Consejo Superior de Investigaciones Científicas

Estación Experimental del Zaidín

Abengoa Research S.L.

Área de Biotecnología

ABENGOA
RESEARCH

Universidad de Granada

Programa de Doctorado en Bioquímica y Biología Molecular

Analysis of the pangenome of *Pseudomonas putida*

Tesis Doctoral

Zulema Udaondo Domínguez

2016

Editor: Universidad de Granada. Tesis Doctorales
Autora: Zulema Udaondo Domínguez
ISBN: 978-84-9125-919-0
URI: http://hdl.handle.net/10481/43836

Análisis del Pangenoma de *Pseudomonas putida*

Memoria que presenta la Licenciada en Biología

Zulema Udaondo Domínguez

para aspirar al Título de Doctor

Fdo.: Zulema Udaondo Domínguez

VºBº del Director	VºBº del Director
Fdo.: Juan Luis Ramos Martín	Fdo.: Estrella Duque Martín de Oliva
Doctor en Biología	Doctora en Biología
Profesor de Investigación del CSIC y Director del Área de Biotecnología en Abengoa Research	Científico Titular del CSIC e Investigadora Senior en Abengoa Research

Abengoa Research S.L./EEZ-CSIC/Universidad de Granada

2016

La parte del trabajo de esta Tesis Doctoral que se realizó en el grupo de Degradación de Tóxicos Orgánicos del Departamento de Protección Ambiental de la Estación Experimental del Zaidín (Consejo Superior de Investigaciones Científicas) Granada, fue financiada por los siguientes proyectos:

- Análisis post-genómico functional de *Pseudomonas putida* KT2440.
  Plan Nacional de I+D+i: BIO-2010-17227.2011-2014
- Bases moleculares de la degradación y resistencia a tolueno.
  Proyecto de Excelencia de la Junta de Andalucía y fondos FEDER: CVI-7391. 2012-2015

La parte del trabajo realizado en el Área de Biotecnología de Abengoa Research S.L. (Campus Palmas Altas) Sevilla, fue financiado por los programas:

- Empower: Re-programming the lifestile of *Pseudomonas putida* for bespoke biocatalysis. European Union horizon PP research and innovation programme under grant agreement Nº 635536.
- Metafluidics. European Union horizon PP research and innovation programme under grant agreement Nº 685474
- Biorefineríes. Abengoa Research Project. (Abengoa Research S.L.)

La doctorando **Zulema Udaondo Domínguez**, y los directores de la tesis **Juan Luis Ramos Martín** y **Estrella Duque Martín de Oliva**, garantizamos, al firmar la tesis doctoral, que el trabajo ha sido realizado por la doctorando bajo la dirección de los directores de tesis y hasta nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados y publicaciones.

Sevilla, 1 de Febrero del 2016

Director de la Tesis

Doctorando

Fdo.: Juan Luis Ramos Martín

Fdo.: Estrella Duque Martín de Oliva

Fdo.: Zulema Udaondo Domínguez

“If I have seen further, it is by standing

on the shoulders of giants”

Isaac Newton

A mi genial familia,

A Sergio

Agradecimientos

Después de casi 5 años de intenso crecimiento tanto profesional como personal no puedo mas que recordar a todas las personas que han hecho que hoy pueda estar escribiendo estos agradecimientos, como broche final a todo el esfuerzo plasmado en las páginas que vienen a continuación. ¡¡Y la lista es larga!!

En primer lugar, quiero agradecer a mi tutor y director de tesis, el Prof. Juan Luis Ramos Martín, por darme la oportunidad de realizar la tesis en su grupo, por todo el trabajo dedicado y todos los conocimientos que me ha regalado durante todos estos años. Muchísimas gracias por el trato tan cercano y por facilitarme tanto las cosas, pero sobre todo por la confianza en mi trabajo que siempre me has demostrado.

A mi codirectora de tesis, la Dra. Estrella Duque Martín de Oliva. Estrella, muchísimas gracias por tu apoyo, por tu esfuerzo y por tus conocimientos, pero sin duda alguna, gracias por haber sido tan amable conmigo durante todos estos años, por todo tu cariño, tus risas y por todo el ánimo que me has dado siempre; siempre has tenido palabras positivas hacia mí y hacia mi trabajo y te lo agradezco enormemente.

También me gustaría agradecer al resto de jefes del grupo de Degradación de Tóxicos Orgánicos de la EEZ: Silvia Marqués (actual jefa de grupo), Pieter Van Dillewijn, Tino Krell, Manuel Espinosa, Regina Wittich, Maribel Ramos, Marian Llamas, Ana María Fernández, muchas gracias a todos por vuestro buen trato, por vuestra ayuda y por vuestros consejos.

Al Prof. José Luis García, y al Dr. Eduardo Díaz, muchísimas gracias por acogerme, por enseñarme y por darme la oportunidad de comenzar mi andadura por el mundo de la bioinformática con vosotros. A mis compañeros del CIB, Tere, Manuel, Bea, Ana, Valle, Ife, Javier, Carlos, Zaira, Virginia, Lorein, Carmen, Miriam… muchísimas gracias por los meses que pasé con vosotros, me hicisteis sentir una más, y me llevé muchos buenos recuerdos... sobre todo... ¡de la fiesta de primavera! Muchas gracias en especial a la Dra. Laura Isabel de Eugenio, por todo el apoyo y las conversaciones sobre música, pelis, videojuegos... el mundo del frikismo en general y lo mejor, ¡¡gatitoosss!! ¡Eres genial Lau!

Al Dr. Manuel J. Gómez, muchísimas gracias por todas las horas dedicadas en mi estancia en el CIB enseñándome los "entresijos" de Pathway Tools, y por todas las charlas sobre bioinformática y ciencia en general.

También quiero agradecer a todos los profesores que desde niña me han ayudado e inculcado el amor por la ciencia y por el conocimiento en general. En especial a mi profesor del cole, Don Jesús, que además fue mi primer profesor de guitarra, al que siempre he recordado con muchísimo cariño. A mi profesor de Biología del Instituto, Andrés Ajo, porque sus clases eran las mejores del instituto y siempre se esforzaba por sacarnos de excursión a lugares increíbles que siempre recordaré, como Búbal. A mi querido profesor de la UAM el Dr. Máximo Sandín. Máximo, podría dedicarte una página entera. Muchísimas gracias por todas las conversaciones en las que siempre

me llevaba trabajo para pensar en casa, por tus clases, de las pocas en las que se discutían artículos científicos; por enseñarme que nunca hay que conformarse con lo que uno sabe o cree saber… eres un ejemplo para mi, muchas gracias por todo.

A mis compañeros de BioIliberis R&D, mi primera parada en mis más de dos años y medio en Granada. Muchísimas gracias a Mati, Amalia, Jose Luis, Jennifer, Cristina y Silvia, por acogerme y por ayudarme en mis primeros meses en Granada, en especial a Paloma y a Germán, por vuestra amistad, todas las conversaciones sobre series y pelis y las tardes de cañas y de tés. Paloma quiero que sepas que te admiro muchísimo y que eres una persona súper especial, ¡muchas gracias por apoyarme siempre!

A todos mis compañeros de mi grupo en AR que habéis hecho las jornadas interminables mucho más llevaderas y por los buenos y "dulces" momentos en los que nos juntábamos todos en el "vending" y celebraciones varias. A Jacobo, Jonás, Toni y Dani, por acogerme, por vuestra ayuda y amistad en mis primeros meses en AR, en el edificio G, los más difíciles quizás, sin contar con la recta final. Al resto de mis compis de AR, constituyentes del multifacético, multidisciplinar, multisensorial grupo... "Martes de Cañas", por hacer el día a día mucho más ameno, por las charlas multidisciplinares, por dejarme sin batería el móvil en dos horas y… ¡porque sois todos geniales! muchas gracias por todo, os deseo todo lo mejor.

A mis bonitis, mis compañeras y amigas en esta última etapa en AR. A Laura, gracias por todos los "scagliette" jajaj, por las risas y por ser tan buena compañera. A Sol, María, Lorena y Bea, gracias por todo vuestro apoyo en todos los momentos, buenos y malos, por todo vuestro cariño y las risas diarias, por hacer más fácil el madrugón y el venir a trabajar, por los momentos de Biodanza (jajaj), pero sobre todo por vuestra amistad, sois lo mejor que me llevo de AR ¡Muchas gracias por todo!

A mis amigos de la facultad de Biología de la UAM. Los mejores compañeros de clase que se puede tener. Gracias por todo vuestro apoyo y amistad durante todos estos años, en especial a Mónica, Sandra, Iker, Natalia, Paula, Peter, Noelia, Edu, Muri, Borja, David, Aida, Clara, Gema, Guille y Marta. ¡Sois los mejores Biologueños, os quiero!

A mis compañeros del Máster de Bionformática de la UCM, Clara, Kris y Álvaro. Muchas gracias por vuestra amistad todos estos años, por vuestras visitas en las que siempre hay momentos para recordar, por todas las risas y por vuestra ayuda durante el máster en mis comienzos con la bioinformática, espero que sigamos teniendo momentos que compartir, ¡os echo de menos!

A mis compis de la EEZ. Muchas gracias a Lázaro, Verónica, Patri B., Óscar, Sophie, Dani, Alejo, Alicia, Patri M., Andrés, Josemi, Cristina, Miriam, Georg, Ana María… Gracias por todos los buenos momentos en la EEZ, me llevo buenísimos recuerdos y son gracias a todos vosotros. En especial a María, mi zubiética. María te echo de menos, tus lecciones sobre zubiético no tienen parangón, espero que nos veamos pronto y que vengas a visitarme de verdad mujeerr! Te deseo todo lo mejor.

A mi grupo de desayuno de la EEZ, los mejores!!! jejej César, Javi, Inés, Nené, Jesús y Adela. Gracias por hacer del desayuno uno de los momentos más divertidos del día!! Os echo de menos y no solo por las tostadas de la cafetería de la EEZ, si no también por las cañas de los viernes!! A Adela, mi primera amiga en la EEZ, muchas gracias por todas las risas y los ratos al sol cuando comíamos detrás de la casa blanca, hablando de todo un poco, te deseo todo lo mejor! A César, por las charlas sobre cosas científicas y no tanto jeje, eres súper genial muchas gracias por todo. A Javi, por estar siempre dispuesto a echarme una mano cada vez que he tenido un problema, ¡muchísimas gracias! A Nené y a Jesús, por tener siempre un ratillo para mi cuando me pasaba a veros por el laboratorio, por hacerme sentir una más del grupo, ¡muchas gracias! A Inés, por compartir mi cariño por los gatetes y por las charlas sobre gimnasia rítmica y viajes, ¡muchas gracias a todos!

A Mai y a Maria Angustias, por ser tan amables siempre y por vuestras palabras de cariño, por estar siempre ahí cuando necesitaba cualquier cosa y por toda vuestra ayuda ¡muchas gracias!

A Mari Ángeles, la alegría de todas las tardes en la EEZ, muchas gracias por tus charlas, por tu buen humor que siempre nos contagiabas ¡y por dejarlo todo tan limpito! A Pedro de seguridad, por tu amabilidad y por tus charlas sobre minerales.

Mirando atrás me doy cuenta de los incontables buenos momentos que viví en Granada y la mayoría de ellos fueron gracias a tres personas. A Marta, mi Paris, jejej, Marta, muchas gracias por todos los momentos en Granada, sobre todo porque contigo compartí una de las mejores cosas que conocí allí que fueron nuestros momentos en la pisci y los jueves de Diamantes!! A Saray, gracias por todos los jueves, los viernes… los sábados… los domingos.. mmm no, ¡¡no había más fuelle para los domingos!!jeje Gracias por tu amistad y por tu apoyo, por todos los buenísimos momentos, ¡¡sin ti no habría sido lo mismo!! Te escribiría tres páginas, ¡te mereces todo lo mejor! A Carlos, uno de los mayores sufridores de esta tesis (jeje). Carlos muchas gracias por ser tan buen amigo, no sabes lo mucho que me has ayudado durante estos casi 5 años, gracias por aguantarme en Granada y por ser uno de mis mayores apoyos aquí en Sevilla. Gracias por todas las charlas de series y videojuegos... aunque a veces te haya hecho un poco de spoiler sin querer… ¡te has ganado el cielo!

A Ali, muchas gracias por estos casi 5 años de amistad. Muchísimas gracias por todos tus consejos, por tener siempre palabras positivas hacia mí y hacia mi trabajo, por apoyarme ¡¡y por todos los vinillos!!

A Ana Segura, por todos los consejos y por ayudarme siempre. Muchas gracias por estos meses en Sevilla en los que te he podido conocer mejor y por tener siempre un rato para escucharme, creo que eres una gran profesional y mejor persona ¡te deseo todo lo mejor!

A mis amigos “del local”, Carmen, Juancar, Botello, Pol, Aitor, Juan, Cris, Ricar, Negro… por todos los momentos compartidos, son ya muchos años y no puedo si

no alegrarme de poder seguir compartiendo mi tiempo con gente tan buena a mi lado. Negrito gracias por todo lo anterior y además por los momentos Destiny!!y porque tus cumples son los mejores!

A mi familia más reciente, Reme, Pablo, Mercedes, Miguel, Estefa, Paula, Anca y Carlos. Gracias por hacerme sentir una más de la familia desde el primer momento y por todos los momentos tan divertidos que he disfrutado con vosotros. En especial a Reme por su cariño y por su apoyo desde el primer momento y a Paula, la alegría de la familia, por todas las sonrisas regaladas a cientos de kilómetros de distancia ¡Muchas gracias por todo!

A mis vecinos de Bellavista, Rafael y Chari, por cuidar de Sergio y de mí desde que llegamos a Sevilla. Muchas gracias por vuestra amistad y por vuestra enorme generosidad; por las cenas tan ricas y en tan buena compañía, por hacernos sentir que también tenemos familia en Sevilla ¡Muchísimas gracias!

A mi Tata, por todos los veranos en los que cuidaste de mí en Villafranca, y a mi familia de Villafranca de los Barros por todos los recuerdos inolvidables que tengo de mis veranos en el pueblo rodeada de mis primos y tíos.

A mi tía Gelu, por haber estado siempre ahí, por tus palabras de cariño siempre, por hacerme sentir especial

A mi "Nievis" y a Carlos. Muchas gracias por vuestra amistad después de todos estos años. Nieves, eres un ejemplo de fuerza y de superación para mi, y además ahora de súper mama!! Gracias por todos los momentos compartidos, por cuidarme tanto y por quererme. Gracias por mis dos pinchiclines preciosos, Olivia y Erik que son mis dos tesoretes y me regalan tantas risas.

A Ruth, Rober y Aitana. Por regalarme vuestra amistad durante todos estos años. Ruth, eres como mi hermana, pero sin el como, jeje. Gracias por taaaanntos años de amistad de la buena, gracias por apoyarme siempre tanto y por quererme tal y como soy, y eso que tú me conoces bien. A mi Aitanilla, por ser la primera persona de este mundo a la que quise desde el momento cero en el que supe que existía. No puedo esperar a ver en que maravillosa persona te vas a convertir

A mi familia de Jerusalén 21, 2B. La familia Tortajada Castro. A Manoli, Santiago, Poli, Ana, Susana, Santi, Alba, Jorge y Santi pequeño (que ya no es tan pequeño). Por cuidarme desde que nací, por hacerme entender que la familia también se elige. En especial quiero dedicarle esta tesis a Manoli, por ser una de las personas más sabias de este mundo. Porque tus consejos me acompañarán siempre, y porque te echo de menos.

A las personas que dan sentido a mi vida, a los mejores amigos que se puede tener en este mundo y los que me hacen sentir que soy la persona más afortunada del mundo por teneros a mi lado; mis hermanos, César, Susana, Alex y Kilian. Gracias por apoyarme siempre y por todas las risas que siempre surgen cuando estamos juntos. ¡No puedo estar más orgullosa de vosotros! A mi madre, Merche, por ser mi

ejemplo de mujer luchadora y trabajadora, nunca podré agradecerte todo lo que has hecho por mi, gracias por apoyarme y cuidarme siempre y por hacerme saber que pase lo que pase mi familia siempre estará conmigo. A la benjamina de la familia, Nora, gracias por todas las ilusiones desde el momento cero en el que supe que existías, gracias por hacerme vivir el regalo de ser tita, ¡y por ser la cosita más bonita que he visto en la vida! A Kimbo, por todo tu cariño y los momentos divertidos, porque no hay un ser de cuatro patas más bueno y más cariñoso. Os echo de menos todos los días. No podría tener una familia mejor, esta tesis también es vuestra ¡Os quiero!

Y finalmente, a Sergio. Gracias por regalarme todo tu cariño día tras día, por tu apoyo incondicional, por estar siempre a mi lado, por mucho que llueva; por hacerme sentir la persona más especial del mundo. Gracias por hacer que todo merezca la pena. Esta tesis también es tuya. ¡Te quiero infinito!

CONTENTS

Contents

Figure index ... i
Table Index ... v
List of abbreviations ... vii
RESUMEN ... xi
GENERAL INTRODUCTION ... 1
Background ... 3
1. Bacterial genomes ... 4
2. Pseudomonas putida ... 11
3. Bioinformatics ... 16
4. The pangenome concept ... 31
References ... 33
AIM OF THIS THESIS AND SPECIFIC OBJETIVES ... 53
RESULTS ... 57
Chapter I. Metabolic potential of the organic-solvent tolerant Pseudomonas putida DOT-T1E deduced from its annotated genome ... 59
Summary ... 63
Introduction ... 64
Materials and Methods ... 65
Results and Discussion ... 66
Genome sequencing, assembly, annotation and bioinformatic analysis ... 66
Metabolic potential ... 68
Biotransformation potential ... 80
References ... 85
Supplementary material ... 92
Chapter II. Bioinformatic analysis of the properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1 ... 99
Summary ... 101
Introduction ... 102
Materials and Methods ... 103
Results and Discussion ... 104
Adhesion and colonization of plant roots ... 104
Evidence of plant growth-promoting activities encoded by BIRD-1 ... 107
Concluding remarks ... 116
References ... 117

Supporting information **124**

Chapter III. Antibiotic resistance determinants in a *Pseudomonas putida* strain isolated from a hospital **135**

Summary **137**

Introduction **138**

Materials and Methods **139**

Results and Discussion **141**

Identification of strain HB3267 141

Antibiotic resistance profile of P. putida strain HB3267 142

Gaining insight into antibiotic resistance through whole genome sequencing of HB3267 145

Antibiotic resistance determinants encoded only on the HB3267 chromosome 148

Antibiotic resistance determinants encoded only on the pPC9 plasmid 154

References **157**

Supporting information **164**

Chapter IV.Analysis of the core genome and the pangenome of *Pseudomonas putida* **167**

Summary **169**

Introduction **170**

Materials and Methods **171**

Results and Discussion **174**

Evolutionary relationships 174

Descriptive statistics of OGs 179

The size of Pseudomonas putida's core genome and pangenome 180

Functional analysis of core genes 182

Availability of a mini-Tn5 mutant collection for P. putida KT2440 187

References **190**

Supporting information **198**

Chapter V.Accessory and unique genes of *Pseudomonas putida* pangenome **205**

Summary **207**

Introduction **208**

Materials and Methods **210**

Results and Discussion **211**

Descriptive analysis of the pangenome of P. putida 211

Functional analysis of the accessory genes 212

Functional analysis of the unique genes 216

References **219**

GENERAL DISCUSSION **225**
References **236**
CONCLUSIONS/CONCLUSIONES **243**

Figure index

GENERAL INTRODUCTION

Figure 1..Distribution of genome sizes among prokaryotes. 4

Figure 2. Mechanisms of gene transfer. 6

Figure 3. Scan Electron Microscopy picture of *Pseudomonas putida* KT2440.. 11

Figure 4. Schema indicating ways to identify new processes to improve the bioremediation of contaminated environments using microbes from *Pseudomonas putida* species. 12

Figure 5. Number of prokaryotic genomes sequenced each year and submitted to NCBI...16

Figure 6. Sequencing and assembly scheme 17

RESULTS

Chapter I. Metabolic potential of the organic-solvent tolerant *Pseudomonas putida* DOT-T1E deduced from its annotated genome

Figure 1. 1. *Pseudomonas putida* DOT-T1E chromosome GC skew analysis..........67

Figure 1. 2 Circular genome of *Pseudomonas putida* DOT-T1E.............................68

Figure 1. 3. Distribution of enzyme activities of *P. putida* DOT-T1E. 69

Figure 1. 4. Pathway for utilization of urea as an N source by *P. putida*. 72

Figure 1. 5. Biotransformation of tyrosine by *P. putida* through metabolic blockage or gene recruitment 81

Figure 1. 6. Potential synthesis of different alcohols from keto acids by DOT-T1E.82

Suppl. Figure 1. 1. .. Catabolism of glucose as an example of sugar metabolism based on annotated genes. 92

Suppl. Figure 1. 2. Catabolism of *L*- and *D*-lysine as an example of an amino acid being used as a source of carbon through multiple confluent pathways....92

Suppl. Figure 1. 3. Catabolism of lactic acid by *P. putida* DOT-T1E as an example of the metabolism of an organic acid...93

Suppl. Figure 1. 4. Catabolism of aromatic compounds to a set of central intermediates..94

Suppl. Figure 1. 5. Genetic organization of the genes for the protocatechuete central catabolic pathway. 95
Suppl. Figure 1. 6. Genetic organization of the duplicated *cat* genes located in two different chromosomal regions.................... 95
Suppl. Figure 1. 7. Genes and enzymes for the utilization of methionine as an S source by DOT-T1E.................... 96
Suppl. Figure 1. 8. Set of biotransformation reactions to achieve the synthesis of adipate / 3-oxoadipyl-CoA based on gene content in DOT-T1E strain........ 97

Chapter II. . Bioinformatic analysis of the properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1

Figure 2. 1. Viable *P. putida* KT2440, BIRD-1 and a *treZ* mutant in rhizosphere soil. 106
Figure 2. 2. Root development of corn seeds germinated on agar plates. 108
Figure 2. 3. Solubilization of phosphate and growth of *P. putida* BIRD-1. 109
Figure 2. 4. Synthesis of IAA based on the annotation of the BIRD-1 genome..... 115
Suppl. Figure 2. 1. Growth of *Pseudomonas putida* BIRD-1 with different sources of inorganic phosphorous. 124
Suppl. Figure 2. 2. Pikovskaya test with BIRD-1 and mutants in the Entner–Doudoroff pathway. 125
Suppl. Figure 2. 3. Growth of *P. putida* with phytate. 126
Suppl. Figure 2. 4. Production of IAA by wild-type BIRD-1 and different isogenic mutants.. 127
Suppl. Figure 2. 5. Corn root development in a control seed without bacteria (A), and inoculated seeds with BIRD-1 (B) and a mutant deficient in the acc deaminase gene (C) 128

Chapter III. Antibiotic resistance determinants in a *Pseudomonas putida* strain isolated from a hospital

Figure 3. 1 The pPC9 map. 146
Figure 3. 2. Genetic organization of the pPC9 ‘‘insert.’’ 147

Figure 3. 3. Location of the antibiotic and solvent efflux ttgGHI pumps within a genomic island in the chromosome of strain HB3267. 152
Figure 3. 4. Potential chromosomal determinants for chloramphenicol resistance of HB3267. 154
Suppl. Figure 3. 1. Phylogenetic tree comparing *gyrB* genes of *Pseudomonas strains*. 165
Suppl. Figure 3. 2 Protein alignment of GyrA (A), GyrB (B) and PhoQ (C) 166

Chapter IV. Analysis of the core genome and the pangenome of *Pseudomonas putida*

Figure 4. 1. Maximum likelihood phylogenetic tree of fourteen *Pseudomonas putida* strains with *Escherichia coli* as outgroup strain, based on a multilocus sequence alignment of four housekeeping genes (16S rRNA*, gyrB*, *rpoB* and *rpoD*)..174
Figure 4. 2 Proteome comparison represented by a Blast-matrix performed with the percent of orthologous genes shared between two strains. 175
Figure 4. 3 Pie chart showing the distribution of the three sets of genes present in the pangenome of nine *Pseudomonas putida* strains...................................176
Figure 4. 4. Codon usage of the nine *Pseudomonas putida* strains as calculated using the cusp program from the EMBOSS package. 177
Figure 4. 5. Number of unique genes present in each strain that contribute to the pangenome of *Pseudomonas putida.* 178
Figure 4. 6. Size of the core genome of *Pseudomonas putida*.. 180
Figure 4. 7 Size of the *Pseudomonas putida* pangenome obtained upon adding a new genome 181
Figure 4. 8 Classification of core genes in *Pseudomonas putida*. 182
Figure 4. 9 Representation of the arginine and proline catabolism in *Pseudomonas putida* core genome. 184
Figure 4. 10. Core genome mutants..187
Suppl. Figure 4. 1 Maximum likelihood phylogenetic tree of the core genes from nine *Pseudomonas putida* strains based on a multilocus sequence alignment of 3326 orthologous group of genes.. 198
Suppl. Figure 4. 2. . Flow diagram for glucose metabolism in *Pseudomonas putida*..199

Suppl. Figure 4. 3. Representation of the tricarboxylic acid cycle (black line) and glyoxylate cycle (green line) in *Pseudomonas putida* core genome 200
Suppl. Figure 4. 4. Representation of the GO "Cellular component" core genome mutants.). 201
Suppl. Figure 4. 5. Representation of the GO "Biological process" core genome mutants. 202

Chapter V. Accessory and unique genes of *Pseudomonas putida* pangenome

Figure 5. 1. Pangenome gene distribution of the number of genes across the nine *P. putida* strains of the study. 211
Figure 5. 2. Functional annotation of accessory genes using KEGG Automatic Annotation Server (KAAS). 212
Figure 5. 3. Functional annotation with core and accessory genes GO terms. 215
Figure 5. 4. Circular representation of the genome of eight *P. putida* strains. 216
Figure 5. 5. Funtional analysis of the set of unique genes with GO terms. 217

Table Index

GENERAL INTRODUCTION

Table 1. The genetic code 8

Table 2. Codon usage table of *Pseudomonas putida KT2440* genome. 9

Table 2. Codon usage table of *Pseudomonas putida KT2440* genome. 10

Table 3. Summary of the main sequencing techniques. 20

Table 4. Blast programs summary 26

Table 5. Homology: terms and definitions 27

RESULTS

Chapter I. Metabolic potential of the organic-solvent tolerant *Pseudomonas putida* DOT-T1E deduced from its annotated genome

Table 1. 1. Proteins found in *P. putida* DOT-T1E that are associated with metal resistance and homeostasis……........ 75

Chapter II. . Bioinformatic analysis of the properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1

Table 2. 1. Genes involved in adhesion to biotic and abiotic surfaces in *Pseudomonas putida* strain BIRD-1 deduced from the identification of reciprocal genes in *P. putida* KT2440. 105

Table 2. 2. Physical linkage of the *fecA* gene to *fecI/fecR* regulatory system. 112

Chapter III. Antibiotic resistance determinants in a *Pseudomonas putida* strain isolated from a hospital

Table 3. 1. Antibiogram assay. 143

Table 3. 2. MIC and MBEC assays with different *Pseudomonas putida* strains 144

Suppl. Table. 3. 1. RND efflux pumps in *P. putida* KT2440 and their orthologs in the HB3267 strain. 164

Chapter IV. Analysis of the core genome and the pangenome of *Pseudomonas putida*

Table 4. 1. General features of the nine *Pseudomonas putida* genomes used in the pangenome study 179

Suppl. Table 4. 1. . Complete list of mutated functions in *P. putida* KT2440.. 203

Chapter V. Accessory and unique genes of *Pseudomonas putida* pangenome

Table 5. 1. Summary of the number of unique genes harbored by each strain. 217

GENERAL DISCUSSION

Table 6. PGPR properties found in *Pseudomonas putida* BIRD1 genome........................229

Table 7. Number of clusters of core genes..231

List of abbreviations

ABC	ATP-binding cassette
AOV	Analysis of Variance
API	Analytical profile index
bp	Base pair
C	Carbon
c.f.u.	Colony forming units
CAI	Codon adaptation index
CDS	Coding DNA sequence
COG	Clusters of Orthologous Groups
contig	Short continuous sequence
DDH	DNA-DNA hybridization
Gb	Giga bytes
GC	Guanine-cytosine pair
GO	Gene Ontology
GS	Glutamine synthetase
GTA	Gene transfer agent
HGT	Horizontal gene transfer
ISR	Induced systemic resistence
IT	Information Technology
KB	Kilobase
LB	Luria-Bertani medium
Mb	Million of base pairs
MIC	Minimum inhibitory concentration
MLSA	Multi locus sequence analysis
NCBI	National Center of Biotechnology Information
ORF	Open reading frame
Pfam	Protein families database
PGPR	Plant growth-promoting rhizobacteria
Prk	Related protein sequences (clusters) database
PTS	Phosphotransferase system
RAM	Random access memory

Rfam	RNA Families Database
RND	Resistance nodulation cell division
sRNA	Small ribonucleic acid
TCA	Tricarboxylic acid
TCP	Transmission Control Protocol
UniProt	Universal Protein Resource
XGD	Xenologous gene displacement

RESUMEN

El continuo desarrollo que han sufrido las técnicas de secuenciación masiva en los últimos años permite obtener hoy día millones de secuencias de ADN, en un único evento de secuenciación, en tiempos y costes cada vez más reducidos. Este crecimiento exponencial de datos biológicos conllevó la necesidad de contar con formas cada vez más eficientes, primero de almacenar la información obtenida en las técnicas experimentales de alto rendimiento, y en segundo lugar de procesar dicha información. La bioinformática surgió de la necesidad de almacenar, organizar, analizar y visualizar los datos obtenidos por técnicas biológicas. La aplicación de la bioinformática en el ámbito de la genómica permite además comparar y relacionar la información genética, siendo capaz de obtener respuestas que no son deducibles a partir de resultados experimentales. Estos mismos resultados experimentales han demostrado que incluso genomas pequeños y relativamente simples, como son los genomas bacterianos, tienen una plasticidad y un repertorio génico mucho más amplio del que se consideraba en un principio. Desde que en 1995 se publicó la secuencia completa del genoma de *Haemophilus influenzae,* hasta la fecha, se han secuenciado miles de genomas bacterianos haciendo posible la identificación de genes esenciales que llevan a cabo procesos comunes en todos los seres vivos, tales como la replicación del DNA, sin embargo es mucho más complicado definir que genes son los que determinan una especie bacteriana. Esto es debido a que en algunas especies se descubren genes nuevos con cada secuenciación de un genoma perteneciente a dicha especie.

Una especie bacteriana puede ser definida por el conjunto de todos los genes que conforman la especie o lo que es lo mismo, su pangenoma. El término pangenoma, fue utilizado por primera vez en el año 2005 por Medini. Los estudios pangenómicos surgen de la posibilidad de disponer de múltiples genomas pertenecientes a una misma especie y ofrecen una nueva aproximación a la definición de especie. El estudio del pangenoma permite determinar cuales son el conjunto de reacciones que confieren a un conjunto de microoganismos diferenciación con respecto a microrganismos de otra especie. Así mismo el pangenoma ayuda a identificar los procesos moleculares de adaptación a un nicho, que confieren las características particulares a cada cepa. Hoy en día el concepto de especie en microbiología es un concepto principalmente fenético, el cual concibe la especie como una unidad basada en el grado de semejanza entre los microrganismos que conforman dicha especie. Se tienen en cuenta tanto caracteres fenotípicos de los microorganismos, como la naturaleza del nicho ecológico y la secuencia del gen que codifica el RNA 16S. Sin

embargo este concepto de especie es más real en microorganismos que tienen un nicho ecológico aislado, en el cual no prime el intercambio genético con otros microorganismos y con unas condiciones ambientales muy estables. Este no es el caso de la mayoría de los microorganismos que habitan la biosfera. La especie utilizada en el presente estudio es *Pseudomonas putida*, como especie modelo de microrganismo cosmopolita con cepas que habitan en distintos nichos edáficos y de agua dulce (suelo, rizosfera, lagunas, sedimentos). Esta especie está formada por bacilos Gram negativos, aerobios, móviles gracias a sus flagelos polares.

Para deducir el pangenoma de esta especie se eligieron nueve cepas intentando abarcar diferentes nichos. Además también se tuvo en cuenta el grado de curación de las secuencias y el estado de su genoma en las bases de datos públicas (borrador o genoma cerrado) intentando contar con aquellas que tuviesen el grado de curación más alto. Además también se utilizaron tres cepas que fueron secuenciadas, ensambladas y reanotadas en nuestro grupo de la EEZ-CSIC en Granada. Estas tres cepas fueron reanotadas manualmente atendiendo a la anotación depositada en las bases de datos de referencia de otras cepas ampliamente estudiadas como son las cepas *Pseudomonas putida* KT2440 y F1. El proceso de reanotación y estudio de las cepas según sus características más relevantes se ha descrito en los tres primeros capítulos que conforman esta tesis.

El capitulo I versa sobre la cepa *Pseudomonas putida* DOT-T1E, un microorganismo procedente de una planta de tratamiento de aguas residuales, caracterizado por ser capaz de proliferar en presencia de altas concentraciones de disolventes orgánicos y de utilizar hidrocarburos aromáticos como única fuente de carbono. Debido al importante potencial biotecnológico de esta cepa para su uso en reacciones de biotransformación en sistemas bifásicos, se procedió al diseño de un atlas metabólico que recabase información tan amplia como posible del conjunto de reacciones enzimáticas que lleva a cabo esta cepa, atendiendo a su código enzimático. El *Enzyme Commision number* (E.C.) es una nomenclatura que permite clasificar las enzimas dependiendo de la reacción que catalizan, constando de 4 números separados por puntos. Estos números representan una clasificación progresiva de mayor a menor especificidad. El atlas metabólico de la cepa DOT-T1E permitió por ejemplo inferir que más de un 41 % de las reacciones son de oxidación/reducción.

Uno de los inconvenientes que tienen lugar a la hora de resolver problemas biológicos es contar con un software que permita visualizar la información biológica de forma

integrada y fácilmente accesible. Por ello además se construyó una base de datos de la cepa DOT-T1E utilizando el programa Pathway Tools. Este software integra datos genómicos con anotaciones funcionales y otros datos tales como descripciones de rutas metabólicas, de señalización y redes de regulación. Con ayuda de Pathway-Tools se procedió a reanotar la cepa de forma manual y además se utilizó para visualizar las rutas metabólicas implicadas en los procesos de degradación de compuestos aromáticos de interés, al igual que para el diseño de nuevas rutas no presentes en la cepa silvestre, tales como la bioconversión de tirosina en compuestos de valor añadido.

El capitulo II se centra en *Pseudomonas putida* BIRD-1, una rizobacteria promotora del crecimiento en plantas (PGPR). Esta cepa es capaz de adherirse a la raíz y de proliferar en la rizosfera de la planta incluso a porcentajes de humedad muy bajos. Para poder estudiar las características genotípicas que le confieren a BIRD-1 su capacidad como PGPR, se elaboró una base de datos para esta cepa utilizando Pathway Tools, donde se procedió a su reanotación manual. El análisis del genoma de BIRD-1 reveló que esta cepa es capaz de utilizar como fuente de carbono y de nitrógeno productos secretados por la planta, como son los aminoácidos, prolina, glutamato, glutamina, asparragina, tirosina, valina, serina y alanina, entre otros; varios ácidos orgánicos como el ácido cítrico y el láctico; azúcares como la glucosa y la fructosa; y varios flavonoides. También comprobamos que el genoma de BIRD-1 alberga una gran cantidad de genes relacionados con la resistencia al estrés provocado por especies reactivas de oxígeno, que se generan en la rizosfera de la planta y que la resistencia a la desecación que exhibe esta cepa, está ligada a su capacidad para producir trehalosa. Las propiedades PGPR de *P. putida* BIRD-1 se derivan de varios factores. Esta cepa es capaz de solubilizar formas insolubles de fósfato inorgánico a través de la producción de ácidos y de mineralizar fosfato orgánico mediante la producción de fosfatasas y al menos una fitasa, la cual facilita la utilización del ácido fítico, que es la principal forma de almacenamiento de fósforo orgánico en la planta. También se comprobó la existencia de una ruta de síntesis del sideróforo pioverdina, que es capaz de quelar el hierro no biodisponible en formas asimilables tanto por el microorganismo que produce el sideróforo, como por las raíces de las plantas en las que habitan. BIRD-1 produce un exceso de ácido indolacético que es una hormona vegetale que promueve el desarrollo de raíces laterales y adventicias, estimula el desarrollo de frutos y promueve fototropismo positivo. Las rutas de síntesis de esta fitohormona a partir de triptófano se elucidaron y se estudiaron a nivel informático.

El capítulo III es un estudio de los determinantes de resistencia a antibióticos de la cepa *Pseudomonas putida* HB3267, microorganismo patógeno oportunista que se aisló de un paciente en el hospital de Besaçon (Francia). Esta cepa es capaz de matar insectos y tiene un amplio patrón de resistencia, siendo resistente a la mayoría de los antibióticos utilizados en laboratorios y hospitales incluyendo aminoglicósidos, ß-lactamicos, péptidos catiónicos, antibióticos formados por grupos funcionales del tipo enediyne, inhibidores de hidrofolato reductasa, fluoroquinolonas y quinolonas, antibióticos tipo glucopéptidos, macrolidos, policétidos y sulfamidas. Una vez creada la base de datos para Pathway Tools, la cepa fue reanotada manualmente y su genoma se analizó para identificar los determinantes de resistencia a múltiples antibióticos. El análisis reveló que los genes implicados en el fenotipo multiresistente están localizados en su mayoría en el cromosoma de la cepa, pero también en su plásmido pPC9. El estudio de estos determinantes de resistencia estableció la participación de genes que codifican bombas de extrusión RND y ABC, mutaciones puntuales en genes esenciales sobre los que actúan los antibióticos y también duplicaciones de genes implicados en la resistencia a antibióticos.

El pangenoma de *Pseudomonas putida* se describe en los capítulos IV y V, en los cuales se determinaron los distintos grupos de genes que comprende el pangenoma de la especie. El análisis de estos grupos de genes, accesorios, únicos y del *core*, permite la investigación a distintos niveles del genoma de la especie ya que suponen una clasificación tanto estructural como funcional del mismo. Estudios estructurales de composición del DNA de los tres distintos grupos de genes (composición G+C y uso de codones) analizados en el capítulo IV confirmaron la existencia de diferencias estadísticamente significativas entre dichos grupos de genes. Los genes del *core* analizados en el capítulo IV, comprende aquellos genes comunes a todas las cepas del análisis. Estos genes codifican funciones metabólicas muy relevantes y por ello, están conservados en todas las cepas. Los genes del *core* en la especie *Pseudomonas putida* son un grupo amplio de genes que abarcan más de la mitad de los genes de cada una de las cepas. El pangenoma de *Pseudmonas putida* se puede clasificar como un pangenoma del tipo “abierto”, ya que este aumenta en número de genes con la incorporación al estudio de los genomas de nuevas cepas.

El capítulo V presenta un estudio funcional de los genes accesorios y únicos. Los genes accesorios son aquellos que se encuentran en al menos dos, pero no en todas las cepas del estudio, mientras que los genes únicos, son aquellos que tan solo están presentes

en una de las cepas. Este análisis funcional permitió verificar que los genes accesorios y únicos codifican funciones específicas del nicho en el que habitan cada una de las cepas, confiriéndole a cada una de ellas su propia “idiosincrasia”.

GENERAL INTRODUCTION

Background

Bacteria are basically everywhere, with an estimated $6x10^{30}$ individuals, bacteria are the most abundant Domain of life that exists on earth (Whitman *et al.,* 1998). Most bacteria are single-cell microorganisms that are only a few micrometers (µm) in length. However, they have been found in all explored fresh water, marine and terrestrial ecosystems. Bacteria inhabit temperate and mild niches as well as the most hostile environments, such as the Antarctic ice (Price, 1999), hot springs (Yim *et al*., 2006), deep seas (Nogi *et al*., 1998) and deep below the surface of earth (Amils *et al*., 2015). They range from the most benign inhabitants of earth to some of the most deadly, multidrug-resistant human pathogens. Microbes catalyze fundamental steps in geochemical cycles, and participate in key ecological relationships (i.e., symbiosis, protocooperation, commensalism, etc.) that determine the diversity and distribution of higher organisms in most, if not all, environments. In addition to the impact that bacteria have on ecosystems, they are also important for human health and our daily life (Consortium The Integrative HMP (iHMP) Research Network, 2014). The gene content of our gut microbiota outnumbers our own human genes by a factor of at least 100 (Qin *et al*., 2010), and encodes metabolic capabilities that are not represented in the genome of *Homo sapiens* (Goodman and Gordon, 2010). In addition to the beneficial contribution of bacterial life in the environment and human health, bacteria are also of great economic importance due to their biocatalytic abilities. Microbes are used in the production of food and beverage products, including cheese, sourdough, beer, potable spirits and vinegar, and the manufacture of commodities, such as leather and linen (Jemli *et al*., 2014). More recently, single enzymes or enzymatic complexes have been used in industrial applications and processes ranging from chemo-enzymatic synthesis to the generation of novel biofuels from renewable biomass (Kim *et al*., 2008; Wen *et al*., 2009; Guldhe *et al*., 2015).

The success of bacterial life is illustrated by the wide variation in their genetic content, not just across phyla, but even among members of the same species (Lan and Reeves, 2000; Ochman *et al*., 2000). In addition, they contain an abundance of gene regulatory mechanisms, that ensure the appropriate genes are expressed when required (Martínez-Antonio and Collado-Vides, 2003). We have only just begun to

learn how to interpret diversity among closely related genotypes in the context of the ecological and evolutionary processes that affect populations (Cordero and Polz, 2014); for this reason, to unveil the bacterial genome of a given species, we need to analyze the genome sequences of several strains — this approximates the "full species genome" or ***bacterial pangenome***. The pangenome is defined as a repertoire of genetic sequences found in a given bacterial species (Medini *et al*., 2005) and that are distinct from an individual genome of the species.

1. Bacterial genomes

Bacteria exhibit a large variability in genome size among all completely sequenced 4,108 prokaryotic (Archaea and Bacteria) genomes (Figure 1) (NCBI, National Center for Biotechnology Information, Complete Microbial Genomes; 2015. http://www.ncbi.nlm.nih.gov/genome/browse/). Thus far, the 0.13 Mb of the Alphaproteobacteria, *Candidatus Hodgkinia cicadicola* TETUND1 (McCutcheon *et al*., 2009) represents the smallest genome and the delta/epsilon proteobacteria *Sorangium cellulosum* So0157-2 (Han *et al*., 2013), the largest complete genome sequenced with 14.8 Mb. Bacterial genomes can be organized in a number of ways but generally contain a single, closed, circular double helix of DNA, the chromosome (Cairns, 1963; Toro and Shapiro, 2010). A number of microorganisms, i.e. *Burkholderia* (Kwak and Shin, 2015), and *Rhizobium* (Crook *et al*., 2013), contain two or more chromosomes. Linear replicons have been found in many bacteria, e.g., *Streptomycetes* (Leblond *et al*., 1993; Bentley *et al*., 2002; Pang *et al*.,

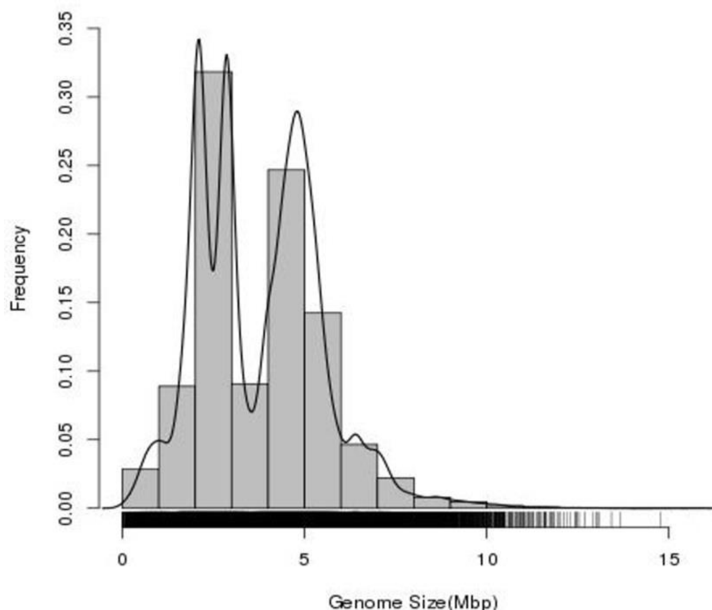


Figure 1..Distribution of genome sizes among prokaryotes. The density curve was obtained by Gaussian-kernel smoothing of the individual data points with R. Data obtained from NCBI, Complete Microbial Genomes;2015.http://www.ncbi.nlm.nih.gov/genome/browse/

2002), and a mixture of linear and circular replicons are present in *Borrelia burgdoferi* (Fraser *et al*., 1997), *Agrobacterium tumefaciens* (Allardet-Servent *et al*., 1993), *Rhodococcus fascians* (Pisabarro *et al*., 1998), and related species. It should be noted that, linear chromosomes are often kept as circular ones inside the cell because the ends are bound by terminal proteins (Yang *et al*., 2002). The number and geometry of chromosomes is just part of the "gene scenario" because in addition to chromosomes, many bacterial isolates also carry dynamic DNA molecules; so called plasmids. Plasmids are extra-chromosomal DNA molecules that are capable of both vertical and horizontal transmission and are commonly found in naturally isolated bacteria. Plasmids share many characteristics with chromosomes but the main difference between chromosomes and plasmids is that plasmids are not essential for the survival of a particular bacterial species. Chromosomes harbor essential genes as well as characteristic cell cycle-linked replication kinetics; unlike plasmids, chromosomes initiate replication once per cell cycle. Where chromosome loss inevitably leads to the death of a bacterial cell, the loss of a plasmid seems only to be disadvantageous for the host cell (Dionisio *et al*., 2005; Egan *et al*., 2005). Both chromosome and plasmids, comprise the bacterial genome, however the definition of the genome as the entire collection of replicons in the cell carries practical difficulties. Plasmids can be lost from a strain during laboratory cultivation on artificial medium and different strains of a bacterial species might be very similar at the chromosomal level yet have a completely different complement of plasmids. Thus, the term "complete genome" must be taken with the caveat that it can only refer to the DNA sequence of a particular strain, at the time it was sequenced.

Together with the main chromosome and plasmids, a number of bacteria have a "second chromosome" or megaplasmid often called a chromid (Harrison *et al*., 2010). The main characteristics of chromids are, a) their considerable size, b) the presence of plasmid-type maintenance and replication systems, c) the presence of adaptive genes typical for plasmids that are useful in a particular ecological niche, d) a nucleotide composition close to that of the chromosome, and maybe most importantly, f) chromids carry conserved genes that are found on the chromosome in other species (Harrison *et al*., 2010; Dziewit and Bartosik, 2015). Chromids have been claimed as genera-specific elements, whereas megaplasmids, because of their possible more recent origin, may be defining interspecies boundaries inside a genus.

However, it is still unclear whether chromosomes, chromids, and megaplasmids have common or distinct evolutionary routes at the intraspecies level (Galardini *et al*., 2013).

Another significant component of bacterial success which is responsible for their fast adaptation to the environment and which is probably the most important mechanism for functional novelty and adaptation in prokaryotes, is their ability to incorporate foreign DNA through horizontal gene transfer (HGT) (Ochman *et al*., 2000; Koonin *et al*., 2001; Jain *et al*., 2002; Lerat *et al*., 2005; Lercher and Pál, 2008; Juhas *et al*., 2009; Syvanen, 2012). Several studies have revealed that its occurrence is much more frequent than previously thought (Ochman *et al*., 2000; McDaniel *et al*., 2010; Syvanen, 2012; Soucy *et al*., 2015). Although duplications play an important role in the short-term adaptation of bacteria (Andersson *et al*., 1998; Hendrickson *et al*., 2002), protein family studies, suggest that HGT and not

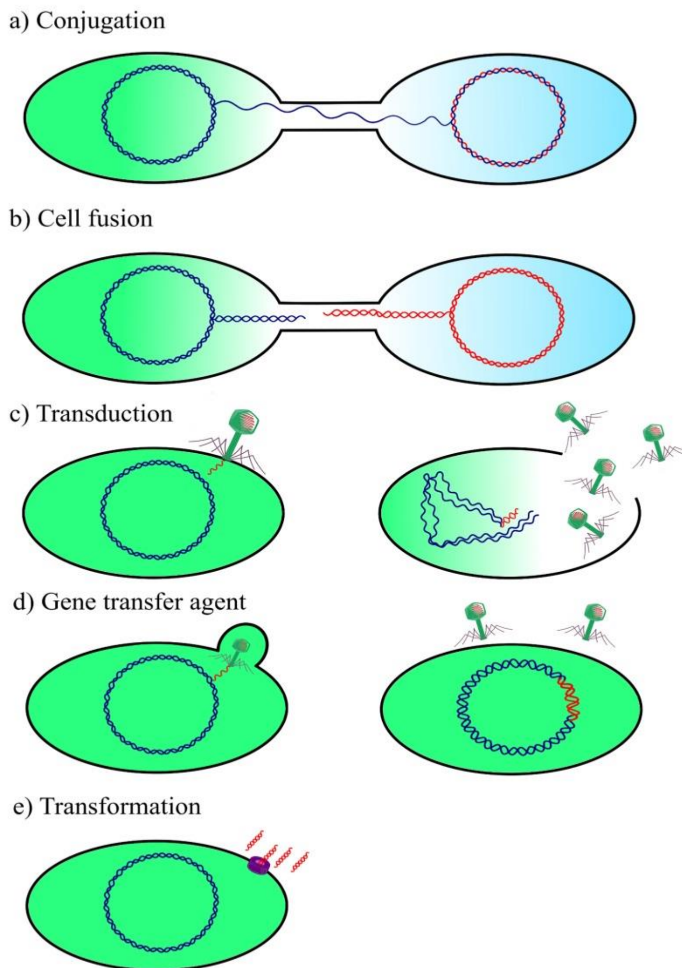


Figure 2. Mechanisms of gene transfer. (a) Conjugation occurs through donor–recipient cell contact, and single-stranded DNA is transferred from the donor cell to the recipient cell. (b) Cell fusion, DNA is exchanged bi-directionally after cell contact and bridge formation between two cells. (c) Gene transfer mediated by phage is known as transduction. In the case of generalized transduction, any piece of genomic DNA may be loaded into the phage head; (d) Gene transfer agents (GTAs) are phages that no longer recognize their own DNA and only carry random fragments of host DNA. Like prophage, they reside in the host cell genome. (e) During transformation DNA is taken up from the surrounding environment. Figure made with Inkscape (The Inkscape Team, 2015) and GIMP (Kimball et al., 2014) programs.

duplication has driven protein expansion and functional novelty in prokaryotes. This is because, the maintenance of duplicated genes requires the gradual evolution of change in the sequence, that confers differences in expression or functionality to the duplicated genes, whereas genes acquired through HGT, are likely to be operationally distinct from those already present in a genome, and thus, immediately able to contribute unique functions and to be maintained in the genome if they bring a beneficial function (Lerat *et al.*, 2005; Treangen and Rocha, 2011).

