

RESEARCH

Open Access



Characterisation of the *Paenarthrobacter nicotinovorans* ATCC 49919 genome and identification of several strains harbouring a highly syntenic *nic*-genes cluster

Amada El-Sabeh¹, Andreea-Mihaela Mlesnita¹, Iustin-Tiberius Munteanu¹, Iasmina Honceriu¹, Fakhri Kallabi^{1,2}, Razvan-Stefan Boiangiu¹ and Marius Mihasan^{1*}

Abstract

Background *Paenarthrobacter nicotinovorans* ATCC 49919 uses the pyridine-pathway to degrade nicotine and could provide a renewable source of precursors from nicotine-containing waste as well as a model for studying the molecular evolution of catabolic pathways and their spread by horizontal gene transfer via soil bacterial plasmids.

Results In the present study, the strain was sequenced using the Illumina NovaSeq 6000 and Oxford Nanopore Technology (ONT) MinION platforms. Following hybrid assembly with Unicycler, the complete genome sequence of the strain was obtained and used as reference for whole-genome-based phylogeny analyses. A total of 64 related genomes were analysed; five *Arthrobacter* strains showed both digital DNA-DNA hybridization and average nucleotide identity values over the species threshold when compared to *P. nicotinovorans* ATCC 49919. Five plasmids and two contigs belonging to *Arthrobacter* and *Paenarthrobacter* strains were shown to be virtually identical with the pAO1 plasmid of *Paenarthrobacter nicotinovorans* ATCC 49919. Moreover, a highly syntenic *nic*-genes cluster was identified on five plasmids, one contig and three chromosomes. The *nic*-genes cluster contains two major locally collinear blocks that appear to form a putative catabolic transposon. Although the origins of the *nic*-genes cluster and the putative transposon still elude us, we hypothesise here that the ATCC 49919 strain most probably evolved from *Paenarthrobacter* sp. YJN-D or a very closely related strain by acquiring the pAO1 megaplasmid and the nicotine degradation pathway.

Conclusions The data presented here offers another snapshot into the evolution of plasmids harboured by *Arthrobacter* and *Paenarthrobacter* species and their role in the spread of metabolic traits by horizontal gene transfer among related soil bacteria.

Keywords Paenarthrobacter, Nicotine metabolism, Evolution, Plasmids, Catabolic transposon

Background

Paenarthrobacter nicotinovorans ATCC 49919 is an aerobic Gram-positive soil bacterium that can use nicotine as a carbon source. On media containing this alkaloid, the strain produces a characteristic blue pigment known as nicotine blue (NB or 4,4',5,5'-tetrahydroxy-3,3'-diazadiphenone-(2,2')). First referenced under

*Correspondence:

Marius Mihasan
marius.mihasan@uaic.ro

¹ Faculty of Biology, Alexandru Ioan Cuza University of Iași, Iași, Romania

² Laboratory of Human Molecular Genetics, Faculty of Medicine of Sfax, University of Sfax, Sfax, Tunisia



the name *Arthrobacter oxidans* [1, 2], the strain was sequentially renamed *Arthrobacter nicotinovorans* [3] and finally *Paenarthrobacter nicotinovorans* [4]. The pyridine-pathway used by this bacterium to degrade nicotine is historically the first, hence the most thoroughly studied bacterial catabolic pathway for this alkaloid [5, 6]. All the intermediates, steps and enzymes involved in the conversion of nicotine to NB are well established and there is available data on the nicotine pathway regulation. The genes responsible for nicotine degradation are grouped in a *nic*-genes cluster located on pAO1, a 165 kb catabolic megaplasmid [7]. The cluster is organized into several gene modules [8] responsible for the catabolism of L- and D-nicotine to α -ketoglutarate, succinate, methylamine and NB. While NB and methylamine accumulate in the growth medium, we have shown that α -ketoglutarate and succinate are integrated into the Krebs cycle and support cell growth [9, 10].

The relevance of this strain and its nicotine catabolic pathway is twofold. First, the pathway is a source of functionalized pyridine intermediates such as 6-hydroxy-L-nicotine [11–13] and could be used as a renewable source of precursors to synthesize drugs and insecticides from nicotine or nicotine-containing waste [14]. Secondly and more relevant to this work due to being so well studied, the pAO1 megaplasmid and its nicotine degradation pathway provide a model for studying the molecular evolution of catabolic pathways and their spread by horizontal gene transfer via soil bacterial plasmids. The *nic*-genes cluster from pAO1 has a lower G+C content than the megaplasmid's average and integrases belonging to the tyrosine family of recombinases have been described at the cluster's 5'-end. These two findings suggest that the *nic*-genes DNA fragment could be a catabolic transposon [8] acquired by horizontal gene transfer [15].

Because pAO1 is a conjugative plasmid, reaching transfer frequencies of 10^{-3} – 10^{-2} per donor [7], it can also serve as a model for studying the molecular evolution of conjugative plasmids transmissible between related soil bacteria. Several attempts were made to classify the available plasmids associated with *Arthrobacter*-related species and to identify a backbone of core genes responsible for major plasmid functions such as replication, partition and conjugation [16–18]. Although such a core backbone remains elusive, several advancements have been made: four different clades of plasmids were described based on the sequence of the *ParA*-like protein [18], and a Type IV secretion system (T4SS) coupled with conserved DNA-repeats were identified within a subset of plasmids believed to have evolved from a common ancestor [19].

Nowadays, there are many sequencing projects looking at the microbial diversity of soil, especially in toxic or polluted samples [20]. *Arthrobacter*-related species

have a high metabolic versatility and often constitute an important and even dominant culturable fraction of the microbial communities [21]. This has led to the accumulation of many *Arthrobacter*-related bacterial genomes. However, until 2016, within the *Arthrobacter* genus were grouped many species which were sequentially classified under 5 novel genera [4], including *Paenarthrobacter*. Therefore, classification of a given genome as belonging to the *Arthrobacter* or *Paenarthrobacter* species is dependent on when the data was deposited and if this taxonomic change was propagated in the databases containing rRNA sequences, leading to some confusion. Several comparative genomics studies dealing with *Arthrobacter* and *Paenarthrobacter* strains and focusing on identifying specific catabolic traits are available [22–29], but a genome-based phylogenetic study is missing so far.

In our previous work [30], we announced the availability of the complete genome sequence of *Paenarthrobacter nicotinovorans* ATCC 49919 and of a closely related lab strain named *nic*- lacking the pAO1 plasmid and hence unable to degrade nicotine. The current study is building upon this previously published work, further characterising the *P. nicotinovorans* ATCC 49919 complete genome and using it as reference to identify related strains containing the *nic*-genes cluster. Whole-genome-based phylogeny was performed using 64 *Paenarthrobacter*, *Arthrobacter*, *Rhodococcus* and *Nocardioideis* genomes, which led to the identification of several mislabelled strains in the database. Moreover, we showed that five plasmids and two contigs belonging to *Arthrobacter* and *Paenarthrobacter* strains are virtually identical with pAO1, and we identified a highly syntenic *nic*-genes cluster on 5 plasmids, one contig and 3 different chromosomes. Further data is provided indicating that the *nic*-genes cluster consists of a putative catabolic transposon and two additional locally colinear blocks containing accessory genes. The data reported here can be used for transcriptomics and proteomics studies of nicotine degradation in *P. nicotinovorans* ATCC 49919 and opens the way for engineering the strain for improved conversion of nicotine and nicotine-containing waste into valuable chemicals. Moreover, the complete genome provides a much-needed reference sequence for nicotine degrading microorganisms (NDMs) that use the pyridine-pathway for the catabolic process and can be used for the assembly and comparative genome analysis of other strains belonging to the *Paenarthrobacter* genus.

Results

Paenarthrobacter nicotinovorans ATCC 49919 genome features

Hybrid assembly using Unicycler [31] yielded the complete genome sequence of *Paenarthrobacter nicotinovorans*

ATCC 49919 with the general features listed in Supplementary Table 1. The genome consists of two replicons: a 4,316,184 bp circular chromosome with an overall G+C content of 63.2%, and a 165,141 bp circular megaplasmid with an overall G+C content of 59.7%. A total of 4,026 genes encoding 3,930 proteins, 23 pseudogenes, 54 tRNAs, 2 ncRNAs, 1 tmRNA, and 6 identical ribosomal operons were identified on the chromosome. In two instances, each three genes of two ribosomal operons are in proximity, being separated either by a single or three non-rRNA-related CDS, respectively. On the plasmid, a total of 145 genes encoding 138 proteins were identified.

eggNOG assigned 2421 Gene Ontology (GO) terms to 626 (15%) annotated genes, and 1,334 PFAM protein families to 3,338 (83%) of the total proteins. Two ABC-transporter related protein families make up 12% of the total number of genes with PFAMs assigned: 3-TM domain of the amino acid permease protein (IPR010065) and the ATP-binding domain of an ABC transporter-like protein (IPR003439). Clusters of Orthologous Groups (COGs) were assigned for 3,383 genes (93%), most being categorized as involved in amino acid transport (E), carbohydrate transport and metabolism (G) and transcription (Table 1).

