Megaplasmids 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. Partridge[4], Vera van Noort[2,5], Rob Lavigne[3], Filipa Grosso[1], Luísa Peixe[1*]

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

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

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

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

[5]Institute 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.

Abstract

Horizontal transfer of plasmids plays a pivotal role in the dissemination of antibiotic resistance genes and emergence of multidrug-resistant bacteria. Sequencing of plasmids is thus paramount for the success of accurate epidemiological tracking strategies in the hospital setting and routine surveillance. Here, we combine Nanopore and Illumina sequencing to fully assemble a carbapenemase-encoding megaplasmid carried by a clinical isolate belonging to a putative novel *Pseudomonas* species. FFUP_PS_41 has a multidrug resistance phenotype and was initially identified as *Pseudomonas putida*, but an average nucleotide identity below the cut-off for species delineation suggests a new species related to the *P. putida* phylogenetic group. FFUP_PS_41 harbors a 498,516-bp untypable megaplasmid (pJBCL41) with low similarity compared with publicly available plasmids. pJBCL41 contains a full set of genes for self-transmission and genes predicted to be responsible for plasmid replication, partitioning, maintenance and heavy metal resistance. pJBCL41 carries a class 1 integron with the |*aacA7*|*bla*$_{\text{VIM-2}}$|*aacA4*| cassette array (In103) located within a defective Tn*402*-like transposon that forms part of a 50,273-bp mosaic region bound by 38-bp inverted repeats typical of the Tn*3* family and flanked by 5-bp direct repeats. This region is composed of different elements, including additional transposon fragments, five insertion sequences and a Tn*3*-Derived Inverted-Repeat Miniature Element. The hybrid Nanopore/Illumina approach resulted in contiguous assemblies and allowed us to fully resolve a carbapenemase-encoding megaplasmid from *Pseudomonas* spp. The identification of novel megaplasmids will shed a new light on the evolutionary effects of gene transfer and the selective forces driving AR.

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

Introduction

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

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

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

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

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

Material and Methods

Bacterial Isolate

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

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

Whole-Plasmid Sequencing and Bioinformatics

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

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

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

Accession Number

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

Results

Antimicrobial Susceptibility and Taxonomy testing

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

Comparative Megaplasmidomics Between pJBCL41 and Related Pseudomonas Plasmids

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

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

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

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

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

pJBCL41 Carries a Complex 50 kb Multidrug Resistance Region

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

associated with gentamicin resistance [51]. The same cassette array has been observed previously among isolates from Portuguese hospitals [22]. The integron is of the In4 type, with a complete 5'-CS bounded by the 25 bp inverted repeat IRi, 2,239 bp of the 3'-CS and IS*6100* flanked by two fragments of the IRt end of Tn*402* [9, 52]. As the region between IRi and IRt lacks *tni* transposition genes, this constitutes a Tn*402*-like transposon that would be defective in self-transposition.

This defective Tn*402*-like transposon is flanked by 5-bp direct repeats (5´-CTGCT-3´) (**Figure 3**), suggesting integration by transposition close to the predicted resolution (*res*) site of a Tn*3*-family transposon. About 300 bp at the IR_L end of the transposon are related (~86% identical) to Tn*As1* (ISfinder), followed by a region containing a gene which may encode a methyl-accepting chemotaxis protein. From the predicted recombination crossover point in the *res* site the sequence matches Tn*Pa40* (ISfinder). This "hybrid" transposon is not flanked by characteristic 5 bp DR but the 5 bp adjacent to IR_L (5´-AGGTA-3´) are repeated 50,273 bp away, immediately adjacent to the 38 bp repeat of a 1,100 bp transposon fragment ~97% identical to part of both Tn*1721* (GenBank accession no. X61367.1, [53]) and Tn*As1* (**Figure 3**). This transposon is truncated by 261 bp region that apparently corresponds to a Tn*3*-Derived Inverted-Repeat Miniature Element (designated TIME-262.1 here). TIMEs are non-autonomous mobile elements commonly found in *Pseudomonas* spp. [54].

