

Megaplastids on the Rise: Combining Sequencing Approaches to Fully Resolve a Carbapenemase-Encoding Plasmid in a Proposed Novel *Pseudomonas* Species

João Botelho^{1*}, Cédric Lood^{2,3}, Sally R. Partridge⁴, Vera van Noort^{2,5}, Rob Lavigne³, Filipa Grosso¹, Luísa Peixe^{1*}

¹UCIBIO/REQUIMTE, Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal

²Centre of Microbial and Plant Genetics, Department of Microbial and Molecular Systems, KU Leuven, Leuven, Belgium

³Laboratory of Gene Technology, Department of Biosystems, KU Leuven, Leuven, Belgium

⁴Centre for Microbiology and Infectious Diseases, The Westmead Institute for Medical Research, The University of Sydney, Westmead Hospital, NSW 2145, Australia

⁵Institute of Biology, Leiden University, Leiden, The Netherlands

*Addresses for correspondence: João Botelho, Laboratório de Microbiologia. Faculdade de Farmácia da Universidade do Porto, Rua Jorge Viterbo Ferreira nº 228, 4050-313 Porto, Portugal; phone: +351 220428500; e-mail: joaobotelho9@hotmail.com; Luísa Peixe, Laboratório de Microbiologia. Faculdade de Farmácia da Universidade do Porto, Rua Jorge Viterbo Ferreira nº 228, 4050-313 Porto, Portugal; phone: +351 220428500; e-mail: lpeixe@ff.up.pt.

27 **Abstract**

28 Horizontal transfer of plasmids plays a pivotal role in the dissemination of antibiotic
 29 resistance genes and emergence of multidrug-resistant bacteria. Sequencing of plasmids
 30 is thus paramount for the success of accurate epidemiological tracking strategies in the
 31 hospital setting and routine surveillance. Here, we combine Nanopore and Illumina
 32 sequencing to fully assemble a carbapenemase-encoding megaplasmid carried by a
 33 clinical isolate belonging to a putative novel *Pseudomonas* species. FFUP_PS_41 has a
 34 multidrug resistance phenotype and was initially identified as *Pseudomonas putida*, but
 35 an average nucleotide identity below the cut-off for species delineation suggests a new
 36 species related to the *P. putida* phylogenetic group. FFUP_PS_41 harbors a 498,516-bp
 37 untypable megaplasmid (pJBCL41) with low similarity compared with publicly
 38 available plasmids. pJBCL41 contains a full set of genes for self-transmission and genes
 39 predicted to be responsible for plasmid replication, partitioning, maintenance and heavy
 40 metal resistance. pJBCL41 carries a class 1 integron with the *|aacA7|bla_{VIM-2}|aacA4|*
 41 cassette array (In103) located within a defective Tn402-like transposon that forms part
 42 of a 50,273-bp mosaic region bound by 38-bp inverted repeats typical of the Tn3 family
 43 and flanked by 5-bp direct repeats. This region is composed of different elements,
 44 including additional transposon fragments, five insertion sequences and a Tn3-Derived
 45 Inverted-Repeat Miniature Element. The hybrid Nanopore/Illumina approach resulted in
 46 contiguous assemblies and allowed us to fully resolve a carbapenemase-encoding
 47 megaplasmid from *Pseudomonas* spp. The identification of novel megaplasmids will
 48 shed a new light on the evolutionary effects of gene transfer and the selective forces
 49 driving AR.

50 Keywords: *Pseudomonas*, megaplasmids, Nanopore, Illumina, antibiotic
 51 resistance

52 **Introduction**

53 Bacteria can become resistant to antibiotics through chromosomal mutations and by the
54 acquisition of resistance genes carried on mobile genetic elements, including plasmids
55 and integrative and conjugative elements [1]. Plasmids are autonomous self-replicating
56 elements that drive the horizontal transfer (HGT) of antibiotic resistance genes from cell
57 to cell by conjugation [2–5]. The mobility of a plasmid depends on the set of genes that
58 it carries, and these extrachromosomal elements may be conjugative, mobilizable or
59 non-transmissible [2, 3]. Conjugative plasmids carry all the machinery necessary for
60 self-propagation: i) a relaxase, a key protein in conjugation; ii) an origin of transfer
61 (*oriT*); iii) a set of genes encoding for the type-IV secretion system (T4SS); and iv) a
62 gene encoding a type-IV coupling protein (T4CP) [2, 3]. Mobilizable plasmids lack the
63 complete set of genes encoding the T4SS and may use the conjugative apparatus of a
64 helper plasmid present in the cell to be successfully transferred. Conjugative plasmids
65 tend to be low copy number and large, whereas mobilizable plasmids are frequently
66 high copy number and smaller (<30 kb) [2, 3]. The term megaplasms [6] has been
67 used for very large replicons (>350 kb) which, in contrast to chromids [7], do not carry
68 essential core genes. Megaplasms frequently have mosaic structures, carrying genetic
69 modules that originate from different ancestral sources [8]. The formation of mosaic
70 plasmids may be influenced by several factors, such as the abundance of conjugative
71 plasmids and transposons, selection pressures, incompatibility groups and the host's
72 tolerance of foreign DNA. According to the plasmid hypothesis, megaplasms are the
73 evolutionary precursors of chromids, due to the amelioration of genomic signatures to
74 those of the chromosomal partner and the acquisition of essential genes [7].