CRISPRCasFinder reported four genes encoding putative Cas3 proteins (Class 1, type III) and 3 CRISPR arrays. CARD returned hits to 16 antibiotic classes which were tested using the disk diffusion method. The strain was shown to be resistant to nalidixic acid (30 µg/disk), ceftriaxone (10 µg/disk), neomycin (30 IU/disk), spectinomycin (10 µg/disk), and susceptible to ampicillin (10 µg/disk), benzylpenicillin (10 IU/disk), cloxacillin (10 µg/disk), cefuroxime (30 µg/disk), chloramphenicol (30 µg/disk), erythromycin (15 µg/disk), gentamicin (10 µg/disk), kanamycin (70 µg/disk), vancomycin (30 µg/disk) and tetracycline (30 µg/disk).

Other *Paenarthrobacter nicotinovorans* genomes

The NCBI Genome database also lists eight more genomes from various *P. nicotinovorans* strains at contig or scaffold level (Supplementary table 2). As shown in Fig. 1, pairwise comparisons of the *Paenarthrobacter nicotinovorans* ATCC 49919 genome against these genomes indicated digital DNA-DNA hybridization (dDDH) values over the species threshold (calculated with formula d_4) as well as average nucleotide identity (ANI) values over the proposed 96% species threshold [32] for 4 strains: *Paenarthrobacter nicotinovorans* nic-, *Paenarthrobacter nicotinovorans* JCM3874, *Paenarthrobacter*

Table 1 Distribution of COG categories in the complete genome of *P. nicotinovorans* ATCC 49919

COG Categories	Number of genes		
	Chromosome	pAO1	Total
Amino acid transport and metabolism (E)	440	11	451
Carbohydrate transport and metabolism (G)	414	10	424
Transcription (K)	402	13	415
Inorganic ion transport and metabolism (P)	236	5	241
Energy production and conversion (C)	196	11	207
Replication, recombination, and repair (L)	176	11	187
Translation, ribosomal structure, and biogenesis (J)	180	0	180
Signal transduction mechanisms (T)	167	6	173
Lipid transport and metabolism (I)	171	1	172
Cell wall/membrane/envelope biogenesis (M)	161	2	163
Coenzyme transport and metabolism (H)	136	7	143
Posttranslational modification, protein turnover, chaperones (O)	109	3	112
Secondary metabolites biosynthesis, transport, and catabolism (Q)	98	3	101
Nucleotide transport and metabolism (F)	93	1	94
Defense mechanisms (V)	74	2	76
Intracellular trafficking, secretion, and vesicular transport (U)	46	4	50
Cell cycle control, cell division, chromosome partitioning (D)	45	2	47
Cell motility (N)	11	1	12
Chromatin structure and dynamics (B)	2	0	2
Cytoskeleton (Z)	1	0	1
Function Unknown (S)	693	21	714

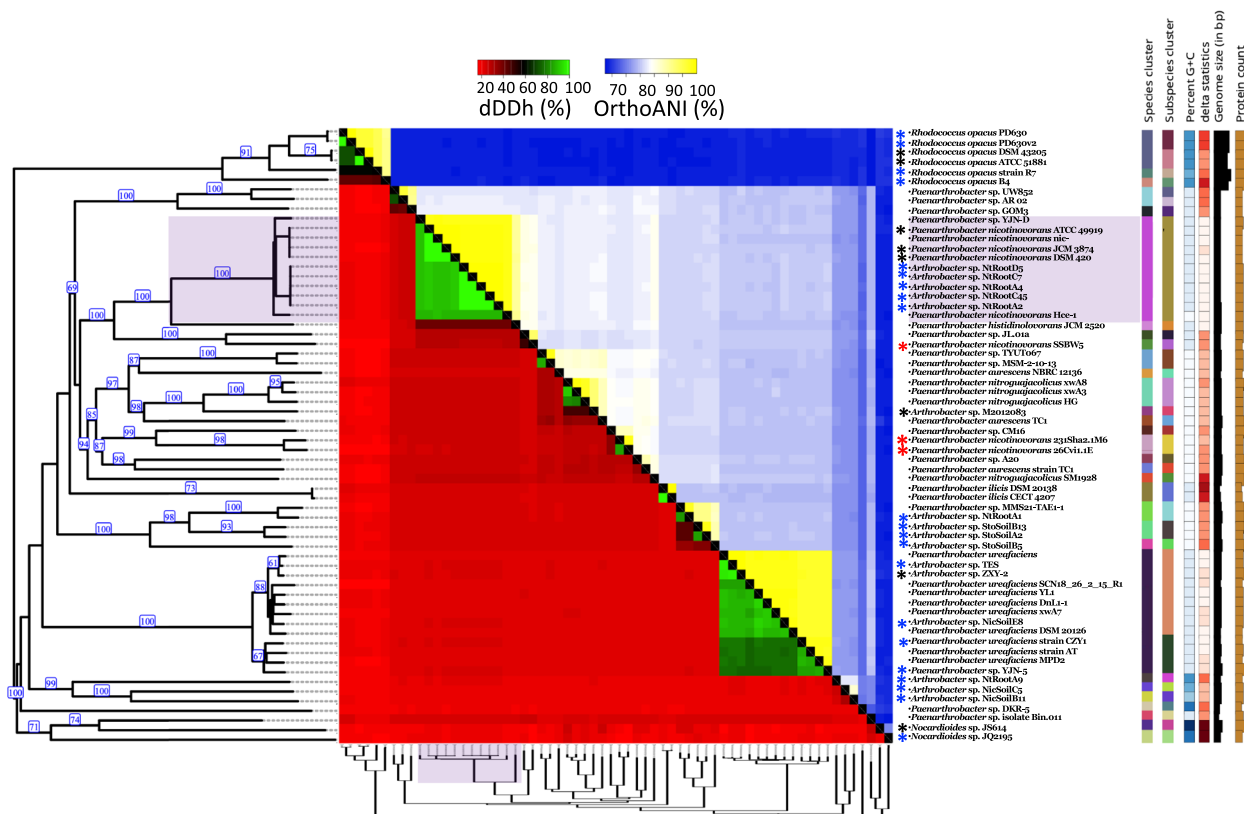


Fig. 1 Phylogenetic tree and pairwise comparisons based on both dDDH (d_d) and ANI values of 64 genomes of *Paenarthrobacter* strains and of other *Arthrobacter*, *Nocardioideis*, and *Rhodococcus* strains which possess the *nic*-genes. The tree was inferred with FastME 2.1.6.1 [33] from GBDP distances calculated from genome sequences. The branch lengths were scaled in terms of GBDP distance formula d_s . Values in blue represent GBDP pseudo-bootstrap support values > 60% from 100 replications, with an average branch support of 67.3%. The tree was rooted at the midpoint. The same tree clustering was used for both columns and rows in the heatmap. Cut-off values for species clustering was 70% for dDDH (d_d) [34] and 96% for ANI. *Paenarthrobacter* strains mislabelled as *nicotinovorans* are marked with a red asterisk. Strains previously reported to harbour *nic*-genes are marked with a black asterisk. Strains first reported here to possess the *nic*-genes are marked with a blue asterisk

nicotinovorans DSM420 and *Paenarthrobacter nicotinovorans* Hce-1. The *nic*- strain is a cured derivative from ATCC 49919 lacking the pAO1 plasmid [30]. The strains JCM 3874, DSM 420 and ATCC 49919 are designations of the same strain and, as shown in Fig. 2, the corresponding deposited genomes contain a contig practically identical with pAO1 at nucleotide level. The draft genome of *Paenarthrobacter nicotinovorans* Hce-1 is highly similar at chromosome level with *Paenarthrobacter nicotinovorans* ATCC 49919, but not one of its 103 contigs is similar with the pAO1 plasmid or contains *nic*-genes. In line with these observations, GTDB-Tk assigned these strains as the same species based both on topological placement and ANI.