Most of the region between these transposon elements consists of a 16,782 bp segment flanked by directly oriented copies of IS*Pst3* (IS*21* family). This region, except for insertion of IS*Pa82* (IS*66* family) and an adjacent deletion in pJBCL41, matches several *Pseudomonas* chromosomes (e.g. *P. aeruginosa* PA7 in **Figure S4**) and different parts of it are found in plasmids in *Pseudomonas*, *Acinetobacter* and *Enterobacteriaceae*, sometimes also flanked by IS. The sequence between Tn*Pa40* and the left-hand IS*Pst3*

in pJBCL41 is a duplication of part of the 16,782 bp region, with IS*Pa1635* (IS*4* family) inserted, flanked by characteristic 8 bp DR, instead of IS*Pa82* and ends with a partial IS*Pa1635*. The right-hand IS*Pst3* truncates a transposon related to Tn*As2* [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 Tn*5393* insertion and a different region in place of the IS*Pst3*-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 Tn*5503*-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 Tn*5503* 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 IS*Pst3*, five IS*Ppu7* (IS*21* family) and one IS*Pa41* (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* megaplasmids. 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

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

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

The DR flanking the 50-kb region in pJBCL41 and the related 59-kb region in the *P. aeruginosa* AR_0440 chromosome could reflect insertion of each region by transposition, possibly mediated by the intact transposase and resolvase of Tn*Pa40*.

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

M360, which carries a complete Tn*2* flanked by the 5 bp DR, and the highly-related pCTX-M3, in which the ends of Tn*2* are present in the same position but the central part of the transposon has undergone extensive rearrangements [60]. The identification of all or part of the 16,782 bp segment found within the 50 kb region in pJBCL41 in other locations also suggests that some of the genes it carries may encode advantageous functions, but this needs further analysis. Identification of other sequences related to parts of these 50-kb and 59-kb region segments may also shed light on how they have arisen and evolved.

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

Acknowledgments

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

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

Conflict of interest

SRP is responsible for updating the Galileo™ AMR database for Arc Bio.

References

1. Alekshun MN, Levy SB. Molecular Mechanisms of Antibacterial Multidrug Resistance. *Cell* 2007; **128**: 1037–1050.

2. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of plasmids. *Microbiol Mol Biol Rev* 2010; **74**: 434–52.

3. Garcillán-Barcia MP, Alvarado A, de la Cruz F. Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiol Rev* 2011; **35**: 936–956.

4. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Front Microbiol* 2015; **6**: 242.

5. Orlek A, Stoesser N, Anjum MF, Doumith M, Ellington MJ, Peto T, et al. Plasmid Classification in an Era of Whole-Genome Sequencing: Application in Studies of Antibiotic Resistance Epidemiology. *Front Microbiol* 2017; **8**: 182.

6. diCenzo GC, Finan TM. The Divided Bacterial Genome: Structure, Function, and Evolution. *Microbiol Mol Biol Rev* 2017; **81**: e00019-17.

7. Harrison PW, Lower RPJ, Kim NKD, Young JPW. Introducing the bacterial

'chromid': not a chromosome, not a plasmid. *Trends Microbiol* 2010; **18**: 141–148.

8. Pesesky MW, Tilley R, Beck DAC. Mosaic Plasmids are Abundant and Unevenly Distributed Across Prokaryotic Taxa. *bioRxiv* 2018; 428219.

9. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clin Microbiol Rev* 2018; **31**: e00088-17.

10. Thomas CM, Haines AS. Plasmids of the Genus Pseudomonas. *Pseudomonas*. 2004. Springer US, Boston, MA, pp 197–231.

11. San Millan A. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. *Trends Microbiol* 2018; **0**.

12. Diene SM, Rolain J-M. Carbapenemase genes and genetic platforms in Gram-negative bacilli: Enterobacteriaceae, Pseudomonas and Acinetobacter species. *Clin Microbiol Infect* 2014; **20**: 831–838.

13. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J Antimicrob Chemother* 2018; **73**: 1121–1137.

14. Greig DR, Dallman TJ, Hopkins KL, Jenkins C. MinION nanopore sequencing identifies the position and structure of bacterial antibiotic resistance determinants in a multidrug-resistant strain of enteroaggregative Escherichia coli. *Microb Genomics* 2018.

