Progress report

Project title: **Sequencing the genome of a useful bacteria: Paenarthrobacter nicotinovorans – next step in extending it's biotechnological applications**

Contract no: **PCE 152 / 2021**

Proposal code: **PN-III-P4-ID-PCE-2020-0656**

Project Manager: **Marius MIHĂŞAN, Prof. Dr.**

Timeframe: **01 JANUARY - 31 DECEMBER 2022**

Abstract

The complete genome of *Paenarthrobacter nicotinovoran*s ATCC 49199 has been sequenced, annotated, and deposited in GenBank database under the fallowing IDs: CP089293 for chromosome and CP089294 for the pAO1 megaplasmid. RNA-seq experiments that would validate the annotation are underway, with protocols for extraction and direct RNA sequencing using a MinION device being already established. The complete genome can already be used as a useful tool in bacterial taxonomy to re-analyze the *Paenarthrobacter* genus.

Results

Whole Genome Sequencing of any bacterial strains involves 3 different steps: genomic DNA isolation, DNA sequencing and finally genome assembly and functional characterization. At the end of 2021, *Paenarthrobacter nicotinovorans* gDNA was isolated and sequenced using two different technologies and a draft genome was available.

In the beginning of 2022, the final genome was assembled using the hybrid assembler Unicycler v.0.4.9 (Wick et al., 2017). Overlapping sequences at contig ends were removed so that each contig's sequence led directly into its neighbors. The genomes were rotated to start with dnaA on the forward strand, assessed for completeness and contamination using CheckM v.1.0.9 (Parks et al., 2015), and uploaded to the NCBI Prokaryote Genome Annotation Pipeline (PGAP v.5.3) (Tatusova et al., 2016) for automatic annotation. The genome consists of two replicons: a 4 316 184 bp circular chromosome with an overall GC content of 63.2% and the 165 141 bp pAO1 circular megaplasmid with an overall GC content of 59.7% - Figure 1. A total of 4026 genes encoding 3930 proteins, 23 pseudogenes, 54 tRNAs, 2 ncRNAs, 1 tmRNA and 6 identical ribosomal operons were identified on the chromosome. EggNOG (Cantalapiedra et al., 2021) assigned 2421 Gene Ontology (GO) terms to 626 (15%) of the annotated genes and 1334 PFAM protein families to 3338 (83%) of the total proteins.

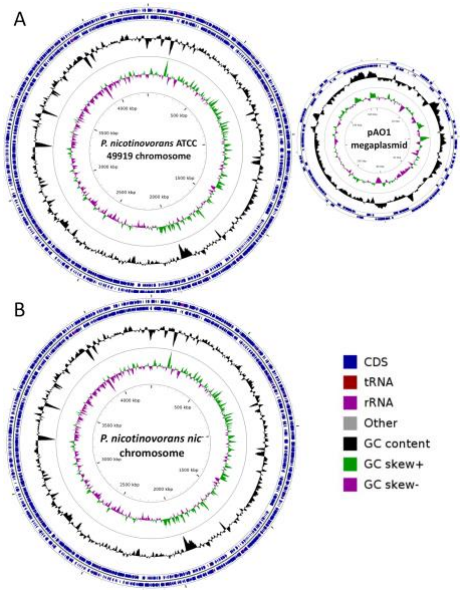

Figure 1. Circular maps of the genomes sequenced: (A) The P. nicotinovorans ATCC 49919 chromosome (left) and its megaplasmid, pAO1 (right); (B) the P. nicotinovorans nic- chromosome.

The maps were generated using Circular Genome Viewer (CGView) v.1.14 (Cantalapiedra et al., 2021).

P. nicotinovorans ATCC 49919 genome as a taxonomic tool

The complete genome of the strain (chromosome and pAO1 plasmid) provides a model for studying molecular evolution of catabolic pathways and their spread by horizontal gene transfer by soil bacterial plasmids. It has been shown that the *nic*-genes DNA fragment from pAO1 plasmid has a lower G+C content than the G+C content of the megaplasmid. The presence of integrases belonging to the tyrosine family of recombinases at the 5′ end of this DNA fragment have suggested that this might actually be a catabolic transposon (Brandsch and Mihasan, 2020) acquired by horizontal gene transfer (Mihasan and Brandsch, 2013).

Although not included in the original activities planned in this project, we have considered that this is a direction that needs to be considered. Hence 65 genomes (complete and draft) from *Paenarthrobacter*, *Arthrobacter*, *Nocardioides* și *Rhodococus* have been downloaded two indicators usefull for species clustering have been calculate: dDDH (Digital DNA-DNA hybridization) using Genome-to-Genome Distance calculator (GGDC) 3.0 (Meier-Kolthoff et al., 2022, 2013) and ANI (Average nucleotide identity) using the OrthoANI algorithm (Lee et al., 2016). This way, we have established that de 4 strains labeled as *Paenarthrobacter* and 5 strains labeled as *Arthrobacter* in the databases can be considered as belonging to the same species – figure 2.