The strains below the threshold limit indicated by the TYGS phylogenomic analysis and GTDB-Tk as belonging to different species are *Paenarthrobacter nicotinovorans* SSBW5, *Paenarthrobacter nicotinovorans* 231Sha2.1M6 and *Paenarthrobacter nicotinovorans* strain 26Cvi1.1E (Fig. 1). The key characteristic of *Paenarthrobacter nicotinovorans*,

i.e., the *nic*-genes, is also absent from the deposited genomes of these strains, indicating that these strains are mislabelled in the database. *Paenarthrobacter nicotinovorans* SSBW5 is more closely related to *Paenarthrobacter* sp. JL.01a (dDDH value 90%, ANI value 49%), while *Paenarthrobacter nicotinovorans* 231Sha2.1M6 and *Paenarthrobacter nicotinovorans* strain 26Cvi1.1E form a branch of their own. To remove further confusion, we propose that “*nicotinovorans*” should be dropped from the nomenclature of these three strains, and that they should be referred to as *Paenarthrobacter* sp. SSBW5, *Paenarthrobacter* sp. 231Sha2.1M6 and respectively, *Paenarthrobacter* sp. 26Cvi1.1E.

Genomes of other nicotine-degrading microorganisms using the pyridine-pathway

The pyridine-pathway for nicotine degradation is not unique to *P. nicotinovorans* ATCC 49919, having been previously described in other NDMs [6] such as: *Arthrobacter* sp. ZXY-2 [8], *Arthrobacter* sp. M2012083 [25, 35], *Arthrobacter* sp. HF-2 [36], *Rhodococcus* sp. Y22 [37], *Rhodococcus*

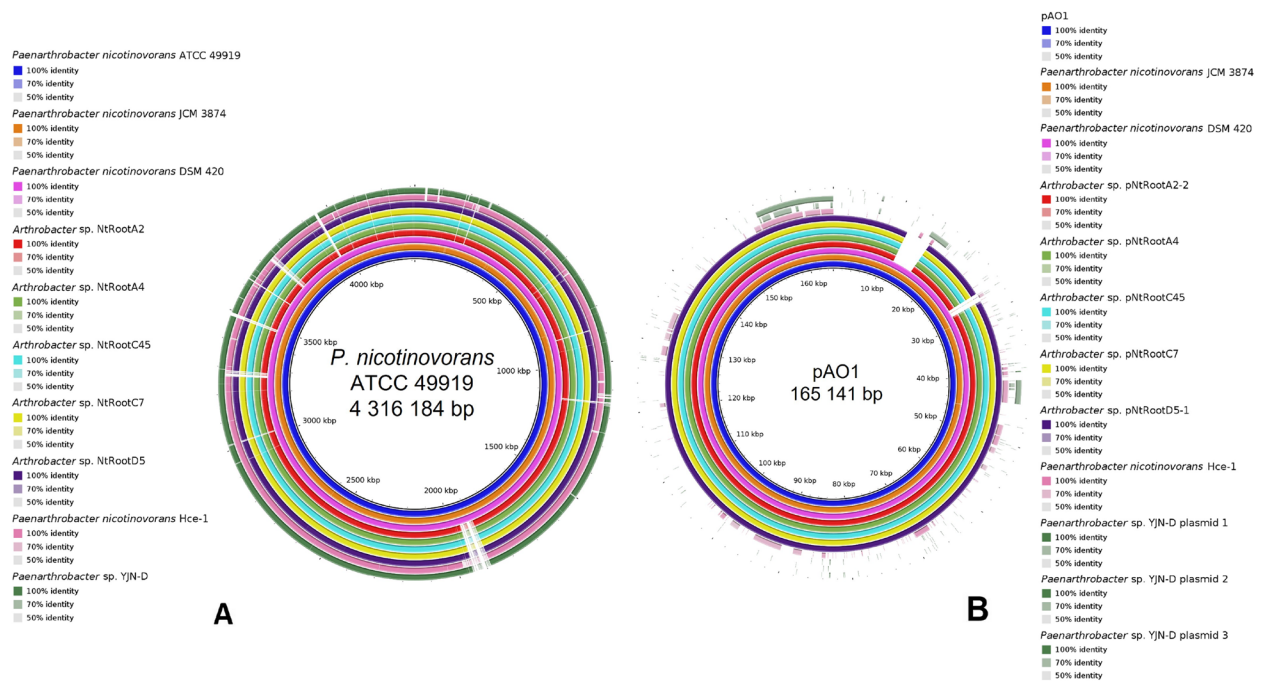


Fig. 2 BRIG comparative genomic analysis showing the high similarity of the *P. nicotinovorans* ATCC 49919 genome with nine genomes of related *Arthrobacter* and *Paenarthrobacter* strains at the (A) chromosome and (B) plasmid level. The assembly IDs used for each strain are listed in Supplementary Table 2

opacus DSMZ 43205 [38] and *Nocardioides* sp. JS614 [38]. Using the sequence of the 20 *nic*-genes experimentally related to nicotine degradation from pAO1 (bold entries in Supplementary Table 3), 21 more genomes (marked with a blue asterisk in Fig. 1 and listed in Table 2) were identified as containing at least 5 different *nic*-genes. As shown in Fig. 1, the phylogenomic analysis of the 21 strains indicated that 5 *Arthrobacter* sp. strains (NtRootD5, NtRootC7, NtRootA4, NtRootC45, NtRootA2) have both dDDH and ANI values over the species threshold when compared to *P. nicotinovorans* ATCC 49919. Moreover, all five *Arthrobacter* sp. strains contain a ~160.5 kb plasmid with high identity at the nucleotide level with pAO1 (Supplementary Table 4, pNtRootD5-1, pNtRootC7-1, pNtRootA4, pNtRootC45, and pNtRootA2-2, respectively), further suggesting that these strains actually belong to the same species as ATCC 49919 (Fig. 2, B). This was confirmed by the taxonomic assignment performed with GTDB-Tk, which also identified these strains as the same species.

Highly syntenic *nic*-genes clusters identified in nine *Arthrobacter* and *Paenarthrobacter* genomes

Using pAO1 *nic*-genes as query, BLAST searches against the NCBI Nucleotide collection database identified a minimum of five *nic* genes on 12 plasmids, 14 chromosomes and 3 contigs from draft genomes belonging to strains from *Rhodococcus*, *Nocardioides*, *Arthrobacter*, and *Paenarthrobacter*

genera. Plasmids pNtRootA2-2, pNtRootA4, pNtRootC7-1, pNtRootC45, pNtRootD5-1 as well as contig JAG-INZ01000002.1 from the draft genome of *P. nicotinovorans* strain DSM 420 and contig BMRR01000008.1 from the draft genome of *P. nicotinovorans* strain JCM 3874 harbour a *nic*-genes cluster that is identical with the pAO1 one.

On the other hand, despite having lower overall nucleotide identity with pAO1, the following share a highly syntenic *nic*-cluster: plasmids pNicSoilB11-2, pNicSoilE8-2, pTES1, pCZY, and pZXY21, the chromosomes from *Arthrobacter* sp. NicSoilC5, *Arthrobacter* sp. NtRootA9 and *Paenarthrobacter* sp. YJN-5, as well as contig NZ_AKkk01000058 from the draft genome of *Arthrobacter* sp. M2012083 (Fig. 3, A). Comparative analysis performed with progressiveMauve indicated the presence of several locally collinear blocks (LCBs). Most of the *nic*-genes are placed within two LCBs and flanked by mobile genetic elements. At one end there are either transposases related to those of the *S. aureus* Tn554 transposon [8], IS21 or IS256 family insertion sequences. At the other end the two LCBs are always flanked by the IS481 family insertion sequence (Fig. 3, B).

Discussion

The sequence of the two replicons comprising the complete genome of *P. nicotinovorans* ATCC 49919 offers a great opportunity to study the molecular evolution of the

Table 2 Genomes of strains harbouring at least 5 different *nic*-genes and their localization

