Título del Trabajo Final

“Bacterial degradation of petroleum hydrocarbons; a comparative genomics study of the genes involved in the catabolic pathways of naphthalene”

Nombre Estudiante: Athanasía Varsaki

Plan de Estudios del Estudiante: Bioinformática

Área del trabajo final: Microbiología, biotecnología y biología molecular

Nombre Consultor/a: Paloma Pizarro Tobías

Nombre Profesor/a responsable de la asignatura: Paloma Pizarro Tobías

Fecha Entrega: 02/01/2018

Licencias alternativas (elegir alguna de las siguientes y sustituir la de la página anterior)

A) Creative Commons:

Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada 3.0 España de Creative Commons

Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-CompartirIgual 3.0 España de Creative Commons

Esta obra está sujeta a una licencia de Reconocimiento-NoComercial 3.0 España de Creative Commons

Esta obra está sujeta a una licencia de Reconocimiento-SinObraDerivada 3.0 España de Creative Commons

Esta obra está sujeta a una licencia de Reconocimiento-CompartirIgual 3.0 España de Creative Commons

B) GNU Free Documentation License (GNU FDL)

C) Copyright

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Bacterial degradation of petroleum hydrocarbons; a comparative genomics study of the genes involved in the catabolic pathways of naphthalene*
Nombre del autor:	*Athanasia Varsaki*
Nombre del consultor/a:	*Paloma Pizarro Tobías*
Nombre del PRA:	*Paloma Pizarro Tobías*
Fecha de entrega (mm/aaaa):	01/2018
Titulación::	*Bioinformática*
Área del Trabajo Final:	*Microbiología, biotecnología y biología molecular*
Idioma del trabajo:	*Ingles*
Palabras clave	*Bioremediation; biodegradation; catabolic genes*

Resumen del Trabajo (máximo 250 palabras):

Los compuestos aromáticos se encuentran entre los contaminantes más prevalentes y persistentes en el medio ambiente. El suelo y los sedimentos contaminados con petróleo suelen contener una mezcla de hidrocarburos aromáticos policíclicos (PAHs) y heterocíclicos. La información sobre el catabolismo de compuestos aromáticos xenobióticos naturales por bacterias ha aumentado considerablemente en los ultimos años. Los estudios detallados más recientes incluyen en general pocas rutas metabólicas, en un rango relativamente limitado de bacterias, en su mayoría del género *Pseudomonas*, y no hay estudios disponibles con respecto a la genómica comparativa de estas

bacterias. En este trabajo proponemos un estudio genético comparativo de los genes involucrados en la ruta catabólica del naftaleno. Usando la enzima naftaleno 1,2-dioxigenasa ferrodoxin-NAD(P)$^+$ reductasa (NahAa, Acc. Number: AAS79488.1) de *Pseudomonas putida* como marcador filogenético, intentamos identificar las proteínas homólogas presentes en las bases de datos de NCBI. Por genómica comparativa, pretendemos identificar los posibles grupos de las vías catabólicas del naftaleno y la estructura de los operones. Los resultados de este estudio indican que las bacterias degradan naftaleno principalmente a través de la vía "clásica" de naftaleno, descrita para el género *Pseudomonas* o a través de la ruta alternativa del gentisato-salicilato.

Abstract (in English, 250 words or less):

Aromatic compounds are among the most prevalent and persistent pollutants in the environment. Petroleum-contaminated soil and sediment commonly contain a mixture of polycyclic aromatic hydrocarbons (PAHs) and heterocyclic aromatics. In recent years there has been a quantum leap in the information published about the catabolism of natural xenobiotic aromatic compounds by bacteria. At present the most detailed studies have been carried out on relatively few metabolic pathways, in a rather limited range of bacteria, mostly of the genus *Pseudomonas*, but there is no study available regarding the comparative genomics of these bacteria. Here we propose a comparative genetics study of the genes involved in the catabolic pathway of naphthalene, using the naphthalene 1,2-dioxygenase system ferredoxin-NAD(P)$^+$ reductase enzyme (NahAa, Acc. Number: AAS79488.1) from *Pseudomonas putida* as a phylogenetic marker. By that way we intent to identify the homologous proteins present in the NCBI databases and through the proteins make conclusions about possible groups of the catabolic pathways of naphthalene and their operon structure. Results of this study indicate that bacteria mainly degrade naphthalene through either the “classic” *nah* pathway described for *Pseudomonas* genus or the alternative gentisate-salicylate pathway.

Index

Lista de figuras

1. Introducción

1.1 Contexto y justificación del Trabajo

Petroleum is a naturally occurring, yellow-to-black liquid found in geological formations beneath the Earth's surface. The name *petroleum* covers both naturally occurring unprocessed crude oil and petroleum products that are made up of refined crude oil and is one of the most important raw materials for the production of petroleum fuels (gasoline and kerosene) and chemical reagents products (pharmaceuticals, solvents, fertilizers, and plastics). Today, about 90 percent of vehicular fuel needs are met by oil. Due to the increasing global petroleum demand, more and more activities such as exploitation, extraction, refining, and transportation of petroleum have been conducted. At the same time, about half of the petroleum products are transported across the world by sea. Consequently, there is a dramatic increase of the petroleum spill or leakage accidents all over the world, such as spills from oil wells, pipelines, drilling rigs, or transport tankers.

The first large marine oil spill occurred in 1907 with the sinking of the Thomas W. Lawson, which released 7,400 tons of paraffin oil off the coast of the United Kingdom. Since then, estimates indicate that more than 7 million tons of oil have been released into the environment, with the Deep Water Horizon (DWH) disaster releasing more than 700,000 tons of crude oil in the Gulf of Mexico (1). Oil spill accidents have a serious impact on the marine/terrestrial ecosystems, economy and human health. For example, oil slick forms an anaerobic condition in the sea water and leads to the death of flora and fauna, with disastrous consequences in the fishing industry. Oil spills can cause hypothermia of marine birds and mammals by reducing/destroying the insulating ability of the plumage of birds and the fur of mammals. Meanwhile, the toxic constituents in petroleum could poison or kill birds, mammals, fish and other marine organisms and damage the fragile underwater ecosystems which lead to a vicious effect on the global food chain, and eventually harm human health by damaging internal organs, such as kidneys, lungs, and liver (2).

Generally, oil spills are being dealt with conventional physical and chemical *in situ* and *ex situ* clean-up technologies (3). However, experience has demonstrated that these strategies are expensive, and often only result in incomplete decomposition of the pollutants of concern. Research over the last two decades has focused on offering remediation schemes that are moving away from the conventional ones and are mainly based on biological methods.

Biodegradation is a viable bioremediation technology for organic pollutants. It has been long known (more than a century ago) that microorganisms degrade environmental pollutants in various matrices and environments (4). The goal of bioremediation is to transform organic pollutants into harmless metabolites or mineralize the pollutants into carbon dioxide and water (5). The biodegradation of petroleum hydrocarbons is a complicated process and the type of bioremediation approach and operational conditions depends on the amount and types of hydrocarbon compounds in the contaminated site and sludge (6).

Petroleum is mainly composed by three hydrocarbon fractions from few up to >60 carbons. These three hydrocarbon fractions are: (i) Paraffin, which is usually the most abundant fraction and contains linear and branched aliphatic hydrocarbons; (ii) naphthenes, which are alicyclic hydrocarbons composed by one or more saturated rings with or without lateral aliphatic branches; and (iii) the aromatic fraction, which is composed by hydrocarbons containing at least one aromatic ring (7). Many bacteria and fungi can degrade petroleum hydrocarbons but not any single strain can afford biodegradation of all compounds found in petroleum sludge (8). Biodegradation of petroleum hydrocarbons is realized by sequential reactions where bacteria in a microbial consortium collaborate with each other for degradation of oily sludge (9) and by that way it is possible to degrade completely all petroleum hydrocarbons apart from heavy polyaromatic's presence in the asphaltenes and resins (10).

Naphthalene has often been used as a model compound to investigate the ability of bacteria to degrade polycyclic aromatic hydrocarbons (PAHs) because it is the simplest and the most soluble PAH (11). The proposed catabolic pathways of naphthalene by bacteria is given in Figure 1 (12). It has been obtained from studies on various strains of *Pseudomonas putida* (13, 14).

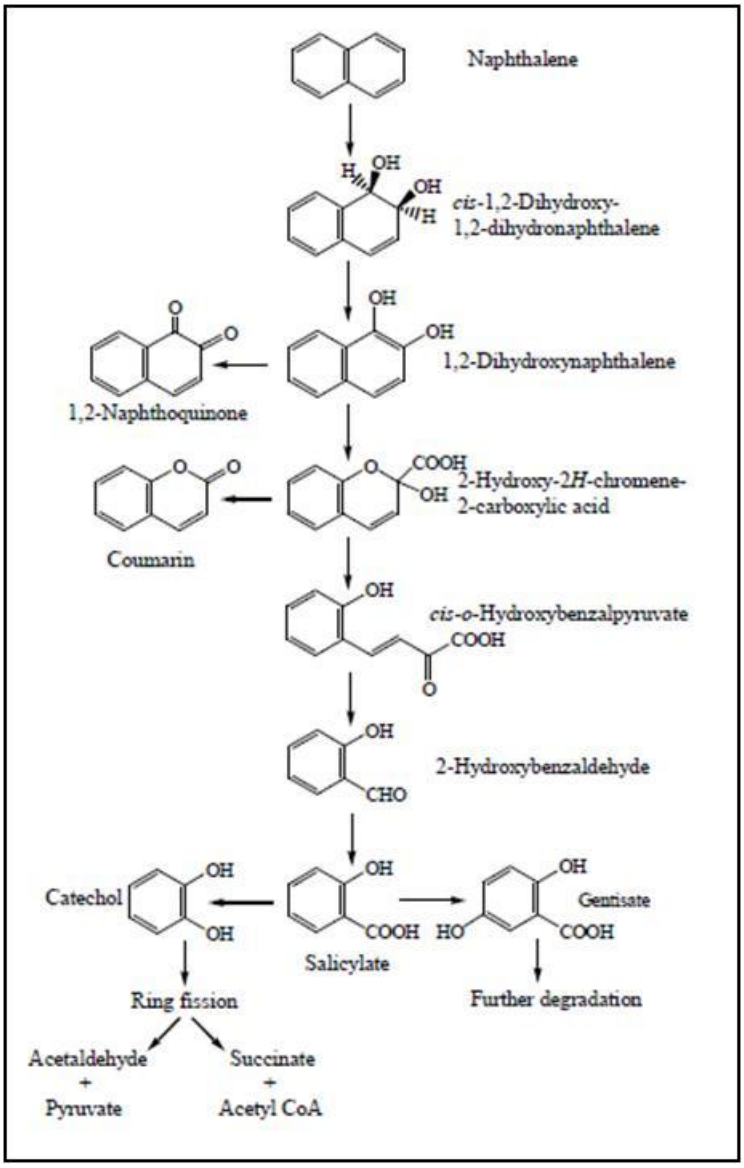

***Figure 1**: Proposed bacterial catabolic pathway of naphthalene*

The bacterial degradation of naphthalene has been well characterized for the catabolic enzyme system encoded by the plasmid NAH7 in *Pseudomonas putida* G7 (15). NAH7 has two operons that contain the structural genes for naphthalene degradation. One operon contains the genes for the upper catabolic pathway (*nahAaAbAcAdBFCED*) encoding the necessary enzymes for the conversion of naphthalene to salicylate (Figure 2). The second operon contains the genes for the lower catabolic pathway (*nahGTHINLOMKJ*) encoding the necessary enzymes for the conversion of salicylate through the catechol meta-cleavage pathway to pyruvate and acetaldehyde (15).

Many studies have been published regarding the ability of isolated bacteria to degrade petroleum hydrocarbons and the pathways they use (16) (see 14 for review and references within). Surprisingly, up to date, there is no study available regarding the comparative genomics of these bacteria. This study tries to fill-in this gap and provide a comparative genomics study of the genes involved in the catabolic pathways of naphthalene.

1.2 Objetivos del Trabajo

This study originally was planned to be divided in five (5) sub-projects (referred as Tasks in the chronogram given in Fig. 2 and Table1).

1. Construction of the dataset using online databases and tools (e.g. BLAST). The databases were constructed using the naphthalene 1,2-dioxygenase system ferredoxin-NAD(P)$^{+}$ reductase enzyme (NahAa, Acc. Number: AAS79488.1) from *Pseudomonas putida* in a PSI-BLAST.
2. Construction of protein profiles, alignments and phylogenetic trees using the Molecular Evolutionary Genetics Analysis (MEGA) program.
3. Localize the genes/operons involved in the catabolic pathway of naphthalene. Determine whether they are located on the chromosome or on plasmids.
4. Determine if conjugative elements are present. In case of presence of conjugative elements, debate the possibility that the presence of the catabolic genes are being spread by horizontal gene transfer.
5. Make a statistical analysis using R (find the % of Gram-positive, Gram-negative bacteria involved, how many of them are α-, β-, γ-, δ-, ε-proteobacteria)

The last sub-project (sub-project No5) was not realized because the members of the dataset constructed from sub-project No1 were mainly Gram-negative β-proteobacteria and conclusions could be made without a statistical analysis with a sophisticated program as R.

1.3 Enfoque y método seguido

The most common petroleum hydrocarbons include aliphatic, branched and cycloaliphatic alkanes, as well as monocyclic and polycyclic aromatic

hydrocarbons (PAHs). PAHs include naphthalene, fluorene, phenanthrene, anthracene, fluoranthene, pyrene, benzo[α]-anthrene and benzo[α]pyrene. Most of the bacterial degradation pathways of the above aromatic compounds have been described in detail (12, 17). As stated in the “Introduction”, although there are several studies regarding the elucidation of the biodegradation mechanisms, there is no study thus far that covers the comparative genomics of the genes involved. In this study we propose to fill-in this gap. Since the biodegradation field is very wide field to cover, this study is focused in the comparative genomics of the genes involved in the catabolic pathway of naphthalene. This compound has been chosen because it has often been used as a model compound, its bacterial degradation has been well characterized and information of bacterial degradation of naphthalene has been used to understand and predict pathways in the degradation of three- or more ring PAHs. The first step of this study is a PSI-Blast using as a bait the naphthalene 1,2-dioxygenase system ferredoxin-$NAD(P)^+$ reductase enzyme (NahAa, Acc. Number: AAS79488.1) from *Pseudomonas putida.* This enzyme has been selected as it acts during the first step of the catabolic pathway and it seems a logical point to start with. After this first step, the protein obtained from the PSI-Blast were used in order to construct alignments, phylogenetic trees and then continue with the rest of the proposed sub-projects.

The informatics tools that were used in this study, have been chosen under the following criteria:

i. They are available free on-line
ii. They are being currently used and cited in studies of comparative genomics
iii. They are easy-to-use
iv. The purpose of this work is mainly an exercise that synthesizes the knowledge and skills acquired throughout the master's courses and is part of the learning process. For this reason, most of the programs/informatics tools used have been introduced during the lectures of the master “Bioinformatics and biostatistics” of the UOC and they consist of a direct proof of good practice.

In particular:

1. For the construction of the alignments and the phylogenetic trees, the MEGA program was selected since it includes many sophisticated methods and tools for phylogenomics, it is easy-to-use and include solid statistical methods only. With millions of downloads, MEGA is cited in more than 85,000 papers.

2. The Phyre[2] server was used for protein structure prediction. (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). The use of this server is free for non-commercial use. Phyre is among the most popular methods for protein structure prediction having been cited over 1500 times. It is able to regularly generate reliable protein models using the principles and the techniques of homology modelling.

3. For the visualization of the sequences two programs where used: Artemis (Release 16.0.0) and Vector NTI (Release 10.3.0). Artemis is a free genome viewer and annotation tool developed by the Sanger Institute, provides a flexible interface where users can upload and view any sequence/annotation file available in Genbank or EMBL format. On the other hand, Vector NTI Advance 10 software (ThermoFisher Scientific) is a completely integrated suite of sequence analysis and design tools that helps manage, view, analyze, transform, share, and publish diverse types of molecular biology data, all within one graphically rich analysis environment.

4. For the prediction of putative operons, the softberry sever was used (www.softberry.com) and in particular the FGENESB Suite of bacterial operon and gene finding program. The FGENESB is a package for automatic annotation of bacterial genomes and its algorithm is based on Markov chain models of coding regions and translation and termination sites. It performs operon prediction based on distances between ORFs and frequencies of different genes in neighbouring each other in known bacterial genomes, as well as on promoter and terminator predictions.

5. For the comparison of the operons the EasyFig2.2.2 was used. EasyFig is a Python application for creating linear comparison figures of multiple genomic loci with an easy-to-use graphical user interface and is freely

available. BLAST comparisons between multiple genomic regions, ranging from single genes to whole prokaryote chromosomes, can be generated, visualized and interactively coloured, enabling a rapid transition between analysis and the preparation of publication quality figures.

1.4 Planificación del Trabajo

The schedule of the study followed the chronogram shown in Table 1 and in Figure 2.

Table 1: Chronogram

Dates	Sub-project
03/10/2017-16/10/2017	Prepare the proposal
17/10/2017-20/11/2017	Task1, Task 2: Construction of protein profiles, alignments and phylogenetic trees
21/11/2017-18/12/2017	Task3, Task4, Task5 : Localize the genes/operons, Determine the presence of conjugative elements, Statistical analysis
19/12/2017-02/01/2018	Preparation of the manuscript
03/01/2018-10/01/2018	Preparation of the presentation
11/01/2018-22/01/2018	Public presentation

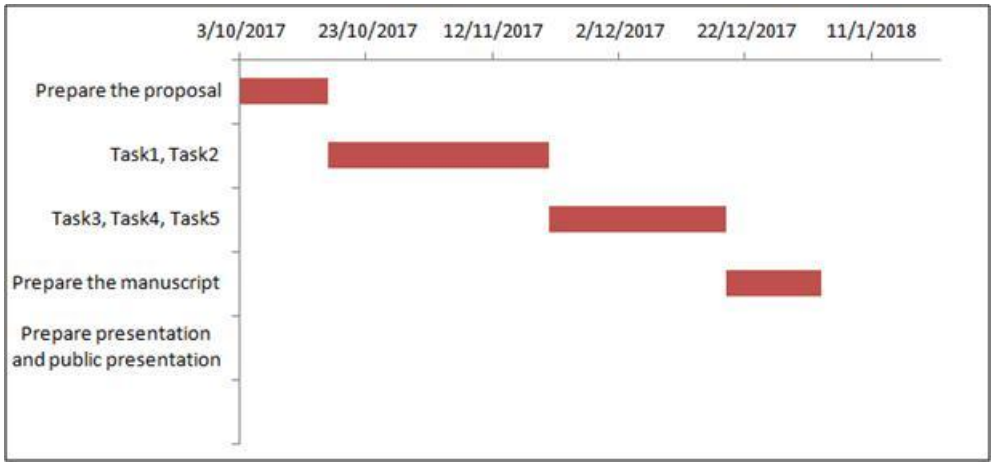

Figure 2: *Chronogram of this study*

1.5 Breve sumario de los productos obtenidos

The putative *nah* operons retrived in this study have a high level of variability being composed by 3 up to 12 CDS. Nevertheless the apparent diversity, all cases indicate that there is only two pathways used by the bacteria to catabolise naphthalene; they use either the “classic” *nah* pathway described for *Pseudomonas* genus, or the alternative gentisate-salicylate pathway.

1.6 Breve descripción de los otros capítulos de la memoria

This study is composed by the following chapters:

1. Introduction: A brief introduction in the state of art of the research topic, objectives, methodology and time-table.
2. Results and Discussion: Presentation of the results obtained and are being discussed in relevance with what is already known from previous studies.
3. Conclusions and general discussion: Presentation of the conclusions of the study and a general discussion.

2. Resultados y Discusión

2.1 Construction of the dataset

A PSI-Blast search was conducted, using naphthalene 1,2-dioxygenase system ferredoxin-NAD(P)$^{+}$ reductase component (NahAa, Acc. Number: AAS79488.1) from *Pseudomonas putida* as a query, in order to identify the homologous proteins present in databases up to 8th October 2017. This component forms part of the naphthalene dioxygenase (NDO) multi-component enzyme system (EC 1.14.12.12) which catalyzes the incorporation of both atoms of molecular oxygen into naphthalene to form cis-(1R,2S)-dihydroxy-1,2-dihydronaphthalene (18) (Figure 3).

