Universitat
de les Illes Balears

Doctoral Thesis

2016

UNDERSTANDING THE AROMATIC HYDROCARBON DEGRADATION POTENTIAL OF *PSEUDOMONAS STUTZERI*: A PROTEO-GENOMIC APPROACH

Isabel Brunet Galmés

Universitat
de les Illes Balears

Doctoral Thesis
2016

Doctoral Program of "Microbiologia Ambiental i Biotecnologia"

UNDERSTANDING THE AROMATIC HYDROCARBON DEGRADATION POTENTIAL OF *PSEUDOMONAS STUTZERI*: A PROTEO-GENOMIC APPROACH

Isabel Brunet Galmés

Thesis Supervisor: Dr. Rafael Bosch

Thesis Supervisor: Dra. Balbina Nogales

Doctor by the Universitat de les Illes Balears

A mumpare i a mumare

Agraïments

Gràcies Rafel i Balbina per dirigir aquesta tesis, que és tant meva com vostre. A en Rafel per l'oportunitat que em vares donar, ja fa 7 anys, d'entrar al laboratori. Així com també per engrescar-me dins el món de la ciència, i ensenyar-me a treballar tant dins com fora del laboratori. I a na Balbina, pels mil consells que m'ha donat aquests anys, per ensenyar-me a ser més meticulosa amb el que faig i per ajudar-me amb tot el que ha pogut.

A en Jordi i n'Elena, per acollir-me dins aquest grup de recerca, pels vostres consells i noves idees per continuar aquesta feina. Voldria agrair també a en Toni Bennasar i en Sebastià les crítiques constructives que m'heu anat fent al llarg d'aquests anys, des del projecte final de màster fins ara.

Gràcies també a tots els companys de laboratori, a més de companys sou uns grans amics. Sempre estaré agraïda a na Marga, en Toni Busquets i n'Arantxa, pels grans consells que m'heu donat, tant dins com fora del laboratori, i per estar sempre disposats a donar-me una mà. A na Joana, que tant em va ajudar durant els meus primers anys i a en Sebas, que sempre alegrava el laboratori. A na Cris, na Magda i na Claudia, per escoltar-me sempre que perdia la paciència amb els experiments. A na Maria i en David, per ajudar-me en els mals moments, i acompanyar-me en els bons. Sense oblidar a les darreres incorporacions, en Dani i na Cati, vos desig el millor amb les vostres tesis. A na Lady, en Farid i en Mohamed, pels punts de vista diferents que m'heu aportat. Gràcies Angel, Xisco i Guillem, per mantenir sempre l'ordre en el laboratori, i fer-ho sempre amb un somriure. També agrair a na Trinidad per no perdre els nervis amb les meves seqüències, que han estat moltes. Gràcies també al grup de bioquímica, sense la seva ajuda no hauria acabat mai els gels de proteïnes!

A en Joseph, per ensenyar-me a fer feina dins un laboratori i per tots els seus consells que encara ara em dóna. També per oferir-me un lloc en el seu grup durant la meva estància. Igualment agraeixo a la seva família, na Mar, na Lluc i na Neus, per donar-me tan bona benvinguda a Anglaterra.

My sincere thanks to Sue and Alex for their invaluable help during the proteomic studies at the University of Warwick, I surely appreciate your help. I'm also grateful to Despina and Amandeep, both of you made my stay in Coventry really pleasant, and I hope I will see you soon.

Gràcies a la meva família, tots vosaltres heu fet possible acabar aquesta etapa. Especialment agraeixo a mumpare i a mumare el seu suport, sense vosaltres mai hauria arribat a entrar al laboratori, i menys encara hauria acabat aquesta tesis. Agraeixo molt quant em vàreu animar a ser alumne col·laboradora ara fa 8 anys, enlloc de cercar una feina d'estiu. Igual d'important ha estat en Pere, especialment aquests darrers anys. El millor company que podria tenir, gràcies per fer agradables els moments més durs i pel coratge que sempre m'has donat. A na Magdalena, per tenir sempre unes paraules d'ànim i donar-me coratge per seguir endavant. A na Catalina i en Rafel, per acollir-me com ho heu fet, pels vostres ànims i per fer-me sempre costat. I a en Rafel, com a cunyat, per escoltar amb tant interès els meus dilemes de feina, i com a fisio perquè sense ell, potser aquesta tesis s'hauria acabat, però la meva salut se n'hauria ressentit molt!

A tots els meus amics, i especialment a ses nines, perquè encara que no sempre vos he pogut dedicar tot el temps que voldria, sempre sou allà. Sense oblidar el BioTim, les nostres xerrades no tenen preu, qualque dia arreglarem el món!

Finalment, m'agradaria donar les gràcies a les diferents entitats que han fet possible realitzar aquest treball. Primer de tot, a la Conselleria d'Educació, Cultura i Universitats per concedir-me una beca de formació de personal investigador 2011-2015, amb cofinançament del Fons Social Europeu. El suport financer es va obtenir del Ministeri d'Economia i Competitivitat a través del projecte Consolider CSD2009-00006. El model d'estudi AN10 serveix com a marc comparatiu pel projecte CTM2011-24886, amb cofinançament FEDER.

Thanks also to the European Molecular Biology Organization, for making possible my work at the University of Warwick with the Short-Term Fellowship, which is part of this thesis.

A tots, mil gràcies!

Govern de les Illes Balears
Conselleria d'Educació, Cultura i Universitats
Direcció General d'Universitats,
Recerca i Transferència del Coneixement

Invertim en el seu futur

Unió Europea
Fons Social Europeu

Table of Contents

ABSTRACT 1

INTRODUCTION 7

1. Degradation of aromatic compounds 9

2. What defines a *Pseudomonas stutzeri*? 11

2.1. Taxonomy 11

2.2. Habitat 12

2.3. Denitrification 13

2.4. Nitrogen fixation 13

2.5. Aromatic hydrocarbon degradation 15

3. Genomic studies 17

4. Proteo-genomic studies 19

OBJECTIVES 23

MATERIALS AND METHODS 27

1. Genomic analysis 29

2. Culture conditions 32

3. Selection of a benzoate-degrader derivative of strain *P. stutzeri* AN10 33

4. Proteomic studies 34

4.1. Naphthalene experiment 34

4.2 Experiment with strain BZ4D 36

5. Gene cloning and complementation 37

RESULTS AND DISCUSSION 41

CHAPTER 1: Genomic comparison of *Pseudomonas stutzeri* 43

1. General characterization of *P. stutzeri* genomes 45

2. Characteristic CDSs of *P. stutzeri* 46

3. Phylogenomic analysis of *P. stutzeri* 51

4. Pan- and core- proteome of *P. stutzeri* 55

5. Identification of putative transposases in *P. stutzeri* genomes 57

CHAPTER 2: Aromatic hydrocarbon degradation potential of *Pseudomonas stutzeri* 61

1. Analysis of aromatic hydrocarbon degradation gene clusters 63

1.1. 4-Hydroxyphenylpyruvate degradation gene cluster 65

1.2. 4-Hydroxybenzoate degradation gene cluster 68

1.3. Benzoate degradation gene cluster 71

1.4. Homoprotocatechuate degradation gene cluster 75

1.5. Carbazole degradation gene cluster 77
1.6. Phenol degradation gene cluster 81
1.7. Naphthalene degradation gene cluster 83
2. Dehydroabietic acid degradation gene cluster 91
3. Growth of *P. stutzeri* with aromatic hydrocarbons 93
CHAPTER 3: Proteomic study of naphthalene degradation by *Pseudomonas stutzeri* 101
1. *P.stutzeri* proteins detected by proteomics 103
2. General proteome comparison 103
3. Comparison of proteomes from naphthalene and succinate cultures 107
3.1. Proteomic changes in naphthalene degradation proteins 107
3.2. Putative transporters involved in naphthalene degradation 110
4. Comparison of proteomes from salicylate and succinate cultures 117
CHAPTER 4: Benzoate degradation potential of *Pseudomonas stutzeri* AN10 121
1. *P. stutzeri* AN10 potential to metabolize benzoate 123
2. Genome sequencing of *P. stutzeri* BZ4D 125
3. Proteomic analysis of BZ4D 128
4. Complementation of AN10 with BZ4D benzoate genes 131
GENERAL DISCUSSION 133
CONCLUSIONS 139
REFERENCES 143
SUPPLEMENTARY INFORMATION 165

List of Figures

Figure I1. Example of aerobic aromatic hydrocarbon degradation 10
Figure I2. Organization of the *nar*, *nir*, *nor*, and *nos* genes in *P. stutzeri* 13
Figure I3. Representation of the *P. stutzeri* A1501 *nif* genes compared with *A. vinelandii* AvOP, *K. pneumoniae* M5a1, *Azoarcus* sp. BH72, and *A. brasilense* Sp7 14
Figure I4. Aerobic catabolism of aromatic compounds in *P. stutzeri* 15
Figure I5. Naphthalene degradation pathway and the genes involved in *P. stutzeri* AN10 16
Figure I6. The catechol and protocatechuate branches of the β-ketoadipate pathway and its regulation in *P. stutzeri* A1501 17
Figure I7. Number of genomic sequences submitted to NCBI and GOLD per year 18

Figure M1. pCatA and pBenR constructions used to complement AN10 and primer localization respect from *catA* and *benR* genes 37

Figure 1.1. Structure of genes involved in the flagellar assembly system in *P. stutzeri* and *P. balearica* 49
Figure 1.2. Structure of genes involved in denitrification in *P. stutzeri* and *P. balearica* 49
Figure 1.3. Structure of genes involved in nitrogen fixation in *P. stutzeri* 50
Figure 1.4. Neighbor-Joining tree calculated from the concatenation of 69 shared COGs in the 18 *P. stutzeri* strains and *P. balearica* SP1402^T 54
Figure 1.5. Number of core-COGs assigned to each group of biological functions according to KEGG categories 56
Figure 1.6. Percentage of TnpAs classified on each IS-family 58
Figure 1.7. Number of putative TnpAs of each IS-family identified in the different studied strains 59

Figure 2.1. 4-Hydroxyphenylpyruvate degradation pathway in *Pseudomonas putida* U 66
Figure 2.2. Structure of genes involved in 4-hydroxyphenylpyruvate degradation in *Pseudomonas putida* U 68
Figure 2.3. 4-Hydroxybenzoate degradation pathway in *P. stutzeri* A1501 69
Figure 2.4. Structure of genes involved in 4-hydroxybenzoate degradation in *P. stutzeri* A1501 71
Figure 2.5. Benzoate degradation pathway in *P. stutzeri* A1501 72
Figure 2.6. Structure of genes involved in benzoate degradation in *P. stutzeri* A1501 74
Figure 2.7. Homoprotocatechuate degradation pathway in *P. putida* U 76
Figure 2.8. Structure of genes involved in homoprotocatechuate degradation in *P. putida* U and *P. stutzeri* KOS6 77
Figure 2.9. Carbazole degradation pathway in *P. resinovorans* CA10 78
Figure 2.10. Structure of genes involved in carbazole degradation in *P. resinovorans* CA10, and *P. stutzeri* XLDN-R 80
Figure 2.11. Phenol degradation pathway in *Pseudomonas* sp. CF600 81
Figure 2.12. Structure of genes involved in phenol degradation in *Pseudomonas* sp. CF600 and *P. stutzeri* KOS6 and TS44 strains 83
Figure 2.13. Naphthalene degradation pathway in *P. stutzeri* AN10 84

Figure 2.14. Structure of genes involved in naphthalene degradation in *P. stutzeri* AN10, other *P. stutzeri* strains (19SMN4, ST27MN3, B1SMN1, and KOS6), *P. balearica* SP1402^T, and plasmid NAH7 of *P. putida* G7 86
Figure 2.15. Gene structure of the putative TnpAs present between *nahP* and *nahR* 89
Figure 2.16. Phylogeny of four different enzymes involved in naphthalene degradation 90
Figure 2.17. Proposed dehydroabietic acid degradation pathway in *P. abietaniphila* BKME-9 92
Figure 2.18. Structure of genes involved in dehydroabietic acid degradation in *P. abietaniphila* BKME-9 and *P.stutzeri* MF28 and TS44 strains 93
Figure 2.19. Aerobic catabolism of aromatic compounds in *P. stutzeri* before and after the present study 96
Figure 2.20. Aromatic catabolic pathways and the genes involved in different *Pseudomonas* genomes 99

Figure 3.1. Principal component analysis of NAF values obtained for proteomes from *P. stutzeri* strains B1SMN1, AN10, 19SMN4, ST27MN3, and *P. balearica* SP1402^T 104
Figure 3.2. Sum of NAF values of proteins involved in different global biological functions (KEGG classification) 105
Figure 3.3. Statistically-significant increases in the abundance of naphthalene degradation proteins in proteomes from naphthalene cultures compared to succinate cultures 108
Figure 3.4. Putative role of NahZ in naphthalene metabolism in *P. stutzeri* 109
Figure 3.5. Average of NAF values for the detected NahP proteins in proteomes with succinate or naphthalene as carbon source 112
Figure 3.6. NAF average of the detected proteins from the transport system MexAB-OprM for each strain in proteomes with succinate or naphthalene as carbon source 113
Figure 3.7. NAF average of the detected proteins from the ExbBD-TonB-type transport system SistT7 for each strain in proteomes with succinate or naphthalene as carbon source 114
Figure 3.8. Statistically-significant increase in the abundance of naphthalene degradation proteins in proteomes from salicylate 3 mM pulse cultures compared to succinate cultures 117
Figure 3.9. Statistically-significant abundance increase of ST27MN3 benzoate degradation proteins from salicylate pulse compared to succinate cultures 118

Figure 4.1. Absorbance at 600 nm obtained for each of the successive sub-cultures of strain AN10 124
Figure 4.2. Growth of twelve AN10 derivatives in MMB with benzoate 20 mM as carbon source 124
Figure 4.3. Fragment of the alignment of the *catA* genes and the carboxy-terminal CatA CDSs from *P. stutzeri* and *P. balearica* 127
Figure 4.4. Principal component analysis of NSAF values obtained for BZ4D growing with succinate or benzoate, and AN10 growing with succinate 128
Figure 4.5. Statistically-significant increase in the abundance of benzoate degradation proteins in proteomes from BZ4D from benzoate and succinate cultures 129
Figure 4.6. NSAF average of catechol 2,3-dioxygenase NahH obtained in AN10 growing with succinate, BZ4D growing with succinate, and BZ4D growing with benzoate 130

Figure 4.7. Growth of the AN10 derivative BZ4D; U16, the AN10 derivative complemented with pBenR and pCatA constructions; Z1, the AN10 derivative complemented with the pBenR construction; and V1, the AN10 derivative complemented with the pCatA construction in MMB supplemented with benzoate 5 mM as sole carbon and energy source 131

List of Tables

Table M1. *P. stutzeri* and *P. balearica* strains used in the genomic comparison study 30
Table M2. List of bacteria whose genomes were used as models for the KEGG annotation.... 31
Table M3. Primers designed in this study for *benR* and *catA* complementation 38
Table M4. Composition of SOB broth and transformation buffers used in the preparation of competent cells 39

Table 1.1 General genomic characteristics of the 18 *P. stutzeri* strains and *P. balearica* SP1402^T 45
Table 1.2. Analysis of presence/absence of CDSs coding for characteristic traits of *P. stutzeri* 47
Table 1.3. Number shared COGs and ANIb values obtained between the different studied genomes 53
Table 1.4. Relation of the total number of annotated CDSs, COGs (with a 50 % of identity in the 50 % of the sequence), and SS-COGs per each strain 57
Table 1.5. Number of TnpAs-COGs shared between the studied strains 60

Table 2.1. Number of CDSs described in the annotation of *P. stutzeri* and *P. balearica* genomes as dioxygenases, hydroxylases, and monooxygenases 63
Table 2.2. Presence of aromatic hydrocarbon degradation gene clusters in *P. stutzeri* and *P. balearica* genomes 64
Table 2.3. Percentages of identity and E-values obtained with BLASTp of the 4-hydroxyphenylpyruvate degradation CDSs of *P. putida* U and *P. stutzeri* and *P. balearica* 67
Table 2.4. Percentages of identity and E-values obtained with BLASTp of the 4-hydroxybenzoate degradation CDSs of *P. stutzeri* A1501 and other strains of *P. stutzeri* 70
Table 2.5. Percentages of identity and E-values obtained with BLASTp of the benzoate degradation CDSs of A1501 and other *P. stutzeri* strains and *P. balearica* 73
Table 2.6. Percentages of identity and E-values obtained with BLASTp of the homoprotocatechuate degradation CDSs of *P. putida* U and *P. stutzeri* KOS6 77
Table 2.7. Percentages of identity and E-values obtained with BLASTp of the carbazole degradation CDSs of *P. resinovorans* CA10 and *P. stutzeri* XLDN-R 79
Table 2.8. Percentages of identity and E-values obtained with BLASTp of the phenol degradation CDSs of *Pseudomonas* sp. CF600 and *P. stutzeri* strains TS44 and KOS6 82
Table 2.9. Percentages of identity and E-values obtained with BLASTp of naphthalene degradation CDSs of *P. stutzeri* AN10 and other *P. stutzeri* strains and *P. balearica* 85
Table 2.10. Percentages of identity and E-values obtained with BLASTp of the dehydroabietic acid degradation CDSs of *P. abietaniphila* BKME-9 and *P. stutzeri* strains MF28 and TS44 92
Table 2.11. Comparison of tested and expected growth of *P. stutzeri* and *P. balearica* strains with different carbon sources 94

Table 3.1. Sum of NAF values of proteins involved in different biological functions (KEGG classification) 106

Table 3.2. Fold change obtained for proteins NahZ, acetate kinase, acetyl-CoA synthetase, catalase, and pyruvate dehydrogenase in *P. stutzeri* AN10 naphthalene cultures compared to succinate cultures 110
Table 3.3. Comparison of the number of transport systems from different families identified in each proteome and the number of transport systems whose normalized abundance increased significantly (some of their proteins) in proteomes from naphthalene 111
Table 3.4. Percentages of identity and E-values obtained with BLASTp of MexAB-OprM proteins from P. stutzeri and P. balearica and the toluene tolerance system TtgABC from P. putida DOT-T1E 112

Table 4.1. Generation time and maximum absorbance (600 nm) of AN10 serial cultures with increasing benzoate concentrations (from 0 to 5 mM) and later decreasing succinate concentrations (from succinate 30 mM to 0 mM) 123
Table 4.2. Generation time and maximum absorbance (600 nm) of BZ4D and AN10 with different carbon sources 125
Table 4.3. Differences observed between AN10 and BZ4D genomes 126

The following abbreviations have been used in the present study:

ABC, ATP-Binding cassette system

ANIb, Average nucleotide identity based on BLAST

ANIm, Average nucleotide identity based on MUMmer

Ap, Ampicillin

ATSDR, Agency for Toxic Substances and Disease Registry

BLAST, Basic Local Alignment Search Tool

CDD, Conserved Domain Database

CDS, Coding DNA sequences

CoA, Coenzyme A

COG, Cluster of orthologous groups

DTT, Dithiothreitol

Gm, Gentamicin

GOLD, Genomes online database

Gv, Genomovar

HGT, Horizontal gene transfer

IS, Insertion sequence

JTT, Jones-Taylor-Thornton

KAAS, KEGG automatic annotation server

KEGG, Kyoto Encyclopedia of Genes and Genomes

LB, Luria-Bertani

LC, Liquid chromatography

LDS, Lithium dodecyl sulphate-β-mercaptoethanol

MALDI-TOF, Matrix-Assisted Laser Desorption/Ionization **T**ime-of-Flight

MFS, Major Facilitator Protein system

MLSA, Multilocus sequence analysis

MMB, Minimal medium broth

MS/MS, Tandem mass spectrometry

NAF, Normalized abundance factor

NCBI, National Center for Biotechnology Information

NSAF, Normalized spectral abundance factor

ORF, Open reading frame

PCA, Principal Component Analysis

PCB, Polychlorinated biphenyl

PGAAP, Prokaryotic Genomes Automatic Annotation Pipeline

RND, Resistance-Nodulation-Cell Division Systems

SDS-PAGE, Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SOB, Super optimal broth

SS-COG, Single strain COG

SSS, Solute:Sodium Symporter system

TCA, Tricarboxylic acid cycle

TCDB, Transporter Classification Database

TnpA, Transposase

TRAP- T, Tripartite ATP-independent Periplasmic Transporter system

TsaT, Outer Membrane Anion Porin system

WGS, Whole-Genome sequence

ABSTRACT

Pseudomonas stutzeri is a *Gammaproteobacteria* which is present in a wide range of environments due to its metabolic versatility. *P. stutzeri* strains define a coherent species, whose strains have been grouped into different genomovars according to DNA-DNA hybridizations, MLSA, siderotyping, whole-cell MALDI-TOF mass spectrometry, and ANIb studies, but any whole genome comparison has been performed. This species has been extensively studied for its ability to degrade aromatic compounds. Naphthalene, phenanthrene, and β-ketoadipate degradation pathways (in strains AN10, P16, and A1501 respectively) are the aromatic hydrocarbon degradation pathways most extensively studied in *P. stutzeri* species. However, accessory proteins involved in the metabolisms mentioned above have not been described yet. Therefore, in the present study we analyzed the genomes of 18 *P. stutzeri* strains as well as *P. balearica* SP1402^T in order to (1) confirm the coherence of *P. stutzeri* species and its structure into genomovars; and (2) establish the aromatic hydrocarbon degradation potential of *P. stutzeri* species. We used three phylogenomic approaches (number of shared COGs between genomes, phylogenomic reconstruction calculated with CDSs conserved in all strains, and ANIb values) to analyze the genomic structure of the species. The results obtained contradict the current structure of *P. stutzeri* species, and showed that only strains from genomovar 1 constituted a coherent group. Therefore, we believe that a more exhaustive taxonomic study of these strains is required. The core-proteome of *P. stutzeri*, defined as 2,094 COGs, was lower than the described for other *Pseudomonas* species, suggesting that *P. stutzeri* is genetically more diverse than other species of this genus. Genomic analysis of 19 *P. stutzeri* and *P. balearica* strains also showed the potential of these strains to use a wide range of aromatic hydrocarbon as carbon sources (catechol, protocatechuate, homogentisate, homoprotocatechuate, 4-hydroxyphenylpyruvate, 4-hydroxybenzoate, salicylate, benzoate, carbazole, phenol, and naphthalene). Focusing on naphthalene degradation as a model, we analyzed the changes in the proteomes of five *P. stutzeri* and *P. balearica* strains growing with succinate, succinate with a salicylate pulse, or naphthalene as carbon source. As expected, the previously described naphthalene degradation proteins were over-expressed in naphthalene compared to succinate proteomes. Additionally, their over-expression in salicylate pulse cultures confirmed the induction of these genes by salicylate. Results allowed us to suggest the role of two other proteins (NahX and NahZ) in naphthalene degradation. The over-expression of 14 different transport systems suggests that they might be involved in naphthalene metabolism. Finally, we demonstrated that despite *P. stutzeri* AN10 was previously described as a non-degrading benzoate strain, it presents the potential to grow with benzoate as carbon source, since we could select an AN10 (*P. stutzeri* BZ4D) derivative able to grow on this compound as sole carbon and energy source. Genomic and proteomic comparisons of AN10 and BZ4D revealed that mutations in *benR* and *catA* genes were essential for the acquisition of the BZ4D benzoate degradation ability.

Abstract

Pseudomonas stutzeri es una *Gammaproteobacteria* presente en un amplio rango de ambientes gracias a su versatilidad metabólica. Las cepas de *P. stutzeri* constituyen una especie coherente, y han sido tradicionalmente agrupadas en genomovares en función de hibridaciones ADN-ADN, MLSA, siderotipado, espectrometría de masas de células enteras por MALDI-TOF, y ANIb. A pesar de ello, no han sido objeto de ningún estudio genómico comparativo completo. Esta especie ha sido ampliamente estudiada por su capacidad de degradar compuestos aromáticos. En este sentido, la vía de degradación de naftaleno, fenantreno, y β-ketoadipato (en las cepas AN10, P16, y A1501 respectivamente) son las vías de degradación de compuestos aromáticos más estudiadas de la especie *P. stutzeri*. Sin embargo, todavía no se han descrito proteínas accesorias involucradas en el metabolismo de estos compuestos. En el presente estudio se han analizado los genomas de 18 cepas de *P. stutzeri* así como *P. balearica* SP1402^T con el objetivo de (1) confirmar la coherencia de la especie *P. stutzeri* y su estructura en genomovares; y (2) establecer el potencial de esta especie para degradar compuestos aromáticos. Para analizar la estructura genómica de esta especie se han utilizado tres aproximaciones filogenómicas distintas: número de COGs compartidos, reconstrucción filogenómica calculada con las CDSs conservadas en todas las cepas y valores de ANIb. Los resultados obtenidos contradicen la estructura de la especie *P. stutzeri*, dado que sólo las cepas de la genomovar 1 forman un grupo coherente, lo que sugiere la necesidad de abordar un análisis taxonómico más exhaustivo de las otras cepas estudiadas. El core-proteoma de *P. stutzeri,* definido en 2094 COGs, es inferior al descrito para otras especies del género *Pseudomonas*, lo que sugiere que *P. stutzeri* presenta una mayor diversidad genética que las otras especies de su género. El análisis genómico de las 19 cepas de *P. stutzeri* y *P. balearica* ha puesto de manifiesto el potencial catabólico de estas cepas para degradar un amplio rango de compuestos aromáticos (catecol, protocatecuato, homogentisato, homoprotocatecuato, 4-hidroxifenilpiruvato, 4-hidroxibenzoato, salicilato, benzoato, carbazol, fenol, y naftaleno). Centrándonos en la degradación de naftaleno como modelo, se han analizado los cambios en el proteoma de cinco cepas de *P. stutzeri* y *P. balearica* creciendo con succinato, succinato con un pulso de salicilato, o naftaleno como fuente de carbono. Como se esperaba, las proteínas de degradación de naftaleno previamente descritas fueron sobreexpresadas en los cultivos con naftaleno. Además, la sobreexpresión de estas proteínas en los cultivos tratados con un pulso de salicilato confirmó la inducción de estos genes por salicilato. Los resultados nos permitieron sugerir el rol de otras dos proteínas (NahX y NahZ) en la degradación de naftaleno. La sobreexpresión de 14 sistemas de transporte diferentes sugiere que estos también podrían estar involucrados en el metabolismo del naftaleno. Finalmente, se ha demostrado que, a pesar de ser descrita como una cepa no degradadora de benzoato, *P. stutzeri* AN10 tiene el potencial de utilizar benzoato como fuente de carbono, ya que hemos podido seleccionar un derivado de AN10 (*P. stutzeri* BZ4D) capaz de crecer con benzoato como única fuente de carbono y energía. La comparación genómica y proteómica de ambas cepas han revelado que sendas mutaciones en los genes *benR* y *catA* han sido esenciales para la capacidad de degradar benzoato adquirida por BZ4D.

Pseudomonas stutzeri és una *Gammaproteobacteria* present en un ampli rang d'ambients gràcies a la seva versatilitat metabòlica. Les soques de *P. stutzeri* constitueixen una espècie coherent, i han estat tradicionalment agrupades en genomovars en funció d'hibridacions ADN-ADN, MLSA, siderotipat, espectrometria de mases de cèl·lules per MALDI-TOF, i ANIb, però no s'ha realitzat cap estudi genòmic comparatiu complet. Aquesta espècie ha estat àmpliament estudiada per la seva capacitat de degradar compostos aromàtics. En aquest sentit, la via de degradació de naftalè, fenantrè, i β-ketoadipat (del les soques AN10, P16, i A1501 respectivament) són les vies de degradació de composts aromàtics més ben estudiades de l'espècie *P. stutzeri*. Però, encara no s'han descrit proteïnes accessòries involucrades en el metabolisme d'aquests composts. En el present estudi s'han analitzat els genomes de 18 soques de *P. stutzeri* així com també *P. balearica* SP1402^T amb l'objectiu de (1) confirmar la coherència de l'espècie *P. stutzeri* i la seva estructura en genomovars; i (2) definir el potencial d'aquesta espècie per degradar composts aromàtics. Per analitzar l'estructura genòmica d'aquesta espècie s'han emprat tres aproximacions filogenòmiques (nombre de COGS compartits, reconstrucció filogenòmica calculada amb les CDSs conservades a totes les soques, i valors de ANIb). Els resultats obtinguts contradiuen l'estructura de l' espècie *P. stutzeri*. Donat que només les soques de la genomovar 1 formen un grup coherent, el que suggereix la necessitat d'abordar una anàlisi taxonòmica més exhaustiva de les altres soques estudiades. El core-proteoma de *P. stutzeri*, definit en 2094 COGS, és inferior al descrit per altres espècies del gènere *Pseudomonas*, el que suggereix que *P. stutzeri* presenta una major diversitat genètica que altres espècies del seu gènere. L' anàlisi genòmica de les 19 soques de *P. stutzeri* i *P. balearica* ha posat de manifest el potencial catabòlic d'aquestes soques per degradar un ampli rang de compostos aromàtics (catecol, protocatecuat, homogentisat, homoprotocatecuat, 4-hidroxifenilpiruvat, 4-hidroxibenzoat, salicilat, benzoat, carbazol, fenol, i naftalè). Centrant-nos en la degradació de naftalè com a model, es van analitzar els canvis en el proteoma de cinc soques de *P. stutzeri* i *P. balearica* creixent amb succinat, succinat amb un pols de salicilat, o naftalè com a font de carboni. Com s'esperava, les proteïnes de degradació de naftalè prèviament descrites van ser sobreexpressades en els cultius amb naftalè. A més, la sobreexpressió d'aquestes proteïnes en els cultius amb un pols de salicilat va confirmar la inducció d'aquests gens per salicilat. Els resultats ens van permetre suggerir el paper de dues proteïnes (NahX i NahZ) en la degradació de naftalè. La sobreexpressió de 14 sistemes de transport diferents suggereix que aquests també podrien estar involucrats en el metabolisme del naftalè. Finalment, s'ha demostrat que, malgrat haver estat descrita com una soca no degradadora de benzoat, *P. stutzeri* AN10 té el potencial d'utilitzar benzoat com a font de carboni, ja que hem pogut seleccionar un derivat d'AN10 (*P. stutzeri* BZ4D) capaç de créixer amb benzoat com a única font de carboni i energia. La comparació genòmica i proteòmica d'ambdues soques ha revelat que mutacions en els gens *benR* i *catA* han estat essencials per la adquisició de la capacitat de degradar benzoat de BZ4D.

INTRODUCTION

1. Degradation of aromatic compounds

Human activities such as maritime transport, tourism, coastal engineering, transport and oil refining, urban development, etc. have an important impact in marine ecosystems (Halpern *et al.,* 2007). Related with that, Halpern and co-workers (2007) established that every marine ecosystem is affected by at least nine different threats, being increased sea temperature, demersal destructive fishing, and organic pollution the threats with a higher impact score in the marine environment. Among all the organic pollutants, polycyclic aromatic hydrocarbons are amongst the ten most toxic compounds, according to the Priority List of Hazardous Substances published by the ATSDR (Agency for Toxic Substances and Disease Registry, USA) (http://www.atsdr.cdc.gov/).

Pollutants present in the environment, such as aromatic compounds, can be used by several microorganisms as carbon source (Fuchs *et al*., 2014). The metabolic ability of microorganisms to transform aromatic compounds and other organic pollutants into less harmful substances, which are then integrated into natural biogeochemical cycles, is known as biodegradation (Margesin and Schinner, 2001). The exploitation of the catabolic capabilities of these microorganisms, which is known as bioremediation, can be essential for the attempt to reduce environmental pollution (Fuentes *et al*., 2014).

Microorganisms possess many different pathways for the degradation of a wide range of chemical structures. In the case of aromatic compounds, aerobic degradation normally proceeds via two major steps: many different peripheral pathways, which transform diverse aromatic compounds to few intermediates (catechol, protocatechuate, gentisate, homoprotocatechuate, homogentisate, hydroquinone, or hydroxyquinol); and a few central or ring-cleavage pathways, which degrade those intermediates to the central metabolism of the cell (Díaz, 2004). In peripheral pathways, the aromatic ring is destabilized through mono- or di-oxygenation, typically resulting in the addition of one or two hydroxyl groups. The remainder of the peripheral pathway consists of preparing the ring for cleavage, usually by dehydrogenation, to form catechol or a closely related monocyclic compound such as gentisate, hydroquinone, or salicylate (George and Hay, 2011).

Following the formation of catechol taken as an example, the cleavage of the hydroxylated aromatic ring by a second dioxygenase through *ortho*- or *meta*- cleavage takes place (see Figure I1). During *ortho*-cleavage, dioxygenation takes place at the 1,2-position of the catechol, within the hydroxyl groups. In contrast, during *meta*-cleavage fission takes place outside the hydroxyls and usually occurs at the 2,3-position, although distal ring fission in the 1,6-position has also been reported (Koh *et al*., 1997). *Ortho*- and *meta*-cleavage are catalyzed by intradiol and extradiol dioxygenases that use Fe(III) and Fe(II), respectively as cofactors. Both enzymes have different structures and catalytic mechanisms. Extradiol dioxygenases appear to be more versatile, as they are involved in more catabolic pathways. Further degradation of *meta*-cleaved products occurs via substrate specific enzymes, so the nature of central pathways becomes a key determinant of the substrate range for bacterial degradation (Harayama and Rekik, 1989; Vaillancourt *et al*., 2006). Finally, further transformations channel ring fission products into the tricarboxylic acid cycle (TCA) for energy production (George and Hay, 2011).

Many microorganisms have been reported to degrade aromatic hydrocarbon compounds (Fuchs *et al*., 2011). Examples of them are *Bacillus* sp. (Kim and Oriel, 1995), *Alcaligenes* sp. (Essam *et al*., 2010), *Streptomyces* sp. (Endo *et al*., 2002), *Trichosporon* sp. (Alexieva *et al*., 2008), *Candida* sp. (Tsai *et al*., 2005), *Ochromonas* sp. (Semple and Cain, 1996), and *Pseudomonas* sp., being the last one the most extensively studied (Wasi *et al*., 2013). The genus *Pseudomonas* belongs to the Class *Gammaproteobacteria*, and is one of the most complex bacterial genera since it contains the largest number of species [230 species according to the List of Prokaryotic Names with Standing in Nomenclature on February 2016 (Euzéby (1997)].

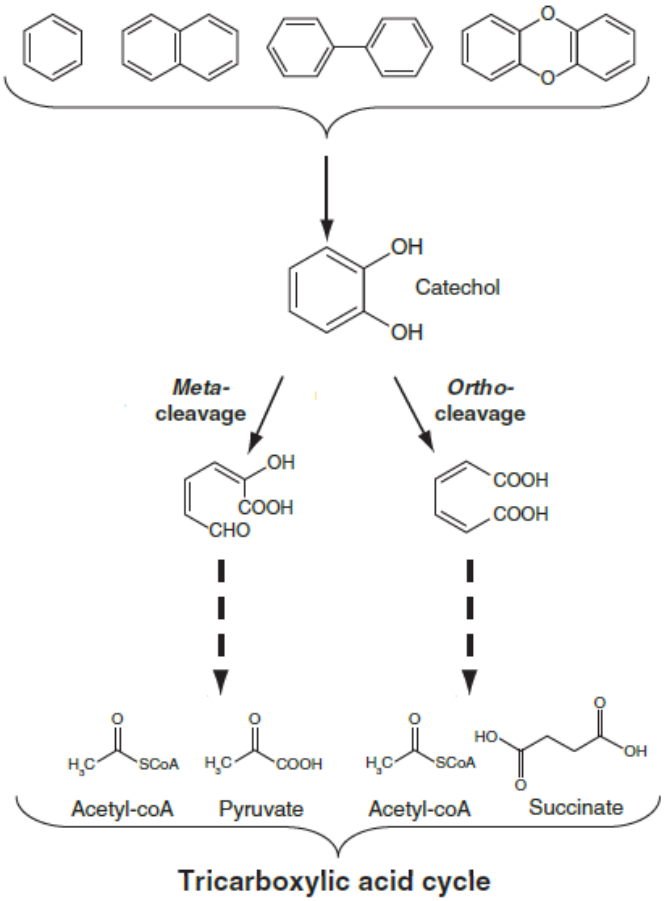

Figure I1. Example of aerobic aromatic hydrocarbon degradation (adapted from George and Hay, 2011).

The genus *Pseudomonas* is widely distributed in nature, inhabiting many different environments, as terrestrial and aquatic habitats, animals, and plants (Silby *et al*., 2011). *Pseudomonas* spp. are important in the recycling of nutrients, but also in clinical and biotechnology. Four *Pseudomonas* species have been studied due to their potential in biodegradation: *P. aeruginosa*, *P. putida*, *P. mendocina*, and *P. stutzeri* (reviewed in Jiménez *et al*., 2004; Kasai and Harayama, 2004; Nishino and Spain, 2004; Wasi *et al*., 2013). In this sense, *P. stutzeri* is a remarkable species of this genus, extensively studied in our laboratory due to their exceptional physiological capacities, such as the capability to metabolize a wide range of aromatic substrates, or the ability to denitrify and fix nitrogen (Lalucat *et al*., 2006).

2. What defines a *Pseudomonas stutzeri*?

P. stutzeri was first described by Burri and Stutzer (1895), and afterwards, van Niel and Allen (1952) defined its phenotypic features, being definitively designated as *P. stutzeri*. Typically, cells are rod shaped, 1–3 μm in length and 0.5 μm in width, and have a single polar flagellum. However, under certain conditions, one or two lateral flagella may also be produced (Lalucat *et al*., 2006). Phenotypic traits of the species include a Gram negative stain, positive catalase and oxidase tests, and a strictly respiratory metabolism. Bacteria of this species can grow on starch and maltose, and have a negative reaction for arginine dehydrolase and glycogen hydrolysis tests. *P. stutzeri* strains are able to grow in minimal media with ammonium or nitrate and an organic carbon as energy source. No additional growth factors are required (Lalucat *et al*., 2006). This species is widely distributed in the environment, occupying diverse ecological niches, as we will discuss later, and has also been isolated as an opportunistic pathogen from humans (Lalucat *et al*., 2006).

2.1. Taxonomy

In the past, mol % G+C content of DNA was a useful characteristic in taxonomy for delineating species (Rosselló-Móra and Amann, 2015). This parameter was determined in *P. stutzeri* strains by thermal denaturation temperature of the DNA and by enzymatically hydrolyzing the DNA. Reported values vary widely: from 60.7 to 66.3 mol % G+C (Palleroni *et al*., 1970) and from 60.9 to 65.0 mol % G+C (Rosselló *et al*., 1991). Despite the heterogeneity between *P. stutzeri* strains, species division was not possible due to the lack of phenotypic differences among the strains. For this reason, Rosselló and co-workers (1991) proposed the term genomovar to be used for distinct genomic groups that are sufficiently different to be classified as different species, but with phenotypes that do not show sufficient robust differences for discriminating them. Only genomovar 6 could be differentiated chemotaxonomically from *P. stutzeri,* due to its ability to grow above 42°C, to grow in the presence of 8.5 % of NaCl, use of xylose and the inability to use ethylene glycol as unique carbon and energy sources (Bennasar *et al.,* 1996). Therefore, strains belonging to genomovar 6 were reclassified as *P. balearica*, the species most closely related to *P. stutzeri*. Currently, at least 21 genomovars have been established within the species (Scotta *et al*., 2013).

Phylogenetic studies of *P. stutzeri* strains based on 16S rDNA sequences demonstrate that they constitute a coherent branch, together with related species within the genus, such as *P. mendocina*, *P. alcaligenes*, *P. pseudoalcaligenes*, and *P. balearica* (Lalucat *et al*., 2006). Several other genes, in addition to 16S rDNA, have been used as phylogenetic markers in *P. stutzeri* studies. The housekeeping genes *gyrB* and *rpoD,* firstly used by Yamamoto and co-workers (2000), had become very useful to discriminate between *Pseudomonas* species, as they are less conserved than the 16S rDNA sequence. The use of *rpoD* partial sequence has been also proposed by Scotta and co-workers (2012) for routine identification and genomovar assignation of *P. stutzeri*. Multilocus Sequence Analysis (MLSA) of genes *rpoD*, *gyrB*, *rpoB*, and 16S rRNA have been also proposed to differentiate *P. stutzeri* strains and to clearly establish the genetic diversity and population structure of the species (Mulet *et al*., 2010). Apart from MLSA, other characteristics have been used to establish genomic groups within the species, such as the

diversity of siderophores produced (siderotyping), and whole-cell MALDI-TOF mass spectrometry (Sikorski *et al*., 2005; Mulet *et al*., 2008; Scotta *et al*., 2013).

New genome sequencing technologies allow genome comparisons that might help the taxonomy and phylogeny. Nowadays, the only genomic comparison of *P. stutzeri* strains was performed by Gomila and co-workers (2015). In that study, MLSA and whole-genome comparison indices that have been proposed for species delineation were calculated for 112 *Pseudomonas* genomes (including 12 *P. stutzeri* genomes): tetranucleotide usage patterns (Teeling *et al*., 2004), average nucleotide identity based on MUMmer and BLAST (ANIm and ANIb, respectively) (Goris *et al*., 2007), and genome-to-genome distance (Meier-Kolthoff *et al*., 2013). The results of Gomila and co-workers (2015) showed a coherence of the *P. stutzeri* group but also that several *P. stutzeri* strains that have been described as genomovars might be viewed as taxonomic outliers of the species. Apart from the whole genome comparison analysis performed by Gomila and co-workers, no other phylogenomic study of *P. stutzeri* has been published yet.

2.2. Habitat

P. stutzeri is an ecologically relevant species, occupying various niches. Some of its strains have been isolated from marine environments (García-Valdés *et al*., 1988), soil (Ma *et al*., 2007), the rhizosphere (Qiu *et al*., 1981), samples from industry (Grigoryeva *et al*., 2013), and even we can find clinical isolates (Gilardi, 1971; Scotta *et al*., 2012).

Soil and rhizosphere are some of the most common *P. stutzeri* environments. The members of this species are frequently described in association with plants such as wheat and barley, due to its nitrogen fixation ability (Lovell *et al*., 2000; Demba Diallo *et al*., 2004). Many *P. stutzeri* strains have been isolated from contaminated soils, such as strain TS44, isolated from an arsenic-contaminated soil (Li *et* al., 2012). The isolation of *P. stutzeri* strains from marine environments is also very common. The most relevant marine strains studied in detail are ZoBell (CCUG 16156), isolated from the water column in the Pacific ocean (Zobell and Upham, 1944); AN10 (CCUG 29243), isolated from polluted marine sediment from the Mediterranean sea (García-Valdés *et al*., 1988); NF13, isolated from a sample taken at 2,500 m depth in the Galapagos rift from near a hydrothermal vent (Ruby *et al*., 1981); and strains MT-1, isolated from deep-sea samples from the Mariana Trench at 11.000 m depth (Tamegai *et al*., 1997).

Apart from natural environments, *P.* stutzeri strains have been isolated from anthropogenic environments such as wastewater (B1SMN1, Rosselló *et al*., 1991) and industrial hydrocarbon sludge (KOS6, Naumova *et al*., 2009). Naphthalene degraders, thiosulfate oxidizers, chlorobenzoate degraders, and cyanide oxidizers have been isolated in wastewater treatment plants (Rosselló *et al*., 1991; Lalucat *et al*., 2006).

2.3. Denitrification

Denitrification is a respiration process in which bacteria use nitrate instead of oxygen as the electron acceptor under anaerobic conditions for the generation of an electrochemical gradient across the cytoplasm membrane (Zumft, 1997). *P. stutzeri* is one of the most active denitrifying heterotrophic bacteria and it has been considered a model system for the denitrification process.This, together with the fact that all well-characterized *P. stutzeri* strains can use nitrate as terminal electron acceptor, defines denitrification as a stable trait for *P. stutzeri* species (Lalucat *et al*., 2006).

The denitrification process involves four successive steps, performed by several metalloproteins (see Figure I2): respiratory nitrate reductase catalyzes the reduction of nitrate to nitrite; nitrite reductase catalyzes the reduction of nitrite to nitric oxide; nitric oxide reductase catalyzes the reduction of nitric oxide to nitrous oxide; and finally nitrous oxide reductase catalyzes the reduction of nitrous oxide to N_2 (Knowles, 1982; Zumft, 1997). In contrast to the assimilatory reduction of nitrate or nitrite to ammonia for biosynthetic purposes, denitrification in bacteria is a dissimilatory transformation, associated with energy conservation.

In *P. stutzeri* the genes encoding functions for nitrite respiration (*nir*), nitric oxide respiration (*nor*), and nitrous oxide respiration (*nos*) seem to be preferentially organized in a single denitrification super cluster of about 30 kb (Lalucat *et al*., 2006) (see Figure I2). In strain CCUG 16156 (ZoBell) this cluster contains 33 genes, which are arranged in subclusters sorted as *nos-nir-nor* (Zumft, 1997). Conversely, the genes encoding for nitrate reductase (*nar*) are located independently from the *nos-nir-nor* gene cluster (Härtig *et al*., 1999). Recent studies have identified similar gene structures in A1501 strain (Yan *et al*., 2005).

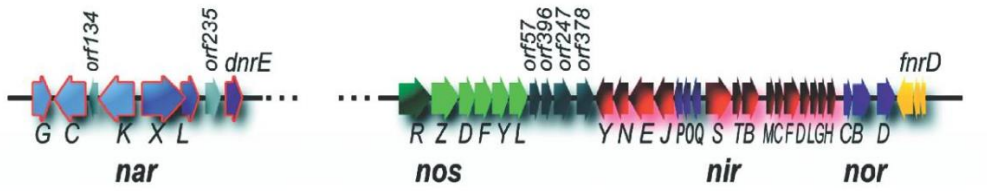

Figure I2. Organization of the *nar*, *nir*, *nor*, and *nos* genes in *P. stutzeri* proposed by Zumft (1997) and adapted by Lalucat *et al*., 2006.

2.4. Nitrogen fixation

P. stutzeri strains are not only able to denitrify, some of them are also able to fix nitrogen (Yan *et al*., 2008). Biological nitrogen fixation is the major route for the conversion of atmospheric nitrogen gas (N_2) to ammonia. This reaction is carried out by the enzyme nitrogenase, which is composed of two different subunits: the Fe protein (encoded by *nifH*) and the MoFe protein (encoded by *nifDK*) (Dean and Jacobson, 1992; Dean *et al*., 1993). However, the primary translation products of the nitrogenase structural genes are not active. The Fe protein needs four iron atoms organized into a single Fe_4S_4 metallocluster. The MoFe protein contains two different types of metalloclusters: two P clusters, constituted of 30 iron atoms; and two FeMo-cofactors, formed by molybdenum and iron atoms. As a result, a group of *nif* genes is required

for processing immature nitrogenase structural components to active forms. Previous studies have identified the proteins that participate in this maturation: NifM, NifE, NifN, NifB, NifV and NifQ (Dean and Jacobson, 1992; Dean *et al*., 1993). In addition, NifS and NifU proteins have been postulated as involved in inorganic iron and sulfide acquisition required for the formation of the nitrogenase-Fe-S cluster (Dos Santos *et al*., 2012).

The ability to fix nitrogen is limited to a small subset of prokaryotes (diazotrophs), which belong to different taxonomic groups. *P. stutzeri* simultaneous ability for nitrogen fixation and denitrification may be of relevance to overall nitrogen cycling in several ecosystems (Lalucat *et al*., 2006). *P. stutzeri* strain A15 (from whom the derivative strain A1501 was later obtained) is a predominant diazotrophic strain isolated from the paddy soil rice rhizosphere (Vermeiren *et al*., 1999). It has been shown that this strain is able to colonize and infect rice roots and to grow endophytically (Desnoues *et al*., 2003). By doing so, it can provide rice plants with fixed nitrogen and hence promote plant growth (Lalucat *et al*., 2006). Another example of described *P. stutzeri* diazotrophic strains is DSM 4166, which was isolated from the roots of a *Sorghum nutans* cultivar (Krotzky and Werner, 1987).

A previous study analyzed the set of genes involved in nitrogen fixation in *P. stutzeri* A1501 (Yan *et al*., 2008). All the genes described above were found, as well as the positive and negative gene regulators (*nifA* and *nifL*), the flavodoxin encoding gene *nifF*, and five other genes with unknown function (*nifT*, *nifY*, *nifW*, *nifZ*, and *nifX*). Moreover, this study revealed a strong homology of the A1501 gene structure with the conventional nitrogen fixation system of *Azotobacter vinelandii* AvOP (see Figure I3), except that the *nif* genes were not contiguous in *A. vinelandii*, but were distributed into two portions of the genome (Yan *et al*., 2008).

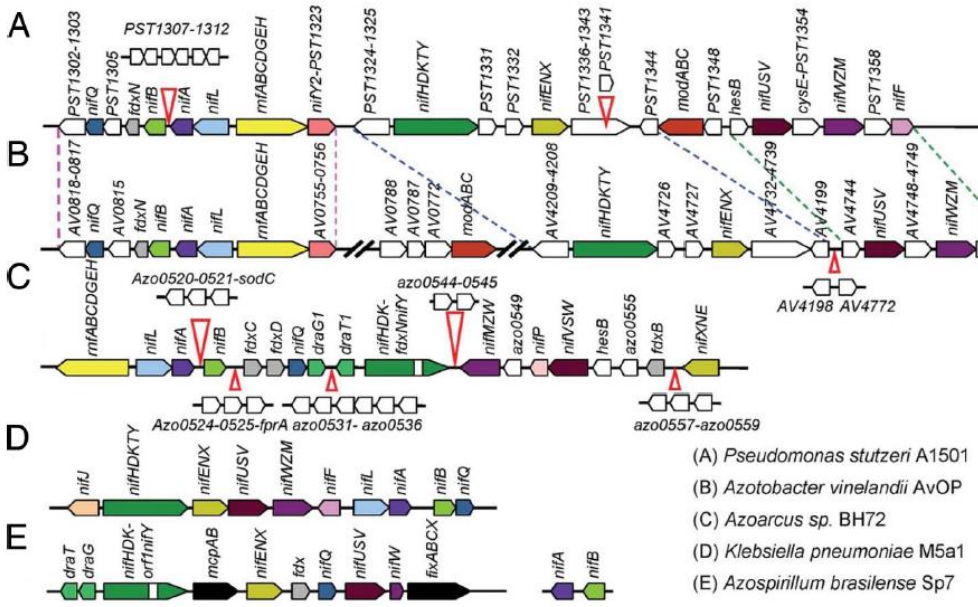

Figure I3. Representation of the *P. stutzeri* A1501 *nif* genes compared with *A. vinelandii* AvOP, *Klebsiella pneumoniae* M5a1, *Azoarcus* sp. BH72, and *A. brasilense* Sp7 (Yan *et al*., 2008).

2.5. Aromatic hydrocarbon degradation

The ability of *P. stutzeri* strains to degrade aromatic compounds aerobically is well known (Lalucat *et al*., 2006). As shown in Figure I4, the aerobic catabolism of these compounds involves a wide variety of peripheral degradation pathways. *P. stutzeri* strains are able to metabolize: aromatic alkyl derivatives (butylbenzene, *sec*-butylbenzene, *tert*-butylbenzene and isobutylbenzene); benzoate; mono- and di -halogen (Br, Cl, I, or F) benzoates; 4-hydroxybenzoate; benzenesulfonate and 4-methyl-benzenesulfonate; carbazole; cresol; dibenzothiophene; fluoranthene; fluorene; indan (or benzocyclopentane); naphthalene and its methyl- and chloro- derivatives; polychlorinated biphenyls (PCBs); phenanthrene; phenol and dimethylphenol; pyrene; quinoline; salicylate and its methyl- and chloro- derivatives; tetralin; toluate; toluene; and xylene (reviewed in Lalucat *et al*., 2006; Kaczorek *et al*., 2013). These pathways channel substrates into a small number of common intermediates such as catechol, methylcatechol, chlorocatechol, protocatechuate, and gentisate. The intermediates are further processed by a few central pathways to TCA intermediates (Lalucat *et al*., 2006).

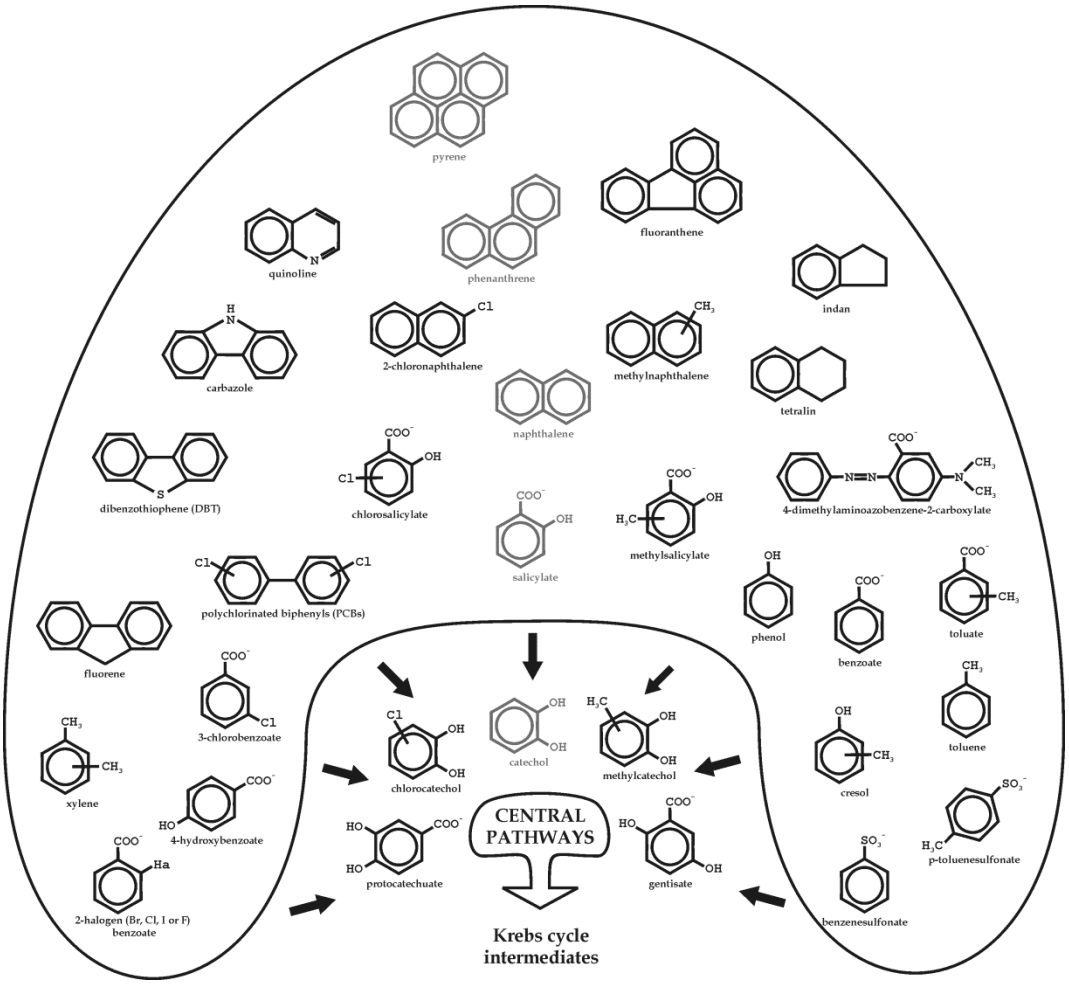

Figure I4. Aerobic catabolism of aromatic compounds in *P. stutzeri* (Lalucat *et al*., 2006).

Three different *P. stutzeri* strains have been well studied due to their biological and biotechnological interest: P16, AN10, and A1501. *P. stutzeri* P16 is a polycyclic aromatic hydrocarbon degrading bacterium isolated from a phenanthrene enrichment culture of a creosote contaminated soil (Stringfellow and Aitken, 1994). This strain is able to grow, via

salicylate, using phenanthrene, fluorene, and naphthalene and its methyl-derivatives as the only carbon and energy sources. Growth rates and kinetic coefficients of the enzymes involved have also been determined (Stringfellow and Aitken, 1995). This strain is also able to transform pyrene to cis-4,5-dihydro-4,5-dihydroxypyrene (Kazunga and Aitken, 2000). However, its genome has not been published, and the genes involved in those pathways have not been identified yet.

P. stutzeri AN10 is a naphthalene degrading bacterium isolated from polluted marine sediments from the western Mediterranean Sea (García-Valdés *et al*., 1988). Strain AN10 is able to degrade naphthalene, 2-methylnaphthalene, and salicylate as sole carbon and energy sources. The catabolic genes involved in naphthalene degradation (Bosch *et al*., 1999a; Bosch *et al*., 1999b; Bosch *et al*., 2000) (see Figure I5), and their enzymatic activities have been previously determined (naphthalene dioxygenase, salicylaldehyde dehydrogenase, salicylate hydroxylase, and catechol 2,3-dioxygenase) (Rosselló-Mora *et al*., 1994). As shown in Figure I5, naphthalene transformation to salicylate is carried out by the gene products of the operon *nahABFCED* (Bosch *et al*., 1999a). Then, salicylate is attacked by two different salicylate hydroxylases (NahG and NahW), which decarboxylate it to catechol (Bosch *et al*., 1999b). Subsequently, catechol is attacked by a catechol 2,3-dioxygenase (NahH) that linearizes the aromatic ring generating hydroxymuconic semialdehyde as product. Finally, this compound is channeled into the TCA cycle by the successive action of the remaining gene products of the operon *nahGTHINLOMKJ* (Bosch *et al*., 2000). All this genetic system, consisting in three operons (*nahAaAbAcAdBFCED*, *nahGTHINLOMKJ* and *nahW*), is regulated by the LysR transcriptional regulator NahR in response to salicylate (Bosch *et al*., 2000) (see Figure I5). In contrast to the usual plasmid location of the naphthalene catabolic pathway, degradation genes are chromosomally encoded in strain AN10 (Rosselló-Mora *et al*., 1994). Salicylate, although it is an intermediate of naphthalene degradation and the inducer of this pathway, is toxic for *P. stutzeri* AN10 at concentrations above 6 mM (Lanfranconi *et al*., 2009). Despite the good knowledge we have of the catabolic pathway for naphthalene degradation, we still do not know why some strains have different

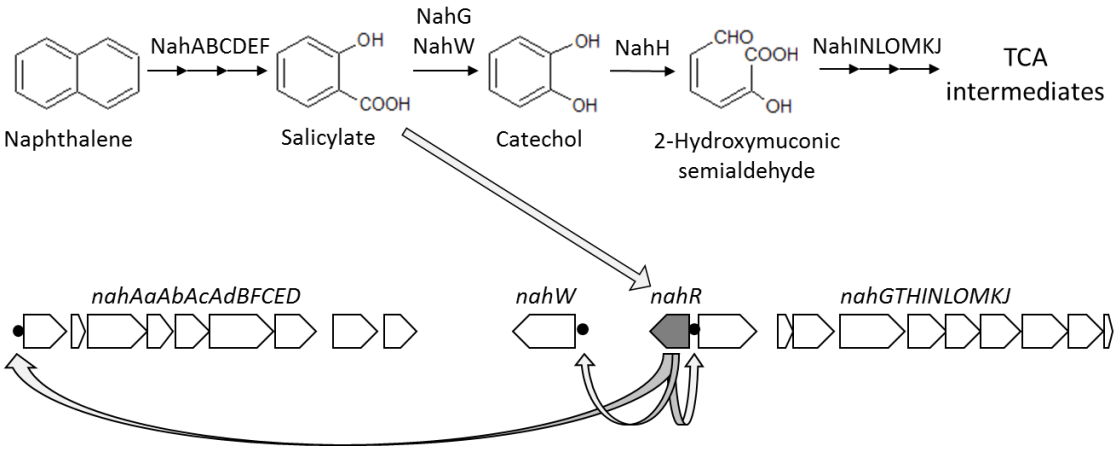

Figure I5. Naphthalene degradation pathway and the genes involved in *P. stutzeri* AN10 (drawn the information in Bosch *et al*., 1999a; Bosch *et al*., 1999b; Bosch *et al*., 2000). Gene encoding for the transcriptional regulator NahR is shown in gray, and genes encoding for enzymes involved in naphthalene degradation are shown in white. The transcriptional regulation of naphthalene degradation genes by NahR in presence of salicylate is represented with arrows and promoters for the transcription of naphthalene degradation genes are shown with black dots.

tolerance to salicylate and which are the accessory proteins (i.e. intake transporters and efflux pumps) involved in the metabolism of naphthalene.

Last years, some aromatic hydrocarbon degradation studies have focused their attention in *P. stuzeri* A1501. As mentioned above, this strain was isolated from a rice rhizosphere (Vermeiren *et al*., 1999). Its interest not only remains on its denitrification and nitrogen fixation capabilities, but also on its aromatic hydrocarbon degradation abilities. Recently Li and co-workers (2010) described the presence of the two branches of the β-ketoadipate pathway in A1501. They have also identified the genes involved in the degradation of benzoate (*ben*) via catechol (*cat*), as well as those involved in 4-hydroxybenzoate degradation (*pob*) via protocatechuate (*pca*). They have also described the absence of the LysR transcriptional regulator (*catR*) and the 4-hydroxybenzoate permease (*pcaK*). Moreover, the expression of genes involved in benzoate degradation was shown to be tightly modulated in this bacterium. The transcriptional activation of *benABCD* operon via the regulator BenR was demonstrated. On the other hand, the transcriptional regulation of the *catBCA* operon could be activated directly in response to benzoate (see Figure I6).

