“ALEXANDRU IOAN CUZA” UNIVERSITY OF IAȘI
FACULTY OF BIOLOGY
DOCTORAL SCHOOL OF BIOLOGY

Integrated omics study of the nicotine degradation pathway from the microorganism *Paenarthrobacter nicotinovorans*

SUMMARY OF THE DOCTORAL THESIS

Scientific coordinator:
PROF. DR. HABIL. MIHĂȘAN MARIUS

Graduate student:
EL-SABEH AMADA

IAȘI
2025

THESIS TABLE OF CONTENTS

INTRODUCTION 5
ACKNOWLEDGEMENTS 6
LIST OF ABBREVIATIONS 7
KEYWORDS 10

MOTIVATION, AIM AND OBJECTIVES OF THE STUDY 11
Scientific motivation 12
Economic and ecological motivation 13
Medical motivation 14
Knowledge gaps relevant to the study 16
Aim 18
Main objectives 19

CHAPTER 1 - TAXONOMY, BIOCHEMISTRY AND MOLECULAR BIOLOGY OF THE *PAENARTHROBACTER NICOTINOVORANS* SPECIES 21
1.1 Taxonomy of the *Paenarthrobacter nicotinovorans* species 21
1.2 Morphology, metabolism and physiology of the *Paenarthrobacter* genus 21
1.3 The necessity to study microbial nicotine degradation 22
1.4 Microbial nicotine degradation pathways 25
1.5 Nicotine degradation in *Paenarthrobacter nicotinovorans* ATCC 49919 26
1.5.1 The pAO1 megaplasmid 26
1.5.2 The nicotine degradation pathway from *P. nicotinovorans* ATCC 49919 28

CHAPTER 2 - OMICS AND THEIR APPLICABILITY TO THE STUDY OF MICROORGANISMS 32
2.1 Defining elements of omics 32
2.2 Current and potential applications of omics 34
2.3 Challenges in omics 36
2.3.1 Reproducibility and replicability 38
2.3.2 Confidence in the accurate representation of biological molecules 39
2.3.3 Storage of big omics data 40
2.3.4 Challenges particular to multi-omics 47
2.4 Evolution of sequencing technologies 49
2.4.1 First generation sequencing 49
2.4.2 Second generation sequencing 51
2.4.3 Third generation sequencing 55
2.5 Genomics and its applicability to the study of microorganisms 61
2.5.1 Advances in the study of prokaryotes driven by the development of genomics 61
2.5.2 Opportunities provided by genomic analyses 62
2.5.3 Difficulties in bacterial genomics 63
2.5.4 Approaches to complete genome assemblies 64
2.6 Transcriptomics and its applicability to the study of microorganisms 71
2.6.1 The evolution of transcriptomics 71
2.6.2 Opportunities provided by transcriptomic analyses 72
2.6.3 Approaches in transcriptomics 73
2.6.4 Difficulties in bacterial transcriptomics 77
2.6.5 Transcriptomic studies of microbial nicotine degradation pathways 81
2.7 Proteomics and its applicability to the study of microorganisms 82
2.7.1 The evolution of proteomics 82
2.7.2 Opportunities provided by proteomic analyses 83
2.7.3 Approaches in proteomics 83
2.7.4 Difficulties in bacterial proteomics 85
2.7.5 Proteomic studies of the *Arthrobacter* genus 87

CHAPTER 3 - METHODOLOGY 89
3.1 *In vitro* methods 89
3.1.1 Bacterial strains and growth conditions 89
3.1.2 Identification of the key moments of nicotine catabolism 91
3.1.3 Extraction, purification, enzymatic treatment and quantification of nucleic acids prior to sequencing library preparation 96
3.1.4 Preparation of sequencing libraries 113
3.1.5 Short- or long-read sequencing of nucleic acids 121
3.2 *In silico* methods 124
3.2.1 Basecalling 124
3.2.2 DNA sequencing data processing, genome assembly and analysis 126
3.2.3 RNA sequencing data processing and differential gene expression analysis 142
3.2.4 Multi-omic data integration and visualization 172
3.2.5 Statistical analyses 175

CHAPTER 4 - RESULTS AND DISCUSSION 178
4.1 Assembly, annotation and analyses of the *Paenarthrobacter nicotinovorans* genome 178
4.1.1 Genomic data processing 179
4.1.2 Hybrid *de novo* genome assembly and annotation with NCBI PGAP 184
4.1.3 Genomic features of *P. nicotinovorans* ATCC 49919 185
4.1.4 Validation of the *P. nicotinovorans* ATCC 49919 genome 187
4.1.5 Functional analyses of the *P. nicotinovorans* ATCC 49919 genome 197
4.2 Transcriptomic analysis of nicotine degradation in *Paenarthrobacter nicotinovorans* 205
4.2.1 The key moments of nicotine catabolism 206
4.2.2 Optimisation of ONT direct RNA sequencing with the Flongle for bacteria 208
4.2.3 Transcriptomic data processing 212
4.2.4 Analysis of the differentially expressed genes during nicotine degradation 224
4.2.5 Functional enrichment analyses of the sDEGs 227
4.3 Multi-omic analysis of nicotine degradation in *Paenarthrobacter nicotinovorans* 229
4.3.1 Results of the proteomic data reanalysis using the whole genome 229
4.3.2 Plasmid genes differentially expressed during nicotine degradation 229
4.3.3 Chromosomal genes differentially expressed during nicotine degradation 233
4.3.4 Integration of the nicotine degradation pathway with the cell's overall metabolism 234
4.3.5 Genes involved in transcription regulation 236
4.3.6 Genes involved in transmembrane transport 237
4.3.7 Genes involved in oxidative stress responses 238
4.4 Contributions to international reference databases and software development platforms 241

CONCLUSIONS 242
PROSPECTS FOR FUTURE RESEARCH 246
SELECTED BIBLIOGRAPHY 248
APPENDIX. 271

KEYWORDS

Paenarthrobacter nicotinovorans ATCC 49919, *Arthrobacter oxydans*, *Arthrobacter nicotinovorans*, pAO1, degradation, integrated study, megaplasmid, bacterium, actinomycetes, actinobacterium, microorganism, NDM, NDB, nicotine, citrate, TCA, 6-hydroxy-nicotine, catabolism, metabolism, genome, genomic, transcriptome, transcriptomic, proteome, proteomic, multi-omic, multiomic, differential gene expression, differential protein expression, functional analysis, DEG, DEP, DGE, DPE, DESeq2, sequencing, next-generation sequencing, short-read sequencing, *de novo* sequencing, long-read sequencing, DNA sequencing, RNA sequencing, direct-RNA sequencing, nanopore, Oxford Nanopore Technologies, ONT, MinION, Flongle, Mk1B, nanopore sequencing, hybrid assembly, Unicycler, bioinformatics, mass spectrometry, liquid chromatography, UPLC, microbial biotechnologies

MOTIVATION, AIM AND OBJECTIVES OF THE STUDY

Nicotine is a secondary metabolite and a natural insecticide, synthesised by plants of the *Solanaceae* family as a defence mechanism against insects. In numerous countries, *Nicotiana* species are economically important crops due to their high nicotine content, which accounts for 98% of the plant's total alkaloids and 2-8% of its dry mass. Nicotine is also present in other *Solanaceae* family members (potatoes, tomatoes, aubergines, chilli peppers, etc.). Although in smaller quantities than in tobacco, nicotine is even present in the varieties of *Camellia sinensis* which are consumed globally as teas.

Because of its toxicity and rapid activity in its pure form, nicotine has been widely used in agriculture as an insecticide. Although it is naturally synthesised by plants, nicotine-based insecticides have been banned in the US since 2001 by the Environmental Protection Agency to prevent contamination of food with the toxic alkaloid. Despite these local efforts, globally and in a single year (2014), the tobacco industry generated over 32 million tonnes of green tobacco, which yielded 6.5 million tonnes of dry tobacco, 6 trillion cigarettes and significant amounts of nicotine-containing waste. These wastes have various forms, ranging from the dry plant material with a nicotine content of 0.83% (w/w) in stems and 1.5% - 2.22% (w/w) in leaves, to the wastewater and plastics contaminated with nicotine. When the nicotine content exceeds 0.05% (w/w), EU Regulations classify the waste as "toxic and hazardous".

Due to the worldwide consumption of tobacco and nicotine-containing foods, the organic compound itself and its active metabolites are constantly introduced into the terrestrial and aquatic environment via numerous and varied mechanisms. Some of these means are represented by the municipal wastewater or the improper disposal of the nicotine products and industrial waste involved in tobacco processing. Therefore, due to human industrial activities, nicotine has become a major pollutant, and the environmental toxicity of the compound and its derivatives has been recognized as a major concern in recent decades, since the waste requires appropriate treatment for safe disposal[1,2].

Moreover, considering the substantial economic burden required for the safe disposal of tobacco waste, it is not surprising that methods to exploit these renewable biomass resources have been intensively and consistently discussed since the turn of the century. Because the chemical structure of the alkaloid has a pyridine and a pyrrolidine ring, nicotine is a substrate for synthesis or biotransformation into its functional and renewable derivatives.

These range from substances of pharmaceutical and agricultural interest to precursors for the synthesis of numerous compounds of economic interest.

Recalling that nicotine is a naturally occurring metabolite, it is not surprising that microorganisms have evolved complete gene sets that allow them to degrade the compound and use it as a carbon and nitrogen source. The discovery in the 1950s of bacteria and fungi that can degrade the alkaloid despite its toxicity was revolutionary, because it suggested that the biologically mediated catabolic process could be an environmentally friendly, economical and efficient strategy for the treatment of nicotine-containing waste. Currently, nicotine-degrading microorganisms and their enzymatic repertoires are of major interest for the decontamination and reuse of nicotine waste[3].

Scientific motivation

Due to its ability to degrade and convert nicotine into non-toxic and useful compounds, the focus of this study is the actinobacterium *Paenarthrobacter nicotinovorans* ATCC 49919. Of primary importance is the fact that the strain has become the reference model for microorganisms that perform nicotine catabolism using the pyridine pathway. Among these microorganisms and since its isolation, *P. nicotinovorans* ATCC 49919 has remained the best studied. The gathered knowledge on the biochemical and genetic mechanisms that regulate nicotine degradation via the pyridine pathway in *P. nicotinovorans* ATCC 49919 revealed that the processes is encoded by the *nic*-gene cluster located on the native pAO1 megaplasmid[4].

Secondly, the strain is a study model for the molecular evolution of catabolic pathways and their spread via horizontal gene transfer in plasmids present in soil bacteria. An analysis of the relationship between genotype and ecotype demonstrated that the *Arthrobacter* pangenome is particularly open, being associated with a volatile accessory genome compared to previous studies of the pangenomes of other genera. Regarding the representative *P. nicotinovorans* ATCC 49919, there is strong evidence to support that the *nic*-gene cluster was probably acquired by pAO1 or by a precursor plasmid from the chromosome of a bacterium possibly related to *Rhodococcus opacus* or to a yet unknown *Arthrobacter* species. In addition, pAO1 was even involved in the horizontal transfer of genes from the chromosome of strain ATCC 49919 to other species of the *Arthrobacter* genus[5].

Economic and ecological motivation

Thirdly, there is a growing global interest towards using the biological agents represented by nicotine-degrading bacteria (NDB) to decontaminate the alkaloid found in the large quantities of waste generated and improperly disposed of by the tobacco industry annually. Therefore, *P. nicotinovorans* ATCC 49919 is a strain of ecological and economic importance due to its ability to catabolise toxic nicotine. Its use for the decontamination of nicotine and bioremediation of natural resources would easily adhere to the principles of green, biodegradable, non-toxic and environmentally friendly chemistry[3,4].