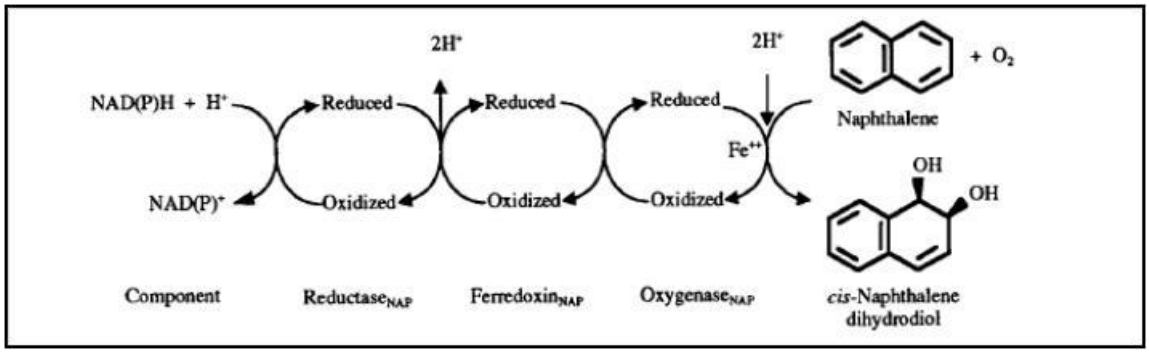

Figure 3: *Reaction catalyzed by the three-component naphthalene dioxygenase (NDO) system.*

The NDO multicomponent enzyme system is composed by an electron transfer component and a dioxygenase component iron sulfur protein. The electron transfer component is composed by a ferredoxin reductase (NahAa) and a ferredoxin (NahAb) (18, 19), and the dioxygenase component is formed by a heterohexamer (trimer of heterodimers) of three large alpha subunits (NahAc) and three small beta subunits (NahAd) (20). As stated before, this enzyme has been selected because it acts during the first step of the catabolic pathway (Figure 1 and Figure 3) and it seemed a logical point to start with. For the PSI-Blast, the following algorithm parameters were used: max target 1,000 and a threshold $P<10^{-80}$. This threshold was selected upon realizing that

oxydoreductases not belonging to degradation of aromatic components were retrieved in an initial search. Therefore, as selection criteria for the P-value was established the one that was giving as hits in the first iteration of the PSI-Blast, only oxydoreductases related to aromatic components degradation.

The PSI-Blast search converged at iteration 4, retrieving 256 non-redundant hits. Thirty-two (32) hits had a query coverage less than 98% and therefore, were not analyzed any further. As a result, a list of 224 sequences were selected (Table ST1 in the section of “Anexos”), six of them (2.7 %) belonging to plasmids and the rest of them (218) to chromosomal DNA.

2.2 Construction of protein profiles, alignments and phylogenetic trees

The proteins retrieved from the PSI-Blast were used in multiple alignments. The alignments were performed with MEGA version 6 (21) using the MUSCLE algorithm (22) and showed that all 224 proteins are more conserved at the N-terminal domain, having eleven (11) conserved residues on the first 50 amino acids, which means that the 20% of the N-terminal amino acids is conserved (Figure 4).

Although that the 3D structure of the NDO has been resolved (20) and the 3D structures of the NahAa, NahAc and NahAd have also been resolved (23, 24), the NahAa 3D structure is not available yet, even though the protein has been purified and characterized (18). A Phyre^2 analysis (25) of the NahAa gave as first hit with 100% confidence the toluene-4-monooxygenase ferredoxin oxidoreductase (T4moF) (26). In T4moF, the N-terminal domain contains a single [2Fe-2S] cluster coordinated by four cysteines in the Cys-X4-Cys-X2-Cys//Cys plant-type ferredoxin motif (27). Cysteine residues 36, 41, 44 and 76 coordinate the iron atoms of the [2Fe-2S] cluster, while residues 35, 37 and 39-42 provide hydrogen bonds to the sulfur atoms in the [2Fe-2S] cluster. The [2Fe-2S] cluster is also flanked by Tyr34 and Leu74, which protect it from the solvent. The presence of a putative [2Fe-2S] cluster could explain the conserved amino acids in the N-terminal domain of NahAa. In Figure 4, the analogous cystein residues are marked with red arrows, the analogous tyrosine residue is marked with a green arrow and the analogous leucine residues is marked with a blue arrow.

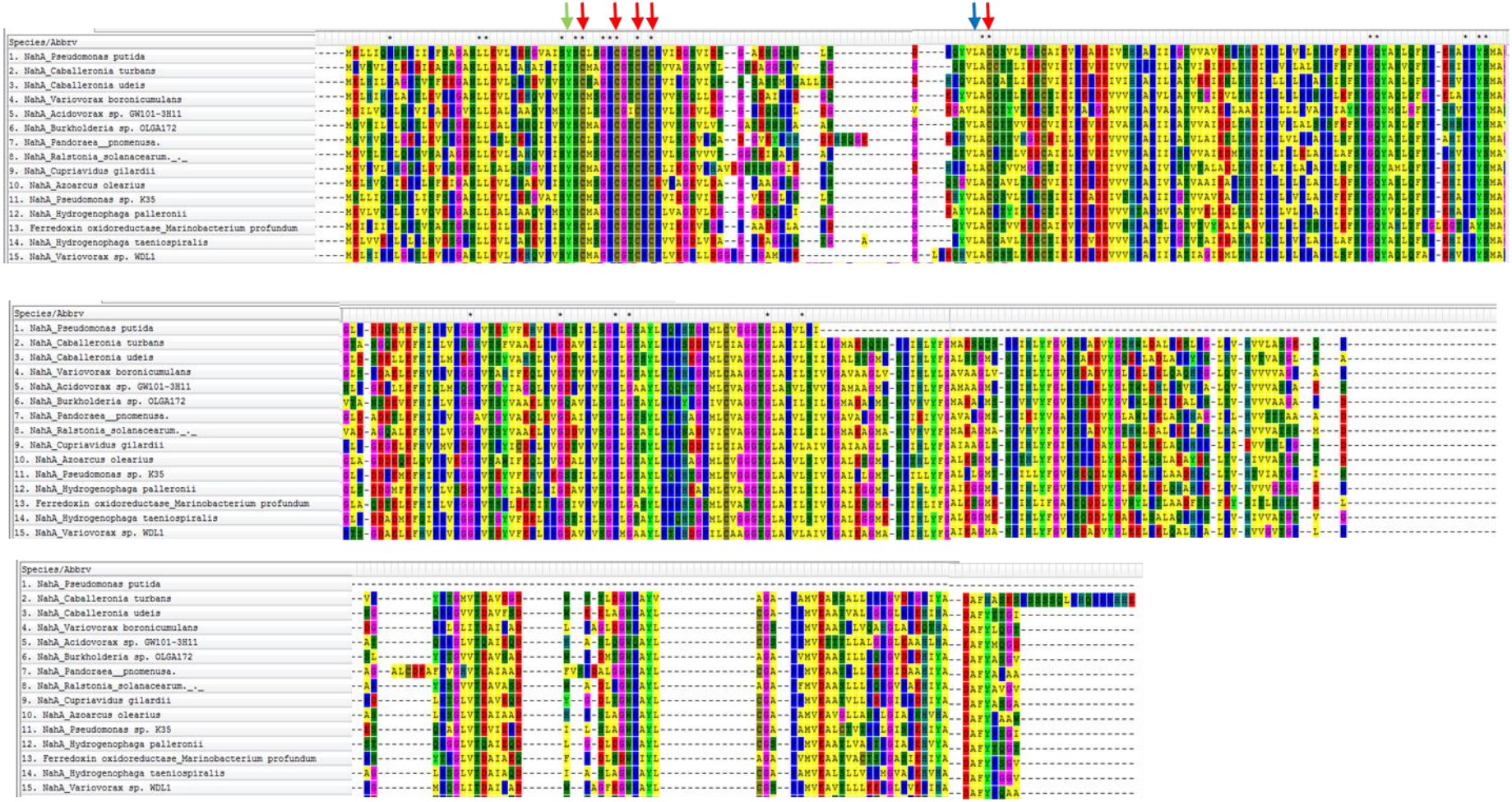

Figure 4: *A representative alignment (it is shown only 15 protein from the 224 used) of the proteins retrieved from the PSI-Blast. Conserved amino acids are shown with an asterisk.*

In order to create the phylogeny, an Maximum Likelihood (ML) tree was constructed using the multiple alignments of the 224 proteins retrieved from the PSI-Blast. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (28). The bootstrap consensus tree inferred from 100 replicates is taken to represent the evolutionary history of the proteins analyzed (29). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated proteins clustered together in the bootstrap test (100 replicates) are shown next to the branches (29). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites [4 categories (+*G*, parameter = 1.3815)]. The analysis involved 224 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 326 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (21). Proteins with less than 50% identity with the NahAa_*P. putida* used as bait in the PSI-Blast, have grouped in a single clade and they were used to root the tree (Blue box in Figure 5). The resulting tree is shown in Figure 5.

The phylogenetic tree includes proteins exclusively from β- (87.5%) and γ-proteobacteria (12.5%), which indicates that the metabolic route of naphthalene is present in a very specific range of bacteria (at least, taking into consideration the data available at the moment). Only 6 out of the 224 proteins (2.7%) are present in plasmids with the rest of them (97.3%) present in the chromosomes of the corresponding bacteria. The plasmids come either from β- or γ-proteobacteria and due to the low number of the sample, no statistical conclusions can be made. The tree showed that 5 out of 6 of the plasmids clustered in a monophyletic branch, well-supported by the bootstrap value (Figure 5, Clade B, bootstrap value: 92). This branch included the NahAa_*P. putida* used as bait in the PSI-Blast (Figure 5, indicated with a red circle). Ferredoxin oxydoreductases from catabolic routes of other aromatic compounds (eg, MntAa: 3-nitrotoluene dioxygenase ferredoxin reductase component and

DntAa: 2,4-dinitrotoluene dioxygenase) are grouped together (Figure 5, Clade A, bootstrap value: 95). A third set of proteins cluster in a second monophyletic branch (Figure 5, Clade C, bootstrap value 98), where all proteins from *Variovorax* are grouped. Proteins from γ-Proteobacteria appeared in an ancestral, monophyletic group (Figure 5, Clade E, bootstrap value: 94).

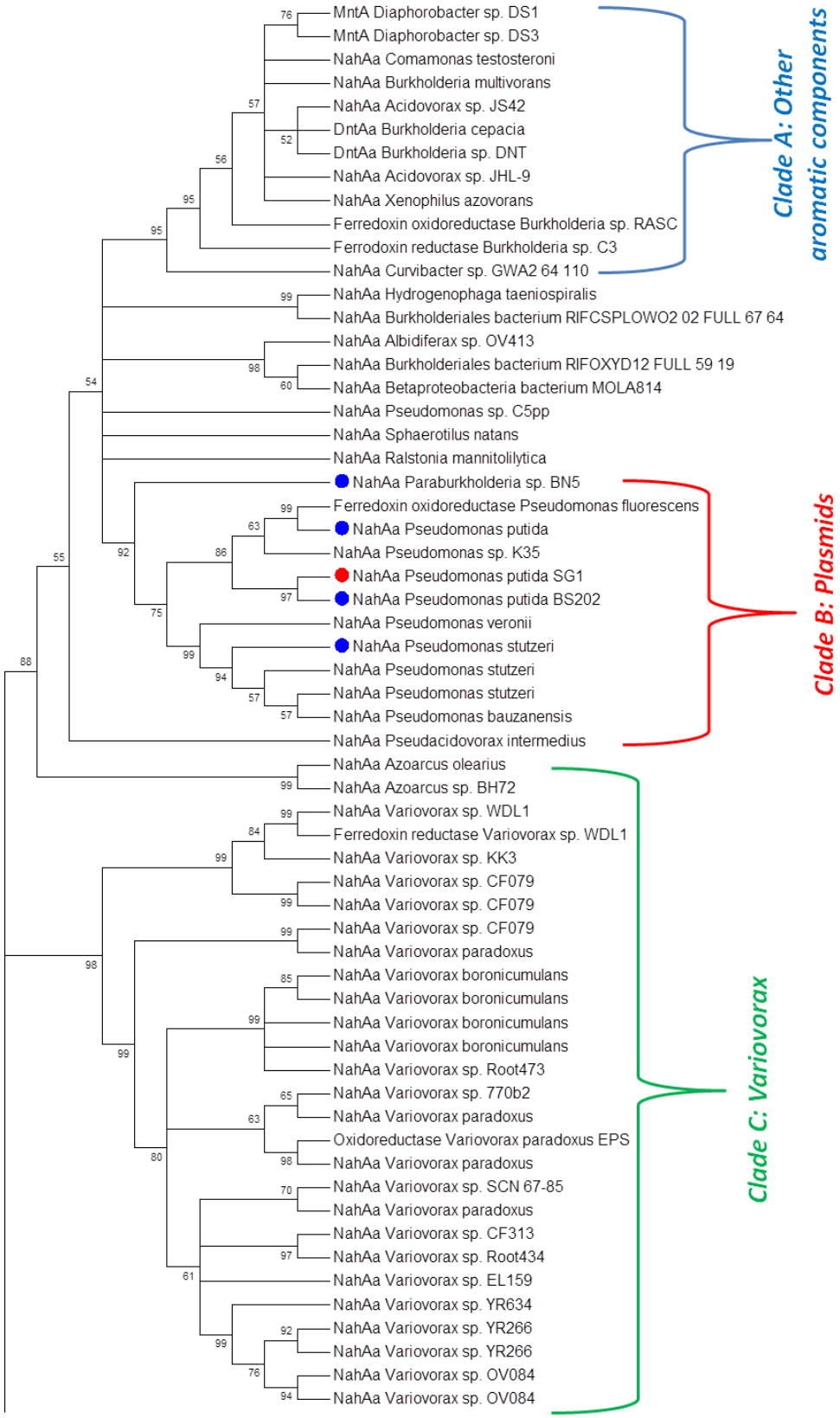

Figure 5: *The maximum-likelihood (ML) phylogenetic tree was built with the 224 NahAa homologues present in NCBI database up to October 2017. Bootstrap values are indicated at the corresponding nodes of the ML tree. The cut-off value for the condensed tree was chosen at bootstrap value=50%. The NahAa from P. putida SG1 is indicated with a red circle. Proteins from plasmids are indicated with a blue circle.*

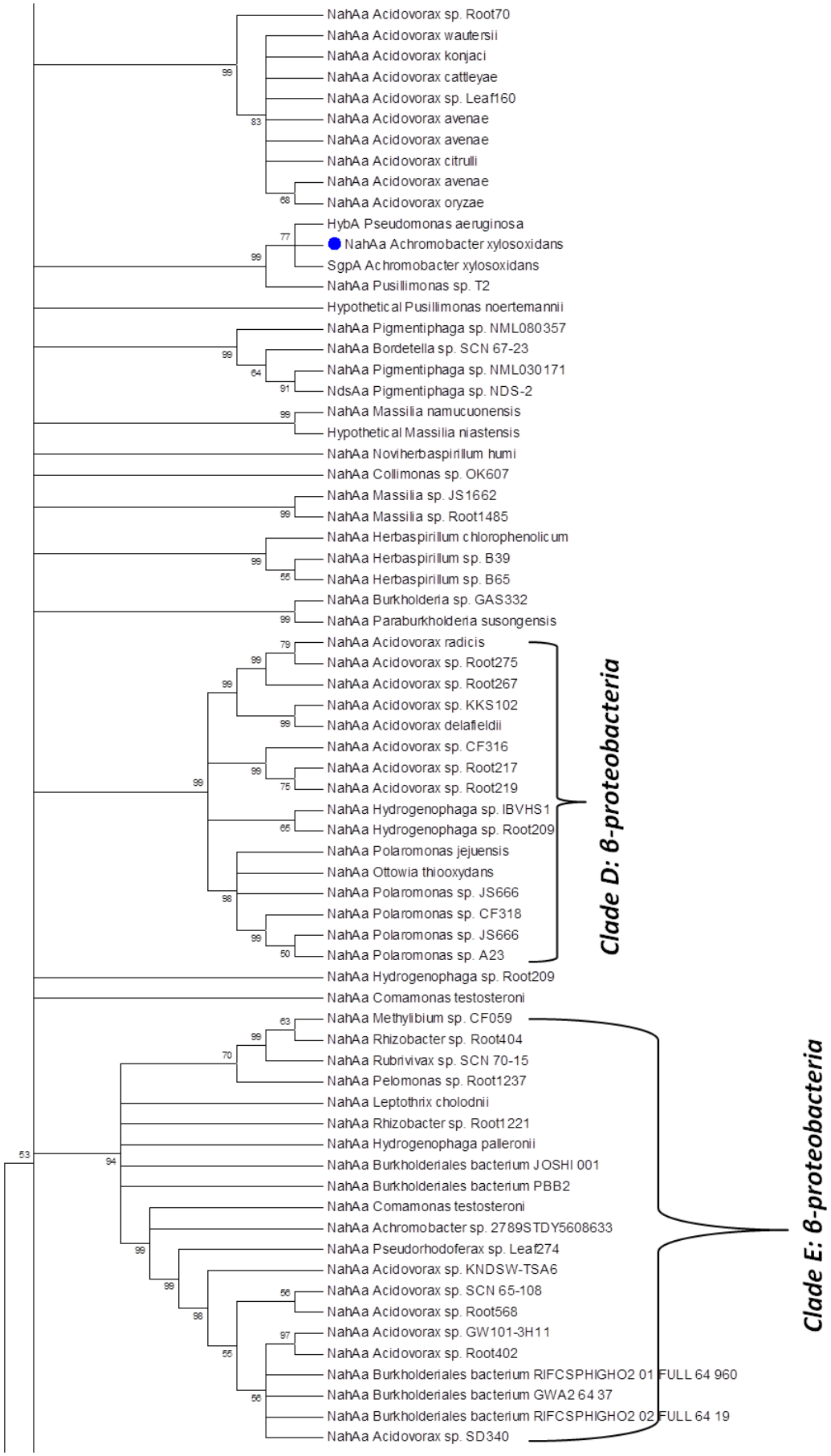

Figure 5: (*Continue)*

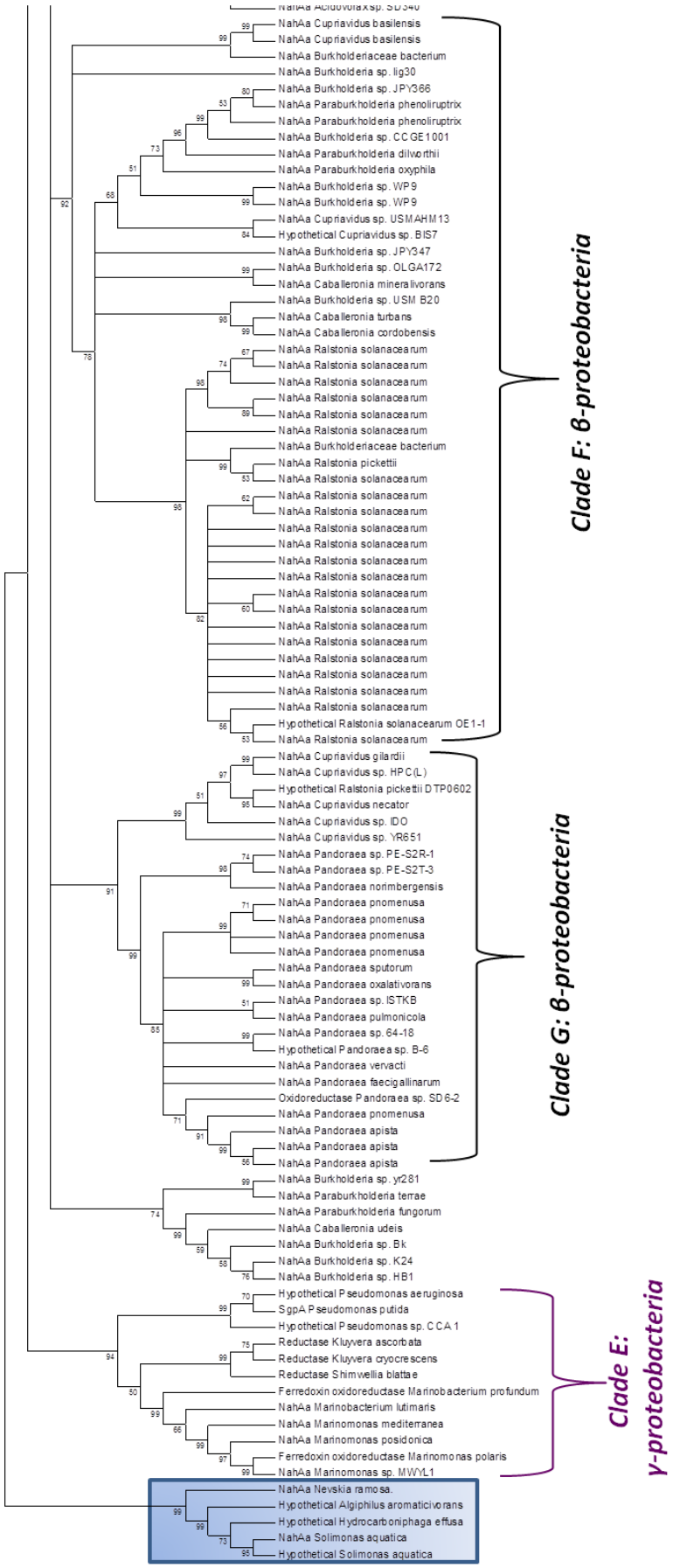
NahAa Cupriavidus basilensis
NahAa Cupriavidus basilensis
NahAa Burkholderiaceae bacterium
NahAa Burkholderia sp. lig30
NahAa Burkholderia sp. JPY366
NahAa Paraburkholderia phenoliruptrix
NahAa Paraburkholderia phenoliruptrix
NahAa Burkholderia sp. CCGE1001
NahAa Paraburkholderia dilworthii
NahAa Paraburkholderia oxyphila
NahAa Burkholderia sp. WP9
NahAa Burkholderia sp. WP9
NahAa Cupriavidus sp. USMAHM13
Hypothetical Cupriavidus sp. BIS7
NahAa Burkholderia sp. JPY347
NahAa Burkholderia sp. OLGA172
NahAa Caballeronia mineralivorans
NahAa Burkholderia sp. USM B20
NahAa Caballeronia turbans
NahAa Caballeronia cordobensis
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Burkholderiaceae bacterium
NahAa Ralstonia pickettii
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
NahAa Ralstonia solanacearum
Hypothetical Ralstonia solanacearum OE1-1
NahAa Ralstonia solanacearum
Clade F: β-proteobacteria
NahAa Cupriavidus gilardii
NahAa Cupriavidus sp. HPC(L)
Hypothetical Ralstonia pickettii DTP0602
NahAa Cupriavidus necator
NahAa Cupriavidus sp. IDO
NahAa Cupriavidus sp. YR651
NahAa Pandoraea sp. PE-S2R-1
NahAa Pandoraea sp. PE-S2T-3
NahAa Pandoraea norimbergensis
NahAa Pandoraea pnomenusa
NahAa Pandoraea pnomenusa
NahAa Pandoraea pnomenusa
NahAa Pandoraea pnomenusa
NahAa Pandoraea sputorum
NahAa Pandoraea oxalativorans
NahAa Pandoraea sp. ISTKB
NahAa Pandoraea pulmonicola
NahAa Pandoraea sp. 64-18
Hypothetical Pandoraea sp. B-6
NahAa Pandoraea vervacti
NahAa Pandoraea faecigallinarum
Oxidoreductase Pandoraea sp. SD6-2
NahAa Pandoraea pnomenusa
NahAa Pandoraea apista
NahAa Pandoraea apista
NahAa Pandoraea apista
Clade G: β-proteobacteria
NahAa Burkholderia sp. yr281
NahAa Paraburkholderia terrae
NahAa Paraburkholderia fungorum
NahAa Caballeronia udeis
NahAa Burkholderia sp. Bk
NahAa Burkholderia sp. K24
NahAa Burkholderia sp. HB1
Hypothetical Pseudomonas aeruginosa
SgpA Pseudomonas putida
Hypothetical Pseudomonas sp. CCA1
Reductase Kluyvera ascorbata
Reductase Kluyvera cryocrescens
Reductase Shimwellia blattae
Ferredoxin oxidoreductase Marinobacterium profundum
NahAa Marinobacterium lutimaris
NahAa Marinomonas mediterranea
NahAa Marinomonas posidonica
Ferredoxin oxidoreductase Marinomonas polaris
NahAa Marinomonas sp. MWYL1
Clade E: γ-proteobacteria
NahAa Nevskia ramosa
Hypothetical Algiphilus aromaticivorans
Hypothetical Hydrocarboniphaga effusa
NahAa Solimonas aquatica
Hypothetical Solimonas aquatica