75 To date, fourteen incompatibility groups (IncP-1 to IncP-14) have been characterized
76 amongst *Pseudomonas* plasmids [9, 10]. Narrow host range plasmids comprise IncP
77 types -2, -5, -7, -10, -12 and -13 and cannot be transferred into *Escherichia coli*. In

78 contrast, other groups display a broad host range, as they are also included in the typing
79 scheme for *Enterobacteriaceae* plasmids: IncP-1 (IncP), IncP-4 (IncQ) and IncP-6
80 (IncG) [9, 10]. Unlike *Enterobacteriaceae* plasmids, no replicon-based PCR typing of
81 *Pseudomonas* plasmids has been created yet.

82 Plasmids may harbor accessory module(s) that provide adaptive advantage(s) for their
83 host, such as virulence-encoding factors and antibiotic resistance genes [9, 11–13].
84 Sequencing of plasmids is thus paramount to the success of accurate epidemiological
85 tracking strategy in the hospital setting and routine surveillance, helping to identify
86 transmission routes and to prevent future outbreaks [14–19]. The advent of WGS has
87 enabled the *in silico* analysis of a wide array of plasmids, most of them from assembly
88 of short-read sequencing data [20–24]. However, fully resolving plasmids with short-
89 read sequencing technologies remains challenging due to the presence of numerous long
90 repeated regions [25], and currently the most accurate approach to assemble these
91 plasmids is to use a combination of short-read and long-read methods [14–19, 26, 27].

92 Here, we combined Nanopore and Illumina sequencing to fully assemble a
93 carbapenemase-encoding megaplasmid carried by an isolate belonging to a putative
94 novel *Pseudomonas* species.

95 **Material and Methods**

96 ***Bacterial Isolate***

97 Isolate FFUP_PS_41 was obtained in 2008 from endotracheal tube secretions of a
98 patient with pneumonia admitted to the Neonatal/Pediatric Intensive Care unit of Centro
99 Hospitalar do Porto - Hospital de Santo António, in Porto, Portugal, as part of regular
100 surveillance of carbapenemase-producers among clinical isolates.

101 FFUP_PS_41 was initially identified as *Pseudomonas putida* by VITEK-2 (bioMérieux)
 102 and later re-classified by pair-wise average nucleotide identity based on BLAST+
 103 (ANiB) using JSpeciesWS v3.0.20 and PyANI v0.2.7
 104 (<https://github.com/widdowquinn/pyani>) [28–30]. Antimicrobial susceptibility testing
 105 was conducted by standard disc diffusion and broth microdilution (for colistin)
 106 methods, according to EUCAST guidelines (<http://www.eucast.org/>).

107 ***Whole-Plasmid Sequencing and Bioinformatics***

108 Genomic DNA from FFUP_PS_41 was extracted using a QIAamp DNA Mini Kit
 109 (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Sequencing
 110 libraries were prepared using Illumina Nextera and the 1D ligation library approach
 111 from Oxford Nanopore Technology (ONT). Libraries were sequenced on the Illumina
 112 HiSeq 2500 sequencer or the MinION sequencer from ONT equipped with a flowcell of
 113 chemistry type R9.4, respectively.

114 Illumina reads were verified for quality using FastQC and Trimmomatic [31, 32], while
 115 MinION reads were processed with ONT's albacore v2.3.0 followed by demultiplexing
 116 using porechop v0.2.3. Both datasets were then combined using the Unicycler assembly
 117 pipeline [33] with a finishing step of Pilon v1.22. The assemblies were visually
 118 inspected using the assembly graph tool Bandage v0.8.1 [34]. Annotation of the
 119 megaplasmid was performed with Prokka v1.13 using default parameters [35]. To
 120 improve annotation, we downloaded additional files of trusted proteins from NCBI
 121 RefSeq plasmids (<ftp://ftp.ncbi.nih.gov/refseq/release/plasmid/>), the NCBI Bacterial
 122 Antimicrobial Resistance Reference Gene Database
 123 (ftp://ftp.ncbi.nlm.nih.gov/pathogen/Antimicrobial_resistance/) and the Antibacterial
 124 Biocide- and Metal-Resistance Genes database (Bac-Met,
 125 <http://bacmet.biomedicine.gu.se/index.html>). EggNOG mapper v4.5.1 and NCBI's

126 Conserved Domain Database CDSEARCH/cdd v3.16 were used for functional
127 annotation and conserved domain search of protein sequences, respectively [36–38].
128 Inference of orthologous groups (OGs) was achieved with OrthoFinder v2.2.6 [39]. The
129 coding sequence (CDS) annotations of the megaplasmid were visualized with Circos
130 v0.69-6 [40]. We used ISfinder [41] to look for insertion sequences (IS). Antimicrobial
131 resistance genes and associated mobile elements were annotated using GalileoTM AMR
132 (<https://galileoamr.arcbio.com/mara/>) (Arc Bio, Cambridge, MA) [42]. Plasmid copy
133 number was estimated based on coverage of the Illumina dataset. GenSkew
134 (<http://genskew.csb.univie.ac.at/>) was used to compute and plot nucleotide skew data to
135 predict the origin of replication.

