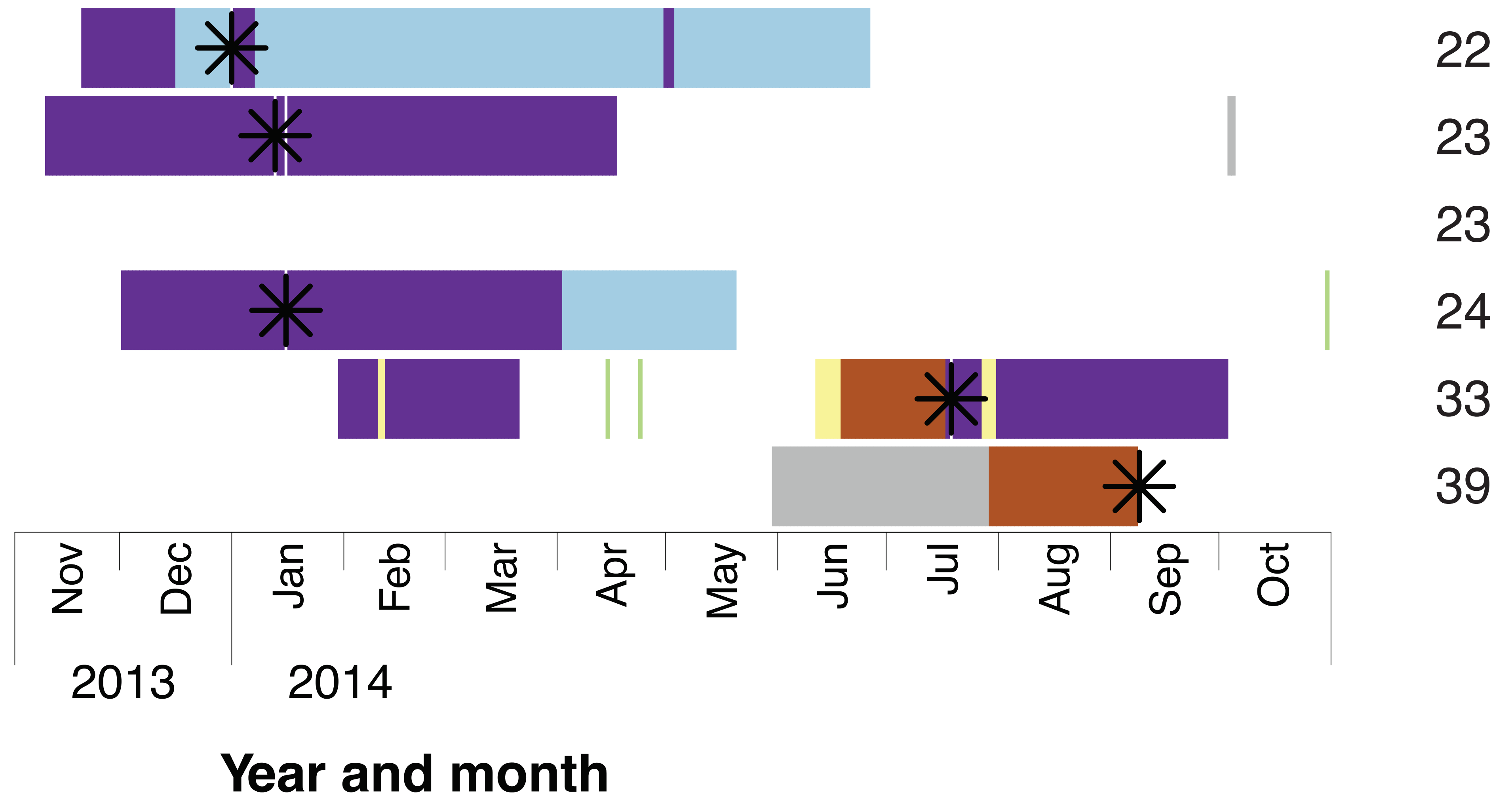
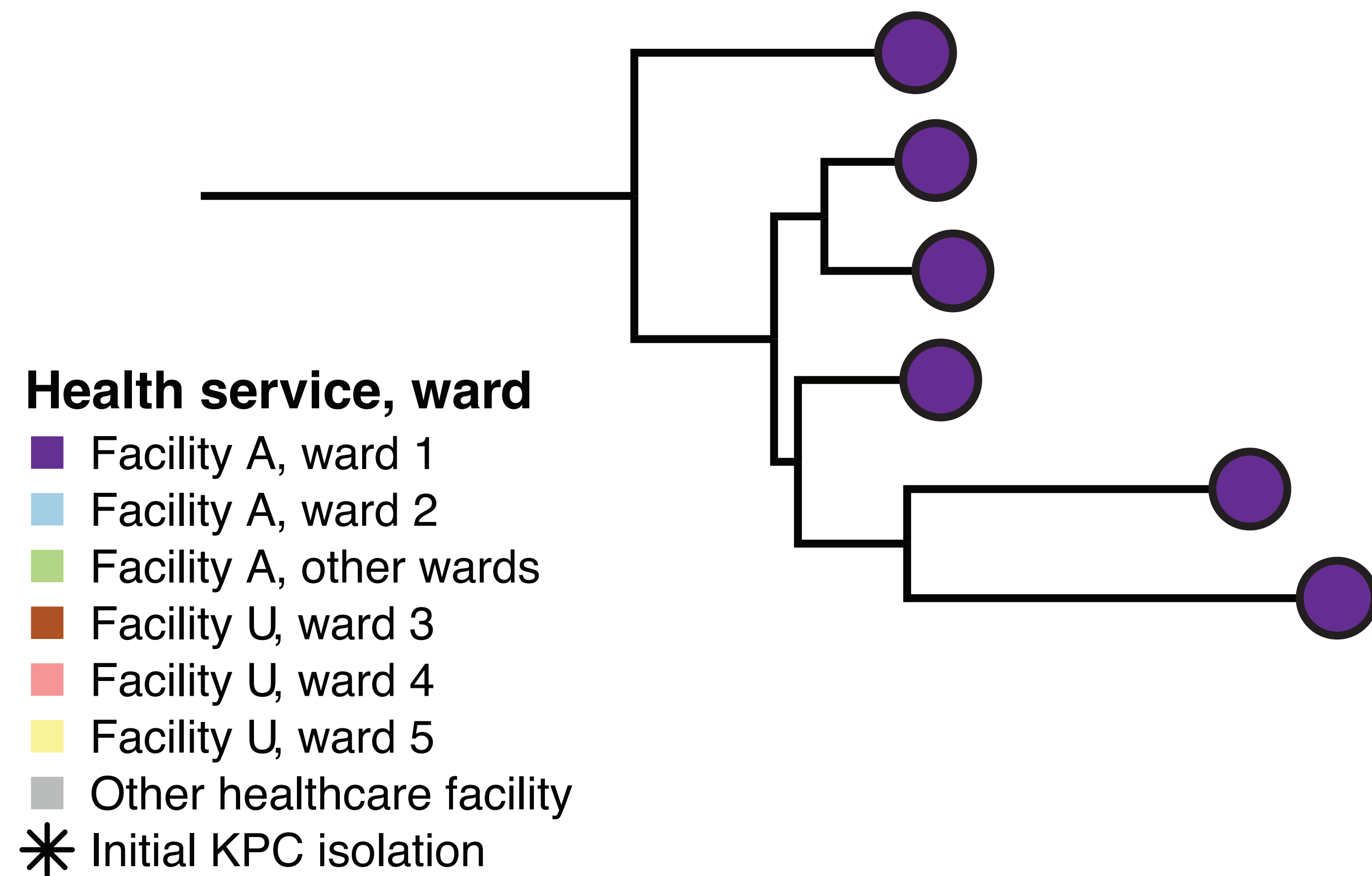
The tree includes 29 ST258 *K. pneumoniae* isolates collected from 29 patients prior to June 2014, with external nodes coloured according to healthcare facility at time of sample collection. Recurrent isolates from each patient have been excluded. Multiple patients in Facility A (purple) and Facility F (orange) were colonised/infected with KPC-producing *K. pneumoniae*, corresponding to known previous outbreaks in those hospitals. Support values (%) from 1000 bootstrap replicates are shown for major branches. Major phylogenetic clades have been labelled cluster A (green shading), B (orange shading), and C (purple shading) in the order that the clades emerged, with the larger clade B comprising two subclades, B1 and B2. Corresponding clusters identified through Bayesian Analysis of the Population Structure (BAPS) are also shown. The tree was rooted using an outgroup isolate (*K. pneumoniae* NJST258_1, GenBank accession: CP006923.1; not shown in the tree) from a different ST258 clade.

