Bacterial gene diversity related to tryptophan metabolism and indole-3-acetic acid (IAA) production in the rhizosphere

Pablo Suárez Pérez
Máster en Bioinformática y Bioestadística
ÁREA 4 – Aula 1

Nombre Consultor/a: Paloma María Pizarro Tobías
Nombre Profesor/a responsable de la asignatura: Carles Ventura Royo

8/01/2020

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Bacterial gene diversity related to tryptophan metabolism and indole-3-acetic acid (IAA) production in the rhizosphere*
Nombre del autor:	*Pablo Suárez Pérez*
Nombre del consultor/a:	*Paloma María Pizarro Tobías*
Nombre del PRA:	*Carles Ventura Royo*
Fecha de entrega (mm/aaaa):	01/2020
Titulación::	*Máster en Bioinformática y Bioestadística*
Área del Trabajo Final:	*Microbiología, biotecnología, filogenética*
Idioma del trabajo:	*Inglés*
Palabras clave	*Triptófano, ácido indoleacético, enzima, gen, operón, ruta metabótica, filogenética, análisis filogenético*

Resumen del Trabajo:

Desde hace años, las rizobacterias han suscitado interés por su capacidad de promover el crecimiento vegetal. Una de las hormonas que estas bacterias producen y que tiene uno de los principales papeles en esta promoción es el ácido indoleacético.

En este trabajo, se recoge toda la evidencia registrada hasta el momento de las rutas de biosintesís de esta hormona en rizobacterias, así como de su precursor, el aminoácido triptófano. Para cada una de las rutas y especies identificadas, se obtuvieron los genes y proteinas que intervienen en los diferentes puntos de cada ruta. A fin de analizar su diversidad, se realizó un estudio filogético de cada una de las enzimas bajo estudio, se documentó su perfil proteico y se analizó la coocurrencia de rutas. Para estudiar su origen y evolución, se realizó a su vez una comparación con un marcador filogenético conocido (*gyrB*).

Los resultados muestran una clara co-occurencia de enzimas y rutas, que en algunos casos puede ser explicada por la presencia de operones, así como ciertas divergencias en algunos casos en las comparaciones con los marcadores.

Abstract:

For years, growth-promoting rizhobacteria have been a topic of interest.. Indole-3-acetic acid or IAA is one of the principal hormones that can be find in these species.

In this study, we have gathered all the evidence so far about IAA byosinthesis. Besides, we have followed the same procedure for the substrate of the IAA pathways (tryptophan). For each one of the identified species and pathways, genes and proteins sequences that are involved have been obtained. We have analyzed these gene pool diversity through phylogenetic analysis, and at the same time we have documented their protein profile and co-ocurrence level. In order to study their origin and evolution, comparisons using the phylogenetic marker *gyrB* have been performed.

Results show high co-occurence (both at enzyme/pathway level), that in some cases can be justified attending to the existence of operons. Besides, several differences of interest have been found in the *gyrB* comparisons.

Table of contents

1. Introduction....1
1.1 Context of the project and justification....1
1.2 Objectives....2
1. Identify the main bacterial species which take part in the tryptophan/IAA metabolism in the rizhosphere....2
....3
1.3 Materials and methods....3
1.4 Project planning....4
1.5 Breve sumario de productos obtenidos....4
1.5 Breve sumario de los productos obtenidos....5
1.6 Breve descripción de los otros capítulos de la memoria....5
2. Chapters....6
2.1. Plant growth promoting rhizobacteria and the importance of the tryptophan metabolism/IAA production in the rhizosphere....6
2.2. Description of the metabolic pathways....7
2.2.1. Indole-3-pyruvate pathway (IPA)....8
2.2.2. Indole-3-acetamide pathway (IAM)....8
2.2.3. Indole-3-acetonitrile pathway (IAN)....9
2.2.4. Tryptophan biosynthesis....10
2.3. Main genes and protein profiles....11
2.3.1. IPA....11
2.3.1.1. Tryptophan aminotransferase....11
2.3.1.2. Indole-3-pyruvate decarboxylase....12
2.3.1.3. Indole-3-acetaldehyde dehydrogenase....13
2.3.2. IAM....14
2.3.2.1. Tryptophan 2-monooxygenase....14
2.3.2.2. Indole-3-acetamide hydrolase....15
2.3.3. IAN....16
2.3.3.1. Nitrile hydratase....16
2.3.3.2. Nitrilase....17
2.3.4. Tryptophan biosynthesis....18
2.3.4.1. Anthranilate synthase....18
2.3.4.2. Tryptophan synthase....19
2.4. Main IAA rhizobacteria species, known pathways and coocurrence....20
2.5. Results: diversity analysis I....24
2.6. Results: diversity analysis II....38
3.Conclussions....51
4. Glossary....53
5. References....54
6. Annexes....65
6.1. Annex 1: Protein accesion numbers list....65
6.2. Annex 2: R pipeline....75
6.3. Annex 3: Tanglegram abbreviation....78

Figures

Figure 1. Fungal and bacterial rhizosphere microbiome ……1
Figure 2. Project timeline (Gant diagram)……..4
Figure 3. Indole-3-acetic acid (IAA)……6
Figure 4. IAA and tryptohan biosynthesis…….……7
Figure 5. Indole-3-pyruvate (IPA) pathway……8
Figura 6. Indole-3-acetamide (IAM) pathway……9
Figure 7. Indole-3-acetonitrile (IAN) pathway……9
Figure 8. Tryptophan biosynthesis pathway……10
Figure 9. Tryptophan aminotransferase reaction……...11
Figure 10. Indole-3-pyruvate decarboxylase reaction……...12
Figure 11. Indole-3-acetaldehyde dehydrogenase reaction……13
Figure 12. Tryptophan-2-monooxygenase reaction……...14
Figure 13. Indole-3-acetamide hydrolase reaction……15
Figure 14. Nitrile hydratase reaction……16
Figure 15. Nitrilase reaction……...17
Figure 16. Anthranilate synthase reaction……...18
Figure 17. Tryptophan synthase reaction……19
Figure 18. Anthranilate synthase phylogeny……...25
Figure 19. Tryptophan synthase (alpha subunit) phylogeny……26
Figure 20. Tryptophan synthase (beta subunit) phylogeny……..28
Figure 21. Trp operon……29
Figure 22. Genetic region trpF, trpE, trpD, trpC, trpB and trpA from *Rhodococcus jostii*……...29
Figure 23. trpA/trpB region comparison……..30
Figure 24. Tryptophan transaminase phylogeny……31
Figure 25. Indole-3-pyruvate decarboxylase phylogeny……...32
Figure 26. Indole-3-acetaldehyde dehydrogenase phylogeny……33
Figure 27. Tryptophan-2-monooxygenase phylogeny……...34
Figure 28. Indole-3-acetamide hydrolase phylogeny……35
Figure 29. Nitrile hydratase (alpha subunit) phylogeny……36
Figure 30. Nitrile hydratase (beta subunit) phylogeny……..37
Figure 31. Nitrilase phylogeny……..38
Figure 32. *gyrB* marker phylogeny……..39
Figure 33. Anthranilate synthase vs *gyrB*……...40
Figure 34. Tryptophan synthase(alpha subunit) vs *gyrB*……...42
Figure 35. Tryptophan synthase(beta subunit) vs *gyrB*……43
Figure 36. Tryptophan transaminase vs *gyrB*……..45
Figure 37. Indole-3-pyruvate decarboxylase vs *gyrB*……46
Figure 38. Indole-3-acetaldehyde dehydrogenase vs *gyrB*……...46
Figure 39. Tryptophan-2-monooxygenase vs *gyrB*……47
Figure 40. Indole-3-acetamide hydrolase vs *gyrB*……..48
Figure 41. Nitrile hydratase (alpha subunit) vs *gyrB*……48
Figure 42. Nitrile hydratase (beta subunit) vs *gyrB*……...49
Figure 43. Nitrilase vs *gyrB*……...50

Tables

Table 1. Tryptophan aminotransferase (protein profile)……12
Table 2. Indole-3-pyruvate decarboxylase (protein profile)……..13
Table 3. Indole-3-acetaldehyde (protein profile)……14
Table 4. Tryptophan-2-monooxygenase (protein profile)……..15
Table 5. Indole-3-acetamide hydrolase (protein profile)……16
Table 6. Nitrile hydratase (protein profile)……17
Table 7. Anthranilate synthase (protein profile)……....19
Table 8. Tryptophan synthase (protein profile)……...20
Table 9. IAA rhizobacteria species and known pathways……23
Table 10. Co-occurrence matrix……23

1. Introduction

1.1 Context of the project and justification

Plant growth in soils is a process that depends on multiple number of factors, either biotic or abiotic[1]. The soil surface and the layer which surrounds a plant's root constitute an extremely important area for root activity and plant metabolism, which is known as rhizosphere[2]. The rhizosphere is home to numerous species of bacteria, fungi, protozoa, algae and other taxonomic groups, being bacteria the most frequent one, although they represent only a small portion of the biomass due to their size (it is a consensus that 1 g sample of rhizosphere's soil contains around 10^8-10^{12} bacterial cells, being gram-negative bacteria those that predominate, mainly *Pseudomonas*, and being *Actinomycetes* the principal group of gram-positive)[3].

In order to show this diversity, we can take a look to a recent research[4] about grapevine's microbiome:

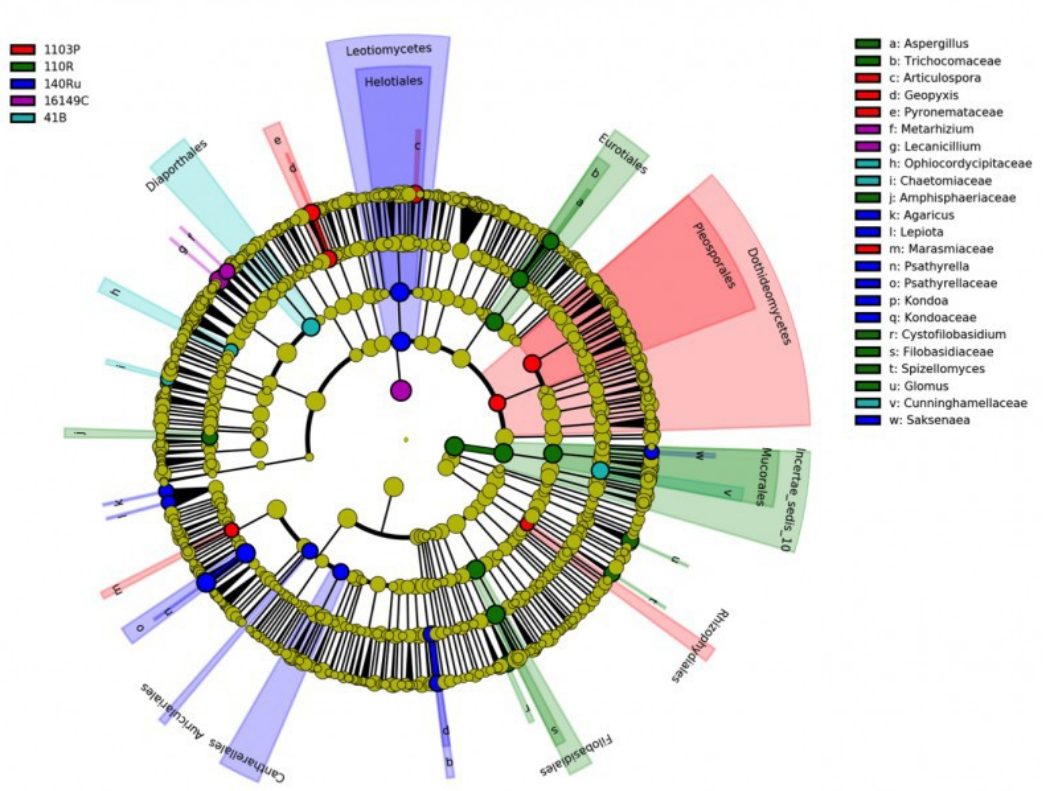


Fig 1.Berlanas et al, 2019[4]. *Fungal and bacterial rhizosphere microbiome associated with grapevine rootstock genotypes in mature and young vineyards. Phylogenetic relations are shown.*

Rhizosphere bacterial species (or rhizobacteria) are known for several interactions that affect both plants and soils[5]. Some of these are activities that ensure the stability and productivity of numerous systems, such as agricultural and natural systems. In this sense, it has been proved that a bunch of bacterial activities could have potential industrial applications, mainly as biotechnological tools in sustainable agriculture and other agrotechnological areas.

One of the activities that promote plant growth in the rhizosphere is the secretion of growth hormones, being IAA one of them[5]. Thus, identifying the gene diversity associated to the metabolic pathway of this hormone could be a crucial step in order to understand and develop new strategies concerning sustainable agriculture and plant other activities. According to this, this project aims to analyze the gene diversity associated to the tryptophan and indole-3-acetic acid (hereinafter, IAA) metabolic pathways. As a first step, we will perform an exhaustive bibliographic research of all the available information about these pathways and the main bacterial species that take part in these processes in the rhizosphere. Then, we will gather information about the substrates and products of the pathways, analyzing every element that may be involved in each step, such as enzymes and co-factors. Therefore, we will also get every gene and protein sequence of the identified enzymes, being these the biological data we will use as elements of comparison to study the diversity. After this preliminary process, we will perform a phylogenetic analysis of each coding gene, in order to establish an evolutionary relationship among the identified bacterial species that share that gene. We will also give more information about protein profiles, distribution and co-ocurrence. We will lately modify this approach using gene markers instead, allowing us to get a general overview of the evolutionary diversification for each analyzed gene through the comparison of the phylogenetic relationship obtained using markers and those obtained by using the selected coding genes. Finally, we will discuss our results.

1.2 Objectives

The objectives of this research are:

1. To identify the main bacterial species which take part in the tryptophan/IAA metabolism in the rizhosphere.

- To collect all the information available concerning the identified genes and proteins that are known for taking part in a step of the tryptophan/IAA biosynthesis and, in some cases, also those that are not yet identified but that have been proposed.
- To describe the identified genes and proteins and their implications in the different metabolic pathways.
- To describe the metabolic pathways.

2. To perform a comparison of the selected proteins (diversity analysis I).
    - To determine the protein profile, co-ocurrence and distribution of each protein.
    - To make a phylogenetic study of each protein according to the identified sequences.

3. To perform a comparison of the bacterial species using the selected markers (diversity analysis II).
    - To study the phylogenetic differences according to gyrase B (*gyrB)* marker.
    - To compare the results from the two approaches (diversity analysis I and II) in order to find differences and similarities.

1.3 Materials and methods

We have obtained all our data (species, known pathways and enzymes) from the information available in books, papers and metabolic databases, mainly KEEG pathway(*Kyoto encyclopedia of genes and genomes,* described as *"a collection of manually drawn pathway maps representing the knowledge on the molecular interaction, reaction and relation networks for metabolism and other topics"*)[106]. Then, we have created a list with these species, gene references and protein accession numbers (*annex 1*). After that, and following the same criteria, we have identified for each species which IAA biosynthesis pathways were present (since there are several, as we will discuss) and which were absent.

In order to perform the phylogenetic analysis that we have proposed, we have used *R* as the main tool. The R *pipeline* we have designed has been built using an amount of well-known packages in these evolutionary studies, such as *ape* (used in the construction and plotting of the trees), s*eqinr* (used when computing distance matrices), *biostrings* (for a

faster manipulation of large sets of biological sequences), *msa* (used in order to perform the required multiple sequence alignment) and *reutils* (to retrieve biological sequences). Besides, comparisons of trees have been performed using the *dendextend* package. The *pipeline* structure and a step by step guidance can be found in the *annex* 2. Finally, we have determined the presence of operons of interest using *Softberry*[107], and then we have compared some genomic regions related to these operons using *Easyfig* [108].

1.4 Project planning

The resources that we have used in this project are:

- Publications (books and *papers* about different topics, such as rhizobacteria, rhizosphere, IAA, tryptophan, *R* packages and software tools).
- NCBI (*National Center for Biotechnology Information*)[109], from where we have collected the available information of interest about the identified species. This includes protein accession number, obtained from *Refseq* through NCBI.
- KEGG (*Kyoto encyclopedia of genes and genomes*), where the metabolic pathways of the species are described, which have served us as a tool to check if any doubtful gene or protein obtained through bibliographic research has been finally identified as a part of the metabolic pathway that we are studying.
- BRENDA[110] and PDB[111] databases, for information about enzymes.
- R packages (*ape, seqinr, reutils, phylotools, dendextend, bionstrings* and *msa*).
- *Softberry*, for the identification of operons, and *Easyfig*, for comparing similar genetic regions among a group of species.

Concerning the structure of the project, we have followed a working approach according to the Gantt chart we show below:

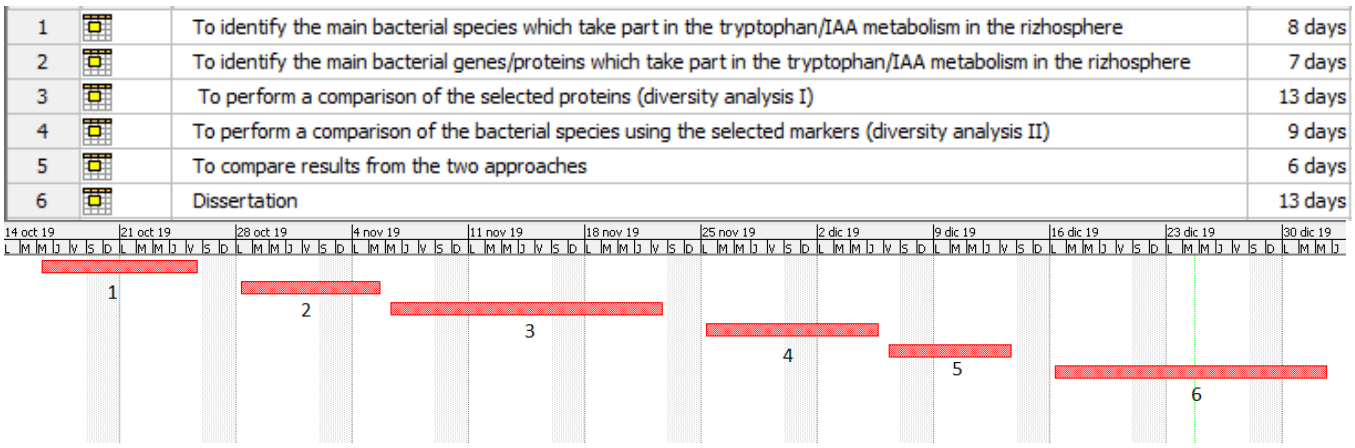


Figure 2. Project timeline (Gant diagram)

1.5 Products summary

- List of rhizobacteria that show evidence of IAA biosynthesis (starting from tryptophan) and tryptophan biosynthesis (starting from chorismate).
- List of enzymes that catalyze each step in each described pathway.
- *R* pipeline (for creating and comparing trees).
- Phylogeny of each enzyme under study.

1.6 Chapters (brief description)

- **Chapter 1**: Introduction of the topic. Description of the rizhosphere, its characteristics and importance. Description of the associated microbiome, characteristics and importance. Tryptophan and IAA characterization: Importance and contributions to the rhizosphere ecosystem. Economic and industrial applications.
- **Chapter 2**: Description of the metabolic pathways. Genes and proteins of interest.
- **Chapter 3**: Gene characterization. Protein structure and functions.
- **Chapter 4**: List of bacterial species that are able to produce IAA. Relationship among them. Co-occurence matrix.
- **Chapter 5:** Diversity analysis I. Phylogeny of each enzyme under study.
- **Chapter 6**: Diversity analysis I. *gyrB* marker phylogeny. Comparisons enzyme vs marker.

2. Chapters

2.1. CHAPTER 1. Plant growth promoting rhizobacteria and the importance of the «tryptophan metabolism - IAA biosynthesis» in the rhizosphere

As mentioned before, rhizobacteria play essential roles in plant nutrition, growth promotion and disease interactions[3]. Plants select these bacteria using specific organic compounds including those that exudate through their roots, creating a selective environment where only a few bacterial species can survive.

Thus, rhizosphere ecosystems behave as an ecological niche for each and every plant and those beneficial bacterial species to which they are associated. When referrring to growth promotion activities, these bacterial species are know as *plant growth promoting rhizobacteria* (PGPR)[5] or simply *plant growth promoting bacteria* (PGPB). In recent years, the utilization of PGPR as fertilizers and pesticides has started to be a topic of interest related to agriculture and biological production[6], since these bacterial species play a role in enhancing nutrient use efficiency and ensuring their availability[7]. PGPR have a high potential in the production of several plant hormones (know as phytohormones), such as auxins (involved in growth and behavioral processes in plant life cycles, such as phototropism, geotropism, hydrotropism, wound response and root growth and development), gibberellins (stem elongation, germination, dormancy, flowering, flower development, and leaf and fruit senescence), cytokinins (promoting cell division), ethylene (ripening of fruits) and absisic acid (seed and bud dormancy, control of organ size and stomatal closure)[8].

The main auxin we can find in plants is indole-3-acetic acid (IAA)[9]. This hormone can be synthesized by the plant itself using tryptophan as substrate, but can be also provided by rhizobacteria and some other groups of microbes[9].

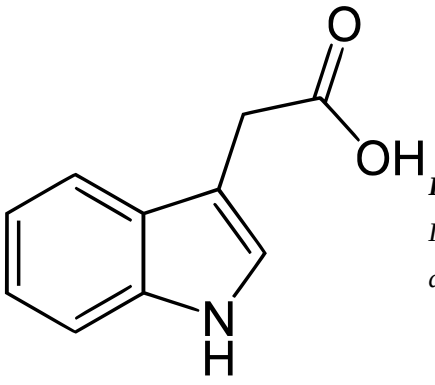


Figure 3.
Indole-3-acetic acid (BRENDA database)

Therefore, the implications of IAA produced by rhizobacteria are known as a key part in plant-growth processes, from cell elongation to cell division an even tissue differentiation[10]. This, in addition to the inherent elevation of size and surface area of root systems in contact with soils by rhizobacteria, which leads to an increased in nutrients and water uptake[11], makes IAA PGPR one of the most important species of every soil ecosystem.

Since IAA has been proved to be one of the most important hormones in plants, understanding tryptophan metabolism and IAA biosynthesis in rhizobacteria has become one topic of economic interest due to huge impacts that the use of these species could make in agrotechnological industries as potential natural fertilizers.

2.2. CHAPTER 2: Description of the metabolic pathways

According to the evidence, tryptophan and IAA biosynthesis are two related processes, since tryptophan is the substrate of almost every IAA pathway that has been described[12].

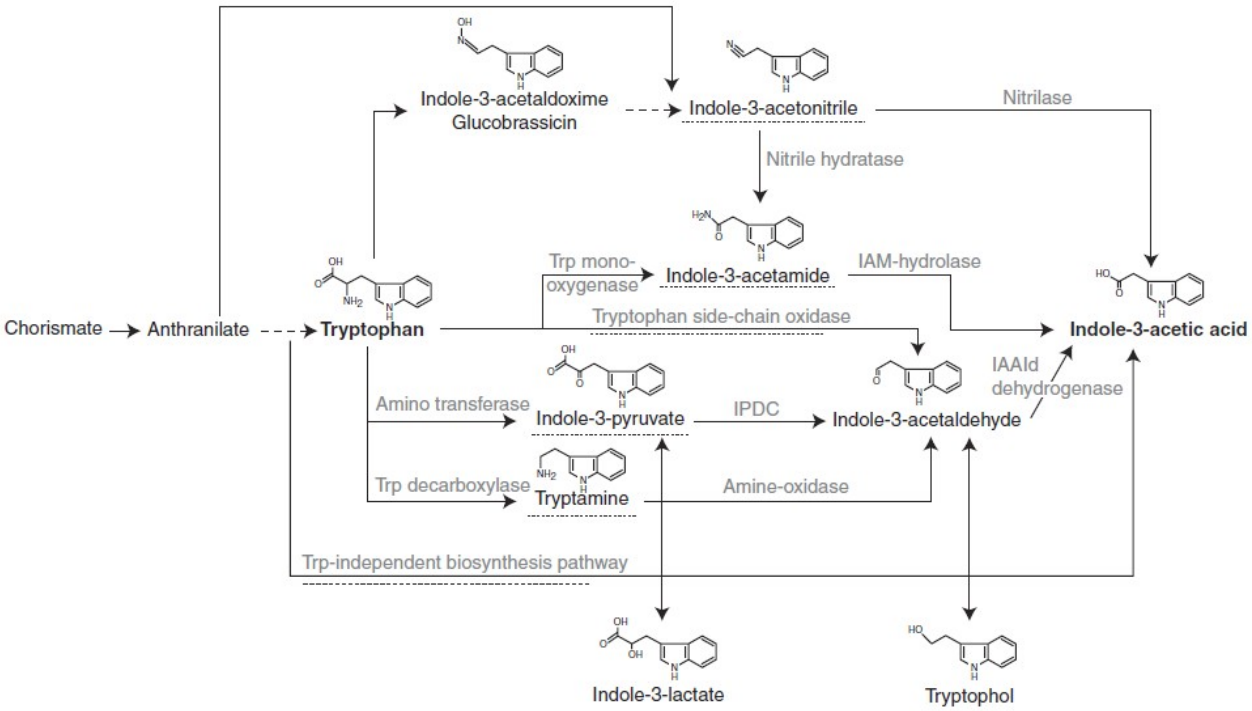


Figure 4. Spaepen & Vanderleyden (2011)[15]. IAA and tryptophan biosynthesis pathways , starting from chorismate.

There are several pathways that start with tryptophan and end with IAA. According to Spaepen & Vanderleyden (2011)[15], these are the indole-3-pyruvate pathway (IPA), the indole-3-acetonitrile pathway (IAN), the indole-3-acetamide pathway (IAM), the tryptamine pathway and the tryptophan side-oxidase pathway. Besides, the existence of a tryptophan independent pathway has been suggested, but there are no evidence so far. Since only three of them (IPA, IAN and IAM) are present in rhizobacteria, these are the ones we are going to study and describe.

2.2.1. Indole-3-pyruvate pathway (IPA)

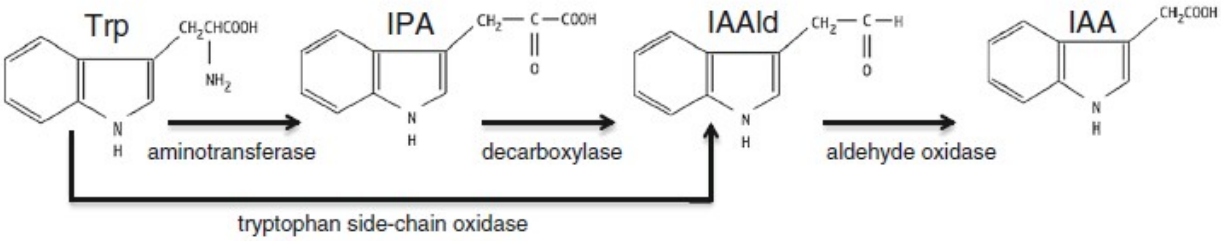


Figure 5. Patten, Cheryl & Glick, Bernard. (1996)[14]. Indole-3-pyruvate pathway. Trp =tryptophan, IPA = indole pyruvate acid, IAAId = Indole-3-acetaldehyde, IAA = indole-3-acetic acid.

The indole-3-pyruvate (IPA) pathway is a major auxin pathway in plants[15]. The IPA pathway is present in many bacteria such as phytopathogens (*Pantoea agglomerans*), plant beneficial bacteria (*Azospirillum, Bacillus, Bradyrhizobium, Enterobacter cloacae*, *Paenibacillus*, *Pseudomonas*, and *Rhizobium*), and in some cyanobacteria species[15].

In this pathway, tryptophan is converted in indole pyruvate acid (IPA) by lossing an amino group. This step is catalized by a transaminase (tryptophan aminotransferase). Then, IPA is decarboxylated by and indole-3-pyurave decarboxylase (IPDC), and finally converted to IAA by an aldehyde oxidase. This pathway is proposed to be connected to the tryptophan side-chain oxidase pathway, which we are not going to discuss since it has not been found in rhizobacteria.

2.2.2. Indole-3-acetamide pathway (IAM)

The indole-3-acetamide (IAM) pathway has been described mainly in phytopathogenic bacteria, although it does occur in phytosymbiotic bacteria as well[14].