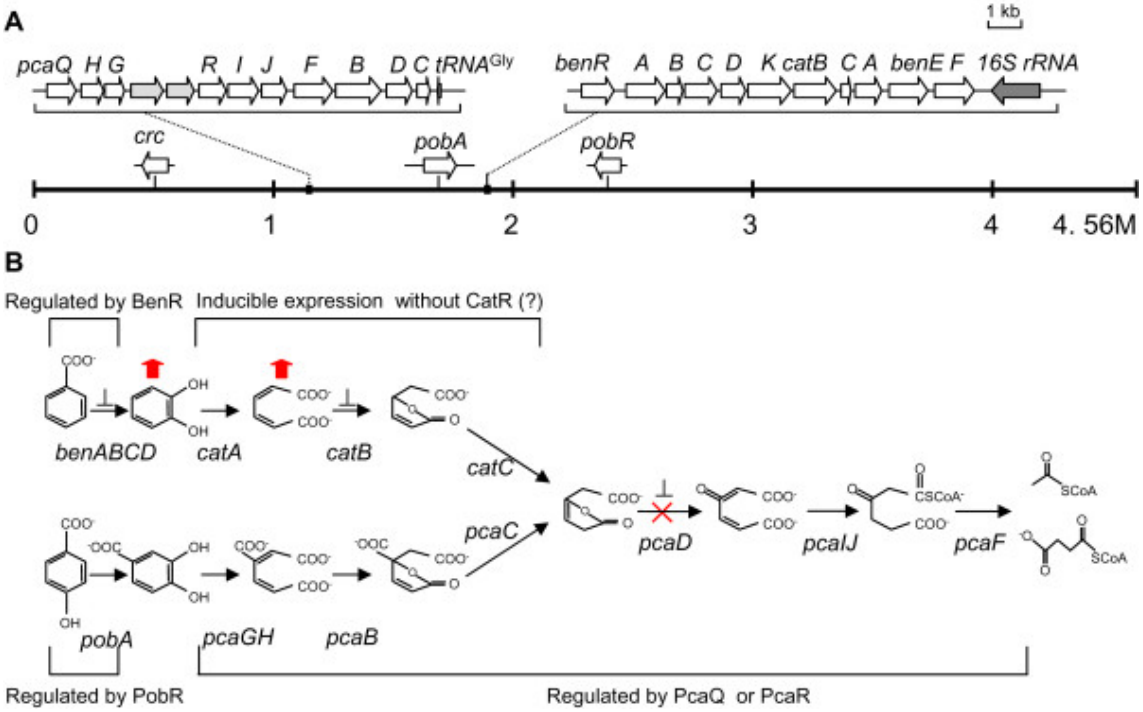

Figure I6. The catechol and protocatechuate branches of the β-ketoadipate pathway and its regulation in *P. stutzeri* A1501. (A) Localization of the gene clusters involved in this pathway. (B) Predicted biochemical steps for the catechol and protocatechuate pathways in *P. stutzeri* A1501. The question mark indicates an unknown mechanism that may be involved in the regulation of *cat* genes. Inactivation of *pcaD* is shown by "×" and accumulation of the intermediates catechol and *cis, cis*-muconate in the supernatants of the *pcaD* mutant are shown by red vertical arrows. Genes whose expression is under catabolite repression control (Crc) are indicated by "⊥" (Li *et al*., 2010).

3. Genomic studies

In the last years, next generation sequencing methods have made possible to obtain whole genome sequence of microorganisms in a short period of time at affordable prices. Due to this

fact, the number of sequences submitted to public databases has increased vertiginously last decade. The most important public sequence database is probably the National Centre for Biotechnology Information (NCBI; Barrett *et al*., 2011). According to this database, on July 31st, 2015, it held more than 258 billion whole genome sequences (WGS), as well as more than 185 billion GenBank sequences (www.ncbi.nlm.nih.gov/genbank/statistics). Moreover, taking into account the escalation of these values the last decades, they will probably continue increasing (see Figure I7 A).

Another important public sequence database is the Genomes Online Database (GOLD; Reddy *et al*., 2015), which on July 31st, 2015, presented 67,874 sequencing projects submitted. Nearly 63 % of these projects belong to species from the domain *Bacteria*, and focusing on bacterial genomes, the most abundant phylum is the *Proteobacteria*, which represent 34.44 % of the sequences. Similarly to the WGS sequences of the NCBI database, the GOLD database also experienced a rapid increment of genome projects, mainly since 2011 (see Figure I7 B).

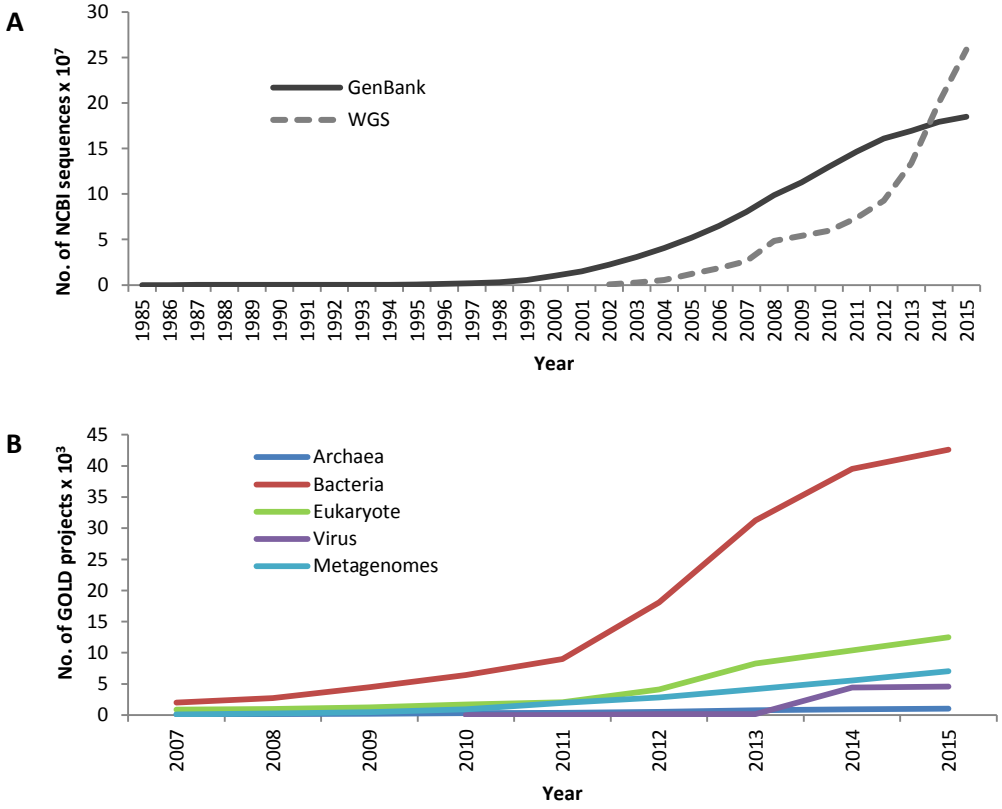

Figure I7. Number of genomic sequences submitted to NCBI (A) and GOLD (B) per year. Extracted from www.ncbi.nlm.nih.gov/genbank/statistics and gold.jgi-psf.org/statistics, on July 31st, 2015.

The availability of this large number of genome sequences allow to perform genomic studies. In the particular case of *P. stutzeri*, on November 2014 there were 17 different genomes available in the NCBI database. These are the genomes we selected for our comparative genomic analysis. Most of them have been analyzed separately and published in brief genome announcements: CGMCC 1.1803^T (ATCC 17588^T) (Chen *et al*., 2011); DSM 4166 (Yu *et al*., 2011); T13 (Li *et al*., 2012); XLDN-R (Liu *et al*., 2012); AN10 (CCUG 29243) (Brunet-Galmés *et al*., 2012); SDM-LAC (Jiang *et al*., 2012); TS44 (Li *et al*., 2012); DSM 10701 and B1SMN1 (Busquets *et al*., 2012, 2013); CCUG 16156 and NF3 (Peña *et al*., 2012, 2013); KOS6 (Grigoryeva *et al*., 2013); MF28 (Chauhan

et al., 2013); 28a24 (Smith *et al*., 2014); . However, no comparative analysis of these genomes had been done. On April 2016 the genomes of 10 more *P. stutzeri* strains were available at the NCBI database.

4. Proteo-genomic studies

The emergence of advanced liquid chromatography mass spectrometry technologies for characterizing very complex mixtures of proteins has launched the field of proteomics. This level of global protein information about an organism such as a bacterium can be combined with genomic data to identify its proteins on a genome-wide scale (VerBerkmoes *et al*., 2004). Moreover, nowadays it also allows an extensive characterization of protein primary structure, topology, interaction network, function, and regulation (Doherty and Whitfield, 2011). In the 1990s, proteins were identified by a mass-fingerprint of a 2D-PAGE gel spot. But this method was restricted to single polypeptide identification each time (Armengaud, 2013). Nowadays, tandem mass spectrometers such as Orbitrap allow to establish the global mass of each peptide in a mixture, as well as to obtain a MS/MS spectrum of each one, were intensities of the fragment ions are reported with their m/z ratio. Such instruments, coupled to a nano liquid chromatography, are able to resolve complex peptide mixtures and record thousands of MS/MS spectra per hour, generating data for thousands of peptides (Armengaud, 2013; Steen and Mann, 2004). Thus, it is no longer necessary to resolve proteins in a 2D-PAGE gel. The resulting MS/MS spectra are then compared with theoretical spectra obtained for all peptide sequences of the protein database (usually the annotated CDSs of the microorganism genome). This process might be performed with different proteomic search engines, such as MASCOT and SEQUEST (Steen and Mann, 2004). The peptide sequences are then used to identify the proteins present in the sample.

Recent studies have shown that counting the number of MS/MS spectra of a given protein in different samples running under strictly equal nanoLC-MS/MS conditions led to a reasonable comparison of their respective abundances. This spectral count method has become quite popular and several studies have used this semi-quantitative approach (Liu *et al*., 2004; Christie-Oleza *et al*., 2015). Other studies take into account the intensity of the corresponding parent ion with the software Progenesis QI for Proteomics in order to increase the number of quantifiable proteins (Bracht *et al*., 2015). Extracted ion-chromatographic intensities show a linear response to protein quantity and such an approach does not need a MS/MS scan of the same peptide in all of the conditions analyzed. It is considered as more rigorous than the quantification of spectral counts (Armengaud, 2013).

An example of possibilities offered by shotgun proteomics is that of Hartmann and Armengaud (2014) regarding the degradation of dibenzo-p-dioxin by *Sphingomonas wittichii* RW1. Although this degradation pathway was biochemically and genetically well characterized (Nojiri *et al*., 2001a; Colquhoun *et al*., 2012), the study of the proteome allowed to the identification of new proteins involved in the degradation of this compound. Another example of the usefulness of a proteo-genomic study was carried out by Christie-Oleza and co-workers (2012) that analyzed the proteome of *Ruegeria pomeroyi* DSS-3 under 30 different culture conditions. These authors were able to detect the expression of the 46 % of the theoretical proteome of this strain. The

results of this study suggest that approximately 30 % of the coding sequences of a genome are the minimum required to live in a virtual environment without stress, energy and specific metabolite requirements. Another 20% of the proteins is dedicated to adaptation to general cellular physiological variations and main specific physiological traits. The remaining CDSs may be expressed only under very specific environmental conditions.

Pseudomonas proteomic studies have also been published in recent years. For example, Toyofuku and co-workers (2012) identified proteins associated with *P. aeruginosa* biofilm formation, in particular with those located into the biofilm extracellular matrix. Another *Pseudomonas* proteomic study was carried out by Manara and co-workers (2012), who analyzed the proteomic changes of *P. putida* in response to cadmium.

In addition, many proteomic studies have been described in relation to hydrocarbon degradation by the genus *Pseudomonas*. One of them is the described by Hemamalini and Khare (2014) about the effect of alkanes such as cyclohexane, octadecane, and dodecane on the outer membrane proteome of *P. aeruginosa* PseA. They observed a differential regulation of different porins (as an example, OprF and OprD were down-regulated while OprE and OprH were up-regulated in the presence of alkanes), which indicates their role in adaptation to solvent exposure. Similarly, Wijte and co-workers (2010) described the proteomic changes suffered by *P. putida* S12 to different toluene concentrations. Particularly, they described a significant increase in energy metabolism proteins, as well as the solvent efflux pumps such as SrpABC, explaining its solvent tolerance. Simon and co-workers (2015) characterized the molecular response of *P. putida* KT2440 to n-butanol by quantitative proteomics. As a result, n-butanol degradation pathway in this strain was elucidated. Additionally, they described an abundance increase of enzymes involved in the TCA cycle and the solvent efflux proteins TtgC and TolC. They associated the up-regulation of these transport proteins with the necessity to decrease the concentration of toxic degradation products like butyraldehyde. In addition, Kasahara and co-workers (2012) carried out a proteomic analysis of *P. putida* F1 growing with different aromatic hydrocarbons as the sole carbon source (such as toluene, ethylbenzene, benzene, cymene, and cumate). In all cases the key enzymes involved in the different degradation pathways were identified, as well as regulators and transporters involved. However, most of the up-regulated proteins detected had no direct relation to the degradation pathway of the tested substrates, although might be related to secondary or tertiary metabolism. Finally, Li and co-workers (2015) described the proteomic changes of *P. putida* ND6 caused by the use of naphthalene as the sole carbon source in comparison to glucose. As a result, they were able to detect the expression of 74 different proteins, 10 of which showed a statistically significant abundance increase, and 21 decreased. They showed the up-regulation of proteins involved in heat shock and universal stress response (such as GroEL), naphthalene degradation (CatA and CatB), cell envelope synthesis (OmpH1 and PhaK-like), and motility (FliC). In contrast, the expression levels of proteins involved in protein and fatty acid synthesis, carbon compound, nucleotide, amino acid metabolism, and small molecule transport were down-regulated in naphthalene cultures. However, they could not observe the up-regulation of either proteins involved in the upper naphthalene degradation pathway, or transport proteins which might play a role in the transport of naphthalene or their catabolic intermediates.

Nowadays, no proteomic study of *P. stutzeri* in response to aromatic hydrocarbons has been published. Therefore, we believe that this sort of study would provide information about the response of this species to the presence of aromatic compounds. In particular, we are now interested in accessory proteins that play a role in this metabolism.

OBJECTIVES

P. stutzeri strains define a coherent species, whose members have been grouped into different genomic groups (genomovars) according to MLSA, siderotyping, whole-cell MALDI-TOF mass spectrometry, and whole-genome comparison studies. Nevertheless, not all *P. stutzeri* strains have been classified into genomovars, and the affiliation of some genomovars into this species is not clear. The new shotgun genome sequencing methods allows us to have many *P. stutzeri* genomes at our disposal. This fact makes possible to perform a whole genomic comparison of this species. A genomic analysis of this species will provide new information that will clarify the phylogenomic relation between the different strains. We hypothesized that a comparative genomic analysis of *P. stutzeri* strains will verify the coherence of the *P. stutzeri* species. We believe that this analysis will show genomic differences between strains classified into different genomovars, which will confirm the structure of this species.

On the other hand, a few *P. stutzeri* strains able to degrade aromatic compounds have been isolated. Some of their catabolic pathways have been previously characterized, as the naphthalene degradation pathway of strain AN10 and the β-ketoadipate pathway in strain A1501. But any genomic analysis of the potential catabolic capabilities of this species has been performed yet. Moreover, no accessory proteins such as efflux pumps and intake membrane transporters involved in metabolism of aromatic compounds have been described in *P. stuzeri*. We hypothesized that a comparative genomic analysis of *P. stutzeri* genomes will confirm the presence of aromatic hydrocarbon degradation genes in all strains, and that they will be generally conserved in all the studied strains. Additionally, a proteomic study of *P. stutzeri* exposed to aromatic hydrocarbons will provide information about its response to these compounds.

Therefore, we have defined three specific objectives for the present study:

1. Perform a comparative analysis of *P. stutzeri* genomes and analyze the *intra*-species phylogenomic relationships.

2. Establish the catabolic potential of *P. stutzeri* for the degradation of aromatic hydrocarbon compounds based on genome analysis.

3. Identify the proteins expressed by *P. stutzeri* aromatic hydrocarbon-degrading strains after exposure to naphthalene, as model compound, with special focus on catabolic as well as accessory proteins.

MATERIALS AND METHODS

1. Genomic analysis

All *P. stutzeri* sequenced genomes available at the NCBI database on November 2014 were selected to perform the genomic comparison described in Chapter 1. Moreover, we decided to include *Pseudomonas balearica* SP1402^T, the type strain of the closest species to *P. stutzeri*. Seven of these genomes sequences have been generated by the Group of Microbiology at the UIB prior to or during this thesis. The genome sequence of strain AN10, 19SMN4, and ST27MN3 (not available at NCBI database yet) have been generated in this thesis. All these studied strains are listed in Table M1, and their more relevant characteristics are shown in Table 1.1.

Cells of *P. stutzeri* AN10, ST27MN3 and 19SMN4 from LB cultures were collected by centrifugation at 13,400 x g. Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega), following the manufacturer's instructions. Genome sequence of *P. stutzeri* AN10 was generated using Celera Assembler vs. 7.0 (Miller *et al*., 2008) from 322,269 8-kb mate-pair reads obtained with the 454 GS FLX Titanium system at LifeSequencing (Spain). Gaps were closed using the assembly information obtained with Velvet program (Zerbino and Birney, 2008) from 4,285,714 500-bp paired-end reads generated with Illumina HiSeq 2000 platform at LifeSequencing (Spain). The order of contigs in the scaffold was confirmed by PCR with specific primers. A single scaffold of 4,709,064-bp in length was obtained for strain AN10 (Brunet-Galmés *et al*., 2012). Genome sequence of *P. stutzeri* 19SMN4 was generated using Newbler v. 2.9 (Roche) from 433,843 8-kb mate-pair reads obtained with the 454 GS FLX Titanium system. Gaps were closed using the assembly information obtained with Celera Assembler vs. 7.0 (Miller *et al*., 2008) from 5,246,472 500-bp paired-end reads generated with Illumina HiSeq 2000 platform. Both sets of reads were obtained at LifeSequencing (Spain). In the case of strain 19SMN4, two scaffolds were obtained, a chromosome of 4,725,662-bp in length, and a plasmid of 107,733-bp in length. Genome sequence of ST27MN3 was generated using the Velvet program (Zerbino and Birney, 2008) with a Kmer of 57 from 4,000,000 100-bp paired-end reads generated with Illumina HiSeq 2000 platform at Macrogen (Korea), previously filtered with Trimmomatic (Bolger *et al.,* 2014). The obtained genome sequence of ST27MN3 included 29 contigs (> 500-bp) with a total length of 4,665,023-bp. Genome annotation of AN10 and 19SMN4 were done using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP, http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html). Coding DNA sequences (CDSs) of ST27MN3 were obtained with the PROKKA pipeline (Seemann, 2014).

The annotated CDSs of the eighteen *P. stutzeri* genomes and the genome of *P. balearica* were classified within the different functional categories using the KEGG Automatic Annotation Server (KAAS) (http://www.genome.jp/tools/kaas) described by Moriya *et al*. (2007). For this annotation, a set of 40 genomes of aromatic hydrocarbon degraders were used as models (see Table M2). When it was necessary this classification was refined by local BLAST (Johnson *et al*., 2008) against the NCBI non-redundant database, using a minimum E-value of 1E-5 for positive identification. Additionally, the amino acid sequences of some particular CDSs were also analyzed with the Conserved Domain Database (CDD) NCBI tool (Marchler-Bauer *et al.*, 2015).

The presence of CDSs involved in seven different *P. stutzeri* characteristics (catalase, cytochrome *c* oxidase, arginine dihydrolase, glycogen hydrolase, the flagellum, nitrogen fixation, and

Table M1. *P. stutzeri* and *P. balearica* strains used in the genomic comparison study.

Strain	Source	References[a]
P. stutzeri CGMCC 1.1803^T (ATCC 17588^T)	Human spinal fluid	Chen *et al*., 2011
P. stutzeri A1501	Rice roots inoculated with strain A15 in a rice paddy soil in China	Li *et al*., 2010
P. stutzeri B1SMN1	Lagooning wastewater treatment plant in Menorca, enriched with 2-methylnaphtalene	Busquets *et al*., 2013
P. stutzeri CMT.9.A (DSM 4166)	Rhizosphere of a *Sorghum nutans* cultivar	Yu *et al*., 2011
P. stutzeri T13	Activated sludge from a municipal wastewater treatment plant	Li *et al*., 2012
P. stutzeri XLDN-R (CCTCC AB 2012149)	Soil sample enriched with carbazole	Liu *et al*., 2012
P. stutzeri ZoBell (CCUG 16156)	Marine sample from the Pacific Ocean (California)	Peña *et al*., 2012
P. stutzeri AN10 (CCUG 29243)	Sediments of the West Mediterranean Sea, enriched with naphthalene	Brunet-Galmés *et al*., 2012
P. stutzeri 19SMN4 (DSM 6084)	Marine sediments of Barcelona coast, enriched with 2-methylnaphtalene	Brunet-Galmés *et al*., this thesis
P. stutzeri ST27MN3	Sediments of the West Mediterranean Sea, enriched with 2-methylnaphtalene	Brunet-Galmés *et al*., this thesis
P. stutzeri JM300 (DSM 10701)	Anaerobic enrichment from soil with succinate and nitrous oxide	Busquets *et al*., 2012
P. stutzeri 28a24	Soil near Tel Aviv airport	Smith *et al*., 2014
P. stutzeri NF13	Deep sea hydrothermal vent in the Galapagos rift	Peña *et al*., 2013
P. stutzeri KOS6	Chemical hydrocarbon sludge in Kazan	Grigoryeva *et al*., 2013
P. stutzeri MF28	Oyster mantle fluid from polluted water	Chauhan *et al*., 2013
P. stutzeri RCH2	Chromium-contaminated aquifer	Lucas *et al.* unpublished
P. stutzeri SDM-LAC	Soil sample enriched with DL-lactate as the carbon source	Jiang *et al*., 2012
P. stutzeri TS44	Arsenic-contaminated soil Huangshi metal mines	Li *et al*., 2012
P. balearica SP1402^T (DSM 6083^T)	Wastewater	Bennasar-Figueras *et al*., 2016

a: References of the Genome Announcements.

Table M2. List of bacteria whose genomes were used as models for the KEGG annotation.

Organism	Code	Metabolic interest	References
Achromobacter xylosoxidans A8	axy	(Halo)aromatic compounds degradation	Jencova *et al*., 2008
Acidovorax sp. JS42	ajs	2-Nitrotoluene and 3-nitrotoluene degradation	Mahan *et al*., 2015
Agrobacterium radiobacter K84	ara	Agrocinopines degradation	Hayman and Farrand, 1990
Alcanivorax borkumensis SK2	abo	Petroleum oil hydrocarbons degradation	Golyshin *et al*., 2003
"Aromatoleum aromaticum" EbN1	eba	Petroleum hydrocarbons, phenolic solvents, and 3-phenylpropanoids degradation	Trautwein *et al*., 2012
Arthrobacter aurescens TC1	aau	s-Triazine compounds degradation	Strong *et al*., 2002
Arthrobacter chlorophenolicus A6	ach	Phenolic compounds degradation	Unell *et al*., 2008
Burkholderia multivorans ATCC 17616	bmj	Anthranilate degradation	Nishiyama *et al*., 2012
Burkholderia xenovorans LB400	bxe	Polychlorinated biphenyls degradation	Tehrani *et al*., 2014
Candidatus Desulfococcus oleovorans Hxd3	dol	Alkanes degradation	So and Young, 1999
Cupriavidus metallidurans CH34	rme	Metal resistance	Nies *et al*., 2006
Dechloromonas aromatica RCB	dar	Benzene degradation	Salinero *et al*., 2009
Delftia acidovorans SPH-1	dac	Taurocholate degradation	Rösch *et al*., 2008
Desulfatibacillum alkenivorans AK-01	dal	Alkanes degradation	Callaghan *et al*., 2012
Dinoroseobacter shibae DFL 12	dsh	Aromatic compounds degradation	Moran *et al*., 2007
Escherichia coli K-12 MG1655	eco	General model strain	Blattner *et al*., 1997
Jannaschia sp. CCS1	jan	Aromatic compounds degradation	Moran *et al*., 2007
Methanolacinia petrolearia DSM 11571	mpi	Methane production	Göker *et al*., 2014
Methylibium petroleiphilum PM1	mpt	Methyl tert-butyl ether degradation	Chen *et al*., 2008
Nostoc sp. PCC 7120	ana	Carotenoids degradation	Marasco *et al*., 2006
Novosphingobium aromaticivorans DSM 12444	nar	p-cresol, dibenzothiophene, naphthalene, toluene, and xylene degradation	Fredrickson *et al*., 1991
Oligotropha carboxidovorans OM5	oca	Carbon monoxide or hydrogen and carbon dioxide assimilation	Paul *et al*., 2008
Petrotoga mobilis SJ95	pmo	Xylan degradation	Lien *et al*., 1998
Polaromonas naphthalenivorans CJ2	pna	Naphthalene degradation	Jeon *et al*., 2008
Polaromonas sp. JS666	pol	cis-1,2-dichloroethene degradation	Mattes *et al*., 2008
Pseudomonas putida F1	ppf	Toluene degradation	Zylstra *et al*., 1988
Pseudomonas putida GB-1	ppg	Manganese oxidation	Geszvain and Tebo, 2010
Pseudomonas putida KT2440	ppu	Aromatic compounds degradation	Jiménez *et al*., 2002
Pseudomonas putida W619	ppw	Trichloroethylene degradation	Weyens *et al*., 2010
Ralstonia eutropha H16	reh	Lignocellulose degradation	Volodina *et al*., 2015
Ralstonia eutropha JMP134	reu	Aromatic and alicyclic compounds degradation	Lykidis *et al*., 2010
Roseobacter denitrificans OCh 114	rde	Aromatic compounds degradation	Moran *et al*., 2007
Ruegeria pomeroyi DSS-3	sil	Aromatic compounds degradation	Moran *et al*., 2007
Ruegeria sp. TM1040	sit	Aromatic compounds degradation	Moran *et al*., 2007
Saccharophagus degradans 2-40^{T}	sde	Cellulose degradation	Taylor *et al*., 2006
Sphingomonas wittichii RW1	swi	Polychlorinated dibenzo-p-dioxins degradation	Nam *et al*., 2005
Thermotoga naphthophila RKU-10	tnp	Naphthalene degradation	Takahata *et al*., 2001
Variovorax paradoxus S110	vap	Aromatic sulfonates degradation	Satola *et al*., 2013
Xylanimonas cellulosilytica DSM 15894	xce	Cellulose and xylan degradation	Foster *et al*., 2010

denitrification) described by Lalucat and co-workers (2006) was confirmed. For this, the NCBI and KEGG annotations were used.

In order to compare all the annotated *P. stutzeri* and *P. balearica* CDSs, we clustered them according to their homology using the CD-HIT server (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi) (Huang *et al*., 2010). One of the most used criterion for grouping protein sequences sharing the same biological function is that they have at least 50 % of amino acid identity in at least a 50 % of the sequence (Lukjancenko *et al*., 2010, 2012; Vesth *et al*., 2010; Jacobsen *et al*., 2011; Karlsson *et al*., 2011; Leekitcharoenphon *et al*., 2012). Therefore, and using this criterion, we defined the *P. stutzeri* pan-proteome (all COGs –cluster of orthologous groups– obtained using this criterion) and the core-proteome (COGs which presented at least one CDS of each *P. stutzeri* strain). For the definition of the pan- and the core-proteome of *P. stutzeri* we did not consider the genome *P. balearica* SP1402^T. A more stringent criterion (95 % of amino acid identity in at least a 90 % of the sequence) was used to select a subset of the core-proteome for the phylogenomic analysis of *P. stutzeri* and for the analysis of transposases (TnpAs) shown below.

Putative TnpAs were identified by two different approaches: using the published NCBI annotation; and using BLAST with a database generated in this study of TnpAs from 21 different IS-families (see Supplementary Table S1), using an E-value lower than 1E-10. TnpAs classification within IS-families was done using the IS-Finder database (https://www-is.biotoul.fr/) (Siguier *et al*., 2006).

A phylogenetic analysis was performed using the shared COGs (with at least a 95 % of identity in at least a 90 % of the sequence) present in single copy in all the genomes. Each of these proteins were concatenated and later aligned with ClustalX 2.0.12 software (Thompson *et al*., 1997) using default parameters. Conserved regions of resulting alignment were later filtered with the Gblocks web server (Talavera and Castresana, 2007) (https://molevol.cmima.csic.es/castresana/Gblocks_server.html). Finally, using MEGA 6.06 program (Tamura *et al*., 2013), distance matrixes were calculated with the Jones-Taylor-Thornton (JTT) model of evolution (Jones *et al*., 1992). Phylogenetic trees were constructed with Neighbor-Joining (Saitou and Nei, 1987) using a bootstrap of 100 replicates. Phylogenetic analysis of 8 different proteins involved in aromatic hydrocarbons degradation (NahAc, NahF, NahD, NahR, NahG, NahH, NahI, and NahM) were also performed using the same approach.

Finally, BLAST average of nucleotide identity (ANIb) values between the 19 *P. stutzeri* and *P. balearica* genomes were calculated using the JSpecies software, with the default parameters (Richter and Rosselló-Móra, 2009).

2. Culture conditions

All strains were grown at 30 °C and 180 rpm in Luria-Bertani (LB) (Sambrook and Russell, 2001) or in minimal medium (MMB) (Aragno and Schlegel, 1981). Carbon and energy sources used with MMB and concentrations (final), if not explicitly-indicated, were as follows: benzoate (5 mM), 4-hydroxyphenylpyruvate (13.9 mM), homogentisate (10 mM), protocatechuate (5 mM), salicylate

(3 mM), or succinate (30 mM). Since naphthalene is insoluble in water, we dissolved it in acetone at 25 % (wt/vol) and sterilized the solution by filtration through 0.22 µm Teflon filters (Millipore). Then, this solution was added aseptically to sterile culture flasks (1 ml for a culture volume of 100 ml), and acetone was let to evaporate in a biosafety cabinet, leading to the formation of naphthalene crystals at the bottom of the flask. When necessary, the antibiotics gentamicin (20 µg/ml) and/or ampicillin (75 µg/ml) were added to LB or MMB.

3. Selection of a benzoate-degrader derivative of strain *P. stutzeri* AN10

In order to obtain a benzoate-degrader derivative of AN10, serial cultures of AN0 in MMB with 30 mM succinate and increasing benzoate concentrations (from 1 to 5 mM, at increases of 1 mM in each step) were performed. After reaching 5 mM benzoate in the presence of 30 mM succinate, a stepwise decrease of succinate was done (cultures at 22.5, 15, 7.5, 3.5 and 0 mM succinate). Thus, the final culture contained only 5 mM benzoate. In a second approach, MMB medium supplemented with benzoate 3 mM was inoculated with an overnight culture of strain AN10 in LB broth (1/50 inoculum). From this point, successive cultures at increasing benzoate concentrations from 4 to 20 mM were performed (from 4 mM to 14 mM at increases of 1 mM benzoate in each step, then transfer to 16 mM, and finally 20 mM benzoate). The subcultures were done when cells reached the late-exponential phase (0.5–0.8 OD_{600}). An aliquot from each of these successive cultures at increasing benzoate concentrations was plated in LB and MMB plates supplemented with benzoate. Four isolates from MMB plates supplemented with benzoate were randomly selected to be stored at –80 °C.

For the phylogenetic identification of all AN10 derivatives obtained in these cultures, 16S rRNA and *rpoD* genes were amplified by PCR using a recombinant *Taq* DNA polymerase from Invitrogen. The PCR mix used was composed of 1X PCR Buffer supplemented with $MgCl_2$ 1.5 mM, dNTPs mix 1 mM each, primers 0.4 µM each, and 3 U of Taq DNA polymerase per 100 µl of reaction. For amplification of 16S rRNA genes we used primers F27 (5'-AGAGTTTGATCMTGGCTCAG-3') and R1492 (5'TACGGYTACCTTGTTACGACTT-3') (Lane, 1991). *rpoD* genes were amplified using primers PsEG30F (5'-ATYGAAATCGCCAARCG-3') and EG790R (5'-CGGTTGATKTCCTTGA-3'), previously described by Mulet and co-workers (2010). In both cases, we used an annealing temperature of 55 °C, an extension time of two minutes, and a total of 30 cycles. The amplified products were purified with Multiscreen PCR 96-well plates (Millipore) and sequenced using BigDye terminator cycle sequencing kit version 3.1 (Applied Biosystems) in an automatic sequence analyzer (3130 genetic analyzer; Applied Biosystems) at the Scientific-Technical services of the university.

In order to select the AN10 derivative which presented the best growth with benzoate, we grew 4 isolates from three different sub-cultures (4 mM, 16 mM, and 20 mM) in MMB supplemented with different benzoate concentrations (5 mM, 10 mM, 15 mM, 20 mM, and 25 mM). Strain AN10 and the AN10 derivative able to use benzoate as carbon source more efficiently (BZ4D) were grown in duplicate in MMB supplemented different carbon sources (succinate 0.5 % wt/vol, salicylate 3 mM, naphthalene, and benzoate 3 mM).

The genome of the AN10 derivative BZ4D was extracted from cells grown in LB broth using the Wizard® Genomic DNA Purification Kit, following the manufacturer's instructions; and later sequenced using the Ilumina technology at GATC Biotech (Germany), obtaining 6.86 million of reads. The genome assembly was performed using the GS De Novo Assembler (Roche), calculated with 4 million reads. As a result, 48 contigs with a 4,678,705-bp in total length were obtained. In order to compare the BZ4D genome with the AN10 genome, three different strategies were performed: (1) a comparison of 4 million BZ4D reads with the AN10 genome using the GS Reference Mapper (Roche); (2) a comparison of BZ4D contigs with the AN10 genome using BLASTn; and (3) a comparison of annotated CDSs from both strains using CD-HIT.

BenR CDS from BZ4D was compared by local BLAST (Johnson *et al*., 2008) with the BenR sequences identified in the 19 *P. stutzeri* and *P. balearica* genomes listed in Table M1. The CatA sequence from BZ4D was aligned using ClustalX 2.0.12 software (Thompson *et al*., 1997) with the CatA sequences identified in the 19 *P. stutzeri* and *P. balearica* genomes listed on Table M1. We also performed a BLASTp comparison of AN10 CatA against the non-redundant NCBI database. The amino acid sequences of BenR and CatA were also analyzed with the Conserved Domain Database (CDD) NCBI tool (Marchler-Bauer *et al.*, 2015).

4. Proteomic studies

4.1. Naphthalene experiment

With the aim of comparing the proteome of different strains exposed to aromatic hydrocarbons, we selected five strains isolated by the Group of Microbiology at the UIB that were characterized as naphthalene degraders: B1SMN1, 19SMN4, AN10, ST27MN3 and SP1402^T. All these strains were grown in triplicate at 30 °C and 180 rpm in MMB supplemented with 3 different carbon and energy sources: (1) 30 mM succinate, (2) 30 mM succinate with a pulse of salicylate 3 mM during the late-exponential phase (0.5–0.8 OD_{600}) for 4 hours, and (3) naphthalene vapor (to avoid problems during proteomic analysis). In these later cultures, naphthalene crystals were added inside the cotton lids of each flask, and the flasks were placed in a closed container to avoid the loss of naphthalene vapors. Cells were collected during the late-exponential phase (0.5–0.8 OD_{600}) by centrifugation at 13,400 x g for 3 minutes, obtaining from 50 to 100 mg of cells per sample, and stored at –80 °C until processed.

Samples were then resuspended in 2 volumes (2 µl per 1 mg of biomass) of lithium dodecyl sulfate-β-mercaptoethanol sample buffer (Invitrogen) and the cells were lysed by 3 cycles of incubations at 95 °C for 5 min followed by 5 min of sonication (Branson 1210 Ultrasonic Cleaner). Cell lysates were directly loaded onto 10 % Tris-Bis NuPAGE gels (Invitrogen) and the gels were run at 200 V for 10 min, in order to perform a short in-gel migration. Gels were stained with Coomassie Blue Safe stain (Invitrogen). Then the gel area containing the proteome fraction was cut and later washed twice with ammonium bicarbonate 50 mM in ethanol 50% vol/vol for 20 minutes at 50 °C. Gel fractions were then dehydrated with absolute ethanol for 5 minutes at 55 °C and disulfide bonds of the in-gel proteins were reduced with DL-dithiothreitol (DTT) 10 mM in ammonium bicarbonate 50 mM for 40 minutes at 56 °C. At that point, cysteine residues were alkylated with iodoacetamide 55 mM in ammonium bicarbonate 50 mM for 30 minutes at room

temperature in the dark. Gel fractions were then washed and dehydrated as described above, and afterwards proteins were digested in-gel with 100 ng of trypsin (Promega) in ammonium bicarbonate 50 mM overnight at 37 °C. In order to extract the peptides, the gel fractions were submerged in formic acid 5 % vol/vol in acetonitrile 50 % vol/vol and they were sonicated three times for 10 minutes (Fisherbrand FB15062 ultrasonic bath). Gel fractions were discarded and supernatants (containing the peptides) were later incubated overnight at 40 °C to evaporate the solvent. Eventually, peptides were resuspended in formic acid 1 % vol/vol in acetonitrile 2 % vol/vol. An aliquot containing 20 µl of peptide solution from each sample was analyzed by nanoLC-ESI-MS/MS using the Ultimate 3000/Orbitrap Fusion instrumentation (Thermo Scientific) using a 120 minute LC separation in a 25 cm column (Dionex). This analysis was performed at the School of Life Sciences of the University of Warwick (United Kingdom).

The MS/MS spectra obtained were aligned with the software Progenesis QI for Proteomics (version 2.0) from Nonlinear Dynamics. Then, peptides were identified using Mascot (version 2.5.0) from Matrix Science, based on the predicted CDSs of each genome. A fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 20 ppm were used. Peptide identifications were accepted if they could be established at greater than 95 % probability. Protein identifications were accepted if they could be established at greater than 99 % probability and presented at least 2 of its peptides identified. Data from each strain was normalized with the software Progenesis QI for Proteomics, obtaining a normalized abundance value for each protein. Using the same software we performed a statistical comparison between the proteomic data obtained from each of the growth condition. The criterion to define a variation as statistically significant was a p-value below 0.05 and a minimum fold change in normalized abundance of 2. In order to be able to compare abundance values from different strains, normalized abundance values were standardized according to the total normalized abundance obtained for each cultures, obtaining a Normalized Abundance Factor (NAF).

Principal component analysis (PCA) of NAF values was performed, in order to study the variability of the different samples, using the software PAST (Hammer *et al*., 2001). First of all, the detected CDSs were grouped according to their homology with CD-HIT server (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi; Huang *et al*., 2010). The criterion used in this case was at least a 75 % of identity in at least a 90 % of the sequence. Once the detected CDSs were grouped, we constructed a matrix with the total sum of NAF values of each cluster in each proteome, which was used as input for the analysis.

All detected proteins were classified within the different functional categories using the KEGG Automatic Annotation Server (KAAS) (http://www.genome.jp/tools/kaas), as mentioned in section 1. The sum of NAF values of all the proteins classified in the different biological functions obtained from cultures with different carbon sources were compared.

Finally, we identified several proteins that presumably formed part of different transport systems. These were classified using the Transporter Classification DataBase (TCDB) into different families (http://www.tcdb.org/) (Saier *et al*., 2014). Transport systems were assigned to the family with E-values lower than 1E-5.

4.2 Experiment with strain BZ4D

A comparison of BZ4D (an AN10 derivative able to use benzoate as carbon and energy source) and AN10 proteomes was performed in order to clarify which proteins were involved in benzoate degradation. AN10 was grown in triplicate at 30 °C and 180 rpm in MMB supplemented with succinate 0.5 % wt/vol. BZ4D was grown in triplicate in the same conditions, but in this case the medium was supplemented either with 30 mM succinate or with 10 mM benzoate. Cells were collected during the late-exponential phase (0.6-0.8 OD_{600}), by 3 cycles of centrifugation at 13,400 x g for 5 min followed by washes of the pellet with Tris-HCl buffer 20 mM pH 7. At that point, 130 mg of cells per each sample were stored at –20 °C until processed. They were then resuspended in 5 volumes (5 µl per 1 mg of biomass) of 100 mM sodium phosphate buffer at pH 8.2 containing Complete Protease Inhibitor (Roche). Cells were lysed at 4 °C by sonication applying a 40 J dose with amplitude of vibration of 20 % and pulses of 5 seconds followed by resting intervals of 5 seconds. Lysates were centrifuged for 20 min at 14,000 x g at 4 °C to remove cellular debris. Protein content from the soluble fractions was quantified using the Bio-Rad protein assay kit (BioRad), and they were further concentrated by trichloroacetic acid precipitation as described previously (Christie-Oleza and Armengaud, 2010). Resulting pellets were resuspended in lithium dodecyl sulphate-β-mercaptoethanol (LDS) protein gel sample buffer (Invitrogen) and incubated at 99 °C for 5 min. Protein mixtures were loaded on 10 % Tris-Bis NuPAGE gel (Invitrogen) for short SDS-PAGE electrophoresis migration. Gels were then stained with Coomassie Blue Safe stain (Invitrogen). At that point, three gel bands per sample were cut from the top to the bottom of the gel, in order to resolve proteins of different molecular weight. They were processed for in-gel proteolysis with trypsin (Roche) following the ProteaseMax protocol (Promega) as previously described (Clair *et al*., 2010). NanoLC-MS/MS experiments were performed using a LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings) using conditions as those previously described (de Groot *et al*., 2009).

Peak lists of detected peptides were generated with the Mascot Daemon software (version 2.2) from Matrix Science using the extract_msn.exe data import filter (ThermoFisher) from the Xcalibur FT package (version 2.0.7, ThermoFisher). MS/MS spectra were searched with the Mascot 2.2 software (Matrix Science) against a database containing all the annotated CDSs from AN10 and BZ4D. Search parameters were as follows: tryptic peptides with a maximum of 2 miss-cleavages, mass tolerances of 5 ppm for the parent ion and 0.5 Da for the MS/MS, fixed modification for carboxyamidomethylated cysteine and variable modification for oxidized methionine. Mascot results were parsed using the IRMa 1.26.1 software (Dupierris *et al*., 2009) filtering assigned peptides with a p-value below 0.05. A protein was considered valid when at least 2 different peptides above the significance threshold were detected. Protein semi-quantitation by spectral abundance was considered as previously described (Liu *et al*., 2004). Statistical comparison between the proteomic data obtained from each of the growth condition was carried out with the PatternLab program (Carvalho *et al*., 2012). Data were normalized attending to the length of each protein and the total spectral counts detected, calculating the normalized spectral abundance factor (NSAF) (Zybailov *et al*., 2006). Spectral counts were also analyzed using the ACFold method with a p-value below 0.05 and a minimum fold of 2 with the PatternLab program (Carvalho *et al*., 2012). This analysis was performed at the Life Science

Division from the French Alternative Energies and Atomic Energy Commission, CEA Marcoule (France).

PCA of NSAF values was performed, in order to study the variability of the different samples using the software PAST (Hammer *et al*., 2001). For this, we constructed a matrix with the total NSAF values of each protein in each proteome, which was used as input for the analysis.

5. Gene cloning and complementation

Two different genes predicted in the genome of strain BZ4D (*benR* and *catA*) were identified putatively as responsible for the benzoate degradation ability acquisition of this strain. Therefore, in order to verify which of them were determinant for this metabolic property, we decided to introduce them in *P. stutzeri* AN10 wild type strain. For this purpose, *benR* and *catA* genes of BZ4D were cloned in pBBR-type replicative plasmids (Kovach *et al*., 1995). In order to be able to distinguish both constructions, two different vectors were used: pBBRMCS-4, with an ampicillin gene resistance; and pBBRMCS-5, with a gentamicin gene resistance (Figure M1 A).

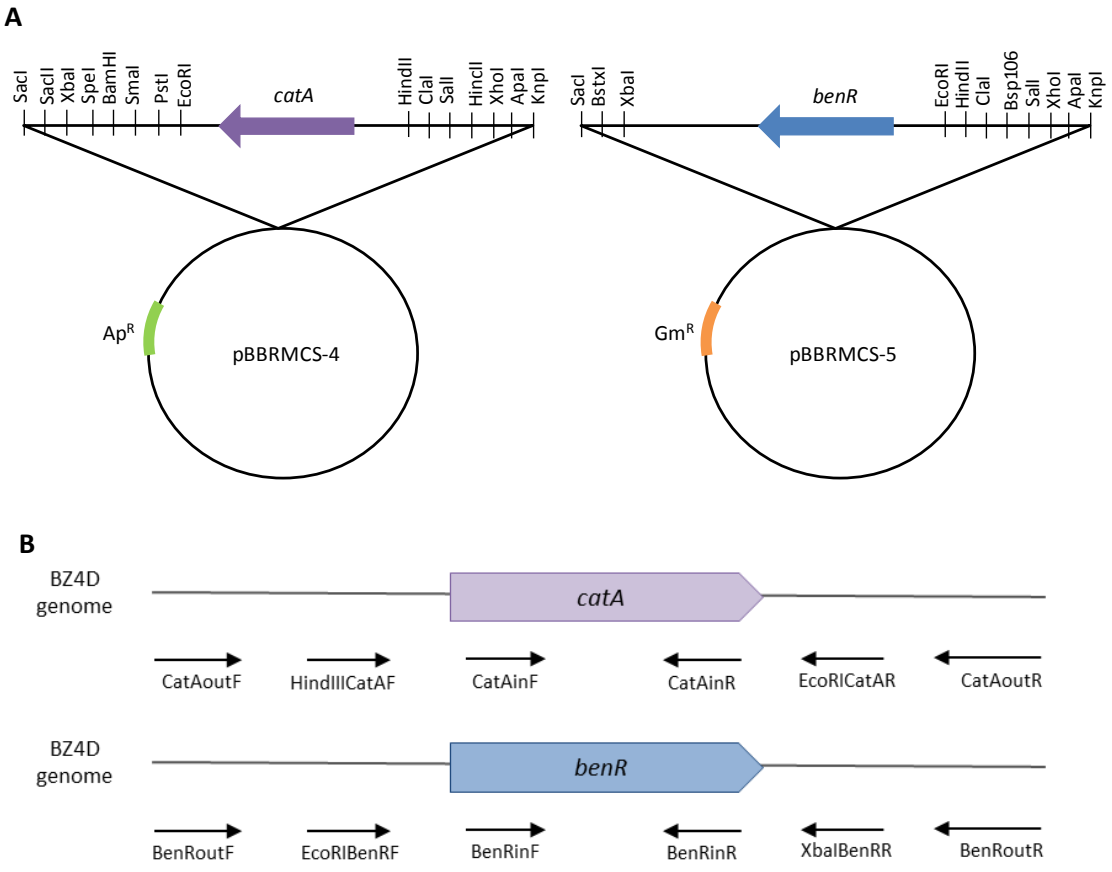

Figure M1. pCatA and pBenR constructions used to complement AN10 (A) and primer localization respect from *catA* and *benR* genes (B).

Genes were amplified from genomic DNA of BZ4D by PCR using the proof-reading polymerase included in the KAPA HiFi PCR Kit (Kapa Biosystems). The PCR mix used was composed of 1X KAPA HiFi Buffer, dNTPs mix 0.3 mM each, primers 0.3 µM each, and 2 U of KAPA HiFi DNA

Polymerase per 100 µl of reaction. Primers EcoRIBenRF and XbaIBenRR for *benR* amplification, and HindIIICatAF and EcoRICatAR for *catA* amplification were designed in this study (see Table M3 and Figure M1 B) and purchased from Invitrogen. All of them presented a thymine tail at the 5'-end followed by a restriction site sequence, which allowed us to insert the amplification product in a specific direction in the plasmids. For *benR* amplification, we performed two first cycles with an annealing temperature of 58 °C and an extension time of 130 seconds; and later 28 cycles with an annealing temperature of 70 °C and an extension time of 130 seconds. For *catA* amplification, we performed two first cycles with an annealing temperature of 60 °C and an extension time of 120 seconds; and later 28 cycles with an annealing temperature of 72 °C and an extension time of 120 seconds. Both amplification products were then purified using PureLink® PCR Purification Kit (Invitrogen), following manufacturer's instructions. Plasmid DNA extractions of the replicative cloning vectors pBBRMCS-4 and pBBRMCS-5 were performed by alkaline lysis using the QIAprep Spin Miniprep Kit (Qiagen), and following manufacturer's instructions.

Table M3. Primers designed in this study for *benR* and *catA* complementation. The restriction site sequences are underlined.

Primer	Sequence (5'-3')	Purpose[a]
pBBRMCSF	aacgacggccagtgagcg	Amplification of the multi cloning site of pBBR-type vectors
pBBRMCSR	ttgtgtggaattgtgagcgg	
EcoRIBenRF	ttttgaattcgcacacgagcggtaaatgg	Cloning of *benR*
XbaIBenRR	tttttctagacgtgatgaaggctacggc	
BenRoutF	caatgggtgctcagtctcg	Amplification of *benR* outside the cloning point
BenRoutR	ccagttgccctcgaagatg	
BenRinF	ctcctatgctcgaacgaagcg	Amplification of *benR* inside the cloning point
BenRinR	gacagattgacgcaggcacg	
HindIIICatAF	ttttaagcttgcatccgtcctcgattcgtg	Cloning of *catA*
EcoRICatAR	ttttgaattcagtacgtcttggggcaaccc	
CatAoutF	ggctgacgagaaggaactgg	Amplification of *catA* outside the cloning point
CatAoutR	tcattcagcgtgatgtcggg	
CatAinF	agttctggaaagcggtggac	Amplification of *catA* inside the cloning point
CatAinR	tcagtggctcgacaacacag	

a: See figure M1 B for annealing positions.

Amplified gene products and plasmids (approximately 120 ng each) were digested for 30 minutes with 50 U of the corresponding restriction enzymes: pBBRMCS-4 and *catA* amplification product were both digested with FastDigest® HindIII and FastDigest® EcoRI (ThermoFisher); and pBBRMCS-5 and *benR* amplification product were digested with FastDigest® XbaI and FastDigest® EcoRI (ThermoFisher), following manufacturer's instructions. After digestion, all products were purified using the PureLink® PCR Purification Kit (Invitrogen), and then ligated using T4 DNA Ligase (Invitrogen) with a ratio 1:3 (vector:PCR-product volumes), following manufacturer's instructions. Resulting constructions described above, designated as pCatA (*catA* with pBBRMCS-4) and pBenR (*benR* with pBBRMCS-5), are shown in Figure M1 A. Correct insertions of both genes were checked by PCR amplification from a plasmid DNA extraction and subsequent DNA sequencing. PCRs were performed using a recombinant *Taq* DNA polymerase from Invitrogen, using the mixed composition detailed in section 3, with primers pBBRMCSF and pBBRMCSR (see Table M3). We used an annealing temperature of 55 °C, an extension time of

two minutes, and a total of 30 cycles. PCR products were purified and further sequenced using forward and reverse primers, as detailed in section 3.

Competent *E. coli* S.17 λ_{pyr} cells (Herrero *et al.*, 1990) were prepared. For this purpose 30 ml of SOB broth (see Table M4) in a 250-ml flask was inoculated with bacteria from a single colony, and incubated overnight at 30 °C and 180 rpm. Then, 8 ml of the overnight culture were added to a 2-l flask which contained 200 ml of SOB medium, and it was incubated at 30 °C and 180 rpm until an OD_{600} of 0.3 was achieved. The culture was further split in four different 50-ml sterile tubes, and chilled on ice for 15 minutes. Cells were pelleted by centrifugation at 3,000 x g for 15 minutes at 4 °C, and they were completely drained. Then, cell pellets were resuspended in 16 ml of cold transformation buffer 1 (see Table M4) by mild vortexing, and they were incubated on ice for 15 minutes. Cells were then pelleted as before, resuspended in 4 ml of cold transformation buffer 2 (see Table M4), and stored in aliquots at –80 °C.

Table M4. Composition of SOB broth and transformation buffers used in the preparation of competent cells.

SOB medium	Transformation buffer 1 pH 5.8	Transformation buffer 2
Tryptone 20 g/l	RbCl 100 mM	RbCl 10 mM
Yeast extract 5 g/l	$MnCl_2$ $4H_2O$ 60 mM	MOPS 10 mM
NaCl 10 mM	Potassium acetate 30 mM	$CaCl_2$ $2H_2O$ 74.8 mM
KCl 6.7 mM	$CaCl_2$ $2H_2O$ 10.2 mM	Glycerol 150 ml/l
$MgCl_2$ 10 mM	Glycerol 150 ml/l	
$MgSO_4$ 10 mM		

Ligation products were transformed into competent *E. coli* S.17 λ_{pyr} cells by heat shock. For this, competent cells were thawed on ice, and the ligation product was added. Then, cells were subjected to a heat shock at 42 °C for 45 seconds, followed by incubation on ice for 15 minutes. Afterwards 1 ml of LB broth was added and the cells were incubated at 30 °C and 180 rpm for 1 hour. Transformed cells were selected by plating them into LB agar supplemented with ampicillin, Ap (for pCatA) or gentamicin, Gm (for pBenR). The presence of plasmids pCatA or pBenR in transformed *E. coli* S.17 λ_{pyr} cells was confirmed by PCR with the same protocol used to check the correct insertion of *catA* and *benR* genes into the replicative plasmids (described above in this section).

E. coli S.17 λ_{pyr} clones containing pCatA and pBenR constructions were conjugated with AN10 cells. For this, 600 µl of late-exponential-phase cultures (approximately 0.8 OD_{600}) of donor and recipient strains were mixed (cell ratio of 1:1). The mixture was centrifuged for 20 seconds at 13,000 x g and the supernatant was removed. Finally, as Christie-Oleza and co-workers (2009) previously described, cells were collected, spotted onto LB agar plates and incubated at 30 °C overnight. Transconjugant AN10 cells were selected by culturing them into MMB agar plates supplemented with succinate 0.5 % wt/vol and Ap or Gm (for *catA* and *benR* transconjugants, respectively). Eventually, in order to select an AN10 clone containing both constructions simultaneously (pCatA and pBenR) an AN10 clone containing the pBenR construction was then conjugated with an *E. coli* S.17 λ_{pyr} clone containing the pCatA construction. All conjugations were confirmed by amplification and further sequencing of *catA* and/or *benR.* PCRs were performed using a recombinant *Taq* DNA polymerase from Invitrogen, using the mixed

composition detailed in section 3. For the amplification of the plasmidic *catA* gene cloned from BZ4D we used a plasmidic DNA extraction product as DNA template, primers pBBRMCSF and CatAinF (see Table M3 and Figure M1 B), an annealing temperature of 58 °C, an extension time of 90 seconds, and a total of 30 cycles. For the amplification of the plasmidic *benR* gene cloned from BZ4D we used a plasmidic DNA extraction product as DNA template, primers pBBRMCSR and BenRinR (see Table M3 and Figure M1 B), an annealing temperature of 60 °C, an extension time of 90 seconds, and a total of 30 cycles. For the amplification of the genomic *catA* and *benR* genes from AN10 wild type we used a genomic DNA extract as DNA template, and primers CatAoutF and CatAoutR (for *catA*) or primers BenRoutF and BenRoutR for *ben R* (see Table M3 and Figure M1 B). Annealing temperatures were 58 °C for *catA* and 55 °C for *benR*. An extension time of two minutes and a total of 30 cycles were used in both cases. All these PCR products were purified and further sequenced using forward and reverse primers, as detailed in section 3. Growth with benzoate of derivatives resulting from gene complementation and the AN10 derivative BZ4D was tested. For this, the different derivatives were grown in triplicate at 30 °C and 180 rpm in MMB supplemented with 5 mM benzoate as carbon and energy source.

RESULTS AND DISCUSSION

CHAPTER 1: Genomic comparison of *Pseudomonas stutzeri*

1. General characterization of *P. stutzeri* genomes

On November 2014, 18 different *P. stutzeri* genomes were available in the public genome database of the National Center for Biotechnology Information (NCBI). From those, strains AN10 and 19SMN4 were sequenced and later submitted by our research group during this study. The genome of strain ST27MN3 was generated in this thesis (see Materials and Methods for details). In addition, the genome sequence of *Pseudomonas balearica* SP1402^T (formerly belonging to *P. stutzeri* genomovar 6) was also included in the analysis, as it is the closest species to *P. stutzeri* (see Table 1.1). Ten of these genomes were complete and the rest were in draft status.

All these strains were isolated from different environments, as shown in Table M1. We can find clinical specimens (such as ATCC 17588^T), strains associated with plants (A1501 and DSM 4166) or animals (MF28), strains isolated from marine environments (CCUG 16156, AN10, 19SMN4, ST27MN3, and NF13), wastewater (B1SMN1, T13, and SP1402^T), an aquifer (RCH2), and soils (XLDN-R, DSM 10701, 28a24, KOS6, SDM-LAC, and TS44), some of them polluted.

Table 1.1 General genomic characteristics of the 18 *P. stutzeri* strains and *P. balearica* SP1402^T.

Strain	Gv[a, b, c]	Sequencing status	Length (Mb)	mol % GC[d]	CDSs	BioProject[b]
P. stutzeri ATCC 17588^T	1	Complete	4.6	63.9	4218	PRJNA68131
P. stutzeri A1501	1	Complete	4.6	63.9	4127	PRJNA16817
P. stutzeri B1SMN1	1	Draft	5.3	63.4	5104	PRJNA170978
P. stutzeri DSM 4166	1	Complete	4.7	64.0	4301	PRJNA63543
P. stutzeri T13	1	Draft	4.7	63.9	4255	PRJNA170692
P. stutzeri XLDN-R	1	Draft	4.7	63.9	4217	PRJNA168597
P. stutzeri CCUG 16156	2	Draft	4.5	61.4	4126	PRJNA74687
P. stutzeri AN10	3	Complete	4.7	62.7	4299	PRJNA167996
P. stutzeri 19SMN4	4	Complete	4.8	62.2	4371	PRJNA242326
P. stutzeri ST27MN3	4	Draft	4.7	62.4	4300	n.d.
P. stutzeri DSM 10701	8	Complete	4.2	63.2	3815	PRJNA78211
P. stutzeri 28a24	11	Complete	4.7	60.6	4147	PRJNA240232
P. stutzeri NF13	19	Draft	4.7	63.0	4321	PRJNA170977
P. stutzeri KOS6	n.d.	Complete	4.9	62.9	4466	PRJNA171881
P. stutzeri MF28	n.d.	Draft	4.9	62.3	4540	PRJNA202932
P. stutzeri RCH2	n.d.	Complete	4.6	62.5	4264	PRJNA60029
P. stutzeri SDM-LAC	n.d.	Draft	4.2	60.5	3727	PRJNA74741
P. stutzeri TS44	n.d.	Draft	4.3	64.4	4000	PRJNA162447
P. balearica SP1402^T	n.a.	Complete	4.4	64.7	3926	PRJNA305687

a: Gv, genomovar.

b: n.d., not determined.

c: n.a., not applicable.

d: data obtained from NCBI.

The 18 *P. stutzeri* strains used in this study belong to, at least, 7 different genomovars or genomic variants. *P. stutzeri* genomovars are genomic groups of strains, proposed by Rosselló and co-workers (1991) to distinguish genetically different strains with no phenotypic differences. Most of strains belonged to genomovar 1 (six strains); followed by genomovar 4 with two representatives; and one strain from genomovars 2, 3, 8, 11 and 19. The remaining 5 sequenced *P. stutzeri* strains have not been assigned to any genomovar yet (see Table 1.1).

The total length of the 18 *P. stutzeri* genomes varied between 4.2 Mb (DSM 10701) and 5.3 Mb (B1SMN1), with an average of 4.7 $\pm$ 0.3 Mb. This range was higher than the previously described

by Lalucat and co-workers (2006), who described a range of 3.8 to 4.6 Mb in length. These genomes have a mol % GC-content of 62.8 ± 1.1 %, which is consistent, but narrower, with the range described by Lalucat and co-workers (2006), of 60 to 66 %. *P. stutzeri* genomes showed an average of 4,255 ± 292 coding sequences (CDSs) per genome. Strain B1SMN1 was the one with a greater number of CDSs, with 5,104 plausible proteins; and the strain with fewer identified CDSs was SDM-LAC, with 3,727 CDSs. All those CDSs are listed in Supplementary Table S2, and their classification into KEGG biological functions is shown in Supplementary Table S3. According to the information of the complete genomes available at NCBI, only four different plasmids have been described in *P. stutzeri* strains: one in 19SMN4 genome (pLIB119, 107 kb) and three in RCH2 genome (pPSEST01, 13 kb; pPSEST02, 10 kb; and pPSEST03, 3 kb). In the genome description of strain B1SMN1 (genome in draft status) two plasmids were described, corresponding to contigs 29 (44 kb) and 20 (56 kb) (Busquets *et al*., 2013). This finding agrees with the results of Ginard and co-workers (1997) who described the presence of two plasmids in B1SMN1 and in strain ST27MN3. However, the plasmids of strain ST27MN3 have not been confirmed so far in the genome draft available for this strain. The possibility that the other strains whose genome information is in draft status (T13, XLDN-R, NF13, MF28, SDM-LAC, and TS44) might contain plasmids cannot be excluded, since there are plasmid proteins annotated in some genomes.

2. Characteristic CDSs of *P. stutzeri*

The species *P. stutzeri* has been described as positive for oxidase and catalase tests, negative for arginine dihydrolase (or arginine deiminase) and for glycogen hydrolase test (see section 2 of the Introduction). Morphologically it has a single polar flagellum (van Niel and Allen, 1952). Over the years isolates of this species have been widely studied due to their denitrification, nitrogen fixation, and aromatic compound degradation capabilities (Lalucat *et al*., 2006). Therefore CDSs encoding proteins or groups of proteins involved in the phenotypic characteristics mentioned above were identified in the genomes studied.