Figure 5: (End)

2.3 Localization of the genes/operons

From the list of 224 proteins retrieved from the PSI-Blast (Table ST1 in the section of “Anexos”), only 50 corresponded to completely sequenced plasmids or genomes (Table ST2 and ST3 respectively, in the section of “Anexos”). The rest of them (174 proteins) corresponded to partially sequenced/Scaffolds or contigs and were not analysed any further. Sequences were visualised with Artemis16.0.0 (30) and the naphthalene 1,2-dioxygenase gene was spotted in the nucleotide sequence. Putative operons were predicted using the FGENEB package of the www.softberry.com server (31) and the graphics of the operon genetic organization were designed using the Vector NTI Advance 10 (ThermoFisher Scientific). Genome-to-genome analysis was performed with EasyFig2.2.2 (32). The analysis showed that the putative naphthalene operons do not have the same genetic organization, varying from 3 up to 12 genes per operon (Table ST2 and ST3 in the section of “Anexos”). In particular were found:

- two putative operon with 12 genes
- two putative operons with 11 genes
- four putative operons with 9 genes
- sixteen putative operons with 7 genes (10 of them belonging to the genus *Ralstonia*)
- three putative operons with 6 genes
- twenty one putative operons with 4 genes (12 of them belonging to the genus *Pandoraea*)
- two putative operons with 3 genes

2.3.1 Putative nah catabolising operons composed by 12 genes

The two putative operons with 12 genes were from *Polaromonas sp.* JS666 and *Burkholderia multivorans* strain DDS 15A-1. A genome-to-genome analysis showed that the putative operon from *Polaromonas sp.* JS666 shows a very low similarity to the nah operon described in *P. putida* (Figure 6).

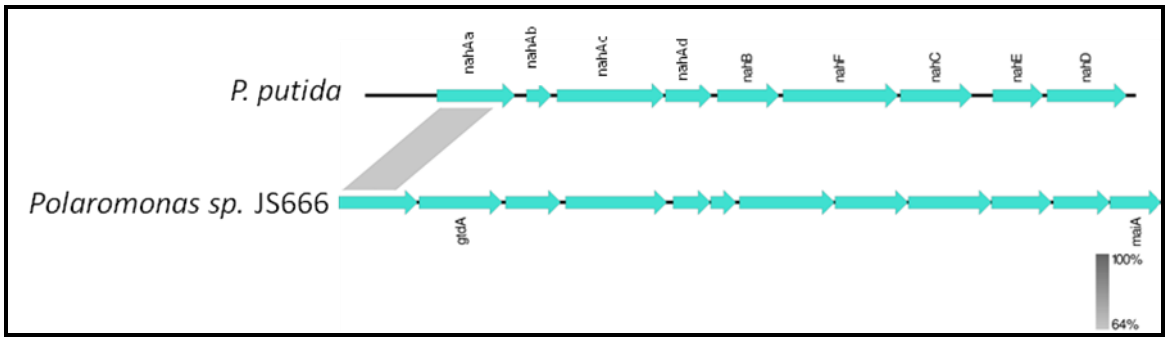

Figure 6: *Genome-to-genome comparison of the nah operon from P. putida strain BS202 with the putative operon from Polaromonas sp. JS666. Analysis was performed with EasyFig2.2.2.*

The first CDS codes for a naphthalene 1,2-dioxygenase (Figure 7), was the one retrieved from the PSI-Blast and is the only gene in common with the nah operon described in *P. putida*.

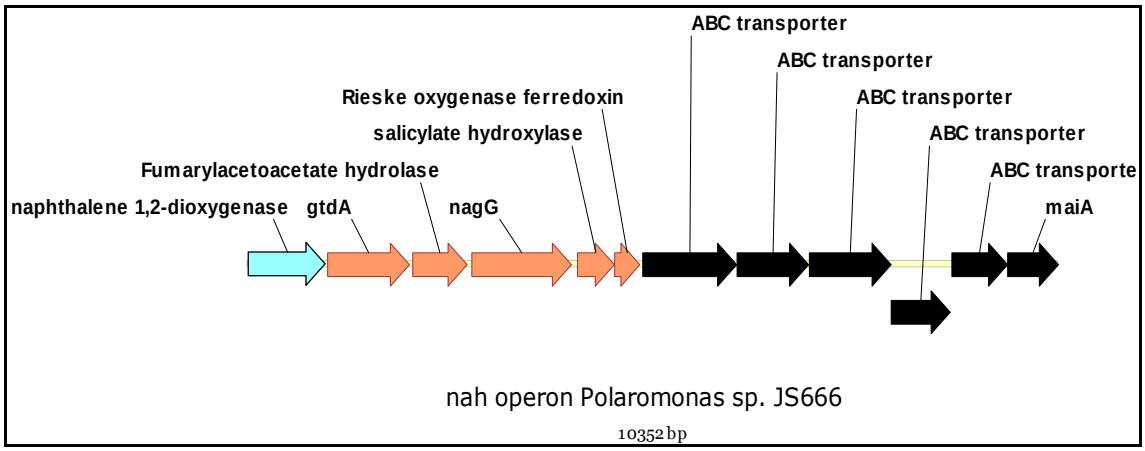

Figure 7: *Genetic organization of the putative nah operon from Polaromonas sp. JS666.* ***Color code****: genes reported as involved in naphthalene degradation are shown in light blue color, genes reported as involved in degradation of other aromatic compounds are shown in light orange colour, genes of other metabolic pathways are shown in black.*

The rest of the genes composing the putative operon that show interest from the degradation point of view are:

- *gtdA*: codes for gentisate 1,2-dioxygenase, an enzyme that catalyzes the chemical reaction in which the metabolite 2,5-dihydroxybenzoate is converted to pyruvate.

- The third ORF codes for a protein belonging to the fumarylacetoacetate (FAA) hydrolase family. FAA is the last enzyme in the tyrosine catabolic pathway, it hydrolyses fumarylacetoacetate into fumarate and acetoacetate which then join the citric acid cycle. This family also includes various hydratases and 4-oxalocrotonate decarboxylases which are involved in the bacterial meta-cleavage pathways for degradation of aromatic compounds (33).
- The forth and the fifth CDS code for the two subunits of the salicylate hydroxylase, whereas the sixth CDS codes for a Rieske oxygenase ferredoxin. The Rieske domain is a binding domain commonly found in Rieske [2Fe-2S] non-heme iron oxygenase systems, such as naphthalene and biophenyl dioxygenases.

The above indicate that *Polaromonas* sp. JS666 probably is using an alternative metabolic route, converting naphthalene into fumarate and pyruvate via salicylate (2-hydroxybenzoate) and gentisate (34) rather than the *meta* cleavage pathway of catechol. The steps of this pathway are shown in Figure 8 (34).

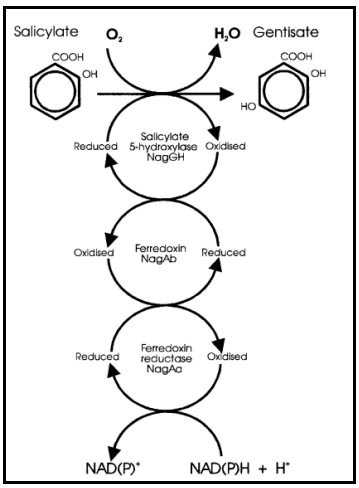

Figure 8: *Proposed model for the conversion of salicylate to gentisate.*

This pathway has been proposed for *Polaromonas naphthalenivorans* CJ2 and *Pseudomonas* sp. strain U2 (now called *Ralstonia* sp. strain U2) (34, 35).

On the other hand, the putative operon from *B. multivorans* strain DDS 15A-1 shows a higher similarity with the nah operon from *P. putida* (Figure 9).

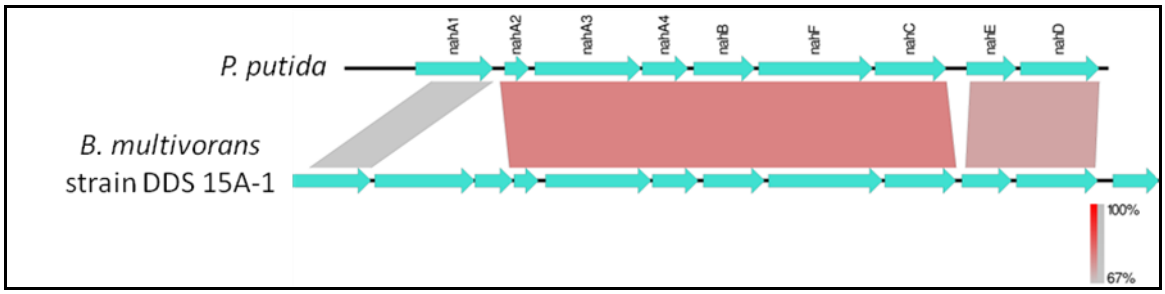

Figure 9: *Genome-to-genome comparison of the nah operon from P. putida strain BS202 with the putative operon from B. multivorans strain DDS 15A-1.*

Although it is very similar to the nah operon of *P. putida*, they are not identical. It contains additionally (Figure 10):

- *nag* and *nagH*, involved in the gentisate pathway mentioned for *Polaromonas* sp. JS666
- a CDS that codes for 3-(cis-5,6-dihydroxycyclohexa-1,3-dien-1-yl)propanoate dehydrogenase, an enzyme involved in the biphenyl degradation (36).
- an ORF coding for an OmpW family protein

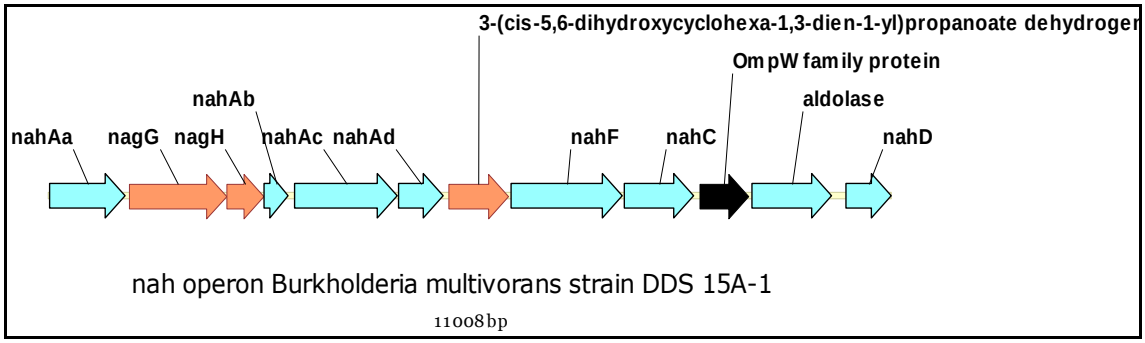

Figure 10: *Genetic organization of the putative nah operon from B. multivorans strain DDS 15A-1.* ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes; black: genes of other metabolic pathways.*

With a better eye-inspection of the genetic organization of the putative nah operon of *B. multivorans* strain DDS 15A-1 it can be observed that the putative CDS that codes for 3-(cis-5,6-dihydroxycyclohexa-1,3-dien-1-yl)propanoate dehydrogenase lays in the place of *nahB.* A protein–Blast of the putative transcriptional product of this CDS puts it in the cis-biphenyl-2,3-dihydrodiol-2,3-

dehydrogenase (BphB)-family, a classical short-chain dehydrogenase/reductase (SDR) of particular importance for its role in the degradation of biphenyl/polychlorinated biphenyls (PCBs). This includes *Pseudomonas* sp. C18 putative 1,2-dihydroxy-1,2-dihydronaphthalene dehydrogenase (*doxE* gene) which participates in the upper naphthalene catabolic pathway (37) and was present in the protein-Blast, along with other *nahB*-analogues. The second operon with the genes for the lower catabolic pathway (*nahGTHINLOMKJ*) encoding the necessary enzymes for the conversion of salicylate through the catechol meta-cleavage pathway to pyruvate and acetaldehyde was not found in the genome of *B. multivorans* strain DDS 15A-1. Nevertheless, just downstream of the operon shown in Figure 10, the putative genes *nagI* and *nagK* were found. Even though prediction with the FGENESB does not put them in the same operon, the putative operon could be composed by 17 genes and represent a combination of *nah* and *nag* CDS (Figure 11).

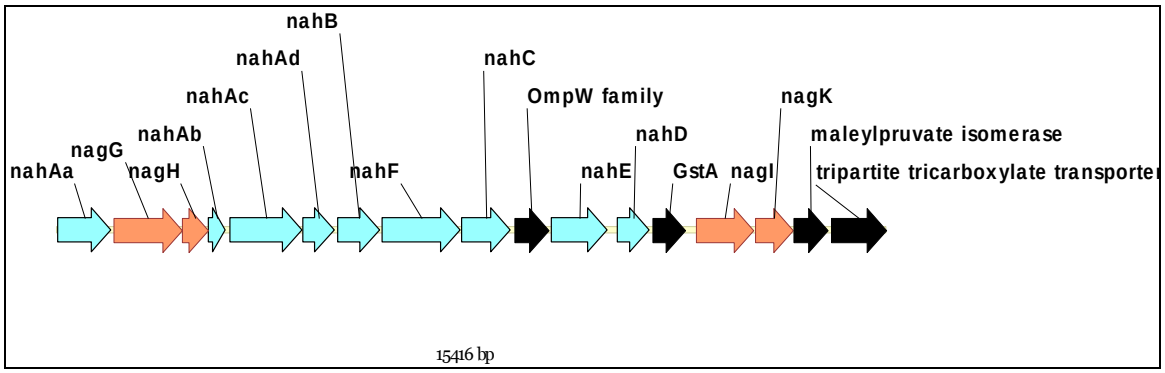

Figure 11: *Genetic organization of the putative nah operon from B. multivorans strain DDS 15A-1 composed by 17 putative CDS.* ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes; black: genes of other metabolic pathways.*

All the above could lead to the conclusion that *B. multivorans* strain DDS 15A-1 might be using an alternative naphthalene catabolic route, which combines the classical nah and the alternative gentisate pathway. Up-to-date, there is no information available regarding studies about naphthalene degradation form any *B. multivorans* strain.

2.3.2 Putative nah catabolising operons composed by 11 genes

The two putative operons with 11 genes were from *Leptothrix cholodnii* SP-6 and from the plasmid pBN2 from *Paraburkholderia* sp. BN5. *Leptothrix cholodnii* SP-6 shows the same organization as *Polaromonas* sp.JS666, but with the difference of having 11 genes instead of 12 (Figure 12). The 11 genes are the nag(AaIKGHAb), five ABC transporters and the gene missing is *maiaA*, coding for a maleylacetoacetate isomerase which catalyzes a cis-trans isomerization. The enzyme converts the first product of homogentisate oxidation, maleylacetoacetate, to the trans isomer, fumarylacetoacetate (38). Up-todate it is not clear if *maiaA* is necessary for the naphthalene degradation. Even though there is no experimental data available in leterature regarding the naphthalene degradation from *Leptothrix cholodnii* SP-6, based on the above information it could be proposed that it probably uses the "alternative" gentisate pathway for the degradation of naphthalene.

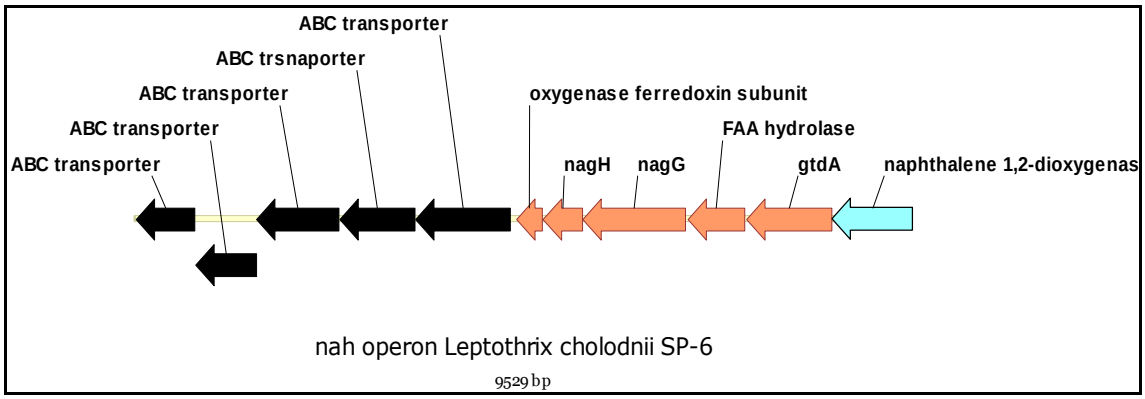

Figure 12: *Genetic organization of the putative nah operon from L. cholodnii SP-6.* ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes; black: genes of other metabolic pathways.*

The putative operon from the plasmid pBN2 from *Paraburkholderia* sp. BN5 (Figure 13) is very similar with the putative operon from *B. multivorans* strain DDS 15A-1 (Figure 7) mentioned before containing both genes from the upper nah operon from *P. putida* (*nahAaAbAcAdBFCED*) and the *nagG* and *nagH* from the gentisate pathway.

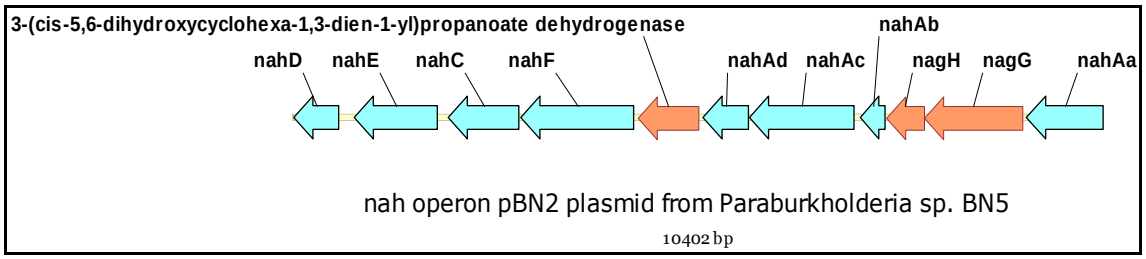

Figure 13: *Genetic organization of the putative nah operon from pBN2 plasmid from Paraburkholderia sp. BN2.* ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes; black: genes of other metabolic pathways.*

As in the case of *B. multivorans* strain DDS 15A-1, the second operon with the genes for the lower catabolic pathway (*nahGTHINLOMKJ*) encoding the necessary enzymes for the conversion of salicylate through the catechol meta-cleavage pathway to pyruvate and acetaldehyde was not found. Nevertheless, just downstream of the operon shown in Figure 13, the putative genes *nagI* and *nagK* were found. Even though prediction with the FGENESB does not put them in the same operon, as in the case of *B. multivorans* strain DDS 15A-1, the putative operon could be composed by 17 genes and represent a combination of *nah* and *nag* CDS (Figure 14).