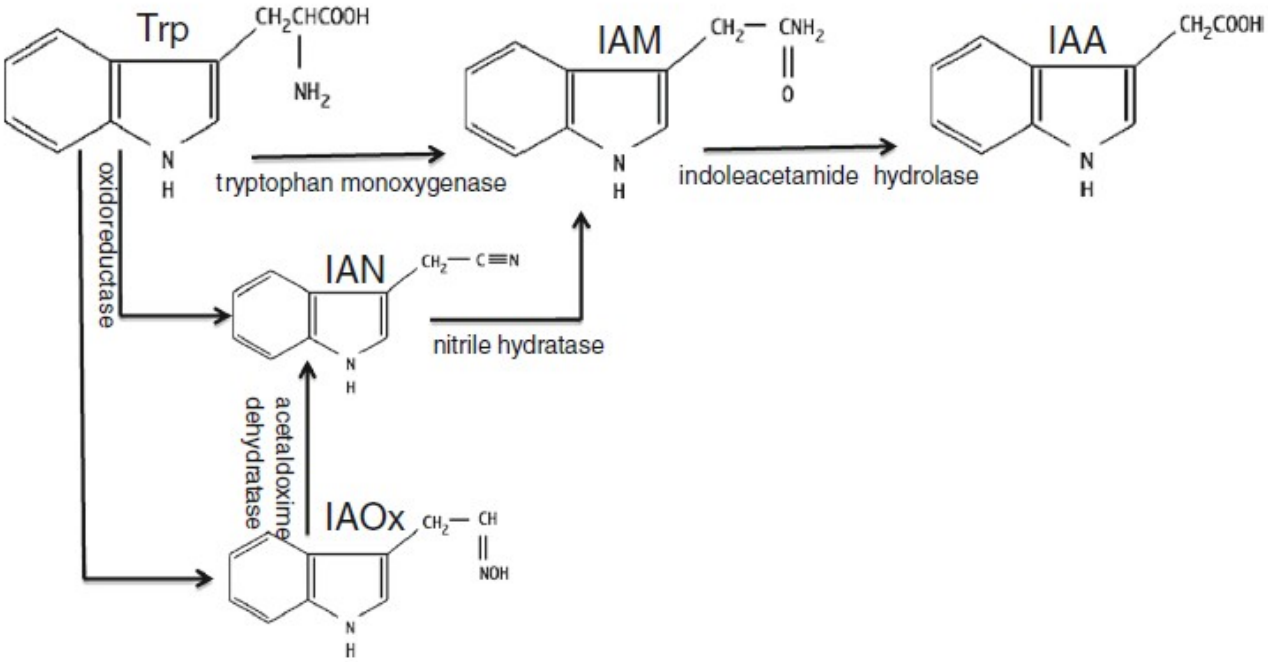


Figure 6. Patten, Cheryl & Glick, Bernard. (1996)[14]. Indole-3-acetamide pathway. IAM indole-3-acetamide, IAN indole-3-acetonitrile, IAA = indole-3-acetic acid, IAOx = indole-3-acetaldoxime.

The pathway has two different steps[15]. In the first one, a tryptophan monooxygenase catalyzes the conversion of tryptophan to IAM; then, IAM is hydrolyzed to IAA and by an indole-acetamide hydrolase.

Besides, IAM and IAN, two of the three pathways described in rhizobacteria, are connected in some species. In the cases where it is present, this connection is mediated by an enzyme with nitrile hydratase activity.

2.2.3. Indole-3-acetonitrile pathway (IAN)

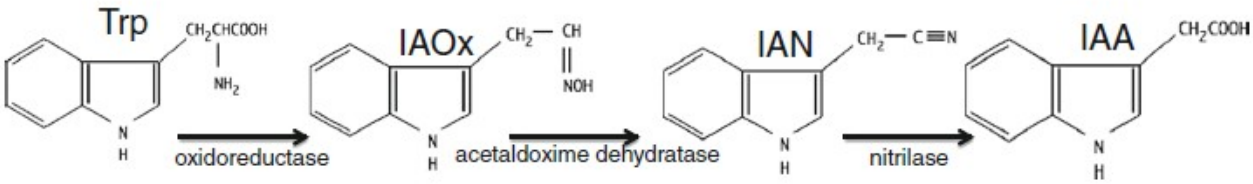


Figure 7. Patten, Cheryl & Glick, Bernard. (1996)[14]. Indole-3-acetonitrile pathway. Trp tryptophan, IAOx indole-3-acetaldoxime, IAN indole-3- acetonitrile, IAA indole3-acetic acid.

The first step of this pathway is the conversion of tryptophan into indole-3-acetaldoxime (IAOx) by an oxidoreductase. However, this enzyme has not been identified in bacteria so far[15]. The second step is the conversion of IAOx in indole-3-acetonitrile (IAN) by an acetaldoxime dehydratase. As in the previous step, the available information about this enzyme is scarse. Thus, for the analysis of this pathway we have study the phylogeney for the enzyme that mediates the third step, nitrilase, since the conversion of IAN in IAA is well documented.

2.2.4. Tryptophan biosynthesis

As part of this project, we have considered the possibility of study the tryptophan biosynthesis of some of the described species, as a way to have more information available when discussing our IAA results. Thus, we show the last steps of tryptophan biosynthesis, which are well described is several rhizosphere species.

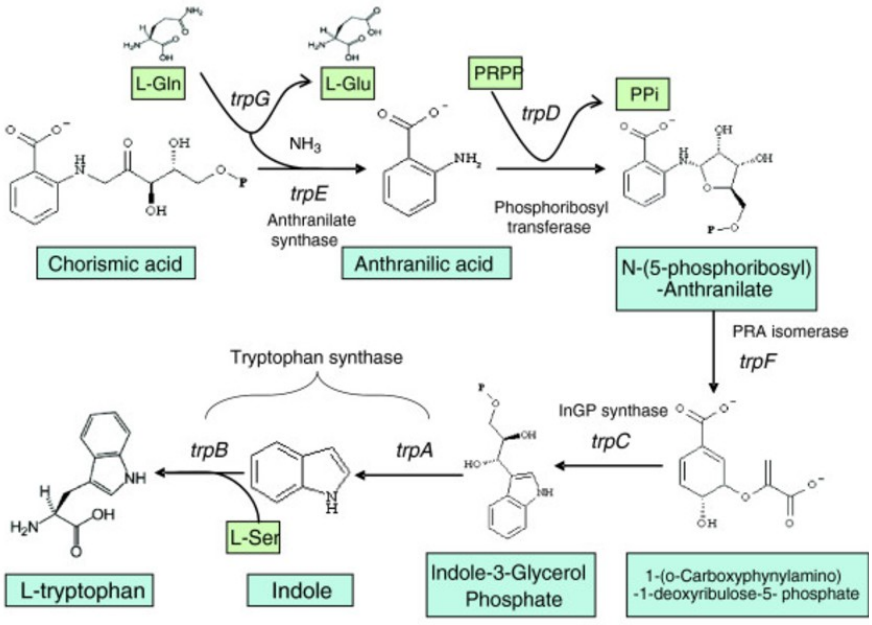


Figure 8. Kagan et al. (2008)[97]. Tryptophan biosynthesis pathway.

Antrhanilate synthase synthesize anthraninilic acid using chorismate as a substrate. Then, it is transformed to N-5-phosphorybosyl anthranilate by a phosphoribosyl transferase. Then, in the third step, N-5-phosphoribosyl anthranilate is converted to indole-3-glycerol phospate, which is transformed to tryptophan by tryptophan synthase in the last step.

2.3. CHAPTER 3: Main genes and protein profiles

2.3.1. IPA

2.3.1.1. Tryptophan aminotransferase

Tryptophan transaminase or simply aminotransferase (EC 2.6.1.27) is an enzyme that catalyzes the chemical reaction:

L-tryptophan + 2-oxoglutarate = (indol-3-yl)pyruvate + L-glutamate

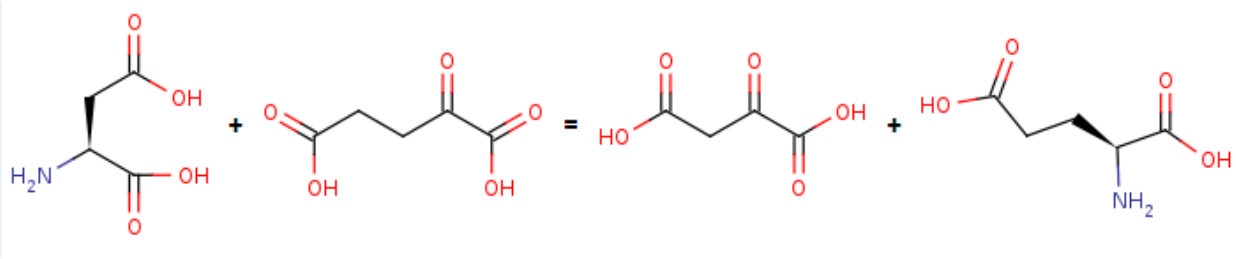


Figure 9. Tryptophan aminotransferase reaction (BRENDA database).

According to BRENDA database[110], this enzyme belongs to the family of transferases, specifically the transaminases, which transfer nitrogenous groups. The systematic name of this enzyme class is L-tryptophan:2-oxoglutarate aminotransferase. This enzyme needs one cofactor, pyridoxal phosphate.

A scarce number of genes that codify this transaminase have been identify. For instance, *patB* in some *Bacillus* species, *tyrB* in *Pseudomonas, tatA* in *Sinorhizobium*, and *phhC* in *Rhizobium.* In some of the rest of the species in our list, it is suggested that a general aromatic aminoacid transferase could be the main enzyme regulating this step instead of tryptophan transaminase[85]. Therefore, we have included both cases in our list (*annex 1),* where the different amount of genes that codify these transaminases are shown.

For describing the structure of trytophan transaminase, we have selected the well-known structure of this enzyme in the plant model *Arabidopsis thaliana*, which IAA biosynthesis pathway is the best documented among all living organisms. There, we summarize several aspects of the proteic profile of this enzyme, according to *PDB database:*

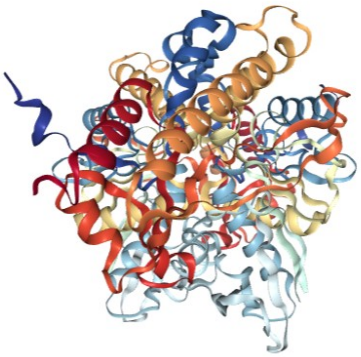
PDB code	*Organism*
3BWO	*Arabidopsis thaliana*
General view	***Structure***
	• *6 identical chains (391 residues each one)* • *Secondary structure:* *-34% helical (15 helices, 134 residues)* *-17% beta sheet (19 strands, 68 residues)*

Table 1 . Tryptophan transaminase (PDB protein profile)

2.3.1.2. Indole-3-pyruvate decarboxylase

Indolepyruvate decarboxylase (EC 4.1.1.74) is an enzyme that catalyzes the chemical reaction:

3-(indol-3-yl)pyruvate = 2-(indol-3-yl)acetaldehyde + CO2

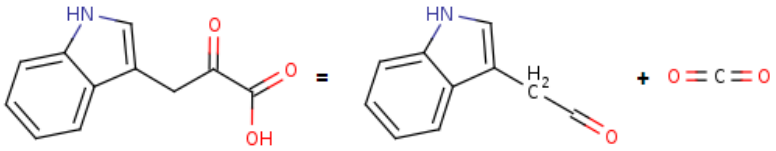


Figure 10. Indole-3-pyruvate decarboxylase reaction (BRENDA database)

This enzyme, which belongs to the family of lyases, has one substrate, 3-(indol-3-yl)pyruvate, and two products, 2-(indol-3-yl)acetaldehyde and CO2.

Other names for in enzyme are indol-3-yl-pyruvate carboxy-lyase and 3-(indol-3-yl)pyruvate carboxy-lyase.

In rhizobacteria, this enzyme is codified by the *ipdC* gene. Thus, other genes, such as *pdc1*, have been indentified. In order to show the protein structure of the enzyme, we have selected the indolepyruvate-3-decarboxylase from *Enterobacter cloacae*:

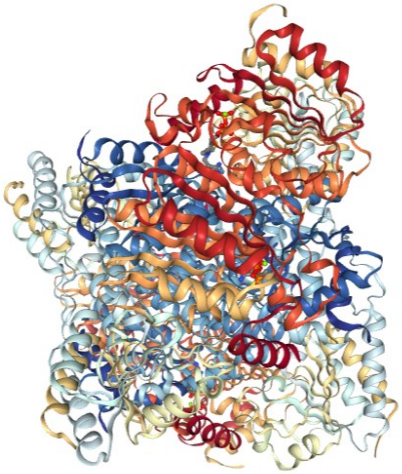
PDB code	***Organism***
1OVM	*Enterobacter cloacae*
General view	***Structure***
	• *4 identical chains (552 residues each one)* • *Secondary structure:* *-40% helical (28 helices, 224 residues)* *-16% beta sheet (22 strands, 93 residues)* • *Ligands:* *-Thiamine diphosphate* *-Magnesium ion*

Table 2. Indole-3-pyruvate decarboxylase (PDB protein profile)

2.3.1.3. Indole-3-acetaldehyde dehydrogenase

The enzyme indole-3-acetaldehyde dehydrogenase (EC 1.2.3.7) catalyzes the chemical reaction:

indole-3-acetaldehyde + NAD+ + H2O = IAA + NADH + H+

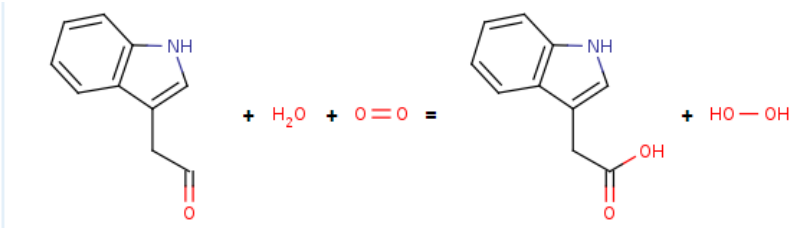


Figure 11. Indole-3-acetaldehyde dehydrogenase reaction (BRENDA database)

Several genes that codify this enzyme have been identified in rhizobacteria. These are: the *ald* family (*aldA*, *aldX*, *aldY*), *dhaS* and *ywdH*. Further details of the prevalence of these genes are provided in the phylogenetic analysis.

For the characterization of this enzyme, we have analyzed the protein profile of this protein in *Bacillus cereus*.

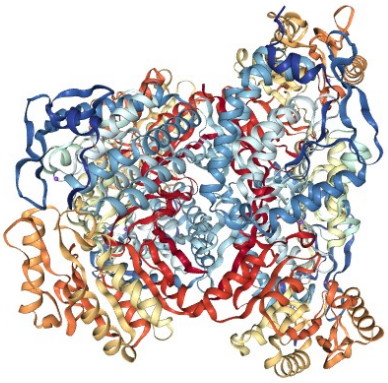
PDB code	***Organism***
4QET	*Bacillus cereus*
General view	***Structure***
	• *4 identical chains (494 residues each one)* • *Secondary structure:* *- 41 %helical (21 helices, 206 residues)* *-21 % beta sheet (26 strands, 104 residues)* • *Ligands:* *- Sodium ion*

Table 3. Indole-3-acetaldehyde dehydrogenase (PDB protein profile)

2.3.2. IAM

2.3.2.1. Tryptophan 2-monooxygenase

Tryptophan 2-monooxygenase (EC 1.13.12.3) is an enzyme that catalyzes the chemical reaction:

L-tryptophan + O2 = (indol-3-yl)acetamide + CO2 + H2O

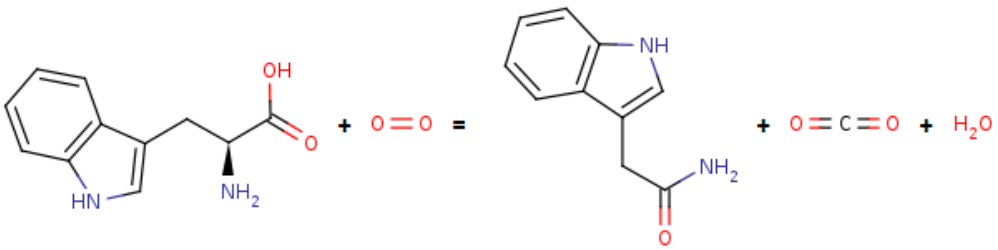


Figure 12. Tryptophan 2-monooxygenase reaction (BRENDA database)

Tryptophan-2-monooxygenase is an oxidoreductase wich incorporates two atoms of oxygen into the substrate. The main genes that codify this enzyme are *iaaM* and *tam1*. For the characterization of this enzyme, we have analyzed the protein profile of this protein in *Pseudomonas savastanoi*.

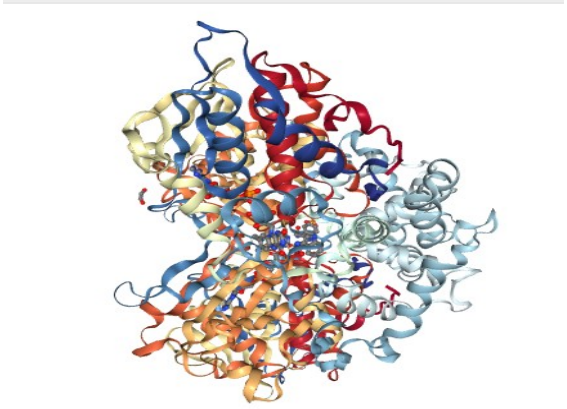
PDB code	*Organism*
4IV9	*Pseudomonas savastanoi*
General view	***Structure***
	• *2 identical chains (557 residues each one)* • *Secondary structure:* *- 36%helical (26 helices, 204 residues)* *-19% beta sheet (37 strands,107 residues)* • *Ligands:* *-Phosphate ion* *-Flavin – adenin dinucleotide* *-1,2 – ethanediol* *- 2 - (1H-indol-3-yl)acetamide*

Table 4. Tryptophan-2-monooxygenase (PDB protein profile)

2.3.2.2. Indole-3-acetamide hydrolase

Indole-3-acetamide hydrolase (EC 3.5.1.4) is an enzyme that catalyzes the chemical reaction:

indole-3-acetamide + H2O= IAA + NH3

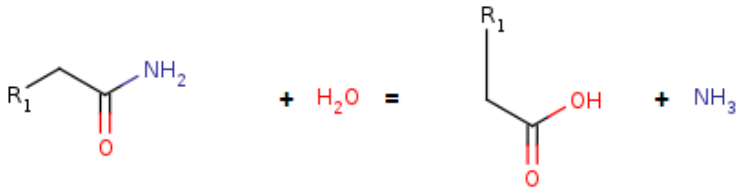


Figure 13. Indole-3-acetamide hydrolase reaction (BRENDA database)

This enzyme belongs to the family of amidases. In rhizobacteria, it is codified mainly by two genes (*iaaH* and *tms2*).

For the characterization of this enzyme, we have analyzed the protein profile of this protein in *Rhodococcus sp.*.

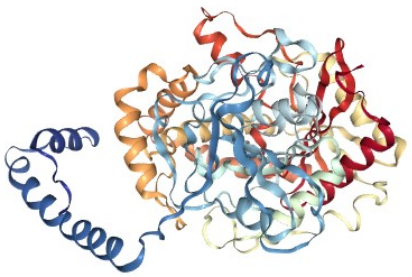
PDB code	*Organism*
3A1K	*Rhodococcus sp. N771*
General view	***Structure***
	• *1 chain (521 residues each one)* • *Secondary structure:* *- 38 %helical (20 helices, 198 residues)* *-15 % beta sheet (19 strands,80 residues)*

Table 5. Indole-3-acetamide hydrolase (PDB protein profile)

2.3.3. IAN

2.3.3.1. Nitrile hydratase

Nitrile hydratases (EC 4.2.1.84) is an enzyme that catalyzes the chemical reaction:

$$R\text{-}C{\equiv}N + H2O \rightarrow R\text{-}C(O)NH2$$

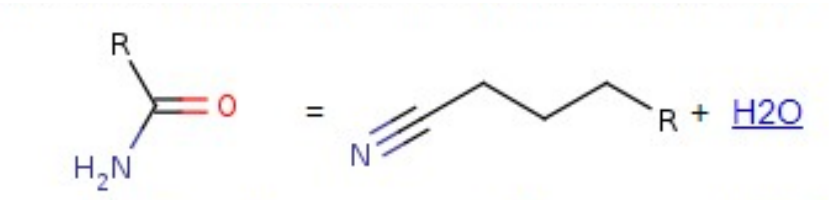


Figure 14. Nitrile hydratase reaction (BRENDA database)

This enzyme, which needs iron or cobalt as cofactors, catalyze the hydration of diverse nitriles to their corresponding amides. In rhizobacteria, it is codified by the genes *nthA* (alpha subunit) and *nthB* (beta subunit).

For the characterization of this enzyme, we have analyzed the protein profile of this protein in *Rhodococcus erythropolis*.

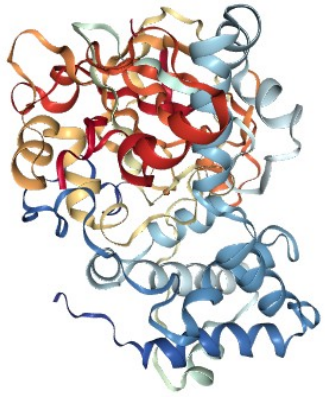
PDB code	*Organism*
1AHJ	*Rhodococcus erythropolis*
General view	***Structure***
	• *2 subunits (A and B)* • *4 identical chains in each subunit, and different among subunits (207 residues in subunit A, 212 residues in subunit B)* • *Secondary structure sub.A:* *- 38 %helical (9 helices, 80 residues)* *-11 % beta sheet (6 strands, 24 residues)* • *Secondary structure sub.B:* *- 32 %helical (10 helices, 68 residues)* *-14 % beta sheet (7 strands, 30 residues)* • *Ligands:* *-Fe ion*

Table 6. Nitrile hydratase (PDB protein profile)

2.3.3.2. Nitrilase

Nitrilase (EC 3.5.5.1) is an enzyme that catalyzes the chemical reaction:

c

Nitrile + H2O = Carboxylate + NH3

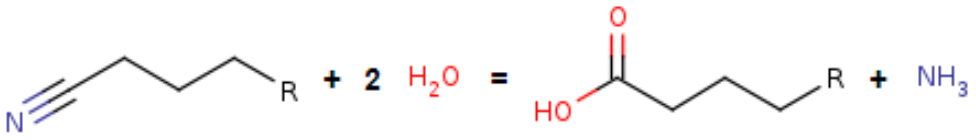


Figure 15. Nitrilase reaction (BRENDA database)

As shown in the picture, nitrilases catalyze the hydrolysis of nitriles to carboxylic acids (in our case, IAA) and ammonia. In rhizobacteria, they are codified by several genes, mainly *nit* and *yhcX*. There are no available structural models for this enzyme in *PDB*.

2.3.4. Tryptophan biosynthesis

2.3.4.1. Anthranilate synthase

Anthranilate synthase (EC 4.1.3.27) is an enzyme that catalyzes the chemical reaction:

chorismate + L-glutamine = anthranilate + pyruvate + L-glutamate

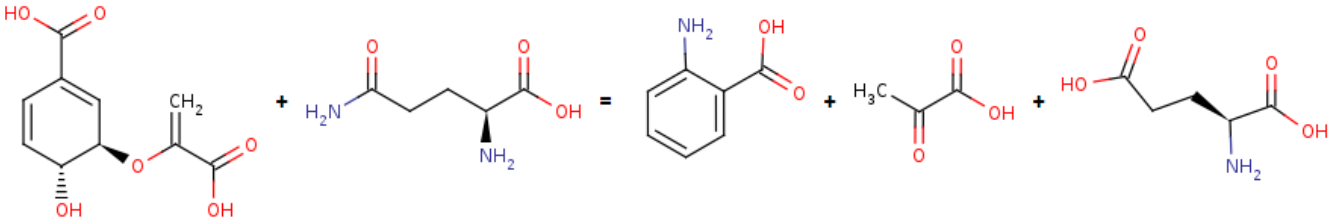


Figure 16. Anthranilate synthase reaction (BRENDA database)

This enzyme takes part in different pathways, such as the biosynthesis of antibiotics, biosynthesis of secondary metabolites, phenazine biosynthesis and phenylalanine, tyrosine and tryptophan biosynthesis. This enzyme is codified by the gene *trpE*.

For the characterization of this enzyme, we have analyzed the protein profile of this protein in *Serratia marcescens*.

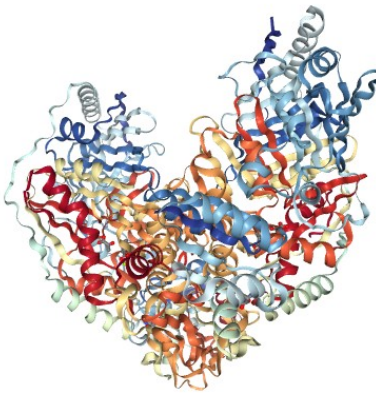
PDB code	***Organism***
1I7Q	*Serratia marcescens*
General view	***Structure***
	• *2 subunits (A and B)* • *2 identical chains in each subunit, and different among subunits (519 residues in subunit A, 193 residues in subunit B)* • *Secondary structure sub.A:* *- 31 %helical (15 helices, 163 residues)* *-31 % beta sheet (33 strands, 166 residues)* • *Secondary structure sub.B:* *- 29 %helical (7 helices, 56 residues)* *-36 % beta sheet (11 strands, 70 residues)* • *Ligands:* *-Pyruvic acid -Magnesium ion* *-Benzoic acid -Glutamic acid*

Table 7. Anthranilate synthase (PDB protein profile)

2.3.4.2. Tryptophan synthase

Tryptophan synthase (EC 4.2.1.20) is an enzyme that catalyzes the chemical reaction:

L-serine + 1-C-(indol-3-yl)glycerol 3-phosphate = L-tryptophan + D-glyceraldehyde 3-phosphate + H2O

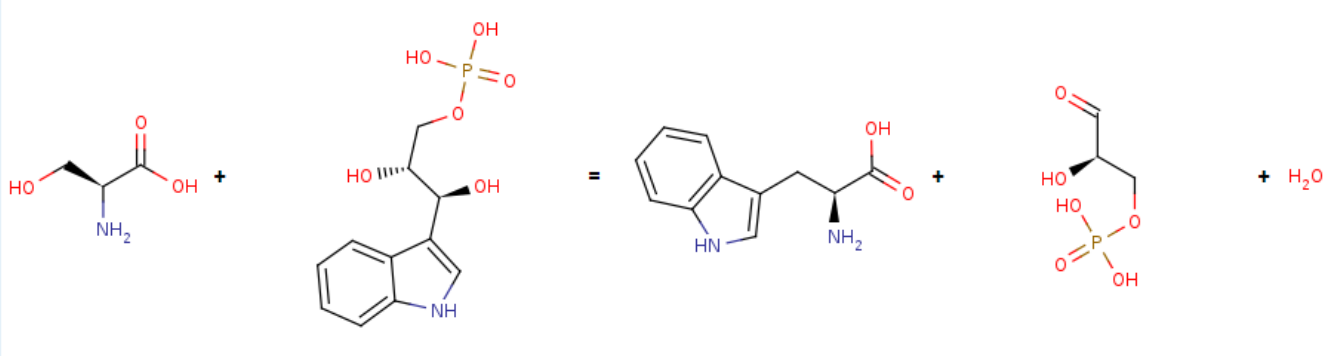


Figure 17. Tryptophan synthase reaction (BRENDA database)

This enzyme has two subunits (alpha and beta).The first one catalyzes the conversion of 1-C-(indol-3-yl)glycerol 3-phosphate to indole and D-glyceraldehyde 3-phosphate This indole is taken by the beta subunit, where it is converted to L-tryptophan.

The main two genes that codify this enzyme are *trpA* and *trpB.* For the characterization of this enzyme, we have analyzed the protein profile of this protein in *Escherichia coli*.

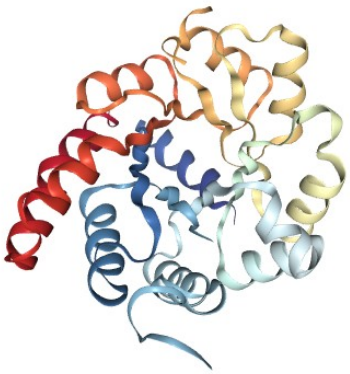
PDB code	***Organism***
1V7Y (alpha subunit)	*Escherichia coli*
General view	***Structure***
	• *2 identical chains (268 residues)* • *Secondary structure:* *- 45 %helical (12 helices, 122 residues)* *- 13 % beta sheet (9 strands, 35 residues)* • *Ligands:* *-Sulfate ion*

Table 8. Tryptophan synthase (PDB protein profile)

2.4. CHAPTER 4: Main IAA rhizobacteria species, known pathways and coocurrence

According to our bibliographic research, the species which show IAA biosynthesis ability are summarized in the following table. We have also mentioned which are the known pathways for each genus. Besides, all of them are able to synthesize tryptophan starting from chorismate. Complementary information to this table is available in *annex 1*.