As shown in Table 1.2, CDSs coding for catalase and cytochrome *c* oxidase enzymes were identified in all strains. The number of CDSs annotated as catalase varied between the genomes from a single protein (as for ST27MN3, DSM 10701, SDM-LAC, and TS44) to 5 different copies (as for A1501, DSM 4166, and RCH2). In the case of cytochrome *c* oxidase, *P. stutzeri* and *P. balearica* genomes presented from 1 to 12 CDSs annotated as cytochrome *c* oxidase, with a total amount of 115 sequences. In contrast to what was expected, glycogen hydrolase enzymes were identified in all strains (see Table 1.2). Similarly, 16 CDSs annotated as arginine hydrolase were found, which belonged to 13 different strains (see Table 1.2). In general, all of them presented this sequence in a single copy, except from the strains A1501, 19SMN4, and ST27MN3, which present two gene copies encoding for this enzyme (all these sequences are listed in Supplementary Table S4).

The formation and structure of flagellum is given by 37 different proteins, whose function is shown in Supplementary Figure S1. As expected, the analysis of 19 *P. stutzeri* and *P. balearica* genomes revealed the presence of the genes encoding for all these proteins in all the genomes

Table 1.2. Analysis of presence/absence of CDSs coding for characteristic traits of *P. stutzeri*. Searched genes are specified in the text.

Strain	Catalase	Cytochrome *c* oxidase	Arginine dihydrolase	Glycogen hydrolase	Flagellum	Nitrogen fixation	Denitrification
ATCC 17588^T	+	+	–	+	+	–	+
A1501	+	+	+	+	+	+	+
B1SMN1	+	+	+	+	+	+	+
DSM 4166	+	+	+	+	+	+	+
T13	+	+	+	+	+	–	+
XLDN-R	+	+	+	+	+	–	+
CCUG 16156	+	+	+	+	+	–	+
AN10	+	+	–	+	+	–	+
19SMN4	+	+	+	+	+	–	+
ST27MN3	+	+	+	+	+	–	+
DSM 10701	+	+	+	+	+	–	+
28a24	+	+	–	+	+	–	–
NF13	+	+	–	+	+	+	+
KOS6	+	+	–	+	+	+	+
MF28	+	+	+	+	+	–	+
RCH2	+	+	+	+	+	–	+
SDM-LAC	+	+	–	+	+	–	+
TS44	+	+	+	+	+	–	+
SP1402^T	+	+	+	+	+	–	+

(listed on Supplementary Table S5). As shown in Figure 1.1, four different flagellar gene structures were observed in the 19 studied genomes. The simplest organization was observed in strains ATCC 17588^T, B1SMN1, CCUG 16156, AN10, NF13, KOS6, TS44, and SP1402^T (see Figure 1.1 A), with 15 gene structures: *motAB*, *fliEFGHIJ*, *cheY*, *fliKLMNOPQR*, *flhB*, *flhAF*, *fleN*, *fliA*, *cheYZAY*, *fliTDS*, *flaG*, *fliC*, *flgLKJIHGFEDCB*, *cheRW* and *flgAMN*. A similar gene structure was found in the A1501 genome, although in this case we observed a second copy of the genes *fliD*, *flaG* and *fliC* (coordinates 1502782..1506196 from NC_009434.1; see Supplementary Table S2) near the operons *fliTSD*, *flaG* and *fliC* (see Figure 1.1 B). Moreover, some genomes presented repetitions of these gene structures. In this sense, the genome of strain DSM 4166 presented repetitions of *flgLK*, *flgIHGFEDCB*, *flgA*, *fliC*, *motBA*, *fliLA*, *fliDS*, *flgM*, *fliJIHGFE*, *cheY*, *fliNPQR* and *flhBA* (see Figure 1.1 C). The absence of the gene encoding for FlaG, a flagellar protein whose function remains unknown in strain DSM 4166 was also remarkable. Instead, we found a second copy of FliD (YP_005938527.1, see Supplementary Table S2) in this strain. Finally, strains T13, XLDN-R, 19SMN4, ST27MN3, DSM 10701, 28a24, MF28, and RCH2 showed the repetitions described above for DSM 4166 (see Figure 1.1 D). These results confirm the potential of this species to assemble the flagellar structure, as is one of the traits of *P. stutzeri*.

The flagellar gene structures found in *P. stutzeri* and *P. balearica* genomes are consistent with those described in *E. coli* by Kalir and co-workers (2001). In that study, highly similar operons were identified: *fliFGHIJK*, *fliLMNOPQR*, *fliE*, *flhBAE*, *flgBCDEFGHIJ*, *flgAMN*, *flgKL*, *fliTSD*, *flgMN*, *fliC* and *motAB*. In addition, they also described the presence in *E. coli* of operons *flhDC* and *fliAZY*, which were not identified in our genomes. In *P. putida* KT2440 the flagellar assembly operons *fliLMNOPQR*, *fliEFGHIJ*, *fliTSD*, *flgFGHIJKL* and *flgBCDE* have been described as well (Martínez-García *et al*., 2014).

Denitrification was described as a wide extended phenotypic characteristic in *P. stutzeri* strains (Lalucat *et al*., 2006). This process involves the use of nitrate as a terminal electron acceptor for obtaining energy in anaerobic, microaerophilic or even aerobic conditions. It involves 4 reductases, encoded in the genome of *P. stutzeri* A1501 by different operons: respiratory nitrate reductase (*narGCKXL*), nitrite reductase (*nirYNEJPOQSTBMCFDLGH*), nitric oxide reductase (*norCBD*), and nitrous oxide reductase (*nosRZDFYL*) (Zumft, 1997).

A BLAST search of the *P. stutzeri* A1501 genes with the 19 *P. stutzeri* and *P. balearica* genomes was performed in this study. The analysis revealed the presence of 570 CDSs encoded by the denitrification operons distributed in 18 genomes (see Supplementary Table S6). No denitrification operons were found in the genome of the strain 28a24, suggesting that it is not able to carry out this process. As expected, the operon structure described by Zumft (1997) was the most abundant in the studied genomes, as it was found in strains A1501, B1SMN1, DSM 4166, and CCUG 16156, previously described as denitrifying bacteria (Yan *et al*., 2008; Yu *et al*., 2011; Peña *et al*., 2012, 2013; Busquets *et al*., 2013), and in strains T13, XLDN-R, ST27MN3, NF13, and RCH2 (see Figure 1.2 A). A similar gene structure was found in strains ATCC 17588^T, AN10, 19SMN4, and KOS6; despite the absence of genes *nirT* and *nirB* in all of them (see Figure 1.2 B). Each of the remaining strains presented different denitrification gene organization. Thus, strain DSM 10701, previously described as a denitrifying bacterium (Busquets *et al*., 2012), despite presenting the *nar*, *nor* and *nos* operons as the A1501 group, showed a different organization of *nir* operon (coordinates 714055..728857 from NC_018177.1; see Supplementary Table S2) (see Figure 1.2 C). The genome of MF28 presented the same denitrification gene structure than DSM 10701 of the operons *nos*, *nir* and *nor*. However, it lacked the operon *nar* (see Figure 1.2 D). Strain SDM-LAC presented on its genome the *norCBD* gene structure between genes *nirE* and *nirP*, together with the absence of *nirJ*. Moreover, the genome of this strain also lacked the *narJIHGCKXL* gene structure (see Figure 1.2 E). The genome of strain TS44 showed the same gene structure than the ATCC 17588^T group, but lacking *narC* (see Figure 1.2 F). Finally, the denitrification gene structure of the SP1402^T genome was also very similar to the ATCC 17588^T group, although lacking *narCKXL* (see Figure 1.2 G).

The lack of *nar* genes in MF28 and SDM-LAC genomes might prevent them from denitrifying nitrate, as these genes encode for the nitrate reductase involved in the first step of denitrification. Similarly, SP1402^T and TS44 might not be capable of denitrifying, due to the lack of *narCKXL* and *narC* respectively, as well as *nirTB*. Proteins NarC and NarK have been described as putative nitrite transporters, and NarX and NarL constitute a nitrate-responsive transcription factor (Zumft, 1997). Surprisingly, strain SP1402^T was previously described as a denitrifying bacterium (Rosselló et al., 1991). Furthermore, there were 6 other strains without *nirTB*: ATCC 17588^T, AN10, 19SMN4, KOS6, MF28, and SDM-LAC. The gene *nirT* encodes a tetraheme cytochrome with a putative electron donor function, and *nirB* encodes cytochrome *c552*. None of them have been yet identified in other well characterized denitrifiers such as *P. aeruginosa* and *Paracoccus denitrificans*. Furthermore, previous results showed that a mutation in *nirT* does not affect the synthesis of an in vitro catalytically competent nitrite reductase (Zumft, 1997). This suggests that both genes might not be essential for the denitrification process, and consequently, strains that are defective only for *nirTB* would also retain the potential for denitrification.

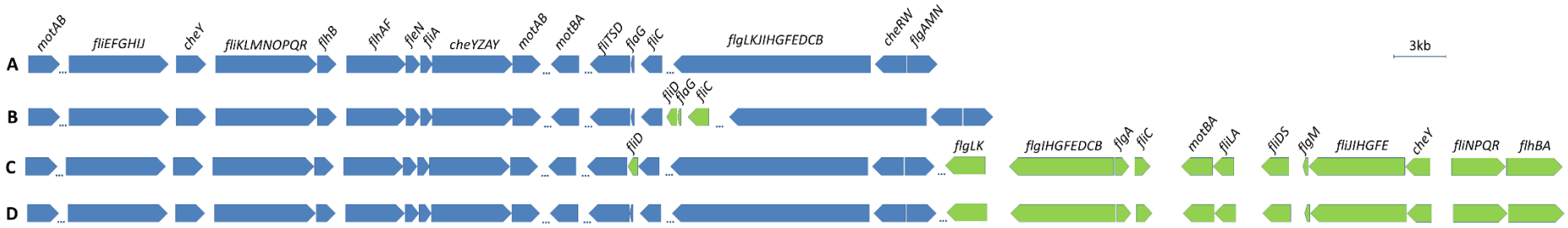

Figure 1.1. Structure of genes involved in the flagellar assembly system in *P. stutzeri* and *P. balearica*. (A) ATCC 17588^{T}, B1SMN1, AN10, NF13, CCUG 16156, KOS6, TS44, and SP1402^{T}. (B) A1501. (C) DSM 4166. (D) T13, XLDN-R, 19SMN4, ST27MN3, RCH2, DSM 10701, 28a24, MF28, and SDM-LAC. In blue the gene structure conserved in all strains. In green non-conserved gene structures. Noncontiguous genes are separated with dots.

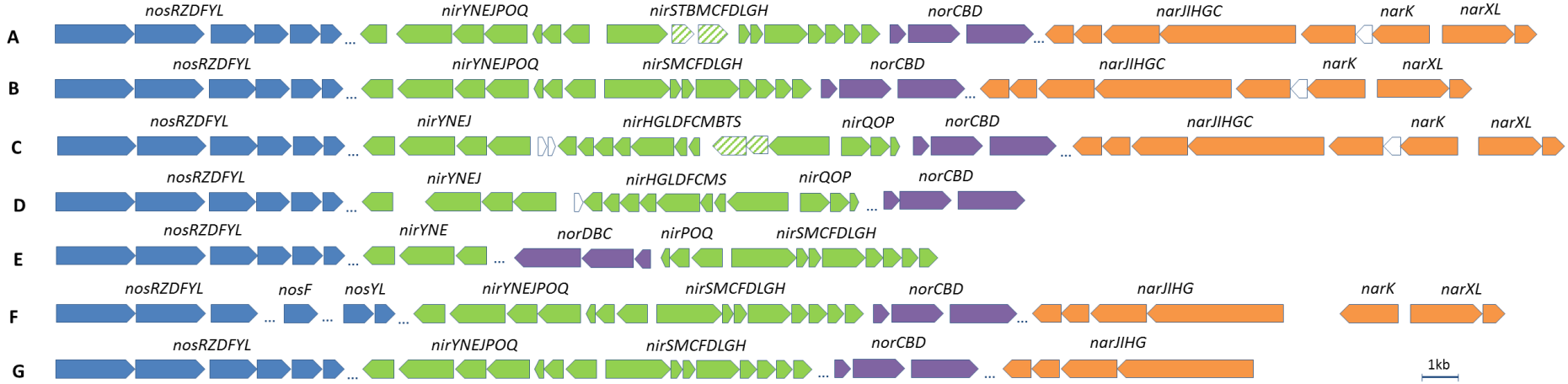

Figure 1.2. Structure of genes involved in denitrification in *P. stutzeri* and *P. balearica*. (A) A1501, B1SMN1, DSM 4166, T13, XLDN-R, CCUG 16156, ST27MN3, NF13, and RCH2. (B) ATCC 17588^{T}, AN10, 19SMN4, and KOS6. (C) DSM 10701. (D) MF28. (E) SDM-LAC. (F) TS44. G, SP1402^{T}. In blue the *nos* genes (nitrous oxide reductase), in green the *nir* genes (nitrite reductase), in purple the *nor* genes (nitric oxide reductase) and in orange the *nar* genes (nitrate reductase). In white non-conserved genes. Striped genes show identities below 50 %. Noncontiguous genes are separated with dots.

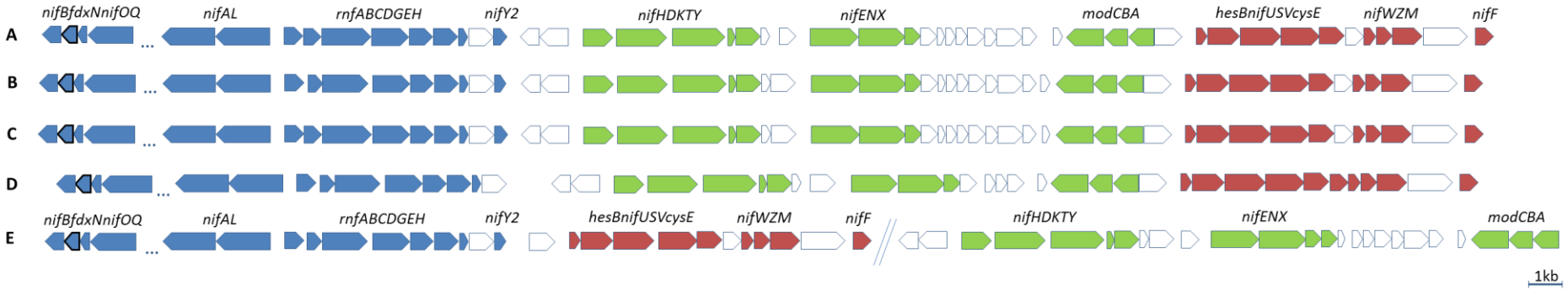

Figure 1.3. Structure of genes involved in nitrogen fixation in *P. stutzeri*. (A) A1501. (B) B1SMN1. (C) DSM 4166. (D) KOS6. (E) NF13. The three different gene structures are colored: blue, *nifQOfdxNnifB* (*nifO* highlighted), *nifLA*, *rnfABCDGEH*, and *nifY2*), green (*nifHDKTY*, *nifENX*, and *modABC*), and red (*hesBnifUSVcysE*, *nifWZM*, and *nifF*). In white non-conserved genes. Long distances (>10 kb) within the same contig are represented with dots. Genes encoded in different contigs are separated with a double slash.

Another interesting characteristic of some strains of *P. stutzeri* is the ability to fix nitrogen. Nitrogen fixation consists on the formation of ammonia from molecular nitrogen. We analyzed the presence of the set of genes involved in nitrogen fixation described by Yan and co-workers (2008) in A1501, in the rest of *P. stutzeri* strains and in *P. balearica*. All these genes were identified in 4 strains: B1SMN1, DSM 4166, NF13, and KOS6 (see Figure 1.3 and Supplementary Table S7). The gene structure described for *nif*, *rnf* and *mod* genes in A1501 was also observed in all these strains. However, the NF13 genome presented genes *nifUSV*, *nifWZM*, and *nifF* just besides *nifY2* instead of *modABC*, as the other strains (coordinates 6998..17024 from AOBS01000070.1; see Supplementary Table S2).

From the 5 different *P. stutzeri* strains identified as potential diazotrophs (nitrogen fixers) A1501 and DSM 4166 have been previously described as diazotrophs (Krotzky and Werner, 1987; Yan *et al*., 2008). The main enzyme involved in nitrogen fixation, nitrogenase, is extremely oxygen sensitive, which means that organisms would be able to fix nitrogen only under anoxic conditions or developing an ability to protect the nitrogen fixing system from oxygen (Dos Santos *et al*., 2011). Strains A1501 and DSM 4166 were both isolated from rice and *Sorghum nutans* roots respectively, environment that confer them the anoxic conditions required for nitrogen fixation. B1SMN1 was isolated from wastewater, KOS6 from a chemical sludge, and NF13 from a hydrothermal vent in the deep Galapagos rift. In all cases, an anoxic environment might be present due to the limited oxygen exchange.

As mentioned above, the 5 identified potential diazotrophs also showed the genes involved in denitrification. It is known that the main enzymes involved in these pathways compete for molybdenum, as they both use it as a cofactor. Because of this, the denitrification and nitrogen fixation pathways cannot occur simultaneously (Bosch *et al*., 1997). However, Bosch and Imperial (2000) previously described the coexistence of denitrification and nitrogen fixation in *Azotobacter vinelandii*. In this species, the nitrogenase-associated protein NifO (AAA22150.1) is capable of channeling the molybdenum to both pathways, and consequently, denitrification and nitrogen fixation are both operative. In *P. stutzeri*, a homologous protein to NifO in *A. vinelandii* was also observed in the 5 potential diazotrophs, with a percentage of identity higher than a 62 % (see Figure 1.3). This suggests that A1501, B1SMN1, DSM 4166, NF13, and KOS6 might be capable of fixing nitrogen while using nitrate as a terminal electron acceptor for obtaining energy in anoxic conditions.

3. Phylogenomic analysis of *P. stutzeri*

With the purpose of studying the relationships between *P. stutzeri* strains as well as *P. balearica* SP1402^T a phylogenomic analysis was performed. For this, sequences were grouped by similarity in clusters of orthologous groups (COGs). The most extended criterion to group CDSs with the same biological function is to share at least 50 % of identity in at least a 50 % of the sequence (Lukjancenko *et al*., 2010, 2012; Vesth *et al*., 2010; Jacobsen *et al*., 2011; Karlsson *et al*., 2011; Leekitcharoenphon *et al*., 2012). However, a phylogenomic analysis performed with this criterion would not be resolutive enough for the comparison of *P. stutzeri* strains, since approximately half of the COGs of each genome were shared with all strains. Because of this, a more stringent criterion was used to study

the dissimilarities between the different genomes, i.e. to share at least 95 % of amino acid identity in at least 90 % of the sequence. Using this criterion, we obtained a total of 42,079 COGs, containing 69 of them one CDS of each strain. CDSs from the 69 COGs which presented one CDS of each strain are listed on Supplementary Table S8. Clusters obtained with both criterions are shown in Supplementary Table S2.

The analysis of the number of shared COGs clearly showed 3 different groups of strains that shared more than 2,500 COGs, i.e. 60 % of the annotated CDSs in *P. stutzeri* genomes on average (see Table 1.3). The first includes the six strains of genomovar 1 (ATCC 17588^T, A1501, B1SMN1, DSM 4166, T13, and XLDN-R). These results are consistent with previous phylogenetic analysis based on a Multilocus Sequence Analysis (MLSA) of 16S rRNA, *gyrB, rpoB,* and *rpoD* genes (Gomila *et al*., 2015). The second group is formed by the two strains belonging to genomovar 4 (19SMN4 and ST27MN3). These two strains are clonal variations (Ginard *et al*., 1997), so it is not surprising that they shared the greater number of COGs. The third group is constituted by strains AN10 and NF13, belonging to genomovar 3 and 19 respectively. A close phylogenetic relation between AN10 and NF13 strains has been also described by Gomila and co-workers (2015), who defined a 93.3 % of MLSA identity between them.

On the other hand, strains CCUG 16156 (genomovar 2), AN10 (genomovar 3), NF13 (genomovar 19), KOS6, and RCH2 (genomovar not determined) presented more than 800 of shared COGs with genomovars 1 and 4; although not belonging to those genomovars (see Table 1.3). This suggests that they are phylogenetically closer to those genomovars rather than to the remaining *P. stutzeri* strains (DSM 10701, 28a24, MF28, SDM-LAC, and TS44). This is also consistent with the phylogenetic analysis of Gomila and co-workers (2015).

As for the rest of *P. stutzeri* strains (DSM 10701, 28a24, MF28, SDM-LAC, and TS44), they did not share more than 400 of COGs with the strains mentioned above, i.e. 10 % of the annotated CDSs in *P. stutzeri* genomes on average. Nevertheless, it is noteworthy that strains 28a24, MF28, and SDM-LAC shared more than 400 COGs, suggesting a closer relation between them. Instead, DSM 10701 and TS44 do not share more than 400 COGs with any of the other studied strains. As shown in Table 1.3, *P. balearica* SP1402^T did not share more than 400 COGs with any of the 18 *P. stutzeri* strains, which is consistent with the taxonomic classification of this strain outside the *P. stutzeri* species. According to this reasoning, the results of the shared COGs evidenced that the taxonomic position of strains DSM 10701, 28a24, MF28, SDM-LAC, and TS44 within the species *P. stutzeri* should be revised (see below).

On the other hand, a phylogenetic analysis was also performed using the 69 shared COGs present in single copy in all genomes (see Supplementary Table S8). Most of these CDSs encoded ribosomal proteins (44.1 %), and proteins involved in carbohydrate and amino acid metabolism (22.1 % and 20.1 %, respectively). From the concatenated amino acid sequence of these shared COGs a Neighbor-Joining phylogeny tree was calculated. As shown in Figure 1.4, the resulting tree presented bootstrap values (100 replicates) greater than 50 at all nodes. Strains from genomovars 1 and 4 formed coherent groups, and both were grouped

together with strains CCUG 16156, AN10, NF3, KOS6, and RCH2. On the other hand, the group constituted by strains DSM 10701, 28a24, MF28, SDM-LAC, and TS44 separated clearly from the *P. stutzeri* branch, as well as from *P. balearica* SP1402^T. All of it is very consistent with the number of COGs shared between strains previously described (see Table 1.3), as well as the phylogenetic analysis based on the concatenate of the partial sequences of 16S rRNA, *gyrB* and *rpoD* genes performed by Gomila and co-workers (2015).

Table 1.3. Number shared COGs (upper diagonal) and ANIb (lower diagonal) values obtained between the different studied genomes. The affiliation of each strain into genomovars is shown in brackets (when it has been determined). Pairs of strains with more than 2,500 shared COGs (i.e. 60 % of the annotated CDSs on average in *P. stutzeri* genomes) are represented in green, pairs of strains with less than 400 shared COGs (10 % of the annotated CDSs on average in *P. stutzeri* genomes) are shown in red, and the intermediate values (from 400 to 2,500) are not colored. The pairs of strains with ANIb values higher than 96 % are represented in green. The pairs of strains with ANIb values between 93 and 96 % are represented in blue (see the text for details).

	ATCC1 7588^T (1)	A1501 (1)	B1SMN1 (1)	DSM 4166 (1)	T13 (1)	XLDN-R (1)	CCUG 16156 (2)	AN10 (3)	19SMN4 (4)	ST27MN3 (4)	DSM 10701 (8)	28a24 (11)	NF13 (19)	KOS6	MF28	RCH2	SDM-LAC	TS44	SP1402^T
ATCC 17588^T (1)	-	2738	3133	3258	3208	3185	1233	1246	1263	1313	188	173	1219	984	191	1273	178	390	381
A1501 (1)	96.8	-	2992	2887	2866	2815	1106	1101	1161	1125	179	153	1147	884	168	1146	168	326	293
B1SMN1 (1)	96.5	97.5	-	3325	3293	3254	1181	1308	1304	1292	198	175	1262	1009	194	1274	190	381	388
DSM 4166 (1)	97.2	97.3	97.4	-	3406	3358	1152	1197	1285	1268	185	161	1261	935	186	1291	185	366	345
T13 (1)	97.6	97.6	97.6	98.1	-	3483	1142	1195	1332	1278	184	168	1236	931	197	1320	183	360	331
XLDN-R (1)	97.2	97.0	97.1	97.4	97.9	-	1144	1298	1326	1336	184	170	1249	999	201	1318	183	351	393
CCUG 16156 (2)	85.9	85.7	85.7	85.5	85.4	85.4	-	1653	1483	1481	179	193	1447	919	206	1468	211	348	289
AN10 (3)	86.4	86.2	86.7	86.2	86.2	86.7	87.9	-	1771	1777	185	193	2536	1171	224	1638	211	352	370
19SMN4 (4)	86.2	86.4	86.4	86.3	86.2	86.3	86.8	88.4	-	3700	210	187	1564	1064	209	2148	199	364	321
ST27MN3 (4)	86.4	86.1	86.2	86.0	86.1	86.2	86.8	88.3	99.4	-	190	180	1533	1082	198	2145	192	406	388
DSM 10701 (8)	80.7	80.8	80.7	80.7	80.6	80.6	79.5	80.2	80.2	80.0	-	244	206	195	252	193	262	180	138
28a24 (11)	78.2	78.2	78.2	78.2	78.1	78.2	77.6	78.2	78.0	78.0	79.5	-	183	193	902	182	685	143	122
NF13 (19)	86.4	86.6	86.6	86.6	86.3	86.5	87.1	93.1	87.8	87.9	80.3	78.3	-	1062	225	1496	195	384	317
KOS6	85.2	85.3	85.6	85.1	85.1	85.4	84.4	86.1	85.4	85.6	80.1	78.1	86.1	-	221	1031	187	357	361
MF28	78.9	78.9	78.8	78.9	79.0	79.2	78.1	78.9	78.6	78.5	79.9	83.3	79.0	79.1	-	246	729	224	161
RCH2	86.4	86.3	86.3	86.5	86.5	86.6	86.7	87.9	89.9	89.9	79.9	78.1	87.4	85.6	79.1	-	196	388	328
SDM-LAC	78.2	78.4	78.3	78.5	78.3	78.4	77.7	78.3	78.0	78.0	79.4	81.8	78.2	78.1	82.0	78.1	-	166	124
TS44	83.0	82.7	82.7	82.8	82.7	82.5	81.3	82.0	82.0	82.2	81.0	78.4	82.4	82.3	80.0	82.1	78.5	-	254
SP1402^T	81.2	80.7	81.2	80.9	80.7	81.3	79.3	80.7	80.1	80.6	79.1	77.2	80.4	80.9	78.2	80.1	77.1	81.0	-

A third phylogenomic approach was performed calculating the ANIb values between these strains using the JSpecies program, as done by Gomila *et al*. (2015), in order to compare the results with the other two approaches followed in this study. Rosselló-Móra and Amann (2015) have defined the ANIm values for delineation of species boundaries after analyzing more than 60,000 genomes from 195 genera. ANIm and ANIb values are nearly identical for genomes with high level of identity (Rosselló-Móra and Amann, 2015), and therefore we have used the proposed values of ANIm for the interpretation of our ANIb data. According

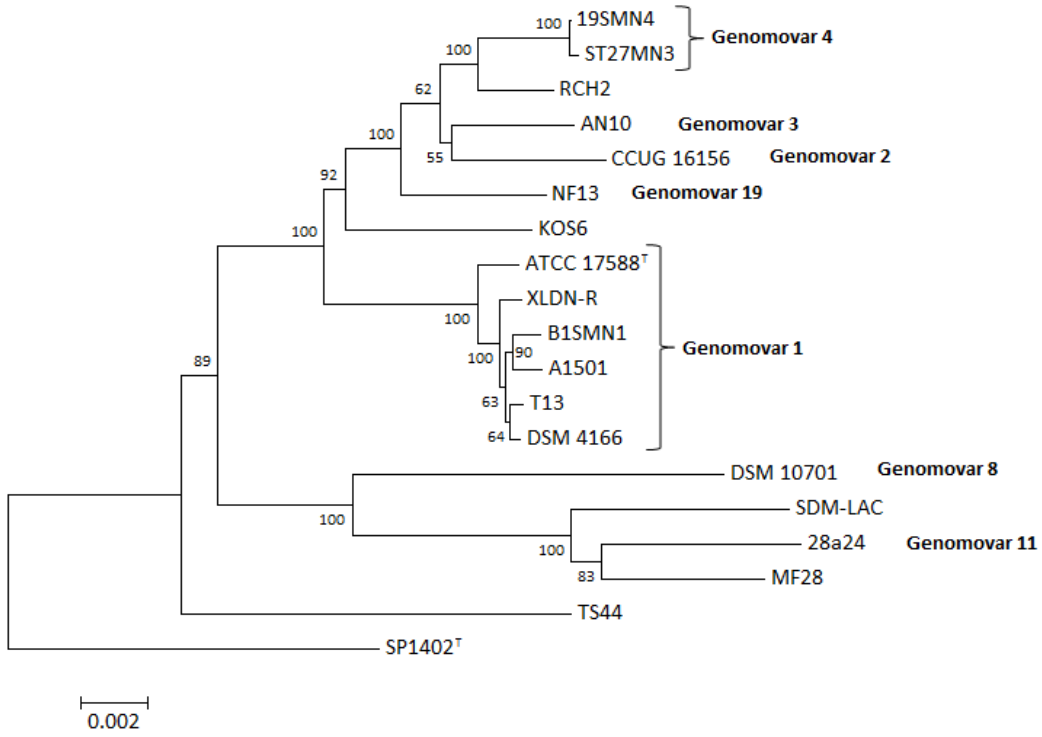

Figure 1.4. Neighbor-Joining tree calculated from the concatenation of 69 shared COGs in the 18 *P. stutzeri* strains and *P. balearica* SP1402^T. Bootstrap of 100 replicates are shown. The genomovar of each strain (if it has been assigned) is also shown.

to Rosselló-Móra and Amann (2015) strains of the same species have ANIm values over 96 %. Values below 93 % correspond to different species of the same genus. The range between 93-96 % ANIm is a fuzzy zone were we can find bacteria that have been assigned to the same or different species, according to the criterion of the taxonomist who described it, and it would be a range were different genomovars of a species could be allocated. As shown in Table 1.3, the results of ANIb were similar to the ones obtained with the number of shared COGs. There was a first group of strains belonging to genomovar 1 (ATCC 17588^T, A1501, B1SMN1, DSM 4166, T13, and XLDN-R) with ANIb values ranging from 96.5 to 98.1 %. A second group of strains was formed by the two clonal variants of genomovar 4: strains 19SMN4 and ST27MN3, which presented the highest ANIb value of all (99.4 %). Strains AN10 (genomovar 3) and NF13 (genomovar 19) shared an ANIb value of 93.1 % (see Table 1.3). The genomes of the other studied strains did not show ANIb values above 93 %. Consequently, according to the limits described by Rosselló-Móra and Amann (2015), strains belonging to genomovar 1 should be considered as *P. stutzeri* strains, forming a monophyletic group (see Figure 1.4) and with rather homogeneous G+C mol% content (see Table 1.1). However, based on genomic data, the taxonomic position of the other strains might be considered doubtful. The results of the phylogenomic analysis indicate that the rest of the strains should be considered as different genomic species (strains 19SMN4 and ST27MN3 would belong to the same genomic species), with the only doubt of strains AN10 and NF13, whose ANIb values range in the 93-96 % and might be considered as genomovars of a species. This is consistent with the values of shared COGs (Table 1.3) and with the phylogenetic reconstruction shown in Figure 1.4. These results are also consistent with the MLSA of *P. stutzeri* performed by Gomila and co-workers (2015). A high MLSA similarity of

TS44 strain was described with the non-characterized strain *Pseudomonas* sp. Chol1, with which presented a 95.2 % of MLSA identity (97 % is established as the minimum intraspecific identity). In relation with strain SDM-LAC, according to an MLSA, the closest-related strain was *P. xanthomarina* CCUG 45643^T instead of other *P. stutzeri* strains, supporting the idea that this strain should be considered as different genomic species.

Thus, the three phylogenomic approaches performed do not support the current structure of *P. stutzeri* species into different genomovars, and only strains from genomovar 1 constituted a coherent group. Since the *P. stutzeri* type strain (ATCC 17588^T) belongs to genomovar 1, strains from this genomovar might be considered as true *P. stutzeri* strains. However, any modification of the taxonomy of these strains requires further additional studies by expert taxonomists and cannot be based in genomic data only. Therefore, for the purpose of this study these strains have been considered as belonging to *P. stutzeri*, since this is their current taxonomic position.

4. Pan- and core- proteome of *P. stutzeri*

In order to be able to compare the total of 80,524 CDSs annotated in *P. stutzeri* sequences were grouped by similarity in COGs. The criterion for grouping CDSs with the same biological function was at least a 50 % of identity in at least a 50 % of the sequence. By this a total amount of 11,548 COGs were obtained, and 2,094 of them were present in all analyzed *P. stutzeri* isolates, with at least one CDS per strain (*P. balearica* SP1402^T was not considered in this study). Therefore, we defined CDSs from those 2,094 COGs as the core-proteome of *P. stutzeri* (see Supplementary Table S9). This means that, using this criteria, almost the 50 % of the annotated CDSs per genome constituted the core-proteome of *P. stutzeri* (as each genome presented an average of 4,240 $\pm$ 292 annotated CDSs). Similar approaches have been followed in other studies using the same criteria. Tettelin and co-workers (2005) defined the core-genomes of 8 *Streptococcus agalactiae* genomes as 1,806 COGs. Schoen and co-workers (2008) described the core-genomes of 7 *Neisseria meningitidis* genomes as 1,333 COGs. In all these cases, the core-proteome defined is smaller than the obtained for *P. stutzeri*. In contrast, Willenbrock and co-workers (2007) described the core-genomes of 32 *Escherichia coli* genomes as 2,241 COGs, more than the *P. stutzeri* core-proteome, indicating that *E. coli* genomes present a lower genetic diversity. In the case of other *Pseudomonas* species, Udaondo and co-workers (2015) analyzed the core-genome of 9 *P. putida* genomes. A core-genome of 3,326 COGs was defined, which represents approximately 60 % of the average of annotated CDSs per genome. This percentage is higher than the obtained for *P. stutzeri* genomes (although it is closer to number of COGs shared by strains of genomovar 1 as shown in Table 1.3.). This might suggest that *P. stutzeri* genomes present a higher genetic diversity than *P. putida*. However, this difference might also be due to the lower number of genomes used in the *P. putida* analysis (9 genomes) compared to the present study (19 genomes), or to the evidences discussed above about the phylogenomic structure of *P. stutzeri*.

The core-proteome was then analyzed according to the KEGG biological functions classification (see Supplementary Table S3), in order to assign each COG to the different

group of biological functions defined by this server. Results showed that they were mainly proteins involved in metabolism (see Figure 1.5). In particular, proteins associated with metabolism were mainly classified into five categories: carbohydrate metabolism (11.5 %) and amino acid metabolism (11.3 %), metabolism of cofactors and vitamins (6.3 %), energy metabolism (5.7 %), and nucleotide metabolism (5.1 %). As shown in Figure 1.5, the second most abundant biological function associated with core-COGs was the related with genetic information processing (9.8 %), followed by those related to environmental information processing (9.6 %). Udaondo and co-workers (2015), also classified the core-genome of *P. putida* into the KEGG functional categories. In this case, the most abundant genes within the core-genome were those that encode nutrient transport proteins, but as in our case, genes involved in carbon and amino acid metabolism were also abundant.

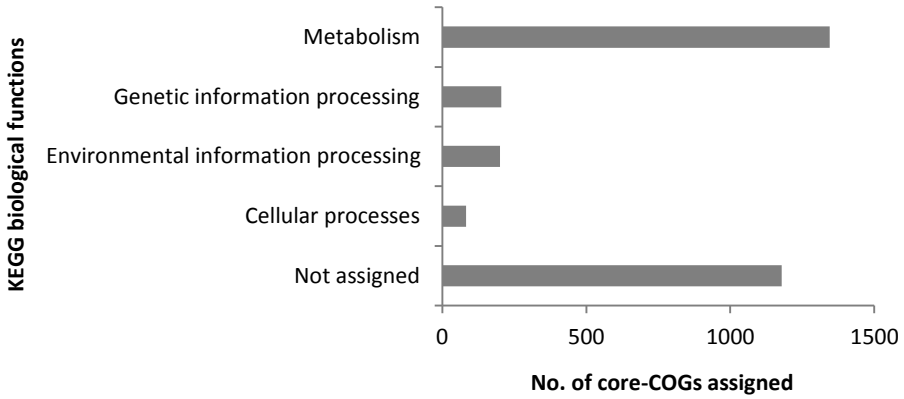

Figure 1.5. Number of core-COGs assigned to each group of biological functions according to KEGG categories.

The strain-specific COGs of each strain (SS-COGs) were also analyzed using the criterion mentioned above, in order to identify those CDSs clearly different from the others. A total of 5,221 groups of COGs were SS-COGs (see Supplementary Table S10). The disparate distribution of those SS-COGs within the different analyzed strains is also remarkable; it goes from 126 to 757 SS-COGs per genome. Strains B1SMN1 and KOS6 showed higher percentages of SS-COGs (> 10 %). The high percentage of SS-COGs in the genome of strain B1SMN1 might be related with the fact that this genome has more annotated CDSs than the other studied strains. Instead, strains DSM 4166, ST27MN3, RCH2, XLDN-R, 19SMN4, AN10, A1501, and ATCC 17588^T presented percentages lower than 5 % (see Table 1.4). These values are consistent with those described by Udaondo and co-workers (2015) for *P. putida*, who showed from 202 to 917 SS-COGs per genome. In this study, *P. putida* Idaho had the most SS-COGs (917), which represented a 16.1 % of its total genome.

The SS-COGs were also analyzed according to the KEGG classification (see Supplementary Table S3), although only 7.5 % of them could be classified. From those, 63.8 % were involved in metabolism, 3.9 % were involved in environmental information processing, and 5.4 % were involved in genetic information processing. In particular, proteins associated with metabolic pathways were mainly classified into three categories: amino acid metabolism (20.6 %), xenobiotic metabolism (15.7 %), and carbohydrate metabolism (14.7 %). The SS-COGs associated with xenobiotic degradation pathways were heterogeneously distributed between the different strains. Strains TS44 and 28a24 presented the higher amount of

them, with 19 and 18 xenobiotic metabolism SS-COGs, respectively. The former was isolated from an arsenic-contaminated soil (Cai *et al*., 2009), and the later, is known as a highly transformable bacteria (Smith *et al.,* 2014). In addition, strains KOS6, AN10, and MF28, all isolated from polluted environments, also presented a remarkable number of xenobiotic metabolism SS-COGs (from 9 to 12 SS-COGs).

Table 1.4. Relation of the total number of annotated CDSs, COGs (with a 50 % of identity in the 50 % of the sequence), and SS-COGs per each strain. Percentages are shown in brackets.

Strain	CDSs	COGs	SS-COGs
ATCC 17588^T	4218	4151	201 (4.8 %)
A1501	4127	4065	191 (4.7 %)
B1SMN1	5104	5045	757 (15.0 %)
DSM 4166	4301	4228	126 (2.9 %)
T13	4255	4230	246 (5.8 %)
XLDN-R	4217	4201	164 (3.9 %)
CCUG 16156	4126	4101	240 (5.9 %)
AN10	4300	4263	172 (4.0 %)
19SMN4	4371	4323	173 (4.0 %)
ST27MN3	4299	4278	137 (3.2 %)
DSM 10701	3815	3782	289 (7.6 %)
28a24	4147	4121	376 (9.1 %)
NF13	4321	4297	219 (5.1 %)
KOS6	4466	4433	511 (11.5 %)
MF28	4540	4507	399 (8.9 %)
RCH2	4264	4200	159 (3.8 %)
SDM-LAC	3727	3712	217 (5.9 %)
TS44	4000	3957	354 (8.9 %)
SP1402^T	3926	3898	292 (7.5 %)

5. Identification of putative transposases in *P. stutzeri* genomes

The presence of transposons and insertion sequences (ISs) has been postulated as an indicative of the existence of horizontal gene transfer (HGT), and consequently, it has been considered of high ecological interest (Sobecky and Hazen, 2009). Furthermore, several authors have previously described the importance of transposons in aromatic compounds degradation (Wyndham *et al*., 1994; Tan, 1999; Top and Springael, 2003; Nojiri *et al*., 2004). This is due to the large number of catabolic transposons involved in the degradation of aromatic compounds described such as naphthalene (Tn4655), toluene (Tn4656, Tn4651, Tn4653), chlorobenzoate (Tn5271, Tn5707), and chlorobenzene (Tn5280) (Tan, 1999; Top and Springael, 2003). In *P. stutzeri*, Christie-Oleza and co-workers (2009) described an ISL3-like TnpA (IS*Pst9*) in the genome of strain AN10, whose transposition was induced by conjugative interaction. Additionally, Bosch and co-workers (1999a, 1999b) previously described the presence of several genes encoding TnpAs flanking the naphthalene degradation operons in the genome of strain AN10. Therefore, we decided to evaluate the

presence of sequences coding for transposases (TnpAs) in the 19 genomes of *P. stutzeri* and *P. balearica*.

Of the total of 42,079 COGs with at least a 95 % of identity in at least a 90 % of the sequence present in the 19 genomes, 296 of them were identified as putative TnpAs. These were identified either because they were annotated as TnpAs, or by presenting a BLAST hit with a database of TnpAs generated in this study. Then these COGs identified as potential TnpAs were assigned to the different described IS-families using the IS-Finder server (Siguier *et al.*, 2006). As shown in Figure 1.6, the most common IS-families were the relatives to IS3 (25.7 %) and IS5 (21.3 %). This is consistent with Siguier and co-workers (2014), who described that those IS-families were the most abundant in the IS-Finder database itself. However, there was also a wide representation of sequences classified into other IS-families: IS66 (11.2 %), IS21 (5.4 %), ISL3 (4.7 %), and Tn3 (4.7 %). These results are consistent with those obtained by Wu and co-workers (2011) with *P. putida*, who described the IS-family IS5 as the most abundant of the 4 studied genomes (strains W619, KT2442, F1, and GB-1) of these species, but differ from the obtained by Kung and co-workers (2010) in *P. aeruginosa*, in which Tn3-like elements were the most frequently found.

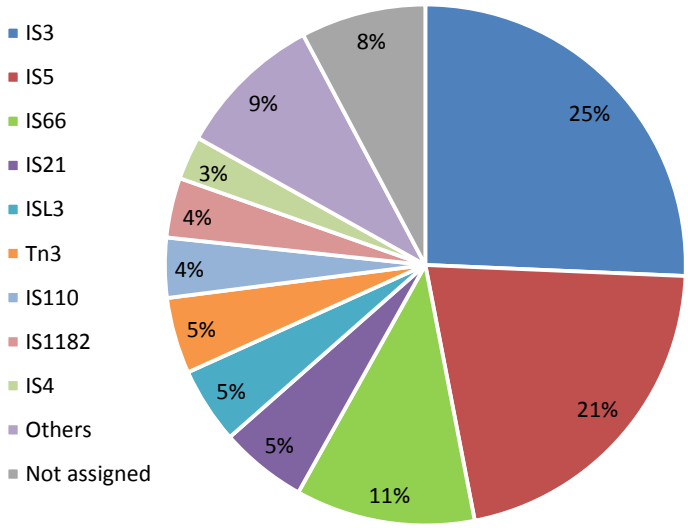

Figure 1.6. Percentage of TnpAs classified on each IS-family. The 296 COGs identified as putative TnpAs were considered. IS-families grouped as "others" include: IS1595, IS200/IS605, IS256, IS30, IS360, IS481, IS6, IS630, IS91, ISAs1, and ISKra4.

Analyzing the distribution of CDSs identified as putative TnpAs between the different strains, we observed that these genomes present a disparate number of TnpAs. As shown in Figure 1.7, the 19 genomes of *P. stutzeri* and *P. balearica* harbor between 6 to 76 sequences identified as putative TnpAs, with an average of 36 ± 22 sequences. The *P. stutzeri* strains with higher numbers of identified TnpAs were: A1501, with 76 sequences; ATCC 17588^{T}, with 69 sequences; B1SMN1, 67 sequences; and DSM 4166, with 62 sequences (see Figure 1.7). In contrast, strains SDM-LAC, T13, RCH2, and 28a24 only presented from 6 to 8 putative TnpAs. This suggests that the first group of strains have suffered more horizontal gene transfer events than the other strains. Similarly, Wu and co-workers (2011) described a great deviation in transposase content in *P. putida*, ranging

between the 36 IS elements of KT2440 genome and the 2 harbored in strain F1. Despite the diverse number of TnpAs per genome, the proportion of IS-families on each strain remained constant. As shown in Figure 1.7, families IS3, IS5, and IS66 were the most common IS-families. Conversely, strains ATCC 17588^{T} and A1501 showed predominantly TnpAs from the IS21 family.

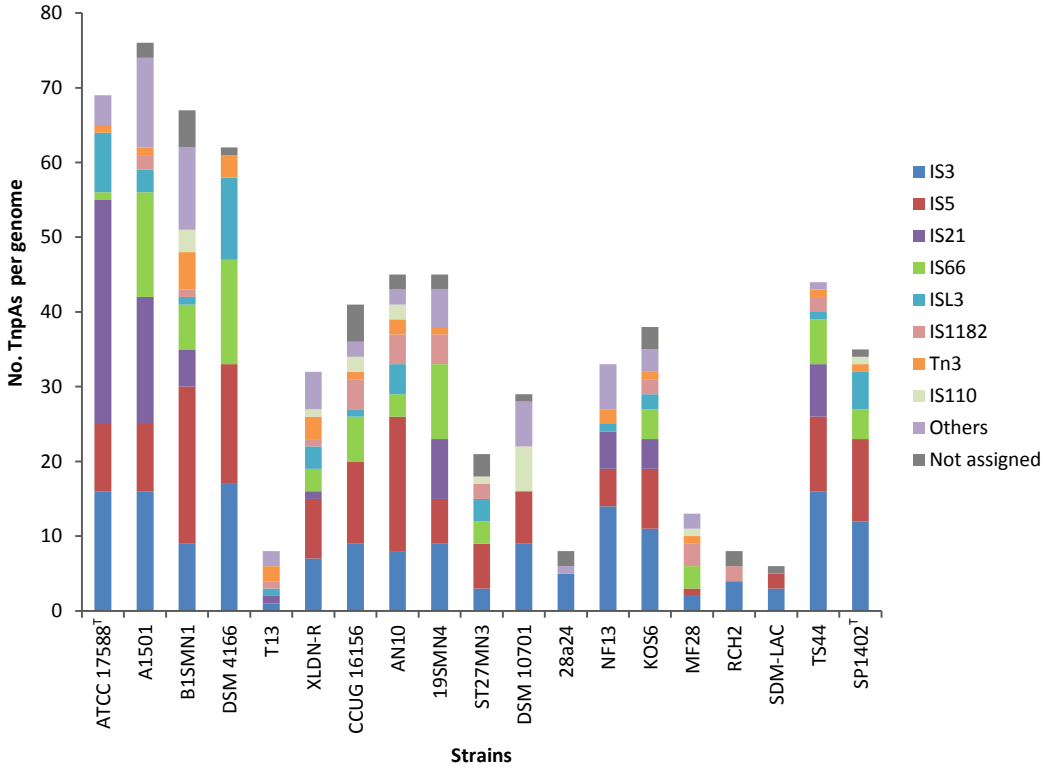

Figure 1.7. Number of putative TnpAs of each IS-family identified in the different studied strains. IS-families grouped as "others" include: IS110, IS1595, IS200/IS605, IS256, IS30, IS360, IS4, IS481, IS6, IS630, IS91, ISAs1, and ISKra4.

In order to analyze the horizontal gene transfer suffered within the genomes of *P. stutzeri* strains along its evolution, we calculated the number of shared COGs identified as possible TnpAs between the studied strains. As shown in Table 1.5, the number of shared TnpA-COGs was from 0 to 13. We also observed that the number of shared TnpA-COGs was not higher between strains of the same genomovar, compared with strains of different genomovars. So, it seems that horizontal gene transfer occurs regardless of the genomovar they belong (see Table 1.5). It is remarkable the higher number of shared TnpAs in *P. stutzeri* genomes in comparison with the analysis of 4 *P. putida* genomes performed by Wu and co-workers (2011). In that case, none of the identified IS elements was shared with more than 2 of the 4 analyzed genomes.

Thirteen TnpA-COGs were shared between strains B1SMN1 and A1501. These COGs have been analyzed in more detail in the genome of strain A1501, and we have seen that 9 of them are concentrated in a region of 56.03 kb-pairs (coordinates 717297..773327 from NC_009434.1; see Supplementary Table S2). In this region there were also 9 CDSs annotated as phage resistance proteins (YP_001171180.1, YP_001171181.1, YP_001171182.1, YP_001171183.1, YP_001171187.1, YP_001171188.1, YP_001171194.1,

YP_001171195.1, and YP_001171198.1; see Supplementary Table S2). Unfortunately, we were not able to analyze this region in strain B1SMN1, since the genome is not closed.

Table 1.5. Number of TnpAs-COGs shared between the studied strains. The affiliation of each strain into genomovars is shown in brackets (when it has been determined). Values equal or higher than 7 are shown in green.

	ATCC17588^T (1)	A1501 (1)	B1SMN1 (1)	DSM4166 (1)	T13 (1)	XLDN-R (1)	CCUG16156 (2)	AN10 (3)	19SMN4 (4)	ST27MN3 (4)	DSM10701 (8)	28a24 (11)	NF13 (19)	KOS6	MF28	RCH2	SDM-LAC	TS44	SP1402^T
ATCC 17588^T (1)		7	5	5	0	3	5	1	4	0	2	0	5	7	1	2	1	4	1
A1501 (1)			13	4	2	9	3	0	3	0	0	0	5	7	0	2	0	3	0
B1SMN1 (1)				4	3	9	4	9	10	8	0	0	3	7	2	2	0	7	4
DSM 4166 (1)					1	4	3	1	4	0	0	0	4	4	1	2	1	4	1
T13 (1)						3	0	2	2	1	0	0	2	1	2	1	0	1	1
XLDN-R (1)							2	4	5	4	0	0	2	4	2	3	0	3	1
CCUG 16156 (2)								5	4	1	1	0	1	2	1	1	2	2	1
AN10 (3)									10	8	1	0	2	12	3	2	2	3	5
19SMN4 (4)										9	3	0	4	11	1	3	1	6	4
ST27MN3 (4)											0	0	0	5	1	2	0	3	3
DSM 10701 (8)												0	3	3	0	0	1	0	1
28a24 (11)													0	0	0	0	0	0	0
NF13 (19)														8	1	0	1	6	2
KOS6															2	3	2	4	5
MF28																1	0	0	2
RCH2																	0	1	0
SDM-LAC																		0	0
TS44																			0
SP1402^T																			

Strains B1SMN1, AN10, 19SMN4, ST27MN3, and KOS6, all isolated from polluted environments, shared from 7 to 12 TnpA-COGs. As an example, strain AN10 shared 9, 10, 8, and 12 TnpA-COGs with strains B1SMN1, 19SMN4, ST27MN3, and KOS6 respectively. Analyzing it in detail, we observed that most of them were located between the naphthalene degrading operons. This will be more extensively analyzed in Chapter 2.

CHAPTER 2: Aromatic hydrocarbon degradation potential of *Pseudomonas stutzeri*

1. Analysis of aromatic hydrocarbon degradation gene clusters

The identification of aromatic hydrocarbon degradation gene clusters in the 18 sequenced genomes of *P. stutzeri* and *P. balearica* SP1402^T was performed by two different strategies: (1) analyzing the annotation by searching those CDSs annotated as monooxygenases, dioxygenases, or hydroxylases; and (2) the analysis of a second annotation of these genomes using the KEGG Automatic Annotation Server (KAAS) (Moriya *et al*., 2007). With the first approach, we identified a total of 299 putative dioxygenases, 139 hydroxylases, and 74 monooxygenases. As seen in Table 2.1, the three types of enzymes were uniformly distributed in the different strains, with an average of 15.7 ± 3.7 dioxygenases, 7.3 ± 1.8 hydroxylases, and 3.9 ± 1.8 monooxygenases.

Table 2.1. Number of CDSs described in the annotation of *P. stutzeri* and *P. balearica* genomes as dioxygenases, hydroxylases, and monooxygenases.

Strain	Dioxygenases	Hydroxylases	Monooxygenases
ATCC 17588^T	17	8	4
A1501	17	8	4
B1SMN1	22	7	5
DSM 4166	17	8	5
T13	13	6	4
XLDN-R	15	6	4
CCUG 16156	12	7	3
AN10	23	7	3
19SMN4	17	7	3
ST27MN3	21	11	2
DSM 10701	16	5	3
28a24	14	7	9
NF13	16	6	4
KOS6	15	9	7
MF28	10	7	4
RCH2	12	12	2
SDM-LAC	9	6	1
TS44	18	6	4
SP1402^T	15	6	3

All CDSs identified as dioxygenases, hydroxylases and monooxygenases were analyzed in more detail. In order to determine if they constituted part of aromatic hydrocarbon degradation gene clusters, the flanking genes were analyzed in the genomic annotation provided by KAAS. With this strategy, we identified gene clusters involved in the degradation of the following aromatic hydrocarbons: homoprotocatechuate 4-hydroxyphenylpyruvate via homogentisate; 4-hydroxybenzoate via protocatechuate *ortho*-cleavage; benzoate via catechol *ortho*-cleavage; carbazole via catechol *ortho*-cleavage; phenol via catechol *meta*-cleavage; and naphthalene via salicylate and catechol *meta*-cleavage; (see Table 2.2). Apart from genes involved in aromatic hydrocarbon degradation, two of the CDSs annotated as dioxygenases in the genomes of strains MF28 and TS44 were predicted to be involved in dehydroabietic acid degradation. This will be described in

section 2. CDSs corresponding to all these gene clusters are shown in Supplementary Table S11.

Table 2.2. Presence of aromatic hydrocarbon degradation gene clusters in *P. stutzeri* and *P. balearica* genomes.

Strain	Catechol *meta*- cleavage	Catechol *ortho*- cleavage	Protocatechuate *ortho*- cleavage	Homogentisate	Homoprotocatechuate	4-Hydroxyphenylpyruvate	4-Hydroxybenzoate	Salicylate	Benzoate	Carbazole	Phenol	Naphthalene	Total pathways in strain
ATCC 17588^T	–	+	+	+	–	+	–	–	+	–	–	–	5
A1501	–	+	+	+	–	+	+	–	+	–	–	–	6
B1SMN1	+	+	+	+	–	+	+	+	+	–	–	+	9
DSM 4166	–	+	+	+	–	+	+	–	+	–	–	–	6
T13	–	+	+	+	–	+	+	–	+	–	–	–	6
XLDN-R	+	+	+	+	–	+	+	–	+	+	–	–	8
CCUG 16156	–	–	–	+	–	+	–	–	–	–	–	–	2
AN10	+	+	–	+	–	+	–	+	+	–	–	+	7
19SMN4	+	+	–	+	–	+	–	+	+	–	–	+	7
ST27MN3	+	+	–	+	–	+	–	+	+	–	–	+	7
DSM 10701	–	+	–	+	–	+	–	–	+	–	–	–	4
28a24	–	–	+	+	–	+	+	–	–	–	–	–	4
NF13	–	+	–	+	–	+	–	–	+	–	–	–	4
KOS6	+	–	–	+	+	+	–	+	–	–	+	+	7
MF28	–	–	–	+	–	+	–	–	–	–	–	–	2
RCH2	–	+	–	+	–	+	–	–	+	–	–	–	4
SDM-LAC	–	+	–	+	–	+	–	–	+	–	–	–	4
TS44	+	+	–	+	–	+	–	–	+	–	+	–	6
SP1402^T	+	+	–	+	–	+	–	+	+	–	–	+	7
Total strains with pathway	8	15	7	19	1	19	6	6	15	1	2	6	

The aromatic ring fission of catechol might be performed by two different dioxygenases: catechol 1,2-dioxygenase (*ortho*-cleavage) and catechol 2,3-dioxygenase (*meta*-cleavage) (George and Hay, 2011). In *P. stutzeri* and *P. balearica* genomes both enzymes have been identified. The former was identified in all strains except: CCUG 16156, 28a24, KOS6, and MF28; and in all cases in a single copy. The latter was not so widely distributed in *P. stutzeri*, as it was identified only in strains: B1SMN1, XLDN-R, AN10, 19SMN4, ST27MN3, KOS6, and TS44; also in single copy (except from KOS6, which presented two genes encoding catechol

2,3-dioxygenase). Both dioxygenases were found in *P. balearica*. The presence of CDSs annotated as catechol 1,2-dioxygenase in the genomes of strains AN10 and 19SMN4 is consistent with the results of Ginard and co-workers (1997), who described the presence of *catA* in both genomes. However, no catechol 1,2-dioxygenase activity was detected by Rosselló-Mora and co-workers (1994) in cultures of 19SMN4 growing with benzoate as carbon source. This activity was not determined in AN10, since this strain was not able to grow with benzoate as carbon source. This will be discussed in section 2 from Chapter 4).

The β-ketoadipate pathway is a convergent pathway for aromatic compound degradation widely distributed in soil bacteria. One branch converts protocatechuate derivatives and the other branch catechol derivatives to β-ketoadipate. The presence of genes involved in both branches (catechol and protocatechuate *ortho*- cleavage) in all strains belonging to genomovar 1 (ATCC 17588^T, A1501, B1SMN1, DSM 4166, T13, and XLDN-R) is also noteworthy. In the genomes of strains AN10, 19SMN4, ST27MN3, DSM 10701, NF13, RCH2, SDM-LAC, TS44, and SP1402^T only the catechol branch was observed. Conversely, in the genome of strain 28a24, only the protocatechuate branch was found. None of the β-ketoadipate branches was found in strains CCUG 16156, KOS6, and MF28. Therefore, the catechol branch of the β-ketoadipate pathway is more frequent than the protocatechuate branch in *P. stutzeri* genomes.

1.1. 4-Hydroxyphenylpyruvate degradation gene cluster

The aromatic hydrocarbon 4-hydroxyphenylpyruvate is an intermediate of the catabolism of the amino acid tyrosine. In *P. putida* U 4-hydroxyphenylpyruvate is channeled to the central route of homogentisate by a 4-hydroxyphenylpyruvate dioxygenase (HppD). As shown in Figure 2.1, homogentisate is further catabolized producing fumarate and acetoacetate by the successive action of the products of operon *hmgABC* (Arias-Barrau *et al*., 2004).

Arias-Barrau and co-workers (2004) previously characterized the genes involved in the 4-hydroxyphenylpyruvate catabolism in *P. putida* U, by using a library of mutants unable to transform this compound. Moreover, these authors described the transcriptional regulation of *hmg* genes by the release of HmgR from the promotor of these genes in the presence of homogentisate.

We were able to identify sequences encoding the HppD, HmgA, HmgB, and HmgC enzymes in all the *P. stuzeri* and *P. balearica* genomes (see Supplementary Table S11). These protein sequences were compared by BLASTp with the CDSs involved in the degradation of tyrosine in *P. putida* U (Arias-Barrau *et al*., 2004). As shown in Table 2.3, the highest similarity was obtained for HppD (more than 50 % of amino acid identity and E-values lower than 1E-80 to its homologous in *P. putida* U). HmgC from *P. stuzeri* and *P. balearica* showed considerable homology with maleylacetoacetate isomerase (HmgC) of the model strain, showing in all cases E-values lower than 1E-50. On the contrary, homology between HmgA and HmgB sequences and those of *P. putida* U was not homogeneous for all *P. stutzeri* strains. As seen in Table 2.3, results showed a clear homology between the HmgA and HmgB

sequences of 5 strains: DSM 10701, 28a24, MF28, SDM-LAC, and TS44; with percentages of identity around 70 % or higher with *P. putida* U. The homology with other strains was lower. In those cases, the only strong homology found against the Swiss-Prot database was with proteins from *Vibrio cholerae* O1 biovar El Tor str.: N16961 for HmgA (Q9KSB4.1) and HmgB (Q9KSB3.1), respectively, with which they shared identities higher than 64 %. The function of these *V. cholerae* proteins, has not been described properly yet. Finally, homologous sequences for the regulatory protein HmgR of *P. putida U* was only identified in 5 strains (DSM 10701, 28a24, MF28, SDM-LAC, and TS44) which were the same strains with high levels of similarity with the model sequences HmgA and HmgB. According to the homologies described above for the proteins involved in 4-hydroxyphenylpyruvate degradation, it is clear that the 19 analyzed genomes present 2 different sequences coding for HmgA and HmgB, one very similar to the model *P. putida* U and another one more similar to *V. cholerae*. This suggests that they might have been acquired from different origins.

