Final 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. Direct RNA-seq experiments using nanopores were used to validate the genome and the resulting transcriptomics data has been deposited in GEO under the identifier GSE240220. The annotated genome allowed for a reanalysis of the proteomics data available which were correlated with the transcriptomics data in order to offer a multi-omics view of the nicotine metabolism in *Paenarthrobacter nicotinovoran*s ATCC 49199.

At the end of the project, the results have been presented at 17 conferences (7 international and 8 natioal) as a total of 15 lectures and 21 posters. Also, the results have been published in 3 articles in impact factor journals. Within the frame of the project, a number of 3 BSc and 5 MSc thesis have been defended.

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.

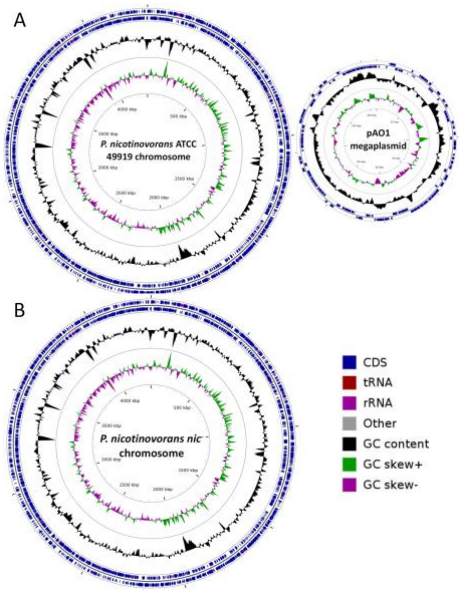


Figure 1. Circular maps of the genomes sequenced: (A) The P. nicotinovorans ATCC 49919 chromosome (left) and its megaplasmid, pAO1 (right); (B) the P. nicotinovorans nic- chromosome. The maps were generated using Circular Genome Viewer (CGView) v.1.14 (Cantalapiedra et al., 2021).

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.