HGT is of such significance that it existence blurs the distinction of bacterial species (Figure 2), indeed the importance of HGT in bacterial evolution has been elevated to such a degree, that numerous bacteriologists now question the concept of bacterial species (Rossello-Mora and Amann, 2001; Cohan, 2002; Gevers *et al.*, 2005; Doolittle and Papke, 2006; Fraser *et al.*, 2009; Riley and Lizotte-Waniewski, 2009). Many HGT events involve individual genes, but transfer of larger portions of genomes, particularly operons, is common as well. Indeed, operons are often subject to horizontal gene transfer (Price *et al.*, 2005; Homma *et al.*, 2007).

In 1961, Jacob and Monod described the bacterial operon as a cluster of functionally interacting genes whose expression is tightly coordinated (Jacob and Monod, 1961). Operons are groups of adjacent genes that typically form a single transcriptional unit. Most operons contain 2 to 4 genes; however, much larger uber-operons, such as the one that encodes many ribosomal proteins or the TOL plasmid catabolic genes are also capable of undergoing HGT (Ramos *et al.*, 1997, 2013; Lathe *et al.*, 2000; Che *et al.*, 2006). The gene content of an operon may be conserved between two genomes; this may be due to the selective pressure for co-expression of genes encoded therein, nevertheless, the order in which the genes that constitute the operon are transcribed may not be kept.

Several studies have attempted to elucidate the selective advantages provided by the transfer of complete operons between bacterial genomes (Homma *et al.*, 2007; Lercher and Pál, 2008). Although successful transfer and initial integration of foreign DNA, become less likely for larger fragments and hence, large modules of interacting genes are unlikely to be transferred in a single step, the transfer of complete operons, is feasible because of the proximity of the genes and because this event may enable the recipient to acquire a useful new capability. Operons allow the

regulation of a set of genes under a common control mechanism, this in turn allows bacteria to rapidly adapt to changes in the environment e.g., nutrient availability may increase or decrease radically, pH could suddenly change or the temperature may fluctuate (Bahl *et al*., 1995; Ramos *et al*., 2013; Barbier *et al*., 2014; Huang *et al*., 2014). However, how genetic networks maintain and transmit phenotypic states between generations in rapidly fluctuating environments is still unclear (Lambert and Kussel, 2014).

An overview of codon bias and codon usage in bacterial genomes

The pioneering work "Studies on the chemical nature of the substance inducing transformation of Pneumococcal types III" from Avery, MacLeod and McCarty in 1944 (Avery *et al*., 1944) revealed that DNA is the macromolecule from which living organisms are constructed. The DNA is compound by directional strings of discrete length, constructed from four different chemical compounds. The arrangement of these four nucleotides or nitrogen-bases called, adenine (A) and guanine (G) (the larger bases called purines) and thymine (T) and cytosine (C) (the smaller bases or pyrimidines), into chains of varying sizes known as Deoxyribonucleic Acid (DNA) sequences or strands is enough to produce the great variety of life forms. The disposition of nucleotides into triplets of bases, called codons, gives rise to a new higher level alphabet of 20 chemical compounds called amino acids.

Table 1. The genetic code

First letter	Second letter: U		C		A		G		Third letter
U	UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine	UCA		UAA	Stop codon	UGA	Stop codon	A
	UUG		UCG		UAG		UGG	Tryptophan	G
C	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lysine	AGA	Arginine	A
	AUG	Methionine	ACG		AAG		AGG		G
G	GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		AAA	Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

Green cells represent start codons, orange cells represent stop codons

The genetic code is said to be degenerate as multiple nucleotide triplets can code for one amino acid (Table 1). Degeneracy of the genetic code was identified by Lagerkvist (Lagerkvist, 1978) and most of the degeneracy was found at the third nucleotide position of the codons. Since 61 codons specify 20 amino acids, most amino acids are encoded by more than one codon, such as serine, which is coded by 6 codons. Synonymous codons are those that code for the same amino acid. There are amino acids that are encoded by a single codon, which is the case for the amino acids methionine and tryptophan. In total 18 of the 20 amino acids are coded by more than one codon. Nevertheless, in genes some synonymous codons are more abundant than others. Grantham's genome hypothesis proposed that each species systematically uses synonymous codons with different frequencies (Table 2) (Grantham, 1980). This phenomenon is called codon bias. Many studies have confirmed this phenomenon (Sharp and Li, 1987; Dong *et al*., 1996; Krause *et al*., 2006; Hershberg and Petrov, 2008; Roymondal *et al*., 2009; Angov, 2011; Plotkin and Kudla, 2011). Likewise, the strength of codon bias varies across genes within each genome, with some genes using a highly biased set of codons and others using the different synonymous codons with similar frequencies (Gouy and Gautier, 1982; Ikemura, 1985). Evolutionary forces drive these differences in codon preference, the codon distribution responds to genome G+C content and the changes in codon usage are at least, partly explained by mutation–selection equilibrium between the different synonymous codons in each organism (Grosjean and Fiers, 1982). Codon bias among genes from the same organism is most likely related to several parameters (Chen *et al*., 2004), such as protein expression level, or growth efficiency (Gouy and Gautier, 1982; Ikemura, 1985; Kurland and Ehrenberg, 1987; Gustafsson *et al*., 2004; Quax *et al*., 2015); and suggests an optimization of the translation machinery to use a small subset of optimal codons and anticodons in fast-growing bacteria and in highly expressed genes (Rocha, 2004). From these pioneering studies, a positive correlation between codon usage, gene expression level, and growth efficiency of prokaryotic cells (Kurland and Ehrenberg, 1987) was established and used to generate the codon adaptation index (CAI) (Sharp and Li, 1987).

Table 3. Codon usage table of *Pseudomonas putida KT2440* genome.

Codon	AA	Fraction	Frequency	Number	Codon	AA	Fraction	Frequency	Number
GCA	A	0.119	13.228	23657	CCA	P	0.142	6.966	12458
GCC	A	0.525	58.534	104679	CCC	P	0.233	11.435	20449
GCG	A	0.253	28.245	50513	CCG	P	0.504	24.703	44178
GCT	A	0.103	11.513	20590	CCT	P	0.120	5.872	10502
TGC	C	0.817	8.482	15169	CAA	Q	0.230	10.800	19314
TGT	C	0.183	1.904	3405	CAG	Q	0.770	36.087	64536
GAC	D	0.702	37.189	66507	AGA	R	0.015	0.989	1768
GAT	D	0.298	15.803	28262	AGG	R	0.033	2.181	3900
GAA	E	0.484	27.203	48649	CGA	R	0.046	3.016	5393
GAG	E	0.516	29.048	51949	CGC	R	0.558	36.843	65889
TTC	F	0.804	28.571	51096	CGG	R	0.154	10.193	18228
TTT	F	0.196	6.943	12417	CGT	R	0.194	12.802	22894
GGA	G	0.032	2.594	4639	AGC	S	0.424	23.751	42475
GGC	G	0.619	49.729	88933	AGT	S	0.091	5.084	9092
GGG	G	0.147	11.787	21079	TCA	S	0.050	2.791	4991
GGT	G	0.202	16.258	29075	TCC	S	0.149	8.339	14914
CAC	H	0.666	15.666	28016	TCG	S	0.246	13.822	24718
CAT	H	0.334	7.873	14079	TCT	S	0.041	2.285	4087
ATA	I	0.041	1.861	3328	ACA	T	0.069	3.291	5886
ATC	I	0.756	34.384	61491	ACC	T	0.664	31.666	56630
ATT	I	0.203	9.224	16495	ACG	T	0.176	8.373	14974
AAA	K	0.278	9.339	16702	ACT	T	0.091	4.350	7780
AAG	K	0.722	24.270	43404	GTA	V	0.102	7.406	13244
CTA	L	0.021	2.418	4324	GTC	V	0.291	21.129	37786
CTC	L	0.125	14.625	26155	GTG	V	0.514	37.320	66742
CTG	L	0.643	75.394	134831	GTT	V	0.092	6.699	11980
CTT	L	0.058	6.801	12162	TGG	W	1.000	14.431	25808
TTA	L	0.011	1.270	2271	TAC	Y	0.729	18.420	32942
TTG	L	0.142	16.659	29792	TAT	Y	0.271	6.834	12222
ATG	M	1.000	22.729	40647	TAA	*	0.216	0.645	1153
AAC	N	0.793	23.454	41944	TAG	*	0.112	0.336	601
AAT	N	0.207	6.132	10967	TGA	*	0.672	2.011	3596

The table was prepared by taking into account the 5,350 genes in the strain and was calculated with the CUSP program from EMBOSS package (Rice et al., 2000). The table shows the existing bias between codons, for example in the use of the ATC codon for isoleucine, AAC for asparagine and TAC for tyrosine.

The CAI was proposed as a quantitative way of predicting the expression level of a gene based on its codon sequence and defines a “relative adaptiveness” to each of the 61 codons (stops codons are not included). The relative adaptiveness of a

codon is defined as its frequency relative to the most often used synonymous codon (Sharp and Li, 1987).

A caveat and limitation is that the predictive value of CAI is directly dependent on the genes used to establish the “reference set” and may have relatively limited value for predicting expression from genes not reflected by the codon bias found in the reference set. Recognizing these limitations, investigators have developed more “universal” CAI, which measure codon bias based on reference sets that are not necessarily derived from preferred, “highly” expressed genes, but from all known coding sequences for a specific organism (Carbone *et al*., 2003). Even assuming these limitations, CAI does not necessarily reflect all possible factors that influence gene expression levels per se, for example, the efficiency of ribosome binding and translation initiation (Berg and Kurland, 1997; Angov, 2011). In addition to being widely used to predict expression levels from genes, the codon adaptation index is also used to approximate the success of heterologous gene expression (Gustafsson *et al*., 2004).

2. *Pseudomonas putida*

Pseudomonas putida (Figure 3) is a Gram-negative, rod-shaped, saprotrophic bacterium with polar flagella (Palleroni, 1984). These chemoorganotrophic bacteria were first described at the end of the 19th century by Migula (1894); later, according to the classification of Bergey's Manual of Systematic Bacteriology, *Pseudomonas* were described as "rod-shaped cells, curved or straight, but not spiral, from 0.5 to 1 µm in diameter by 1.5 - 5 µm in length". The development of techniques such as DNA / DNA hybridization or rRNA / DNA allowed a clear division of *Pseudomonas species* (Palleroni *et al*., 1972). Following this latest

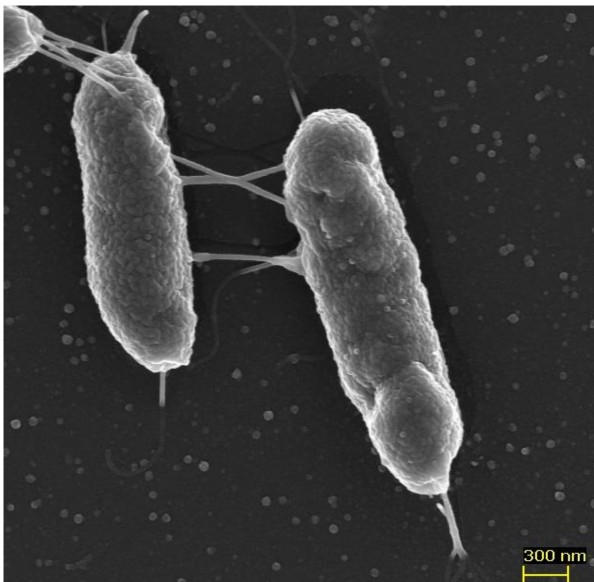


Figure 3. Scan Electron Microscopy picture of *Pseudomonas putida* KT2440. Courtesy of Estrella Duque.

ranking, Palleroni proposed five taxa (RNA-I RNA-V). Finally, 16S rRNA sequencing techniques reflected the diversity of these groups, and it was the RNA I group, within the subclass-γ Proteobacteria, which today is recognized as covering the true *Pseudomonas* (Bergey *et al*., 2005).

Pseudomonas putida are aerobic microorganisms, which use a type of strict respiratory metabolism with oxygen as the terminal electron acceptor, although in some cases nitrate can be used as an alternative electron acceptor. *P. putida* are catalase and oxidase positive and grow at neutral pH, at temperatures in the mesophilic zone and do not require growth factors. Due to its ability to produce fluorescence-emitting pigments they belong to the group of fluorescent *Pseudomonas*.

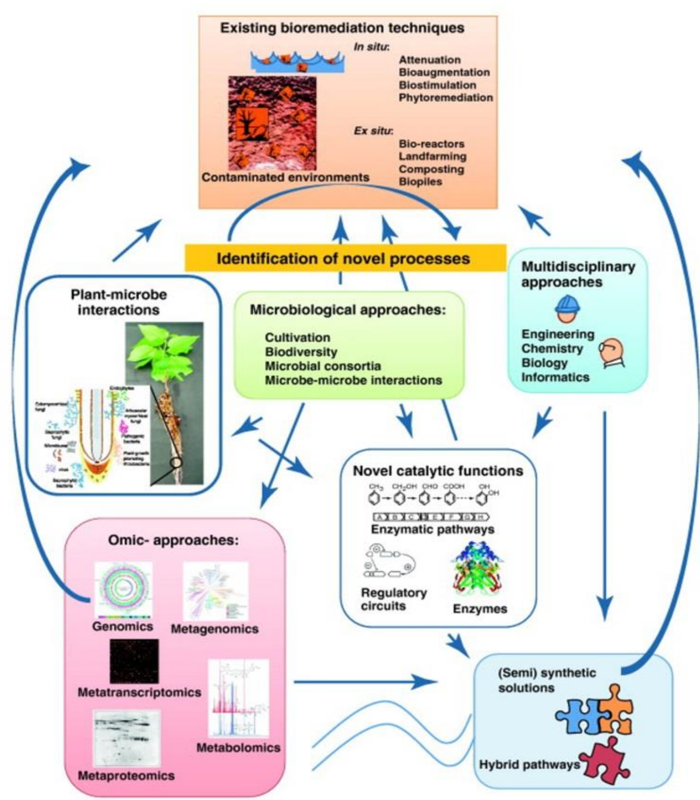


Figure 4. Schema indicating ways to identify new processes to improve the bioremediation of contaminated environments using microbes from *Pseudomonas putida* species. Adapted from *Trends in Biotechnology* Review (Ramos et al., 2011)

Fourteen *Pseudomonas putida* strains have been fully sequenced and completely annotated, while another thirty-one genomes are available in partial scaffolds or contigs (NCBI, National Centre for Biotechnology Information, Genome assembly and annotation report; 2015. http://www.ncbi.nlm.nih.gov/genome/genomes/174). *Pseudomonas putida* has a single circular chromosome that encodes approximately 5000 genes. Its genome size has been determined to be approximately 6 Mb in length (Ramos-Díaz and Ramos, 1998; Nelson, *et al*., 2002), with a GC content of

about 60%. *Pseudomonas putida* strains are widespread in the environment; this is because of a very adaptable metabolism that allows them to colonize different habitats, either pristine or polluted (Weinel *et al*., 2002; Wu *et al*., 2010). They are of great importance from an ecological point of view, due to their involvement in the environmental cycles of the major elements and in the degradation of biogenic and anthropogenic pollutants (Figure 4) (Timmis, 2002). Several strains have been proven to be good colonizers of a variety of plant rhizospheres, these include, *Pseudomonas putida* KT2440 (Espinosa-Urgel and Ramos, 2004) and *Pseudomonas putida* BIRD1 (Roca *et al*., 2013). At the present time, some strains are being used as a base for biological fertilizers and remediation treatments (Segura *et al*., 2009; Roca *et al*., 2013; Segura and Ramos, 2013). Their use as biocontrol agent against plant pathogens has also been studied (Costerton, 1999; Rainey, 1999; Ramos-González *et al*., 2005; Setubal *et al*., 2005).

Strains of *P. putida* are a paradigm of metabolically versatile microorganisms because they are able to use aliphatic and aromatic hydrocarbons as a sole carbon and energy source and a wide variety of inorganic and organic nitrogen sources (Daniels *et al*., 2010; Udaondo *et al*., 2013). This species is indeed, one of the most well-studied aromatic hydrocarbon degrading bacteria (Ramos *et al*., 1995; Segura *et al*., 2012; Udaondo *et al*., 2012). They are also capable of degrading gasoline BTEX (Benzene, Toluene, Ethylbenzene, and Xylene) (Mazzeo *et al*., 2010; Kim and Lee, 2011; You *et al*., 2013), although the efficiency in their ability to degrade aromatic compounds can vary among strains. Due to their solvent-tolerant characteristics, some strains of this species are considered to be true extremophiles, which allow them to thrive in the presence of high concentrations of toxic solvents (Isken and de Bont, 1998; Ramos *et al*., 2002, 2015; Neumann *et al*., 2005).

Well over 2000 articles have been written about various aspects of *P. putida* physiology, enzymology, and genetics by microbiologists and biochemists, in addition to more applied studies by chemists and environmental engineers utilizing *P. putida* and its enzymes for green chemistry applications and bioremediation (PubMed Database from NCBI, National Centre for Biotechnology Information; 2015.
http://www.ncbi.nlm.nih.gov/pubmed?term=Pseudomonas%20putida%5BTitle%5D. Articles published between 1956 and September 2015).

Genome sequencing has continuously increased our knowledge of this species and it enabled researchers to explain their ability to adapt and survive in different environments and to use biogenic and xenobiotic compounds as carbon sources Recent studies in the area of transcriptomics, have elucidated the intricate mechanisms that this species use to defend themselves against a broad spectrum of antibiotics (Molina-Santiago *et al*., 2015), and the regulatory inputs that orchestrate the complex network of metabolic responses observed after solvent addition (Ramos *et al*., 2015). The information derived from the analysis of the genome of several *P. putida* strains along with the ease of genetic manipulation of *P. putida*, have made this species a model system for the study of bioremediation, biocatalysis and the phenomena involved in root colonization (Udaondo *et al*., 2012; Segura and Ramos, 2013).

Pseudomonas putida KT2440

Pseudomonas putida KT2440, represents the best-characterized strain from this species and because of its rapid growth, robustness and its amenability to genetic analysis and manipulation, it is considered a workhorse for *Pseudomonas* research (Timmis, 2002; Martins Dos Santos *et al*., 2004; Wu *et al*., 2010). This rhizospheric bacteria is the plasmid-free derivative of the strain *P. putida* mt-2, isolated from garden soil in Japan based on its ability to use 3-methylbenzoate (Nakazawa, 2002). *P. putida* KT2440 shares 85% of its predicted coding sequences with the opportunistic pathogen *Pseudomonas aeruginosa*, even so, *P. putida* KT2440 lacks key virulence factors including exotoxin A and type III secretion systems (Nelson *et al*., 2002). KT2440 is inefficient in restriction of foreign DNA, a trait that has made this strain very well studied in the expansion of degradative carbon pathways (Bagdasarian and Timmis, 1982; Ramos *et al*., 1986, 1994) and as a host for cloning and expression of heterologous genes to be used in biotransformation processes (Delgado *et al*., 1992; Kraak *et al*., 1997; Tan *et al*., 1997; Kellerhals *et al*., 1999).

Pseudomonas putida BIRD1

Pseudomonas putida BIRD-1, isolated from garden soil, is an efficient plant growth-promoting rhizobacteria (PGPR). It mobilizes soil nutrients that become available for uptake by plants and stimulates proliferation of secondary roots (Matilla *et al*., 2011). In turn compounds present in plant exudates are also used by

P. putida BIRD1 to grow. BIRD1 is a good colonizer of several plant roots e.g., maize, tomato, pepper, zucchini and strawberry (Roca *et al*., 2013). This strain is able to solubilize organic and inorganic phosphate through the production of gluconic acid; a process which makes phosphate more accessible to the plant. In addition, BIRD1 has the ability to synthesize phytohormones as indoleacetic acid, secrete siderophores (as pyoverdine) in iron–deficient conditions and survive the oxidative stress imposed by the rhizosphere (Roca *et al*., 2013).

***Pseudomonas putida* DOT-T1E**

Pseudomonas putida DOT-T1E is a solvent-tolerant strain isolated from a wastewater treatment plant in Granada, Spain (Ramos *et al*., 1995) based on its capability to use several aromatic compounds such as benzene, ethylbenzene and toluene as a carbon and energy source (Udaondo *et al*., 2012, 2013). These aromatic hydrocarbons are oxidized to their corresponding catechols and upon *meta*-cleavage, the resulting alkylmuconic acid semialdehydes, are directed towards the Krebs cycle (Mosqueda *et al*., 1999), where generation of NADH takes places and electrons are channeled to a wealth of respiratory chains (Rojo, 2010). In addition, this strain is highly resistant to organic solvents such as octanol, decanol, benzene, toluene and others (Ramos *et al*., 1995; Mosqueda *et al*., 1999; Rojas *et al*., 2001; Segura and Ramos, 2014), which has led the scientific community to consider this microorganism as a model system for the study of the response of Gram negative bacteria to toxic organic chemicals. It should be noted that utilization of solvents as a carbon source and tolerance to solvents are independent events (Mosqueda *et al*., 1999). Strain DOT-T1E has also been used for the biosynthesis of chemicals such as the development of biotechnological biotransformation processes in two-phase systems for the production of *p*-hydroxybenzoate, alkylcatechols and other chemicals (Ramos-González *et al*., 2001; Rojas *et al*., 2001).

***Pseudomonas putida* HB3267**

Pseudomonas putida HB3267 (Molina *et al*., 2013) was isolated from an in-patient of the University Hospital of Besacon, France. This strain has the ability to kill insects; an unusual feature that has only been reported for strains of other species of this genus (Mahar *et al*., 2005). A distinctive characteristic of the HB3267 strain is its resistance to the majority of the antibiotics used in laboratories and

hospitals, including aminoglycosides, ß-lactams, cationic peptides, chromoprotein enediyne antibiotics, dihydrofolate reductase inhibitors, fluoroquinolones and quinolones, glycopeptide antibiotics, macrolides, polyketides and sulfonamides (Molina *et al*., 2013).

3. *Bioinformatics*

Determining the genome sequence of microorganisms is the basis and prerequisite for understanding their biology and functional characterization (Liao *et al*., 2015). The first prokaryotic DNA was sequenced in 1977 following the development of the chain-termination method by Sanger's team (Sanger *et al*., 1977a). Subsequently, in the same year, the first full DNA genome of phage phiX174 was resolved, harboring 5,000 bp (Sanger *et al*., 1977b). The next milestone was reached 20 years later when a number of technological advances in DNA sequencing allowed the completion of the sequence of the chromosome of *Haemophilus influenzae* Rd. which is about 1.8 Mb (Fleischmann *et al*., 1995). This task involved 40 researchers and took one whole year. Since 1995 the number of

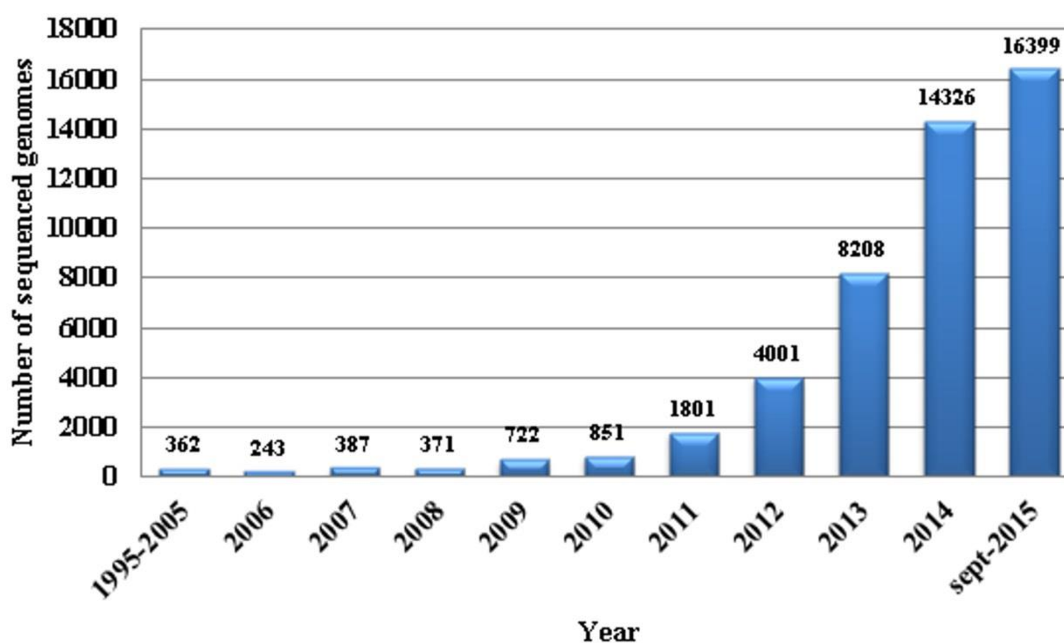


Figure 5. Number of prokaryotic genomes sequenced each year and submitted to NCBI.
Source: http://www.ncbi.nlm.nih.gov/genome/browse/ September 2015

bacterial genomes sequenced has grown exponentially. There are more than 40,000 sequenced bacterial genomes currently publically available in 2015 (NCBI, National Center for Biotechnology Information Genome Browser; 2015. http://www.ncbi.nlm.nih.gov/genome/browse/), and thousands of metagenome projects (GOLD, Genomes OnLine Database. https://gold.jgi-psf.org/) (Figure 5).

The advent of next-generation sequencing (NGS) instruments, the so-called second-generation sequencers, changed the traditional Sanger sequencing paradigm (Liao *et al*., 2015), providing large volumes of data compared with conventional Sanger sequencers.

Today, thousands of genomes are being sequenced worldwide. The most significant factor influencing this fact is the vast reduction in the price of sequencing as a result of technical developments, along with the number of programs available that ease the handling of massive data sets (Shendure and Ji, 2008). While sequencing becomes faster and cheaper, this rate of genomic data generation poses significant challenges for genomics, such as speed and complexity of analysis, data quality assessment, along with result visualization and interpretation. Thousands of genomes are available on public databases to be analyzed and compared, and the analysis reveals a surprising set of ORFs that provide new insights into the unlimited diversity and plasticity of bacterial genomes.

The term ***bioinformatic*** is used for a variety of different fields which are common in that they address biological problems using computational approaches. Basically, the bioinformaticians' ability to analyze, compare, interpret, and visualize the vast number of bacterial genomes, transcriptomes, proteomes, metatranscriptomics, etc., is simply trying to keep up with these developments (Luscombe *et al*., 2001).

Genome assembly

Sequencing reads are the typical result of a sequencing run. Second-generation sequencers generate massive amounts of overlapping reads of 30 to 1000 base pairs, depending on the technology employed. To ensure that every region of the genome is represented in the set of reads, the genome is oversampled several times, a number referred to as the coverage of the genome. Genomic libraries can be produced and

sequenced from one end or both ends of the fragments, commonly referred to as single end or paired-end sequencing, respectively (Figure.6). The assembly process uses these sources of read information to computationally reconstruct the genome, it basically involves the merging and the ordering of these short fragments with the aim of getting as close as possible to the original sequence. *Genome depth,* evenness of *coverage*, *read length* and *read quality,* are the four major factors that determine the ability to reconstruct genome sequences from sequence data (Loman *et al*., 2012).

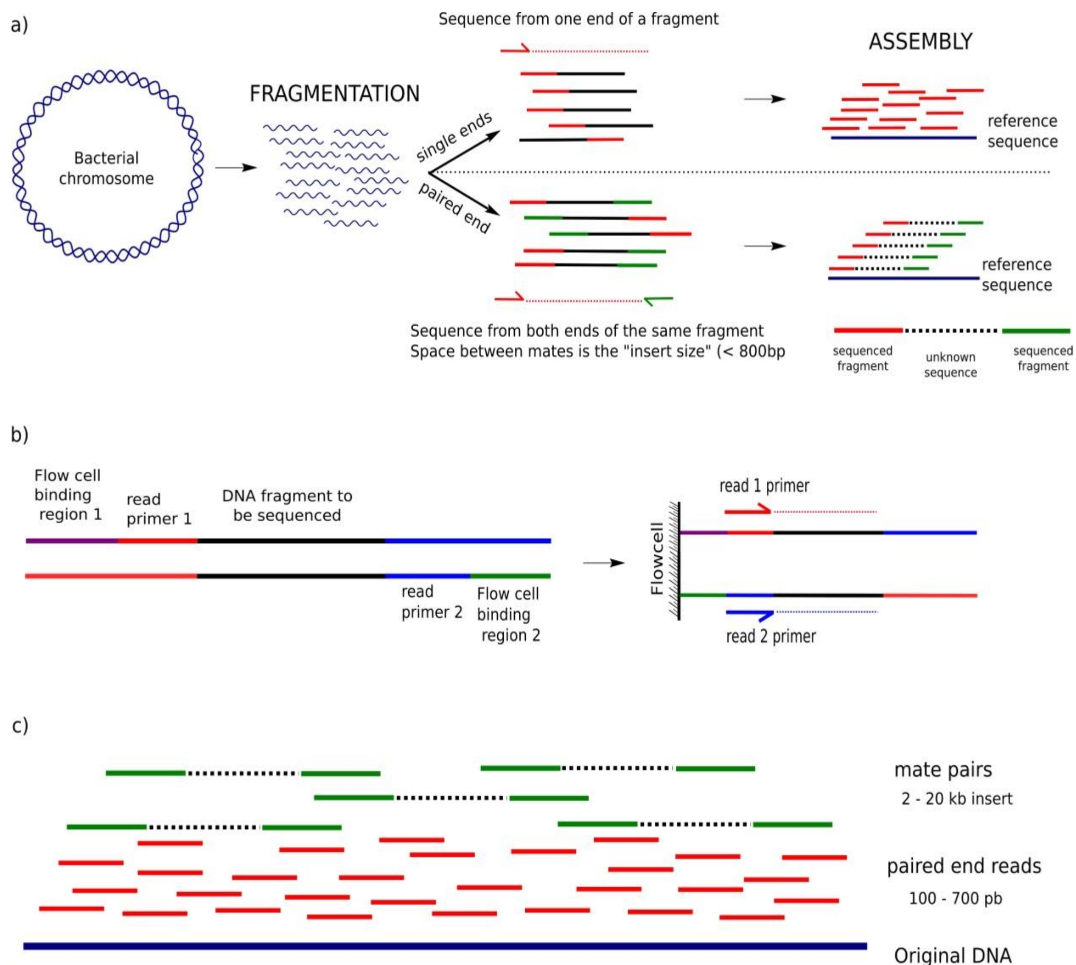


Figure 7. Sequencing and assembly scheme a) Diagram of a single-end or paired-end fragment of a whole-genome DNA shotgun library, whose average legth is 100-800-bp. b) Schematic paired-end read. If two separate read cycles occur in both directions (using both read primer 1 and 2), this kind of read will provide data about both sides of the DNA fragment of interest (black fragment). c) Mate paired end sequencing involves generates long-insert paired-end DNA libraries to improve genome assembly.

All assembly approaches rely on the simple assumption that highly similar DNA fragments originate from the same position within a genome. This similarity

between DNA fragments is used to join the fragments into larger sequences (contigs) using various assembly strategies optimized for read size and amount of coverage supplied by a host of different tools (Miller *et al.*, 2010; Nagarajan and Pop, 2013). Genomic repeats make the puzzle more complicated and at this point, the length of the sequences that are read by the sequencing instrument is crucial, because if the fragments of repeated DNA are smaller than the sequencing read, the fragments can be resolved. In the absence of very long reads (>1 Kbp), several sequencing technologies also allow for the generation of mate-pairs, where reads come in pairs with a known approximate distance between them. A standard use of mate-pairs is to link contigs together into scaffold sequences that may have gaps but represent an ordered and oriented region of the genome. Paired ends and mate-pairs have been used to tackle the problem of repetitive regions, but the resulting assembly is usually far from ideal. The third-generation Pacific Biosciences RS sequencer (PacBio) system (Eid *et al.*, 2009), solves the problem of short sequenced fragments, generating long reads with a mean length of 4.5 kbp and with randomly distributed sequencing errors (Miyamoto *et al.*, 2014). The long reads have made this technology ideal for the completion of *de novo* genome assemblies, however, this new technology still needs to address some disadvantages, such as its high cost per Mb (US $2-17 per Mb) and its high overall error rate (~13%) (van Dijk *et al.*, 2014) (Table 3). As a result of all of this, today, the length of the sequences generated by modern sequencing instruments is considerably shorter (hundreds to thousands of base pairs) than that of the genomes or genomic features being studied (which commonly span tens of thousands to billions of base pairs). Thus, many analyses start with the computational process of sequence assembly that joins together the multiple sequence fragments generated by the sequencing instrument

.

Table 4. Summary of the main sequencing techniques. PE – paired end. CG – Citosine-Guanine. AT – Adenina-Timina

Sequencer	454 GS FLX	HiSeq 2000	SOLiD v4	Sanger ABI 3730xl	PacBio Rs	Ion Torrent 316
Company	Roche (USA)	Illumina (USA)	Applied Biosystems (Life Technologies) (USA)	Applied Biosystems (Life Technologies) (USA)	Pacific Biosciences (USA)	Applied Biosystems (Life Technologies) (USA)
Sequencing principle	Pyrosequencing	Sequencing by synthesis	Sequencing by ligation and two-base coding	Dideoxy chain termination	Single molecule, realtime sequencing	Synthesis (H^+ detection) on the chip
Read length	700 - 1000 bp	2 x 100 PE reads	35 -75 pb	400 – 1,000 bp	13,000 bp	200 - 400 bp
Accuracy (%)	99. 9%	99%	99.90%	99.9%	88%	98.5%
Error rate	1%	0.26%	0.01%	0.10%	13.00%	1.70%
Run time	23 hours	10 days	7 days for SE 14 days for PE	20 mins ~3 hours	2 - 4 hours	2 hours
Cost per Mb	US $10	US $0.07	US $0.13	US $2,400	US $8	US $4,25

Advantage	Long read length, short run time	High throughput and low cost per Mb. High capacity of multiplexing	Accuracy, high capacity multiplexing	High quality, long read length	Longest read length, no amplification error	Low cost per sample
Disvantage	Error rate with polybase mora than 6 , high cost, low throughput	Short read assembly, high computation needs	Short read assembly	High cost, low throughput	Error rates, comparatively small outputs, high cost per Mb	High cost per Mb
Primary error	Indels in homopolymeric regions	Substitution in particular at the end of the read	AT bias, substitution in particular at the end of the read	Substitution	CG deletion, Indels	Indels

Contigs can principally be scaffolded by using three methods: (i) by comparison to one or multiple related reference replicons, (ii) by mate-pair sequencing of both ends of a larger insert library which allows the determination of order and distance of two reads relative to each other, and (iii) by optical mapping based on digestion of immobilized DNA molecules and determination of the size and order of fragments (Darling *et al.*, 2010). In the presence of a closely related complete replicon, reference mapping is probably the optimal solution, since little additional experimentation is required to close the gaps. In order to completely sequence replicons, gaps between contigs have to be filled by Sanger sequencing of PCRs overlapping both contigs. This requires the identification of primers, which specifically bind only inside the sequence bordering the gaps.

Determining whether an assembly is correct, is difficult given that the correct answer is usually unknown. The output of assemblers is usually fragmented and often contains mistakes that range from small nucleotide changes, to copy number changes in tandem repeats and large-scale rearrangements of the genome structure. To detect assembly errors, scientists have relied on independently derived information about the genome being assembled. The iterative refinement of assemblies through mapping and recruitment of unassembled reads has been shown to be an effective approach for improving assembly quality.

Genome annotation

Once a genome has been sequenced and assembled into a set of consensus contigs and/or scaffolds, annotation is the next step in any sequence analysis pipeline, because it provides biological meaning to the DNA sequence. At minimum, the annotated genome should provide the coordinates of predicted coding sequences and their putative products (Seemann, 2014).

Functional annotations are mostly based on the comparison of new features to databases of genetic elements with known function, such as the Universal Protein Resource (UniProt) database for proteins (The Uniprot Consortium, 2014) or the RNA Families Database (Rfam) (Gardner *et al.*, 2009). As more genomes are sequenced and published in public databases, it becomes easier to find the open reading frames from a new genome by comparative methods (Mathé *et al.*, 2002). In

this method the given DNA is compared with a data base containing DNA or protein sequences annotated from different species and then the genes are predicted based on the assumption that coding regions will be well conserved.

Nonetheless, not all the coding sequences from a genome share homology with other genes; in this case the first structural annotation is performed by *ab initio* gene prediction algorithms that are based on motifs or patterns that flank sequence coding regions. Prokaryotic protein-coding DNA has certain periodicities and statistical properties that are relatively easy to detect. These characteristics make prokaryotic gene finding relatively straightforward, so well-designed systems are able to achieve high levels of accuracy. One of these signals are the promoter sequences. Promoters consist of two short sequences at -10 (Pribnow box) and -35 (Gilbert box) positioned upstream from the transcription start site. These specific DNA sequences are recognized by subunit σ of RNA polymerase that binds to DNA to initiate the transcription. The σ subunit binds specifically to sequences in both the -35 and -10 promoter regions, substantiating the importance of these sequences in promoter function. Genes with promoters that differ from the consensus sequences, which require accessory transcription factors that bind to target cognate sequences and favor recruitment of RNA polymerase are transcribed less efficiently than genes whose promoters match the consensus sequences more closely.

Another sign is the statistics related to the stop codons that are found in an open reading frame translated in its six frames. Since 3 of the 64 possible codons in the genetic code are stop codons, one would expect a stop codon approximately every 20–25 codons, or 60–75 base pairs, in a random sequence (Lambros *et al.*, 2007). Likewise, other factors taken into account are the ORF length and the codon usage of the predicted coding sequence.Gene finding in eukaryote genomes is considerably more challenging than in prokaryotes for several reasons. First, the promoter and other regulatory signals in eukaryotes are more complex and less well-understood than in prokaryotes, making them more difficult to be reliably recognized. Also, unlike the prokaryotic genomes, eukaryotic genes have introns and the percentage of coding DNA is much lower (Mathé *et al.*, 2002). Two classic examples of signals identified by eukaryotic gene finders are CpG islands and binding sites for a poly(A) tail.

Commonly used *ab initio* gene finding programs, are trained with a set of predicted genes of strains related to the query genome using Hidden Markov Model algorithms. These Markov models are probabilistic models. They are a well-known tool for analyzing biological sequence data, and the predominant model for microbial sequence analysis is a fixed-order Markov chain (Borodovsky and McIninch, 1993; Borodovsky *et al*., 1995; Salzberg *et al*., 1998). Their effectiveness in modeling the correlations between adjacent symbols, domains, or events, have been extensively used in various fields, especially in the analysis of language text to recognize seminal words (Yoon, 2009). They are at the heart of a diverse range of programs, including gene finding, profile searches, multiple sequence alignment and regulatory site identification (Eddy, 2004). For the purpose of gene finding, it consists of states corresponding to a biological meaning (e.g. intron, exon, splice site) and allows transitions between these states in a biologically meaningful way (e.g. an acceptor splice site must follow an intron). The model defines a probability distribution on DNA sequences together with their gene structure (Stanke and Waack, 2003). Examples of *ab initio* programs are GENSCAN (Burge and Karlin, 1997), GENEID (Parra *et al*., 2000), GeneMarks (Besemer *et al*., 2001), Easygene (Larsen and Krogh, 2003) and Glimmer (Delcher *et al*., 1999).

The accuracy of gene prediction programs is generally assessed using two parameters, sensitivity and specificity. Sensitivity is the percentage of real genes in a test genome that are correctly identified by the program. Specificity is the percentage of predicted genes that match with the real genes. These two parameters have a reciprocal relationship; increase in sensitivity in general causes a decrease in specificity, and vice versa. The prediction program uses a threshold to balance these two parameters. Gene prediction algorithms such as GeneMarkS, and EasyGene have attained sensitivity and specificity of over 90% on nearly all available prokaryotic genomes. A previous version of the Glimmer program reported a sensitivity of 98–99%, however, it generated many false positives; that is, the program predicted more genes than the genome actually has. Recent versions of Glimmer (Glimmer v.3) have improved these results (Delcher *et al*., 2007).

Once the ORFs are determined, the annotation of the gene/gene prediction can be done in a number of ways using either the DNA or protein sequences. As public databases are growing and diversifying their contents, more complete information is

able to be assigned to an ORF, in a way that today it is possible to assign different levels of annotation to a single coding region. Information about protein domains, their interactions with other proteins, spatial location in the cell and alterations during evolutionary time can be represented in the different levels of genome annotation (Reed *et al.*, 2006). Several ontologies have been developed for genome annotation and expression analysis such as the Gene Ontology (GO) (Ashburner *et al.*, 2000; The Gene Ontology Consortium, 2010) and KEGG Orthology (KO) (Mao *et al.*, 2005). These ontologies try to provide not only a consistent vocabulary, but also provide a mechanism to support the comparison of genes and annotations across many genomes.

The quality of genome annotation, in general, varies significantly, depending on whether the sequence was exclusively automatically annotated or manually curated by a skilled scientist. Manually curated annotations generally show a higher level of accuracy because the automatic annotation pipelines are often based on simplified rules to determine the accuracy of an annotated CDS. Therefore, re-annotation of genomes becomes more and more important for any genomics project that includes sequences from various sources and dates. At this point, it is important to choose a good tool for the hosting and curating of a genome database. The Pathway Tools software suite is a comprehensive system to identify, curate, store, and publish biochemical pathways on the Web in the form of pathway genome databases (PGDB) (Karp *et al.*, 2002). The pathway genome database from Pathway Tools, can integrate information about genes and proteins at different levels, ,including the metabolic compounds, reactions, biochemical pathways, enzymes, and enzyme complexes (Mueller *et al.*, 2003). The ability to integrate all of this information using the same tool allows the possibility to go beyond simple gene annotation and to reconstruct the full metabolic/biochemical network of an organism. With the reconstructed metabolic network, a better understanding of the functioning organism can be obtained by relating its phenotypic characteristics to metabolic functional pathways. Metabolic pathways also provide information that can be utilized in metabolic engineering for the optimization or suppression of certain pathways as to control and prevent undesirable reactions. Finally, metabolic network comparisons between different strains or species can also lead to the discovery of new pathways or functional proteins.

Sequence homology search methods

It is practically impossible to speak about bioinformatics without speaking of the BLAST (Basic Local Alignment Search Tool) algorithm (Altschul *et al*., 1997). The BLAST programs (Table 4) are widely used tools for searching protein and DNA databases for sequence similarities. It uses a rapid search algorithm and gives a list of pairwise alignments which are ranked by a score. The reason why this method has become an essential step in bioinformatic analysis, is because, sequence comparison is a reliable method for identification of putative biological functions for newly sequenced genes.

Table 5. Blast programs summary

BLAST program	query vs. database
blastn	nucleotide vs. nucleotide
blastp	protein vs. protein
psi-blast	protein vs. distant relative protein
blastx	six-frame nucleotide translated sequence vs. protein
tblastx	nucleotide six-frame translation vs. nucleotide six-frame translation
tblastn	protein vs. Nucleotide six-frame translation
megablast	large number of query sequences

The study of biological sequence data has revealed that genes in an organism have significant similarity with those in other organisms that diverged many years ago. This is possible because genomes of related species exhibit similarities in functional elements that have undergone little change since the species common ancestor. In a biological context, two entities are homologous if they are related because they share a common ancestor (Wagner, 1989). When homology is applied to genes (Fitch, 1970), at least two fundamentally different subclasses must be distinguished (Table 5): paralogy, the relationship between genes that have originated by a gene duplication event; and orthology, which refers to genes that originated by a speciation event or vertical evolutionary descent (Theissen, 2002; Koonin, 2005).

Table 6. Homology: terms and definitions

Homologs	**Genes sharing common origing**
Orthologs	Genes originating from a single ancestral gene in the last common ancestor of the compared genomes
Pseudoorthologs	Genes that currently are paralogs but appear to be orthologous due to differential, lineage-specific gene loss
Xenologs	Homologous genes acquired via XGD by one or both of the compared species but appearing to be orthologous in pairwise genome comparisons
Co-orthologs	Two or more genes in one lineage that are, collectively, orthologous to one or more genes in another lineage due to a lineage-specific duplication(s). Members of a co-orthologous gene set are inparalogs relative to the respective speciation event
Paralogs	**Genes related by duplication**
Inparalogs (symparalogs)	Paralogous genes resulting from a lineage-specific duplication(s) subsequent to a given speciation event (defined only relative to a speciation event, no absolute meaning)
Outparalogs (alloparalogs)	Paralogous genes resulting from a duplication(s) preceding a given speciation event (defined only relative to a speciation event, no absolute meaning)
Pseudoparalogs	Homologous genes that come out as paralogs in a single-genome analysis but actually ended up in the given genome as a result of a combination of vertical inheritance and HGT

In xenologous gene displacement (XGD), the original gene from a given set of orthologs is displaced by a member of the same set of orthologs from a different lineage (Koonin et al., 2001; Novichkov et al., 2004; Koonin, 2005)

When a BLAST analysis is performed, using any of its variants, the best hit of a query sequence is found (if it exists) in a database of sequences, by scoring the local alignments between the query and the matched sequences in the database. Along with the list of ordered matches, the program provides other output data, as well as the percent of similarity that both sequences share. This similarity measure reflects the likelihood of two sequences being evolutionarily related; thus, when a similarity search finds a statistically significant match, we can confidently infer that the two sequences are homologous; but if no statistically significant match is found in a database, we cannot be certain that no homologs are present, because homologous sequences do not always share significant sequence similarity (Pearson, 2013).

There are several sequence similarity search tools besides the BLAST program, like FASTA and SSEARCH (Pearson and Lipman, 1988). These programs are implementations of the Smith-Waterman algorithm that was first proposed by these scientists in 1981 (Smith and Waterman, 1981) and optimized by Gotoh (Gotoh, 1982).

To determine similar regions between two nucleotide or protein sequences, the Smith-Waterman algorithm, instead of looking at the total sequence, compares segments of all possible lengths and optimizes the similarity measure (Huang *et al.*, 2015). Global alignments seek to align every residue in one sequence with every residue in a second sequence. The advantage of using local alignment tools is that local alignments allow one to identify conserved domains in proteins, which may not extend over the entire sequence and in addition, the statistics of local similarity scores are easy to interpret and estimate whether an alignment similarity score would be expected by chance. The other key output of a local alignment is the E-value (expected value) which is used to This parameter describes the number of hits one can "expect" to see by chance when searching a database of a particular size and, as the closer it is to zero, the more significant the match will be. The calculation of the E-value takes into account the length of the query sequence, hence high E-values make sense because shorter sequences have a higher probability of occurring in the database purely by chance. E-values lower than 10^{-8} indicate homology with a high degree of confidence.

Phylogenetic methods

The study of phylogenetics looks at the evolutionary relationship among groups of organisms and is based on the fundamental principle that different species descended from common ancestors (Edwards and Cavalli-Sforza, 1965). At the DNA sequence level, these evolutionary relationships are determined using molecular differences. The comparison of related sequences is typically represented as a multiple sequence alignment which is an alignment constructed with more than two sequences. The similarities between all pairs of sequences can be used to construct a *phylogenetic tree*, summarizing the most likely ancestry of the sequences, linking them hierarchically from the most closely related pair to the most distantly related group. The phylogenetic tree thus, represents a hypothesis of the

order in which evolutionary events are assumed to have occurred (Edwards and Cavalli-Sforza, 1965). Four steps are necessary in phylogenetic analyses of molecular sequences, namely, i) the selection of molecular markers, ii) obtaining sequences from selected molecules, iii) a multiple sequence alignment and iv) tree reconstruction.

Different mechanisms of molecular evolution in bacteria have led to today's biodiversity, these mechanisms include, mutation, gene duplication, genome reorganization, genetic exchange through recombination, reassortment and horizontal gene transfer (Vandamme, 2003; Lerat *et al.*, 2005). All this suggests that genes are good evolutionary markers, but all sequences are not of equal value in determining phylogenetic relationships. To be a useful chronometer, a molecule has to meet certain specifications as to clock-like behavior (changes in its sequence have to occur as randomly as possible), range (rates of change have to be commensurate with the spectrum of evolutionary distances being measured), size (the molecule has to be large enough to provide an adequate amount of information), and to be a "smooth-running" chronometer (Woese, 1987).

The 16S ribosomal RNA gene has been, for many decades, considered the "ultimate molecular chronometer" (Fox *et al.*, 1977; Woese and Fox, 1977; Woese, 1987; Patel, 2001; Coenye and Vandamme, 2003). The gene contains highly conserved regions, as well as hypervariable regions that can be used to measure evolutionary distances through multiple sequence alignments. Its analysis is relatively easy has enabled the accurate classification of bacteria into families and genera. However, the major advantage of 16S rRNA, the presence and conservation among all known organisms, is also their biggest drawback, because it doesn't have sufficient resolving power to separate very closely-related organisms (>97 % similarity) and therefore, analyses of this gene are not suitable for distinguishing prokaryotes at the species level (Ludwing, 2010). Considering the limitations of the 16S rRNA gene phylogenies in bacterial taxonomy, a multi locus sequence analysis (MLSA) using different protein coding genes, has been preferred for studying closely-related species of bacteria (Adékambi and Drancourt, 2004; Mutch and Young, 2004; Konstantinidis and Tiedje, 2005; Naser *et al.*, 2005; Martens *et al.*, 2008; Pascual *et al.*, 2010; Serrano *et al.*, 2010; Fernández *et al.*, 2015). The MLSA consists of the concatenation of multiple gene sequences. The genes included in the

analysis have to be conserved among all the investigated genomes, therefore the genes supposedly acquired by HGT are outside of the analysis, as well as the genes with more than one copy (paralogs) (Lerat *et al*., 2003; Thompson *et al*., 2009). However, even with its advantages over other approaches MLSA, just as 16S rRNA analysis, disregards certain important aspects of bacterial evolution, such as HGT, lineage-specific gene expansion or lineage-specific gene loss. Therefore, even if conserved genes reflect the phylogeny of the *genetic backbone* of a species it does not reflect its complete evolutionary history.