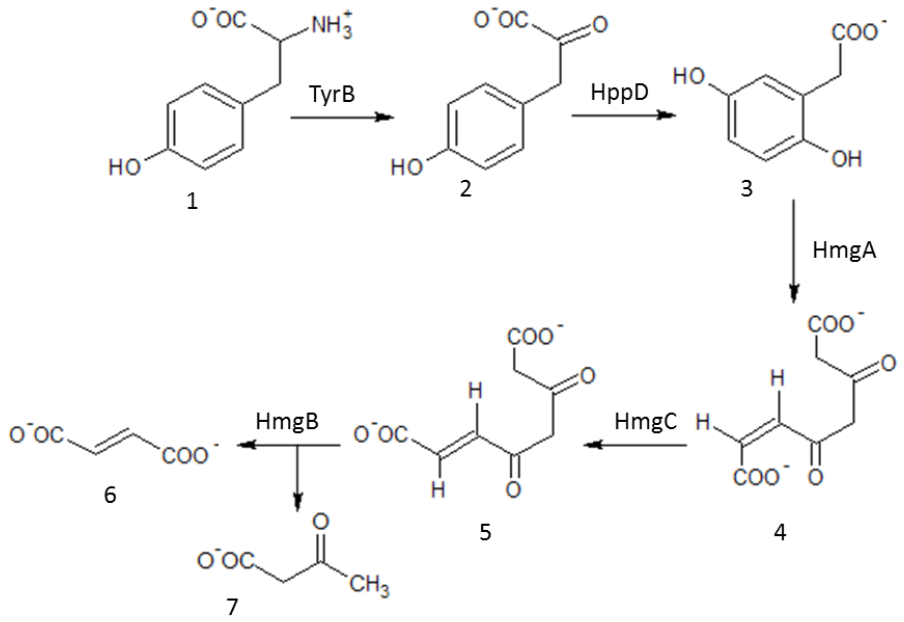

Figure 2.1. 4-Hydroxyphenylpyruvate degradation pathway in *Pseudomonas putida* U (Arias-Barrau *et al*., 2004). Enzymes: TyrB, tyrosine aminotransferase; HppD, 4-hydroxyphenylpyruvate dioxygenase; HmgA, homogentisate dioxygenase; HmgC, maleylacetoacetate isomerase; and HmgB, fumarylacetoacetate hydrolase. Compounds: 1, tyrosine; 2, 4-hydroxyphenylpyruvate; 3, homogentisate; 4, maleylacetoacetate; 5, fumarylacetoacetate; 6, fumarate; and 7, acetoacetate.

The organization of the genes involved in the degradation of 4-hydroxyphenylpyruvate was also compared with the model strain *P. putida* U. In this strain, genes are organized in two closely adjacent operons, the first encoding *hmgABC* and *hmgR*, transcribed in opposite directions; and a second for the *hppD* gene, located at an unknown distance from the other genes (see Figure 2.2; Arias-Barrau *et al*., 2004). The complete gene structure observed in *P. putida* U was not found in any of the strains analyzed in this study (see Figure 2.2). Two different gene organizations were found in *P. stuzeri* and *P. balearica* genomes. The first one, observed in strains which presented *hmgR* and the *hmgA* and *hmgB* genes more

Table 2.3. Percentages of identity and E-values obtained with BLASTp of the 4-hydroxyphenylpyruvate degradation CDSs of *P. putida* U (Arias-Barrau *et al*., 2004) and *P. stutzeri* and *P. balearica*.

		ATCC 17588^T	A1501	B1SMN1	DSM 4166	T13	XLDN-R	CCUG 16156	AN10	19SMN4	ST27MN3	DSM 10701	28a24	NF13	KOS6	MF28	RCH2	SDM-LAC	TS44	SP1402^T
HppD	% Id	54.3	54.6	54.8	54.6	54.6	54.8	54.1	54.6	54.4	54.4	54.3	53.9	53.4	54.4	83.8	54.4	82.4	54.8	54.8
	E-value	3E-107	7E-108	5E-109	4E-108	4E-108	4E-109	9E-108	4E-108	7E-109	7E-109	2E-108	3E-107	2E-107	1E-107	0	4E-108	3E-179	3E-109	8E-108
HmgA	% Id	24.6	24.6	24.6	24.6	24.6	24.6	24.6	24.6	24.3	24.6	74.5	76.8	24.9	24.6	75.4	24.3	75.9	77.1	24.6
	E-value	5E-17	8E-17	1E-16	8E-17	8E-17	5E-17	3E-17	1E-16	5E-16	5E-17	0	0	1E-17	6E-17	0	2E-16	0	0	3E-16
HmgC	% Id	48.1	48.6	47.7	48.6	48.6	47.7	48.6	48.1	48.6	49.1	48.1	48.1	48.1	47.7	49.1	49.1	49.5	47.4	46.5
	E-value	2E-51	9E-52	2E-51	2E-52	2E-52	2E-51	2E-52	2E-51	1E-52	7E-53	7E-50	5E-50	6E-52	2E-50	3E-52	5E-53	1E-51	4E-52	2E-49
HmgB	% Id	23.9	23.3	22.7	23.3	23.3	23.9	23.3	23.9	23.3	23.3	69.3	72.8	23.3	23.3	72.3	23.3	71.8	71.4	23.3
	E-value	2E-04	4E-04	9E-04	4E-04	4E-04	2E-04	5E-04	6E-05	5E-04	3E-04	7E-180	0	3E-04	9E-04	0	3E-04	0	0	2E-03
HmgR[a]	% Id	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	62.5	62.7	n.p.	n.p.	63.6	n.p.	63.8	65.3	54.8
	E-value	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	4E-83	8E-82	n.p.	n.p.	4E-81	n.p.	6E-83	3E-83	8E-108

a: n.p., not predicted.

similar to *P. putida* U (DSM 10701, 28a24, MF28, SDM-LAC, and TS44), presented the same gene structure of genes *hmgR*, *hmgA*, and *hmgB* than *P. putida* U, and *hppD* was also located separately (see Figure 2.2). However, in these strains *hmgC* was not located downstream *hmgB*, as in *P. putida* U. A second gene organization was identified in the other *P. stuzeri* and *P. balearica* genomes, which lacked of gene *hmgR* and whose *hmgA* and *hmgB* genes differ from *P. putida* U. In these strains, an *hppDhmgABC* gene cluster was observed (see Figure 2.2), equal to the structure present in the genome of *Vibrio cholerae* O1 biovar El Tor str.: N16961 (coordinates 1431464..1435372 from AE003852.1).

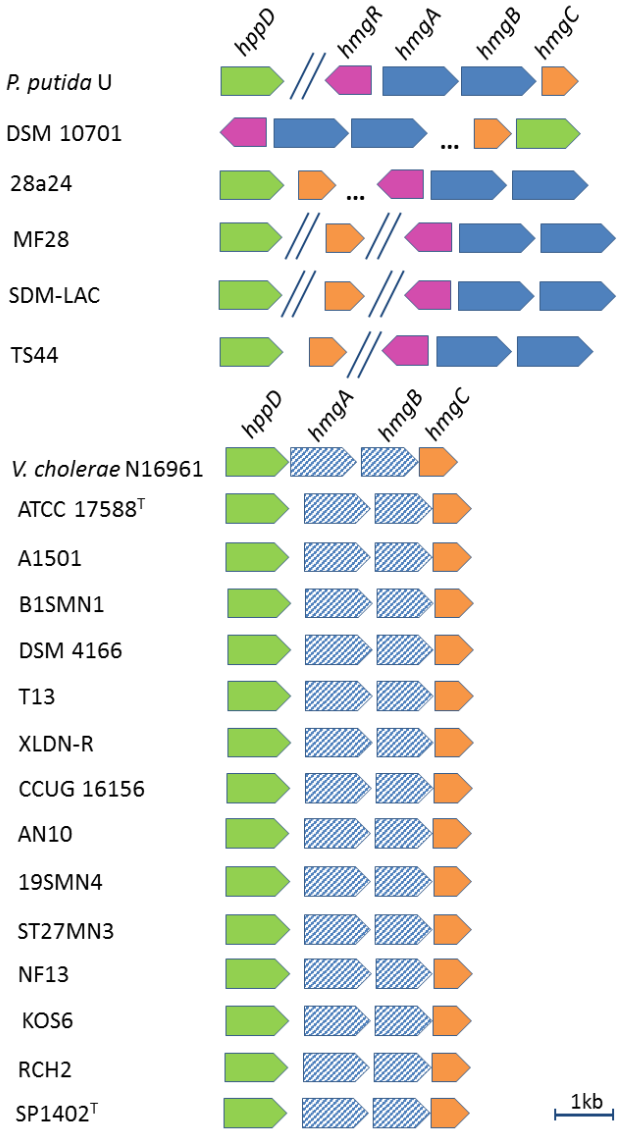

Figure 2.2. Structure of genes involved in 4-hydroxyphenylpyruvate degradation in *Pseudomonas putida* U (Arias-Barrau *et al.*, 2004), *Vibrio cholerae* O1 biovar El Tor str. N16961, and *P. stutzeri* and *P. balearica* strains. Common gene structures between the model strain and *P. stutzeri* and *P. balearica* genomes are represented with the same color. Striped genes represent those with identities lower than 30 % with the model strain *P. putida* U. Unknown distances between genes are shown with two forward slashes and long distances (>10 kb) within the same contig are represented with dots.

1.2. 4-Hydroxybenzoate degradation gene cluster

The compound 4-hydroxybenzoate is a phenolic derivative of benzoate. *P. stutzeri* strain A1501 has been previously described as a 4-hydroxybenzoate degrader, via the protocatechuate branch of the β-ketoadipate pathway. Li and co-workers (2010) hypothesized that degradation of 4-hydroxybenzoate started with the addition of a

hydroxyl group by the 4-hydroxybenzoate 3-monooxygenase (PobA) resulting in protocatechuate. Later, the aromatic ring of this compound would be cleaved in *ortho* position by protocatechuate 3,4-dioxygenase (PcaGH), producing β-carboxymuconate. Then this would be transformed by the successive action of several enzymes (PcaB, PcaC, and PcaD), resulting in β-ketoadipate. Afterwards, this compound would be channeled to TCA intermediates by the action of PcaIJ and PcaF (see Figure 2.3; Harwood and Parales, 1996). A genome-wide analysis of the aromatic catabolism of *P. stutzeri* A1501 revealed the presence of genes involved in 4-hydroxybenzoate degradation. Moreover, they demonstrated that PcaD and the transcriptional regulator PcaR were essential for the growth with 4-hydroxybenzoate, since mutants in *pcaR* and *pcaD* genes failed to grow with this compound as carbon source (Li *et al*., 2010).

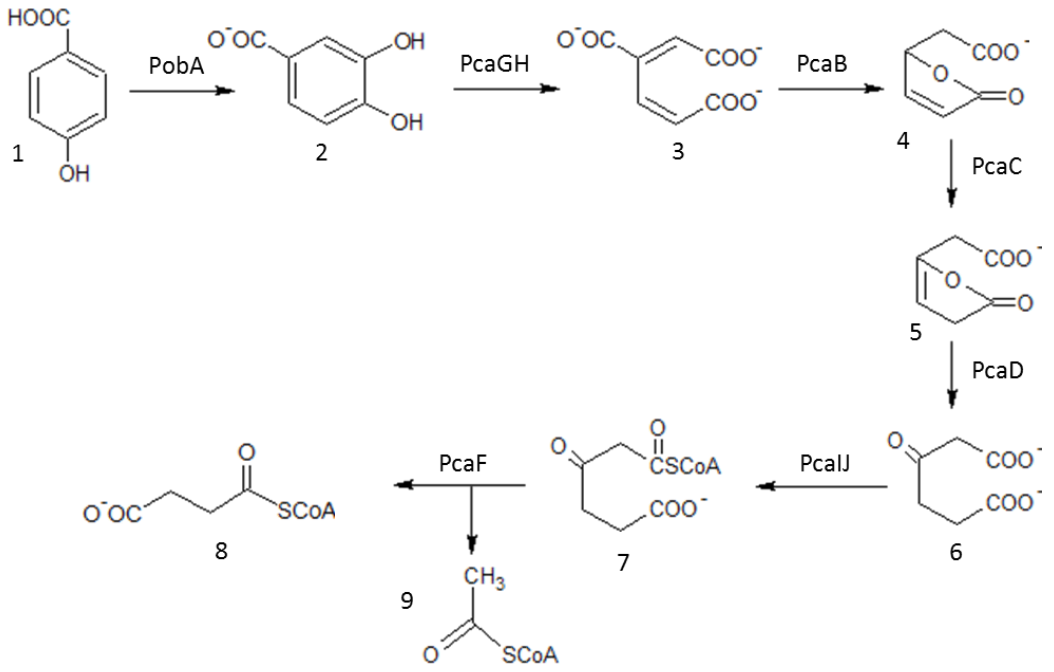

Figure 2.3. 4-Hydroxybenzoate degradation pathway in *P. stutzeri* A1501 (Li *et al*., 2010). Enzymes: PobA, 4-hydroxybenzoate 3-monooxygenase; PcaGH, protocatechuate 3,4-dioxygenase; PcaD, 3-oxoadipate enol-lactonase; PcaB, 3-carboxy-cis,cis-muconate cycloisomerase; PcaC, 4-carboxymuconolactone decarboxylase; and PcaIJ, 3-oxoadipate CoA-transferase. Compounds: 1, 4-hydroxybenzoate; 2, protocatechuate; 3, β-carboxymuconate; 4, γ-carboxymuconolactone; 5, β-ketoadipate enol-lactone; 6, β-ketoadipate; 7, β-ketoadipyl-CoA; 8, succinyl-CoA; and 9, acetyl-CoA.

CDSs involved in 4-hydroxybenzoate degradation were identified, with BLASTp using the A1501 CDSs as a model, in 6 *P. stuzeri* strains: ATCC 17588^T, B1SMN1, DSM 4166, T13, XLDN-R, and 28a24 (see Supplementary Table S11), except from PobA, which was not annotated in ATCC 17588^T. Surprisingly, we could identify the transcriptional regulator of PobA (PobR) in the genome of this strain. Therefore, we performed a more exhaustive research of *pobA* in the ATCC 17588^T genome, which revealed the presence of a 272-bp sequence with a 99 % of identity with the *pobA* gene from A1501. The remaining 912-bp sequence of *pobA* from A1501 was not found in the genome of strain ATCC 17588^T, concluding that this gene is truncated in the genome of this strain. Finally, the genome of strain 28a24 presented a different localization of genes *pcaB*, *pcaC*, *pobA*, and *pobR* (see Figure 2.4). As seen in Table 2.4, in the genome of the first 5 strains (all belonging to genomovar 1) percentages of

identity greater than 90 % were obtained for all proteins. In contrast, in the case of strain 28a24 (genomovar 11), identities were generally lower than other strains (from 65 to 90 %), although they still presented a clear homology with the model A1501.

Table 2.4. Percentages of identity and E-values obtained with BLASTp of the 4-hydroxybenzoate degradation CDSs of *P. stutzeri* A1501 (Li *et al*., 2010) and other strains of *P. stutzeri*.

		ATCC 17588^T	B1SMN1	DSM 4166	T13	XLDN-R	28a24
PobA[a]	% Id	t	98.0	97.7	99.0	99.0	81.9
	E-value	t	0	0	0	0	0
PcaG	% Id	97.0	100	96.0	97.5	98.5	78.6
	E-value	7E-117	8E-121	3E-116	3E-118	2E-119	2E-96
PcaH	% Id	97.5	100	98.3	98.7	97.9	87.0
	E-value	2E-142	1E-145	6E-144	3E-144	2E-142	2E-129
PcaB	% Id	98.7	100	98.0	98.7	98.7	80.9
	E-value	0	0	0	0	0	0
PcaC	% Id	99.2	100	98.5	100	99.2	90.8
	E-value	1E-74	6E-75	5E-74	6E-75	1E-74	4E-69
PcaD	% Id	98.5	100	99.6	98.9	98.5	75.2
	E-value	7E-152	5E-154	3E-153	7E-152	7E-152	3E-116
PcaI	% Id	98.3	100	99.7	99.7	99.3	88.0
	E-value	4E-166	8E-169	1E-168	5E-168	3E-168	4E-150
PcaJ	% Id	100	100	100	99.6	99.6	89.2
	E-value	5E-151	5E-151	5E-151	2E-150	2E-150	4E-136
PcaF	% Id	98.0	100	98.0	98.0	97.3	88.8
	E-value	0	0	0	0	0	0
PcaQ	% Id	99.6	100	99.67	99.3	99.3	80.2
	E-value	2E-139	2E-172	9E-172	1E-171	1E-171	2E-139
PcaR	% Id	100	100	100	99.6	99.6	90.36
	E-value	2E-151	3E-158	2E-151	2E-157	1E-157	2E-143
PobR	% Id	98.9	99.3	98.9	98.6	98.6	65.6
	E-value	6E-162	2E-162	6E-162	2E-161	2E-161	7E-110

a: t, truncated.

The organization of the genes involved in the 4-hydroxybenzoate degradation pathway was previously described in A1501 by Li and co-workers (2010). In this strain, genes involved in the transformation of protocatechuate to TCA intermediates are organized as follows: a *pcaQHG* cluster, two different open reading frames (orfs) with unknown role in this pathway (an ATPase and a permease), and a *pcaRIJFBDC* cluster (see Figure 2.4). Harwood and Parales (1996) described the regulation of most of the genes of the protocatechuate branch of the β-ketoadipate pathway by PcaR in *P. putida*. Additionally, these authors described the transcriptional regulation induction of the *pca* genes by PcaR and PcaQ in *Agrobacterium tumefaciens*. The presence of genes coding for PcaQ have been also reported in different *Pseudomonas* strains, such as *P. aeruginosa* PAO1 (Jiménez *et al*., 2002) and *P. putida* KT2440 (Li *et al*., 2010). Therefore, the transcription of *pca* genes in A1501 might be regulated by PcaR and PcaQ (Li *et al*., 2010). Finally, genes involved in the transformation of 4-hydroxybenzoate to protocatechuate, PobA and its transcriptional regulator PobR, are separated in the A1501 genome (at 232 kb and 925 kb from *pcaC* respectively). As shown in Figure 2.4, in four of the *P. stutzeri* strains (B1SMN1, DSM 4166, T13, and XLDN-R) the gene structure of A1501 was fully maintained. A similar structure was observed in the genome of strain ATCC 17588^T, on which the gene encoding for the 4-hydroxybenzoate-3-monooxygenase (PobA) was interrupted. Finally, the genome of strain

28a24 presented a different localization of genes *pcaB*, *pcaC*, *pobA*, and *pobR* (see Figure 2.4).

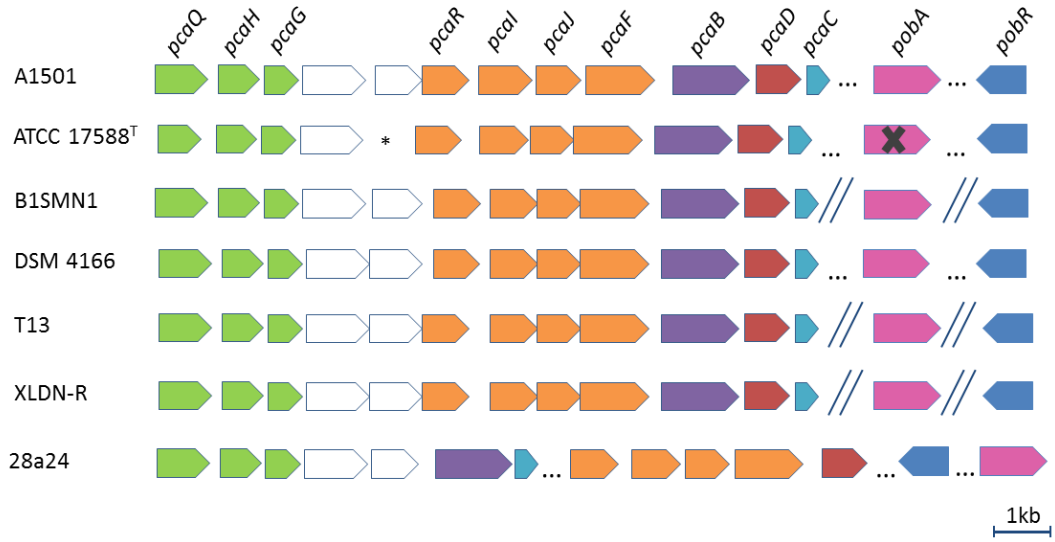

Figure 2.4. Structure of genes involved in 4-hydroxybenzoate degradation in *P. stutzeri* A1501 (Li *et al*., 2010) and other *P. stutzeri* strains. Common gene structures between the model strain and *P. stutzeri* genomes are represented with the same color. Genes in white have not been characterized as involved in this pathway. Disrupted genes are marked with an "X". The unannotated CDSs are shown with an asterisk. Unknown distances between genes are shown with two forward slashes and long distances (>10 kb) within the same contig are represented with dots.

1.3. Benzoate degradation gene cluster

The ability of *P. stutzeri* A1501 to degrade benzoate has been previously described, and the degradation of this compound via the β-ketoadipate pathway has been characterized by Li and co-workers (2010). In this strain, benzoate is catabolized to catechol by the successive action of the products of operon *benABCD*, a benzoate 1,2dioxygenase. Afterwards, the aromatic ring of catechol is cleaved in *ortho* position by a catechol 1,2-dioxygenase (CatA) resulting in cis,cis-muconate. Then, cis,cis-muconate is transformed to β-ketoadipate by the successive action of CatB, CatC, and PcaD. As mentioned for the 4-hydroxybenzoate degradation, β-ketoadipate is later catabolized by the action of PcaIJ and PcaF to TCA intermediates (see Figure 2.5; Harwood and Parales, 1996).

Genes involved in benzoate degradation has been identified in *P. stutzeri* A1501 (Li *et al*., 2010). Gene disruption analysis and genetic complementation tests demonstrated that genes *benR*, *pcaR*, and *pcaD* were required for the growth of A1501 in benzoate. BenR protein functions as a transcriptional activator of the *ben* operon in response to benzoate, PcaR as a regulator of the *pca* operon, while transcription of the *catBC* promoter was activated directly in response to benzoate in A1501 (Li *et al*., 2010).

In order to identify these genes in other *P. stutzeri* genomes, a comparison with BLASTp was performed, using the *pca-ben-cat* operons from A1501 as a model. These genes were identified (E-values below 1E-50 and percentages of identity greater than a 60 %, see Table

2.5) in most of the strains (see Table 2.5 and Supplementary Table S11). Exceptions were strains CCUG 16156, 28a24, KOS6, and MF28.

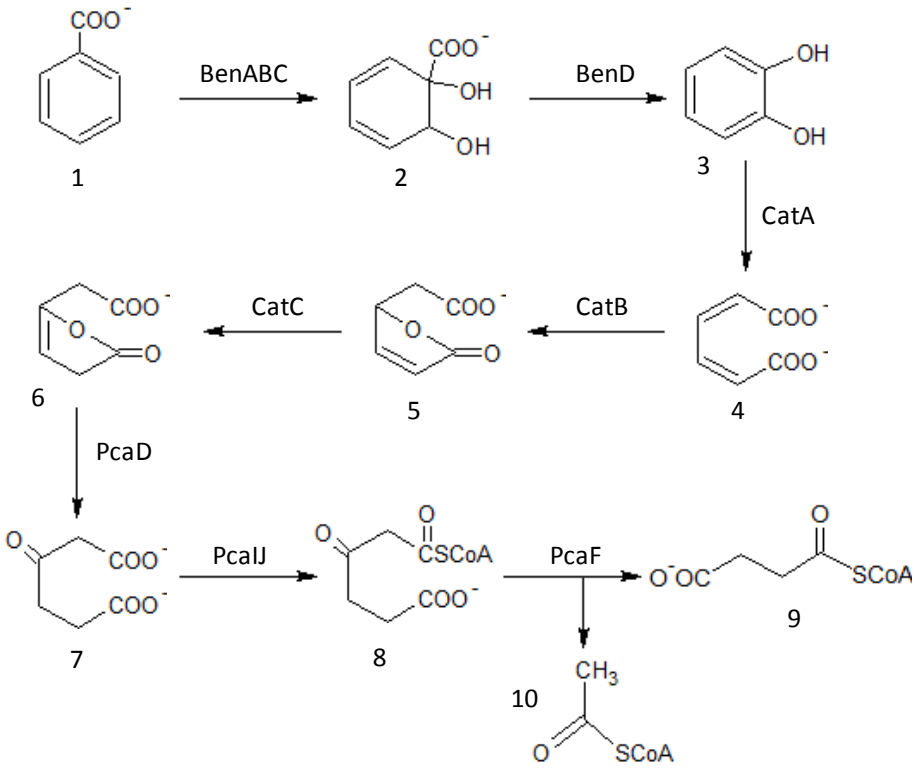

Figure 2.5. Benzoate degradation pathway in *P. stutzeri* A1501 (Li *et al*., 2010). Enzymes: BenABC, benzoate 1,2-dioxygenase; BenD, dihydroxycyclohexadiene carboxylate dehydrogenase; CatA, catechol 1,2-dioxygenase; CatB, muconate cycloisomerase; CatC, muconolactone D-isomerase; PcaD, 3-oxoadipate enol-lactonase; PcaIJ, 3-oxoadipate CoA-transferase; and PcaF, 3-oxoadipil-CoA thiolase. Compounds: 1, benzoate; 2, benzoate diol; 3, catechol; 4, cis,cis-muconate; 5, muconolactone; 6, β-ketoadipate enol-lactone; 7, β-ketoadipate; 8, β-ketoadipyl-CoA; 9, succinyl-CoA; and 10, acetyl-CoA.

On the other hand, gene sinteny was also analyzed. In the genome of strain A1501, the genes involved in benzoate degradation are organized in two separate gene structures *ben-cat* and *pca,* separated by 0.4 Mb (Li *et al*., 2010). As shown in Figure 2.6, the *ben-cat* gene cluster is constituted by: the transcriptional regulator *benR*; the *benABCD* genes; a gene encoding a transporter involved in benzoate degradation, *benK* (Collier *et al*., 1997); the *catBCA* genes, and two other genes involved in benzoate uptake, *benE* and *benF* (Haussmann *et al*., 2009). The organization of these genes was highly conserved in *P. stutzeri* and *P. balearica* genomes, as shown in Figure 2.6. In all the studied genomes the structure of the *ben-cat* gene cluster was basically identical. Exceptions for that are the disruption of *benD, benR, and benK* by a stop codon in ATCC 17588^T, AN10, and 19SMN4, respectively; and the disruption of *benE* in T13 by the end of the contig. The genome of SDM-LAC presented a histidine kinase encoding-gene (WP_029404594.1, see Supplementary Table S11) between *benR* and *benA*, and lacked gene *benF*.

As mentioned in section 1.2, strains ATCC 17588^T, B1SMN1, DSM 4166, T13, and XLDN-R showed the same *pca* gene structure than A1501. The other strains (AN10, 19SMN4, ST27MN3, DSM 10701, NF13, RCH2, SDM-LAC, TS44, and SP1402^T) lacked of the initial

Table 2.5. Percentages of identity and E-values obtained with BLASTp of the benzoate degradation CDSs of A1501 (Li *et al.*, 2010) and other *P. stutzeri* strains and *P. balearica*.

		ATCC 17588^T	B1SMN1	DSM 4166	T13	XLDN-R	AN10	19SMN4	ST27MN3	DSM 10701	NF13	RCH2	SDM-LAC	TS44	SP1402^T
BenA	% Id	100	99.8	99.3	100	99.5	95.7	95.9	95.4	91.4	96.2	86.8	94.7	95.9	87.7
	E-value	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BenB	% Id	100	100	100	100	100	88.9	87.7	88.9	85.8	90.1	88.3	87.0	94.4	85.8
	E-value	2E-98	2E-98	2E-98	2E-98	2E-98	4E-89	2E-88	2E-89	3E-87	4E-90	4E-90	9E-88	3E-94	9E-87
BenC	% Id	98.8	98.5	98.8	98.2	98.5	87.8	88.7	87.8	86.0	87.	93.5	88.4	92.3	82.4
	E-value	0	0	0	0	0	2E-176	6E-178	4E-176	3E-173	3E-175	0	2E-178	0	4E-164
BenD[a]	% Id	t	98.8	99.6	99.2	99.2	93.4	91.1	93.8	94.2	93.1	93.1	93.8	95.7	86.1
	E-value	t	2E-148	3E-149	6E-149	2E-148	1E-140	2E-138	2E-140	1E-142	7E-141	5E-140	2E-142	1E-144	9E-127
CatA	% Id	95.8	98.2	97.5	97.2	97.2	91.6	91.2	91.2	86.9	91.2	87.3	84.5	91.8	83.4
	E-value	4E-163	6E-166	3E-165	5E-165	5E-165	1E-132	1E-155	1E-155	2E-148	3E-156	1E-147	1E-143	2E-155	3E-140
CatB	% Id	98.1	97.05	98.1	99.5	97.9	93.3	96.0	96.0	91.2	94.9	94.6	92.0	93.6	87.4
	E-value	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CatC	% Id	99.0	99.0	99.0	97.9	97.9	91.6	93.7	93.7	92.7	96.9	96.9	92.6	94.8	83.2
	E-value	2E-54	9E-55	2E-54	4E-54	6E-54	1E-49	3E-51	3E-51	9E-52	2E-53	2E-53	2E-50	2E-52	2E-44
PcaD	% Id	98.5	100	99.6	98.9	98.5	79.8	82.4	82.4	75.5	79.8	82.8	74.8	75.2	73.3
	E-value	9E-152	8E-154	5E-153	9E-152	9E-152	2E-123	9E-129	9E-129	2E-120	8E-124	3E-129	2E-116	7E-118	4E-112
PcaI	% Id	100	100	100	99.6	99.6	96.4	95.0	96.3	91.4	96.1	93.9	91.1	96.1	93.2
	E-value	3E-151	4E-158	3E-151	3E-157	2E-157	9E-153	1E-150	2E-146	9E-145	2E-152	4E-149	2E-144	6E-152	4E-155
PcaJ	% Id	100	100	100	99.6	99.6	93.5	93.4	93.4	93.0	98.1	93.4	91.4	96.2	94.2
	E-value	8E-151	8E-151	8E-151	2E-150	2E-150	7E-142	1E-139	1E-139	7E-142	2E-148	4E-139	2E-138	4E-146	3E-144
PcaF	% Id	98.0	100	98.0	98.0	97.3	89.5	90.8	90.8	89.0	91.3	90.8	89.5	91.8	87.8
	E-value	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BenR[a]	% Id	100	100	100	100	99.7	t	92.7	93.1	84.2	94.3	92.7	82.3	92.1	73.9
	E-value	0	0	0	0	0	t	7E-179	3E-179	4E-162	0	7E-178	1E-163	4E-177	3E-139
PcaR	% Id	100	100	100	99.6	99.6	96.4	95.0	96.3	91.4	96.1	93.9	91.1	96.1	88.5
	E-value	3E-151	4E-158	3E-151	3E-157	2E-157	9E-153	1E-150	2E-146	9E-145	2E-152	4E-149	2E-144	6E-152	1E-135
BenK[a]	% Id	98.2	98.9	99.1	99.1	98.9	95.1	t	93.9	87.2	95.5	92.8	90.8	88.4	80.0
	E-value	0	0	0	0	0	0	t	0	0	0	0	0	0	0
BenE	% Id	96.0	99.3	99.8	100	98.8	88.2	89.2	89.2	82.5	94.0	90.2	86.7	90.9	76.0
	E-value	0	0	0	1E-58	0	0	0	0	0	0	0	0	0	5E-175
BenF[b]	% Id	92.9	98.3	98.1	98.3	97.0	80.7	80.9	80.9	69.9	91.8	82.8	n.p.	74.2	70.4
	E-value	0	0	0	0	3E-172	0	0	0	3E-175	0	0	n.p.	0	2E-174

a: t, truncated.

b: n.p.: not predicted

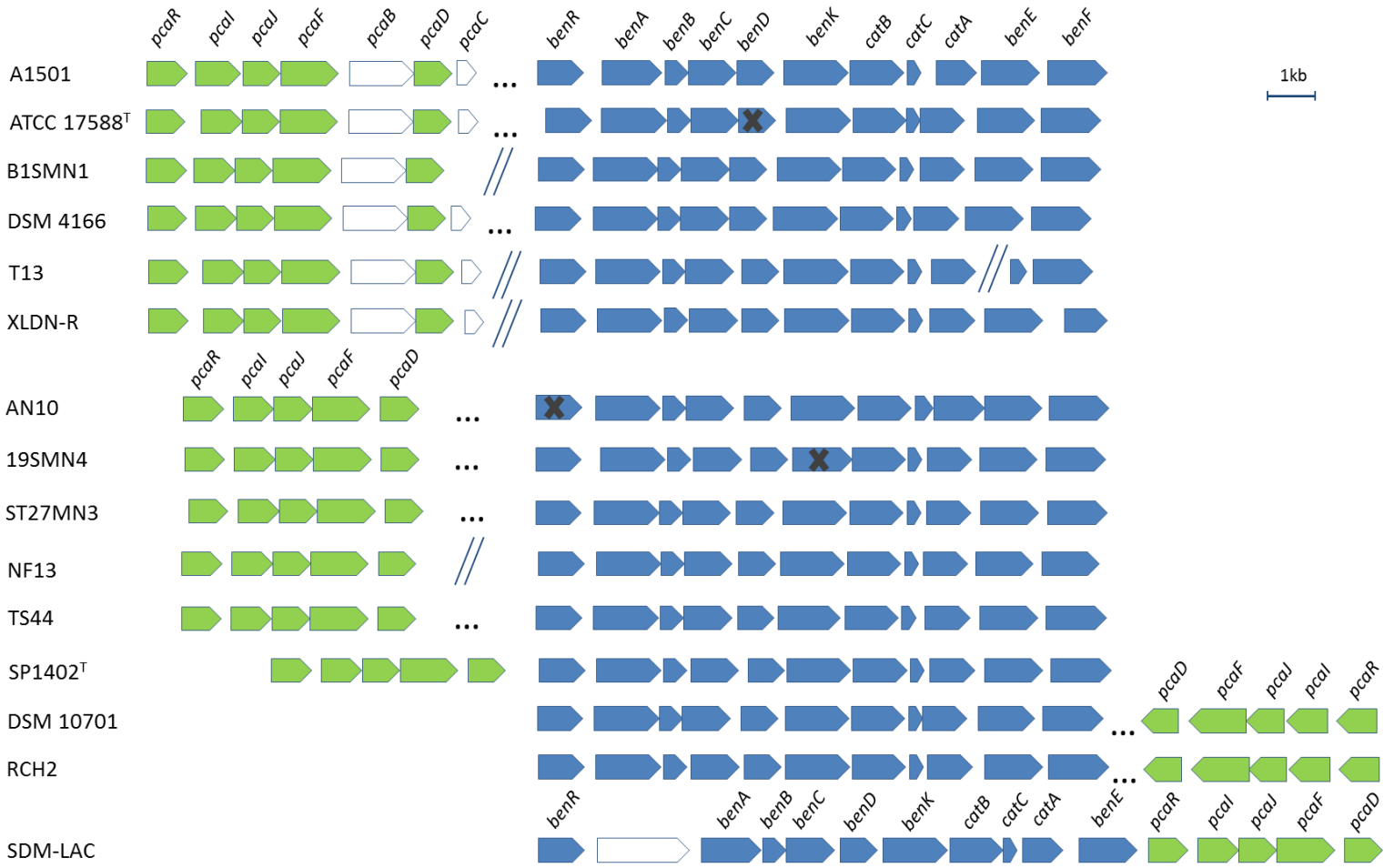

Figure 2.6. Structure of genes involved in benzoate degradation in *P. stutzeri* A1501 (Li *et al.*, 2010) and other *P. stutzeri* strains and *P. balearica*. Common gene structures between the model strain and *P. stutzeri* and *P. balearica* genomes are represented with the same color. Genes in white have not been characterized as involved in this pathway. Disrupted genes are marked with an "X". Unknown distances between genes are shown with two forward slashes and long distances (>10 kb) within the same contig are represented with dots.

genes involved in the protocatechuate branch of β-ketoadipate pathway, including *pcaB* and *pcaC*, (see section 1.2). But they presented the genes involved in the transformation of β-ketoadipate (*pcaRIJFD*), and hence the catechol branch of β-ketoadipate pathway. In the *P. stutzeri* genomes the *pca* gene cluster was generally not located near the *ben-cat* cluster. Exceptions were *P. stutzeri* SDM-LAC and *P. balearica* SP1402^T, in which the *ben-cat* gene cluster was located upstream and downstream *pcaRIJFD,* respectively. In RCH2 and DSM 10701 genomes, *pca* gene cluster was located downstream instead of upstream the *ben-cat* gene cluster. In the genomes of strains T13, XLDN-R, and NF13 we were not able to determine the genetic distance between the *pca* and the *ben-cat* gene clusters, as they are located in separate contigs.

The structure of genes involved in benzoate degradation suggests that the studied strains could catabolize benzoate through catechol *ortho*-cleavage (by catechol 1,2-dioxygenase). In this sense, Rosselló-Mora and co-workers (1994) quantified catechol 1,2-dioxygenase and catechol 2,3-dioxygenase activities of strains B1SMN1 and 19SMN4 growing with benzoate. They observed that although B1SMN1 presented mainly catechol 1,2-dioxygenase activity, this activity was not observed in 19SMN4 cultures, and only catechol 2,3-dioxygenase activity was quantified in 19SMN4. However, our genomic analysis cannot explain the lack of catechol 1,2-dioxygenase activity in 19SMN4 cultures growing with benzoate, since the genome of 19SMN4 only lacked the gene encoding a putative transporter (*benK*).

1.4. Homoprotocatechuate degradation gene cluster

Homoprotocatechuate is a monoaromatic compound that acts as an intermediate in the central degradation pathways of tyramine and dopamine and an intermediate of the degradation of 4-hydroxyphenylacetic acid. The degradation pathway of homoprotocatechuate in *Pseudomonas putida* U has been previously characterized by Arcos and co-workers (2010). In this pathway, homoprotocatechuate is initially attacked by homoprotocatechuate 2,3-dioxygenase (HpaD), which breaks the aromatic ring forming 5-carboxymethyl-2-hydroxymuconic semialdehyde (see Figure 2.7). Then, this product is transformed by HpaE to 5-carboxymethyl-2-hydroxymuconic acid. Afterwards, this is converted by the successive action of HpaF, HpaG, HpaH, and HpaI into pyruvate and succinate semialdehyde. Eventually, succinate semialdehyde is transformed by succinate semialdehyde dehydrogenase (GabD) to succinate (Arcos *et al*., 2010).

A search of genes involved in homoprotocatechuate degradation in the 19 genomes of *P. stutzeri* and *P. balearica* was performed by BLASTp, using the previously described *P. putida* U pathway as a model (Arcos *et al.,* 2010). The homoprotocatechuate degradation genes were identified only in the genome of strain KOS6 (see Supplementary Table S11). The identified sequences of this strain showed high homology with the model sequences, with percentages of identity ranging from 57 to 92 % and, in all cases, with E-values below 1E-60 (see Table 2.6).

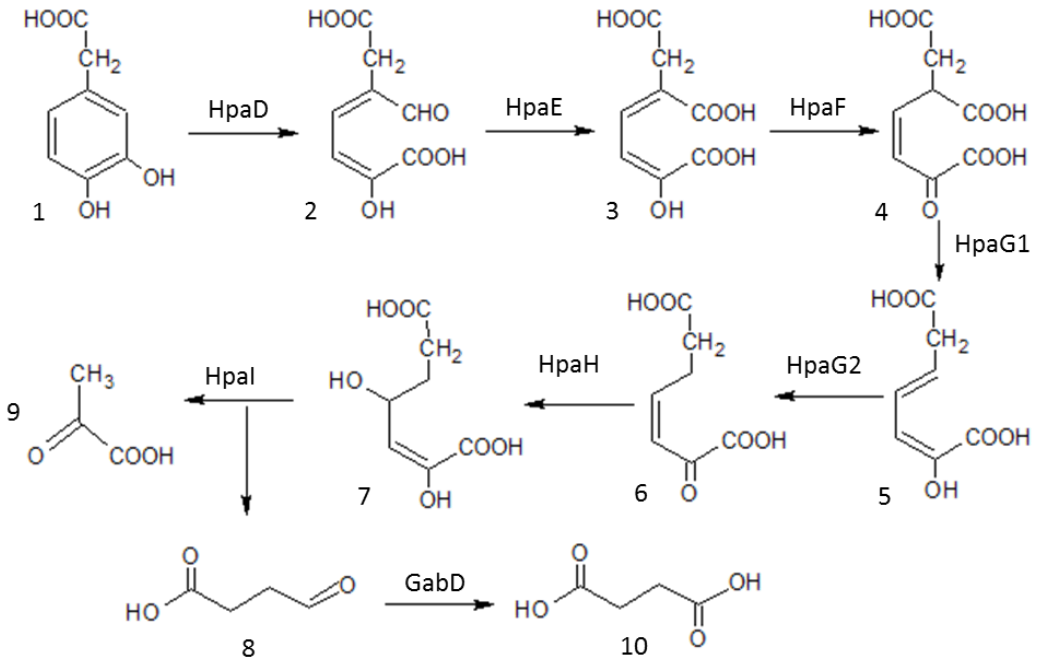

Figure 2.7. Homoprotocatechuate degradation pathway in *P. putida* U (Arcos *et al*., 2010). Enzymes: HpaD, homoprotocatechuate 2,3-dioxygenase; HpaE, 5-carboxy-2-hydroxymuconate semialdehyde dehydrogenase; HpaF, 5-carboxymethyl-2-hidyroxymuconate isomerase; HpaG1 5-oxo-pent-3-ene-1,2,5-tricarboxylate decarboxylase; HpaG2, 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase; HpaH, 2-oxo-hepta-3-ene-1,7-dioate hydratase; HpaI, 2,4-dihydroxyhepta-2-ene-1,7-dioate aldolase; and GabD, succinate semialdehyde dehydrogenase. Compounds: 1, homoprotocatechuate; 2, 5-carboxymethyl-2-hydroxymuconic semialdehyde; 3, 5-carboxymethyl-2-hydroxymuconic acid; 4, 5-oxo-pent-3-en-1,2,5-tricarboxylic acid; 5, 2-hydroxy-hept-2,4-diene-1,7-dioic acid; 6, 2-oxo-hept-3-en-1,7-dioic acid; 7, 2,4-dihydroxy-hepta-2-ene-1,7-dioic acid; 8, succinate semialdehyde; 9, pyruvate; and 10, succinate.

The genes coding for proteins involved in homoprotocatechuate degradation in *P. putida* U are arranged in a single gene structure (*hpaG1G2EDFXHI*). Furthermore, as shown in Figure 2.8, in *P. putida* U near this genes we find also *hpaBC* genes, involved in the transformation of 4-hydroxyphenylacetic acid to homoprotocatechuate; three transcriptional regulation genes, *hpaY*, *hpaA, and hpaR*, as well as *tetR*, which function remains unknown (Arcos *et al*., 2010). The genome of strain KOS6 also presented a gene cluster similar to *hpaG1G2EDFXHI*; although it lacked gene *hpaX*, which encodes for a 4-hydroxyphenylacetic acid transporter (see Figure 2.8). It also lacked *hpaBC*; and the transcriptional regulators *hpaY*, *hpaA,* and *hpaR*. A gene encoding for a transcriptional regulator (GntR) was annotated near the homoprotocatechuate degradation gene cluster in KOS6 genome. No homology was found (E-values above 1) between GntR from KOS6 and any of the three homoprotocatechuate transcriptional regulators from *P. putida* U (HpaA, HpaR, and HpaY). Additionally, downstream *hpaI* a gene annotated as a succinate-semialdehyde dehydrogenase (*gabD*), which transforms succinate semialdehyde to succinate (see Figure 2.8) was present in the genome of strain KOS6. Any *gabD* gene has been described for *P. putida* U, since its genome is not completely sequenced yet. Therefore, KOS6 seems to present the essential genes for the degradation of homoprotocatechuate but not for the degradation of 4-hydroxyphenylacetic acid.

Table 2.6. Percentages of identity and E-values obtained with BLASTp of the homoprotocatechuate degradation CDSs of *P. putida* U (Arcos *et al*., 2010) and *P. stutzeri* KOS6.

	% Id	E-value
HpaD	57.3	7E-116
HpaE	92.4	0
HpaF	69.4	4E-67
HpaG1	71.2	1E-116
HpaG2	85.6	2E-148
HpaH	86.9	2E-175
HpaI	79.8	1E-158

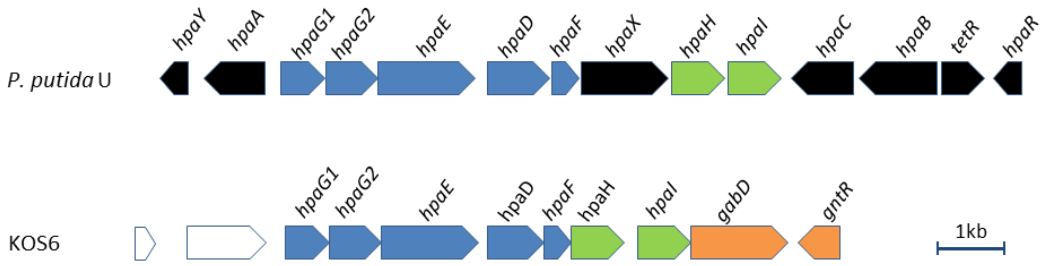

Figure 2.8. Structure of genes involved in homoprotocatechuate degradation in *P. putida* U (Arcos *et al*., 2010) and *P. stutzeri* KOS6. The two common gene structures between the model strain and KOS6 are represented in blue or green. Genes in black represent those genes from the model strain that have not been found in KOS6. Genes in orange represent those genes from the model strain that have not been found in KOS6. Genes in white have not been characterized as involved in this pathway.

1.5. Carbazole degradation gene cluster

Carbazole is a heterocyclic aromatic compound derived from coal tar and shale oil, whose mutagenic and toxic effects are well known (Jha and Bharti, 2002). *P. stutzeri* XLDN-R has been described as carbazole degrader, since it was isolated from a soil sample enriched with carbazole as sole carbon and nitrogen source (Liu *et al*., 2012). These authors described the presence of the genes involved in carbazole degradation in the genome of strain XLDN-R, which were homologous to the *car* gene structure of *P. resinovorans* CA10. The carbazole degradation pathway of *P. resinovorans* CA10 has been previously characterized by Nojiri (2012). As shown in Figure 2.9, in strain CA10 carbazole is oxygenated at angular (C9a) and adjacent (C1) carbons in a *cis* configuration. This type of dioxygenation is termed angular dioxygenation, and is catalyzed by carbazole-1,9a-dioxygenase (Nojiri, 2012). The angular dioxygenation produce an unstable hemiaminal compound, which is spontaneously cleaved to form 2'-aminobiphenyl-2,3-diol. This product is later transformed to anthranilate and 2-hydroxypenta-2,4-dienoate through *meta*-cleavage, and hydrolysis. The anthranilate produced is attacked by an anthranilate 1,2-dioxygenase producing catechol, which is degraded by the catechol branch of the β-ketoadipate pathway (*ortho*-cleavage). On the other hand, 2-hydroxypenta-2,4-dienoate is converted to pyruvate and acetyl-CoA by the action of CarD, CarE, and CarF (see Figure 2.9).

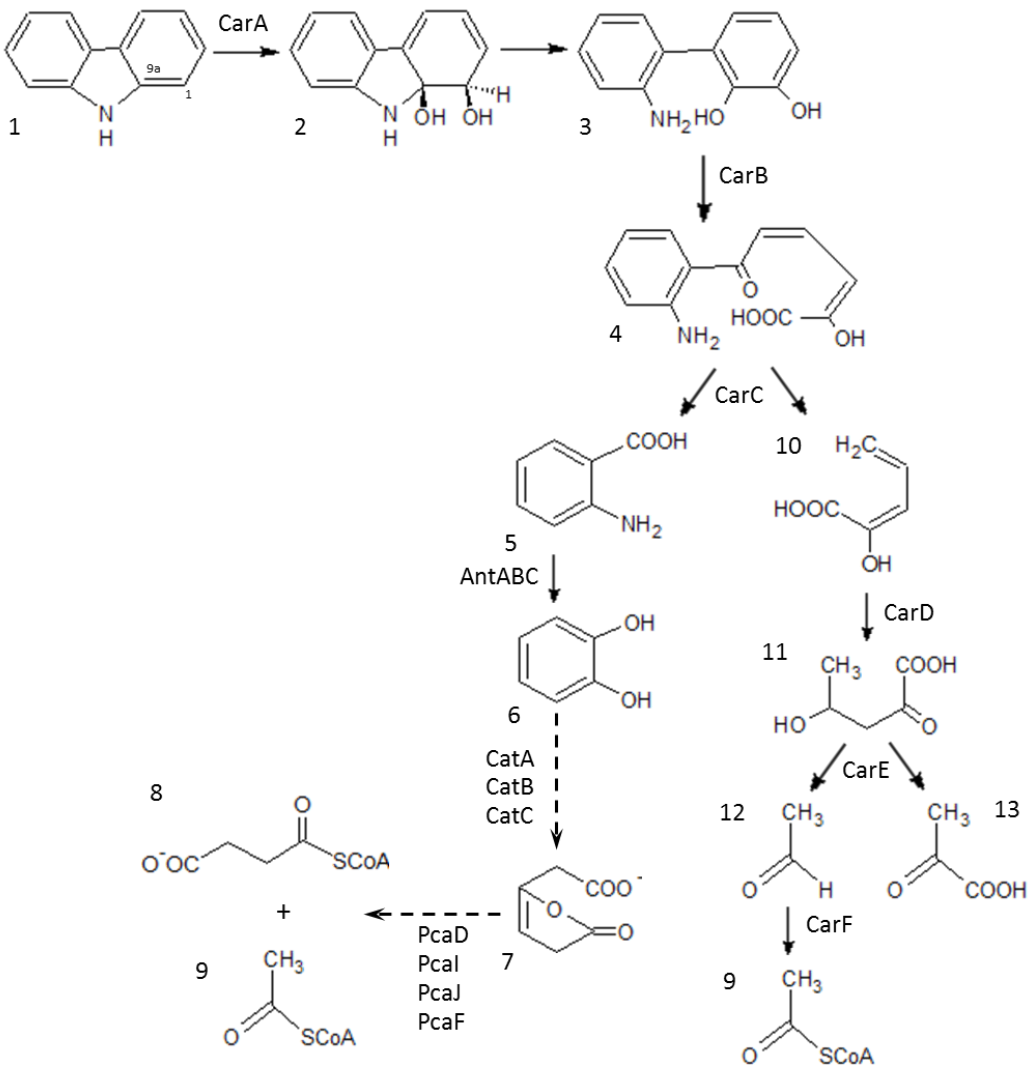

Figure 2.9. Carbazole degradation pathway in *P. resinovorans* CA10 (Nojiri, 2012). Enzymes: CarA, carbazole 1,9a-dioxygenase; CarB, 2'-aminobiphenyl-2,3-diol 1,2-dioxygenase; CarC, 2-hydroxy-6-oxo-6-(2'-aminophenyl)-hexa-2,4-dienoic acid hydrolase; AntABC, anthranilate 1,2-dioxygenase; CatA, catechol 1,2-dioxygenase; CatB, muconate cycloisomerase; CatC, muconolactone D-isomerase; CarD, 2-hydroxypenta-2,4-dienoate hydratase; CarE, 4-hydroxy-2-oxovalerate aldolase; and CarF, acetaldehyde dehydrogenase. Compounds: 1, carbazole (C9a and C1 are shown); 2, 1-hydro-1,9-diohydroxycarbazole; 3, 2'-aminobiphenyl-2,3-diol; 4, 2-hydroxy-6-oxo-6-(2'-aminophenyl)-hexa-2,4-dienoic acid; 5, anthranilate; 6, catechol; 7, β-ketoadipate enol-lactone; 8, succinyl-CoA; 9, acetyl-CoA; 10, 2-hydroxypenta-2,4-dienoate; 11, 4-hydroxy-2-oxovalerate; 12, acetaldehyde; and 13, pyruvate. Compound 2 is transformed to compound 3 through a spontaneous reaction. Dashed arrows represent several consecutive reactions.

A search of the genes involved in carbazole degradation in the 19 genomes of *P. stutzeri* and *P. balearica* was performed by BLASTp, using the previously described *P. resinovorans* CA10 pathway as a model (Nojiri *et al*., 2001b). The carbazole degradation genes were identified only in the genome of strain XLDN-R (see Supplementary Table S11). The identified sequences of this strain showed high homology with the model sequences, with percentages of identity ranging from 72 to 100 % and, in all cases, with E-values below 1E-50 (see Table 2.7).

Table 2.7. Percentages of identity and E-values obtained with BLASTp of the carbazole degradation CDSs of *P. resinovorans* CA10 (Nojiri et al., 2001b) and *P. stutzeri* XLDN-R.

	% Id	E-value
CarAa	99.7	0
CarAc	99.1	2.E-60
CarAd	99.7	0
CarBa	100.0	2.E-50
CarBb	99.3	5.E-156
CarC	99.7	2.E-170
CarD	100.0	1.E-147
CarE	99.7	0
CarF	99.7	8.E-177
AntA	88.9	0
AntB	72.4	1.E-73
AntC	74.6	3.E-154
AntR	99.4	1.E-180

As described by Liu and co-workers (2012), the genes involved in carbazole degradation in the genome of strain XLDN-R were found to be structured as in *P. resinovorans* CA10. As shown in Figure 2.10, in strain CA10 four gene structures have been described: *antCBA*, *carAaAaBaBbCAc*, *carAdD*, and *carFE* (Nojiri *et al*., 2001b). These authors also described the presence of three different *orfs* that putatively encoded for transport proteins, named as *orf24*, *orf30*, and *orf31*. Additionally, Urata and co-workers (2004) described the presence of a gene encoding the transcriptional regulator of *ant* and *car* operons (AntR). All the gene structures described above were also found in the genome of strain XLDN-R. Although we cannot verify the presence of two consecutive *carAa* genes in the genome of XLDN-R, this possibility cannot be excluded, since this gene is localized in a small contig (NZ_AKYE01000015.1, 1205 bp), which only contain this gene.

The presence of four identical TnpAs named as IS*Pre*1 (BAB32779.1, BAB32763.1, BAB32748.1, and BAB32744.1), encoded between *ant* and *car* genes in the genome of strain CA10 was also remarkable. A similar protein was annotated in the genome of strain XLDN-R (WP_026006650.1; see Supplementary Table S11), sharing 100 % of identity in 52 % of the sequence. However, the distribution of carbazole degradation genes in many contigs prevent us to hypothesize about the localization of *tnpA* genes between the genes involved in carbazole described above.

In *P. resinovorans* CA10, genes involved in carbazole degradation are located in megaplasmid pCAR1 (Nojiri *et al*., 2001b). In particular, these genes are located in transposon Tn4676, which has been described as an active mobile element for the HGT of carbazole degradation genes (Shintani *et al*., 2005). This, together with the high homology found between XLDN-R and CA10 carbazole degradation proteins suggest that their genes might have been acquired for strain XLDN-R by an HGT event.

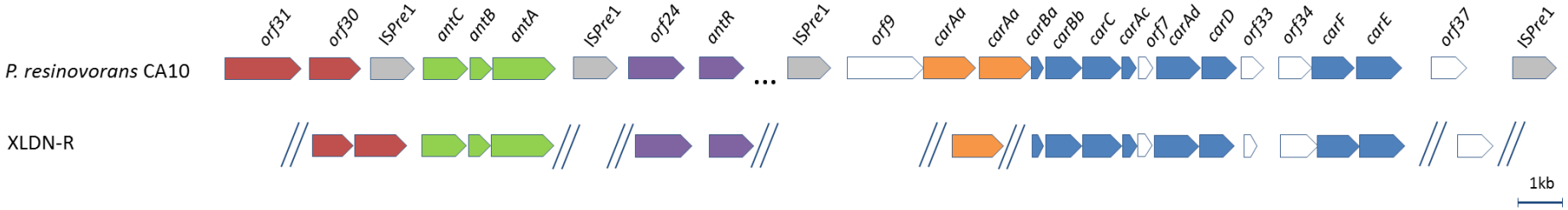

Figure 2.10. Structure of genes involved in carbazole degradation in *P. resinovorans* CA10 (Nojiri et al., 2001b), and *P. stutzeri* XLDN-R. Common gene structures between the model strain CA10 and XLDN-R are shown in different colors. Genes coding for putative transposases are shown in gray. Genes in white have not been characterized as involved in this pathway. Unknown distances between genes are shown with two forward slashes and long distances (>10 kb) within the same contig are represented with dots.

1.6. Phenol degradation gene cluster

Phenol is an aromatic hydrocarbon composed by a benzene ring and a hydroxyl group, which typically results from the degradation of organic material by microorganisms (Jeong *et al*., 2003). Phenol degradation by *P. stutzeri* has been previously described in strain SPC2 (Ahamad and Kunhi, 1996), whose genes are not available in public databases yet. The phenol degradation pathway has been characterized in *Pseudomonas* sp. CF600 by Shingler and co-workers (1992). As shown in Figure 2.11, in this bacterium phenol is firstly attacked by a phenol hydroxylase (DmpKLMNOP), which converts this compound to catechol. Catechol is later transformed by catechol 2,3-dioxygenase (DmpB), which breaks the aromatic ring and produces 2-hydroxymuconic semialdehyde. Afterwards, this compound can be catabolized to 2-oxo-pent-4-dienoate by DmpD or the successive action of DmpC, DmpI, and DmpH. Finally, this product is transformed to acetyl-CoA by DmpE, DmpG, and DmpF (see Figure 2.11).

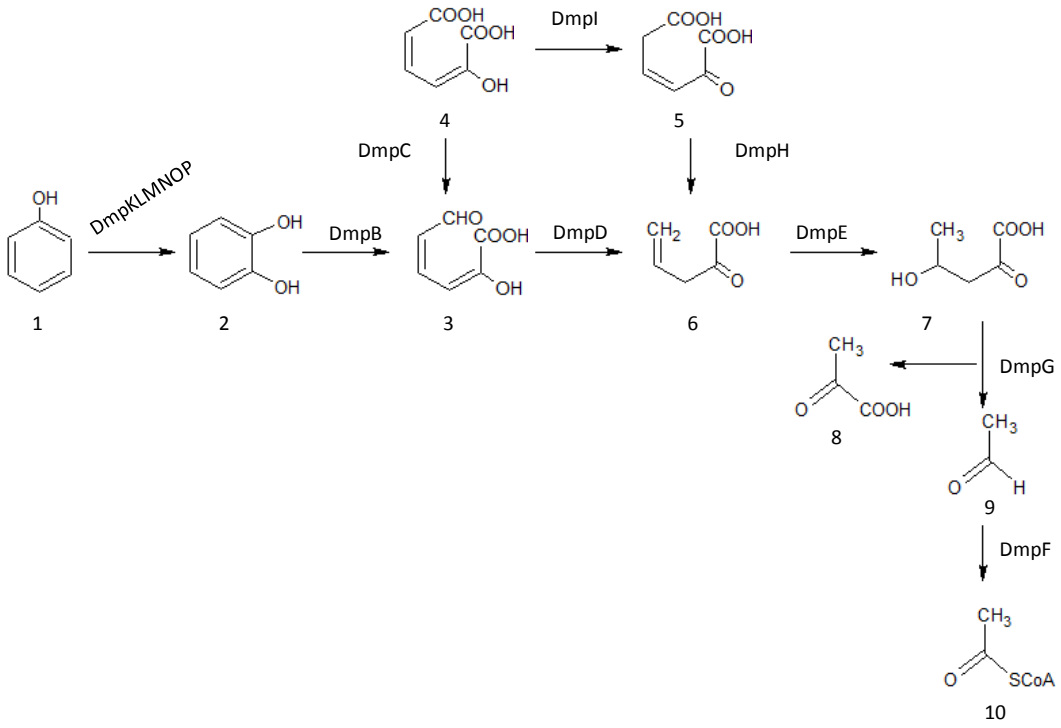

Figure 2.11. Phenol degradation pathway in *Pseudomonas* sp. CF600 (Shingler *et al*., 1992). Enzymes: DmpKLMNOP, phenol hydroxylase; DmpB, catechol 2,3-dioxygenase; DmpD, 2-hydroxymuconic semialdehyde hydrolase; DmpC, 2-hydroxymuconic semialdehyde dehydrogenase; DmpI, 4-oxalocrotonate isomerase; DmpH, 4-oxalocrotonate decarboxylase; DmpE, 2-oxo-pent-4-dienoate hydratase; DmpG, 4-hydroxy-2-oxovalerate aldolase; and DmpF, acetaldehyde dehydrogenase. Compounds: 1, phenol; 2, catechol; 3, 2-hydroxymuconic semialdehyde; 4, 2-hydroxyhexa-2,4-diene-1,6-dioate; 5, 2-oxohex-3-ene-1,6-dioate; 6, 2-oxo-pent-4-dienoate; 7, 4-hydroxy-2-oxovalerate; 8, pyruvate; 9, acetaldehyde; and 10, acetyl-CoA.

In order to identify the CDSs involved in the phenol degradation pathway in *P. stutzeri* and *P. balearica* genomes, a BLASTp search using the CDSs previously in *Pseudomonas* sp. CF600 was performed. With this approach, we identified most of these CDSs in the genomes of strains KOS6 and TS44 (see Supplementary Table S11). In general, they showed identity

percentages around 40 % and E-values lower than 1E-10 (see Table 2.8). The only sequence that was not identified was those that encoded DmpD and DmpQ. The former, as shown in Figure 2.11, can be replaced by DmpC, DmpI, and DmpH; and the latter has been described as involved in the activation of the catechol 2,3-dioxygenase DmpB (Powlowski and Shingler, 1994).

Table 2.8. Percentages of identity and E-values obtained with BLASTp of the phenol degradation CDSs of *Pseudomonas* sp. CF600 (Shingler *et al*., 1992) and *P. stutzeri* strains TS44 and KOS6.

	KOS6		TS44	
	% Id	**E-value**	**% Id**	**E-value**
DmpK	40.3	1E-13	43.7	4E-14
DmpL	53.8	2E-97	56.0	3E-101
DmpM	51.1	2E-24	50.0	5E-24
DmpN	62.6	0	63.0	0
DmpO	42.1	6E-26	44.7	6E-28
DmpP	62.3	3E-124	62.5	7E-125
DmpQ[a]	n.p.	n.p.	n.p.	n.p.
DmpB	53.3	2E-91	52.9	6E-91
DmpC	70.1	0	70.6	0
DmpD[a]	n.p.	n.p.	n.p.	n.p.
DmpI	47.3	9E-12	45.5	3E-11
DmpH	57.6	1E-89	58.0	8E-89
DmpE	60.2	5E-88	60.2	5E-88
DmpG	54.7	1E-108	55.3	5E-109
DmpF	54.3	2E-84	54.3	3E-85
DmpR	45.3	3E-130	45.0	2E-132

a: n.p., not predicted.