Fourthly, strain ATCC 49919 can be utilised for the conversion of nicotine-containing waste into a whole range of "green chemicals". The group of actinobacteria, which includes *P. nicotinovorans* ATCC 49919, is responsible for the production of 45% of the documented secondary metabolites and of many lytic enzymes known for their functions in various physiological, cellular and biological processes, including the signal transduction which allows cells to interact with their environment and adapt accordingly[6]. Various species of actinobacteria of the *Rhodococcus*, *Streptomyces*, *Nocardioides* and *Corynebacterium* genera are used as cell factories for the synthesis of useful industrial products. It is therefore not surprising that many intermediary metabolites of the nicotine degradation pathway from *P. nicotinovorans* ATCC 49919 are of industrial and pharmaceutical importance.

For example, following the degradation of nicotine by *P. nicotinovorans* ATCC 49919, methylamine accumulates in the growth medium. In addition, other nicotinic derivatives such as gamma-aminobutyric acid (GABA), α-ketoglutaric acid, succinic acid, etc. are synthesised in this species. Methylamine is a major intermediate compound in the chemical industry, used as a synthesis precursor for other organic substances. Methylamine is used for the synthesis of theophylline and ephedrine in the pharmaceutical industry and in the production of pesticides and solvents, an estimated 115,000 tonnes of methylamine being chemically synthesized in 2005. Succinic acid was recognised by the US Department of Energy as one of the top 12 value-added products from biomass, representing chemicals with multiple functional groups that have the potential to be transformed into new families of useful molecules. In 2021, the global market size for succinic acid was estimated at ~$223 million.

Medical motivation

What distinguishes *P. nicotinovorans* ATCC 49919 from other microorganisms capable of using nicotine as a carbon source is the strain's potential in the pharmaceutical

industry, through the secondary metabolites it synthesises. Among these, of major relevance are GABA, α-ketoglutaric acid, 6-hydroxy-L-nicotine (6HLN) and even methylamine.

The effects of 6HLN on memory impairment, anxiety and oxidative stress were assessed in *Danio rerio* models of scopolamine-induced Alzheimer's disease[7]. The behaviour assays, biochemical determination of acetylcholinesterase activity and oxidative stress enzymes expressed in the zebrafish brain, plus the real-time quantitative polymerase chain reaction (RT-qPCR) analysis of the expression of genes involved in modulating the learning and memory processes indicated that 6HLN attenuated anxiety-like behaviour and memory deficits, reduced oxidative stress and acetylcholinesterase activity in the brain and increased the expression of the tested marker genes[7]. Therefore, it was postulated that 6HLN could compensate the depletion of acetylcholine and restore normal brain functions, making it a possible target for the development of drugs useful in the treatment of disorders associated with a decrease in brain acetylcholine levels, such as Alzheimer's disease[7].

Knowledge gaps relevant to the study

From the afore mentioned, it can be concluded that the importance of *P. nicotinovorans* ATCC 49919 is multifactorial. In addition to its value for the study of nicotine degradation via the pyridine pathway, by involving more sustainable industrial processes and starting from the renewable resources represented by the large amounts of nicotine produced annually or already present in the contaminated waters and soils, the strain could become a cell factory for compounds synthesized following the practices of green chemistry[8]. The products resulting from the nicotine degradation encoded by the pAO1 megaplasmid, of which were highlighted methylamine, succinic acid and 6-hydroxy-L-nicotine, have intrinsic economic value. Additionally, they can be used as precursors for the synthesis of a wide range of compounds with applications in multiple areas of industry, including the pharmaceutical, agricultural, food and cosmetics ones.

Considering the potential of *P. nicotinovorans* ATCC 49919 coupled with the fact that the pAO1 megaplasmid sequence has been known[9] since 2003, the absence of a complete *P. nicotinovorans* ATCC 49919 genome in sequence databases is surprising. Because of the lack of genomic information, any efforts to harness the ecological and biotechnological potential of any *Paenarthrobacter* strains are severely hindered and even made impossible by the scarcity of precise genetic tools necessary for improving the rate of catabolism of any substrate and increasing the yield of the metabolic intermediates of interest.

The absence of a complete and functionally annotated genome is accompanied by the sparseness of data on the genetic mechanisms which regulate the metabolic pathways encoded by the bacterial chromosome itself and, if applicable, by the corresponding plasmids. In addition, if any plasmids are present, which is of relevance for strain ATCC 49919, a comprehensive knowledge of the mechanisms through which the metabolic pathways encoded by the extrachromosomal and chromosomal DNA interact and integrate is essential. Acquiring this knowledge requires, besides the implicit exploration of the genome, the assessment of any changes induced by the environmental conditions, adaptations which must be analysed at the transcriptome and proteome level. Moreover, the lack of the complete bacterial genome sequence is accompanied by the absence of key information on replication origins, promoters, operon structure, transcription start and stop sites and CRISPR systems in the genus *Paenarthrobacter*, further limiting the development of engineering tools. In addition, plasmids of the *Paenarthrobacter* genus are also insufficiently studied, with only two publications[5,10] focusing on their comparative genomics and molecular biology.

Removing the obstacle represented by the incomplete genome of any *P. nicotinovorans* strain is an important step towards the application of these microorganisms. However, sequencing and assembling the genome of *P. nicotinovorans* ATCC 49919 specifically is of even greater scientific relevance and timeliness, being an important step in complementing the efforts already made towards characterising nicotine degradation via the pyridine pathway. The "classical" approach of cloning, purification and characterisation used until 2019 to study nicotine metabolism in *P. nicotinovorans* ATCC 49919 was found to be limited in its potential to provide robust experimental data on the function of numerous genes comprising the pyridine pathway encoded by pAO1[11].

To address this problem, Mihășan et al. utilised a proteomic approach using liquid chromatography coupled with tandem mass spectrometry detection (nanoLC-MS/MS) to identify the proteins expressed by *P. nicotinovorans* ATCC 49919 in three different growth media[11–13]. Although this approach provided important insights into the expression of genes encoding nicotine catabolism in pAO1, there are still numerous questions regarding the functionality of these sequences, which have not been fully addressed by the proteomic experiments either[11–13]. Specifically, for half of the pAO1 megaplasmid genes presumed to be involved in nicotine catabolism, their function is either unknown or there is no experimental evidence of involvement in nicotine metabolism. Even more, it is possible that

at least some of these genes are artefacts of the evolutionary processes that contributed to the assembly of the *nic* cluster[13].

Aim

The recent development of "gene-to-metabolite" and "metabolite-to-gene" omic technologies has had a major impact in revealing the prevalence of silent gene clusters in the genomes of microorganisms. Moreover, the discovery of precise genome editing tools and the use of artificial operons for metabolic pathway engineering have become key players in activating silent gene clusters to obtain novel metabolites on a large-scale, which previously would have been poorly expressed and difficult to characterise under laboratory conditions. In comparison to the traditional method of metabolite extraction, access to the diversity of distinct and previously uncharacterised biosynthetic gene clusters plus the use of modern gene editing tools to modulate operon expression are expected to contribute to the discovery of new compounds and to their diversification. Considering that out of 23,000 bioactive metabolites synthesised by microorganisms, approximately 10,000 were identified in actinobacteria, there is an overwhelming potential in exploring this group of microorganisms to improve the quality of human life and beyond.

Therefore, to reveal the unknown or insufficiently explored aspects of nicotine degradation via the pyridine pathway in *P. nicotinovorans* ATCC 49919, the implementation of a multi-omic approach[14] was chosen, based on the recent developments in the field. This study aimed to fully assemble and analyse the genome of the strain of interest, sequence and analyse its transcriptome following nicotine treatment, re-analyse the pre-existing time-dependent proteomic data using the novel complete genome and, finally, integrate and analyse the omics data on nicotine degradation by *P. nicotinovorans* ATCC 49919.

Thus, the first complete genome of a *P. nicotinovorans* strain and the first transcriptome of a microorganism that performs nicotine catabolism using the pyridine pathway were assembled and analysed. These achievements have enabled the first multi-omic study of bacterial (and even microbial) nicotine catabolism, integrating genomic, transcriptomic and proteomic data. The multi-omic analysis of nicotine degradation in *P. nicotinovorans* ATCC 49919 should further our understanding of the molecular biology and biochemistry of nicotine degradation via the pyridine pathway, providing the foundation for the further development of a catabolic model in the reference species. Starting from this model and with the help of genetic engineering, *P. nicotinovorans* ATCC 49919 has the potential to become an important biological agent in the decontamination of nicotine-

polluted resources, and a real chemical factory through processes adhering to the principles of green and sustainable chemistry.

Main objectives

The aim of this study is to generate novel genomic and transcriptomic data and integrate it with a re-analysed proteomic dataset describing nicotine degradation via the pyridine pathway from *P. nicotinovorans*. The number of main study objectives corresponds to the types of macromolecules analysed. Following, the main objectives of the study and their associated activities are presented.

O.1. Obtain and analyse the fully assembled, structurally and functionally annotated genome of *P. nicotinovorans* ATCC 49919.

O.1.1. Validate and implement the methodology for the extraction and purification of high molecular weight genomic DNA (gDNA) for long-read sequencing.
O.1.2. Prepare libraries and perform long-read sequencing of gDNA.
O.1.3. Primary process the gDNA short-read sequencing data.
O.1.4. Primary process the gDNA long-read sequencing data.
O.1.5. Assemble the *P. nicotinovorans* ATCC 49919 genome via a hybrid *de novo* approach.
O.1.6. Validate and annotate the *P. nicotinovorans* ATCC 49919 genome assembly.
O.1.7. Perform the functional analyses of the *P. nicotinovorans* ATCC 49919 genome.
O.1.8. Disseminate the results obtained by fulfilling this objective.

O.2. Obtain the transcriptome of *P. nicotinovorans* ATCC 49919 induced by nicotine and analyse the differential gene expression between the key time points of the alkaloid's degradation.

O.2.1. Identify the key moments of nicotine catabolism in the strain of interest.
O.2.2. Establish the methodology for total RNA extraction for long-read sequencing.
O.2.3. Prepare the libraries and perform direct RNA sequencing.
O.2.4. Primary process the direct RNA sequencing data.
O.2.5. Process the direct RNA reads for the statistical analysis of the differential gene expression following the nicotine treatment of *P. nicotinovorans* ATCC 49919.
O.2.6. Interpret the differential gene expression analysis results.
O.2.7. Disseminate the results obtained by fulfilling this objective.

O.3. Perform the integrated omic analysis of *P. nicotinovorans* ATCC 49919 treated with nicotine.

O.3.1. Reinterpret the nicotine-related proteome of *P. nicotinovorans* ATCC 49919 using the complete genome of the strain (result of **O.1.**).

O.3.2. Statistically analyse the differential protein expression at the three key time points of nicotine degradation by *P. nicotinovorans* ATCC 49919.

O.3.3. Integrate the transcriptomic data (result of **O.2.**) with the reinterpreted proteomic data (result of **O.3.1.**) and analyse the genes and their products involved in the response of *P. nicotinovorans* ATCC 49919 to nicotine treatment.

O.3.4. Disseminate the results obtained by fulfilling this objective.

CHAPTER 3 - METHODOLOGY

Figure 3.1. summarises **O.1.1.** – **O.1.7.**, the main activities pursued to achieve the first objective of the study, i.e. obtain, annotate and analyse the complete genome of *Paenarthrobacter nicotinovorans* ATCC 49919.