Genus / Species	*Evidence of IAA biosynthesis*	*Known pathways*
Bacillus sp.	*Chagas et al. (2019) [14] Spaepen S, Vanderleyden J. (2011) [15] Patten C, Glick B (1996) [16]*	*IPA, IAN*
Bacillus cereus	*Ozdal et al. (2016)[17]*	*IPA*
Bacillus subtilis	*Wagi S, Ahmed A. (2019)[18]*	*IPA, IAN*
Bacillus mycoides	*Ghazal et al. (2013) [19]*	*IPA*
Bacillus thuringiensis	*Chagas et al. (2019) [14]*	*IPA*
Bacillus amyloliquefaciens	*Shao et al. (2015)[20]*	*IPA*
Bacillus filamentosus	*Yahaghi et al. (2018)[21]*	*IPA*
Bacillus megaterium	*Lee et al. (2016) [22]*	*IPA*
Rhodopseudomonas sp.	*Patten C, Glick B (1996)[16]*	*IPA, IAN*
Rhodopseudomonas palustris	*Lo et al. (2018) [23]*	*IPA, IAN*
Dickeya sp.	*Pauline B. (2017) [24]*	*IPA*
Dickeya zeae	*Zhou et al. (2015)[25]*	*IPA*
Pantoea sp.	*Spaepen S, Vanderleyden J. (2011)[15] Patten C, Glick B (1996)[16]*	*IPA, IAM*
Pantoea ananatis	*Coutinho TA, Venter SN (2009) [26]*	*IPA*
Pantoea rwandensis	*Brady C et al. (2012) [27] Estenson K et al. (2018) [28]*	*IPA*
Pantoea dispersa	*Kulkarni G et al. (2013) [29]*	*IPA, IAM*
Pantoea agglomerans	*Apine OA, Jadhav JP (2011) [30]*	*IPA*
Serratia sp.	*Ouyang et al. (2016)[31]*	*IPA*
Serratia liquefaciens	*Zelaya-Molina et al. (2016)[32]*	*IPA*
Serratia marcescens	*Hasuty A, Choliq A (2018)[33] Khan AR et al. (2017)[34]*	*IPA*
Serratia fonticola	*Jung et al. (2017)[35]*	*IPA*
Acinetobacter sp.	*Spaepen S, Vanderleyden J. (2011)[15] Patten C, Glick B (1996)[16]*	*IPA*

Acinetobacter baumannii	*Lin et al. (2018)*[36] *Lin et al. (2012)*[37]	*IPA*
Acinetobacter indicus	*Sachdev et al. (2010)*[38]	*IPA*
Acinetobacter bohemicus	*Sachdev et al. (2010)*[38]	*IPA*
Acinetobacter junii	*Huddedar et al. (2002)*[39]	*IPA*
Rhizobium sp.	*Spaepen S, Vanderleyden J. (2011)*[15] *Datta C and Basu P (2000)*[40] *Bhattacharyya R and Pati B (2000)*[41] *Kobayashi et al. (1995)*[42] *Patten C, Glick B (1996)*[16]	*IPA, IAM, IAN*
Rhizobium etli	*Spaepen et al. (2009)*[43]	*IPA, IAM, IAN*
Rhizobium miluonense	*Ghosh et al. (2013)*[44]	*IPA, IAM, IAN*
Rhizobium leguminosarium	*Bhattacharjee et al. (2011)*[45]	*IPA, IAM, IAN*
Rhizobium lusitanum	*Dubey et al. (2011)*[46]	*IPA, IAM, IAN*
Rhizobium acidisoli	*Cruz-González et al. (2017)*[47]	*IPA, IAM, IAN*
Rhizobium favelukessii	*Del Papa et al. (0216)*[48]	*IPA, IAM, IAN*
Rhizobium tropicii	*Eddie L et al. (2017)*[49]	*IPA, IAM*
Rhizobium mesoamericanum	-	*IAM, IAN*
Azospirillum sp.	*Spaepen S, Vanderleyden J. (2011)*[15] *El-Khawas H, Adachi K (1999)*[50] *Akbari et al. (2007)*[51] *Baca et al. (1994)*[52] *Patten C, Glick B (1996)*[16]	*IPA*
Azospirillum brasiliense	*Molina et al. (2018)*[53]	*IPA*
Enterobacter sp.	*Spaepen S, Vanderleyden J. (2011)*[15] *Patten C, Glick B (1996)*[16]	*IPA*
Enterobacter cloacae	*Bose et al. (2016)*[54]	*IPA*
Erwinia sp.	*Spaepen S, Vanderleyden J. (2011)*[15] *Bradl MT, Lindow SE (1996)*[55] *Clark et al. (1993)*[56]	*IPA*
Erwinia amylovora	*Yang et al. (2007)*[57]	*IPA*
Klebsiella sp.	*Patten C, Glick B (1996)*[16] *LU ZX, Song W (1999)*[58]	*IPA, IAN*
Klebsiella oxytoca	*Celloto et al. (2012)*[59]	*IPA*
Klebsiella pneumoniae	*Sachdev et al. (2009)*[60]	*IPA*
Klebsiella michiganensis	*Mitra et al. (2018)*[61]	*IPA, IAN*
Kitasotaspora sp.	*Shrivastava et al. (2008)*[62]	*IPA*
Kitasotaspora setae	*Shrivastava et al. (2008)*[62]	*IPA*
Acetobacter sp.	*Spaepen S, Vanderleyden J. (2011)*[15] *Bastian et al. (1998)*[63]	*IPA*

	Patten C, Glick B (1996)[16]	
Acetobacter diazotrhropicus	*Bastian et al. (1998)* [63]	*IPA*
Sinorhizobium sp.	*Spaepen S, Vanderleyden J. (2011)*[15] *Patten C, Glick B (1996)*[16]	*IPA, IAN*
Sinorhizobium fredii	*Vinardell et al. (2015)*[64]	*IPA, IAN*
Sinorhizobium meliloti	*Imperlini et al. (2009)*[65]	*IPA, IAN*
Sinorhizobium medicae	*Kallala et al. (2018)*[66]	*IPA, IAN*
Pseudomonas sp.	*Spaepen S, Vanderleyden J. (2011)*[15] *Ahmad et al. (2005)*[67] *Balaji et al. (2012)*[68] *Patten C, Glick B (1996)*[16]	*IPA, IAM, IAN*
Pseudomonas fluorescens	*Suzuki et al. (2003)*[69]	*IPA, IAN*
Pseudomonas enthomophila	*Ansari F, Ahmad I (2018)*[70]	*IPA*
Pseudomonas aeruginosa	*Marathe et al. (2017)* [71]	*IPA*
Pseudomonas putida	*Patten CL, Glick BR (2002)*[72]	*IPA*
Pseudomonas rhizosphaerae	*Elena A et al. (2007)* [73]	*IPA*
Pseudomonas syringae	*Surico et al. (1985)*[74] *Flores et al. (2018)*[75]	*IPA*
Pseudomonas protegens	*Andreolli et al. (2018)*[76]	*IAM*
Bradyrhizobium sp.	*Spaepen S, Vanderleyden J. (2011)*[15] *Sekine et al.(1988)*[77] *Patten C, Glick B (1996)*[16]	*IPA, IAM, IAN*
Bradyrhizobium japonicum	*Egebo et al. (1991)*[78] *Jensen et al. (1995)*[79] *Siqueira et al. (2014)*[80]	*IPA, IAM, IAN*
Bradyrhizobium elkanii	*Yagi et al. (2000)*[81]	*IPA, IAM, IAN*
Bradyrhizobium diazoefficiens	*Siqueira et al. (2014)* [80]	*IPA, IAM, IAN*
Ralstonia sp.	*Patten C, Glick B (1996)*[16]	*IPA, IAN*
Ralstonia pickettii	*Nair S, Vakil B (2015)* [82]	*IPA*
Ralstonia mannitolilytica	*Abhishek et al. (2016)*[83]	*IPA*
Ralstonia solanacearum	*Kurosawa et al. (2009)*[84]	*IAN*
Rhodococcus sp.	*Patten C, Glick B (1996)*[16] *Spaepen S, Vanderleyden J. (2011)*[15]	*IPA, IAN*
Rhodococcus jostii	*Daiana et al. (2014)*[85]	*IPA, IAN*
Rhodococcus erythropolis	*Daiana et al. (2014)*[85]	*IPA, IAN*
Burkholderia sp.	*Spaepen S, Vanderleyden J. (2011)*[15] *Patten C, Glick B (1996)*[16] *Zuñiga et al. (2013)*[86]	*IPA, IAN, IAM*
Burkholderia mallei	*Johan et al. (2008)*[87]	*IPA, IAN, IAM*
Burkholderia cepacia	*Castanheira et al. (2015)* [88]	*IPA, IAN, IAM*

Burkholderia pseudomallei	*Johan et al. (2008)* [87]	*IPA, IAN, IAM*
Paraburkholderia sp.	*Donoso et al. (2016)*[89]	*IPA, IAM*
Paraburkholderia phymatum	*Mannaa et al. (2018)*[90]	*IPA, IAM*
Paraburkholderia xenovorans	*Mannaa et al. (2018)*[90]	*IPA. IAM*
Variovorax sp.	*Patten C, Glick B (1996)*[16]	*IPA, IAM*
Variovorax paradoxus	*Jiang et al. (2012)*[91]	*IPA, IAM*
Paenarthrobacter sp.	*Asano et al. (1982)*[92]	*IAN*
Paenarthrobacter aurescens	*Cai et al. (2014)*[93]	*IAN*
Agrobacterium sp.	*Spaepen S, Vanderleyden J. (2011)*[15] *Mano Y (2012)*[94] *Patten C, Glick B (1996)*[16]	*IAM, IAN*
Agrobacterium tumefaciens	*Kutaeek M, Rovenska J (1991)*[95]	*IAM, IAN*
Agrobacterium fabrum	*Kutaeek M, Rovenska J (1991)*[95]	*IAM, IAN*
Agrobacterium vitis	-	*IAM, IAN*
Agrobacterium rhizogenes	*Schaerer S, Pilet P (1993)*[96]	*IAM, IAN*

Table 9. IAA rhizobacteria species and known pathways

We can perform a co-ocurrence analysis of the different pathways, in order to determine how often do they coexist in the same species. According to all the data gathered in the *annex* 1, we find that:

	IAM	***IAN***	***IPA***
IAM	*22*	*15*	*14*
IAN	*15*	*31*	*21*
IPA	*14*	*21*	*65*

Table 10. Co-occurrence matrix

As mentioned before, IPA is the most frequent pathway (65 species), followed by IAN (31 species) and IAM (22 species). For the 32,2% of the species where IPA is present, IAN coexist. The same applies for the 21,5% of IPA/IAM species. We can also see the high correlation between IAN and IAM (15 of the 31 IAN species have an active IAM pathway, and the same applies for the 22 IAM species where IAN is found). Since this information comes from a bibliographical research it could be possible that, due to lack of evidence, some pathways are present in species where we say they are not. Thus, this co-ocurrence analysis should be interpreted as a tool that gives meaning in the context of our study, taking into account the inherent *bias* associated to this type of research.

2.5. CHAPTER 5. Results: diversity analysis I

In order to perform the diversity analysis I described in the chapter *Objectives*, we have obtained all the information available concerning genes and proteins of the identified rhizobacteria species. Therefore, protein accession numbers of the sequences we are working with can be found in the *annex 1*. Since the protein sequence is under selective constraint for protein function and protein structure, and these are conserved over much longer periods than the individual codon choices, we have decided to carry out our diversity analysis using the protein sequences instead of DNA sequences. If we were looking for differences within a closely related group of species (for instance, a bunch of species from the same strain), DNA would be a better option, but in our case we are working with species that are far away from each other from an evolutionary point of view, and the higher conservation status of protein sequences in comparison to DNA sequences will be a plus in our analysis.

In order to run our analysis, we have created several *csv* document, each of them with the accession numbers of the identified proteins for each enzyme under study. These *csv* worked as an input for the *R* pipeline we have designed. Further information about the structure of these *csv* documents and the *R* pipeline can be found in *annex 2*.

2.5.1. Anthranilate synthase

The gene that regulates the synthesis of anthranilate has been identified in 72 species of rhizobacteria that have IAA activity. In all of them, the gene has been identified as *trpE*, or suggested to be *trpE* according to comparative analysis.

As we can see in our phylogenetic tree (figure 18) using *trpE* sequences, there is a clear stratification by genus. Starting by the top of the tree, we can found a *cluster* constituted by *Paenibacillus* and *Bacillus* species. Then, another *cluster* includes those species of *Serratia*, *Enterobacter*, *Klebsiella* and *Pantoea*. Following, we can find a group integrated by *Agrobacterium*, *Rhizobium* and *Sinorhizobium*, which are genus that are closely related (as we will see later, according to the results of the *gyrB* phylogenetic marker). *Azospirillum* and *Bradyrhizobium* species are a branch of this group.

At the bottom of the tree, we find the group with the highest number of branches. The fist one includes those species from *Rhodococcus*, *Pseudarthobacter* and *Paenarthrobacter*. Then we find *Acinetobacter*, and also the *trpE* sequence from *Bacillus thuringiensis YBT – 1518*, that we expected to find in the top of the tree with the other *Bacillus* species since there is a clear stratification for this enzyme. The next

branch also goes according to the expected evolutionary relation observed, and is configurated by *Ralstonia*, *Burkholderia* and *Paraburkholderia* species. Finally, the last branch gather all the species from the genus *Pseudomonas*.

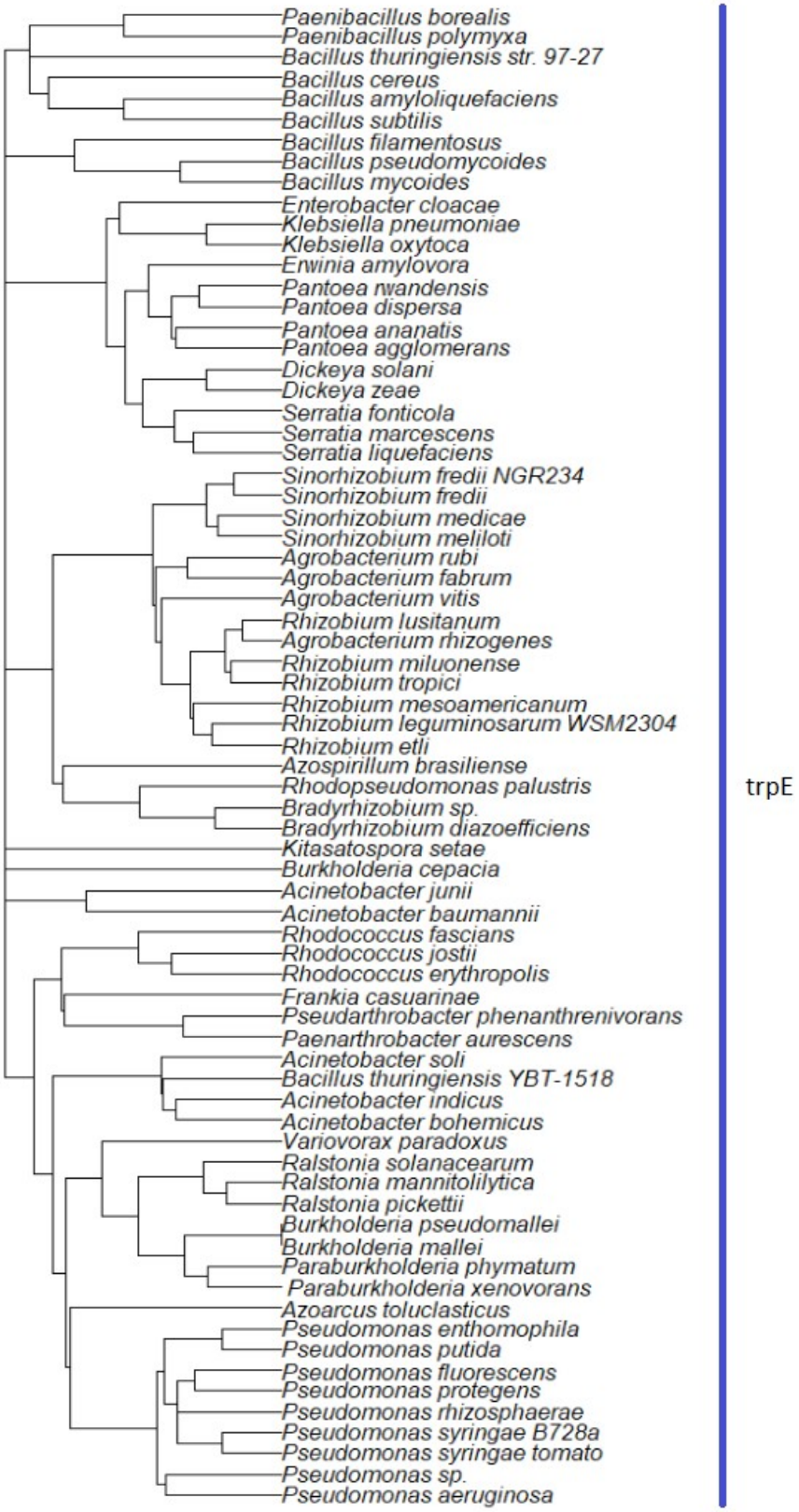


Figure 18. Anthranilate synthase phylogeny.

2.5.2. Tryptophan synthase (alpha subunit)

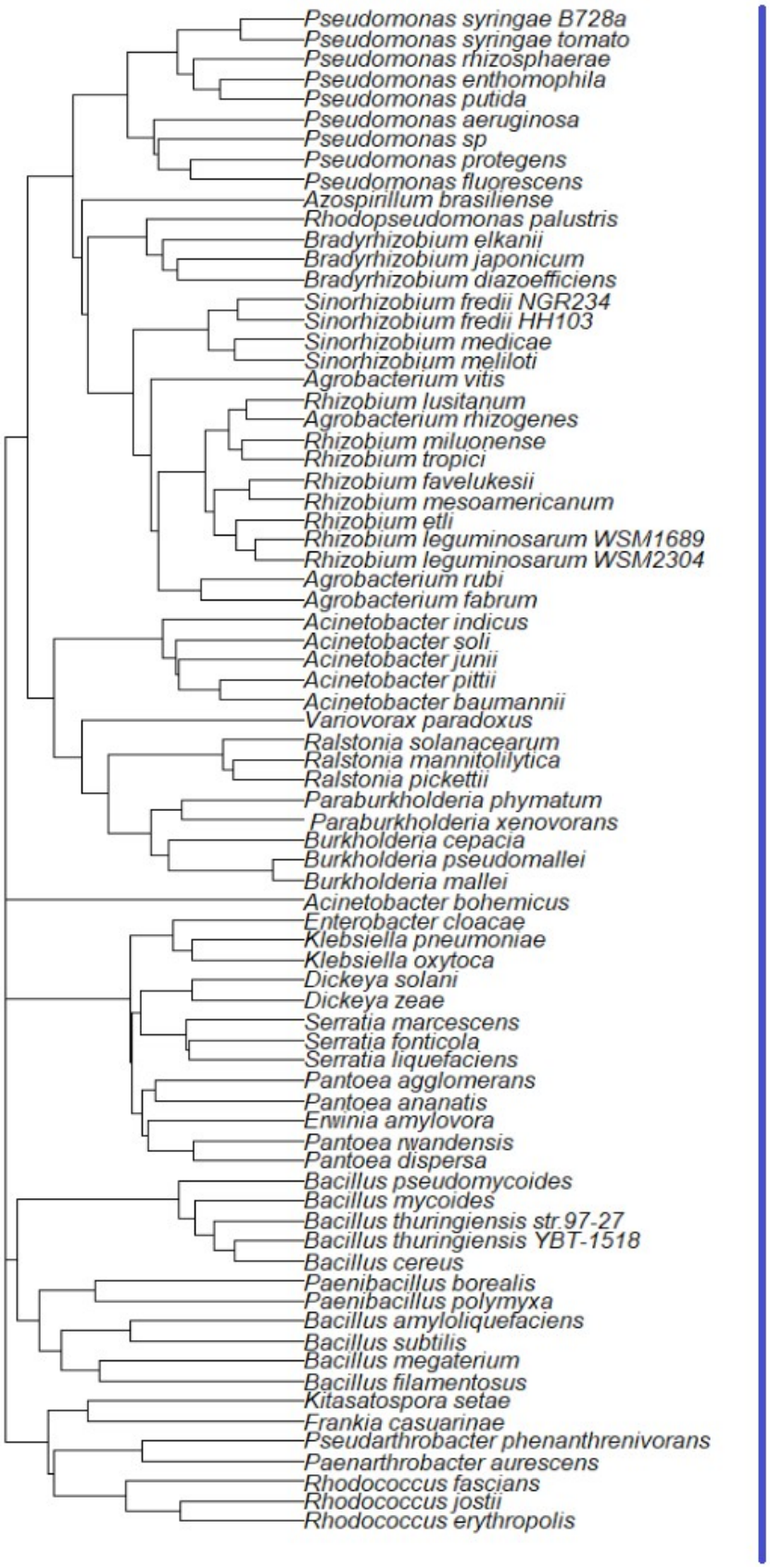


Figure 19. Tryptophan synthase (alpha subunit) phylogeny.

The gene that regulates the synthesis of the tryptophan synthase's alpha subunit has been identified in 76 species of rhizobacteria that also show IAA activity. In all of them, this gene has been identified as *trpA*, or suggested to be *trpA* according to comparative analysis. The reason why we have four 76 instead of 72 as in the previous case is because there are four species where *trpE* sequences have not been identified or at least mentioned in the bibliography.

This phylogenetic tree (figure 19) is similar to the one observed for anthranilate synthase. As we well explain later after showing the beta subunit's tree, there is reason for this resemblance. Starting from the top of the tree, we observe a well-defined *cluster* constituted by the species of the genus *Pseudomonas*. In the next branch, we find that *Azospirillum, Rhodopseudomonas, Bradyrhizobium, Sinorhizobium, Rhizobium* and *Agrobacterium* species are close from a phylogenetic point of view according to *trpA*. These two branches were unrelated in the anthranilate synthase tree.

The first branch of the next cluster contains all the *Acinetobacter* species under analysis. Then, *Variovorax, Ralstonia, Burkholderia* and *Paraburkholderia* appear in the next branch. If we continue our path to the bottom of the tree, we will find once again the *cluster* of *Enterobacter, Klebsiella, Dickeya, Pantoea, Serratia* and *Erwinia*. Once again, we have our group of *Bacillus* and *Paenibacillus* (this time, with the expected result of *Bacillus thuringiensis* YBT – 1518 being close to *Bacillus thuringiensis str.97-27*) and, as the last branch of the tree, the group *Kitasatospora, Frankia, Pseudarthrobacter, Paenarthrobacter* and *Rhodococcus*.

2.5.3. Tryptophan synthase (beta subunit)

The gene that regulates the synthesis of the tryptophan synthase's beta subunit has been identified in 76 species of rhizobacteria which also show IAA activity. In all of them, this gene has been identified as *trpB*, or suggested to be *trpB* according to comparative analysis.

As the tree shows (figure 20), the groups or *clusters* observed are very similar to those in the alpha subunit, with slight differences. Starting from the top of the tree, we find the group *Erwinia – Pantoea -Klebsiella – Serratia – Dickeya.* Then, there is a new *cluster* similar to those observed before, constituted by *Paenarthrobacer, Pseudarthrobacter, Kitasatosporae, Rhodococcus* and *Frankia*. After that, we find the big group of *Agrobacterium, Sinorhizobium, Rhizobium*, followed by the *Pseudomonas* group, the *Bradyrhizobium-Azospirillum-Rhodopseudomonas* group and then by the

Acinetobacter-Burkholderia-Paraburkholderia group. Finally, we can see how all the *Bacillus* and *Paenibacillus* species are closely related according to the *trpB* phylogeny.

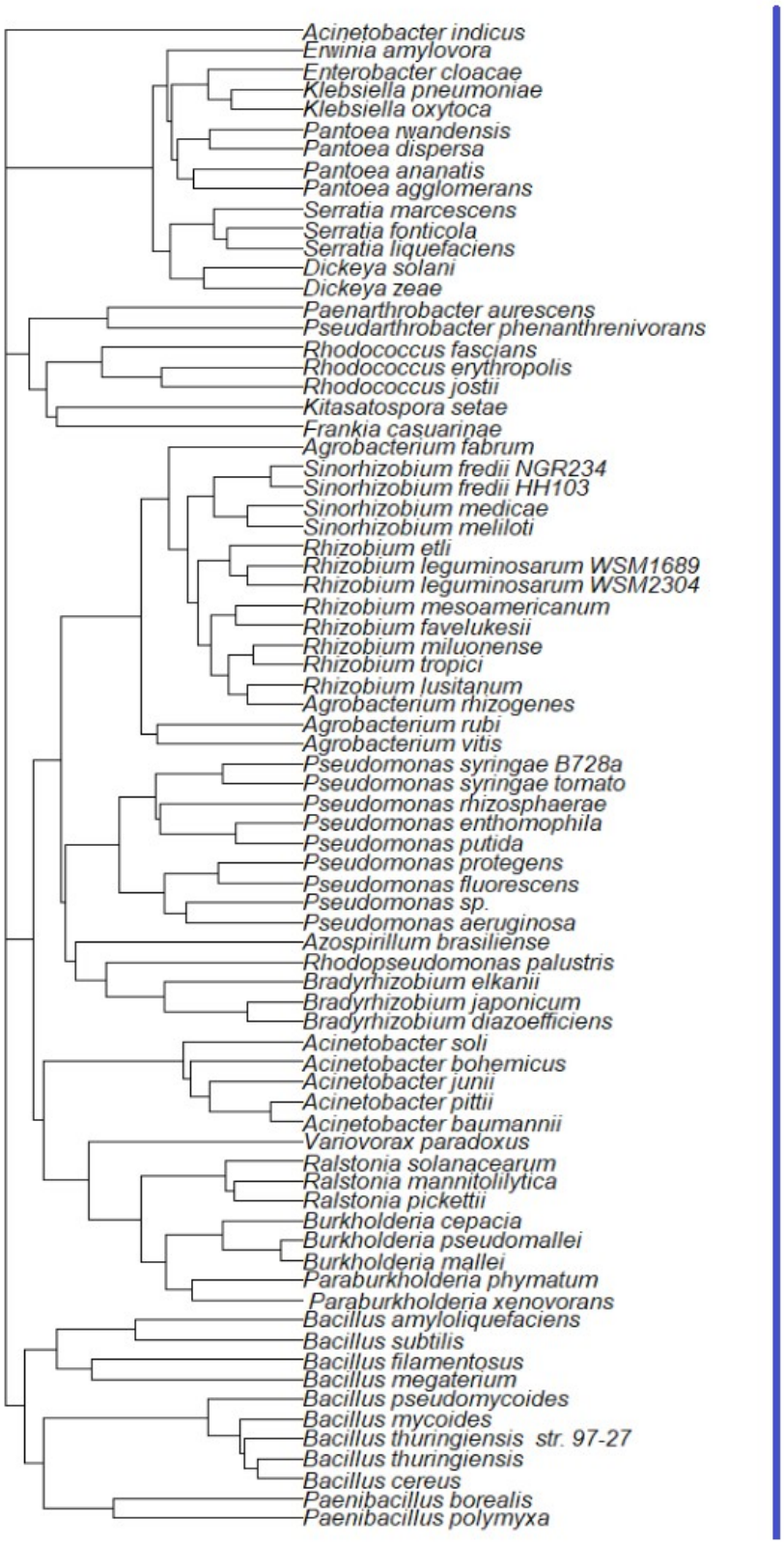


Figure 20. Tryptophan synthase (beta subunit) phylogeny.

2.5.4. Trp genes co-ocurrence: the trp operon

Having reached this point, it could be suggested that the high degree of co-ocurrence and similarity between the phylogenetic structures of the three enzymes described has a reasonable explanation. As we have mentioned before, these three enzymes take part in the tryptophan biosynthesis pathway, which transform the available chorismate into tryptophan that will be used by some rhizobacteria as the substrate in the production of IAA through different pathways. Nevertheless, we do not see this level of co-occurence in none of the enzymes of these routes, so we could suggest that the reason for the pattern we have observed might be due to the presence of a well-conserved operon. As it is know, there is a *trp* operon well-described in some groups of bacteria[98], and we suggest that this is present in almost all the species in our study.

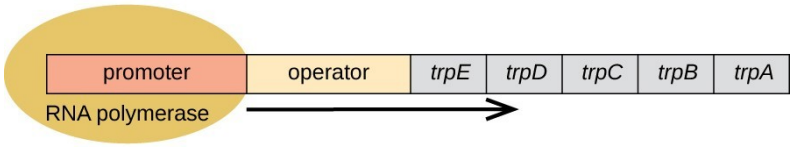


Figure 21. Merino *et al.* (2008)[98] : Suggested trp orperon structure

Thus, we have analyzed, using *Softberry* (a tool that allows us to predict the presence or absence of operons), the genetic region where *trpE*, *trpB* and *trpA* are located in some of the species we are working with in order to find evidence for the presence of this operon. For instance, if we take *Rhodococcus jostii*, one of the species from the bottom branch of the *trpA* tree, and introduce the sequence for this genomic region in the S*oftberry* browser, we find that this data suggest that there is one functional operon:

```
Length of sequence - 5548 bp
Number of predicted genes - 6
Number of transcription units - 1, operons - 1
    N      Tu/Op   Conserved  S             Start          End    Score
                   pairs(N/Pv)
    1     1 Op  1      .       +    CDS          1 -      1548    985
    2     1 Op  2      .       +    CDS       1520 -      2536    681
    3     1 Op  3      .       +    CDS       2529 -      3278    631
    4     1 Op  4      .       +    CDS       3283 -      3930    381
    5     1 Op  5      .       +    CDS       3911 -      5113   1238
    6     1 Op  6      .       +    CDS       5106 -      5547    384
```

Figure 22 (from Softberry). Genetic region trpF, trpE, trpD, trpC, trpB and trpA from Rhodococcus jostii.