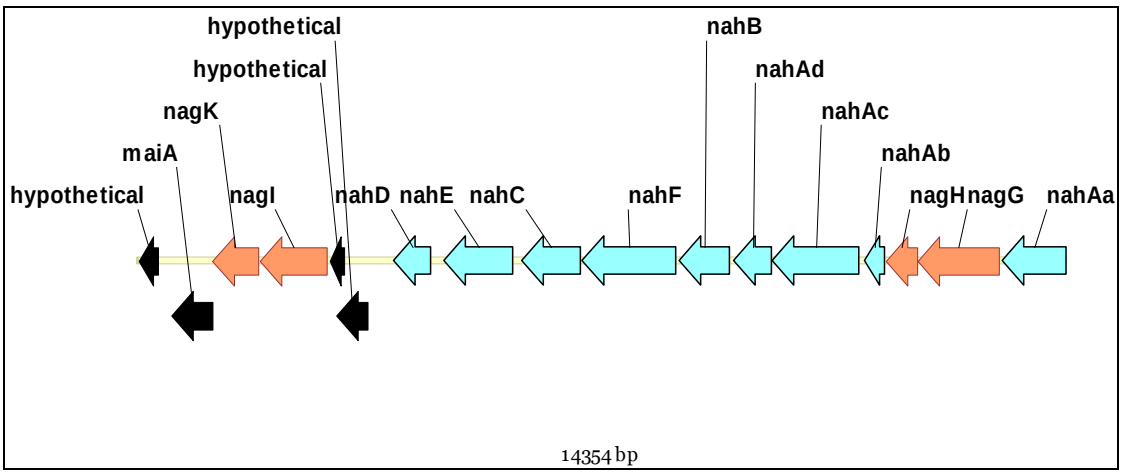

Figure 14: *Genetic organization of the putative nah operon from pBN2 plasmid from Paraburkholderia sp. BN2 composed by 17 DCS.* ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes; black: genes of other metabolic pathways.*

The similarity between the molecules mentioned before was visualized by a genome-to-genome comparison (Figure 15).

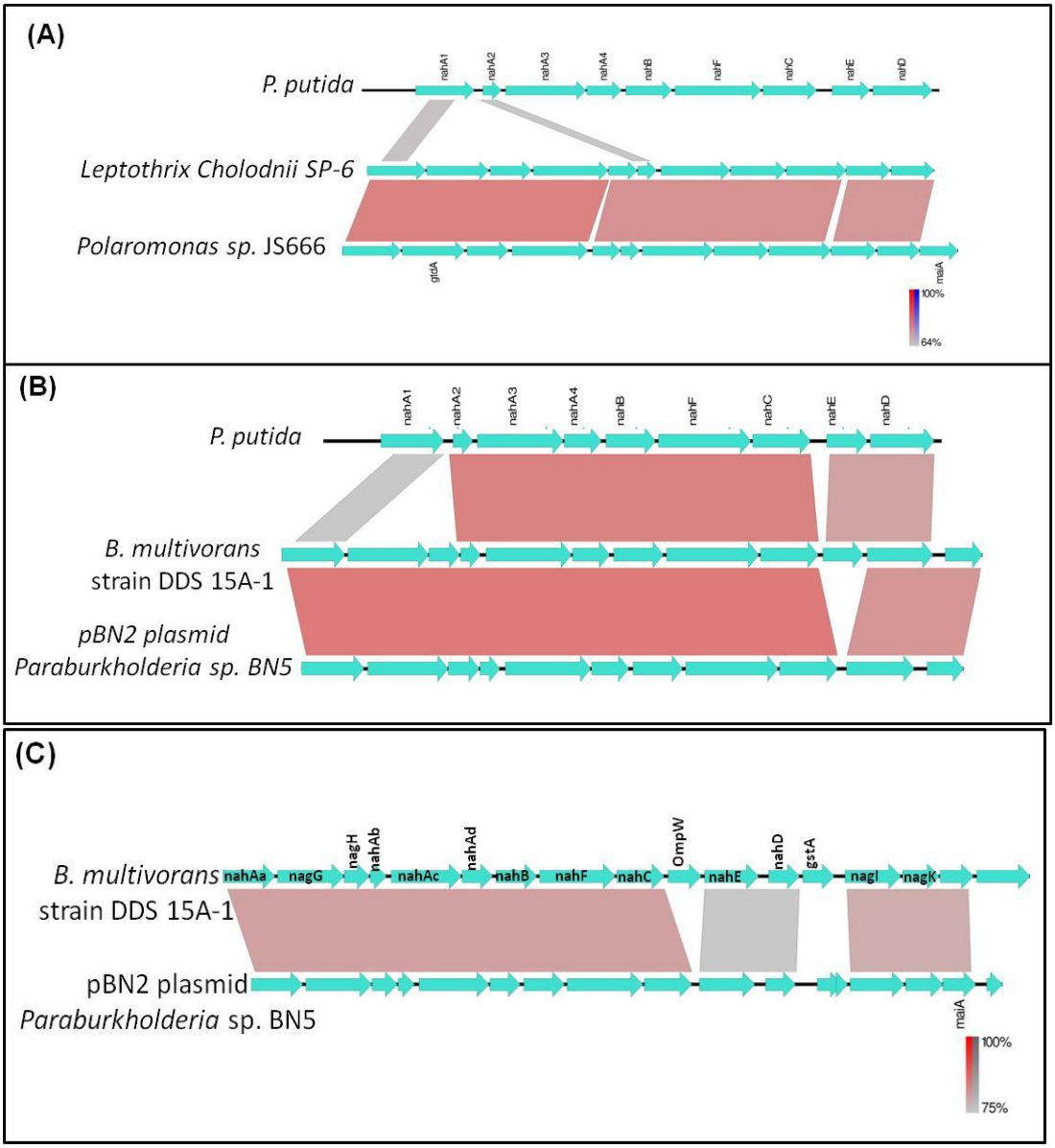

Figure 15: *Genome-to-genome comparison of the putative operons composed by 12 and 11 genes*

2.3.3 Putative nah catabolising operons composed by 9 genes

The members of this group are putative operons from four γ-proteobacteria: *Pseudomonas putida* strain BS202 (pNPL1 plasmid), *Pseudomonas stutzeri* strain 19SMN4, *Marinomonas* sp. MWYL1 and *Marinomonas posidonica* IVIA-Po-181. Genome-to-genome comparison showed that the putative nah operon from *P. putida* strain BS202 and *Pseudomonas stutzeri* strain 19SMN4 are very similar, whereas the *Marinomonas* are quite different from the other two but very similar among them (Figure 16).

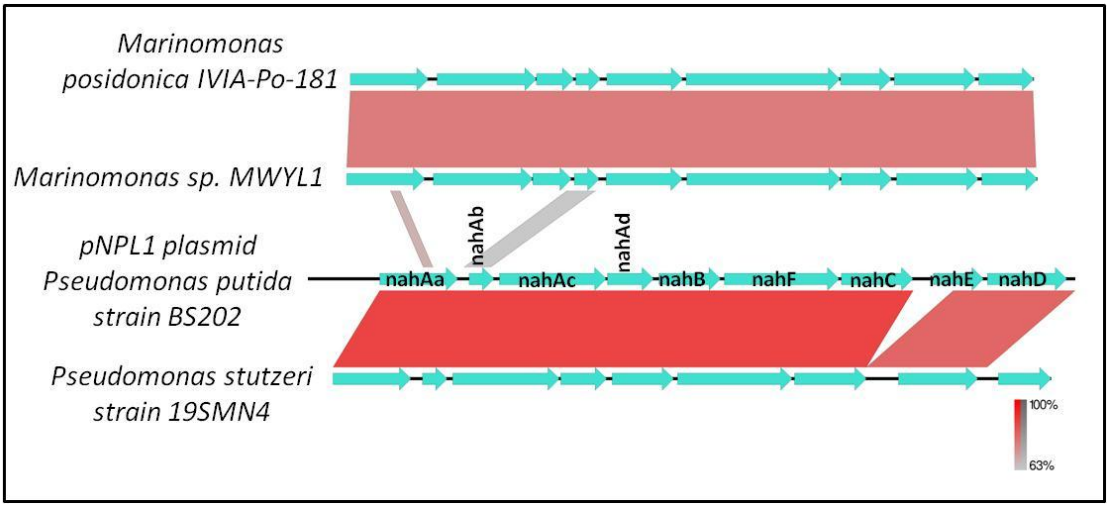

Figure 16: *Genome-to-genome comparison of the putative nah operons composed by 9 genes*

The putative nah operon from the pNPL1 plasmid from *P. putida* BS202 is identical to the “classical” nah operon described in literature (Figure 17).

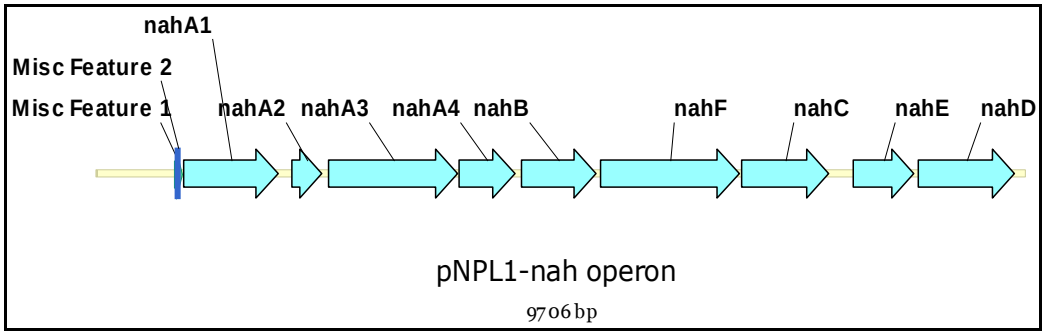

Figure 17: *Genetic organization of the putative nah from the pNPL1 plasmid from P. putida strain BS202.* ***Color code****: light blue: genes reported as involved in naphthalene degradation*

Almost identical is the putative nah operon from *Pseudomonas stutzeri* strain 19SMN4 (Figure 18).

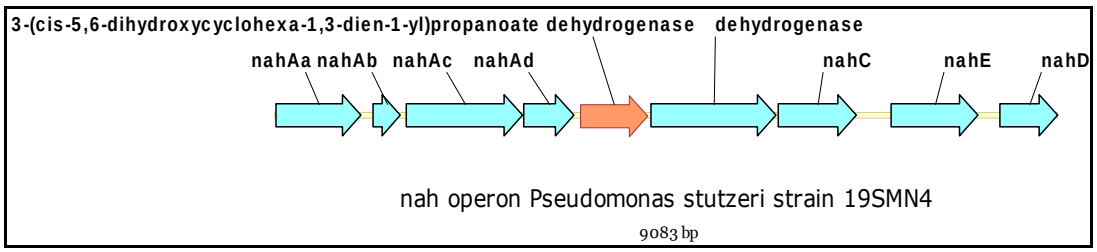

Figure 18: *Genetic organization of the putative nah operon from P. stutzeri strain 19SMN4.* ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes.*

As in the case of the putative operon from *B. multivorans* strain DDS 15A-1, the putative CDS that codes for 3-(cis-5,6-dihydroxycyclohexa-1,3-dien-1-yl)propanoate dehydrogenase lays in the place of *nahB.* A protein–BLAST of the putative transcriptional product of this CDS belongs it in the same family as *nahB*.

The third member of this group is the putative nah operon from *Marinomonas* sp. MWYL1 and as shown in Figure 16, is quite different from the other two from *Pseudomonas*. Its genetic organization is shown in Figure 19 and it is identical to the genetic organization of the putative nah operon from *M. posidonica* IVIA-Po-181 (not shown).

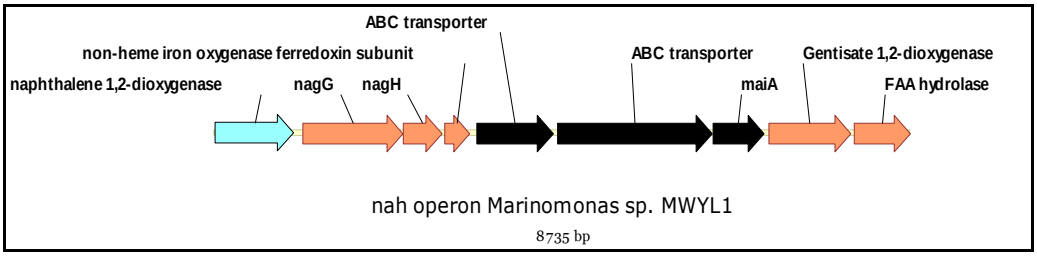

Figure 19: *Genetic organization of the putative nah operon from Marinomonas sp. MWYL1.* ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes; black: genes of other metabolic pathways.*

From an eye-inspection of both *Marinomonas* putative operons, they seem similar to the putative nah operons from *Polaromonas* JS666. A genome-to-genome comparison confirmed this observation (Figure 20), showing that the putative operons from the *Marinomonas* strains contain the six gene of the gentisate pathway, leading to the conclusion that they might share the same mechanism of catabolising naphthalene. Up-to-date there is no information available in the literature about any studies of naphthalene degradation from any *Marinomonas* strain.

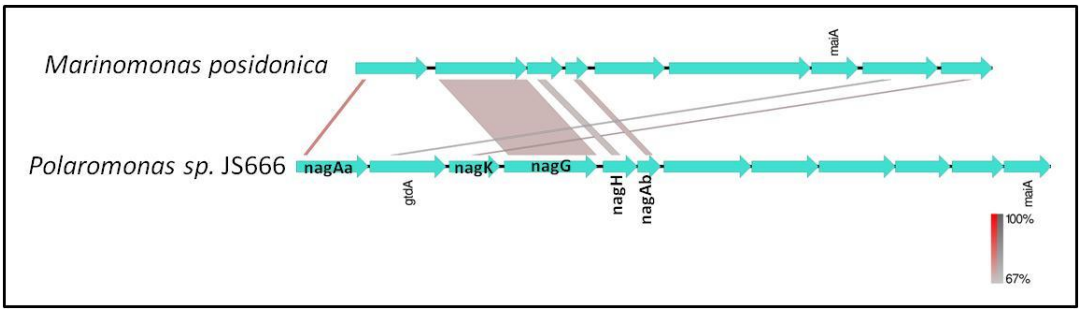

Figure 20: *Genome-to-genome comparison between the putative nah operons from strains M. posidonica MWYL1, M. posidonica IVIA-Po-181 and Polaromonas JS666*

2.3.4 Putative nah catabolising operons composed by 7 genes

This group contains sixteen members of putative operons, 10 of them from bacteria belonging to the *Ralstonia* genus. The other six are: two from *Burkholderia,* one from *Marinomonas*, two from *Paraburkholderia* and one from *Variovorax*.

All the putative nah operons from *Ralstonia solanacearum* strains have the same genetic organization (Figure 21) and genome-to-genome analysis showed that they are almost identical (Figure 22A).

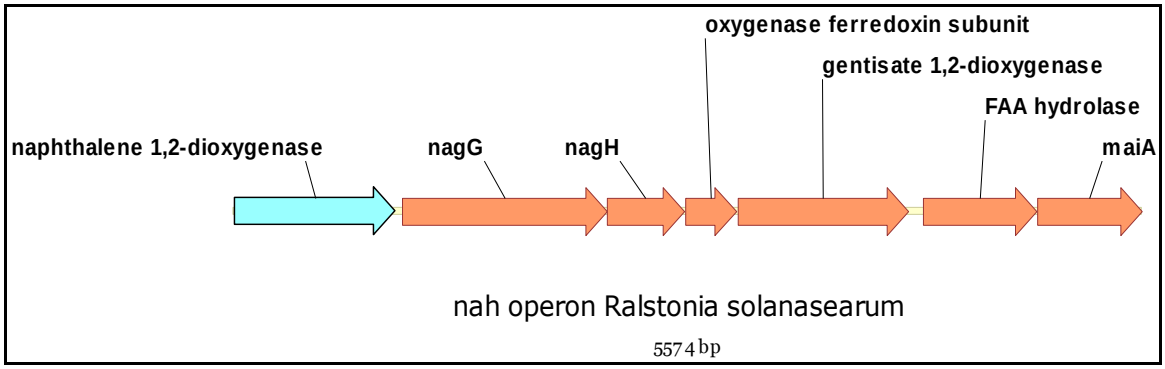

Figure 21: *Genetic organization of the putative nah operon from Ralstonia solanasearum.* ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes.*

The putative nah operons from the *R. solanasearum* strains are similar to the *Polaromonas* putative nah operon (Figure 22B). Although that up-to-date there is no information available in the literature about any studies of naphthalene degradation from any *Ralstonia* strain, the similarity of the putative operons with that of *Polaromonas* might indicate that they also follow the gentisate pathway for the degradation of naphthalene.

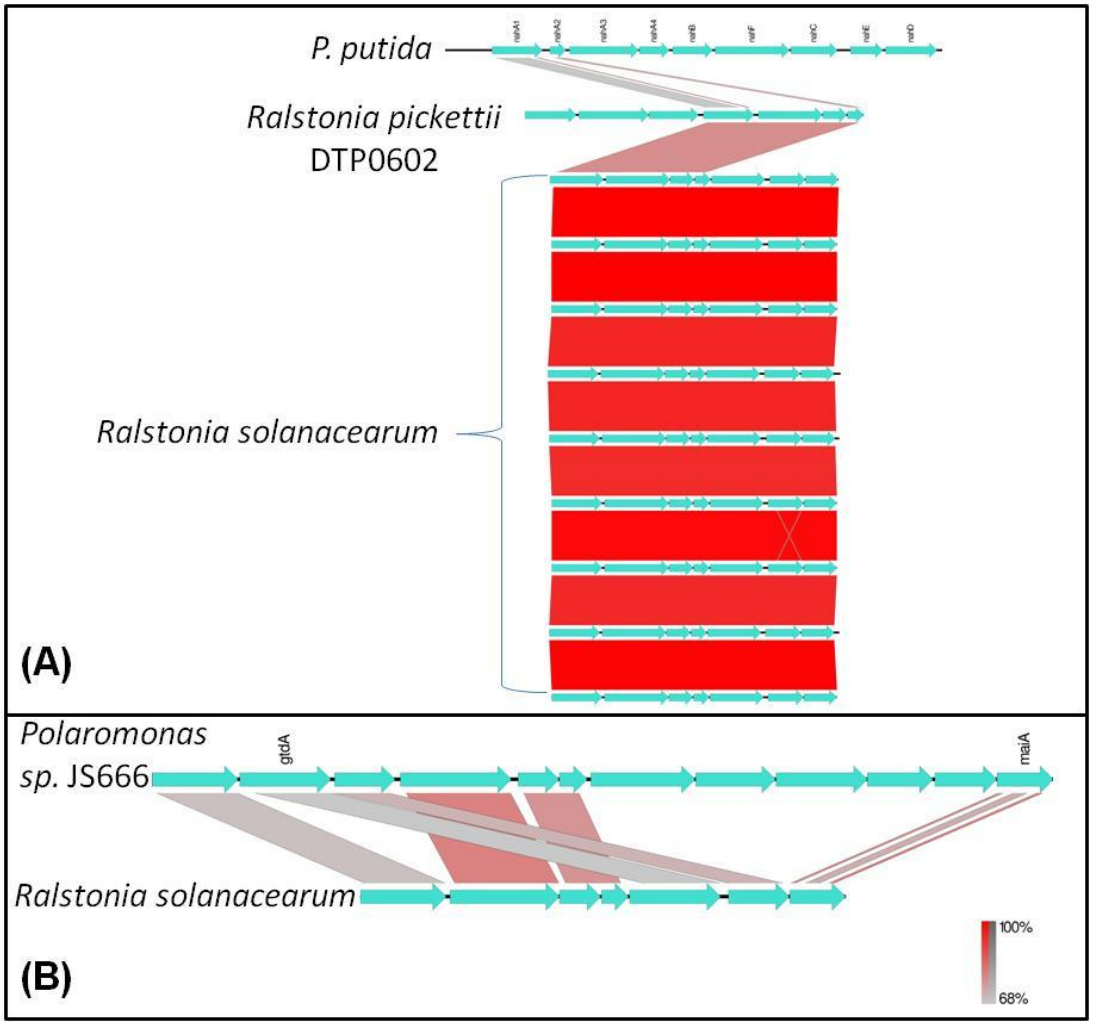

Figure 22: (A) *Genome-to-genome comparison of the putative nah operons (7 genes) from the R. solanacearum* ***(B)*** *Genome-to-genome comparison of the putative nah operons between R. solanacearum and Polaromonas sp JS666.*

The putative nah operon from from *Ralstonia pickettii* DTP0602 is different from the other *Ralstonia* putative operons (Figure 22A). Its genetic organization is shown in Figure 23. Three of the putative CDS are annotated as "hypothetical". A Protein-Blast of those three putative CDS showed that:

- The first CDS annotated as "hypothetical" belongs to the Bug (*Bordetella* uptake gene) protein family of periplasmic solute-binding receptors
- The second CDS annotates as "hypothetical" belongs to the D-3-Phosphoglycerate Dehydrogenases family
- The third CDS annotated as "hypothetical" belongs to the Class I oxygenase reductases, enzymes are that contain a reductase with Rieske type [2Fe-2S] redox center and an oxygenase

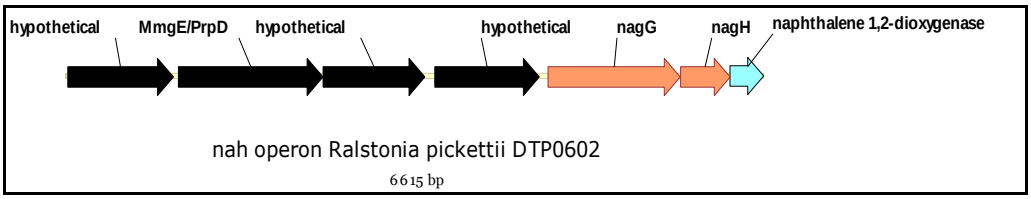

Figure 23: *Genetic organization of the putative nah operon from R. pickettii DTP0602.* ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes; black: genes of other metabolic pathways.*

Genome-to-genome comparison with the putative operon of *Polaromonas* sp JS666 showed that the putative nah operon of *R. pickettii* DTP0602 contains only four out of six of the genes that compose the gentisate-salicylate pathway used to catabolise naphthalene (operon 1 in Figure 24A). The other two genes were found in a different operon located at positions 935371-940587 on the chromosome of *R. pickettii* DTP0602 (operon 2 in Figure 24B). All the above indicate that *R. pickettii* DTP0602 might use the gentisate-salicylate pathway in order to catabolise naphthalene.