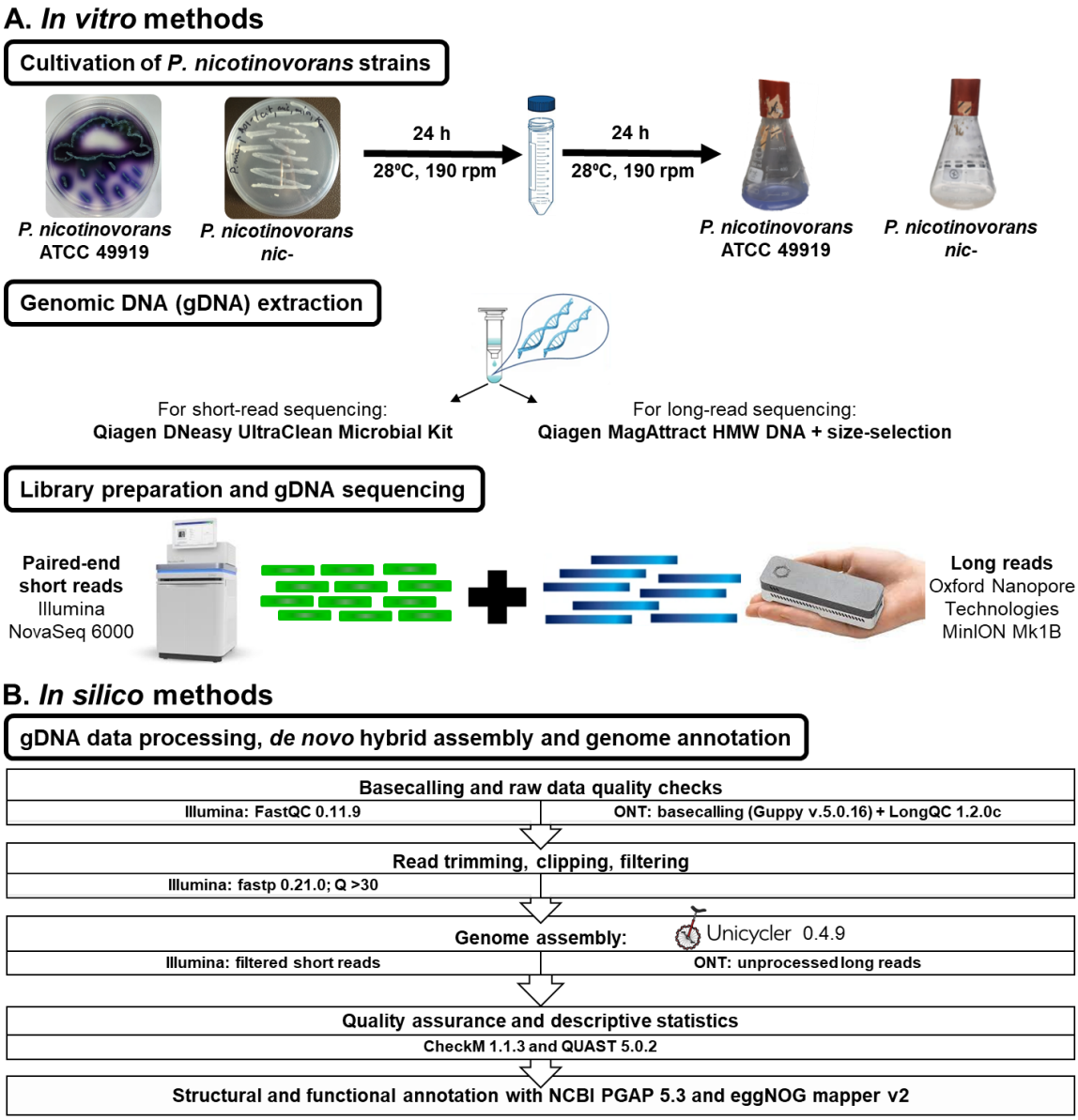

Figure 3.1. Schematic representation of activities O.1.1. - O.1.7., carried out to assemble, annotate, validate and functionally analyse the complete genome of *Paenarthrobacter nicotinovorans* ATCC 49919.

Figure 3.2. - A. depicts **O.2.1. - O.2.3.**, representing the main wet-lab activities performed to accomplish the second objective of the study, i.e. sequence the *P. nicotinovorans* ATCC 49919 nicotine-induced transcriptome. Figure 3.2. - B. corresponds to activities **O.2.4. - O.2.6.** and provides a summary of the main steps involved in processing the transcriptomic data to analyse the differential gene expression of *P. nicotinovorans* ATCC 49919 during nicotine catabolism.

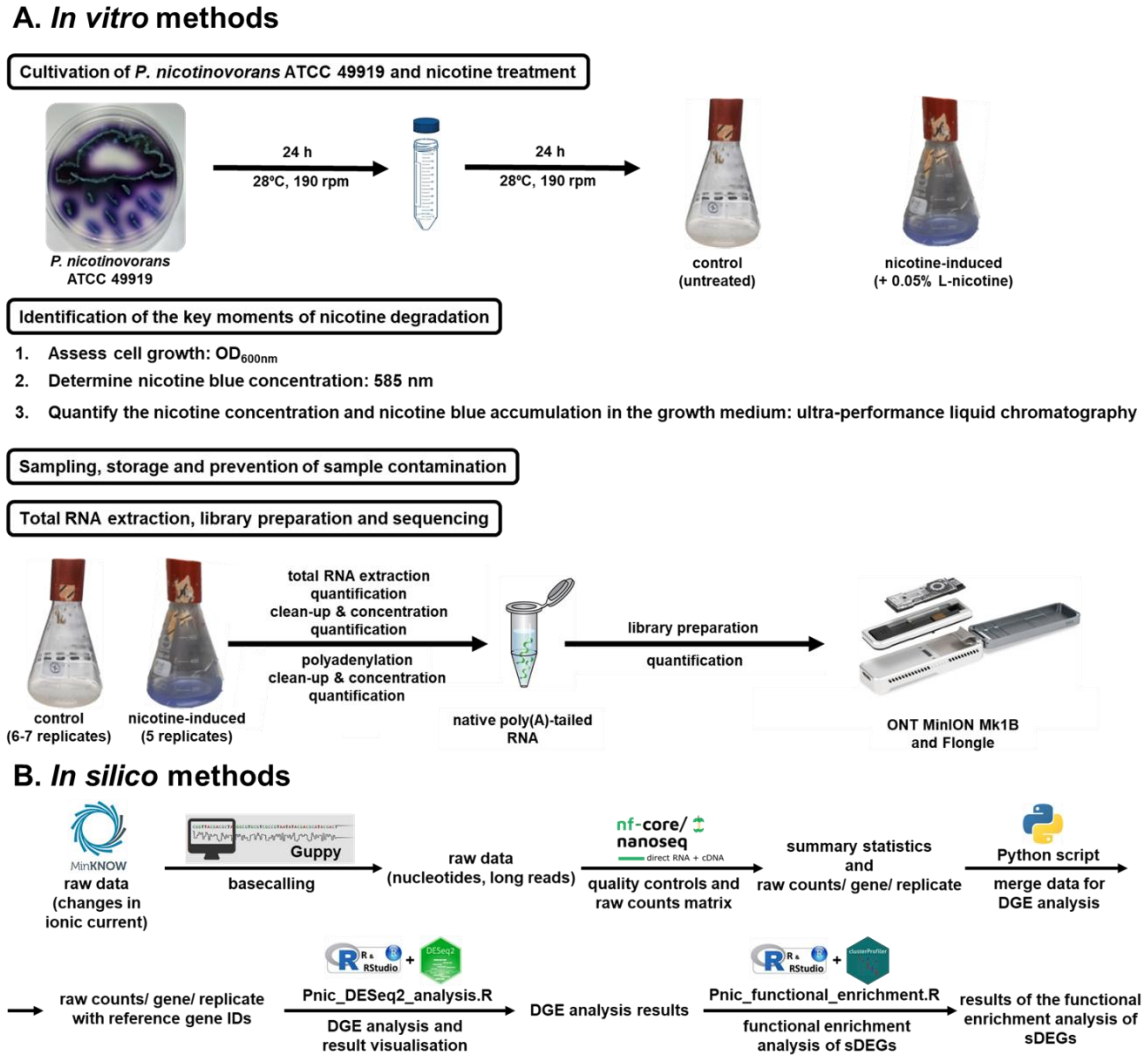

Figure 3.2. Schematic representation of the main steps undertaken to analyse the differential gene expression (DGE) during nicotine degradation in *Paenarthrobacter nicotinovorans* ATCC 49919, associated with activities A. O.2.1. - O.2.3. and the *in vitro* methods, respectively B. O.2.4. - O.2.6. and the *in silico* methods.

OD_{600nm}: the optical density of a sample measured at 600 nm; sDEGs: statistically significant differentially expressed genes.

Regarding activity **O.3.1.**, the proteomic dataset available in the ProteomeXchange with ID PXD012577 was re-analysed using as reference the *P. nicotinovorans* ATCC 49919 genome annotation. The Scaffold programme with the MudPIT analysis option was used to validate the peptides and proteins identified in the nanoLC-MS/MS data. For activity **O.3.2.**, Fisher's test was used to analyse the protein differential expression between the three key time points of nicotine degradation in *P. nicotinovorans* ATCC 49919.

The results tables for each pairwise comparison were further processed with the Pnic_proteome_FC_rescale.R script[15] to transform the fold change (FC) into log2 (LFC) values and re-scale them against the minimum and maximum LFC values obtained from the transcriptomic data. These modifications were necessary to allow the integrated visualisation of the data resulted from the individual transcriptomic and proteomic analyses.

For activity **O.3.3.**, the metabolic pathways enriched during nicotine catabolism were identified by implementing the enrichKEGG function from the clusterProfiler package, with the statistical threshold $p = 0.1$ and the Benjamini-Hochberg (BH) method to correct for multiple hypothesis testing. The pathview package 1.42.0, available through Bioconductor, was used to visualise the most enriched biochemical pathways available in KEGG PATHWAY.

CHAPTER 4 - RESULTS AND DISCUSSION

4.1 Assembly, annotation and analyses of the Paenarthrobacter nicotinovorans genome

Figure 4.1. shows the circular maps of the replicons identified in the two genomes sequenced and assembled in this study, namely *P. nicotinovorans* ATCC 49919 and the *nic*- derivative strain lacking the pAO1 megaplasmid. Two replicons were identified in the genome of *P. nicotinovorans* ATCC 49919 (Figure 4.1. – A.). The longest replicon of strain ATCC 49919 measures 4 316 184 bp and has an average guanine and cytosine (GC) content of 63.2%, being assigned to the bacterial chromosome. A total of 4173 genes and 4098 coding sequences (CDSs) for 4068 proteins and 30 pseudogenes were identified on the chromosome. The genome displays 54 CDSs for transfer RNA, three CDSs for non-coding RNA and six identical and complete ribosomal operons, each containing all three genes encoding the prokaryotic rRNA subunits (5S, 16S and 23S, coloured purple in Figure 4.1. – A.). In two instances, two ribosomal operons are in proximity, separated only by three non-rRNA-encoding sequences. The second replicon has a length of 165 141 bp and an average GC content of 59.7%, representing a plasmid with 145 CDSs. BLAST analysis of the extrachromosomal DNA present in strain ATCC 49919 showed 99.9% identity to the pAO1 megaplasmid sequence previously described by Igloi and Brandsch (GenBank record AJ507836.1) and known[9] to be present in the genome of strain ATCC 49919.

To contribute with reliable data to the community, multiple technical validations of the assembled genomes were performed to facilitate the study and use of strain ATTC 49919, of the *P. nicotinovorans* species, and of the whole *Paenarthrobacter* genus. To ensure access to the resulting genomic data in adherence to the FAIR principles in the omic sciences (detailed in the thesis subchapters 2.3.1 and 2.3.3.2), the complete genomes were registered and are accessible in the international reference database NCBI GenBank. In February 2022, NCBI GenBank selected the *P. nicotinovorans* ATCC 49919 genome assembled here as the reference sequence for the *P. nicotinovorans* species, in further recognition of its quality. The significance of the genomes included in NCBI RefSeq was described in the thesis subchapter 2.3.3.2 on the main databases used to store omics data.