15. Phan HTT, Stoesser N, Maciuca IE, Toma F, Szekely E, Flonta M, et al. Illumina

short-read and MinION long-read WGS to characterize the molecular epidemiology of an NDM-1 Serratia marcescens outbreak in Romania. *J Antimicrob Chemother* 2018; **73**: 672–679.

16. Dong N, Lin D, Zhang R, Chan EW-C, Chen S. Carriage of blaKPC-2 by a virulence plasmid in hypervirulent Klebsiella pneumoniae. *J Antimicrob Chemother* 2018.

17. Ludden C, Reuter S, Judge K, Gouliouris T, Blane B, Coll F, et al. Sharing of carbapenemase-encoding plasmids between Enterobacteriaceae in UK sewage uncovered by MinION sequencing. *Microb Genomics* 2017; **3**: e000114.

18. George S, Pankhurst L, Hubbard A, Votintseva A, Stoesser N, Sheppard AE, et al. Resolving plasmid structures in Enterobacteriaceae using the MinION nanopore sequencer: assessment of MinION and MinION/Illumina hybrid data assembly approaches. *Microb Genomics* 2017; **3**: e000118.

19. Lemon JK, Khil PP, Frank KM, Dekker JP. Rapid Nanopore Sequencing of Plasmids and Resistance Gene Detection in Clinical Isolates. *J Clin Microbiol* 2017; **55**: 3530–3543.

20. Botelho J, Grosso F, Peixe L. Characterization of the pJB12 Plasmid from Pseudomonas aeruginosa Reveals Tn6352, a Novel Putative Transposon Associated with Mobilization of the blaVIM-2-Harboring In58 Integron. *Antimicrob Agents Chemother* 2017; **61**: e02532-16.

21. Botelho J, Grosso F, Quinteira S, Mabrouk A, Peixe L. The complete nucleotide sequence of an IncP-2 megaplasmid unveils a mosaic architecture comprising a putative novel blaVIM-2-harbouring transposon in Pseudomonas aeruginosa. *J*

Antimicrob Chemother 2017; **72**: 2225–2229.

22. Botelho J, Grosso F, Quinteira S, Brilhante M, Ramos H, Peixe L. Two decades of blaVIM-2-producing Pseudomonas aeruginosa dissemination: an interplay between mobile genetic elements and successful clones. *J Antimicrob Chemother* 2018; **73**: 873–882.

23. Bonnin RA, Poirel L, Nordmann P, Eikmeyer FG, Wibberg D, Pühler A, et al. Complete sequence of broad-host-range plasmid pNOR-2000 harbouring the metallo-β-lactamase gene blaVIM-2 from Pseudomonas aeruginosa. *J Antimicrob Chemother* 2013; **68**: 1060–1065.

24. Xiong J, Alexander DC, Ma JH, Déraspe M, Low DE, Jamieson FB, et al. Complete sequence of pOZ176, a 500-kilobase IncP-2 plasmid encoding IMP-9-mediated carbapenem resistance, from outbreak isolate Pseudomonas aeruginosa 96. *Antimicrob Agents Chemother* 2013; **57**: 3775–82.

25. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb Genomics* 2017; **3**: e000128.

26. T'Syen J, Raes B, Horemans B, Tassoni R, Leroy B, Lood C, et al. Catabolism of the groundwater micropollutant 2,6-dichlorobenzamide beyond 2,6-dichlorobenzoate is plasmid encoded in Aminobacter sp. MSH1. *Appl Microbiol Biotechnol* 2018; **102**: 7963–7979.

27. Albers P, Lood C, Özturk B, Horemans B, Lavigne R, van Noort V, et al. Catabolic task division between two near-isogenic subpopulations co-existing in a herbicide-degrading bacterial consortium: consequences for the interspecies

consortium metabolic model. *Environ Microbiol* 2018; **20**: 85–96.

28. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, et al. Microbial species delineation using whole genome sequences. *Nucleic Acids Res* 2015; **43**: 6761–71.

29. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 2016; **32**: 929–931.

30. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 2016; **8**: 12–24.

31. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/* .

32. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–20.

33. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol* 2017; **13**: e1005595.

34. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies: Fig. 1. *Bioinformatics* 2015; **31**: 3350–3352.

35. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **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* $_{\text{DIM-2}}$ -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 Megaplasmids 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, Copty 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

Secondary Chromosomes in Bacteria. *PLoS Comput Biol* 2010; **6**: e1000732.