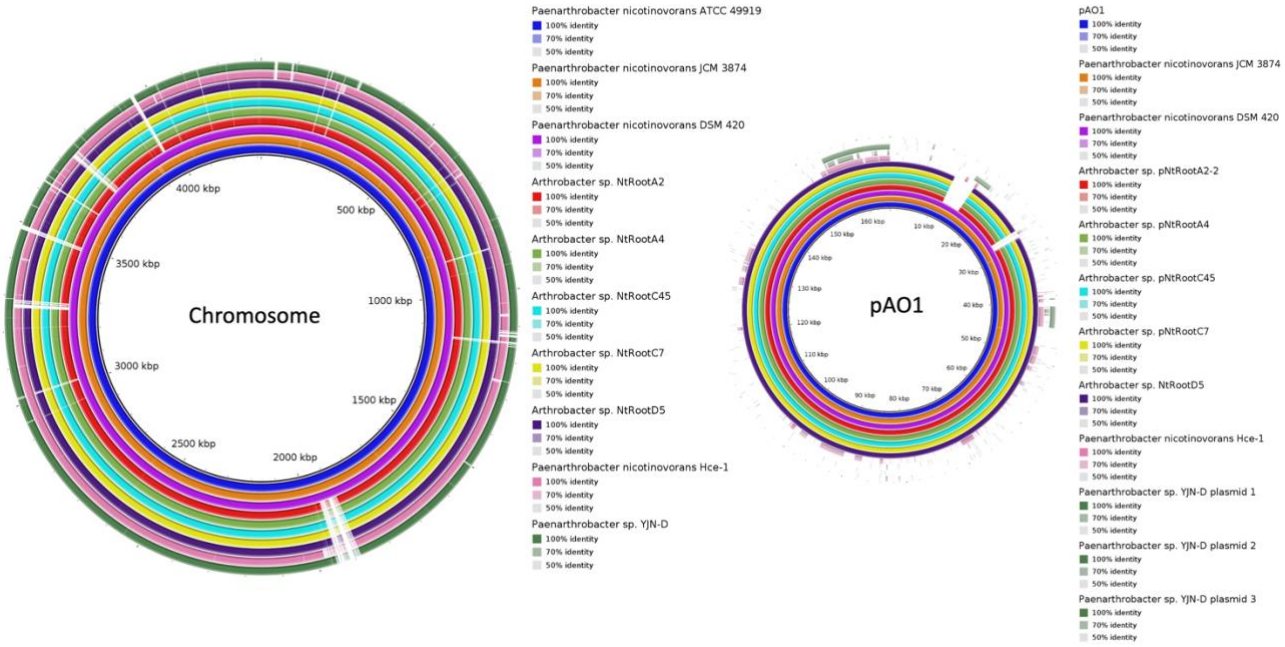

Figura 2. Comparative analysis of highly similar *Paenarthrobacter nicotinovorans* genomes. Right - 4 strains labeled as *Paenarthrobacter* and 5 strains labeled as *Arthrobacter* are almost identical at chromosome level with *Paenarthrobacter nicotinovorans ATCC 49199.* Left - 7 strains also harbor plasmids that are almost identical with pAO1.

Direct RNA-seq experiments on MinION devices

A protocol for preparing total RNA and TEX treated RNA libraries have been established in the lab that produces reasonable results. The libraries have been sequenced producing a usefull amount of data presented (Table 1) with high quality (Figure 3).

Table 1. Comparative data on the direct RNA-seq experiments performed so far in the project.

Seq. Date	Lib. Type	MUX SCAN	RNA quant.	No. active pores	Seq. Time (h)	Reads	data	N50
22_08_25	pA	yes	<400 ng	84		27.25 k	power outage	589
22_08_29	TEX	yes	<400 ng	63	4,2	18.5 k	489.89 MB	579
22_09_06	TEX	yes	952 ng	96	3,48	16.24 k	1010.88 MB	1.55 kb
22_09_07	pA	yes	528 ng	80	10,3	40.25 k	2.31 GB	1.46 kb
22_09_08	pA	no	500 ng	82	24	78.72 k	4.76 GB	1.46 kb
22_09_09	pA	no	200 ng	59	23	32.05 k	2.01 GB	1.56 kb
22_10_13	TEX	yes	160 ng	79	20,48	64.33 k	2.63 GB	1.2 kb
22_10_14	pA	yes	160 ng	81	21,27	66.91 k	4.28 GB	1.58 kb
22_10_20	TEX	yes	584 ng	52	14,57	21.77 k	1.33 GB	1.55 kb
22_10_21	pA	yes	632 ng	84	23,43	59.57 k	3.62 GB	1.49 kb