Therefore, we can perform a comparison of some of these regions using *Easyfig* in order to understand the structure of this operon:

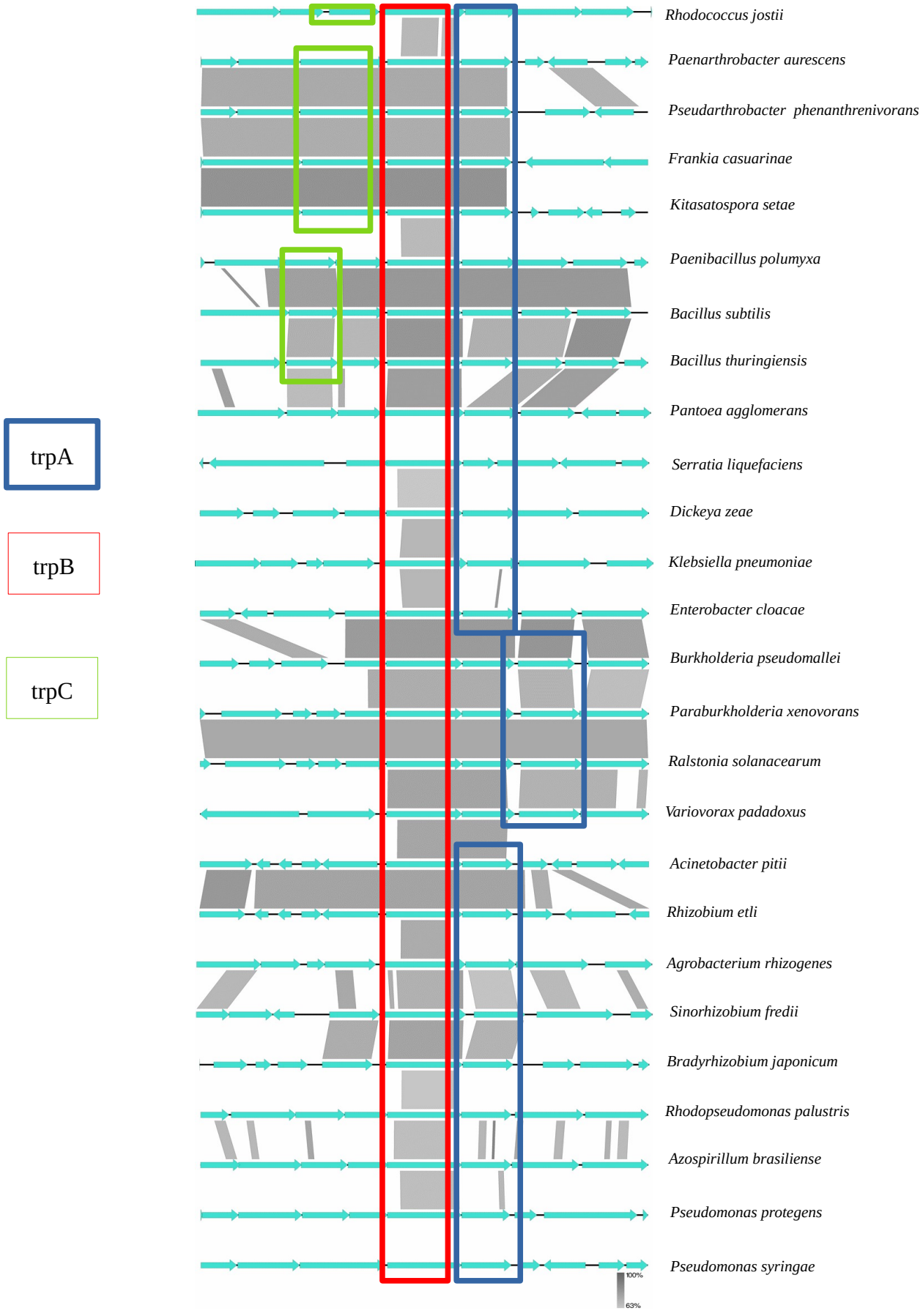


Figure 23 **.**trpA/trpB region comparison

As we can see in almost all the species tested, with the exceptions of *Burkholderia, Paraburkholderia*, *Variovorax* and *Ralstonia*, *trpB* is located immediately near to *trpA*. Nevertheless, in these exceptions only one gene is among them.

We see that, from the top to the bottom of the comparison matrix (figure 23), the distance from the *trpB-trpA* tandem to the rest of the operon *(trpC, trpE* and *trpD*) tends to get bigger. The pattern observed for this changes has some similarities to that observed for the structure of both *trpB* and *trpA* trees, were species where *trpB-trpA* are near to *trpC-trpD* are closer in these trees, and the same works for those were *trpB-trpA* region is far away from the rest of the operon's genes.

2.5.5. Tryptophan transaminase

The information about the genes that codify this enzyme is scarce. We have identified 41 species where these genes have been determined or suggested. At this point, 4 different genes are responsible for the synthesis of this transaminase (*tyrB1*, *tyrB* and *phhC* in *Pseudomonas*, *patB* in *Bacillus* and *tatA* in *Sinorhizobium).*

As we can see in the phylogenetic tree (figure 24), those species where *tyrB1* or *tyrB* is know or suggested are displayed together in the upper part. This could suggest that both *tyrB1* and *tyrB* are closer genes from an evolutionary point of view. In the middle part, *Bacillus thuringiensis str. 97-27*, where *patB* has been identified, shows almost no relation to the rest of the tree, suggesting that it is probably the only sequence of this gene in our tree.

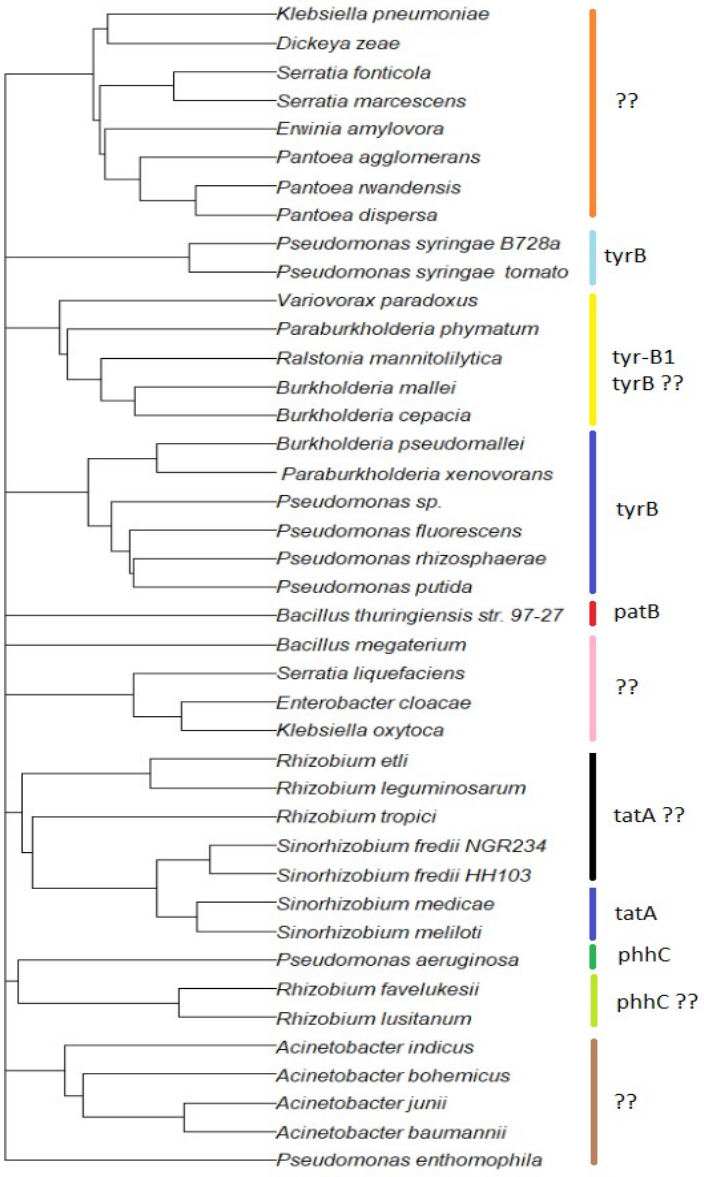


Figure 24. Tryptophan transaminase phylogeny.

Then, *Sinorhizobium* species where *tatA* is know as the gene that codifies this transaminase (*Sinorhizobium meliloti* and *Sinorhizobium medicae*) are grouped together, as expected. They form a *cluster* with the rest of the species from the same *genus*, and also with those of *Rhizobium*. Thus, and due to the evolutionary proximity suggested by the marker phylogeny, we could think that the suggested genes in *Rhizobium* are also *tatA*. Nevertheless, and since some species of *Rhizobium* where the transaminase gene has been suggested but not identified are in the same branch as *Pseudomonas aeruginosa*, we could think that *phhC* is also an active transaminase that takes part in this step of the IPA patwhay.

2.5.6.Indole-3-pyruvate decarboxylase

We have 31 species where the genes that codify this enzyme have been identified or suggested. Almost all the identified genes correspond to *ipdC*, but those from *Acinetobacter* are known to be *pdc1*. With the exception of *Rhizobium*, *Sinorhizobium* and *Pseudomonas*, the rest of the species where indole-3-pyruvate decarboxylase sequences have been identified or suggested but not associated to *ipdC* or *pdc1 (*this is, *Dickeya*, *Acetobacter*, *Azospirillum* and *Rhodopseudomonas)* are suggested to be *ipdC*, or at least very close to it in terms of evolutionary proximity. According to the tree (figure 25), *Rhizobium* and *Sinorhizobium* are related, and the same applies for the following *cluster* (*Azospirilum*, *Rhodopseudomonas* and *Bradyrhizobium)*.

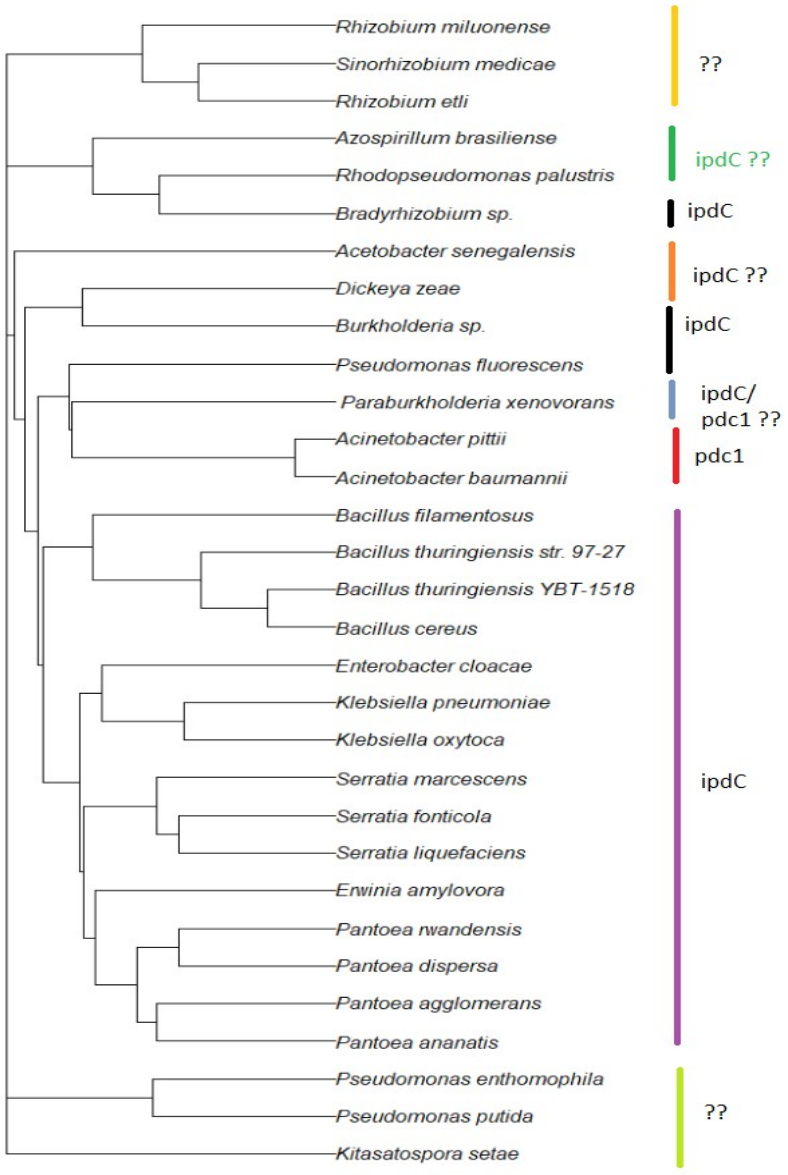


Figure 25. Indole-3-pyruvate decarboxylase phylogeny.

Then, there is a big *cluster* of *ipdC/pdc1* genes integrated by *Acetobacter, Dickeya, Burkholderia, Acinetobacter, Bacillus, Klebsiella, Enterobacter, Serratia, Erwinia* and *Pantoea*. Finally, indole-3-pyruvate decarboxylases from *Pseudomonas* and *Kitasatosporae* seem to have evolved different from those of the rest of species codified by *ipdC/pdc1*, suggesting another gene to be defined.

2.5.7.Indole-3-acetaldehyde dehydrogenase

We have 36 species where the genes that codify this enzyme have been identified or suggested. Since one species can have more than one dehydrogenases mediating the same process according to KEGG, we have decided to compare those codified by the *dhaS* gene, common to all the species of the list. Nevertheless, there are some genes that are proposed to take part in this step, mainly in *Bacillus*, such as *aldA* and *ywdH.*

According to our three, there is a big cluster starting from the bottom, which includes those *dhaS* sequences from *Rhizobium, Sinorhizobium* and *Pseudomonas*.

In the center of the tree, there is another well-defined group (*Serratia, Klebsiella Burkholderia, Ralstonia, Enterobacter)*.

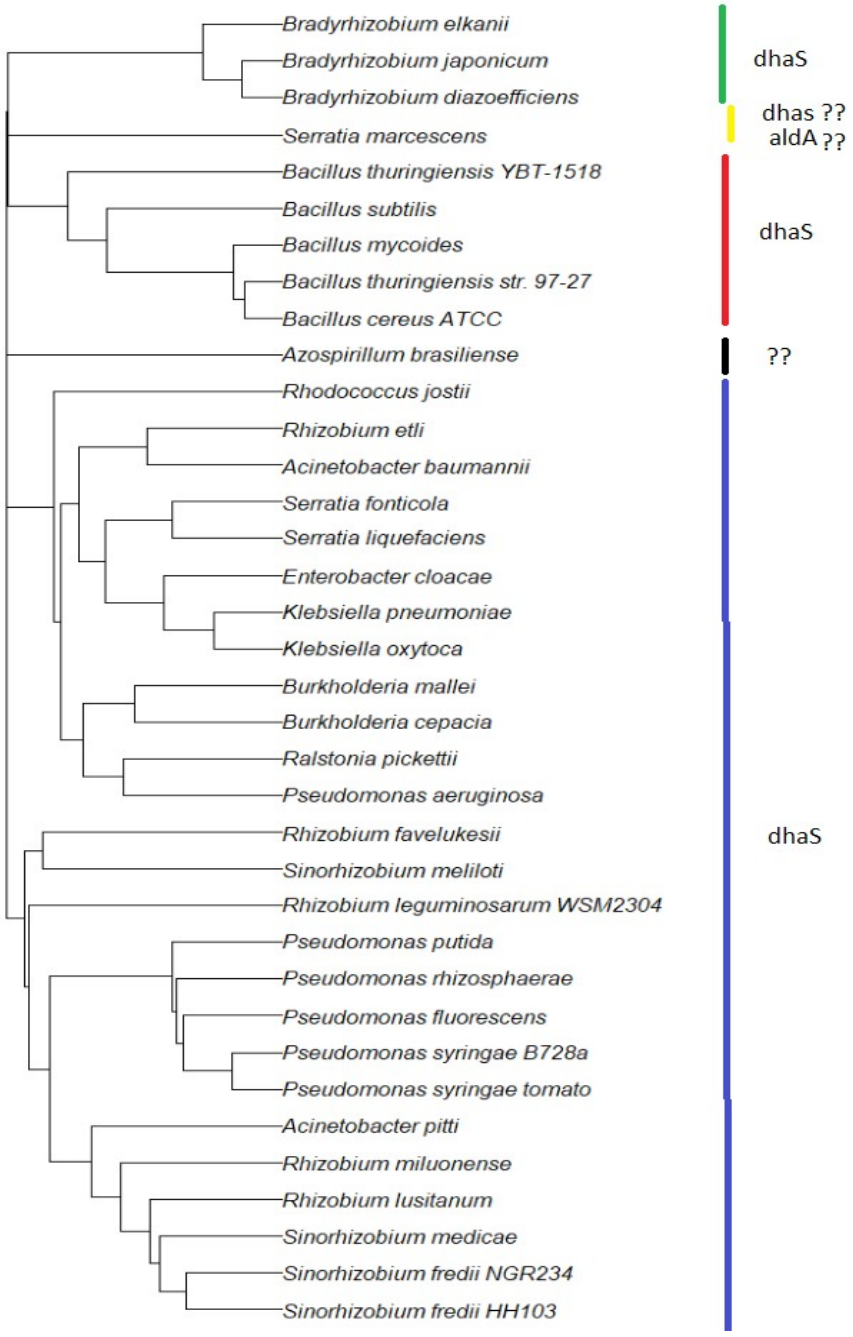


Figure 26. Indole-3-acetaldehyde dehydrogenase phylogeny.

In the upper part, there are two groups, constituted each one of them for only one genus (*Bradyrhizobium* and *Bacillus*). *Serratia marcescens*, *Pseudomonas aeruginosa* and *Azospirillum brasiliense* did not show the expected relation observed previously in the rest of the trees. This could mean that there is a problem with the sequence (an error in its identification as *dhaS*) or more probably that they have evolved differently due to natural selection.

2.5.8.Tryptophan-2-monooxygenase

The information about this enzyme is scarce. Thus, only genes for 7 species have been identified. The main gene that codifies this monooxygenase is *iaaM* and, in some species, *tms1*.

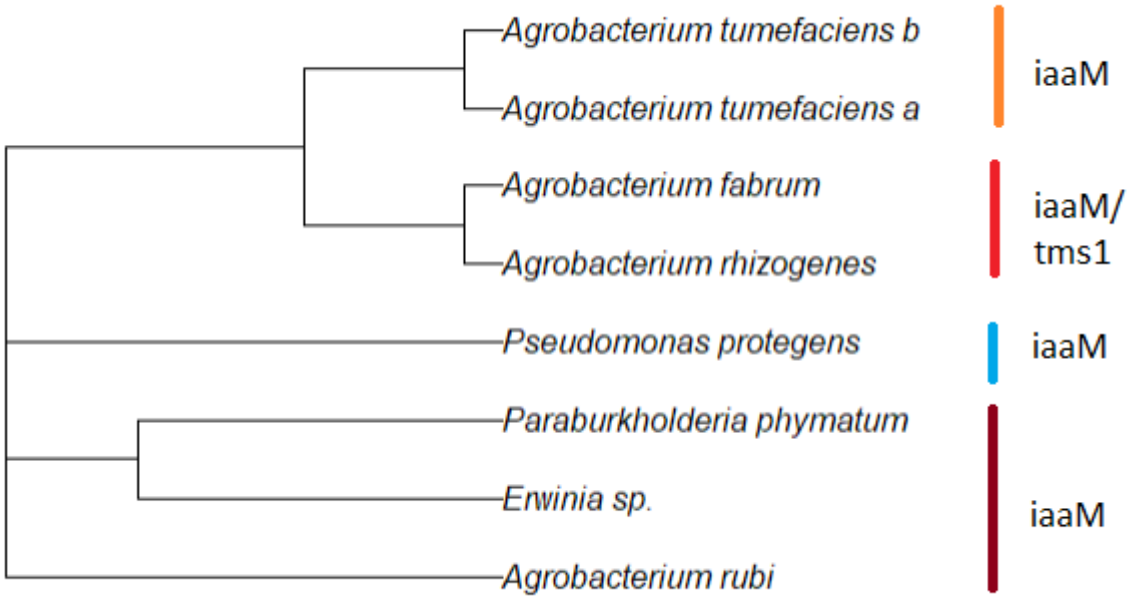


Figure 27. Tryptophan-2-monooxygenase phylogeny.

As shown in the tree (figure 27), *Agrobacterium tumefaciens*, *Agrobacterium fabrum*, and *Agrobacterium rhizogenes* are closely related, being *Agrobacterium rubi* the species which shows the higher distance among the *genus*. As we could expect, *Paraburkholderia* and *Erwinia* are clearly different from the *Agrobacterium* group.

2.5.9.Indole acetamide hydrolase

Genes for 7 species have been identified. The main gene that codifies this hydrolase is *iaaH.*

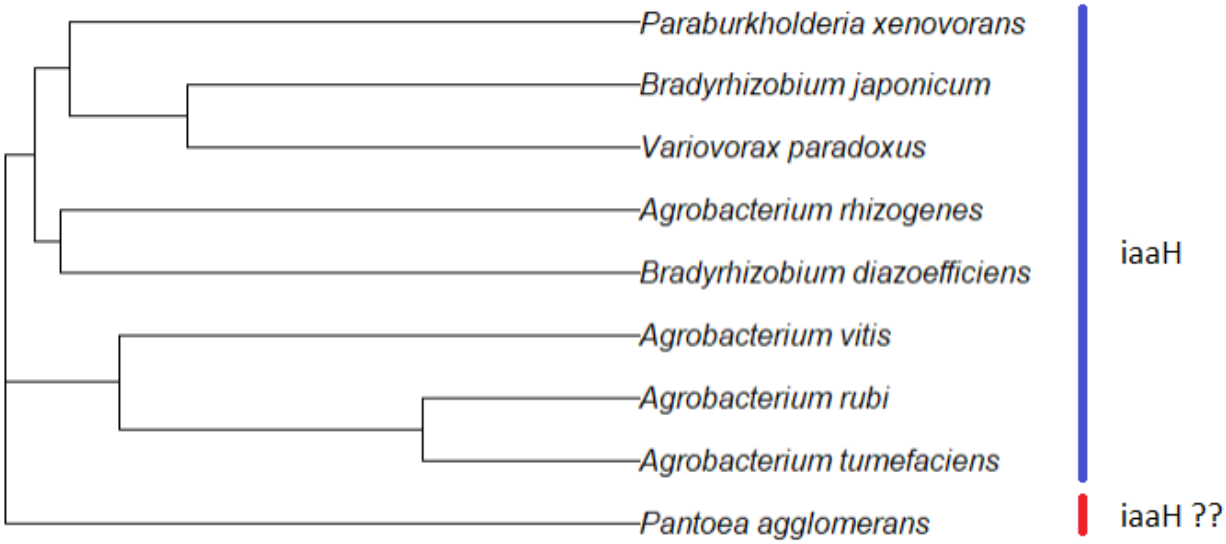


Figure 28. Indole acetamide hydrolase phylogeny.

As for the monooxygenase, we can see a *cluster* of *Agrobacterium* species (figure 28). Nevertheless, *Agrobacterium rhizogenes* seems to be closer to *Bradyrhizobium diazoefficiens*. Even though this could seem difficult to explain, both species are very close from an evolutionary perspective according to the information shown by the *gyrB* marker, and therefore this could explain why *Agrobacterium rhizogenes* is displayed in that position, since it could be possible that selective procedures lead to the observed divergence.

2.5.10. Nitrile hydratase (alpha subunit)

Genes for 16 species have been identified, being *nthA* the gene that codifies this enzyme.

We can clearly see (figure 29) 4 different *clusters* (*Rhodococcus*, *Bradyrhizobium, Sinorhizobium,* and *Rhizobium* group, from which *Rhizobium leguminosarum* is not included). Genus represented by a single species (*Agrobacterium, Paenarthrobacter*, *Klebsiella*) seem to have evolved differently to these 4 groups.

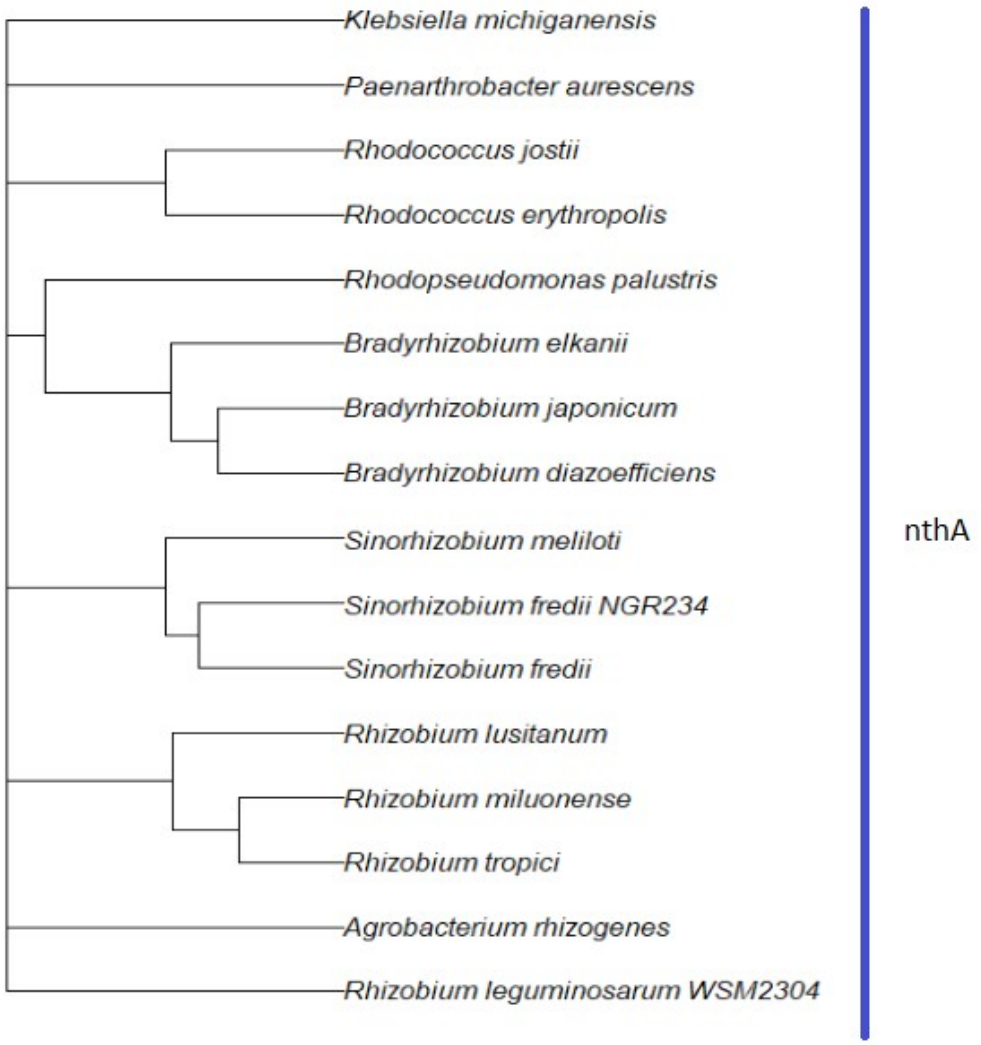


Figure 29. Nitrile hydratase (alpha subunit) phylogeny.

2.5.11. Nitrile hydratase (beta subunit)

Genes for 19 species have been identified. The main gene that codifies this hydratase is *nthB.*

As for *nthA*, *Paenarthrobacter aurescens* and *Klebsiella michiganensis* show the highest differences in comparison to other species (figure 30). Nevertheless, *Agrobacterium rhizogenes* appears as a part of the *Rhizobium* group. Besides, and since there were more information available concerning *Rhizobium*, there are two groups of this genus, each of them with two species.