Another important issue in phylogenetic analysis is the choice between DNA or amino acid sequences. Taxonomic comparisons show that the genes of closely related species usually only differ from one another by point mutations. As already mentioned herein, these differences are usually found in the third codon position of ORFs. The 3rd codon position has a faster evolutionary rate than the 1st and 2nd codon positions. The redundancy of the genetic code ensures that nucleotide sequences usually evolve more quickly than the proteins they encode. Thus, differences between closely related species are assessed most sensitively by analysis of their nucleotide sequences. Once the multisequence alignment of the selected molecular markers has been performed, the phylogenetic tree must be constructed (Peplies *et al*., 2008; Koonin *et al*., 2011). To that end, multiple sequence alignment is further implemented by phylogenetic reconstruction methods that can be divided into two main categories; parsimony methods and model-based methods. Parsimony, the most widely used method, is based on the principle that the simplest explanation is the explanation best supported by the current data. Accordingly, the optimal phylogenetic tree is the one that minimizes the number of *ad-hoc* hypotheses required to explain the data. Model-based methods, such as maximum likelihood and Bayesian inference assume a model of DNA sequence evolution and then find the tree that best fits the model (Holder and Lewis, 2003; Johnson and Omland, 2004; Shapiro, 2005; Rizzo and Rouchka, 2007). The main advantage of tree-based methods is their sensitivity; however trees are generally computationally expensive (time consuming) when the dataset is too large. Tree construction performance depends on the accuracy of multiple sequence alignments, which cannot be assured when a larger number of sequences are introduced, or when dealing with multi-domain proteins. Also, they are sensitive to the number of gaps in the alignment,

which can lead to a reduced amount of information, from which the evolution model will be created.

4. *The pangenome concept*

Experiments of pulse-field gel electrophoresis in the 1990s, highlighted the fact that genomes of the same species can vary in size (Bergthorsson and Ochman, 1995; Thong *et al*., 1997). Also, as mentioned earlier, whole genome sequence comparison of isolates of the same species reveals a degree of variability of intra-species genomic content much higher than expected (Mira *et al*., 2010; Laing *et al*., 2011). This could be due to the absence of preserved molecular mechanisms of sexual reproduction, which makes bacterial genomes prone to constant genomic rearrangement, lineage-specific gene loss, duplication of genes, and horizontal acquisition of foreign DNA. Such is the case for the species *Pseudomonas putida*, in which there are strains with quite different metabolic and ecological capabilities. Basically, it appears that the genome complex that characterizes the bacterial species is much larger than that contained by any single cell. The pangenome defines the entire genomic repertoire of a given phylogenetic clade and encodes for all possible lifestyles carried out by the organisms (Vernikos *et al*., 2014). A bacterial species therefore, can be defined by its pangenome or the set of all genes belonging to it (Medini *et al*., 2005). The pangenome can be split into three sets of genes; the core genes, accessory genes and unique genes. Pangenome core genes are those shared by all the genomes. These genes likely encode functions related to the basic cellular biology, and build the genetic backbone of the bacterial species. The core genome mainly consists of housekeeping genes, i.e., genes involved in maintaining basic metabolic functions, replication of DNA, the synthesis of the cell envelope and transport binding proteins. They are thus mostly vertically transferred from parent to progeny (Tettelin *et al*., 2005; Hiller *et al*., 2007). The accessory genome is the set of genes constituted by genes that are present in at least two, but not all the genomes of the study. They are presumably not involved in essential metabolic functions but provide an important pool for genetic variability. Accessory genes of a species' pangenome are often involved in adaptation to a specific niche or manifestation of a specific phenotype (Medini *et al*., 2005; Tettelin *et al*., 2005; Mira *et al*., 2010; Udaondo *et al*., 2015). Finally, the unique genes are those present in only one isolate of a group of bacteria. Thus, these genes show no or only weak homology to genes

of any other isolate of the same species, i.e., they are specific to one genome. Where the functional annotation of unique genes is possible it reveals a high percentage of genes related to phage genes and mobile genetic elements.

A pangenome can also be classified as open or closed. In an open pangenome its size increases indefinitely when adding new genomes; thus sequencing additional strains will likely yield novel genes. Conversely, in a closed pangenome, adding new genomes will not lead to the discovery of new coding capabilities. The closed pangenomes are highly conserved, and are typically associated with bacterial species that live in select niches, where they are excluded from the overall microbial gene pool or have a diminished capacity to acquire genes, such as *Bacillus anthracis* and *Mycobacterium tuberculosis* (Medini *et al*., 2005; Mira *et al*., 2010). By contrast an open pangenome is observed in bacterial species that can colonize and exploit several different environmental niches and can expand their accessory and pangenome through different means of lateral gene transfer (Medini *et al*., 2005; Tettelin *et al*., 2005; Mira *et al*., 2010). This is the case of the pangenome from the species *Streptococcus agalactiae, S. pyogenes, Staphylococcus aureus, Mycobacterium tuberculosis* and *Escheriria coli* (D'Auria *et al*., 2010; Dobrindt *et al*., 2010).

A pangenome study can provide a more comprehensive definition of the niches occupied by a given species in a community and provide insight about how variations in gene repertoires and functions contribute to the fitness of that species in an ecosystem and to the overall ecosystems robustness.

In this PhD I have explored the genome of single strains of the species *P. putida* and I have approached the *P. putida* pangenome for the first time. Strains of this species share a robust set of core genes that determine their capacity to move and colonize niches of planktonic and sessile cells; they have an aerobic metabolism efficient in energy generation that allows them to grow fast and gain and colonization advantages. Niche specific properties are linked to accessory genes that encode functions that confer some strains, but not all of them, specific ability to thrive as a key member of a niche.

References

Adékambi, T. and Drancourt, M. (2004) Dissection of phylogenetic relationships among 19 rapidly growing Mycobacterium species by 16S rRNA, *hsp65*, *sodA*, *recA* and *rpoB* gene sequencing. *International journal of systematic and evolutionary microbiology* **54**: 2095–105.

Allardet-servent, A., Michaux-charachon, S., Jumas-bilak, E., Karayan, L., and Ramuz, M. (1993) Presence of One Linear and One Circular Chromosome in the *Agrobacterium tumefaciens* C58 *Genome*. **175**: 7869–7874.

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., and Dj, L. (1997) Gapped BLAST and PSI- BLAST: a new generation of protein database search programs. *Nucleic acids Res* **25**: 3389–3402.

Amils, R., Biología, C. De, and Severo, M. (2015) Technological challenges to understanding the microbial ecology of deep subsurface ecosystems. *Environ Microbiol Reports* **7**(1): 9–10.

Andersson, D.I., Slechta, E.S., and Roth, J.R. (1998) Evidence that gene amplification underlies adaptive mutability of the bacterial lac operon. *Science (New York, N.Y.)* **282**: 1133–1135.

Angov, E. (2011) Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnology journal* **6**: 650–659.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**: 25–9.

Avery, O.T., MacLeod, C.M., and McCarty, M. (1944) Studies on the chemical nature of the substance inducing trnasformation of *Pneumococcal* types: Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *Journal of Experimental Medicine* **79**: 137–158.

Bagdasarian, M. and Timmis, K.N. (1982) Host: vector systems for gene cloning in *Pseudomonas*. *Current topics in microbiology and immunology* **96**: 47–67.

Bahl, H., Miiller, H., Behrens, S., Joseph, H., Narberhaus, F., Mikrobiologie, I., and Gi, D. (1995) Expression of heat shock genes in *Clostridium acetobutylicum*. **17**: 341–348.

Barbier, M., Damron, F.H., Bielecki, P., Suárez-Diez, M., Puchałka, J., Albertí, S., *et al*. (2014) From the environment to the host: re-wiring of the transcriptome of *Pseudomonas aeruginosa* from 22°C to 37°C. *PloS one* **9**: e89941.

Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., *et al*. (2002) Complete genome sequence of the model actinomycete *Streptomyces*. *Nature* **417**: 141–147.

Berg, O.G. and Kurland, C.G. (1997) Growth rate-optimised tRNA abundance and codon usage. *Journal of molecular biology* **270**: 544–50.

Bergey, D.H., Garrity, G.M., Krieg, N.R., and Staley, J.. (2005) The Proteobacteria. In, Brenner,D.J., Krieg,N.R., Staley,J.., and Garrity,G.M. (eds), *Bergey´s Manual of Systematic Bacteriology. Vol. 2*. Springer, New York, pp. 15–19.

Bergthorsson, U. and Ochman, H. (1995) Heterogeneity of genome sizes among natural isolates of Escherichia coli. *Journal of bacteriology* **177**: 5784–9.

Besemer, J., Lomsadze, A, and Borodovsky, M. (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic acids research* **29**: 2607–2618.

Borodovsky, M. and McIninch, J. (1993) Recognition of genes in DNA sequence with ambiguities. *Bio Systems* **30**: 161–171.

Borodovsky, M.Y., McIninch, J., Koonin, E. V, Rudd, K.E., Médigue, C., and Danchin, A. (1995) Detection of new genes in a bacterial genome using Markov models for three gene classes. **23**: 3554–3562.

Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology* **268**: 78–94.

Cairns, J. (1963) The bacterial chromosome and its manner of replication as seen by autoradiography. *Journal of molecular biology* **6**: 208–213.

Carbone, A., Zinovyev, A., and Kepes, F. (2003) Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* **19**: 2005–2015.

Che, D., Li, G., Mao, F., Wu, H., and Xu, Y. (2006) Detecting uber-operons in prokaryotic genomes. *Nucleic Acids Research* **34**: 2418–2427.

Chen, S.L., Lee, W., Hottes, A.K., Shapiro, L., and McAdams, H.H. (2004) Codon usage between genomes is constrained by genome-wide mutational processes. *Proceedings of the National Academy of Sciences of the United States of America*

101: 3480–5.

Coenye, T. and Vandamme, P. (2003) Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiology Letters* **228**: 45–49.

Cohan, F.M. (2002) What are bacterial species? *Annual review of microbiology* **56**: 457–87.

Consortium The Integrative HMP (iHMP) Research Network (2014) The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. *Cell host & microbe* **16**: 276–89.

Cordero, O.X. and Polz, M.F. (2014) Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Reviews Microbiology* **12**: 263–273.

Costerton, J.W. (1999) Introduction to biofilm. *International journal of antimicrobial agents* **11**: 217–21; discussion 237–9.

Crook, M.B., Mitra, S., Ané, J.-M., Sadowsky, M.J., and Gyaneshwar, P. (2013) Complete Genome Sequence of the Sesbania Symbiont and Rice Growth-Promoting Endophyte *Rhizobium* sp. Strain IRBG74. *Genome announcements* **1(6)**: pii: e00934-13

D'Auria, G., Jiménez-Hernández, N., Peris-Bondia, F., Moya, A., and Latorre, A. (2010) *Legionella pneumophila* pangenome reveals strain-specific virulence factors. *BMC genomics* **11**: 181.

Daniels, C., Godoy, P., Duque, E., Molina-Henares, M.A., de la Torre, J., Del Arco, J.M., *et al*. (2010) Global regulation of food supply by *Pseudomonas putida* DOT-T1E. *Journal of bacteriology* **192**: 2169–81.

Darling, A.E., Mau, B., and Perna, N.T. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one* **5**: e11147.

Delcher, A.L., Bratke, K. a., Powers, E.C., and Salzberg, S.L. (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**: 673–679.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic acids research* **27**: 4636–41.

Delgado, A., Duque, E., and Ramos, J.L. (1992) Behavior in agricultural soils of a recombinant *Pseudomonas* bacterium that simultaneously degrades alkyl- and haloaromatics. *Microbial releases : viruses, bacteria, fungi* **1**: 23–28.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014) Ten years of next-

generation sequencing technology. *Trends in Genetics* **30.**:

Dionisio, F., Conceic, I.C., Marques, A.C.R., Fernandes, L., and Gordo, I. (2005) The evolution of a conjugative plasmid and its ability to increase bacterial fitness. *Biol Lett* **1**:250–252.

Dobrindt, U., Chowdary, M.G., Krumbholz, G., and Hacker, J. (2010) Genome dynamics and its impact on evolution of *Escherichia coli. Medical Microbiology and Immunology* **199**: 145–154.

Dong, H., Nilsson, L., and Kurland, C.G. (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *Journal of molecular biology* **260**: 649–663.

Doolittle, W.F. and Papke, R.T. (2006) Genomics and the bacterial species problem. *Genome biology* **7**: 116.

Dziewit, L. and Bartosik, D. (2015) Comparative Analyses of Extrachromosomal Bacterial Replicons, Identification of Chromids, and Experimental Evaluation of their Indispensability. In, Mengoni, A., Galardini, M., and Fondi, M. (eds), *Bacterial Pangenomics SE 2*, Methods in Molecular Biology. Springer New York, pp. 15–29.

Eddy, S.R. (2004) What is a hidden Markov model? *Nature biotechnology* **22**: 1315–1316.

Edwards, A.W. and Cavalli-Sforza, L.L. (1965) A method for cluster analysis. *Biometrics* **21**: 362–375.

Egan, E.S., Fogel, M.A., and Waldor, M.K. (2005) MicroReview Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Molecualr Microbiology* **56**(5): 1129–1138.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., *et al*. (2009) Real-time DNA sequencing from single polymerase molecules. *Science (New York, N.Y.)* **323**: 133–138.

Espinosa-Urgel, M., Kolter, R., and Ramos, J.-L. (2002) Root colonization by *Pseudomonas putida:* love at first sight. *Microbiology (Reading, England)* **148**: 341–3.

Espinosa-Urgel, M. and Ramos, J.L. (2004) Cell density-dependent gene contributes to efficient seed colonization by *Pseudomonas putida* KT2440. *Applied and environmental microbiology* **70**: 5190–5198.

Fernández, M., Porcel, M., de la Torre, J., Molina-Henares, M.A., Daddaoua, A., Llamas, M.A., *et al*. (2015) Analysis of the pathogenic potential of nosocomial

Pseudomonas putida strains. *Frontiers in microbiology* **6**: 871.

Fitch, W.M. (1970) Distinguishing homologous from analogous proteins. *Systematic zoology* **19**: 99–113.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., *et al*. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science (New York, N.Y.)* **269**: 496–512.

Fox, G.E., Magrum, L.J., Balch, W.E., Wolfe, R.S., and Woese, C.R. (1977) Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proceedings of the National Academy of Sciences* **74**: 4537–4541.

Fraser, C., Alm, E.J., Polz, M.F., Spratt, B.G., and Hanage, W.P. (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**: 741–746.

Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., *et al*. (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. **390.**:

Galardini, M., Pini, F., Bazzicalupo, M., Biondi, E.G., and Mengoni, A. (2013) Replicon-dependent bacterial genome evolution: the case of *Sinorhizobium meliloti*. *Genome biology and evolution* **5**: 542–58.

Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., *et al*. (2009) Rfam: updates to the RNA families database. *Nucleic acids research* **37**: D136–40.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., *et al*. (2005) Opinion: Re-evaluating prokaryotic species. *Nature reviews. Microbiology* **3**: 733–9.

Godoy, P., Molina-Henares, A.J., de la Torre, J., Duque, E., and Ramos, J.L. (2010) Characterization of the RND family of multidrug efflux pumps: in silico to in vivo confirmation of four functionally distinct subgroups. *Microbial biotechnology* **3**: 691–700.

Goodman, A.L. and Gordon, J.I. (2010) Our unindicted co-conspirators: human metabolism from a microbial perspective. *Cell Metab.* **12**: 111–116.

Gotoh, O. (1982) An improved algorithm for matching biological sequences. *Journal of molecular biology* **162**: 705–708.

Gouy, M. and Gautier, C. (1982) Codon usage in bacteria: correlation with gene

expressivity. *Nucleic acids research* **10**: 7055–7074.

Grantham, R. (1980) Working of the genetic code. *Trends in Biochemical Sciences* **5**: 327–331.

Grosjean, H. and Fiers, W. (1982) Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* **18**: 199–209.

Guldhe, A., Singh, B., Mutanda, T., Permaul, K., and Bux, F. (2015) Advances in synthesis of biodiesel via enzyme catalysis: Novel and sustainable approaches. *Renewable and Sustainable Energy Reviews* **41**: 1447–1464.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004) Codon bias and heterologous protein expression. *Trends in biotechnology* **22**: 346–53.

Han, K., Li, Z., Peng, R., Zhu, L., Zhou, T., Wang, L., *et al*. (2013) Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline. *Nature Scientific Reports* **3**:2101

Harrison, P.W., Lower, R.P.J., Kim, N.K.D., and Young, J.P.W. (2010) Introducing the bacterial “ chromid ”: not a chromosome, not a plasmid. *Trends in Microbiology* **18**: 141–148.

Hendrickson, H., Slechta, E.S., Bergthorsson, U., Andersson, D.I., and Roth, J.R. (2002) Amplification-mutagenesis: evidence that “directed” adaptive mutation and general hypermutability result from growth with a selected gene amplification. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 2164–2169.

Hershberg, R. and Petrov, D. A (2008) Selection on codon bias. *Annual review of genetics* **42**: 287–299.

Hiller, N.L., Janto, B., Hogg, J.S., Boissy, R., Yu, S., Powell, E., *et al*. (2007) Comparative Genomic Analyses of Seventeen *Streptococcus pneumoniae* Strains: Insights into the Pneumococcal Supragenome. *Journal of Bacteriology* **189**: 8186–8195.

Holder, M. and Lewis, P.O. (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nature reviews. Genetics* **4**: 275–284.

Homma, K., Fukuchi, S., Nakamura, Y., Gojobori, T., and Nishikawa, K. (2007) Gene cluster analysis method identifies horizontally transferred genes with high reliability and indicates that they provide the main mechanism of operon gain in 8 species of γ-Proteobacteria. *Molecular biology and evolution* **24**: 805–13.

Huang, L.T., Wu, C.-C., Lai, L.-F., and Li, Y.-J. (2015) Improving the Mapping of Smith-Waterman Sequence Database Searches onto CUDA-Enabled GPUs. *BioMed Research International* **2015**: 1–10.

Huang, S.C., Burne, R. a., and Chen, Y.Y.M. (2014) The pH-Dependent expression of the urease operon in *Streptococcus salivarius* is mediated by CodY. *Applied and Environmental Microbiology* **80**: 5386–5393.

Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular biology and evolution* **2**: 13–34.

Isken, S. and de Bont, J.A.M. (1998) Bacteria tolerant to organic solvents. *Extremophiles* **2**: 229–238.

Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3**: 318–356.

Jain, R., Rivera, M.C., Moore, J.E., and Lake, J.A. (2002) Horizontal gene transfer in microbial genome evolution. *Theoretical population biology* **61**: 489–495.

Janda, J.M. and Abbott, S.L. (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of clinical microbiology* **45**: 2761–4.

Jemli, S., Ayadi-Zouari, D., Hlima, H. Ben, and Bejar, S. (2014) Biocatalysts: application and engineering for industrial purposes. *Critical Reviews in Biotechnology* **8551**: 1–13.

Johnson, J. and Omland, K. (2004) Model selection in ecology and evolution. *Trends in Ecology & Evolution* **19**: 101–108.

Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., and Crook, D.W. (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews* **33**: 376–393.

Karp, P.D., Paley, S., and Romero, P. (2002) The Pathway Tools software. *Bioinformatics (Oxford, England)* **18 Suppl 1**: S225–S232.

Kellerhals, M.B., Kessler, B., and Witholt, B. (1999) Closed-loop control of bacterial high-cell-density fed-batch cultures: production of mcl-PHAs by *Pseudomonas putida* KT2442 under single-substrate and cofeeding conditions. *Biotechnology and bioengineering* **65**: 306–15.

Kim, J., Jeon, C.O., and Park, W. (2008) Dual regulation of zwf-1 by both 2-keto-3-deoxy-

6-phosphogluconate and oxidative stress in *Pseudomonas putida*. *Microbiology* **154**: 3905–3916.

Kim, L.-H. and Lee, S.-S. (2011) Isolation and characterization of ethylbenzene degrading *Pseudomonas putida* E41. *Journal of microbiology (Seoul, Korea)* **49**: 575–84.

Kimball, S., Mattis, P., and and The Gimp Team (2014) Gimp 2.8.14.

Konstantinidis, K.T. and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 2567–72.

Koonin, E. V (2005) Orthologs, paralogs, and evolutionary genomics. *Annual review of genetics* **39**: 309–338.

Koonin, E. V, Makarova, K.S., and Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annual review of microbiology* **55**: 709–42.

Koonin, E. V., Puigbò, P., and Wolf, Y.I. (2011) Comparison of Phylogenetic Trees and Search for a Central Trend in the "Forest of Life." *Journal of Computational Biology* **18**: 917–924.

Kraak, M.N., Smits, T.H.M., Kessler, B., and Witholt, B. (1997) Polymerase C1 Levels and Poly (R-3-Hydroxyalkanoate) Synthesis in Wild-Type and Recombinant *Pseudomonas* Strains. **179**: 4985–4991.

Krause, A., Ramakumar, A., Bartels, D., Battistoni, F., Bekel, T., Boch, J., *et al*. (2006) Complete genome of the mutualistic, N2-fixing grass endophyte *Azoarcus* sp. strain BH72. *Nature biotechnology* **24**: 1385–91.

Kurland, C.G. and Ehrenberg, M. (1987) Growth-optimizing accuracy of gene expression. *Annual review of biophysics and biophysical chemistry* **16**: 291–317.

Kwak, Y. and Shin, J.-H. (2015) Complete genome sequence of *Burkholderia pyrrocinia* 2327(T), the first industrial bacterium which produced antifungal antibiotic pyrrolnitrin. *Journal of biotechnology* **211**: 3–4.

Lagerkvist, U. (1978) "Two out of three": an alternative method for codon reading. *Proceedings of the National Academy of Sciences of the United States of America* **75**: 1759–62.

Laing R. C., Zhang Y., Thomas, J. E., Gannon, V. P. J. (2011) Everything at once: Comparative analysis of the genomes of Bacterial pathogens. *Veterinary*

Microbiology **153**: 13-26

Lambert, G. and Kussel, E. (2014) Memory and Fitness Optimization of Bacteria under Fluctuating Environments. *PLoS Genetics* **10**: e1004556.

Lambros, S., Panos, I., and Spiridon, L. (2007) Coding potential prediction in *Wolbachia* using artificial neural networks. *In silico biology* **7**: 105–13.

Lan, R. and Reeves, P.R. (2000) Intraspecies variation in bacterial genomes : the need for a species genome concept. *Trends in Microbiology* **8**: 396–401.

Larsen, T.S. and Krogh, A. (2003) EasyGene-a prokaryotic gene finder that ranks ORFs by statistical significance. *BMC bioinformatics* **4**: 21.

Lathe, W.C., Snel, B., and Bork, P. (2000) Gene context conservation of a higher order than operons. *Trends in Biochemical Sciences* **25**: 474–479.

Leblond, P., Redenbach, M., and Cullum, J. (1993) Physical Map of the *Streptomyces lividans* 66 Genome and Comparison with That of the Related Strain *Streptomyces coelicolor* A3 (2). **175**: 3422–3429.

Lerat, E., Daubin, V., and Moran, N.A. (2003) From Gene Trees to Organismal Phylogeny in Prokaryotes:The Case of the γ-Proteobacteria. *PLoS Biology* **1**: e9.

Lerat, E., Daubin, V., Ochman, H., and Moran, N.A. (2005) Evolutionary origins of genomic repertoires in bacteria. *PLoS biology* **3**: e130.

Lercher, M.J. and Pál, C. (2008) Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Molecular biology and evolution* **25**: 559–67.

Liao, Y. C., Lin, S.-H., and Lin, H.-H. (2015) Completing bacterial genome assemblies: strategy and performance comparisons. *Scientific Reports* **5**: 8747.

Loman, N.J., Constantinidou, C., Chan, J.Z.M., Halachev, M., Sergeant, M., Penn, C.W., *et al*. (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology* **10**: 599–606.

Ludwing, W. (2010) Molecular Phylogeny of Microorganisms: Is rRNA still a usseful marker? In, Oren,A. and Papke,R.T. (eds), *Molecular Phylogeny of Microorganisms*. Caister Academic Press, Norfolk, pp. 65–83.

Luscombe, N.M., Greenbaum, D., and Gerstein, M. (2001) What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine* **40**: 346–358.

Mahar, A.N., Darban, D.A., Gowen, S.R., Hague, N.G.M., Jan, N.D., Munir, M., and Mahar, A.Q. (2005) Use of Entomopathogenic Bacterium *Pseudomonas putida* (Enterobacteriaceae) and its Secretion Against Greater Wax Moth, Galleria mellonella Pupae. *Journal of Entomo* **2**: 77–85.

Mao, X., Cai, T., Olyarchuk, J.G., and Wei, L. (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**: 3787–3793.

Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P., and Willems, A. (2008) Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former Sinorhizobium). *International journal of systematic and evolutionary microbiology* **58**: 200–14.

Martínez-Antonio, A. and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol* **6**(5):482–489.

Martins Dos Santos, V. a P., Heim, S., Moore, E.R.B., Strätz, M., and Timmis, K.N. (2004) Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. *Environmental Microbiology* **6**: 1264–1286.

Mathé, C., Sagot, M.-F., Schiex, T., and Rouzé, P. (2002) Current methods of gene prediction, their strengths and weaknesses. *Nucleic acids research* **30**: 4103–4117.

Matilla, M.A., Pizarro-Tobias, P., Roca, A., Fernández, M., Duque, E., Molina, L., *et al*. (2011) Complete genome of the plant growth-promoting rhizobacterium *Pseudomonas putida* BIRD-1. *Journal of Bacteriology* **193**: 1290.

Mazzeo, D.E., Levy, C.E., de Angelis, D. de F., and Marin-Morales, M.A. (2010) BTEX biodegradation by bacteria from effluents of petroleum refinery. - PubMed - NCBI. *Sci Total Environ* **408**: 4334–4340.

McCutcheon, J.P., McDonald, B.R., and Moran, N. A. (2009) Origin of an Alternative Genetic Code in the Extremely Small and GC–Rich Genome of a Bacterial Symbiont. *PLoS Genetics* **5**: e1000565.

McDaniel, L.D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K.B., and Paul, J.H. (2010) High frequency of horizontal gene transfer in the oceans. *Science (New York, N.Y.)* **330**: 50.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005) The microbial pan-genome. *Current opinion in genetics & development* **15**: 589–94.

Migula W (1894) Ueberein neues system der bakterien. *Arb. Bakteriol. Ins. Karisruhe***. 1**: 235-238

Miller, J.R., Koren, S., and Sutton, G. (2010) Assembly algorithm for Next-Ganeration Sequencing data. *Genomics* **95**: 315–327.

Miller, S.H., Browne, P., Prigent-Combaret, C., Combes-Meynet, E., Morrissey, J.P., and O'Gara, F. (2010) Biochemical and genomic comparison of inorganic phosphate solubilization in *Pseudomonas* species. *Environmental Microbiology Reports* **2**: 403–411.

Mira, A., Martín-Cuadrado, A.B., D'Auria, G., and Rodríguez-Valera, F. (2010) The bacterial pan-genome: A new paradigm in microbiology. *International Microbiology* **13**: 45–57.

Miyamoto, M., Motooka, D., Gotoh, K., Imai, T., Yoshitake, K., Goto, N., *et al*. (2014) Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes. *BMC genomics* **15**: 699.

Molina, L. (2000) Survival of *Pseudomonas putida* KT2440 in soil and in the rhizosphere of plants under greenhouse and environmental conditions. *Soil Biology and Biochemistry* **32**: 315–321.

Molina, L., Bernal, P., Udaondo, Z., Segura, A., and Ramos, J. (2013) Complete genome sequence of a *Pseudomonas putida* clinical isolate, strain H8234. *Genome Announcements* **1**: 2–3.

Molina-Henares, M.A., Ramos, J.L., and Espinosa-Urgel, M. (2006) A two-partner secretion system is involved in seed and root colonization and iron uptake by *Pseudomonas putida* KT2440. *Environmental Microbiology* **8**: 639–647.

Molina-Santiago, C., Daddaoua, A., Gómez-Lozano, M., Udaondo, Z., Molin, S., and Ramos, J.-L. (2015) Differential transcriptional response to antibiotics by *Pseudomonas putida* DOT-T1E. *Environmental microbiology*.

Mosqueda, G., Ramos-González, M.I., and Ramos, J.L. (1999) Toluene metabolism by the solvent-tolerant *Pseudomonas putida* DOT-T1 strain, and its role in solvent impermeabilization. *Gene* **232**: 69–76.

Mueller, L. a, Zhang, P., and Rhee, S.Y. (2003) AraCyc: a biochemical pathway database for Arabidopsis. *Plant physiology* **132**: 453–460.

Mutch, L.A. and Young, J.P.W. (2004) Diversity and specificity of *Rhizobium leguminosarum* biovar viciae on wild and cultivated legumes. *Molecular ecology*

13: 2435–44.

Nagarajan, N. and Pop, M. (2013) Sequence assembly demystified. *Nature reviews. Genetics* **14**: 157–67.

Nakazawa, T. (2002) Travels of a *Pseudomonas*, from Japan around the world. *Environmental Microbiology* **4**: 782–786.

Naser, S.M., Thompson, F.L., Hoste, B., Gevers, D., Dawyndt, P., Vancanneyt, M., and Swings, J. (2005) Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. *Microbiology (Reading, England)* **151**: 2141–50.

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A.P., *et al*. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environmental microbiology* **4**: 799–808.

Nogi, Y., Masui, N., and Kato, C. (1998) Photobacterium profundum sp. nov., a new, moderately barophilic bacterial species isolated from a deep-sea sediment. *Extremophiles* **2**: 1–7.

Novichkov, P.S., Omelchenko, M. V, Gelfand, M.S., Mironov, A. a, Wolf, Y.I., and Koonin, E. V (2004) Genome-wide molecular clock and horizontal gene ransfer in bacterial evolution. *Journal of Bacteriology* **186**: 6575–6585.

Ochman, H., Lawrence, J.G., and Groisman, E. a (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.

Pal, K.K., Tilak, K. V, Saxena, a K., Dey, R., and Singh, C.S. (2001) Suppression of maize root diseases caused by *Macrophomina phaseolina*, *Fusarium moniliforme* and *Fusarium graminearum* by plant growth promoting rhizobacteria. *Microbiological research* **156**: 209–223.

Palleroni, N.J. (1984) Gram negative aerobic rods and cocci, family I: *Pseudomonadaceae*. in Bergey's manual of systematic bacteriology, Williams a. N. R. Krieg and J. G. Holt (ed) Baltimore, Md.

Palleroni, N.J., Ballard, R.W., Ralston, E., and Doudoroff, M. (1972) Deoxyribonucleic acid homologies among some *Pseudomonas* species. *Journal of bacteriology* **110**: 1–11.

Pang, X., Zhou, X., Sun, Y., and Deng, Z. (2002) Physical Map of the Linear Chromosome of *Streptomyces hygroscopicus* 10-22 Deduced by Analysis of Overlapping Large Chromosomal Deletions. *Journal of Bacteriology* **184**: 1958–1965.

Parra, G., Blanco, E., and Guigó, R. (2000) GeneID in *Drosophila. Genome research* **10**: 511–5.

Pascual, J., Macián, M.C., Arahal, D.R., Garay, E., and Pujalte, M.J. (2010) Multilocus sequence analysis of the central clade of the genus *Vibrio* by using the 16S rRNA, *recA*, *pyrH, rpoD*, *gyrB*, *rctB* and *toxR* genes. *International journal of systematic and evolutionary microbiology* **60**: 154–65.

Patel, J.B. (2001) 16S rRNA Gene Sequencing for Bacterial Pathogen Identification in the Clinical Laboratory. *Molecular Diagnosis* **6**: 313–321.

Pearson, W.R. (2013) An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis [et al.]* **Chapter 3**: Unit3.1.

Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 2444–2448.

Peplies, J., Kottmann, R., Ludwig, W., and Glöckner, F.O. (2008) A standard operating procedure for phylogenetic inference (SOPPI) using (rRNA) marker genes. *Systematic and Applied Microbiology* **31**: 251–257.

Pisabarro, A., Correia, A., and Martin, J.F. (1998) Pulsed-field gel electrophoresis analysis of the genome of *Rhodococcus fascians*: genome size and linear and circular replicon composition in virulent and avirulent strains. *Curr Microbiol* **36**: 302–308.

Plotkin, J.B. and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nature reviews. Genetics* **12**: 32–42.

Price, M.N., Huang, K.H., Arkin, A.P., and Alm, E.J. (2005) Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome research* **15**: 809–19.

Price, P.B. (1999) A habitat for psychrophiles in deep Antarctic ice.*PNAS* 97(3):1247-1251

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, S., Manichanh, C., *et al*. (2010) A human gut microbial gene catalog established by metagenomic sequencing. *Nature* **464**: 59–65.

Quax, T.E.F., Claassens, N.J., Söll, D., and van der Oost, J. (2015) Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular Cell* **59**: 149–161.

Rainey, P.B. (1999) Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. *Environmental microbiology* **1**: 243–57.

Ramos, J., Duque, E., Huertas, M., and Haidour, A. (1995) Isolation and expansion of the catabolic potential of a *Pseudomonas putida* strain able to grow in the presence of high concentrations of aromatic hydrocarbons. *J. Bacteriol.* **177**: 3911–3916.

Ramos, J.L., Díaz, E., Dowling, D., de Lorenzo, V., Molin, S., O'Gara, F., *et al*. (1994) The behavior of bacteria designed for biodegradation. *Bio/technology (Nature Publishing Company)* **12**: 1349–1356.

Ramos, J.L., Duque, E., Gallegos, M.T., Godoy, P., Ramos-Gonzalez, M.I., Rojas, A., *et al*. (2002) Mechanisms of solvent tolerance in gram-negative bacteria. *Annual review of microbiology* **56**: 743–68.

Ramos, J.-L., García-Salamanca, A., Molina-Santiago, C., and Udaondo, Z. (2013) Operon. In, *Encyclopedia of Genetics*. Elsevier Inc., pp. 176–180.

Ramos, J.L., Marqués, S., van Dillewijn, P., Espinosa-Urgel, M., Segura, A., Duque, E., *et al*. (2011) Laboratory research aimed at closing the gaps in microbial bioremediation. *Trends in Biotechnology* **29**: 641–647.

Ramos, J.L., Marqués, S., and Timmis, K.N. (1997) Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators. *Annual review of microbiology* **51**: 341–373.

Ramos, J.-L., Sol Cuenca, M., Molina-Santiago, C., Segura, A., Duque, E., Gómez-García, M.R., *et al*. (2015) Mechanisms of solvent resistance mediated by interplay of cellular factors in *Pseudomonas putida*. *FEMS microbiology reviews* **39**: 555–66.

Ramos, J.L., Stolz, A., Reineke, W., and Timmis, K.N. (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. *PNAS* **83**: 8467–8471.

Ramos-Díaz, M.A. and Ramos, J.L. (1998) Combined physical and genetic map of the *Pseudomonas putida* KT2440 chromosome. *Journal of Bacteriology* **180**: 6352–6363.

Ramos-González, M.I., Campos, M.J., and Ramos, J.L. (2005) Analysis of *Pseudomonas putida* KT2440 gene expression in the maize rhizosphere: in vivo [corrected] expression technology capture and identification of root-activated promoters. *Journal of bacteriology* **187**: 4033–41.

Ramos-González, M.I., Godoy, P., Alaminos, M., Ben-Bassat, A., and Ramos, J.L. (2001)

Physiological characterization of *Pseudomonas putida* DOT-T1E tolerance to p-hydroxybenzoate. *Applied and environmental microbiology* **67**: 4338–4341.

Reed, J.L., Famili, I., Thiele, I., and Palsson, B.O. (2006) Towards multidimensional genome annotation. *Nature reviews. Genetics* **7**: 130–141.

Riley, M.A. and Lizotte-Waniewski, M. (2009) Population genomics and the bacterial species concept. *Methods in Molecular Biology* **532**: 367–377.

Rizzo, J. and Rouchka, E.C. (2007) Review of Phylogenetic Tree Construction. University of Louisville. *Bioinformatics Review (report).*

Roca, A., Pizarro-Tobías, P., Udaondo, Z., Fernández, M., Matilla, M. a, Molina-Henares, M.A., *et al*. (2013) Analysis of the plant growth-promoting properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1. *Environmental microbiology* **15**: 780–794.

Rocha, E.P.C. (2004) Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome research* **14**: 2279–86.

Rojas, A., Rojas, A., Duque, E., Duque, E., Mosqueda, G., Mosqueda, G., *et al*. (2001) Three Ef ux Pumps Are Required To Provide Ef cient Tolerance to Toluene in *Pseudomonas putida* DOT-T1E *Jorunal of bacteriology* **183**: 3967–3973.

Rojo, F. (2010) Carbon catabolite repression in *Pseudomonas*: Optimizing metabolic versatility and interactions with the environment. *FEMS Microbiology Reviews* **34**: 658–684.

Rossello-Mora, R. and Amann, R. (2001) The species concept for procaryotes. *FEMS Microbiology Ecology*. **25**: 39–67.

Roymondal, U., Das, S., and Sahoo, S. (2009) Predicting gene expression level from relative codon usage bias: an application to *Escherichia coli* genome. *DNA research* **16**: 13–30.

Salzberg, S.L., Deicher, A.L., Kasif, S., and White, O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Research* **26**: 544–548.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., *et al*. (1977) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**: 687–695.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *PNAS* **74**: 5463–5467.

Seemann, T. (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.

Segura, A., Molina, L., Fillet, S., Krell, T., Bernal, P., Muñoz-Rojas, J., and Ramos, J.-L. (2012) Solvent tolerance in Gram-negative bacteria. *Current opinion in biotechnology* **23**: 415–21.

Segura, A. and Ramos, J.L. (2013) Plant-bacteria interactions in the removal of pollutants. *Current opinion in biotechnology* **24**: 467–473.

Segura, A. and Ramos, J.L. (2014) Biodegradative Bacteria Nojiri,H., Tsuda,M., Fukuda,M., and Kamagata,Y. (eds) Springer Japan, Tokyo.

Segura, A., Rodríguez-Conde, S., Ramos, C., and Ramos, J.L. (2009) Bacterial responses and interactions with plants during rhizoremediation. *Microbial biotechnology* **2**: 452–64.

Serrano, W., Amann, R., Rosselló-Mora, R., and Fischer, U. (2010) Evaluation of the use of multilocus sequence analysis (MLSA) to resolve taxonomic conflicts within the genus *Marichromatium*. *Systematic and Applied Microbiology* **33**: 116–121.

Setubal, J.C., Moreira, L.M., and da Silva, A.C.R. (2005) Bacterial phytopathogens and genome science. *Current opinion in microbiology* **8**: 595–600.

Shapiro, B. (2005) Choosing Appropriate Substitution Models for the Phylogenetic Analysis of Protein-Coding Sequences. *Molecular Biology and Evolution* **23**: 7–9.

Sharp, P.M. and Li, W.-H. (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* **15**: 1281–1295.

Shendure, J. and Ji, H. (2008) Next-generation DNA sequencing. *Nature biotechnology* **26**: 1135–45.

Siddiqui, Z.A. (2006) PGPR: Biocontrol and biofertilization Siddiqui,Z.A. (ed). Dordrecht (The Netherlands):Springer.

Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences. *Journal of Molecular Biology* **147**: 195–197.

Soucy, S.M., Huang, J., and Gogarten, J.P. (2015) Horizontal gene transfer: building the web of life. *Nature Reviews Genetics* **16**: 472–482.

Stanke, M. and Waack, S. (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**: 215–225.

Syvanen, M. (2012) Evolutionary Implications of Horizontal Gene Transfer. *Annual Review of Genetics* **46**: 341–358.

Tan, Y., Xu, M., Tan, X., Wang, X., Saikawa, Y., Nagahama, T., *et al*. (1997) Overexpression and large-scale production of recombinant L-methionine-alpha-deamino-gamma-mercaptomethane-lyase for novel anticancer therapy. *Protein expression and purification* **9**: 233–45.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., *et al*. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *PNAS* **102**: 13950–13955.

The Gene Ontology Consortium (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic acids research* **38**: D331–335.

The Inkscape Team (2015) Inkscape 0.91.

The Uniprot Consortium (2014) UniProt: a hub for protein information. *Nucleic Acids Research* **43**: D204–D212.

Theissen, G. (2002) Secret life of genes. *Nature* **415**: 741.

Thompson, C.C., Vicente, A.C.P., Souza, R.C., Vasconcelos, A.T.R., Vesth, T., Alves, N., *et al*. (2009) Genomic taxonomy of Vibrios. *BMC evolutionary biology* **9**: 258.

Thong, K.L., Puthucheary, S.D., and Pang, T. (1997) Genome size variation among recent human isolates of *Salmonella typhi*. *Research in microbiology* **148**: 229–235.

Timmis, K.N. (2002) *Pseudomonas putida*: a cosmopolitan opportunist par excellence. *Environmental microbiologyicrobiology* **4**: 779–781.

Toro, E. and Shapiro, L. (2010) Bacterial Chromosome Organization and Segregation. *Cold Spring Harb Perspect Biol* **2**: 1–15.

Treangen, T.J. and Rocha, E.P.C. (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genetics* **7**:e1001284**.**

Udaondo, Z., Duque, E., Fernández, M., Molina, L., de la Torre, J., Bernal, P., *et al*. (2012) Analysis of solvent tolerance in *Pseudomonas putida* DOT-T1E based on its genome sequence and a collection of mutants. *FEBS letters* **586**: 2932–8.

Udaondo, Z., Molina, L., Daniels, C., Gómez, M.J., Molina-Henares, M.A., Matilla, M.A., *et al*. (2013) Metabolic potential of the organic-solvent tolerant *Pseudomonas putidaz* DOT-T1E deduced from its annotated genome. *Microbial biotechnology* **6**: 598–611.

Udaondo, Z., Molina, L., Segura, A., Duque, E., and Ramos, J.L. (2015) Analysis of the core genome and pangenome of *Pseudomonas putida*. *Environmental microbiology*. DOI: 10.1111/1462-2920.13015.

Vernikos, G., Medini, D., Riley, D.R., and Tettelin, H. (2014) Ten years of pan-genome analyses. *Current opinion in microbiology* **23C**: 148–154.

Wagner, G.P. (1989) The Biological Homology Concept. *Annual Review of Ecology and Systematics* **20**: 51–69.

Wei, Q., Tarighi, S., Dötsch, A., Häussler, S., Müsken, M., Wright, V.J., *et al*. (2011) Phenotypic and genome-wide analysis of an antibiotic-resistant small colony variant (SCV) of *Pseudomonas aeruginosa*. *PloS one* **6**: e29276.

Weinel, C., Nelson, K.E., and Tümmler, B. (2002) Global features of the *Pseudomonas putida* KT2440 genome sequence. **4**: 809–818.

Wen, F., Nair, N.U., and Zhao, H. (2009) Protein engineering in designing tailored enzymes and microorganisms for biofuels production. *Current opinion in biotechnology* **20**: 412–9.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Perspective Prokaryotes : The unseen majority. **95**: 6578–6583.

Woese, C.R. (1987) Bacterial Evolution. *Microbiology* **51**: 221–271.

Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5088–90.

Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J., and van der Lelie, D. (2010) Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS microbiology reviews* **35**: 299–323.

Yang, C., Huang, C., Li, C., Tsay, Y., Lee, S., and Chen, C.W. (2002) The terminal proteins of linear *Streptomyces* chromosomes and plasmids : a novel class of replication priming proteins. *Molecular Microbiology* **43**(2): 297–305.

Yim, L.C., Hongmei, J., Aitchison, J.C., and Pointing, S.B. (2006) Highly diverse community structure in a remote central Tibetan geothermal spring does not display monotonic variation to thermal stress. **57**: 80–91.

Yoon, B. J. (2009) Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current genomics* **10**: 402–415.

You, Y., Shim, J., Cho, C.-H., Ryu, M.-H., Shea, P.J., Kamala-Kannan, S., *et al*. (2013) Biodegradation of BTEX mixture by *Pseudomonas putida* YNS1 isolated from oil-contaminated soil. *Journal of basic microbiology* **53**: 469–75.

AIM OF THIS THESIS
AND SPECIFIC OBJETIVES

The overall objective of this PhD is to define and analyze the pangenome of nine bacterial strains of the species *Pseudomonas putida* using bioinformatics approaches and e-tools. In a first step three *P. putida* strains, previously sequenced and automatically annotated by the group at the Experimental Station of the Zaidín, were manually re-annotated and analyzed to for enhanced quality. These three *P. putida* strains were the PGPR *P. putida* BIRD-1 strain, the solvent tolerant *P. putida* DOT-T1E strain and the opportunistic pathogen *P. putida* HB3267. The results of these analyses are presented in chapters I, II and III.

To obtain a more comprehensive analysis embracing the different phenotypic characteristics of members of the *P. putida* species, six other strains from the NCBI public database, were included in the analysis. These strains were chosen taking into account the high level of accuracy in the assembling and annotation process. The only strain in the study that failed to meet these characteristics (its genome was still divided into contigs) was the *P. putida* strain Idaho, which was selected because of its solvent tolerant trait. The analysis of the pangenome of the species *P. putida* can be found in chapters IV and V of this thesis.

In short, this study has focused on the following specific objectives:

1. Sequence analysis of three *Pseudomonas putida* strains isolated from different habitats and exhibiting specific phenotypes.
2. Determination of the *P. putida* pangenome: core genes, accessory genes and unique genes.

RESULTS

Chapter I

Metabolic potential of the organic-solvent tolerant *Pseudomonas putida* DOT-T1E deduced from its annotated genome

This chapter was published as:

Zulema Udaondo, Lazaro Molina, Craig Daniels, Manuel J. Gómez, María A. Molina-Henares, Miguel A. Matilla, Amalia Roca, Matilde Fernández, Estrella Duque, Ana Segura and Juan Luis Ramos (2013) Metabolic potential of the organic solvent tolerant *Pseudomonas putida* DOT-T1E deduced from its annotated genome. *Microb Biotech* **6**(5):598-611

Summary

Pseudomonas putida DOT-T1E is an organic solvent tolerant strain capable of degrading aromatic hydrocarbons. Here we report the DOT-T1E genomic sequence (6,394,153 bp) and its metabolic atlas based on the classification of enzyme activities. The genome encodes for at least 1,751 enzymatic reactions that account for the known pattern of C, N, P and S utilization by this strain. Based on the potential of this strain to thrive in the presence of organic solvents and the subclasses of enzymes encoded in the genome, its metabolic map can be drawn and a number of potential biotransformation reactions can be deduced. This information may prove useful for adapting desired reactions to create value-added products. This bioengineering potential may be realized via direct transformation of substrates, or may require genetic engineering to block an existing pathway, or to re-organize operons and genes, as well as possibly requiring the recruitment of enzymes from other sources to achieve the desired transformation.

Introduction

Bacteria of the genus *Pseudomonas* are motile Gramnegative bacteria characterized by high metabolic versatility, and aerobic respiration, although a few strains of different species are able to use nitrate as a final electron acceptor (Palleroni, 2010). Pseudomonads are ubiquitous soil and water microorganisms that colonize many different environments and, consequently, have diverse lifestyles. Strains of the species *Pseudomonas putida* are frequently soil inhabitants and are important in organic matter recycling in nature; they have a high bioremediation potential because they often carry genes to deal with natural and xenobiotic chemicals (Nelson *et al*., 2002; Caballero *et al*., 2005; van Dillewijn *et al*., 2007; Arias *et al*., 2008; Segura *et al*., 2009a,b). The key to the ubiquitous distribution of these bacteria is not only their metabolic potential, but also the arsenal of regulatory genes that allow them to adapt to changes in the environment (Sashidhar and Podile, 2009; Wu *et al*., 2010). A few *P. putida* strains, namely, S12, Idaho and DOT-T1E (Weber *et al*., 1993; Ramos *et al*., 1995; Pinkart *et al*., 1996), are able to thrive in the presence of toxic solvents (e.g. decanol, octanol, toluene, styrene), and these strains are considered extremophile microorganisms with great potential in bioremediation and in biocatalysis in biphasic systems (Ramos *et al*., 1995; Isken and de Bont, 1996; Molina *et al*., 2011; Ramos *et al*., 2011; Tao *et al*., 2011; Udaondo *et al*., 2012). Organic solvents are toxic to most microorganisms because they dissolve in the cell membranes, disorganize their structures and impair vital functions such as respiration, and the collapse in energy generation lead to cell death (Sikkema *et al*., 1995; Ramos *et al*., 2002; 2011). Solvent tolerance in *P. putida* DOT-T1E is a multifactorial trait that involves chromosomal and plasmid encoded functions (Ramos *et al*., 2002; Segura *et al*., 2005; 2009b; Rodríguez-Herva *et al*., 2007; García *et al*., 2010; Molina *et al*., 2011). The first barrier to solvents involves a reduction in the permeability of the cell membrane via a fast *cis* to *trans* isomerization of unsaturated fatty acids followed by a slower modification of phospholipid head groups (Keweloh and Heipieper, 1996; Junker and Ramos, 1999; Heipieper *et al*., 2001; 2003; Bernal *et al*., 2007; Pini *et al*., 2009). However, this reduction in permeability does not prevent entry of the solvents, which results in unfolding of proteins and the consequential function of a number of chaperones (Segura *et al*., 2005; Domínguez-Cuevas *et al*., 2006; Volkers *et al*., 2006). The

main mechanism underlying solvent tolerance lies in the action of RND (resistance-nodulation-cell division) efflux pumps encoded on the host chromosome and on the pGRT1plasmid (Kieboom *et al*., 1998; Kim *et al*., 1998; Ramos *et al*., 1998; Mosqueda and Ramos, 2000; Rojas *et al*., 2001; Rodríguez-Herva *et al*., 2007; Godoy *et al*., 2010; Udaondo *et al*., 2012). Resistance to solvents is also modulated by the action of chromosomally encoded ABC efflux transporters that use energy to remove solvents from the cells to the outer medium (Kim *et al*., 1998; García *et al*., 2010). Here, we present, the genome of the solvent tolerant *P. putida* DOT-T1E strain obtained using the *454* technology. This microorganism uses a wide range of carbon, nitrogen, sulfur and phosphorous sources due to its wide metabolic potential. In addition we summarize previous knowledge on the biotransformation potential of this strain and how the properties and genomic information can be used to design new biotechnological processes.