136 ***Accession Number***

137 The sequence of plasmid pJBCL41 was deposited in GenBank with accession number
138 MK496050.

139 **Results**

140 ***Antimicrobial Susceptibility and Taxonomy testing***

141 Clinical isolate FFUP_PS_41 has a multidrug resistance (MDR) phenotype, showing
142 resistance to imipenem, meropenem, ceftazidime, cefepime, aztreonam,
143 piperacilin+tazobactam, gentamicin, tobramycin, amikacin, ciprofloxacin but remains
144 susceptible to colistin (MIC=1 mg/L). FFUP_PS_41 was initially identified as *P. putida*
145 by VITEK-2. However, it displays an ANI value below the cut-off for species
146 identification (95%) [28] when compared with the complete genome of type strains
147 belonging to the *Pseudomonas* genus, suggesting that it represents a new species related
148 to the *P. putida* phylogenetic group.

149 **Comparative Megaplasmidomics Between pJBCL41 and Related *Pseudomonas*** 150 **Plasmids**

151 Using a hybrid assembly approach, we were able to fully resolve a mosaic megaplasmid
152 (named pJBCL41) carried by *Pseudomonas* sp. FFUP_PS_41 (**Figure S1**). pJBCL41 is
153 498,516 bp and a total of 608 predicted CDS were annotated (**Figure 1**). It has an
154 average GC content of 56.0%, which is lower than that observed for the chromosome
155 (62.6%) and the mean content for strains identified as *P. putida* (62.0%, according to
156 information retrieved on the 08/03/2019 on
157 <https://www.ezbiocloud.net/taxon?tn=Pseudomonas%20putida>).

158 NCBI's CDD calls 42.1% (256) of the predicted CDS for pJBCL41 (**Table S1**),
159 indicating that most genes encode proteins of unknown function. The backbone of this
160 megaplasmid harbours genes predicted to be responsible for plasmid replication, heavy
161 metal resistance and carries two predicted type-II toxin-antitoxin (TA) systems and
162 genes encoding for partition systems (**Figure 1**) [43]. Several genes encoding transport
163 and metabolic processes, as well as transposable elements and CDS associated with
164 transcription, regulatory, chemotaxis signal transduction and mobility functions could
165 be identified. These traits are frequently overrepresented on large plasmids (**Figure 2**)
166 [6, 44]. Also, pJBCL41 harbours several genes coding for the synthesis of DNA
167 precursors, which may promote replication and transcription processes to alleviate the
168 burden that this acquired element may impose on the host cell.

169 pJBCL41 has low nucleotide sequence identity with *Pseudomonas* megaplasmid
170 deposited in public databases (**Table 1** and **Figure S2**). OrthoFinder assigned 59.4% of
171 proteins encoded by pJBCL41 and the most closely-related plasmid, pQBR103 from
172 *Pseudomonas fluorescens* [45] into 335 OGs (**Table S2**). pQBR103 was found in
173 *Pseudomonas* populations colonizing the leaf and root surfaces of sugar beet plants
174 growing at Wytham, United Kingdom and carried no antimicrobial resistance genes

175 [45]. Curiously, a blastp analysis between the proteins encoded by these megaplasמידs
176 revealed that the average amino acid sequence identity is 72.8% among sequences
177 producing significant alignments.

178 Large plasmids identified among the *Pseudomonas* genus usually belong to the IncP-2
179 incompatibility group [10, 21, 24]. However, the IncP-2-type
180 stability/replication/conjugal transfer system is absent from pJBCL41 as previously
181 observed for other megaplasמידs carried by different *Pseudomonas* species [46, 47].
182 Two replication proteins could be identified here. One replicase gene is located at
183 458,679 bp on the plasmid and is close to predicted the origin of replication (**Figure**
184 **S3**). pJBCL41 is estimated to be present as a single copy, from read coverage vs. the
185 chromosome. Like many megaplasמידs, pJBCL41 appears to possess a full set of genes
186 for self-transmission [2, 3]. We identified a cluster of genes encoding an F-type T4SS,
187 encompassing i) a gene encoding a TraD homolog, an AAA+ ATPase of the pfamVirD4
188 type, known as the T4CP and which is a key protein in conjugation; ii) a gene encoding
189 a TraI relaxase homolog, which together with accessory proteins is responsible for
190 cleaving the plasmid in a site-specific manner to initiate DNA transfer and iii) a set of
191 genes (*traEFGKNV* homologues) coding for the mating pair formation system
192 responsible for pilus assembly and retraction (**Figure 1**) [2, 3, 48].

193 ***pJBCL41 Carries a Complex 50 kb Multidrug Resistance Region***

194 The plasmid pJBCL41 carries genes typically found on IncP-2 encoding resistance to
195 tellurite, which could allow co-selection and enrichment of bacteria with MDR plasmids
196 [49]. It also harbours a class 1 integron with the *|aacA7|bla_{VIM-2}|aacA4|* cassette array
197 (named In103 by INTEGRALL [50]) (**Figure 3**): *aacA7* confers resistance to
198 aminoglycosides and *bla_{VIM-2}* encodes resistance to β -lactams (including carbapenems).
199 The *aacA4* gene cassette has a C residue at nucleotide position 329, encoding a serine

200 associated with gentamicin resistance [51]. The same cassette array has been observed
 201 previously among isolates from Portuguese hospitals [22]. The integron is of the In4
 202 type, with a complete 5'-CS bounded by the 25 bp inverted repeat IRi, 2,239 bp of the
 203 3'-CS and IS6100 flanked by two fragments of the IRt end of Tn402 [9, 52]. As the
 204 region between IRi and IRt lacks *tni* transposition genes, this constitutes a Tn402-like
 205 transposon that would be defective in self-transposition.