Epidemiological links based on overlapping patient admissions were unable to resolve where and when transmission was occurring for most isolates.

Links in the above network connect patients based on overlapping patient admissions in the same hospital ward at the same time (minimum of one day overlap). Links that occurred after detection of bla_{KPC} in both patients have been excluded. Nodes are numbered by patient, and coloured by healthcare facility (as in Figure 3) at the time of sample collection. The epidemiological network connecting patients 22, 23, 24 and 33 correlates with the closely related genomes in Cluster C (Figure 3).

Combined analysis, with genomic relationships between isolates overlaid upon epidemiological data, delineated multiple transmission networks.

A maximum likelihood phylogenetic tree is shown on the left, labelled with major genomic clusters and supporting branch bootstrap values (%) from 1000 replicates for major branches. Nodes of the tree are coloured by healthcare facility at the time of sample collection as in Figures 3 and 4. Coloured horizontal bars on the right indicate healthcare facility admissions over time (x axis), with different colours representing different healthcare networks. Black diamonds (¿) indicate first detection of KPC for each patient.

Combined analysis of genomic and epidemiological data of the cluster C network identified secondary transmission.

Example case study: Patients 22 and 23 both reported overseas hospitalisation in the 12 months prior to first detection of KPC – patient 22 in Vietnam for spinal surgery following a motor vehicle accident, and patient 23 in Greece for stem cell therapy. Both patients had undergone rectal screening on admission to Ward 1 in Facility A, with patient 22 being placed in intensive contact precautions for the duration of his hospital admissions after isolating another multidrug-resistant organism, though was required to use shared bathroom facilities with patients in the adjacent room. Having required treatment with meropenem for both hospital-acquired pneumonia and a surgical wound infection, patient 22 was later diagnosed with a KPC-producing *K. pneumoniae* indwelling catheter-associated urinary infection in January 2014. Twelve days later, patient 23 was subsequently found to have a polymicrobial sacral wound infection, with cultures including KPC-producing *K. pneumoniae* from sacral tissue. In response to this, all patients on the ward who had been admitted to the same room and/or shared bathroom facilities with patients 22 and 23 were screened, with the subsequent identification of patient 24. Alerts were placed on the records of patients meeting the criteria who had been previously discharged. Environmental screening of the rooms and bathrooms was conducted, with no KPC-producing organisms identified, and extended bleach cleaning with changes of curtains, chairs and other furnishings was conducted for the entire ward. Patient 33 was also admitted to Ward 1 in Facility A in February 2014, subsequent to the identification of KPC-2 in patients 22, 23 and 24. This patient was not screened as he had not been admitted to the same room, nor had he shared a bathroom with the identified cases. He also reported no recent history of overseas travel. However, he was identified in July 2014 through routine screening at Facility A following transfer from Ward 3, Facility U, located 25 km away. A KPC-producing isolate from Patient 39 was identified in September 2014, and although the isolate genomically clustered with isolates from Patients 22, 23, 24

and 33 identified at Facility A, she had no previous presentations to that healthcare facility. However, immediately prior to identification of KPC, she had also been in Facility U on Ward 3, though she was admitted there 13 days after Patient 33's discharge. Given this was the only plausible epidemiological link to the other cluster C patients, secondary transmission was presumed to have occurred in Ward 3, Facility U.

Bayesian evolutionary analysis indicates each of the phylogenetic clades corresponding to the genomic clusters emerged prior to the detection of KPC in Victoria.

A maximum clade credibility timed phylogeny from Bayesian evolutionary analysis of local CC258 *K. pneumoniae* isolate genomes is shown, with median node heights displayed. The thin red bars indicate 95% highest posterior density (HPD) intervals for the most recent common ancestor (MRCA) for major clades and defined genomic clusters (indicated on the right). The shaded grey region indicates the recent period when KPC isolates were detected in Victoria.