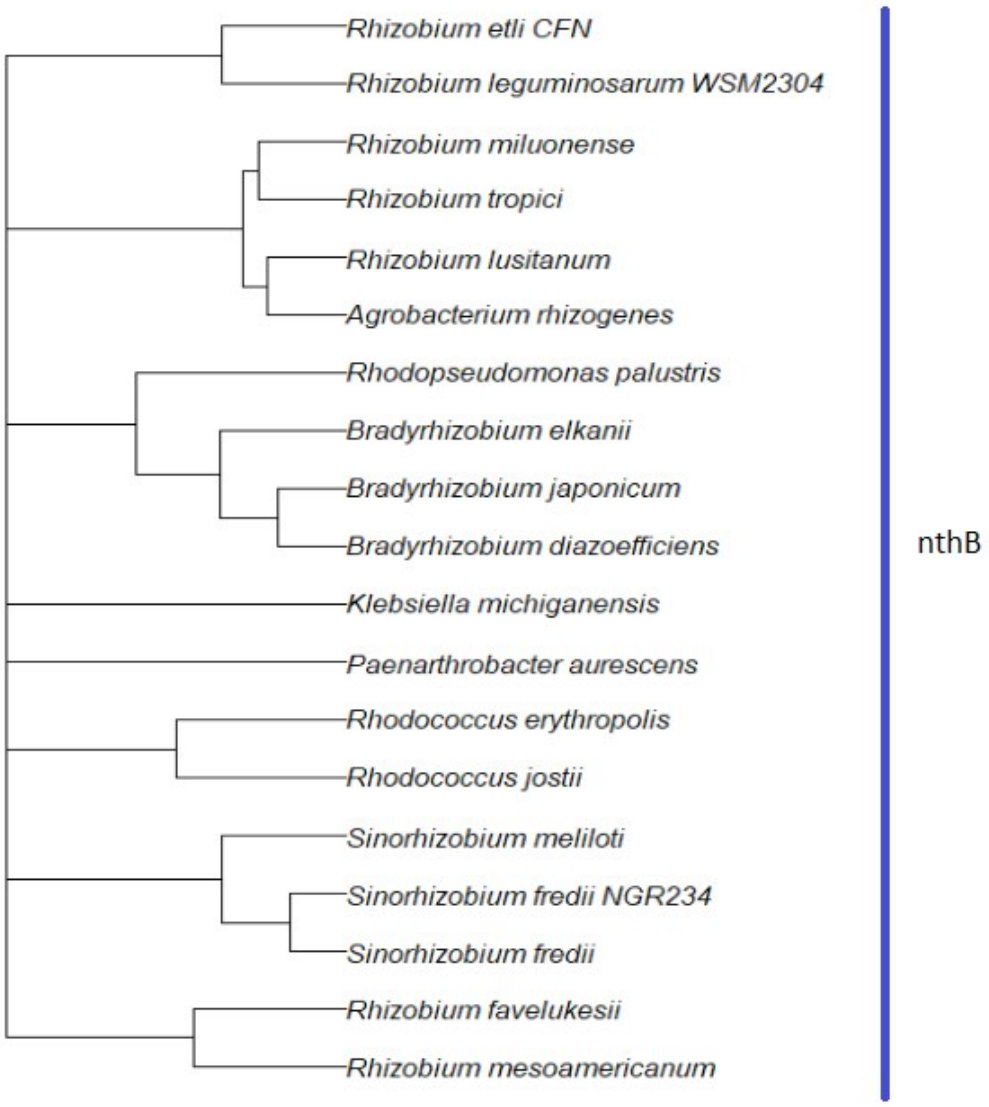


Figure 30. Nitrile hydratase (beta subunit) phylogeny.

2.5.12. Nitrilase

Genes for 14 species have been identified. The main genes that codify this nitrilase are *nit* and *yhcX*.

Since only two of the species we are working with show strong evidence for *nit* or *yhcX* as the gene that codify this nitrilase, we are working mainly with sequences that have been suggested but not identified as *nit* or *yhcX*. According to the phylogeny (figure 31), we suggest that these sequences could be also *yhcX* or *nit*.

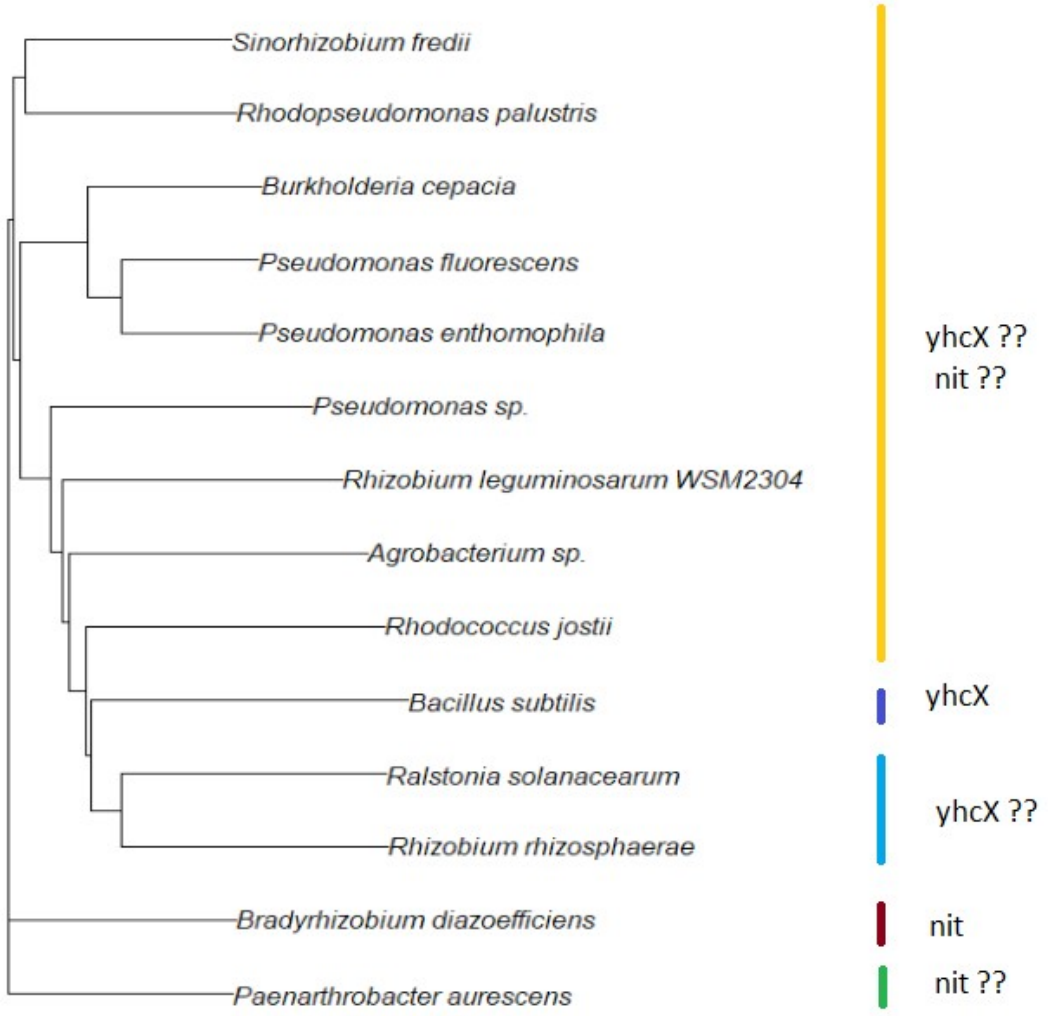


Figure 31. Nitrilase phylogeny.

2.6. CHAPTER 6: Results: diversity analysis II

Comparative analysis of the phylogenetic trees represented in the previous points has been done using DNA gyrase B (*gyrB*) as a marker. The reasons for choosing this marker instead of the multiple that are available (*recA, rpoB, RNA 16S,...)* are*:*

- *GyrB* is one of the most used, widely spread bacterial markers, and this allow us to compare our results to a higher number of previous research.
- Concerning the species that we have identified, there are at least several studies where this marker was used to establish phylogenetic relationships, being in some cases more discriminative than RNA 16S. For instance, *Sinorhizobium*[99],

Rhizobium[100], *Agrobacterium*[101], *Bacillus*[102], *Pseudomonas*[103], *Burkholderia*[104] and *Acinetobacter*[105].

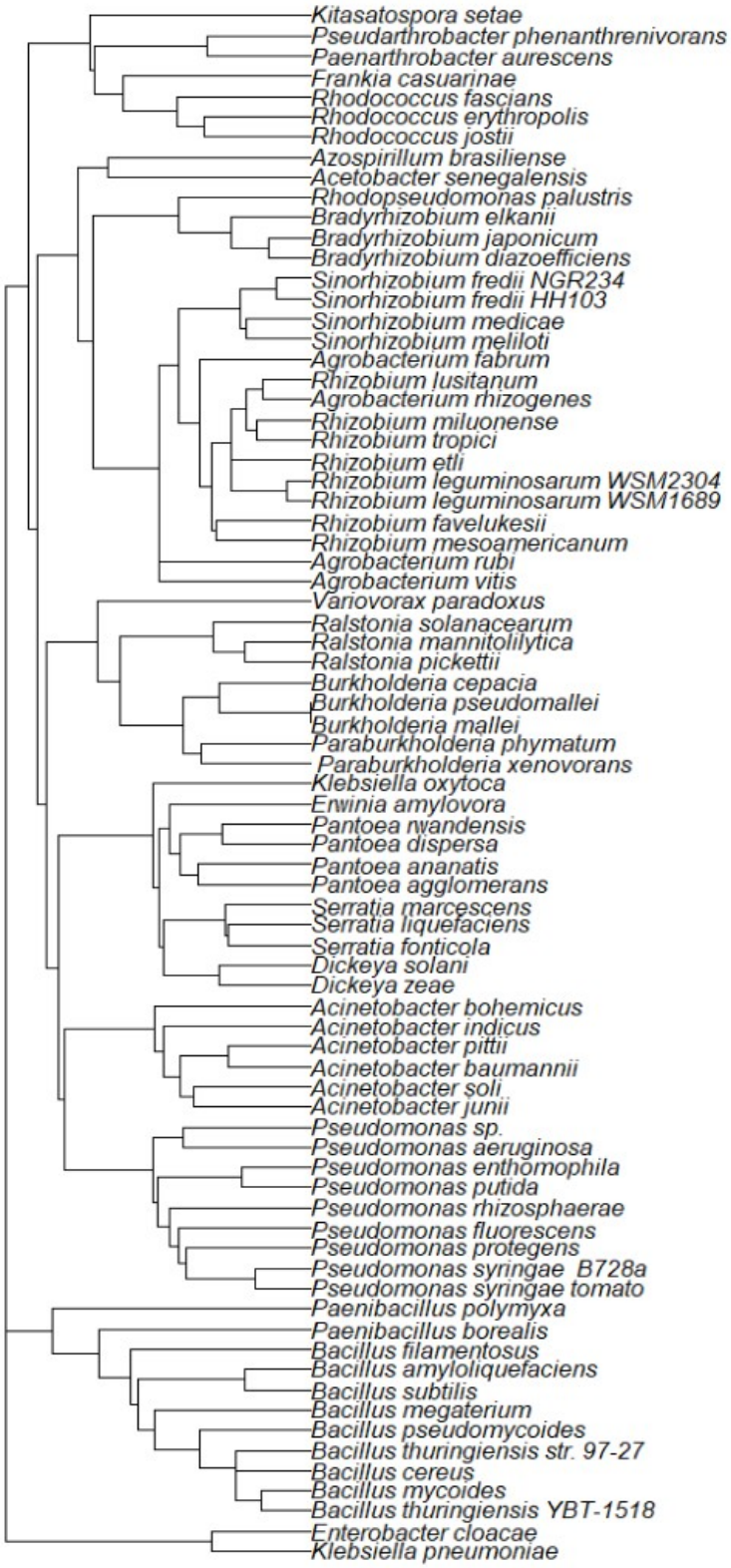


Figure 32. *gyrB* phylogeny.

As we can see (figure 32), the phylogenetic tree described by the *gyrB* marker gives us an insight into the evolutionary distance among each species and genus under study. As we can see starting from the bottom, there is a *Bacillus* group. In the upper part, of the tree, there is the group *Kitasatospora-Arthrobacter-Frankia- Rhodococcus.*

Going down, *Bradyrhizobium, Sinorhizobium, Rhizobium* and *Agrobacterium* form another related group. The same works for *Variovorax, Ralstonia, Burkholderia* and *Paraburkholderia*. Following the same pattern, *Erwinia, Pantoea, Serratia* and *Dickeya* are closer among them. Finally, *Acinetobacter* and *Pseudomonas* are the last two groups of the tree.

The *R pipeline* we have used to perform the comparisons can be found in the *annex 2*. Abbreviations can be found in *annex 3.*

2.6.1. Anthranilate synthase vs *gyrB*

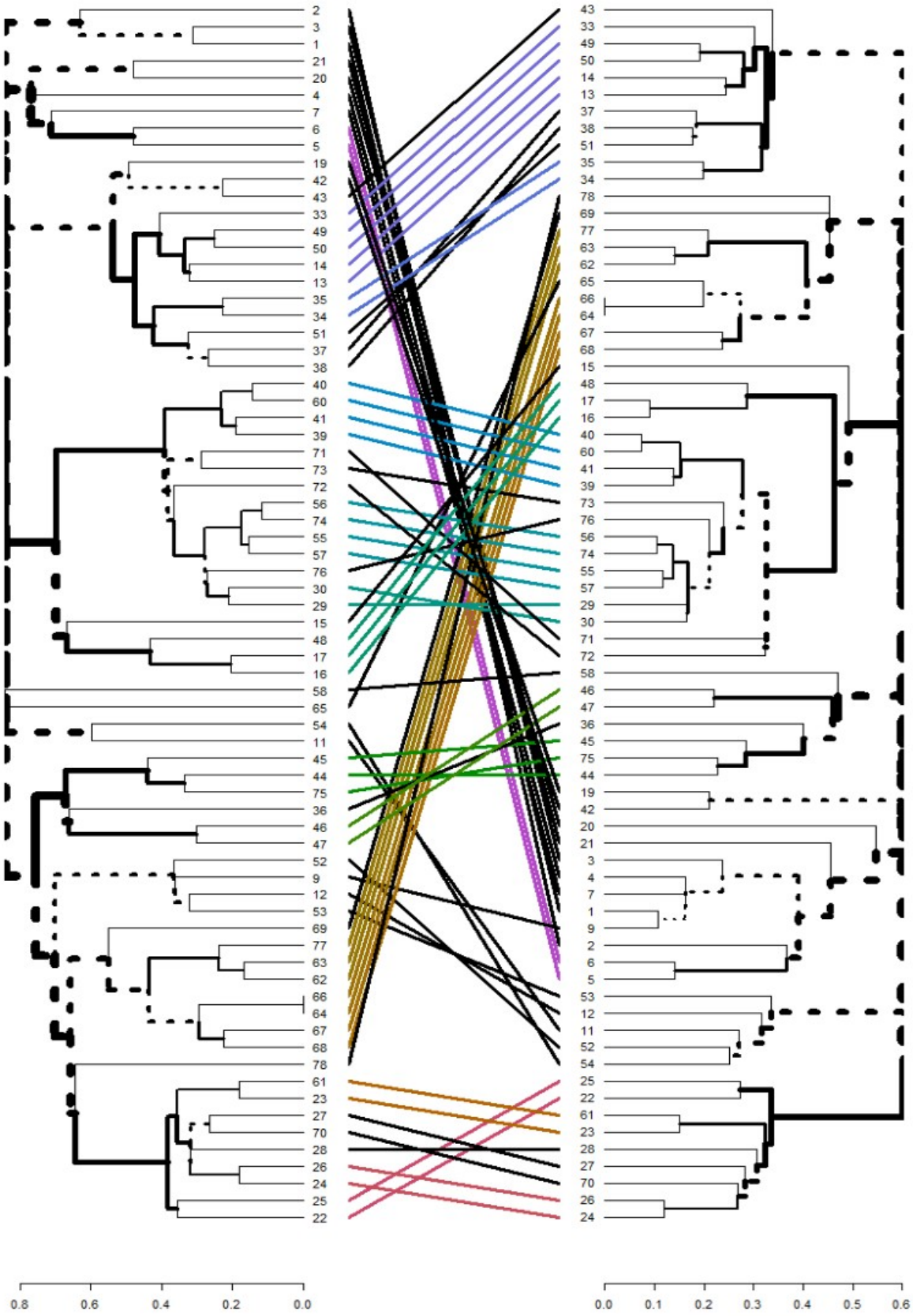


Figure 33. Anthranilate synthase vs *gyrB* comparison (*dendextend* tanglegram).

Agrobacterium (71,72)	*Pseudomonas (27, 28, 70)*
Bacillus (1,2,3,4,7,9)	*Rhizobium (76)*
Klebsiella (42,43)	*Acinetobacter (12,53)*
Serratia (37,38,51)	*Pseudarthrobacter (47)*

While color lines show that a species is in the same branch in both trees, black lines show species and genus which have different locations in both trees (figure 33). Genus and species that have different classifications depending on the tree are summarized in the previous chart.

Nevertheless, an as we can see, these differences are not huge, since the main part of them refer to a different position of these species but inside the same *clusters* in both trees.

2.6.2. Tryptophan synthase (alpha subunit) vs *gyrB*

Several differences of position are observed (figure 34), mainly referring to changes due to the topology and disposition of the tree (color).

For the changes of position inside *clusters*, we have:

Bacillus (2,5,7,8,9,1,4,3)
Serratia (38)
Kitasatospora (58)
Pseudomonas (28,22,25,27)
Agrobacterium (71,72,73)
Acinetobacter (12,52,54)
Klebsiella (42, 43)
Enterobacter (19)
Erwinia (33)

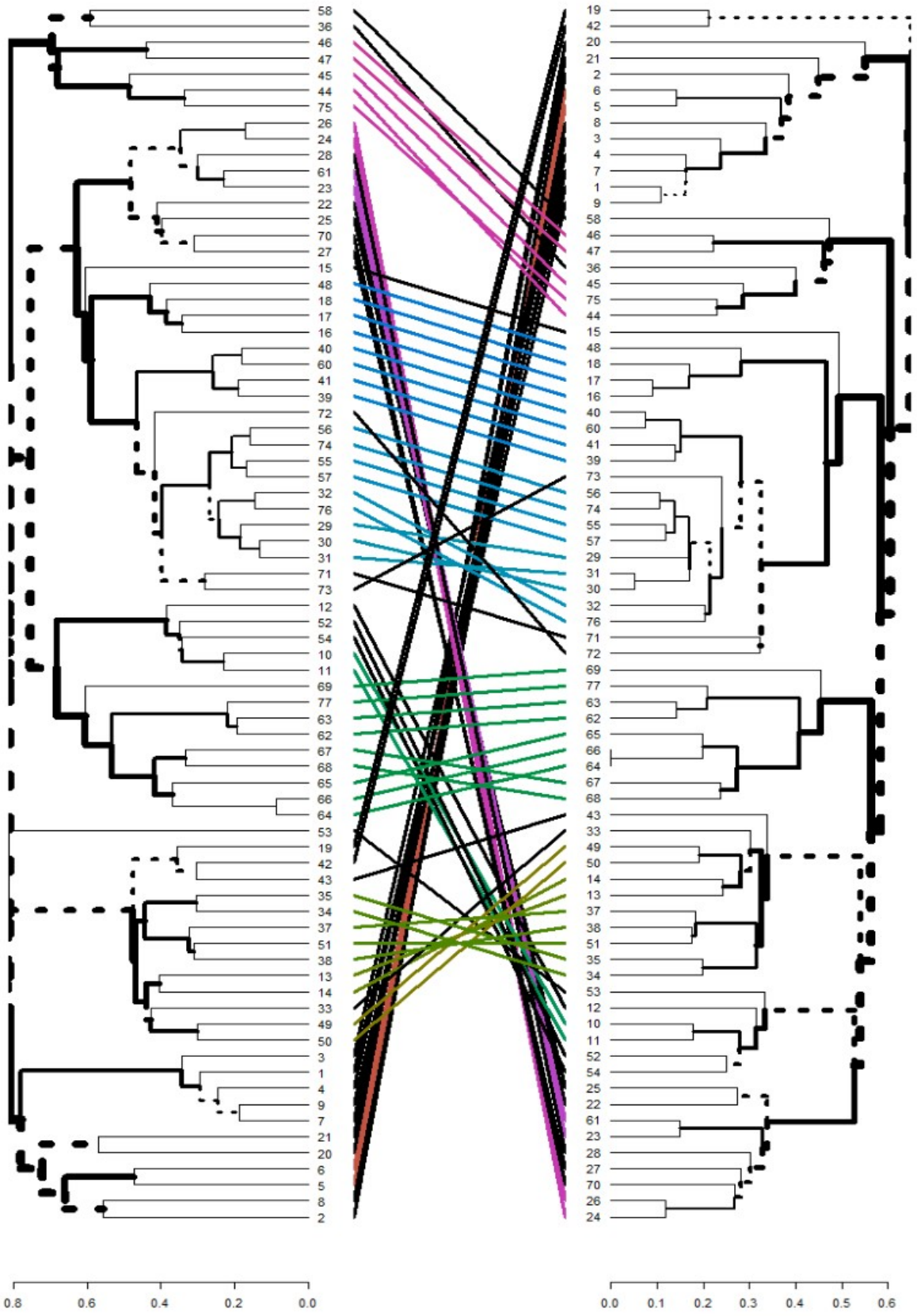


Figure 34. Tryptophan synthase (alpha subunit) vs *gyrB* comparison (*dendextend* tanglegram).

2.6.3. Tryptophan synthase (beta subunit) vs *gyrB*

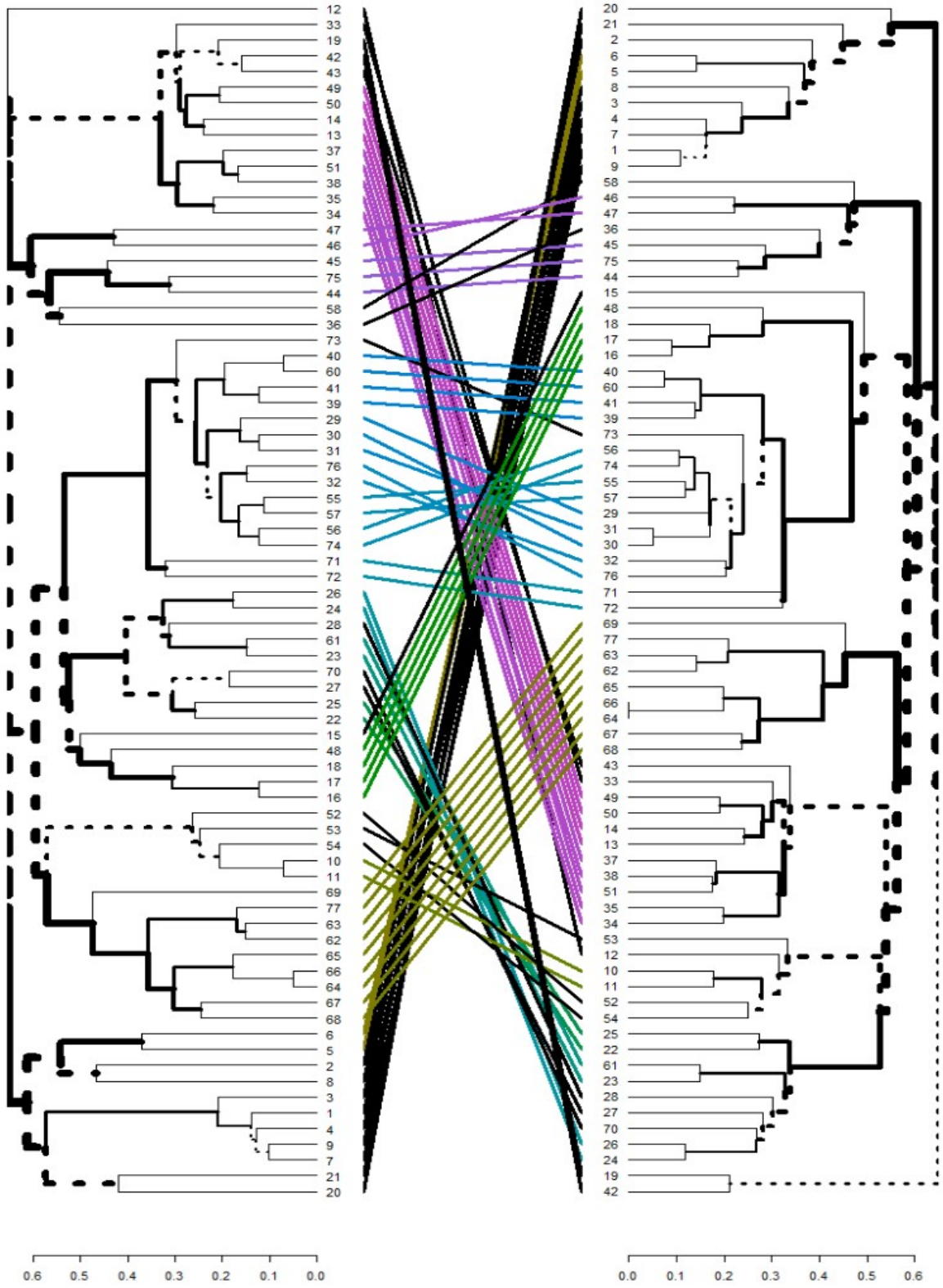


Figure 35. Tryptophan synthase (beta subunit) vs *gyrB* comparison (*dendextend* tanglegram).

Several differences of position are observed (figure 35), mainly referring to changes due to the topology and disposition of the tree (color).

For the changes of position inside *clusters*, we have:

Bacillus (2,5,7,8,9,1,4,3)	*Klebsiella (42, 43)*
Erwinia (33)	*Pseudomonas (28, 27, 70)*
Enterobacter (19)	*Azospirillum (15)*
Serratia (38)	*Acinetobacter (52, 53, 54)*
Kitasatospora (58)	*Paenibacillus (21,20)*
Agrobacterium (73)	*Klebsiella (42, 43)*

2.6.4. Tryptophan transaminase vs *gyrB*

As we can see (figure 36), there are important differences for each species. This could be expected since we have been working with different transaminase genes that are not a reflection of the evolutionary relationship among species but those genes. For having a clear tanglegram, we would need to compare each *tyrB*, *patB*, *tyr*, *tatA* and *phcC* separately. Nevertheless, this could lead us to think that the different number of genes observed could be a reflection of the different ways this group of species has acquired the capability of metabolize tryptophan.

2.6.5. Indole-3-pyruvate decarboxylase vs *gyrB*

Several differences of position are observed (figure 37), mainly referring to changes due to the topology and disposition of the tree (color).

For the changes of position inside *clusters*, we have:

Kitasatospora (58)	*Bacillus* (2,4,7,9)
Erwinia (33)	*Rhizobium (55, 29)*
Pantoea (49, 50)	*Serratia* (37, 38, 51)
Burkholderia (64)	*Sinorhizobium (41)*
Paraburkholderia (68)	*Azospirillum (15)*
Dickeya (34)	*Acetobacter (59)*
Pseudomonas (27)	

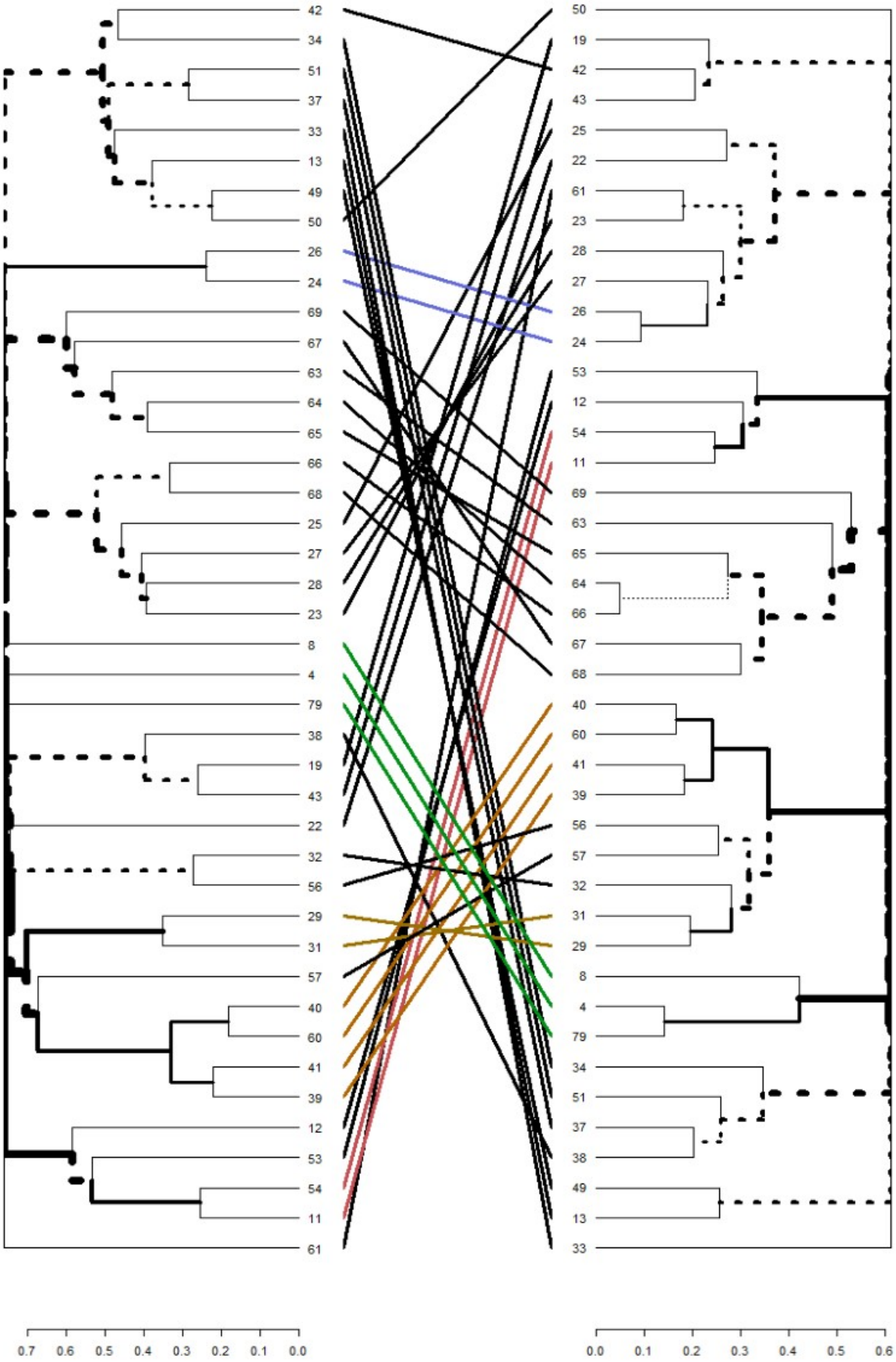


Figure 36. Tryptophan transaminase vs *gyrB* comparison (*dendextend* tanglegram).

