Megaplasmids on the rise: combining sequencing approaches to fully resolve a carbapenemase-encoding plasmid in a novel *Pseudomonas* species

João Botelho[1*¥], Cédric Lood[2,3], Sally 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

*joaobotelho9@hotmail.com

¥From 24th april 2019: Schulenburg fellow group
Max-Planck Institute for Evolutionary Biology
Ploen, Germany

INTRODUCTION AND PURPOSE

- 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.
- Fully resolving plasmids using short-read sequencing technologies remains challenging due to the presence of numerous long repeated regions, and currently the most accurate approach to assemble these plasmids is to use a combination of short-read and long-read methods.
- 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.

METHODS

- Isolate FFUP_PS_41 – collected from endotracheal tube secretions of a patient with pneumonia admitted to the Neonatal/Pediatric Intensive Care unit of Hospital de Santo António, in Porto, Portugal in 2008;
- Genomic DNA from FFUP_PS_41 – extracted by QIAamp DNA Mini Kit,
- Illumina – Nextera, HiSeq 2500, FastQC and Trimmomatic;
- Nanopore - 1D ligation library approach from Oxford Nanopore Technology (ONT), MinION sequencer from ONT equipped with a flowcell of chemistry type R9.4, ONT's albacore v2.3.0 followed by demultiplexing using porechop v0.2.3;
- Datasets combined using Unicycler with a finishing step of Pilon v1.22 and assemblies visually inspected using Bandage v0.8.1;
- Megaplasmid – annotation with Prokka v1.13, functional annotation and conserved domain search of protein sequences using EggNOG mapper v4.5.1 and NCBI's Conserved Domain Database CDSEARCH/cdd v3.16, respectively, orthologous groups inferred using OrthoFinder v2.2.6, visualization with Circos v0.69-6, annotation of antimicrobial resistance genes and associated mobile elements using Galileo™ AMR and prediction of the origin of replication using GenSkew.

CONCLUSIONS

- 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 gene transfer mechanisms and the selective forces driving antibiotic resistance.

RESULTS

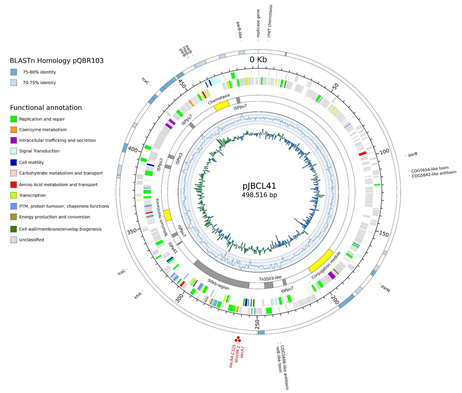

Fig. 1. Circular representation of features of pJBCL41. The innermost circle is a histogram of the GC skew, the next a graph of GC content. The next circle displays selected regions of interest (yellow) and IS and transposons or related elements (grey). The next two circles represent the coding regions 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. Red dots highlight genes coding for antibiotic resistance.

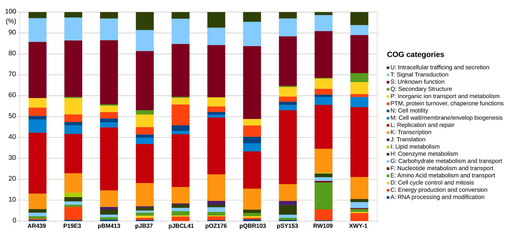

Fig. 2. Functional characterization of pJBCL41 and related megaplasmids. COG stands for Cluster of Orthologous Groups.

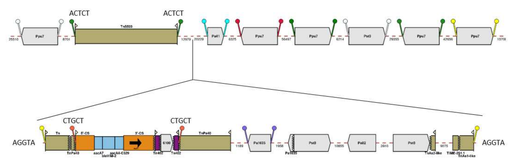

Fig. 3. Map of resistance genes and mobile genetic elements inserted in the backbone of pJBLC41. Gene cassettes are shown as blue boxes labelled with the cassette name and are oriented in the 5'-CS to 3'-CS direction. IS are shown as block arrows labelled with the IS name/number, with the pointed end corresponding to IR_R. TIME-261.1 and fragments of Tn*3*-family transposons are shown as beige boxes with 38 bp IR represented by flags. The fragment annotated as "Tn*As1*-like" is ~97% identical to a region in common between Tn*1721* and Tn*As1* in ISfinder. The fragment annotated as "TnAs2-like" is ~94% identical to Tn*As2*. The integron is inserted in a proposed hybrid transposon, apparently created by *res*-mediated recombination between a *tnp* region matching Tn*Pa40* and another transposon, 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 indicate more than one pair of DR), with sequences indicated for DR of transposons. Mobile elements are shown to scale and numbers below dashed red lines indicate the lengths of intervening regions in bp.