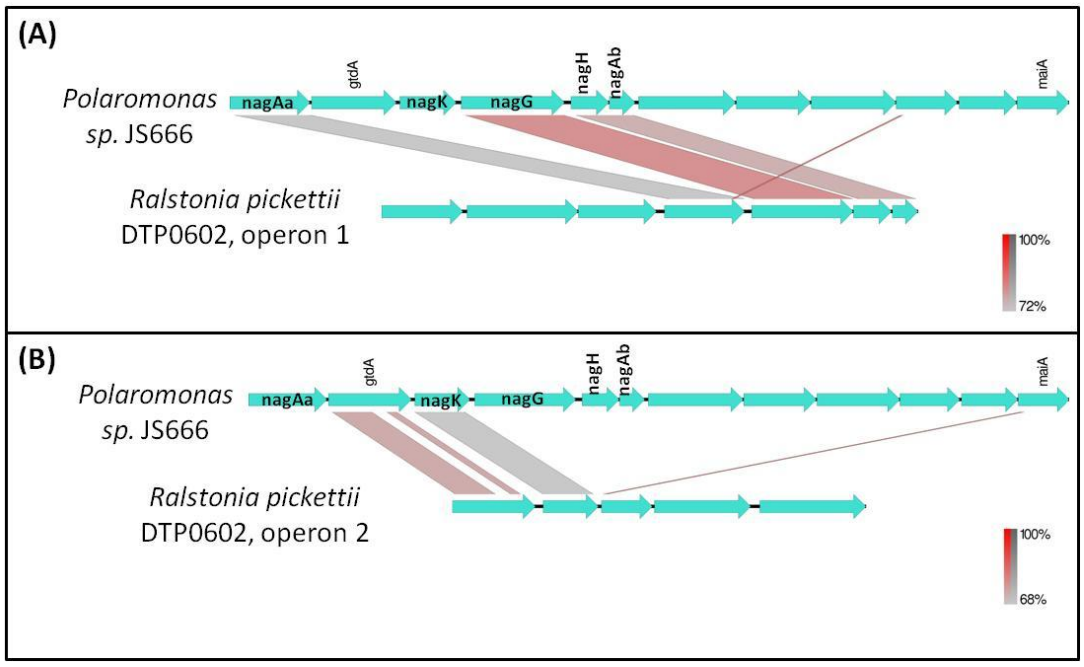

Figure 24: *Genome-to-genome comparison of the putative nah operons from Polaromonas sp JS666 and putative operons from the R. pickettii DTP0602* ***(A)*** *putative nah operon (operon 1) and* ***(B)*** *putative operon 2 (located at positions 935371-940587).*

The other six putative operons of this group are two from *Burkholderia,* one from *Marinomonas*, two from *Paraburkholderia* and one from *Variovorax*. The *Burkholderia* and the *Paraburkholderia* putative operons are a “classical” example of the gentisate-salicylate pathway, even though they do not strictly respect the order of the genes (Figure 25).

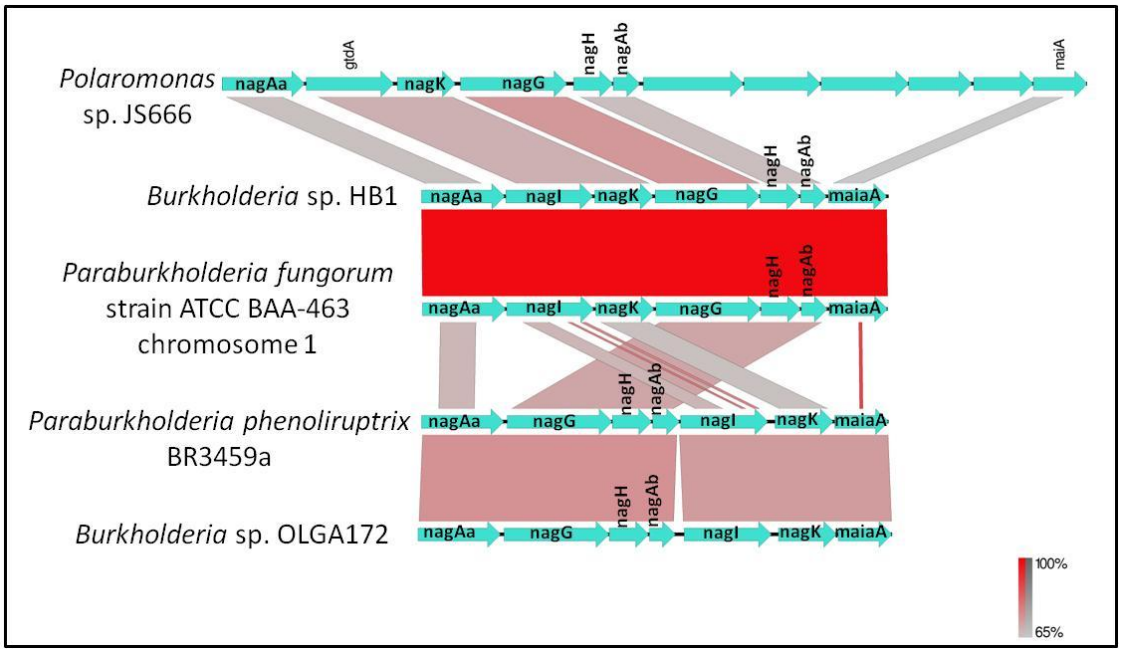

Figure 25: *Genome-to-genome comparison of the putative nah operons from Polaromonas sp JS666 and the Burkholderia and Paraburkholderia members of operon group containing 7 genes.*

In the case of the putative operons from *Marinomonas mediterranea* MMB-1 and *Variovorax boronicumulans* strain J1, they are different from all mentioned above. The genetic organization of the putative operon from *Variovorax boronicumulans* strain J1 is shown in Figure 26. From an eye-inspection it can be assumed that there are four out of six genes composing the gentisate-salicylate pathway. The two genes missing are *nagI* and *nagK*, coding for gentisate 1,2-dioxygenase and salicylate 5-hydroxylase, respectively.

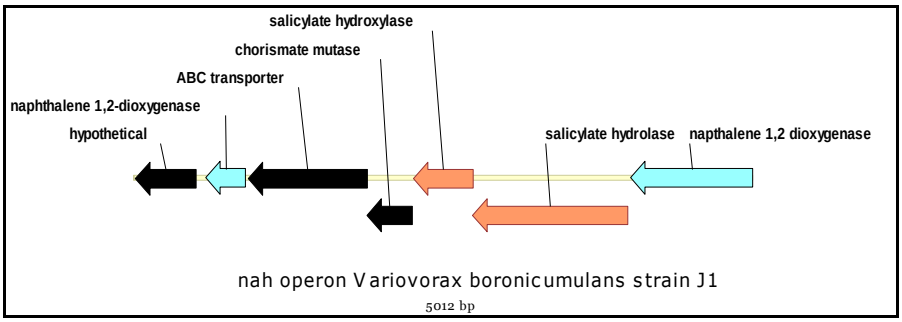

Figure 26: *Genetic organization of the putative nah operon from V. boronicumulansi* strain J1. ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes; black: genes of other metabolic pathways.*

As in the case of *R. pickettii* DTP0602 described before, the other two genes were found in a different operon located at positions 437409-448537 on the chromosome of *V. boronicumulans* strain J1 (operon 2 in Figure 27) and by that way the gentisate-salicylate pathway is completed and could be used in order to catabolise naphthalene.

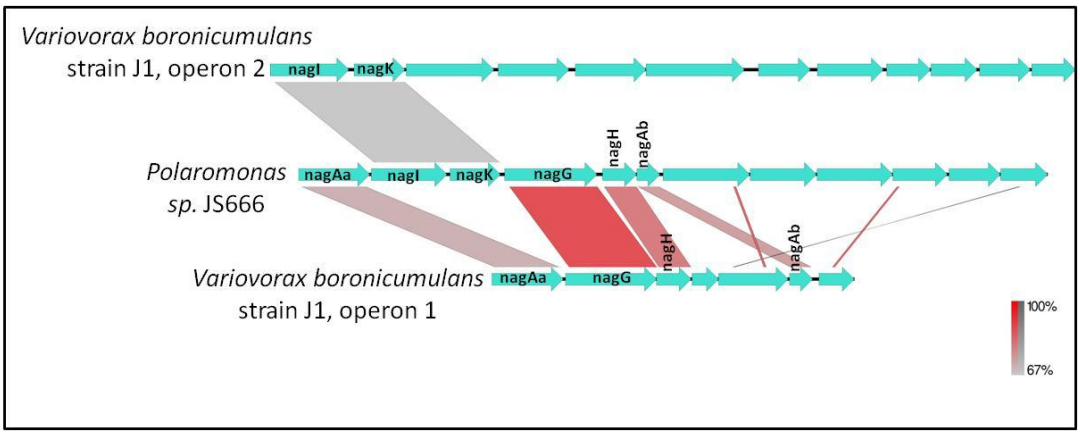

Figure 27: *Genome-to-genome comparison of the putative nah operons from Polaromonas sp JS666 and putative operons from the V. boronicumulans strain J1 putative nah operon (operon 1) and putative operon 2 (laying in position 437409-448537).*

The genetic organization of the putative nah operon from *M. mediterranea* MMB-1 is given in Figure 28.

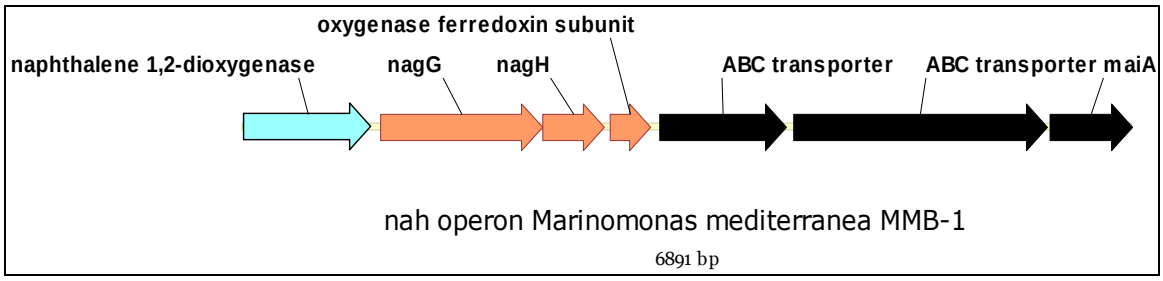

Figure 28: *Genetic organization of the putative nah operon from M. mediterranea MMB-1.* ***Color code****: light blue: genes reported as involved in naphthalene degradation; orange: genes reported as involved in other degradation routes; black: genes of other metabolic pathways.*

It is composed by four of the six genes of the gentisate-salicylate pathway. The two genes missing are *nahI* and *nagK*, which they were spotted in the positions 2403839-2410843 in the *M. mediterranea* MMB-1 genome, completing in this way the gentisate-salicylate pathway.

2.3.5 Putative nah catabolising operons composed by 6 genes

This group of operons is composed of three members only and they are the putative operons from *Acidovorax* sp. JS42, *Burkholderia* sp. CCGE1001 and from the plasmid pAK5 of *Pseudomonas putida* AK5. The only one experimentally described in the literature is the case of the spg-operon from plasmid pAK5 of *P. putida* (39). Its gene organization is shown in Figure 29 and includes six open reading frames (ORFs) (*sgpAIKGHB*). The four ORFs code for the entire salicylate 5-hydroxylase oxidoreductase component (*sgpA*), large and small subunits of the oxigenase component (*sgpG* and *sgpH*) and [2Fe-2S] ferredoxin (*sgpB*). Genes for gentisate 1, 2-dioxygenase (*sgpI*) and fumarylpyruvate hydrolase (*sgpK*) are located in salicylate 5-hydroxylase genes clustering between *sgpA* and *sgpG*. The salicylate 5-hydroxylase ferredoxin reductase (SgpA) is the protein retrieved from the PSI-BLASTP search conducted, using naphthalene 1,2-dioxygenase system ferredoxin-NAD(P)$^+$ reductase component (NahAa, Acc. Number: AAS79488.1) from *Pseudomonas putida* as a query.

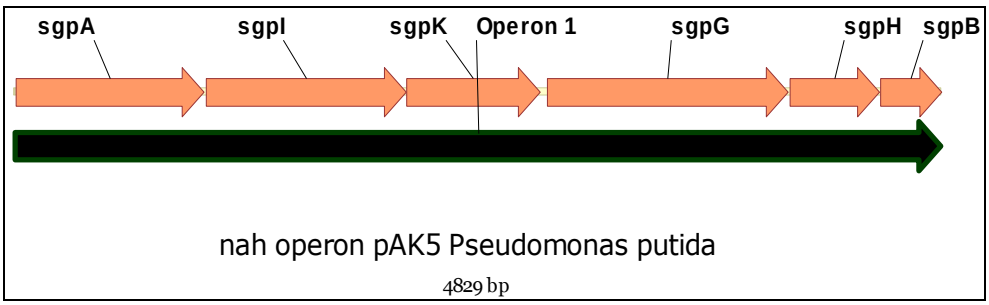

Figure 29: *The spg operon from pAK5 plasmid of P. putida, involved in the salicylate degradation through gentisate.*

Surprisingly, pAK5 plasmid also contains the classical nah-operon and *P. putida* AK5 is the first naturally occurring pseudomonas strain to be described that combines an operon of salicylate degradation via gentisate with a “classical” nah1 operon (*nahAaAbAcAdBFCED*). Degradation of naphthalene to salicylate in strain AK5 is encoded by the nah1 operon similar to nah1 operons in plasmid pNAH7. In contrast to plasmid mentioned above, the nah2 operon (*nahGTHINLOMKJ*) controlling degradation of salicylate through catechol is absent. Genes for salicylate catabolism are localized separately from the nah1

genes and are organized in an operon. The new operon was designated the sgp-operon (salicylate-gentisate pathway).

Genome-to-genome comparison of the putative operons of this group (Figure 30) revealed that the putative operon from *Burkholderia* sp. CCGE1001 is shares homology with the sgp-operon from the pAK5 plasmid of *P. Putida*, but is not a sgp-operon. It is a “classical” nag operon of salicylate degradation via gentisate. Up-to-date there are no experimental information of naphthalene degradation from any *Burkholderia* strain, but considering the above genome comparison, it can be proposed that *Burkholderia* sp. CCGE1001 could catabolise naphthalene through the gentisate pathway.

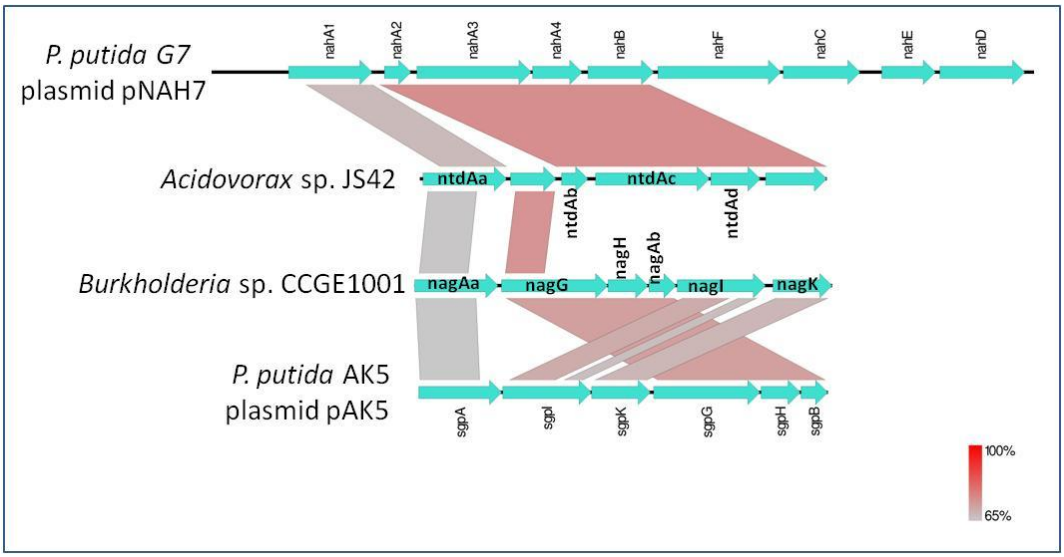

Figure 30: *Genome-to-genome comparison of the putative operons members of group containing 6 genes.*

In the case of the operon from *Acidovorax* sp. JS42, even thought the first CDS was retrieved from the PSI-BLAST and has a 66% identity with naphthalene 1,2-dioxygenases, it is not a nah catabolising operon. This operon codes for 2-nitrotoluene 2,3-dioxygenase (2NTDO), a multicomponent enzyme system that adds both atoms of molecular oxygen to nitroarene substrates, forming nitrite and (methyl)catechol and forms part of the 2-nitrotoluene degradation (40). The 2-nitrotoluene degradation pathway is shown in Figure 31 and in the first step degradation occurs by dioxygenation of the aromatic ring by 2-nitrotoluene 2,3-dioxygenase (2NTDO, encoded by *ntdAaAbAcAd*), resulting

in formation of 3-methylcatechol (3MC) and nitrite. The aromatic ring of 3MC is then cleaved at the meta position by a catechol 2,3-dioxygenase, and the product is further degraded to compounds that enter the tricarboxylic acid (TCA) cycle. The catechol degradation (Ctd) enzymes were identified based on the analysis of the *Acidovorax* sp. JS42 genome sequence (41).

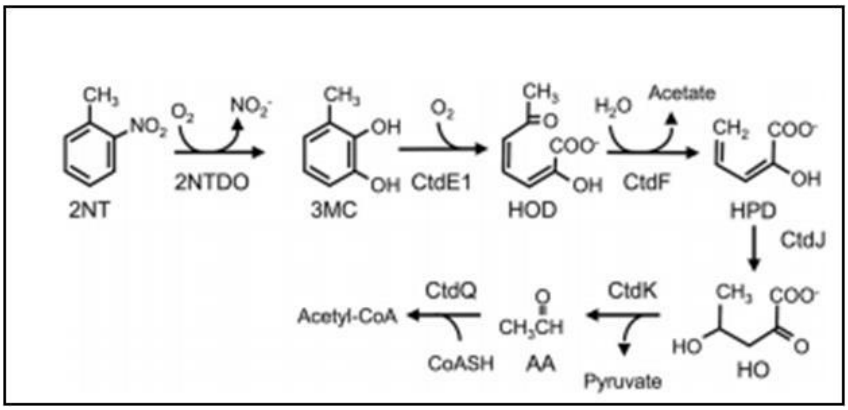

Figure 31: *2-Nitrotoluene degradation pathway in Acidovorax sp. strain JS42. CtdE1, catechol 2,3-dioxygenase; CtdF, 2-hydroxymuconate semialdehyde hydrolase; CtdJ, 2-oxopent-4-dienoate hydratase; CtdK, 4-hydroxy-2-oxovalerate aldolase; CtdQ, acetaldehyde dehydrogenase (acylating); 3MC, 3-methylcatechol; HOD, 2-hydroxy-6-oxohepta-2,4-dienoate; HPD, 2-hydroxypenta-2,4-dienoate; HO, 4-hydroxy-2-oxovalerate; AA, acetaldehyde.*

Multicomponent dioxygenases are used by many bacteria to catalyze the initial step in degradation of various compaounds. All of the nitroarene dioxygenases identified to date fall within the naphthalene dioxygenase family of Rieske nonheme iron oxygenases (42) and that is the reason why it was retrieved by the initial PSI-BLAST of this study, where the naphthalene 1,2-dioxygenase system ferredoxin-NAD(P)$^+$ reductase enzyme (NahAa, Acc. Number: AAS79488.1) from *P. putida* was used as bait. Interestingly, the identified nitroarene dioxygenases are most similar to the naphthalene dioxygenase from *Ralstonia* sp. U2, a strain that converts naphthalene to central metabolites via gentisate rather than using the meta cleavage of catechol used by *P. putida* G7 (34). Thus, it has been suggested that nitroarene dioxygenases have evolved from a naphthalene dioxygenase system similar to that in *Ralstonia* sp. strain U2 (40).

2.3.6 Putative nah catabolising operons composed by 4 genes

This group is composed by 21 members, 12 of them belonging to the genus of *Pandoraea*. All the putative operons from the *Pandoraea* genus have the genetic organization shown in Figure 32, having four of the six genes composing the nag operon of the gentisate pathway.