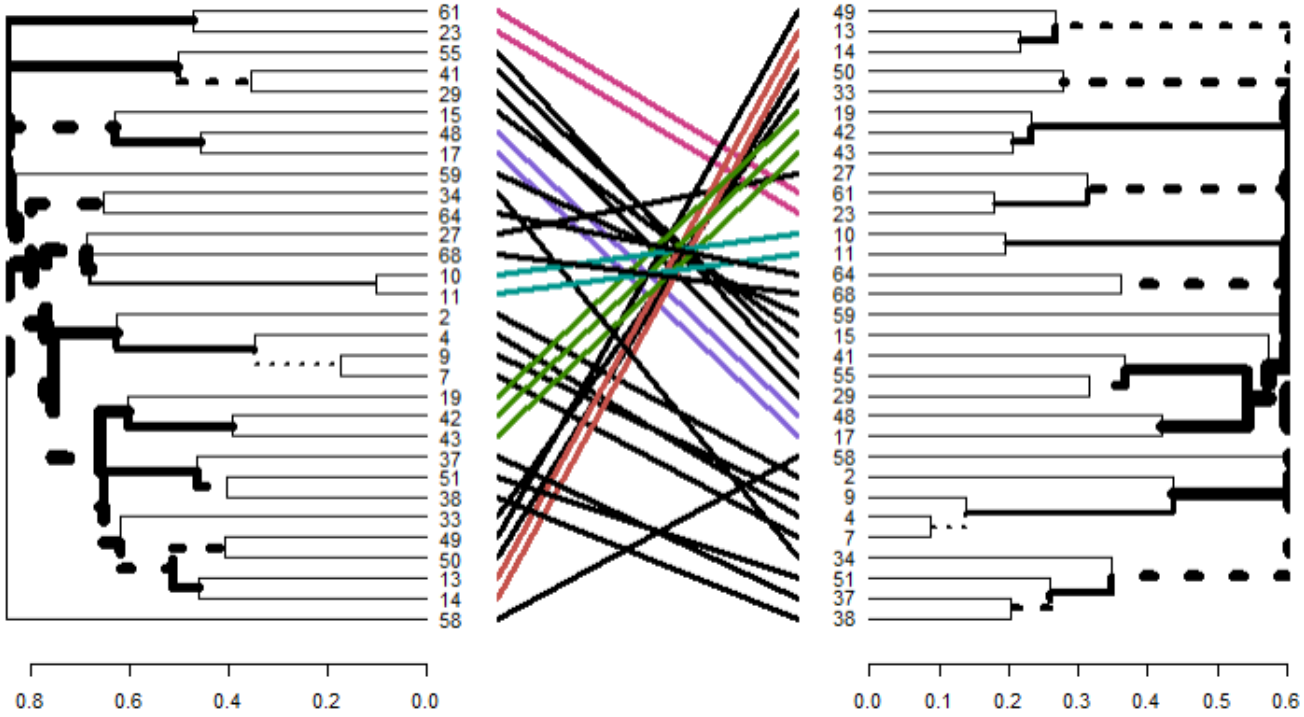


Figure 37. Indole-3-pyruvate decarboxylase vs *gyrB* comparison (*dendextend* tanglegram).

2.6.6. Indole-3-acetaldehyde dehydrogenase vs *gyrB*

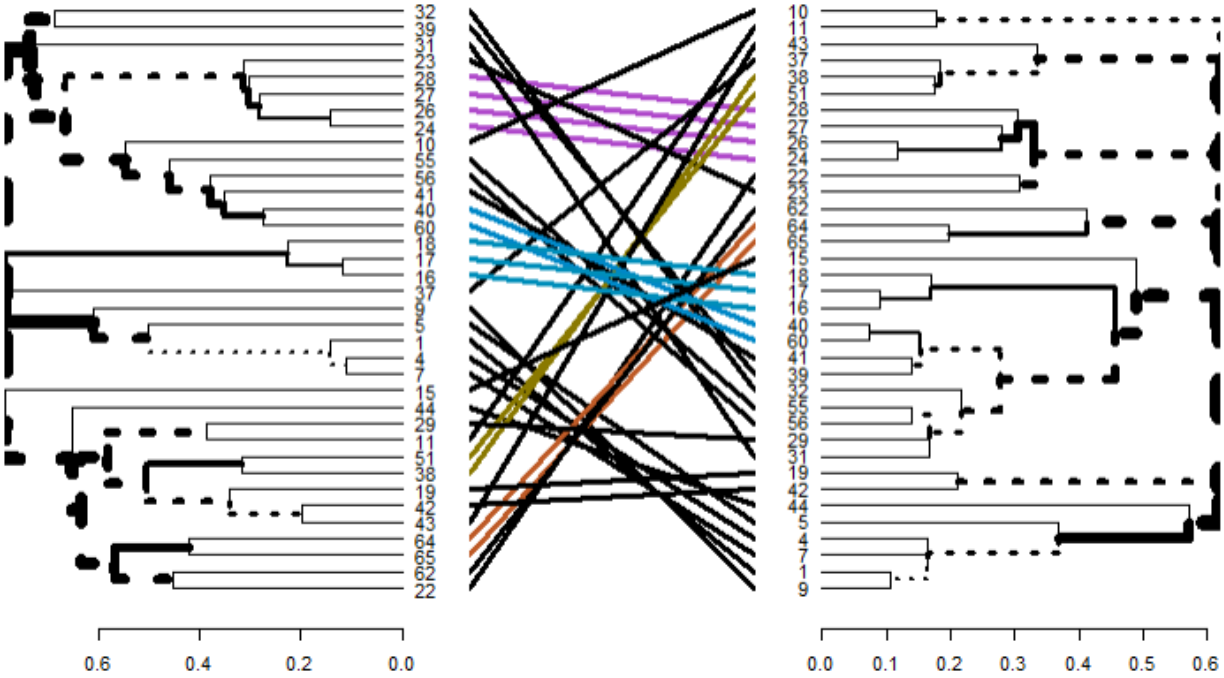


Figure 38. Indole-3-acetaldehyde dehydrogenase vs *gyrB* comparison (*dendextend* tanglegram).

Several differences of position are observed (figure 38), mainly referring to changes due to the topology and disposition of the tree (color).

For the changes of position inside *clusters*, we have:

Rhizobium (32,31,55, 56)
Sinorhizobium (39, 41)
Pseudomonas(23,22)
Acinetobacter (10)
Serratia (37)
Bacillus (9,4,1,5,7)
Klebsiella (19,42,43)
Ralstonia (62)

2.6.7. Tryptophan-2-monooxygenase vs *gyrB*

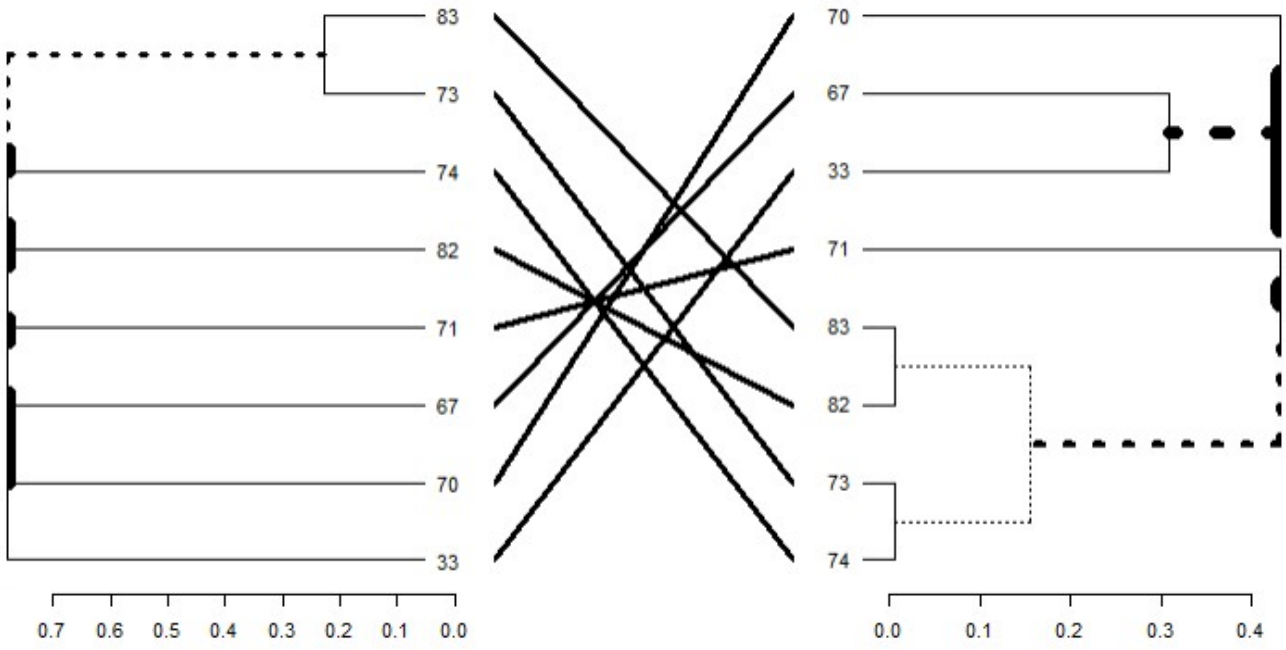


Figure 39. Tryptophan-2-monooxygenase vs *gyrB* comparison (*dendextend* tanglegram).

As we can see in the tanglegram (figure 39), there are huge differences between the pylogheny described by the *gyrB* marker and the phylogeny of the enzyme under study. Since the species that synthetize IAA through IAM pathway are closely related from an evolutionary point of view (distances are short attending to the proximity of the branches to the root), this differences could be due to the similarity of all the tryptophan-2-monooxygenase proteins under study. Besides, having only 8 species under study could have led to this result.

2.6.8. Indole-3-acetamide hydrolase vs *gyrB*

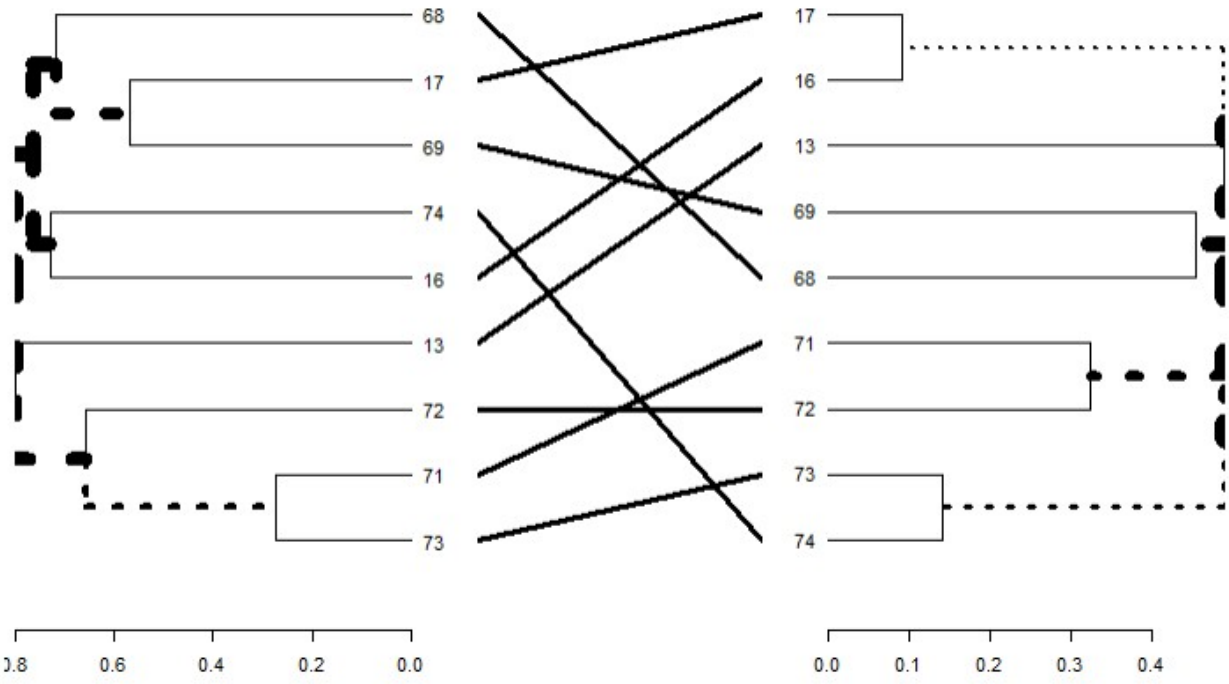


Figure 40. Indole-3-acetamide hydrolase vs *gyrB* comparison (*dendextend* tanglegram).

As in the previous point, differences are noticeable in every species under study.

2.6.9. Nitrile hydratase (alpha subunit) vs *gyrB*

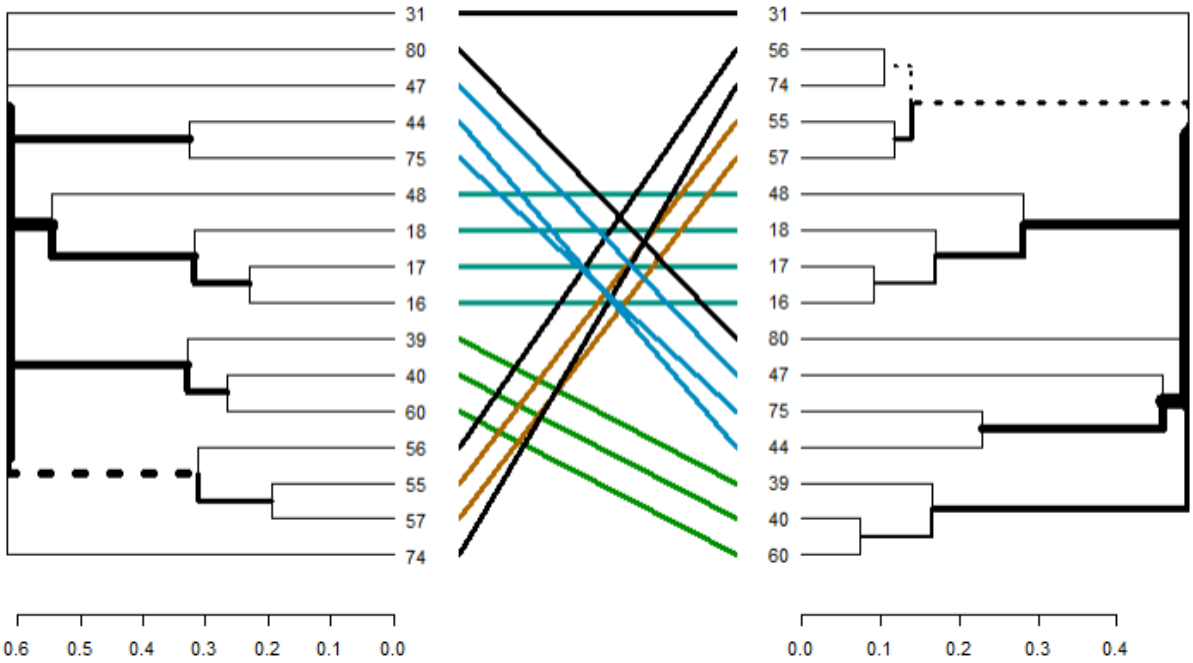


Figure 41. Nitrile hydratase (alpha subunit) vs *gyrB* comparison (*dendextend* tanglegram).

Several differences of position are observed (figure 41), mainly referring to changes due to the topology and disposition of the tree (color).

For the changes of position inside *clusters*, we have:

Rhizobium (31, 56)
Klebsiella (80)
Agrobacterium (74)

2.6.10. Nitrile hydratase (beta subunit) vs *gyrB*

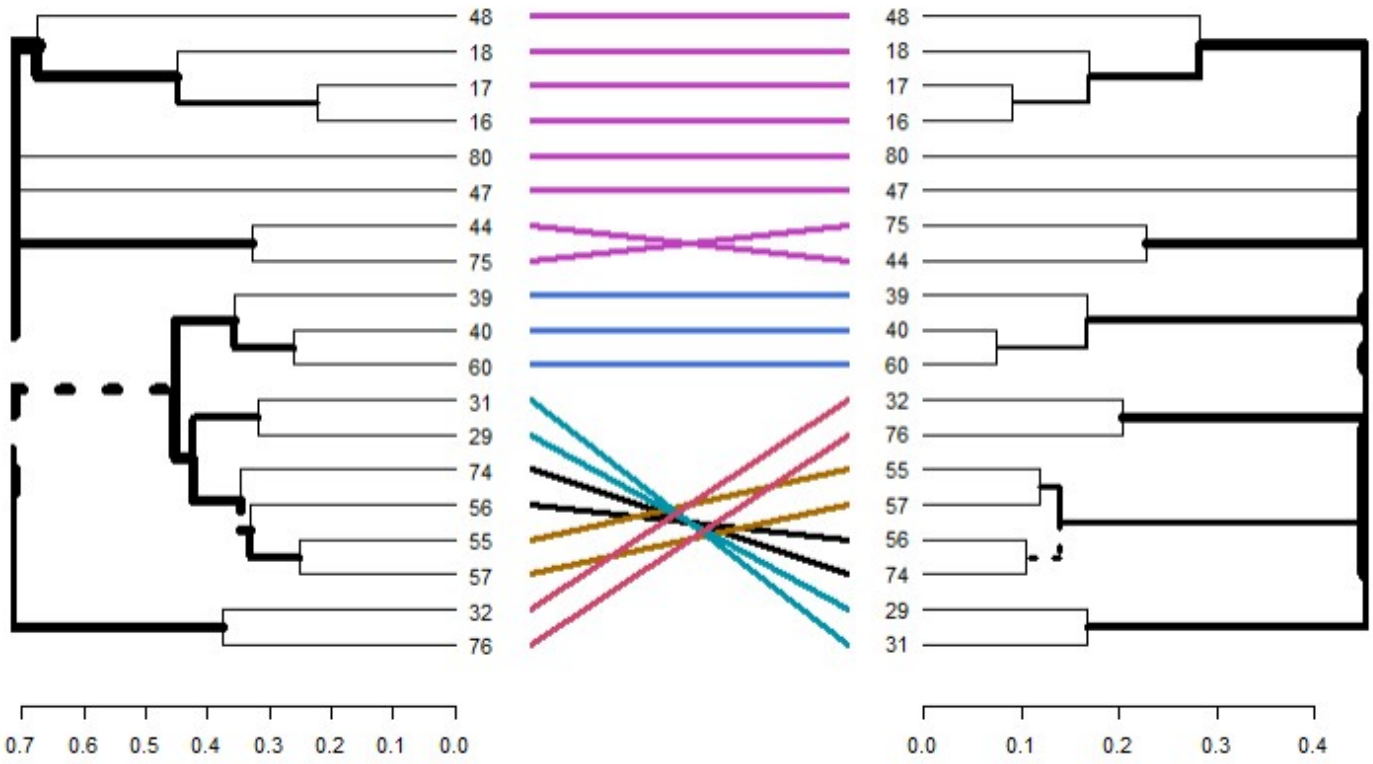


Figure 42. Nitrile hydratase (alpha subunit) vs *gyrB* comparison (*dendextend* tanglegram).

Several differences of position are observed (figure 42), mainly referring to changes due to the topology and disposition of the tree (color).

For the changes of position inside *clusters*, we have:

Rhizobium (56)
Agrobacterium (74)

2.6.11. Nitrilase vs *gyrB*

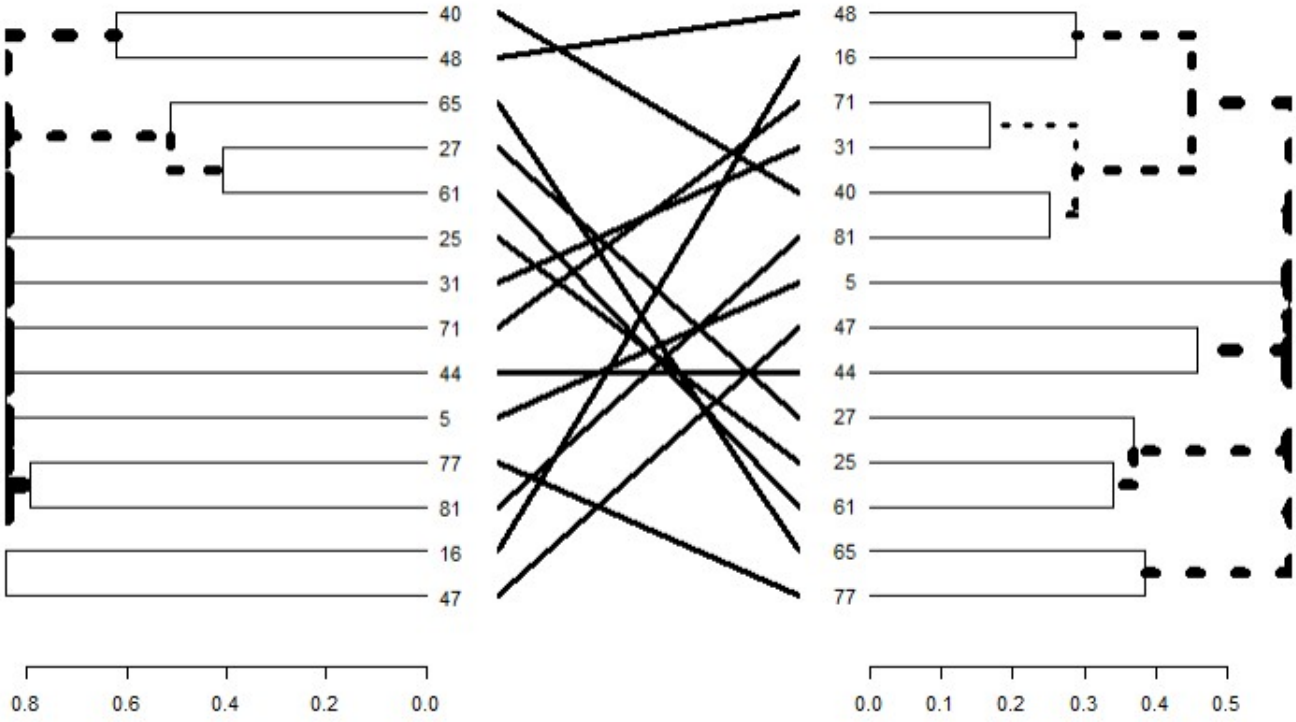


Figure 43. Nitrilase vs *gyrB* comparison (*dendextend* tanglegram).

As shown in the tanglegram (figure 43), there are huge differences between the pylogheny described by the *gyrB* marker and the phylogeny of the enzyme under study.

3.Conclusions

Concerning the research itself, we can conclude that:

- Tryptophan and IAA biosynthesis are two related pathways. Attending to this, and for every rhizobacteria able to produce IAA that we have found, there is also evidence of tryptophan production.
- The high levels of co-occurence observed for anthranilate synthase and tryptophan synthases (alpha and beta) are due to the existence of the *trp* operon.
- IPA is the main IAA pathway in rhizobacteria. In relation to this, evidence suggest that this is the first pathway that appeared in rhizobacteria.
- IAM is a marginal pathway in comparison to IPA. In relation to this, evidence suggest that this pathway appeared after IPA in several related species (*Agrobacterium, Pseudomonas, Bradyrhizobium, Pantoea, Paraburkholderia, Burkholderia)*, as phylogeny described by *gyrB* suggests.
- IAN is a marginal pathway in comparison to IPA. In relation to this, evidence suggest that this pathway appeared after IPA in several related species (*Agrobacterium, Pseudomonas, Bradyrhizobium,, Rhizobium, Klebsiella, Rhodococcus, Sinorhizobium, Burkholderia)*, as phylogeny described by *gyrB* suggests.
- IAN/IAM relation is strong, as previously suggested by the bibliographic research.
- Information about several enzymes is scarce (tryptophan-2-monooxygenase, indole-3-acetamide hydrolase), or some of the suggested enzymes are not clearly related to the pathway since they differ too much from the expected phylogeny described by *gyrB* and these differences can not be explained by evolution, since the species are very close (nitrilase). Thus, we can not consider these cases as solid evidence in order to articulate a discussion.

In relation to the project planning, we think we have accomplished our expected objectives. Even though in some cases information was scarce, we have been able to synthesize almost all the available data about the topic and analyze it, obtaining for some of the cases under study strong evidences which lead us to reaffirm several hypothesis we have formulated when starting the project.

Besides, we successfully reached every objective and deadline we set in PEC 0 and PEC 1 on scheduled time. Moreover, we have been able to introduce new perspectives, such as the operon analysis, which was incorporated as a part of the project after PEC 2 attending to the results of the co-ocurrence matrix draft.

Finally, we consider this project as a first step into further analysis in the near future, as we expect that the lack of information about some enzymes decreases in a few years.

4. Glossary

B

Binary (tree): *tree data structure in which each node has at most two children.*

Biosynthesis*: multi-step, enzyme-catalyzed process where substrates are converted into more complex products*

C

Co-occurence*: coexistence within the same species.*

Cofactor*: non-protein chemical compound or metallic ion that is required for an enzyme's activity*

E

Enzyme*: Protein that regulates a chemical reaction.*

G

Gibberellin*: plant hormone that regulates several developmental processes*

I

Indole-3-acetid acid*: Plant, growth-promoting hormone.*

L

Ligands*: substance that forms a complex with a biomolecule (for instance, an enzyme) to serve a biological purpose*

M

Marker (phylogenetic): *DNA fragment which is used in phylogenetic reconstructions, with predictable variation within a given species, and with available seqeunces for most or all species of a genus.*

O

Operon*: a unit constituted by linked genes that regulates its own expression.*

P

.***Phylogenetics****: study of the evolutionary history and relationships among individuals or groups of organisms*

R

Rhizosphere*: region of soil that is directly influenced by root secretions, and associated soil microorganisms*

Rhizobacteria*: rhizosphere associated bacteria.*

S

Secondary structure (protein)*: Three dimensional form described by the aminoacids.*

T

Tryptophan*: aminoacid, used in the biosynthesis of proteins.*

U

Ultrametric (tree)*: rooted and weighted tree with leaves at the same depth.*

5. References

[1] Vandegehuchte ML, de la Peña E, Bonte D. Relative importance of biotic and abiotic soil components to plant growth and insect herbivore population dynamics. PLoS One.2010; 5(9):e12937. doi:10.1371/journal.pone.0012937

[2] Jones DG. The Rhizosphere. Edited by J. M. Lynch. Chichester, Sussex, UK: John Wiley & Sons, (1990), pp. 458, £65.00, ISBN 0-471-925489. Experimental Agriculture. 1992;28(2):238-238. doi:10.1017/S0014479700019700

[3] Kennedy AC, de Luna LZ, in Encyclopedia of Soils in the Environment, 2005. 399-406.

[4]Berlanas C, Berbegal M, Elena G, Laidani M , Cibriain J, Sagües A, Gramaje D. The fungal and bacterial rhizosphere microbiome associated with grapevine rootstock genotypes in mature and young vineyards. Frontiers in Microbiology. 2019; 10-1142. doi:10.3389/fmicb.2019.01142.

[5] Beneduzi A, Ambrosini A, Passaglia LM. Plant growth-promoting rhizobacteria (PGPR): Their potential as antagonists and biocontrol agents. Genet Mol Biol.; 2012;35(4 (suppl)):1044–1051.doi:10.1590/s1415-47572012000600020

[6] Naveed M, Qureshi M, Zahir Z, Hussain MB, Sessitsch A, Mitter B . L-Tryptophan dependent biosynthesis of indole-3-acetic acid (IAA) improves plant growth and colonization of maize by Burkholderia phytofirmans PsJN. Annals of Microbiology. 2014; 65. doi: 10.1007/s13213-014-0976-y.