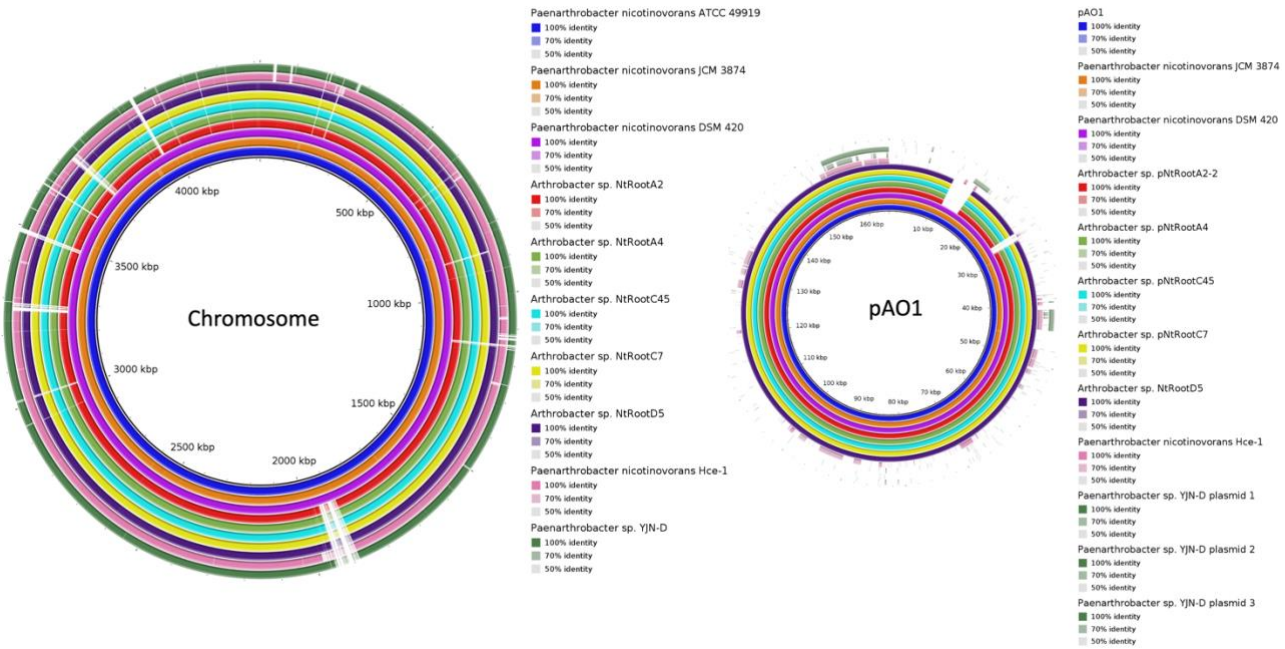


Figura 2. Comparative analysis of highly similar *Paenarthrobacter nicotinovorans* genomes. Right - 4 strains labeled as *Paenarthrobacter* and 5 strains labeled as *Arthrobacter* are almost identical at chromosome level with *Paenarthrobacter nicotinovorans ATCC 49199.* Left - 7 strains also harbor 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. A number of 3 key-points in the development of the culture have been selected as fallows (figure 3):

- 5 hours after inoculation (HAI) start of the exponential phase, nicotine metabolism is active
- 10 HAI in between the exponential and the stationary phase, nicotine concentration is low;
- 24 HAI stationary phase, nicotine is depleted, nicotine metabolism is inactive.

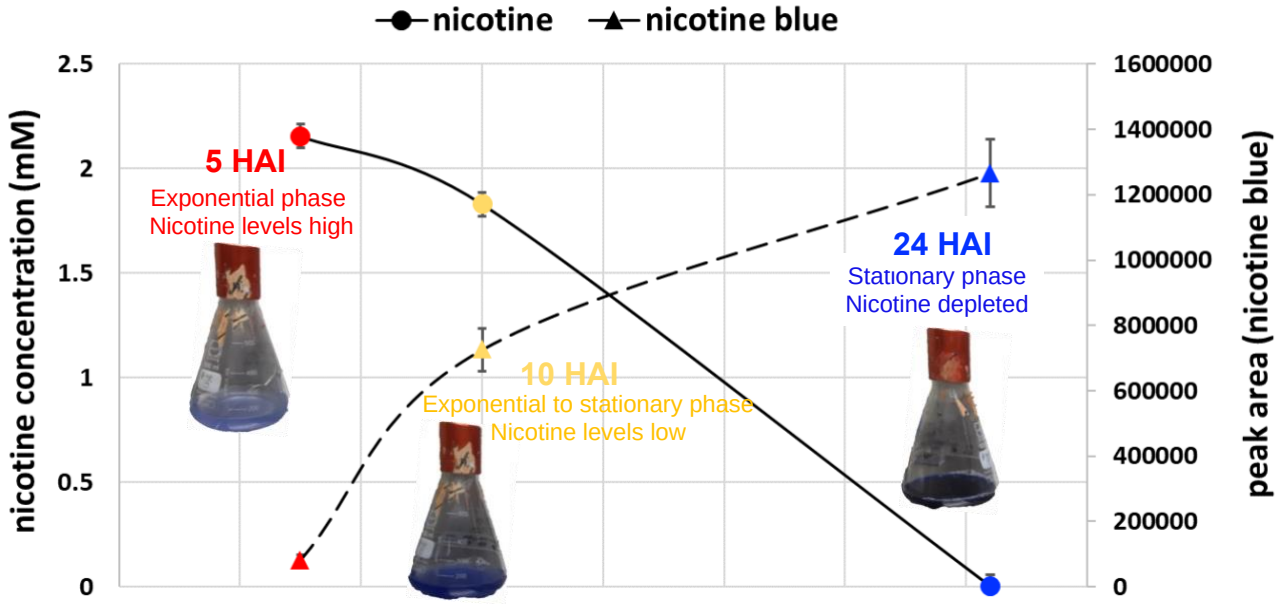


Figure 3. *P. nicotinovorans* ATCC 49919 growth curves on nicotine containing media and the key moments used for sampling. HAI - hours after inoculation.

Total RNA was extracted, polyadenylated and sequenced on a nanopore device. The resulting data was processed using nf-core/nanoseq, including differential gene analysis (DE) using DESeq2. The main statistical indicators for evaluating the quality and completeness of the acquired data are presented in Table 1.