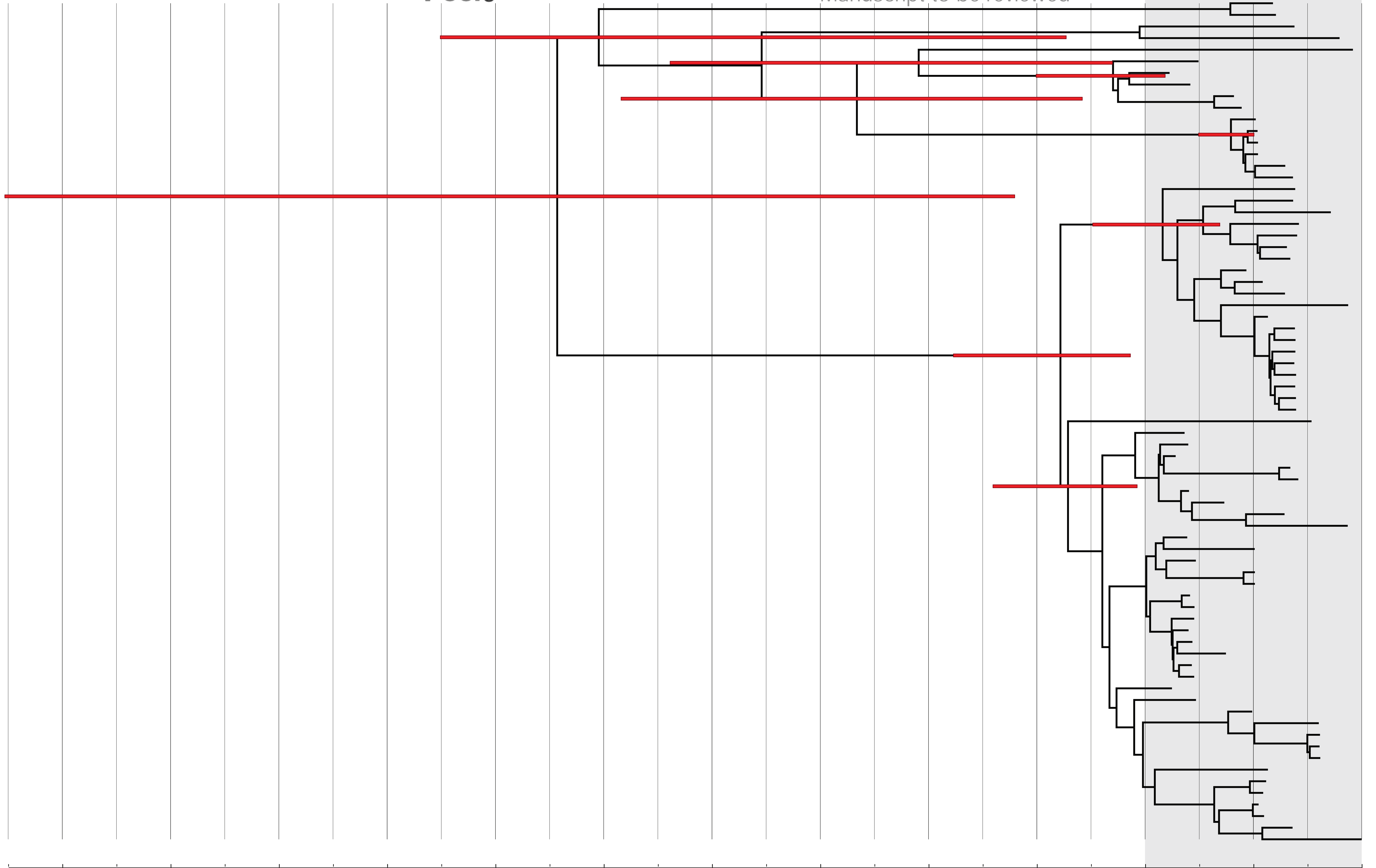
GENOMIC
CLUSTER

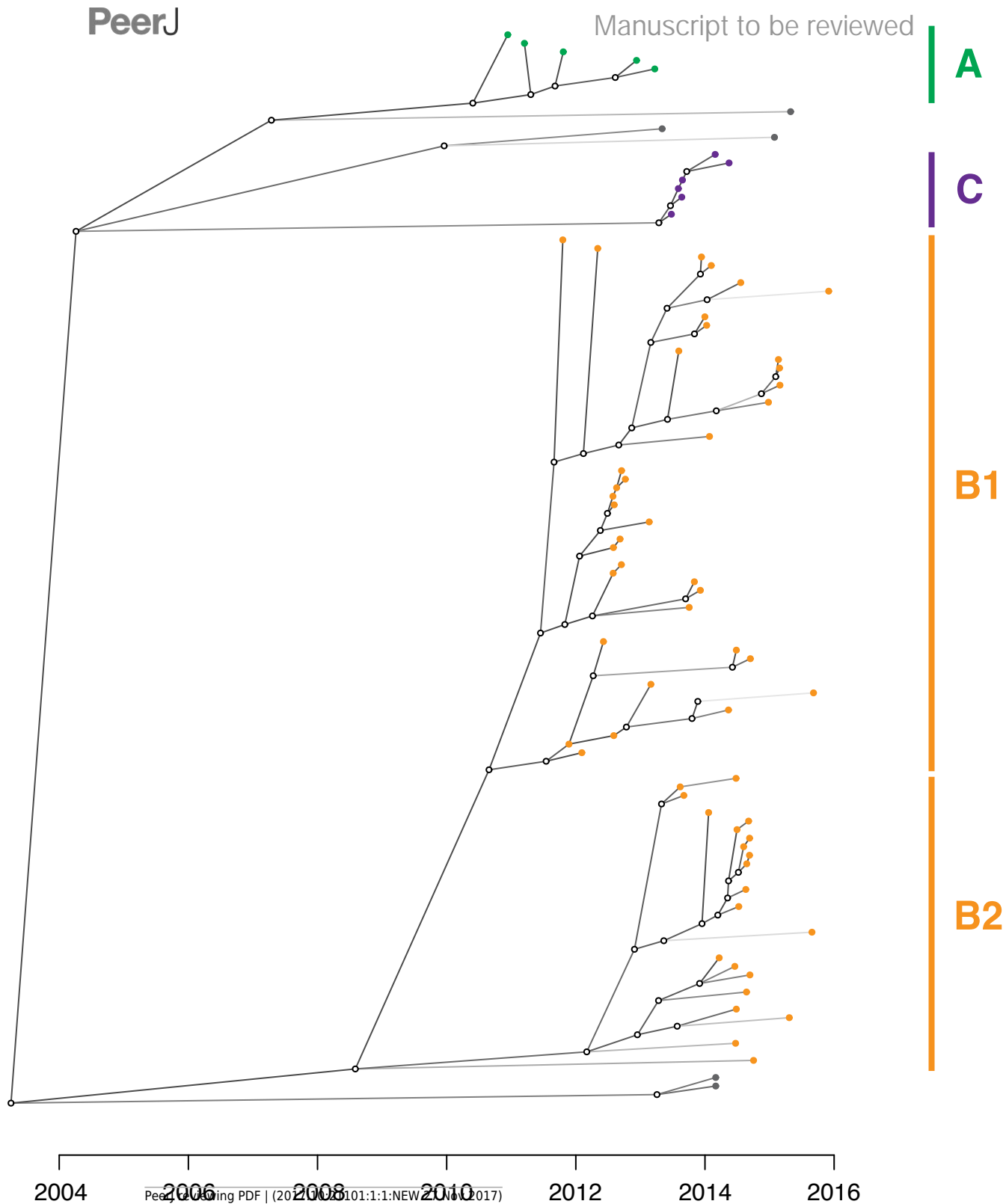
A

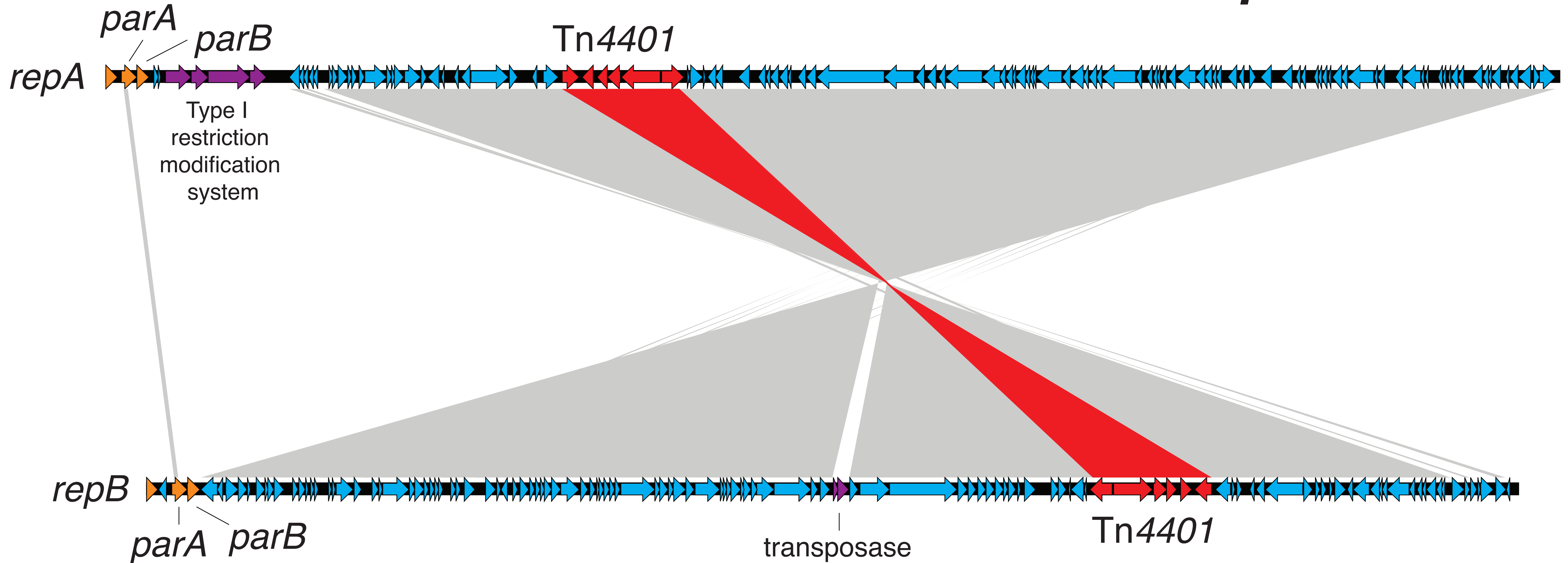
C

B2

B1

1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016

An inferred transmission tree shows that undetected colonisation was significant in propagating the outbreak.

Solid nodes represent the posterior mean time of KPC acquisition by individuals and are coloured by the corresponding genomic cluster, with empty circles representing inferred unsampled individuals contributing to the transmission tree. Braches are shaded by number of missing links in the transmission tree, with lighter branches representing increasing numbers of missing links implicated.

KPC plasmids from the *C. farmeri* isolates were almost identical to KPC plasmids from *K. pneumoniae* outbreak isolates, despite differing replication proteins.