Materials and Methods

The complete genome sequence of *P. putida* DOT-T1E (GenBank accesion number CP003734) was determined by applying a strategy combining whole-genome-shotgun *454*-pyrosequencing on the genome sequencer FLX platform (20x coverage with 305 contigs, 257 354 nt biggest contig) and Sanger sequencing of PCR amplicons covering gaps between contigs. In addition, after a first round of annotation, regions of lower quality as well as regions with putative frameshifts were resequenced from PCR amplification of the dubious regions and the complete genome sequence was established. The genome of *P. putida* DOT-T1E has two circular replicons; a single chromosome of 6,260,702 base pairs with a G+C content of 61 % (Figure 1.1) and another of 133,451 base pairs corresponding to a plasmid previously named pGRT1 (Molina *et al*., 2011; Genebank HM626202.1, NCBI Reference Sequence: NC_015855.1).

Using a combination of the Glimmer v3.03 software (Salzberg *et al*., 1998; Delcher *et al*., 1999), BLAST analysis and manual curation, a total of 5,803 ORFs were predicted and annotated in the chromosome, of which 170 have no significant homology (at E-value of 10^{-5}) to any ORF present among the sequenced *Pseudomonas* genomes. The analysis of ORFs of pGRT1 was performed previously and revealed it encodes for 126 proteins (Molina *et al*., 2011). Gene names were

taken from Best Blast Hit when available, and gene products were classified into COG category (Tatusov *et al*., 2003), Pfam, Prk and Smart families with RPSBlast. Putative ribosomal binding sites and tRNA genes were identified with Rfam (Griffiths-Jones *et al*., 2003) and tRNAscan-SE (Lowe and Eddy, 1997). Manual validation and visualization of the entire metabolic potential of DOT-T1E was performed using the Pathway Tools Program v16.0 (Letunic *et al*., 2008; Karp *et al*., 2010), which allows graphic visualization of the *P. putida* annotations. Analyses were performed using an Intel(R) Core (TM)i 7-2600 CPU 3.40 GHz with 8 Gb of RAM memory running a linux Ubuntu 11.04 operating system. Gene products were analysed, compared and assigned to metabolic pathways according to the MetaCyc scheme (Caspi *et al*., 2008), and published research articles.

The cut-off criteria for identifying orthologous proteins were compiled by protein–protein pairwise analysis and reciprocal tBLASTN analysis to identify mutual best hits as potential orthologues. The functional annotations of DOT-T1E genes were corrected for consistency with their counterparts in *P. putida* KT2440 and *P. putida* F1. The coordinates of numerous genes were adjusted according to the criteria of full-length alignment, plausible ribosome binding sites, and minimal overlap between genes on opposite DNA strands.

Results and Discussion

Genome sequencing, assembly, annotation and bioinformatic analysis

In total the genome of DOT-T1E encodes 5,721 proteins and 82 RNAs of which 58 corresponded to tRNAs. We analysed the GC skews of the T1E chromosome, which is defined as the value of [G-C]/[G+C] where G and C represent the local base frequencies of G and C respectively. In prokaryotic genomes the GC skew tends to have a positive value on the leading strand of DNA synthesis and a negative value on the lagging strand, resulting in polarity changes at the origin and terminus of replication (Bentley and Parkhill, 2004). The putative position of the replicative terminus is therefore operationally defined as the peak of the cumulative GC skew and it typically resides opposite to the origin of replication in bacterial genomes (Figure. 1.1) (Bentley and Parkhill, 2004). For T1E, the peak GC skew was indeed

mapped opposite the replication origin (at 49.2 %) of the genome. We also found that the *oriC* site locates between the *rpmH* and *dnaA* genes and contains two identical boxes (5′- TTATCCACA-3′) with the first T of the first box corresponding to position 991,824 while the last A of second box is 991,893.

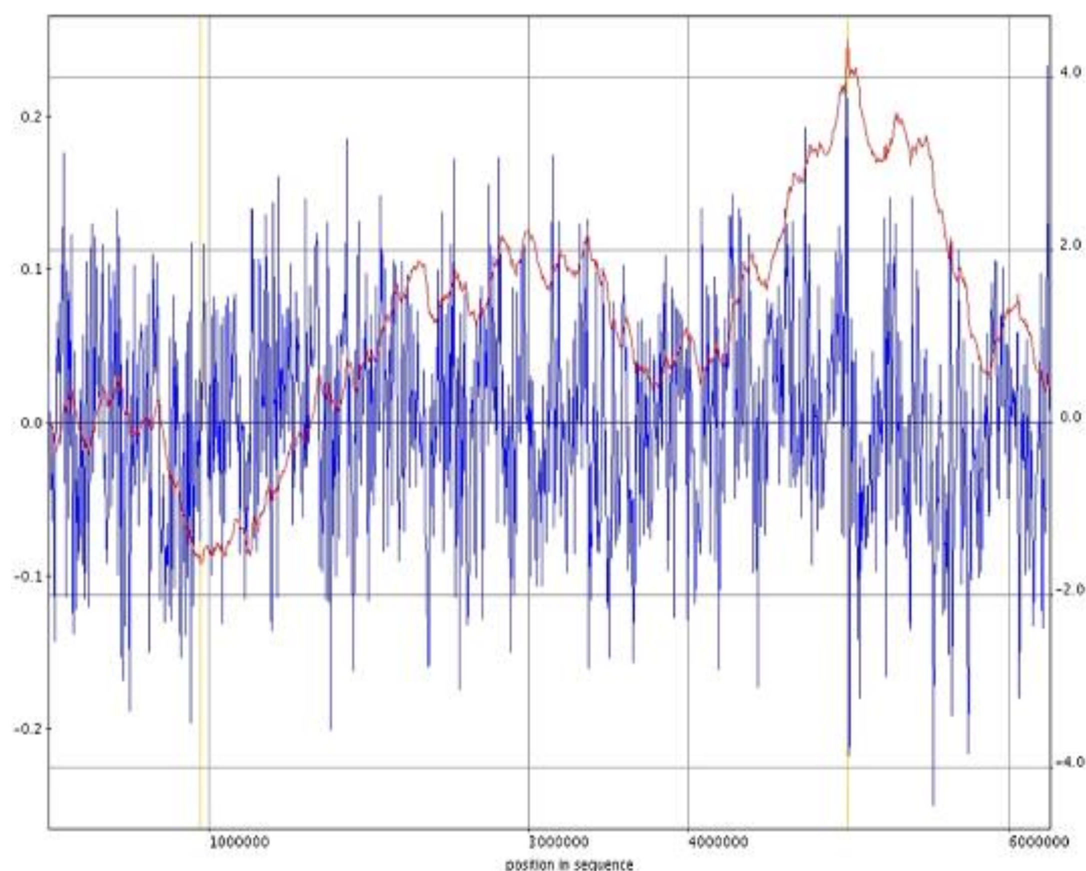


Figure 1 1. ***Pseudomonas putida*** **DOT-T1E chromosome GC skew analysis.** GC skew is defined as the normalized excess of G over C in a given sequence. It is given by (G−C)/(G+C), and it is calculated with a sliding window of 1,000 nucleotides along the genome. It is represented in blue. The cumulative GC skew is the sum of the values of neighbouring sliding windows from an arbitrary start to a given point in the sequence and it is represented in red. GC-skew is positive in the leading strand and negative in the lagging strand.

Figure 1.2 shows the Genome Atlas of *P. putida* (Ussery *et al*., 2009). We analysed the genome to identify potential genomic islands using three different algorithms based on: (i) lack of continuity in the genome, (ii) alignment to other *P. putida* strains and (iii) G+C content and codon usage. This yielded four island regions, 1,504,914–1,553,486; 3,046,659–3,066,609; 4,526,081–4,539,056 and 4,945,609–4,985,959. Most ORFs in these four islands encode hypothetical proteins of unknown function. ORFs in islands 1 and 4 exhibit no homology with any other known sequence, although significant homology was found with transposases.

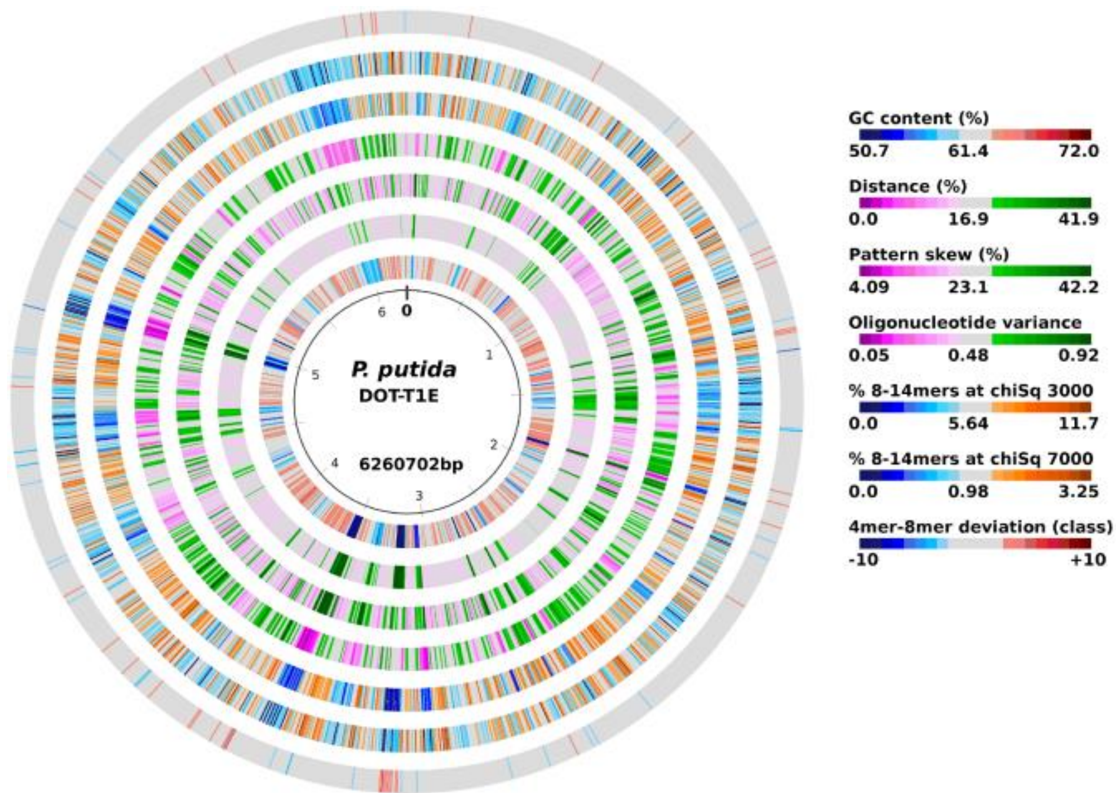


Figure 1 2 Circular genome of *Pseudomonas putida* DOT-T1E. G+C content and the three tetranucleotide parameters are plotted on the innermost four rings. Distance (second innermost circle) is the distance between global and local sliding window tetranucleotide patterns. Pattern skew (third inner most circle) is the distance between tetranucleotide rankings on direct and reverse strands. Oligonucleotide variance (fourth inner most circle) is the numerical variance of oligomers, where a lower value indicates tetramer usage and is more highly restricted (for example in repeat regions) (Klockgether et al., 2011). The third and second outermost circles show the frequency of distribution of overrepresented ($\chi^2 > 3000$) and highly overrepresented ($\chi^2 > 7000$) 8–14 mers in the genome of *P. putida* DOT-T1E. The outermost ring visualizes differences between tetranucleotide usage and the frequency of the overrepresented longer oligomers. Figures were created with JcircleGraph (Davenport et al., 2009).

ORFs in island 3 and 2 are conserved in *P. putida* ND6, a strain that degrades naphthalene (Li *et al*., 2012).

Metabolic potential

As indicated above analysis of the entire metabolic potential of DOT-T1E was performed using the Pathway Tools Program v.16.0 (http://bioinformatics.ai.sri.com/ptools/) (Karp *et al*., 2002; Letunic *et al*., 2008). In the genome of *P. putida* DOT-T1E we identified up to 1,751 enzymatic reactions performed by approximately 1,686 enzymes with 1,268 unique potential substrates. A numerical classification for the enzymes based on the chemical reactions they

carried out according to the Enzyme Commission number (EC number) was elaborated in order to understand the metabolic potential of this strain. According to EC nomenclature (Bairoch, 2000), oxidoreductases (EC 1) were the most abundant enzymes, representing 41 % of the total (Figure 1.3A). Enzymes belonging to EC classes 2 (transferases), EC classes 3 (hydrolases) and 4 (lyases) represented 21 %, 17 % and 10 % of all enzymes respectively, while isomerases (EC 5) and ligases (EC 6) were the least abundant, with 5 % and 6 % of total enzymes respectively. This is consistent with the scenario of a high metabolic versatility described for Pseudomonads (Daniels *et al*., 2010; Palleroni, 2010).

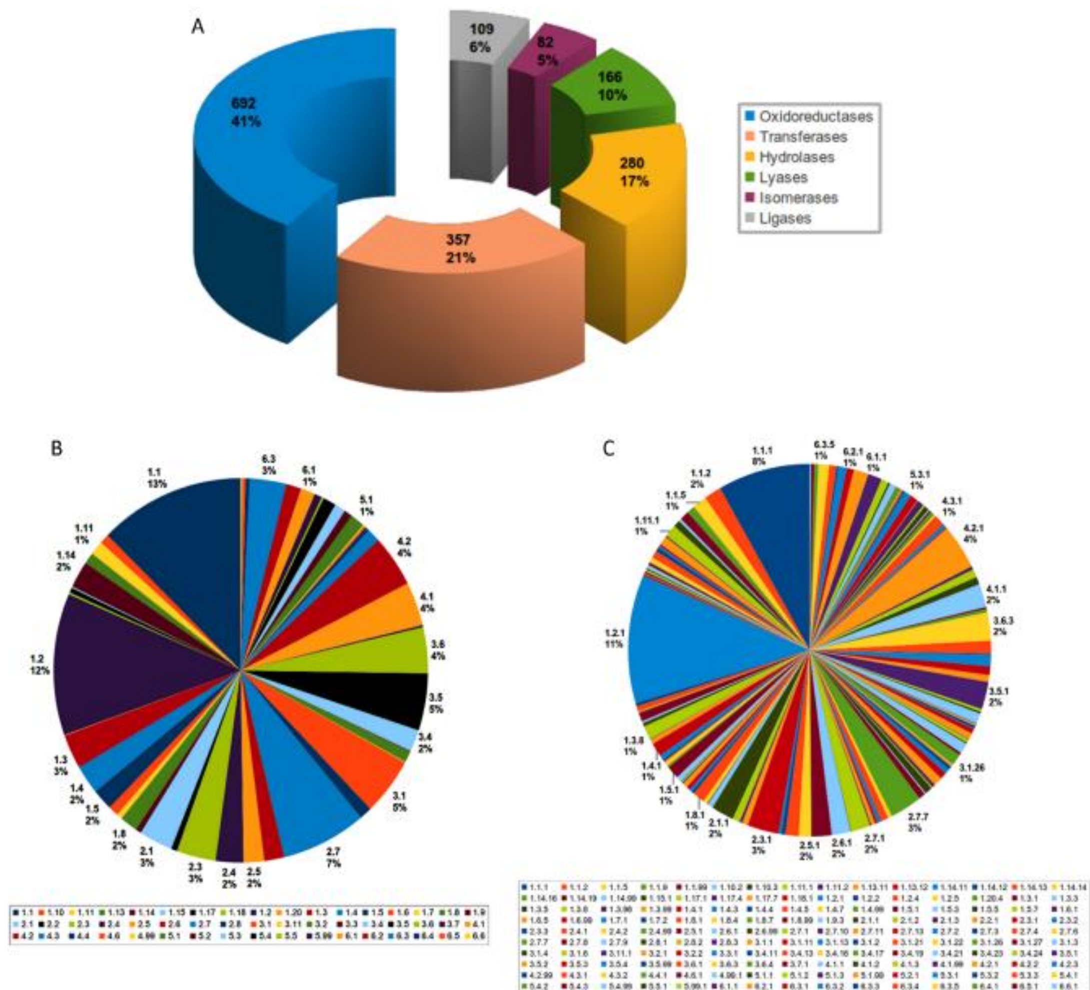


Figure 1 3. Distribution of enzyme activities of *P. putida* **DOT-T1E classified according to the EC nomenclature.** (A) EC X; (B) EC XX; and (C) EC XXX. Colour code for classes and subclasses by numbers are indicated. For full details of the EC classification the reader is referred to http://www.chem.qmul.ac.uk/iubmb/enzyme/.

The second level of EC nomenclature (EC X.X) includes a total of 65 subclasses, of which 51 are present in *P. putida* DOT-T1E (Figure 1.3B). As expected, from the high number of oxidoreductases, two subclasses of this group were among the most

abundant with enzymes that use the CH-OH group as donor (EC 1.1) and those using aldehyde as donors (EC 1.2) representing nearly 12 % of the total for each group. A striking observation was the presence of certain abundant enzyme classes, such as for example phosphotransferases (EC 2.7, 7 % of total); and a series of hydrolases acting on carbon-nitrogen bonds (EC 3.5, 5 % of total), or acting on ester bonds and anhydrides (EC 3.1; about 5 % of total). Figure 3B presents the enzymes of DOT-T1E grouped based on their subclasses. We further classified the enzymes identified in functional subclasses according to the EC X.X.X nomenclature to focus on the potential donors and acceptors in the case of oxidoreductase enzymes or potential groups of substrates in other enzymes (Figure 1.3C). Among a total number of 269 subclasses in the third level of EC nomenclature (EC X.X.X), 150 were present in *P. putida* the most abundant (11 % of the total), also numerically important were the carbon-oxygen lyases (EC 4.2.1, 4% of total), nucleotidyl phosphotransferases (EC 2.7.7, 3 % of total) and acyltransferases (EC 2.3.1, 3% of total).

The enzyme data sets were additionally used to analyse potential substrates and to generate a complete list of enzyme distribution per functional category EC X.X.X.X, the data for which is shown in Suppl.Table 1.1. Using the Pathway Tool platform, the set of phenomics assays previously described by our group (Daniels *et al*., 2010), and the EC X.X.X classification allowed us to explain the pattern of growth of strain DOT-T1E with 65 different carbon sources, 60 nitrogen sources, and 15 sulfur sources used as nutrients (Suppl. Table 1.2). In total 425 pathways for the metabolism of different compounds were delineated. This analysis confirms the limited ability of *P. putida* to use sugars as a C source, which is restricted to glucose, gluconate and fructose.

Pseudomonas putida DOT-T1E has a complete Entner–Doudoroff route for utilization of glucose and other hexoses, but lacks the 6-phosphofructokinase of the glycolytic pathway, in agreement with the genome analysis of others Pseudomonads (del Castillo *et al*., 2007). A large number of sugars were found to not be metabolized by T1E including xylulose, xylose, ribulose, lyxose, mannose, sorbose, D-mannose, alginate, rhamnose, rhamnofuranose, galactose, lactose, epimelibiose, raffinose, sucrose, stachyose, manninotriose, melibiose, tagatose, starch and cello-oligosaccharides, to cite some, in agreement with the lack of genes for the metabolism of these chemicals after the genome analysis of this strain.

The results also confirmed the ability of *P. putida* to use as a C source organic acids (such as acetic, citric, glutaric, quinic, lactic and succinic among others), certain *L*-amino acids (Ala, Arg, Asn, Glu, His, Ile, Lys, Pro, Tyr and Val), and various amino organic compounds. (See Suppl. Figures 1.1–1.4 for examples of catabolic pathways for sugars, amino acids, organic acids and aromatic compounds catabolism). Strain T1E harbours genes for a limited number of central pathways for metabolism of aromatic compounds and numerous peripheral pathways for funnelling of aromatic compounds to these central pathways. As in other Pseudomonads one of the strategies exploited by this microbe for the degradation of different aromatic compounds is to modify their diverse structures to common dihydroxylated intermediates (Dagley, 1971); another strategy is to generate acyl-CoA derivatives such as phenylacetyl-CoA (Fernández *et al*., 2006). Regarding peripheral pathways the *P. putida* DOT-T1E genome analysis has revealed determinants for putative enzymes able to transform a variety of aromatic compounds

The DOT-T1E strain is able to use aromatic hydrocarbons such as toluene, ethylbenzene, benzene and propylbenzene to cite some (Mosqueda *et al*., 1999). The strain also uses aromatic alcohols such as conyferyl- and coumaryl-alcohols and their aldehydes; a range of aromatic acids such as ferulate, vanillate, *p*-coumarate, *p*-hydroxybenzoate, *p*-hydroxyphenylpyruvate, phenylpyruvate, salicylate, gallate and benzoate (see Suppl. Figure 1.4). These chemicals are channelled to central catabolic pathways. Upon oxidation of these chemicals they are metabolized through one of the three central pathways for dihydroxylated aromatic compounds present in this strain. The β-ketoadipate pathway is a convergent pathway for aromatic compound degradation widely distributed in soil bacteria. This pathway consists of a catechol branch (*cat*) and protocatechuate branch (*pca*). The *pca* genes in *P. putida* DOT-T1E are arranged in three operons [*pca- RKFTBDC* (T1E_0230 through T1E_0238), *pcaGH* (T1E_0829 and T1E_830), *pcaJI* (T1E_2058 and T1E_2059)], as is also the case in other *P. putida* and *P. syringae* strains (Suppl. Figure 1.5). The *cat* genes encode the proteins responsible for catechol degradation and are organized in two clusters [*catRBCA* (T1E_5502 through T1E_5505) and *catBCA* (T1E_1744 through T1E_ 1746)] (Suppl. Figure 1.6), maintaining the gene order found in others *P. putida* strains and also in *P. aeruginosa*. The identity of the *catBC* and *catA* genes in both clusters is in the range of 79–82 %. In addition, we should mention that two

other *catA* genes were found, one of them with a high degree of similarity to the KT2440 *catA2* gene, which corresponded to ORF T1E_1057, that is adjacent to the *benRABCDK* genes (T1E_1055 to T1E_1064) for benzoate degradation; while the other *catA* allele corresponded to ORF T1E_5511. It should be noted that this allele is within a cluster of genes that are transcribed in the same direction and which encode genes for salycilate metabolism (T1E_5510 through T1E_5513). The genes involved in phenylacetate degradation were also identified in *P. putida* DOT-T1E. There are 16 genes encoding for phenylacetate degradation organized in a cluster (ORFs T1E_5587 to T1E_5603) and within the cluster a series of potential operons were identified, i.e. the *paaGHIJK* genes (T1E_5590 through T1E_5594) that encode the ring-hydroxylating oxygenase enzyme, the *paaABCDE* genes that encode the β-oxidation enzymes, a potential phenylacetate transport system (*paaLM*) and the regulatory system made of *paaXY*, that correspond to T1E_5587 and T1E_5588 respectively.

Homologous genes for degradation of homogentisate are also present in strain DOT-T1E. Homogentisate is catabolized by a central catabolic pathway that involves three enzymes, homogentisate dioxygenase (T1E_1557), a newly identified putative maleylacetoacetate isomerase (T1E_1555) and a fumarylacetoacetate hydrolase (T1E_1558). In this pathway homogentisate is funneled to yield fumarate and acetoacetate. A search for *hpa* and *gtd* genes that encode genes belonging to the homoprotocatechuate and gentisate pathways yielded no results from the DOT-T1E genome, which suggests the absence of a *meta* ring-cleavage pathway for the degradation of homoprotocatechuate and gentisate.

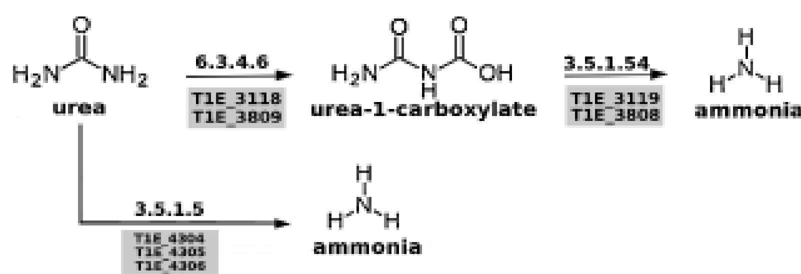


Figure 1 4. Pathway for utilization of urea as an N source by *P. putida*. The genes that encoded the enzymes of these two pathways were identified based on BLAST analysis and comparison to proteins that carry out the indicated reactions

Pseudomonads strains are able to use a range of inorganic nitrogen sources. In this regard three predicted transporters involved in the uptake of ammonium were identified. DOT-T1E incorporates ammonium into C

skeletons using mainly the ATP-dependent activity of glutamine synthetase (GS) followed by the action of glutamate synthase (GOGAT). The genome of DOT-T1E encodes four GS (T1E_0118, 1260, 2050 and 4444) and four GOGAT enzymes (T1E_1644, 2053, 2506 and 3293). Strain DOT-T1E can use nitrate as an N source, which is reduced to ammonium using an assimilatory nitrate reductase (EC: 1.7.99.4) encoded by the T1E_4793 gene, that is in a cluster with *nirB* and *nirD* which encodes an assimilatory nitrite reductase (EC1.7.1.4). (The ORFs encoding these proteins correspond to T1E_4793 through T1E_4795.) The strain also has the complement of genes for utilization of urea either through direct conversion to ammonia (T1E_4304 through T1E_4306, *ureABC*) or via conversion first to urea-1-carboxylate (T1E_3118 through and 3809) and then conversion to ammonia (T1E_3119 and T1E_3808) (Figure 1.4).

Details for the utilization of *D*- and *L*-amino acids as N sources were published by Daniels and colleagues (2010). It was found that the wild-type DOT-T1E strain was able to use a number of either *D*- or *L*-amino acids (i.e. *D*-ornithine, *D*-alanine, *D*-arginine, *D*-asparagine, *D*-lysine and *D*-valine), dipeptides, ethanolamine, and adenine as an N source (Daniels *et al.*, 2010). It is of interest to highlight that this strain can use several *D*-amino acids for which racemases are needed. We have found that the genome of DOT-T1E encodes at least five broad-substrate racemases (T1E_2780, TIE_3429, TIE_1731, TIE_0166, TIE_4880) that can convert D-amino acids into L-amino acids which upon transamination allow the catabolism of these compounds to provide nitrogen for growth (Daniels *et al.*, 2010). Eight aminopeptidases (TIE_3567, TIE_2564, TIE_4792, TIE_1957, TIE_2243, TIE_3241, TIE_3898, TIE_0833) also allows this bacterium to utilize a number of dipeptides and tripeptides as C- and N- sources, in agreement with the saprophytic character of strains of this species (Daniels *et al.*, 2010). Strain DOT-T1E has a number of genes that may encode enzymes/transporters needed for the acquisition of inorganic phosphate, namely: (i) two low-affinity Pit type transporters (T1E_0227 and T1E_0045), (ii) two putative ABC-type inorganic phosphate high-affinity transporter (T1E_2661 through 2663 and T1E_3987 through 3989) and (iii) a PstS type (T1E_2660) high-affinity transporter system regulated by the *phoBR* (T1E_3994 and 3993) response regulator system. This strain uses organic phosphate ester compounds under phosphorous-limiting conditions (Daniels *et al.*, 2010). DOT-T1E also uses organic phosphonates that are transported by a high-affinity

ABC transport system consisting of the *phnD*, *phnE* and *phnC* gene products (T1E_4609 through 4612). Members of the *pseudomonadaceae* have been reported to play a key role in mineralization of carbon bound sulfur in rhizosphere soils. Organic sulfur in soils is comprised mostly of sulfonates and sulfate esters; hence, many soil bacteria carry genes that encode enzymes for utilization of alkanesulfonates. Metabolism of these compounds is achieved through the action of the Ssu enzymes, which are encoded by a set of genes that form an operon, namely, *ssuA* through *F* (T1E_2976 through 2982). This organization is similar to that in other *Pseudomonas* (Kahnert and Kertesz, 2000). The strain DOT-T1E is also endowed with at least one putative arylsulfatase (T1E_5507) which may explain the ability of the strain to use aromatic sulfate esters (Daniels *et al.*, 2010). The DOT-T1E strain is also endowed with four genes that may encode the enzymes required to make sulfur available from methionine (T1E_0568, T1E_2981, T1E_4829 and T1E_4830), which is released as sulfite (Suppl. Figure 1.7). The set of reactions is initiated by MdeAas in other pseudomonads and the pathway is depicted in Suppl. Figure 1.7. A relevant characteristic of DOT-T1E is its capability to grow on minimal medium without the need of vitamins or other cofactors. We found 165 genes encoding enzymatic reactions mediating the biosynthesis of a number of cofactors, i.e. nicotinate, nicotinamide, vitamin B6, riboflavin, ubiquinone, porphyrin, biotin, thiamine, folate, pantothenate and CoA which amounts for 74 distinct biosynthetic pathways. This is consistent with a metabolism in which different enzymes have been described to use these molecules as cofactors.

Based on phenotypic analysis using the BIOSCREEN growth test system described by Daniels and colleagues (2010), it was shown that *P. putida* T1E tolerated various heavy metals. Based on the strain's genome sequence, 64 genes were identified that encode proteins putatively involved in heavy metal resistance and homeostasis (Table 1.1). The majority of the *P. putida* DOT-T1E heavy metal resistance genes are found spread throughout the genome, and they are conserved among all sequenced *P. putida* strains. Up to three different systems potentially involved in simultaneous cobalt, zinc and cadmium resistance were found. One of the cation efflux systems is the CzcD (T1E_2808) immersed in a cluster with the corresponding response regulator CzcR (T1E_2811) and the sensor histidine kinase encoded by the *czcS* gene (T1E_2812). Another family of transporters that may mediate the extrusion of these three heavy metal ions are the one encoded by the

cadA1 (T1E_2820) and *cadA2* (T1E_4489) genes; as well as by the resistance-nodulation-cell division (RND) pump CzcABC (T1E_5270, T1E_5271, T1E_5272).

Table 1. 1. Proteins found in *P. putida* DOT-T1E that are associated with metal resistance and homeostasis.

Gene location	Protein name	Metal	Family/ domain	Predicted role	Definition	e-value
T1E_0296	CzcS2	Me^2	TC reg	Sensor protein	Hypothetical protein	0.0
T1E_0297	CzcR2	Me^2	TC reg	Response regulator	DNA-binding response regulator CzrR	e^{-126}
T1E_0503	TPMT	Te, Se	TPMT	Te and Se Se/Te detoxificatio n	Thiopurine *S*-methyltransfer ase	e^{-122}
T1E_0621	ZnuA2	Zn/Mn(?)	PBD	Zn/Mn(?) uptake	Periplasmic solute-binding protein	e^{-167}
T1E_0622	ZnuC2	Zn/Mn(?)	ATP-binding protein	Zn/Mn(?) uptake	Cation ABC transporter, AP-binding protein	e^{-122}
T1E_0658	Fur	Fe	Fur	Fe regulation	Ferric uptake regulator, Fur family	$7e^{-73}$
T1E_0727	CopA	Cu	HMA	Copper exporting ATPase	Heavy metal translocating P-type ATPase	0.0
T1E_1144	ArsR3	As, Sb	ArsR	Transcriptio nal regulator	ArsR family transcriptional regulator	$5e^{-152}$
T1E_1232	CopS2	Cu	TC reg	Response regulator	Heavy metal sensor signal transduction histidine kinase	0.0
T1E_1233	CopR2	Cu	TC reg	Sensor protein	Two-component heavy metal response transcriptional regulator	e^{-122}
T1E_1234	T1E_1234	Cu	Cupredoxi ne	Copper homeostasis	Plastocyanin/a zurin family copper-binding protein	$4e^{-72}$

T1E_1474	ModC	Mo	ATP-binding protein	Mo uptake	Molybdate ABC transporter ATPase	0.0
T1E_1475	ModB	Mo	I M pore	Mo uptake	Molybdate ABC transporter inner membrane protein	e^{-128}
T1E_1476	ModA	Mo	PBD	Mo uptake	Molybdenum ABC transporter periplasmic molybdate-binding protein	e^{-136}
T1E_1824	NikE	Ni	ATP-binding protein	Ni uptake	Nickel transporter ATP-binding protein NikE	$3e^{-41}$
T1E_1827	NikB	Ni	I M pore	Ni uptake	Nickel transporter permease NikB	$7e^{-50}$
T1E_2011	CumA	Cu	Cu oxidase	Copper homeostasis	Multicopper oxidase	0.0
T1E_2070	NikC	Ni	I M pore	Ni uptake	Nickel transporter permease NikC	$4e^{-59}$
T1E_2193	ModR	Mo	modE	Mo uptake regulation	Mode family transcriptional regulator	e^{-127}
T1E_2274	T1E_2274	Cu	(MFS) transporter s	Copper homeostasis	Bcr/CflA family multidrug resistance transporter	0.0
T1E_2279	T1E_2279	Co/Zn/Cd	OEP	Cobalt-zinc-cadmium resistance	Heavy metal RND efflux outer membrane protein, CzcC family	0.0
T1E_2577	ZnuB2	Zn/Mn(?)	I M pore	Zn/Mn uptake	hypothetical protein	e^{-160}
T1E_2719	ArsH2	As, Sb	ArsH	Arsenical resistant	ArsH protein	e^{-132}
T1E_2720	ArsC2	As, Sb	ArsC	As(V) reduction	Arsenate reductase	$8e^{-86}$

T1E_2721	ArsB2	As, Sb	ArsB	As(III), Sb(III) efflux	Arsenite efflux transporter	0.0
T1E_2722	ArsR2	As, Sb	ArsR	Transcriptional repressor	Arsenic resistance transcriptional regulator	$8e^{-63}$
T1E_2794	NikA	Ni	PBR	Ni uptake	Nickel ABC transporter, periplasmic nickel-binding protein	$1e^{-39}$
T1E_2808	CzcD	Me^2	CDF	Transport and regulation	CDF family cobalt/cadmium/zinc transporter	e^{-166}
T1E_2811	CzcR1	Me^2	TC reg	Response regulator	DNA-binding response regulator CzrR	e^{-124}
T1E_2812	CzcS1	Me^2	TC reg	Sensor kinase	Sensor histidine kinase	0.0
T1E_2820	CadA1	Zn/Cd	P-type ATPase	Me2+ efflux	Heavy metal translocating P-type ATPase	0.0
T1E_2933	TetR	Drug(?)	TetR	Transcriptional regulator	TetR family transcriptional regulator	e^{-117}
T1E_3354	ChrA	Cr	ChrA	Chromate efflux	Chromate transporter	0.0
T1E_3756	PacR(CuR)	Cu/Ag	MerR	Transcriptional regulaor	MerR family transcriptional regulaor	$1e^{-75}$
T1E_3757	PacS	Cu	P-type	Cooper uptake	Heavy metal translocating P-type ATPase	0.0
T1E_3759	PacZ(Cop)	Cu	HMA	Activator	heavy metal transport/detoxification protein	$7e^{-31}$
T1E_3760		Cu		Cooper homeostasis	Multidrug resistance transporter, Bcr/CflA family	$1e^{-163}$
T1E_4452	MfpII	Me^2/drug	RND MFP/HlyD	Me2+/drug efflux	efflux transporter, RND family, MFP subunit	0.0
T1E_4453	MfpI	Me^2/drug	RND MFP/HlyD	Me2/drug efflux	RND efflux transporter	0.0

T1E_4454	CzcA4	Me^{2}/drug	RND	Me2/drug efflux	Acriflavin resistance protein	0.0
T1E_4488	CadR	Zn/Cd	MerR	Cd, Zn efflux	MerR family transcriptional regulator	$8e^{-82}$
T1E_4489	CadA2	Cd/Zn	P-type ATPase	Cd, Zn efflux	Heavy metal translocating P-type ATPase	0.0
T1E_4513	CopA1	Cu	MultiCU oxidases	Cu chelation	Copper resistance protein A	0.0
T1E_4672	ZnuC1	Zn	ATP-binding protein	Zn uptake	Zinc ABC transporter ATP-binding protein	e^{-147}
T1E_4694	CusA	Me^{2}	RND	$Me2^{+}$ efflux	CzcA family cobalt/zinc/cadmium efflux transporter permease	0.0
T1E_4695	CusB	Me^{2}	RND MFP/HlyD	$Me2^{+}$ efflux	CzcB family cobalt/zinc/cadmium efflux transporter membrane fusion protein	0.0
T1E_4696	CusC	Me^{2}	OEP	$Me2^{+}$ efflux	CzcC family cobalt/zinc/cadmium efflux transporter outer membrane protein	0.0
T1E_4697	PorD		Porin	Channel basic amino acids	Porin, putative	0.0
T1E_4698	CzcR3	Me^{2}	TC reg	Response regulator	DNA-binding heavy metal response regulator, putative	e^{-127}
T1E_4760	ZnuA1	Zn	PBD	Zn uptake	Periplasmic solute-binding protein	e^{-171}
T1E_4761	Zur	Zn	Fur	Regulator	FUR family transcriptional regulator	$2e^{-74}$
T1E_4763	ZnuB1	Zn	I M pore	Zn uptake	Hypothetical protein	e^{-138}

T1E_4936		Cu	CBS	Copper homeostasis	CBS domain containing protein	$1e^{-130}$
T1E_4939	ArsR1	As, Sb	ArsR	Transcriptio nal regulator	ArsR family transcriptional regulator	$5e^{-48}$
T1E_4996	ArsC3	As, Sb		Arsenate reductase	Arsenate reductase	$1e^{-60}$
T1E_5088	CzcA5	Me^2/drug	RND	Me2+/drug efflux	Acriflavin resistance protein	0.0
T1E_5089	Mtrc2	Me^2/drug	RND MFP/HlyD	Me2+/drug efflux	Efflux transporter, RD family, MFP subunit	0.0
T1E_5270	CzcA2	Me^2	RND	Cation efflux	Cobalt-zinc-cadmium resistance protein CzcA	0.0
T1E_5271	CzcB2	Me^2	RND MFP/HlyD	Cation efflux	Cobalt-zinc-cadmium resistance protein CzcB, putative	0.0
T1E_5272	CzcC2	Me^2	OEP	Cation efflux	Cobalt-zinc-cadmium resistance protein CzcC, putative	0.0
T1E_5277	OmpR		Response regulator	Copper homeostasis, Cobalt-zinc-cadmium resistance	DNA-binding heavy metal response regulator	$6e^{-124}$
T1E_5753	CopB1	Cu	OM protein	Cu chelation(?)	Copper resistance B precursor	e^{-159}

The CusABC efflux system (T1E_4694, T1E_ 4695, T1E_4696) is involved in resistance to silver and copper ions. Seven genes involved in resistance to arsenite–arsenate–antimonite efflux were annotated. Four of them *arsHCBR* made an operon (T1E_2719– 2722), and the three other genes related to arsenite resistance (T1E_4939, T1E_4996 and T1E_1144) are scattered throughout the genome. Finally one chromate resistance protein ChrA (T1E_3354) was found in the genome of DOT-T1E suggesting it is the responsible for chromate efflux in this strain.

Biotransformation potential

As mentioned above DOT-T1E has the ability to thrive in the presence of toxic organic solvents that normally form a biphasic system with water. This property can be exploited to develop double-phase biotransformation systems (organic solvent and water) in which water insoluble chemicals, toxic substrates or chemical products are kept in the organic phase. The main advantages of these systems are that the product(s) is(are) continuously removed by a solvent phase, their toxic effects are decreased and the lifespan of the biocatalytic system is longer. In addition, if the concentration of the product increases in the organic phase, product recovery is easier and less costly (Bruce and Daugulis, 1991; Leon *et al*., 1998). Rojas and colleagues (2004) demonstrated that *P. putida* DOT-T1E was tolerant to different aliphatic alcohols such as decanol, nonanol and octanol. These aliphatic alcohols are useful in double-phase biotransformation systems to deliver hydrophobic or toxic compounds or to recover added value products that partition preferentially in the organic phase. This concept was exploited by Neumann and colleagues (2005) who showed that DOT-T1E in the presence of 1-decanol tolerated up to 200-fold higher concentrations of the model substrate 3-nitrotoluene than in aqueous medium. In the same line Wierckx and colleagues (2008) showed that phenol production from glucose by *P. putida* S12, another solvent tolerant strain, increased up to 10-fold using a biphasic system. This set of results is the bases that support the potential of DOT-T1E as a useful biocatalyst for biphasic systems.

As described herein, a wide range of oxido-reductase enzymes are encoded in the genome of DOT-T1E, a number of which are of commercial interest. Among these are a numerous dioxygenases that might selectively hydroxylate the aromatic rings at positions 1 and 2; 2 and 3; 2 and 5; 3 and 4; and 4 and 5 (see Suppl. Table 1.3). These dioxygenases may catalyse the stereo-specific dioxygenation of hydrocarbons and could yield secondary commodity chemicals such as adipic acid and γ-caprolactam. At least 16 monooxygenases that may act on diverse chemicals have also been annotated (Suppl. Table 1.4). Some of these enzymes have the potential to oxidize alkanes to their pertinent alcohols, and are of interest to generate added-value products such as linear branched alcohols, aromatic alcohols, diols, hydroxypropionic acid and others. Since the produced chemicals are not metabolized

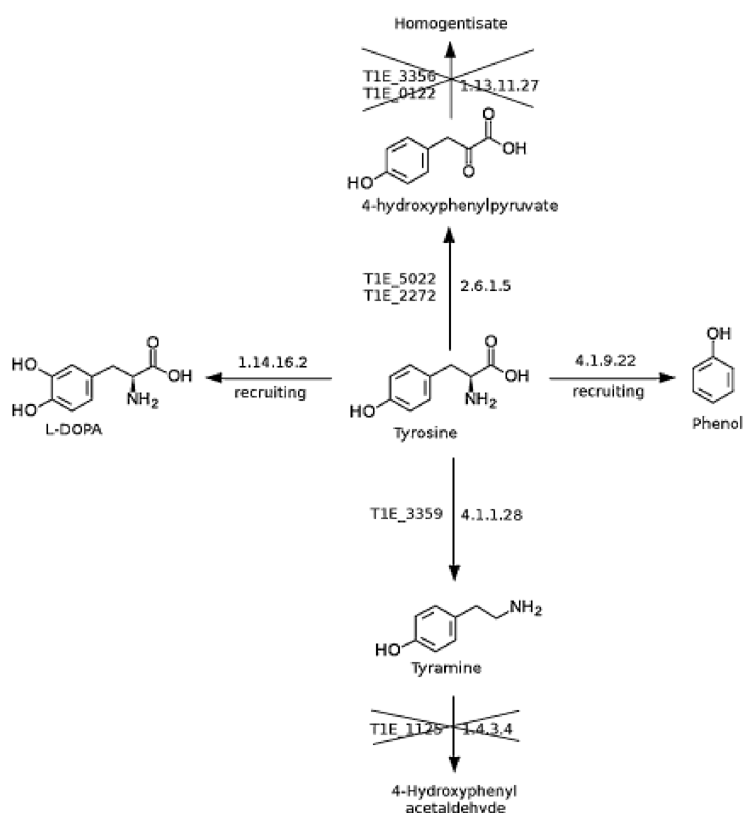


Figure 1 5. Biotransformation of tyrosine by *P. putida* through metabolic blockage or gene recruitment. The EC XXXX of the enzymes needed for the listed biotransformation are indicated. The text describes the approaches used by different research groups to achieve the indicated products.

by *P. putida* DOT-T1E, they accumulate in culture supernatants and high yields can be achieved via extraction of the second phase (Suppl. Figure 1.8).

Biotransformations based on genetically engineering production strains for certain compounds may require either blocking an existing pathway, recruiting new enzymes, or a combination of both approaches. Some of these approaches have been used before with DOT-T1E or related solvent-tolerant strains. An example of biotransformation through inhibition of a single gene is the production of 3-methylcatechol from toluene. Thus, Rojas and colleagues (2004) showed that a catechol 2,3- dioxygenase knockout mutant of DOT-T1E in a 1-octanol/water bioreactor produced 20-fold higher amounts of the compound than in the aqueous medium; these results demonstrate the usefulness of double-phase systems. Ramos-González and colleagues (2001) developed a system for transformation of toluene into 4-hydroxybenzoate which involves the use of a double mutant of DOT-T1E in which the toluene dioxygenase and *p*-hydroxybenzoate hydroxylase genes were first inactivated. Then, a set of genes for sequential oxidation of toluene via toluene 4-monooxygenase were incorporated and the recombinant strain system produced up to 35 Mm of the product. Efficient bioconversion of glucose to phenol or cinnamic acid was achieved by Wierckx and colleagues (2005) and Nijkamp and colleagues (2005), respectively, with *P. putida* S12 to this end a tyrosine phenol lyase (EC

4.1.9.22) from *Pantoea agglomerans* was recruited and it enabled the S12 strain to produce phenol (Figure 1.5). In fed-batch assays in water, the productivity was limited by accumulation of 5 mM phenol in the medium; above this concentration phenol was toxic.

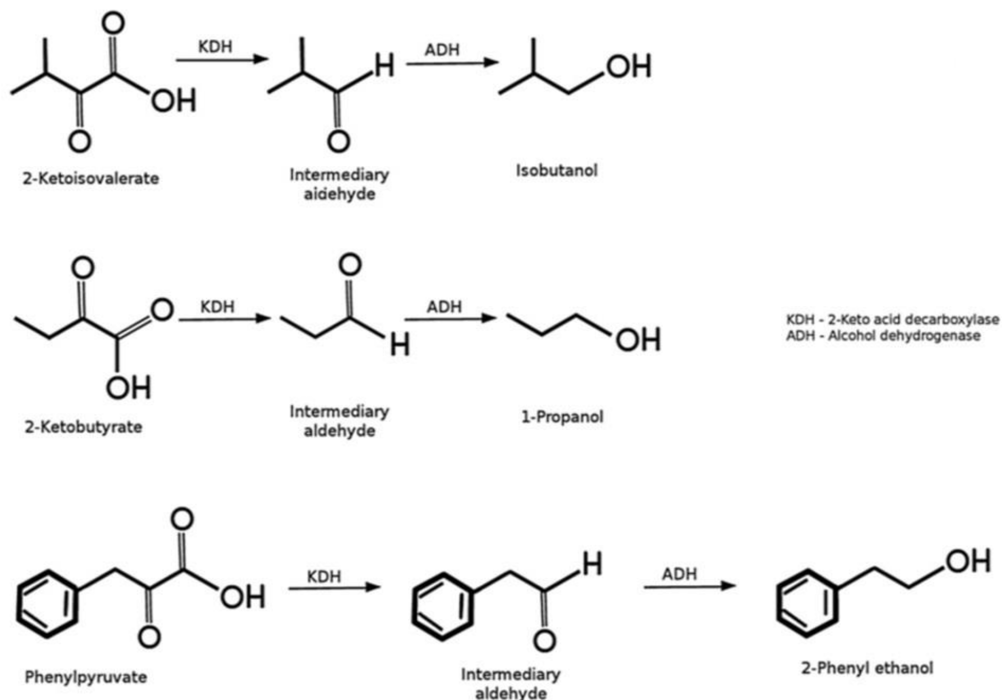


Figure 1 6. Potential synthesis of different alcohols from keto acids by DOT-T1E. 2-Ketoisovalerate, 2-ketobutyrate and phenylpyruvate are produced in the catabolism of isoleucine, threonine and tryptophane respectively. According to Atsumi and colleagues colleagues (2008) recruitment of a broad substrate range keto acid decarboxylase (KDH) yields an aldehyde, which along with one of the multiple alcohol dehydrogenase enzymes encoded in the genome of this strain can lead to the synthesis of the corresponding alcohol (see Suppl. Table 1.1)

However, this toxicity was overcome by use of 1-octanol as a second phase and as an extractant for phenol in a biphasic system. This approach resulted in accumulation of nearly 50 mM phenol in the octanol phase (Wierckx *et al.*, 2008). Other possibilities for the production of added-value molecules with DOT-T1E are their synthesis from tyrosine; for example, DOT-T1E can produce *L*-DOPA from tyrosine (Figure 1.5). This can be achieved by recruiting one of the following activities: a polyphenol oxidase (EC 1.10.3.1), a tyrosinase (1.14.18.1) or a tyrosine 3-monooxygenase (E 1.14.16.2) (Krishnaveni *et al.*, 2009; Surwase and Jadhav, 2011). It should also be noted that with tyrosine as a substrate DOT-T1E can produce tyramine (via an internal aromatic amino acid decarboxylase, EC 4.1.1.28) and 4-hydroxyphenylpyruvate using a tyrosine amino transferase (EC 2.6.1.5).

However, accumulation of the products of these biotransformations requires the inhibition of further catabolism of the products because they can be used as a C source by DOT-T1E (Daniels *et al*., 2010).

One of our aims is to customize strains for the production of aromatic alcohols for biofuel production. In this regard DOT-T1E can be used to produce alkyl and aromatic alcohols (Figure 1.6) through blocking the catabolic pathways for amino acid degradation, in which keto acid intermediates are converted into their corresponding alcohols, as reported for *Escherichia coli* (Atsumi *et al*., 2008) a process that requires the recruitment of a keto acid decarboxylase to produce an intermediate aldehyde that is subsequently transformed into its corresponding alcohol. *D*-xylose is the second most abundant sugar in lignocellulosic materials and its utilization by industrial organisms to produce biofuels and added-value aromatic compounds is of interest (Octave and Thomas, 2009). As described above, strains of the *P. putida* species cannot use pentose sugars, but this was overcome via the engineered addition of *xylAB* genes, which allow the conversion of *D*-xylose in *D*-xylulose and xylulose-5-P, to allow metabolism of *D*-xylose via the pentose phosphate pathway (Meijnen *et al*., 2008; 2009). However, growth was rather slow and fast growers were isolated after enrichment in fermentors. In a recent omics-based study, the authors have shown that high yield growth involved inactivation of glucose dehydrogenase and rearrangement of central carbon catabolism to allow for more efficient decarboxylation of 6-phosphogluconate for the catabolism of the sugar via the pentose phosphate pathway (Meijnen *et al*., 2012). Since *P. putida* S12 is as tolerant to solvents as DOT-T1E (Segura *et al*., 2003), we hypothesize that a similar strain of DOT-T1E could be engineered. In summary, the analysis of the genome of the solventtolerant DOT-T1E strain explains the catabolic potential of this microorganism in accordance with previously published physiological studies. The use of the Pathway Tool platform together with the identification of enzymes using the international EC codes not only support the metabolic reactions, but also provide an opportunity to design biotransformation reactions to produce value-added products in high concentration. Due to the proven ability of DOT-T1E to thrive in the presence of a second organic phase in biphasic systems, only minimal genetic manipulation will be required in order to reap substantial reward from the genomic analysis reported here.

Acknowledgements

We thank Manuel Ferrer for early collaboration in this project and Maria del Mar Fandila for secretarial assistance.