206 This defective Tn402-like transposon is flanked by 5-bp direct repeats (5'-CTGCT-3')
 207 (**Figure 3**), suggesting integration by transposition close to the predicted resolution
 208 (*res*) site of a Tn3-family transposon. About 300 bp at the IR_L end of the transposon are
 209 related (~86% identical) to TnAs1 (ISfinder), followed by a region containing a gene
 210 which may encode a methyl-accepting chemotaxis protein. From the predicted
 211 recombination crossover point in the *res* site the sequence matches TnPa40 (ISfinder).
 212 This “hybrid” transposon is not flanked by characteristic 5 bp DR but the 5 bp adjacent
 213 to IR_L (5'-AGGTA-3') are repeated 50,273 bp away, immediately adjacent to the 38 bp
 214 repeat of a 1,100 bp transposon fragment ~97% identical to part of both Tn1721
 215 (GenBank accession no. X61367.1, [53]) and TnAs1 (**Figure 3**). This transposon is
 216 truncated by 261 bp region that apparently corresponds to a Tn3-Derived Inverted-
 217 Repeat Miniature Element (designated TIME-262.1 here). TIMEs are non-autonomous
 218 mobile elements commonly found in *Pseudomonas* spp. [54].

219 Most of the region between these transposon elements consists of a 16,782 bp segment
 220 flanked by directly oriented copies of ISPst3 (IS21 family). This region, except for
 221 insertion of ISPa82 (IS66 family) and an adjacent deletion in pJBCL41, matches several
 222 *Pseudomonas* chromosomes (e.g. *P. aeruginosa* PA7 in **Figure S4**) and different parts
 223 of it are found in plasmids in *Pseudomonas*, *Acinetobacter* and *Enterobacteriaceae*,
 224 sometimes also flanked by IS. The sequence between TnPa40 and the left-hand ISPst3

in pJBCL41 is a duplication of part of the 16,782 bp region, with *ISPa1635* (IS4 family) inserted, flanked by characteristic 8 bp DR, instead of *ISPa82* and ends with a partial *ISPa1635*. The right-hand *ISPst3* truncates a transposon related to *TnAs2* [55], which is separated from TIME-261.1 by a 9,075 bp region that also matches *Pseudomonas* chromosomes and includes a putative aminoglycoside phosphotransferase gene. Blast searches with the complete 50 kb region identified a 59 kb region in the chromosome of *P. aeruginosa* AR_0440 (GenBank accession no. CP029148.1) that has similar ends, but lacks an integron, with an additional *Tn5393* insertion and a different region in place of the *ISPst3*-bounded segment (**Figure S4**). This 59 kb region is flanked by 5 bp DR (5'-AATGA-3') and an uninterrupted version of the flanking sequence matches other *Pseudomonas* chromosomes. A *Tn5503*-like transposon encoding a type-II TA system and two metal dependent phosphohydrolases is also inserted in pJBCL41 [56] and is flanked by 5-bp DR (5'-ACTCT-3'), indicating that this element transposed independently of the 50-kb region (**Figure 3**). It has only 10 nucleotide differences from the original *Tn5503* on plasmid Rms149, the archetype of *Pseudomonas* plasmid incompatibility group IncP-6 [56], and additional copies of short repeats in a GC-rich region within a gene encoding an ATP-utilizing enzyme. An additional *ISPst3*, five *ISPpu7* (IS21 family) and one *ISPa41* (IS5 family) all flanked by DR of characteristic length, are also inserted in the pJBCL41 backbone (**Figures 1 and 3**).

Discussion

In this study, we took advantage of a hybrid assembly approach to fully resolve and characterize a carbapenemase-encoding megaplasmid and to compare it with related *Pseudomonas* megaplasms. The lower GC content of pJBCL41 compared with the FFUP_PS_41 chromosome and strains identified as *P. putida* may be related to a more

250 relaxed selection acting on these secondary replicons, as the maintenance of GC-rich
 251 genomes is energetically more demanding [57, 58]. Ongoing studies will help to
 252 characterize the biology and genomic signatures related to this new putative
 253 *Pseudomonas* species (Botelho *et al*, unpublished data).

254 Since secondary replicons are under strong pressure to undergo genomic reshuffling
 255 [57], the observed low nucleotide sequence identity between pJBCL41 megaplasms
 256 and large *Pseudomonas* plasmids deposited in public databases might be expected. Even
 257 though pJBCL41 and pQBR103 plasmids are similar in size and functionalities, there is
 258 a high level of divergence between genes encoding related proteins. Indeed, it is rare to
 259 identify megaplasms with a similar nucleotide sequence in strains belonging to
 260 different species within the same genus [6, 47]. These results suggest that pJBCL41 and
 261 pQBR103 may share a common ancestor, but independent evolutionary trajectories
 262 have led to significant diversification among related genes. The presence of different
 263 replicons suggests that pJBCL41 may have resulted from co-integration of distinct
 264 plasmid modules. The replication module defines plasmid copy number and plasmid
 265 survival in several hosts. Low copy-number plasmids are more frequently lost, due to
 266 random assortment at cell division [2, 3] and extra stability modules, such as TA and
 267 partition systems, may be required to ensure that large plasmids such as pJBCL41 are
 268 maintained [43, 59].