BLAST comparison between an IncFIB (pQIL-like) plasmid genome from a ST258 *Klebsiella pneumoniae* isolate, AUSMDU00008079 (above), and an IncR plasmid genome from a *Citrobacter farmeri* isolate, AUSMDU00008141 (below), from the outbreak. The grey shading indicates corresponding DNA regions of high nucleotide identity transcribed in opposing directions, with the Tn*4401* transposon harbouring bla_{KPC-2} highlighted in red. The plasmid genomes have been orientated to their respective replicons and downstream plasmid partitioning genes, *parA* and *parB*.

Genomes of isolates from the same host group together in the phylogeny.

Maximum likelihood tree of the study isolates with additional isolates obtained up to 1 November 2016 included. Multiple isolates from the same patient have been coloured by patient, with the accompanying graph indicating the collection dates for the corresponding isolates. Thirty-two isolates from six clinical samples obtained from patient 70 (blue) over 8 months have been highlighted to illustrate the within-host lineages emerging in this patient. Six environmental isolates from the room of patient 75 are also shown (light pink). Bootstrap values (%) from 1000 replicates have been displayed for major branches in the tree.

100

100

100

100

100

100

100

100

96

100

91

0.02

● Environmental sampling
  isolates

● Isolates from patients
  with a single isolate

Patient code

74
72
71
70
43
40
34
30
27
25
23
19
18
10
9
8

2012   2013   2014   2015   2016   2017

Date of collection