58. Rocha EPC, Danchin A. Base composition bias might result from competition for metabolic resources. *Trends Genet* 2002; **18**: 291–294.

59. Baxter JC, Funnell BE. Plasmid Partition Mechanisms. *Microbiol Spectr* 2014; **2**.

60. Zhu W, Luo L, Wang J, Zhuang X, Zhong L, Liao K, et al. Complete nucleotide sequence of pCTX-M360, an intermediate plasmid between pEL60 and pCTX-M3, from a multidrug-resistant Klebsiella pneumoniae strain isolated in China. *Antimicrob Agents Chemother* 2009; **53**: 5291–3.

61. Romanchuk A, Jones CD, Karkare K, Moore A, Smith BA, Jones C, et al. Bigger is not always better: Transmission and fitness burden of ~1 MB Pseudomonas syringae megaplasmid pMPPla107. *Plasmid* 2014; **73**: 16–25.

62. Dougherty K, Smith BA, Moore AF, Maitland S, Fanger C, Murillo R, et al. Multiple Phenotypic Changes Associated with Large-Scale Horizontal Gene Transfer. *PLoS One* 2014; **9**: e102170.

63. San Millan A, MacLean RC. Fitness Costs of Plasmids: a Limit to Plasmid Transmission. *Microbiol Spectr* 2017; **5**.

64. Rankin DJ, Rocha EPC, Brown SP. What traits are carried on mobile genetic elements and why? *Heredity (Edinb)* 2011; **106**: 1–10.

499

500

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

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%	*P. fluorescens*	NC_009444.1	425094	2008	Environment	United Kingdom
XWY-1	10059	1.22E+05	9%	0	99.59%	*P. sp.*	NZ_CP026333.1	394537	2016	Environment	China
pJB37	7285	43666	17%	0	99.90%	*P. aeruginosa*	KY494864.1	464804	2008	Clinical	Portugal
pSY153	6131	1.41E+05	20%	0	99.85%	*P. putida*	KY883660.1	468170	2012	Clinical	China
pOZ176	6129	77698	19%	0	99.71%	*P. aeruginosa*	KC543497.1	500839	2000	Clinical	China
pBM413	5306	51713	17%	0	99.93%	*P. aeruginosa*	CP016215.1	423017	2012	Clinical	China
RW109	4728	41531	17%	0	71.75%	*P. aeruginosa*	NZ_LT969519.1	555265	NA	Industrial	NA
P19E3	4715	28494	15%	0	71.79%	*P. koreensis*	NZ_CP027478.1	467568	2014	Environment	Switzerland
AR439	4697	39223	16%	0	71.79%	*P. aeruginosa*	NZ_CP029096.1	437392	NA	Clinical	NA