Organism Name	Strain	Assembly Accession	<i>nic</i> genes found on	
<i>Paenarthrobacter nicotinovorans</i>	ATCC 49919	GCA_021919345.1	pAO1	Identical <i>nic</i> -genes cluster
<i>Paenarthrobacter nicotinovorans</i>	JCM 3874	GCA_014648735.1	contig, BMRR01000008.1	
<i>Paenarthrobacter nicotinovorans</i>	DSM 420	GCA_017876445.1	contig, JAGINZ010000002.1	
<i>Arthrobacter</i> sp.	NtRootD5	GCA_019977175.1	pNtRootD5-1	syntenic <i>nic</i> -genes cluster
<i>Arthrobacter</i> sp.	NtRootC7	GCA_019977135.1	pNtRootC7-1	
<i>Arthrobacter</i> sp.	NtRootA4	GCA_019977095.1	pNtRootA4	
<i>Arthrobacter</i> sp.	NtRootC45	GCA_019977155.1	pNtRootC45	
<i>Arthrobacter</i> sp.	NtRootA2	GCA_019976995.1	pNtRootA2-2	
<i>Arthrobacter</i> sp.	NicSoilB11	GCA_019977375.1	pNicSoilB11-2	
<i>Arthrobacter</i> sp.	NicSoilE8	GCA_019977415.1	pNicSoilE8-2	
<i>Arthrobacter</i> sp.	TES	GCA_014863565.1	pTES1	
<i>Paenarthrobacter ureafaciens</i>	CZY1	GCA_016694995.1	pCZY	
<i>Arthrobacter</i> sp.	ZXY-2	GCA_001854365.1	pZXY21	
<i>Arthrobacter</i> sp.	M2012083	GCA_000281065.1	contig, NZ_AKKK01000058	<i>nic</i> genes spread within the replicon
<i>Paenarthrobacter</i> sp.	YJN-5	GCA_015040095.1	chromosome	
<i>Arthrobacter</i> sp.	NtRootA9	GCA_019977115.1	chromosome	
<i>Arthrobacter</i> sp.	StoSoilB13	GCA_019977255.1	pStoSoilB13-1	
<i>Arthrobacter</i> sp.	StoSoilA2	GCA_019977195.1	chromosome	
<i>Arthrobacter</i> sp.	NicSoilC5	GCA_019977395.1	chromosome	
<i>Arthrobacter</i> sp.	NtRootA1	GCA_019976975.1	chromosome	
<i>Arthrobacter</i> sp.	StoSoilB5	GCA_019977235.1	chromosome	
<i>Nocardioides</i> sp.	JQ2195	GCA_012272695.1	chromosome	
<i>Nocardioides</i> sp.	JS614	GCA_000015265.1	chromosome	
<i>Rhodococcus opacus</i>	R7	GCA_000736435.1	chromosome	
<i>Rhodococcus opacus</i>	B4	GCA_000010805.1	chromosome	
<i>Rhodococcus opacus</i>	PD630	GCA_000599545.1	chromosome	
<i>Rhodococcus opacus</i>	PD630	GCA_020542785.1	chromosome	
<i>Rhodococcus opacus</i>	DSM 43205	GCF_001646735.1	multiple contigs	
<i>Rhodococcus opacus</i>	ATCC 51881	GCF_012396235.1	multiple contigs	

(See figure on next page.)

Fig. 3 Nine highly syntenic *nic*-genes clusters identified in various *Arthrobacter* and *Paenarthrobacter* genomes. **A** BRIG comparative genomic analysis of the pAO1 megaplasmid against the strains harboring the *nic*-genes listed in the legend. External black circle – the localisation of the *nic*-genes cluster on pAO1; blue labels – pAO1 genes experimentally related to nicotine metabolism; red labels – genes related to recombination events. **B** Mauve alignments of the *nic*-genes clusters. Plasmids and genomes are listed on the left with the corresponding location of the *nic*-genes below. Coloured boxes – linear collinear blocks (LCB); white gaps – insertions and deletions; position atop or below the horizontal line represents the direction of LCB; red rectangles and text – recombination related ORFs; blue rectangles and text – pAO1 genes experimentally related to nicotine metabolism; *nbr* – nicotine blue oxidoreductase; *mao* – monoamine-oxidase; *sad* – succinic semi aldehyde dehydrogenase; *folD* – methylene-tetrahydrofolate dehydrogenase/ cyclohydrolase; *abo* – γ -N-methylaminobutyrate oxidase; *purU* – formyl-tetrahydrofolate deformylase; *pmfR* – transcriptional regulator; *mobA* – MobA, related to molybdopterin cytosine dinucleotide cofactor biosynthesis; *nit* – ω -amidase; *hph* – 2,6-dihydroxypyridine-3-hydroxylase; *pkc* – putative polyketide cyclase; *kdhL* – ketone dehydrogenase, large subunit; *pnh* – 2,6-dihydroxypseudooxynicotine hydrolase; *kdhMS* – ketone dehydrogenase, medium and subunits; *6hlnO* – 6-hydroxy-L-nicotine oxidase; *ndhLSM* – nicotine dehydrogenase, large, small and medium subunits; *moaA* – molybdopterin cofactor synthesis protein; *6hdno* – 6-hydroxy-D-nicotine oxidase; *hdnoR* – transcriptional regulator; LCB2 – locally colinear block containing genes associated with the synthesis of the molybdopterin cytosine dinucleotide cofactor; LCB1 – locally colinear block containing genes for processing 6-hydroxy-D-nicotine. For an overview of the nicotine degradation pathway and role of each gene product, please see Supplementary Fig. 2

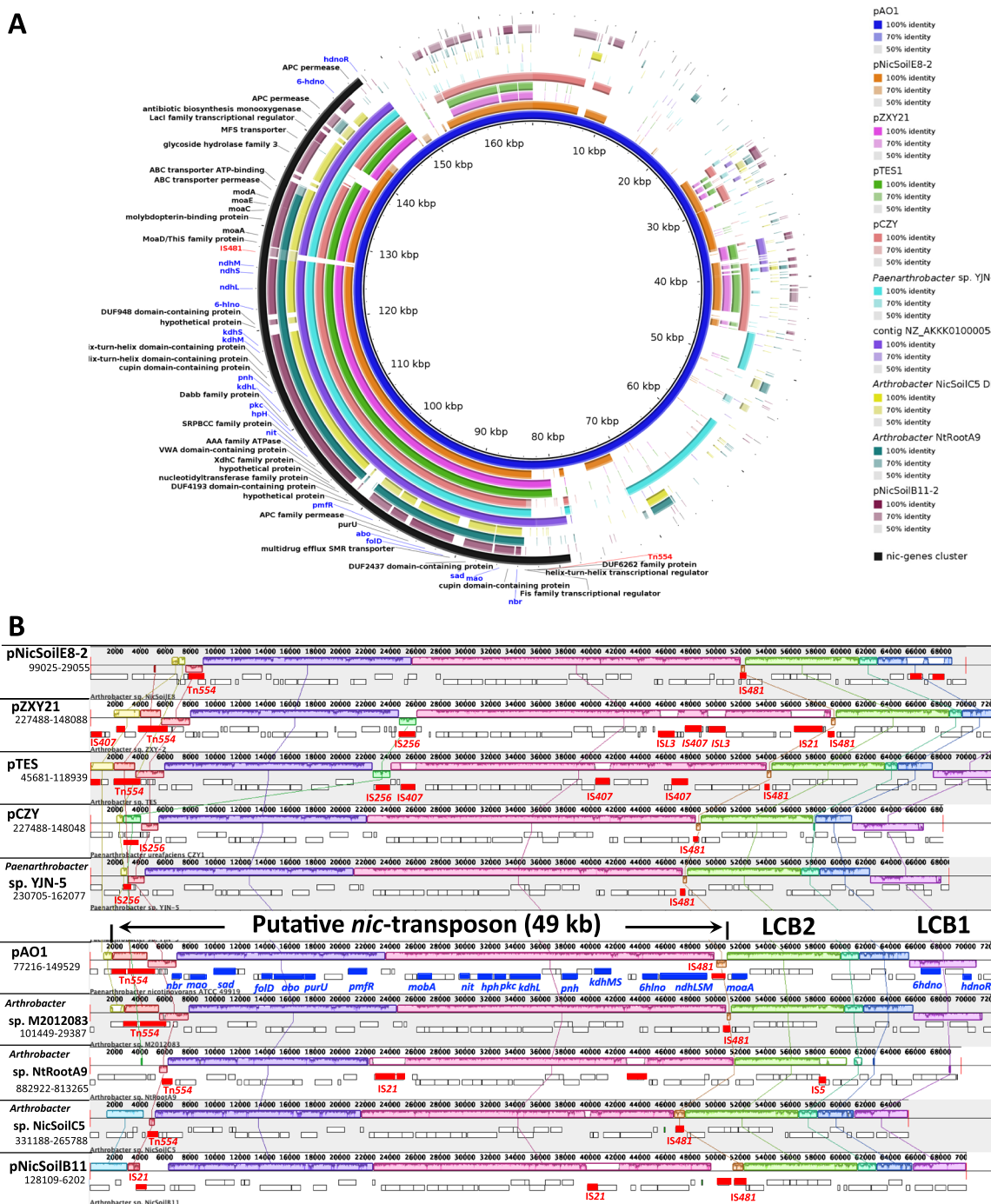


Fig. 3 (See legend on previous page.)

plasmid-encoded nicotine catabolic pathway and how it was spread to soil bacteria by horizontal gene transfer. The different G + C content of the pAO1 plasmid compared to the *P. nicotinovorans* ATCC 49919 chromosome suggests that the two replicons might have evolved independently [39] up to some point when the plasmid was acquired. Phylogenomic analysis of 64 related genomes

identified 9 *Paenarthrobacter* and *Arthrobacter* genomes that have both dDDH and ANI values over the species threshold when compared to *P. nicotinovorans* ATCC 49919. Interestingly, two of these genomes, namely *Paenarthrobacter nicotinovorans* Hce-1 and *Paenarthrobacter* sp. YJN-D, do not contain any plasmids or contigs similar to pAO1. The phylogenetic tree depicted in Fig. 1

shows that *Paenarthrobacter* sp. YJN-D and *Paenarthrobacter nicotinovorans* Hce-1 share a common ancestor with all the strains harbouring the pAO1 megaplasmid. At some point, *Paenarthrobacter* sp. YJN-D or a closely related strain appears to have acquired the plasmid and the ability to metabolise nicotine, thus leading to ATCC 49919. Similar plasmid acquisition events detected by whole genome sequencing and by comparative genomics have been reported for the 167 kb plasmid pCF-SAN061771_01 in *Escherichia coli* ST1485 [40], for two ~40 kb plasmids in *Sulfurospirillum* sp. ACS_{DCE} and strain ACS_{TCE} [41], the ~240 kb plasmid pS810b in *Pseudomonas aeruginosa* [42] and the much smaller 3.7 kb plasmid pPVER1 in *Providencia vermicola* [43].