[7] Khalid M, Arshad M, Shaharoona B, Mahmood T. Plant growth promoting rhizobacteria and sustainable agriculture. In: Khan MS, Zaidi A, Musarrat J (eds) Microbial Strategies for Crop Improvement. Springer-Verlag Berlin Heidelberg, Germany. 2009;133–160. doi:10.1007/978-3-642-01979-1_7

[8] Kaushal M. Climatic Resilient Agriculture for Root, Tuber, and Banana Crops using Plant Growth-Promoting Microbes. 2019; 307-329. doi:10.1016/B978-0-12-816483-9.00012- 8.

[9] Patten C , Glick B. Bacterial biosynthesis of indole-3-acetic acid. Canadian journal of microbiology.1996;42. 207-20. doi:207-20. 10.1139/m96-032.

[10] Aeron A, Pandey P, Kumar S, Maheshwari DK. Emerging role of plant growth promoting rhizobacteria. In: Maheshwari DK (ed) Bacteria in agrobiology: crop ecosystem. Springer Verlag Berlin/Heidelberg.2011; 1–26. doi: doi.org/10.1007/978-3-642-18357-7_1

[11] Kaushal M. Rhizobacterial Efficacy for Sustainable Crop Productivity in Agroecosystems. 2017;50-61

[12] Idris EE, Iglesias DJ, Talon M, Borriss R. Tryptophan-dependent production of indole-3-acetic acid (IAA) affects level of plant growth promotion by Bacillus

amyloliquefaciens FZB42. Mol Plant- Microbe Interact. 2007;20: 619-626. doi:10.1094/MPMI-20-6-0619.

[13] Wang B, Chu J, Yu T, et al. Tryptophan-independent auxin biosynthesis contributes to early embryogenesis in Arabidopsis. Proc Natl Acad Sci USA.2015;112(15):4821–4826. doi:10.1073/pnas.1503998112

[14] Chagas A, Oliveira A, Costa J. Production of indole-3-acetic acid by bacillus isolated from different soils. Bulgarian Journal of Agricultural Science. 2015;21. 282-287.

[15] Spaepen S, Vanderleyden J. Auxin and plant-microbe interactions. Cold Spring Harb Perspect Biol. 2011;3(4):a001438. doi:10.1101/cshperspect.a001438

[16] Patten C & Glick B. Bacterial biosynthesis of indole-3-acetic acid. Canadian journal of microbiology. 1996;42. 207-20. doi:10.1139/m96-032.

[17] Ozdal M, Özlem O, Sezen A, Algur O. Biosynthesis Of Indole-3-Acetic Acid By Bacillus cereus Immobilized Cells. Cumhuriyet Science Journal. 2016;37. 212. doi:10.17776/csj.34085.

[18] Wagi S, Ahmed A. Bacillus spp.: potent microfactories of bacterial IAA. PeerJ. 2019;7:e7258. doi:10.7717/peerj.7258

[19] Ghazal M, Moussa L, Fayed S, Mostfa A. The use of bacillus mycoides and bacillus subtlis as biocontrol agents for the fungi causing root rot disease in common bean . Egypt J. Biotechnol. 2013;44. 37-43. doi:10.13140/RG.2.1.2674.5841.

[20] Shao J, Li S, Zhang N. Analysis and cloning of the synthetic pathway of the phytohormone indole-3-acetic acid in the plant-beneficial Bacillus amyloliquefaciens SQR9. Microb Cell Fact. 2015;14:130. doi:10.1186/s12934-015-0323-4.

[21] Yahaghi Z, Shirvani M, Nourbakhsh F, Coba T, Pueyo J, Talebi M. Isolation and Characterization of Pb-Solubilizing Bacteria and Their Effects on Pb Uptake by Brassica juncea: Implications for Microbe-Assisted Phytoremediation. Journal of microbiology and biotechnology. 2018;28. doi:10.4014/jmb.1712.12038.

[22] Lee JC, Whang KS. Optimization of Indole-3-acetic Acid (IAA) Production by Bacillus megaterium BM5. Korean Journal of Soil Science and Fertilizer. 2016;49: 461-468. doi:10.7745/KJSSF.2016.49.5.461.

[23] Lo KJ, Lin SS, Lu CW, Kuo CH, Liu CT. Whole-genome sequencing and comparative analysis of two plant-associated strains of Rhodopseudomonas palustris (PS3 and YSC3).Sci Rep. 2018;8(1):12769. doi:10.1038/s41598- 018-31128-8

[24] Blin P. Dynamique des interactions entre Dickeya dianthicola, Dickeya solani et leur hôte Solanum tuberosum. 2017.

[25] Zhou J, Cheng Y, Lv M. The complete genome sequence of Dickeya zeae EC1 reveals substantial divergence from other Dickeya strains and species. BMC Genomics.2015;16(1):571. doi:10.1186/s12864-015-1545-x

[26] Coutinho TA, Venter SN. Pantoea ananatis: an unconventional plant pathogen. Mol Plant Pathol. 2009;10(3):325–335. doi:10.1111/j.1364-3703.2009.00542.x

[27] Brady CL, Cleenwerck I, Westhuizen L, Venter SN, Coutinho TA, Vos PD. Pantoea rodasii sp. nov., Pantoea rwandensis sp. nov. and Pantoea wallisii sp. nov., isolated from eucalyptus. Int J Syst Evol Micr.2012;62: 1457–1464. doi: 10.1099/ijs.0.032615-0.

[28] Estenson K, Hurst G, Standaert R, Bible A, Garcia D, Chourey K, & Doktycz, Mitchel & Morrell-Falvey, Jennifer. Characterization of Indole-3-acetic Acid Biosynthesis and the Effects of This Phytohormone on the Proteome of the Plant-Associated Microbe Pantoea sp. YR343. Journal of Proteome Research.2018;17. doi:10.1021/acs.jproteome.7b00708.

[29] Kulkarni G, Nayak A, Sajjan S, Oblesha A, Karegoudar, TB. Indole-3-acetic acid biosynthetic pathway and aromatic amino acid aminotransferase activities in Pantoea dispersa strain GPK. Letters in applied microbiology.2013;56. doi:10.1111/lam.12053.

[30] Apine OA, Jadhav JP. Optimization of medium for indole-3-acetic acid production using Pantoea agglomerans strain PVM. J Appl Microbiol.2011;110:1235–1244.doi: 10.1111/j.1365-2672.2011.04976.x

[31] Ouyang LM, Pei H, Xu Z. Low nitrogen stress stimulating the indole-3-acetic acid biosynthesis of Serratia sp. ZM is vital for the survival of the bacterium and its plant growth-promoting characteristic. Archives of microbiology.2016;199. doi:10.1007/s00203-016-1312-7

[32] Zelaya-Molina L, Hernandez-Soto LM, Guerra-Camacho J, Monterrubio R , Patiño-Siciliano A, Villa-Tanaca L, Hernández-Rodríguez C. Ammonia-Oligotrophic and Diazotrophic Heavy Metal-Resistant Serratia liquefaciens Strains from Pioneer Plants and Mine Tailings. Microbial Ecology.2016;72. doi:10.1007/s00248-016- 0771-3.

[33] Hasuty A, Choliq A. Production of Indole Acetic Acid (IAA) by serratia marcescens subsp. marcescens and rhodococcus aff. qingshengii. International Journal of Agricultural Technology.2018;14. 299-312.

[34] Khan AR, Park GS, Asaf S, Hong SJ, Jung BK, Shin JH. Complete genome analysis of Serratia marcescens RSC-14: A plant growth-promoting bacterium that alleviates cadmium stress in host plants. PLoS One.2017;12(2):e0171534. doi:10.1371/ journal.pone.0171534

[35] Jung BK, Khan AR, Hong SJ, Park GS, Park YJ, Park CE, Jeon HJ, Lee SE, Shin JH. Genomic and phenotypic analyses of Serratia fonticola strain GS2: a rhizobacterium isolated from sesame rhizosphere that promotes plant growth and

produces N-acyl homoserine lactone. Journal of Biotechnology.2017;241:158-162. doi: 10.1016/j.jbiotec.2016.12.002.

[36] Lin HR, Shu HY, Lin GH. Biological roles of indole-3-acetic acid in Acinetobacter baumannii. Microbiological Research.2018;216. 10.1016/j.micres.2018.08.004.

[37] Lin G, Chen H, Huang J, Liu T, Lin T, Wang S, Tseng C, Shu H. Identification and characterization of an indigoproducing oxygenase involved in indole 3-acetic acid utilization by Acinetobacter baumannii. Antonie Van Leeuwenhoek.2012;101:881–890. doi:doi: 10.1007/s10482-012-9704-4

[38] Sachdev D, Nema P, Dhakephalkar P, Zinjarde S, Chopade B. Assessment of 16S rRNA gene-based phylogenetic diversity and promising plant growth-promoting traits of Acinetobacter community from the rhizosphere of wheat. Microbiological research.2010;165. 627-38. doi:10.1016/j.micres.2009.12.002.

[39] Huddedar S, Shete A, Tilekar J, Gore S, Dhavale D, Chopade B. Isolation, Characterization, and Plasmid pUPI126-Mediated Indole-3-Acetic Acid Production in Acinetobacter Strains from Rhizosphere of Wheat. Applied biochemistry and biotechnology.2002;102-103. 21-39. doi:10.1385/ABAB:102-103:1-6:021.

[40] Datta C, Basu P. Indole acetic acid production by a Rhizobium species from root nodules of a leguminous shrub, Cajanus cajan. Microbiological research.2000;155. 123-7. doi:10.1016/S0944-5013(00)80047-6.

[41] Bhattacharyya R, Pati B. Growth behaviour and indole acetic acid (IAA) production by a Rhizobium isolated from root nodules of Alysicarpus vaginalis DC. Acta microbiologica et immunologica Hungarica.2000;47. 41-51.

[42] Kobayashi M, Suzuki T, Fujita T, Masuda M, Shimizu S. Occurrence of enzymes involved in biosynthesis of indole-3-acetic acid from indole-3-acetonitrile in plantassociated bacteria, Agrobacterium and Rhizobium. Proc Natl Acad Sci USA.1195; 92:714–718. doi:10.1073/pnas.92.3.714

[43] Spaepen S, Das F, Luyten E, Michiels J, Vanderleyden J. Indole-3-acetic acid-regulated genes in Rhizobium etli CNPAF512. FEMS microbiology letters.2009;291. 195-200. doi:10.1111/j.1574-6968.2008.01453.x.

[44] Ghosh P, Saha P, Mayilraj S, Maiti T. Role of IAA metabolizing enzymes on production of IAA in root, nodule of Cajanus cajan and its PGP Rhizobium sp. Biocatalysis and Agricultural Biotechnology.2013;2. 234-239. doi: 10.1016/j.bcab.2013.04.002

[45] Bhattacharjee R, Jourand P, Chaintreuil C, Dreyfus B, Singh A, Mukhopadhyay S. Indole acetic acid and ACC deaminase-producing Rhizobium leguminosarum bv. trifolii SN10 promote rice growth, and in the process undergo colonization and chemotaxis. Biology and Fertility of Soils.2011;48. doi:10.1007/s00374-011-0614-9.

[46] Dubey RC, Maheshwari D, Kumar H, Choure K. Assessment of diversity and plant growth promoting attributes of rhizobia isolated from Cajanus cajan L. African Journal of Biotechnology.2011;9. 8619-8629. doi:10.5897/AJB10.1127

[47] Cruz-González X, Laza-Pérez N, Mateos PF, Rivas R. Analysis and effect of the use of biofertilizers on Trifolium rubens L., a preferential attention species in Castile and Leon, Spain, with the aim of increasing the plants conservation status. AIMS Microbiol.2017;3(4) 733-746.doi:10.3934/microbiol.2017.4.733

[48] Del Papa F, Schlueter A, Althabegoiti M, Rogel M, Martínez-Romero E, Nilsson JF, Ormeño-Orrillo E, Pistorio M, Pühler A, Niehaus K, Tejerizo G, Lagares A. Rhizobium favelukesii sp. nov., isolated from the root nodules of alfalfa (Medicago sativa L). International Journal of Systematic and Evolutionary Microbiology.2016;66. doi:10.1099/ijsem.0.001373.

[49] Eddie LI, Alves de Paiva Rolla dos Santos A, Luiz Martinez de Oliveira A, Hungria M, Pains Rodrigues E. Indole-3-acetic acid production via the indole-3-pyruvate pathway by plant growth promoter Rhizobium tropici CIAT 899 is strongly inhibited by ammonium, Research in Microbiology2017;168(3), 283-292. doi:10.1016/j.resmic.2016.10.010.

[50] El-Khawas H, Adachi K. Identification and quantification of auxins in culture media of Azospirillum and Klebsiella and their effect on rice roots. Biol Fertility Soils.1999;28:377–38. doi:10.1007/s003740050507

[51] Akbari GA, Arab SM, Alikhani HA, Allahdadi I, Arzanesh MH. Isolation and selection of indigenous Azospirillum spp. and the IAA of superior strains effects on wheat roots. World J Agric Sci.2007;3:523–529.

[52] Baca BE, Soto-Urzua L, Xochihua-Corona YG, Cuervo-Garcia A. Characterization of two aromatic amino acidaminotransferases and production of indoleacetic acid in Azospirillum strains. Soil Biol Biochem.1194;26:57–63. doi:10.1016/0038-0717(94)90195-3

[53] Molina R, Rivera D, Mora V, Lopez G, Rosas S, Spaepen S,Vanderleyden J, Cassan F. Regulation of IAA Biosynthesis in Azospirillum brasilense Under Environmental Stress Conditions. Current Microbiology.2018;75. doi:10.1007/s00284-018- 1537-6.

[54] Bose A, Kher M, Nataraj M, Keharia H. Phytostimulatory effect of indole-3-acetic acid by Enterobacter cloacae SN19 isolated from Teramnus labialis (L. f.) Spreng rhizosphere. Biocatalysis and Agricultural Biotechnology.2016;6. doi:10.1016/j.bcab.2016.03.005.

[55] Brandl MT, Lindow SE. Cloning and characterization of a locus encoding an indolepyruvate decarboxylase involved in indole-3-acetic acid synthesis in Erwinia herbicola. Appl Environ Microbiol.1996;62:4121–4128.

[56] Clark E, Manulis S, Ophir Y, Barash I, Gafni Y. Cloning and characterization of iaaM and iaaH from Erwinia herbicola pathovar gypsophilae. Phytopathology.1993;83:234– 240.

[57] Yang S, Zhang Q, Guo JH, Charkowski A, Glick B, Ibekwe A, Cooksey D, Yang CH. Global Effect of Indole-3-Acetic Acid Biosynthesis on Multiple Virulence Factors of Erwinia chrysanthemi 3937. Applied and environmental microbiology.2007;73. 1079-88. doi:10.1128/AEM.01770-06.

[58] Lü Z, Song W. Research of indole-3-acetic acid biosynthetic pathway of Klebsiella oxytoca SG-11 by HPLC and GC-MS. Se pu = Chinese journal of chromatography / Zhongguo hua xue hui. 2000;18:328-31.

[59] Celloto V, Oliveira A, Gonçalves J, Watanabe C, Matioli G, Gonçalves R. Biosynthesis of Indole-3-Acetic Acid by New Klebsiella oxytoca Free and Immobilized Cells on Inorganic Matrices. TheScientificWorldJournal. 2012. doi: 10.1100/2012/495970. doi:10.1100/2012/495970.

[60] Sachdev D, Chaudhari H, Kasture V, Dhavale D, Chopade B. Isolation and characterization of indole acetic acid producing Klebsiella pneumoniae strains from the rhizosphere of wheat (Triticum aestivum) and their effect on plant growth. Indian J. Exp. Biol.2009;47: 993-1000.

[61] Mitra S Pramanik K, & Ghosh P, Soren T, Sarkar A, Dey R, Pandey S, Maiti T. Characterization of Cd-resistant Klebsiella michiganensis MCC3089 and its potential for rice seedling growth promotion under Cd stress. Microbiological Research.2018;210. 12-25. doi:10.1016/j.micres.2018.03.003.

[62] Shrivastava S, D'Souza S, Desai P. Production of indole-3-acetic acid by immobilized actinomycetes (Kitasatospora Sp.) for soil application. Curr. Sci.2008;94.

[63] Bastian F, Cohen A, Piccoli P, Luna M, Bottini R, Baraldi R. Production of indole-3-acetic acid and gibberellins A_1 and A_3 by Acetobacter diazotrophicus and Herbaspirillum seropedicae in chemically-defined culture media. Plant Growth Regulation.1998;24:7–11. doi: 10.1023/a:1005964031159

[64] Vinardell J. The Sinorhizobium fredii HH103 Genome: A Comparative Analysis With S. fredii Strains Differing in Their Symbiotic Behavior With Soybean. Molecular plant-microbe interaction.2015;28(7):811-24. doi:10.1094/MPMI-12-14-0397-FI

[65] Imperlini E, Bianco C, Lonardo E, Camerini S, Cermola M, Moschetti G, Defez R. Effects of indole-3-acetic acid on Sinorhizobium meliloti survival and on symbiotic nitrogen fixation and stem dry weight production. Applied microbiology and biotechnology.2009;83. 727-38. doi:10.1007/s00253-009-1974-z.

[66] Kallala N, M'sehli W, Jelali K, Kais Z, Mhadhbi H. Inoculation with Efficient Nitrogen Fixing and Indoleacetic Acid Producing Bacterial Microsymbiont Enhance Tolerance of the Model Legume Medicago truncatula to Iron Deficiency. BioMed Research International. 2018. 1-14. doi:10.1155/2018/9134716.

[67] Ahmad F, Ahmad I, Khan MS. Indole acetic acid production by the indigenous isolates of Azotobacter and fluorescent Pseudomonas in the presence and absence of tryptophan. Turk J Biol.2005;29:29–34.

[68] Balaji N, Lavanya SS, Muthamizhselvi S, Tamilarasan K. Optimization of fermentation conditions for indole acetic acid production by Pseudomonas species. Int J Adv Biotechnol Res.2012;3:797–803.

[69] Suzuki S, He Y, Oyaizu H. Indole-3-acetic acid production in Pseudomonas fluorescens HP72 and its association with suppression of creeping bentgrass brown patch. Curr Microbiol.2003;47:138–143. doi:10.1007/s00284-002-3968-2

[70] Ansari F, Ahmad I. (2018). Biofilm Development, Plant Growth Promoting Traits and Rhizosphere Colonization by Pseudomonas entomophila FAP1: A Promising PGPR. Advances in Microbiology. 08. 235-251. doi:10.4236/aim.2018.83016.

[71] Marathe R, Phatake YB , Shaikh A, Shinde BP, Gajbhiye M. Effect of IAA produced by Pseudomonas aeruginosa 6a (bc4) on seed germination and plant growth of Glycin max. Journal of Experimental Biology and Agricultural Sciences.2017;5. 351-358. doi:10.18006/2017.5(3).351.358. doi:10.18006/2017.5(3).351.358

[72] Patten CL, Glick BR. Role of Pseudomonas putida indoleacetic acid in development of the host plant root system. Appl Environ Microbiol.2002;68(8):3795–3801. doi:10.1128/aem.68.8.3795-3801.2002

[73] Tsavkelova EA, Cherdyntseva TA, Botina SG, Netrusov AI.Bacteria associated with orchid roots and microbial production of auxin, Microbiological Research.2007;162(1),69-76. doi:10.1016/ j.micres.2006.07.014.

[74] Surico G, Iacobellis NS, Sisto A. Studies on the role of indole-3-acetic acid and cytokinins in the formation of knots on olive and oleander plants by Pseudomonas syringae pv. savastanoi. Physiol Plant Pathol.1985;26:309–320. Doi:10.1016/0048-4059(85)90006-2.

[75] Flores O, Prince C, Nuñez M, Vallejos A, Mardones C, Yáñez C, Besoain X, Bastías R. Genetic and Phenotypic Characterization of Indole-Producing Isolates of Pseudomonas syringae pv. actinidiae Obtained From Chilean Kiwifruit Orchards. Frontiers in Microbiology.2018;9. doi:10.3389/fmicb.2018.01907.

[76] Andreolli M, Zapparoli G, Angelini E, Lucchetta G, Lampis S, Vallini G. Pseudomonas protegens MP12: A plant growth-promoting endophytic bacterium with broad-spectrum antifungal activity against grapevine phytopathogens. Microbiological Research.2018;219. doi:10.1016/j.micres.2018.11.003.

[77] Sekine M, Ichikawa T, Kuga N, Kobayashi M, Sakurai A, Syono K. Detection of the IAA biosynthetic pathway from tryptophan via indole-3-acetamide in Bradyrhizobium spp. Plant Cell Physiol.1988;29:867–874. doi:10.1093/oxfordjournals.pcp.a077574.

[78] Egebo L, Nielsen S, Jochimsen B. Oxygen-dependent catabolism of indole-3-acetic acid in Bradyrhizobium japonicum. J Bacteriol.1991;173:4897–4901. doi:10.1128/jb.173.15.4897-4901.1991

[79] Jensen JB, Egsgaard H, Van Onckelen H, Jochimsen BU. Catabolism of indole-3-acetic acid and 4- and 5-chloroindole-3-acetic acid in Bradyrhizobium japonicum. J Bacteriol. 1995;177(20):5762–5766. doi:10.1128/jb.177.20.5762-5766.1995

[80] Siqueira AF, Ormeño-Orrillo E, Souza RC, et al. Comparative genomics of Bradyrhizobium japonicum CPAC 15 and Bradyrhizobium diazoefficiens CPAC 7: elite model strains for understanding symbiotic performance with soybean.BMC Genomics. 2014;15(1):420. doi:10.1186/1471-2164-15-420

[81] Yagi K, Matsumoto T, Chujo T, Nojiri H, Omori T, Minamisawa K, Nishiyama M, Yamane H. Isolation and Characterization of Low-indole-3-acetic Acid-producing Mutants from Bradyrhizobium elkanii. Bioscience, biotechnology, and biochemistry.2000;64. 1359-64. doi:10.1271/bbb.64.1359.

[82] Kailasan N, Vamanrao VB. Isolation and Characterization of Ralstonia Pickettii - A Novel Phosphate Solubilizing Bacterium from Pomegranate Rhizosphere from Western India, International Journal of Life Sciences Biotechnology and Pharma Research.2015;4(1),1-9.

[83] Paul A, Datta B. Cadmium-resistant Ralstonia mannitolilytica Relieved Cadmium Toxicity in Mustard Plant Through Root Colonization and Growth Promoting Activity. Research Journal of Microbiology.2016;11. 157-168. doi:10.3923/jm.2016.157.168.

[84] Kurosawa N, Hirata T, Suzuki H. Characterization of putative tryptophan monooxygenase from Ralstonia solanasearum. J Biochem.2009;146:23–32. doi: 10.1093/jb/mvp040.

[85] Duca D, Lorv J, Patten C, Rose D, Glick B. Indole-3-acetic acid in plant-microbe interactions. Antonie van Leeuwenhoek.2014;106. doi:10.1007/s10482-013-0095-y.

[86] Zuñiga A, Poupin MJ, Donoso R, Ledger T, Guiliani N, Gutierrez RA, Gonzalez B. Quorum sensing and indole-3-acetic acid degradation play a role in colonization and plant growth promotion of Arabidopsis thaliana by Burkholderia phytofirmans PsJN. Mol Plant-Microbe Interact.2013;26:546–553. doi:10.1094/MPMI-10-12-0241-R.

[87] Johan HJ, Gerards SL. Discovery of a bacterial gene cluster for catabolism of the plant hormone indole 3-acetic acid, FEMS Microbiology Ecology.2008;65(2):238–250. doi: 10.1111/j.1574-6941.2008.00436.x.

[88] Castanheira N, Dourado AC, Kruz S, Alves PIL, Delgado-Rodríguez AI, Pais I, Semedo JN, Campos P, Sánchez C, Borges N, Carvalho G, Crespo MT, Fareleira P. Plant growth promoting Burkholderia species isolated from annual ryegrass in Portuguese soils. Journal of Applied Microbiology.2015;120. doi:10.1111/jam.13025.

[89] Donoso R, Leiva-Novoa P, Zúñiga A, Timmermann T, Recabarren G, Gonzalez B. Biochemical and Genetic Bases of Indole-3-Acetic Acid (Auxin Phytohormone)

Degradation by the Plant-Growth-Promoting Rhizobacterium Paraburkholderia phytofirmans PsJN. Applied and Environmental Microbiology.2016;83. doi:10.1128/AEM.01991-16.

[90] Mannaa M, Park I, Seo YS. Genomic Features and Insights into the Taxonomy, Virulence, and Benevolence of Plant-Associated Burkholderia Species. Int J Mol Sci. 2018;20(1):121. doi:10.3390/ijms20010121.

[91] Jiang F, Chen L, Belimov AA, et al. (2012). Multiple impacts of the plant growth-promoting rhizobacterium Variovorax paradoxus 5C-2 on nutrient and ABA relations of Pisum sativum. J Exp Bot.63(18):6421–6430. doi:10.1093/jxb/ers301.

[92] Asano Y, Fujishiro K, Tani Y, Yamada H. Aliphatic nitrile hydratase from Arthrobacter sp. J-1 purification and characterization. Agric Biol Chem.1982;46:1165–1174. doi:10.1271/bbb1961.46.1165.

[93] Cai W, Su E, Zhu S, Ren Y, Wei D. Characterization of a nitrilase from Arthrobacter aurescens CYC705 for synthesis of iminodiacetic acid. The Journal of General and Applied Microbiology. 2014;60(6):207–14. doi:10.2323/jgam.60.207.

[94] Mano Y, Nemoto K. The pathway of auxin biosynthesis in plants. Journal of experimental botany.2012;63:2853-72. doi:10.1093/jxb/ers091.

[95] Kutáèek M, Rovenská J. Auxin synthesis in Agrobacterium tumefaciens and A. tumefaciens-transformed plant tissue. Plant Growth Regul. 1991;10:313–327 doi:10.1007/BF00024591.

[96] Schaerer S, Pilet PE. Quantification of indole-3-acetic acid in untransformed and Agrobacterium rhizogenes-transformed pea roots using gas chromatography mass spectrometry. Planta.1993;189:55-59. doi:10.1007/BF00201343.

[97] Kagan J, Sharon I Beja O, Kuhn J. The tryptophan pathway genes of the Sargasso Sea metagenome: New operon structures and the prevalence of non-operon organization. Genome biology.2008;9. doi:10.1186/gb2008-9-1-r20.

[98] Merino E, Jensen RA, Yanofsky C. Evolution of bacterial trp operons and their regulation. Curr Opin Microbiol.2008;11(2):78–86. doi:10.1016/j.mib.2008.02.005.

[99] Martens M, Dawyndt P, Coopman R, Gillis M, De Vos P, Willems A. Advantages of multilocus sequence analysis for taxonomic studies: A case study using 10 housekeeping genes in the genus Ensifer (including former Sinorhizobium). International journal of systematic and evolutionary microbiology.2008;58. 200-14. doi:10.1099/ijs.0.65392-0.

[100] Roberts R, Jackson RW, Mauchline TH. Is there sufficient Ensifer and Rhizobium species diversity in UK farmland soils to support red clover (Trifolium pratense), white clover (T. repens), lucerne (Medicago sativa) and black medic (M. lupulina) Appl Soil Ecol.2017;120:35–43. doi:10.1016/j.apsoil.2017.06.030.

[101] Puławska J, Kałużna M.. Phylogenetic relationship and genetic diversity of Agrobacterium spp. isolated in Poland based on gyrB gene sequence analysis and RAPD. European Journal of Plant Pathology.2011;133. doi:10.1007/s10658-011-9911-2.

[102] Duc M, Satomi M, Agata N, Venkateswaran K. GyrB as a phylogenetic discriminator for members of the Bacillus anthracis-cereus-thuringiensis group. Journal of microbiological methods. 2004; 56:383-94. doi: 10.1016/j.mimet.2003.11.004.

[103] Yamamoto S, Harayama S. PCR Amplification and Direct Sequencing of gyrB Genes with Universal Primers and Their Application to the Detection and Taxonomic Analysis of Pseudomonas putida Strains. Applied and environmental microbiology.1995;61(10):3768. doi:10.1128/AEM.61.3.1104-1109.1995.

[104] Tabacchioni S, Ferri L, Manno G, Mentasti M, Cocchi P, Campana S, Ravenni N, Taccetti G, Dalmastri C, Chiarini L, Bevivino A, Fani R. Use of the gyrB gene to discriminate among species of the Burkholderia cepacia complex. FEMS microbiology letters. 2018;281:175-82. doi:10.1111/j.1574-6968.2008.01105.x.

[105] Yamamoto S, Bouvet P, Harayama S. Phylogenetic structures of the genus Acinetobacter based on gyrB sequences: comparison with the grouping by DNA-DNA hybridization. International journal of systematic bacteriology. 1999;49(1):87-95 . doi:10.1099/00207713-49-1-87.