References

Arias, S., Olivera, E.R., Arcos, M., Naharro, G., and Luengo, J.M. (2008) Genetic analyses and molecular characteriza- tion of the pathways involved in the conversion of 2-phenylethylamine and 2-phenylethanol into phenylacetic acid in *Pseudomonas putida* U. *Environ Microbiol* **10**: 413– 432.

Atsumi, S., Cann, A.F., Connor, M.R., Shen, C.R., Smith, K.M., Brynildsen, M.P., *et al*. (2008) Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab Eng* **10**: 305–311.

Bairoch, A. (2000) The ENZYME database. Nucleic Acids Res 28: 304–305. Bentley, S.D., and Parkhill, J. (2004) Comparative genomic structure of prokaryotes. Annu Rev Genet 38: 771–792.

Bentley S. and Parhkhill J. (2004) Comparative genomics structure of Prokaryotes. *Annu. Rev . Genet.* **38:** 771-791

Bernal, P., Muñoz-Rojas, J., Hurtado, A., Ramos, J.L., and Segura, A. (2007) A *Pseudomonas putida* cardiolipin synthesis mutant exhibits increased sensitivity to drugs related to transport functionality. Environ Microbiol **9**: 1135–1145.

Bruce, L.J., and Daugulis, A.J. (1991) Solvent selection strategies for extractive biocatalysis. *Biotechnol Prog* **7**: 116–124.

Caballero, A., Esteve-Nuñez, A., Zylstra, G.J., and Ramos, J.L. (2005) Assimilation of nitrogen from nitrite and trinitro-toluene in *Pseudomonas putida* JLR11. *J Bacteriol* **187**: 396–399.

Caspi, R., Foerster, H., Fulcher, C.A., Kaipa, P., Krummen- acker, M., Latendresse, M., *et al*. (2008) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **36**: D623–D631.

del Castillo, T., Ramos, J.L., Rodríguez-Herva, J.J., Fuhrer, T., Sauer, U., and Duque, E. (2007) Convergent peripheral pathways catalyze initial glucose catabolism in *Pseudomonas putida*: genomic and flux analysis. *J Bacteriol* **189**: 5142–5152.

Dagley, S.C. (1971) Catabolism of aromatic compounds by microorganism. *Adv Microb Physiol* **6**: 1–46.

Daniels, C., Godoy, P., Duque, E., Molina-Henares, M.A., de la Torre, J., del Arco, J.M., *et al*. (2010) Global regulation of food supply by *Pseudomonas putida* DOT-T1E. *J Bacteriol* **192**: 2169–2181.

Davenport, C.F., Wiehlmann, L., Reva, O.N., and Tümmler, B. (2009) Visualization of *Pseudomonas* genomic structure by abundant 8–14 mer oligonucleotides. *Environ Microbiol* **11**: 1092–1104.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**: 4636–4641.

Domínguez-Cuevas, P., González-Pastor, J.E., Marqués, S., Ramos, J.L., and de Lorenzo, V. (2006) Transcriptional tradeoff between metabolic and stress-response programs in *Pseudomonas putida* KT2440 cells exposed to toluene. *J Biol Chem* **281**: 11981–11991.

Fernández, C., Ferrández, A., Miñambres, B., Díaz, E., and García, J.L. (2006) Genetic characterization of the phenylacetyl-coencime a oxygenase from the aerobic phenylacetic acid degradation pathway *of Escherichia coli. Appl Environ Microbiol* **72**: 7422–7426.

García, V., Godoy, P., Daniels, C., Hurtado, A., Ramos, J.L., and Segura, A. (2010) Functional analysis of new trans- porters involved in stress tolerance in *Pseudomonas putida* DOT-T1E. *Environ Microbiol Rep* **2**: 389–395.

Godoy, P., Molina-Henares, A.J., Torre, J., Duque, E., and Ramos, J.L. (2010) Characterization of the RND family of multidrug efflux pumps: in silico to in vivo confirmation of four functionally distinct subgroups. *Microb Biotechnol* **3**: 691–700.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., Eddy S.R. (2003) Rfam: an RNA family database. *Nucleic Acids Res* **31**: 439–441.

Heipieper, H.J., de Ward, P., van der Meer, P., Killian, J.A., Isken, S., de Bont, J.A.M., *et al*. (2001) Regiospecific effect of 1-octanol on cis-trans isomerization of unsaturated fatty acids in the solvent-tolerant strain *Pseudomonas putida* S12. *Appl Microbiol Biotechnol* **57**: 541–547.

Heipieper, H.J., Meinhardt, F., and Segura, A. (2003) The cis-trans isomerase of unsaturated fatty acids in *Pseudomonas* and *Vibrio*: biochemistry, molecular biology and physiological function of a unique stress adaptive mechanism. *FEMS Microbiol Lett* **229**: 1–7.

Isken, S., and de Bont, J.A.M. (1996) Active efflux of toluene in a solvent-resistant bacterium. *J Bacteriol* **178**: 6056–6058.

Junker, F., and Ramos, J.L. (1999) Involvement of the cis/ trans isomerase Cti in solvent resistance of *Pseudomonas putida* DOT-T1E. *J Bacteriol* **181**: 5693–5700.

Kahnert, A., and Kertesz, M.A. (2000) Characterization of a sulfur-regulated oxygenative alkylsulfatase from *Pseudomonas putida* S-313. *J Biol Chem* **275**: 31661– 31667.

Karp, P.D., Paley, S., and Romero, P. (2002) The Pathway Tools software. *Bioinformatics* **18** (Suppl. 1): S225–S232.

Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T.J., *et al*. (2010) Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* **11**: 40–79. Keweloh,

Keweloh, H., and Heipieper, H.J. (1996) Trans unsaturated fatty acids in bacteria. *Lipids* **31**: 129–137.

Kieboom, J., Dennis, J.J., de Bont, J.A.M., and Zylstra, G.J. (1998) Identification and molecular characterization of an efflux pump involved in *Pseudomonas putida* S12 solvent tolerance. *J Biol Chem* **273**: 85–91.

Kim, K., Lee, S., Lee, K., and Lim, D. (1998) Isolation and characterization of toluene-sensitive mutants from the toluene-resistant bacterium *Pseudomonas putida* GM73. *J Bacteriol* **180**: 3692–3696.

Klockgether, J., Cramer, N., Wiehlmann, L., Davenport, C.F., and Tümmler, B. (2011) *Pseudomonas aeruginosa* genomic structure and diversity. *Front Microbiol* **2**: 150.

Krishnaveni, R., Rathod, V., Thakur, M.S., and Neelgund, Y.F. (2009) Transformation of L-tyrosine to L-DOPA by a novel fungus, *Acremonium rutilum*, under submerged fermentation. *Curr Microbiol* **58**: 122–128.

Leon, R., Fernandes, P., Pinheiro, H.M., and Cabral, J.M.S. (1998) Whole-cell biocatalysis in organic media. *Enzyme Microb Technol* **23**: 483–500.

Letunic, I., Yamada, T., Kanehisa, M., and Bork, P. (2008) iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci* **33:** 101–103.

Li, S., Zhao, H., Li, Y., Niu, S., and Cai, B. (2012) Complete genome sequence of the naphthalene-degrading *Pseudomonas putida* strain ND6. *J Bacteriol* **194**: 5154–5155.

Lowe, T.M., and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955–964.

Meijnen, J.P., de Winde, J.H., and Ruijssenaars, H.J. (2008) Engineering *Pseudomonas putida* S12 for efficient utilization of D-xylose and L-arabinose. *Appl Environ Microbiol* **74**: 5031–5037.

Meijnen, J.P., de Winde, J.H., and Ruijssenaars, H.J. (2009) Establishment of oxidative D-xylose metabolism in *Pseu-domonas putida* S12. *Appl Environ Microbiol* **75**: 2784–2791.

Meijnen, J.-P., and de Winde, J.H., and Ruijssenaars, H.J. (2012) Metabolic and regulatory rearrangements underlying efficient D-xylose utilization in engineered *Pseudomonas putida* S12. *J Biol Chem* **287**: 14608–14614.

Molina, L., Duque, E., Gómez, M.J., Krell, T., Lacal, J., García-Puente, A., *et al*. (2011) The pGRT1 plasmid of *Pseudomonas putida* DOT-T1E encodes functions relevant for survival under harsh conditions in the environment. *Environ Microbiol* **13**: 2315–2327.

Mosqueda, G., and Ramos, J.L. (2000) Aset of genes encod- ing a second toluene efflux system in *Pseudomonas putida* DOT-T1E is linked to the tod genes for toluene metabolism. *J Bacteriol* **182**: 937–943.

Mosqueda, G., Ramos-González, M.I., and Ramos, J.L. (1999) Toluene metabolism by the solvent-tolerant *Pseudomonas putida* DOT-T1 strain, and its role in solvent impermeabilization. *Gene* **232**: 69–76.

Nelson, K.E.,Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A., *et al*. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* **4**: 799–808.

Neumann, G., Kabelitz, N., Zehnsdorf, A., Miltner, A., Lippold, H., Meyer, D., *et al*. (2005) Prediction of the adaptability of *Pseudomonas putida* DOT-T1E to a second phase of a solvent for economically sound two-phase biotransformations. *Appl Environ Microbiol* **71**: 6606–6612.

Nijkamp, K., van Luijk, N., de Bont, J.A., andWery, J. (2005) The solvent-tolerant *Pseudomonas putida* S12 as host for the production of cinnamic acid from glucose. *Appl Microbiol Biotechnol* **69**: 170–177.

Octave, S., and Thomas, D. (2009) Biorefinery: toward an industrial metabolism. *Biochimie* **91**: 659–664.

Palleroni, N.J. (2010) The *Pseudomonas* story. *Environ Microbiol* **12**: 1377–1383.

Pini, C.V., Bernal, P., Godoy, P., Ramos, J.L., and Segura, A. (2009) Cyclopropane fatty acids are involved in organic solvent tolerance but not in acid stress resistance in *Pseudomonas putida* DOT-T1E. *Microb Biotechnol* **2**: 253– 261.

Pinkart, H.C., Wolfram, J.W., Rogers, R., and White, D.C. (1996) Cell envelope changes in solvent-tolerant and solvent-sensitive *Pseudomonas putida* strains following exposure to o-xylene. *Appl Environ Microbiol* **62**: 1129– 1132.

Ramos, J.L., Duque, E., Huertas, M.J., and Haidour, A. (1995) Isolation and expansion of the catabolic potential of a *Pseudomonas putida* strain able to grow in the presence of high concentrations of aromatic hydrocarbons. *J Bacteriol* **177**: 3911–3916.

Ramos, J.L., Duque, E., Godoy, P., and Segura, A. (1998) Efflux pumps involved in toluene tolerance in *Pseudomonas putida* DOT-T1E. *J Bacteriol* **180**: 3323–3329.

Ramos, J.L., Duque, E., Gallegos, M.T., Godoy, P., Ramos-González, M.I., Rojas, A., *et al*. (2002) Mechanisms of solvent tolerance in gram-negative bacteria. *Annu Rev Microbiol* **56**: 743–768.

Ramos, J. L., Marqués, S., van Dillewijn, P., Espinosa-Urgel, M., Segura, A., Duque, E., *et al*. (2011) Laboratory research aimed at closing the gaps in microbial bioremediation. *Trends Biotechnol* **29**: 642–647.

Ramos-González, M.I., Godoy, P., Alaminos, M., Ben-Bassat, A., and Ramos, J.L. (2001) Physiological characterization of *Pseudomonas putida* DOT-T1E tolerance to *p*-hydroxybenzoate. *Appl Environ Microbiol* **67**: 4338– 4341.

Revelles, O., Wittich, R.M., and Ramos, J.L. (2007) Identification of the initial steps in D-lysine catabolism in *Pseudomonas putida. J Bacteriol* **189**: 2787–2792.

Rodríguez-Herva, J.J., García, V., Hurtado, A., Segura, A., and Ramos, J.L. (2007) The *ttgGHI* solvent efflux pump operon of *Pseudomonas putida* DOT-T1E is located on a large self-transmissible plasmid. *Environ Microbiol* **9**: 1550–1561.

Rojas, A., Duque, E., Mosqueda, G., Golden, G., Hurtado, A., Ramos, J.L., and Segura, A. (2001) Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. *J Bacteriol* **183**: 3967– 3973.

Rojas, A., Duque, E., Schmid, A., Hurtado, A., Ramos, J.L., and Segura, A. (2004) Biotransformation in double-phase systems: physiological responses of *Pseudomonas putida* DOT-T1E to a double phase made of aliphatic alcohols and biosynthesis of substituted catechols. *Appl Environ Microbiol* **70**: 3637–3643.

Salzberg, S.L., Delcher, A.L., Kasif, S., and White, O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* **26:**544-548

Sashidhar, B., and Podile, A.R. (2009) Transgenic expression of glucose dehydrogenase in *Azotobacter vinelandii* enhances mineral phosphate solubilization and growth of sorghum seedlings. *Microb Biotechnol* **2**: 521–529.

Segura, A., Rojas, A., Hurtado, A., Huertas, M.J., and Ramos, J.L. (2003) Comparative genomic analysis of solvent extrusion pumps in *Pseudomonas* strains exhibiting different degrees of solvent tolerance. *Extremophiles* **7**: 371–376.

Segura, A., Godoy, P., van Dillewijn, P., Hurtado, A., Arroyo, N., Santacruz, S., and Ramos, J.L. (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**: 5937–5945.

Segura, A., Rodríguez-Conde, S., Ramos, C., and Ramos, J.L. (2009a) Bacterial responses and interactions with plants during rhizoremediation. *Microb Biotechnol* **2:** 452– 464.

Segura, A., de Wit, P., and Preston, G.M. (2009b) Life of microbes that interact with plants. *Microb Biotechnol* **2:** 412–415.

Sikkema, J., de Bont, J.A.M., and Poolman, B. (1995) Mechanisms of membrane toxicity of hydrocarbons. *Microbiol Rev* **59**: 201–222.

Surwase, S.N., and Jadhav, J.P. (2011) Bioconversion of L-tyrosine to L-DOPA by a novel bacterium *Bacillus* sp. JPJ. *Amino Acids* **41**: 495–506.

Tao, F., Tang, H., Gai, Z., Su, F.,Wang, X., He, X., and Xu, P. (2011) Genome sequence of *Pseudomonas putida Idaho*, a unique organic-solvent-tolerant bacterium. *J Bacteriol* **193**: 7011–7012.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**: 41.

Udaondo, Z., Duque, E., Fernández, M., Molina, L., de la Torre, J., Bernal, P., *et al*. (2012) Analysis of solvent tolerance in *Pseudomonas putida* DOT-T1E based on its genome sequence and a collection of mutants. *FEBS Lett* **586**: 2932–2938.

Ussery, D.W., Kiil, K., Lagesen, K., Sicheritz-Ponten, T., Bohlin, J., andWassenaar, T.M. (2009) The genus *Burkholderia*: analysis of 56 genomic sequences. *Genome Dyn* **6**: 140–157.

Van Dillewijn, P., Caballero, A., Paz, J.A., González-Pérez, M.M., Oliva, J.M., and Ramos, J.L. (2007) Bioremediation of 2,4,6-trinitrotoluene under field conditions. *Environ Sci Technol* **41**: 1378–1383.

Volkers, R.J.M., de Jong, A.L., Hulst, A.G., van Baar, B.L.M., de Bont, J.A.M., and Wery, J. (2006) Chemostat-based proteomic analysis of toluene-affected *Pseudomonas putida* S12. *Environ Microbiol* **8:** 1674–1679.

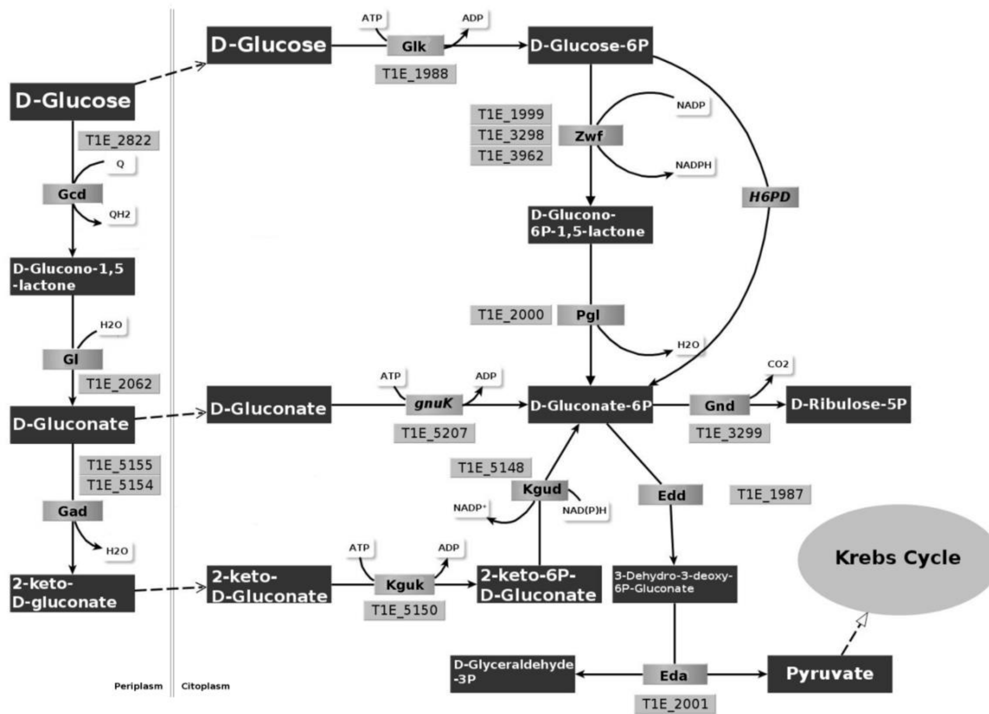
Weber, F.J., Ooijkaas, L.P., Schemen, R.M., Hartmans, S., and de Bont, J.A. (1993) Adaptation of *Pseudomonas putida* S12 to high concentration of styrene and other organic solvents. *Appl Environ Microbiol* **59**: 3502–3504.

Wierckx, N.J., Ballerstedt, H., de Bont, J.A., and Wery, J. (2005) Engineering of solvent-tolerant *Pseudomonas putida* S12 for bioproduction of phenol from glucose. *Appl Environ Microbiol* **71**: 8221–8227.

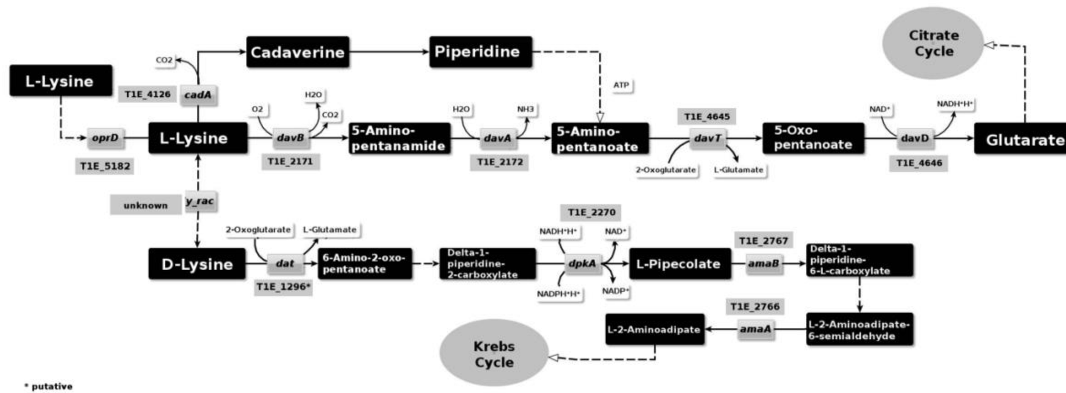
Wierckx, N.J., Ballerstedt, H., de Bont, J.A., de Winde, J.H., Ruijssenaars, H.J., and Wery, J. (2008) Transcriptome analysis of a phenol-producing *Pseudomonas putida* S12 construct: genetic and physiological basis for improved production. *J Bacteriol* **190**: 2822–2830.

Wu, X., Monchy, S., Taghavi, S., Zhu,W., Ramos, J., and van der Lelie, D. (2010) Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS Microbiol Rev* **35:** 299–323.

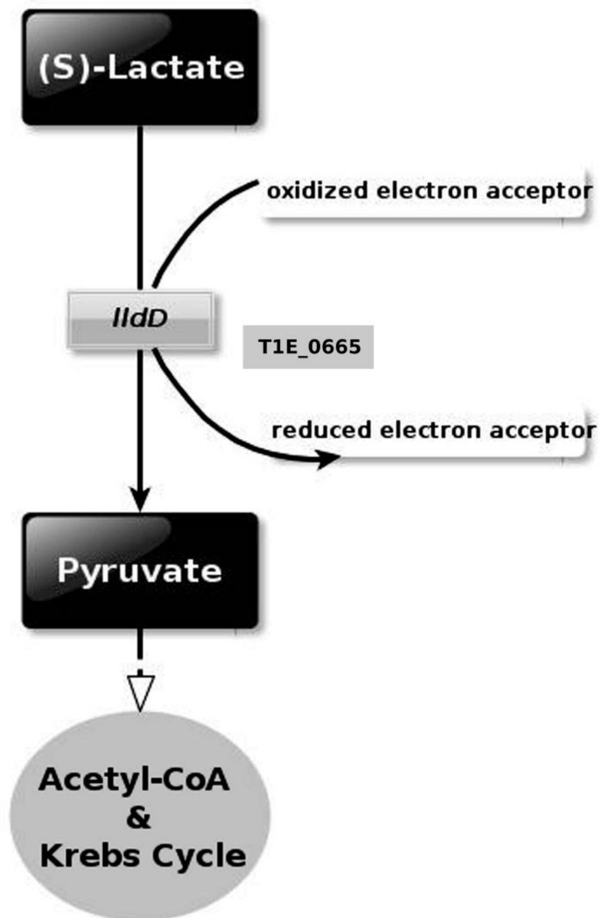
Supplementary material



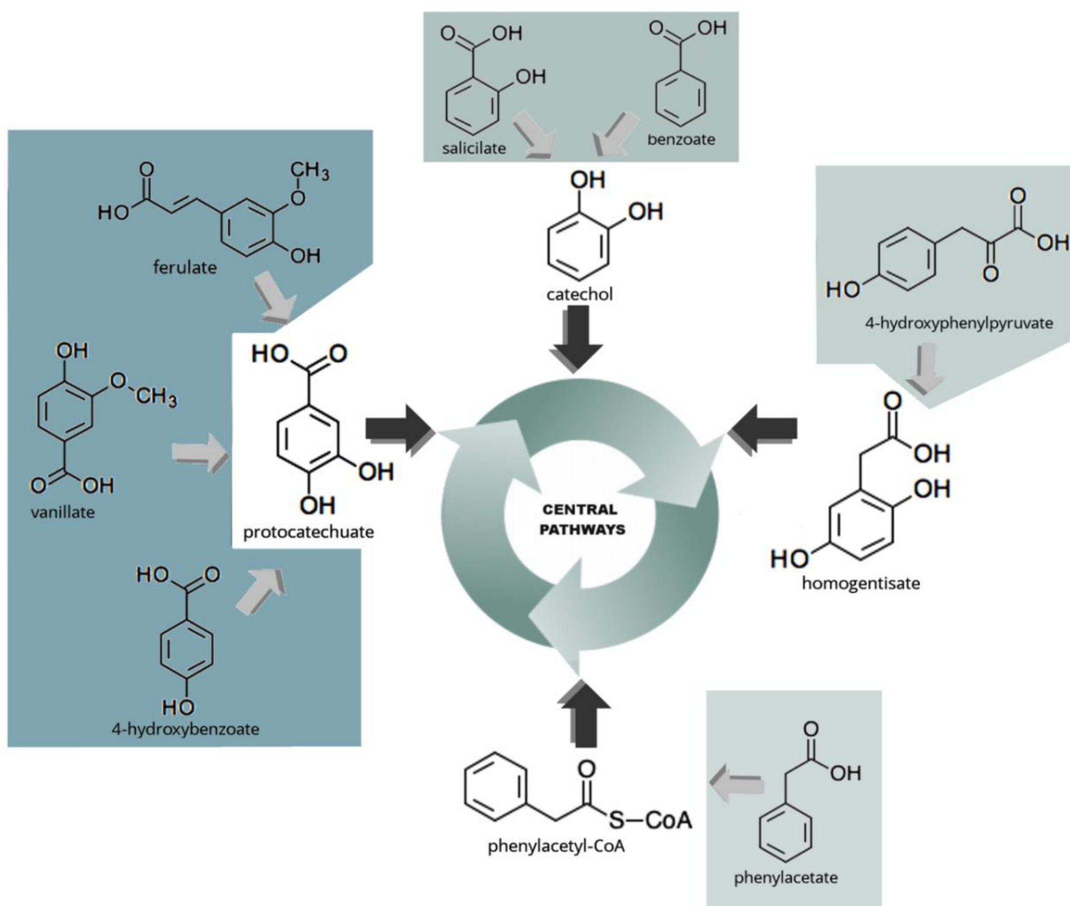
Suppl. Figure 1. 1. .. Catabolism of glucose as an example of sugar metabolism based on annotated genes. Glucose can be oxidized in the periplasm to yield gluconate or ketogluconate or in the cytoplasm upon phosphorylation. All three pathways converge at the level of 6-phosphogluconate that is metabolized via the Entner-Doudoroff pathway that yields chemicals that feed the Krebs cycle. The genes are annotated according to del Castillo *et al*. (2007)



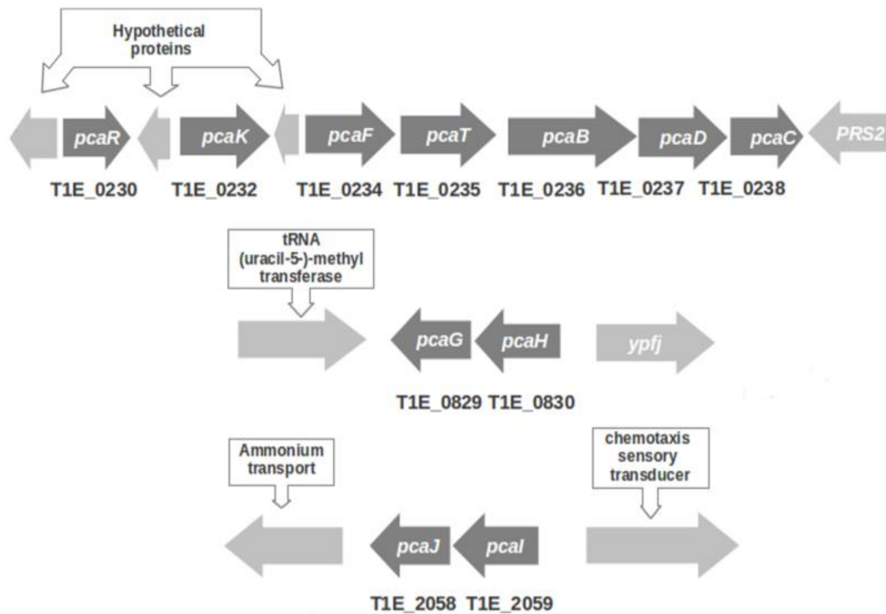
Suppl. Figure 1. 2. Catabolism of *L*- and *D*-lysine as an example of an amino acid being used as a source of carbon through multiple confluent pathways. The set of genes and most of the intermediates were analyzed in detail by Revelles *et al*. (2007) for the KT2440 strain. All genes shown in this pathway for DOT-T1E are based on the identification of the homologous genes



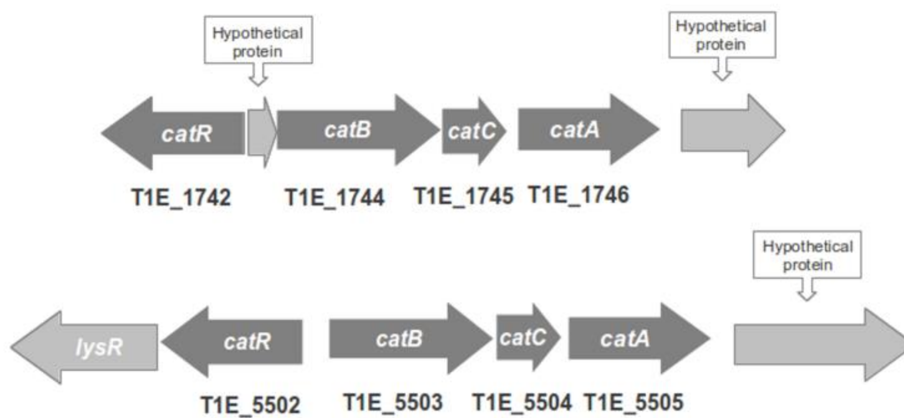
Suppl. Figure 1. 3. Catabolism of lactic acid by *P. putida* DOT-T1E as an example of the metabolism of an organic acid. Information was derived from Nelson *et al*. (2002). Lactate enables the growth of DOT-T1E in minimal medium with doubling times in the range of 120±10 min. Inactivation of the *lldD* gene blocks the use of lactate as a C-source



Suppl. Figure 1. 4. Catabolism of aromatic compounds to a set of central intermediates. For the described reaction the corresponding enzyme(s) were identified as homologous to those described in detail by Jiménez *et al*. (2002) for KT2440 strains. A set of peripheral enzymes lead to the formation of a number of catechol-related compounds, which upon *ortho* or *meta*-cleavage yielded Krebs cycle intermediates.

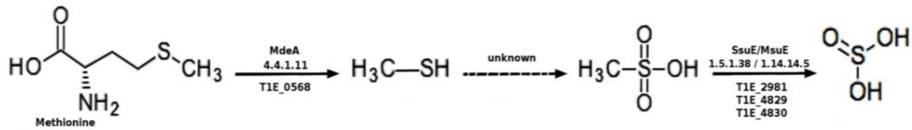


Suppl. Figure 1. 5. Genetic organization of the genes for the protocatechuete central catabolic pathway. The genes were identified by BLAST analysis and the operon structure of the genes deduced from the overlapping nature of all the genes, except *pcaF* and *pcaT*; however, these genes form an operon based on RT-PCR analysis with RNA isolated from strains grown in 4-hydroxybenzoate (unpublished).

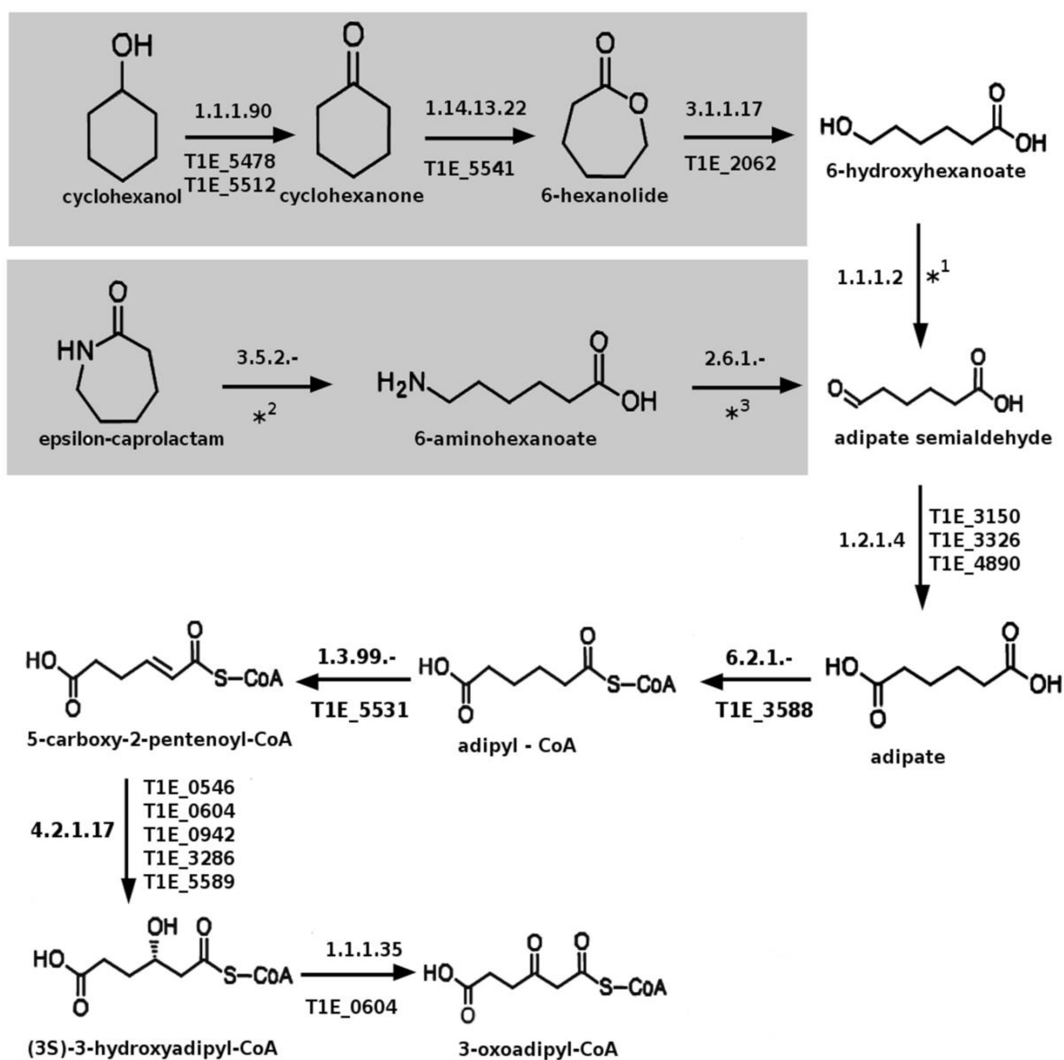


Suppl. Figure 1. 6. Genetic organization of the duplicated *cat* genes located in two different chromosomal regions. One of the *cat* clusters includes the *catR* gene that probably controls the expression of the two *catABC* operons based on high sequence conservation of the promoter region in the two *catB* genes.

	Product Name	Gene Id	EC Number
1	MdeA - Methionine gamma-lyase	T1E_0568	4.4.1.11
2	SsuE - FMN reductase	T1E_2981	1.5.1.38
3	MsueD - Alkanesulfonate monooxygenase	T1E_4829 T1E_4830	1.14.14.5



Suppl. Figure 1. 7. Genes and enzymes for the utilization of methionine as an S source by DOT-T1E. The gene products involved in the catabolism of methionine were identified based on BLAST analyses.



Suppl. Figure 1. 8. Set of biotransformation reactions to achieve the synthesis of adipate / 3-oxoadipyl-CoA based on gene content in DOT-T1E strain. Multiple steps are represented by X^1, X^2 and X^3. The loci are:

X^1.- T1E_0314, T1E_0333, T1E_0690, T1E_0898, T1E_1004, T1E_1466, T1E_1534, T1E_2226, T1E_2233, T1E_2397, T1E_3843, T1E_3853, T1E_3860, T1E_4416, T1E_5253, T1E_5478;

X^2.- T1E_0103, T1E_1864, T1E_1899, T1E_2485, T1E_2953, T1E_3631, T1E_4716, T1E_4861, T1E_4862;

X^3.- T1E_0369, T1E_1262, T1E_4182, T1E_4445

Chapter II

Bioinformatic analysis of the properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1

The information of this chapter was published as:

Amalia Roca*, Paloma Pizarro-Tobias*, Zulema Udaondo*, Matilde Fernández, Miguel A. Matilla, M. Antonia Molina Henares, Lázaro Molina, Ana Segura, Estrella Duque and Juan Luis Ramos (2013). Analysis of the plant growth promoting properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1 **15**(3):780-794

Summary

Pseudomonas putida BIRD-1 is a plant growth-promoting rhizobacteria with the ability to adher to the plant roots and colonize at high cell densities the rhizosphere, even in soils with low moisture. Genome analysis of BIRD-1 revealed that this strain carries the genetic information that makes it able to use as a source of carbon and nitrogen a wide range of plant secreted products. The BIRD-1 genome encodes a wide range of proteins that help this strain to deal with reactive oxygen stress generated in the plant rhizosphere and to synthesize the disaccharide trehalose, involved with desiccation tolerance. The plant growth-promoting properties of BIRD-1 derive from its ability to enhance phosphorous and iron solubilization and to produce a phytohormone precursor such as indole-3-acetic acid. *Pseudomonas putida* BIRD-1 genome also harbour genes that encode at least five phosphatases related to phosphorous solubilization, one of them being a phytase that facilitates the utilization of phytic acid, the main storage form of phosphorous in plants. The biosynthetic pathways for the production for trehalose, the siderophore pyoverdine and the indole-3-acetic acid in BIRD-1 strain were inferred from bioinformatics analysis.

Introduction

The rhizosphere, or the layer of soil influenced by the root, is a very changing environment where complex interactions occurs between plant roots and microorganisms (Molina, 2000; Uroz *et al*., 2010; Blom *et al*., 2011). Bacteria that efficiently colonize the rhizosphere and stimulate plant growth through direct or indirect mechanisms are referenced as plant growth-promoting rhizobacteria (PGPR) (Lugtenberg and Kamilova, 2009; Blom *et al*., 2011; Matilla *et al*., 2011). PGRP microorganisms are of great interest in the area of biotechnology because they can increase the yields of crops. PGRP are used as inoculants for biofertilization, phytostimulation and biocontrol of plant pathogens and allow to reduce the use of fertilizers and agrochemicals used for crops. Plant growth promotion can be mediated by the synthesis of phytohormones, by enhancing the uptake of mineral nutrients as phosphorous and iron, antagonizing soilborne pathogenic microorganism or inducing systemic resistance, among others (Lugtenberg and Kamilova, 2009; Matilla *et al*., 2011). PGPR strains are able to produce several bioactive compounds as antibiotics, siderophores and indole-3-acetic acid (Siddiqui, 2006). Siderophores are soluble molecules which chelate iron with hight affinity and other metals with low affinity (Wandersman and Delepelaire, 2004; Schalk *et al*., 2011). Several evidences indicate that microbial siderophores play an important role in the biocontrol of some soil-borne plant diseases by avoiding the action of soil-borne pathogens through competition for iron (Antoun *et al*., 1998; Wandersman and Delepelaire, 2004; Beneduzi *et al*., 2012) and in plant iron nutrition. Some non-pathogenic rhizobacteria have been reported to suppress diseases in plants by triggering induced systemic resistance (ISR), a plant response that depends on jasmonic acid and on the plant's ethylene signalling pathways (Bakker *et al*., 2007). *Pseudomonads* are well known as frequent rhizosphere colonizers and some strains of the species *Pseudomonas fluorescens, Pseudomonas putida*, *Pseudomonas mendocina* and *Pseudomonas stutzeri* have been shown to possess PGPR capabilities (Bakker *et al*., 2007; Wu *et al*., 2010; Beneduzi *et al*., 2012). Strains from *Pseudomonas putida* species, are also among the most powerful phosphate solubilizers (Rodríguez and Fraga, 1999; Vyas and Gulati, 2009; Daniels *et al*., 2010).

Pseudomonas putida BIRD-1 strain is an efficient PGPR microorganism through the mobilization of nutrients and stimulation of secondary root proliferation (Matilla *et al*., 2011; Roca *et al*., 2013). The strain was isolated from a garden soil in a culture medium without iron addition and with insoluble inorganic phosphate as a source of phosphorous. The complete genome sequence of the BIRD-1 strain was sequenced by 454 pyrosequencing technology comprising a circular chromosome of 5,731,541 bp with no plasmid, and with a G+C content of 61.7 %. BIRD-1 final genome encodes 4960 proteins, 65 tRNAs and 22 rRNAs. The size of the genome of the strain BIRD-1 is similar to other *P. putida* strains as *P. putida* W619 (5,774,330 Mb) (Wu *et al*., 2010) but is smaller than those of *P. putida* KT2440 (6,181,860 Mb) (Nelson, Weinel, Paulsen, Dodson, Hilbert, Martins dos Santos, *et al*., 2002)*, P. putida F1* (5,959,964 Mb), and *P. putida* GB-1 (6,078,430 Mb), the strain also shows a high homology and synteny with that of KT2440 and that of the F1 strain.

In this chapter we analysed the genome of *P. putida* BIRD-1 in relation to its PGPR properties and the most relevant data are presented. Wet assays were performed by our collaborator at BioIliberis R&D and are described in the Supporting information.

Materials and Methods

The complete genome sequence of *Pseudomonas putida* BIRD-1 (GenBank accession number CP002290) was obtained by using 454 pyrosequencing technology with a total coverage of 20x. The reads sequences were assembled into 106 contigs, 52 of them were longer than 3,000 bp. Closing gaps was achieved by sequencing PCR products covering gaps between adjacent contigs (Matilla *et al*., 2011). Functional annotation of the entire genome was performed using a combination of the Glimmer 3.03 software (Salzberg *et al*., 1998; Delcher *et al*., 1999, 2007), BLAST analysis and manual curation. The genbank file of the BIRD-1 genome was used to construct the data base with the functional annotations based on the Pathway-Tools software (Karp *et al*., 2002) where the manual validation and visualization of the BIRD-1 genome was performed. Analyses were made using an Intel(R) Core (TM)i 7-2600 CPU 3.40 GHz with 8 Gb of RAM memory running a linux Ubuntu 11.04 operating system. Gene products were analysed, compared and

assigned to metabolic pathways according to the MetaCyc scheme (Caspi *et al*., 2008) and published research articles. The cut-off criteria for identifying orthologous proteins were compiled by protein–protein pairwise BLASTp analysis and reciprocal tBLASTN analysis to identify mutual best hits as potential orthologues. The functional annotations of BIRD-1 genes were corrected for consistency with their counterparts in *P. putida* KT2440 and *P. putida* F1. The coordinates of numerous genes were adjusted according to the criteria of full-length alignment, plausible ribosome binding sites, and minimal overlap between genes on opposite DNA strands.

Results and Discussion

Adhesion and colonization of plant roots

The assays performed by our collaborator (see related article (Roca *et al*., 2013)) showed that approximately 10^4 to 10^5 c.f.u. of *P. putida* BIRD-1 cells adhered per corn seed, an amount similar to that reported for KT2440 before (Weinel *et al*., 2002; Espinosa-Urgel and Ramos, 2004; Molina-Henares *et al*., 2006; Yousef-Coronado *et al*., 2008). Eleven genes were found in the genome of BIRD-1 that have been previuos identified as able to mediate adhesion of KT2440 to abiotic and biotic surfaces (Duque *et al*., 2012), Table 2.1 shows the degree of identity between the two strains which was in the range of 40 % to 100 %. Mutants deficient in *lapA* gene were 10-fold less efficient in adhesion than the parental strain and exhibited decreased ability in rhizosphere colonization in competition with the wild-type strain (Pizarro-Tobias; personal communication). Proliferation of BIRD-1 in the plant rhizosphere is favoured by the ability of the strain to use compounds secreted by the plant, among which are the 20 proteinogenic amino acids, organic acids such as citric, lactic, sugars such as glucose and fructose and flavonoids. Phenomic assays with a wide battery of carbon sources revealed that *P. putida* BIRD-1 is able to use a wide range of amino acids as carbon or nitrogen sources and it has the complete set of genes required for the metabolism of proline, glutamate, glutamine, asparagine, tyrosine, valine, branched amino acids, serine and alanine among the amino acids.

Table 2. 1. Genes involved in adhesion to biotic and abiotic surfaces in *Pseudomonas putida* strain BIRD-1 deduced from the identification of reciprocal genes in *P. putida* KT2440.

	P. putida KT2440			*P. putida* BIRD1		
Group	**Locus**	**Gene Name**	**Product Name**	**Locus**	**e-val**	**Iden %**
Large adhesion proteins	PP_0168	*lapA*	Surface adhesion protein	PPU_BIRD_0199	0.0	91
	PP_0806	*lapF*	Surface adhesion protein	PPU_BIRD_0852	0.0	87
	PP_0805	*lapH*	Component of an ABC transport system for LapF	PPU_BIRD_0851	0.0	90
Regulators	PP_0952	*rpoN*	RNA polymerase factor sigma-54	PPU_BIRD_1005	0.0	98
	PP_1650	*gacS*	Global sensor kinase GacS	PPU_BIRD_3966	0.0	90
Flagella	PP_4352	*flhB*	Flagella protein FlhB	PPU_BIRD_1506	$1e^{-180}$	84
	PP_4369	*fliF*	Flagella MS-ring protein	PPU_BIRD_1490	0.0	88
	PP_4376	*fliD*	Flagellar hook-associated protein	PPU_BIRD_1467	$2e^{-86}$	40
	PP_4378	*fliC*	Flagellin	PPU_BIRD_1465	$2e^{-55}$	65
Others	PP_1633	--	Hypothetical secreted protein	PPU_BIRD_3983	$2e^{-66}$	100
	PP_5329	*pstS*	PstS	PPU_BIRD_5121	$1e^{-170}$	89

Reciprocal genes were identified in the *P. putida* genome by BLAST search. The corresponding PP numbers in the genome of KT2440 and BIRD-1 are given together with the identity of the corresponding proteins in the two strains. The genes are named as annoted in the KT2440 genome. E-val (e-value). Iden% (percent of sequence identity)

The strain can use citrate and succinate, compounds exudated by roots, as well as a wide range of lignin-derived compounds, including hydroxylated aromatic acids and methoxylated compounds. Bioinformatics analysis showed that all of these pathways are common with the strain KT2440 and gene sequence comparison analysis showed

that the sequence and the order of the genes were preserved in the two strains. Metabolism of aromatic compounds has been reviewed before by Jiménez and colleagues (2002) and Bielecki and colleagues (2011).

Plant growth-promoting rhizobacteria properties are particularly relevant in marginal soils or when plants are exposed to abiotic or biotic stresses. Water scarcity is a common problem in soils in the Mediterranean basin. The survival of *P. putida* BIRD-1 was tested in soils with different degrees of humidity using *P. putida* KT2440 as a reference strain for these studies (Roca *et al*., 2013). The experimental assays showed that when the soil moisture was kept at 2 %, BIRD-1 survived at a higher density than KT2440 (Figure 2.1). Therefore, BIRD-1 is more tolerant to desiccation than *P. putida* KT2440 (Muñoz-Rojas *et al*., 2006). Trehalose has been suggested to be important for the survival of *P. putida* at low-humidity rates. Biosynthesis of trehalose in *Pseudomonas* can take place through two pathways that involve the conversion of maltose into trehalose via TreS [trehalose synthases (PPU_BIRD-1_1765) and PPU_BIRD-1_2817 or from maltodextrine catalysed by the TreY (PPU_BIRD-1_1772)/TreZ (PPU_BIRD-1_1774) proteins].

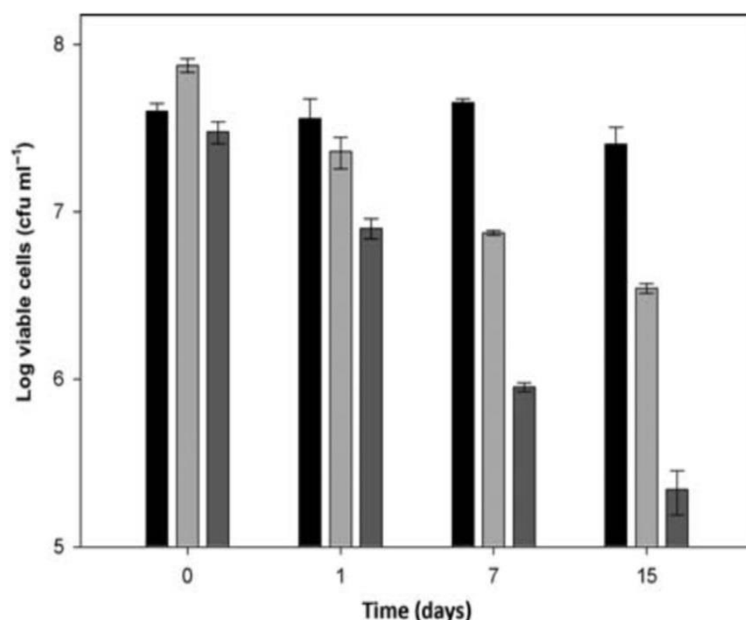


Figure 2 1. Viable *P. putida* KT2440, BIRD-1 and a *treZ* mutant in rhizosphere soil. Soils were inoculated with ~5x10^{7} c.f.u.g^{-1} and sterile pre-germinated corn seeds sown. At the indicated time c.f.u.g^{-1} rhizosphere soil was determined using three different pots. Black bars (BIRD-1), light grey (KT2440) and dark grey (*treZ* BIRD-1 mutant) (Roca et al., 2013)

To test the role of trehalose production in resistance to desiccation in BIRD-1, mutants in *treZ* and *treS* genes were contrusted. Survival assays have shown that the initial survival of the *treZ* mutant in the low-humidity soil was inferior than that of the parental strain or that of mutants in the *treS* gene (Fig. 2.1),

suggesting that the maltodextrine pathway for the production of trehalose is relevant for colonization of soils by *P. putida* BIRD-1.

Root colonization by bacteria requires the ability to overcome the oxidative stress produced by the plant root cell respiration (Matilla *et al*., 2007). In a series of phenomic analyses carried out *in P. putida,* was shown to reach high cell densities in cultures without the stressor or with 1 mM methylviologen, 4 mM H_2O_2, 1 mM Cr_2O_7 and 100 $mgml^{-1}$ ampicillin (Roca *et al*., 2013). The genome of BIRD-1 was then interrogated for the set of genes involved in oxidative stress responses, and it was found to encode two iron–manganese SodA superoxide dismutases (PPU_BIRD-1_0966 and PPU_BIRD-1_0999) and four catalases (PPU_BIRD-1_0143, 0518, 2060 and 2848). In addition, we searched for the presence of alkylhydroper-oxidases, cytochrome-c peroxidase and glutathione peroxidase. We found six homologues for the first group (PPU_BIRD-1_1273, 2010, 2150, 2706, 2708, 3265 and 3243), a single cytochrome peroxidase (PPU_BIRD- 1_2732) and two glutathione peroxidases (PPU_BIRD- 1_3741 and 3931). The redundancy of the genes encoding different isoenzymes related to oxidative stress in the genome of *P. putida* BIRD-1 probably points towards the importance of these proteins and their likely synergic action in the removal of active oxygen species.