Genes coding for the enzymes involved in phenol degradation in *Pseudomonas* sp. CF600 are organized in a single gene cluster, *dmpRKLMNOPQBCDEFGHI*, located in plasmid pVI150 (Shingler *et al.,* 1992). The gene structure observed in strains TS44 and KOS6 was slightly different from the CF600 structure (see Figure 2.12). Genes *dmpKLMNOP*, *dmpFG*, and *dmpHI* maintained a similar structure. In contrast, *dmpB* appeared between the transcriptional regulator *dmpR* and the phenol hydroxylase component *dmpK*. Similarly, genes *dmpHI* appeared between *dmpE* and *dmpF*. Furthermore, in the genomes of KOS6 and TS44 a gene encoding a protein of unknown function was present between *dmpP* and *dmpC,* which did not show homology with proteins in the genome of CF600. All these genes are chromosomally located in KOS6 genome. In TS44 genome we could not know the localization of these genes, as this genome is constituted by 78 contigs.

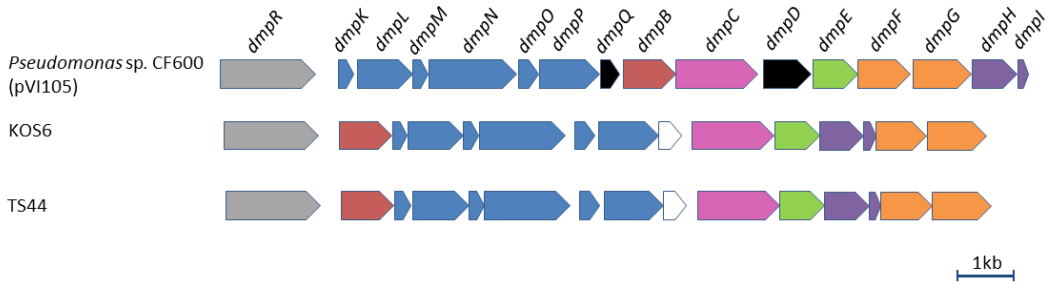

Figure 2.12. Structure of genes involved in phenol degradation in *Pseudomonas* sp. CF600 (Shingler *et al*., 1992) and *P. stutzeri* KOS6 and TS44 strains. Common gene structures between the model strain and *P. stutzeri* genomes are represented with the same color. Genes in black represent those genes from the model strain that have not been found in KOS6. Genes in white have not been characterized as involved in this pathway.

1.7. Naphthalene degradation gene cluster

Naphthalene is a polycyclic aromatic hydrocarbon formed by two benzene rings. Previous studies have described the ability of some *P. stutzeri* strains (AN10, B1SMN1, 19SMN4, ST27MN3) and *P. balearica* LS401 for using this compound as the sole source of carbon and energy source (Rosselló-Mora *et al*., 1994). In addition, the genes involved in naphthalene degradation in strain AN10 have been previously described (Bosch *et al*., 1999a; Bosch *et al*., 1999b; Bosch *et al*., 2000). In AN10, naphthalene is transformed to salicylate by the products of the operon *nahAaAbAcAdBFCED*. Then, salicylate can be degraded by two distinct salicylate hydroxylases (NahG and NahW) which decarboxylate salicylate to catechol. Subsequently, the aromatic ring of catechol is broken by a catechol 2,3-dioxygenase (NahH) generating hydroxymuconic semialdehyde. Finally, this compound is channeled to the TCA cycle by the successive action of the remaining gene products of the operon *nahGTHINLOMKJ* (see Figure 2.13). As a result, for each molecule of naphthalene 2 molecules of pyruvate, 1 acetyl-CoA and 2 molecules of CO_2 are obtained.

In order to identify all CDSs involved in naphthalene degradation in the genomes of *P. stutzeri* and *P. balearica* a BLASTp homology search was performed using the AN10 sequences as model. Homologous sequences were identified in the four other strains described as naphthalene degraders (B1SMN1, 19SMN4, ST27MN3, and SP1402^T). In addition, homologous CDSs were also identified in the genome of strain KOS6 (see Supplementary Table S11). In all cases amino acid identities were higher than 80 %, and E-values lower than 1E-30 (see Table 2.9). As mentioned above, we identified the genes involved in phenol degradation via catechol *meta*-cleavage in the genome of strain KOS6. The coexistence in the KOS6 chromosome of two different gene structures for catechol *meta*-cleavage (DmpB for phenol and NahH for naphthalene degradation) is noteworthy. Both set of genes presented identities from 40 to 72 %, suggesting a different origin. Genes homologous to AN10 genes involved in the catechol *meta*-cleavage of naphthalene degradation were also found in the genome of strain XLDN-R, all located in contig NZ_AKYE01000001.1 (see Table 2.9 and Supplementary Table S11). But genes involved in naphthalene, salicylate, or phenol were not found in this genome. Therefore, the genome of this strain showed genes involved in the *ortho*- and the *meta*-catechol cleavage.

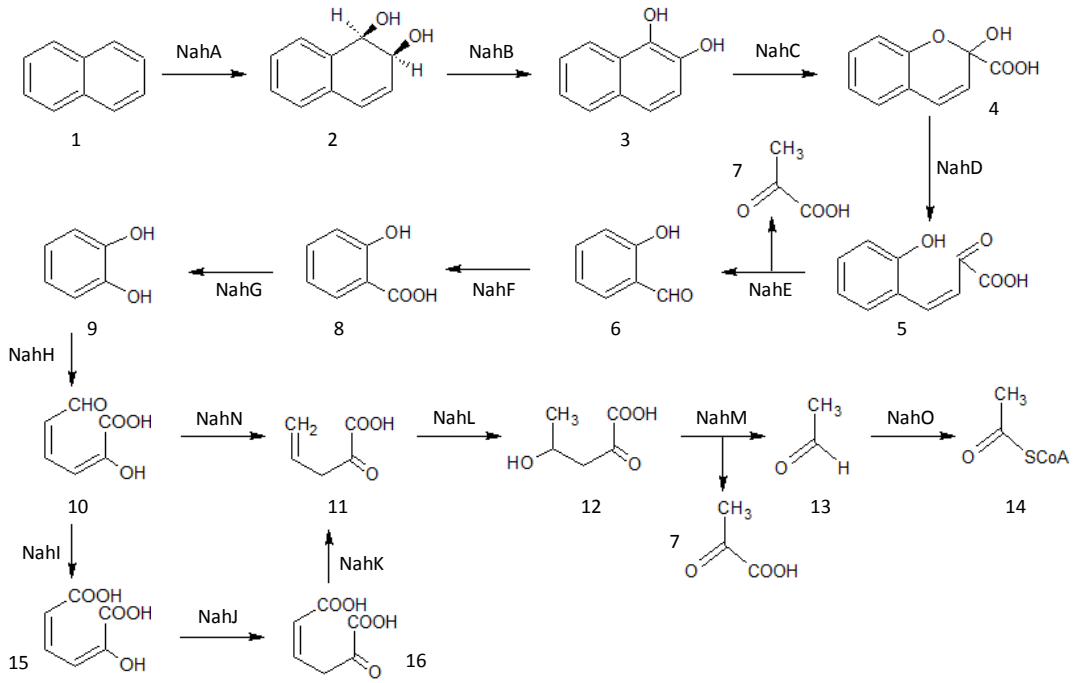

Figure 2.13. Naphthalene degradation pathway in *P. stutzeri* AN10 (Bosch *et al*., 1999a; Bosch *et al*., 1999b; Bosch *et al*., 2000). Enzymes: NahA, naphthalene 1,2-dioxygenase; NahB, cis-1,2-dihydro-1,2-dihydroxynaphthalene-1,2 dehydrogenase; NahC, 1,2-dihydroxynaphthalene dioxygenase; NahD, 2-hydroxychromene-2-carboxylate isomerase; NahE, trans-o-hydroxybenzylidenepyruvate hydratase-aldolase; NahF, salicylaldehyde dehydrogenase; NahG and NahW, salicylate hydroxylase; NahH, catechol 2,3-dioxygenase; NahN, 2-hydroxymuconic semialdehyde hydrolase; NahL, 2-oxopent-4-dienoate hydratase; NahM, 4-hydroxy-2-oxovalerate aldolase; NahO, acetaldehyde dehydrogenase; NahI, 2-hydroxymuconic-semialdehyde dehydrogenase; NahJ, 4-oxalocrotonate isomerase; and NahK, 4-oxalocrotonate decarboxylase. Compounds: 1, naphthalene; 2, cis-1,2-dihydro-1,2-dihydroxynaphthalene; 3, 1,2-dihydroxynaphthalene; 4, 2-hydroxychromene-2-carboxylate; 5, cis-o-hydroxybenzalpyruvate; 6, salicylaldehyde; 7, pyruvate; 8, salicylate; 9, catechol; 10, 2-hydroxymuconic semialdehyde; 11, 2-oxopent-4-dienoate; 12, 4-hydroxy-2-oxovalerate; 13, acetaldehyde; 14, acetyl-CoA; 15, 2-hydroxyhexa-2,4-diene-1,6-dioate; and 16, 2-oxohexa-3-ene-1,6-dioate.

The genetic system involved in naphthalene degradation in AN10 is organized in three different operons: the first one corresponding to the upper pathway (*nahAaAbAcAdBFCED*), the corresponding to the lower pathway (*nahGTHINLOMKJ*), and the third to the second salicylate hydroxylase (*nahW*). All of them are transcriptionally activated by NahR in response to salicylate (Bosch *et al.*, 1999a; Bosch et al., 1999b; Bosch *et al*., 2000). Sinteny analysis demonstrated that all *P. stutzeri* strains and *P. balearica* showed the same gene structure than AN10 (see Figure 2.14). Strains KOS6 and AN10 presented the gene cluster *nahAaAbAcAdBFCED* separated from the rest (15.6-kb and 21.8-kb respectively). In these regions (617639..633263 from KK020677.1 and 1442524..1464319 from NC_018028.1; see Supplementary Table S2), both genomes presented mainly genes encoding putative TnpAs. In contrast, 19SMN4 and SP1402^T presented these gene clusters closer in their genome (11.5-kb and 9.8-kb respectively), despite having genes coding for TnpAs in this region as well (see Figure 2.14). This distance was unknown in the other two strains (B1SMN1 and ST27MN3), because their genomes were not completely closed. Strain 19SMN4 was the only one with the naphthalene degradation genes located in a plasmid (pLIB119), in agreement with the results of Rosselló-Mora and co-workers (1994).

Table 2.9. Percentages of identity and E-values obtained with BLASTp of naphthalene degradation CDSs of *P. stutzeri* AN10 (Bosch *et al*., 1999a; Bosch *et al*., 1999b; Bosch *et al*., 2000) and other *P. stutzeri* strains and *P. balearica*.

		B1SMN1	XLDN-R[a]	19SMN4	ST27MN3	KOS6	SP1402^T
NahAa	% Id	99.7	n.p.	100	99.7	99.7	99.4
	E-value	0	n.p.	0	0	0	0
NahAb	% Id	100	n.p.	100	100	99.0	100
	E-value	3E-60	n.p.	3E-60	4E-60	4E-59	6E-60
NahAc	% Id	99.1	n.p.	97.8	98.0	96.7	98.0
	E-value	0	n.p.	0	0	0	0
NahAd	% Id	98.9	n.p.	100	100	99.5	99.5
	E-value	1E-111	n.p.	3E-112	1E-111	8E-112	2E-111
NahB	% Id	99.2	n.p.	100	100	99.2	99.6
	E-value	1E-147	n.p.	3E-149	3E-149	5E-148	9E-148
NahC	% Id	99.7	n.p.	99.0	99.0	99.3	99.7
	E-value	0	n.p.	0	0	0	0
NahD	% Id	99.5	n.p.	99.6	100	99.6	99.6
	E-value	3E-125	n.p.	1E-131	5E-132	1E-131	3E-135
NahE	% Id	100	n.p.	99.7	99.6	99.7	99.7
	E-value	0	n.p.	0	2E-162	0	0
NahF	% Id	99.6	n.p.	99.8	100	100	100
	E-value	0	n.p.	0	0	0	0
NahG	% Id	99.8	n.p.	99.8	99.8	100	99.5
	E-value	0	n.p.	0	0	0	0
NahW	% Id	100	n.p.	100	100	100	100
	E-value	0	n.p.	0	0	0	0
NahH	% Id	100	97.4	100	100	100	99.7
	E-value	0	1E-179	0	0	0	0
NahI	% Id	100	100	100	100	100	99.4
	E-value	0	0	0	0	0	0
NahJ	% Id	98.4	98.4	95.2	98.4	98.4	98.4
	E-value	1E-31	2E-29	4E-31	1E-31	1E-31	2E-31
NahK	% Id	98.5	98.1	97.4	98.5	98.1	98.5
	E-value	9E-150	3E-147	7E-149	9E-150	2E-149	2E-149
NahN	% Id	99.7	99.7	99.7	99.7	99.7	96.8
	E-value	4E-167	1E-164	4E-167	4E-167	4E-167	7E-160
NahL	% Id	100	99.2	100	100	100	96.9
	E-value	3E-149	2E-146	3E-149	3E-149	3E-149	1E-144
NahM	% Id	99.7	100	100	100	99.7	99.7
	E-value	0	0	0	0	0	0
NahO	% Id	100	100	100	100	100	100
	E-value	9E-178	1E-175	9E-178	9E-178	9E-178	2E-177
NahT	% Id	100	n.p.	100	100	100	100
	E-value	1E-67	n.p.	1E-67	1E-67	1E-67	2E-67
NahX	% Id	99.3	n.p.	100	99.3	100	99.3
	E-value	1E-83	n.p.	1E-84	1E-83	1E-84	2E-83
NahZ	% Id	99.2	n.p.	99.4	99.2	99.2	99.2
	E-value	0	n.p.	0	0	0	0
NahR	% Id	100	n.p.	100	100	100	100
	E-value	3E-177	n.p.	3E-177	3E-177	3E-177	7E-177
NahP	% Id	97.6	n.p.	97.0	97.4	97.4	97.0
	E-value	0	n.p.	0	0	0	0

a: n.p., not predicted.

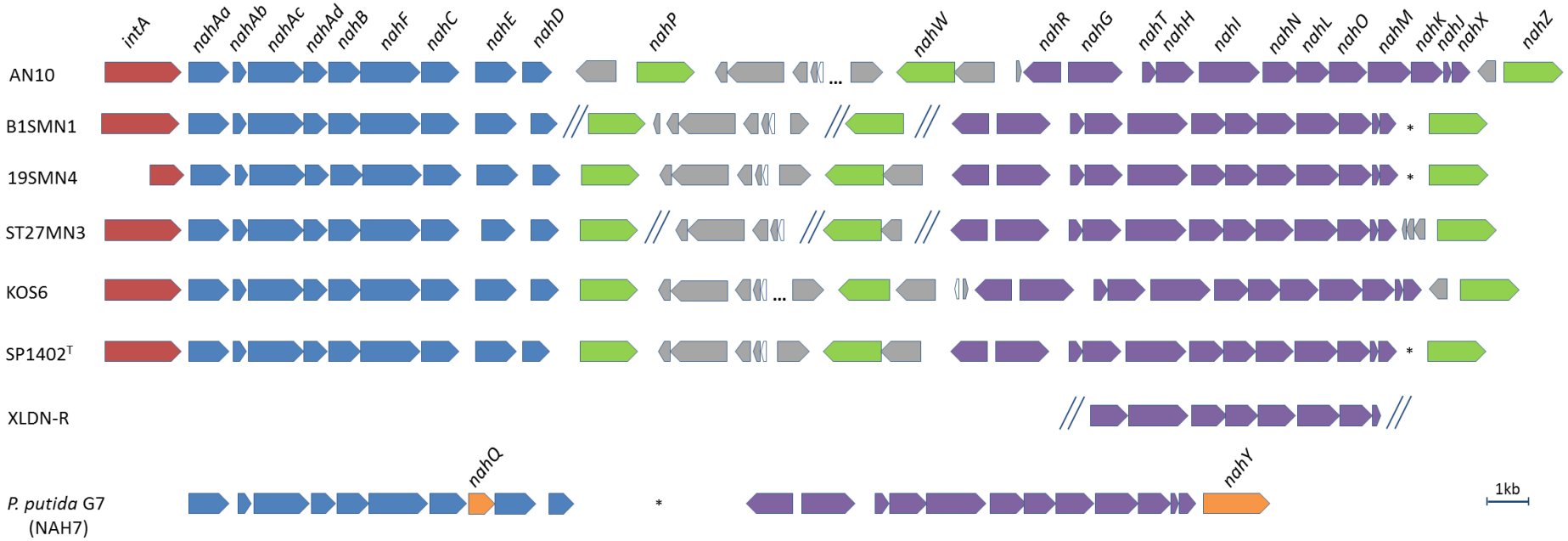

Figure 2.14. Structure of genes involved in naphthalene degradation in *P. stutzeri* AN10 (Bosch *et al.*, 1999a; Bosch *et al.*, 1999b; Bosch *et al.*, 2000), other *P. stutzeri* strains (19SMN4, ST27MN3, B1SMN1, and KOS6), *P. balearica* SP1402^{T}, and plasmid NAH7 of *P. putida* G7 (Sota *et al.*, 2006). Common gene structures between the model strain AN10 and the other analyzed strains are shown in blue and purple. Genes in green represent those genes from the model strain AN10 that have not been found in *P. putida* G7 (NAH7). Genes in orange represent those genes from the model strain *P. putida* G7 (NAH7) that have not been found in *P. stutzeri* and *P. balearica*. Genes in white have not been characterized as involved in this pathway. Genes coding for putative transposases are shown in gray, and genes coding for putative integrases are shown in red. XLDN-R genes homologous to AN10 catechol *meta*-cleavage genes are also shown. The unannotated CDSs are shown with an asterisk. Unknown distances between genes are shown with two forward slashes and long distances (>10 kb) within the same contig are represented with dots. A more detailed representation of the region between *nahP* and *nahR* is shown in Figure 2.15.

In addition to the degradation genes previously described, other conserved genes were detected in this genomic region in all strains. We identified a putative integrase encoding-gene (*intA*), upstream the naphthalene degradation operon *nahAaAbAcAdBFCED* (see Figure 2.14). In addition, we identified a gene coding for a putative aromatic hydrocarbon transporter, *nahP*, located downstream *nahAaAbAcAdBFCED*. The amino acid sequence of NahP was then analyzed with the Conserved Domain Database (CDD) NCBI tool (Marchler-Bauer *et al.*, 2015). This protein was classified within family "Toluene_X" (pfam03349), a family of outer membrane transport proteins which is involved in toluene transport. This suggests that NahP could play a role in transport of naphthalene or its metabolites through the membrane.

Finally, two homologous genes located downstream operon *nahGTHINLOMKJ, nahX and nahZ,* were also conserved in all the strains. NahX gene was annotated by the NCBI pipeline as an ATP: corrinoid adenosyltransferase, which is an enzyme responsible for converting vitamin B12 to coenzyme B12. The NahX sequence of *P. stutzeri* and *P. balearica* genomes presented 67 % of identity and an E-value below 1E-47 with a NahX sequence from the NAH7 naphthalene degradation plasmid. This gene has been previously described in the naphthalene degradative plasmid NAH7, although its role in naphthalene degradation remains still unknown (Sota *et al*., 2006). Lee and co-workers (1999) described that coenzyme B12 has an important role for the function of a 1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase (XylL), encoded by the TOL plasmid pWW0, in toluene degradation. XylL catalyzes a methyl transfer reaction from toluene cis-dihydrodiol to adenosylcobalamin (coenzyme B12), producing catechol and methylcobalamin (Lee *et al*., 1999). The toluene catabolic reaction of XylL is homologous to the naphthalene degradation step performed by the cis-1,2-dihydro-1,2-dihydroxynaphthalene-1,2-dehydrogenase NahB, which transforms cis-1,2-dihydro-1,2-dihydroxynaphthalene to 1,2-dihydroxynaphthalene (see Figure 2.13). Thus, we can hypothesize that NahB could be involved in the methyl transfer reaction when methylnaphthalenes are catabolized. However, reports on methylnaphthalene degradation reveal that 1- and 2- methylnaphthalene are channeled to 3- and 4- methylsalicylate, respectively (Williams *et al*., 1975; Mahajan *et al*., 1994; Dutta *et al*., 1998; Sharanagouda and Karegoudar, 2001), and thus there is no methyl transferase reaction at that point. Other enzymes in the naphthalene degradation pathway that generate catechol-like compounds are both salicylate hydroxylases, NahG and NahW (see Figure 2.13). However, previous studies demonstrate salicylate hydroxylase activity of both enzymes (NahG and NahW) cloned in *E. coli* JM109 without the presence of NahX nor adenosylcobalamin (Bosch *et al*., 1999b; Zhao *et al*., 2005). Additionally, methylsalicylate degradation reports also revealed absence of methyl transferase activity, since methylcatechol is the final product of salicylate hydroxylase reaction (Bosch *et al*., 1999b; Zhao *et al*., 2005). On the other hand, adenosylcobalamin has been reported as essential for other types of enzymatic reactions, such as reductions and isomerizations (Giedyk *et al*., 2015). Two different isomerases are involved in naphthalene degradation pathway: NahD, a 2-hydroxychromene-2-carboxylate isomerase; and NahJ, a 4-oxalocrotonate isomerase (see Figure 2.13). The in vitro activity of the first one has been also described in the absence of adenosylcobalamin (Ohmoto *et al*., 1998), but there is no information on NahJ. Therefore, we hypothesize that NahX might provide the cofactor adenosylcobalamin to NahJ. Similar NahX sequence (63 % identity and E-value of 1E-47) were also identified at 1.25 kb downstream of *dmpG*, involved in phenol degradation, in KOS6 and

TS44 genomes, but it was not found in plasmid pVI150 from *Pseudomonas* sp. CD600, the model used for phenol degradation.

With respect to *nahZ*, this gene did not present any homologous in the NAH7 naphthalene degradation plasmid, and it has no assigned role in naphthalene degradation. It was annotated by the NCBI pipeline as a FAD-dependent oxidase. A putative role of NahZ in naphthalene metabolism will be discussed in section 3.1 of Chapter 3.

The presence of a large number of putative transposase encoding-genes in the region of naphthalene degradation operons, conserved in all six strains, was noteworthy. As shown in section 5 from Chapter 1 (Table 1.5), we observed high numbers of shared TnpA-COGs between the strains isolated from marine sediments contaminated with hydrocarbons (AN10, 19SMN4, and ST27MN3) and strains B1SMN1 and KOS6, isolated from sewage and an industrial hydrocarbon sludge, respectively. Interestingly, all of them present on their genomes the naphthalene degradation operons. These results suggest that the high number of shared TnpA-COGs between these strains is most likely due more to the presence of naphthalene degradation operons in all of them rather than their origin.

Between *nahP* and *nahR* there were 6 different highly conserved CDSs encoding putative TnpAs (shown in Figure 2.15 numbered from 1 to 6). These were mainly classified as IS5-type (TnpAs 3–6), but also IS66-type (TnpAs 1 and 2). Their classification into IS-families was not surprising, as IS5 and IS66 were the second and the third most abundant IS-families identified in *P. stutzeri* and *P. balearica* genomes (see Figure 1.6). In particular, the sequence of the TnpA 1 presented the best homology (E-value = 0) with the IS*Ppu*13 sequence of *P. putida* KT2440 from the IS-Finder database, for which no more information is available. TnpAs 4 and 5 presented the best sequence homology (E-values below 1E-100) with IS*Pre*1 sequence of *P. resinovorans* CA10 from the IS-Finder database. As mentioned in section 1.5, IS*Pre*1 is located in the transposon Tn4676 from the carbazole catabolic plasmid pCAR1 (Shintani *et al*., 2005). The presence of IS*Pre*1 has also been described in the naphthalene catabolic plasmid pDTG1 from *P. putida* NCIB 98164, between *nahG* and *nahT* genes from the lower naphthalene degradation operon, interrupting *nahG* transcription. Due to this, strain NCIB 98164 is forced to degrade naphthalene through the chromosomally encoded β-ketoadipate pathway, rather than the expected catechol *meta*-cleavage pathway (Dennis and Zylstra, 2004). In the studied genomes of *P. stutzeri* and *P. balearica* strains, genes homologous to IS*Pre*1 were observed flanking *nahW*, which encodes for a second salicylate hydroxylase. These observations suggest that IS*Pre1* might be involved in the genetic rearrangement of naphthalene catabolic operons in *P. stutzeri*, as the acquisition of the salicylate hydroxylase NahW. Naphthalene degradation plasmid NAH7 also showed 3 truncated genes encoding TnpAs between *nahD* and *nahR* (located at the unannotated region shown in Figure 2.14), which were not found between the naphthalene degradation operons of *P. stutzeri*. Catabolic genes might easily be transferred across bacteria by catabolic transposons (Wyndham *et al*., 1994; Tan, 1999; Top and Springael, 2003; Nojiri *et al*., 2004). Despite the large number of putative transposases found in the naphthalene degradation operons, none of the TnpAs were contained in a complete structure of a transposon. However, their abundance might be a residue from past HGT events occurred during the adaptation of these strains to naphthalene degradation.

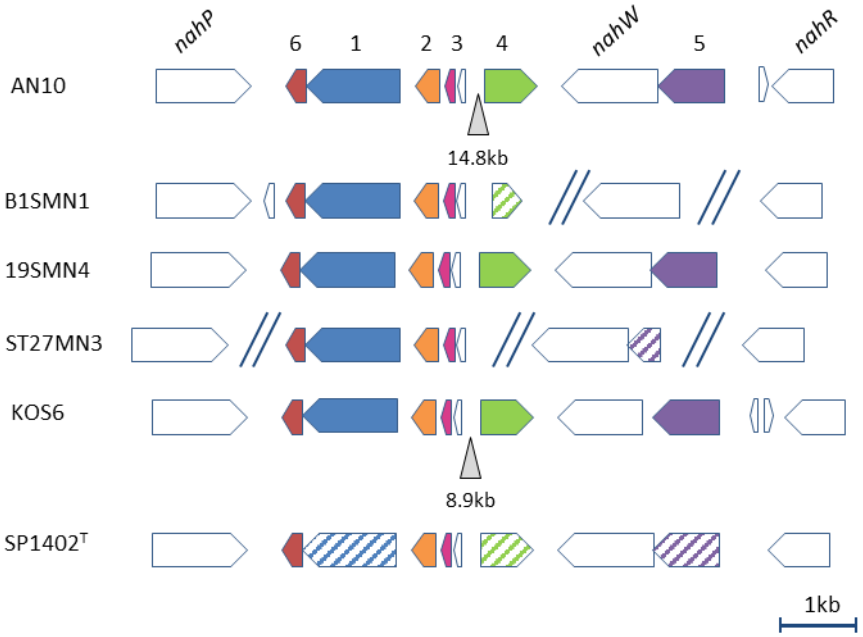

Figure 2.15. Gene structure of the putative TnpAs present between *nahP* and *nahR*. These genes have been listed from 1 to 6 and shown in different colors. Striped genes represent those with an identity below 95 % in 90 % of the sequence with their homologous genes in AN10 genome. Remaining genes are shown in white.

Comparing the gene structure of naphthalene degradation operons of *P. stutzeri* and *P. balearica* with the structure of these genes in the naphthalene degradation plasmid NAH7, clear similarities could be observed. As *P. stutzeri* and *P. balearica* strains, the structure in plasmid NAH7 consists in an operon for the upper naphthalene pathway (*nahAaAbAcAdBFCQED*), although in this case we found gene *nahQ* (whose function remains unknown) between *nahC* and *nahE* (see Figure 2.14). We confirmed the absence of *nahQ* in the *P. stutzeri* and *P. balearica* genomes by BLASTn, using the *nahQ* sequence from plasmid NAH7. On the other hand, the operon for lower pathway of plasmid NAH7, which is 3.96-kb downstream the upper naphthalene operon, lacks *nahZ* and has *nahY* instead, a gene involved in naphthalene chemotaxis (Sota *et al*., 2006). In addition, plasmid NAH7 lacks *nahW*, the second salicylate hydroxylase which is conserved in *P. stutzeri* and *P. balearica* naphthalene degradation pathways (see Figure 2.14).

Ferrero and co-workers (2002) previously described that closely related naphthalene degrading *Pseudomonas* isolated from western Mediterranean Sea presented two distinct *nahAc* genes (from the upper naphthalene degradation pathway): the *P. stutzeri* AN10 type (Bosch et al., 1999a; Bosch et al., 1999b; Bosch et al., 2000), and the *Pseudomonas* sp. C18 type (Denome *et al*., 1993). Both types showed the same *nah* gene structure, with the exception of the presence of gene *nahQ* in strain C18 (as in plasmid NAH7), that was not annotated in AN10. Furthermore, these authors demonstrated that in some *Pseudomonas* isolates both types of *nah* genes coexist, such as strains S1MN2 and 11NH. In order to evaluate the origin of the different *nah* genes in the genomes analyzed in our study, we performed a Neighbor-Joining tree of several *nah* gene products. For this, we selected eight different CDSs whose genes were located at different points of the upper (NahAc, NahD, NahF) and lower naphthalene degradation operons (NahG, NahH, NahI and NahM) plus the regulator NahR. Furthermore, apart from the CDSs of *P.*

stutzeri and *P. balearica* we included homologous CDSs from other naphthalene degraders: *Pseudomonas* sp. C18 (only the upper pathway has been described), *Pseudomonas* sp. NCIB 98164, and the plasmid NAH7 (Denome *et al*., 1993; Sota *et al*., 2006). Additionally, XylE (homologous to NahH), XylG (homologous to NahI), and XylK (homologous to NahM) from the toluene degradation plasmids pWW0 and pWW53 (which contain two different copies of the lower xylene degradation operon) (Burlage *et al*., 1989; Osborne *et al*., 1988) were also included in this study.

Analyzing phylogenetic trees of the enzymes from the upper naphthalene degradation pathway (NahAc, NahD, NahF) we could observe that the sequences were grouped in two groups: the first one with *P. stutzeri* and *P. balearica* sequences, and another one containing the sequences from strains C18 and NCIB 98164, as well as the plasmid NAH7. Figure 2.16 A shows the phylogenetic tree of NahD (trees of NahAc and NahF are shown in Supplementary Figure S2).

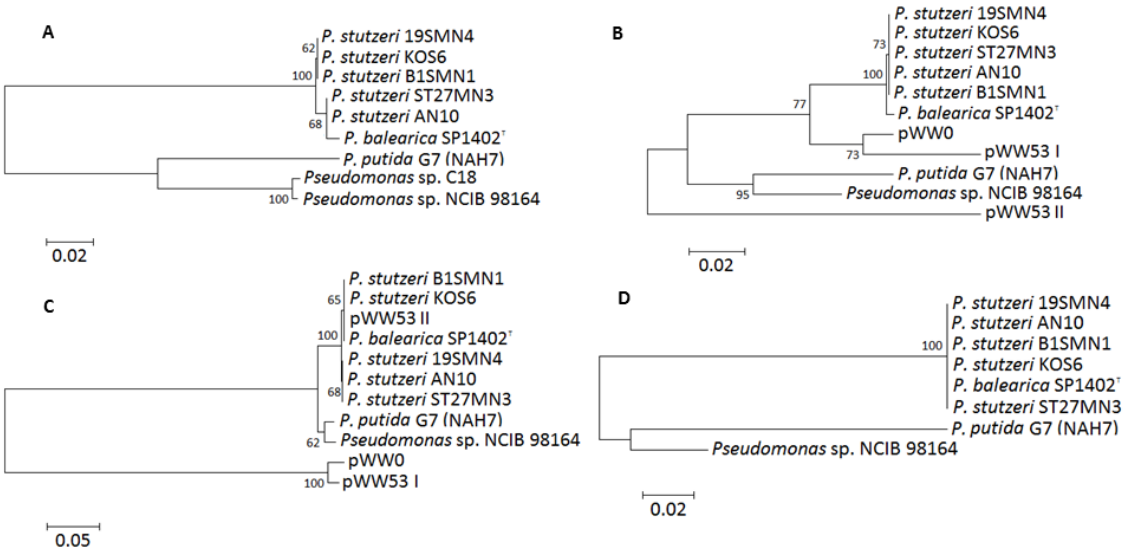

Figure 2.16. Phylogeny of four different enzymes involved in naphthalene degradation. (A) NahD. (B) NahH and XylE for plasmids pWW0 and pWW53. (C) NahM and XylK for plasmids pWW0 and pWW53. (D) NahR. Bootstrap numbers below 50 (from 100 replicates) were omitted.

Regarding the sequences involved in the lower naphthalene degradation pathway (NahG, NahH, NahI, and NahM), the NahG phylogeny showed the same groups described for genes involved in the upper pathway (see Supplementary Figure S2). In the case of NahH, NahI and NahM, three groups were observed: (1) sequences from *P. stutzeri* and *P. balearica*; (2) sequences from the NCIB 98164 strain, plasmid NAH7, and the second sequence from plasmid pWW53; and (3) sequence from plasmid pWW0 and the first sequence from plasmid pWW53. In NahH phylogeny (Figure 2.16 B), NahH sequences from *P. stutzeri* and *P. balearica* (group 1) were closer to the third group of sequences rather than to the second group. In NahM phylogeny (Figure 2.16 C), sequences from *P. stutzeri* and *P. balearica* (group 1) were closer to the second group of sequences rather than to the third group. In NahI phylogeny pWW53 XylGII sequence was as close to NCIB 98164 and NAH7 NahI sequences than to those of *P. stutzeri* and *P. balearica* (see Supplementary Figure S2). Finally, the phylogeny of the transcriptional regulator NahR (Figure 2.16 D) also reflects the grouping of *P. stutzeri* and *P. balearica* sequences. The results obtained for NahM suggest that the lower naphthalene degradation operon of *P. stutzeri* and *P. balearica*

and the second copy of the lower xylene degradation operon from pWW53 may have a common origin. The lower xylene degradation operon from pWW53 might be the initial gene cluster. Then, a strain with this initial *xyl* gene cluster may have acquired the lower naphthalene degradation operon, and in the case of *P. stutzeri* strains, a recombinant operon from both lower operons would be selected. The coexistence of two different *nah* genes in strains S1MN2 and 11NH described by Ferrero and co-workers (2002) support this hypothesis. Finally, the upper naphthalene degradation operon would have also been acquired.

2. Dehydroabietic acid degradation gene cluster

Abietane is a diterpene hydrocarbon (a terpene is composed of 20 carbon atoms) which constitutes the basis for a wide variety of natural compounds such as resin acids from coniferous plant and legumes (Findeisen *et al*., 2007). *Pseudomonas abietaniphila* BKME-9, a bacterium isolated from a pulp and paper wastewater treatment system, is able to use dehydroabietic acid, an abietane-type acid, as sole carbon source (Mohn, 1995). As shown in Figure 2.17, in strain BKME-9 dehydroabietic acid is first hydroxylated by the cytochrome P450 DitQ (Smith *et al.*, 2004). These authors suggest that the product of this reaction might be 7-hydroxy dehydroabietic acid, an intermediate which would be later transformed to 7-oxodehydroabietic acid. The presence of a second cytochrome P450 involved in the degradation of abietane to 7-oxodehydroabietic acid was also suggested by Smith and co-workers (2004), although they were not able to identify it. A second cytochrome P450 has been described as DitU in *Burkholderia xenovorans* LB400 by Smith and co-workers (2007). Then a three-component ring-hydroxylating dioxygenase (DitA1A2A3) is able to transform 7-oxodehydroabietic acid to the catecholic intermediate 7-oxo-11,12-dihydroxy-8,13-abietadien acid (Martin and Mohn, 1999). The following enzymatic steps have not been characterized (Martin and Mohn, 1999, 2000; Smith *et al.*, 2004; see Figure 2.17).

The gene cluster involved in this degradation pathway in *P. abietaniphila* BKME-9 has been described by Smith and co-workers (2004), and comprises catabolic genes: *ditQ*, cytochrome P450; *ditA*, ring-hydroxylating dioxygenase; *ditC*, extradiol cleavage dioxygenase; *ditB*, putative dehydrogenase/reductase; *ditD*, putative isomerase/decarboxylase; *ditF*, putative sterol carrier-like protein; *ditG*, putative dehydrogenase/reductase; *ditH*, putative isomerase/decarboxylase; *ditI*, putative dehydrogenase/reductase; *ditJ*, coenzyme A ligase; *ditL*, putative amidohydrolase; *ditM*, putative fumarylacetoacetate hydrolase; *ditN*, putative 3-hydroxyacyl-CoA dehydrogenase; *ditO*, putative acetyl-CoA acetyltransferase. Additionally the dehydroabietic degradation gene cluster also presented two regulatory genes (*ditR* and *ditK*), a permease (*ditE*), and a protein with unknown function (*ditP*). The structure of genes involved in abietane degradation in *B. xenovorans* LB400 is different to the described for *P. abietaniphila* BKME-9 (Smith *et al*., 2007).

Analyzing the *P. stutzeri* and *P. balearica* genomes, CDSs homologous to those described by Smith and co-workers (2004) involved in the dehydroabietic acid degradation were found in strains MF28 and TS44 (see Supplementary Table S11). These CDSs were analyzed with BLASTp against the dehydroabietic acid degradation operon from BKME-9. Results showed that they

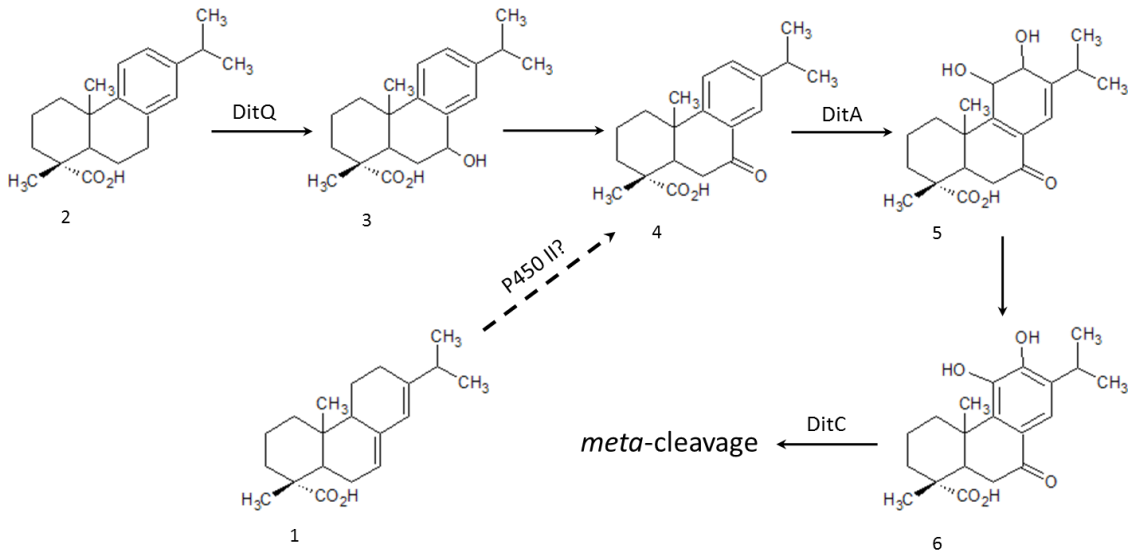

Figure 2.17. Proposed dehydroabietic acid degradation pathway in *P. abietaniphila* BKME-9 (Smith *et al.*, 2004). Enzymes: DitQ, cytochrome P450; DitA, 7-oxodehydroabietic acid dioxygenase; P450 II, a second cytochrome P450; and DitC, extradiol cleavage dioxygenase. Compounds: 1, abietic acid; 2, dehydroabietic acid; 3, 7-hydroxy dehydroabietic acid; 4, 7-oxodehydroabietic acid; 5, 7-oxo-11,12-dihydroxy-8,13-abietadien acid; 6, 7-oxo-11,12-dihydroxydehydroabietic acid. Dashed arrows represent hypothetical steps in the pathway.

Table 2.10. Percentages of identity and E-values obtained with BLASTp of the dehydroabietic acid degradation CDSs of *P. abietaniphila* BKME-9 (Smith *et al*., 2004) and *P. stutzeri* strains MF28 and TS44.

	MF28		TS44	
	% Id	**E-value**	**% Id**	**E-value**
DitQ	86.1	0	86.6	0
DitA1	91.5	0	92.2	0
DitA2	85.7	9E-123	86.2	4E-123
DitA3	87.3	3E-42	85.9	7E-42
DitC	85.3	0	85.6	0
DitB	81.8	4E-157	81.8	2E-157
DitD	67.4	5E-141	67.7	1E-141
DitE	64.9	0	64.5	0
DitR	78.7	7E-146	78.7	1E-145
DitF	88.6	0	88.9	0
DitG	76.9	1E-122	75.3	9E-118
DitH	82.7	0	79.5	0
DitI	84.9	4E-168	86.5	3E-170
DitJ	82.6	0	83.2	0
DitK	83.5	5E-129	79.7	6E-129
DitL	77.9	0	76.9	0
DitM	81.7	2E-171	81.0	8E-171
DitN	81.1	2E-175	81.1	2E-175
DitO	85.7	0	85.9	0
DitP	75.7	6E-79	70.8	1E-57

presented high similarities, with E-values lower than 1E-40 and identities above 60 % (see Table 2.10). The second cytochrome P450 (DitU; BxeC0631) described by Smith and co-workers (2007) in *B. xenovorans* LB400 was not found in the genomes of TS44 and MF28. However, genes annotated as cytochrome P450 were observed in TS44 (EIK52895.1; see Supplementary Table S2) and MF28 (EQM77671.1; see Supplementary Table S2), located at 75 and 18-kb to *ditA3* (see Supplementary Table S2), which might function as the second cytochrome P450.

The organization of genes involved in dehydroabietic acid degradation in both *P. stutzeri* strains (MF28 and TS44) presented almost the same gene structure than in *P. abietaniphila* BKME-9, with the following differences: (1) both *P. stutzeri* strains had a gene encoding an histidine kinase between the permease and *ditA3*, which was not present in the BKME-9 operon; (2) *P. stutzeri* strains presented a gene between *ditF* and *ditG*, which was not present in BKME-9, encoding an hypothetical protein, and (3) between *ditA2* and *ditI* strains MF28 and TS44 had a hypothetical protein not present in BKME-9 (see Figure 2.18).

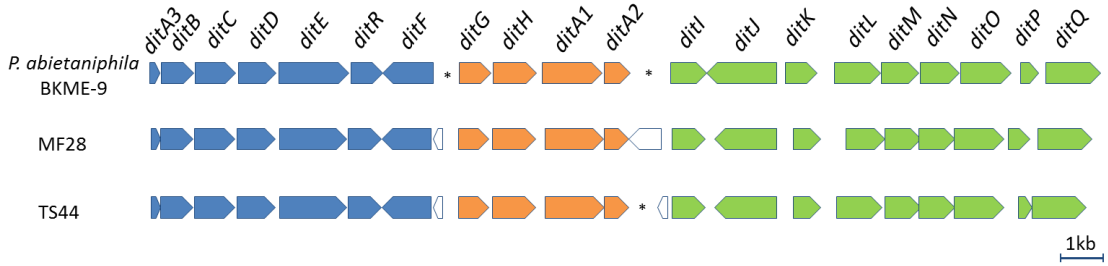

Figure 2.18. Structure of genes involved in dehydroabietic acid degradation in *P. abietaniphila* BKME-9 (Smith *et al*., 2004) and *P. stutzeri* MF28 and TS44 strains. Common gene structures between the model strain and *P. stutzeri* genomes are represented with the same color. Genes in white have not been characterized as involved in this pathway. The unannotated CDSs are shown with an asterisk.

3. Growth of *P. stutzeri* with aromatic hydrocarbons

Ten of the 19 *P. stutzeri* and *P. balearica* studied strains included in this analysis were available in the culture collection of our laboratory: ATCC 17588^T, B1SMN1, DSM 4166, CCUG 16156, AN10, 19SMN4, ST27MN3, DSM 10701, NF13, and SP1402^T. We also had strain A15 in our culture collection. A15 is the ancestor of strain A1501, which is the one whose genome has been sequenced. As shown above (see for example Table 2.2), these strains presented in their genomes genes involved in the catabolism of 4-hydroxyphenylpyruvate, protocatechuate, benzoate, or naphthalene. For this reason, the ability of these strains for growing with these aromatic compounds as sole carbon and energy source was tested. Cultures were performed in MMB supplemented with these four aromatic compounds (see section 2 from Materials and Methods).

Despite the fact that all *P. stutzeri* and *P. balearica* tested strains seemed to have all genes required for the degradation of 4-hydroxyphenylpyruvate (see Figure 2.2), none of them were able to grow with this compound as carbon and energy source (see Table 2.11) in the conditions tested. The same result was obtained using homogentisate, an intermediate of the degradation of 4-hydroxyphenylpyruvate, as sole carbon and energy source. As the genes involved in this pathway were observed in all studied *P. stutzeri* and *P. balearica* genomes, they probably have

an active role in cell metabolism. All the tested strains (except DSM 10701) presented on their genomes the same gene structure than *Vibrio cholerae* O1 biovar El Tor str.: N16961, and they lacked of a transcriptional regulator. In the case of strain DSM 10701, it showed a similar gene structure than *P. putida* U, but *hmgC* and *hppD* were located at more than 10 kb from the other genes. Either the lack of transcriptional regulator or the distant location of the genes involved would prevent their expressions. Moreover, the culture conditions might not have been adequate to allow the growth with 4-hydroxyphenylpyruvate or homogentisate.

In the case of growth with protocatechuate as carbon source, three of the four tested strains which had the corresponding genes were able to grow (A15, B1SMN1, and DSM 4166). As shown in Table 2.11, the only strain with the genes involved in protocatechuate degradation unable to grow was ATCC 17588^T. This strain presented an interruption in the gene encoding for PobA (see Figure 2.4), a 4-hydroxybenzoate-3 monooxygenase which theoretically is involved in the degradation of 4-hydroxybenzoate, but that is not necessary for protocatechuate degradation (see Figure 2.3). Hardwood and Parales (1996) described the induction of the transcription of an enzyme involved in protocatechuate degradation (PcaB) by 4-hydroxybenzoate in *Burkholderia cepacia* DBO1. Kemp and Hegeman (1968) described the capability of 4-hydroxybenzoate to induce protocatechuate 3,4-dioxygenase activity (PcaGH) in *P. aeruginosa* ATCC 15692, suggesting that *P. stutzeri* ATCC 17588^T might require 4-hydroxybenzoate as an inducer to grow with protocatechuate. As expected, no growth was observed for the strains lacking the genes for protocatechuate degradation (see Table 2.11).

Table 2.11. Comparison of tested and expected growth of *P. stutzeri* and *P. balearica* strains with different carbon sources. Positive growth is shown as "+". No growth is represented as "– ". Bold type symbols show discrepancies between the growth observed and the expected growth.

Strains	Protocatechuate	Homogentisate	4-hydroxyphenylpyruvate	Benzoate	Naphthalene
ATCC 17588^T	**–**	**–**	**–**	–	–
A15	+	**–**	**–**	+	–
B1SMN1	+	**–**	**–**	+	+
DSM 4166	+	**–**	**–**	+	–
CCUG 16156	–	**–**	**–**	–	–
AN10	–	**–**	**–**	–	+
19SMN4	–	**–**	**–**	+	+
ST27MN3	–	**–**	**–**	+	+
DSM 10701	–	**–**	**–**	+	–
NF13	–	**–**	**–**	+	–
SP1402^T	–	**–**	**–**	–	+

The results of the cultures with benzoate were also consistent with the genetic analysis in most of the cases. As shown in Table 2.11, growth was observed for 7 of the 10 strains which had the genes for the degradation of benzoate (see Figure 2.6): A15, B1SMN1, DSM 4166, 19SMN4, ST27MN3, DSM 10701, and NF13. B1SMN1 and 19SMN4 had been previously described as benzoate degrader strains (Rosselló-Mora *et al*., 1994). As expected, no growth was observed for strains AN10 and ATCC 17588^T, as previously described Rosselló-Mora and co-workers (1994). As shown in Figure 2.6, these two strains had a disruption by stop codons of *benR* (transcriptional regulator of *ben* genes) and *benD* (dihydroxycyclohexadiene carboxylate

dehydrogenase) genes, respectively. We hypothesized that the disruption of these two genes may cause their inability to use benzoate as carbon source. Besides these two strains, no growth was also observed for *P. balearica* SP1402^T, despite having all the genes involved in this process (see Figure 2.6). We compared the DNA sequences of the BenR binding site for *benA* described in *P. stutzeri* A1501 by Li and co-workers (2010) with the genomes of the different studied strains. As shown in Supplementary Figure S3, the BenR binding site in A1501 is composed by two TGCA sequences and two GGATA sequences that alternate in the sequence. All the studied strains which presented the benzoate degradation genes showed the same structure than A1501 except the genome of strain SP1402^T. The sequence in this strain presented an inversion in the first GGATA sequence, as well as in the -10/-35 promoter consensus sequences (see Supplementary Figure S3). These mutations might prevent the transcriptional regulation activity of BenR. Therefore, we hypothesize that the mutations in the BenR binding site might be responsible of the inability of this strain to grow with benzoate as carbon source.

On the other hand, the ability to grow with benzoate of 19SMN4 is remarkable, despite the fact that this strain lacks the transporter BenK (whose gene is disrupted also by a stop codon, see Figure 2.6). This suggests that this protein might not be essential for benzoate degradation, or that other proteins (most likely the transporters BenE and BenF) might replace the role of BenK.

Finally in the experiments with naphthalene, as described by Rosselló-Mora and co-workers (1994), the five strains tested which presented the naphthalene degrading gene structure (B1SMN1, AN10, 19SMN4, ST27MN3, and SP1402^T) (see Figure 2.14), were able to grow on naphthalene as carbon and energy source (see Table 2.11). Finally, no growth was observed in strains in which no naphthalene degradation gene structure was identified.

To summarize the results of this chapter, we have found the genes involved in the degradation of several aromatic compounds in 19 *P. stutzeri* and *P. balearica* genomes (see Table 2.2). However, the *P. stutzeri* strains analyzed did not cover the whole aromatic hydrocarbon degradation potential described by Lalucat and co-workers (2006) (see figure I4). This is not surprising, since from the 34 strains reviewed by these authors, only five of them were included in the present study (ATCC 17588^T, B1SMN1, AN10, 19SMN4, and ST27MN3). As shown in Figure 2.19, we identified the genes involved in catechol and protocatechuate degradation. García-Valdés and co-workers (2003) described the ability of *P. stutzeri* AN10 to grow with 4-chlorocatechol or 3-methylcatechol as sole carbon and energy source, which are *meta*-cleaved by catechol 2,3-dioxygenase (NahH). Therefore, we believe that the *P. stutzeri* strains which presented the genes involved in the catechol *meta*-cleavage are also able to catabolize chloro and methyl catechol derivatives. The genes involved in gentisate degradation were not found in any of the 18 *P. stutzeri* studied strains. Instead, we identified the genes involved in two additional central pathways: homoprotocatechuate and homogentisate degradation (see Figure 2.19). As for the peripheral pathways, we identified the genes involved in six (4-hydroxybenzoate, salicylate, benzoate, carbazole, phenol, and naphthalene) of the 28 peripheral pathways described by these authors. Additionally, we identified the genes involved in 4-hydroxyphenylpyruvate degradation, which were not described by Lalucat and co-workers (2006). We could not identify genes involved in the degradation of other compounds described by Lalucat and co-workers (2006) such as pyrene, dibenzothiophene, fluorine, and phenanthrene in the genomes of the 18 *P. stutzeri* strains analyzed (see Figure 2.19). In contrast,

we identified genes involved in the degradation of the diterpene dehydroabietic acid. The genes involved in its catabolism of this compound had not been described previously in *P. stutzeri* genomes. Sequencing other *P. stutzeri* genomes will be necessary to establish a more complete aromatic hydrocarbon degradation potential of this species.

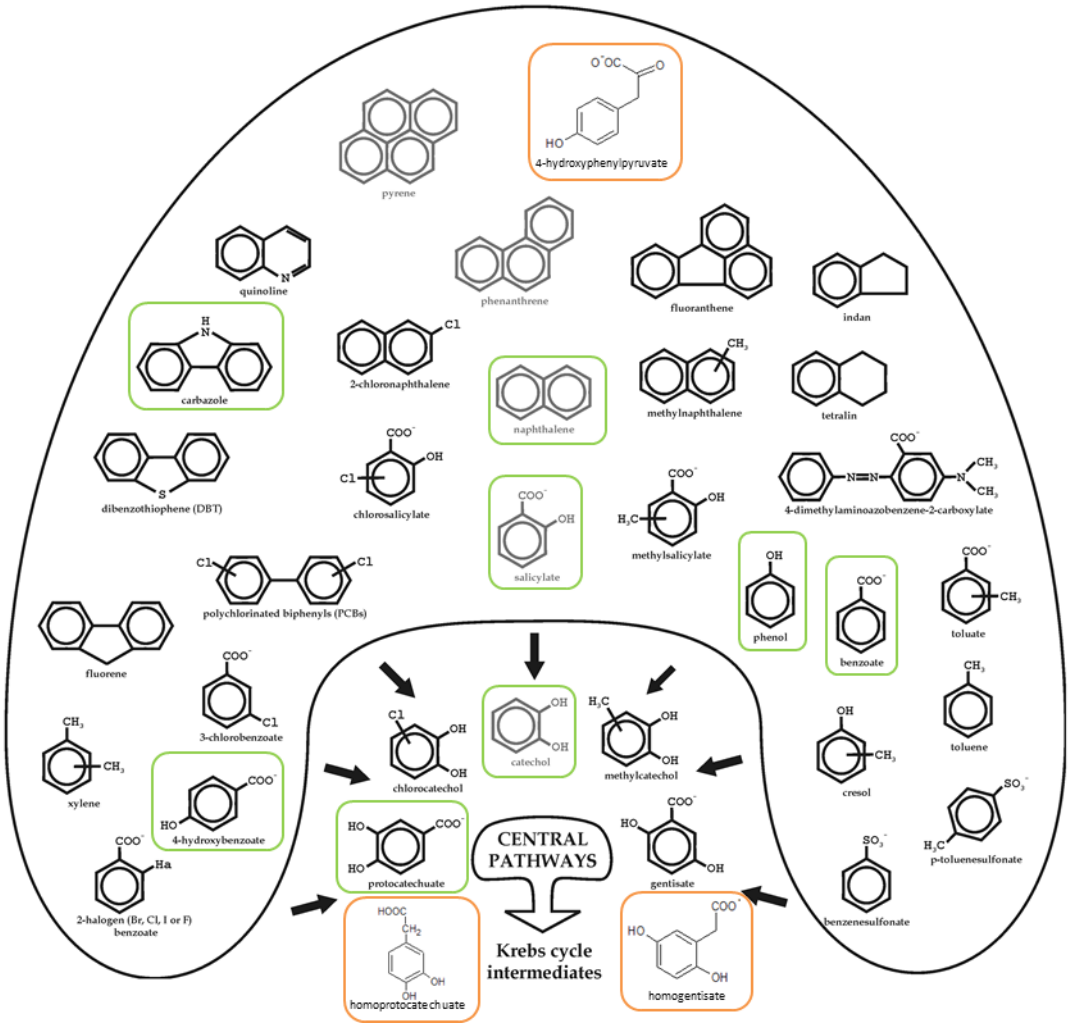

Figure 2.19. Aerobic catabolism of aromatic compounds in *P. stutzeri* before and after the present study (modified from Lalucat *et al*., 2006). In green, compounds which can be catabolized by *P. stutzeri* according to Lalucat and co-workers (2006), and whose degradation genes have been identified in the present study. In orange, compounds whose degradation genes have been identified in the present study, and which were not described by Lalucat and co-workers (2006).

Genes involved in the protocatechuate branch from the β-ketoadipate pathway were found almost exclusively in the genomes of strains belonging to genomovar 1. The exception is strain 28a24, which presented genes homologous to the protocatechuate degradation genes from genomovar 1, but organized in a different structure (see Figure 2.4). Moreover, strains belonging to genomovar 1 are the only ones which presented genes involved in the two branches from the β-ketoadipate pathway. As mentioned in section 3 from Chapter 1, strains belonging to genomovar 1 were the only ones which constituted a coherent group according to the three phylogenomic approaches used in this study. Therefore, the presence of genes involved in the two branches from the β-ketoadipate pathway might be considered as a distinguishing characteristic of this group of strains. From the 11 *P. stutzeri* strains whose growth with protocatechuate was tested in this study, only strains from genomovar 1 except ATCC 17588^T

(A15, B1SMN1, and DSM 4166) were able to use this compound as carbon source. Therefore, the growth with protocatechuate might be used as a phenotypic characteristic for the differentiation of *P. stutzeri* species. Additionally, strains from genomovar 1 are the only ones which showed in their genome genes involved in 4-hydroxybenzoate degradation (except 28a24). In the case of strain ATCC 17588^T, its genome presented both genes involved (*pobA* and *pobR*), although the first one appeared truncated. Both characteristics could be considered for the taxonomists to define the species.

Genes involved in catechol *ortho*- and *meta*-cleavage coexist in the same genome in strains B1SMN1, AN10, 19SMN4, ST27MN3, TS44, and SP1402^T. Genes involved in catechol *meta*-cleavage were found only in the genomes with genes involved in phenol or naphthalene or carbazole (in the case of XLDN-R) degradation. Surprisingly, despite the low abundance of genes involved in catechol *meta*-cleavage in *P. stutzeri* genomes, we identified two different genes encoding for the enzyme catechol 2,3-dioxygenase in the genome of strain KOS6. One of them is located with naphthalene degradation genes and the other one is located with phenol degradation genes. The well-organized gene structure of genes involved in both pathways, as well as the presence of genes coding for transcriptional regulators, suggest that the two catechol 2,3-dioxygenases of KOS6 would not act indistinctly in naphthalene and phenol degradation.

Our results revealed the presence of genes involved in 4-hydroxyphenylpyruvate via homogentisate in all the 18 *P. stutzeri* genomes. This is consistent with the described by Jiménez and co-workers (2010), who also identified genes involved in this pathway in 18 *Pseudomonas* genomes studied belonging to 7 different species (*P. putida*, *P. entomophila*, *P. stutzeri*, *P. fluorescens*, *P. aeruginosa*, *P. mendocina*, and *P. syringae*). In the genomes of strains CCUG 16156 and MF28, we only identified genes involved in these two pathways (4-hydroxyphenylpyruvate and homogentisate degradation). Therefore, strains CCUG 16156 and MF28 have a very limited range of aromatic compounds that can be used. Genes involved in benzoate degradation were also generally found in many of the studied genomes (14 *P. stutzeri* genomes). In contrast, genes involved in salicylate, carbazole, phenol, and naphthalene degradation seem to be rare in *P. stutzeri* genomes, and they were only found in strains either isolated under selective pressure or from polluted environments. Therefore, these results suggest that some aromatic hydrocarbon catabolic pathways are widespread in *P. stutzeri*, and other pathways are more unusual, although they might be found under selective pressure. In relation to *P. balearica*, its aromatic hydrocarbon catabolic potential was found to be similar to the described for *P. stutzeri* species. The only *P. balearica* studied genome, from strain SP1402^T, presented the genes involved in the catabolism of 4-hydroxyphenylpyruvate via homogentisate, benzoate via catechol *ortho*-cleavage, and naphthalene via salicylate and catechol *meta*-cleavage. We could not find genes involved in aromatic hydrocarbon degradation unique to SP1402^T genome. However, an extended analysis of *P. balearica* genomes should be performed to characterize its catabolic potential.

All the genes involved in aromatic hydrocarbon degradation pathways identified in the present study were previously described in *Pseudomonas* spp. genomes by Jiménez and co-workers (2010). These authors analyzed the genes involved in aromatic hydrocarbon degradation presented in the genomes of 18 *Pseudomonas* strains belonging to 7 different species, including *P. stutzeri* A1501 (see Figure 2.20). The study revealed the presence of at least 11 central

pathways: β-ketoadipate (catechol and protocatechuate branches), homogentisate, homoprotocatechuate, gentisate, 2,3-dihydroxy-*p*-cumate, aminophenol, alkylcatechols, phenylacetyl-CoA, 2,5-dihydroxypiridine, and hydroquinone degradation, to which many different peripheral pathways converged. Homogentisate and β-ketoadipate (catechol and protocatechuate branches) pathways were the most widespread in *Pseudomonas*, according to the results described by Jiménez and co-workers (2010). They attributed this abundance to the fact that their substrates are common carbon sources in the environment, suggesting that their acquisition was an old evolutionary event. *P. putida, P. aeruginosa,* and *P. entomophila* were the *Pseudomonas* species with the largest number of different central pathways described, with 8, 6, and 6 pathways respectively. Genes involved in the β-ketoadipate (both branches), and homogentisate central pathways were identified by Jiménez and co-workers (2010) in the genome of *P. stutzeri* A1501. Additionally, the genomic study presented in this thesis revealed the presence of genes involved in the homoprotocatechuate central degradation pathway in *P. stutzeri* genomes (see Figure 2.20, Table 2.2, and section 1.4 for details). However, we could not identify genes involved in gentisate, cumate, aminophenol, alkylcatechols, phenylacetyl-CoA, dihydroxypiridine, and hydroquinone degradation in the *P. stutzeri* genomes studied in this thesis. Considering the aromatic hydrocarbon degradation peripheral pathways, genes involved in 16 different pathways were described by Jiménez and co-workers (2010) in the genomes of 18 *Pseudomonas* strains, being 4-hydroxyphenylpyruvate, 4-hydroxybenzoate, and benzoate degradation the most commonly found (see Figure 2.20). *P. putida*, *P. fluorescens*, and *P. entomophila* were the *Pseudomonas* species with the largest number of different peripheral pathways described (with 12, 9 and 9 different pathwaysrespectively). Genes involved in benzoate, 4-hydroxybenzoate, and 4-hydroxyphenylpyruvate degradation were identified by Jiménez and co-workers (2010) in the genome of *P. stutzeri* A1501. Additionally, the genomic study presented in this thesis revealed the presence genes involved in the catabolism of salicylate, carbazole, phenol, and naphthalene in *P. stutzeri* genomes. However, we could not identify genes involved in many peripheral pathways such as phenylethanol, phenylethylamine, benzene, nicotinate, vanillin, mandelate, and coumarate degradation in the *P. stutzeri* genomes studied in this thesis. These results suggest a wider aromatic hydrocarbon degradation potential of *P. putida*, *P. fluorescens*, and *P. entomophila* compared with *P. stutzeri*.