The pAO1 plasmid sequence described here shows 99.99% identity with the previously available sequence of the pAO1 megaplasmid (GenBank entry AJ507836) but is 4 bp longer (165,141 bp vs 165,137 bp). The length difference provides a strong validation of our assembly by being in good agreement with the observation of the original pAO1 sequence submitters that the irresolvable compression they encountered during assembly could have increased the actual size of the plasmid by up to 5 bp [7]. Moreover, a closer look at the *nic*-genes cluster from the two pAO1 sequences indicated that all the genes that were experimentally related to nicotine metabolism were correctly annotated, with only minor differences for the position of the start or stop codons. The more distinct differences in annotations can be observed in the case of hypothetical and putative proteins, a situation that is due to improvements in gene detection algorithms (Supplementary Table 3).

BLAST searches using the sequences of the *nic*-genes from pAO1 allowed the identification of 21 more genomes containing at least 5 different genes putatively related to nicotine metabolism. In an effort to identify a common ancestor for the *nic*-genes cluster from all these genomes, a phylogenetic tree was reconstructed from concatenated sequences of *purU*, *pnh*, *pmfR*, *ndhL*, *kdhL* (Fig. 4) and gene synteny was assessed (for an overview, see Supplementary Fig. 1). In 5 cases, the *nic*-genes are placed on plasmids that are almost identical with pAO1: pNtRootD5-1, pNtRootC7-1, pNtRootA4, pNtRootC45 and pNtRootA2-2 (dddH values ≥ 99.7%, ANI values ≥ 99.9%, nucleotide identity levels in Supplementary Table 4). At about ~160.5 kb, all these plasmids are approximately 4.6 kb smaller than pAO1, lacking two stretches of DNA that are outside of the *nic*-genes cluster: position 11,920 to position 15,907 encoding 5 hypothetical proteins, and position 26,501 to position 28,385 encoding a putative DprA DNA-processing protein reported as being involved in bacterial natural transformation [44].

A highly syntenic *nic*-genes cluster was identified in 9 *Arthrobacter* and *Paenarthrobacter* strains, all genes experimentally related to nicotine metabolism being located on three locally collinear blocks (LCBs) (Fig. 3, B). The two largest LCBs contain most of the *nic*-genes and are always flanked by integrases, further supporting the hypothesis that this is a catabolic transposon [8] which is spreading the *nic*-genes from/to chromosomes and plasmids in soil bacteria. The size of the transposon is 49 kb. There are other known similarly large transposons: the

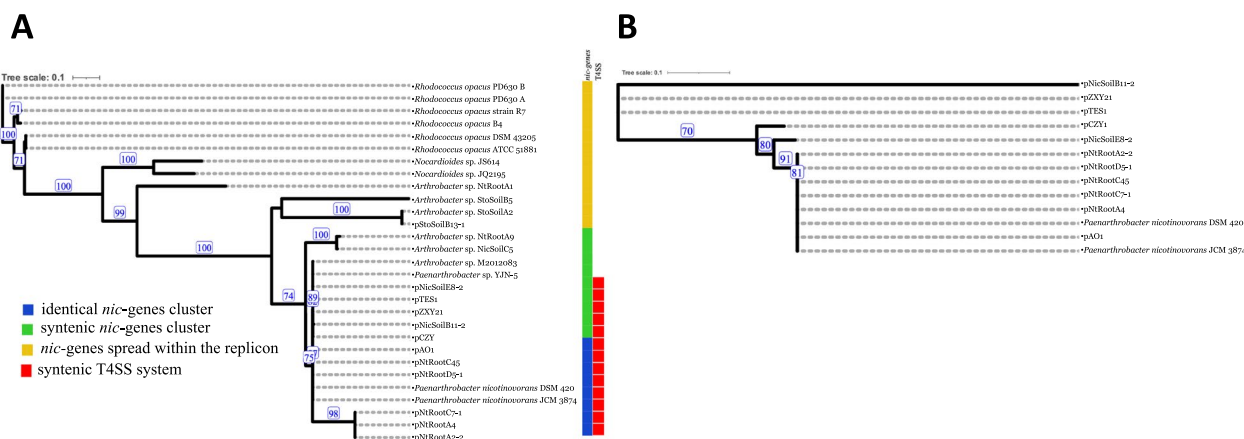


Fig. 4 Maximum Likelihood phylogeny of the microbial pyridine pathway for nicotine degradation (A) and of the plasmids harbouring the *nic*-genes cluster B. The nicotine pathway phylogeny was reconstructed from the concatenated *purU*, *pnh*, *pmfR*, *ndhL*, *kdhL* gene sequences. The plasmids phylogeny was reconstructed from features postulated to be shared by *Paenarthrobacter* [19]: the syntenic T4SS system and DNA repeats located 5' of 3 key ORFs—Duf4192, DprA and ParB. Numbers on branches indicate bootstrap support percentages and values > 70% are shown. The size bar corresponds to 0.1 nucleotide substitutions per site; the length of the dashed lines is not true to scale. Blue colour indicates identical *nic*-genes cluster; green indicates syntenic *nic*-genes cluster and orange indicates the presence of 5 *nic* genes, but the *nic*-cluster is not syntenic; red indicates plasmids that share a syntenic T4SS and DNA repeats located 5' of 3 key ORFs—Duf4192, DprA and ParB

56 kb Tn4651 from *Pseudomonas putida* mt-2 responsible for toluene and xylene degradation, the 38 kb Tn4655 from *P. putida* G7 responsible for naphthalene degradation and the 55 kb Tn4371 from *Ralstonia oxalatica* A5 responsible for biphenyl/4-chlorobiphenyl degradation [45]. Mobilisation of the *nic*-transposon is controlled by 3 site-specific recombinases and transposases similar to Tn544A, Tn554B and Tn554C of the *S. aureus* transposon 544 and other predicted site-specific recombinases [8]. It is still unknown whether the mobilisation process of the *nic*-catabolic transposon is random or driven by other evolutionary forces. Nevertheless, it is generally accepted that the decision for transfer is under the control of the mobile genetic elements (MGEs) and that MGEs can sense when the cell is no longer a promising host for vertical transmission and leave the chromosome for horizontal transmission [46]. The invasion of this transposon by mobile elements that have been previously reported for plasmid pZXY21 [8] can also be seen for plasmids pTES and pNicSoilB11, and the chromosome of *Arthrobacter* sp. NtRootA9 (Fig. 3, B). Although the origin of this catabolic transposon remains elusive, the lower mean G + C content of these LCBs (58.03% ± 0.09) compared to the G + C content of the corresponding replicons/contigs (59.7% up to 67.2%) indicate that the *nic*-genes have been acquired by horizontal gene transfer by all these hosts. So far, we report the presence of this *nic*-transposon only in members of *Paenarthrobacter*, *Arthrobacter*, *Rhodococcus*, and *Nocardioideis* genera. The high (51% – 75%) G + C content that members of Actinobacteria are known for might be a limiting factor for further horizontal travel of this transposon within the soil bacterial communities. The high GC content in a genome is usually correlated with specific promoters sequences as well as specific codon usage [47] which together might hinder the survival of the *nic* transposon and pAO1 megaplasmid in strains with lower GC content.

The third LCB (labelled LCB1 in Fig. 3, B) is placed outside of the putative *nic*-genes-containing transposon and contains only two genes known to be involved in nicotine degradation. One is *6hdno* which encodes a 6-hydroxy-D-nicotine oxidase involved in processing the D-nicotine that might form through the racemisation of the naturally occurring L-nicotine during the decay of the tobacco plant in the soil. The second gene encodes the transcriptional repressor of *6hdno* which discriminates poorly between 6-hydroxy-D-nicotine and 6-hydroxy-L-nicotine as inducers [48]. As L-nicotine is the main product in tobacco leaves, this LCB is not essential for nicotine catabolism, but might offer an advantage in ecological niches where both D- and L-nicotine are present.