Currently, there are 21 *P. nicotinovorans* genomes registered in NCBI Genome, of which 13 were deposited between February 2022 and March 2025. Even to date, of the 21 total genomes categorised as *P. nicotinovorans*, only three assemblies are complete, two of the three being the ones resulted from this study for ATCC 49919 and *nic*-. Furthermore, the result of the first study objective, namely the complete, circular and structurally and

functionally annotated genome of *P. nicotinovorans* ATCC 49919, has retained its status as the reference sequence for the *Paenarthrobacter nicotinovorans* species since February 2022, which emphasises its quality and impact.

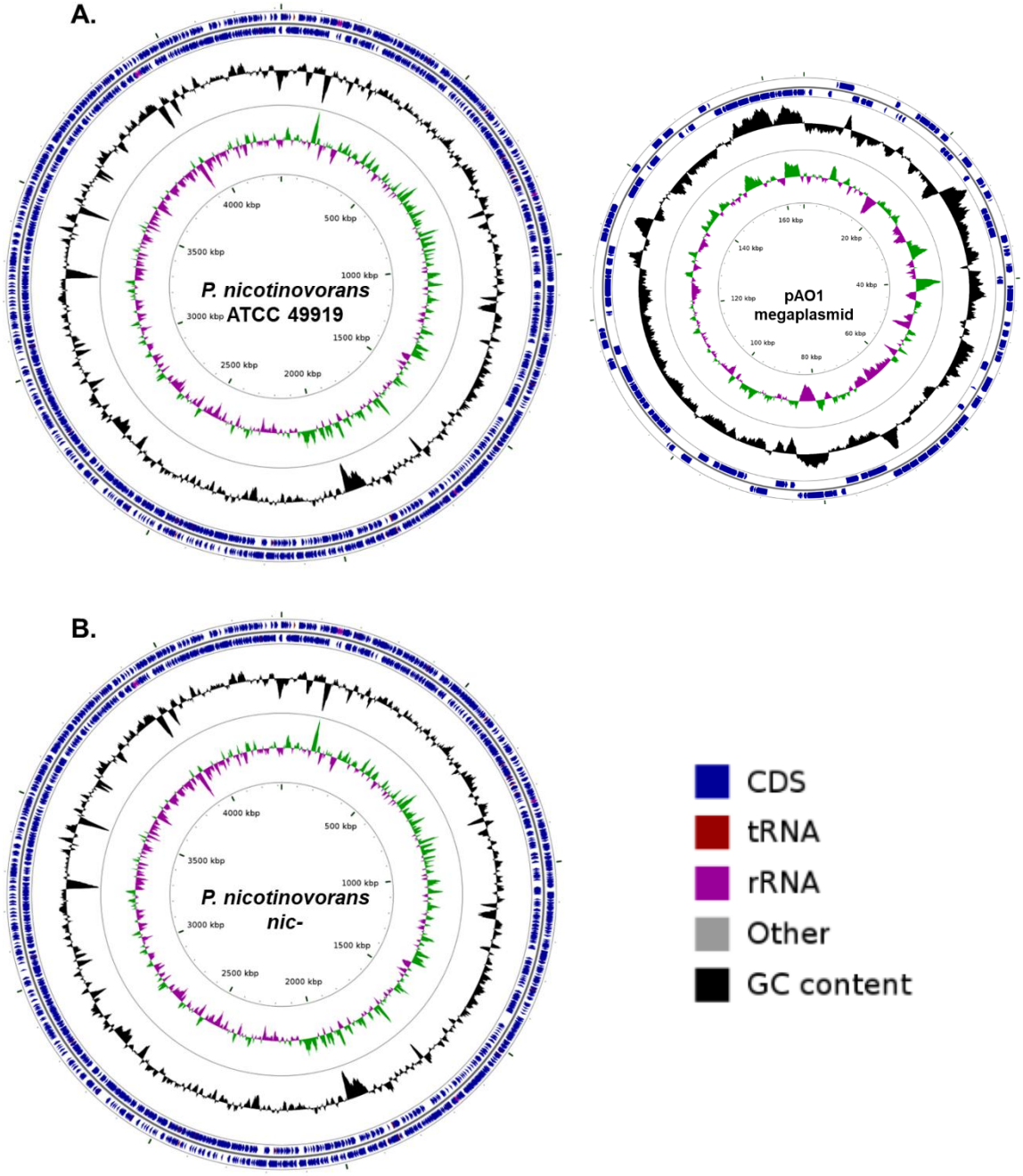

Figure 4.1. Circular maps of the assembled *P. nicotinovorans* genomes.

A. Replicons identified in *P. nicotinovorans* ATCC 49919; left: the bacterial chromosome, right: the pAO1 megaplasmid. **B.** The replicon identified in *P. nicotinovorans nic-*, representing the bacterial chromosome. The figures were generated with CGView, according to the thesis subchapter 3.2.2.6; kbp: kilo base-pairs; CDS: coding sequences; GC: guanine and cytosine.

4.2 Transcriptomic analysis of nicotine degradation in Paenarthrobacter nicotinovorans

4.2.1 The key moments of nicotine catabolism

Figure 4.2. shows the correlation between the main growth stages of *P. nicotinovorans* ATCC 49919 with the key moments of nicotine catabolism for the samples used for the transcriptomic analysis. The key time points for sampling *P. nicotinovorans* ATCC 49919 cells were set at approximately 5, 10 and 24 hours after inoculation (HAI).

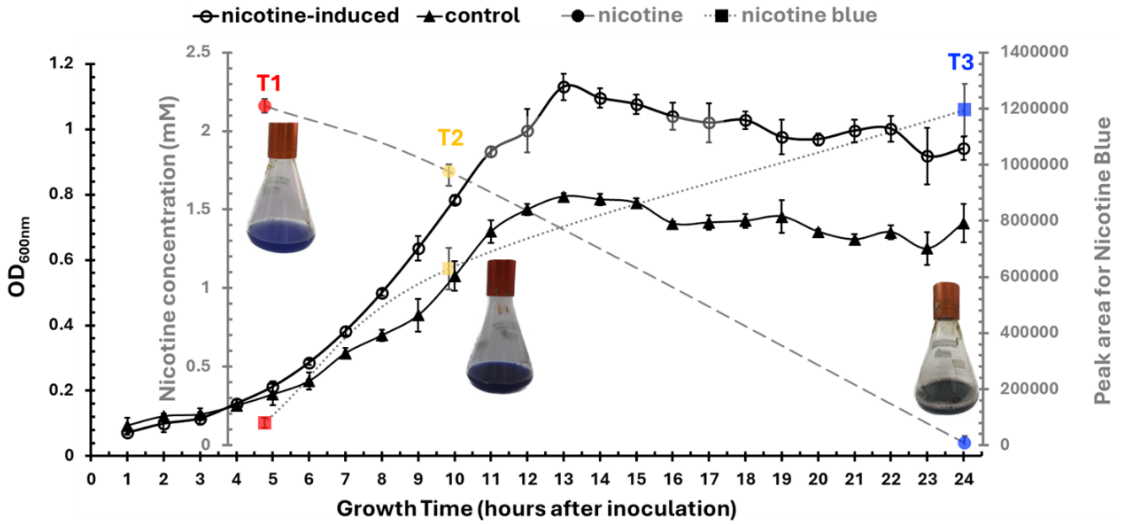

Figure 4.2. Dynamics of the nicotine concentration and nicotine blue accumulation at the three key time points of nicotine catabolism in *P. nicotinovorans* ATCC[15].

Red: samples taken at 5 h after inoculation (HAI), corresponding to the start of the log phase; yellow: samples taken at 10 HAI, corresponding to the log to stationary phase transition; blue: samples taken at 24 HAI, corresponding to the stationary phase. Error bars are the standard deviation for a minimum of three measurements per biological replicate[15].

In *P. nicotinovorans* ATCC 49919, the lag phase lasts until ~ 5 HAI and corresponds to an OD_{600nm} = 0.25. This timepoint, hereafter referred to as T1, represents the onset of the logarithmic/ exponential/ log growth phase, indicated by the debut of the accumulation of the final product of nicotine catabolism in *P. nicotinovorans* ATCC 49919, nicotine blue, in the growth medium (Figure 4.2.).

At ~10 HAI (OD_{600nm} = 0.6), hereafter referred to as T2, the transition from the logarithmic to the stationary phase is captured. At this stage, the amount of nicotine in the growth medium starts to decrease (Figure 4.2.). At T2, *P. nicotinovorans* ATCC 49919

cultures reach peak cell density, nicotine catabolism is intense, and the accumulation of intermediate metabolites in the growth medium continues.

At 24 HAI, hereafter referred to as T3, the cells are in the stationary phase. This starts at ~18 HAI and is characterised by a constant number of bacterial cells, the complete consumption of nicotine from the culture medium and the end of nicotine catabolism, as shown in Figure 4.2. Depending on the growth medium supplementation with nicotine, at 24 HAI the OD_{600nm} ranges from 0.7 to 1, the cell growth being more intense when nicotine is available as an additional carbon source to citrate (Figure 4.2.).

4.2.2 Analysis of the differential gene expression during nicotine degradation

Figure 4.3. – A. shows that under nicotine supplementation at T1, T2 and T3 were identified 29, 35 and 17 sDEGs (p-adj < 0.1 and |LFC| > 1), respectively. As shown in Figure 4.3. – B., 25 of the sDEGs at T1 were also identified at T2, all belonging to the *nic*-genes cluster on the pAO1 megaplasmid. The only four sDEGs unique to T1 are all chromosomal, whereas the ten genes uniquely differentially expressed at T2 are six *nic* and four chromosomal genes. With a single exception represented by *JMY29_20035*, a chromosomal gene encoding an ABC transporter permease which was downregulated at T1 in the presence of nicotine (T1 nic-induced vs control: LFC < -4.4, p-adj < 0.1), all of the sDEGs were upregulated during nicotine catabolism at T1 and T2[15].

Regarding T3, all the sDEGs by the cells grown on nicotine were specific to this timepoint (Figure 4.3. – B.) and are encoded by the bacterial chromosome – three sDEGs were upregulated, while the remaining 14 genes were downregulated. These results fit well with the sample clustering analysis (thesis subchapter 4.2.2.3), as T1 and T2 samples are from the early and, respectively, late logarithmic stage, therefore they share most of the sDEGs affected by nicotine. On the other hand, the distinct T3 samples represent the late stationary phase, by which nicotine has already been depleted, hence they share no sDEGs with the cells sampled at T1 or T2[15].

A time-course analysis of the transcripts from the cells grown on the nicotine-supplemented medium shows no significant differences (p-adj < 0.1) between the genes expressed in the early log and log to stationary phase (nic-induced T2 vs T1). However, when comparing the cells in the late stationary phase (T3, nicotine is depleted) to the ones in the late logarithmic phase (T2), a higher number of sDEGs was detected than for the comparison between the nicotine-induced cells in the stationary versus early log phase (with p-adj < 0.1 and |LFC| > 1, 1171 sDEGs for T3 vs T1, respectively 1249 sDEGs for T3 vs

T2). 1035 sDEGs were found to be common to both the early log and the log to stationary phase, whereas 136 sDEGs were specific to T1 and 214 sDEGs were only identified at T2. The sDEGs encoded by pAO1 were all downregulated at T3, confirming that nicotine catabolism ceased in the stationary phase because of the complete consumption of nicotine from the growth medium, as was expected from the UPLC results (Figure 4.2.)[15].