Evidence of plant growth-promoting activities encoded by BIRD-1

The stimulatory effect of BIRD-1 on growth of maize roots was tested on agar plates with sterilized corn seeds deposited on the surface of a water agar plate either inoculated or not with *P. putida* BIRD-1 (10^6 c.f.u. cm^{-2} of surface plate). Germination of corn seeds was monitored over time showing that the size of the primary root and the secondary roots were larger for seeds on agar surfaces with bacteria than in the absence of microbes at any incubation time (Figure 2.2). The PGPR properties of *Pseudomonas putida* BIRD-1 could be due to one or several of the following explanations: (i) increased solubilization and mineralization of nutrients, particularly insoluble phosphates (Richardson, 2001; Miller *et al*., 2010); (ii) facilitation of iron uptake by plants by producing higher amounts of siderophores (Pal *et al*., 2001; Bakker *et al*., 2007; Fernández-Piñar *et al*., 2011); (iii) the ability to produce higher levels of indole-3-acetic acid (IAA) (Patten and Glick, 2002; Dobbelaere *et al*., 2003; Dey *et al*., 2004; Viruel *et al*., 2011); (iv) the ability to

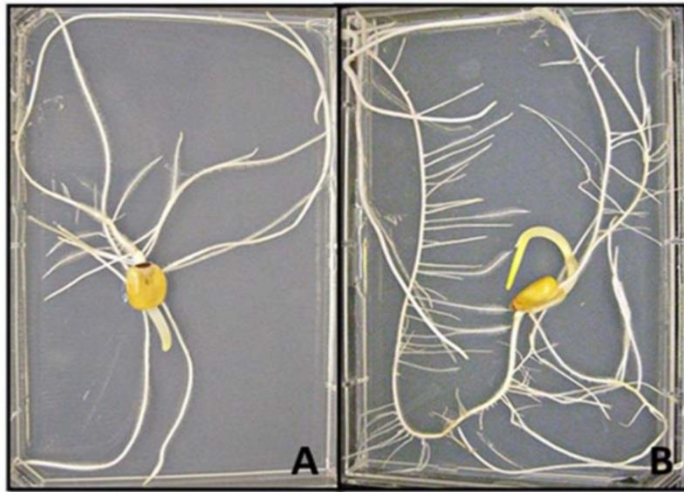


Figure 2 2. Root development of corn seeds germinated on agar plates. Water agar plates were prepared, one plate was kept as a control (A) and 0.1 ml of a suspension containing 108 c.f.u.ml^{-1} was spread on top of the other plate (B). Once the surfaces were dry, a single sterile corn seed was deposited per agar plate and root development monitored over time. The picture corresponds to a typical situation after 1 week incubation at 22ºC (Roca et al., 2013)

produce 1-aminocyclo-propane-1-carboxylate (ACC) deaminase to reduce the level of ethylene in the root of developing plants thereby increasing the root length and growth (Li *et al*., 2000; Penrose and Glick, 2001); (v) the ability to cope with oxidative stress generated in the rhizosphere of plants (Stajner *et al*., 1997; Matilla *et al*., 2010). Several experimental assays were performed according to each of these possibilities.

Solubilization of inorganic phosphate. Several studies have attributed the phosphate solubilization trait to the production of organic acids (Jones and Darrah, 1994; Nautiyal, 1999; Rodríguez and Fraga, 1999; Miller *et al*., 2010). To test BIRD-1's ability to solubilize different phosphorous sources, the strain BIRD1 was cultured on modified medium *A* with various insoluble inorganic phosphate sources (tricalcium phosphate, dicalcium phosphate and phosphate rock) in the presence of different C sources and monitored growth as c.f.u. ml^{-1}. BIRD-1 growth occurred with the three P sources with glucose, fructose, glycerol and aromatic acids such as benzoate, although growth with fructose led to higher cell densities (Suppl. Fig. 2.1). Figure 2.3 shows the results with phosphate rock as a source of phosphorous, and 5 mM sodium benzoate as a carbon source. As it can be observed BIRD-1 cells can grow exponentially and the phosphate in solution reached a concentration of 0.5 mg l^{-1}, which indicated that the strain solubilized phosphate.

Organic acids such as gluconic acid and 2-ketogluconic acid have been proposed to be the most frequent agent of mineral phosphate solubilization by bacteria such as *Pseudomonas* sp., *Erwinia herbicola*, *Burkholderia* sp., *Rhizobium leguminosarum*, *Bacillus firmus* and other soil bacteria (Duff and Webley, 1959; Banik and Dey, 1982; Halder *et al.*, 1990; Halder and Chakrabarty, 1993; Rodríguez and Fraga, 1999). *Pseudomonas putida* BIRD-1 produces gluconic acid in the periplasmic space after the oxidation of glucose. In this study a BIRD-1 mutant that fails to use glucose as a C source and that exhibits a knockout in Entner–Doudoroff pathway Δ*eda* (2-dehydro-3-deoxyphosphogluconate aldolase) was identified. This mutant converted most glucose to 2-ketogluconate (50 % of glucose was converted into the acid), and as a consequence, in the Pikovskaya's assay the mutant strain exhibited increased ability to solubilize insoluble tricalcium phosphate (Suppl. Figure 2.2).

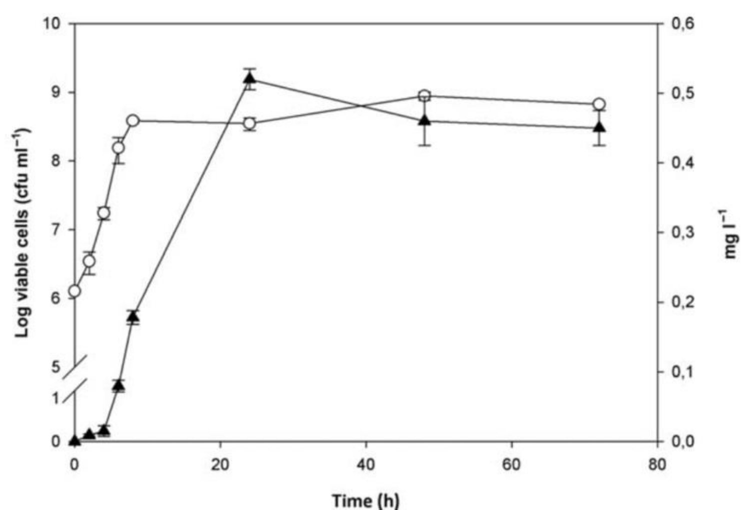


Figure 2 3. Solubilization of phosphate and growth of *P. putida* BIRD-1.Was used *A* medium with 5g of phosphate rock per 100 ml as a source of inorganic phosphorous, and inoculated cultures with an initial cell density of $2x10^6$c.f.u.ml^{-1}. At the indicated times phosphate in solution in mgl^{-1} (○) and c.f.u.ml-1(▲) were determined as described in Experimental procedures of supplementary material).

Several enzymes have also been shown to be involved in making insoluble phosphorous compounds available for cell growth. These processes are achieved via the action of phosphatases, such as phosphohydrolases (Gügi *et al.*, 1991; Rodríguez and Fraga, 1999), phytases (Richardson and Hadobas, 1997), phosphonoacetate hydrolases (McGrath *et al.*, 1998), D-α-glycerophosphatases (Skrary and Cameron, 1998) and C-P lyases (Ohtake *et al.*, 1996). We screened in the genome of *P. putida* BIRD-1 to identify genes encoding these activities, and we found five potential phosphatases encoded by the genome of BIRD-1, namely a 4-phytase family member (PPU_BIRD-1_5077), a phosphonoacetate hydrolase of the PhnA family

(PPU_BIRD-1_0727), two PAP2 acid phosphatases (PPU_BIRD-1_2395 and PPU_BIRD- 1_0951), an alkaline phosphatase (PPU_BIRD-1_0932) and an exopolyphosphatase (PPU_BIRD-1_5012). In agreement with *P. putida* BIRD-1 encoding a phytase is the fact that the strain is able to grow reaching high cell densities with phytate, the main form of phosphorous stored in plants, as a source of phosphorous (Suppl. Figure 2.3).

Production of siderophores. *Pseudomonas putida* BIRD-1 secretes the siderophore pyoverdine when grown under iron-deficient conditions. To identify genes involved in pyoverdine biosynthesis, a mini-Tn5 mutant library from BIRD-1 was screened for selection of clones unable to grow on liquid minimal medium without supplemented iron. A single clone with an insertion in the *pvdD* gene (PPU-BIRD-1_1630) was found. Some genes related to pyoverdine synthesis were grouped (i.e. PPU_BIRD-1_1630 through to 1634), while other *pvd* genes for pyoverdine synthesis were found to be scattered throughout the *P. putida* BIRD-1 genome (e.g. *pvdA* which corresponded to PPU_BIRD-1_1983). The genome of BIRD-1 encodes a TonB-dependent pyoverdine-iron receptor, FpvA, which is highly specific for the pyoverdine produced by this microorganism. The *fvpA* gene is in an operon with *pvdS* which encodes a sigma-24 family factor and its corresponding anti-sigma (Matilla *et al*., 2011). Expression of this cluster is known to be under the control of the PvdS and FpvR genes both of which are present in the genome of BIRD-1 and which guarantee the regulated expression of siderophore production. Transport of iron from the periplasm to the cytoplasm is most likely mediated by a three-component ABC transporter made up of the FecBCE proteins (PPU_BIRD-1_3269/3271) whose genes form an operon, although we cannot rule out the possibility of other ABC transporters playing this role.

Many *P. putida* strains are capable of using a wide range of structurally unrelated exogenous siderophores by means of multiple receptors that are located in the outer membrane. In the genome of the KT2440 strain, Martínez-Bueno and colleagues (2002) identified 10 TonB-dependent siderophore receptors (FecA-like) for which the strain lacked the corresponding biosynthetic genes for the siderophore, although the strain had its corresponding adjacent FecI/FecR control system (Table 2.2). This was taken as evidence that KT2440 was an efficient iron-sequestering strain. We carried out a similar bioinformatics analysis with the BIRD-1 genome and

found that BIRD-1 also possesses these 10 TonB-like iron receptors (Table 2.2) as well as 15 other FecA-like proteins that were not linked to *fecI/fecR*-like genes (Table 2.2). Eleven of these 15 FecA-like proteins have a best hit in BLAST analysis with the corresponding protein system of KT2440 while the others exhibited > 95 % identity with iron receptors of *P. putida* F1.

The unusual systems of iron acquisition probably contribute to broaden the ecological competence of *P. putida* in terms of colonization and persistence in the rhizosphere. Since BIRD-1 has multiple iron receptors we hypothesized that the *fvpD* mutant will grow and survive in soils through the acquisition of iron via the series of alternative receptors encoded in its genome, what it was experimentally confirmed (Roca *et al*., 2013).

Table 2. 2. Physical linkage of the *fecA* gene to *fecI/fecR* regulatory system.

Receptor	Gene ID *P. putida* BIRD-1	Protein name (Blast hit)	Identity %	Organism
FecA-like associated to FecI/FecR				
Pp-rec-5	Ppu_BIRD-1_4301	Probable TonB-dependent receptor	89	*P. putida* KT2440
Pp-rec-7	Ppu_BIRD-1_0385	TonB-dependent siderophore receptor	95	*P. putida* KT2440
Pp-rec-8	Ppu_BIRD-1_0190	TonB-dependent siderophore receptor	98	*P. putida* KT2440
Pp-rec-14	Ppu_BIRD-1_1056	TonB-dependent siderophore receptor	95	*P. putida* KT2440
Pp-rec-17	Ppu_BIRD-1_2217	TonB-dependent siderophore receptor	98	*P. putida* KT2440
Pp-rec-18	Ppu_BIRD-1_0918	FecA-like outer membrane receptor	99	*P. putida* KT2440
Pp-rec-19	Ppu_BIRD-1_3458	TonB-dependent siderophore receptor	99	*P. putida* KT2440
Pp-rec-20	Ppu_BIRD-1_3580	Ferric-pseudobactin M114 receptor *pbuA*	97	*P. putida* F1
Pp-rec-21	Ppu_BIRD-1_3262	TonB-dependent siderophore receptor	89	*P. putida* GB1
Pp-rec-22	Ppu_BIRD-1_0868	TonB-dependent siderophore receptor	99	*P. putida* F1
FecA solo				
Pp-rec-23	Ppu_BIRD-1_1698	TonB-dependent siderophore receptor	99	*P. putida* F1
Pp-rec-24	Ppu_BIRD-1_2432	Outer membrane ferric siderophore receptor, putative	99	*P. putida* KT2440
Pp-rec-25	Ppu_BIRD-1_2577	Outer membrane ferric siderophore receptor, putative	98	*P. putida* KT2440
Pp-rec-26	Ppu_BIRD-1_3090	Outer membrane ferric siderophore receptor, putative	99	*P. putida* KT2440
Pp-rec-27	Ppu_BIRD-1_3412	Ferric enterobactin receptor	98	*P. putida* KT2440

Pp-rec-28	Ppu_BIRD-1_2426	TonB-dependent siderophore receptor	97	*P. putida* KT2440
Pp-rec-29	Ppu_BIRD-1_0575	TonB-dependent siderophore receptor	98	*P. putida* KT2440
Pp-rec-30	Ppu_BIRD-1_3267	TonB-dependent siderophore receptor	97	*P. putida* KT2440
Pp-rec-31	Ppu_BIRD-1_0912	TonB-dependent siderophore receptor	99	*P. putida* KT2440
Pp-rec-32	Ppu_BIRD-1_0300	TonB-dependent siderophore receptor	98	*P. putida* KT2440
Pp-rec-33	Ppu_BIRD-1_0294	TonB-dependent siderophore receptor	99	*P. putida* KT2440
Pp-rec-34	Ppu_BIRD-1_3765	TonB-dependent siderophore receptor	99	*P. putida* KT2440
Pp-rec-35	Ppu_BIRD-1_2177	TonB-dependent siderophore receptor	99	*P. putida* KT2440
Pp-rec-36	Ppu_BIRD-1_2178	TonB-dependent siderophore receptor	99	*P. putida* KT2440
Pp-rec-37	Ppu_BIRD-1_3161	TonB-dependent siderophore receptor	98	*P. putida* F1
Pp-rec-38	Ppu_BIRD-1_4461	TonB-dependent siderophore receptor	95	*P. putida* F1

The upper table shows *fecA* genes physically linked to *fecI/fecR* regulatory systems**.** The lower table shows non-linked FecA-like. The PP_BIRD-1_xxxx number locates the corresponding receptor in the genome of BIRD-1 strain. The annotation refers to the best blast hit with sequences deposited in the SWISSPROT/TrEMBL database and we indicate the degree of identity and the microorganism source that gave the best hit.

Biosynthesis of IAA. Indole-3-acetic acid is the most physiologically active auxin in plants, which is known to stimulate both rapid (e.g. increases in cell elongation) and long-term (e.g. cell division and differentiation) responses in plants (Cleland, 1990; Spaepen *et al*., 2007). IAA is the most common and best-characterized phytohormone. It has been estimated that 80 % of bacteria isolated from the rhizosphere can produce IAA, particularly in medium with tryptophan (Patten and Glick, 1996). Roca *et al*. (2013) found that *P. putida* BIRD-1 produced and excreted IAA to the outer medium (up to 120 ppm per unit of OD_{660}) and that production was at least two fold higher when the medium was supplemented with 3 mM tryptophan (Suppl. Figure 2.4).

In BIRD-1, IAA could be produced through at least two putative tryptophan-dependent IAA biosynthetic pathways that are also present in other *P. putida* strains (Wu *et al*., 2010) (Figure 2.4). BIRD-1 possesses two copies of tryptophan 2-monooxygenase which is involved in the conversion of tryptophan into indole-3-acetamide in one of the pathways (PPU_BIRD-1_0418 and PPU_BIRD-1_1202), while in the other a single gene is present in the initial step that involves the conversion of tryptophan into tryptamine (PPU_BIRD-1_3125). Regarding the secretion of IAA it should be noted that BIRD-1 possesses three genes encoding a putative auxin efflux carrier (PPU_BIRD-1_2233, PPU_BIRD-1_2634 and PPU_BIRD-1_0977), similar to that that facilitates the excretion of the plant hormone to the surrounding medium or soil.

To assign appropriately the IAA production to one of the two pathways or both, each of the two pathways were individually inactivated through the generation of mutants by site-directed mutagenesis using pCHESI-Km and pCHESI-Gm. A single mutant in tryptophan decarboxylase was also constructed using pCHESI-Km, and a double mutant inactivating tryptophan decarboxylase and one of the tryptophan 2-monooxygenases was also constructed. The surprising result was that both the single mutant and double mutants produced similar amounts of IAA which were half of that produced by the wild-type strain. These results suggest that a third IAA biosynthesis pathway may exist in BIRD-1.

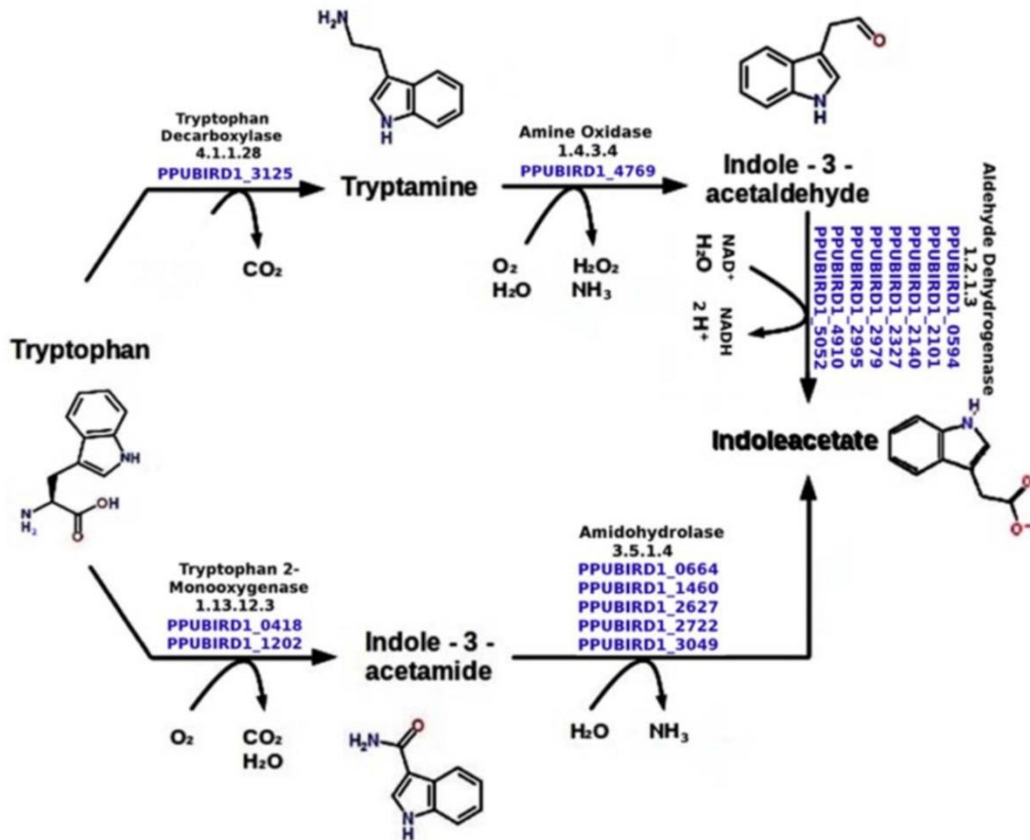


Figure 2 4. Synthesis of IAA based on the annotation of the BIRD-1 genome

Inactivation of the ACC deaminase. Ethylene is a plant growth regulator that influences plant growth, development and senescence. ACC is present in plant root exudates (Reid, 1987) and some PGPR have been proposed to promote plant growth by lowering the levels of ethylene through the activity of the enzyme ACC deaminase, which hydrolyses ACC, the immediate biosynthetic precursor of ethylene in plants (Yang and Hoffman, 1984). The products of this hydrolysis, ammonia and ketobutyrate, can be used by the bacterium as a source of nitrogen and carbon for growth (Klee *et al*., 1991). In this way, the bacterium acts as a sink for ACC and thus lowers ethylene levels in plants, preventing some of the potentially deleterious consequences of high ethylene concentrations (Glick *et al*., 1998; Steenhoudt and Vanderleyden, 2000; Saleem *et al*., 2007). We found that BIRD-1 has an ORF (PPU_BIRD-1_3642) that exhibits high homology to the ACC deaminase of different microorganisms. A mutant with a knock-out in the PPU_BIRD-1_3642 ORF did not influence plant root development under standard growth conditions in soil in controlled chambers (Roca *et al*., 2013) (Suppl. Figure 2.5). Then it was suggested that ACC deaminase does not play a significant role in the development of maize roots.

Concluding remarks

Several mechanisms by which PGPR promote plant growth have been described in the literature. In the genome of *P. putida* BIRD-1 we have identified a set of the genes that are involved in this phenotype. For instance, BIRD-1 is able to solubilize organic and inorganic phosphate, to synthesize plant hormones, to produce siderophores and to survive the oxidative stress imposed by the rhizosphere. Furthermore, the BIRD-1 genome also encodes all of the functions required to adhere to seeds and to colonize roots. BIRD-1 possesses a complete set of chemotaxis genes (Z. Udaondo and J.L. Ramos, unpubl. results) and responds to amino acids in root exudates as a signal (A. Roca and P. Pizarro-Tobías, unpubl. results). The combination of all of these properties makes it an excellent microbe for utilization in agriculture. BIRD-1 promotes the growth of the plant by synthesizing IAA and other chemicals. Furthermore, its ability to survive desiccating conditions grants this strain an additional property of exceptional utility for its use in non-irrigated farming. We view the *Pseudomonas*–plant pair as a mutualistic interaction in which the microorganism solubilizes P and Fe and also produces plant growth hormones that favour plant growth, while the plants release up to 20 % of fixed CO_2 as root exudates that can be used by *Pseudomonas* to proliferate, so that at the end a beneficial interaction is established due to the concomitant high microbial cell densities in the plant rhizosphere. *Pseudomonas putida* BIRD-1 possesses an exceptional arsenal of mechanisms to flourish in the rhizosphere of plants and to promote plant growth. More studies are necessary to firmly establish these abilities in field tests and to determine the PGPR efficiency in a less controlled environment.

References

Antolín, M.C., Pascual, I., García, C., Polo, A., and Sánchez-Díaz, M. (2005) Growth, yield and solute content of barley in soils treated with sewage sludge under semiarid Mediterranean conditions. *Field Crops Res* **94**: 224–237.

Antoun H., Beauchamp, C. J., Goussard N., Chabot, R., Lalande, R. (1998) Potential of *Rhizobium* and *Bradyrhizobium* species as plant growth promoting rhizobacteria on non-legumes: effect on radishes (*Raphanus sativus* L.*). Plant and Soil* **204**:57-67

Bakker, P.A., Pieterse, C.M., and van Loon, L.C. (2007) Induced systemic resistance by fluorescent *Pseudomonas* spp. *Phytopathology* **97**: 239–243.

Banik, S., and Dey, B.K. (1982) Available phosphate content of an alluvial soil as influenced by inoculation of some isolated phosphate solubilizing microorganisms. *Plant Soil* **69**: 353–364.

Beneduzi, A., Ambrosini, A., and Passaglia, L.M.P. (2012) Plant growth-promoting rhizobacteria (PGPR): Their potential as antagonists and biocontrol agents. *Genetics and molecular biology* **35**: 1044–1051

Bielecki, P., Puchalka, J., Wos-Oxley, M.L., Loessner, H., Glik, J., Kawecki, M., *et al*. (2011) In vivo expression profiling of *Pseudomonas aeruginosa* infections reveals niche- specific and strain-independent transcriptional programs. *PLoS One* **6**: e24235.

Blom, D., Fabbri, C., Connor, E.C., Schiestl, F.P., Klauser, D.R., Boller, T., *et al*. (2011) Production of plant growth modulating volatiles is widespread among rhizosphere bacteria and strongly depends on culture conditions. *Environ Microbiol* **13**: 3047–3058.

Caspi, R., Foerster, H., Fulcher, C.A., Kaipa, P., Krummen- acker, M., Latendresse, M., *et al*. (2008) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **36**: D623–D631.

Choi, K.H., Kumar, A., and Schweizer, H.P. (2006) A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: application for DNA fragment transfer between chromosomes and plasmid transformation. *J Microbiol Methods* **64**: 391–397.

Cleland, R.E. (1990) Auxin and cell elongation. *In Plant Hormones and Their Role in Plant Growth and Development.* Davies, P.J. (ed.). Dordrecht, the Netherlands: Kluwer, pp. 132–148.

Daniels, C., Godoy, P., Duque, E., Molina-Henares, M.A., Torre, J., del Arco, J.M., *et al*. (2010) Global regulation of food supply by *Pseudomonas putida* DOT-T1E. *J Bacteriol* **192**: 2169–2181.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**: 4636–4641.

Delcher, A.L, Bratke, K. A., Powers, E.C., Salzberg, S.L. (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer *Bioinformatics* **23**(6):763-679

Dey, R., Pal, K.K., Bhatt, D.M., and Chauhan, S.M. (2004) Growth promotion and yield enhancement of peanut (*Arachis hypogaea* L) by application of plant growth promoting rhizobacteria. *Microbiol Res* **159:** 371–394.

Dobbelaere, S., Vanderleyden, J., and Okon, Y. (2003) Plant growth promoting effects of diazotrophs in the rhizosphere. *Crit Rev Plant Sci* **22**: 107–149.

Duff, R.B., and Webley, D.M. (1959) 2-Ketogluconic acid as a natural chelator produced by soil bacteria. *Chem Ind* **1939**: 1376–1377.

Duque, E., Torre, J., Bernal, P., Molina-Henares, M.A., Ala- minos, M., Espinosa-Urgel, M., *et al*. (2012) Identification of reciprocal adhesion genes in pathogenic and nonpathogenic *Pseudomonas*. *Environ Microbiol* [Epub ahead of print]. DOI: 10.1111/j.1462-2920.2012.02732.x.

Enderle, P.J., and Farwell, M.A. (1998) Electroporation of freshly plated *Escherichia coli* and *Pseudomonas aeruginosa* cells. *Biotechniques* **25**: 954–958.

Espinosa-Urgel, M., and Ramos, J.L. (2004) Cell density-dependent gene contributes to efficient seed colonization by *Pseudomonas putida* KT2440. *Appl Environ Microbiol* **70**: 5190–5198.

Espinosa-Urgel, M., Salido, A., and Ramos, J.L. (2000) Genetic analysis of functions involved in adhesion of *Pseudomonas putida* to seeds. *J Bacteriol* **182**: 2363–2369.

Fernández-Piñar, R., Cámara, M., Dubern, J.F., Ramos, J.L., and Espinosa-Urgel, M. (2011) The *Pseudomonas aeruginosa* quinolone quorum sensing signal a the multicellular behaviour of *Pseudomonas putida* KT2440. *Res Microbiol* **162**: 773–781

Glick, B.R., Penrose, D.M., and Li, J. (1998) A model for lowering plant ethylene concentration by plant growth promoting rhizobacteria. *J Theor Biol* **190**: 63–68.

Mügi, B., Orange, N., Hellio, F., Burini, J.F., Guillou, C., Leriche, F., and Guespin-Michel, J.F. (1991) Effect of growth temperature on several exported enzyme activities in the psychrotropic bacterium *Pseudomonas fluorescens*. *J Bacteriol* **173**: 3814–3820.

Halder, A.K., and Chakrabarty, P.K. (1993) Solubilization of inorganic phosphate by Rhizobium. *Folia Microbiol* **38**: 325–330.

Halder, A.K., Mishra, A.K., Bhattacharya, P., and Chakrabarthy, P.K. (1990) Solubilization of rock phosphate by *Rhizobium* and *Bradyrhizobium*. *J Gen Appl Microbiol* **36**: 1–92.

Jiménez, J.I., Miñambres, B., García, J.L., and Díaz, E. (2002) Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ Microbiol* **4**: 824–841.

Jones, D.L., and Darrah, P.R. (1994) Role of root derived organic acids in the mobilization of nutrients from the rhizosphere. *Plant Soil* **166**: 247–257.

Jorquera, M.A., Hernández, M.T., Renger, Z., Marschner, P., and Mora, M. (2008) Isolation of culturable phosphobacteria with both phytate-mineralization and phosphate-solubilization activity from the rhizosphere of plants grown in a volcanic soil. *Biol Fertil Soils* **44**: 1025–1034.

Klee, H.J., Hayford, M.B., Kretzmer, K.A., Barry, G.F., and Kishore, G.M. (1991) Control of ethylene synthesis by expression of a bacterial enzyme in transgenic tomato plants. *Plant Cell* **3**: 1187–1193.

Li, J., Ovakin, D.H., Charles, T.C., and Glick, B.R. (2000) An ACC deaminase minus mutant of *Entreobacter cloacae* UW4 no longer promotes root elongation. *Curr Microbiol* **41**: 101–105.

Lugtenberg, B., and Kamilova, F. (2009) Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol* **63**: 541– 556.

McGrath, J.W., Hammerschmidt, F., and Quinn, J.P. (1998) Biodegradation of phosphonomycin by *Rhizobium huakuii* PMY1. *Appl Environ Microbiol* **64:** 356–358.

Martínez-Bueno, M., Tobes, R., Rey, M., and Ramos, J.L. (2002) Detection of multiple extracitoplasmatic function (ECF) sigma factors in the genome of *Pseudomonas putida* KT2440 and their counterparts in *Pseudomonas aeruginosa* PAO1. *Environ Microbiol* **4**: 842–855.

Matilla, M.A., Espinosa-Urgel, M., Rodriguez-Herva, J.J., Ramos, J.L., and Ramos-Gonzalez, M.I. (2007) Genomic analysis reveals the major driving forces of bacterial life in the rhizosphere. *Genome Biol* **8**: R179.

Matilla, M.A., Ramos, J.L., Bakker, P.A.H., Doornbos, R., Badri, D.V., Vivanco, J.M., and Ramos-González, M.I. (2010) *Pseudomonas putida* KT2440 causes induced systemic resistance and changes in *Arabidopsis* root exudation. *Environ Microbiol Rep* **2**: 381–388.

Matilla, M.A., Pizarro, P., Roca, A., Fernández, M., Duque, E., Molina, L., *et al*. (2011) Complete genome of the plant-growth promoting rhizobacterium *Pseudomonas putida* BIRD-1. *J Bacteriol* **193**: 1290.

Miller, S.H., Browne, P., Prigent-Combaret, C., Combes- Meynet, E., Morrissey, J.P., and O'Gara, F. (2010) Biochemical and genomic comparison of inorganic phosphate solubilization in *Pseudomonas* species. *Environ Microbiol Rep* **2**: 403–411.

Molina, L., Ramos, C., Duque, E., Ronchel, M.C., García, J.M., Wyke, L., and Ramos, J.L. (2000) Survival of *Pseudomonas putida* KT2440 in soil and in the rhizosphere of plants under greenhouse and environmental conditions. *Soil Biol Biochem* **32:** 315–321.

Molina-Henares, M.A., Ramos, J.L., and Espinosa-Urgel, M. (2006) A two-partner secretion system is involved in seed and root colonization and iron uptake by *Pseudomonas putida* KT2440. *Environ Microbiol* **8**: 639–747.

Molina-Henares, M.A., García-Salamanca, A., Molina- Henares, A.J., Torre, J., Herrera, M.C., Ramos, J.L., and Duque, E. (2009) Functional analysis of aromatic biosynthetic pathways in *Pseudomonas putida* KT2440. *Microb Biotechnol* **2**:91-100

Muñoz-Rojas, J., Bernal, P., Duque, E., Godoy, P., Segura, A., and Ramos, J.L. (2006) Involvement of cyclopropane fatty acids in the response of *Pseudomonas putida* KT2440 to freeze-drying. *Appl Environ Microbiol* **72**: 472–477.

Murphy, J., and Riley, J.P. (1958) Asingle-solution method for the determination of soluble phosphate in sea water. *J Mar Biolog Assoc UK* **37**: 9–14.

Naik, P.R., Sahoo, N., Goswani, D., Ayyadurai, N., and Sak- thivel, N. (2008) Genetic and functional diversity among fluorescent *Pseudomonads* isolated from the rhizosphere of Banana. *Microb Ecol* **56**: 492–504.

Nautiyal, C.S. (1999) An efficient microbiological growth medium for screening phosphate solubilizing microorganisms. *FEMS Microbiol Lett* **170**: 265–270.

Nelson, K.E.,Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A., *et al*. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* **4**: 799–808.

Ohtake, H.,Wu, H., Imazu, K., Anbe, L., Kato, J., and Kuroda, A. (1996) Bacterial phosphonate degradation, phosphite oxidation and polyphosphate accumulation. *Resour Conserv Recy* **18**: 125–134.

Pal, K.K., Tilak, K.V.B., Saxena, A.K., Dey, R., and Singh, C.S. (2001) Suppression of maize root diseases caused by *Macrophomina phaseolina*, *Fusarium moniliforme* and *Fusarium germinearum* by plant growth promoting rhizobacteria. *Microbiol Res* **156:** 209–223.

Patten, C.L., and Glick, B.R. (1996) Bacterial biosynthesis of indole-3-acetic acid. *Can J Microbiol* **42**: 207–220.

Patten, C.L., and Glick, B.R. (2002) Role of *Pseudomonas putida* indole-3-acetic acid in development of the host plant root system. *Appl Environ Microbiol* **68**: 3795– 3801.

Penrose, D.M., and Glick, B.R. (2001) Levels of ACC and related compounds in exudate and extracts of canola seeds treated with ACC deaminase-containing plant growth-promoting bacteria. *Can J Microbiol* **47**: 368– 372.

Prentki, P., and Krisch, H.M. (1984) In vitro insertional mutagenesis with a selectable DNA fragment. *Gene* **29**: 303– 313.

Ramos-González, M.I., Campos, M.J., and Ramos, J.L. (2005) Analysis of *Pseudomonas putida* KT2440 gene expression in the maize rhizosphere: in vivo expression technology capture and identification of root-activated promoters. *J Bacteriol* **187**: 4033–4041.

Reid, M. (1987) Ethylene in plant growth, development and senescence. *In Plant Hormones and Their Role in Plant Growth and Development*. Davies, P.J. (ed.). Boston, MA, USA: Martinus Nijhoff, pp. 257–279.

Richardson, A.E. (2001) Prospect for using soil microorganisms to improve to improve the acquisition of phosphorus by plants. *Aust J Plant Physiol* **28**: 897–906.

Richardson, A.E., and Hadobas, P.A. (1997) Soil isolates of *Pseudomonas* spp. that utilize inositol phosphates. *Can J Microbiol* **43**: 509–516.

Rodríguez, H., and Fraga, R. (1999) Phosphate solubilizing bacteria and their role in plant growth promotion. *Biotechnol Adv* **17**: 319–339.

Saleem, M., Arshad, M., Hussain, S., and Bhatti, A.S. (2007) Perspective of plant growth promoting rhizobacteria (PGPR) containing ACC deaminase in stress agriculture. *J Ind Microbiol Biotechnol* **34**: 635–648.

Salzberg, S.L., Deicher, A.L., Kasif, S., and White, O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Research* **26**: 544–548.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. New York, NY, USA: Cold Spring Harbor Laboratory Press.

Schalk, I.J., Hannauer, M., and Braud, A. (2011) New roles of bacterial siderophores in metal transport and tolerance. *Environ Microbiol* **13**: 2844–2854.

Segura, A., Rodríguez-Conde, S., Ramos, C., and Ramos, J.L. (2009) Bacterial responses and interactions with plants during rhizoremediation. *Microb Biotechnol* **4**: 452–464.

Shekar, C., and Nautiyal, C.S. (1999) An efficient microbiological growth medium for screening phosphorous solubilizing microorganisms. *FEMS Microbiol Lett* **170**: 265–270.

Siddiqui, Z.A. (2006) PGPR: Biocontrol and biofertilization Siddiqui,Z.A. (ed). Dordrecht (The Netherlands):Springer.

Skrary, F.A., and Cameron, D.C. (1998) Purification and characterization of a *Bacillus licheniformis* phosphatase specific for D-alphaglycerophosphate. *Arch Biochem Biophys* **349**: 27–35.

Spaepen, S., Vanderleyden, J., and Remans, R. (2007) Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiol Rev* **31**: 425–448.

Stajner, D., Kevreaan, S., Gasaic, O., Mimica-Dudic, N., and Zongli, H. (1997) Nitrogen and *Azotobacter chroococcum* enhance oxidative stress tolerance in sugar beet. *Biol Plant* **39**: 441–445.

Steenhoudt, O., and Vanderleyden, J. (2000) *Azospirillum*, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. *FEMS Microbiol Rev* **24**: 487–506.

Uroz, S., Buée, M., Murat, C., Frey-Klett, P., and Martin, F. (2010) Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ Microbiol Rep* **2**: 281–288.

Viruel, E., Lucca, M.E., and Sineriz, F. (2011) Plant growth promotion traits of phosphobacteria isolated from puna, Argentina. *Arch Microbiol* **193**: 489–496.

Vyas, P., and Gulati, A. (2009) Organic acid production in vitro and plant growth promotion in maize under controlled environment by phosphate-solubilizing fluorescent *Pseudomonas. BMC Microbiol* **9**: 174.

Weinel, C., Nelson, K.E., and Tümmler, B. (2002) Global features of the *Pseudomonas putida* KT2440 genome sequence. *Environ Microbiol* **4**: 809–818.

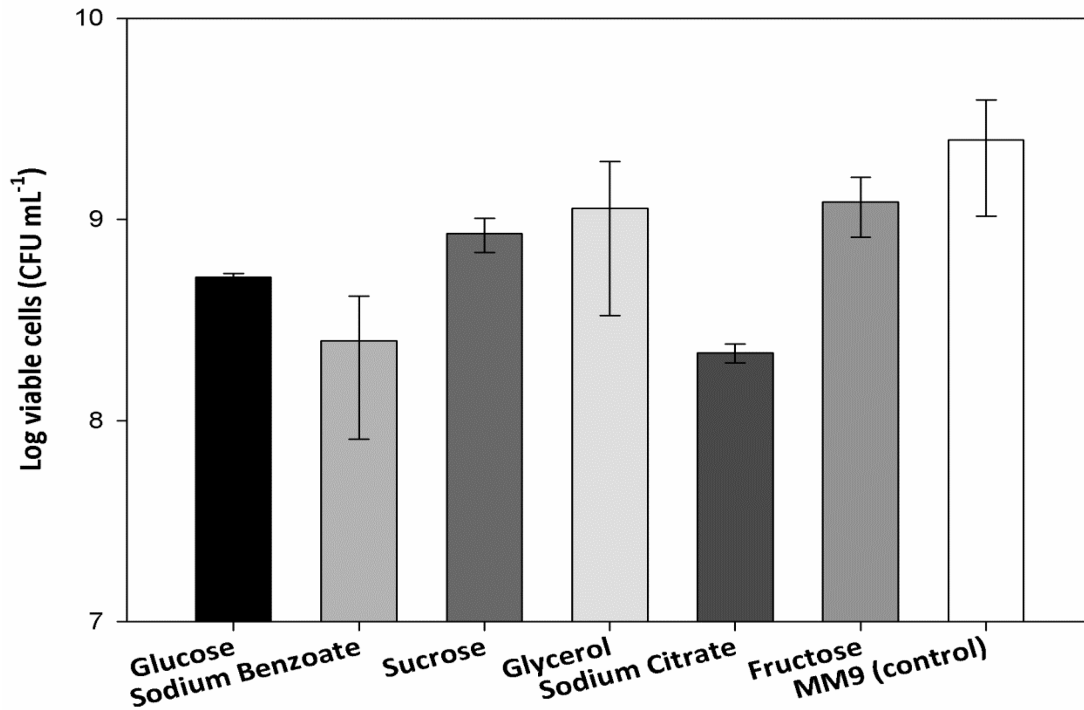
Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J.L., and van der Lelie, D. (2010) Comparative genomics and functional analysis of niche specific adaptation in *Pseudomonas putida*. *FEMS Microbiol Rev* **35**: 299–323.

Yang, S.F., and Hoffman, N.E. (1984) Ethylene biosynthesis and its regulation in higher plants. *Annu Rev Plant Physiol* **35**: 155–189.

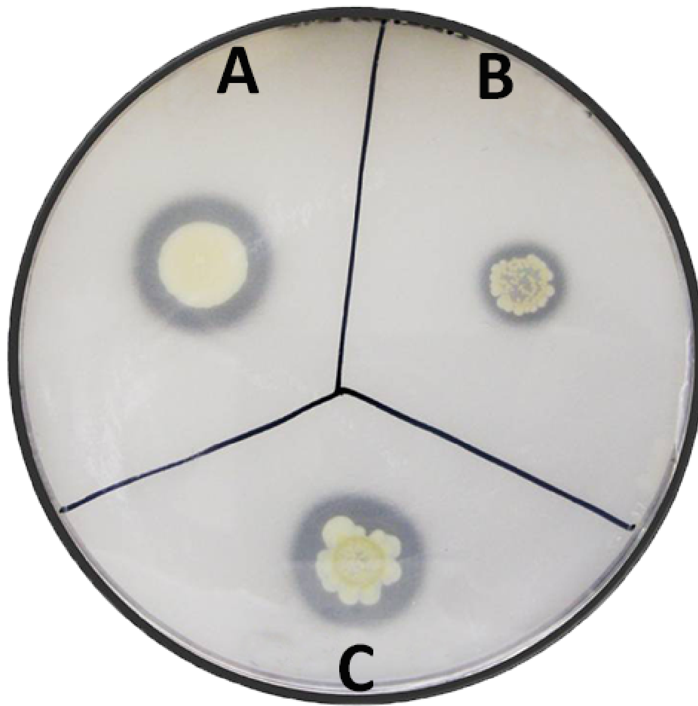
Yousef-Coronado, F., Travieso, M.L., and Espinosa-Urgel, M. (2008) Different, overlapping mechanisms for colonization of abiotic and plant surfaces by *Pseudomonas putida*. *FEMS Microbiol Lett* **288**: 118–124.

.

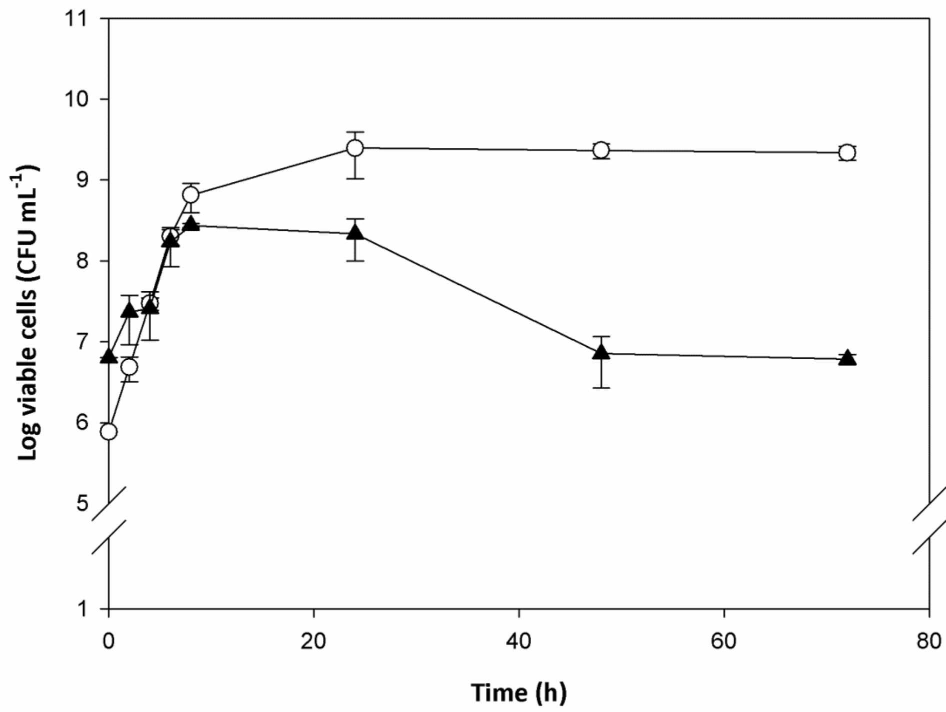
Supporting information



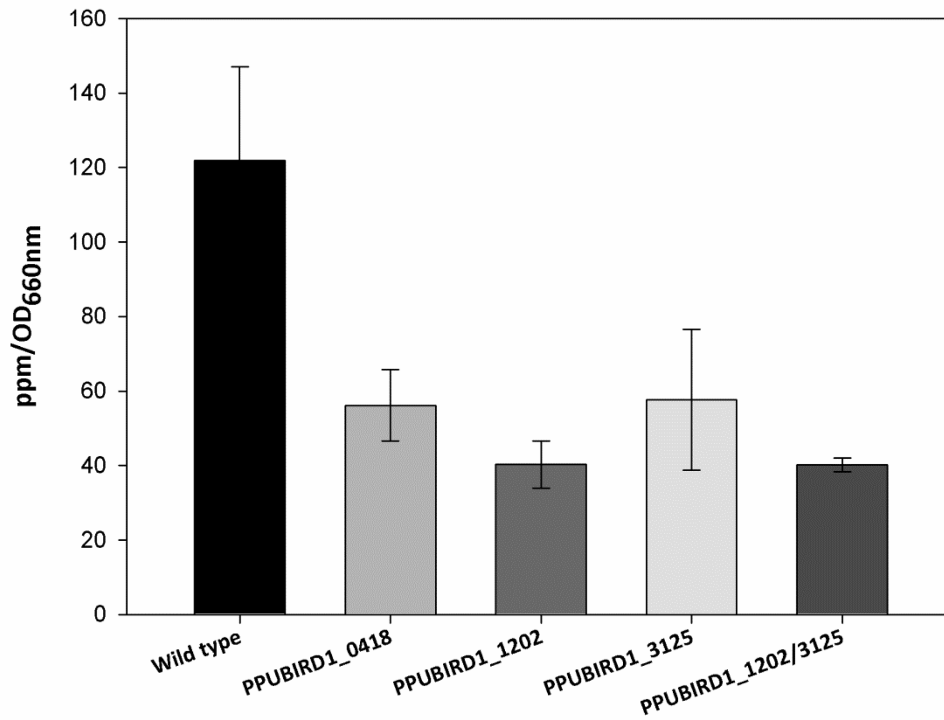
Suppl. Figure 2. 1. Growth of *Pseudomonas putida* BIRD-1 with different sources of inorganic phosphorous. We used *A* medium with the indicated C source and insoluble tricalcium phosphate (500 mg $^{l-1}$) as a source of phosphate. Viable cells at the beginning of the assay were 2 to 3x10^7 c.f.u. ml^{-1}. After 24 h incubation at 30°C the c.f.u. ml^{-1} was determined after spreading serial dilutions on LB medium. The control is M9 medium with 50 mM soluble phosphate.



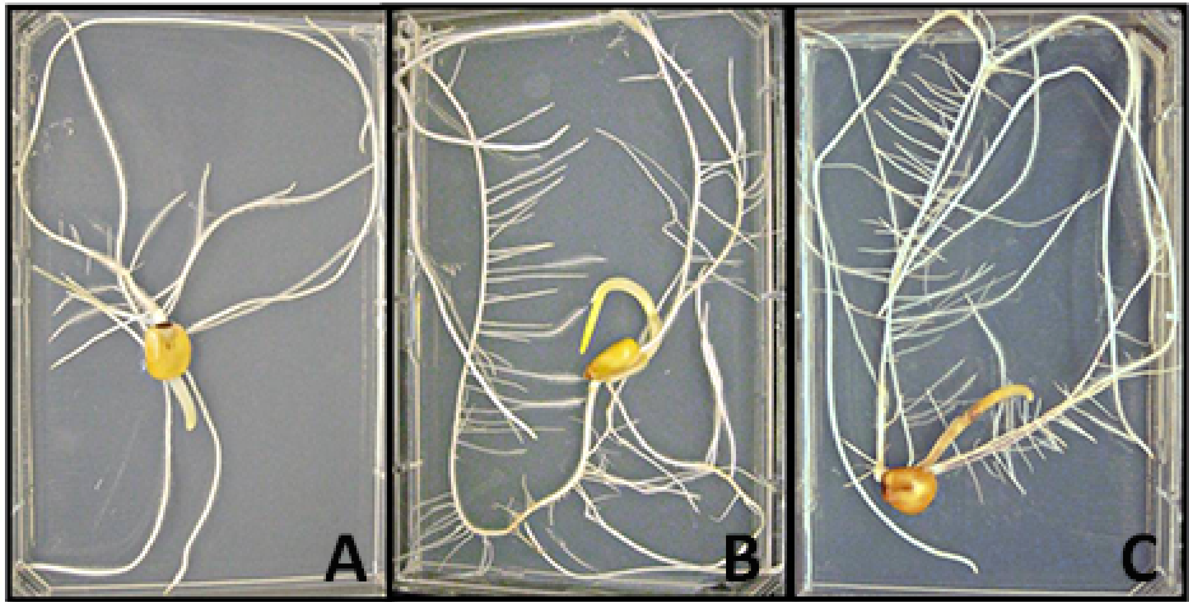
Suppl. Figure 2. 2. Pikovskaya test with BIRD-1 and mutants in the Entner–Doudoroff pathway. Plates were prepared as described in the experimental procedures section of supporting information. We inoculated cells using a toothpick in a single point on the agar surface. Plates were incubated for 48 h at 30°C and then the solubilization halo was visualized. (A) Wild-type; (B) *zwf* mutant and (C) *eda* mutant



Suppl. Figure 2. 3. Growth of *P. putida* with phytate. Assays were performed with *A* medium containing 1 gl^{-1}phytate. At the indicated times c.f.u. ml^{-1}were determined. As a control, cells growing on M9 minimal medium were used. ○, growth in M9 medium; ▲, growth with phytate.



Suppl. Figure 2. 4. Production of IAA by wild-type BIRD-1 and different isogenic mutants. IAA concentration was determined as described in the experimental procedures of supporting information. The solid back bar represents the wild-type strains and the other bars the indicated mutants.



Suppl. Figure 2. 5. Corn root development in a control seed without bacteria (A), and inoculated seeds with BIRD-1 (B) and a mutant deficient in the acc deaminase gene (C). Conditions were as described for Figure 2.2 except that a knockout (PPU_BIRD-1_3642) *acc* mutant was used

.

Experimental procedures

Strains, plasmids and culture media. The bacterial strains cloning vectors and plasmids used in this study have been described before (Molina-Henares *et al*., 2009; Matilla *et al*., 2010). *Pseudomonas putida* BIRD-1 and its mutant derivatives were routinely grown in M9 minimal medium supplemented with glucose [0.5 % (w/v)] as the carbon source (Abril *et al*., 1989). Cultures were incubated at 30°C and shaken on an orbital platform operating at 200 strokes per minute. When required, antibiotics were used at the following final concentrations (in µg ml^{-1}): ampicillin, 100; chloramphenicol, 30; kanamycin, 50; and tetracycline, 20. *Escherichia coli* strains were grown at 37°C in LB broth (Sambrook *et al*., 1989) with shaking.