Table 1. Key indicators for the generated RNA-seq data.

The DE-analysis showed a number of 15 significantly differentially expressed transcripts at 5 HAI and 19 transcripts at 10 HAI, out of which 17 had a log2FC>1 (volcano plots in Figure 4).

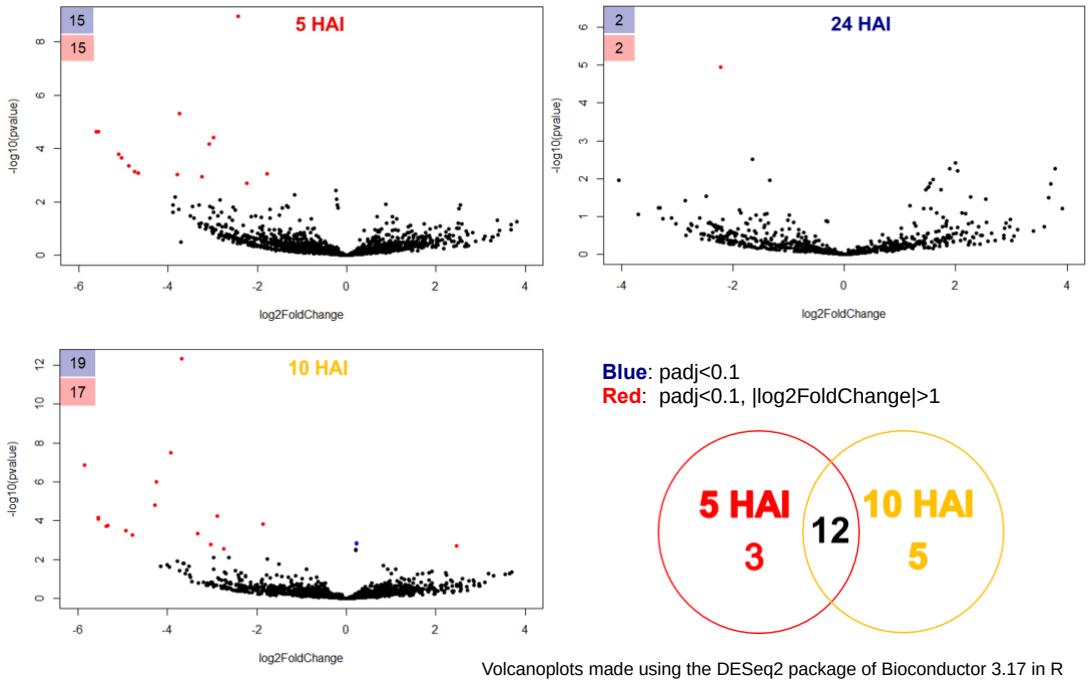


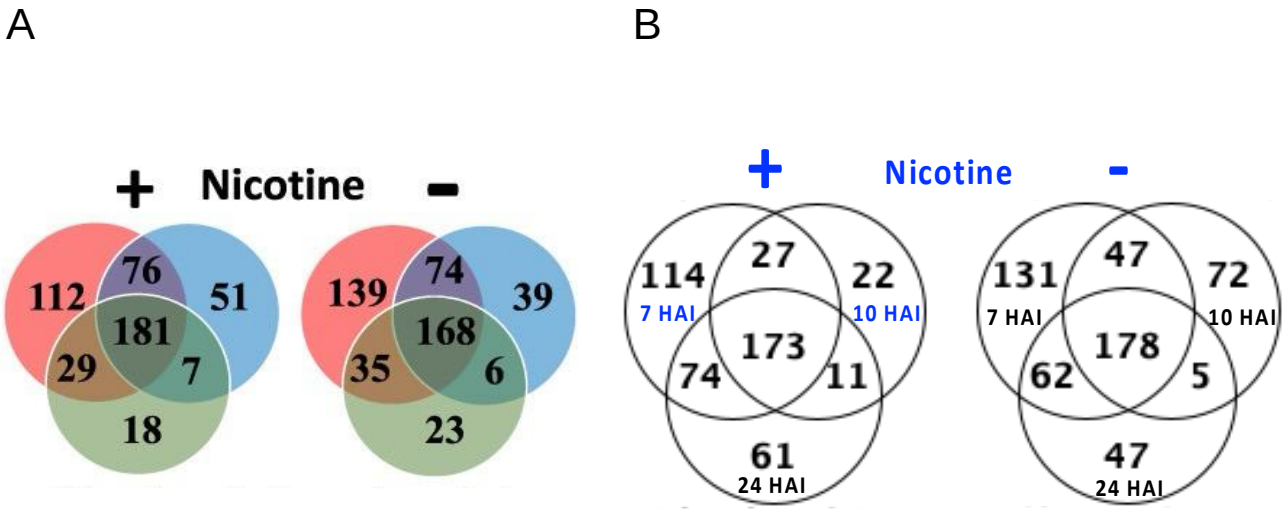
Figura 4. Volcano plots and Venn diagram indicating the number of differentially expressed transcripts when cells are grown with or without nicotine.