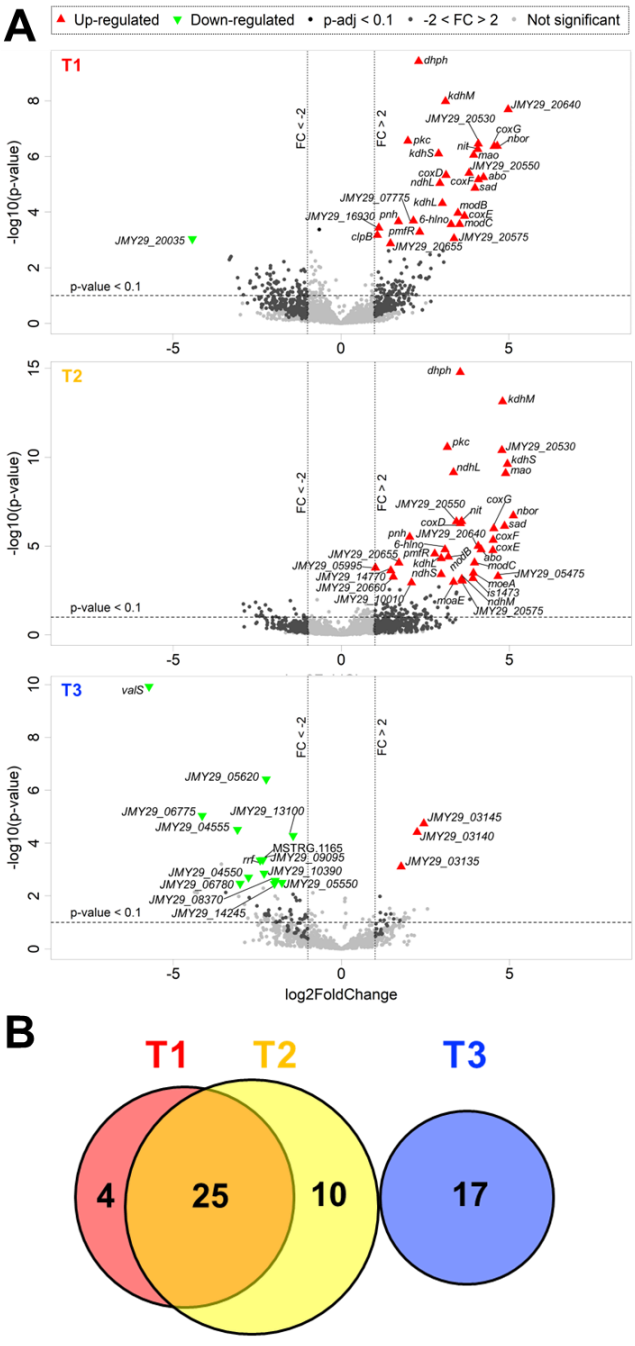

Figure 4.3. Nicotine-related differentially expressed genes[15].

A. Volcano plots of the DGE analysis of nicotine-induced versus control conditions at T1, T2 and T3; **B.** Venn diagram showing the overlap of the sDEGs between the nicotine-induced versus control conditions at each of the three key timepoints[15].

4.3 Multi-omic analysis of nicotine degradation in Paenarthrobacter nicotinovorans

4.3.1 Plasmid genes differentially expressed during nicotine degradation

The DGE analysis of *P. nicotinovorans* ATCC 49919 at the three timepoints of nicotine catabolism revealed that 25 of the 40 annotated *nic*-genes (Figure 4.4., Figure 4.5.) were upregulated in the logarithmic phase in the presence of nicotine (Figure 4.3. – A.). Of these 25 *nic*-genes, 14 (*nbor, mao, sad, coxD, nit, coxG, dhph, pkc, pnh, kdhL, kdhM, 6-hlno, ndhL, modC*) had products that were detected by nano LC-MS/MS in the growth medium supplemented with nicotine[13,15]. In addition, the enzymes encoded by *nbor, mao, sad, nit, dhph, pnh, kdhL, kdhM, 6-hlno* and *ndhL* have known and experimentally proven functions related to nicotine catabolism in *P. nicotinovorans* ATCC 49919[15–17].

On the other hand, *coxD, coxG, pkc* and *modC* have thus far been only putatively involved in nicotine degradation based on their products being detected by nanoLC-MS/MS in the growth medium supplemented with nicotine[13]. Direct RNA sequencing data confirms *coxD, coxG, pkc* and *modC* have nicotine-induced expression, as they are significantly downregulated (LFC < -4, p-adj < 0.05) in the stationary phase when nicotine is no longer detected in the growth medium (Figure 4.2.)[15].

The other 11 *nic*-genes shown here to be upregulated in the logarithmic phase in the presence of nicotine (Figure 4.3. – A.) are *JMY29_20530* (hypothetical cupin domain-containing protein), *JMY29_20550* (hypothetical fumarylacetoacetate hydrolase family protein), *abo*, *perm* (putative APC family permease), *pmfR*, *coxF*, *coxE*, *JMY29_20640* (hypothetical Dabb family protein), *JMY29_20655* (cupin domain-containing protein), *kdhS* and *modB*. The corresponding proteins encoded by these 11 *nic*-genes could not be identified in the proteomic data, most probably due to their low abundance, sample preparation or other unknown factors[15].

For the remaining *nic*-genes annotated initially[9], no statistically significant (p-adj < 0.1 and |LFC| > 1) differential expression that would suggest their involvement in nicotine catabolism was detected. Most of the genes that were experimentally correlated with nicotine metabolism prior to the current study were correctly annotated in the pAO1 sequence published in 2003[9]. However, the analysis of the pAO1 sequence also revealed differences in the length of the annotated genes, which were caused by changes in the coordinates describing start or stop codons. Additionally, there are also discrepancies between the computational structural annotation with NCBI PGAP and the start and stop coordinates indicated by the direct reads of the transcripts expressed by *P. nicotinovorans* ATCC 49919 in the tested experimental conditions. Therefore, all the *nic*-genes for which there was at

least one read in at least one RNA library sequenced here are labelled in Appendix 6 of the thesis. Also included are the inconsistencies in the annotation of the *nic*-cluster observed between the data obtained by Igloi and Brandsch[9] and those generated by short- and long-read DNA sequencing and direct RNA sequencing.

For eight of the hypothetical or putative *nic*-genes detected in the original[9] pAO1 sequence, the analysis of the newly generated gDNA or RNA reads did not provide further evidence that any of these genes are expressed or at least encoded by the plasmid sequence assembled here using the hybrid *de novo* approach. This observation is substantiated by the lack of any structural and functional computational annotation of these genes, performed with NCBI PGAP 5.3. Therefore, several differences in the annotation of hypothetical and putative proteins between the original and new annotation of pAO1 were also noticed and highlighted in Appendix 6 of the thesis, these inconsistencies owing to the improvement over the last decades in gene prediction algorithms[18].

Moreover, neither the direct RNA sequencing data nor its bioinformatic analyses (alignment against the complete genome with Minimap2, transcript reconstruction and quantification with StringTie2 and featureCounts), indicated the existence or expression of the eight putative or hypothetical genes, marked in Appendix 6 of the thesis as “nedetectat de PGAP sau prin secvențiere”. The discrepancies between these coordinates could have complicated and adversely affected the success of applications requiring genetic engineering of the pAO1 megaplasmid and of *P. nicotinovorans* ATCC 49919.

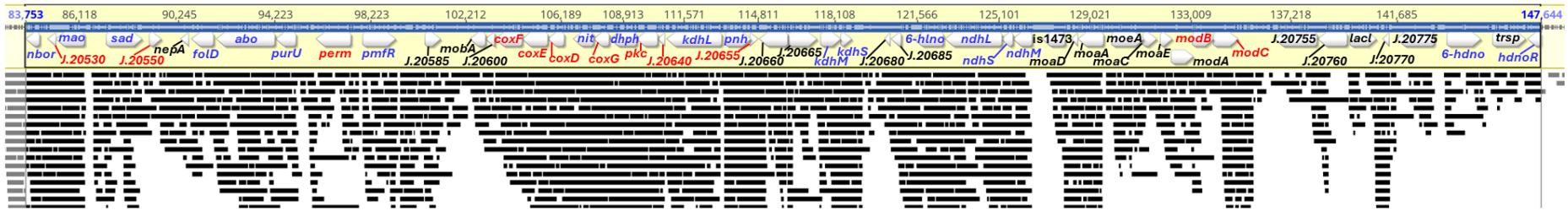

Figure 4.4. Genetic organisation of the *nic*-gene cluster (top) and representative RNA reads (bottom) aligned using Geneious Prime 2024.0.7 RNA with the medium sensitivity setting[15].

Blue text: genes with known functions in nicotine degradation and previously shown to have nicotine-related expression; red text: genes shown here to have nicotine-related expression; black text: putative genes.

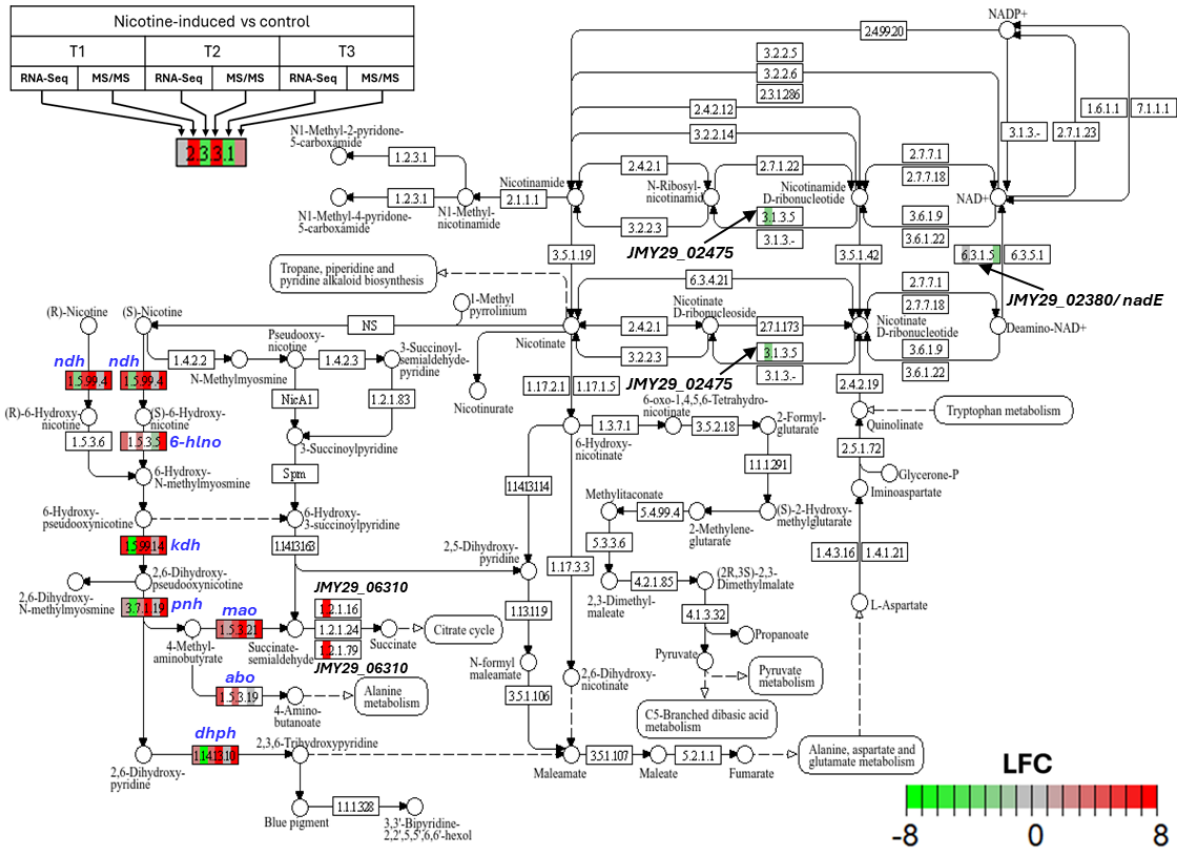

Figure 4.5. The nicotine degradation pathway encoded by pAO1 in *Paenarthrobacter nicotinovorans* ATCC 49919[15].