When indicated *P. putida* cells were grown on modified *A* medium (Nautilaya, 1999) that lacks a phosphorous source and whose specific composition per litre was NH_4Cl, 267 mg; $MgSO_4$ x $7H_2O$, 410 mg; KCl, 300 mg; NaCl, 200 mg, 1 ml of an aqueous solution of iron citrate (6 gl^{-1}) and 0.5 % (w/v) glucose, or 10 mM sodium benzoate as the carbon sources. As a phosphorous source we added either phosphate rock (5gl^{-1}) or a mixture of insoluble inorganic phosphates made up of $CaHPO_4$ x $2H_2O$ and $Ca_5(OH)(PO_4)_3$.

To screen the ability of *P. putida* strains to use inorganic and organic insoluble phosphate as a phosphorous source Pikovskaya agar plates (Naik *et al*., 2008) and PSM (phytate screening medium) agar solid medium were used (Jorquera *et al*., 2008). The reaction was positive when a clear halo surrounding the bacterial colonies was formed after 3–7 days of incubation at 30°C. Chrome azurol S agar solid medium was used to screen siderophore production (Alexander and Zuberer, 1991). The reaction was considered positive when an orange halo surrounding the bacterial colony was formed after 3–7 days of incubation at 30ºC.

Chrome azurol S agar solid medium was used to screen siderophore production (Alexander and Zuberer, 1991). The reaction was considered positive when an orange halo surrounding the bacterial colony was formed after 3–7 days of incubation at 30°C.

To screen the ability of *P. putida* strains to produce IAA on solid medium, amended LB and the Salkowski's reagent were used (Naik *et al*., 2008). The composition of the amended LB per litre was 10 g of bactotryptone, 5 g of yeast

extract, 10 g of NaCl, 5 mM L-tryptophan, 600 mg of sodium dodecyl sulfate (SDS), 10 ml of glycerol and 15 g of agar. The composition of the Salkowski's reagent per litre was 2 % v/v of 0.5 M $FeCl_3$ dissolved in 35 % perchloric acid. Quantification of IAA in liquid medium was performed after growth of the strains for 24 h at 30°C with different concentrations of L-tryptophan (0, 100, 250, 500 and 1000 mg l^{-1}, as described) (Patten and Glick, 2002).

Quantification of soluble phosphate in culture medium. Determination of soluble phosphate in solution was per- formed as described by Murphy and Riley (1958). Superna- tants at 0, 2, 4, 8, 24, 48 and 72 h from *P. putida* BIRD-1 cultures grown in phosphate rock (5 gl^{-1}) were mixed with a reagent whose composition per 250 ml was 125 ml of H_2SO_4 (5N), 37.5 ml of 4 % ammonium molybdate solution and 75 ml of 0.1 M ascorbic acid. Forty millilitres of each supernatant was mixed with 8 ml of the reagent solution and incubated for 30 min in a 60°C water bath. After cooling, absorbance at 620 nm was measured. The amount of soluble phosphate (mgl^{-1}) present in the culture supernatants was determined by comparing absorbance measures with a PO_4^{3-} standard curve.

Siderophore production. Strains were inoculated in M9 minimal medium supplemented with glucose (25 mM). Iron citrate (6 mgl^{-1}) was added to the control cultures (plus iron) but not to the iron- deprived ones. When the cultures reached an OD at 660 nm of around 1, they were illuminated with ultraviolet light in a LKB 2011-002 MacroVue transilluminator (LKB Bromma, Sweden) and photographed (GelDoc 2000, Bio-Rad). To record the UV-visible spectrum (220–700 nm) of the culture supernatant, 1 ml of culture was withdrawn, cells removed by centrifugation (14,000 g x 5 min) and the supernatant used to measure the absorption spectrum.

Toxic compound resistance assays. Individual colonies of *P. putida* BIRD-1 and mutant strains were picked from freshly cultured LB plates, streaked onto LB medium plates, supplemented with the suitable antibiotic and grown overnight at 30°C. The biomass of this overnight plate was recovered from the plate surface and resuspended in 15 ml of LB liquid medium to an OD_{660} of 0.1. Microwell plate wells were filled with 180 ml of the above cell suspension and with 20 ml of the different 10 x concentrated solutions of the stressors. Samples were incubated and data

recordings were processed using a Bioscreen C MBR analyser as described before (Daniels *et al*., 2010). Toxic compounds tested were: $K_2Cr_2O_7$ (12.5 $mgml^{-1}$), H_2O_2 (0.004 %), methylviologen (100 mM), tert-butyl hydroperoxide (0.00078 %), KCN (0.325 $mgml^{-1}$) and ampicillin (100 $mgml^{-1}$).

Surface sterilization, germination of seeds and root colonization assay. These assays were performed as described by Ramos-González and colleagues (2005). Corn seeds were surface-sterilized by rinsing with sterile deionized water, washing for 10 min with 70 % (v/v) ethanol, then for 15 min with 10 % (v/v) bleach, and followed thorough rinsing with sterile deionized water. Surface-sterilized seeds were pre-germinated on water agar [2.25 % agar (w/v)] at 30°C in the dark for 48 h. For root colonization assays seeds were inoculated with approximately $5x10^6$ c.f.u. ml^{-1} from a LB medium overnight culture and suspended in M9 salts medium (Sambrook *et al*., 1989). After incubation without shaking for 30 min at 30°C, seeds were washed in sterile distilled water and planted in 50 ml Falcon tubes containing 35 ml of a mixture of sterile peat and washed sand (3:1 v/v) and were then maintained in a controlled chamber at 24°C and 55–65 % humidity with a daily light period of 16 h. At the indicated times plants were collected, shoots discarded and the roots placed in 50 ml Sterilin tubes containing 10 ml of M9 minimal salts medium (Sambrook *et al*., 1989) and 4 g of glass beads (0.3 mm diameter). Tubes were vortexed for 1 min and the number of c.f.u. attached to the surface of the root was estimated by drop-plating serial dilutions of the supernatant in selective media.

Corn seeds adhesion assays. Strains were cultured overnight at 30°C in LB broth and culture turbidity was adjusted to OD_{660} = 1 in a final volume of 1 ml of M9 minimal salts and the assays conducted as described by Espinosa-Urgel and Ramos (2004).

Corn seeds growth on agar plates. Seeds were sterilized, as described above, and submerged for 30 min without shaking on overnight cultures (108 c.f.u. ml^{-1}) of the different strains under test. Seeds were rinsed with sterile distilled water to wash out non-attached cells, and then placed on water agar plates (agar 2.25 % w/v). Growth was checked after 7 days of incubation in the dark at 28°C. Roots were then weighed, for fresh weight, and stored in a stove at 90°C for 48 h.

Competitive root colonization assays. Surface sterilization, germination of seeds and bacterial inoculation were performed as described above, except that seedlings were inoculated with a mixture of a SmR BIRD-1 strain, as the wild type, and the mutant strain in the specified gene. Inocula size differences between the wild-type and mutant strains were less than 2 %. At the indicated times, bacterial cells were recovered from the rhizosphere, as specified above. LB agar supplied with rifampin and streptomycin (or kanamycin) was used to select BIRD-1 or the mutant strain respectively.

Microcosm assays. Sterile maize seedlings were either inoculated or not with approximately 10^6 *P. putida* BIRD-1 cells per entire seedling (Espinosa-Urgel *et al.*, 2000). The seeds were then planted in sterilized potting material (Compo Sana Universal) and grown for 10 days. Then the plantlets were transplanted into plant pots containing 600 g of sterilized loamy silt soil classified as Humic Haploxerept with the following characteristics: 38 % sand, 43 % silt and 19 % clay, pH 7.9, the organic matter content was 2.1 % and its $CaCO_3$ content was 8 %. To determine drought tolerance of wild-type and mutant strains, soil water content was adjusted to different degrees. Viable microorganisms in the rhizosphere soil (soil closely associated to root) and bulk soil from four selected plants were determined, as described above.

Germination of vegetable seeds. Non-sterile vegetable seeds (>30 per treatment) were sown in seed beds, previously filled with a mixture of non-sterile peat and washed sand (3:1 v/v) for control treatments. For the treated seeds, a 109 c.f.u. ml^{-1} culture of the assayed strain was diluted (1:10 v/v), and mixed (1:250 v/wt) with the peat sand mixture; vegetable seeds were then sown and watered with distilled water to achieve appropriate substrate moisture. Seed beds were covered to allow germination in the dark. Approximately 72 h later, when germination was accomplished and germination rate was calculated, seed beds were uncovered to allow stem development, and watered, with tap water, when necessary.

Greenhouse assays. Vegetable seeds of horticultural importance in the Mediterranean area (corn, cucumber, zucchini, lettuce, chard, pepper, melon, onion, tomato and bean) were chosen for greenhouse assays. Seeds were germinated as described above, and plants were watered with tap water when necessary along the

assay. Monitoring of growth was performed every month by measuring 30 plants of each treatment, stem length and gauge, when necessary. The root length was measured after 3 months, and dry weight was calculated for each treatment. Survival of inoculated cells was screened every 15 days, by drop-plating serial dilutions of rhizosphere soil (soil closely associated to roots) in selective media, as described before.

Measurement of phosphatase activity in soil. Phosphatase activity was determined as described by Antolín and colleagues (2005). The amount of *p*-nitrophenol (PNP) released from 0.5 g of soil from each treatment (bulk soil and non-inoculated corn rhizosphere soil taken as controls versus BIRD-1 inoculated corn rhizosphere soil) was measured after incubation at 37°C for 120 min with 0.115 M 4-nitrophenyl phosphate-disodium (PNPP) as substrate for the enzymatic reaction, in 2 ml of maleate buffer (0.1 M, pH 6.5). Then, samples were cooled at 2°C for 15 min to stop enzymatic reaction, and 0.5 ml of 0.5 M $CaCl_2$ and 2 ml of 0.5 M NaOH were added and well mixed. Each sample was centrifuged at 2000 g for 10 min. A blank experiment was performed for each assay, in which the substrate was added to the soil sample after incubation and before stopping the reaction. The amount of PNP per hour released from each soil sample (mg PNP $g^{-1}h^{-1}$) was determined by comparing absorbance measures with a PNP standard curve.

In vitro nucleic acid techniques. Total DNA extraction was performed as described previously (Rodríguez-Herva *et al.*, 2007). Plasmid DNA was isolated using the Qiagen spin miniprep kit. For DNA digestion, the manufacturer's instructions were followed (Roche and New England Biolabs). Ligation of DNA and Southern blots were performed following the standard protocols (Sambrook *et al.*, 1989). Highly electro-competent cells were prepared as previously reported (Choi *et al.*, 2006) and transformed using an EC100 electroporator according to the manufacturer's instructions. Expand high-fidelity Taq polymerase was used in the the amplification of PCR fragments for cloning.

Construction of mutants. For the construction of mutants, gene fragments between 400 and 1000 bp were amplified with the appropriate oligonucleotides and cloned into pMBL1-T plasmid (Dominion MBL). Plasmids containing the corresponding inserts were then cut with BamHI (that cuts into the poly-linker of the

vector, but not into the insert sequence), and the ΩKm resistance cassette of plasmid pHP45Ω-Km (Prentki and Kirsch, 1984), previously excised with BamHI, was ligated into the BamHI restriction site. Approximately 600 ng of the resulting plasmid was electroporated into *P. putida* BIRD-1 (Enderle and Farwell, 1998) to allow a single recombination event between the cloned DNA fragment and its homologous sequence. The resulting kanamycin (25 $\mu g ml^{-1}$) resistant clones were selected and analysed by Southern blot. Clones that contained the insertion in the appropriate location were kept for further analysis.

Chapter III

Antibiotic resistance determinants in a *Pseudomonas putida* strain isolated from a hospital

The information of this chapter was published as:

Lázaro Molina*, Zulema Udaondo*, Estrella Duque, Matilde Fernández, Carlos Molina-Santiago, Amalia Roca, Mario Porcel, Jesus de la Torre, Ana Segura, Patrick Plesiat, Katy Jeannot, Juan Luis Ramos (2014) Antibiotic resistance determinants in a *Pseudomonas putida* strain isolated from a hospital *PloS One*.**9**(1): e81604

Summary

Environmental microbes harbor an enormous pool of antibiotic and biocide resistance genes that can impact the resistance profiles of animal and human pathogens via horizontal gene transfer. *Pseudomonas putida* strains are ubiquitous in soil and water but have been seldom isolated from humans. We have established a collection of *P. putida* strains isolated from in-patients in different hospitals in France. One of the isolated strains (HB3267) kills insects and is resistant to the majority of the antibiotics used in laboratories and hospitals, including aminoglycosides, ß-lactams, cationic peptides, chromoprotein enediyne antibiotics, dihydrofolate reductase inhibitors, fluoroquinolones and quinolones, glycopeptide antibiotics, macrolides, polyketides and sulfonamides. Similar to other *P. putida* clinical isolates the strain was sensitive to amikacin. To shed light on the broad pattern of antibiotic resistance, which is rarely found in clinical isolates of this species, the genome of this strain was sequenced and analysed. The study revealed that the determinants of multiple resistances are both chromosomally-borne as well as located on the pPC9 plasmid. Further analysis indicated that pPC9 has recruited antibiotic and biocide resistance genes from environmental microorganisms as well as from opportunistic and true human pathogens. The pPC9 plasmid is not self-transmissible, but can be mobilized by other bacterial plasmids making it capable of spreading antibiotic resistant determinants to new hosts.

Introduction

Human disease outbreaks are increasing at an alarming rate. One of the most recent and serious occurred in Germany, involving a Stx2a-producing *Escherichia coli* (STEC) strain of serotype O104:H4 that caused more than 4000 cases of illness and 50 deaths. This strain exhibited resistance to numerous antibiotics making it difficult to eradicate (Bielaszewska *et al*., 2011). Horizontal gene transfer has been proposed as the most likely genetic event for the spread of multidrug resistant phenotypes in pathogens (Rasko *et al*., 2011); however, a question that still needs to be answered is 'What is the origin of these acquired antibiotic resistant determinants?'

Pseudomonas putida strains are typically found in soil and water and members of this species have a broad metabolic versatility, which allows them to adapt to different habitats and nutritional environments (Fernández *et al*., 2009; Dogan *et al*., 2011; Tang *et al*., 2011; Wu *et al*., 2010). Strains of this species have occasionally been isolated from patients in hospitals in Japan, the USA, Italy and France. Infections by these microorganisms have been reported to be linked to insertion of catheters or drainage tubes (von Graevenitz and Weinstein, 1971; Yoshino *et al*., 2011). Hospital isolates of *P. putida* are often resistant to β-lactams (Loiseau-Marolleau and Malarre, 1977; Docquier *et al*., 2003; Yomoda *et al*., 2003), and instance Yomoda *et al*. (Yomoda *et al*., 2003) reported that of 32 *P.* putida strains isolated in hospitals in Japan twenty two of them harbored plasmids transferable to *P. aeruginosa* by conjugation or transformation. The same study also indicated that a number of plasmids from these clinical isolates were responsible for resistance to aminoglycosides. Apart from the fact that opportunistic microbes could become 'specialized' pathogens able to attack the most vulnerable immunocompromised patients (Levy, 2002; Yomoda *et al*., 2003); the ability to transfer antibiotic resistant determinants from non-pathogenic species to pathogens in hospital environments is a serious concern (Levy, 2002; Yomoda *et al*., 2003). The Hospital of Besançon in France has established a collectionof *P. putida* isolates from in-patients, and in agreement with von Greenitz and Weinstein (von Graevenitz and Weinstein, 1971), it has been found that these strains have a low pathogenic potential when compared with *P. aeruginosa* PAO1 using virulence assays in a insect model (our unpublished results). We analyzed 15 of these isolates and found that one of them, *P. putida*

HB3267 (Hospital of Bensançon 3267), was able to kill insects and exhibited resistant to a large number of antibiotics. To shed light on the unusual pattern of antibiotic resistance of the strain HB3267 we sequenced and analysed the genome. The analysis revealed that a number of genes involved in multi-drug resistant phenotypes are located in a non-self-transmissible plasmid that was shown to be an efficient vehicle for spreading antibiotic resistance between different *Pseudomonas* strains.

Materials and Methods

DNA analysis and identification of the HB3267 strain. Amplification of 16S rDNA using HB3267 chromosomal DNA was performed with the F8 and R798 primers and the complete sequence of the gene compared with 16S rDNA sequences in databases (Janda and Abbott, 2007). Aranda-Olmedo *et al*. (Aranda-Olmedo *et al*., 2002) showed that *P. putida* strains are characterized by the presence of a highly conserved 35-mer REP sequence. A primer based on the KT2440 REP sequence was used to amplify HB3267 DNA. Positive *(P. putida* KT2442, (Bagdasarian *et al*., 1981)) and negative (*Escherichia coli*) DNA controls were included. The REPc method allows identification of *P. putida* strains (Aranda-Olmedo *et al*., 2002) since this primer amplifies only DNA from this species, producing products of different sizes for each strain. For multilocus sequence typing (MLST) we used a set of primers to amplify RNA polymerase sigma factor *rpoD*, DNA gyrase subunit B *gyrB*, N-(59- phosphoribosyl) anthranilate isomerase *trpF*, 6-phosphogluconate dehydratase *edd*, and recombinase A *recA* genes (Frapolli *et al*., 2007; Khan *et al*., 2008). The complete gene sequences were obtained, translated into the protein sequence and compared as described (Khan *et al*., 2008).

Antibiograms. For these assays 31 different antibiotics were used (Biomerieux commercial disk). Overnight cultures of HB3267 were spread on 240x240 mm LB plates, air dried in a laminar flow and then discs containing antibiotics were placed on the LB plates. Plates were incubated at 30ºC for 16 h. Halos surrounding the discs were measured as an indication of inhibition of growth. This assay was repeated at least three times in duplicate.

Minimal inhibitory concentration (MIC) assay. These assays were performed in 96-well plates, using LB medium and the following stock antibiotic solutions: tetracycline (Tc), 10 mgml^{-1}; kanamycin (Km), 25 mgml^{-1}; gentamicin (Gm), 100 mgml^{-1}; nalidixic acid (Nal), 10 mgml^{-1}; spectinomycin (Sp), 100 mgml^{-1}; rifampicin (Rif), 10 mgml^{-1}; chloramphenicol (Cm), 30 mgml^{-1}, ampicillin (Ap), 100 mgml^{-1}, norfloxacin (Nor), 20 mgml^{-1}l and ceftriaxone (Cro), 25 ml. Serial 10-fold dilutions of the stock antibiotic solutions were prepared and 10 µl of each of these dilutions added to 190 µl of LB, minimal medium M9 and Mueller-Hinton broth medium. Optically standardized 18 hour cultures (10 mL) of *P. putida* strains were used as inoculum. The 96-well plates were incubated at 30uC and 200 rpm overnight and culture turbidity was measured as an indication of growth. The MIC value was established as the lowest concentration at which an antibiotic inhibits growth >90 % (Andrews, 2001). The assays were repeated three times in duplicate.

Biofilm susceptibility testing. Biofilm assays were performed in 96-well flat-bottomed polystyrene microtitre plates. An aliquot of 100 µl of a bacterial suspension contains 105 c.f.uml^{-1} was added to each well and incubated for 5–6 h at 30ºC. Subsequently, liquid culture medium was removed; the wells of the plates washed twice to eliminate all planktonic cells and finally serially diluted antibiotics in LB medium added. These plates were incubated overnight at 30ºC, and after removing planktonic cells as above 100 ml LB was added, and the biofilm cells released by 5 minutes low intensity sonication (Branson 1510 waterbath ultrasonicator). The minimal biofilm eradication concentration (MBEC) was defined as the minimal concentration of antibiotic required to eradicate the biofilm (Ceri *et al*., 2001).

Conjugation experiments. *Pseudomonas putida* KT2440 (Tel), a tellurite resistant strain and *P. putida* HB3267 were grown overnight on LB medium. For biparental matings 1 ml cultures with a turbidity of around 1 at 660 nm were mixed, harvested by centrifugation, wasted with LB and resuspended in 50 ml LB that was laid on nitrocellulose filter disks placed on LB plates. After overnight incubation, transconjugants were selected on LB medium containing Tel, Sm and Tc. For triparental mating experiments, receptor and donor strains were used as previously described, but the pWW0 (Williams and Murray, 1974) and pRK600 plasmids were used as helper plasmids.

Sequencing. Genomic DNA containing both the chromosome and pPC9 plasmid was purified from strain *P. putida* HB3267, using the Wizard® Genomic DNA Purification Kit and sequenced using *454* technology by Macrogen (Seoul, Korea), and assembled into 278 contigs, providing 25x coverage. These contigs were ordered by comparison (BLASTN) with the genomic sequences from other *P. putida* available in the database (KT2440, NC_002947.3; F1, CP000712.1; GB-1, CP000926.1; W619, CP000949.1; BIRD1, CP002290.1), as well as with a close relative *Pseudomonas entomophila* L48 (NC_008027.1). Genomic gaps were closed by designing primers at the contig ends, followed by PCR and further sequencing of the junction sequences. Genomic DNA was automatically annotated using a program pipeline based on Glimmer 3.0 (Delcher *et al.*, 1999) for gene prediction, and BLAST and RPSBLAST for functional assignment of ORFs, based on sequence similarity to sequences deposited in the NR, Swissprot, COG, Pfam, Smart and Prk databases (Altschul *et al.*, 1997). Finally, automatic annotations were manually curated. The chromosome and plasmid sequences are available through Genbank under accession numbers CP003738 and CP003739, respectively.

Results and Discussion

Identification of strain HB3267

A collection of *Pseudomonas putida* strains isolated from humans has been established in the Bacteriology Laboratory of the University Hospital of Besançon (France). Among these, only one strain named HB3267 which was isolated from an in-patient who died from unknown causes, was able to kill insects (Porcel, M. and Duque, E., In preparation). Since *P. putida* strains are seldom isolated from humans a series of molecular analysis were carried out to unequivocally assign this strain to the putida species. Aranda-Olmedo *et al*. (Aranda-Olmedo *et al.*, 2002) showed that a conserved 35 bp repetitive extragenic palindromic (REP) sequence is specifically associated with *P. putida* strains. PCR analysis was performed using chromosomal DNA of HB3267 as a template and a primer based on the previously defined REP sequence (Aranda-Olmedo *et al.*, 2002); and a positive amplicon was obtained. The same assay was performed with KT2440 as a positive control and E. coli as a negative control, with the expected results. The presence of the REP sequence in the

genome of strain HB3267 suggested the original assignment of this strain to the species *P. putida* based on API classification was correct. To further confirm this result 16S rDNA amplification was carried out and the whole gene encoding the 16S sequence was obtained. Sequence comparison confirmed that HB3267 closest 16S genetic homologues all belong to the *P. putida* species. To establish a relationship with other *P. putida* strains, multi-locus sequence typing assays (MLST assays) were carried out using the housekeeping genes *rpoD*, *gyrB*, *trpF*, *edd* and *recA*; these analyses confirmed that the gene products of HB3267 exhibited the highest identity with the gene products of several *P. putida* strains. Phylogenetically, strain HB3267 was closest to *P. putida* S16, a nicotine degrader isolated from a field under continuous tobacco cropping (Wang *et al*., 2007; Dereeper *et al*., 2008) (Suppl. Figure 3.1).

Antibiotic resistance profile of P. putida strain **HB3267**

A distinctive characteristic of the HB3267 strain was its apparent multidrug resistance compared to other *P. putida* from the same hospital collection. Disk inhibition and MIC assays were performed with the most frequently used laboratory/clinical antibiotics to obtain quantitative data; for comparison we used *P. putida* KT2440R, a well characterized strain as a control (Table 3.1). Antibiogram assays showed that *P. putida* HB3267 was resistant to most of the 31 antibiotics tested in this study (Table 3.1). The exceptions were the aminoglycoside amikacin, as well as rifampicin and nitrofurantoin. A remarkable discovery was that HB3267 was resistant to all fluoroquinolones tested (ciprofloxacin, norfloxacin, pefloxacin and ofloxacin), while KT2440R showed high sensitivity to these antibiotics. The same was true for most aminoglycoside antibiotics tested (gentamycin, kanamycin, neomycin, streptomycin and netilmicin), with the exception of amikacin as mentioned. The two strains also differed in their resistance to polymyxin B, colistin, cefotaxime, amoxicillin, imipenem, cephalosporin and ceftazidime, being that HB3267 was resistant to all of them while KT2440 was sensitive. MIC assays revealed that HB3267 was highly resistant to aminoglycoside antibiotics such as gentamycin, kanamycin and spectinomycin being, depending of the medium employed, at least 220-fold, 75-fold and 333-fold more resistant than KT2440R, respectively.

Table 3. 1. Antibiogram assay.

Antibiotic, concentration (mg)	Strain		Antibiotic group
	KT2440R*	HB3267*	
Ciprofloxacin, 5	2.8	0	fluoroquinolone
Norfloxacin, 10	2.5	0	fluoroquinolone
Pefloxacin, 5	1.9	0	fluoroquinolone
Ofloxacin, 5	1.8	0	fluoroquinolone
Nalidixic acid, 30	0	0	quinolone
Erithromycin, 15	0	0	macrolide
Gentamycin, 10	1.8	0	aminoglycoside c
Kanamycin, 30	2	0	aminoglycoside
Neomycin, 30	1.8	0	aminoglycoside
Streptomycin, 10	0.9	0	aminoglycoside
Amikacin, 30	0	1.7	aminoglycoside
Netilmicin, 30	1.4	0	aminoglycoside
Tetracycline, 30	0	0	polyketide
Polymyxin B, 300	1.3	0	polymyxin
Colistin, 50	1.5	0	polymyxin
Trimethoprim, 20	0	0	dihydrofolate reductase inhibitors
Chloramphenicol, 30	0	0	bacteriostatic
Amoxycillin, 25	1	0	ß-lactam (penicillin)
Carbenicillin, 100	0	0	ß-lactam (penicillin)
Ticarcillin, 70	0	0	ß-lactam (penicillin)
Piperacillin, 10	0	0	ß-lactam (penicillin)
Ampicillin, 10	0	0	ß-lactam (penicillin)
Imipemen, 10	2.8	0	ß-lactam (carbapenem)
Cefotaxime, 30	1.5	0	ß-lactam (cephalosporin)
Ceftazidime, 30	1.6	0	ß-lactam (cephalosporin)
Ceftriaxone, 30	0	0	ß-lactam (cephalosporin)
Sulfonamide G, 20	0	0	sulfonamides
Rifampicin, 30	0.6	1.6	rifamycin
Vancomycin, 30	0	0	glycopeptide
Esperamicin, 100	0	0	chromoprotein enediyne
Nitrofurantoin, 300	1.8	2.4	

*Numbers indicate the size of the inhibition halo in cm. Data average of 3 assays performed in duplicate with standard deviation below 5% of the given values.

The HB3267 strain was also much more resistant to polyketide antibiotics such as tetracycline (23-fold more resistant), quinolone antibiotics such as nalidixic acid (67-fold), β-lactams such as ampicillin (30-fold), and bacteriostatic antibiotics such as chloramphenicol (3-fold). *Pseudomonas putida* KT2440R is a spontaneous mutant obtained by exposure to rifampicin (Espinosa-Urgel *et al*., 2002), which explains why the MIC concentration for this antibiotic was 32-fold higher for KT2440R than for HB3267 (Table 3.2). Compared to other pseudomonad clinical isolates, the HB3267 strain was more resistant (in terms of range of antibiotic and MICs) than *P. aeruginosa* PAO1. For example, HB3267 was 5000 times more resistant to gentamicin than *P. aeruginosa* PAO1 (Muller *et al*., 2011).

Table 3. 2. MIC and MBEC assays with different *Pseudomonas putida* strains

	MIC (LB)		MIC (M9)		MIC (M-H)		MBEC	
Antibiotic	**HB3267**	**KT2440R**	**HB3267**	**KT2440R**	**HB3267**	**KT2440R**	**HB3267**	**KT2440R**
Tetracycline	200	8	450	30	350	15	>10000	2500
Kanamycin	3125	10	1500	10	1500	20	>25000	1600
Gentamicin	10000	20	5000	2	5500	25	>10000	1250
Nalidixic acid	2000	30	>1800	25	>1800	30	>10000	625
Spectinomycin	10000	30	>10000	10	>10000	15	>50000	1250
Rifampicin	62	2000	5	600	10	800	1250	5000
Choramphenicol	1000	376	1200	250	1000	300	>30000	1800
Ampicillin	10000	625	10000	600	10000	600	>100000	12500
Fluoroantimicin	1560	>3000	nt	nt	nt	nt	25000	nt
Amikacin	16	>100	1	2.5	1	2	50	625
Ceftriaxone	325	10	300	7.5	300	10	12500	1600
Norfloxacin	220	10	220	1	240	10	5000	75

Numbers indicate the MIC concentration (mg/ml) and MBEC concentrations (mg/ml) required to inhibit 90 % growth and for biofilm eradication. nt, not tested

Horii *et al*. (Horii *et al*., 2005) analyzed the susceptibility of five clinical isolates of *P. putida* (from patients with acute, repetitive or chronic urinary tract infections) to fluoroquinolones. Similar to HB3267, four of the five isolates were resistant to fluoroquinolones, but in contrast with HB3267, all isolates were susceptible to aminoglycosides. MICs assays revealed that the resistance of HB3267 was at the same range than the described *P. putida* clinical isolates and at least 22-fold more than KT2440 (Horii *et al*., 2005). Treviño *et al*. (Treviño *et al*., 2010) also isolated two *P. putida* strains from immunocompromised patients. These two strains, as for the five isolates of Horii *et al*. (Horii *et al*., 2005), and the HB3267 strain we describe here, are susceptible to amikacin; these results indicate that this antibiotic could be used for treatment of multidrug-resistant *P. putida* strains. Resistance of *P. putida* clinical isolates to β-lactams (Horii *et al*., 2005; Saha *et al*., 2010; Treviño *et al*., 2010; Muller *et al*., 2011), aminoglycosides (Horii *et al*., 2005; Mendes *et al*., 2007; Muller *et al*., 2011), and fluoroquinolones has been described (Horii *et al*., 2005; Rolston *et al*., 2005; Kumita *et al*., 2009; Muller *et al*., 2011); however, it should be noted that the minimal inhibitory concentrations found for HB3267 are much higher than what has been reported for other clinical isolates. HB3267 is the only *P. putida* clinical isolate reported to be resistant to tetracycline. The HB3267 strain is also the only *P. putida* clinical isolate described to be resistant to the biocide sulfonamide, trimethoprim/sulfamethoxazole and colistin, which is a characteristic often associated to *P. aeruginosa* (Poirel *et al*., 2000; Iyer

Parameswaran and Murphy, 2009; Moskowitz *et al*., 2012). It is known that *P. putida* strains are able to form biofilms on biotic and abiotic surfaces (Espinosa-Urgel *et al*., 2002; 2012). Bacterial cells in biofilms are often more resistant to antibiotics than planktonic cells. We used the O'Toole and Kolter approach to produce biofilms of HB3267 which were flat and dense and similar to those produced by the KT2440 strain (Duque *et al*., 2012). We then determined the Minimal Biofilm Eradication Concentration (MBEC) as described by Ceri *et al*. (Ceri *et al*., 2001). We found that both strains, KT2440 and HB3267 were much more resistant forming biofilms than living as planktonic cells, but most of the antibiotic tested, such as gentamicin, ampicillin, tetracycline, kanamycin, nalidixic acid, spectinomicyn and chloramphenicol, that were able to eradicate at high concentrations the KT2440 biofilms, were unable to eradicate those formed by HB3267. Antibiotics that were effective (rifampicin, fluoroantimicin, amikacin, ceftriaxone, and norfloxacin were required in 3 to 40-fold higher concentrations than those required to inhibit more than 90 % growth of planktonic cells (Table 3.2)

Gaining insight into antibiotic resistance through whole genome sequencing of **HB3267**

The total genome size of the *P. putida* HB3267 strain is 5,908,671 bp with a G+C content of 63.61 %, which is similar to the published genomes of other *P. putida* strains. The HB3267 genome comprises a chromosome of 5,829,171 bp (Gene Bank CP003738) and a plasmid of 80,360 bp named pPC9 (Gene Bank CP003739).

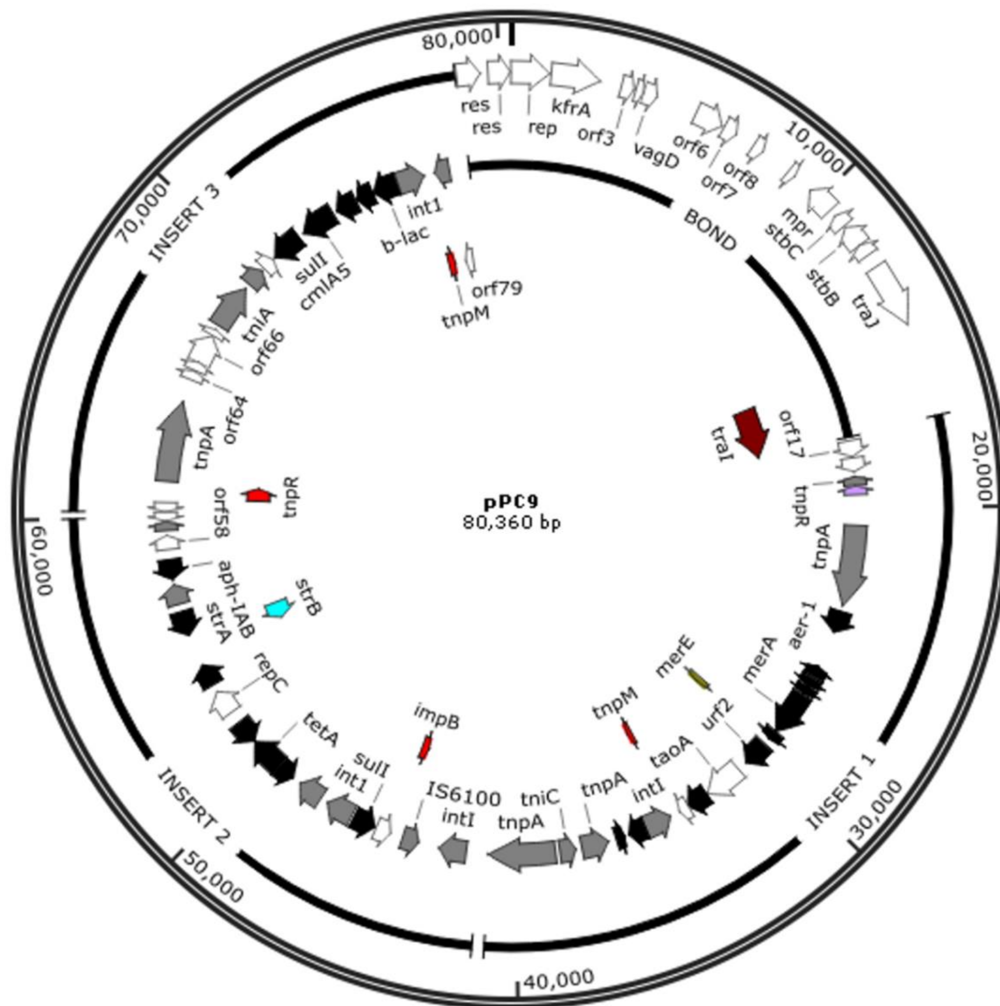


Figure 3. 1 The pPC9 map. Genetic organization of pPC9, in white, genes forming the backbone of pPC9, in grey genes from the insert with homology to genes related to transposition, in black genes from the insert with antibiotic resistance function.

The genome of HB3267 contains 5,261 ORFs of which 5,196 encode proteins. The other ORFs code for 61 tRNAs, 4 5S rRNAs, 4 16S rRNAs and 4 23S rRNAs. All essential conditional genes identified in KT2440 (Molina-Henares *et al*., 2010; Wu *et al*., 2010) are present in the genome of HB3267, and functions were assigned to almost 75 % of the total genes which encode proteins. Analysis of the sequence of

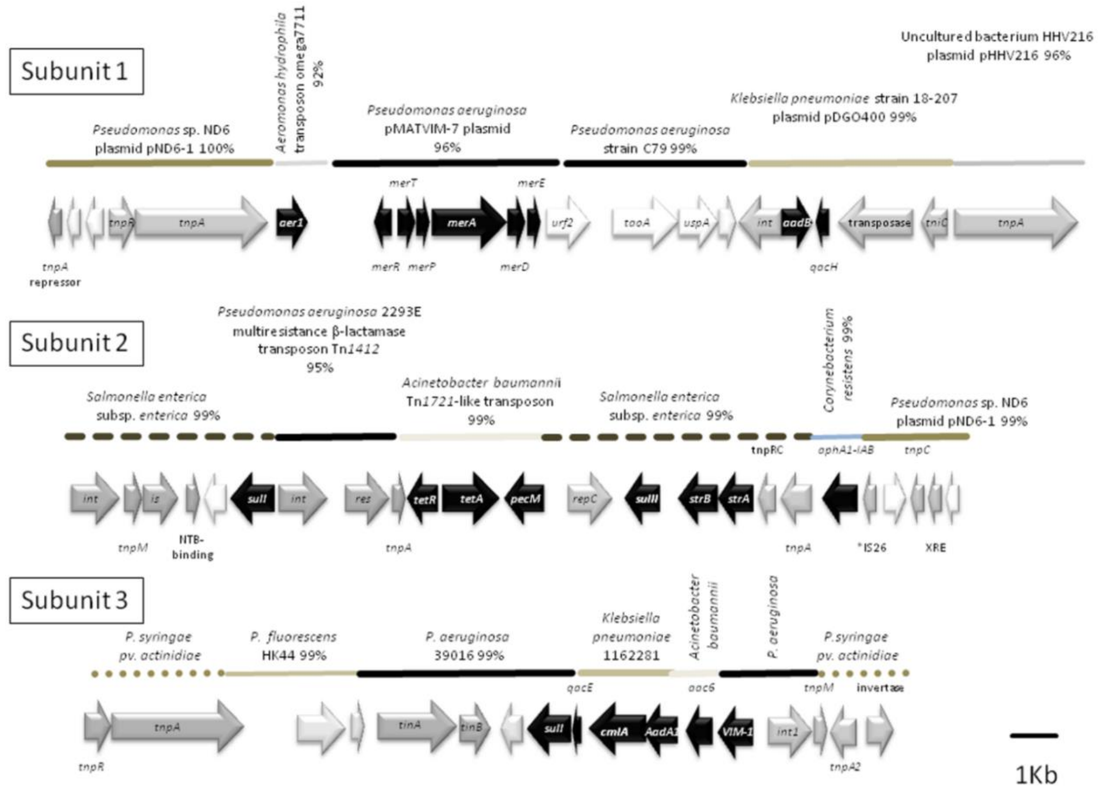


Figure 3. 2. Genetic organization of the pPC9 ‘‘insert.’’ Black arrows represent genes with functions related to antibiotic resistance. In grey are genes with functions related to transposition and insertion machinery. Non-colored genes are those that encode hypothetical proteins, those with unknown function, or those with functions unrelated to antibiotic resistance, transposition or integration. Horizontal lines over genes represent DNA homology to different microorganisms, and percentages indicate the degree of homology

the pPC9 plasmid revealed that this plasmid has a modular structure with 90 open reading frames (ORFs) that are distributed across several domains (Figure 3.1).

The pPC9 backbone (38 kb) primarily contains genes for plasmid-related functions such as those required for replication and partitioning (Bramucci *et al*., 2006). (Figure 3.1), and it exhibits high similarity to the backbone of pCT14 from *Pseudomonas* sp. CT14, a strain isolated from activated sewage sludge, while the ‘‘insert’’ region is made of several transposon-like elements bearing genes related to

transposition and resistance to multiple antibiotics (Figure 3.2). The backbone of pPC9 (Figure 3.1), bears only two putative conjugation genes (*traJ,I*), lacking most of the genes that would be necessary for self-mobilization. In agreement with this is that in biparental conjugation experiments this plasmid was unable to be transferred from HB3267 to other *P. putida* strains. Triparental mating experiments using the *Escherichia coli* pRK600 plasmid or the *P. putida* pWW0 plasmid as a helper also showed the inability of pPC9 to be transferred to *P. putida* strains. However, in experiments with pWW0, recombination events between pPC9 and pWW0 were detected; with transconjugants bearing a plasmid that conferred the ability to grow on toluene from pWW0, and streptomycin and tetracycline resistance from pPC9. This event occurred at a rate of 10^{-4} transconjugants per donor. This result confirms the presumption made by Yomoda *et al*. (Yomoda *et al*., 2003) that the existence of *P. putida* resident species in in-patients can facilitate the spread of drug resistance genes via horizontal gene transfer. The ‘‘insert’’ region of the pPC9 plasmid is a mosaic of recruited DNA from different microorganisms. The antibiotic resistant determinants of pPC9 are grouped within three subregions as illustrated in Figure 3.2, surrounded by genes related to transposition and/or integration. A potential source of this DNA were closely related *Pseudomonas* strains such as the nonhuman pathogen *Pseudomonas sp*. ND6, isolated from industrial wastewater (Zhao *et al*., 2005); *Pseudomonas syringae* pv. *actinidiae*, which is the causal agent of bacterial canker of green-fleshed kiwifruit (Marcelletti *et al*., 2011); and *P. fluorescens* HK44, which colonizes plant roots and degrades phenolic compounds (Kamath *et al*., 2004). Within this ‘‘insert’’ region, DNA fragments could have also been acquired from human opportunistic pathogens or pathogens, such as *P. aeruginosa*, *Aeromonas hydrophila*, *Acinetobacter baumannii*, *Corynebacterium resistens* and Enterobacteriaceae (*Klebsiella pneumoniae*, *Salmonella enterica*) (Figure 3.2). This mosaic suggests that HB3267 may have been a resident of different environments, such as water, soil, plant rhizospheres and the human body, or that it has exchanged DNA with these potential microorganisms in different habitats.

Antibiotic resistance determinants encoded only on the **HB3267** *chromosome*

Quinolones and fluoroquinolones. Quinolones and fluoroquinolones are chemotherapeutic bactericidal drugs that eradicate bacteria by interfering with DNA

replication. Quinolones enter into cells through porins and their targets are DNA gyrases and topoisomerases (Hernández *et al.*, 2011). Point mutations in *gyrA* and *gyrB* (DNA gyrase subunits), and *parC* and *parE* (topoisomerase IV subunits) are associated with resistance to fluoroquinolone antibiotics (Eaves *et al.*, 2004). In the HB3267 strain, GyrA showed a series of amino acid replacements with respect to the GyrA protein of fluoroquinolone-sensitive strains; HB3267 shows a Thr83Iso mutation in GyrA, which is known to lead to fluoroquinolone resistance in other clinic isolates of *P. putida* as HU2001-412 strain (Horii *et al.*, 2005), that is present also in *P. aeruginosa* LESB58; only a unique difference in the case of HB3267 was a Val to Gly change at residue 894 (Suppl. Figure 3.2A). In the case of GyrB, a polymorphism unique to HB3267 and the resistant HU2001-412 strain was Glu468Asp, also known to lead fluoroguinolone resistance (Horii *et al.*, 2005) (Suppl. Figure 3.2B). For ParC, the substitution Ser87Leu was present in HB3267 and *P. aeruginosa* PA7; however, for the other topoisomerase IV subunit ParE no substitution was found. The unique polymorphism in GyrA and/ or GyrB and/or ParC may be responsible for the resistance of HB3267 against these antibiotics, but this hypothesis requires further testing using site specific mutants and gene complementation experiments.

Cationic antimicrobial peptides. Strain HB3267 is more resistant to polymyxin B than KT2440R (Tables 3.1 and 3.2) and other strains (McPhee *et al.*, 2003). LPS is one of the surface elements that influence resistance to cationic peptides in *P. aeruginosa* (Pescaretti *et al.*, 2011). Vaara (1993) described that the *firA* gene product, involved in lipid A biosynthesis was relevant in the resistance of *E. coli* and *S. typhimurium* to polymyxin B (Vaara, 1993). Strain HB3267 harbours two copies of this gene which encodes for a UDP-3-O-3-hydroxymyristoyl glucosamine N-acyltransferase that is 70 % identical to the FirA of *Klebsiella pneumoniae*. The higher copy number of *firA* may provide an explanation for the HB3267 strain's higher resistance to polymyxin B than in other strains of *P. putida*. Antibiotic resistance determinants encoded on the chromosome of HB3267 and its resident pPC9 plasmid.

Aminoglycosides. Aminoglycosides bind to the A-site of the 30S subunit of bacterial ribosomes disturbing elongation of the peptide chain. Wei *et al*. (2011) demonstrated using data obtained from phenomics, transcriptomics and proteomic

analysis that resistance to aminoglycosides in the *P. aeruginosa* PAO1 strain is multifactorial including the presence of mutations in chromosomal genes such as the *phoP* and *phoQ*, as well as in the *mexZ* gene encoding a repressor of the mexXY genes encoding an efflux pump (Wei *et al*., 2011). When all the sequences of *phoQ* from aminoglycoside sensitive *Pseudomonas* strains were compared to the resistant HB3267 strain, the Gly365Arg polymorphism was exclusive to HB3267 (Suppl. Figure 3.2C). In the case of PhoP, Ser21Gly (exception *P. putida* S16 and *P. entomophila* L48), and Arg204His (exception *P. putida* S16) changes were found but none were exclusive of HB3267. No clear homolog to *mexZ* was present in the chromosome of HB3267.

HB3267 sensitivity and KT2440 resistance to amikacin may be explained by multifactorial differences in expression of chromosomal genes involved in cell permeability, LPS synthesis, efflux pumps and chemical modification (Struble and Gill, 2006). Vaziri *et al*. (2011) described the existence of aminoglycoside modifying enzymes that are encoded by plasmids as the primary resistance mechanism employed by *P. aeruginosa* against these antibiotics (Vaziri *et al*., 2011). The pPC9 plasmid carries genes that encode 6 aminoglycoside modifying enzymes that are not present in the genome of *P putida* strains sensitive to aminoglycosides. Therefore, these aminoglycosidases likely contribute to the resistance phenotype of HB3267. One of these aminoglycosidases was 100 % identical to the *aadB* gene product of *P. aeruginosa* (Figure 3.2 subunit 1). This protein has 2'-aminoglycoside nucleotidyltransferase activity, and has been proposed to be responsible for bacterial resistance to the aminoglycosides dibekacin, gentamicin, kanamycin, sisomicin and tobramycin (Schmidt *et al*., 1988). Another protein was 100 % identical to StrA from *Salmonella enterica* that has aminoglycoside 3'-phospho-transferase activity and a third protein was 100 % identical to StrB from *Acinetobacter baumannii*, which has aminoglycoside 6'-phosphotransferase activity (Figure 3.2 subunit 2). Both of these proteins have been traditionally associated with resistance to the aminoglycoside streptomycin (Han *et al*., 2004). We also found a gene that encodes for a protein that is 100 % identical to AphA1-IAB (Figure 3.2 subunit 2) from *Corynebacterium striatum*, which is an aminoglycoside 3'- phosphotransferase involved in the inactivation of aminoglycoside antibiotics such as kanamycin, neomycin, neamine, and

ribostamycin (Hainrichson *et al*., 2007). Another gene coding for a protein with 99 % identity to the *aadA1* gene product from *Escherichia coli* (Figure 3.2 subunit 3) is also present. AadA1 is an aminoglycoside-3'-adenylyltransferase that confers resistance to streptomycin and spectinomycin (Dahshan *et al*., 2011). Finally, there is also a gene that codes a protein which is 98 % identical to Aac6 from *Enterobacter cloacae*, an aminoglycoside phosphotransferase that confers resistance to netilmicin and tobramycin (Santos *et al*., 2010). It therefore appears that the pPC9 plasmid has recruited a number of aminoglycoside modifying enzymes from different origins.

Tetracyclines. Tetracyclines bind to the 30S subunit of microbial ribosomes. They inhibit protein synthesis by blocking the attachment of charged aminoacyl-tRNAs to the A site on the ribosome. Thus, they prevent introduction of new amino acids to the nascent peptide chain (Goldman *et al*., 1983).

Several mechanisms have been described by which bacteria gain resistance to tetracycline, namely, extrusion of tetracycline via efflux pumps, changes in ribosome proteins so that tetracycline no longer binds, and chemical inactivation of tetracyclines. Tetracycline efflux is the most efficient mechanism of resistance to this antibiotic for Gram negative bacteria (Levy, 2002b). The resistance of *P. putida* KT2440 to tetracycline is linked to the RND TtgABC efflux pump (Godoy *et al*., 2010). The TtgABC pump is also responsible for resistance to a broad range of antibiotics such as β-lactams, nalidixic acid, and chloramphen icol (Rojas *et al*., 2001; Godoy *et al*., 2010). The genes which encode the TtgABC pump are located on the chromosome of HB3267.