269 The DR flanking the 50-kb region in pJBCL41 and the related 59-kb region in the *P.*
 270 *aeruginosa* AR_0440 chromosome could reflect insertion of each region by
 271 transposition, possibly mediated by the intact transposase and resolvase of TnPa40.

272 However, the size, complexity and differences the internal parts of these related regions
 273 may be more consistent with initial insertion of a simple transposon followed by further
 274 insertions, deletions and rearrangements. A similar situation is seen in plasmid pCTX-

275 M360, which carries a complete Tn2 flanked by the 5 bp DR, and the highly-related
276 pCTX-M3, in which the ends of Tn2 are present in the same position but the central part
277 of the transposon has undergone extensive rearrangements [60]. The identification of all
278 or part of the 16,782 bp segment found within the 50 kb region in pJBCL41 in other
279 locations also suggests that some of the genes it carries may encode advantageous
280 functions, but this needs further analysis. Identification of other sequences related to
281 parts of these 50-kb and 59-kb region segments may also shed light on how they have
282 arisen and evolved.

283 To sum up, we show that a hybrid Nanopore/Illumina approach is useful for producing
284 contiguous assemblies and allowed full resolution of a carbapenemase-encoding
285 *Pseudomonas* megaplasmid. The presence of this large plasmid may provide a selective
286 advantage to the host cell. However, given their size and gene content, acquisition of
287 these secondary replicons may pose a significant cost [61–63]. The high level of gene
288 variation when compared to publicly available megaplasms suggests that these
289 secondary replicons frequently undergo gene loss and gain through HGT. The reduced
290 purifying selection and the high prevalence of transposable elements frequently
291 observed on megaplasms may help to explain why these elements readily acquire
292 foreign DNA [6, 57, 64]. In fact, mosaic plasmids such as pJBCL41 and the majority of
293 megaplasms have a high proportion of mobile genetic elements [8]. The identification
294 of novel megaplasms may shed light on the evolutionary effects of gene transfer and
295 the selective forces driving antibiotic resistance.

296 **Acknowledgments**

297 This work was supported by the Applied Molecular Biosciences Unit- UCIBIO which is
298 financed by national funds from FCT/MCTES (UID/Multi/04378/2019). JB and FG were
299 supported by grants from Fundação para a Ciência e a Tecnologia (SFRH/BD/104095/2014 and

300 SFRH/BPD/95556/2013, respectively). CL is supported by an SB PhD fellowship from FWO
301 Vlaanderen (1S64718N).

302 **Conflict of interest**

303 SRP is responsible for updating the GalileoTM AMR database for Arc Bio.

304

305 **References**

- 306 1. Alekshun MN, Levy SB. Molecular Mechanisms of Antibacterial Multidrug
307 Resistance. *Cell* 2007; **128**: 1037–1050.
- 308 2. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility
309 of plasmids. *Microbiol Mol Biol Rev* 2010; **74**: 434–52.
- 310 3. Garcillán-Barcia MP, Alvarado A, de la Cruz F. Identification of bacterial
311 plasmids based on mobility and plasmid population biology. *FEMS Microbiol*
312 *Rev* 2011; **35**: 936–956.
- 313 4. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids:
314 classification and identification based on replication and transfer systems and
315 host taxonomy. *Front Microbiol* 2015; **6**: 242.
- 316 5. Orlek A, Stoesser N, Anjum MF, Doumith M, Ellington MJ, Peto T, et al.
317 Plasmid Classification in an Era of Whole-Genome Sequencing: Application in
318 Studies of Antibiotic Resistance Epidemiology. *Front Microbiol* 2017; **8**: 182.
- 319 6. diCenzo GC, Finan TM. The Divided Bacterial Genome: Structure, Function, and
320 Evolution. *Microbiol Mol Biol Rev* 2017; **81**: e00019-17.
- 321 7. Harrison PW, Lower RPJ, Kim NKD, Young JPW. Introducing the bacterial

- 322 ‘chromid’: not a chromosome, not a plasmid. *Trends Microbiol* 2010; **18**: 141–
323 148.
- 324 8. Pesesky MW, Tilley R, Beck DAC. Mosaic Plasmids are Abundant and
325 Unevenly Distributed Across Prokaryotic Taxa. *bioRxiv* 2018; 428219.
- 326 9. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements
327 Associated with Antimicrobial Resistance. *Clin Microbiol Rev* 2018; **31**: e00088-
328 17.
- 329 10. Thomas CM, Haines AS. Plasmids of the Genus *Pseudomonas*. *Pseudomonas*.
330 2004. Springer US, Boston, MA, pp 197–231.
- 331 11. San Millan A. Evolution of Plasmid-Mediated Antibiotic Resistance in the
332 Clinical Context. *Trends Microbiol* 2018; **0**.
- 333 12. Diene SM, Rolain J-M. Carbapenemase genes and genetic platforms in Gram-
334 negative bacilli: Enterobacteriaceae, *Pseudomonas* and *Acinetobacter* species.
335 *Clin Microbiol Infect* 2014; **20**: 831–838.
- 336 13. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B,
337 Guerra B, et al. Plasmids carrying antimicrobial resistance genes in
338 Enterobacteriaceae. *J Antimicrob Chemother* 2018; **73**: 1121–1137.
- 339 14. Greig DR, Dallman TJ, Hopkins KL, Jenkins C. MinION nanopore sequencing
340 identifies the position and structure of bacterial antibiotic resistance determinants
341 in a multidrug-resistant strain of enteroaggregative *Escherichia coli*. *Microb*
342 *Genomics* 2018.
- 343 15. Phan HTT, Stoesser N, Maciucă IE, Toma F, Szekely E, Flonta M, et al. Illumina