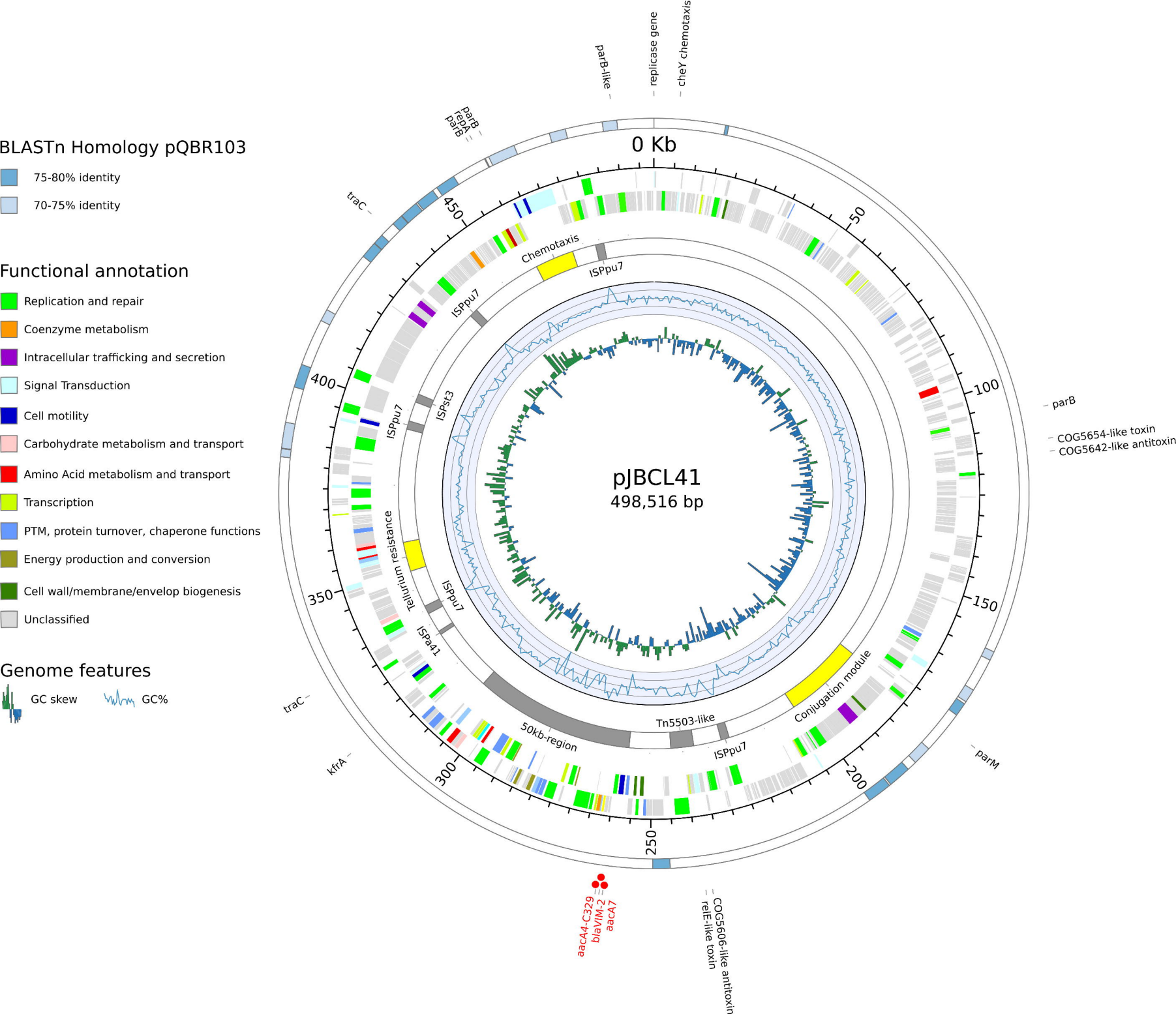
NA stands for no available data

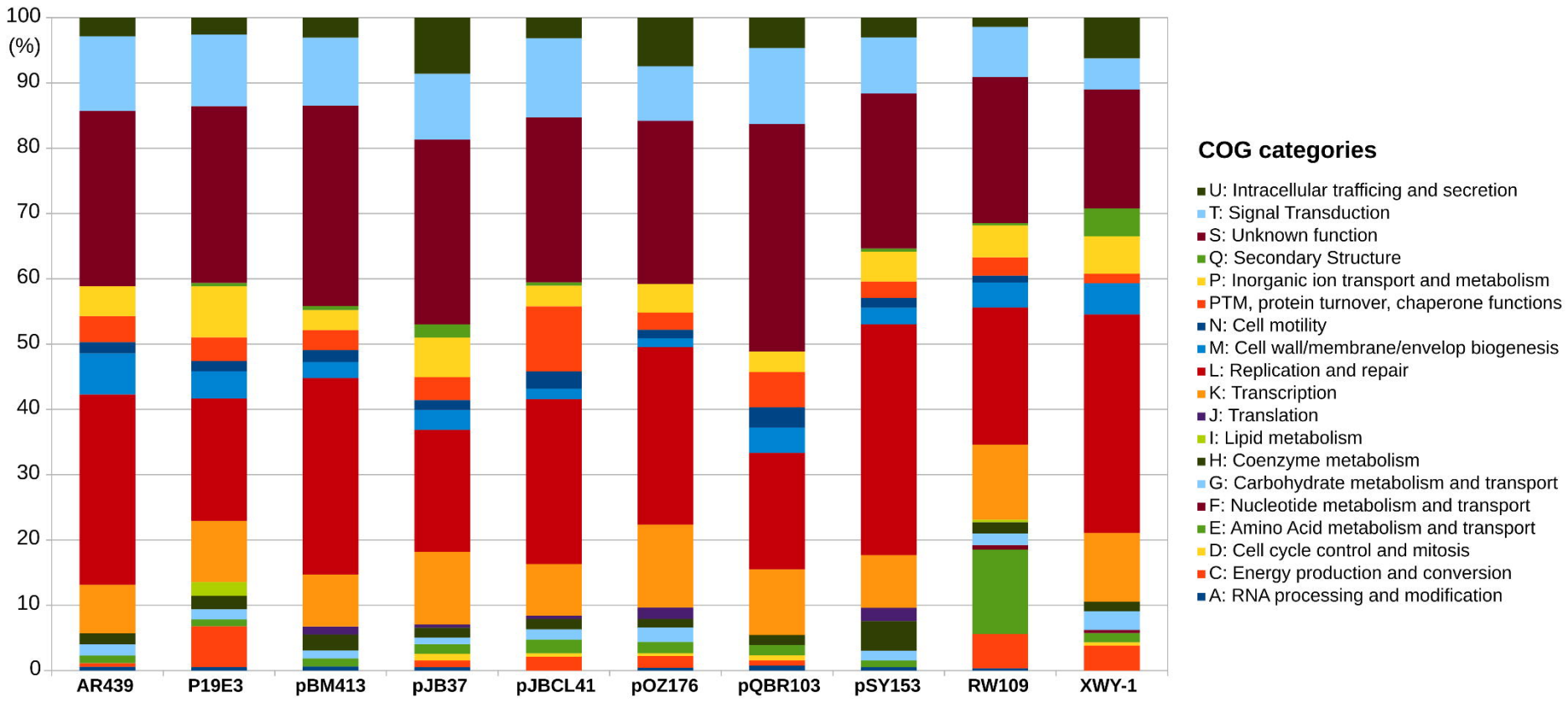
Figure 1. Circular representation of genomic features of pJBCL41. The innermost circle is a histogram of the GC skew, the next a graph of GC content. The next circle displays selected regions of interest (yellow) and IS and transposons or related elements (grey). Red dots highlight genes encoding for 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 megaplasmids. 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 Tn*3*-family transpospons 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 Tn*1721* (GenBank accession no. X61367.1) and Tn*As1* in ISfinder. The fragment annotated as "TnAs2-like" is ~94% identical to Tn*As2* in ISfinder. The integron is inserted in a proposed hybrid transposon, apparently created by *res*-mediated recombination between a *tnp* region matching Tn*Pa40* and another tranpsoson, labelled "Tn", that is ~86% identical to Tn*As1* 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 indictate 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 Galileo™
529 AMR.


BLASTn Homology pQBR103
75-80% identity
70-75% identity