The genus *Burkholderia* has been also extensively studied as a model for aromatic hydrocarbon degradation, as described by O'Sullivan and Mahenthiralingam (2005). Pérez-Pantoja and co-workers (2012) reviewed the genomes of 35 *Burkholderia* strains, belonging to at least 16 different species, able to degrade aromatic compounds. The more widespread central pathways in bacteria of this genus were the phenylacetyl-CoA, the β-ketoadipate pathway (protocatechuate and catechol branches), and the homogentisate and homoprotocatechuate degradation pathways, which were found in most of the genomes analyzed. As for the peripheral degradation pathways, genes involved in 4-hydroxybenzoate, benzoate, anthranilate, and 4-hydroxyphenylpyruvate degradation were found in most of the 35 *Burkholderia* genomes. The wide range of central and peripheral pathways in *Burkholderia* compared to *P. stutzeri* revealed that *Burkholderia* might be able to transform a larger number of aromatic compounds than *P. stutzeri*. However, it has to take into account that Pérez-Pantoja and co-workers (2012) analyzed a higher number of *Burkholderia* strains (35) compared to the *P. stutzeri* strains used in this thesis (18).

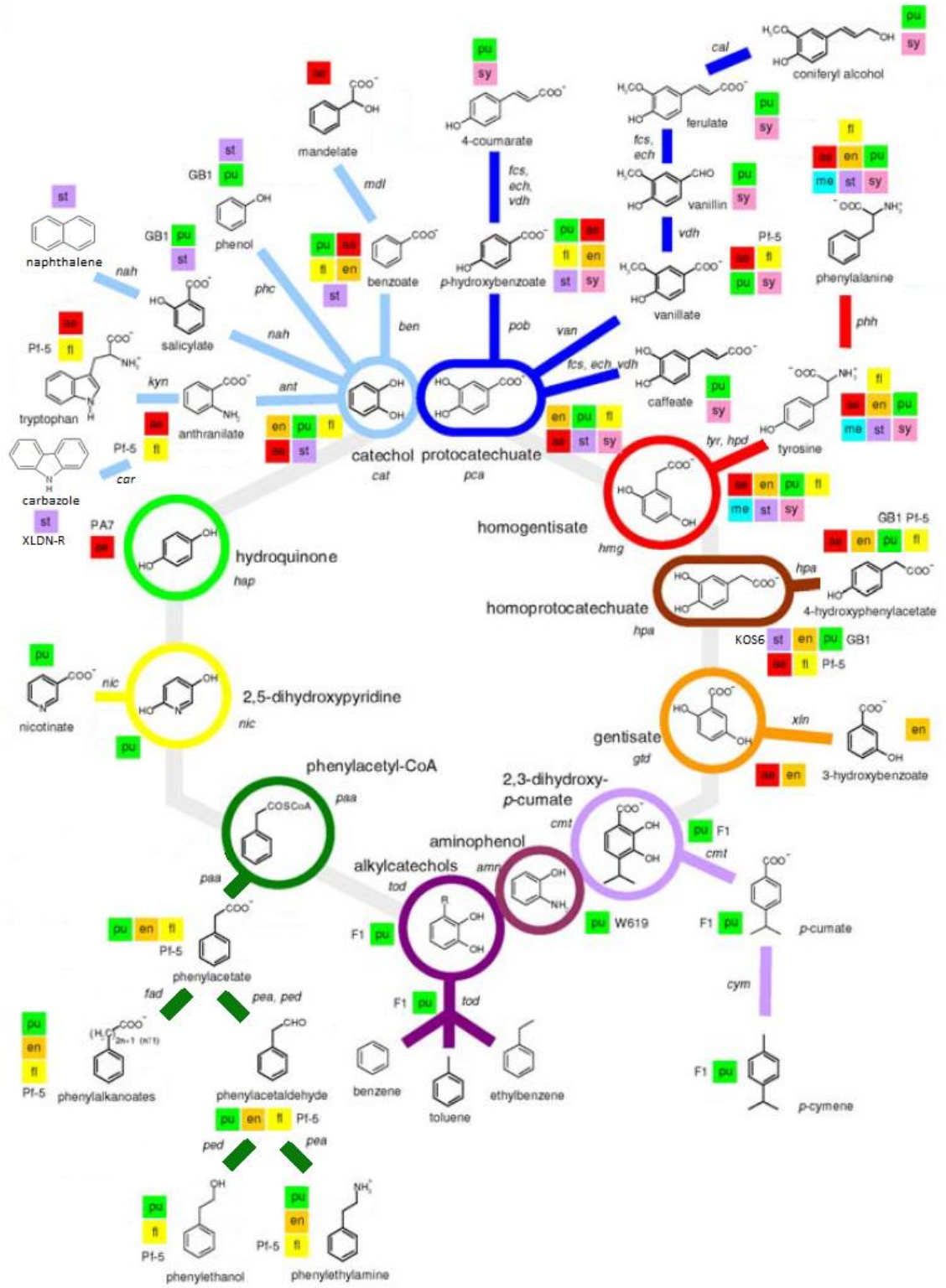

Figure 2.20. Aromatic catabolic pathways and the genes involved in different *Pseudomonas* genomes. The different peripheral pathways that converge in the same central intermediate are indicated. The name of the genes coding for the peripheral and central pathways is also shown. Genomes used in this analysis: *P. aeruginosa* strains PAO1, PA7, PA14, PACS2, and 2192 (ae); *P. fluorescens* strains SBW25, Pf-01 and Pf-5 (fl); *P. entomophila* strain L48 (en); *P. mendocina* strain ymp (me); *P. putida* strains KT2440, GB1, F1 and W619 (pu); the 18 *P. stutzeri* strains of the present study (st); *P. syringae* strains 1448A, B728a, and DC3000 (sy). The names of the strains are specified in those cases where the genes/pathways were identified only in some strains (but not in all available strains) of a particular species. Genes located in catabolic plasmids have not been taken in consideration. Modified from Jimenez and co-workers (2010).

CHAPTER 3: Proteomic study of naphthalene degradation by *Pseudomonas stutzeri*

1. *P. stutzeri* proteins detected by proteomics

Catabolic genes involved in naphthalene degradation by *P. stutzeri* AN10 were previously described (Bosch *et al*., 1999a; Bosch *et al*., 1999b; Bosch *et al*., 2000). Moreover, the enzymatic activities involved in this pathway (naphthalene dioxygenase, salicylaldehyde dehydrogenase, salicylate hydroxylase, catechol 1,2 dioxygenase, and catechol 2,3 dioxygenase) were measured in strains B1SMN1, AN10 and 19SMN4 (Rosselló-Mora *et al*., 1994). Additionally, as described in Chapter 2, we identified the set of genes involved in naphthalene degradation in these four *P. stutzeri* strains, in strain KOS6 and in *P. balearica* SP1402^T (see Figure 2.14). However, accessory proteins such as membrane transporters and efflux pumps involved in the metabolism of naphthalene remain unknown. For this reason, a proteomic analysis of four naphthalene degrading *P. stutzeri* strains (AN10, B1SMN1, 19SMN4 and ST27MN3) and *P. balearica* SP1402^T was performed. Cultures were done in MMB supplemented with 3 different carbon sources: (1) succinate 30 mM; (2) succinate 30 mM with a pulse of salicylate 3 mM during 4 hours, and (3) naphthalene vapors (see section 4.1 from Material and Methods for details). With the study of the proteome of the cultures with a pulse of salicylate we wanted to identify proteins induced by salicylate, which is the inducer of the naphthalene degradation pathway described so far (Schell and Wender, 1986). In contrast, the proteomic study of cultures with naphthalene should allow us to identify the proteins either induced by naphthalene or as a result of the chemical stress caused by naphthalene.

A total of 7,585 different proteins were detected in the five strains, which represents an average of 34.6 % of the annotated CDSs per genome. Protein detectability ranged between 29.8 % (B1SMN1) and 37.3 % (19SMN4) of annotated CDSs. The percentage of detected proteins in B1SMN1 was slightly lower due to the greater amount of annotated CDSs in its genome (5,103 annotated CDSs). However, the number of detected proteins was similar to the average of the detected proteins of the remaining strains. In contrast, SP1402^T was the strain with less detected proteins (1,381 proteins), although the proportion of detected proteins remained similar to the other strains. All detected proteins are shown in Supplementary Table S12. In a similar study, Kasahara and co-workers (2012) were able to detect a 30 – 45 % of the known proteins analyzing the proteome of *P. putida* F1 growing with different aromatic hydrocarbons as the sole carbon source (such as toluene, ethylbenzene, benzene, cymene, and cumate).

2. General proteome comparison

In order to study the variability of the different samples, all detected CDSs were grouped according to their homology with at least a 75 % of identity in at least the 90 % of the sequence. In this case we used this criteria since it was previously described that a sequence identity above 60 % in at least 80 % of the sequence was necessary for an accurate functional classification (Rost, 2002; Tian and Skolnick, 2003; Espadaler *et al*., 2008). Once the detected CDSs were grouped, we constructed a matrix with the total sum of NAF values of each group of proteins in each proteome, which was used as input for PCA analysis (see section 4.1 from Materials and Methods for details). As shown in Figure 3.1, principal component 1 (PC1) and PC2 explained 49 % of the variance in total proteome observed in the samples. Triplicates from the same culture condition appeared grouped, except from one replicate of SP1402^T growing with salicylate. PCA

also showed that proteomes from succinate and naphthalene cultures were the two conditions which presented the higher differences. In contrast, proteomes obtained from salicylate cultures were more related with those obtained from succinate cultures rather than those from naphthalene cultures. Therefore, the proteomes of the studied strains seemed to be more affected by the use of naphthalene as carbon source rather than by the presence of the inductor salicylate.

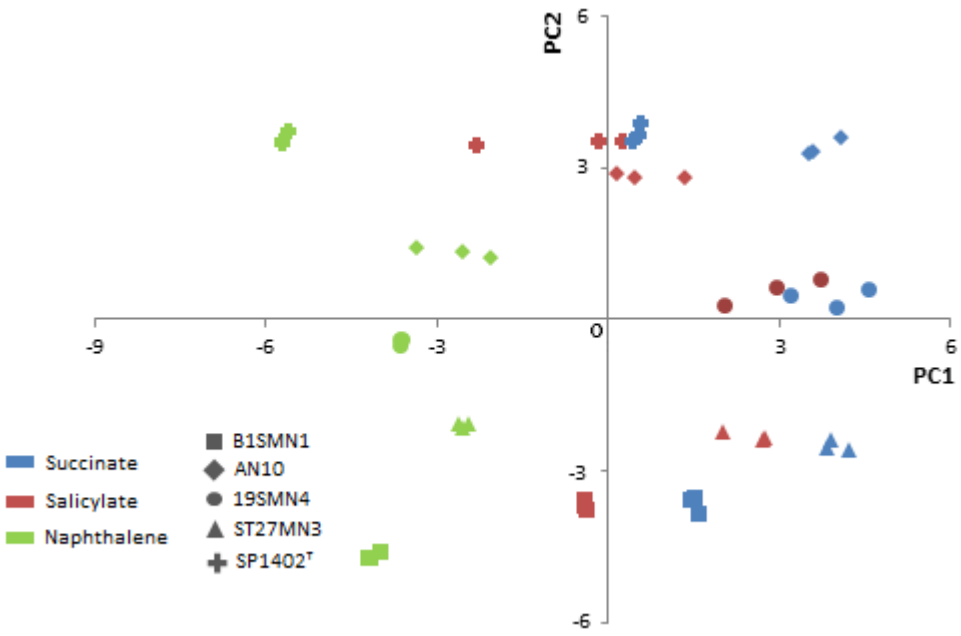

Figure 3.1. Principal component analysis of NAF values obtained for proteomes from *P. stutzeri* strains B1SMN1, AN10, 19SMN4, ST27MN3, and *P. balearica* SP1402^T. Percentages of variance explained: PC1 26 %; PC2 23 %.

In order to establish general differences between the different proteomes, we analyzed the NAF sum of all proteins classified in the different biological functions described by KEGG. Classification of all the studied proteins into the KEGG biological functions is shown in Supplementary Table S3. As shown in Figure 3.2, the group of biological functions which presented greater variations in abundance between the different conditions corresponded to xenobiotics metabolism. In this case, proteomes from naphthalene cultures, and to a lesser extent cultures with a salicylate pulse, presented higher NAF values than proteomes from succinate cultures for proteins involved in xenobiotics metabolism. This result was expected since it is due to the fact that this functional group includes proteins involved in naphthalene degradation pathway (see next section for details).

Besides xenobiotics metabolism, carbohydrate metabolism was also altered. In this case, proteins involved in carbohydrate metabolism increased in abundance in proteomes from naphthalene cultures, compared to succinate and salicylate cultures, which remained similar (see Figure 3.2). Statistically significant variations (p-value < 0.05) were found for proteins from different pathways of carbohydrate metabolism: gluconeogenesis and glycolysis, pyruvate metabolism, and butanoate metabolism (see Table 3.1). These results contradict the described by Li and co-workers (2015). These authors analyzed the proteomic changes of *P. putida* ND6 growing with glucose or naphthalene as carbon source, and in the case of proteins involved in

carbon metabolism they described a down-regulation of proteins involved in glycolysis and TCA cycle. In particular, they observed down-regulation of a phosphopyruvate hydratase (AFK68149), a dihydrolipoamide succinyltransferase (AFK71337), a NADP-dependent isocitrate dehydrogenase (AFK71507), and an isocitrate lyase (AFK71425).

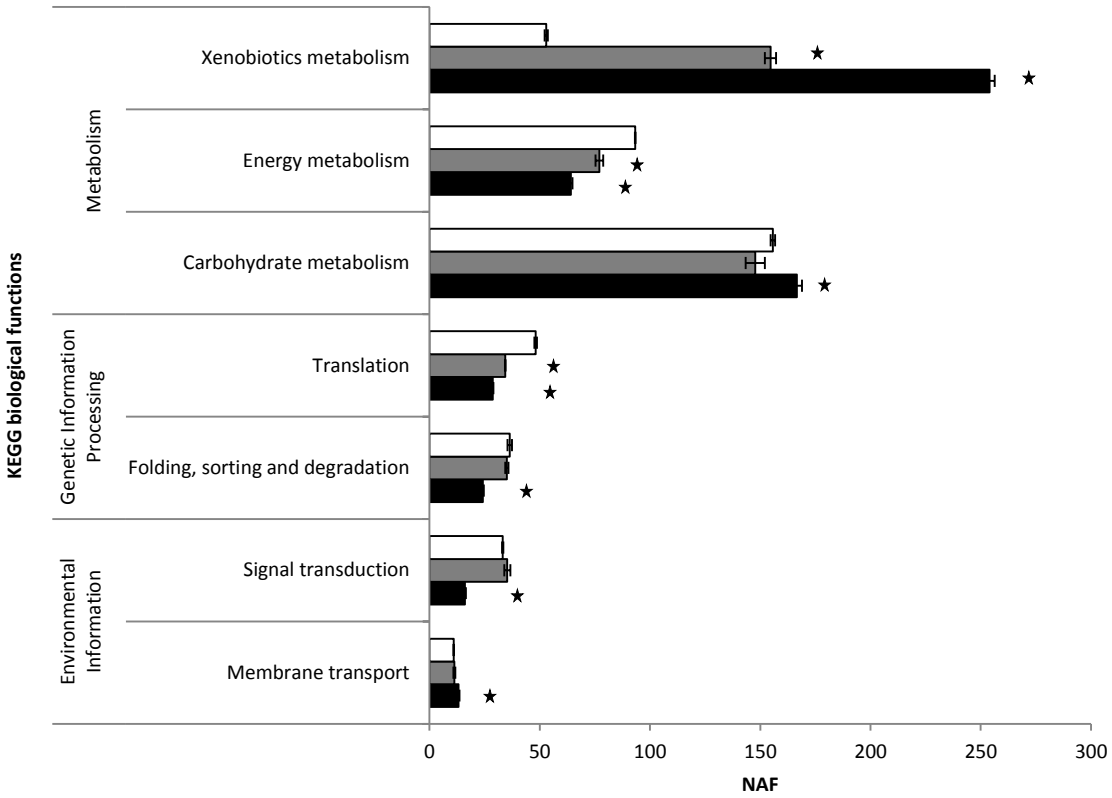

Figure 3.2. Sum of NAF values of proteins involved in different global biological functions (KEGG classification). Only functions which presented statistically significant variations between the different culture carbon sources (succinate, white; salicylate pulse, gray; and naphthalene, black) are shown. Statistically-significant abundance fold changes between salicylate pulse and naphthalene proteomes respect with succinate proteomes are marked with a star. Error bars represent the standard deviation between the three replicates of each culture condition.

In contrast, proteins involved in energy metabolism presented lower NAF values in proteomes from naphthalene cultures compared to salicylate and, specially, succinate cultures (see Figure 3.2). Analyzing all pathways involved in energy metabolism, we observed that this decrease was statistically significant in proteins involved in oxidative phosphorylation (see Table 3.1). A more exhaustive analysis of this group of proteins revealed that from the 207 proteins classified by KEGG into the oxidative phosphorylation pathway 113 of them were detected in our proteomes. The results showed that 43 % of the detected proteins presented a significant decrease in abundance in naphthalene cultures. A particular example of this normalized reduction in abundance is the cytochrome *cbb*$_3$ oxidase. This enzyme represents a distinctive class of proton-pumping respiratory heme-copper oxidases. This protein has been reported to have a very high affinity for oxygen, which give *cbb*$_3$ oxidases an essential role in the energy metabolism in microaerobic environments (Pitcher *et al.,* 2002). *P. stutzeri* cytochrome *cbb*$_3$ oxidase was previously described to be a complex formed by 3 subunits (CcoN, CcoO, and CcoP), being CcoN the catalytic subunit of the enzyme (Pitcher *et al*., 2002). The normalized abundance of these proteins was significantly reduced in proteomes from naphthalene cultures compared to

succinate cultures (fold changes above 2, p-value <0.05), except for CcoN in B1SMN1, and CcoO in 19SMN4 and ST27MN3 proteomes. Domínguez-Cuevas and co-workers (2006) previously described a reduction in the abundance of proteins involved in energy metabolism, and particularly of 21 cytochrome cbb_3 oxidase subunits, in *P. putida* KT2440 cells exposed to toluene as well. These authors hypothesized that the reduction of these proteins is due to the membrane damage caused by toluene, which then caused an oxidative stress in *P. putida* KT2440 cells.

Additionally, a significant decrease in the abundance of proteins involved in denitrification was detected in proteomes from naphthalene cultures compared to succinate cultures. A total of 167 proteins were encoded by denitrification operons (see Figure 1.2, from Chapter 1) in the 5 studied genomes, and 43 of them were detected by proteomics. Fifty-six percent of these proteins, all of them nitrite reductase subunits, had significantly lower normalized abundances in proteomes from naphthalene cultures compared to succinate cultures (fold changes above 2, p-value <0.05).

Table 3.1. Sum of NAF values of proteins involved in different biological functions (KEGG classification). Values from salicylate pulse and naphthalene cultures statistically different (p< 0.05) from values from succinate cultures are shown in bold.

KEGG biological function		Succinate	Salicylate	Naphthalene
Carbohydrate metabolism	Glycolysis / Gluconeogenesis	20.6 ± 0.2	18.6 ± 0.6	**23.9 ± 0.4**
	Pyruvate metabolism	35.5 ± 0.4	34.8 ± 0.8	**41.7 ± 0.8**
	Butanoate metabolism	11 ± 0.4	**13.7 ± 0.2**	**15 ± 0.2**
Energy metabolism	Oxidative phosphorylation	23.7 ± 0.2	**18.4 ± 0.6**	**10.5 ± 0.2**
Membrane transport	ABC transporters	9.1 ± 0.2	9.5 ± 0.5	**11.6 ± 0.5**
Signal transduction	Two-component system	33.2 ± 0.2	35.3 ± 1.4	**16.1 ± 0.4**
Translation	Ribosomal proteins	40.5 ± 0.7	**28.4 ± 0.1**	**23.2 ± 0.3**
Folding, sorting and degradation	RNA degradation	28.9 ± 0.9	**26.6 ± 0.3**	**18.2 ± 0.2**

The reduction in the abundance of proteins involved in energy metabolism does not agree with what was observed by Wijte and co-workers (2010) in a proteomic study of *P. putida* S12 growing with toluene. They observed an increase of more than 25 energy metabolism proteins (involved in TCA cycle, ATP protein motive force, and glycolysis) in cells from 3 mM and 5 mM toluene cultures. These authors deduced that the increase in the abundance of energy metabolism proteins was due to the increase in energy consumption observed in *P. putida* S12 growing with toluene. This energy is required for the expression and action of different solvent efflux pump systems responsible of the solvent tolerance of this bacterium (such as SrpABC). In contrast, our results are consistent with the proteomic study of *P. aeruginosa* SJTD-1 performed by Liu and co-workers (2015), in which the abundance of 31 proteins involved in energy production decreased in the presence of n-octadecane. Moreover, Domínguez-Cuevas and co-workers (2006) previously described a reduction in the transcription of all detected genes involved in energy metabolism (22 of them annotated as cytochromes) in *P. putida* KT2440 cells exposed to toluene. These authors hypothesized that the exposition to aromatic compounds caused damage at the level of the cell envelope, which lead to oxidative damage that is observed as a reduction in electron transport chain activity, causing a general oxidative stress. Li and co-

workers (2015) hypothesized that the decreased expression of proteins involved in glycolysis and TCA cycle in *P. putida* ND6 growing with naphthalene indicates that energy production was inhibited due to the toxicity caused by higher naphthalene concentration.

Apart from the variations in the abundance of proteins involved in basal metabolism mentioned above, we detected variations in proteins involved in other biological functions. The NAF values of proteins involved in membrane transport (ABC transporters) significantly increased (fold changes above 2, p-value <0.05) in the proteomes from naphthalene cultures compared to salicylate and succinate cultures (see Table 3.1). The changes in abundances of membrane transport proteins will be discussed later in section 3.2. In contrast, proteins involved in signal transduction, mainly two-component system proteins (see Table 3.1) presented statistically-significant lower NAF values in naphthalene cultures compared to succinate and salicylate cultures.

On the other hand, in the proteomes of naphthalene cultures we observed lower normalized abundances of proteins involved in: translation, mainly ribosomal proteins (see Table 3.1); and folding and degradation of genetic material, mainly RNA degradation (see Table 3.1). The decrease of ribosomal proteins contradicts the previous results of Domínguez-Cuevas and co-workers (2006) in *P. putida* KT2440 cells exposed to toluene, which described the increase of the abundance of 10 different ribosomal proteins. However, our results are consistent with the down-regulation of proteins involved in protein synthesis in *P. putida* ND6 naphthalene cultures compared to glucose cultures described by Li and co-workers (2015). These authors attributed this down-regulation to an ATP deficiency resulting from naphthalene toxicity. Examples related to the other protein changes observed have not been found in the literature.

3. Comparison of proteomes from naphthalene and succinate cultures

3.1. Proteomic changes in naphthalene degradation proteins

As shown in Figure 3.2, the normalized abundance of proteins involved in xenobiotics metabolism in proteomes from naphthalene cultures was 5-fold larger than the normalized abundance for succinate cultures. Analyzing it in detail, we observed that this increase was due to an increase of the normalized abundance of proteins involved in naphthalene degradation, which are encoded by *nah* operons. Analyzing each strain separately, we observed that 20 of the 23 proteins encoded by the *nah* operons (see Figure 2.14 from Chapter 2) were detected by proteomics in all strains. The exceptions were: NahAb, the ferredoxin subunit of naphthalene 1,2-dioxygenase (104 amino acids, 11.5 KDa), which was not detected in B1SMN1; the ferredoxin NahT (112 amino acids, 12.5 KDa), which was not detected in any strain; and the 4-oxalocrotonate isomerase NahJ (63 amino acids, 6.9 KDa), which was only detected in AN10. The probability of detecting a protein by LC-MS/MS decreases for low molecular weight proteins, as it is less probable that the trypsin digestion produces peptides with the proper mass/charge ratio to be detected. In the particular case of this study, detected proteins presented an average of 379 ± 228 amino acids in length, and only 9.6 % of the proteins with less than 100 amino acids were detected by proteomics. Consequently, the lack of detection of these 3 proteins (NahAb, NahT, and NahJ) was most likely related to their small size.

As shown in Figure 3.3, the normalized abundance of 16 of the 23 naphthalene degradation proteins presented statistically-significant (p-value < 0.05) 2-fold increases in proteomes from naphthalene cultures compared to succinate cultures in all the studied strains. As for the rest, we found several situations. In the case of the salicylate hydroxylase NahW and the acetaldehyde dehydrogenase NahO, there was a significant increase in all strains except B1SMN1 and AN10 respectively, in which despite been detected, their normalized abundance did not change significantly. In contrast, NahZ and NahR, despite being detected in all strains, only increased significantly in AN10. Moreover, the abundance of NahAb increased in 19SMN4 and ST27MN3, while in SP1402^T and AN10 its abundance remained constant.

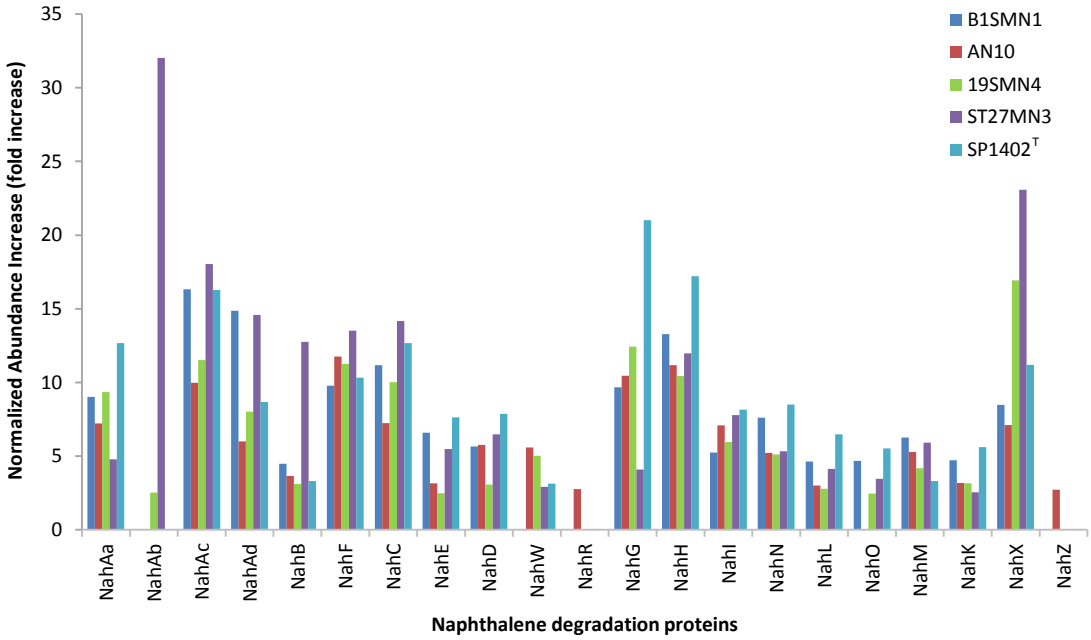

Figure 3.3. Statistically-significant increases in the abundance of naphthalene degradation proteins in proteomes from naphthalene cultures compared to succinate cultures. NahT is not shown because it was not detected in any strain. NahJ is not shown because its normalized abundance did not vary significantly in the only strain in which it was detected (AN10).

In a similar study, Li and co-workers (2015) analyzed the protein expression of *P. putida* ND6 growing with naphthalene or glucose as carbon source. This strain is able to metabolize naphthalene via the catechol *meta*- and *ortho*- cleavage pathways. In the case of proteins involved in naphthalene degradation, they observed an up-regulation of proteins involved in the catechol *ortho*- cleavage pathway (CatA, CatB, and CatC), while the abundance of proteins involved in the catechol *meta*- cleavage pathway did not change. Therefore, in contrast to the *P. stutzeri* and *P. balearica* strains analyzed in this thesis, in *P. putida* ND6 the catechol *ortho*-cleavage pathway played a predominant role in naphthalene degradation.

As mentioned in the previous chapter, the role of most of the proteins involved in naphthalene degradation has been previously described (Bosch *et al.,* 1999a; Bosch *et al.,* 1999b; Bosch *et al.,* 2000; Lanfranconi, *et al.*, 2009). The only exceptions for that are NahX and NahZ, whose role in the naphthalene metabolism has not been described yet. As shown in Figure 3.3, the normalized abundance of NahX increased significantly in proteomes from naphthalene cultures compared to succinate cultures in all strains. As mentioned in Chapter 2 (section 1.7), NahX has

been annotated by NCBI as an ATP: corrinoid adenosyltransferase, an enzyme responsible of the conversion of cobalamin (vitamin B12) to adenosylcobalamin (coenzyme B12). This cofactor has been reported as essential for a wide range of enzymatic reactions, such as methylations, reductions and isomerizations (Lee *et al*., 1999; Giedyk *et al*., 2015). We hypothesized that NahX might provide the cofactor adenosylcobalamin for the isomerization of 4-oxalocrotonate by NahJ (see section 1.7 from Chapter 2 for details).

In the case of NahZ, this protein was annotated by the NCBI pipeline as a FAD-dependent oxidase. As shown in Figure 2.13 (Chapter 2), one molecule of naphthalene is transformed subsequently to one molecule of acetyl-CoA, which directly enters the TCA cycle, and two molecules of pyruvate. An additional enzyme, pyruvate dehydrogenase, is needed to channel pyruvate to the TCA cycle (Recny *et al.,* 1982) (Figure 3.4). The gene for pyruvate dehydrogenase is not encoded in the *nah* operons but it is present in the genome of all *P. stutzeri* strains analyzed (see Supplementary Table S2). But there is another possibility to channel pyruvate to the TCA cycle. A FAD-dependent pyruvate oxidase which transforms pyruvate to acetyl phosphate has been previously described in *E. coli* (Williams and Hager, 1966). Hydrogen peroxide is also produced, which is further transformed into water by catalase. The acetyl phosphate can then be transformed to acetyl-CoA by the consecutive action of an acetate kinase and an acetyl-CoA synthetase (Recny *et al.,* 1982). Therefore, we hypothesized that NahZ could be involved in the transformation of pyruvate to acetyl phosphate during naphthalene degradation (Figure 3.4). To test this hypothesis we analyzed the proteomes to see if the rest of putative enzymes leading to the formation of acetyl-CoA via this route were also detected. We observed a significant increase in the normalized abundance of NahZ in AN10 naphthalene cultures respect to succinate cultures. As shown in Table 3.2, two AN10 proteins identified as acetate kinase (YP_006458872.1; see Supplementary Table S2) and acetyl-CoA synthetase (YP_006458092.1; see Supplementary Table S2) showed higher normalized abundance in cells from naphthalene cultures. The proteomes from naphthalene cultures of the other strains did not show significant variation in normalized abundance of NahZ or acetate kinase proteins.

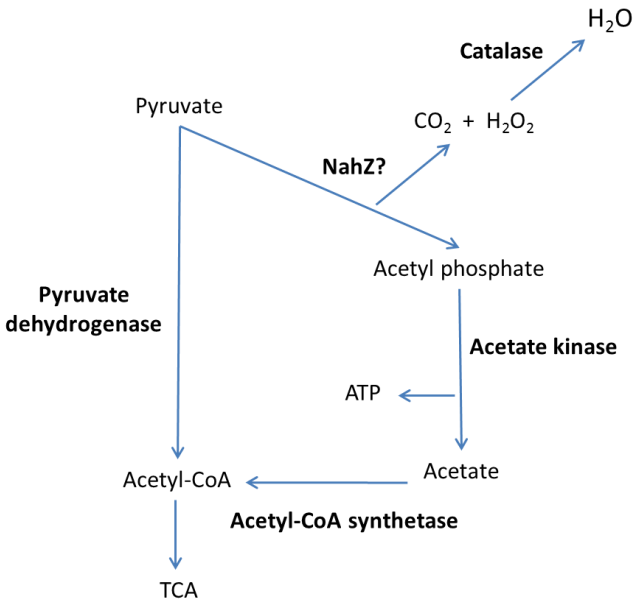

Figure 3.4. Putative role of NahZ in naphthalene metabolism in *P. stutzeri*.

There were also two different proteins identified as pyruvate dehydrogenase and two proteins identified as catalase detected by proteomics in AN10 cultures. For both enzymes there were one protein which increased its normalized abundance in naphthalene cultures (YP_006456119.1 and YP_006457434.1; see Supplementary Table S2) and one protein which decreased. This suggests that this FAD-dependent pyruvate oxidase pathway could be also involved in naphthalene degradation, helping in the channeling of pyruvate to acetyl-CoA. We can hypothesize that NahZ would act as a FAD-dependent pyruvate oxidase (see Figure 3.4). However, NahZ sequence from AN10 do not show sequence homology with pyruvate oxidase sequences from *E. coli*. Therefore, the role of NahZ as a pyruvate oxidase needs to be experimentally proved.

Table 3.2. Fold change obtained for proteins NahZ, acetate kinase, acetyl-CoA synthetase, catalase, and pyruvate dehydrogenase in *P. stutzeri* AN10 naphthalene cultures compared to succinate cultures.

Protein	ID	Fold change	p-value
NahZ	YP_006456966.1	2.7	1E-02
Acetate kinase	YP_006458872.1	3.7	2E-03
Acetyl-CoA synthetase	YP_006458092.1	12.7	1E-04
Catalase	YP_006457434.1	2.6	1E-03
	YP_006459239.1	-4.3	5E-04
Pyruvate dehydrogenase	YP_006456118.1	-1.3	1E-01
	YP_006456119.1	1.6	2E-02

3.2. Putative transporters involved in naphthalene degradation

Additionally to proteins directly involved in naphthalene degradation, statistically-significant increases in the normalized abundance of proteins annotated as transporters or transport systems were also detected in proteomes from naphthalene compared to succinate cultures (2-fold increase, p-value < 0.05; see Supplementary Table S13). A total of 14 transport systems were classified using the Transporter Classification DataBase (TCDB) into 8 different families: ATP-Binding Cassette systems (ABC), Resistance-Nodulation-Cell Division systems (RND) Tripartite ATP-independent Periplasmic Transporter systems (TRAP-T), FadL system, Outer Membrane Anion Porin system (TsaT), Solute:Sodium Symporter system (SSS), TonB system, and Major Facilitator Protein system (MFS) (see Table 3.3).

Of all the transport systems identified that might have a role in naphthalene metabolism, some of them highlighted due to their homology with transporters related to aromatic hydrocarbon metabolism previously described. The first of them is NahP whose encoding gene is located downstream *nahD* in the genomes of strains AN10, 19SMN4, ST27MN3 and SP1402^T (see Figure 2.14 in Chapter 2). In B1SMN1 genome, we were not able to define the location in the genome, as *nahP* and *nahD* are located in different contigs. NahP proteins were classified into the FadL family, as they showed an E-value lower than 1E-5 with a FadL-type protein from the TCDB database. This family of transporters belongs to the superfamily of porins, but in contrast to most described porins, FadL is monomeric (van den Berg *et al*., 2004). The FadL protein of *E. coli*

gives name to the family. This protein is involved in long-chain fatty acids transport through the outer membrane (Bae *et al.,* 2014). Several FadL proteins involved in biodegradation have been described, such as XylN (m-xylene transport in the TOL plasmid pWW0 of *P. putida* F1) and TbuX (toluene transport in *Ralstonia pickettii* PKO1) (Wang *et al*., 1995; Kasai *et al*., 2001). The NahP sequences identified in *P. stutzeri* and *P. balearica* had 83 % identity with the XylN protein encoded in the TOL plasmid pWW0. The normalized abundance of NahP was significantly higher in proteomes from naphthalene cultures in 4 of the studied strains: B1SMN1, 19SMN4, ST27MN3, and SP1402^T (Figure 3.5). In the case of strain AN10, a 27.96-fold increase was observed, although this value was not statistically-significant, as its p-value was greater than 0.05, probably due to the variability observed between the different replicates. These results suggests that NahP might have a role in naphthalene metabolism, probably allowing the entrance of naphthalene through the outer membrane, as XylN and TodX with m-xylene and toluene, respectively; or removing toxic naphthalene degradation intermediates (such as salicylate) from the cell.

Table 3.3. Comparison of the number of transport systems from different families identified in each proteome and the number of transport systems whose normalized abundance increased significantly (some of their proteins) in proteomes from naphthalene (shown in brackets).

Family	No. per family	B1SMN1	AN10	19SMN4	ST27MN3	SP1402^T
ABC	5	3 (3)	5 (5)	5 (5)	5 (5)	3 (2)
RND	2	2 (2)	1 (0)	3 (1)	2 (2)	1 (0)
TRAP-T	2	2 (2)	1 (1)	2 (2)	1 (1)	0 (0)
FadL	1	1 (1)	1 (0)	1 (1)	1 (1)	1 (1)
TsaT	1	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)
SSS	1	1 (1)	1 (1)	1 (1)	1 (1)	0 (0)
TonB	1	1 (0)	1 (1)	1 (1)	1 (1)	1 (1)
MFS	1	0 (0)	0 (0)	1 (1)	1 (1)	1 (0)
Total	14	11 (10)	10 (8)	14 (12)	12 (12)	7 (4)

The second transport system of interest was a MexAB-OprM system (RND family) identified in all the studied strains. Strains 19SMN4 and ST27MN3 presented two copies of these genes in their genomes, which were similar but not identical (see Supplementary Table S13). In the genome of 19SMN4, the first set of genes (*mexA*, *mexB*, and *oprM*) are chromosomally located, and the second set of genes (*mexAb*, *mexBb*, and *oprMb*) are located in its plasmid (pLIB119). We found that, all proteins of this system presented E-values lower than 1E-5 with RND-type transporters from the TCDB database. RND transport systems have been widely studied in gram negative bacteria, mainly because of their role in antibiotic resistance. These transport systems consist of 3 proteins: an inner membrane protein (for example MexB), an outer membrane protein (such as OprM) and a membrane fusion protein (such as MexA). The activity of this system is coupled to a proton gradient (Alvarez-Ortega *et al*., 2013). The most studied systems of this family of transporters are AcrAB-TolC and MexAB-OprM involved in transport of wide range of different substrates; they can be anionic, cationic, or neutral, and include detergents, antibiotics, and other toxins (Wong *et al*., 2014).

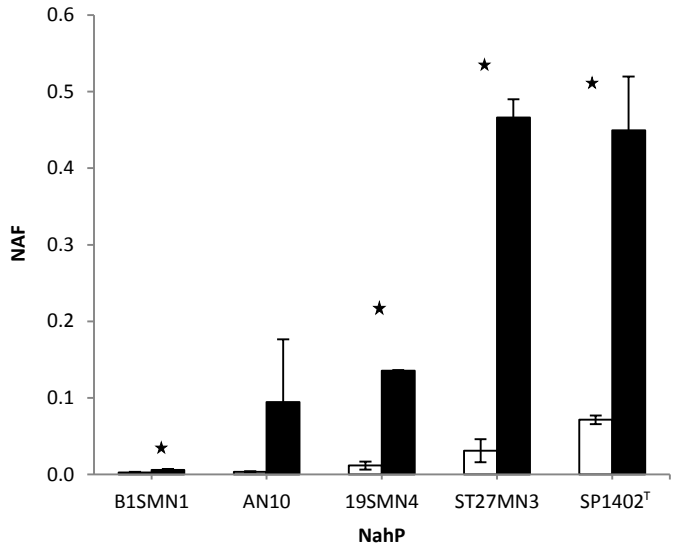

Figure 3.5. Average of NAF values for the detected NahP proteins in proteomes with succinate (white) or naphthalene (black) as carbon source. Error bars represent the standard deviation between the three replicates of each strain and culture condition. Strains with significant normalized abundance variations are represented with a star.

Table 3.4. Percentages of identity and E-values obtained with BLASTp of MexAB-OprM proteins from *P. stutzeri* and *P. balearica* and the toluene tolerance system TtgABC from *P. putida* DOT-T1E (Ramos *et al*., 1998).

	TtgA		TtgB		TtgC	
	% Id	E-value	% Id	E-value	% Id	E-value
B1SMN1	67.4	3E-139	75.6	0	65.2	3E-172
AN10	66.6	9E-143	75.6	0	65.4	1E-172
19SMN4[a]	66.2	2E-140	77.9	0	64.9	1E-172
19SMN4[b]	58.6	4E-119	63.9	0	61.7	2E-165
ST27MN3	66.2	2E-140	77.9	0	64.7	4E-172
ST27MN3	58.6	4E-119	63.9	0	61.7	2E-165
SP1402^T	60.9	5E-124	65.6	0	60.8	1E-158

a: chromosome localization.

b: plasmid localization.

A 73 % of homology between MexAB-OprM system from *P. aeruginosa* and the toluene tolerance system TtgABC from *P. putida* DOT-T1E (a strain with high tolerance to organic solvents such as toluene) has been previously described by Ramos and co-workers (1998). TtgABC was involved in removing the solvent from bacterial cell membranes, enhancing toluene tolerance (Ramos *et al*., 1998). The amino acid sequences of the MexAB-OprM system of the *P. stutzeri* strains were compared with the toluene tolerance system TtgABC from strain DOT-T1E. As shown in Table 3.4, E-values lower than 1E-100 and identities around 60–80 % were obtained in all cases. By proteomics we were able to detect the three proteins of MexAB-OprM system in all strains (see Figure 3.6). In the case of strains with two copies of these systems, only one of them was detected. Particularly, in 19SMN4 all the proteins encoded by the chromosomal copy were detected by proteomics, but only OprM protein encoded by the plasmidic copy (see as

OprMb in Supplementary Table S13) was detected. In B1SMN1 and ST27MN3 a statistically-significant increase of OprM normalized abundance was detected in naphthalene cultures (see Figure 3.6). However, the fold changes in abundance of the other MexAB-OprM proteins were not statistically significant. In *P. putida* DOT-T1E, the TtgABC transport system involved in the extrusion of aromatic solvents is constitutively expressed (Duque *et al*., 2001). Since MexAB-OprM is homologous to the TtgABC transport system and its proteins generally did not show statistically significant abundance fold changes, we suggest that its expression might be also constitutive. The detection of all these proteins in all the studied strain, together with the homology with the TtgABC system of strain DOT-T1E, suggest that *P. stutzeri* MexAB-OprM pump might have a plausible role in naphthalene metabolism.

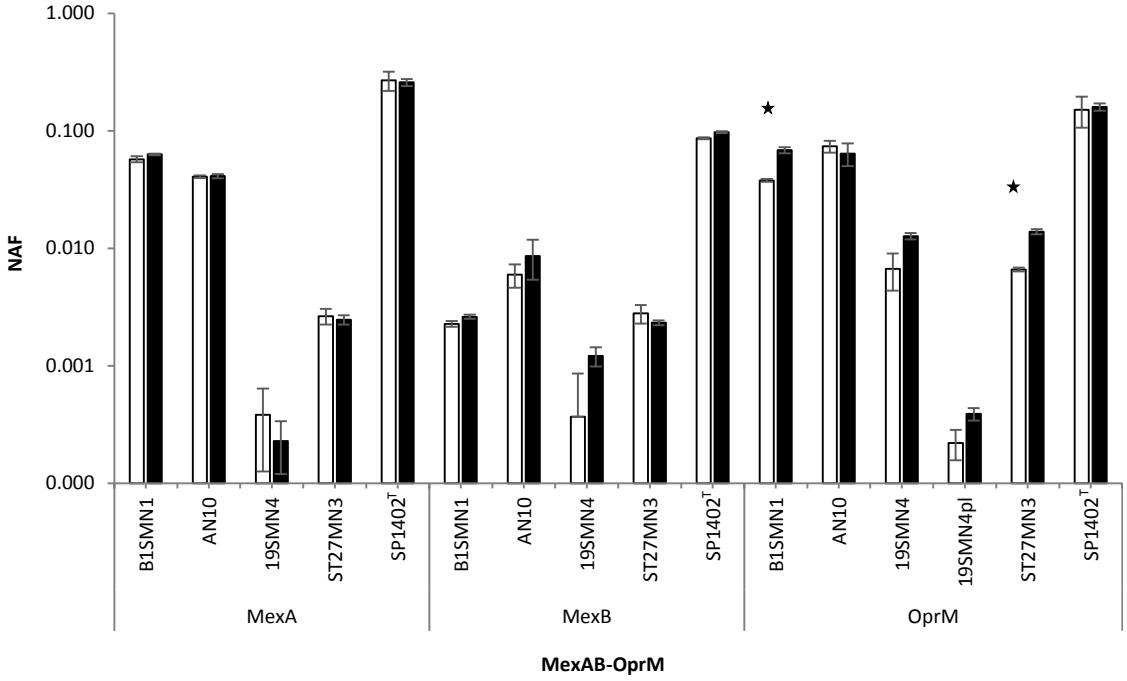

Figure 3.6. NAF average of the detected proteins from the transport system MexAB-OprM for each strain in proteomes with succinate (white) or naphthalene (black) as carbon source. Values are shown on a logarithmic scale. Error bars represent the standard deviation between the three replicates of each culture condition. Strains with significant normalized abundance variations are represented with a star.

Apart from the MexAB-OprM system, we identified a second RND transport system, which was named in this study as SistT5. This system consists of 3 different proteins: SistT5a, SistT5b, and SistT5c. All of them were identified in all analyzed genomes except from SistT5c, which was not present in the AN10 annotation (see Supplementary Table S13). Protein SistT5c was detected by proteomics in B1SMN1, 19SMN4, and ST27MN3, and protein SistT5a was detected in 19SMN4. In all cases, these proteins had a statistically significant higher normalized abundance in naphthalene cultures compared to succinate cultures (see Supplementary Table S13). Proteins SistT5a and SistT5c presented high homology (more than a 70 % of identity and E-values lower than 1E-143) with proteins MexF and MexE respectively, belonging to the MexEF-OprN transport system from *P. aeruginosa* PAO1, described by Maseda and co-workers (2002) as a xenobiotic efflux transporter. Therefore, it can be suggested that this transport system might have a role in naphthalene metabolism.

Thirdly, a TonB transport-system, named as SistT7 in this study, was identified in all studied genomes. It is constituted by 4 different proteins: SistT7a, SistT7b, SistT7c, and SistT7d. All of them showed E-values lower than 1E-20 with ExbBD-TonB-type proteins from the TCDB database. TonB transport systems were firstly identified in *E. coli* and they are involved in the transport of voluminous nutrients such as vitamin B12, siderophores, nickel complexes, and carbohydrates (Noinaj *et al*., 2010). These transporters span the cytoplasmic membrane and the periplasm, and interact with outer membrane receptors in gram-negative bacteria (Ahmer *et al*., 1995). The stoichiometry, structure, and mechanism of action of the TonB transport system remains unknown (Sverzhinsky *et al.,* 2015). This system has also been described in *P. putida* DOT-T1E and proposed to be involved in its tolerance to organic compounds (Godoy *et al*., 2001). In our study we detected proteins SistT7a and SistT7c in all strains, as well as SistT7d in strains ST27MN3 and SP1402^{T}. SistT7b was not detected in the proteomes of any of the studied strains. Moreover, a statistically-significant increase in the abundance of SistT7c was obtained in proteomes from naphthalene cultures of strains AN10, 19SMN4, ST27MN3, and SP1402^{T} (see Figure 3.7). Thus, it can be also suggested that the ExbBD-TonB transport system detected in *P. stutzeri* is involved in naphthalene tolerance.

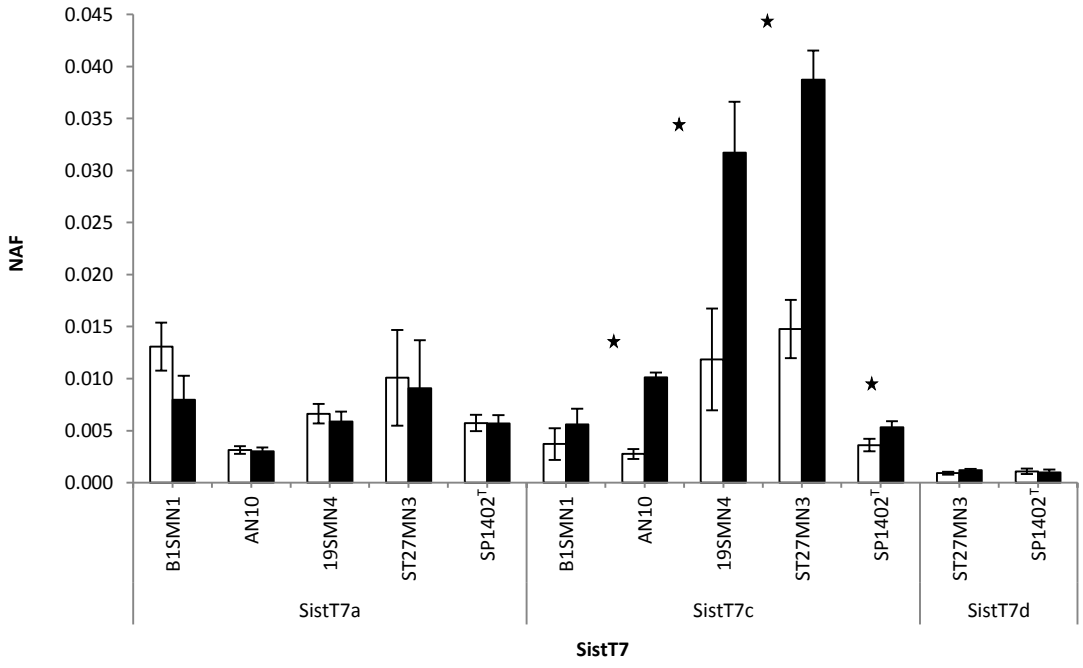

Figure 3.7. NAF average of the detected proteins from the ExbBD-TonB-type transport system SistT7 for each strain in proteomes with succinate (white) or naphthalene (black) as carbon source. Error bars represent the standard deviation between the three replicates of each culture condition. Strains with significant normalized abundance variations are represented with a star.

Finally, a TsaT-like transporter, named as SistT11, in this study was identified in the B1SMN1 genome annotation. The gene for this transporter is located 2,348-bp upstream *benR* (from the benzoate degradation operon). SistT11 is constituted by a single protein, SistT11a. The first described member of TsaT family of transport systems was a transporter for 4-toluenesulfonate from *Comamonas testosteroni* T-2 (Mampel *et al*., 2004). In this bacterium, protein TsaT forms an anion-selective outer membrane channel. Acting in co-operation with the cytoplasmic membrane protein TsaS, is capable of transporting 4-toluenesulfonate outside the cell (Mampel

et al., 2004). The protein SistT11a of strain B1SMN1 presented an identity of 25 % with an E-value of 1E-19 with the TsaT sequence from *C. testosteroni* T-2. Any protein homologous to TsaS was found in the studied *P. stutzeri* and *P. balearica* genomes. By proteomics we detected a normalized abundance fold change of 20.83 (p-value < 0.05) for the B1SMN1 TsaT-like protein SistT11a in proteomes from naphthalene cultures compared to succinate cultures. This suggests a plausible role of this protein in pumping naphthalene or any of its intermediates outside the cell.

There were other transport systems that were identified in the analysis that showed statistically-significant increases of normalized abundance in naphthalene cultures compared to succinate. These transport systems were not homologous to well-characterized aromatic-hydrocarbon transport-systems. All these proteins are listed in Supplementary Table S13. Five different ABC transport systems were identified, which were named SistT1, SistT3, SistT4, SistT9, and SistT12. ABC transport systems are widely distributed and they are involved in the transport of different substrates through the membrane powered by the direct hydrolysis of ATP (Davidson *et al.,* 2008). As shown in Supplementary Table S13, an abundance increase of proteins belonging to the transport system SistT1 in naphthalene proteomes was observed in all the studied strains. These proteins have been annotated as amino acid transporters, however the fold change observed suggests that they might play a role in naphthalene metabolism. Michalska and co-workers (2012) described that ABC transporters from several species of soil bacteria previously annotated as amino acid transporters were able to bind aromatic compounds, such as benzoate, 3-hydroxybenzoate, 4-hydroxybenzoate, and salicylate. This suggests that the putative role of SistT1 in naphthalene metabolism could be either as an amino acid or aromatic hydrocarbon transporter. We also identified two different TRAP-T transport systems, which were named SistT6 and SistT10. TRAP-T transport systems are secondary active transporters that use the proton motive force (Forward *et al*., 1997) and work in conjunction with an extracytoplasmatic solute-binding receptor (Rabus *et al*., 1999). An SSS transport system was also identified in all studied genomes, which was named SistT2. The SSS transport systems are widely distributed in prokaryotes and eukaryotes, and they transport a variety of solutes (sugars, amino acids, vitamins, ions, phenylacetate, urea, and water) through the inner membrane using a sodium motive source (Jung, 2002). Finally, a MFS transport system was identified in all studied genomes, which was named SistT8. MFS transport systems are the most abundant and diverse systems. The MFS systems can specifically transport a wide range of substrates: sugars, ions, nucleosides, and amino acids (Law *et al*., 2008). In this case, a significant abundance increase of protein SistT8b was detected in naphthalene cultures of strains 19SMN4 and ST27MN3 only. Taking together all these results we can then hypothesize that these transport proteins might also have a role in naphthalene metabolism, acting as aromatic hydrocarbon pumps, as mentioned for the previously discussed ones.

As we have described for *P. stutzeri*, increases in abundance of transport proteins in *Pseudomonas* cultures with aromatic hydrocarbons have been also observed in previous proteomic studies. As previously mentioned, the toluene tolerance of *P. putida* DOT-T1E and *P. putida* KT2440 has been extensively studied, in particular the role of the transport system TtgABC (Ramos *et al*., 1998). Transcription analysis of KT2440 cells grown on glucose in the presence and absence of toluene revealed transport systems specifically induced in response to toluene: a TtgABC transport system (PP1384, PP1385, PP1386) homologous to the MexAB-OprM

system identified in *P. stutzeri* and *P. balearica* genomes; a protein from an ABC transport system (PP0958) and two MFS-type transporters, a Bcr/CflA protein (PP3588) and PP3349 (Segura *et al*., 2005; Domínguez-Cuevas *et al*., 2006) with no homologous in *P. stutzeri* genomes. The transcriptional induction of the last three proteins was found to be much higher than the *ttgABC* induction levels (Domínguez-Cuevas *et al*., 2006). Conversely, the TtgABC transport system is constitutively expressed in *P. putida* DOT-T1E (Duque *et al*., 2001).

Another study was performed by Wijte and co-workers (2010), who analyzed proteomic changes of *P. putida* S12 growing with and without phenol (3 and 5 mM). An increase of transport proteins SrpABC (RND-like) and PP1272 (MFS-like) was observed in 3 mM phenol cultures, and even further increases were observed in 5 mM phenol cultures. The most prominent increases in abundance were for proteins of the SrpABC efflux system (fold change higher than 15), which was directly linked to the solvent tolerance of strain S12. The SrpA sequence from *P. putida* S12 showed high homology with the MexA proteins identified in *P. stutzeri* and *P. balearica* genomes (identities above 55 % and E-value below 1E-114), which were also detected in the proteomes analyzed in this chapter (see Figure 3.6). In contrast, the study of Roma-Rodrigues and co-workers (2010) in *P. putida* KT2440 cultures growing with phenol revealed an increase in the abundance of the periplasmic and outer membrane subunits from the TtgABC efflux pump in cultures with phenol (fold change up to 35). A third study performed by Hemamalini and Khare (2014) with *Pseudomonas aeruginosa* PseA described the up-regulation of the outer membrane porin OprE in cultures with different alkanes such as cyclohexane, n-heptane, and n-dodecane. This protein showed a 46 % of identity and an E-value of 6E-103 with protein YP_006458185.1 from AN10 (see Supplementary Table S2), which showed a statistically significant abundance increase (fold change of 2.4 with a p-value of 7E-4) in naphthalene compared to succinate proteomes (see Supplementary Table S12). Homologous proteins were detected in the proteomes of strains B1SMSN1, 19SMN4, and ST27MN3, with no statistically significant abundance changes.

It has been demonstrated that exposure to aromatic hydrocarbons severely increases the fluidity of the outer membrane, as the divalent cations that cross-bridge adjacent lipopolysaccharide molecules are displaced (Wijte *et al.,* 2010). As a result, membrane permeability also increases and therefore, the intrusion of aromatic hydrocarbons into the cell increases. In order to stabilize the outer membrane, *P. putida* strain S12 increase the expression of outer membrane proteins, as a response to the increase in membrane permeability (Wijte *et al.,* 2010; Ramos *et al*., 2015). Moreover, the entry of organic solvents across the outer membrane into the cell results in the leakage of protons across the outer membrane and therefore in a decrease in the proton motive force across the inner membrane. The change in membrane protein abundance in presence of organic solvents is likely to be an attempt of the cell to restore this proton imbalance (Hemamalini and Khare, 2014). In addition, the induction of specific efflux pumps that extrude toxic solvents to the medium is also essential in the exposure to toxic solvents such as aromatic hydrocarbons, as reviewed by Ramos and co-workers (2015). In particular, Rojas and co-workers (2001) described the important role of three efflux systems (TtgABC, TtgDEF, and TtgGHI) in the toluene tolerance of *P. putida* DOT-T1E. Therefore, we believe that the increase in the abundance of the transport proteins detailed in this section in *P. stutzeri* naphthalene proteomes might be due to the need of the cells to import naphthalene, export naphthalene derivatives toxic for the cell (such as salicylate), or transport

other molecules required in naphthalene metabolism. These possibilities should be tested experimentally.

4. Comparison of proteomes from salicylate and succinate cultures

From the comparison between proteomes from naphthalene and succinate cultures, it was clear that the growth with naphthalene as carbon source had multiple effects on *P. stutzeri* and *P. balearica* proteomes. We analyzed also the effects in the proteome induced by salicylate (the inducer of the expression of the naphthalene degradation genes), comparing the proteomes of cultures exposed to a pulse of salicylate 3 mM with the proteomes in cultures with succinate (see section 4.2 from Materials and Methods for details).

Results showed a significant increase (fold changes above 2 and a p-value < 0.05) of the normalized abundance of most of the naphthalene degradation proteins in salicylate 3 mM pulse cultures (see Figure 3.8). However, this increase was lower (fold-increase up to 23) than the one observed in proteomes from naphthalene compared to succinate cultures (fold-increase up to 32, see Figure 3.3).

As expected, we detected the proteins involved in the transformation of naphthalene to salicylate (NahA, NahB, NahC, NahD, NahE, and NahF) in salicylate pulse cultures in all strains (except for NahB in B1SMN1). These proteins are not necessary for salicylate degradation but its expression is also induced by salicylate due to the presence of the P_{nah} promoter that interacts with the NahR protein (Bosch *et al*., 1999a; Bosch *et al*., 1999b).

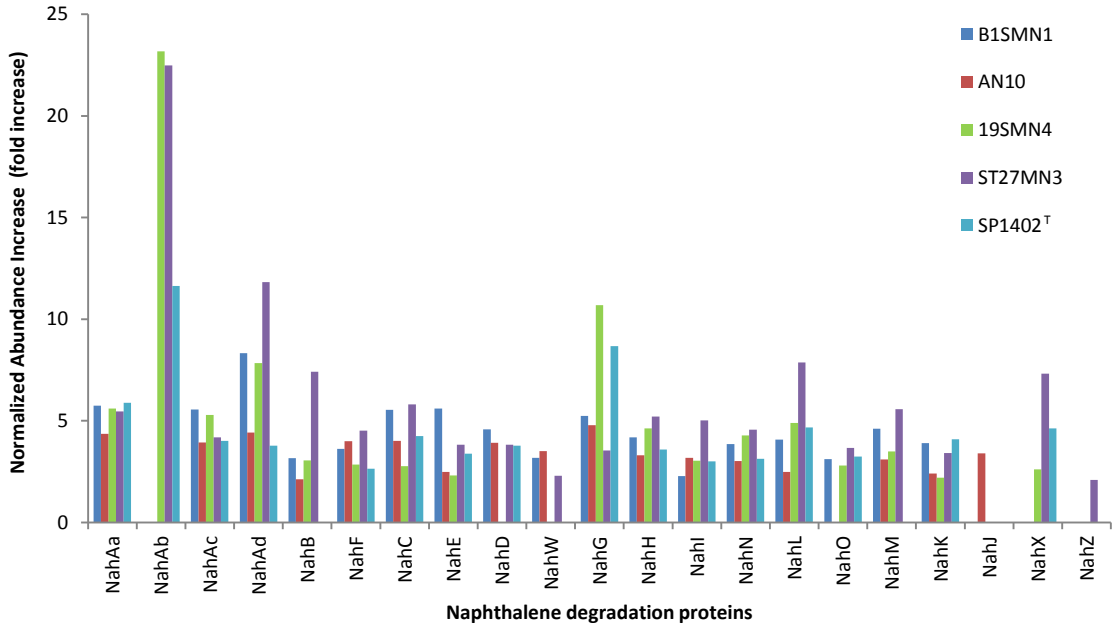

Figure 3.8. Statistically-significant increase in the abundance of naphthalene degradation proteins in proteomes from salicylate 3 mM pulse cultures compared to succinate cultures. NahT is not shown because it was not detected in any strain. NahR is not shown because its normalized abundance did not vary significantly although it was detected in all strains.