- 344 short-read and MinION long-read WGS to characterize the molecular
345 epidemiology of an NDM-1 *Serratia marcescens* outbreak in Romania. *J*
346 *Antimicrob Chemother* 2018; **73**: 672–679.
- 347 16. Dong N, Lin D, Zhang R, Chan EW-C, Chen S. Carriage of blaKPC-2 by a
348 virulence plasmid in hypervirulent *Klebsiella pneumoniae*. *J Antimicrob*
349 *Chemother* 2018.
- 350 17. Ludden C, Reuter S, Judge K, Gouliouris T, Blane B, Coll F, et al. Sharing of
351 carbapenemase-encoding plasmids between Enterobacteriaceae in UK sewage
352 uncovered by MinION sequencing. *Microb Genomics* 2017; **3**: e000114.
- 353 18. George S, Pankhurst L, Hubbard A, Votintseva A, Stoesser N, Sheppard AE, et
354 al. Resolving plasmid structures in Enterobacteriaceae using the MinION
355 nanopore sequencer: assessment of MinION and MinION/Illumina hybrid data
356 assembly approaches. *Microb Genomics* 2017; **3**: e000118.
- 357 19. Lemon JK, Khil PP, Frank KM, Dekker JP. Rapid Nanopore Sequencing of
358 Plasmids and Resistance Gene Detection in Clinical Isolates. *J Clin Microbiol*
359 2017; **55**: 3530–3543.
- 360 20. Botelho J, Grosso F, Peixe L. Characterization of the pJB12 Plasmid from
361 *Pseudomonas aeruginosa* Reveals Tn6352, a Novel Putative Transposon
362 Associated with Mobilization of the blaVIM-2-Harboring In58 Integron.
363 *Antimicrob Agents Chemother* 2017; **61**: e02532-16.
- 364 21. Botelho J, Grosso F, Quinteira S, Mabrouk A, Peixe L. The complete nucleotide
365 sequence of an IncP-2 megaplasmid unveils a mosaic architecture comprising a
366 putative novel blaVIM-2-harboursing transposon in *Pseudomonas aeruginosa*. *J*

- 367 *Antimicrob Chemother* 2017; **72**: 2225–2229.
- 368 22. Botelho J, Grosso F, Quinteira S, Brilhante M, Ramos H, Peixe L. Two decades
369 of blaVIM-2-producing *Pseudomonas aeruginosa* dissemination: an interplay
370 between mobile genetic elements and successful clones. *J Antimicrob Chemother*
371 2018; **73**: 873–882.
- 372 23. Bonnin RA, Poirel L, Nordmann P, Eikmeyer FG, Wibberg D, Pühler A, et al.
373 Complete sequence of broad-host-range plasmid pNOR-2000 harbouring the
374 metallo- β -lactamase gene blaVIM-2 from *Pseudomonas aeruginosa*. *J Antimicrob*
375 *Chemother* 2013; **68**: 1060–1065.
- 376 24. Xiong J, Alexander DC, Ma JH, Déraspe M, Low DE, Jamieson FB, et al.
377 Complete sequence of pOZ176, a 500-kilobase IncP-2 plasmid encoding IMP-9-
378 mediated carbapenem resistance, from outbreak isolate *Pseudomonas aeruginosa*
379 96. *Antimicrob Agents Chemother* 2013; **57**: 3775–82.
- 380 25. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. On the
381 (im)possibility of reconstructing plasmids from whole-genome short-read
382 sequencing data. *Microb Genomics* 2017; **3**: e000128.
- 383 26. T'Syen J, Raes B, Horemans B, Tassoni R, Leroy B, Lood C, et al. Catabolism of
384 the groundwater micropollutant 2,6-dichlorobenzamide beyond 2,6-
385 dichlorobenzoate is plasmid encoded in *Aminobacter* sp. MSH1. *Appl Microbiol*
386 *Biotechnol* 2018; **102**: 7963–7979.
- 387 27. Albers P, Lood C, Öztürk B, Horemans B, Lavigne R, van Noort V, et al.
388 Catabolic task division between two near-isogenic subpopulations co-existing in
389 a herbicide-degrading bacterial consortium: consequences for the interspecies

- 390 consortium metabolic model. *Environ Microbiol* 2018; **20**: 85–96.
- 391 28. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K,
392 Kyrpides NC, et al. Microbial species delineation using whole genome
393 sequences. *Nucleic Acids Res* 2015; **43**: 6761–71.
- 394 29. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web
395 server for prokaryotic species circumscription based on pairwise genome
396 comparison. *Bioinformatics* 2016; **32**: 929–931.
- 397 30. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and
398 taxonomy in diagnostics for food security: soft-rotting enterobacterial plant
399 pathogens. *Anal Methods* 2016; **8**: 12–24.
- 400 31. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data.
401 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> .
- 402 32. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
403 sequence data. *Bioinformatics* 2014; **30**: 2114–20.
- 404 33. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome
405 assemblies from short and long sequencing reads. *PLOS Comput Biol* 2017; **13**:
406 e1005595.
- 407 34. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of
408 *de novo* genome assemblies: Fig. 1. *Bioinformatics* 2015; **31**: 3350–3352.
- 409 35. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;
410 **30**: 2068–2069.