Also, outside of the putative transposon and always associated with *nic*-genes in these genomes is LCB2,

shown in Fig. 3, B. This block of DNA contains genes involved in the synthesis of the molybdopterine cytosine dinucleotide cofactor [49, 50] required for the assembly of two key enzymes in the pyridine-pathway for nicotine degradation: nicotine dehydrogenase and keto-dehydrogenase.

It was previously shown that a subset of 12 *Arthrobacter*/*Paenarthrobacter* plasmids share a common mechanism of plasmid partitioning and conjugation—a syntenic T4SS and DNA repeats located 5' of 3 key ORFs—Duf4192, DprA and ParB [19]. The same markers are present on all plasmids harbouring the *nic*-genes cluster reported here and the tree depicted in Fig. 4 indicates their evolutionary relationships. The group of plasmids identical with pAO1 (pNtRootA2-2, pNtRootA4, pNtRootC7-1, pNtRootC45, pNtRootD5-1, contig JAG-INZ010000002.1 from the draft genome of *P. nicotivorans* strain DSM 420 and contig BMRR01000008.1 from the draft genome of *P. nicotivorans* strain JCM 3874) have a common ancestor with pNicSoilE8-2.

Conclusions

The complete genome of *Paenarthrobacter nicotivorans* ATCC 49919 was established using a combination of short- and long-read sequencing followed by hybrid assembly, providing a much-needed reference genome for *P. nicotivorans* strains as well as for other NDMs that use the pyridine pathway for nicotine degradation. Whole-genome-based phylogeny of 64 related genomes available in the NCBI Genome database identified five *Arthrobacter* strains that show both dDDH and ANI values over the species threshold when compared to *P. nicotivorans* ATCC 49919, hence they are mislabelled in the database. Three genomes labelled as *Paenarthrobacter nicotivorans* strains fall below the species threshold limits and, as indicated by both the TYGS phylogenomic analysis and GTDB-Tk assignment, belong to different bacterial species. Moreover, we have shown that the ATCC 49919 strain most probably evolved from *Paenarthrobacter* sp. YJN-D or a very closely related strain by acquiring the pAO1 megaplasmid.

Five plasmids and two contigs belonging to *Arthrobacter* and *Paenarthrobacter* strains were shown here to be virtually identical with the pAO1 plasmid of *Paenarthrobacter nicotivorans* ATCC 49919. Moreover, a highly syntenic *nic*-genes cluster was identified on four plasmids, one contig and four chromosomes. The *nic*-genes cluster contains two major LCBs that apparently form a putative catabolic transposon. The other two LCBs harbour accessory genes that are either involved in the assembly of rare co-factor-containing proteins or related to a more efficient use of the nicotine stereoisomers available in the soil.

Although the origins of the *nic*-genes cluster and the putative transposon still elude us, the data presented here offers another snapshot into the evolution of plasmids harboured by *Arthrobacter* and *Paenarthrobacter* species and their role in the spread of metabolic traits by horizontal gene transfer among related soil bacteria.

Material and methods

Genome sequencing, assembly, annotation, and functional analysis

The complete *Paenarthrobacter nicotinovorans* ATCC 49919 genome was sequenced and assembled using a hybrid approach based on Illumina short-reads and ONT MinION long-reads. Complete protocols and technical details on genomic DNA extraction, sequencing and data processing are provided in our previously published genome announcement paper [30]. Functional analysis of the protein-coding genes was performed using the eggNOG [51, 52] mapper v2 online service available at: <http://eggnog-mapper.embl.de/>. Genome-wide comparison and visualization were performed with BLAST Ring Image Generator (BRIG) [53]. CRISPRCasFinder [54] available at <https://crisprcas.i2bc.paris-saclay.fr/> was used to scan for CRISPR arrays and associated Cas proteins. The Comprehensive Antibiotic Resistance Database (CARD) [55] was used to identify putative antibiotic resistance genes. Resistance to the identified antibiotic classes was further confirmed in the lab using the disk diffusion method on Mueller Hinton agar plates as previously described [56].

Nucleotide sequence accession numbers

The complete and functionally annotated genome is deposited in NCBI Genome with Accession numbers: CP089293 [57] for the chromosome and CP089294 [58] for the pAO1 megaplasmid. All SRA entries are available under the BioSample SAMN17383832 [59] in NCBI BioProject PRJNA693273 [60]. MIGS [61] mandatory information for the complete genome of *Paenarthrobacter nicotinovorans* ATCC 49919 is provided in Supplementary Table 1 and our genome announcement paper [30].

Identification of other genomes of nicotine-degrading microorganisms using the pyridine-pathway

The sequences of the 22 *nic*-genes experimentally related to nicotine degradation in pAO1 were used to perform BLAST searches against the NCBI Nucleotide collection database (max E-value 0.0001, min identity 60%). The genomes containing hits for at least 5 different *nic*-genes were further analysed.

Comparative analysis of *nic*-genes cluster was performed with progressiveMauve [62] using the HOXD

scoring matrix [63]. Mobilome analysis of the *nic*-genes clusters and surrounding areas was performed using MobileElementFinder v1.0.3 (Database version: v1.0.2) [64]. Concatenated sequence tree using the sequences of 5 *nic*-genes (*purU*, *pnh*, *pmfR*, *ndhL*, *kdhL*; Supplementary table 3 for details) were done in PhyML 3.0 [65] using the Jukes-Cantor substitution model.

Whole-genome-based phylogeny

A total of 64 genomes were downloaded from NCBI Genomes, representing all the available *Paenarthrobacter* genomes plus the *Arthrobacter*, *Nocardioideis* and *Rhodococcus* strains previously or herein demonstrated as having *nic*-genes (details and accession numbers in Supplementary Table 2). Taxonomic classifications were performed based on the Genome Database Taxonomy (GTDB, r207) [66] using the GTDB-Tk (v.2.2.4) toolkit [67]. The genomes were uploaded to the Type (Strain) Genome Server (TYGS) [34] for whole genome-based pairwise comparisons and phylogenetic inference. All pairwise comparisons among the set of genomes were conducted using Genome Blast Distance Phylogeny (GBDP) and accurate intergenomic distances inferred under the algorithm 'trimming' and distance formula d_5 [68]. Digital DNA-DNA hybridization (dDDH) values and confidence intervals were calculated using the recommended settings of the Genome-to-Genome Distance calculator (GGDC) 3.0 [68, 69]. The resulting intergenomic distances were used to infer a balanced minimum evolution tree with branch support via FASTME 2.1.6.1 including SPR postprocessing [33]. Species clustering was done using a 70% dDDH cut-off [34], while subspecies clustering was done using a 79% dDDH threshold as previously introduced [33]. Average nucleotide identity by orthology (ANI) values were calculated using the OrthoANI algorithm [70]. All phylogenetic trees were visualized using iTOL [71] while heatmaps were generated with Heatmapper [72]. Sequence manipulations and alignment visualisation were done in Geneious Prime 2022.2.2 (<https://www.geneious.com/>).

Abbreviations

ANI	Average nucleotide identity
COG	Clusters of Orthologous Group
dDDH	Digital DNA-DNA hybridization
GO	Gene Ontology
LCB	Locally collinear block
NB	Nicotine Blue
NDM	Nicotine degrading microorganism
ONT	Oxford Nanopore Technology
T4SS	Type IV secretion system

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09644-3>.

Additional file 1: Supplementary Table 1. MIGS mandatory information for the complete genome of *Paenarthrobacter nicotinovorans* ATCC 49919. **Supplementary Table 2.** Genomes of strains belonging to the *Paenarthrobacter*, *Arthrobacter*, *Nocardioideis* and *Rhodococcus* genera evaluated in the study. The table lists all the *Paenarthrobacter* strains available in NCBI Genomes database as well as *Arthrobacter*, *Nocardioideis* and *Rhodococcus* strains known to have the *nic*-genes. **Supplementary Table 3.** Annotations of the *nic*-genes cluster on the two pAO1 plasmid sequences. Bold indicates genes experimentally related to nicotine degradation. **Supplementary Table 4.** MEGA BLAST results of different plasmids and contigs against the pAO1 megaplasmid of *P. nicotinovorans* ATCC 49919.