In contrast to what was observed in naphthalene cultures, a statistically-significant increase in the abundance of proteins involved in benzoate degradation was observed in salicylate pulse cultures of strain ST27MN3 (see Figure 3.9). In this proteome all proteins involved in benzoate degradation were detected except for: a muconolactone D-isomerase (CatC) and two putative benzoate transporters (BenK and BenE). Strains B1SMN1, 19SMN4, AN10, and SP1402^{T} also presented in their genomes the genes involved in benzoate degradation (see Figure 2.6 from Chapter 2). However, only the catechol1,2-dioxygenase (CatA) from B1SMN1 (EPL60843.1; see Supplementary Table S12) was detected with a significant increase in abundance in naphthalene and salicylate pulse cultures respect to succinate cultures.

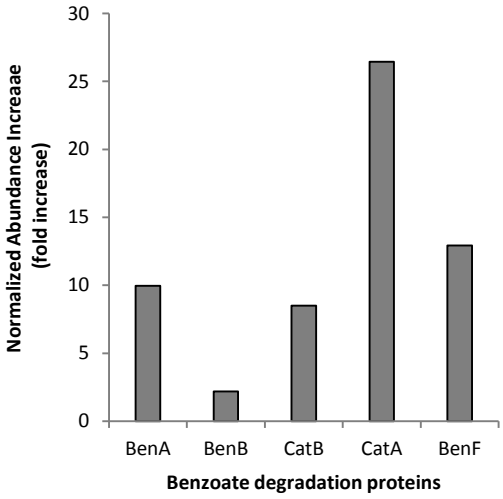

Figure 3.9. Statistically-significant abundance increase of ST27MN3 benzoate degradation proteins from salicylate pulse compared to succinate cultures. Significant increase in the abundance of CatA was also observed in the proteome of strain B1SMN1.

Benzoate and salicylate are similar molecules that can be degraded by *P. putida* mt-2 and *P. putida* NAH7, respectively, through very similar catabolic pathways. However, the activator of the benzoate degradation operon is the AraC-like factor BenR (homologous to XylS), while for the salicylate operon is the LysR-type protein NahR (Cases and de Lorenzo, 2005). Ramos and co-workers (1986) analyzed the capability of XylS to induce the expression of toluate 1,2-dioxygenase gene from TOL plasmid in *P. putida* KT2440, whose expression is normally regulated by XylS in response to benzoate. Results showed that the wild type transcriptional regulator XylS, which usually presents specificity for benzoate derivatives (such as 2-methylbenzoate 2,3-dimethylbenzoate, 3,4-dimethylbenzoate, and 3,4-dichlorobenzoate), did not respond to the presence of salicylate (2-hydroxybenzoate). However, they isolated spontaneous *xylS* mutants which exhibited high specificity to benzoate derivatives, particularly salicylate. For all of this, we postulated that the *benR* transcriptional regulator of ST27MN3 might present a mutation that allows its induction by salicylate, due to cross-induction. A second explanation for the salicylate induction of benzoate degradation proteins would be that catechol resulting from salicylate degradation induced the synthesis of these catabolic enzymes. In fact, Feist and Hegeman (1969) previously described the induction of catechol 1,2-dioxygenase CatA and the muconate cycloisomerase CatB by catechol in *P. putida* U. However, in this case we would expect similar inductions in all the studied strains, since they all presented the benzoate degradation operon.

In order to study the possibility that the *benR* gene of ST27MN3 presented a mutation that allowed its induction by salicylate, we compared the BenR amino acid sequence of ST27MN3 with the BenR sequences of the other studied strains. As expected, the most similar sequence to ST27MN3 was BenR from 19SMN4, which differed only in position 235, where ST27MN3 presented glutamic acid and 19SMN4 aspartic acid. The BenR sequence from strain B1SMN1 (whose CatA also increased significantly its abundance in salicylate cultures) also presented glutamic acid in position 235, as ST27MN3. Like most AraC family members, XylS appears to consist of two domains: the N-terminal domain, involved in effector recognition and dimerization; and the C-terminal domain, responsible for DNA binding (Ruiz and Ramos, 2002; Ruíz *et al*., 2003; Domínguez-Cuevas *et al.,* 2008). Several mutations in the XylS gene from plasmid TOL (homologous to BenR) which affect its activity have been previously described. Six different XylS positions (41, 45, 137, 153, 193, and 194) have been described as crucial for effector recognition and dimerization of XylS (Ruiz and Ramos, 2002; Ruíz *et al*., 2003). Domínguez-Cuevas and co-workers (2008) described eight positions in XylS from plasmid TOL responsible for DNA binding, five of them (242, 243, 246, 249, and 250) bind the DNA sequence TGCA, and the other three (296, 299, and 302) bind the sequence GGNTA. The only difference between the BenR sequence from ST27MN3 and 19SMN4 (position 235) is located at 7 amino acids from the TGCA binding site (from 242 to 250) but far from the domain involved in effector recognition. Therefore, no differences have been found to justify the expression of benzoate degradation proteins in ST27MN3 salicylate pulse cultures.

Thinking on the possibility that the *benR* transcriptional regulator of ST27MN3 might be induced by salicylate due to cross-induction, we compared the DNA sequences of the BenR binding site for *benA* described in *P. stutzeri* A1501 by Li and co-workers (2010) with the genomes of the different studied strains. The BenR binding site in A1501 is composed by two TGCA sequences and two GGATA sequences that alternate in the sequence (see Supplementary Figure S3), and strains B1SMN1, AN10, 19SMN4 and ST27MN3 presented the same structure than A1501.

As discussed above, 14 transport systems were over expressed in proteomes from naphthalene cultures compared to succinate cultures. The normalized abundance of these proteins was also analyzed in salicylate pulse cultures. Only 8 of the 72 different proteins over expressed in naphthalene cultures were significantly over-expressed in salicylate pulse cultures: SistT1b and SistT1c from 19SMN4; SistT3c from 19SMN4 and ST27MN3; NahP from ST27MN3 and SP1402^{T} ; MexA from 19SMN4 and ST27MN3, and OprM from ST27MN3; SistT5c from ST27MN3; and SistT10d from 19SMN4, ST27MN3, and B1SMN1. These results suggest that either the mentioned 14 transport systems are not inducible by salicylate or the 4 hours salicylate 3 mM pulse was not sufficient to induce them significantly.

CHAPTER 4: Benzoate degradation potential of *Pseudomonas stutzeri* AN10

1. *P. stutzeri* AN10 potential to metabolize benzoate

As mentioned in Chapter 2 the complete set of genes involved in benzoate degradation was identified in *P. stutzeri* AN10 genome, although the gene for the transcriptional regulator *benR* was truncated by a stop codon (see Figure 2.6). In addition, this strain was not able to grow with benzoate as sole carbon and energy source (see Table 2.11), as previously described by Rosselló-Mora and co-workers (1994). We were interested in selecting an AN10 mutant able to express the *ben* genes and allowing its growth with benzoate as unique carbon and energy source. To do so, we performed serial cultures of AN10 at increasing benzoate concentrations and decreasing succinate concentrations (see section 3 of Materials and Methods for details). The growth of AN10 was not much affected by increasing benzoate concentrations (up to 5 mM) if succinate was also provided, since the maximum absorbance at 600 nm remained between 1.2 to 1.3 and the generation time from 9.6 to 13.5 hours (see Table 4.1). Nevertheless, a reduction in the growth of AN10 was observed when succinate concentration decreased. Finally, no growth was observed in cultures with benzoate as a sole carbon and energy source (see Table 4.1). Although this experiment did not allow us to obtain an AN10 derivative able to use benzoate as sole carbon source, it showed that benzoate was not toxic for AN10 strain, at least at concentrations up to 5 mM.

Table 4.1. Generation time and maximum absorbance (600 nm) of AN10 serial cultures with increasing benzoate concentrations (from 0 to 5 mM) and later decreasing succinate concentrations (from succinate 30 mM to 0 mM).

Carbon source	Generation time (h)	A_{600}max
Succinate 30 mM	11.65	1.33
Succinate 30 mM and Benzoate 1 mM	9.60	1.30
Succinate 30 mM and Benzoate 2 mM	13.46	1.22
Succinate 30 mM and Benzoate 3 mM	11.46	1.34
Succinate 30 mM and Benzoate 4 mM	12.14	1.21
Succinate 30 mM and Benzoate 5 mM	12.98	1.18
Succinate 22.5 mM and Benzoate 5 mM	23.10	0.91
Succinate 15 mM and Benzoate 5 mM	24.93	0.73
Succinate 7.5 mM and Benzoate 5 mM	41.26	0.37
Succinate 3.5 mM and Benzoate 5 mM	106.64	0.19
Benzoate 5 mM	no growth	no growth

In order to isolate an AN10 benzoate-degrading derivative, we used a strong inoculum of AN10 (10^8 CFU/ml) from a culture in LB broth to inoculate MMB medium with benzoate 3 mM as carbon source. Growth was observed in this culture after 40 days of incubation. At this point, a sequence of sub-cultures at increasing benzoate concentrations (from 4 mM up to 20 mM), were performed (see section 3 from Materials and Methods for details). As shown in Figure 4.1, while we increased benzoate concentration of the sub-cultures from 4 mM to 10 mM, the maximum absorbances increased from 0.43 to 1. In sub-cultures with 11 to 20 mM of benzoate, the maximum absorbance at 600 nm remained similar. This result suggested that benzoate concentration was a limiting growth factor up to a benzoate concentration between 10 and 11 mM.

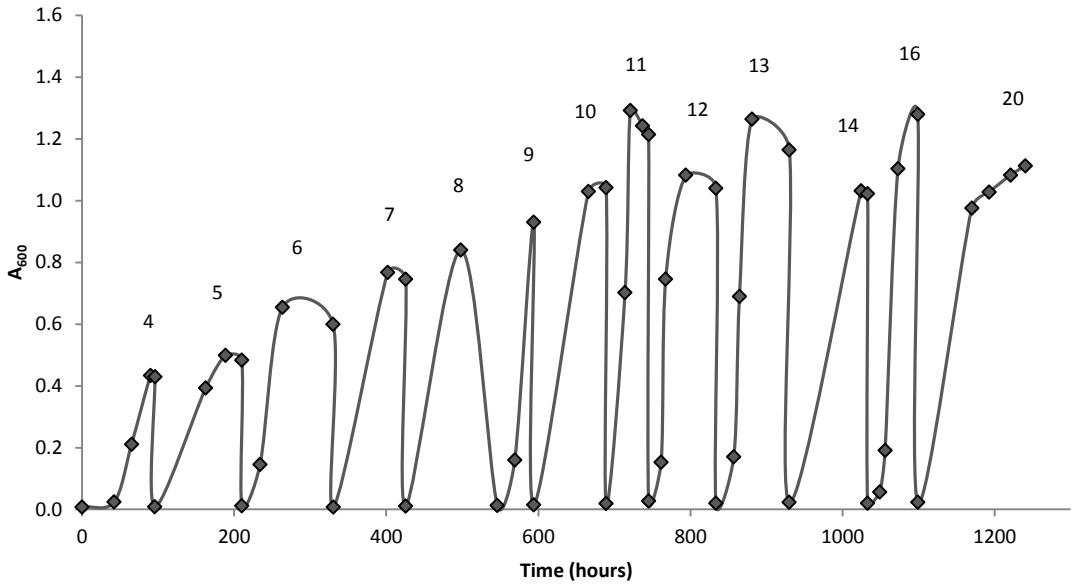

Figure 4.1. Absorbance at 600 nm obtained for each of the successive sub-cultures of strain AN10. Benzoate concentration (mM) used for each sub-culture is shown over the maximum absorbance reached.

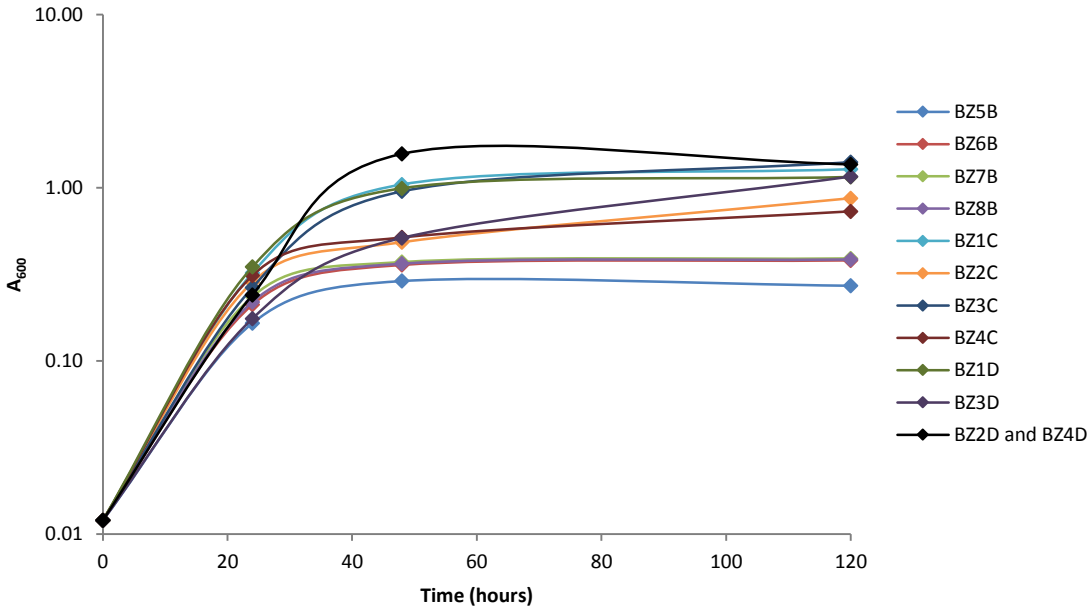

Figure 4.2. Growth of twelve AN10 derivatives in MMB with benzoate 20 mM as carbon source. Derivatives BZ5B-BZ8B, BZ1C-BZ4C, and BZ1D-BZ4D were isolated from sub-cultures with 4, 16 and 20 mM benzoate, respectively.

For each of the sub-cultures with increasing benzoate concentration, 4 isolates were randomly selected and stored at –80 °C (see section 3 in Material and Methods for details). This approach allowed us to obtain AN10 derivatives able to grow on benzoate isolated from sub-cultures at different benzoate concentrations. We checked that the selected isolates were AN10 derivatives by amplification and sequencing their 16S rRNA and *rpoD* genes. All sequences obtained were 100 % identical with AN10 wild type genes. In order to select the AN10 derivative which presented the best growth with benzoate, we grew the four derivatives isolated from three

different sub-cultures (4 mM, 16 mM, and 20 mM benzoate concentrations) in MMB supplemented with different benzoate concentrations (5 mM, 10 mM, 15 mM, 20 mM, and 25 mM). Derivatives BZ4D and BZ2D, both isolated from the sub-culture with benzoate 20 mM, showed the best growth in MMB supplemented with benzoate 20 mM (see Figure 4.2).

Of all the isolated AN10 derivatives, we selected BZ4D for further studies. Its growth and the growth of the parental AN10 strain in MMB supplemented with different carbon sources was evaluated. As shown in Table 4.2, both strains presented similar generation times and maximum absorbances growing with succinate. However, in cultures with naphthalene as carbon source, BZ4D presented higher generation time and lower maximum absorbances than AN10. Surprisingly, BZ4D was unable to grow with salicylate as carbon source and, as expected, AN10 was unable to grow with benzoate as carbon source (see Table 4.2). The problems of BZ4D to grow on naphthalene, and mainly salicylate, suggested that this derivative suffered some change during the selection experiment that prevented its growth with those carbon sources. This will be discussed in the section below.

Table 4.2. Generation time and maximum absorbance (600 nm) of BZ4D and AN10 with different carbon sources. Average values and standard deviations of two replicate cultures are shown.

Carbon source	Strain	Generation time (h)	A_{600}max
Succinate 30 mM	AN10	9.62 ± 0.14	1.25 ± 0.01
Succinate 30 mM	BZ4D	8.72 ± 0.07	1.34 ± 0
Salicylate 3 mM	AN10	38.44 ± 3.30	0.37 ± 0.01
Salicylate 3 mM	BZ4D	no growth	no growth
Benzoate 3 mM	AN10	no growth	no growth
Benzoate 3 mM	BZ4D	43.58 ± 3.28	0.40 ± 0.02
Naphthalene [a]	AN10	11.69 ± 0.64	1.45 ± 0.06
Naphthalene [a]	BZ4D	17.70 ± 0.03	0.82 ± 0.03

a: see Materials and methods section 2 for details.

2. Genome sequencing of *P. stutzeri* BZ4D

The genome of the AN10 derivative BZ4D was sequenced using the Ilumina technology, obtaining 6.86 million of reads. Forty-eight contigs with 4,678,705-bp in total length were obtained. No structural rearrangements were detected in the BZ4D genome in comparison with the AN10 genome. However, 73 point mutations were observed between the two genomes: 44 silent mutations (no change in amino acid) and 29 missense mutations (causing changes in amino acids). Additionally, an insertion of 7 base-pairs was identified in BZ4D at position 1,979,521 of the AN10 genome; as well as a 3 base-pairs deletion in BZ4D at position 3,054,628 of the AN10 genome. These 31 missense mutations caused changes in the amino acidic sequence of 17 different CDSs (see Table 4.3). It is remarkable that 9 of the mutations were in a gene coding for a RarD permease (YP_006458742.1, see Supplementary Table S2). No role of this protein in aromatic degradation has been described yet. It was not detected in any of the *P. stutzeri* proteomes shown in chapter 3 or in the proteomic study described in next section.

Of all these 17 proteins, we focused on BenR and CatA due to their role in benzoate degradation: BenR as the transcriptional regulator of operon *benABCD;* and CatA as the enzyme responsible of the *ortho*-cleavage of catechol. Additionally, we focused on NahR, the transcriptional regulator of naphthalene degradation operons, due to the inability of BZ4D for growing using salicylate as carbon source (see Table 4.2).

Table 4.3. Differences observed between AN10 and BZ4D genomes. Silent mutations are not shown.

AN10 position	AN10 base	BZ4D base	Amino acid substitution	Protein id AN10	Protein id BZ4D	NCBI Annotation
862,699	C	T	P38L	YP_006456455.1	BZ4D_01439	Transcriptional regulator MalT
862,705	C	A	P40Q			
1,442,356	T	G	I264S	YP_006456979.1	BZ4D_04377	Transcriptional regulator NahR
1,733,220	T	C	Q439H	YP_006457273.1	-	Transposase
1,733,226	G	C	T437M			
1,733,227	G	A				
1,971,208	T	G	stop171E	-	BZ4D_02777	Transcriptional regulator BenR
1,979,521	-	GCCGCGG	Frameshift	YP_006457505.1	BZ4D_02769	Catechol 1,2-dioxygenase CatA
2,121,507	C	A	G422V	YP_006457628.1	BZ4D_02402	Xylulokinase
2,548,946	G	C	V249L	YP_006457982.1	BZ4D_01687	*cbb*$_3$-type cytochrome *c* oxidase
2,552,268	G	A	A216T	YP_006457985.1	BZ4D_01684	*cbb*$_3$-type cytochrome *c* oxidase
2,552,270	A	C				
2,552,285	C	G	I222M			
2,996,465	G	C	P257A	YP_006458395.1	BZ4D_00823	Flagellar protein FliH
3,054,628	TCT	-	EN328D	YP_006458455.1	BZ4D_00763	Na^+ translocating NADH-Quinone reductase
3,244,907	C	A	R34S	YP_006458641.1	BZ4D_00572	LysE transporter
3,346,505	C	G	G71R	YP_006458742.1	BZ4D_00470	RarD permease
3,346,507	C	A	R70L			
3,346,515	G	C	R67L			
3,346,516	C	A				
3,346,518	C	G				
3,346,519	G	A	A66I			
3,346,520	C	T				
3,346,522	C	T	G65E			
3,346,531	G	T	A62E			
3,758,287	G	T	P280T	YP_006459140.1	BZ4D_00068	HppD
3,785,909	G	A	R234Q	YP_006459166.1	BZ4D_00040	Exonuclease
3,785,912	C	T	A235V			
3,813,636	C	A	G213W	YP_006459197.1	BZ4D_00008	Chromosome segregation ATPase
4,289,512	C	T	P43S	YP_006459642.1	BZ4D_03237	Hypothetical protein
4,583,378	G	T	G208V	YP_006459879.1	BZ4D_03850	Transaminase

The transcriptional regulator protein of naphthalene degradation, NahR, presented a serine (amino acid with a polar chain) instead of an isoleucine (non-polar chain) in position 264 in BZ4D. This is due to a single base-pair mutation in position 1,442,356 of the AN10 genome. This mutation is localized into the C-terminal substrate binding domain of LysR-type transcriptional regulators (domain cd08459) involved in specific inducer binding (Park *et al.,* 2005). Therefore, this mutation might prevent the growth of BZ4D with salicylate as carbon source.

As shown in Table 4.3, the substitution of the thymine at position 1,971,208 for a guanine in the BZ4D genome caused the substitution of a stop codon for a glutamate codon in gene *benR* (codon 171). Due to this variation, the translation of protein BenR, previously truncated in AN10, would be possible in BZ4D. As expected, this protein sequence in BZ4D was annotated by PROKKA as a transcriptional regulator protein. Moreover, it presented 93.4 % of identity with

the previously described *benABCD* operon transcriptional regulator BenR from A1501 (Li *et al*., 2010). We compared by BLASTp the BZ4D BenR sequence with the fourteen *P. stutzeri* and *P. balearica* BenR CDSs identified in Chapter 2 (see Figure 2.6, Chapter 2). The most similar sequences were BenR from ST27MN3 and NF13, which presented an identity of 99.1 %. If the BenR protein, which was truncated in AN10 genome, could be expressed in BZ4D, this would allow the transcriptional activation of *benABCD* genes in presence of the inductor benzoate. The BenR inducer binding sites proposed by Li and co-workers (2010) for A1501 were also found 103 base-pairs before *benA* gene in BZ4D and AN10 genomes (see Supplementary Figure S3). Consequently, BZ4D cells were able to grow using benzoate as sole carbon source.

In the case of *catA*, the catechol 1,2-dioxygenase gene, a 7 base-pair insertion was detected after codon 274 in BZ4D. As shown in Figure 4.3 A, the sequence GCCGCGG in AN10 genome in position 1,979,521 appeared to be directly repeated in the genome of BZ4D. As a result, there was a frame shift from amino acid position 278 to the end in BZ4D CatA (see Figure 4.3 B). According to an analysis of different biological systems, Levinson and Gutman (1987) revealed the presence of sequences exchange among slipped DNA strands, resulting in mutations that would be a significant source of evolutionary divergence. These acquired DNA sequence repetitions have been interpreted as components of three-dimensional structures formed by hybridization between slipped strands (Ornston *et al*., 1990). In particular, sequence repetitions of 6 to 10 base-pairs have been reported as frequent events involved in the evolutionary divergence of *catA* gene in *Acinetobacter calcoaceticus* BD413 and *P. putida* PRS2015 (Neidle *et al*., 1988). Therefore, we might hypothesize that the insertion of the sequence GCCGCGG in BZ4D genome might be a result of slipped DNA strands.

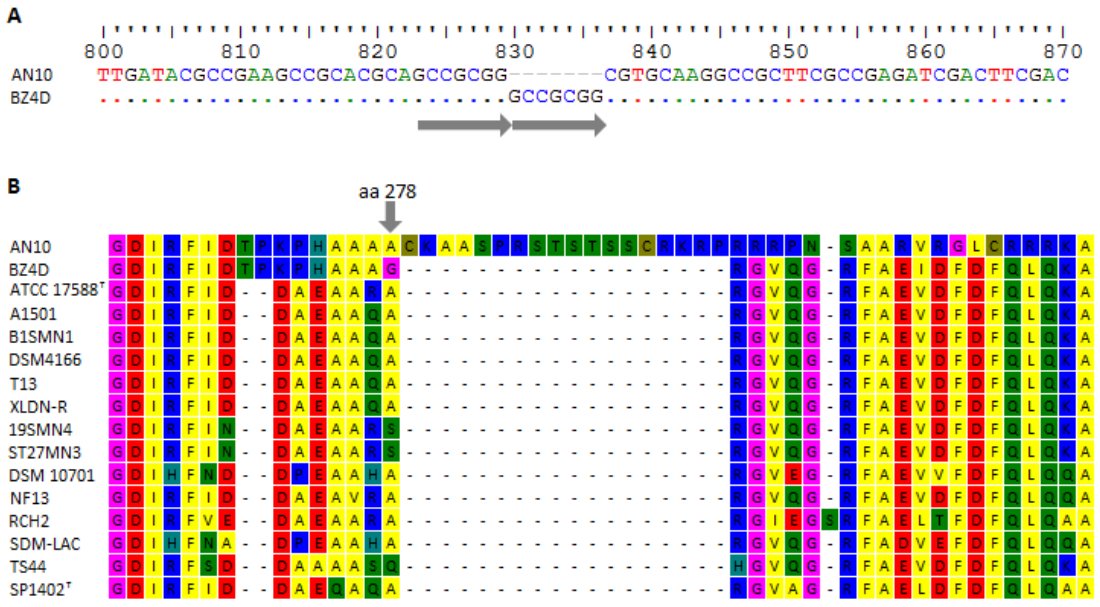

Figure 4.3. Fragment of the alignment of the *cat*A genes (A) and the carboxy-terminal CatA CDSs (B) from *P. stutzeri* and *P. balearica*. The sequence repetition in BZ4D catA *gene* is marked with arrows. The amino acid position where the CatA frameshift starts is also shown.

The CatA modified amino acid sequence is localized within the "1,2-CTD" domain (cd03460), associated to catechol 1,2-dioxygenase activity. Comparing both CatA carboxy-terminal ends (AN10 and BZ4D) with the other fourteen CatA sequences identified in *P. stutzeri* and *P.*

balearica (see Chapter 2), it is clear that the BZ4D CatA carboxy-terminal sequence is more similar to the other *P. stutzeri* sequences than the original AN10 sequence (see Figure 4.3 B). Moreover, we performed a BLASTp comparison of AN10 CatA against the non-redundant NCBI database, and no homologous sequences were found with the carboxy-terminal end sequence. This suggests that BZ4D CatA might be the functional one, instead of AN10 CatA.

3. Proteomic analysis of BZ4D

In order to define the proteins involved in the benzoate catabolic pathway of BZ4D, a proteomic experiment was performed. In this experiment the proteome of AN10 growing with succinate was compared with the proteome of the derivative strain BZ4D growing with succinate or with benzoate as sole carbon and energy sources. This approach allowed us to identify a total of 1,630 proteins (listed on Supplementary Table S14). A PCA of the Normalized Spectral Abundance Factor (NSAF) of the detected proteins showed that the proteomes of the bacteria in the 3 analyzed conditions were clearly different. PC1 explained 74.8 % of the variance, and separated clearly the proteomes from benzoate and succinate cultures (see Figure 4.4).

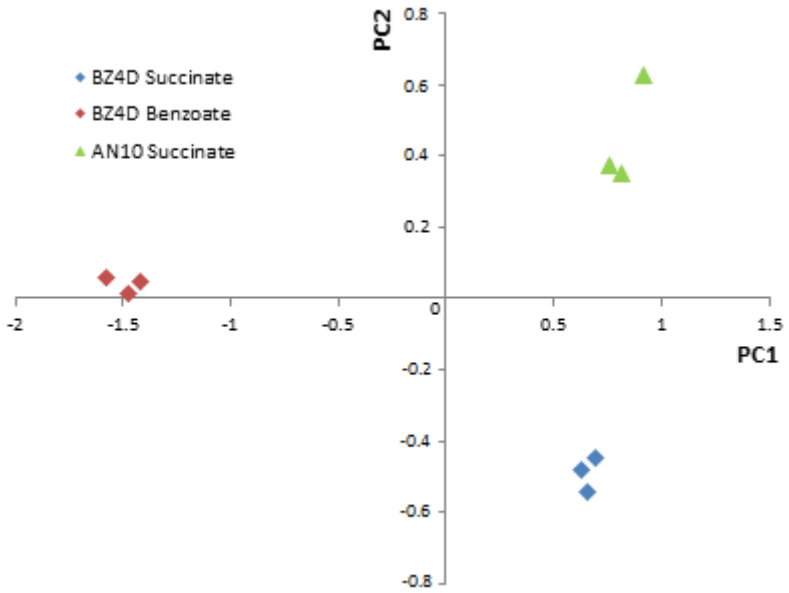

Figure 4.4. Principal component analysis of NSAF values obtained for BZ4D growing with succinate or benzoate, and AN10 growing with succinate. Percentages of variance explained: PC1 75 %; PC2 10 %.

There were 353 proteins which showed statistically-significant normalized abundance variations (fold changes below 2 and p-values < 0.05) between succinate and benzoate BZ4D proteomes. In particular, 216 of these proteins increased in abundance in proteomes from benzoate cultures and 137 of them decreased. In contrast, only 93 proteins presented significant normalized abundance variations between AN10 and BZ4D succinate cultures. Among those, in BZ4D proteomes we observed a statistically significant reduction in the abundance of 7 proteins involved in naphthalene metabolism: NahH, NahK, NahO, NahL, NahX, NahJ, and NahN. The results for NahH will be commented later (see below). The transcription of all these proteins is regulated in AN10 by NahR in the presence of salicylate (Bosch *et al*., 1999a; Bosch *et al*., 1999b; Bosch *et al*., 2000). Therefore, the reduction in the abundance of these proteins might be due

to the mutation observed in the *nahR* gene in BZ4D. This result might explain the different growth parameters observed for AN10 and BZ4D in naphthalene (see Table 4.2).

As described in Chapter 2, in *P. stutzeri* A1501 benzoate is catabolized to catechol by the successive action the products of operon *benABCD*. Afterwards, the aromatic ring of catechol is cleaved in *ortho* position by catechol 1,2-dioxygenase (CatA) and then transformed to β-ketoadipate by the successive action of CatB, CatC, and PcaD. This compound is later catabolized by the action of PcaIJF to TCA intermediates (see Figure 2.5, Chapter 2). The transcriptional regulators BenR and PcaR control the transcription of *ben* and *pca* operons, while transcription of the *catBC* promoter can be activated directly in response to benzoate in A1501 (Li *et al*., 2010, see Figure I6).

All proteins encoded by the operons *ben*, *cat*, and *pca*, presented significant increases in normalized abundance in BZ4D benzoate cultures compared to BZ4D succinate cultures (see Figure 4.5). This included the two transport proteins presumably involved in benzoate metabolism, BenF and BenK (fold changes of 28.9 and 56.6 respectively). It is also remarkable that 5 of these proteins (BenA and BenB, the two different benzoate 1,2-dioxygenase subunits; CatA, catechol 1,2-dioxygenase; CatB, muconate cycloisomerase; and PcaI, 3-oxoadipate CoA-transferase subunit) presented NSAF 50 times higher in benzoate cultures. Exceptions of that were the transcriptional regulator BenR and the putative transporter BenE, not detected by proteomics; as well as the transcriptional regulator PcaR, which did not show a statistically significant fold change in proteomes from BZ4D benzoate cultures compared to succinate cultures.

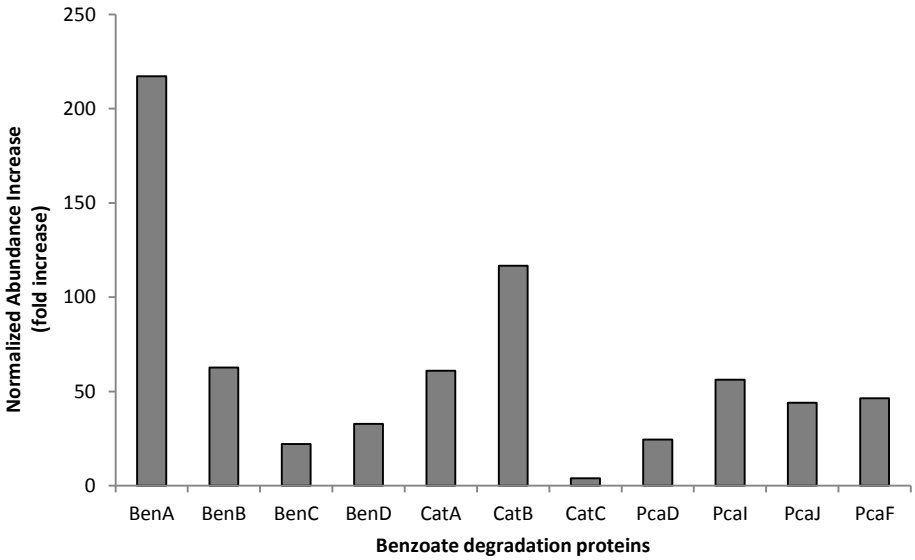

Figure 4.5. Statistically-significant increase in the abundance of benzoate degradation proteins in proteomes from BZ4D from benzoate and succinate cultures. PcaR is not shown, because no significant variation in abundance was observed between the different cultures. BenR is not shown, because it was not detected by proteomics.

As mentioned in Chapter 2, strain AN10 presents in its genome two different dioxygenases for the cleavage of catechol. This compound might be cleaved in *meta*-position by catechol 2,3-dioxygenase (NahH) as in the naphthalene degradation pathway, or in *ortho*-position by catechol 1,2-dioxygenase (CatA), provided that this enzyme is functional in AN10, which seems unlikely

considering the results presented above in section 2 and the results of Rosselló-Mora and co-workers (1994). These two dioxygenases are also present in BZ4D, and in this case we postulated that CatA would be functional, as discussed above. In order to determine which strategy strain BZ4D uses to catabolize the catechol obtained from benzoate degradation (NahH, catechol 2,3-dioxygenase; or CatA, catechol 1,2-dioxygenase), we analyzed these two proteins in particular. NahH in BZ4D was slightly more abundant in benzoate than succinate cultures, although this difference was not statistically significant (p-value=0.052). However, in BZ4D NahH abundance was significantly lower than the basal NahH abundance obtained in AN10 succinate cultures (see Figure 4.6). In contrast, CatA abundance in BZ4D benzoate cultures (NSAF of 0.324 ± 0.016) was more than 3 times higher than NahH abundance in BZ4D benzoate cultures (NSAF of 0.014 ± 0.009). Moreover, NahH was not detected in BZ4D and AN10 proteomes in succinate. These results suggested that the pathway used by BZ4D to catalyze catechol obtained from benzoate degradation was the *ortho*-cleavage (CatA) instead of *meta*-cleavage (NahH), as expected, and as previously described for A1501 (Li *et al.,* 2010). It is also noteworthy that in the proteomes of AN10 in succinate the abundance of NahH was 9 times higher than for BZ4D, which might mean that some of the mutations suffered during the selection experiment (i.e. *nahR*) could affect the expression of NahH in BZ4D.

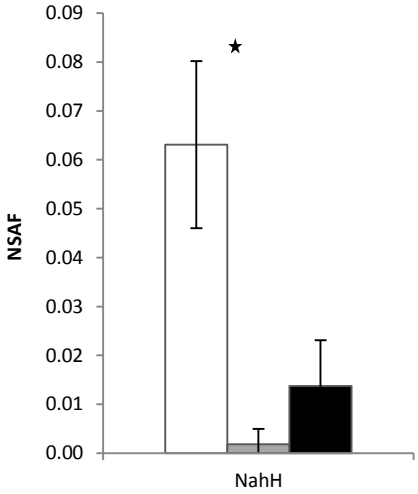

Figure 4.6. NSAF average of catechol 2,3-dioxygenase NahH obtained in AN10 growing with succinate (white), BZ4D growing with succinate (gray), and BZ4D growing with benzoate (black). Error bars represent the standard deviation between the three replicates of each culture condition.

In addition, there was also a statistically-significant difference in the abundance of NahZ, a protein presumably involved in naphthalene metabolism (see section 3.1 from Chapter 3). This protein presented a 5-fold abundance increase (p-value of 0.004) in the proteome of BZ4D growing with benzoate compared to BZ4D growing with succinate. This result indicates that this protein might also have a role in benzoate metabolism in BZ4D.

Finally, apart from the detection of the transporters BenF and BenK, a significant increase in the abundance of MexB and OprM proteins, from the MexAB-OprM transport system (see Supplementary Table S13), was also detected in BZ4D benzoate cultures compared to BZ4D succinate cultures (fold changes of 2.3 and 2.2 respectively). As was previously described in section 3.2 from Chapter 3, this transport system is homologous to the toluene tolerance system TtgABC of *P. aeruginosa* DOT-T1E, and might have a role in naphthalene metabolism (Ramos *et*

al., 1998). This suggests that, together with BenK and BenF, these transporters might have a role in benzoate metabolism in BZ4D.

4. Complementation of AN10 with BZ4D benzoate genes

The genomic and proteomic analysis of AN10 and its benzoate-degrading derivative BZ4D discussed above suggested that two of the mutations occurred in BZ4D during the selection experiment might have been essential for gaining the ability to use benzoate as sole carbon and energy source, i.e. the base change in *benR* that reverted the stop codon and the insertion in *catA* gene. To prove this hypothesis, the two mutated genes were introduced in the parental strain AN10 (see section 5 from Materials and Methods for details). Derivatives named as Z1 with the pBenR construction, V1 with the pCatA construction, and U16 with both constructions simultaneously (pBenR and pCatA) were obtained. The coexistence in these AN10 mutants of the two versions of *benR* and *catA* (the wild type and the sequence from BZ4D in the plasmids) was verified by amplification and further sequencing of both genes (see section 5 from Materials and Methods for details).

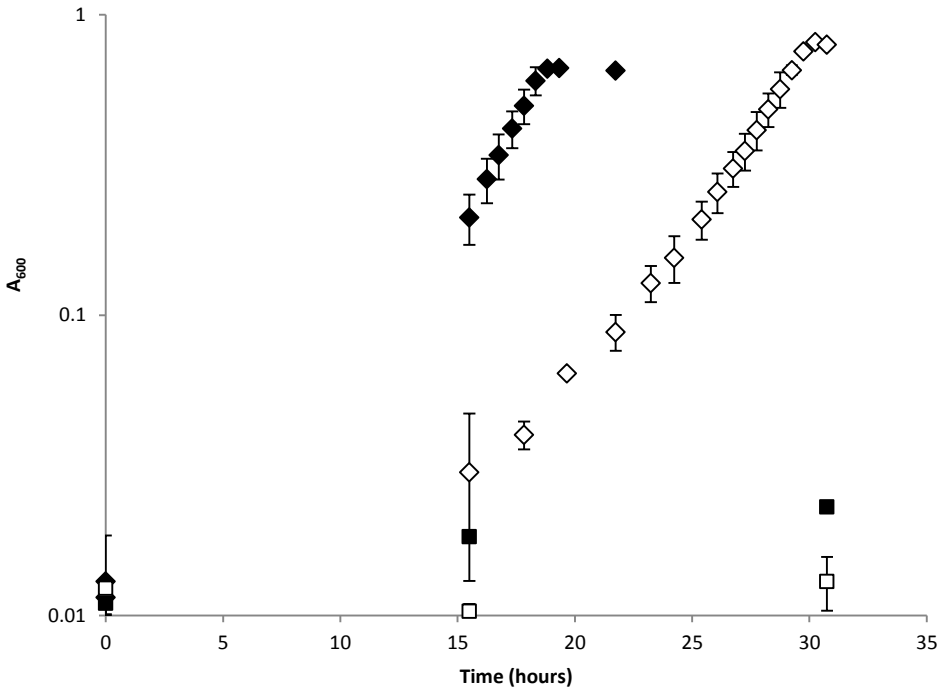

Figure 4.7. Growth of the AN10 derivative BZ4D (black diamonds); U16, the AN10 derivative complemented with pBenR and pCatA constructions, (white diamonds); Z1, the AN10 derivative complemented with the pBenR construction (black squares); and V1, the AN10 derivative complemented with the pCatA construction (white squares) in MMB supplemented with benzoate 5 mM as sole carbon and energy source.

The growth of BZ4D in MMB supplemented with benzoate 5 mM as sole carbon and energy source was compared with the growth of the 3 previously described AN10 derivatives (Z1, V1, and U16). After 5 days of incubation no growth was detected in V1 cultures. Similarly no growth was detected in Z1 cultures. These results indicated that the complementation of AN10 with

only the *catA* or only the *benR* mutated genes of BZ4D was not enough to allow growth of AN10 in benzoate as carbon source. Only strain BZ4D and AN10 derivative U16 grew with benzoate as sole carbon and energy source (see Figure 4.7). Both derivatives presented very similar generation times: approximately 4.7 ± 0.2 hours in U16 and 4.8 ± 0.1 hours in BZ4D. Based on these findings it can be concluded that both mutations were essential and sufficient for the benzoate degradation ability acquired by derivative BZ4D.

GENERAL DISCUSSION

P. stutzeri strains define a coherent species, whose members have been traditionally grouped into genomovars, and previously studied by our group due to their ability to use aromatic hydrocarbons as carbon source (Lalucat *et al*., 2006). In the present study we have focused in analyzing all *P. stutzeri* genomes available on November 2014 to verify their affiliation to the different genomovars described, contribute to classify *P. stutzeri* strains with no genomovar affiliation, and confirm the presence of aromatic hydrocarbon degradation genes in their genomes. Additionally, we performed a proteomic study of different *P. stutzeri* strains exposed to aromatic hydrocarbons to analyze their response to these compounds, and identify putative accessory proteins involved.

The classification of *P. stutzeri* strains into genomovars was proposed by Rosselló and co-workers (1991) to be used for distinct genomic groups that are sufficiently different to be classified as different species, but with phenotypes that do not show sufficient robust differences for discriminating them as different species. Later, genomovar 6 was reclassified as a different species (*Pseudomonas balearica*), since sufficient phenotypic differences were described (Bennasar *et al*., 1996). From then new *P. stutzeri* isolates have been classified into genomovars according to different techniques such as DNA-DNA hybridizations, MLSA, siderotyping, and whole-cell MALDI-TOF mass spectrometry (Sikorski *et al*., 2005; Mulet *et al*., 2012; Scotta *et al*., 2013). Accordingly, the number of genomovars has increased over the years to a minimum of 21 genomovars described so far (Scotta *et al*., 2013). But some *P. stutzeri* isolates have not been classified into genomovars yet, and the affiliation of some of the strains into this species remained unclear (Scotta *et al*., 2013; Gomila *et al*., 2015).

The genomic analysis of 18 *P. stutzeri* strains presented in this thesis revealed the presence of genes involved in the main characteristics that define this species, such as genes coding for the enzymes catalase, cytochrome *c* oxidase, and glycogen hydrolase; as well as genes involved in the formation and structure of flagellum. The analysis also allowed the identification of a core-proteome constituted by approximately the 50 % of the annotated CDSs per genome, mainly involved in metabolic functions. A phylogenomic analysis of the genomes of 18 *P. stutzeri* strains and *P. balearica* SP1402^{T} was performed by three different approaches: the number of shared COGs, a phylogeny based on 69 COGs shared between all strains, and ANIb values. We had hypothesized that a genomic analysis would confirm the coherence of the *P. stutzeri* species and the classification of its species into genomovars. However, our analyses showed that from all the studied strains only those belonging to genomovar 1 formed a coherent group. In contrast, the other strains showed clear differences with strains from genomovar 1 in all the phylogenomic analyses. According to the ANIm values established by Rosselló-Móra and Amann (2015) as the species threshold, strains that do not belong to genomovar 1 should be considered as different genomic species. If we focus on the number of shared COGs (using the criteria of sharing at least 95 % of amino acid identity in at least the 90 % of the sequence), we observed that strains that are likely to belong to the same genomic species (i.e. those in genomovar 1 and the pair 19SMN4/ST27MN3) shared more than 2700 COGs, and strains which might be genomovars or a single species (i.e AN10 and NF13) shared about 2500 COGs (see Table 1.3.). The rest of the strains shared a number of COGs well below these values. Given these results we can propose these boundaries as a criterion to take into account for species and genomovar definition in the *P. stutzeri* group.

Our phylogenomic results show that an exhaustive taxonomic study of the strains analyzed in this study is necessary to confirm their classification within or outside the *P. stutzeri* species. In case that taxonomists propose sufficient phenotypic characteristics to describe the strains that do not belong to genomovar 1 as new species, the core-proteome defined in this thesis should be considered as the core-proteome of the *P. stutzeri* group. The core-proteome of the species *P. stutzeri* would have to be recalculated in basis of the genomes of strains in genomovar 1 Although eventually the 18 *P. stutzeri* studied strains will probably be divided into several different species, the number of shared-COGs between most of them is remarkable if we compare with the core-proteomes defined for *Streptococcus agalactiae* (Tettelin *et al*., 2005), *Neisseria meningitidis* (Schoen *et al*., 2008), *Escherichia coli* (Willenbrock *et al*., 2007), and *P. putida* (Udaondo *et al*., 2015). Therefore, our results clearly shows that despite the taxonomic position of some strains may vary in the future, all these strains form a coherent group that shares a significant portion of its genome.

Several authors have described the importance of HGT in aromatic compounds degradation, as a large number of catabolic plasmids and transposons involved in the degradation of aromatic compounds have been described (Tan, 1999; Top and Springael, 2003). Our genomic study revealed that the studied *P. stutzeri* genomes rarely present aromatic hydrocarbon degradation genes in their plasmids, since from all the gene structures identified, only naphthalene degradation genes from strain 19SMN4 were located in a plasmid (pLIB119). In fact, most of the genes involved in aromatic hydrocarbon degradation, which are a good example of genes subjected to HGT, and that are encoded in plasmids in other *Pseudomonas* (such as genes involved in carbazole, phenol and naphthalene degradation), have been identified in the chromosome in *P. stutzeri* (except naphthalene degradation genes in 19SMN4). Rosselló-Mora and co-workers (1994) previously described the generalized chromosomal location of naphthalene and toluate degradation genes in 11 *P. stutzeri* strains, and proposed the acquisition of these genes due to catabolic transposons in this species. The evaluation of the presence of sequences coding for TnpAs in the genomes of *P. stutzeri* and *P. balearica* revealed a large number of these elements in several strains. The presence of transposons has been postulated as an indicative of the existence of HGT events (Sobecky and Hazen, 2009), so transposition does seem to be a mechanism for the HGT of *P. stutzeri*. The analysis of shared TnpAs revealed a large number of shared TnpAs between naphthalene degrading strains, as several genes encoding TnpAs are located between the upper and lower naphthalene degradation operons. Since the presence of transposons is considered to indicate the existence of HGT events, our results suggest that the ability of these strains to degrade naphthalene might have been acquired by different transposition events.

P. stutzeri has been described as able to degrade a wide range of aromatic compounds (see Figure I4). From them, the most studied catabolic pathways have been the naphthalene degradation pathway of strain AN10 (Bosch *et al*., 1999a; Bosch *et al*., 1999b; Bosch *et al*., 2000), the phenanthrene degradation by strain P16 (Stringfellow and Aitken, 1995), and the β-ketoadipate pathway in strain A1501 (Li *et al*., 2010). We hypothesized that a comparative genomic analysis of this species will confirm the presence of aromatic hydrocarbon degradation genes in all their genomes, which will be mainly conserved in all the studied strains. Therefore, we searched genes involved in aromatic hydrocarbon degradation in *P. stutzeri* and *P. balearica* genomes. Our genomic analysis showed the potential of this species to degrade many aromatic

hydrocarbons through six different peripheral pathways: 4-hydroxyphenylpyruvate, 4-hydroxybenzoate, benzoate, carbazole, phenol, and naphthalene via salicylate. As we hypothesized, all the studied strains showed genes involved in at least one aromatic hydrocarbon degradation pathway. Nevertheless, the aromatic hydrocarbon degradation potential of each *P. stutzeri* strain differ, being the naphthalene degrader strains those with the wider degradation potential. Some of the peripheral pathways identified were found to be common in *P. stutzeri* genomes, as genes involved in 4-hydroxyphenylpyruvate and benzoate degradation, both also common in the genomes of other *Pseudomonas* species (Jiménez *et al*., 2010). In contrast, genes involved in the degradation of aromatic hydrocarbon pollutants (such as carbazole, phenol, and naphthalene) were only present in the genomes of those strains isolated from polluted environments or under selective pressure. All the peripheral pathways identified converge to four central metabolites: catechol (*ortho*- and *meta*- cleavage), protocatechuate (only *ortho*-cleavage), homogentisate, and homoprotocatechuate. The genes involved in these pathways were not homogenously distributed, since only strains from genomovar 1 presented both branches from the β-ketoadipate pathway (catechol and protocatechuate *ortho*-cleavage). Moreover, genes involved in catechol *meta*-cleavage central pathway seem to be linked to the presence of phenol and naphthalene degradation pathways. The gene structures identified as involved in aromatic hydrocarbon degradation seem to be conserved between strains belonging to the same genomovar. Otherwise, the strains phylogenomically more distant to genomovar 1 (DSM 10701, 28a24, MF28, SDM-LAC, and TS44) present a different gene organization than the other strains. The only exception to that is the structure of phenol degradation genes of TS44, which is similar to the gene structure found in strain KOS6.

Testing the functionality of some of the identified gene structures involved in aromatic hydrocarbons degradation, in general we did not observe discrepancies between the observed and the expected growth according to gene prediction. However, some strains were not able to grow with an aromatic compound despite having the gene structure involved in its degradation. This lack of growth might be due to: (1) mutations in some of the genes involved that prevent their proper function; (2) mutations in any of the promoters involved; and (3) the inadequacy of the growth conditions tested, as the lack of the proper inductor. No growth was observed in benzoate cultures of two studied strains which presented the genes involved (AN10 and SP1402^{T}). Although this experiment was not defined in the specific objectives of this thesis, considering the results obtained for strain AN10, we performed a selection experiment which allowed us to isolate an AN10 derivative able to grow with benzoate as carbon source (BZ4D). This reveals that if a strain potentially can use a particular compound, under selective pressure we can select a mutant able to use this compound. In the case of BZ4D, two mutations in two genes involved in benzoate degradation were required, which has been demonstrated by proteomic and complementation experiments. Additionally, other 73 point mutations were observed in the AN10 derivative BZ4D after the selection experiment. One of these mutations was located in the transcriptional regulator of naphthalene degradation genes, which might cause the inefficiency of BZ4D to grow with naphthalene and salicylate. In this sense, it is remarkable that during the short period of time of the selection experiment, the derivative selected (BZ4D) had lost their ability to use salicylate as carbon source. These results are consistent with previous studies that highlighted *P. stutzeri* AN10 as bacteria with a high

genomic plasticity (Martín-Cardona, 2009). This genomic plasticity would facilitate the adaptability of this species to grow with under selective pressure. In the case of SP1402^T, we postulate that its lack of growth with benzoate might be due to the mutations observed in the BenR binding site. This might prevent the transcriptional regulation of *ben* operon by BenR. However, we believe that a selection of a SP1402^T derivative able to grow with benzoate as carbon source could be easily obtained, as it was possible with AN10.

The catabolic pathways of some *P. stutzeri* strains have been previously characterized, being naphthalene degradation by strain AN10 one of the main models. However, differently from other *Pseudomonas* species such as *P. putida* (Segura *et al*., 2005; Domínguez-Cuevas *et al*., 2006; Li *et al*., 2015) no proteomic study of *P. stutzeri* in response to aromatic hydrocarbons has been previously performed. In the present study we have analyzed the proteomic response of four *P. stutzeri* strains and *P. balearica* SP1402^T growing with naphthalene or in response to a salicylate pulse by shot-gun proteomics. This experiment allowed us to examine a considerable fraction of the proteome. We have observed for the first time the expression of naphthalene degradation proteins in *P. stutzeri* and *P. balearica* cultures growing with this polycyclic aromatic hydrocarbon as carbon source. Moreover, the results obtained with proteomics suggest the role of two other proteins (NahX and NahZ) in naphthalene degradation. We hypothesized that the former might provide adenosylcobalamin, a cofactor that might be necessary for the function of a naphthalene degradation protein such as NahJ; and the latter would act as a FAD-dependent pyruvate oxidase, collaborating in the channeling of pyruvate to acetyl-CoA during naphthalene degradation. These two hypotheses should be now tested experimentally. The reduction in the abundance of proteins involved in oxidative phosphorylation, signaling systems, and ribosomal proteins in naphthalene proteomes reveal the general response of *P. stutzeri* cells to naphthalene exposure. In this sense, apart from the up-regulation of naphthalene degradation proteins we observed a down-regulation of proteins involved in carbohydrate and energy metabolism, which would involve a diminution of the energy available for the cell. We believe that this lack of energy might cause the down-regulation of proteins involved in translation and RNA degradation. A similar response was observed in proteomes of BZ4D growing with benzoate as carbon source, suggesting that this might be a generalized response of *P. stutzeri* cells to aromatic compounds. Li and co-workers (2015) described a similar behavior for *P. putida* ND6 growing with naphthalene. An exception is that, contrarily to what we observed, these authors described an increase of heat-shock and universal stress proteins, which was attributed for them to the toxicity caused by high naphthalene concentrations. Additionally, we described for the first time the putative role of several transporters in naphthalene metabolism, and some of them highlight for their homology with previously described aromatic compound transporters. Moreover, genes coding for the transport systems over-expressed in the analyzed proteomes were also annotated in the other *P. stutzeri* and *P. balearica* genomes analyzed. The presence of these transport systems in all analyzed genomes, including genomes of non-degrading naphthalene strains, suggests that they might play a role in the metabolism of other compounds apart from naphthalene and benzoate. Clearly, further research will be required to demonstrate the role of these transport proteins in the degradation and/or tolerance to different aromatic hydrocarbons.

The results presented in this thesis provide new information of the phylogenomic relation between *P. stutzeri* strains. Additionally, it highlights the importance of aromatic hydrocarbon degradation capabilities in this species, as well as suggests new enzymes and accessory proteins involved in this metabolism.

CONCLUSIONS

1. The three phylogenomic approaches of 19 *P. stutzeri* and *P. balearica* genomes (the analysis of the number of shared COGs, the phylogenomic tree using the 69 CDSs conserved in all the studied strains, and the analysis of the ANIb values) do not support the current structure of *P. stutzeri* species, and only strains from genomovar 1 constituted a coherent group. According to the limit for species delineation described by Rosselló-Móra and Amman (2015), the remaining strains would belong to different genomic species. Therefore, an exhaustive taxonomic study of these strains is necessary for this species.

2. We defined the *P. stutzeri* core-proteome as 2,094 COGs (50 % of the annotated CDSs per genome), less than the 3,326 COGs described for *P. putida*. It was also remarkable the high number of shared TnpAs in *P. stutzeri* genomes in comparison with *P. putida* genomes. This suggests that the genetic diversity of *P. stutzeri* genomes is higher than *P. putida*. However, since the number of analyzed genomes and the criteria used are not the equal for both species we should be cautious about this comparison.

3. The 18 analyzed *P. stutzeri* strains and *P. balearica* SP1402^T presented the potential capability to use a wide range of aromatic compounds as carbon sources: catechol, protocatechuate, homogentisate, homoprotocatechuate, 4-hydroxyphenylpyruvate, 4-hydroxybenzoate, salicylate, benzoate, carbazole, phenol, and naphthalene. The homogentisate and catechol *ortho*-cleavage pathways were the most abundant, and the phenol and naphthalene pathways the more infrequent in the *P. stutzeri* studied strains.

4. Naphthalene degradation proteins have been observed for the first time in *P. stutzeri* cultures growing with naphthalene as carbon source, demonstrating the expression of the three operons involved at the protein level. Moreover, the expression of these proteins in succinate cultures exposed to a salicylate pulse confirms the induction of the naphthalene degradation pathway by salicylate.

5. The expression of proteins NahX, a putative ATP: corrinoid adenosyltransferase, and NahZ, a FAD-dependent pyruvate oxidase, in *P. stutzeri* naphthalene cultures suggests that they might play a role in naphthalene metabolism. According to the information available of these enzymes, we propose that NahX might provide adenosylcobalamin, a cofactor that might be necessary for the function of a naphthalene degradation protein such as NahJ; and NahZ might collaborate in the channeling of pyruvate to acetyl-CoA during naphthalene degradation.

6. The expression of 14 different transport systems in *P. stutzeri* naphthalene cultures, some of which highlight for their homology with previously described aromatic compound transporters (such as the toluene tolerance system TtgABC), suggests that they might play a role in the transport and/or the tolerance to naphthalene or its metabolic derivatives.

7. Despite *P. stutzeri* AN10 was previously described as a non-degrading benzoate strain, its potential to grow with benzoate as carbon source has been demonstrated. Therefore, under selective pressure in a short period of time (40 days in the case of the AN10 growth with benzoate) mutations that allow its growth on benzoate can appear. Mutations in *benR* and *catA* genes were essential and sufficient for the benzoate degradation ability acquired for the AN10 derivative BZ4D.

REFERENCES

Ahamad A, Kunhi M (1996) Degradation of phenol through *ortho*-cleavage pathway by *Pseudomonas stutzeri* strain SPC2. Lett Appl Microbiol. 22:26–9.

Ahmer BM, Thomas MG, Larsen RA, Postle K (1995) Characterization of the *exbBD* operon of *Escherichia coli* and the role of ExbB and ExbD in TonB function and stability. J Bacteriol. 177:4742–7.

Alexieva Z, Gerginova M, Manasiev J, Zlateva P, Shivarova N, Krastanov A (2008) Phenol and cresol mixture degradation by the yeast *Trichosporon cutaneum*. J Ind Microbiol Biotechnol. 35:1297–301.

Alvarez-Ortega C, Olivares J, Martínez JL (2013) RND multidrug efflux pumps: what are they good for? Front Microbiol. 4:7.

Aragno M, Schlegel H (1981) The hydrogen-oxidizing bacteria. In: The prokaryotes (Ed by Starr M, Stolp H, Trüper HG, Balows A, Schlegel HG). Springer-Verlag, Berlin Heidelberg. Vol 1, pp 864–93.

Arcos M, Olivera ER, Arias S, Naharro G, Luengo JM (2010) The 3,4-dihydroxyphenylacetic acid catabolon, a catabolic unit for degradation of biogenic amines tyramine and dopamine in *Pseudomonas putida* U. Environ Microbiol. 12:1684–704.

Arias-Barrau E, Olivera ER, Luengo JM, Fernández C, Galán B, García JL, Díaz E, Miñambres B (2004) The homogentisate pathway: a central catabolic pathway involved in the degradation of L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida*. J Bacteriol. 186:5062–77.

Armengaud J (2013) Microbiology and proteomics, getting the best of both worlds! Environ Microbiol. 15:12–23.

Bae JH, Park BG, Jung E, Lee PG, Kim BG (2014) *fadD* deletion and *fadL* overexpression in *Escherichia coli* increase hydroxy long-chain fatty acid productivity. Appl Microbiol Biotechnol. 98:8917–25.

Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I, Kimelman M, Pruitt KD, Resenchuk S, Tatusova T, Yaschenko E, Ostell J (2011) BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. Nucleic Acids Res. 40:D57-63.

Bennasar A, Rosselló-Mora R, Lalucat J, Moore ER (1996) 16S rRNA gene sequence analysis relative to genomovars of *Pseudomonas stutzeri* and proposal of *Pseudomonas balearica* sp. nov. Int J Syst Bacteriol. 46:200–5.

Bennasar-Figueras A, Salvà-Serra F, Jaén-Luchoro D, Seguí C, Aliaga F, Busquets A, Gomila M, Moore ER, Lalucat J (2016) Complete genome sequence of *Pseudomonas balearica* DSM 6083^{T}. Genome Announc. 4:e00217-16.

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science. 277:1453–62.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–20.

Bosch R, Rodríguez-Quiñones F, Imperial J (1997) Identification of gene products from the *Azotobacter vinelandii nifBfdxNnifOQ* operon. FEMS Microbiol Lett. 157:19–25.

Bosch R, García-Valdés E, Moore ER (1999a) Genetic characterization and evolutionary implications of a chromosomally encoded naphthalene-degradation upper pathway from *Pseudomonas stutzeri* AN10. Gene. 236:149–57.

Bosch R, Moore ER, García-Valdés E, Pieper DH (1999b) NahW, a novel, inducible salicylate hydroxylase involved in mineralization of naphthalene by *Pseudomonas stutzeri* AN10. J Bacteriol. 181:2315–22.

Bosch R, Imperial J (2000) Biosynthesis of the nitrogenase iron-molybdenum cofactor. Recent Res Devel Microbiology. 4:131–44.

Bosch R, García-Valdés E, Moore ER (2000) Complete nucleotide sequence and evolutionary significance of a chromosomally encoded naphthalene-degradation lower pathway from *Pseudomonas stutzeri* AN10. Gene. 245:65–74.

Bracht T, Schweinsberg V, Trippler M, Kohl M, Ahrens M, Padden J, Naboulsi W, Barkovits K, Megger DA, Eisenacher M, Borchers CH, Schlaak JF, Hoffmann AC, Weber F, Baba HA, Meyer HE, Sitek B (2015) Analysis of disease-associated protein expression using quantitative proteomics-fibulin-5 is expressed in association with hepatic fibrosis. J Proteome Res. 14:2278–86.

Brunet-Galmés I, Busquets A, Peña A, Gomila M, Nogales B, García-Valdés E, Lalucat J, Bennasar A, Bosch R (2012) Complete genome sequence of the naphthalene-degrading bacterium *Pseudomonas stutzeri* AN10 (CCUG 29243). J Bacteriol. 194:6642–3.

Burlage RS, Hooper SW, Sayler GS (1989) The TOL (pWW0) catabolic plasmid. Appl Environ Microbiol. 55:1323–8.