Overview of the differentially expressed (p-adj < 0.1, |LFC| > 1) genes and proteins related to nicotinate and nicotinamide metabolism (KEGG map pnv00760) during the growth of *P. nicotinovorans* ATCC 49919 on nicotine-supplemented medium as revealed by the transcriptomic and proteomic data; RNA-seq: results from RNA-sequencing data; MS/MS: results from nanoLC-MS/MS data; LFC: log2FoldChange; blue text: genes with known functions in nicotine degradation and previously shown to have nicotine-related expression; black text: genes shown in this study to have nicotine-related expression[15]; image generated according to the thesis subchapter 3.2.4.2.

4.3.2 Nicotine degradation pathway integration into the general cellular metabolism

It was previously suggested that α-keto-glutarate, one of the main final products of nicotine catabolism in *P. nicotinovorans* ATCC 49919, is integrated into the general pathways of the cell via an altered TCA cycle[13]. Consequently, it was postulated that if the bacteria have nicotine available as a carbon source, the additional import of citrate is not of major necessity for cellular growth and metabolism[13]. This is further suggested and supported at the RNA level, by both the chromosomal and pAO1-encoded genes[15] (Figure 4.6.).

JMY29_20550 is a *nic*-gene that encodes a hypothetical protein of the fumaryl-acetoacetate hydrolase family (InterPro Pfam ID: PF01557). In *E. coli* W, an enzyme in the same family catalyses a decarboxylation followed by an isomerization reaction involved in the metabolism of the aromatic compound 4-hydroxyphenylacetic acid. The mRNA levels of *JMY29_20550* were over nine times higher (LFC > 3.4, p-adj < 0.001) when nicotine was present in the growth medium (Figure 4.3. – A.) and were markedly lower (LFC < -6.4, p-adj < 0.001) after nicotine consumption by the induced cells at T3 compared to T1 and T2 (Figure 4.2.). An over four-fold downregulation of *JMY29_20550* (LFC < -2, p-adj < 0.05) was observed when comparing T3 to T2 untreated samples. Given these observations and its strong upregulation in *P. nicotinovorans* ATCC 49919 in the presence of nicotine as an additional carbon source to citrate, it is very likely that the enzyme encoded by *JMY29_20550* has similar roles to its *E. coli* homolog, directly converting α-keto-glutarate into succinic and formic acid, thus connecting nicotine metabolism to the altered TCA cycle reported previously[13,15].

Furthermore, the CitMHS family transporter of citrate , encoded by the chromosomal *JMY29_06775*, is downregulated (LFC < -1.9, p-adj < 0.05) in the stationary phase in both control and nicotine-induced cells. Nonetheless, the expression levels of *JMY29_06775* are remarkably low (LFC = -4.14, p-adj < 0.005) in the stationary phase nicotine-induced cells compared to the untreated bacteria. This is most probably due to nicotine byproducts such as succinic acid and a-keto-glutarate, which persist in the stationary phase and continue to fuel the TCA cycle even after the depletion of nicotine from the growth medium[15] (Figure 4.6.).

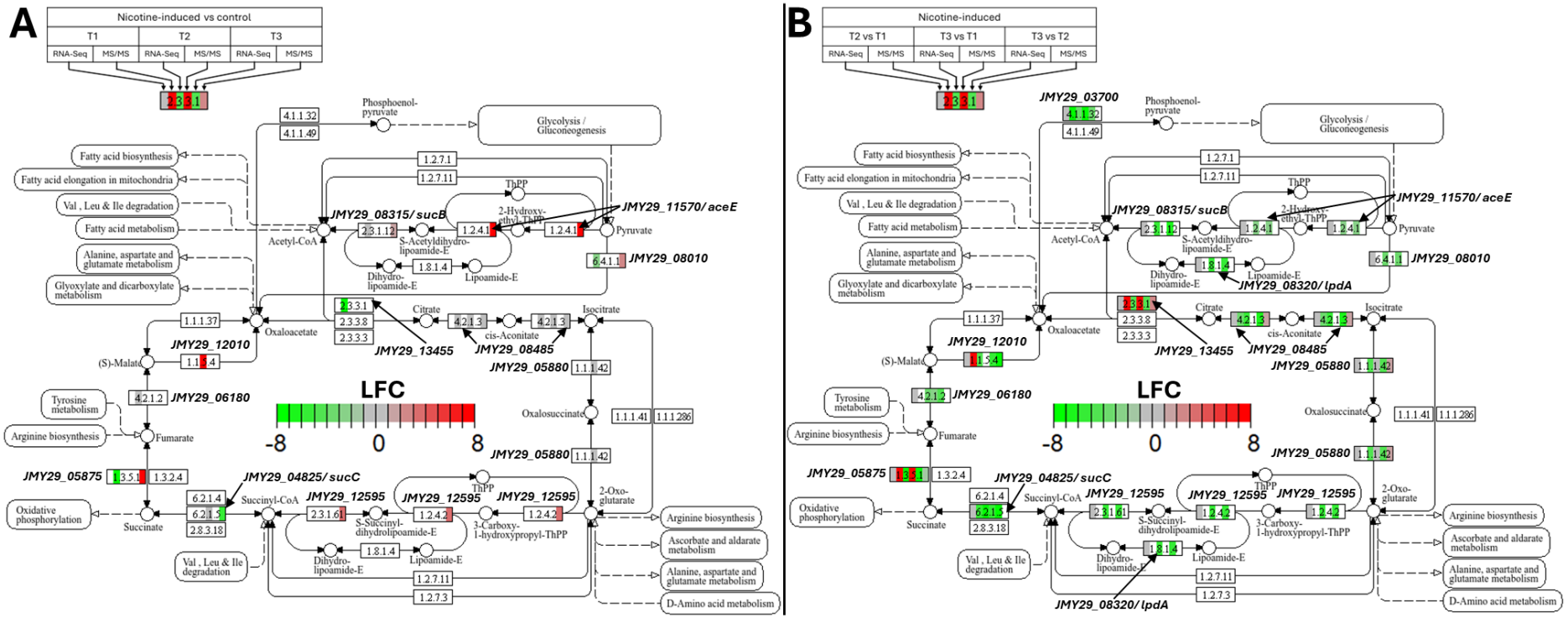

Figure 4.6. Overview of the citrate cycle-related differentially expressed genes (DEGs) and proteins (DEPs) during the growth of nicotine-induced *P. nicotinovorans* ATCC 49919, as revealed by the transcriptomic and proteomic data[15].

A. DEGs and DEPs in the nicotine-induced versus control cells; **B.** time course analysis of DEGs and DEPs in the nicotine-induced cells; RNA-seq: results from RNA-sequencing data; MS/MS: results from nanoLC-MS/MS data; LFC: log2FoldChange; blue text: genes with known functions in nicotine degradation and previously shown to have nicotine-related expression; black text: genes shown in this study to have nicotine-related expression[15]; image generated according to the thesis subchapter 3.2.4.2.

CONCLUSIONS

Considering the importance of developing efficient methods for tobacco waste decontamination and the bioremediation of natural resources polluted by the toxic alkaloid, this study aimed to expand the knowledge of nicotine catabolism in *Paenarthrobacter nicotinovorans* ATCC 49919, the microorganism of reference for the compound's degradation via the pyridine pathway. To complement the already available biochemical, biomolecular and proteomic data on the studied nicotine-degrading microorganism (NDM), an approach based on recent developments in next-generation sequencing technologies was proposed. Hence, the main objectives were derived from and hierarchised following the central dogma of molecular biology and are represented by the generation and analysis of the complete and annotated genome of *P. nicotinovorans* ATCC 49919, sequencing and analysis of its nicotine-related transcriptome, re-analysis of a pre-existing proteomic dataset using the newly assembled genome and, finally, integration of the complementary omics data.

Following a critical review of the latest developments in sequencing technologies and considering the particularities of the studied species, a hybrid approach was chosen for the genome assembly, based on the consensus of the specialised community that this would most likely ensure the desired quality of the resulting genome. In addition, the genome of a derivative strain which has lost its nicotine-degrading capacity, *P. nicotinovorans nic-*, was also sequenced, assembled and analysed.

According to the FAIR principles, the raw genomic DNA sequencing data and the assembled and annotated genomes of *P. nicotinovorans* ATCC 49919 and *nic-* were shared in public reference databases and can be accessed and used by any interested individual. The genome of *P. nicotinovorans* ATCC 49919 assembled using the hybrid *de novo* approach represents the first complete and structurally and functionally annotated sequence of a microorganism capable of nicotine degradation via the pyridine pathway[18].

In addition, the outcome of the first objective allowed several analyses that contributed to the impact of this study. Firstly, the genome allowed the expansion of the limited pre-existing knowledge on the origin and evolution of the *nic* catabolic genes. The use of the complete genome sequence of *P. nicotinovorans* ATCC 49919 as a reference for comparative phylogenetic and phylogenomic analyses against the genomes of 63 species which are related to or share a common set of five *nic* genes was described[19]. Two of the *P. nicotinovorans* sequences (DSM 420 and JCM 3874) from NCBI Genome were shown to be

redundant, while others are not representatives of the *P. nicotinovorans* species since they are not equipped with the genes required for nicotine catabolism, characteristic essential for being classified as "nicotinovorans". Furthermore, a series of replicons belonging to strains related to *P. nicotinovorans* ATCC 49919 and displaying a conserved gene order within the *nic* catabolic cluster were identified[19].

Secondly, the complete genome obtained facilitated the successful execution of the second main objective of the study, represented by the sequencing and analysis of the *P. nicotinovorans* ATCC 49919 transcriptome at three key moments of nicotine degradation. The transcriptomic dataset generated in this study is of high quality, an important feature especially considering that it is the very first to describe microbial nicotine catabolism via the pyridine pathway. In addition, the data produced are among the first to be generated using nanopore technology applied for direct sequencing of bacterial RNA, especially using the Flongle as the sequencing cell.

The raw transcriptomic data interpreted in nucleotides and reads, plus the key files used in the differential gene expression analysis between the key time points of nicotine catabolism were submitted to the relevant database, NCBI GEO, and are freely available. These represent an important source of information on bacterial nicotine catabolism, but also a resource for future studies in the field of transcriptomics or even microbial epitranscriptomics, when the bioinformatic methods are sufficiently developed to allow this type of analysis on prokaryotes.

The results obtained from the DGE analysis of *P. nicotinovorans* ATCC 49919 induced by nicotine at three key time points are consistent with and confirm those previously revealed by the proteomic analysis of the species of interest grown on citrate medium supplemented with nicotine[13]. Furthermore, the direct RNA sequencing data significantly complemented the biochemical, genomic and proteomic information generated previously or reported here for strain ATCC 49919. The successful achievement of the first two objectives allowed the fulfilment of the final one, namely the integration of the original genomic and transcriptomic data with a pre-existing proteomic dataset.

The novel transcriptomic data and the reinterpreted proteome of *P. nicotinovorans* ATCC 49919 using the newly assembled genome confirmed the nicotine-related expression of 25 *nic*-genes and their products. Of these, eight were reported here first as having nicotine-related expression: *JMY29_20530*, *JMY29_20550*, *perm*, *coxF*, *coxE*, *JMY29_20640*, *JMY29_20655* and *modB*. Analysis of the active genes and their corresponding protein levels provided insights into the integration of nicotine catabolism with the overall metabolism of

the bacterial cell and revealed a repertoire of multiple defence mechanisms against the oxidative stress generated during the degradation of the toxic alkaloid. The expression patterns of genes and proteins following nicotine treatment, as indicated by the transcriptomic and proteomic analyses, were corroborated with functional information from close sequence homologues, allowing the identification of several new putative genes involved in key processes of nicotine degradation. It was proposed that *JMY29_05475* putatively encodes a blue pigment exporter of the EamA transporter family, *JMY29_04550* and *JMY29_04555* were postulated to be involved in controlling nicotine transport, and *JMY29_20550* in linking nicotine metabolism to the modified tricarboxylic acid cycle[15].