Additional file 2: Supplementary figure 1. Overview of identity and gene synteny among the *nic*-genes containing strains of *Paenarthrobacter*, *Arthrobacter*, *Nocardioideis* and *Rhodococcus* evaluated in this study. A. The *nic*-genes form a single locally collinear block (LCB, red) in strains harboring an identical *nic*-genes cluster. B. Most *nic*-genes are located in three LCBs (yellow, green and magenta) in strains harboring a syntenic *nic*-genes cluster. C. Five key *nic*-genes (*purU*, *pnh*, *pmfR*, *ndhL*, *kdhL*) could be identified in other strains, but many genes are missing and no syntenic *nic*-genes cluster can be described.

Additional file 3: Supplementary figure 2. Overview of the nicotine catabolic pathway of *Paenarthrobacter nicotinovorans*. CAPS AND BOLD letters indicate the intermediates: 6-HMM – 6-hydroxy-methylmyosmine; 6-HPON – 6-hydroxy-pseudoxyynicotine; 2,6-HPON – 2,6-dihydroxy-pseudoxyynicotine; 2,6-DHP – 6-dihydroxy-pyridine; MGABA - γ -N-methylaminobutyrate ; 2,3,6-THP - 2,3,6-trihydroxy-pyridine; NB -nicotine blue, 4,4',5,5'-tetrahydroxy-3,3'-diazadiphenquinone-(2,2'); CH2 TH4 - methylenetetrahydrofolate; GABA - γ -aminobutyric acid ; SSA - succinic semialdehyde, alpha-KGA - a-keto-glutaramate; alpha-KG - a-keto-glutarate ; CAPS indicate enzymes catalyzing the stepwise degradation of nicotine: NDH - nicotine dehydrogenase; 6HLNO - 6-hydroxy-L-nicotine oxidase; 6HDNO - 6-hydroxy-D-nicotine oxidase; KDH - ketone dehydrogenase; DHPONH - 2,6-dihydroxypseudoxyynicotine hydrolase; DHPH – 2,6-dihydroxypyridine-3-hydroxylase NBOR – nicotine blue oxidoreductase; MABO - γ -N-methylaminobutyrate oxidase; FoID - methylene-tetrahydrofolate dehydrogenase/cyclohydrolase; PurU - formyl-tetrahydrofolate deformylase; MAO - monoamine-oxidase; AO – amine-oxidase; SsaDH - succinic semialdehyde dehydrogenase; PKC – putative polyketide cyclase; NIT - w-amidase.

Acknowledgements

Not applicable.

Authors' contributions

A.E.S. extracted DNA and performed long-read sequencing, data acquisition, data evaluation and hybrid genome assembly, genome wide analysis and critically revised the manuscript. M.A.M. did data evaluation, performed the antibiotic susceptibility testing and critically revised the manuscript. M.I.T. did data evaluation, critically revised the manuscript. I.H. extracted the DNA for the short-read sequencing and did formal analysis, critically reviewed the manuscript. I.H. and F.K. extracted the DNA. R.S.B. extracted DNA and did the library prep for short-read sequencing, performed data acquisition, data evaluation and critically revised the manuscript. M.M. devised the project, conceived, and planned the experiments; provided the computational resources; performed data evaluation and genome wide analysis; took the lead in drafting the manuscript. The final manuscript was approved by all authors.

Funding

This work was supported by a grant from the Romanian Ministry of Education and Research, CNCS—UEFISCDI, project number PN-III-P4-ID-PCE-2020-0656, within PNCDI III. Fakhri Kallabi was a postdoctoral “Eugen Ionescu” fellow supported by the Francophone University Agency (AUF).

Availability of data and materials

Paenarthrobacter nicotinovorans ATCC 49919 is available upon request or at ATCC. The sequencing data is available under NCBI BioProject accession number PRJNA693273. The genome assembled and annotated sequences have been deposited at NCBI GenBank under the following accession numbers: CP089293 for *P. nicotinovorans* ATCC 49919 chromosome and CP089294 for *P. nicotinovorans* ATCC 49919 megaplasmid pAO1. All data analysed during this study is included in this published article and its supplementary files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 21 April 2023 Accepted: 1 September 2023

Published online: 11 September 2023

References

- Eberwein H, Gries FA, Decker K [On the decomposition of nicotine by bacterial enzymes. II. Isolation and characterization of a nicotine-splitting soil bacterium]. Hoppe-Seyler's Zeitschrift für Physiol Chemie. 1961;323:236–48.
- Decker K, Bleeg H. Induction and purification of stereospecific nicotine oxidizing enzymes from *Arthrobacter* oxidans. Biochim Biophys Acta - Enzymol Biol Oxid. 1965;105:313–24.
- Kodama Y, Yamamoto H, Amachi T. Reclassification of two strains of *Arthrobacter oxydans* and proposal of *Arthrobacter nicotinovorans* sp. nov. Int J Syst Bacteriol. 1992;42:234–9.
- Busse H-J. Review of the taxonomy of the genus *Arthrobacter*, emendation of the genus *Arthrobacter sensu lato*, proposal to reclassify selected species of the genus *Arthrobacter* in the novel genera *Glutamicibacter* gen. nov, *Paeniglutamicibacter* gen. nov. Pseudoglut Int J Syst Evol Microbiol. 2016;66:9–37.
- Brandsch R. Microbiology and biochemistry of nicotine degradation. Appl Microbiol Biotechnol. 2006;69:493–8.
- Liu J, Ma G, Chen T, Hou Y, Yang S, Zhang K-Q, et al. Nicotine-degrading microorganisms and their potential applications. Appl Microbiol Biotechnol. 2015;99:3775–85.
- 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;185:1976–86.
- Brandsch R, Mihasan M. A soil bacterial catabolic pathway on the move: transfer of nicotine catabolic genes between *Arthrobacter* genus megaplasmids and invasion by mobile elements. J Biosci. 2020;45:58.
- Mihășan 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;8:16239.
- Mihășan M, Boiangiu RȘ, Guzun D, Babii C, Aslebagh R, Channaveerappa D, et al. Time-dependent analysis of *Paenarthrobacter nicotinovorans* pAO1 nicotine-related proteome. ACS Omega. 2021;6:14242. <https://doi.org/10.2021/acsomega.1c01020>.
- Hritcu L, Stefan M, Brandsch R, Mihasan M. 6-hydroxy-L-nicotine from *Arthrobacter nicotinovorans* sustain spatial memory formation by decreasing brain oxidative stress in rats. J Physiol Biochem. 2013;69:25–34.
- Hritcu L, Mihasan M. 6-Hydroxy-L-Nicotine and Memory Impairment. In: Preedy VR, editor. Neuroscience of Nicotine. Elsevier; 2019. p. 165–72.
- Boiangiu RS, Mihasan M, Gorgan DL, Stache BA, Petre BA, Hritcu L. Cotinine and 6-Hydroxy-L-Nicotine reverses memory deficits and reduces oxidative stress in A β 25-35-induced rat model of alzheimer's disease. Antioxidants. 2020;9:768.