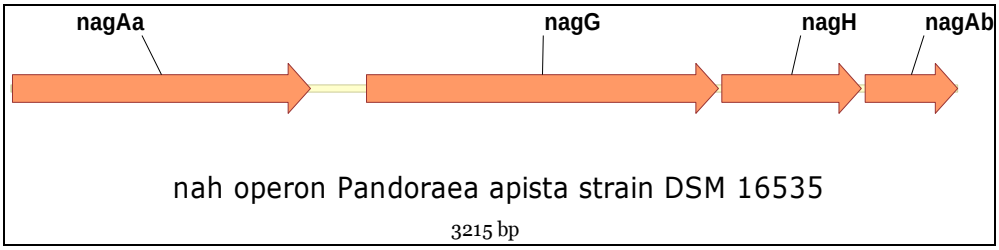

Figure 32: *Genetic organization of the putative operons composed by 4 CDS, belonging to the Pandoraea genus.*

Genome-to-genome analysis showed that they have a high level of identity (more than 75%, data not shown). The two CDs missing are *nagI* and *nagK* and in all 12 *Pandoraea* genomes were found located on a different putative operon (Table ST4). The genetic organization of the second putative operon is shown in Figure 33 and is the same for all 12 members of this group, except *Pandoraea apista* DSM 16537 in which the salicylate hydroxylase CDS is missing and *Pandoraea vervacti* strain NS15 in which there is a IS5-family transposase between the salicylate hydroxylase and the MFS transporter.

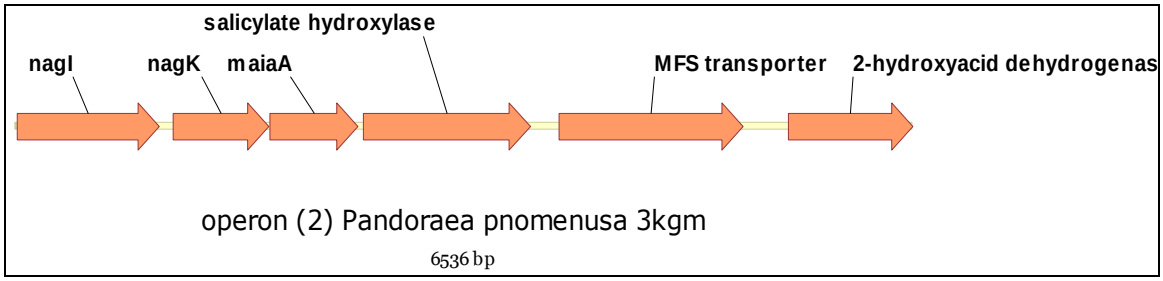

Figure 33: *Genetic organization of the putative operons (2) of the Pandoraea genus, containing the putative nagI and nagK genes.*

To our knowledge there is no experimental information of naphthalene degradation from any *Pandoraea* strain, but considering the above, it can be

proposed that they could catabolise naphthalene through the gentisate pathway.

The remaining 9 putative operons, members of this group are:

- three strains from *Acidovorax* (*A.avenae* subsp. *avenae* ATCC 19860, *A. citrulli* AAC00-1 and *A.* sp. KKS102)
- two strains from *Azoarcus* (*A. olearius* strain DQS4 and *Azoarcus* sp. BH72)
- one strain from *Burkholderiales* (*Burkholderiales* bacterium JOSHI001)
- two strains from *Cupriavidus* (*C. basilensis* strain 4G11 and *Cupriavidus* sp. USMAHM13)
- one strain from *Shimwellia* (*S. blattae* DSM 4481)

The genetic organization of all the 9 members is identical and is shown in Figure 34, having four of the six genes composing the nag operon of the gentisate pathway.

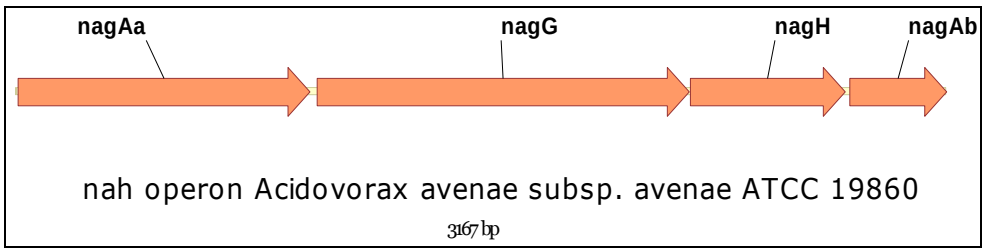

Figure 34: *Genetic organization of the putative nah catabolising operons composed by 4 CDS*

Genome-to-genome analysis showed that, as in the case of the putative operons from *Pandoraea*, they have a high level of identity (more than 65%, data not shown). The two CDs missing are *nagI* and *nagK* and in all 9 genomes were found located on a different putative operon (Table ST5), thus supplying the enzymes needed for the gentisate pathway.

2.3.7 Putative nah catabolising operons composed by 3 genes

In this group there are only two members: the putative operon from the pA81 plasmid of *Achromobacter xylosoxidans* A8 and the putative operon from *Variovorax paradoxus* EPS. Both putative operons have the same genetic organization and is shown in Figure 35.

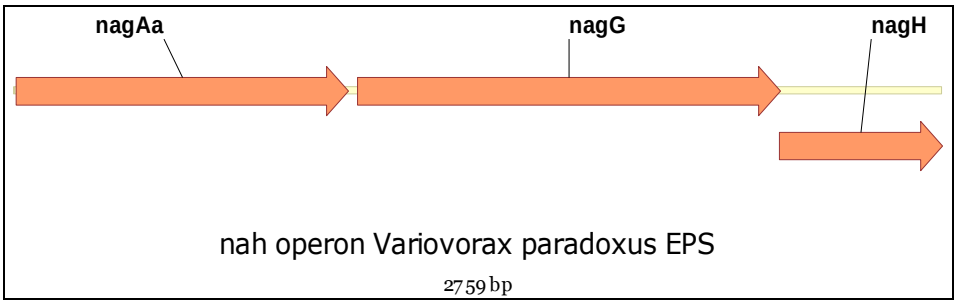

Figure 35: *Genetic organization of the putative nah catabolising operons composed by 3 CDS*

The putative operon from *V. paradoxus* EPS has three of the six genes of the gentisate pathway. The other three were spotted in a second putative operon at positions 388570-397699, on the chromosome of *V. paradoxus* EPS.

The putative operon from the pA81 plasmid of *A. xylosoxidans* A8, is not a naphthalene calabolising operon, but it is a salicylate degradating operon and forms part of a transposon. This transposon is the (halo)aromatic acid transposon TnAxl, which also carries operons for catabolism of ortho-substituted chlorobenzoates (43).

3. Conclusions and general discussion.

Polycyclic aromatic hydrocarbons (PAHs) and their derivatives are widespread in the natural environment (44) and can contaminate the ecosystem for a long time as a result of heir low solubility in water and their absorption to small particles (45). Various bacterial strains have been discovered that degrade low molecular weight PAHs as part of their metabolism. One of the simplest PAHs is naphthalene, which has been widely studied and referred to as a model compound for investigating the mechanisms of bacterial biodegradation. Microbial naphthalene metabolisms and genetic regulations involved in the degradation pathway are extensively characterized in several bacterial strains, particularly the *Pseudomonas* species.

Naphthalene degradation is organized into upper and lower pathways (44). The upper pathway enzymes are involved in the conversion of naphthalene to salicylate. This pathway comprises 10 genes organized in the order *nahAaAbAcAdBFCQED*. The lower pathway enzymes are encoded by *nahGTHINLOMKJY* and are involved in the oxidation of salicylate to pyruvate and acetyl coenzyme A (44). The nucleotide sequences of genes encoding the upper pathway enzymes from several *Pseudomonas* strains have been reported: *ndo* genes (46), *nah* genes (47), *dox* genes (37), *pah* genes (48) and *sgp* genes (39). The genetic regulation of this pathway was also studied in detail for different bacterial strains. In *Ralstonia* sp. U2, the naphthalene dioxygenase genes (*nag* genes) contained all of the genes corresponding to the classical *nah* genes of *Pseudomonas* strains in the same order, with the exception of two extra genes inserted between the ferredoxin reductase gene and ferredoxin gene. The two additional genes, named *nagG* and *nagH*, are structural subunits of salicylate-5-hydroxylase and can help the host convert naphthalene to gentisate (34).

From the 50 putative operons analyzed in this study, not all of them contain the same number of ORFs. But that observations does not seem to be important as all of them follow one of the pathways mentioned above. Forty-five (45) of the putative operons analyzed seem to be analogues of the nag operon

described for *Ralstonia* sp. U2 mentioned above (which catabolizes naphthalene through the gentisate pathway) (34), two putative operons from *Pseudomonas* strains were identical to the “classical” nah operon and one was the sgp operon described in literature (39). Only two (putative operons from *B. multivorans* strain DDS 15A-1 and from pBN2 plasmid of *Paraburkholderia* sp. BN5) were different from the naphthalene catabolizing operons described thus far and they could belong to a different pathway not described yet. The fact that the 90% of the putative Nah proteins retrived from the PSI-Blast belong to putative nag operons could mean two things; either the nag operons are more distributed in naphthalene catabolizing bacteria or it is an artifact because of the protein used as bait in the PSI-Blast.

An interesting observation is that all 50 putative operons studied seem to be regulated by a protein belonging to the LysR-type family of transcriptional regulators. This is a fact already described in literature. Transcriptional control of the classical naphthalene pathway is regulated by NahR, a regulator protein belonging to the LysR-type family of transcriptional regulators (49). NahR is responsible for the regulation of both nah operons and the gene that encodes it is located upstream of and is transcribed divergently from nahG, the first gene of the meta-pathway operon (50). This gene arrangement has been found in several different classical naphthalene genes cloned from different bacteria (51). The nag pathway in *Ralstonia* sp. strain U2 also contains a putative regulator gene, nagR, which has high sequence similarity to nahR. In contrast to nahR, this gene is located upstream of and is divergently transcribed from nagAa (52). Recently, the sgp operon was found to be preceded by the divergently directed *sgpR* gene. The amino acid sequence of the *sgpR* product qualifies it as a LysR-type transcriptional regulator (LTTR) and suggests its potential function (53).

A second very interesting observation is that all 50 putative operons studied belong to plasmids/genomes containing mobile genetic elements (Table ST6 in the part of “Anexos”). The role of mobile genetic elements (MGEs, including transposons, phagerelated elements, genomic islands, conjugative plasmids and combinations/derivatives) and their horizontal transfer in evolution of bacterial genomes and adaptation of microbial populations to specific environmental changes is generally accepted nowadays. MGEs that encode

catabolic genes are considered to play a major role in the adaptation of microbial populations to xenobiotic organic compounds, which have been introduced in the environment during the past century. This occurs either by spreading the genes in a community and thereby increasing the diversity of organisms able to degrade these compounds, or by rearranging and combining pre-existing genes or gene fragments from diffeerent microorganisms to constitute a new pathway dealing with a new compound. Interestingly, both on plasmids and in the chromosome, catabolic genes are often bordered by IS-elements (insertion sequences, Table ST6). These IS-elements may have played a role in recruitment of these genes by the replicon but also increase the potential of further exchange of the genes between different replicons and different hosts.

The main conclusions of this study are:

1. The members of the dataset retrieved from the PSI-Blast include proteins exclusively from β- (87.5%) and γ-proteobacteria (12.5%), which indicates that the metabolic route of naphthalene is present in a very specific range of bacteria.
2. NahA proteins retrieved from the PSI-Blast are coded from ORFs located in putative nah, nag or sgp operons
3. The 90% of the putative operons studied code for enzymes of the gentisate pathway
4. All putative operons studied seem to be regulated by a protein belonging to the LysR-type family of transcriptional regulators.
5. All putative operons studied belong to plasmids/genomes containing mobile genetic elements

During this study most of the programs/informatics tools used have been introduced in the lectures of the master “Bioinformatics and biostatistics” of the UOC and the knowledge adquired was put in practice. More specificaly, the use of the necessary informatics tools for alignments, phylogenetic trees, analysis of genomes and prediction of operons were implemented and applied in a real case scenario. By that way decisions concerning the appropriate programs to be used, conclusions, analysis and predictions had to be made, based on literature. The original objectives were accomplished on time. The proposed

schedule and methodology was sound and was followed without the need to make any changes so that the study could be completed successfully. As a future project, this study could be enriched with the following tasks:

1. Add a phylogenetic study of *nagG* and *nagH* genes. These genes are the ones that diverge the nah from the nag operons and it would be interesting to create their phylogeny
2. Make alignments of the putative LysR-type transcriptional regulators of the operons studied and compare them with the ones described in literature
3. Make a more profound analysis of the MGE of the replicons studied. Locate their exact position and reveal their type (Class I or II transposons, Tra genes)

4. Glosario

ORF: Open reading frame

MEGA: Mega Evolutionary Genetics Analysis

MGE: Mobile genetic elements

NDO: Naphthalene dioxygenase

PAHs: Polycyclic aromatic hydrocarbons

5. Bibliografía

1. **Li P**, **Cai Q**, **Lin W**, **Chen B**, **Zhang B**. 2016. Offshore oil spill response practices and emerging challenges. Mar. Pollut. Bull. **110**:6–27.
2. **Ko J-Y**, **Day WJ**. 2004. A review of ecological impacts of oil and gas development on coastal ecosystems in the Mississippi Delta. Ocean Coast. Manag. **47**:597–623.
3. **Khan FI**, **Husain T**, **Hejazi R**. 2004. An overview and analysis of site remediation technologies. J. Environ. Manage. **71**:95–122.
4. **Söhngen N**. 1913. Benzin, Petroleum, Paraffinöl und Paraffin als Kohlenstoff- und Energiequelle für Mikroben. Zentr Bacteriol Parasitenk Abt II **37**:595–609.
5. **Alexander M**. 1999. Biodegradation and Bioremediation, 2nd ed. Academic Press: San Diego, USA.
6. **Jain PK**, **Gupta VK**, **Gaur RK**, **Lowry M**, **Jaroli DP**, **Chauhan UK**. 2011. Bioremediation of Petroleum oil Contaminated Soil and Water. Res. J. Environ. Toxicol. **5**:1–26.
7. **Speight JG**. 2001. Handbook of petroleum analysis. Wiley-Interscience.
8. **Fan C-Y**, **Krishnamurthy S**. 1995. Enzymes for Enhancing Bioremediation of Petroleum-Contaminated Soils: A Brief Review. J. Air Waste Manage. Assoc. **45**:453–460.
9. **Hu G**, **Li J**, **Zeng G**. 2013. Recent development in the treatment of oily sludge from petroleum industry: a review. J. Hazard. Mater. **261**:470–90.
10. **Liu W**, **Luo Y**, **Teng Y**, **Li Z**, **Ma LQ**. 2010. Bioremediation of oily sludge-contaminated soil by stimulating indigenous microbes. Environ. Geochem. Health **32**:23–29.
11. **Goyal AK**, **Zylstra GJ**. Genetics of naphthalene and phenanthrene degradation by Comamonas testosteroni. J. Ind. Microbiol. Biotechnol. **19**:401–7.
12. **Seo J-S**, **Keum Y-S**, **Li QX**. 2009. Bacterial Degradation of Aromatic Compounds. Int. J. Environ. Res. Public Health **6**:278–309.
13. **Davies JI**, **Evans WC**. 1964. Oxidative metabolism of naphthalene by soil

pseudomonads. The ring-fission mechanism. Biochem. J. **91**:251–61.

14. **Dunn NW**, **Gunsalus IC**. 1973. Transmissible plasmid coding early enzymes of naphthalene oxidation in Pseudomonas putida. J. Bacteriol. **114**:974–9.
15. **Simon MJ**, **Osslund TD**, **Saunders R**, **Ensley BD**, **Suggs S**, **Harcourt A**, **Suen WC**, **Cruden DL**, **Gibson DT**, **Zylstra GJ**. 1993. Sequences of genes encoding naphthalene dioxygenase in Pseudomonas putida strains G7 and NCIB 9816-4. Gene **127**:31–7.
16. **Seo J-S**, **Keum Y-S**, **Li QX**. 2009. Bacterial Degradation of Aromatic Compounds. Int. J. Environ. Res. Public Health **6**:278–309.
17. **Fuentes S**, **Méndez V**, **Aguila P**, **Seeger M**. 2014. Bioremediation of petroleum hydrocarbons: catabolic genes, microbial communities, and applications. Appl. Microbiol. Biotechnol. **98**:4781–4794.
18. **Haigler BE**, **Gibson DT**. 1990. Purification and properties of NADH-ferredoxinNAP reductase, a component of naphthalene dioxygenase from Pseudomonas sp. strain NCIB 9816. J. Bacteriol. **172**:457–64.
19. **Ensley BD**, **Gibson DT**. 1983. Naphthalene dioxygenase: purification and properties of a terminal oxygenase component. J. Bacteriol. **155**:505–11.
20. **Kauppi B**, **Lee K**, **Carredano E**, **Parales RE**, **Gibson DT**, **Eklund H**, **Ramaswamy S**. 1998. Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase. Structure **6**:571–86.
21. **Tamura K**, **Stecher G**, **Peterson D**, **Filipski A**, **Kumar S**. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. **30**:2725–9.
22. **Edgar RC**. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–7.
23. **Lee K**, **Kauppi B**, **Parales RE**, **Gibson DT**, **Ramaswamy S**. 1997. Purification and crystallization of the oxygenase component of naphthalene dioxygenase in native and selenomethionine-derivatized forms. Biochem. Biophys. Res. Commun. **241**:553–7.
24. **Brown EN**, **Friemann R**, **Karlsson A**, **Parales J V.**, **Couture MM-J**, **Eltis LD**, **Ramaswamy S**. 2008. Determining Rieske cluster reduction potentials. JBIC J. Biol. Inorg. Chem. **13**:1301–1313.

25. **Kelley LA**, **Mezulis S**, **Yates CM**, **Wass MN**, **Sternberg MJE**. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. **10**:845–58.
26. **Acheson JF**, **Moseson H**, **Fox BG**. 2015. Structure of T4moF, the Toluene 4-Monooxygenase Ferredoxin Oxidoreductase. Biochemistry **54**:5980–5988.
27. **Rypniewski WR**, **Breiter DR**, **Benning MM**, **Wesenberg G**, **Oh BH**, **Markley JL**, **Rayment I**, **Holden HM**. 1991. Crystallization and structure determination to 2.5-A resolution of the oxidized [2Fe-2S] ferredoxin isolated from Anabaena 7120. Biochemistry **30**:4126–31.
28. **Jones DT**, **Taylor WR**, **Thornton JM**. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8**:275–82.
29. **Felsenstein J.** 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution (N. Y). **39**:783–791.
30. **Rutherford K**, **Parkhill J**, **Crook J**, **Horsnell T**, **Rice P**, **Rajandream MA**, **Barrell B**. 2000. Artemis: sequence visualization and annotation. Bioinformatics **16**:944–5.
31. **Solovyev V**, **Salamov A**. 2011. Automatic Annotation of Microbial Genomes and Metagenomic Sequences, p. 61–78. *In* Li, RW (ed.), Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies. Nova Science Publishers.
32. **Sullivan MJ**, **Petty NK**, **Beatson SA**. 2011. Easyfig: a genome comparison visualizer. Bioinformatics **27**:1009–1010.
33. **Pollard JR**, **Bugg TD**. 1998. Purification, characterisation and reaction mechanism of monofunctional 2-hydroxypentadienoic acid hydratase from Escherichia coli. Eur. J. Biochem. **251**:98–106.
34. **Fuenmayor SL**, **Wild M**, **Boyes AL**, **Williams PA**. 1998. A gene cluster encoding steps in conversion of naphthalene to gentisate in Pseudomonas sp. strain U2. J. Bacteriol. **180**:2522–30.
35. **Jeon CO**, **Park M**, **Ro H-S**, **Park W**, **Madsen EL**. 2006. The Naphthalene Catabolic (nag) Genes of Polaromonas naphthalenivorans CJ2: Evolutionary Implications for Two Gene Clusters and Novel Regulatory Control. Appl. Environ. Microbiol. **72**:1086–1095.

36. **Hofer B**, **Eltis LD**, **Dowling DN**, **Timmis KN**. 1993. Genetic analysis of a Pseudomonas locus encoding a pathway for biphenyl/polychlorinated biphenyl degradation. Gene **130**:47–55.
37. **Denome SA**, **Stanley DC**, **Olson ES**, **Young KD**. 1993. Metabolism of dibenzothiophene and naphthalene in Pseudomonas strains: complete DNA sequence of an upper naphthalene catabolic pathway. J. Bacteriol. **175**:6890–901.
38. **EDWARDS SW**, **KNOX WE**. 1956. Homogentisate metabolism: the isomerization of maleylacetoacetate by an enzyme which requires glutathione. J. Biol. Chem. **220**:79–91.
39. **Izmalkova TY**, **Sazonova OI**, **Nagornih MO**, **Sokolov SL**, **Kosheleva IA**, **Boronin AM**. 2013. The organization of naphthalene degradation genes in Pseudomonas putida strain AK5. Res. Microbiol. **164**:244–253.
40. **Parales RE**, **Huang R**, **Yu C-L**, **Parales J V**, **Lee FKN**, **Lessner DJ**, **Ivkovic-Jensen MM**, **Liu W**, **Friemann R**, **Ramaswamy S**, **Gibson DT**. 2005. Purification, characterization, and crystallization of the components of the nitrobenzene and 2-nitrotoluene dioxygenase enzyme systems. Appl. Environ. Microbiol. **71**:3806–14.
41. **Ju K-S**, **Parales RE**. 2010. Nitroaromatic Compounds, from Synthesis to Biodegradation. Microbiol. Mol. Biol. Rev. **74**:250–272.
42. **Gibson DT**, **Parales RE**. 2000. Aromatic hydrocarbon dioxygenases in environmental biotechnology. Curr. Opin. Biotechnol. **11**:236–43.
43. **Jencova V**, **Strnad H**, **Chodora Z**, **Ulbrich P**, **Vlcek C**, **Hickey WJ**, **Paces V**. 2008. Nucleotide sequence, organization and characterization of the (halo)aromatic acid catabolic plasmid pA81 from Achromobacter xylosoxidans A8. Res. Microbiol. **159**:118–27.
44. **Peng R-H**, **Xiong A-S**, **Xue Y**, **Fu X-Y**, **Gao F**, **Zhao W**, **Tian Y-S**, **Yao Q-H**. 2008. Microbial biodegradation of polyaromatic hydrocarbons. FEMS Microbiol. Rev. **32**:927–955.
45. **Park W**, **Jeon CO**, **Cadillo H**, **DeRito C**, **Madsen EL**. 2004. Survival of naphthalene-degrading Pseudomonas putida NCIB 9816-4 in naphthalene-amended soils: toxicity of naphthalene and its metabolites. Appl. Microbiol. Biotechnol. **64**:429–35.
46. **Yang Y**, **Chen RF**, **Shiaris MP**. 1994. Metabolism of naphthalene,

fluorene, and phenanthrene: preliminary characterization of a cloned gene cluster from Pseudomonas putida NCIB 9816. J. Bacteriol. **176**:2158–64.