By combining nanoLC-MS/MS-based proteomic data with data generated by short-read DNA, long-read DNA and direct RNA sequencing, this study provides the first multi-omic investigation of *Paenarthrobacter nicotinovorans* ATCC 49919. Even more, this work represents the very first multi-omic analysis of a bacterial nicotine degradation pathway[15].

PERSPECTIVES FOR FURTHER STUDY

In the script included in Appendix 4 of the thesis and available in the public domain[15], written for the functional analysis of the transcriptomic and proteomic data, the taxon identifier (tax_id) can be replaced with the NCBI identifier of any species of interest. Therefore, with minimal modifications, the modular script[15] can be used both for the generation of a local database for any species with information available in NCBI, and for the functional enrichment analysis of differentially expressed genes or proteins generated by another study. This script can be developed into an automated pipeline for processing DGE analysis results for functional enrichment, of specific relevance for non-model species for which annotation packages are most likely not available in the public domain.

In addition to being amongst the best characterised nicotine degradation pathways, nicotine catabolism by *P. nicotinovorans* ATCC 49919 also has biotechnological potential, as was highlighted in the ***Economic and ecological motivation*** and ***Medical motivation***. Currently, efforts are proactively focused on the development of CRISPR-based plasmids and tools for engineering *P. nicotinovorans* ATCC 49919 to improve the yield of nicotine degradation and its conversion efficiency, with promising results being already available[15].

The findings reported by this multi-omic study reveal new avenues for genetic and metabolic engineering to further improve the strain to increase the nicotine yield and conversion rates. With this aim, the main targets are *JMY29_04550* and *JMY29_04555* - putatively involved in controlling the transmembrane transport of nicotine, *JMY29_05475* - postulated to be involved in exporting the nicotine blue pigment, plus *JMY29_14770* which encodes a catalase. Hypothetically, the overexpression of these genes would increase the rate of nicotine uptake and export of metabolic derivatives and concomitantly reduce intracellular oxidative stress. Therefore, nicotine tolerance and its degradation rate would be simultaneously improved[15].

Furthermore, the eight *nic*-genes reported for the first time in this study as having nicotine-related expression are possible targets for knock-out experiments to disrupt the catabolic process and isolate metabolic intermediates of economic interest. This approach has proven feasible in other NDMs. In addition, even though engineering the entire cellular metabolism is a more complex task, it was already shown to be feasible in one NDM, namely in *Pseudomonas* sp. JY-Q. The potential main targets for this approach are the *nic*-gene *JMY29_20550*, identified as a key enzyme involved in linking the nicotine degradation pathway to the modified tricarboxylic acid cycle, as well as the four chromosome-encoded

transcription factors with nicotine-related expression (*JMY29_08370*, *JMY29_09095*, *JMY29_04550*, *JMY29_04555*)[15].

The findings based on the transcriptomic analysis results require further validation using methods which are alternative and complementary to direct RNA sequencing, especially for the putative genes indicated to be differentially expressed but whose products were not identified in the proteomic dataset. Exploratory studies are particularly required to assess the impact of these potential target genes for the genetic and metabolic engineering of *P. nicotinovorans* ATCC 49919 with the aim of developing its potential biotechnological applications. If efforts to genetically engineer nicotine degradation via the pyridine pathway in *P. nicotinovorans* ATCC 49919 will not yield the desired results, an alternative strategy that has potential is transferring the pAO1 megaplasmid into strains such as *Pseudomonas putida* S16 or J5 and *Agrobacterium tumefaciens* S33. It has been shown or suggested that these strains utilise the same molybdenum compound essential for nicotine degradation as *P. nicotinovorans* ATCC 49919, a factor that limits the transfer of pAO1 in other species used in genetic and metabolic engineering, such as *E. coli*.

DISSEMINATION OF THESIS RESULTS

List of publications

1. **El-Sabeh A**, Mlesnita AM, Mihasan M. Integrated transcriptomic and proteomic analysis of nicotine metabolism in *Paenarthrobacter nicotinovorans* ATCC 49919. International Biodeterioration & Biodegradation. Volume 199. 2025. DOI: 10.1016/j.ibiod.2025.106017 – **IF 4.1**, **Q2**
2. **El-Sabeh A**, Mlesnita AM, Munteanu IT, Honceriu I, Kallabi F, Boiangiu RS, Mihasan M. Characterisation of the *Paenarthrobacter nicotinovorans* ATCC 49919 genome and identification of several strains harbouring a highly syntenic *nic*-genes cluster. BMC Genomics. 24, 536. 2023. DOI: 10.1186/s12864-023-09644-3 **– IF 3.5**, **Q1**
3. **El-Sabeh A**, Honceriu I, Kallabi F, Boiangiu RS, Mihasan M. Complete genome sequences of two closely related *Paenarthrobacter nicotinovorans* strains. Microbiology Resource Announcements. 16;11(6). 2022. DOI: 10.1128/mra.00133-22 **– IF 0.7**, **Q3**

Contributions to international reference databases and software development platforms

In alignment to the FAIR principles adopted in the omic sciences, the original data and scripts generated and used in this study were deposited in international reference platforms, can be used integrally or partially, and are modifiable by anyone.

1. The raw data generated from sequencing the genomes of *Paenarthrobacter nicotinovorans* strains ATCC 49919 and *nic*- were deposited in NCBI's Sequence Read Archive (SRA) under the accession numbers in Tabel 4.3. - column "Nr. Access SRA" of the thesis.
2. The complete genome assemblies were deposited in NCBI Genome with the following accession numbers:
   - *Paenarthrobacter nicotinovorans* ATCC 49919: ASM2191934v1, correlated with GenBank assembly GCA_021919345.1 and RefSeq assembly GCF_021919345.1, containing the:
     - chromosome sequence: GenBank CP089293.1, RefSeq: NZ_CP089293.1
     - pAO1 megaplasmid sequence: GenBank CP089294.1, RefSeq: NZ_ CP089294.1
   - *Paenarthrobacter nicotinovorans nic-*: ASM2191800v1, correlated with GenBank assembly GCA_021918005.1 and RefSeq assembly GCF_021918005.1; it contains:
     - the chromosome sequence: GenBank CP089515.1, RefSeq NZ_CP089515.1

3. Direct RNA sequencing data from *Paenarthrobacter nicotinovora*ns ATCC 49919 and the key files generated for and from the differential gene expression analysis were deposited in NCBI GEO under accession numbers GSE240220 and GSE279035.
4. The nanoLC-MS/MS dataset re-analysed using on the newly assembled genome was deposited in the ProteomeXchange Consortium through the partner repository PRIDE, with the identification number PXD057439.
5. The Python and R scripts, plus the R log file, can be accessed in Zenodo at DOI: 10.5281/zenodo.14035581, and in GitHub at github.com/amadaes/Pnic_multiomics.

Team member in research projects

2021 - December 2023 – Project „Sequencing the genome of a useful bacterium: *Paenarthrobacter nicotinovorans* - the next step in expanding its biotechnological applications", acronym PaeNicGenom, grant contract no. PCE 152/ 2021, code PN-III-P4-ID-PCE-2020-0656
(cercetare.bio.uaic.ro/grupuri/bioactive/content/grants/PCE2021/PCE2021_objectives.html)

International scholarships received

1. FEBS–YSF grant to attend the 23rd FEBS Young Scientists' Forum (YSF) 2024 and the 48th FEBS Congress
2. Bursary to attend the 2nd International Molecular Biosciences PhD & Postdoc FEBS-IUBMB-ENABLE Conference
3. Young Scientist Bursary to attend the 47th FEBS Congress
4. Young Scientist Bursary to attend the 25th IUBMB-46th FEBS-15th PABMB Congress

List of published abstracts

1. **El-Sabeh A**, Mihasan M, A multi-omic study of nicotine catabolism in *Paenarthrobacter nicotinovorans*, FEBS Open Bio Volume 14, Issue S2 Supplement: Mining biochemistry for human health and well-being, 48th FEBS Congress, 29 June – 3 July 2024, Milano, Italy, June 2024, pg. 83 (SpT-07-1).
   DOI: 10.1002/2211-5463.13836
2. **El-Sabeh A**, Mihasan M, Transcriptome analysis of genes and metabolic pathways associated with nicotine degradation in *Paenarthrobacter nicotinovorans* ATCC 49919, Accelerating human and environmental knowledge discovery through bioinformatics advances. 2nd RoBioinfo Conference 2024, 2024, Poster Session, Editura Universitara.
   DOI: 10.5682/9786062817954

3. **El-Sabeh A**, Mihasan M, Transcriptomic study of nicotine catabolism in *Paenarthrobacter nicotinovorans* ATCC 49919 using direct RNA sequencing, Abstract Book - Annual Scientific Session of Naturalist Students (SSASN), 7th edition, October 26, 2023 and Scientific Session of the Faculty of Biology (SSFB), New Trends in Biology: from Molecules to Complex Systems, 3rd edition, October 27-28, 2023, Journal of Experimental and Molecular Biology (2023) 24(4):215-261, pg. 242.
   DOI: 10.47743/jemb-2023-161
4. **El-Sabeh A**, Mihasan M, Transcriptomic study of nicotine catabolism in *Paenarthrobacter nicotinovorans* ATCC 49919 using long-read direct RNA sequencing, FEBS Open Bio, Volume 14: Supplement: FEBS-IUBMB-ENABLE 2nd International Molecular Biosciences PhD and Postdoc Conference, The emerging challenge, Environmental impacts on human health, 23–25 November 2023, University of Cologne, Germany, May 2024, pg. 14 (ST007).
   DOI: 10.1002/2211-5463.13792
5. **El-Sabeh A**, Mihasan M, Investigation of bacterial nicotine-metabolism using direct RNA sequencing on a MinION Flongle, 2023, FEBS Open Bio, Volume 13: Supplement: Together in bioscience for a better future, 47th FEBS Congress, July 8–12, 2023, Tours, France, July 2023, pg. 146 (P-04.1-16).
   DOI: 10.1002/2211-5463.13646
6. Mihasan M, **El-Sabeh A**, A highly syntenic *nic*-genes cluster is present in several related bacterial strains, 2023, Posters, FEBS Open Bio, Volume 13: Supplement: Together in bioscience for a better future, 47th FEBS Congress, July 8–12, 2023, Tours, France, July 2023, pg. 123 (P-03.1-12).
   DOI: 10.1002/2211-5463.13646
7. **El-Sabeh A**, Mihasan M, Investigation of nicotine-catabolism in *Paenarthrobacter nicotinovorans* ATCC 49919 using long-read direct RNA sequencing, Human Genomics and Microbiome in the Context of Data Federation: RoBioinfo Conference 2023, Editura Universitară, Editia I, 2023.
   DOI: 10.5682/9786062816179
8. **El-Sabeh A**, Honceriu I, Kallabi F, Boiangiu RS, Mihășan M, The Complete Genome Assembly of the Nicotine-Degrading *Paenarthrobacter nicotinovorans* ATCC 49919. New trends in Biology: from molecules to complex systems, Journal of Experimental and Molecular Biology, 23(2):1-69, pg. 46.
   DOI:10.47743/jemb-2022-23-2