Websites and databases

[106] *Kanehisa M, Goto S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res.2000;28:27-30 .* https://www.genome.jp/kegg/pathway.html
Last access: 06/01/2020.

[107] Solovyev V, Salamov A. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies (Ed. R.W. Li), Nova Science Publishers. 2011; 61-78.
http://www.softberry.com/berry.phtml?topic=fgenesb&group=help&subgroup=gfindb
Last access: 06/01/2020

[108] Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualiser. Bioinformatics.2011;27(7):1009-1010. https://mjsull.github.io/Easyfig/
Last access: 06/01/2020

[109]O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio

M, Kitts P, Murphy TD, Pruitt KD. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;4;44(D1):D733-45. https://www.ncbi.nlm.nih.gov/
Last access: 06/01/2020

[110] Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. Nucleic Acids Res. 2002;30(1):47–49. https://www.brenda-enzymes.org/
Last access: 06/01/2020.

[111] The Protein Data Bank HM Berman, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. Nucleic Acids Research.2000;28: 235-242. doi:10.1093/nar/28.1.235 http://www.rcsb.org/
Last access: 06/01/2020.

6. Annexes

6.1. Annex 1: Protein accesion numbers list

Anthranilate synthase	
Species / strain	***Protein accesion number***
Bacillus mycoides	*WP_002130223.1*
Bacillus filamentosus	*WP_046218467.1*
Bacillus pseudomycoides	*WP_006093083.1*
Bacillus thuringiensis str. 97-27	*YP_035472.1*
Bacillus subtilis	*NP_390149.1*
Bacillus amyloliquefaciens	*WP_013352680.1*
Bacillus cereus	*WP_016765469.1*
Bacillus thuringiensis YBT-1518	*YP_004997452.1*
Acinetobacter baumannii	*WP_001134879.1*
Acinetobacter indicus	*WP_016659680.1*
Pantoea agglomerans	*WP_069026348.1*
Pantoea ananatis	*WP_041457665.1*
Azospirillum brasiliense	*WP_059398696.1*
Bradyrhizobium diazoefficiens	*NP_769129.1*
Bradyrhizobium sp.	*WP_063993485.1*
Enterobacter cloacae	*YP_003612237.1*
Paenibacillus polymyxa	*WP_013371669.1*
Paenibacillus borealis	*WP_042215518.1*
Pseudomonas aeruginosa	*NP_249300.1*
Pseudomonas putida	*NP_742583.1*
Pseudomonas syringae tomato	*NP_790415.1*
Pseudomonas sp.	*WP_015475284.1*
Pseudomonas syringae B728a	*YP_237677.1*
Pseudomonas fluorescens	*WP_014340496.1*
Pseudomonas rhizosphaerae	*WP_043192485.1*
Rhizobium etli	*WP_011426312.1*
Rhizobium leguminosarum WSM2304	*WP_025395496.1*
Erwinia amylovora	*WP_004157746.1*
Azoarcus toluclasticus	*WP_018988107.1*
Dickeya zeae	*WP_012884790.1*

Dickeya solani	*WP_022633510.1*
Frankia casuarinae	*WP_011437405.1*
Serratia marcescens	*WP_025303047.1*
Serratia liquefaciens	*WP_020827151.1*
Sinorhizobium meliloti	*NP_386493.1*
Sinorhizobium fredii NGR234	*YP_002826863.1*
Sinorhizobium medicae	*YP_001327962.1*
Klebsiella pneumoniae	*YP_005226463.1*
Klebsiella oxytoca	*WP_032749461.1*
Rhodococcus jostii	*WP_081437410.1*
Rhodococcus fascians	*WP_027497282.1*
Pseudarthrobacter phenanthrenivorans	*WP_013600757.1*
Paenarthrobacter aurescens	*WP_011774530.1*
Rhodopseudomonas palustris	*WP_011160030.1*
Pantoea rwandensis	*WP_038646679.1*
Pantoea dispersa	*WP_031280219.1*
Serratia fonticola	*WP_059200954.1*
Acinetobacter soli	*WP_076033390.1*
Acinetobacter bohemicus	*WP_004650659.1*
Acinetobacter junii	*WP_075696121.1*
Rhizobium miluonense	*WP_092853408.1*
Rhizobium lusitanum	*WP_092575771.1*
Rhizobium tropici	*WP_015340663.1*
Kitasatospora setae	*WP_014140115.1*
Sinorhizobium fredii	*WP_014329157.1*
Pseudomonas enthomophila	*WP_011531855.1*
Ralstonia pickettii	*WP_012436494.1*
Ralstonia mannitolilytica	*WP_045784937.1*
Burkholderia mallei	*YP_105301.1*
Burkholderia cepacia	*WP_027788313.1*
Burkholderia pseudomallei	*YP_109645.1*
Paraburkholderia phymatum	*WP_012402069.1*
Paraburkholderia xenovorans	*WP_011489939.1*
Variovorax paradoxus	*WP_013539095.1*
Pseudomonas protegens	*WP_011063807.1*
Agrobacterium rubi	*WP_045228635.1*

Agrobacterium vitis	*WP_015916754.1*
Agrobacterium fabrum	*NP_355246.1*
Agrobacterium rhizogenes	*WP_034476631.1*
Rhodococcus erythropolis	*WP_019748660.1*
Rhizobium mesoamericanum	*WP_007533998.1*
Ralstonia solanacearum	*WP_011002787.1*

Tryptophan synthase (alpha and beta subunit)		
Species / strain	***Protein accesion number (alpha)***	***Protein accesion number (beta)***
Bacillus mycoides	*WP_002011422.1*	*WP_002086917.1*
Bacillus filamentosus	*WP_019391689.1*	*WP_040056716.1*
Bacillus pseudomycoides	*WP_006093989.1*	*WP_006093988.1*
Bacillus thuringiensis str.97-27	*YP_035478.1*	*YP_035477.1*
Bacillus subtilis	*NP_390144.1*	*NP_390145.2*
Bacillus amyloliquefaciens	*WP_013352675.1*	*WP_013352676.1*
Bacillus cereus	*NP_831022.1*	*NP_831021.1*
Bacillus megaterium	*WP_034649087.1*	*WP_013059002.1*
Bacillus thuringiensis YBT-1518	*WP_000537817.1*	*WP_023521439.1*
Acinetobacter pittii	*YP_004996598.1*	*YP_004996603.1*
Acinetobacter baumannii	*WP_000088559.1*	*WP_000372734.1*
Acinetobacter indicus	*WP_016658658.1*	*WP_016658999.1*
Pantoea agglomerans	*WP_031593124.1*	*WP_010244436.1*
Pantoea ananatis	*WP_013025947.1*	*WP_013025946.1*
Azospirillum brasiliense	*WP_051139993.1*	*WP_014238652.1*
Bradyrhizobium diazoefficiens	*NP_767386.1*	*NP_767385.1*
Bradyrhizobium japonicum	*WP_014490924.1*	*WP_014490923.1*
Bradyrhizobium elkanii	*WP_018269279.1*	*WP_016847382.1*
Enterobacter cloacae	*YP_003612232.1*	*QGN43144.1*
Paenibacillus polymyxa	*WP_013371664.1*	*WP_013371665.1*
Paenibacillus borealis	*WP_042215507.1*	*WP_042215509.1*
Pseudomonas aeruginosa	*NP_248725.1*	*NP_248726.1*
Pseudomonas putida	*NP_742252.1*	*NP_742253.1*

Pseudomonas syringae tomato	*NP_790018.1*	*NP_790017.1*
Pseudomonas sp	*WP_015474850.1*	*WP_015474851.1*
Pseudomonas syringae B728a	*YP_233145.1*	*YP_233146.1*
Pseudomonas fluorescens	*WP_014335912.1*	*WP_014335911.1*
Pseudomonas rhizosphaerae	*WP_043191967.1*	*WP_043191965.1*
Rhizobium etli	*WP_011423427.1*	*WP_011423426.1*
Rhizobium leguminosarum WSM1689	*WP_025396516.1*	*WP_025396515.1*
Rhizobium leguminosarum WSM2304	*WP_012559405.1*	*WP_003589509.1*
Rhizobium favelukesii	*WP_024313191.1*	*WP_024313192.1*
Erwinia amylovora	*WP_004157751.1*	*WP_004157750.1*
Dickeya zeae	*WP_012884795.1*	*WP_012884794.1*
Dickeya solani	*WP_022633505.1*	*WP_022633506.1*
Frankia casuarinae	*WP_011437400.1*	*WP_011437401.1*
Serratia marcescens	*WP_025303052.1*	*WP_025303051.1*
Serratia liquefaciens	*WP_020827156.1*	*WP_020827155.1*
Sinorhizobium meliloti	*NP_384135.1*	*NP_384134.1*
Sinorhizobium fredii NGR234	*YP_002827862.1*	*YP_002827861.1*
Sinorhizobium medicae	*YP_001328897.1*	*YP_001328896.1*
Klebsiella pneumoniae	*YP_005226459.1*	*YP_005226460.1*
Klebsiella oxytoca	*WP_032749463.1*	*WP_032749462.1*
Rhodococcus jostii	*WP_009473666.1*	*WP_005247488.1*
Rhodococcus fascians	*WP_032382845.1*	*WP_045841352.1*
Pseudarthrobacter phenanthrenivorans	*WP_013600762.1*	*WP_013600761.1*
Paenarthrobacter aurescens	*WP_011774535.1*	*WP_014921436.1*
Rhodopseudomonas palustris	*WP_011155641.1*	*WP_011155640.1*
Pantoea rwandensis	*WP_038646670.1*	*WP_038646672.1*
Pantoea dispersa	*WP_021510100.1*	*WP_021510101.1*
Serratia fonticola	*WP_059200957.1*	*WP_021806583.1*
Acinetobacter soli	*WP_076033526.1*	*WP_004933113.1*
Acinetobacter bohemicus	*WP_004649267.1*	*WP_004649514.1*
Acinetobacter junii	*WP_004908864.1*	*WP_004908869.1*
Rhizobium miluonense	*WP_092846528.1*	*WP_092846526.1*
Rhizobium lusitanum	*WP_092573050.1*	*WP_037195558.1*

Rhizobium tropici	*WP_041677136.1*	*WP_015338190.1*
Kitasatospora setae	*WP_014135190.1*	*WP_014135191.1*
Sinorhizobium fredii HH103	*WP_014330399.1*	*WP_014330398.1*
Pseudomonas enthomophila	*WP_011531492.1*	*WP_011531493.1*
Ralstonia pickettii	*WP_012436020.1*	*WP_004634417.1*
Ralstonia mannitolilytica	*WP_045787168.1*	*WP_045787170.1*
Burkholderia mallei	*YP_106282.1*	*YP_106284.2*
Burkholderia cepacia	*WP_027791147.1*	*WP_027791145.1*
Burkholderia pseudomallei	*YP_111702.1*	*YP_111704.1*
Paraburkholderia phymatum	*WP_012403664.1*	*WP_012403662.1*
Paraburkholderia xenovorans	*WP_011490499.1*	*WP_011490497.1*
Variovorax paradoxus	*WP_013540011.1*	*WP_013540012.1*
Pseudomonas protegens	*WP_015633627.1*	*WP_015633628.1*
Agrobacterium rubi	*WP_045231057.1*	*WP_045231056.1*
Agrobacterium vitis	*WP_012654517.1*	*WP_012654516.1*
Agrobacterium fabrum	*NP_353059.2*	*NP_353058.2*
Agrobacterium rhizogenes	*WP_034481355.1*	*WP_034481354.1*
Rhodococcus erythropolis	*WP_003944010.1*	*WP_019748657.1*
Rhizobium mesoamericanum	*WP_007528741.1*	*WP_040676711.1*
Ralstonia solanacearum	*WP_011001917.1*	*WP_016722990.1*

Tryptophan transaminase	
Species / strain	***Protein accesion number***
Bacillus thuringiensis str. 97-27	*YP_038927.1*
Bacillus thuringiensis israelensis	*YP_001573823.1*
Bacillus megaterium	*WP_034653497.1*
Dickeya zeae	*WP_012886073.1*
Pantoea rwandensis	*WP_034826928.1*
Pantoea dispersa	*WP_021507848.1*
Pantoea agglomerans	*WP_039392122.1*
Serratia liquefaciens	*WP_020826149.1*
Serratia marcescens	*WP_025304443.1*
Serratia fonticola	*WP_059199023.1*
Acinetobacter baumannii	*WP_000486246.1*
Acinetobacter indicus	*WP_016658081.1*

Acinetobacter bohemicus	*WP_004651547.1*
Acinetobacter junii	*WP_075695924.1*
Rhizobium etli	*WP_042120054.1*
Rhizobium leguminosarum WSM2304	*WP_012556200.1*
Rhizobium lusitanum	*WP_037199473.1*
Rhizobium favelukesii	*WP_040680899.1*
Rhizobium tropici	*WP_015342501.1*
Enterobacter cloacae	*YP_003613215.1*
Erwinia amylovora	*WP_004160437.1*
Klebsiella oxytoca	*WP_009653304.1*
Klebsiella pneumoniae	*YP_005226639.1*
Sinorhizobium fredii NGR234	*YP_002828121.1*
Sinorhizobium fredii HH103	*WP_014330740.1*
Sinorhizobium meliloti	*NP_384413.1*
Sinorhizobium medicae	*YP_001329162.1*
Pseudomonas fluorescens	*WP_014337444.1*
Pseudomonas enthomophila	*WP_011533853.1*
Pseudomonas aeruginosa	*NP_249561.1*
Pseudomonas putida	*NP_744123.1*
Pseudomonas rhizosphaerae	*WP_043187277.1*
Pseudomonas syringae B728a	*YP_237963.1*
Pseudomonas syringae tomato	*NP_795070.1*
Pseudomonas sp.	*WP_015478084.1*
Ralstonia mannitolilytica	*WP_045218312.1*
Burkholderia mallei	*YP_105416.1*
Burkholderia cepacia	*WP_027791545.1*
Burkholderia pseudomallei	*YP_110375.1*
Paraburkholderia phymatum	*WP_012404685.1*
Paraburkholderia xenovorans	*WP_011487541.1*
Variovorax paradoxus	*WP_026346460.1*

Indole-3-pyruvate decarboxylase	
Species / strain	***Protein accesion number***
Bacillus cereus	*NP_832195.1*
Bacillus thuringiensis str. 97-27	*YP_036605.1*

Bacillus thuringiensis YBT-1518	*WP_043924657.1*
Bacillus filamentosus	*WP_040057813.1*
Dickeya zeae	*WP_012885658.1*
Rhodopseudomonas palustris	*WP_011158661.1*
Pantoea ananatis	*WP_041457742.1*
Pantoea rwandensis	*WP_038645609.1*
Pantoea dispersa	*WP_021508250.1*
Pantoea agglomerans	*WP_069025252.1*
Serratia liquefaciens	*WP_020827955.1*
Serratia marcescens	*WP_025303759.1*
Serratia fonticola	*WP_059201431.1*
Acinetobacter pittii	*YP_004996187.1*
Acinetobacter baumannii	*WP_000469459.1*
Rhizobium etli	*WP_011428754.1*
Rhizobium miluonense	*WP_092843813.1*
Azospirillum brasiliense	*WP_035671558.1*
Enterobacter cloacae	*YP_003614211.1*
Erwinia amylovora	*WP_004158798.1*
Klebsiella oxytoca	*WP_046877197.1*
Klebsiella pneumoniae	*YP_005228112.1*
Kitasatospora setae	*WP_014139775.1*
Acetobacter senegalensis	*WP_006559524.1*
Sinorhizobium medicae	*YP_001314600.1*
Pseudomonas fluorescens	*VVM49029.1*
Pseudomonas enthomophila	*WP_044487944.1*
Pseudomonas putida	*OLS60705.1*
Bradyrhizobium sp.	*CCE00441.1*
Burkholderia sp.	*KVE46961.1*
Paraburkholderia xenovorans	*WP_011493090.1*

Indole-3-acetaldehyde dehydrogenase	
Species / strain	***Protein accesion number***
Bacillus cereus ATCC	*NP_833288.1*
Bacillus thuringiensis str. 97-27	*YP_037635*
Bacillus thuringiensis YBT-1518	*WP_023522195.1*

Bacillus mycoides	*WP_002014355.1*
Bacillus subtilis	*NP_389813.1*
Serratia liquefaciens	*WP_020828536.1*
Serratia marcescens	*WP_025302738.1*
Serratia fonticola	*WP_059201719.1*
Acinetobacter pitti	*YP_004994974.1*
Acinetobacter baumannii	*WP_024437179.1*
Rhizobium etli	*WP_076032845.1*
Rhizobium miluonense	*WP_092854547.1*
Rhizobium leguminosarum WSM2304	*WP_012556268.1*
Rhizobium lusitanum	*WP_037200250.1*
Rhizobium favelukesii	*WP_024318191.1*
Azospirillum brasiliense	*WP_059399623.1*
Enterobacter cloacae	*YP_003614735.1*
Klebsiella oxytoca	*WP_032751220.1*
Klebsiella pneumoniae	*YP_005228727.1*
Sinorhizobium fredii NGR234	*YP_002823906.1*
Sinorhizobium fredii HH103	*WP_014331646.1*
Sinorhizobium meliloti	*NP_436440.1*
Sinorhizobium medicae	*YP_001314783.1*
Pseudomonas fluorescens	*WP_014338079.1*
Pseudomonas aeruginosa	*NP_252194.1*
Pseudomonas putida	*NP_745782.1*
Pseudomonas rhizosphaerae	*WP_043190198.1*
Pseudomonas syringae B728a	*YP_235484.1*
Pseudomonas syringae tomato	*NP_792480.1*
Bradyrhizobium japonicum	*WP_014496032.1*
Bradyrhizobium elkanii	*WP_016842317.1*
Bradyrhizobium diazoefficiens	*NP_770516.1*
Ralstonia pickettii	*WP_012430470.1*
Rhodococcus jostii	*WP_011596502.1*
Burkholderia mallei	*YP_105944.1*
Burkholderia cepacia	*WP_027789375.1*

Tryptophan -2-monoxygenase	
Species / strain	***Protein accesion number***
Pseudomonas protegens	*WP_015637260.1*
Agrobacterium tumefaciens a	*NP_059676.1*
Agrobacterium tumefaciens b	*WP_040132230.1*
Agrobacterium fabrum	*NP_396528.1*
Agrobacterium rubi	*WP_045231697.1*
Paraburkholderia phymatum	*WP_012406795.1*
Agrobacterium rhizogenes	*ASK46546.1*
Erwinia sp.	*PIJ52522.1*

Indole-3-acetamide hydrolase	
Species / strain	***Protein accesion number***
Agrobacterium tumefaciens	*WP_010974823.1*
Agrobacterium rubi	*WP_045231698.1*
Agrobacterium vitis	*WP_012649066.1*
Agrobacterium rhizogenes	*WP_080705517.1*
Paraburkholderia xenovorans	*WP_011492251.1*
Bradyrhizobium japonicum	*WP_014491862.1*
Bradyrhizobium diazoefficiens	*NP_773053.1*
Variovorax paradoxus	*WP_013543727.1*
Pantoea agglomerans	*AAC17186.1*

Nitrile hydratase		
Species / strain	***Protein accesion number (alpha)***	***Protein accesion number (beta)***
Bradyrhizobium diazoefficiens	*NP_771138.1*	*NP_771137.1*
Bradyrhizobium japonicum	*WP_014495325.1*	*WP_014495326.1*
Bradyrhizobium elkanii	*WP_018271696.1*	*WP_016840716.1*
Agrobacterium rhizogenes	*WP_007696956.1*	*WP_012651754.1*
Rhodococcus erythropolis	*WP_003946052.1*	*WP_003946075.1*
Rhodococcus jostii	*WP_009472916.1*	*WP_009472917.1*
Sinorhizobium meliloti	*NP_386213.1*	*NP_386212.1*

Sinorhizobium fredii NGR234	*YP_002826528.1*	*YP_002826527.1*
Sinorhizobium fredii	*WP_014328896.1*	*WP_014328895.1*
Rhizobium lusitanum	*WP_092574114.1*	*WP_092574112.1*
Rhizobium miluonense	*WP_092855403.1*	*WP_092855405.1*
Rhizobium leguminosarum WSM2304	*WP_012558320.1*	*WP_012558319.1*
Rhizobium tropici	*WP_015340279.1*	*WP_015340278.1*
Rhodopseudomonas palustris	*WP_011158356.1*	*WP_011158357.1*
Paenarthrobacter aurescens	*WP_011777211.1*	*WP_011777212.1*
Klebsiella michiganensis	*WP_009652266.1*	*WP_014229359.1*
Rhizobium etli	-	*WP_011426028.1*
Rhizobium favelukesii	-	*WP_024313558.1*
Rhizobium mesoamericanum	-	*WP_007534919.1*

Nitrilase	
Species / strain	***Protein accesion number***
Bacillus subtilis	*NP_388806.2*
Pseudomonas enthomophila	*WP_011534641.1*
Pseudomonas fluorescens	*WP_014338132.1*
Pseudomonas sp.	*WP_015094686.1*
Bradyrhizobium diazoefficiens	*NP_770037.1*
Burkholderia cepacia	*WP_006483427.1*
Agrobacterium sp.	*WP_003521904.1*
Rhodococcus jostii	*WP_011595980.1*
Sinorhizobium fredii	*WP_014332611.1*
Rhizobium leguminosarum WSM2304	*ACS54332.1*
Rhizobium rhizosphaerae	*OQP85201.1*
Rhodopseudomonas palustris	*WP_011159701.1*
Paenarthrobacter aurescens	*WP_011773102.1*
Ralstonia solanacearum	*WP_011002568.1*

6.2. Annex 2: R pipeline

```
library(msa)
library(Biostrings)
library(ape)
library(seqinr)
library(reutils)
library(phylotools)
library(dendextend)
library(phangorn)

phylo_tree <- function(input_file){

#We import a comma-separated csv with the information of interest
#(species first column, protein ids second column), and set the
#protein id we are working we as «id»:
  data<- read.csv(input_file, header = FALSE, sep=",")
  id<-data$V2

#Using efetch from the reutils package, we retrieve from the NCBI
#protein database all the sequences of interest.
   efetch(id, db= "protein", rettype = "fasta", retmode = "text",
outfile = "old.fasta")

#In order to make our tree more readable, we change the name of each
#one of the entries using the phylotools package:
  old_name <- get.fasta.name("old.fasta")
  new_name <- data$V1
  ref2 <- data.frame(old_name, new_name)
   rename.fasta(infile = "old.fasta", ref_table = ref2, outfile =
"new.fasta")

#Then, we use readAAStringSet from Biostrings, in order to read all
#the sequences from our FASTA file:
  sequences <- readAAStringSet("new.fasta")

#After that, we align them as a previous step of the phylogenetic
#analysis. For doing this, we use the msa package, setting the
alignment conditions to CLUSTALW (default):
  alignment <- msa(sequences)

#In order to get the distant matrix, we need to transform our
#alingment to a sequinr compatible object:
  alignment_sequinr <- msaConvert(alignment, type="seqinr::alignment")

 #We calculate the distance matrix:
  distances <- dist.alignment(alignment_sequinr, "identity")

#Using the distance matrix, we create the tree according to Neighbour
joinning (NJ) using ape:
  tree <- nj(distances)
```

```
}
#We will use nitrilase.csv and gyrB_nitrilase.csv as examples:

a<-phylo_tree("csv_nitrilase.csv")
b<-phylo_tree("csv_gyrB_nitrilase.csv")

#Since we are using dendextend to compare trees, we are going to need
#to convert our phylo objects to dendextend compatible ojects. Thus,
#we need to make sure that our phylo trees are binary, ultrametric and
rooted. We can check if the conditions are true as follows:
is.binary(a)
is.ultrametric(a)
is.rooted(a)
is.binary(b)
is.ultrametric(b)
is.rooted(b)

#If not, we can force every one of the three conditions:

#If root is absent, we can set a root:
a$root.edge <- 0
b$root.edge <- 0

#If the tree is not ultrametric, we can convert it to ultrametric. In
#this step, we are going to use the phangorn package:

convert_to_ultra<-function(nonultra,method=c("nnls","extend")){
    method<-method[1]
    if(method=="nnls") nonultra←nnls.tree(cophenetic(nonultra),
nonultra,
        rooted=TRUE,trace=0)
    else if(method=="extend"){
        h<-diag(vcv(nonultra))
        d<-max(h)-h
        ii<-sapply(1:Ntip(nonultra),function(x,y) which(y==x),
            y=nonultra$edge[,2])
        nonultra$edge.length[ii]<-nonultra$edge.length[ii]+d
    } else
        cat("impossible to convert to ultrametric")
    nonultra
}

a<- convert_to_ultra(a)
b<- convert_to_ultra(b)


#Finally, we can force the binary condition too:
a<-multi2di(a)
b<-multi2di(b)
```

```
#After the transformation, we can check the three conditions once
again:
is.binary(a)
## [1] TRUE
is.ultrametric(a)
## [1] TRUE
is.rooted(a)
## [1] TRUE
is.binary(b)
## [1] TRUE
is.ultrametric(b)
## [1] TRUE
is.rooted(b)
## [1] TRUE


#Besides, we can plot our trees:
plot(a)
plot(b)


#Now, we can convert our phylo trees to dendrograms and perform all
the comparisons needed.
dend1 <- as.dendrogram(a)
dend2 <- as.dendrogram(b)

#As mentioned before, we use dendextend to compare dendrograms:
tanglegram(dend1,dend2)
```

6.3. Annex 3: Tanglegram abbreviation

Species	*Nº*	*Species*	*Nº*	*Species*	*Nº*
Bacillus mycoides	*1*	*Paenibacillus borealis*	*21*	*Sinorhizobium medicae*	*41*
Bacillus filamentosus	*2*	*Pseudomonas aeruginosa*	*22*	*Klebsiella pneumoniae*	*42*
Bacillus pseudomycoides	*3*	*Pseudomonas putida*	*23*	*Klebsiella oxytoca*	*43*
Bacillus thuringiensis str. 97-27	*4*	*Pseudomonas syringae tomato*	*24*	*Rhodococcus jostii*	*44*
Bacillus subtilis	*5*	*Pseudomonas sp.*	*25*	*Rhodococcus fascians*	*45*
Bacillus amyloliquefaciens	*6*	*Pseudomonas syringae B728a*	*26*	*Pseudarthrobacter phenanthrenivorans*	*46*
Bacillus cereus	*7*	*Pseudomonas fluorescens*	*27*	*Paenarthrobacter aurescens*	*47*
Bacillus megaterium	*8*	*Pseudomonas rhizosphaerae*	*28*	*Rhodopseudomonas palustris*	*48*
Bacillus thuringiensis YBT-1518	*9*	*Rhizobium etli*	*29*	*Pantoea rwandensis*	*49*
Acinetobacter pittii	*10*	*Rhizobium leguminosarum WSM1689*	*30*	*Pantoea dispersa*	*50*
Acinetobacter baumannii	*11*	*Rhizobium leguminosarum WSM2304*	*31*	*Serratia fonticola*	*51*
Acinetobacter indicus	*12*	*Rhizobium favelukesii*	*32*	*Acinetobacter soli*	*52*
Pantoea agglomerans	*13*	*Erwinia amylovora*	*33*	*Acinetobacter bohemicus*	*53*
Pantoea ananatis	*14*	*Dickeya zeae*	*34*	*Acinetobacter junii*	*54*
Azospirillum brasiliense	*15*	*Dickeya solani*	*35*	*Rhizobium miluonense*	*55*
Bradyrhizobium diazoefficiens	*16*	*Frankia casuarinae*	*36*	*Rhizobium lusitanum*	*56*
Bradyrhizobium japonicum	*17*	*Serratia marcescens*	*37*	*Rhizobium tropici*	*57*
Bradyrhizobium elkanii	*18*	*Serratia liquefaciens*	*38*	*Kitasatospora setae*	*58*
Enterobacter cloacae	*19*	*Sinorhizobium meliloti*	*39*	*Acetobacter senegalensis*	*59*
Paenibacillus polymyxa	*20*	*Sinorhizobium fredii NGR234*	*40*	*Sinorhizobium fredii HH103*	*60*

Species	*Nº*
Pseudomonas enthomophila	*61*
Ralstonia pickettii	*62*
Ralstonia mannitolilytica	*63*
Burkholderia mallei	*64*
Burkholderia cepacia	*65*
Burkholderia pseudomallei	*66*
Paraburkholderia phymatum	*67*
Paraburkholderia xenovorans	*68*
Variovorax paradoxus	*69*
Pseudomonas protegens	*70*
Agrobacterium rubi	*71*
Agrobacterium vitis	*72*
Agrobacterium fabrum	*73*
Agrobacterium rhizogenes	*74*
Rhodococcus erythropolis	*75*
Rhizobium mesoamericanum	*76*
Ralstonia solanacearum	*77*
Azoarcus toluclasticus	*78*
Bacillus thuringiensis israelensis	*79*
klebsiella michiganensis	*80*
Rhizobium rhizosphaerae	*81*