All transcriptomics data generated have been deposited GEO (Gene expression omnibus) under the access number GSE240220.

Proteomics data re-analysis and multi-omics data integration

Part of a Fulbright project the PI has attempted to solve the nicotine-induced proteome of *P. nicotinovorans* ATCC 49199. Without a suitable complete genome of the strain, the analysis performed at that time has its limitations as indicated in the published papers (10.1038/s41598-018-

34687-y and10.1021/acsomega.1c01020). The complete genomes resolved in this project offers a unique opportunity to re-analyze the MS/MS data. This allowed the identification of 607 proteins compared with the 528 initially reported. The distribution of these proteins according to the sampling time is presented in Figure 5.



DOI: (10.1021/acsomega.1c01020)

Figure 5. Comparative analysis of the number of identified proteins identified in the initial proteomics study (A) and fallowing the re-analysis using the genome generated in the project (B).

An integrated analysis of the proteomics and transcriptomics data offers an unique and complete overview of the nicotine metabolism in *P. nicotinovortans* ATCC 49199 – figure 6. Beside the genes already established as being involved in nicotine metabolism, the transcriptomics and proteomics data generated in the project allowed the identification of 3 extra genes that have putative roles in nicotine metabolism as well as 8 genes that have albeit having a nicotine-induced, their precise role is still missing.

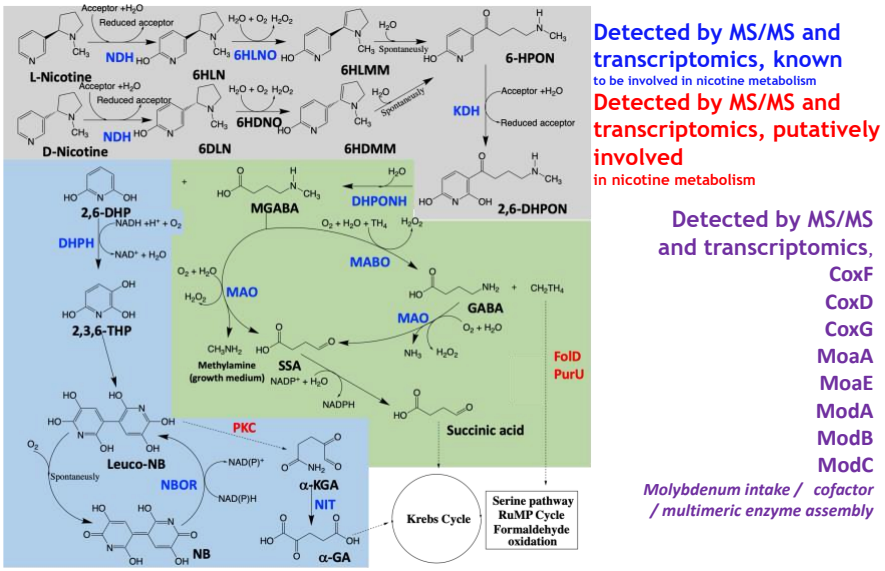


Figura 6. Overview of the nicotine degradation pathway and multi-omics data integration.

Publications and results disseminations

At the end of the project, the results have been presented at 17 conferences (9 international and 8 national) as a total of 15 lectures and 21 posters. Also, the results have been published in 3

articles in impact factor journals. Within the frame of the project, a number of 3 BSc and 5 MSc thesis have been defended.

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/