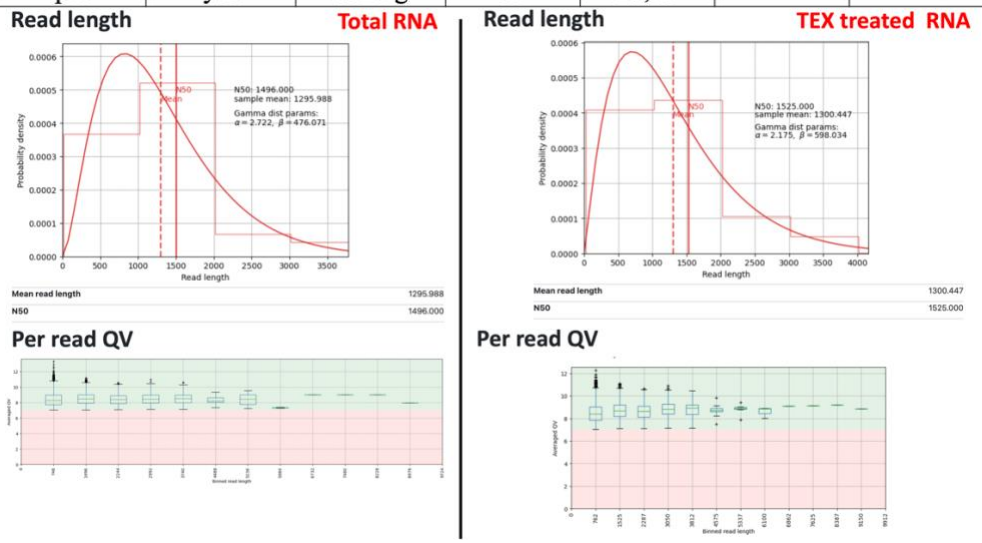

Figure 3. Typical reports of direct RNA-seq experiments generated with LongQC.

The direct RNA-seq data generated so far was mapped directly on the chromosome and pAO1 megaplasmid using Geneious Prime v.2022.2.2 – Figure 4. At the time this report was written, we aim to use the nfcore/nanoseq pipeline to process the RNA long-reads and perform differential gene expression analysis.

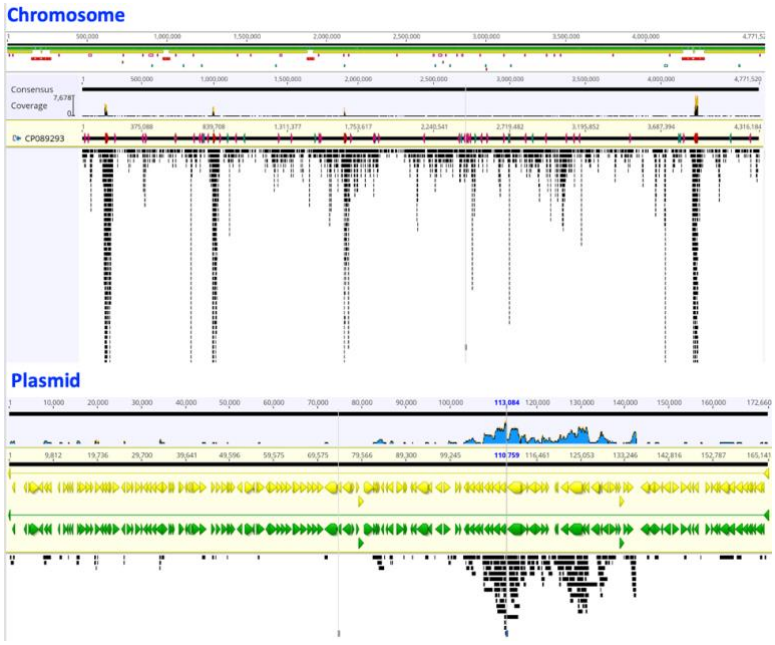

Figura 4. A typical RNA-seq mapping experiment on chromosome (up) and plasmid (down). Mapping is performed with either Geneious Assembler v.2022.2. or BBDuk.

Publications and results disseminations

At this point (24 month after the projects debut) the results were presented at 12 conferences (7 international and 5 national) as 10 lectures and 12 poster presentations. Some results were included in 5 papers in the following journals: ACS Omega, Microbiology Resource Announcements, Plants, Journal of Molecular and Experimental Biology.

News on results, publications, acquisition of novel equipment or key materials were posted in a timely manner on the project webpage (http://cercetare.bio.uaic.ro/grupuri/bioactive/content/grants/PCE2021.html), and group Facebook page: https://www.facebook.com/bioactive.bio.uaic/