47. **Li W**, **Shi J**, **Wang X**, **Han Y**, **Tong W**, **Ma L**, **Liu B**, **Cai B**. 2004. Complete nucleotide sequence and organization of the naphthalene catabolic plasmid pND6-1 from Pseudomonas sp. strain ND6. Gene **336**:231–40.
48. **Takizawa N**, **Iida T**, **Sawada T**, **Yamauchi K**, **Wang YW**, **Fukuda M**, **Kiyohara H**. 1999. Nucleotide sequences and characterization of genes encoding naphthalene upper pathway of pseudomonas aeruginosa PaK1 and Pseudomonas putida OUS82. J. Biosci. Bioeng. **87**:721–31.
49. **Schell MA**, **Poser EF**. 1989. Demonstration, characterization, and mutational analysis of NahR protein binding to nah and sal promoters. J. Bacteriol. **171**:837–46.
50. **Yen KM**, **Gunsalus IC**. 1982. Plasmid gene organization: naphthalene/salicylate oxidation. Proc. Natl. Acad. Sci. U. S. A. **79**:874–8.
51. **Schell MA**. 1993. Molecular Biology of the LysR Family of Transcriptional Regulators. Annu. Rev. Microbiol. **47**:597–626.
52. **Zhou NY**, **Fuenmayor SL**, **Williams PA**. 2001. nag genes of Ralstonia (formerly Pseudomonas) sp. strain U2 encoding enzymes for gentisate catabolism. J. Bacteriol. **183**:700–8.
53. **Filatova IY**, **Kazakov AS**, **Muzafarov EN**, **Zakharova M V**. 2017. Protein SgpR of Pseudomonas putida strain AK5 is a LysR-type regulator of salicylate degradation through gentisate. FEMS Microbiol. Lett. **364**.

6. Anexos

Table ST1: The list of 224 proteins retrieved from the PSI-Blast

	Protein	Acc Number	Host
1	NahAa	AAS79488.1	*Pseudomonas putida* strain SG1
2	NahAa	AAB62705.1	*Pseudomonas putida* strain BS202
3	NahAa	OCX93220.1	*Pseudomonas* sp. K35
4	NahAa	WP_024718189.1	*Pseudomonas putida*
5	ferredoxin oxidoreductase	AAL07270.1	*Pseudomonas fluorescens*
6	NahAa	WP_095423764.1	*Paraburkholderia* sp. BN5
7	NahAa	OPK03992.1	*Pseudomonas veronii* VI4T1
8	NahAa	ADK11285.1	*Pseudomonas stutzeri* NJ
9	NahAa	WP_003292051.1	*Pseudomonas stutzeri*
10	NahAa	WP_014819639.1	*Pseudomonas stutzeri* strain 19SMN4
11	NahAa	EZQ14078.1	*Pseudomonas bauzanensis* strain W13Z2
12	NahAa	WP_068171435.1	*Hydrogenophaga taeniospiralis*
13	NahAa	OGB17027.1	*Burkholderiales bacterium* RIFCSPLOWO2_02_FULL_67_64
14	NahAa	WP_039615382.1	*Pseudomonas* sp. C5pp
15	NahAa	OGO97130.1	*Curvibacter* sp. GWA2_64_110
16	NahAa	WP_045786202.1	*Ralstonia mannitolilytica*
17	ferredoxin oxidoreductase	AAB09763.1	*Burkholderia* sp. RASC strain DNT
18	DntaA	AAL50024.1	*Burkholderia cepacia* strain R34
19	Ferrodoxin reductase	ACT53245.1	*Burkholderia* sp. C3 strain C3
20	MntA	AGH09219.1	*Diaphorobacter* sp. DS1
21	MntA	AGH09229.1	*Diaphorobacter* sp. DS3
22	NahAa	WP_011806220.1	*Acidovorax* sp. JS42
23	DntAa	AAX31152.1	*Burkholderia* sp. DNT
24	NahAa	WP_026437494.1	*Acidovorax* sp. JHL-9
25	NahAa	WP_034359000.1	*Comamonas testosteroni*
26	NahAa	WP_037482963.1	*Sphaerotilus natans*
27	NahAa	WP_038211661.1	*Xenophilus azovoran*s
28	NahAa	WP_048805400.1	*Burkholderia multivorans* strain DDS 15A-1
29	NahAa	OGB52657.1	*Burkholderiales* bacterium RIFOXYD12_FULL_59_19
30	NahAa	WP_023471407.1	*Betaproteobacteria* bacterium MOLA814
31	NahAa	WP_065340771.1	*Azoarcus olearius* strain DQS4
32	NahAa	WP_092440326.1	*Collimonas* sp. OK607
33	NahAa	WP_011766244.1	*Azoarcus* sp. BH72
34	NahAa	WP_058642131.1	*Pseudacidovorax intermedius*
35	NahAa	ODU17846.1	*Variovorax* sp. SCN 67-85
36	NahAa	WP_093054801.1	*Variovorax* sp. YR634
37	NahAa	WP_093239598.1	*Variovorax* sp. EL159
38	NahAa	WP_093432435.1	*Variovorax* sp. 770b2
39	Hypothetical	WP_017524094.1	*Pusillimonas noertemannii*
40	NahAa	WP_036238759.1	*Massilia* sp. JS1662
41	NahAa	WP_042576831.1	*Variovorax paradoxus*
42	NahAa	SDY74659.1	*Variovorax* sp. YR266
43	NahAa	SES94580.1	*Variovorax* sp. OV084
44	NahAa	WP_081270575.1	*Variovorax paradoxus*

45	NahAa	WP_093074554.1	*Variovorax* sp. OV084
46	NahAa	WP_093173846.1	*Variovorax* sp. YR266
47	NahAa	WP_093554787.1	*Massilia namucuonensis*
48	Oxidoreductase	ADU37122.1	*Variovorax paradoxus* EPS
49	NahAa	WP_007832232.1	*Variovorax* sp. CF313
50	NahAa	WP_034398527.1	*Comamonas testosteroni*
51	NahAa	WP_041942906.1	*Variovorax paradoxus* EPS
52	NahAa	WP_056330920.1	*Massilia* sp. Root1485
53	NahAa	WP_056596711.1	*Variovorax* sp. Root434
54	NahAa	WP_062477809.1	*Variovorax boronicumulans*
55	NahAa	WP_070060757.1	*Variovorax boronicumulans*
56	NahAa	SDC34521.1	*Variovorax* sp. CF079
57	NahAa	WP_072633605.1	*Ralstonia solanacearum* strain EP1
58	NahAa	WP_077003436.1	*Variovorax* sp. KK3
59	NahAa	WP_085489316.1	*Paraburkholderia susongensis*
60	NahAa	WP_089401839.1	*Noviherbaspirillum humi*
61	NahAa	WP_093103280.1	*Variovorax* sp. CF079
62	NahAa	WP_093341827.1	*Variovorax* sp. PDC80
63	NahAa	WP_095745371.1	*Variovorax boronicumulans* strain J1
64	NahAa	WP_095948606.1	*Variovorax boronicumulans*
65	NahAa	WP_020653875.1	*Massilia niastensis*
66	NahAa	WP_034331400.1	*Herbaspirillum* sp. B39
67	NahAa	WP_050469262.1	*Herbaspirillum chlorophenolicum*
68	NahAa	WP_056272351.1	*Hydrogenophaga* sp. Root209
69	NahAa	WP_056276036.1	*Hydrogenophaga* sp. Root209
70	NahAa	WP_056580641.1	*Variovorax* sp. Root473
71	NahAa	WP_057594945.1	*Variovorax paradoxus*
72	NahAa	WP_068832131.1	*Polaromonas jejuensis*
73	NahAa	ODS69087.1	*Bordetella* sp. SCN 67-23
74	NahAa	ODV10699.1	*Rubrivivax* sp. SCN 70-15
75	NahAa	WP_071090728.1	*Ralstonia solanacearum*
76	Hypothetical	APC68127.1	*Ralstonia solanacearum* OE1-1
77	NahAa	WP_071623877.1	*Ralstonia solanacearum*
78	NahAa	WP_074287189.1	*Burkholderia* sp. GAS332
79	NahAa	WP_077594648.1	*Polaromonas* sp. A23
80	NahAa	WP_087782021.1	*Pigmentiphaga* sp. NML080357
81	NahAa	WP_087840734.1	*Pigmentiphaga* sp. NML030171
82	NahAa	WP_092756739.1	*Albidiferax* sp. OV413
83	NdsA	BAC53589.1	*Pigmentiphaga* sp. NDS-2
84	NahAa	WP_007862447.1	*Polaromonas* sp. CF318
85	NahAa	WP_011001041.1	*Ralstonia solanacearum*
86	NahAa	WP_016725112.1	*Ralstonia solanacearum* strain YC40-M
87	NahAa	WP_018440041.1	*Burkholderia* sp. JPY347
88	NahAa	WP_019717888.1	*Ralstonia solanacearum*
89	NahAa	WP_020748828.1	*Ralstonia solanacearum*
90	NahAa	WP_028860402.1	*Ralstonia solanacearum*
91	NahAa	WP_046568751.1	*Paraburkholderia fungorum*
92	NahAa	WP_047499156.1	*Methylibium* sp. CF059
93	NahAa	AKZ26929.1	*Ralstonia solanacearum* strain YC45
94	NahAa	WP_056637513.1	*Acidovorax* sp. Root70
95	NahAa	CUV22296.1	*Ralstonia solanacearum*
96	NahAa	CUV26816.1	*Ralstonia solanacearum*
97	NahAa	CUV45591.1	*Ralstonia solanacearum*
98	NahAa	WP_058907081.1	*Ralstonia solanacearum* strain CQPS-1
99	NahAa	WP_062085352.1	*Caballeronia udeis*
100	NahAa	WP_063500190.1	*Burkholderia* sp. OLGA172
101	NahAa	WP_068685341.1	*Variovorax* sp. WDL1
102	NahAa	WP_074579661.1	*Polaromonas* sp. JS666

103	NahAa	WP_075464902.1	*Ralstonia solanacearum* strain KACC 10722
104	NahAa	WP_086123269.1	*Hydrogenophaga* sp. IBVHS1
105	NahAa	WP_087452178.1	*Ralstonia solanacearum* strain SEPPX05
106	NahAa	WP_088176949.1	*Burkholderia* sp. Bk
107	NahAa	WP_092941150.1	*Acidovorax wautersii*
108	NahAa	WP_093971005.1	*Pusillimonas* sp. T2
109	NahAa	WP_009551513.1	*Burkholderiales bacterium* JOSHI_001
110	Oxidoreductase	WP_010807123.1	*Pandoraea* sp. SD6-2
111	SgpA	WP_011255169.1	*Achromobacter xylosoxidans*
112	NahAa	WP_011793354.1	*Acidovorax citrulli* AAC00-1
113	NahAa	WP_012345834.1	*Leptothrix cholodnii* SP-6
114	NahAa	WP_013397092.1	*Achromobacter xylosoxidans* A8
115	NahAa	WP_028603456.1	*Ottowia thiooxydans*
116	NahAa	WP_030101522.1	*Burkholderia* sp. K24
117	NahAa	WP_044528838.1	*Herbaspirillum* sp. B65
118	NahAa	WP_047847830.1	*Caballeronia mineralivorans*
119	NahAa	WP_056464570.1	*Rhizobacter* sp. Root404
120	NahAa	WP_057223733.1	*Acidovorax* sp. Root275
121	NahAa	WP_057272754.1	*Acidovorax* sp. Root267
122	Ferrodoxin reductase	KWT66092.1	*Variovorax* sp. WDL1
123	NahAa	WP_064049638.1	*Ralstonia solanacearum*
124	NahAa	WP_071021068.1	*Cupriavidus* sp. USMAHM13
125	HybA	AAC69483.1	*Pseudomonas aeruginosa*
126	NahAa	WP_003265434.1	*Ralstonia solanacearum* strain UY031
127	NahAa	WP_003277762.1	*Ralstonia solanacearum*
128	NahAa	WP_010461139.1	*Acidovorax radicis*
129	NahAa	WP_011481936.1	*Polaromonas* sp. JS666
130	NahAa	WP_013592708.1	*Acidovorax avenae* subsp. *avenae* ATCC 19860
131	NahAa	WP_015013726.1	*Acidovorax* sp. KKS102
132	NahAa	WP_019699854.1	*Acidovorax avenae* subsp. *avenae* ATCC 19860
133	NahAa	WP_026433434.1	*Acidovorax oryzae*
134	NahAa	WP_027802469.1	*Paraburkholderia dilworthii*
135	NahAa	WP_028222878.1	*Paraburkholderia oxyphila*
136	NahAa	WP_028362659.1	*Burkholderia* sp. JPY366
137	NahAa	WP_038712630.1	*Burkholderia* sp. lig30
138	NahAa	WP_039567721.1	*Ralstonia solanacearum*
139	NahAa	WP_043356437.1	*Cupriavidus basilensis* strain 4G11
140	NahAa	WP_053843817.1	*Acidovorax avenae*
141	NahAa	WP_053858376.1	*Burkholderia* sp. HB1
142	NahAa	WP_055326326.1	*Ralstonia solanacearum*
143	NahAa	WP_056195931.1	*Pelomonas* sp. Root1237
144	NahAa	WP_056660217.1	*Rhizobacter* sp. Root1221
145	NahAa	WP_060987730.1	*Acidovorax delafieldii*
146	NahAa	WP_066269935.1	*Hydrogenophaga palleronii*
147	ferredoxin oxidoreductase	WP_067297403.1	*Marinobacterium profundum*
148	NahAa	SEF08079.1	*Burkholderia* sp. WP9
149	NahAa	WP_091809335.1	*Burkholderia* sp. WP9
150	NahAa	WP_092836818.1	*Acidovorax cattleyae*
151	NahAa	WP_013587300.1	*Burkholderia* sp. CCGE1001
152	NahAa	WP_014617417.1	*Ralstonia solanacearum* strain UW163
153	NahAa	WP_015001775.1	*Paraburkholderia phenoliruptrix* BR3459a
154	NahAa	WP_035485900.1	*Paraburkholderia phenoliruptrix*

155	NahAa	WP_039367362.1	*Pandoraea pnomenusa*
156	NahAa	WP_039400878.1	*Pandoraea sputorum* strain DSM 21091
157	NahAa	WP_039405159.1	*Pandoraea pulmonicola* strain DSM 16583
158	NahAa	WP_042112664.1	*Pandoraea apista* strain TF80G25
159	NahAa	WP_044847058.1	*Burkholderia* sp. USM B20
160	NahAa	WP_046292957.1	*Pandoraea oxalativorans* strain DSM 23570
161	NahAa	WP_048627858.1	*Pandoraea apista* strain DSM 16535
162	NahAa	WP_053572544.1	*Caballeronia cordobensis*
163	NahAa	WP_056669097.1	*Acidovorax* sp. Leaf160
164	NahAa	WP_057201646.1	*Acidovorax* sp. Root217
165	NahAa	WP_058375803.1	*Pandoraea norimbergensis* strain DSM 11628
166	NahAa	WP_061119800.1	*Caballeronia turbans*
167	NahAa	WP_063598471.1	*Pandoraea pnomenusa* strain MCB032
168	NahAa	WP_087691382.1	*Pandoraea* sp. PE-S2R-1
169	NahAa	WP_087722510.1	*Pandoraea* sp. PE-S2T-3
170	NahAa	WP_092137060.1	*Cupriavidus* sp. YR651
171	NahAa	WP_092951498.1	*Acidovorax konjaci*
172	NahAa	WP_094069146.1	*Pandoraea apista*
173	NahAa	OYU25262.1	*Burkholderiales* bacterium PBB2
174	NahAa	WP_023594572.1	*Pandoraea pnomenusa* 3kgm
175	NahAa	WP_025250375.1	Pandoraea pnomenusa strain RB38
176	NahAa	WP_045235879.1	*Burkholderiaceae* bacterium 16
177	NahAa	WP_054432510.1	*Achromobacter* sp. 2789STDY5608633
178	NahAa	WP_056897267.1	*Pseudorhodoferax* sp. Leaf274
179	Hypothetical	WP_084929442.1	*Pseudomonas aeruginosa*
180	NahAa	WP_091779450.1	*Burkholderia* sp. yr281
181	NahAa	WP_094291749.1	*Acidovorax* sp. KNDSW-TSA6
182	SgpA	ACO92374.1	*Pseudomonas putida*
183	NahAa	WP_007855076.1	*Acidovorax* sp. CF316
184	NahAa	WP_013661544.1	*Marinomonas mediterranea* MMB-1
185	NahAa	WP_013796866.1	*Marinomonas posidonica* IVIA-Po-181
186	NahAa	WP_024977952.1	*Ralstonia pickettii*
187	NahAa	WP_038618531.1	*Pandoraea pnomenusa* strain DSM 16536
188	NahAa	WP_042316770.1	*Paraburkholderia terrae*
189	NahAa	WP_042590330.1	*Ralstonia solanacearum*
190	NahAa	WP_042878795.1	*Cupriavidus necator*
191	NahAa	WP_045205718.1	*Burkholderiaceae* bacterium 26
192	NahAa	WP_056743330.1	*Acidovorax* sp. Root568
193	NahAa	WP_057269083.1	*Acidovorax* sp. Root219
194	NahAa	WP_063462015.1	*Acidovorax* sp. GW101-3H11
195	NahAa	WP_064576239.1	*Cupriavidus gilardii*
196	NahAa	ODS60347.1	*Acidovorax* sp. SCN 65-108
197	Hypothetical	WP_069865819.1	*Pseudomonas* sp. CCA 1
198	NahAa	OGA59186.1	*Burkholderiales* bacterium RIFCSPHIGHO2_01_FULL_64_960
199	NahAa	OGA84093.1	*Burkholderiales* bacterium GWA2_64_37
200	NahAa	OGB09710.1	*Burkholderiales* bacterium RIFCSPHIGHO2_02_FULL_64_19
201	NahAa	SEG59731.1	*Marinobacterium lutimaris*
202	NahAa	WP_019451300.1	*Cupriavidus* sp. BIS7
203	Hypothetical	AGW94018.1	*Ralstonia pickettii* DTP0602
204	NahAa	WP_023263257.1	*Cupriavidus* sp. HPC(L)
205	NahAa	WP_034396672.1	*Comamonas testosteroni*
206	NahAa	WP_055398692.1	*Acidovorax* sp. SD340

207	NahAa	WP_056061878.1	*Acidovorax* sp. Root402
208	NahAa	WP_069341944.1	*Pandoraea* sp. ISTKB
209	NahAa	WP_012070357.1	*Marinomonas* sp. MWYL1
210	NahAa	WP_039013644.1	*Cupriavidus* sp. IDO
211	NahAa	WP_044456834.1	*Pandoraea vervacti* strain NS15
212	NahAa	WP_047905736.1	*Pandoraea faecigallinarum*
213	NahAa	OJY18768.1	*Pandoraea* sp. 64-18
214	ferredoxin oxidoreductase	WP_072840530.1	*Marinomonas polaris*
215	NahAa	WP_006157498.1	*Cupriavidus basilensis*
216	Hypothetical	WP_017232765.1	*Pandoraea* sp. B-6
217	Reductase	WP_035896115.1	*Kluyvera ascorbata*
218	Reductase	WP_052283276.1	*Kluyvera cryocrescens*
219	NahAa	SEP73738.1	*Solimonas aquatica*
220	Hypothetical	WP_084191308.1	*Algiphilus aromaticivorans*
221	Hypothetical	WP_093281094.1	*Solimonas aquatica*
222	Hypothetical	WP_040453150.1	*Hydrocarboniphaga effusa*
223	Reductase	WP_002444116.1	*Shimwellia blattae* DSM 4481 = NBRC 105725
224	NahAa	WP_022976508.1	*Nevskia ramosa*

Table ST2: The list of proteins retrieved from the PSI-Blast, that are located in sequenced plasmids.