- 411 36. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et
412 al. CDD: NCBI's conserved domain database. *Nucleic Acids Res* 2015; **43**:
413 D222–D226.
- 414 37. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al.
415 CDD/SPARCLE: functional classification of proteins via subfamily domain
416 architectures. *Nucleic Acids Res* 2017; **45**: D200–D203.
- 417 38. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al.
418 eggNOG 4.5: a hierarchical orthology framework with improved functional
419 annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*
420 2016; **44**: D286–D293.
- 421 39. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome
422 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*
423 2015; **16**: 157.
- 424 40. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al.
425 Circos: an information aesthetic for comparative genomics. *Genome Res* 2009;
426 **19**: 1639–45.
- 427 41. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the
428 reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006; **34**:
429 D32–D36.
- 430 42. Partridge SR, Tsafnat G. Automated annotation of mobile antibiotic resistance in
431 Gram-negative bacteria: the Multiple Antibiotic Resistance Annotator (MARA)
432 and database. *J Antimicrob Chemother* 2018; **73**: 883–890.

- 433 43. Díaz-Orejas R, Espinosa M, Yeo CC. The Importance of the Expendable: Toxin–
434 Antitoxin Genes in Plasmids and Chromosomes. *Front Microbiol* 2017; **8**: 1479.
- 435 44. Janssen PJ, Van Houdt R, Moors H, Monsieurs P, Morin N, Michaux A, et al.
436 The Complete Genome Sequence of *Cupriavidus metallidurans* Strain CH34, a
437 Master Survivalist in Harsh and Anthropogenic Environments. *PLoS One* 2010;
438 **5**: e10433.
- 439 45. Tett A, Spiers AJ, Crossman LC, Ager D, Ciric L, Dow JM, et al. Sequence-
440 based analysis of pQBR103; a representative of a unique, transfer-proficient
441 mega plasmid resident in the microbial community of sugar beet. *ISME J* 2007;
442 **1**: 331–340.
- 443 46. Sun F, Zhou D, Wang Q, Feng J, Feng W, Luo W, et al. Genetic characterization
444 of a novel *bla*_{DIM-2} -carrying megaplasmid p12969-DIM from clinical
445 *Pseudomonas putida*. *J Antimicrob Chemother* 2016; **71**: 909–912.
- 446 47. Smith BA, Leligdon C, Baltrus DA. Just the Two of Us? A Family of
447 *Pseudomonas* Megaplasms Offers a Rare Glimpse Into the Evolution of 2
448 Large Mobile Elements.
- 449 48. Thomson NR, Cerdeño-Tárraga AM, Brown CJ, Frost LS. Annotation of plasmid
450 genes. *Plasmid* 2017; **91**: 61–67.
- 451 49. Gullberg E, Albrecht LM, Karlsson C, Sandegren L, Andersson DI. Selection of
452 a multidrug resistance plasmid by sublethal levels of antibiotics and heavy
453 metals. *MBio* 2014; **5**: e01918-14.
- 454 50. Moura A, Soares M, Pereira C, Leitao N, Henriques I, Correia A. INTEGRALL:

- 455 a database and search engine for integrons, integrases and gene cassettes.
456 *Bioinformatics* 2009; **25**: 1096–1098.
- 457 51. Tsafnat G, Copt J, Partridge SR. RAC: Repository of Antibiotic resistance
458 Cassettes. *Database (Oxford)* 2011; **2011**: bar054.
- 459 52. Partridge SR, Brown HJ, Stokes HW, Hall RM. Transposons Tn1696 and Tn21
460 and Their Integrons In4 and In2 Have Independent Origins. *Antimicrob Agents*
461 *Chemother* 2001; **45**: 1263–1270.
- 462 53. Allmeier H, Cresnar B, Greck M, Schmitt R. Complete nucleotide sequence of
463 Tn1721: gene organization and a novel gene product with features of a
464 chemotaxis protein. *Gene* 1992; **111**: 11–20.
- 465 54. Szuplewska M, Czarnecki J, Bartosik D. Autonomous and non-autonomous Tn 3
466 -family transposons and their role in the evolution of mobile genetic elements.
467 *Mob Genet Elements* 2014; **4**: 1–4.
- 468 55. Pfeiffer F, Zamora-Lagos M-A, Blettinger M, Yeroslaviz A, Dahl A, Gruber S, et
469 al. The complete and fully assembled genome sequence of *Aeromonas*
470 *salmonicida* subsp. *pectinolytica* and its comparative analysis with other
471 *Aeromonas* species: investigation of the mobilome in environmental and
472 pathogenic strains. *BMC Genomics* 2018; **19**: 20.
- 473 56. Haines AS, Jones K, Cheung M, Thomas CM. The IncP-6 plasmid Rms149
474 consists of a small mobilizable backbone with multiple large insertions. *J*
475 *Bacteriol* 2005; **187**: 4728–38.
- 476 57. Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ. Why Genes Evolve Faster on