Burri R, Stutzer A (1895) Ueber Nitrat zerstörende bakterien und den durch dieselben bedingten stickstoffverlust. Zentbl Bakteriol Parasitenkd. 1:257–432.

Busquets A, Peña A, Gomila M, Bosch R, Nogales B, García-Valdés E, Lalucat J, Bennasar A (2012) Genome sequence of *Pseudomonas stutzeri* strain JM300 (DSM 10701), a soil isolate and model organism for natural transformation. J Bacteriol. 194:5477–8.

Busquets A, Peña A, Gomila M, Mayol J, Bosch R, Nogales B, García-Valdés E, Bennasar A, Lalucat J (2013) Draft genome sequence of *Pseudomonas stutzeri* strain B1SMN1, a nitrogen-fixing and naphthalene-degrading strain isolated from wastewater. Genome Announc. 1:e00584–13.

Cai L, Liu G, Rensing C, Wang G (2009) Genes involved in arsenic transformation and resistance associated with different levels of arsenic-contaminated soils. BMC Microbiol. 9:4.

Callaghan AV, Morris BE, Pereira IA, McInerney MJ, Austin RN, Groves JT, Kukor JJ, Suflita JM, Young LY, Zylstra GJ, Wawrik B (2012) The genome sequence of *Desulfatibacillum alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation. Environ Microbiol. 14:101–13.

Carvalho PC, Fischer JS, Xu T, Yates JR 3rd, Barbosa VC (2012) PatternLab: from mass spectra to label-free differential shotgun proteomics. Curr Protoc Bioinformatics. 13:13–9.

Cases I, de Lorenzo V (2005) Genetically modified organisms for the environment: stories of success and failure and what we have learned from them. Int Microbiol. 8:213–22.

Chauhan A, Green S, Pathak A, Thomas J, Venkatramanan R (2013) Whole-genome sequences of five oyster-associated bacteria show potential for crude oil hydrocarbon degradation. Genome Announc. 1:e00802–13.

Chen D, Chen J, Zhong W, Cheng Z (2008) Degradation of methyl *tert*-butyl ether by gel immobilized *Methylibium petroleiphilum* PM1. Bioresour Technol. 99:4702–8.

Chen M, Yan Y, Zhang W, Lu W, Wang J, Ping S, Lin M (2011) Complete genome sequence of the type strain *Pseudomonas stutzeri* CGMCC 1.1803. J Bacteriol. 193:6095.

Christie-Oleza JA, Lanfranconi MP, Nogales B, Lalucat J, Bosch R (2009) Conjugative interaction induces transposition of IS*Pst9* in *Pseudomonas stutzeri* AN10. J Bacteriol. 191:1239–47.

Christie-Oleza JA, Armengaud J (2010) In-depth analysis of exoproteomes from marine bacteria by shotgun liquid chromatography-tandem mass spectrometry: the *Ruegeria pomeroyi* DSS-3 case-study. Mar Drugs. 8:2223–39.

Christie-Oleza JA, Fernandez B, Nogales B, Bosch R, Armengaud J (2012) Proteomic insights into the lifestyle of an environmentally relevant marine bacterium. ISME J. 6:124–35.

Christie-Oleza JA, Armengaud J, Guerin P, Scanlan DJ (2015) Functional distinctness in the exoproteomes of marine *Synechococcus*. Environ Microbiol. 17:3781–94.

Clair G, Roussi S, Armengaud J, Duport C (2010) Expanding the known repertoire of virulence factors produced by *Bacillus cereus* through early secretome profiling in three redox conditions. Mol Cell Proteomics. 9:1486–98.

Collier LS, Nichols NN, Neidle EL (1997) *benK* encodes a hydrophobic permease-like protein involved in benzoate degradation by *Acinetobacter* sp. strain ADP1. J Bacteriol. 179:5943–6.

Colquhoun DR, Hartmann EM, Halden RU (2012) Proteomic profiling of the dioxin-degrading bacterium *Sphingomonas wittichii* RW1. J Biomed Biotechnol. 2012:408690.

Davidson AL, Dassa E, Orelle C, Chen J (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol Mol Biol Rev. 72:317–64.

de Groot A, Dulermo R, Ortet P, Blanchard L, Guérin P, Fernandez B, Vacherie B, Dossat C, Jolivet E, Siguier P, Chandler M, Barakat M, Dedieu A, Barbe V, Heulin T, Sommer S, Achouak W, Armengaud J (2009) Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. PLoS Genet. 5:e1000434.

Dean DR, Jacobson MR (1992) Biochemical genetics of nitrogenase. In: Biological nitrogen fixation (Ed by Stacey G, Burris RH, Evans HJ). Chapman and Hall, New York. pp 762–834.

Dean DR, Bolin JT, Zheng L (1993) Nitrogenase metalloclusters: structures, organization, and synthesis. J Bacteriol. 175:6737–44.

Demba Diallo M, Willems A, Vloemans N, Cousin S, Vandekerckhove TT, de Lajudie P, Neyra M, Vyverman W, Gillis M, Van der Gucht K (2004) Polymerase chain reaction denaturing gradient gel electrophoresis analysis of the N_2-fixing bacterial diversity in soil under *Acacia tortilis* ssp. raddiana and *Balanites aegyptiaca* in the dryland part of Senegal. Environ Microbiol. 6:400–15.

Dennis JJ, Zylstra GJ (2004) Complete sequence and genetic organization of pDTG1, the 83 kilobase naphthalene degradation plasmid from *Pseudomonas putida* strain NCIB 9816-4. J Mol Biol. 341:753–68.

Denome SA, Stanley DC, Olson ES, Young KD (1993) Metabolism of dibenzothiophene and naphthalene in *Pseudomonas* strains: complete DNA sequence of an upper naphthalene catabolic pathway. J Bacteriol. 175:6890–901.

Desnoues N, Lin M, Guo X, Ma L, Carreño-Lopez R, Elmerich C (2003) Nitrogen fixation genetics and regulation in a *Pseudomonas stutzeri* strain associated with rice. Microbiology. 149:2251–62.

Díaz E (2004) Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. Int Microbiol. 7:173–80.

Doherty MK, Whitfield PD (2011) Proteomics moves from expression to turnover: update and future perspective. Expert Rev Proteomics. 8:325–34.

Domínguez-Cuevas P, González-Pastor JE, Marqués S, Ramos JL, de Lorenzo V (2006) Transcriptional tradeoff between metabolic and stress-response programs in *Pseudomonas putida* KT2440 cells exposed to toluene. J Biol Chem. 281:11981–91.

Domínguez-Cuevas P, Marín P, Marqués S, Ramos JL (2008) XylS-Pm promoter interactions through two helix-turn-helix motifs: identifying XylS residues important for DNA binding and activation. J Mol Biol. 375:59–69.

Dos Santos PC, Dean DR (2011) Co-ordination and fine-tuning of nitrogen fixation in *Azotobacter vinelandii*. Mol Microbiol. 79:1132–5.

Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R (2012) Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. BMC Genomics. 13:162.

Dupierris V, Masselon C, Court M, Kieffer-Jaquinod S, Bruley C (2009) A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa. Bioinformatics. 25:1980–1.

Duque E, Segura A, Mosqueda G, Ramos JL (2001) Global and cognate regulators control the expression of the organic solvent efflux pumps TtgABC and TtgDEF of *Pseudomonas putida*. Mol Microbiol. 39:1100–6.

Dutta TK, Selifonov SA, Gunsalus IC (1998) Oxidation of methyl-substituted naphthalenes: pathways in a versatile *Sphingomonas paucimobilis* strain. Appl Environ Microbiol. 64:1884–9.

Endo K, Hosono K, Beppu T, Ueda K (2002) A novel extracytoplasmic phenol oxidase of *Streptomyces*: its possible involvement in the onset of morphogenesis. Microbiology. 148:1767–76.

Espadaler J, Eswar N, Querol E, Avilés FX, Sali A, Marti-Renom MA, Oliva B (2008) Prediction of enzyme function by combining sequence similarity and protein interactions. BMC Bioinformatics. 9:249.

Essam T, Amin MA, El Tayeb O, Mattiasson B, Guieysse B (2010) Kinetics and metabolic versatility of highly tolerant phenol degrading *Alcaligenes* strain TW1. J Hazard Mater. 173:783–8.

Euzéby JP (1997) List of bacterial names with standing in nomenclature: a folder available on the Internet. Int J Syst Bacteriol. 47: 590–2.

Feist CF, Hegeman GD (1969) Phenol and benzoate metabolism by *Pseudomonas putida*: regulation of tangential pathways. J Bacteriol. 100:869–77.

Ferrero M, Llobet-Brossa E, Lalucat J, García-Valdés E, Rosselló-Mora R, Bosch R (2002) Coexistence of two distinct copies of naphthalene degradation genes in *Pseudomonas* strains isolated from the western Mediterranean region. Appl Environ Microbiol. 68:957–62.

Findeisen A, Kolivoska V, Kaml I, Baatz W, Kenndler E (2007) Analysis of diterpenoic compounds in natural resins applied as binders in museum objects by capillary electrophoresis. J Chromatogr A. 1157:454–61.

Forward JA, Behrendt MC, Wyborn NR, Cross R, Kelly DJ (1997) TRAP transporters: a new family of periplasmic solute transport systems encoded by the *dctPQM* genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. J Bacteriol. 179:5482–93.

Foster B, Pukall R, Abt B, Nolan M, Glavina Del Rio T, Chen F, Lucas S, Tice H, Pitluck S, Cheng JF, Chertkov O, Brettin T, Han C, Detter JC, Bruce D, Goodwin L, Ivanova N, Mavromatis K, Pati A, Mikhailova N, Chen A, Palaniappan K, Land M, Hauser L, Chang YJ, Jeffries CD, Chain P, Rohde M, Göker M, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP, Lapidus A (2010) Complete genome sequence of *Xylanimonas cellulosilytica* type strain (XIL07). Stand Genomic Sci. 2:1–8.

Fredrickson JK, Brockman FJ, Workman DJ, Li SW, Stevens TO (1991) Isolation and characterization of a subsurface bacterium capable of growth on toluene, naphthalene, and other aromatic compounds. Appl Environ Microbiol. 57:796–803.

Fuchs G, Boll M, Heider J (2014) Microbial degradation of aromatic compounds - from one strategy to four. Nat Rev Microbiol. 9:803–16.

Fuentes S, Méndez V, Aguila P, Seeger M (2014) Bioremediation of petroleum hydrocarbons: catabolic genes, microbial communities, and applications. Appl Microbiol Biotechnol. 98:4781–94.

García-Valdés E, Cozar E, Rotger R, Lalucat J, Ursing J (1988) New naphthalene-degrading marine *Pseudomonas* strains. Appl Environ Microbiol. 54:2478–85.

García-Valdés E, Castillo MM, Bennasar A, Guasp C, Cladera AM, Bosch R, Engesser KH, Lalucat J (2003) Polyphasic characterization of *Pseudomonas stutzeri* CLN100 which simultaneously degrades chloro- and methylaromatics: a new genomovar within the species. Syst Appl Microbiol. 26:390–403.

George KW, Hay AG (2011) Bacterial strategies for growth on aromatic compounds. Adv Appl Microbiol. 74:1–33.

Geszvain K, Tebo BM (2010) Identification of a two-component regulatory pathway essential for Mn(II) oxidation in *Pseudomonas putida* GB-1. Appl Environ Microbiol. 76:1224–31.

Giedyk M, Goliszewska K, Gryko D (2015) Vitamin B12 catalysed reactions. Chem Soc Rev. 2015 44:3391–404.

Gilardi GL (1971) Characterization of *Pseudomonas* species isolated from clinical specimens. Appl Microbiol. 21:414–9.

Ginard M, Lalucat J, Tümmler B, Römling U (1997) Genome organization of *Pseudomonas stutzeri* and resulting taxonomic and evolutionary considerations. Int J Syst Bacteriol. 47:132–43.

Godoy P, Ramos-González MI, Ramos JL (2001) Involvement of the TonB system in tolerance to solvents and drugs in *Pseudomonas putida* DOT-T1E. J Bacteriol. 183:5285–92.

Göker M, Lu M, Fiebig A, Nolan M, Lapidus A, Tice H, Del Rio TG, Cheng JF, Han C, Tapia R, Goodwin LA, Pitluck S, Liolios K, Mavromatis K, Pagani I, Ivanova N, Mikhailova N, Pati A, Chen A, Palaniappan K, Land M, Mayilraj S, Rohde M, Detter JC, Bunk B, Spring S, Wirth R, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP (2014) Genome sequence of the mud-dwelling archaeon *Methanoplanus limicola* type strain (DSM 2279^{T}), reclassification of *Methanoplanus petrolearius* as *Methanolacinia petrolearia* and emended descriptions of the genera *Methanoplanus* and *Methanolacinia*. Stand Genomic Sci. 9:1076–88.

Golyshin PN, Martins Dos Santos VA, Kaiser O, Ferrer M, Sabirova YS, Lünsdorf H, Chernikova TN, Golyshina OV, Yakimov MM, Pühler A, Timmis KN (2003) Genome sequence completed of *Alcanivorax borkumensis*, a hydrocarbon-degrading bacterium that plays a global role in oil removal from marine systems. J Biotechnol. 106:215–20.

Gomila M, Peña A, Mulet M, Lalucat J, García-Valdés E (2015) Phylogenomics and systematics in *Pseudomonas*. Front Microbiol. 6:214.

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 57:81–91.

Grigoryeva TV, Laikov AV, Naumova RP, Manolov AI, Larin AK, Karpova IY, Semashko TA, Alexeev DG, Kostryukova ES, Müller R, Govorun VM (2013) Draft genome of the nitrogen-fixing

bacterium *Pseudomonas stutzeri* strain KOS6 isolated from industrial hydrocarbon sludge. Genome Announc. 1:e00072–12.

Halpern BS, Selkoe KA, Micheli F, Kappel CV (2007) Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats. Conserv Biol. 21:1301–15.

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological Statistics software package for education and data analysis. Paleontologia Electronica. 4:9.

Harayama S, Rekik M (1989) Bacterial aromatic ring-cleavage enzymes are classified into two different gene families. J Biol Chem. 264:15328–33.

Härtig E, Schiek U, Vollack KU, Zumft WG (1999) Nitrate and nitrite control of respiratory nitrate reduction in denitrifying *Pseudomonas stutzeri* by a two-component regulatory system homologous to NarXL of *Escherichia coli*. J Bacteriol. 181:3658–65.

Hartmann EM, Armengaud J (2014) Shotgun proteomics suggests involvement of additional enzymes in dioxin degradation by *Sphingomonas wittichii* RW1. Environ Microbiol. 16:162–76.

Harwood CS, Parales RE (1996) The beta-ketoadipate pathway and the biology of self-identity. Annu Rev Microbiol. 50:553–90.

Haussmann U, Qi SW, Wolters D, Rögner M, Liu SJ, Poetsch A (2009) Physiological adaptation of *Corynebacterium glutamicum* to benzoate as alternative carbon source a membrane proteome-centric view. Proteomics. 9:3635–51.

Hayman GT, Farrand SK (1990) *Agrobacterium* plasmids encode structurally and functionally different loci for catabolism of agrocinopine-type opines. Mol Gen Genet. 223:465–73.

Hemamalini R, Khare S (2014) A proteomic approach to understand the role of the outer membrane porins in the organic solvent-tolerance of *Pseudomonas aeruginosa* PseA. PLoS One. 9:e103788.

Herrero M, de Lorenzo V, Timmis KN (1990) Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. J Bacteriol. 172:6557–67.

Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 26:680–2.

Jacobsen A, Hendriksen RS, Aaresturp FM, Ussery DW, Friis C (2011) The *Salmonella enterica* pan-genome. Microb Ecol. 62:487–504.

Jencova V, Strnad H, Chodora Z, Ulbrich P, Vlcek C, Hickey WJ, Paces V (2008) Nucleotide sequence, organization and characterization of the (halo)aromatic acid catabolic plasmid pA81 from *Achromobacter xylosoxidans* A8. Res Microbiol. 159:118–27.

Jeon CO, Park M, Ro HS, Park W, Madsen EL (2008) The naphthalene catabolic (*nag*) genes of *Polaromonas naphthalenivorans* CJ2: evolutionary implications for two gene clusters and novel regulatory control. Appl Environ Microbiol. 72:1086–95.

Jeong JJ, Kim JH, Kim CK, Hwang I, Lee K (2003) 3- and 4-alkylphenol degradation pathway in *Pseudomonas* sp. strain KL28: genetic organization of the *lap* gene cluster and substrate specificities of phenol hydroxylase and catechol 2,3-dioxygenase. Microbiology. 149:3265–77.

Jha AM, Bharti MK (2002) Mutagenic profiles of carbazole in the male germ cells of Swiss albino mice. Mutat Res. 500:97–101.

Jiang T, Gao C, Su F, Zhang W, Hu C, Dou P, Zheng Z, Tao F, Ma C, Xu P (2012) Genome sequence of *Pseudomonas stutzeri* SDM-LAC, a typical strain for studying the molecular mechanism of lactate utilization. J Bacteriol. 194:894–5.

Jiménez JI, Miñambres B, García JL, Díaz E (2002) Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. Environ Microbiol. 4:824–41.

Jiménez JI, Miñambres B, García JL, Díaz E (2004) Genomic insights in the metabolism of aromatic compounds in *Pseudomonas*. In: *Pseudomonas* (Ed by Ramos JL). Kluwer Academic/Plenum Publishers, New York. Vol 3, pp 425–62.

Jiménez JI, Nogales J, García JL, Díaz E (2010) A genomic view of the catabolism of aromatic compounds in *Pseudomonas*. In: Handbook of Hydrocarbon and Lipid Microbiology (Ed by Timmis KN). Springer-Verlag, Berlin Heidelberg. Chapter 35, pp 1297–325.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL (2008) NCBI BLAST: a better web interface. Nucleic Acids Res. 36:W5–9.

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8:275–82.

Jung H (2002) The sodium/substrate symporter family: structural and functional features. FEBS Lett. 529:73–7.

Kaczorek E, Sałek K, Guzik U, Jesionowski T, Cybulski Z (2013) Biodegradation of alkyl derivatives of aromatic hydrocarbons and cell surface properties of a strain of *Pseudomonas stutzeri*. Chemosphere. 90:471–8.

Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, Leibler S, Surette MG, Alon U (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. Science. 292:2080–3.

Karlsson FH, Ussery DW, Nielsen J, Nookaew I (2011) A closer look at *Bacteroides*: phylogenetic relationship and genomic implications of a life in the human gut. Microb Ecol. 61:473–85

Kasahara Y, Morimoto H, Kuwano M, Kadoya R (2012) Genome-wide analytical approaches using semi-quantitative expression proteomics for aromatic hydrocarbon metabolism in *Pseudomonas putida* F1. J Microbiol Methods. 91:434–42.

Kasai Y, Inoue J, Harayama S (2001) The TOL plasmid pWW0 *xylN* gene product from *Pseudomonas putida* is involved in m-xylene uptake. J Bacteriol. 183:6662–6.

Kasai Y, Harayama S (2004) Catabolism of PAHs. In: *Pseudomonas* (Ed by Ramos JL). Kluwer Academic/Plenum Publishers, New York. Vol 3, pp 463–90.

Kazunga C, Aitken MD (2000) Products from the incomplete metabolism of pyrene by polycyclic aromatic hydrocarbon-degrading bacteria. Appl Environ Microbiol. 66:1917–22.

Kemp MB, Hegeman GD (1968) Genetic control of the beta-ketoadipate pathway in *Pseudomonas aeruginosa*. J Bacteriol. 96:1488–99.

Kim IC, Oriel PJ (1995) Characterization of the *Bacillus stearothermophilus* BR219 phenol hydroxylase gene. Appl Environ Microbiol. 61:1252–6.

Knowles R (1982) Denitrification. Microbiol Rev. 46:43–70.

Koh S, McCullar MV, Focht DD (1997) Biodegradation of 2,4-dichlorophenol through a distal *meta*-fission pathway. Appl Environ Microbiol. 63:2054–7.

Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM 2nd, Peterson KM (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene. 166:175–6.

Krotzky A, Werner D (1987) Nitrogen fixation in *Pseudomonas stutzeri*. Arch. Microbiol. 147:48–57.

Kung VL, Ozer EA, Hauser AR (2010) The accessory genome of *Pseudomonas aeruginosa*. Microbiol Mol Biol Rev. 74:621–41.

Lalucat J, Bennasar A, Bosch R, García-Valdés E, Palleroni NJ (2006) Biology of *Pseudomonas stutzeri*. Microbiol Mol Biol Rev. 70:510–47.

Lane D (1991) 16S/23S sequencing. In: Nucleic acids techniques in bacterial systematics (Ed by Stackebrandt EG and Goodfellow M). John Wiley and Sons, Chichester. pp 115–75.

Lanfranconi MP, Christie-Oleza JA, Martín-Cardona C, Suárez-Suárez LY, Lalucat J, Nogales B, Bosch R (2009) Physiological role of NahW, the additional salicylate hydroxylase found in *Pseudomonas stutzeri* AN10. FEMS Microbiol Lett. 300:265–72.

Law CJ, Maloney PC, Wang DN (2008) Ins and outs of major facilitator superfamily antiporters. Annu Rev Microbiol. 62:289–305.

Lee JY, Park HS, Kim HS (1999) Adenosylcobalamin-mediated methyl transfer by toluate cis-dihydrodiol dehydrogenase of the TOL plasmid pWW0. J Bacteriol. 181:2953–7.

Leekitcharoenphon P, Lukjancenko O, Friis C, Aarestrup FM, Ussery DW (2012) Genomic variation in *Salmonella enterica* core genes for epidemiological typing. BMC Genomics. 13:88.

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol. 4:203–21.

Li D, Yan Y, Ping S, Chen M, Zhang W, Li L, Lin W, Geng L, Liu W, Lu W, Lin M (2010) Genome-wide investigation and functional characterization of the beta-ketoadipate pathway in the

nitrogen-fixing and root-associated bacterium *Pseudomonas stutzeri* A1501. BMC Microbiol. 10:36.

Li A, Gai Z, Cui D, Ma F, Yang J, Zhang X, Sun Y, Ren N (2012) Genome sequence of a highly efficient aerobic denitrifying bacterium, *Pseudomonas stutzeri* T13. J Bacteriol. 194:5720.

Li X, Gong J, Hu Y, Cai L, Johnstone L, Grass G, Rensing C, Wang G (2012) Genome sequence of the moderately halotolerant, arsenite-oxidizing bacterium *Pseudomonas stutzeri* TS44. J Bacteriol. 194:4473–4.

Li SS, Hu X, Zhao H, Li YX, Zhang L, Gong LJ, Guo J, Zhao HB (2015) Quantitative analysis of cellular proteome alterations of *Pseudomonas putida* to naphthalene-induced stress. Biotechnol Lett. 37:1645–54.

Lien T, Madsen M, Rainey FA, Birkeland NK (1998) *Petrotoga mobilis* sp. nov., from a North Sea oil-production well. Int J Syst Bacteriol. 48:1007–13.

Liu H, Sadygov RG, Yates JR 3rd (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 76:4193–201.

Liu X, Gai Z, Tao F, Yu H, Tang H, Xu P (2012) Genome sequences of *Pseudomonas luteola* XLDN4-9 and *Pseudomonas stutzeri* XLDN-R, two efficient carbazole-degrading strains. J Bacteriol. 194:5701–2.

Liu H, Sun WB, Liang RB, Huang L, Hou JL, Liu JH (2015) iTRAQ-based quantitative proteomic analysis of *Pseudomonas aeruginosa* SJTD-1: A global response to n-octadecane induced stress. J Proteomics. 123:14–28.

Lovell CR, Piceno YM, Quattro JM, Bagwell CE (2000) Molecular analysis of diazotroph diversity in the rhizosphere of the smooth cordgrass, *Spartina alterniflora*. Appl Environ Microbiol. 66:3814–22.

Lukjancenko O, Wassenaar TM, Ussery DW (2010) Comparison of 61 sequenced *Escherichia coli* genomes. Microb Ecol. 60:708–20.

Lukjancenko O, Ussery DW, Wassenaar TM (2012) Comparative genomics of *Bifidobacterium*, *Lactobacillus* and related probiotic genera. Microb Ecol. 63:651–73.

Lykidis A, Pérez-Pantoja D, Ledger T, Mavromatis K, Anderson IJ, Ivanova NN, Hooper SD, Lapidus A, Lucas S, González B, Kyrpides NC (2010) The complete multipartite genome sequence of *Cupriavidus necator* JMP134, a versatile pollutant degrader. PLoS One. 5:e9729.

Ma C, Gao C, Qiu J, Hao J, Liu W, Wang A, Zhang Y, Wang M, Xu P (2007) Membrane-bound L- and D-lactate dehydrogenase activities of a newly isolated *Pseudomonas stutzeri* strain. Appl Microbiol Biotechnol. 77:91–8.

Mahajan MC, Phale PS, Vaidyanathan CS (1994) Evidence for the involvement of multiple pathways in the biodegradation of 1- and 2-methylnaphthalene by *Pseudomonas putida* CSV86. Arch Microbiol. 161:425–33.

Mahan KM, Penrod JT, Ju KS, Al Kass N, Tan WA, Truong R, Parales JV, Parales RE (2015) Selection for growth on 3-nitrotoluene by 2-nitrotoluene-utilizing *Acidovorax* sp. strain JS42 identifies nitroarene dioxygenases with altered specificities. Appl Environ Microbiol. 81:309–19.

Mampel J, Maier E, Tralau T, Ruff J, Benz R, Cook AM (2004) A novel outer-membrane anion channel (porin) as part of a putatively two-component transport system for 4-toluenesulphonate in *Comamonas testosteroni* T-2. Biochem J. 383:91–9.

Manara A, DalCorso G, Baliardini C, Farinati S, Cecconi D, Furini A (2012) *Pseudomonas putida* response to cadmium: changes in membrane and cytosolic proteomes. J Proteome Res. 11:4169–79.

Marasco EK, Vay K, Schmidt-Dannert C (2006) Identification of carotenoid cleavage dioxygenases from *Nostoc* sp. PCC 7120 with different cleavage activities. J Biol Chem. 281:31583–93.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH (2015) CDD: NCBI's conserved domain database. Nucleic Acids Res. 43:D222–6.

Margesin R, Schinner F (2001) Biodegradation and bioremediation of hydrocarbons in extreme environments. Appl Microbiol Biotechnol. 56:650–63.

Martin VJ, Mohn WW (1999) A novel aromatic-ring-hydroxylating dioxygenase from the diterpenoid-degrading bacterium *Pseudomonas abietaniphila* BKME-9. J Bacteriol. 181:2675–82.

Martin VJ, Mohn WW (2000) Genetic investigation of the catabolic pathway for degradation of abietane diterpenoids by *Pseudomonas abietaniphila* BKME-9. J Bacteriol. 182:3784–93.

Martín-Cardona C (2009) Versatilidad genético-fisiológica en poblaciones de *Pseudomonas stutzeri* AN10 sometidas a estrés químico por salicilato. Doctoral thesis. Universitat de les Illes Balears.

Martínez-García E, Nikel PI, Chavarría M, de Lorenzo V (2014) The metabolic cost of flagellar motion in *Pseudomonas putida* KT2440. Environ Microbiol. 16:291–303.

Maseda H, Kitao M, Eda S, Yoshihara E, Nakae T (2002) A novel assembly process of the multicomponent xenobiotic efflux pump in *Pseudomonas aeruginosa*. Mol Microbiol. 46:677–86.

Mattes TE, Alexander AK, Richardson PM, Munk AC, Han CS, Stothard P, Coleman NV (2008) The genome of *Polaromonas* sp. strain JS666: insights into the evolution of a hydrocarbon- and xenobiotic-degrading bacterium, and features of relevance to biotechnology. Appl Environ Microbiol. 74:6405–16.

Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 14:60.

Michalska K, Chang C, Mack JC, Zerbs S, Joachimiak A, Collart FR (2012) Characterization of transport proteins for aromatic compounds derived from lignin: benzoate derivative binding proteins. J Mol Biol. 423:555–75.

Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A, Johnson J, Li K, Mobarry C, Sutton G (2008) Aggressive assembly of pyrosequencing reads with mates. Bioinformatics. 24: 2818–24.

Mohn WW (1995) Bacteria obtained from a sequencing batch reactor that are capable of growth on dehydroabietic acid. Appl Environ Microbiol. 61:2145–50.

Moran MA, Belas R, Schell MA, González JM, Sun F, Sun S, Binder BJ, Edmonds J, Ye W, Orcutt B, Howard EC, Meile C, Palefsky W, Goesmann A, Ren Q, Paulsen I, Ulrich LE, Thompson LS, Saunders E, Buchan A (2007) Ecological genomics of marine Roseobacters. Appl Environ Microbiol. 73:4559–69.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–5.

Mulet M, Gomila M, Gruffaz C, Meyer JM, Palleroni NJ, Lalucat J, García-Valdés E (2008) Phylogenetic analysis and siderotyping as useful tools in the taxonomy of *Pseudomonas stutzeri*: description of a novel genomovar. Int J Syst Evol Microbiol. 58:2309–15.

Mulet M, Lalucat J, García-Valdés E (2010) DNA sequence-based analysis of the *Pseudomonas* species. Environ Microbiol. 12:1513–30.

Mulet M, Gomila M, Lemaitre B, Lalucat J, García-Valdés E (2012) Taxonomic characterization of *Pseudomonas* strain L48 and formal proposal of *Pseudomonas entomophila* sp. nov. Syst Appl Microbiol. 35:145–9.

Nam IH, Hong HB, Kim YM, Kim BH, Murugesan K, Chang YS (2005) Biological removal of polychlorinated dibenzo-p-dioxins from incinerator fly ash by *Sphingomonas wittichii* RW1. Water Res. 39:4651–60.

Naumova RP, Grigoryeva TV, Rizvanov AA, Gogolev JV, Kudrjashova NV, Laikov AV (2009) Diazotrophs originated from petrochemical sludge as a potential resource of waste remediation. World Appl Sci J. 6:154–7.

Neidle EL, Hartnett C, Bonitz S, Ornston LN (1988) DNA sequence of the *Acinetobacter calcoaceticus* catechol 1,2-dioxygenase I structural gene *catA*: evidence for evolutionary divergence of intradiol dioxygenases by acquisition of DNA sequence repetitions. J Bacteriol. 170:4874–80.

Nies DH, Rehbein G, Hoffmann T, Baumann C, Grosse C (2006) Paralogs of genes encoding metal resistance proteins in *Cupriavidus metallidurans* strain CH34. J Mol Microbiol Biotechnol. 11:82–93.

Nishino SH, Spain JC (2004) Catabolism of nitroaromatic compounds. In: *Pseudomonas* (Ed by Ramos JL). Kluwer Academic/Plenum Publishers, New York. Vol 3, pp 575–608.

Nishiyama E, Ohtsubo Y, Yamamoto Y, Nagata Y, Tsuda M (2012) Pivotal role of anthranilate dioxygenase genes in the adaptation of *Burkholderia multivorans* ATCC 17616 in soil. FEMS Microbiol Lett. 330:46–55.

Noinaj N, Guillier M, Barnard TJ, Buchanan SK (2010) TonB-dependent transporters: regulation, structure, and function. Annu Rev Microbiol. 64:43–60.

Nojiri H, Habe H, Omori T (2001a) Bacterial degradation of aromatic compounds via angular dioxygenation. J Gen Appl Microbiol. 47:279–305.

Nojiri H, Sekiguchi H, Maeda K, Urata M, Nakai S, Yoshida T, Habe H, Omori T (2001b) Genetic characterization and evolutionary implications of a *car* gene cluster in the carbazole degrader *Pseudomonas* sp. strain CA10. J Bacteriol. 183:3663–79.

Nojiri H, Shintani M, Omori T (2004) Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. Appl Microbiol Biotechnol. 64:154–74.

Nojiri H (2012) Structural and molecular genetic analyses of the bacterial carbazole degradation system. Biosci Biotechnol Biochem. 76:1–18.

Ohmoto T, Kinoshita T, Moriyoshi K, Sakai K, Hamada N, Ohe T (1998) Purification and some properties of 2-hydroxychromene-2-carboxylate isomerase from naphthalenesulfonate-assimilating *Pseudomonas* sp. TA-2. J Biochem. 124:591–7.

Ornston LN, Neidle EL, Houghton JE (1990) Gene rearrangements, a force for evolutionary change; DNA sequence arrangements, a source of genetic constancy. In: The bacterial chromosome (Ed by Drlica K, Riley M). American Society for Microbiology, Washington D.C. pp 325–34.

Osborne DJ, Pickup RW, Williams PA (1988) The presence of two complete homologous *meta* pathway operons on TOL plasmid pWW53. J Gen Microbiol. 134:2965–75.

O'Sullivan LA, Mahenthiralingam E (2005) Biotechnological potential within the genus *Burkholderia*. Lett Appl Microbiol. 41:8–11.

Palleroni NJ, Doudoroff M, Stanier RY, Solánes RE, Mandel M (1970) Taxonomy of the aerobic pseudomonads: the properties of the *Pseudomonas stutzeri* group. J Gen Microbiol. 60:215–31.

Park HH, Lee HY, Lim WK, Shin HJ (2005) NahR: effects of replacements at Asn 169 and Arg 248 on promoter binding and inducer recognition. Arch Biochem Biophys. 434:67–74.

Paul D, Bridges S, Burgess SC, Dandass Y, Lawrence ML (2008) Genome sequence of the chemolithoautotrophic bacterium *Oligotropha carboxidovorans* OM5T. J Bacteriol. 190:5531–2.

Peña A, Busquets A, Gomila M, Bosch R, Nogales B, García-Valdés E, Lalucat J, Bennasar A (2012) Draft genome of *Pseudomonas stutzeri* strain ZoBell (CCUG 16156), a marine isolate and model organism for denitrification studies. J Bacteriol. 194:1277–8.

Peña A, Busquets A, Gomila M, Mayol J, Bosch R, Nogales B, García-Valdés E, Bennasar A, Lalucat J (2013) Draft genome of *Pseudomonas stutzeri* strain NF13, a nitrogen fixer isolated from the Galapagos rift hydrothermal vent. Genome Announc. 1:e0011313.

Pérez-Pantoja D, Donoso R, Agulló L, Córdova M, Seeger M, Pieper DH, González B (2012) Genomic analysis of the potential for aromatic compounds biodegradation in *Burkholderiales*. Environ Microbiol. 14:1091–117.

Pitcher RS, Cheesman MR, Watmough NJ (2002) Molecular and spectroscopic analysis of the cytochrome cbb_3 oxidase from *Pseudomonas stutzeri*. J Biol Chem. 277:31474–83.

Powlowski J, Shingler V (1994) Genetics and biochemistry of phenol degradation by *Pseudomonas* sp. CF600. Biodegradation. 5:219–36.

Qiu YS, Zhou SP, Mo XZ, Wang D, Hong JH (1981) Study of nitrogen fixing bacteria associated with rice root. Acta Microbial Sinica. 21:468–72.

Rabus R, Jack DL, Kelly DJ, Saier MH Jr (1999) TRAP transporters: an ancient family of extracytoplasmic solute-receptor-dependent secondary active transporters. Microbiology. 145:3431–45.

Ramos JL, Stolz A, Reineke W, Timmis KN (1986) Altered effector specificities in regulators of gene expression: TOL plasmid *xylS* mutants and their use to engineer expansion of the range of aromatics degraded by bacteria. Proc Natl Acad Sci U S A. 83:8467–71.

Ramos JL, Duque E, Godoy P, Segura A (1998) Efflux pumps involved in toluene tolerance in *Pseudomonas putida* DOT-T1E. J Bacteriol. 180:3323–9.

Ramos JL, Sol Cuenca M, Molina-Santiago C, Segura A, Duque E, Gómez-García MR, Udaondo Z, Roca A (2015) Mechanisms of solvent resistance mediated by interplay of cellular factors in *Pseudomonas putida*. FEMS Microbiol Rev. 39:555–66.

Recny MA, Hager LP (1982) Reconstitution of native *Escherichia coli* pyruvate oxidase from apoenzyme monomers and FAD. J Biol Chem. 257:12878–86.

Reddy TB, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos EA, Kyrpides NC (2015) The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta) genome project classification. Nucleic Acids Res. 43:D1099–106.

Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 106:19126–31.

Rojas A, Duque E, Mosqueda G, Golden G, Hurtado A, Ramos JL, Segura A (2001) Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. J Bacteriol. 183:3967–73.

Roma-Rodrigues C, Santos PM, Benndorf D, Rapp E, Sá-Correia I (2010) Response of *Pseudomonas putida* KT2440 to phenol at the level of membrane proteome. J Proteomics. 73:1461–78.

Rösch V, Denger K, Schleheck D, Smits TH, Cook AM (2008) Different bacterial strategies to degrade taurocholate. Arch Microbiol. 190:11–8.

Rosselló R, García-Valdés E, Lalucat J, Ursing J (1991) Genotypic and phenotypic diversity of *Pseudomonas stutzeri*. System Appl Microbiol. 14:150–7.

Rosselló-Mora RA, Lalucat J, García-Valdés E (1994) Comparative biochemical and genetic analysis of naphthalene degradation among *Pseudomonas stutzeri* strains. Appl Environ Microbiol. 60:966–72.

Rosselló-Móra R, Amann R (2015) Past and future species definitions for *Bacteria* and *Archaea*. Syst Appl Microbiol. 38:209–16.

Rost B (2002) Enzyme function less conserved than anticipated. J Mol Biol. 318:595–608.

Ruby EG, Wirsen CO, Jannasch HW (1981) Chemolithotrophic sulfur-oxidizing bacteria from the Galapagos rift hydrothermal vents. Appl Environ Microbiol. 42:317–24.

Ruiz R, Ramos JL (2002) Residues 137 and 153 at the N terminus of the XylS protein influence the effector profile of this transcriptional regulator and the sigma factor used by RNA polymerase to stimulate transcription from its cognate promoter. J Biol Chem. 277:7282–6.

Ruíz R, Marqués S, Ramos JL (2003) Leucines 193 and 194 at the N-terminal domain of the XylS protein, the positive transcriptional regulator of the TOL *meta*-cleavage pathway, are involved in dimerization. J Bacteriol. 185:3036–41.

Saier MH Jr, Reddy VS, Tamang DG, Västermark A (2014) The transporter classification database. Nucleic Acids Res. 42:D251–8.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4:406–25.

Salinero KK, Keller K, Feil WS, Feil H, Trong S, Di Bartolo G, Lapidus A (2009) Metabolic analysis of the soil microbe *Dechloromonas aromatica* str. RCB: indications of a surprisingly complex life-style and cryptic anaerobic pathways for aromatic degradation. BMC Genomics.10:351.

Sambrook J, Russell DW (2001) Molecular cloning, a laboratory manual. 3rd edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

Satola B, Wübbeler JH, Steinbüchel A (2013) Metabolic characteristics of the species *Variovorax paradoxus*. Appl Microbiol Biotechnol. 97:541–60.

Schell MA, Wender PE (1986) Identification of the *nahR* gene product and nucleotide sequences required for its activation of the *sal* operon. J Bacteriol. 166:9–14.

Schoen C, Blom J, Claus H, Schramm-Glück A, Brandt P, Müller T, Goesmann A, Joseph B, Konietzny S, Kurzai O, Schmitt C, Friedrich T, Linke B, Vogel U, Frosch M (2008) Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*. Proc Natl Acad Sci U S A. 105:3473–8.

Scotta C, Mulet M, Sánchez D, Gomila M, Ramírez A, Bennasar A, García-Valdés E, Holmes B, Lalucat J (2012) Identification and genomovar assignation of clinical strains of *Pseudomonas stutzeri*. Eur J Clin Microbiol Infect Dis. 31:2133–9.

Scotta C, Gomila M, Mulet M, Lalucat J, García-Valdés E (2013) Whole-cell MALDI-TOF mass spectrometry and multilocus sequence analysis in the discrimination of *Pseudomonas stutzeri* populations: three novel genomovars. Microb Ecol. 66:522–32.

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30:2068–9.

Segura A, Godoy P, van Dillewijn P, Hurtado A, Arroyo N, Santacruz S, Ramos JL (2005) Proteomic analysis reveals the participation of energy- and stress-related proteins in the response of *Pseudomonas putida* DOT-T1E to toluene. J Bacteriol. 187:5937–45.

Semple KT, Cain RB (1996) Biodegradation of phenols by the alga *Ochromonas danica*. Appl Environ Microbiol. 62:1265–73.

Sharanagouda U, Karegoudar TB (2001) Degradation of 2-methylnaphthalene by *Pseudomonas* sp. strain NGK1. Curr Microbiol. 43:440–3.

Shingler V, Powlowski J, Marklund U (1992) Nucleotide sequence and functional analysis of the complete phenol/3,4-dimethylphenol catabolic pathway of *Pseudomonas* sp. strain CF600. J Bacteriol. 174:711–24.

Shintani M, Yoshida T, Habe H, Omori T, Nojiri H (2005) Large plasmid pCAR2 and class II transposon Tn4676 are functional mobile genetic elements to distribute the carbazole/dioxin-degradative *car* gene cluster in different bacteria. Appl Microbiol Biotechnol. 67:370–82.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34:D32–6.

Siguier P, Gourbeyre E, Chandler M (2014) Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev. 38:865–91.

Sikorski J, Lalucat J, Wackernagel W (2005) Genomovars 11 to 18 of *Pseudomonas stutzeri*, identified among isolates from soil and marine sediment. Int J Syst Evol Microbiol. 55:1767–70.

Silby MW, Winstanley C, Godfrey SA, Levy SB, Jackson RW (2011) *Pseudomonas* genomes: diverse and adaptable. FEMS Microbiol Rev. 35:652–80.

Simon O, Klebensberger J, Mükschel B, Klaiber I, Graf N, Altenbuchner J, Huber A, Hauer B, Pfannstiel J (2015) Analysis of the molecular response of *Pseudomonas putida* KT2440 to the next-generation biofuel n-butanol. J Proteomics. 122:11–25.

Smith BA, Dougherty KM, Baltrus DA (2014) Complete genome sequence of the highly transformable *Pseudomonas stutzeri* strain 28a24. Genome Announc. 2:e00543–14.

Smith DJ, Martin VJ, Mohn WW (2004) A cytochrome P450 involved in the metabolism of abietane diterpenoids by *Pseudomonas abietaniphila* BKME-9. J Bacteriol. 186:3631–9.

Smith DJ, Park J, Tiedje JM, Mohn WW (2007) A large gene cluster in *Burkholderia xenovorans* encoding abietane diterpenoid catabolism. J Bacteriol. 189:6195–204.

So CM, Young LY (1999) Isolation and characterization of a sulfate-reducing bacterium that anaerobically degrades alkanes. Appl Environ Microbiol. 65:2969-76.

Sobecky PA, Hazen TH (2009) Horizontal gene transfer and mobile genetic elements in marine systems. Methods Mol Biol. 532:435–53.

Sota M, Yano H, Ono A, Miyazaki R, Ishii H, Genka H, Top EM, Tsuda M (2006) Genomic and functional analysis of the IncP-9 naphthalene-catabolic plasmid NAH7 and its transposon Tn*4655* suggests catabolic gene spread by a tyrosine recombinase. J Bacteriol. 188:4057–67.

Steen H, Mann M (2004) The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol. 5:699–711.

Stringfellow WT, Aitken MD (1994) Comparative physiology of phenanthrene degradation by two dissimilar pseudomonads isolated from a creosote-contaminated soil. Can J Microbiol. 40:432–8.

Stringfellow WT, Aitken MD (1995) Competitive metabolism of naphthalene, methylnaphthalenes, and fluorene by phenanthrene-degrading pseudomonads. Appl Environ Microbiol. 61:357–62.

Strong LC, Rosendahl C, Johnson G, Sadowsky MJ, Wackett LP (2002) *Arthrobacter aurescens* TC1 metabolizes diverse s-triazine ring compounds. Appl Environ Microbiol. 68:5973–80.

Sverzhinsky A, Chung JW, Deme JC, Fabre L, Levey KT, Plesa M, Carter DM, Lypaczewski P, Coulton JW (2015). Membrane protein complex $ExbB_4$-$ExbD_1$-$TonB_1$ from *Escherichia coli* demonstrates conformational plasticity. J Bacteriol. 197:1873–85

Takahata Y, Nishijima M, Hoaki T, Maruyama T (2001) *Thermotoga petrophila* sp. nov. and *Thermotoga naphthophila* sp. nov., two hyperthermophilic bacteria from the Kubiki oil reservoir in Niigata, Japan. Int J Syst Evol Microbiol. 51:1901–9.

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 56:564–77.

Tamegai H, Li L, Masui N, Kato C (1997) A denitrifying bacterium from the deep sea at 11,000-m depth. Extremophiles. 1:207–11.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30:2725–9.

Tan HM (1999) Bacterial catabolic transposons. Appl Microbiol Biotechnol. 51:1–12.

Taylor LE 2nd, Henrissat B, Coutinho PM, Ekborg NA, Hutcheson SW, Weiner RM (2006) Complete cellulase system in the marine bacterium *Saccharophagus degradans* strain 2-40^T. J Bacteriol. 188:3849–61.

Teeling H, Meyerdierks A, Bauer M, Amann R, Glöckner FO (2004) Application of tetranucleotide frequencies for the assignment of genomic fragments. Environ Microbiol. 6:938–47.

Tehrani R, Lyv MM, Van Aken B (2014) Transformation of hydroxylated derivatives of 2,5-dichlorobiphenyl and 2,4,6-trichlorobiphenyl by *Burkholderia xenovorans* LB400. Environ Sci Pollut Res Int. 21:6346–53.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A. 102:13950–5.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–82.

Tian W, Skolnick J (2003) How well is enzyme function conserved as a function of pairwise sequence identity? J Mol Biol. 333:863–82.

Top EM, Springael D (2003) The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. Curr Opin Biotechnol. 14:262–9.

Toyofuku M, Roschitzki B, Riedel K, Eberl L (2012) Identification of proteins associated with the *Pseudomonas aeruginosa* biofilm extracellular matrix. J Proteome Res. 11:4906–15.

Trautwein K, Grundmann O, Wöhlbrand L, Eberlein C, Boll M, Rabus R (2012) Benzoate mediates repression of C_4-dicarboxylate utilization in "*Aromatoleum aromaticum*" EbN1. J Bacteriol. 194:518–28.

Tsai SC, Tsai LD, Li YK (2005) An isolated *Candida albicans* TL3 capable of degrading phenol at large concentration. Biosci Biotechnol Biochem. 69:2358–67.

Udaondo Z, Molina L, Segura A, Duque E, Ramos JL (2015) Analysis of the core genome and pangenome of *Pseudomonas putida*. Environ Microbiol. *In press*. DOI: 10.1111/1462–2920.13015.

Urata M, Miyakoshi M, Kai S, Maeda K, Habe H, Omori T, Yamane H, Nojiri H (2004) Transcriptional regulation of the *ant* operon, encoding two-component anthranilate 1,2-dioxygenase, on the carbazole-degradative plasmid pCAR1 of *Pseudomonas resinovorans* strain CA10. J Bacteriol. 186:6815–23.

Unell M, Nordin K, Jernberg C, Stenström J, Jansson JK (2008) Degradation of mixtures of phenolic compounds by *Arthrobacter chlorophenolicus* A6. Biodegradation. 19:495–505.

Vaillancourt FH, Bolin JT, Eltis LD (2006) The ins and outs of ring-cleaving dioxygenases. Crit Rev Biochem Mol Biol. 41:241–67.

van den Berg B, Black PN, Clemons WM Jr, Rapoport TA (2004) Crystal structure of the long-chain fatty acid transporter FadL. Science. 304:1506–9.

van Niel CB,Allen MB (1952) A note on *Pseudomonas stutzeri*. J Bacteriol. 64:413–22.

VerBerkmoes NC, Connelly HM, Pan C, Hettich RL (2004) Mass spectrometric approaches for characterizing bacterial proteomes. Expert Rev Proteomics. 1:433–47.

Vermeiren H, Willems A, Schoofs G, de Mot R, Keijers V, Hai W, Vanderleyden J (1999) The rice inoculant strain *Alcaligenes faecalis* A15 is a nitrogen-fixing *Pseudomonas stutzeri*. Syst Appl Microbiol. 22:215–24.

Vesth T, Wassenaar TM, Hallin PF, Snipen L, Lagesen K, Ussery DW (2010) On the origins of a *Vibrio* species. Microb Ecol. 59:1–13.

Volodina E, Raberg M, Steinbüchel A (2015) Engineering the heterotrophic carbon sources utilization range of *Ralstonia eutropha* H16 for applications in biotechnology. Crit Rev Biotechnol. 2:1–14.

Wang Y, Rawlings M, Gibson DT, Labbé D, Bergeron H, Brousseau R, Lau PC (1995) Identification of a membrane protein and a truncated LysR-type regulator associated with the toluene degradation pathway in *Pseudomonas putida* F1. Mol Gen Genet. 246:570–9.

Wasi S, Tabrez S, Ahmad M (2013) Use of *Pseudomonas* spp. for the bioremediation of environmental pollutants: a review. Environ Monit Assess. 185:8147–55.

Weyens N, Truyens S, Dupae J, Newman L, Taghavi S, van der Lelie D, Carleer R, Vangronsveld J (2010) Potential of the TCE-degrading endophyte *Pseudomonas putida* W619-TCE to improve plant growth and reduce TCE phytotoxicity and evapotranspiration in poplar cuttings. Environ Pollut. 158:2915–9.

Wijte D, van Baar BL, Heck AJ, Altelaar AF (2010) Probing the proteome response to toluene exposure in the solvent tolerant *Pseudomonas putida* S12. J Proteome Res. 10:394–403.

Willenbrock H, Hallin PF, Wassenaar TM, Ussery DW (2007) Characterization of probiotic *Escherichia coli* isolates with a novel pan-genome microarray. Genome Biol. 8:R267.

Williams FR, Hager LP (1966) Crystalline flavin pyruvate oxidase from *Escherichia coli*. I. Isolation and properties of the flavoprotein. Arch Biochem Biophys. 116:168–76.

Williams PA, Catterall FA, Murray K (1975) Metabolism of naphthalene, 2-methylnaphthalene, salicylate, and benzoate by *Pseudomonas* PG: regulation of tangential pathways. J Bacteriol. 124:679–85.

Wong K, Ma J, Rothnie A, Biggin PC, Kerr ID (2014) Towards understanding promiscuity in multidrug efflux pumps. Trends Biochem Sci. 39:8–16.

Wu X, Monchy S, Taghavi S, Zhu W, Ramos J, van der Lelie D (2011) Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. FEMS Microbiol Rev. 35:299–323.

Wyndham RC, Cashore AE, Nakatsu CH, Peel MC (1994) Catabolic transposons. Biodegradation. 5:323–42.

Yamamoto S, Kasai H, Arnold DL, Jackson RW, Vivian A, Harayama S (2000) Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. Microbiology. 146:2385–94.

Yan Y, Yang J, Chen L, Yang F, Dong J, Xue Y, Xu X, Zhu Y, Yao Z, Lin M, Wang Y, Jin Q (2005) Structural and functional analysis of denitrification genes in *Pseudomonas stutzeri* A1501. Sci China C Life Sci. 48:585–92.

Yan Y, Yang J, Dou Y, Chen M, Ping S, Peng J, Lu W, Zhang W, Yao Z, Li H, Liu W, He S, Geng L, Zhang X, Yang F, Yu H, Zhan Y, Li D, Lin Z, Wang Y, Elmerich C, Lin M, Jin Q (2008) Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501. Proc Natl Acad Sci U S A. 105:7564–9.

Yu H, Yuan M, Lu W, Yang J, Dai S, Li Q, Yang Z, Dong J, Sun L, Deng Z, Zhang W, Chen M, Ping S, Han Y, Zhan Y, Yan Y, Jin Q, Lin M (2011) Complete genome sequence of the nitrogen-fixing and rhizosphere-associated bacterium *Pseudomonas stutzeri* strain DSM 4166. J Bacteriol. 193:3422–3.

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–9.

Zhao H, Chen D, Li Y, Cai B (2005) Overexpression, purification and characterization of a new salicylate hydroxylase from naphthalene-degrading *Pseudomonas* sp. strain ND6. Microbiol Res. 160:307–13.

Zobell CE, Upham HC (1944) A list of marine bacteria including description of sixty species. Bull Scripps Inst Oceanogr. 5:239–92.

Zumft WG (1997) Cell biology and molecular basis of denitrification. Microbiol Mol Biol Rev. 61:533–616.

Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP (2006) Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. J Proteome Res. 5:2339–47.

Zylstra GJ, McCombie WR, Gibson DT, Finette BA (1988) Toluene degradation by *Pseudomonas putida* F1: genetic organization of the *tod* operon. Appl Environ Microbiol. 54:1498–503

SUPPLEMENTARY INFORMATION

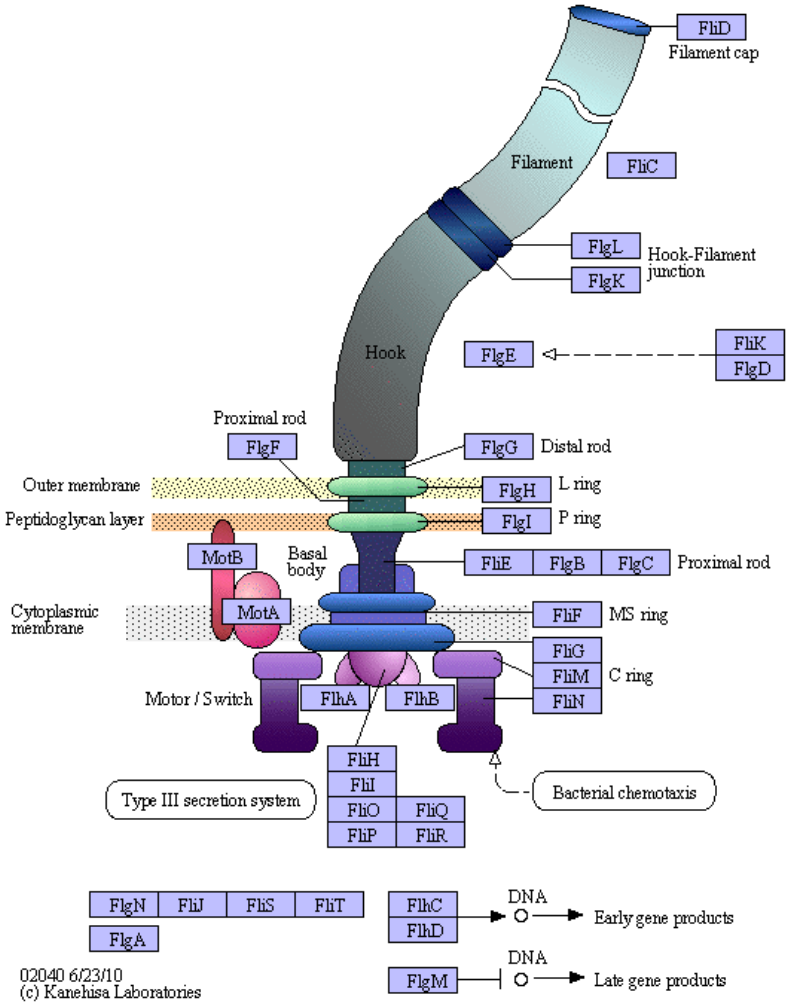

Figure S1. Scheme of proteins involved in the flagellar assembly system (source KEGG).

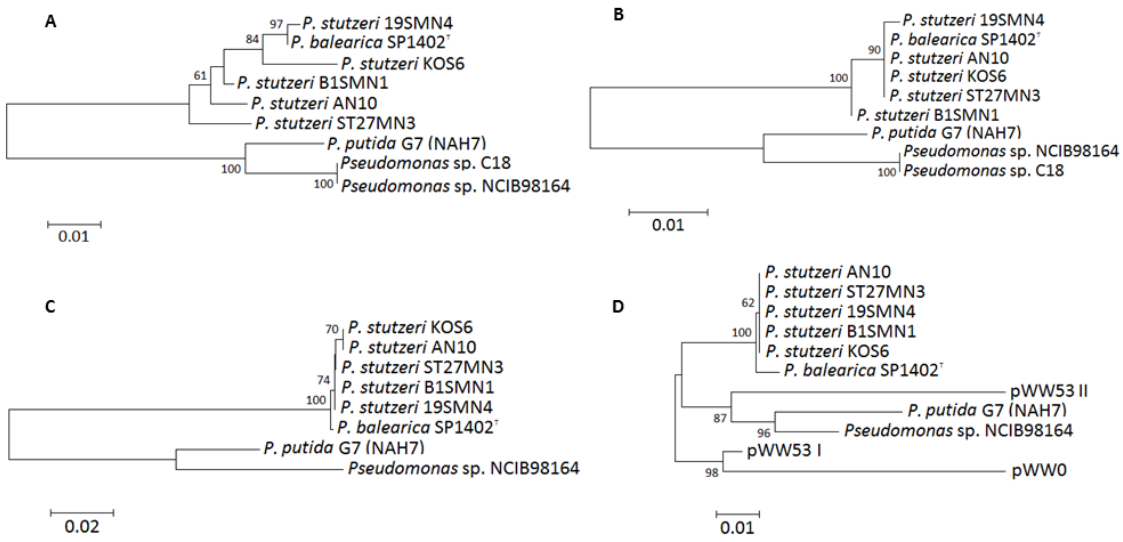

Figure S2. Phylogeny of 4 different enzymes involved in the degradation of naphthalene: A, NahAc; B, NahF; C, NahG; D, NahI. Bootstrap numbers below 60 were omitted.

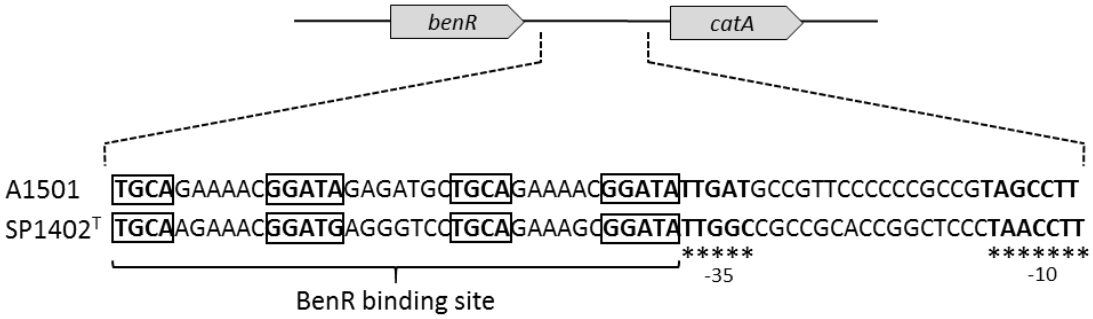

Figure S3. Structure of the BenR binding site in *P. stutzeri* A1501 (Li *et al*., 2010) and *P. balearica* SP1402^T. The BenR binding site sequences are indicated in boxes, and the -10/-35 promoter consensus sequences are indicated by asterisks.

All supplementary tables are attached together with this document in Excel format.

Table S1. IS database generated in this study with TnpAs belonging to 21 different IS-families from the IS-Finder database (https://www-is.biotoul.fr/) (Siguier *et al.*, 2006).

Table S2. CDSs annotated in the 19 studied *P. stutzeri* and *P. balearica* genomes, with their corresponding localization in the genome. The classification into COGs using two different criteria (at least 50 % of identity in at least the 50 % of the amino acidic sequence, and at least 95 % of identity in at least the 90 % of the amino acidic sequence) is also shown.

Table S3. KEGG biological function classification of all CDSs annotated in the 19 studied *P. stutzeri* and *P. balearica* genomes.

Table S4. CDSs annotated as putative catalase, cytochrome *c* oxidase, arginine deiminase, and glycogen hydrolase in the 19 studied *P. stutzeri* and *P. balearica* genomes.

Table S5. CDSs identified as involved in the structure and formation of the flagellum in the 19 studied *P. stutzeri* and *P. balearica* genomes.

Table S6. CDSs identified as involved in denitrification in the 19 studied *P. stutzeri* and *P. balearica* genomes.

Table S7. CDSs identified as involved in nitrogen fixation in the 19 studied *P. stutzeri* and *P. balearica* genomes.

Table S8. CDSs from the 69 COGs which presented one CDS of each *P. stutzeri* and *P. balearica* genome, using the criteria of at least 95 % of identity in at least the 90 % of the amino acidic sequence.

Table S9. Core-proteome of *P. stutzeri*, using the criteria of at least 50 % of identity in at least the 50 % of the amino acidic sequence.

Table S10. Strain specific CDSs using the criteria of at least 50 % of identity in at least the 50 % of the amino acidic sequence.

Table S11. CDSs identified as involved in aromatic hydrocarbon degradation pathways in the 19 studied *P. stutzeri* and *P. balearica* genomes. CDSs identified as involved in dehydroabietic acid degradation pathway are also shown.

Table S12. Proteins detected in cultures of strains B1SMN1, 19SMN4, ST27MN3, AN10, and SP1402^T growing with succinate, succinate with a salicylate pulse, and naphthalene; all in triplicate. Abundance fold changes between succinate and naphthalene cultures and between succinate and salicylate cultures, with their respective p-values, are shown. NAF (F), Normalized Abundance (N), and Raw Abundance (R) obtained are also shown.

Table S13. Transport systems presumably involved in naphthalene metabolism. Normalized abundance fold changes and their respective p-values are shown.

Table S14. Proteins detected in cultures of AN10 growing with succinate and BZ4D growing with succinate and benzoate, all in triplicate. Abundance fold changes between BZ4D succinate and benzoate cultures and between AN10 and BZ4D succinate cultures, with their respective p-values, are shown. Spectral counts (SC) and Normalized Spectral Abundance Factors (NSAF) obtained are also shown.