	Protein	Acc Number	Host	Plasmid	Acc Number	CDS	Identity	Query cover	Threshold	Operon structure	Operon region
1	NahAa	WP_095423764.1	*Paraburkholderia* sp. BN5	pBN2	NZ_CP022992.1	complement (478969-479955)	87%	100%	9,00E-140	11 genes	complement (469554-479955)
2	NahAa	WP_014819639.1	*Pseudomonas stutzeri* strain 19SMN4	pLIB119	NZ_CP007510.1	50930-51916	84%	100%	2,00E-133	9 genes	50930-60012
3	NahAa	WP_013397092.1	*Achromobacter xylosoxidans* A8	pA81	NC_014641.1	complement (39949-40935)	59%	100%	2,00E-142	3 genes	complement (38438-40935)
4	SgpA	ACO92374.1	*Pseudomonas putida*	pAK5	FJ859895.1	Partial, only nah operon	55%	100%	5,00E-128	6 genes	1926-6754

Table ST3: The list of proteins retrieved from the PSI-Blast, that are located in sequenced chromosomes/genomes.

	Protein	Acc Number	Host	Genome	CDS	Acc Number	Identity	Query cover	Threshold	Operon structure	Operon region
1	NahAa	WP_011806220.1	*Acidovorax* sp. JS42	Chromosome	3260037-3261023	NC_008782.1	66%	100%	1,00E-135	6 genes	3260037-3264857
2	NahAa	WP_048805400.1	*Burkholderia multivorans* strain DDS 15A-1	Chromosome 1	1558869-1559855, 1564317-1565303	NZ_CP008730.1	66%	100%	6,00E-136	11 genes	1564317-1575234
3	NahAa	WP_065340771.1	*Azoarcus olearius* strain DQS4	Genome	2879867-2880808	NZ_CP016210.1	64%	100%	2,00E-137	4 genes	2875983-2879755
4	NahAa	WP_011766244.1	*Azoarcus* sp. BH72	Genome	complement (2783542-2784528)	NC_008702.1	64%	100%	5,00E-138	4 genes	complement (2780947-2784528)
5	NahAa	WP_041942906.1	*Variovorax paradoxus* EPS	Chromosome	3162324-3163313	NC_014931.1	62%	100%	2,00E-140	3 genes	3162324-3165082
6	NahAa	WP_072633605.1	*Ralstonia solanacearum* strain EP1	Genome	2650097-2651083	NZ_CP015115.1	61%	100%	2,00E-135	7 genes	2650097-2655670
7	NahAa	ATA54863.1	*Variovorax boronicumulans* strain J1	Genome	complement (3860055-3861044)	CP023284.1	61%	100%	2,00E-140	7 genes	complement (3856033-3861044)
8	Hypothetical	APC68127.1	*Ralstonia solanacearum* OE1-1	Genome	1227375-1228277	CP009764.1	60%	100%	6,00E-134	7 genes	1221668-1227304
9	NahAa	WP_016725112.1	*Ralstonia solanacearum* strain YC40-M	Genome	2605369..2606355	NZ_CP015850.1	60%	100%	7,00E-135	7 genes	2605369-2610933
10	NahAa	WP_046568751.1	*Paraburkholderia fungorum* strain ATCC BAA-463	Chromosome 1	3419546..3420541	NZ_CP010026.1	60%	100%	9,00E-130	7 genes	3419546-3425113

11	NahAa	AKZ26929.1	*Ralstonia solanacearum* strain YC45	Genome	complement (2647167-2648153)	CP011997.1	60%	100%	8,00E-135	7 genes	complement (2642585-2648153)
12	NahAa	WP_058907081.1	*Ralstonia solanacearum* strain CQPS-1	Genome	166383..167369	NZ_CP016914.1	60%	100%	2,00E-134	7 genes	166383-171956
13	NahAa	WP_063500190.1	*Burkholderia* sp. OLGA172	Chromosome 2	2955668-2956570	NZ_CP014579.1	59%	100%	3,00E-137	7 genes	complementary (2949865-2955535)
14	NahAa	WP_075464902.1	*Ralstonia solanacearum* strain KACC 10722	Genome	2486710..2487696	NZ_CP014702.1	59%	100%	2,00E-134	7 genes	2486710-2492423
15	NahAa	WP_087452178.1	*Ralstonia solanacearum* strain SEPPX05	Genome	3548620..3549606	NZ_CP021448.1	59%	100%	2,00E-133	7 genes	3548620-3554281
16	NahAa	WP_009551513.1	*Burkholderiales bacterium* JOSHI_001	Chromosome	complement (3676334-3677320)	NZ_CM001438.1	59%	100%	7,00E-143	4 genes	complement (3674206-3677320)
17	NahAa	WP_011793354.1	*Acidovorax citrulli* AAC00-1	Genome	183480-184475	NC_008752.1	59%	100%	1,00E-135	4 genes	183480-186597
18	NahAa	WP_012345834.1	*Leptothrix cholodnii* SP-6	Genome	complement (860209-861195)	NC_010524.1	59%	100%	2,00E-146	11 genes	complement (851667-861195)
19	NahAa	WP_071021068.1	*Cupriavidus* sp. USMAHM13	Chromosome 2	2960517-2961509	NZ_CP017752.1	58%	100%	1,00E-129	4 genes	2960517-2963801
20	NahAa	WP_003265434.1	*Ralstonia solanacearum* strain UY031	Genome	complement (1620566-1621552)	NZ_CP012687.1	58%	100%	9,00E-134	7 genes	complement (1615987-1621552)
21	NahAa	WP_011481936.1	*Polaromonas* sp. JS666	Genome	998284-999267	NC_007948.1	58%	100%	1,00E-147	12 genes	998284-1008635
22	NahAa	WP_013592708.1	*Acidovorax avenae*	Genome	212885-213880	NC_015138.	58%	100%	2,00E-	4 genes	212885-

			subsp. *avenae* ATCC 19860			1			135		216051
23	NahAa	WP_015013726.1	*Acidovorax* sp. KKS102	Genome	complement (2169169-2170152)	NC_018708.1	58%	100%	3,00E-143	4 genes	complement (2167057-2170152)
24	NahAa	WP_043356437.1	*Cupriavidus basilensis* strain 4G11	Chromosome secondary	3038480-3039466	NZ_CP010537.1	58%	100%	2,00E-138	4 genes	3038480-3041599
25	NahAa	WP_053858376.1	*Burkholderia* sp. HB1	Chromosome 1	complement (2536318-2537313)	NZ_CP012192.1	58%	100%	6,00E-129	7 genes	complement (2531725-2537313)
26	NahAa	WP_013587300.1	*Burkholderia* sp. CCGE1001	Chromosome 1	641561-642547	NC_015136.1	57%	100%	8,00E-137	6 genes	641561-646488
27	NahAa	WP_014617417.1	*Ralstonia solanacearum* strain UW163	Genome	complement (2601757-2602743)	NZ_CP012939.1	57%	100%	1,00E-132	7 genes	complement (2597178-2602743)
28	NahAa	WP_015001775.1	*Paraburkholderia phenoliruptrix* BR3459a	Chromosome 1	630455-631441	NC_018695.1	57%	100%	2,00E-136	7 genes	630455-636043
29	NahAa	WP_039400878.1	*Pandoraea sputorum* strain DSM 21091	Genome	complement (4555318-4556349)	NZ_CP010431.2	57%	100%	6,00E-135	4 genes	complement (4553134-4556349)
30	NahAa	WP_039405159.1	*Pandoraea pulmonicola* strain DSM 16583	Genome	complement (4691650-4692681)	NZ_CP010310.2	57%	100%	4,00E-133	4 genes	complement (4689453-4692681)
31	NahAa	WP_042112664.1	*Pandoraea apista* strain TF80G25	Genome	complement (792947-793960)	NZ_CP011279.1	57%	100%	1,00E-136	4 genes	complement (790746-793960)
32	NahAa	WP_046292957.1	*Pandoraea oxalativorans* strain	Genome	complement (4434850-	NZ_CP011253.3	57%	100%	7,00E-133	4 genes	complement

			DSM 23570		4435881)						(4432631-4435881)
33	NahAa	WP_048627858.1	*Pandoraea apista* strain DSM 16535	Genome	1270391-1271404	NZ_CP013481.2	57%	100%	2,00E-136	4 genes	1270391-1273605
34	NahAa	WP_058375803.1	*Pandoraea norimbergensis* strain DSM 11628	Genome	complement (5001556-5002665)	NZ_CP013480.3	57%	100%	6,00E-134	4 genes	complement (4999358-5002665)
35	NahAa	WP_063598471.1	*Pandoraea pnomenusa* strain MCB032	Genome	complement (1352279-1353307)	NZ_CP015371.1	56%	100%	3,00E-134	4 genes	complement (1350048-1353307)
36	NahAa	WP_023594572.1	*Pandoraea pnomenusa* 3kgm	Genome	910448-911476	NC_022904.2	56%	100%	3,00E-134	4 genes	910448-913704
37	NahAa	WP_025250375.1	*Pandoraea pnomenusa* strain RB38	Genome	complement (4275972-4277000)	NZ_CP007506.3	56%	100%	4,00E-134	4 genes	complement (4273741-4277000)
38	NahAa	WP_013661544.1	*Marinomonas mediterranea* MMB-1	Genome	2643936-2644922	NC_015276.1	55%	100%	5,00E-119	7 genes	2643936-2650826
39	NahAa	WP_013796866.1	*Marinomonas posidonica* IVIA-Po-181	Genome	2588278-2589264	NC_015559.1	55%	100%	3,00E-120	9 genes	2588278-2596225
40	NahAa	WP_038618531.1	*Pandoraea pnomenusa* strain DSM 16536	Genome	1173259-1174287	NZ_CP009553.3	55%	100%	9,00E-134	4 genes	1173259-1176515
41	Hypothetical	AGW94018.1	*Ralstonia pickettii* DTP0602	Chromosome 2	1667352-1668341	CP006668.1	54%	100%	5,00E-129	7 genes	1663866-1670480
42	NahAa	WP_012070357.1	*Marinomonas* sp. MWYL1	Genome	3014106-3015092	NC_009654.1	53%	100%	4,00E-118	9 genes	3014106-3022840
43	NahAa	WP_044456834.1	*Pandoraea vervacti* strain NS15	Genome	complement (4465896-4466951)	NZ_CP010897.2	53%	100%	1,00E-129	4 genes	complement (4463366 9-4466951)
44	NahAa	WP_047905736.1	*Pandoraea*	Genome	complement	NZ_CP0118	53%	100%	4,00E-	4 genes	complem

			faecigallinarum		(4091561-4092619)	07.3			131		ent (4089351-4092619)
45	Reductase	WP_002444116.1	*Shimwellia blattae* DSM 4481 = NBRC 105725	Genome	3538503-3539480	NC_017910.1	46%	100%	1,00E-111	4 genes	3538503-3541568

Table ST4: List of the putative nah operons belonging to the *Pantorea* genus

	Protein	Acc Number	Host	Operon region (1)	Operon region (2)
1	NahAa	WP_039400878.1	*Pandoraea sputorum* strain DSM 21091	complement (4553134-4556349)	complement (719224-725536)
2	NahAa	WP_039405159.1	*Pandoraea pulmonicola* strain DSM 16583	complement (4689453-4692681)	complement (704911-711385)
3	NahAa	WP_042112664.1	*Pandoraea apista* strain TF80G25	complement (790746-793960)	complement (2737000-2743499)
4	NahAa	WP_046292957.1	*Pandoraea oxalativorans* strain DSM 23570	complement (4432631-4435881)	complement (704893-711156)
5	NahAa	WP_048627858.1	*Pandoraea apista* strain DSM 16535	1270391-1273605	4735167-4741701
6	NahAa	WP_058375803.1	*Pandoraea norimbergensis* strain DSM 11628	complement (4999358-5002665)	complement (774479-780813)
7	NahAa	WP_063598471.1	*Pandoraea pnomenusa* strain MCB032	complement (1350048-1353307)	complement (3277194-3283728)
8	NahAa	WP_023594572.1	*Pandoraea pnomenusa* 3kgm	910448-913704	4456704-4463239
9	NahAa	WP_025250375.1	*Pandoraea pnomenusa* strain RB38	complement (4273741-4277000)	complement (660417-666951)
10	NahAa	WP_038618531.1	*Pandoraea pnomenusa* strain DSM 16536	1173259-1176515	4735167-4741701

11	NahAa	WP_044456834.1	*Pandoraea vervacti strain* NS15	complement (44633669-4466951)	complement (739557-746912)
12	NahAa	WP_047905736.1	*Pandoraea faecigallinarum*	complement (4089351-4092619)	complement (671624-677938)

Table ST5: List of the putative nah operons composed by 4 putative genes

	Protein	Acc Number	Host	Operon region (1)	Operon region (2)
1	NahAa	WP_065340771.1	*Azoarcus olearius* strain DQS4	2875983-2879755	2753264-2758108
2	NahAa	WP_011766244.1	*Azoarcus* sp. BH72	complement (2780947-2784528)	2673841-2678685
3	NahAa	WP_009551513.1	*Burkholderiales bacterium* JOSHI_001	complement (3674206-3677320)	complement (3054230-3058856)
4	NahAa	WP_011793354.1	*Acidovorax citrulli* AAC00-1	183480-186597	160851-165863
5	NahAa	WP_071021068.1	*Cupriavidus* sp. USMAHM13	2960517-2963801	complement (382891-388641)
6	NahAa	WP_013592708.1	*Acidovorax avenae* subsp. *avenae* ATCC 19860	212885-216051	197252-201244
7	NahAa	WP_015013726.1	*Acidovorax* sp. KKS102	complement (2167057-2170152)	2730324-2735293

8	NahAa	WP_043356437.1	*Cupriavidus basilensis* strain 4G11	3038480-3041599	complement (2216568-2221802)
9	Reductase	WP_002444116.1	*Shimwellia blattae* DSM 4481 = NBRC 105725	3538503-3541568	1418639-1424648

Table ST6:

	Protein	Protein Acc Number	Host	Locus	Nucleotide Acc Number	Tra genes/Transposases present
1	NahAa	AAB62705.1	*Pseudomonas putida* strain BS202	pNPL1 plasmid	AF010471.1	Partial
2	NahAa	WP_095423764.1	*Paraburkholderia* sp. BN5	pBN2 plasmid	NZ_CP022992.1	TraD, conjugal transfer proteins, Transposases
3	NahAa	WP_014819639.1	*Pseudomonas stutzeri* strain 19SMN4	pLIB119 plasmid	NZ_CP007510.1	Transposases
4	NahAa	WP_013397092.1	*Achromobacter xylosoxidans* A8	pA81 plasmid	NC_014641.1	TrbC,D,F,H,I,J,M,N-VirB4,D4-TraX,C,G,I,L,M-Transposases
5	SgpA	ACO92374.1	*Pseudomonas putida*	pAK5 plasmid	FJ859895.1	partial
6	NahAa	WP_011806220.1	*Acidovorax* sp. JS42	Chromosome	NC_008782.1	TraG,H,F,N,U,V,B,K,D-Transposases
7	NahAa	WP_048805400.1	*Burkholderia multivorans* strain DDS 15A-1	Chromosome 1	NZ_CP008730.1	TraG,F-TrbC,D,E-Transposases
8	NahAa	WP_065340771.1	*Azoarcus olearius* strain DQS4	Genome	NZ_CP016210.1	Transposases

9	NahAa	WP_011766244.1	*Azoarcus* sp. BH72	Genome	NC_008702.1	Transposases
10	NahAa	WP_041942906.1	*Variovorax paradoxus* EPS	Chromosome	NC_014931.1	Transposases
11	NahAa	WP_072633605.1	*Ralstonia solanacearum* strain EP1	Genome	NZ_CP015115.1	TraR, Transposases
12	NahAa	ATA54863.1	*Variovorax boronicumulans* strain J1	Genome	CP023284.1	Transposases
13	Hypothetical	APC68127.1	*Ralstonia solanacearum* OE1-1	Genome	CP009764.1	
14	NahAa	WP_016725112.1	*Ralstonia solanacearum* strain YC40-M	Genome	NZ_CP015850.1	TraR,F-TrF,E-Transposases
15	NahAa	WP_046568751.1	*Paraburkholderia fungorum* strain ATCC BAA-463	Chromosome 1	NZ_CP010026.1	Transposase
16	NahAa	AKZ26929.1	*Ralstonia solanacearum* strain YC45	Genome	CP011997.1	TraR, Transposases
17	NahAa	WP_058907081.1	*Ralstonia solanacearum* strain CQPS-1	Genome	NZ_CP016914.1	TraR, Transposases
18	NahAa	WP_063500190.1	*Burkholderia* sp. OLGA172	Chromosome 2	NZ_CP014579.1	TraR, Transposases
19	NahAa	WP_075464902.1	*Ralstonia solanacearum* strain KACC 10722	Genome	NZ_CP014702.1	TraR, Transposases
20	NahAa	WP_087452178.1	*Ralstonia solanacearum* strain SEPPX05	Genome	NZ_CP021448.1	TraD,R-TrbL-Transposases
21	NahAa	WP_009551513.1	*Burkholderiales bacterium* JOSHI_001	Chromosome	NZ_CM001438.1	TraS,G-TrbL,C-VirB4-Transposases

22	NahAa	WP_011793354.1	*Acidovorax citrulli* AAC00-1	Genome	NC_008752.1	TraG-TrbC,D,E,F,I-Transposases
23	NahAa	WP_012345834.1	*Leptothrix cholodnii* SP-6	Genome	NC_010524.1	TraG-TrbL,C-VirB4-Transposases
24	NahAa	WP_071021068.1	*Cupriavidus* sp. USMAHM13	Chromosome 2	NZ_CP017752.1	Transposases
25	NahAa	WP_003265434.1	*Ralstonia solanacearum* strain UY031	Genome	NZ_CP012687.1	TraG-Transposases
26	NahAa	WP_011481936.1	*Polaromonas* sp. JS666	Genome	NC_007948.1	Transposases
27	NahAa	WP_013592708.1	*Acidovorax avenae* subsp. *avenae* ATCC 19860	Genome	NC_015138.1	Transposases
28	NahAa	WP_015013726.1	*Acidovorax* sp. KKS102	Genome	NC_018708.1	TrbF,D,C,E-TraG,R-Transposases
29	NahAa	WP_043356437.1	*Cupriavidus basilensis* strain 4G11	Chromosome secondary	NZ_CP010537.1	TraF,G,W,C,L-Transposases
30	NahAa	WP_053858376.1	*Burkholderia* sp. HB1	Chromosome 1	NZ_CP012192.1	Transposases
31	NahAa	WP_013587300.1	*Burkholderia* sp. CCGE1001	Chromosome 1	NC_015136.1	TraC,B-TrbJ,L,G,I-Transposase
32	NahAa	WP_014617417.1	*Ralstonia solanacearum* strain UW163	Genome	NZ_CP012939.1	TraR-Transposases

33	NahAa	WP_015001775.1	*Paraburkholderia phenoliruptrix* BR3459a	Chromosome 1	NC_018695.1	TraC,B,H,D-TrbL,G,I-VirB4-Transposases
34	NahAa	WP_039400878.1	*Pandoraea sputorum* strain DSM 21091	Genome	NZ_CP010431.2	Transposases
35	NahAa	WP_039405159.1	*Pandoraea pulmonicola* strain DSM 16583	Genome	NZ_CP010310.2	Transposase
36	NahAa	WP_042112664.1	*Pandoraea apista* strain TF80G25	Genome	NZ_CP011279.1	TraG,R-Transposases
37	NahAa	WP_046292957.1	*Pandoraea oxalativorans* strain DSM 23570	Genome	NZ_CP011253.3	Transposase
38	NahAa	WP_048627858.1	*Pandoraea apista* strain DSM 16535	Genome	NZ_CP013481.2	TrbE,F,I-TraG,R-Transposases
39	NahAa	WP_058375803.1	*Pandoraea norimbergensis* strain DSM 11628	Genome	NZ_CP013480.3	Transposase
40	NahAa	WP_063598471.1	*Pandoraea pnomenusa* strain MCB032	Genome	NZ_CP015371.1	TraG-TrbC,D,E,F,I-Transposases
41	NahAa	WP_023594572.1	*Pandoraea pnomenusa* 3kgm	Genome	NC_022904.2	TrbJ-Transposase
42	NahAa	WP_025250375.1	*Pandoraea pnomenusa* strain RB38	Genome	NZ_CP007506.3	Transposase
43	NahAa	WP_013661544.1	*Marinomonas mediterranea* MMB-1	Genome	NC_015276.1	Transposase
44	NahAa	WP_013796866.1	*Marinomonas posidonica* IVIA-Po-181	Genome	NC_015559.1	Transposase

45	NahAa	WP_038618531.1	*Pandoraea pnomenusa* strain DSM 16536	Genome	NZ_CP009553.3	Transposases
46	Hypothetical	AGW94018.1	*Ralstonia pickettii* DTP0602	Chromosome 2	CP006668.1	Transposase
47	NahAa	WP_012070357.1	*Marinomonas* sp. MWYL1	Genome	NC_009654.1	Transposase
48	NahAa	WP_044456834.1	*Pandoraea vervacti* strain NS15	Genome	NZ_CP010897.2	Transposase
49	NahAa	WP_047905736.1	*Pandoraea faecigallinarum*	Genome	NZ_CP011807.3	TrbJ-TraJ-Transposase
50	Reductase	WP_002444116.1	*Shimwellia blattae* DSM 4481 = NBRC 105725	Genome	NC_017910.1	TraR-Transposase