- 477 Secondary Chromosomes in Bacteria. *PLoS Comput Biol* 2010; **6**: e1000732.
- 478 58. Rocha EPC, Danchin A. Base composition bias might result from competition for
479 metabolic resources. *Trends Genet* 2002; **18**: 291–294.
- 480 59. Baxter JC, Funnell BE. Plasmid Partition Mechanisms. *Microbiol Spectr* 2014; **2**.
- 481 60. Zhu W, Luo L, Wang J, Zhuang X, Zhong L, Liao K, et al. Complete nucleotide
482 sequence of pCTX-M360, an intermediate plasmid between pEL60 and pCTX-
483 M3, from a multidrug-resistant *Klebsiella pneumoniae* strain isolated in China.
484 *Antimicrob Agents Chemother* 2009; **53**: 5291–3.
- 485 61. Romanchuk A, Jones CD, Karkare K, Moore A, Smith BA, Jones C, et al. Bigger
486 is not always better: Transmission and fitness burden of ~1 MB *Pseudomonas*
487 *syringae* megaplasmid pMPPla107. *Plasmid* 2014; **73**: 16–25.
- 488 62. Dougherty K, Smith BA, Moore AF, Maitland S, Fanger C, Murillo R, et al.
489 Multiple Phenotypic Changes Associated with Large-Scale Horizontal Gene
490 Transfer. *PLoS One* 2014; **9**: e102170.
- 491 63. San Millan A, MacLean RC. Fitness Costs of Plasmids: a Limit to Plasmid
492 Transmission. *Microbiol Spectr* 2017; **5**.
- 493 64. Rankin DJ, Rocha EPC, Brown SP. What traits are carried on mobile genetic
494 elements and why? *Heredity (Edinb)* 2011; **106**: 1–10.

495

496

497

498

499

500

501 **Table 1.** Blastn results between pJBCL41 and related megaplasmsids.

Plasmid identifier	Max score	Total score	Query cover	E value	Ident (%)	Species	Genbank accession	Size (bp)	Year of isolation	Source	Country
pQBR103	11073	1.06E+05	44%	0	72.99%	<i>P. fluorescens</i>	NC_009444.1	425094	2008	Environment	United Kingdom
XWY-1	10059	1.22E+05	9%	0	99.59%	<i>P. sp.</i>	NZ_CP026333.1	394537	2016	Environment	China
pJB37	7285	43666	17%	0	99.90%	<i>P. aeruginosa</i>	KY494864.1	464804	2008	Clinical	Portugal
pSY153	6131	1.41E+05	20%	0	99.85%	<i>P. putida</i>	KY883660.1	468170	2012	Clinical	China
pOZ176	6129	77698	19%	0	99.71%	<i>P. aeruginosa</i>	KC543497.1	500839	2000	Clinical	China
pBM413	5306	51713	17%	0	99.93%	<i>P. aeruginosa</i>	CP016215.1	423017	2012	Clinical	China
RW109	4728	41531	17%	0	71.75%	<i>P. aeruginosa</i>	NZ_LT969519.1	555265	NA	Industrial	NA
P19E3	4715	28494	15%	0	71.79%	<i>P. koreensis</i>	NZ_CP027478.1	467568	2014	Environment	Switzerland
AR439	4697	39223	16%	0	71.79%	<i>P. aeruginosa</i>	NZ_CP029096.1	437392	NA	Clinical	NA

NA stands for no available data

Figure 1. Circular representation of genomic features of pJBCL41. The innermost circle is a histogram of the GC skew, the next a graph of GC content. The next circle displays selected regions of interest (yellow) and IS and transposons or related elements (grey). Red dots highlight genes encoding for antibiotic resistance. The next two circles represent the coding regions on the negative and positive strands colored by their functional annotation (when available). The outermost circle displays regions with high levels of identity to pQBR103 (GenBank accession no. NC_009444.1).

Figure 2. Functional characterization of pJBCL41 and related megaplasms. COG stands for Cluster of Orthologous Groups.

Figure 3. Map of resistance genes and mobile genetic elements inserted in the backbone of pJBLC41. Gene cassettes are shown as blue boxes labelled with the cassette name and are oriented in the 5'-CS to 3'-CS direction. IS are shown as block arrows labelled with the IS name/number, with the pointed end corresponding to IR_R. TIME-261.1 and fragments of Tn3-family transposons are shown as beige boxes with 38 bp IR represented by flags. The fragment annotated as “TnAs1-like” is ~97% identical to a region in common between Tn1721 (GenBank accession no. X61367.1) and TnAs1 in ISfinder. The fragment annotated as “TnAs2-like” is ~94% identical to TnAs2 in ISfinder. The integron is inserted in a proposed hybrid transposon, apparently created by *res*-mediated recombination between a *tnp* region matching TnPa40 and another transposon, labelled “Tn”, that is ~86% identical to TnAs1 over the ~300 bp at the IR_L end only. Direct repeats are shown as a pair of ‘lollipops’ of the same colour flanking an IS or a pair of IRs (but note that the same colour may be used to indicate more than one pair of DR), with sequences indicated for DR of transposons. Mobile elements are

527 shown to scale and numbers below dashed red lines indicate the lengths of intervening
528 regions in bp. This figure was constructed from diagrams generated using GalileoTM
529 AMR.

■ 75-80% identity

■ 70-75% identity

- Replication and repair
- Coenzyme metabolism
- Intracellular trafficking and secretion
- Signal Transduction
- Cell motility
- Carbohydrate metabolism and transport
- Amino Acid metabolism and transport
- Transcription
- PTM, protein turnover, chaperone functions
- Energy production and conversion
- Cell wall/membrane/envelop biogenesis
- Unclassified

GC skew GC%