Functional annotation
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
Genome features
GC skew
GC%
pJBCL41
498,516 bp
0 Kb
50
100
150
200
250
300
350
400
450
parB-like
replicase gene
cheY chemotaxis
parB
repA
parB
traC
Chemotaxis
ISPpu7
ISPpu7
ISPst3
ISPpu7
ISPpu7
ISPa41
ISPpu7
Tn5503-like
50kb-region
Conjugation module
ISPpu7
parB
COG5654-like toxin
COG5642-like antitoxin
parM
COG5606-like antitoxin
relE-like toxin
aacA4-C329
blaVIM-2
aacA7
kfrA
traC
Tellurium resistance


100
(%)
90
80
70
60
50
40
30
20
10
0
AR439
P19E3
pBM413
pJB37
pJBCL41
pOZ176
pQBR103
pSY153
RW109
XWY-1
COG categories
U: Intracellular trafficing and secretion
T: Signal Transduction
S: Unknown function
Q: Secondary Structure
P: Inorganic ion transport and metabolism
PTM, protein turnover, chaperone functions
N: Cell motility
M: Cell wall/membrane/envelop biogenesis
L: Replication and repair
K: Transcription
J: Translation
I: Lipid metabolism
H: Coenzyme metabolism
G: Carbohydrate metabolism and transport
F: Nucleotide metabolism and transport
E: Amino Acid metabolism and transport
D: Cell cycle control and mitosis
C: Energy production and conversion
A: RNA processing and modification