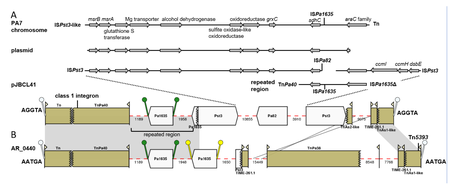

Figure 4. A. Comparison of the IS*Pst3*-flanked region in pJBCL41 with part of the *P. aeruginosa* PA7 chromosome and a representative plasmid (pCAV2018-177). Inserted IS are shown above (all IS*Pa1635* copies have flanking DR, IS*Pa82* does not) and the gap adjacent to IS*Pa82* in pJBCL41 indicates a deletion. The names of IS and Tn defining the ends of the matching regions are shown adjacent to these regions. Predicted gene names/functions of selected orfs are shown (additional hypothetical proteins are not marked). The region between Tn*Pa40* and IS*Pa1635*Δ in pJBCL41 that forms a partial repeat is also shown, **B.** Comparison of the 50 kb insertion in pJBCL41 with the 59 kb insertion in the chromosome of *P. aeruginosa* AR_0440. Diagrams were generated using Galileo™ AMR, with mobile elements shown to scale and numbers below dashed red lines indicating the lengths of intervening regions in bp. Matching regions are indicated by grey shading. The 9,075 bp region between the Tn*As2*-like transposon and TIME-261.1 in pJBCL41 matches positions 9836-762 of the 15,449 bp region between TIME-261.1 and Tn*Pa38* in AR_0440, shown by dotted lines. DR are shown by white 'lollipops' and sequences are indicated. The position of the class 1 integron in pJBCL41 and Tn*5393* in AR_0440, both flanked by 5 bp DR, are indicated by labelled vertical lines.

- pJBCL41 - 498,516-bp long untypable plasmid, 608 predicted coding sequences, 56% GC content, single copy (**Figure 1**);
- Low nucleotide sequence identity with *Pseudomonas* megaplasmids deposited in public databases, but presents traits usually represented in large plasmids: transport and metabolic processes, transposable elements, transcription, synthesis of DNA precursors, regulatory, chemotaxis signal transduction and mobility functions (**Figure 2**);
- Full set of genes responsible for self-transmission (F-type type-IV secretion system);
- Carries a class 1 integron with the |*aacA7*|bla_{VIM-2}|*aacA4*| cassette array (In103) encoding resistance to aminoglycosides and beta-lactams (**Figure 3**);
- In103 is 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 that is flanked by 5-bp direct repeats (**Figure 3**);
- The 50-kb region is composed of different elements, including additional transposon fragments, five insertion sequences (IS) and a Tn*3*-Derived Inverted-Repeat Miniature Element (TIME) (**Figure 3**);
- Most of the region between these transposon elements consists of a 16,782 bp segment flanked by IS*Pst3* that matches several *Pseudomonas* chromosomes (e.g. *P. aeruginosa* PA7 in **Figure 4A**) and different parts of it are found in plasmids in *Pseudomonas*, *Acinetobacter* and *Enterobacteriaceae*;
- The 50 kb region is related to a 59 kb region in the chromosome of *P. aeruginosa* AR_0440. The ends of the two regions are the same, apart from the integron inserted in pJBCL41 and Tn*5393* in AR_0440, but they differ in the central region (**Figure 4B**). The 59 kb region in AR_0440 is flanked by 5-bp DR and an uninterrupted version of the flanking sequence matches other *P. aeruginosa* chromosomes.

ACKNOWLEDGMENTS AND DISCLOSURES

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).

SP is responsible for updating the Galileo™ AMR database for Arc Bio