14. Yu W, Wang R, Li H, Liang J, Wang Y, Huang H, et al. Green route to synthesis of valuable chemical 6-hydroxynicotine from nicotine in tobacco wastes using genetically engineered *Agrobacterium tumefaciens* S33. *Biotechnol Biofuels*. 2017;10:288.
15. Mihasan M, Brandsch R. pAO1 of *Arthrobacter nicotinovorans* and the spread of catabolic traits by horizontal gene transfer in gram-positive soil bacteria. *J Mol Evol*. 2013;77:22–30.
16. Ganas P, Igloi GL, Brandsch R. *Microbial Megaplasmids*. Berlin, Heidelberg: Springer, Berlin Heidelberg; 2009.
17. Jerke K, Nakatsu CH, Beasley F, Konopka A. Comparative analysis of eight *Arthrobacter* plasmids. *Plasmid*. 2008;59:73–85.
18. Mihāşan M. Bioinformatics-based molecular classification of *Arthrobacter* plasmids. *Cell Mol Biol Lett*. 2015;20:612.
19. Mihāşan M, Brandsch R. A predicted T4 secretion system and conserved DNA-repeats identified in a subset of related *Arthrobacter* plasmids. *Microbiol Res*. 2016;191:32–7.
20. Nayfach S, Roux S, Seshadri R, Udway D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. *Nat Biotechnol*. 2020;39:499–509.
21. Stanislauskiene R, Gasparaviciute R, Vaitekunas J, Meskiene R, Rutkiene R, Casaitė V, et al. Construction of *Escherichia coli*-*Arthrobacter*-*Rhodococcus* shuttle vectors based on a cryptic plasmid from *Arthrobacter rhombi* and investigation of their application for functional screening. *FEMS Microbiol Lett*. 2012;327:78–86.
22. Niewerth H, Schuldes J, Parschat K, Kiefer P, Vorholt JA, Daniel R, et al. Complete genome sequence and metabolic potential of the quinaldine-degrading bacterium *Arthrobacter* sp. Rue61a. *BMC Genomics*. 2012;13:534.
23. Vandera E, Samiotaki M, Parapouli M, Panayotou G, Koukoku AI. Comparative proteomic analysis of *Arthrobacter phenanthrenivorans* Sphe3 on phenanthrene, phthalate and glucose. *J Proteomics*. 2015;113:73–89.
24. Kallimanis A, LaButti KM, Lapidus A, Clum A, Lykidis A, Mavromatis K, et al. Complete genome sequence of *Arthrobacter phenanthrenivorans* type strain (Sphe3). *Stand Genomic Sci*. 2011;4:123–30.
25. Yao Y, Tang H, Su F, Xu P. Comparative genome analysis reveals the molecular basis of nicotine degradation and survival capacities of *Arthrobacter*. *Sci Rep*. 2015;5:8642.
26. Wang Y, Hu Y, Wang J, Yu Y, Song Y, Yang G, et al. Isolation and characterization of *Arthrobacter* sp. HW08 capable of biodegrading swainsonine. *AFR J Microbiol Res*. 2010;4:1635–8.
27. Vikram S, Kumar S, Vaidya B, Pinnaka AK, Raghava GPS. Draft genome sequence of the 2-Chloro-4-Nitrophenol-degrading bacterium *Arthrobacter* sp. Strain SJCon Genome Announc. 2013;1:e0005813.
28. Monnet C, Loux V, Gibrat J-F, Spinnler E, Barbe V, Vacherie B, et al. The *arthrobacter arilaitensis* Re117 genome sequence reveals its genetic adaptation to the surface of cheese. *PLoS ONE*. 2010;5:e15489.
29. Meng J, Sun XMM, Li SSS, Liang H. Draft genome sequence of *Paenarthrobacter nicotinovorans* Hce-1. *Genome Announc*. 2017;5:e00727-e817.
30. El-Sabeh A, Honceriu I, Kallabi F, Boiangiu R-S, Mihasan M. Complete genome sequences of two closely related *paenarthrobacter nicotinovorans* strains. *Microbiol Resour Announc*. 2022;11:e0013322.
31. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol*. 2017;13:e1005595.
32. Ciufu S, Kannan S, Sharma S, Badretin A, Clark K, Turner S, et al. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int J Syst Evol Microbiol*. 2018;68:2386.
33. Lefort V, Desper R, Gascuel O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol*. 2015;32:2798–800.
34. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun*. 2019;10:1–10.
35. Yao Y, Tang H, Ren H, Yu H, Wang L, Xu P. Genome sequence of a nicotine-degrading strain of *Arthrobacter*. *J Bacteriol*. 2012;194:5714–5.
36. Ruan A, Min H, Zhu W. Studies on biodegradation of nicotine by *Arthrobacter* sp. strain HF-2. *J Environ Sci Health B*. 2006;41:1159–70.
37. Gong X, Ma G, Duan Y, Zhu D, Chen Y, Zhang KQ, et al. Biodegradation and metabolic pathway of nicotine in *Rhodococcus* sp. Y22. *Mol J Microbiol Biotechnol*. 2016;32:188.
38. Cobzaru C, Ganas P, Mihasan M, Schleberger P, Brandsch R. Homologous gene clusters of nicotine catabolism, including a new ω -amidase for α -ketoglutarate, in species of three genera of Gram-positive bacteria. *Res Microbiol*. 2011;162:285–91.
39. Hülter N, Ilhan J, Wein T, Kadibalban AS, Hammerschmidt K, Dagan T. An evolutionary perspective on plasmid lifestyle modes. *Curr Opin Microbiol*. 2017;38:74–80.
40. Hammad AM, Gonzalez-Escalona N, El Tahan A, Abbas NH, Koenig SSK, Allué-Guardia A, et al. Pathogenome comparison and global phylogeny of *Escherichia coli* ST1485 strains. *Sci Rep*. 2022;12:18495.
41. Yang Y, Schubert T, Lv Y, Li X, Yan J. Comparative genomic analysis reveals preserved features in Organohalide-Respiring *Sulfurospirillum* strains. *mSphere*. 2022;7:0093121.
42. Lopes R, Furlan JPR, Stehling EG. Acquisition of bla_{IMP}-13 on a novel IncP-7 plasmid in XDR VIM-2-positive *Pseudomonas aeruginosa* belonging to the global high-risk clone ST235 in an agricultural ecosystem. *J Glob Antimicrob Resist*. 2022;30:403–5.
43. Andolfo G, Schuster C, Ben GH, Ruocco M, Leclercq A. Genomic analysis of the nomenclatural type strain of the nematode-associated entomopathogenic bacterium *Providencia vermicola*. *BMC Genomics*. 2021;22:1–21.
44. Sharma DK, Misra HS, Bihani SC, Rajpurohit YS. Biochemical properties and roles of DprA protein in bacterial natural transformation, virulence, and pilin variation. *J Bacteriol*. 2023. <https://doi.org/10.1128/JB.00465-22>.
45. Top EM, Springael D, Boon N. Catabolic mobile genetic elements and their potential use in bioaugmentation of polluted soils and waters. *FEMS Microbiol Ecol*. 2002;42:199–208.
46. Haudiquet M, De Sousa JM, Touchon M, Rocha EPC. Selfish, promiscuous and sometimes useful: how mobile genetic elements drive horizontal gene transfer in microbial populations. *Philos Trans R Soc B*. 2022;377:20210234.
47. Connell ND. Expression systems for use in actinomycetes and related organisms. *Curr Opin Biotechnol*. 2001;12:446–9.
48. Sandu C, Chiribau CB, Brandsch R. Characterization of HdnR, the transcriptional repressor of the 6-hydroxy-D-nicotine oxidase gene of *Arthrobacter nicotinovorans* pAO1, and its DNA-binding activity in response to L- and D-nicotine derivatives. *J Biol Chem*. 2003;278:51307–15.
49. Menéndez C, Igloi G, Henninger H, Brandsch R. A pAO1-encoded molybdopterin cofactor gene (moaA) of *Arthrobacter nicotinovorans*: characterization and site-directed mutagenesis of the encoded protein. *Arch Microbiol*. 1995;164:142–51.
50. Menéndez C, Siebert D, Brandsch R. MoaA of *Arthrobacter nicotinovorans* pAO1 involved in Molybdopterin cofactor synthesis is an Fe-S protein. *FEBS Lett*. 1996;391:101–3.
51. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol*. 2017;34:2115–22.
52. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation orthology assignments and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021;38:5825–9.
53. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*. 2011;12:1–10.
54. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res*. 2018;46:W246–51.
55. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res*. 2020;48:D517–25.
56. Bauer AW, Kirby WM, Sherris JC, Turck M. Antibiotic susceptibility testing by a standardized single disk method. *Am J Clin Pathol*. 1966;45(4_ts):493–6.
57. El-Sabeh A, Honceriu I, Boiangiu RS, Kallabi F, Mihasan M. Sequencing the genome of a useful bacteria: *Paenarthrobacter nicotinovorans* - next step in extending its biotechnological applications. *GenBank*. 2021. <https://identifiers.org/ncbi/insdc:CP089293>. Accessed 21 Apr 2023.
58. El-Sabeh A, Honceriu I, Boiangiu RS, Kallabi F, Mihasan M. Sequencing the genome of a useful bacteria: *Paenarthrobacter nicotinovorans* - next step

- in extending its biotechnological applications. GenBank. 2021. <https://identifiers.org/ncbi/insdc:CP089294>. Accessed 21 Apr 2023.
59. NCBI BioSample. 2021. <https://identifiers.org/biosample:SAMN17383832>. Accessed 21 Apr 2023.
 60. NCBI BioProject. 2021. <https://identifiers.org/bioproject:PRJNA693273>. Accessed 21 Apr 2023.
 61. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol*. 2008;26:541–7.
 62. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*. 2010;5: e11147.
 63. Chiaromonte F, Yap VB, Miller W. Scoring pairwise genomic sequence alignments. *Pac Symp Biocomput*. 2002;7:115–26.
 64. Johansson MHK, Bortolaia V, Tansirichaiya S, Aarestrup FM, Roberts AP, Petersen TN. Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: mobileElementFinder. *J Antimicrob Chemother*. 2021;76:101–9.
 65. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21.
 66. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res*. 2022;50:D785–94.
 67. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics*. 2022;38:5315–6.
 68. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform*. 2013;14:1–14.
 69. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res*. 2022;50:D801–7.
 70. Lee I, Kim YO, Park SC, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol*. 2016;66:1100–3.
 71. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49:W293–6.
 72. Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, et al. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res*. 2016;44:147.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