9. **El-Sabeh A**, Honceriu I, Kallabi F, Boiangiu RS, Mihășan M, Secvența completă a genomurilor a două tulpini strâns înrudite de *Paenarthrobacter nicotinovorans*. Volum de rezumate: Sesiunea științifică anuală a studenților naturaliști – ediția VI-a., 2022 (Iași – Universitatea „Alexandru Ioan Cuza"), pg. 13.
   www.bio.uaic.ro/?page_id=12649
10. **El-Sabeh A**, Honceriu I, Kallabi F, Boiangiu RS, Mihasan M, Genome sequence of *Paenarthrobacter nicotinovorans* strain ATCC 49919, a nicotine-degrading soil microorganism, FEBS Open Bio, Volume 12: Supplement: The Biochemistry Global Summit, 25th IUBMB Congress, 46th FEBS Congress, 15th PABMB Congress, July 9–14, 2022, Lisbon, Portugal, July 2022, pg. 17 (ShT-06.1–1).
    DOI: 10.1002/2211-5463.13442
11. **El-Sabeh A**, Honceriu I, Kallabi F, Boiangiu RS, Mihasan M, Complete genome sequence of the well-studied nicotine-degrading microorganism *Paenarthrobacter nicotinovorans* strain ATCC 49919, FEMS Conference on Microbiology: 30 June - 2 July 2022, Belgrade, Serbia, Serbian Society of Microbiology, 2022: Electronic Abstract Book, ISBN 8691489782.
    www.femsbelgrade2022.org/abstract-book

Communications at international conferences and awards received

Oral communications

1. 29.06 - 03.07.2024 – The 48th FEBS Congress, Milan, Italy
   El-Sabeh A, Mihasan M - A multi-omic study of nicotine catabolism in *Paenarthrobacter nicotinovorans*
2. 23 - 25.11.2023 – The 2nd International Molecular Biosciences PhD & Postdoc FEBS-IUBMB-ENABLE Conference, Cologne, Germany
   El-Sabeh A, Mihasan M - Transcriptomic study of nicotine catabolism in *Paenarthrobacter nicotinovorans* ATCC 49919 using long-read direct RNA sequencing
   – awarded one of the two prizes for Short Talk presentations (15 minutes) in the "Novel Experimental Model Systems" session
3. 13 - 15.09.2023 – The Annual International Conference of the Romanian Society of Biochemistry and Molecular Biology, Cluj-Napoca, Romania
   El-Sabeh A, Mihasan M - Long-read direct RNA nanopore sequencing of the nicotine-related transcriptome of *Paenarthrobacter nicotinovorans* ATCC 49919
   – awarded the 2nd prize in the Young Researchers session

4. 11 - 13.05.2023 – The RoBioinfo Conference 2023, Bucharest, Romania

El-Sabeh A, Mihasan M - Investigation of nicotine-catabolism in *Paenarthrobacter nicotinovorans* ATCC 49919 using long-read direct RNA sequencing

5. 21 - 24.10.2022 – International Conference and XXXIX Scientific Sessions of the Romanian Society for Cell Biology (RSCB 2022), Cluj Napoca, Romania

El-Sabeh A, Honceriu I, Kallabi F, Boiangiu RS, Mihasan M - Complete genome sequence of the nicotine-degrading soil microorganism *Paenarthrobacter nicotinovorans* ATCC 49919

6. 09 - 14.07.2022 – The Biochemistry Global Summit 2022: 25th IUBMB-46th FEBS-15th PABMB, Lisbon, Portugal

El-Sabeh A, Honceriu I, Kallabi F, Boiangiu RS, Mihasan M - Genome sequence of *Paenarthrobacter nicotinovorans* strain ATCC 49919, a nicotine-degrading soil microorganism

Poster communications

7. 25 - 27.09.2024 – The Annual International Conference of the Romanian Society of Biochemistry and Molecular Biology, Bucharest, Romania

El-Sabeh A, Mihasan M - A multiomic analysis of nicotine degradation via the pyridine pathway in *Paenarthrobacter nicotinovorans* ATCC 49919

– **awarded one of the three poster session prizes**

8. 29.06 - 03.07.2024 – The 48th FEBS Congress, Milan, Italy

El-Sabeh A, Mihasan M - A multi-omic study of nicotine catabolism in *Paenarthrobacter nicotinovorans*

9. 26.06 – 29.09.2024 – The 23rd FEBS Young Scientists' Forum (YSF 2024), Pavia, Italy

El-Sabeh A, Mihasan M - A multi-omic study of nicotine catabolism in *Paenarthrobacter nicotinovorans*

10. 11 - 13.04.2024 – The 2nd RoBioinfo Conference, Cluj-Napoca, Romania

El-Sabeh A, Mihasan M - Transcriptome analysis of genes and metabolic pathways associated with nicotine degradation in *Paenarthrobacter nicotinovorans* ATCC 49919

– **awarded one of the three poster session prizes**

11. 08 - 12.07.2023 – The 47th FEBS Congress, Tours, France

El-Sabeh A, Mihasan M - Investigation of bacterial nicotine-metabolism using direct RNA sequencing on a MinION Flongle

12. 30.06 - 02.07.2022 – FEMS Conference on Microbiology, Belgrade, Serbia

El-Sabeh A, Honceriu I, Kallabi F, Boiangiu RS, Mihasan M - Complete genome sequence of the well-studied nicotine-degrading microorganism *Paenarthrobacter nicotinovorans* strain ATCC 49919

Communications at national conferences and awards received

Oral communications

13. 27 - 28.10.2023 – Scientific Session of the Faculty of Biology - Trends in Biology: from molecules to complex systems, organised by the "Alexandru Ioan Cuza" University of Iași, Romania

El-Sabeh A, Mihasan M - Transcriptomic study of nicotine catabolism in *Paenarthrobacter nicotinovorans* ATCC 49919 using direct RNA sequencing

14. 27 - 28.10.2022 – Scientific Session of the Faculty of Biology - Trends in Biology: from molecules to complex systems, organised by the "Alexandru Ioan Cuza" University of Iași, Romania

El-Sabeh A, Honceriu I, Kallabi F, Boiangiu RS, Mihasan M - The complete genome assembly of the nicotine degrading *Paenarthrobacter nicotinovorans* ATCC 49919

15. 21.05.2022 – The Annual Scientific Session of the Naturalist Students, organised by the Faculty of Biology of the "Alexandru Ioan Cuza" University of Iași, Romania

El-Sabeh A, Honceriu I, Kallabi F, Boiangiu RS, Mihasan M - The complete genome sequence of two closely related *Paenarthrobacter nicotinovorans* strains

Poster communications

16. 27 - 28.10.2024 – Scientific Session of the Faculty of Biology - Trends in Biology: from molecules to complex systems, organised by the "Alexandru Ioan Cuza" University of Iași, Romania

El-Sabeh A, Mihasan M – A multiomic study of nicotine catabolism in *Paenarthrobacter nicotinovorans* ATCC 49919

– **awarded one of three poster session prizes**

SELECTED BIBLIOGRAPHY

1. Lam J, Schneider J, Shadbegian R, Pega F, St Claire S, Novotny T. Modelling the global economic costs of tobacco product waste. Bull World Health Organ. 2022 Oct 1;100(10):620–7. DOI: 10.2471/BLT.22.288344
2. Zhang LY, Mai J, Shi JF, Ai KB, He L, Zhu MJ, et al. Study on tobacco quality improvement and bacterial community succession during microbial co-fermentation. Ind Crops Prod. 2024 Feb 1;208:117889. DOI: 10.1016/J.INDCROP.2023.117889
3. Deng M, Basak P, Zhang Y, Song J, Suo H. An update in recent research on nicotine contamination and nicotine-degrading microorganisms. Toxicon. 2025 Jan;254:108209. DOI: 10.1016/j.toxicon.2024.108209
4. Mihăşan M. Megaplasmidul pAO1-Structură şi funcţie. Iași, Romania: Editura Universităţii "Alexandru Ioan Cuza" din Iaşi; 2011.
5. Mihaşan M. Bioinformatics-Based Molecular Classification of Arthrobacter Plasmids. Cell Mol Biol Lett. 2015 Dec 1;20(4):612–25. DOI: 10.1515/cmble-2015-0036
6. Demain AL, Sanchez S. Microbial drug discovery: 80 years of progress. J Antibiot (Tokyo). 2009 Jan;62(1):5–16. DOI: 10.1038/ja.2008.16
7. Boiangiu RS, Mihasan M, Gorgan DL, Stache BA, Hritcu L. Anxiolytic, promnesic, anti-acetylcholinesterase and antioxidant effects of cotinine and 6-hydroxy-l-nicotine in scopolamine-induced zebrafish (Danio rerio) model of alzheimer's disease. Antioxidants. 2021 Feb 1;10(2):1–28. DOI: 10.3390/antiox10020212
8. Beer B, Pick A, Sieber V. In vitro metabolic engineering for the production of α-ketoglutarate. Metab Eng. 2017;40:5–13. DOI: 10.1016/j.ymben.2017.02.011
9. Igloi GL, Brandsch R. Sequence of the 165-kilobase catabolic plasmid pAO1 from Arthrobacter nicotinovorans and identification of a pAO1-dependent nicotine uptake system. J Bacteriol. 2003 Mar;185(6):1976–86. DOI: 10.1128/JB.185.6.1976-1986.2003
10. Jerke K, Nakatsu CH, Beasley F, Konopka A. Comparative analysis of eight Arthrobacter plasmids. Plasmid. 2008 Mar 1;59(2):73–85. DOI: 10.1016/J.PLASMID.2007.12.003
11. Mihăşan M, Babii C, Aslebagh R, Channaveerappa D, Dupree E, Darie CC. Proteomics based analysis of the nicotine catabolism in Paenarthrobacter nicotinovorans pAO1. Sci Rep. 2018 Nov 2;8(1):1–9. DOI: 10.1038/s41598-018-34687-y
12. Mihăşan M, Babii C, Aslebagh R, Channaveerappa D, Dupree EJ, Darie CC. Exploration of Nicotine Metabolism in Paenarthrobacter nicotinovorans pAO1 by Microbial Proteomics. Adv Exp Med Biol. 2019;1140:515–29. DOI: 10.1007/978-3-030-15950-4_30
13. Mihăşan M, Boiangiu RŞ, Guzun D, Babii C, Aslebagh R, Channaveerappa D, et al. Time-Dependent Analysis of Paenarthrobacter nicotinovorans pAO1 Nicotine-Related Proteome. ACS omega. 2021 Jun 8;6(22):14242–51. DOI: 10.1021/ACSOMEGA.1C01020
14. Zhang Z, Mei X, He Z, Xie X, Yang Y, Mei C, et al. Nicotine metabolism pathway in bacteria: mechanism, modification, and application. Appl Microbiol Biotechnol. 2022 Feb 1;106(3):889–904.
15. El-Sabeh A, Mlesnita AM, Mihasan M. Integrated transcriptomic and proteomic analysis of nicotine metabolism in Paenarthrobacter nicotinovorans ATCC 49919. Int Biodeterior Biodegradation. 2025 Mar;199:106017. DOI: 10.1016/j.ibiod.2025.106017
16. Brandsch R. Microbiology and biochemistry of nicotine degradation. Appl Microbiol Biotechnol. 2006 Jan 7;69(5):493–8. DOI: 10.1007/S00253-005-0226-0/FIGURES/2
17. Fitzpatrick PF. The enzymes of microbial nicotine metabolism. Beilstein J Org Chem. 2018 Aug 31;14:2295. DOI: 10.3762/BJOC.14.204
18. El-Sabeh A, Honceriu I, Kallabi F, Boiangiu RS, Mihasan M. Complete Genome Sequences of Two Closely Related Paenarthrobacter nicotinovorans Strains. Microbiol Resour Announc. 2022 Jun 16;11(6). DOI: 10.1128/MRA.00133-22
19. El-Sabeh A, Mlesnita AM, Munteanu IT, Honceriu I, Kallabi F, Boiangiu RS, et al. Characterisation of the Paenarthrobacter nicotinovorans ATCC 49919 genome and identification of several strains harbouring a highly syntenic nic-genes cluster. BMC Genomics. 2023 May 4;24(1). DOI: 10.1186/s12864-023-09644-3