A secondary role in tetracycline resistance was assigned to the TtgGHI efflux pump, which is located on the pGRT1 plasmid in *P. putida* DOT-T1E (Rodríguez-Herva *et al*., 2007), although the primary role of this pump in *P. putida* DOT-T1E appears to be solvent extrusion (Rojas *et al*., 2001; Rodríguez-Herva *et al*., 2007; Molina *et al*., 2011). The *ttgGHI* genes are present in the chromosome of HB3267 and the operon is located in a 44 Kb genomic island (Figure 3.3) with no homology to the chromosomal sequences of other *P. putida* strains or with the rest of pGRT1 plasmid (Molina *et al*., 2011). The total G+C content of this island is 55 %, a value

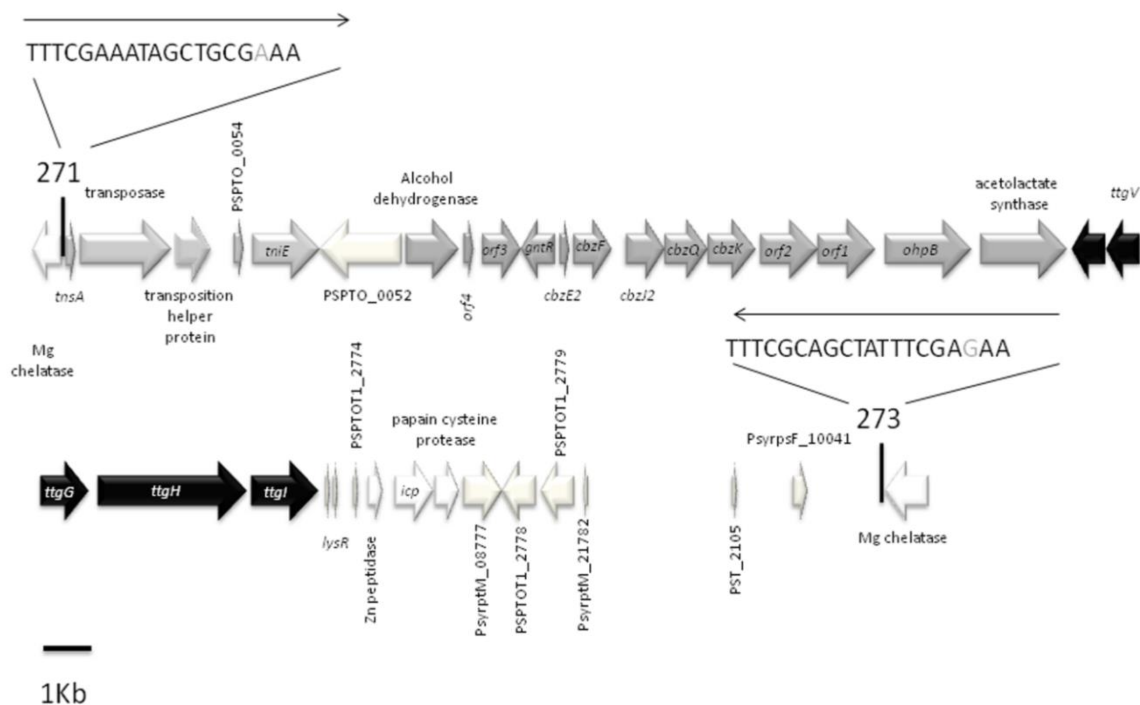


Figure 3. 3. Location of the antibiotic and solvent efflux ttgGHI pumps within a genomic island in the chromosome of strain HB3267. The *ttgGHI* efflux genes are indicated in black; genes involved in transposition events are in light grey; the *cbz* operon, which is involved in chlorobenzene degradation, is in medium grey. Vertical lines indicate the insertion point; arrows above the sequence indicate the inverted repeat sequences of the Tn552-like

lower than that of the rest of chromosome. The gene that encodes the repressor of

this *ttgGHI* operon, the TtgV protein is 89 % identical to that of DOT-T1E.

It should be noted that the pPC9 plasmid also encodes another tetracycline efflux pump of the TetA type, which is 100 % identical to the *Acinetobacter baumannii* Tn1721-like transposon (Figure 3.2 subunit 2) (McMurry *et al.*, 1980). This gene is not present in the genome of KT2440R; the presence of multiple tetracycline efflux pumps supports the data which shows that HB3267 exhibits higher levels of resistance to tetracycline than other strains.

β-lactams. RND efflux pumps and β-lactamases are key players in the resistance of Pseudomonads to β-lactam antibiotics. The HB3267 chromosome carries 10 β-lactamase genes that are also present in KT2440, explaining the resistance of both strains to penicillin-derived antibiotics such as ampicillin, carbenicillin, ticarcillin and piperacillin, and the cephalosporin ceftriaxone. Livermore suggested that the

phenotype of resistance to the cephalosporin cefotaxime and ceftazidime of some *Pseudomonas* clinical isolates was mediated by the action of the chromosomal *ampC* gene that codes for a β-lactamase (Livermore, 1995). This gene is present in KT2440, which is sensitive to these antibiotics. Two single nucleotide polymorphisms were found in the *ampC* gene of HB3267 when compared to the sensitive KT2440 strain, namely Pro148Ala and Gly263Arg, which may explain the HB3267 resistance to these antibiotics. We found that the pPC9 plasmid bears 2 additional β-lactamase genes, one that codes for an enzyme that is 95 % identical to AER-1 from *Aeromonas hydrophila* (Figure 3.2 subunit 1) a protein that can efficiently hydrolyze carbenicillin (Hedges *et al*., 1985); as well as a gene that codes for a protein that is 100 % identical to VIM-1 from *Klebsiella pneumonia* (Figure 3.2 subunit 3), which is involved in the carbapenem-resistant phenotype of that microorganism (Steinmann *et al*., 2011).

Chloramphenicol. Chloramphenicol is a bacteriostatic antimicrobial that functions by inhibiting bacterial protein synthesis. *P. putida* KT2440 is a chloramphenicol-resistant bacterium that is able to grow in the presence of this antibiotic at a concentration of up to 25 mg/ml. Genomic analysis revealed that the TtgABC efflux pump and biosynthesis of pyrroloquinoline (PQQ) were involved in chloramphenicol resistance (Fernandez *et al*., 2012). These genes are present in the chromosome of the HB3267 strain, which also shows high resistance to this antibiotic. An additional *pqqC* (coenzyme PQQ synthesis protein C) gene is present in the chromosome of the HB3267 strain, which is also present in the close relatives *P. putida* S16 and *Pseudomonas* sp. TJI-51, but not in KT2440 (Figure 3.4A). The AgmR regulator (PP_2665) controls the expression of the *pqq* genes and the operon encoding the ABC extrusion pump (Fernandez *et al*., 2012). Up to three polymorphisms were present in PqqC of HB3267 when compared to other *P. putida* strains, namely HisGln44 (Figure 3.4B), His142Leu and Ala116Gly. Per se, these mutations and the presence of an additional copy of the *pqqC* gene could explain in part the high resistance of HB3267 to chloramphenicol. In addition plasmid pPC9 encodes a protein that is 99 % identical to CmlA from *Aeromonas caviae* (Figure 3.1), an efflux pump that expels chloramphenicol and ethidium from the cells (Minek *et al*., 1998).

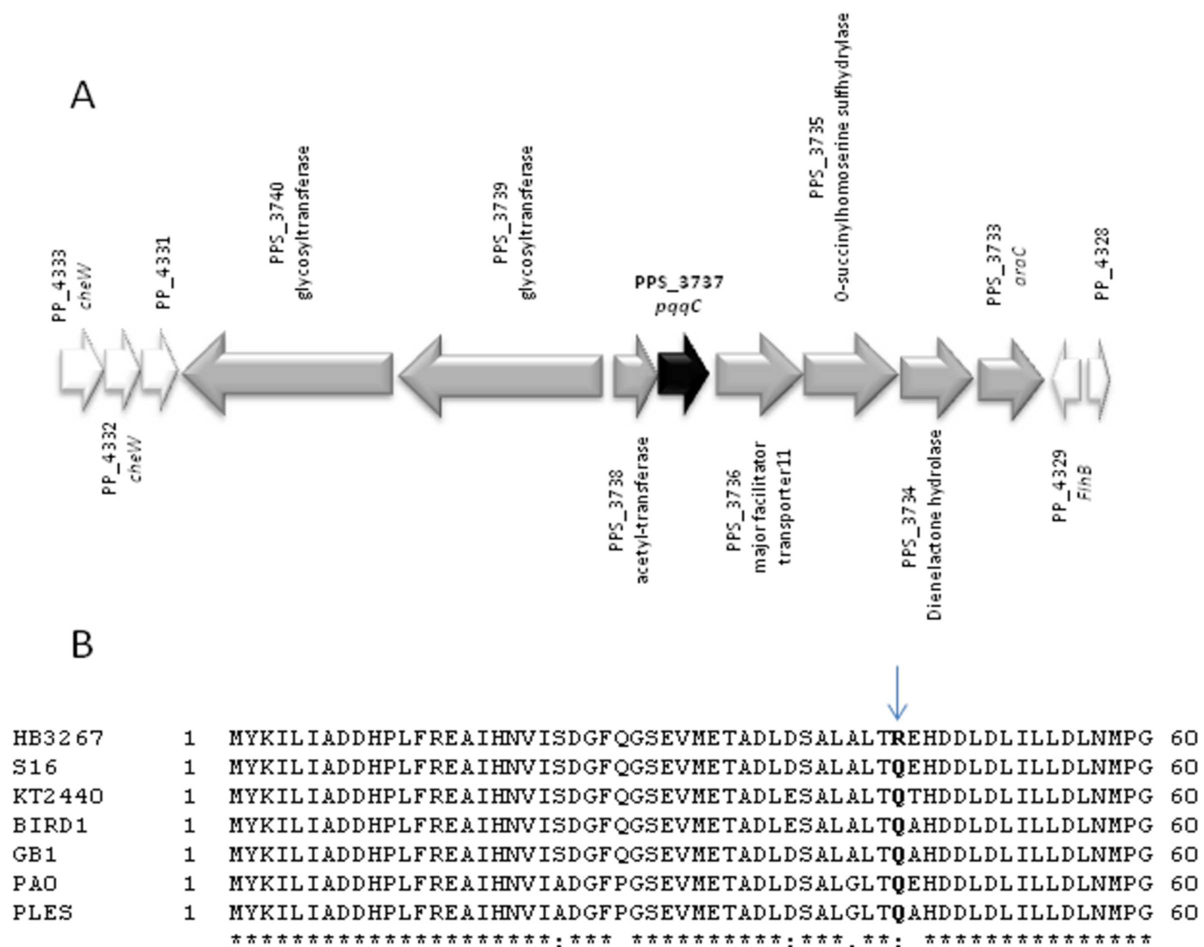


Figure 3. 4. Potential chromosomal determinants for chloramphenicol resistance of HB3267. (A) In black, chromosomal location of the additional pqqC gene of HB3267.; in grey, the region of the HB3267 chromosome that is not present in KT2440; in white, genes in synteny with KT2440. (B) Protein alignment of AgmR from *P. putida* HB3267 (HB3267, B479_11475), *P. putida* S16 (S16, PPS_2213), *P. putida* KT2440 (PPS_2213, PP_2665), *P. putida* BIRD-1 (BIRD1, PPUBIRD1_3011), *P. putida* GB-1 (GB1, PputGB1_3138), *P. aeruginosa* PA7 (PA7, PSPA7_3317), and *P. aeruginosa* PAO1 (PAO1, PA1978) strains. Amino acid mutations referred to in the text are indicated in bold.

Antibiotic resistance determinants encoded only on the pPC9 plasmid

Sulfonamides. Sulfonamides were the first compounds used as chemotherapy drugs and have been used as antibacterial agents since the 1930's. In bacteria, sulfonamides act as competitive inhibitors of the dihydropteroate synthetase (DHPS), an enzyme involved in folate synthesis. As such, the compounds cause

microorganisms to become ‘‘starved’’ of folate and die (Kent, 2000). Plasmid-mediated sulfonamide resistance in Gram-negative bacteria is very frequently found in clinical isolates and often in combination with other antibiotic resistance traits. The plasmids generally express alternative dihydropteroate synthases, the Sul proteins, which confer resistance to the drug (Rådström and Swedberg, 1988). In pPC9, three genes homologous to those encoding Sul proteins were found: one with 97 % identity to the SulI protein of *P. aeruginosa* (Figure 3.2. subunit 2), another with a 100 % identity with SulII of diverse enteric bacteria (Figure 3.2. subunit 2), and finally a gene coding a protein with 99 % identity to SulI of *E. coli* (Figure 3.2. subunit 3).

Our results show that strain *P. putida* HB3267, isolated from a deceased in-patient in a French hospital, is resistant to the majority of antibiotics and biocides used in laboratories and hospitals (aminoglycosides, ß-lactam antibiotics, cationic peptides dihydrofolate reductase inhibitors, fluoroquinolones, quinolones, glycopeptide antibiotics, macrolides, polyketide, and sulfonamides). This broad range of resistance is rarely found in clinical isolates.

Another relevant finding is that MICs for these antibiotics in planktonic cells were much higher for HB3267 than that of multidrug-resistant strains of *Pseudomonas aeruginosa*. Sequencing of the genome of HB3267 revealed that the determinants of multiple resistances are located chromosomally and on the plasmid pPC9. Regions of the plasmid bearing multidrug resistant genes show high homology with DNA from environmental microorganisms as well as from human opportunistic and true human pathogens indicating both contact and DNA exchange between the HB3267 strain and these environmental and clinically relevant microorganisms. The pPC9 plasmid carries integrons and transposons where the antibiotic resistant determinants are grouped.

We have shown that pPC9 is not self-transmissible but transfer of the antibiotic resistant genes from pPC9 to other microbes can be mediated by shuttle vectors, such as the TOL plasmid pWW0. The results presented in this work support the notion that the acquisition of new antibiotic and biocide resistant traits by opportunistic human pathogens, may arise from the cohabitation in the human body of pathogens with new multidrug-resistant ‘‘residents’’, such as the HB3267 strain.

Acknowledgments

We thank M. Mar Fandila for secretarial help and Ben Pakuts for critical reading of the manuscript.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Andrews JM (2001) Determination of minimum inhibitory concentrations. *J. Antimicrob Chemother* **48**(Suppl. 1): 5–16.

Aranda-Olmedo I, Tobes R, Manzanera M, Ramos JL, Marques S (2002) Species-specific repetitive extragenic palindromic (REP) sequences in *Pseudomonas putida. Nucleic Acids Res* **30**: 1826–1833.

Bagdasarian M, Lurz R, Rückert B, Franklin FC, Bagdasarian MM, *et al*. (1981) Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*. *Gene* **16:** 237–247.

Bielaszewska M, Mellmann A, Zhang W, Köck R, Fruth A, *et al*. (2011) Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect Dis* **11**: 671–676.

Bramucci M, Chen M, Nagarajan V (2006) Genetic organization of a plasmid from an industrial wastewater bioreactor. *Appl Microbiol Biotechnol* **71**: 67–74

Ceri H, Olson M, Morck D, Storey D, Read R, *et al*. (2001) The MBEC assay system; Multiple equivalent biofilms for antibiotic and biocide susceptibility testing. *Methods Enzymol* **337**: 377–385.

Dahshan H, Abd-El-Kader MA, Chuma T, Moriki H, Okamoto K (2011) Re-emergence of multidrug resistant *Salmonella enterica* serovar *Stanley* from cattle. *Vet Res Commun* **35**: 55–60.

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**: 4636–4641

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, *et al*. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**: W465–469

Docquier JD, Riccio ML, Mugnaioli C, Luzzaro F, Endimiani A, *et al*. (2003) IMP-12, a new plasmid-encoded metallo-beta-lactamase from a *Pseudomonas putida* clinical isolate. *Antimicrob Agents Chemother* **47**: 1522–1528.

Dogan NM, Kantar C, Gulcan S, Dodge CJ, Yilmaz BC, *et al*. (2011) Chromium(VI) bioremoval by *Pseudomonas* bacteria: role of microbial exudates for natural attenuation and biotreatment of Cr(VI) contamination. *Environ Sci Technol* **45**: 2278–2285.

Duque E, de la Torre J, Bernal P, Molina-Henares MA, Alaminos M, *et al*. (2012) Identication of reciprocal adhesión genes in pathogenic and nonpathogenic *Pseudomnonas*. *Environ Microbiol* and *EMI* R In press. doi: 10.1111/j.1462-2920.2012.02732.x

Eaves DJ, Randall L, Gray DT, Buckley A, Woodward MJ, *et al*. (2004) Prevalence of mutations within the quinolone resistance-determining region of *gyrA*, *gyrB*, *parC*, and *parE* and association with antibiotic resistance in quinolone-resistant *Salmonella enterica*. *Antimicrob Agents Chemother* **48**: 4012–4015.

Espinosa-Urgel M, Kolter R, Ramos JL (2002) Root colonization by *Pseudomonas putida*; love at first sight. Microbiology **148**: 341–343.

Fernández M, Conde S, de la Torre J, Molina-Santiago C, Ramos JL, *et al*. (2012) Mechanisms of resistance to chloramphenicol in *Pseudomonas putida* KT2440. *Antimicrob Agents Chemother* **56**: 1001–1009.

Fernández M, Duque E, Pizarro-Tobías P, Van Dillewijn P, Wittich RM, *et al*. (2009) Microbial responses to xenobiotic compounds. Identification of genes that allow *Pseudomonas putida* KT2440 to cope with 2,4,6-trinitrotoluene. *Microb Biotechnol* **2**: 287–294.

Frapolli M, Défago G, Moënne-Loccoz Y (2007) Multilocus sequence analysis of biocontrol fluorescent *Pseudomonas* spp. producing the antifungal compound 2,4-diacetylphloroglucinol. *Environ Microbiol* **9**: 1939–1955.

Goldman RA, Hasan T, Hall CC, Strycharz WA, Cooperman BS (1983) Photoincorporation of tetracycline into *Escherichia coli ribosomes*. Identification of the major proteins photolabeled by native tetracycline and tetracycline photoproducts and implications for the inhibitory action of tetracycline on protein synthesis. *Biochemistry* **22:** 359–368.

Hainrichson M, Yaniv O, Cherniavsky M, Nudelman I, Shallom-Shezifi D, *et al*. (2007) Overexpression and initial characterization of the chromosomal aminoglycoside 3′-O-phosphotransferase APH(3′)-IIb from *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* **51**: 774–776.

Han HS, Koh YJ, Hur JS, Jung JS (2004) Occurrence of the *strA-strB* streptomycin resistance genes in *Pseudomonas* species isolated from kiwifruit plants. *J Microbiol* **42**: 365–368.

Hedges RW, Medeiros AA, Cohenford M, Jacoby GA (1985) Genetic and biochemical properties of AER-1, a novel carbenicillin-hydrolyzing beta-lactamase from *Aeromonas hydrophila*. *Antimicrob Agents Chemother* 1985 **27**: 479–484.

Hernández A, Sánchez MB, Martínez JL (2011) Quinolone Resistance: Much More than Predicted. *Front Microbiol* **2**: 22.

Horii T, Muramatsu H, Iinuma Y (2005) Mechanisms of resistance to fluoroquinolones and carbapenems in *Pseudomonas putida. J Antimicrob Chemother* **56**: 643–647.

Iyer PG, Murphy TF (2009) Chronic obstructive pulmonary disease: role of bacteria and updated guide to antibacterial selection in the older patient. *Drugs Aging* **26**: 985–995.

Janda JM, Abbott SL (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* **45**: 2761–2764.

Kamath R, Schnoor JL, Alvarez PJ (2004) Effect of root-derived substrates on the expression of nah-lux genes in *Pseudomonas fluorescens* HK44: implications for PAH biodegradation in the rhizosphere. *Environ Sci Technol* **38**: 1740–1745.

Kent M (2000) Advanced Biology, *Oxford University Press*, p. 46 ISBN 9780199141951.

Khan NH, Ahsan M, Yoshizawa S, Hosoya S, Yokota A, *et al*. (2008) Multilocus sequence typing and phylogenetic analyses of *Pseudomonas aeruginosa* isolates from the ocean. *Appl Environ Microbiol* **74**: 6194–6205

Kumita W, Saito R, Sato K, Ode T, Moriya K, *et al*. (2009) Molecular characterizations of carbapenem and ciprofloxacin resistance in clinical isolates of *Pseudomonas putida*. *J Infect Chemother* **15**: 6–12.

Levy SB (2002 a) The 2000 Garrod lecture. Factors impacting on the problem of antibiotic resistance. *J Antimicrob Chemother* **49**: 25–30.

Levy SB (2002 b) Active efflux, a common mechanism for biocide and antibiotic resistance. *J Appl Microbiol* **92:** 65S–71S.

Livermore DM (1995) beta-Lactamases in laboratory and clinical resistance. *Clin Microbiol Rev* **8**: 557–584.

Loiseau-Marolleau ML, Malarre N (1977) *Pseudomonas putida*: identification, antibiotic sensitivity and pathogenicity. *Pathol Biol* (Paris) **25:** 637–645.

Marcelletti S, Ferrante P, Petriccione M, Firrao G, Scortichini M (2011) *Pseudomonas syringae* pv. *actinidiae* draft genomes comparison reveal strain-specific features involved in adaptation and virulence to *Actinidia* species. *PLoS One* **6**: e27297.

McMurry LM, Petrucci R, Levy SB (1980) Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in *Escherichia coli*. *Proc Natl Acad Sci* (USA) **77**: 3974–3977.

McPhee JB, Lewenza S, Hancock REW (2003) Cationic antimicrobial peptides activate a two component regulatory system, PmrA-PmrB, that regulates resistance to polymyxin B and cationic antimicrobial peptides in *Pseudomonas aeruginosa. Mol Microbiol* **50**: 205–217

Mendes RE, Castanheira M, Toleman MA, Sader HS, Jones RN, *et al*. (2007) Characterization of an integron carrying blaIMP-1 and a new aminoglycoside resistance gene*, aac*(6′)-31, and its dissemination among genetically unrelated clinical isolates in a Brazilian hospital. *Antimicrob Agents Chemother* **51**: 2611–2614

Minek Tl, Morita Y, Kataoka A, Mizushima T, Tsuchiya T (1998) Evidence for chloramphenicol/H+ antiport in Cmr (MdfA) system of *Escherichia coli* and properties of the antiporter. *J Biochem* **124**: 187–193.

Molina L, Duque E, Gómez MJ, Krell T, Lacal J, *et al*. (2011) The pGRT1 plasmid of *Pseudomonas putida* DOT-T1E encodes functions relevant for survival under harsh conditions in the environment. *Environ Microbiol* **13**: 2315–2327.

Molina-Henares MA, de la Torre J, García-Salamanca A, Molina-Henares AJ, Herrera MC, *et al*. (2010) Identification of conditionally essential genes for growth of *Pseudomonas putida* KT2440 on minimal medium through the screening of a genome-wide mutant library. *Environ Microbiol* **12**: 1468–1485.

Moskowitz SM, Brannon MK, Dasgupta N, Pier M, Sgambati N, *et al*. (2012) PmrB mutations promote polymyxin resistance of *Pseudomonas aeruginosa* isolated from colistin-treated cystic fibrosis patients. *Antimicrob Agents Chemother* **56**: 1019–1030.

Muller C, Plésiat P, Jeannot K (2011) A two-component regulatory system interconnects resistance to polymyxins, aminoglycosides, fluoroquinolones, and β-lactams in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* **55**: 1211–1221

Pescaretti ML, López FE, Morero RD, Delgado MA (2011) The PmrA/PmrB regulatory system controls the expression of the *wzzfepE* gene involved in the O-antigen synthesis of *Salmonella enterica* serovar *Typhimurium. Microbiology* **157**: 2515–2521

Poirel L, Naas T, Nicolas D, Collet L, Bellais S, *et al*. (2000) Characterization of VIM-2, a carbapenem-hydrolyzing metallo-beta-lactamase and its plasmid- and integron-borne gene from a Pseudomonas *aeruginosa* clinical isolate in France. *Antimicrob Agents Chemother* **44**: 891–897.

Rådström P, Swedberg G (1988) RSF1010 and a conjugative plasmid contain sulII, one of two known genes for plasmid-borne sulfonamide resistance dihydropteroate synthase. *Antimicrob Agents Chemother* **32:** 1684–1692.

Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, *et al*. (2011) Origins of the *Escherichia coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* **365**: 709–717.

Rodríguez-Herva JJ, García V, Hurtado A, Segura A, Ramos JL (2007) The *ttgGH*I solvent efflux pump operon of *Pseudomonas putida* DOT-T1E is located on a large self-transmissible plasmid. *Environ Microbiol* **9**: 1550–1561.

Rojas A, Duque E, Mosqueda G, Golden G, Hurtado A, *et al*. (2001) Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. *J Bacteriol* **183**: 3967–3973.

Rolston KV, Kontoyiannis DP, Yadegarynia D, Raad II (2005) Nonfermentative gram-negative bacilli in cancer patients: increasing frequency of infection and antimicrobial susceptibility of clinical isolates to fluoroquinolones. *Diagn Microbiol Infect Dis* **51**: 215–218

Saha R, Jain S, Kaur IR (2010) Metallo beta-lactamase producing *Pseudomonas* species–a major cause of concern among hospital associated urinary tract infection. *J Indian Med Assoc* **108**: 344–348.

Santos C, Caetano T, Ferreira S, Mendo S (2010) Tn5090-like class 1 integron carrying bla(VIM-2) in a *Pseudomonas putida* strain from Portugal. *Clin Microbiol Infect* **16:** 1558–1561.

Schmidt FR, Nücken EJ, Henschke RB (1988) Nucleotide sequence analysis of 2″-aminoglycoside nucleotidyl-transferase ANT(2″) from Tn4000: its relationship with AAD(3″) and impact on Tn21 evolution. *Mol Microbiol* **2:** 709–717.

Steinmann J, Kaase M, Gatermann S, Popp W, Steinmann E, *et al*. (2011) Outbreak due to *a Klebsiella pneumoniae* strain harbouring KPC-2 and VIM-1 in a German university hospital, July 2010 to January. Euro Surveill **16** pii: 19944.

Struble JM, Gill RT (2006) Reverse engineering antibiotic sensitivity in a multidrug-resistant *Pseudomonas aeruginosa* isolate. *Antimicrob Agents Chemother* **50**: 2506–2515.

Tang H, Yu H, Li Q, Wang X, Gai Z, *et al*. (2011) Genome sequence of *Pseudomonas putida* strain B6-2, a superdegrader of polycyclic aromatic hydrocarbons and dioxin-like compounds. *J Bacteriol* **193:** 6789–6790.

Treviño M, Moldes L, Hernández M, Martínez-Lamas L, García-Riestra C, *et al*. (2010) Nosocomial infection by VIM-2 metallo-beta-lactamase-producing *Pseudomonas putida*. *J Med Microbiol* **59**: 853–855.

Vaara M (1993) Outer membrane permeability barrier to azithromycin, clarithromycin, and roxithromycin in gram-negative enteric bacteria. *Antimicrob Agents Chemother* **37**: 354–356

Vaziri F, Peerayeh SN, Nejad QB, Farhadian A (2011) The prevalence of aminoglycoside-modifying enzyme genes (*aac* (6′)-I, aac (6′)-II, *ant* (2″)-I*, aph* (3′)-VI) in *Pseudomonas aeruginosa*. *Clinics (Sao Paulo)* **66**: 1519–1522.

von Graevenitz A, Weinstein J (1971) Pathogenic significance of *Pseudomonas fluorescens* and *Pseudomonas putida*. *Yale J Biol Med* **44**: 265–273.

Wang SN, Liu Z, Tang HZ, Meng J, Xu P (2007) Characterization of environmentally friendly nicotine degradation by *Pseudomonas putida* biotype A strain S16. *Microbiology* **153**: 1556–1565.

Wei Q, Tarighi S, Dötsch A, Häussler S, Müsken M, *et al*. (2011) Phenotypic and genome-wide analysis of an antibiotic-resistant small colony variant (SCV) of *Pseudomonas aeruginosa*. *PLoS One* **6**: e29276.

Williams PA, Murray K (1974) Metabolism of benzoate and the methylbenzoates *by Pseudomonas putida* (arvilla) mt-2: evidence for the existence of a TOL plasmid. *J Bacteriol* **120**: 416–423.

Wu X, Monchy S, Taghavi S, Zhu W, Ramos J, *et al*. (2010) Comparative genomics and functional analysis of niche-specific adaptation in Pseudomonas putida. FEMS Microbiol Rev 35: 299–323. doi: 10.1111/j.1574-6976.2010.00249.x

Yomoda S, Okubo T, Takahashi A, Murakami M, Iyobe S (2003) Presence of *Pseudomonas putida* strains harboring plasmids bearing the metallo-beta-lactamase gene bla(IMP) in a hospital in Japan. J *Clin Microbiol* **41**: 4246–4251.

Yoshino Y, Kitazawa T, Kamimura M, Tatsuno K, Ota Y, *et al*. (2011*) Pseudomonas putida* bacteremia in adult patients: five case reports and a review of the literature. *J Infect Chemother* **17**: 278–282.

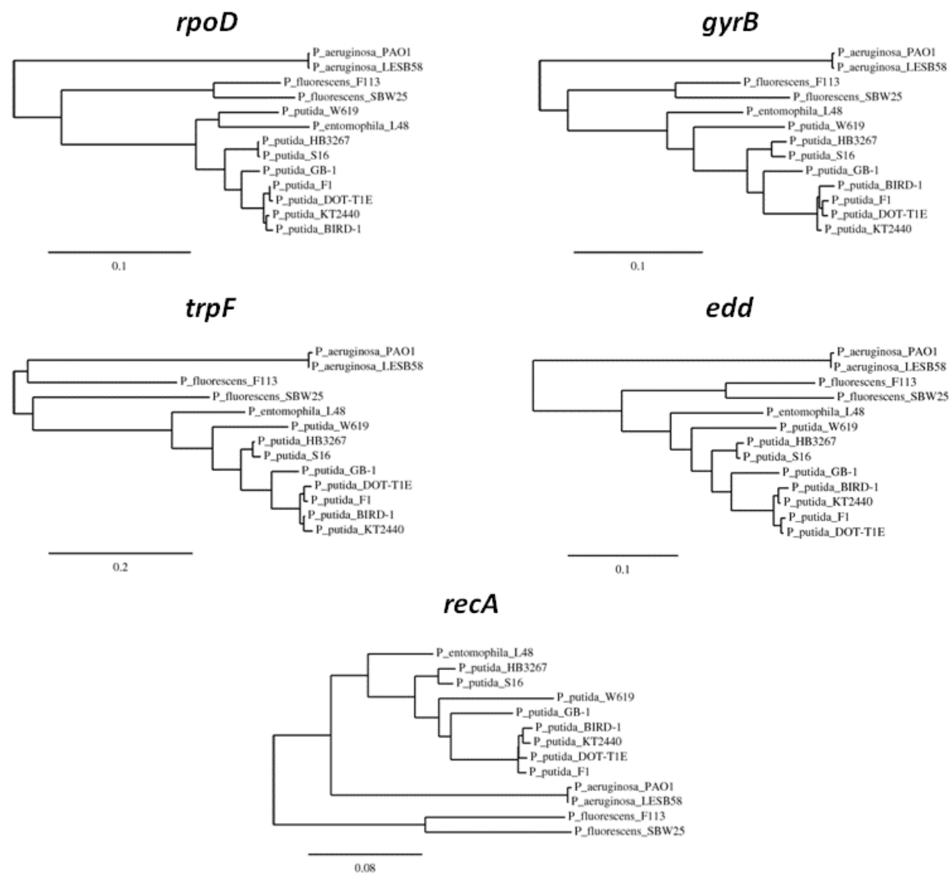
Zhao H, Chen D, Li Y, Cai B (2005) Overexpression, purification and characterization of a new salicylate hydroxylase from naphthalene-degrading *Pseudomonas* sp. strain ND6. *Microbiol Res* **160**: 307–313.

Supporting information

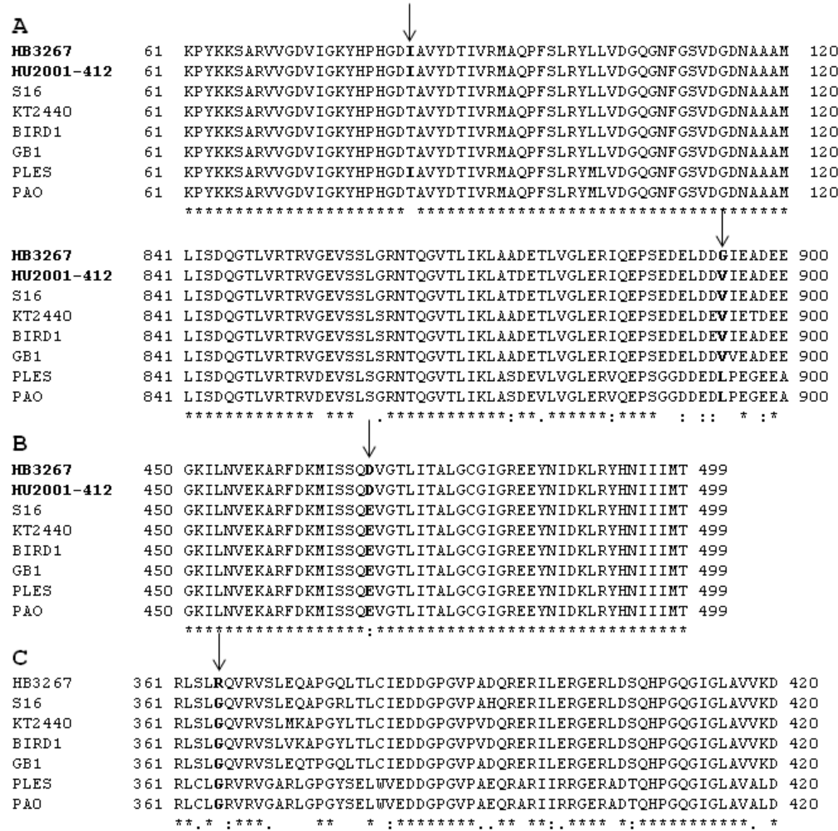
Suppl. Table. 3. 1. RND efflux pumps in *P. putida* KT2440 and their orthologs in the HB3267 strain.

***P. putida KT2440* gene id**	**Group**	***P. putida HB3267* gene id**	**P. putida HB3267 protein**	**e-val**
PP_3426	1	Ppu_PC9_4111	Efflux pump membrane transporter *mexE*	0.0
PP_1385		Ppu_PC9_3653	Efflux memebrane transporter ttgB	0.0
PP_2818		Ppu_PC9_1729	Multidrug resistance protein MexB	0.0
PP_3456		Ppu_PC9_1218	Multidrug resistance protein	0.0
PP_2065	2	Ppu_PC9_2821	Uncharacterized transporter	0.0
PP_0906		Ppu_PC9_0548	Uncharacterized transporter	0.0
PP_3583		Ppu_PC9_3815	Multidrug resistance protein	0.0
PP_3584		Ppu_PC9_3816	Multidrug resistance protein	0.0
PP_5387	3	Ppu_PC9_1983	Cation efflux system protein	0.0
PP_2410		Ppu_PC9_0624	Cation efflux system protein	0.0
PP_0043		Ppu_PC9_2009	Cation efflux system protein	0.0
PP_1517	4	Ppu_PC9_3070	RND family effluz transporter	0.0
PP_5173		Ppu_PC9_2223	Acriflavin resistance protein	0.0

Group 1: extrusion of antibiotics. Group 2: extrusion of organic compounds that generate oxidative stress. Group 3: Metal resistance efflux pumps. Group 4: Extrusion of antibiotics and metal resistance**.** E-val (e-value)



Suppl. Figure 3. 1. Phylogenetic tree comparing *gyrB* genes of *Pseudomonas strains*. Phylogram constructed using the platform Phylogeny.fr. which is a combination of a predefined pipeline using leading programs that include MUSCLE, Gblocks, PhyML and TreeDyn ***P. aeruginosa*** PAO-1 (NC_018080), ***P. fluorescens*** F113 (NC_016830), ***P. monteilii*** BCRC 17520 (FJ418641), ***P. putida*** BIRD-1 (NC_017530), ***P putida*** GB-1 (NC_010322), ***P. putida*** KT2440 (NC_002947), ***P. putida*** HB3267 (CP003738), ***P. putida*** S16 (NC_015733).



Suppl. Figure 3. 2 Protein alignment of GyrA (A), GyrB (B) and PhoQ (C) from *P. putida* HB3267. (HB3267, B479_00265, B479_06830, B479_20445, respectively), *P. putida* S16 (S16, PPS_1408, PPS_0012, PPS_4028), *P. putida* KT2440 (KT2440, PP_1767, PP_0013, PP_1187*),* *P. putida* BIRD-1 (BIRD1, PPU_BIRD1_3846,…., PPU_BIRD1_1228), *P. putida* GB-1 (GB1, PputGB1_1358, PputGB1_0006, PputGB1_4229) strains and *P. aeruginosa* LESB58 (LESB PLES_19001, PLES_00031, PLES_41411), *P. aeruginosa* PAO1 (PAO1, PA3168, PA0004, PA1180) strains. Amino acid changes referred to in the text are indicated in bold; "*****" Identical residues, "**:**" conservative substitutions and "**.**" semiconservative substitutions.

Chapter IV

Analysis of the core genome and pangenome of *Pseudomonas putida*

This chapter was published as

Zulema Udaondo, Lázaro Molina, Ana Segura, Estrella Duque, Juan L. Ramos (2015) Analysis of the core genome and pangenome of *Pseudomonas putida* *Environ Microbiol* DOI: 10.1111/1462-2920.13015

Summary

Pseudomonas putida are strict aerobes that proliferate in a range of temperate niches and are of interest for environmental applications due to their capacity to degrade pollutants and ability to promote plant growth. Furthermore solvent tolerant strains are useful for biosynthesis of added-value chemicals. We present a comprehensive comparative analysis of nine strains and the first characterization of the *Pseudomonas putida* pangenome. The core genome of *P. putida* comprises approximately 3386 genes. The most abundant genes within the core genome are those that encode nutrient transporters. Other conserved genes include those for central carbon metabolism through the Entner-Doudoroff pathway, the pentose phosphate cycle, arginine and proline metabolism, and pathways for degradation of aromatic chemicals. Genes that encode transporters, enzymes and regulators for amino acid metabolism (synthesis and degradation) are all part of the core genome, as well as various electron transporters, which enable aerobic metabolism under different oxygen regimes. Within the core genome are thirty genes for flagella biosynthesis and twelve key genes for biofilm formation. *Pseudomonas putida* strains share 85 % of the coding regions with *Pseudomonas aeruginosa*; however, in *P. putida*, virulence factors such as exotoxins and type III secretion systems are absent.

Introduction

Pseudomonas putida are members of the Pseudomonadales order of the Gammaproteobacteria class and are chemo-organotrophic aerobic, gram-negative rods with polar flagella that use respiratory rather than fermentative metabolism (Palleroni, 1984). These microorganisms are widely distributed in environmental niches in all continents, although they are particularly abundant in temperate soils and waters. They have often been found in polluted soils due to their metabolic versatility (Timmis, 2002). A number of *P. putida* strains grow favourably in plant roots and are used as plant growth promoting rhizobacteria (PGPR) (Matilla *et al*., 2011; Roca *et al*., 2013; Pizarro-Tobías *et al*., 2014). The PGPR properties of *P. putida* derive from a number of characteristics, including: the ability to solubilize phosphate; the inhibition of fungal growth through the production of iron-chelators such as siderophores; and the ability to promote root elongation via secretion of indole-acetic acid (IAA) and other phytohormones. Genes involved in the catabolism of a wide variety of organic compounds, including many naturally-occurring products (eg, metabolites from the partial degradation of lignin), are chromosomally encoded in the *P. putida* genome. As stated above, some strains of this species colonize the rhizosphere of plants (Molina *et al*., 2000; Weyens *et al*., 2010; Li *et al*., 2013; Roca *et al*., 2013). This characteristic, combined with the ability to degrade an array of compounds, provides these strains with a strong potential for bioremediation of contaminated soils in a process known as rhizophytoremediation (van Dillewijn *et al*., 2009; Segura and Ramos, 2013). Because *Pseudomonas putida* is easy to culture in laboratory settings (with a duplication time of around one hour in various culture media) and because there are a wide array of genetic tools available for its manipulation, this microbe has become a 'workhorse' and model organism in soil bacteria research (Timmis, 2002). The ability of *Pseudomonas putida* to thrive in different niches is the result of its genetic arsenal (genome size around of 6Mb), which contains highly sophisticated genes and pathways that mediate adaptation to changes in the local environment. Understanding the physiological and genetic basis of bacterial responses to diverse energy sources or stressors is crucial for the development of applications such as bioremediation, the control of plant diseases, and the promotion of plant growth (Reineke *et al*., 1982; Sashidhar and Podile, 2009). *Pseudomonas putida* KT2440 is

the best characterized saprophytic Pseudomonad with the ability to survive and function in the environment, despite many years of cultivation in different laboratories (Bagdasarian *et al*., 1981; Weinel *et al*., 2002; Santos *et al*., 2004; Frank *et al*., 2011). The first physical and genetic map of the strain was published in 1998 (Ramos-Díaz and Ramos, 1998) and its genome sequence was published by Nelson *et al.* (2002). Also, the strain was used as a model system for horizontal and vertical evolution of catabolic pathways in the degradation of alkylaromatic compounds (Ramos *et al*., 1986, 1987; Timmis, 2002; Li *et al*., 2013). The genetic versatility of this bacterium has been proven to be large, making it suitable for metabolic engineering to create superior strains for a variety of biotechnological applications, including the design of new catabolic pathways for pollutants, the biocatalysis of pharmaceutical intermediates and quality improvement of fossil fuels (Calzada *et al*., 2009; Prakash *et al*., 2010; Tao *et al*., 2011; Fernández *et al*., 2012). For *P. putida* there are twelve complete and eighteen in-progress genomes publicly available on NCBI (http://www.ncbi.nlm.nih.gov/genome/genomes/174, as of June, 2014). In the present study, we carry out a comparative analysis of the complete genomes of nine strains isolated from different environments that were carefully selected to represent a broadest diversity of strains of this species. These data open the way to describe the pangenomes of *P. putida* at the species level.

Materials and Methods

Strain genome sequences. The genomes of *P. putida* strains BIRD1, DOT-T1E and HB3267 were determined previously by our group (Roca *et al*., 2013; Udaondo *et al*., 2013; Molina *et al*., 2014). When this project started, the genomes of other strains (*P. putida* F1 (Phoenix *et al*., 2003), *P. putida* GB1 (Okazaki *et al*., 1997), *P. putida* KT2440 (Nelson *et al*., 2002), *P. putida* S16 (Yu *et al*., 2011) and *P. putida* W619 (Taghavi *et al*., 2005)) were available at ftp://ftp.ncbi.nih.gov/genomes/Bacteria/ and used in the current study. The draft genome sequence of *Pseudomonas putida* Idaho (Tao *et al*., 2011) was downloaded from NCBI and annotated by our group using the RAST annotation server (Aziz *et al*., 2008).

Compositional variations and gene mapping. Open reading frames of the nine genomes were used to calculate the GC content and CAI (codon adaptation index) of each set of genes in the pangenome (core genes, accessory genes and unique genes). The total number of genes in the pangenome was used to prepare the codon usage table. Values were established with GEECEE (fractional GC content calculator), CAI (codon adaptation index calculator) and CUSP (condon usage table calculator) programs from the EMBOSS application suite (Rice *et al*., 2000). Correlation analyses were used to identify the relationship between the fractional GC content and CAI. ANOVA analysis was carried out in R, using the AOV function, followed by the TukeyHSD test for significance. Each gene was mapped on the chromosome with DNAPlotter (Carver *et al*., 2009).

Phylogenetic analysis. Multilocus sequence phylogenetic analyses of the nine strains used for the pangenomic analysis plus five more other *P. putida* strains (DLL- E4, H8234, NBRC-14164, ND6 and S12), with an outgroup reference organism, *Escherichia coli* (Blattner *et al*., 1997), was carried out using the nucleotide sequences of 16S rRNA, DNA gyrase B subunit *gyrB*, RNA polymerase sigma 70 factor *rpoD* and RNA polymerase beta 70 factor *rpoB* genes. The MUSCLE algorithm (Edgar, 2004) was used to align the concatenated gene sequences, then Gblocks program (Castresana, 2000) was used to assist with the removal of highly variable and ambiguously aligned positions. The best nucleotide substitution model for our data was chosen using the ModelGenerator program (Keane *et al*., 2006). Then, a maximum likelihood tree was built with the MEGA5.2 program (Tamura *et al*., 2011) using the Tamura-Nei (TN93) substitution model, which was gamma distributed with invariant sites (G+I parameters) and bootstrap values of each branch were calculated 500 times.

The core genes of the 9 strains were aligned using progressiveMauve (Mauve v2.3.1) (Darling *et al*., 2010), then a Perl script was used to convert the extended multifasta (xmfa) output of Mauve to a fasta format. The Gblocks v0.91b program (Castresana, 2000) was used to select the conserved block from the multiple alignment. The best evolution model that fits with the data was selected using jModelTest2 v1.4 program (Darriba *et al*., 2012). The resulting aligment without gaps containing 3,057,018 bp was used to construct a phylogenetic tree by

maximum likelihood using RAxML v8 program (Stamatakis, 2014) with the GTR+G+I model and 500 bootstrap replicates.

Construction of ortholog groups (OGs). The proteomes of the nine strains used for pangenomics were clustered into OGs using a BBH approach using InParanoid 4.1 (Remm *et al*., 2001) and MultiParanoid (Alexeyenko *et al*., 2006). The BLOSUM 80 matrix was used during the InParanoid run. As the nine strains used in this study are evolutionarily closely related, MultiParanoid was reliably used to merge multiple pairwise orthologs from InParanoid into multi-strain ortholog groups. Genes were defined as unique if they did not have a bidirectional best hit in any of the eight other genomes.

Core-genome extrapolation. The core genome was estimated based on the method described by Tettelin *et al.* (Tettelin *et al*., 2005). The number of shared genes found on sequential addition of each new genome sequence was extrapolated by fitting an exponential decaying function to the amount of core genome genes.

KEGG ontology (KO) and Gene ontology (GO) annotation and analysis. For functional annotation, data mining and statistical analysis of core genome sequences, we used the KAAS automatic genome annotation server (Moriya *et al*., 2007) to assign KO terms. GO categories were inferred for the core proteome through the alignment of protein sequences against Uniprot meeting the similarity criterion of at least 50% identity covering at least 50% of both sequences. Functional profiles were summarized from BLAST results by mapping Uniprot IDs to Gene Ontology terms. For each GO term assigned to a given protein, all parents' terms were considered as well. GO assignments were then grouped together into functional categories based on similarity and overlap.

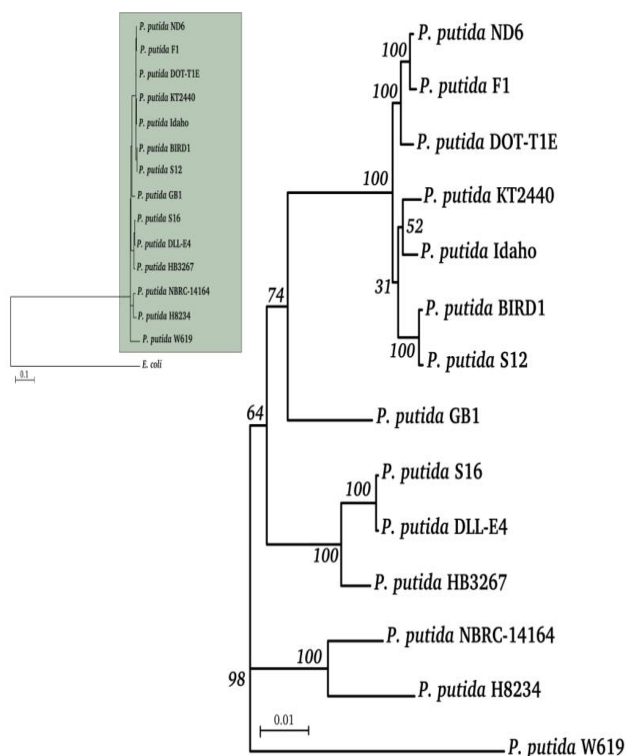
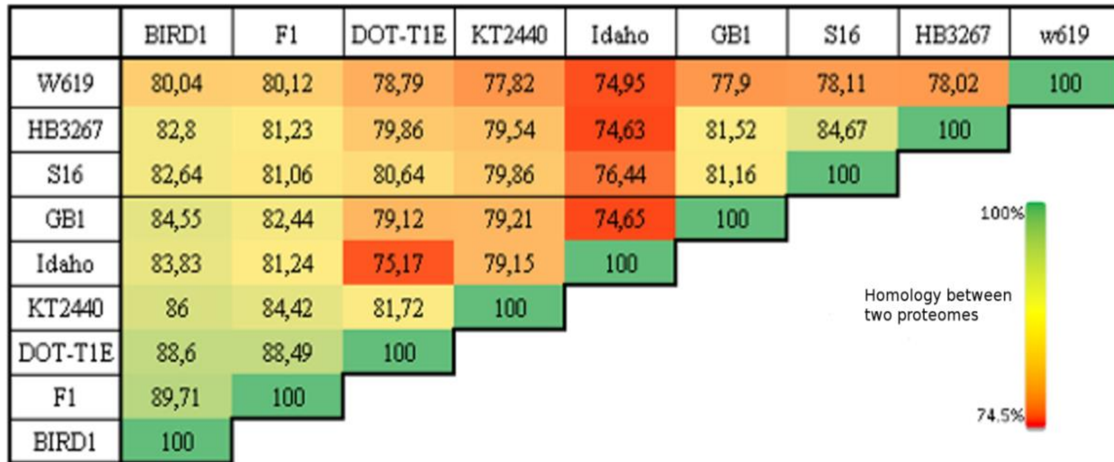


Figure 4. 1. Maximum likelihood phylogenetic tree of fourteen *Pseudomonas putida* strains with *Escherichia coli* as outgroup strain, based on a multilocus sequence alignment of four housekeeping genes (16S rRNA, *gyrB*, *rpoB* and *rpoD*). The monophyletic group above comprises strains that are mainly identified as solvent tolerant and/or as plant growth promoting rhizobacteria.

Results and Discussion

Evolutionary relationships

16S rRNAs are functionally conserved, relatively long molecules, that are relevant for phylogenetic studies at the genus level; however, phylogenetic relationships between closely related microorganisms need to be solved using the sequence of other genes to achieve a higher resolution and a more robust analysis (Yamamoto and Harayama, 1998; Yamamoto *et al.*, 2000; Mulet *et al.*, 2010; Pérez-Yépez *et al.*, 2014). Multilocus sequence analysis using four conserved genes (*gyrB*, *rpoB*, *rpoD* and 16S rRNA) from fourteen *P. putida* strains was generated to establish a phylogenetic tree (Figure 4.1). A close relationship exists between *P. putida* strains isolated as solvent tolerant strains, soil inhabitants with PGPR properties and hospital isolates. This suggests the potential coevolution of strains sharing certain specific niches and that their evolutionary distance is related with their different lifestyles. We also found that the endophytic W619 strain has evolved



	BIRD1	F1	DOT-T1E	KT2440	Idaho	GB1	S16	HB3267	w619
W619	80,04	80,12	78,79	77,82	74,95	77,9	78,11	78,02	100
HB3267	82,8	81,23	79,86	79,54	74,63	81,52	84,67	100	
S16	82,64	81,06	80,64	79,86	76,44	81,16	100		
GB1	84,55	82,44	79,12	79,21	74,65	100			
Idaho	83,83	81,24	75,17	79,15	100				
KT2440	86	84,42	81,72	100					
DOT-T1E	88,6	88,49	100						
F1	89,71	100							
BIRD1	100								

Figure 4. 2 Proteome comparison represented by a Blast-matrix performed with the percent of orthologous genes shared between two strains. The Blast-matrix is ordered based on the results of the phylogenetic tree in Figure 1 and serves to highlight the correlation between the strains which share solvent tolerant and plant growth promoting capabilities. As can be observed in the Blast-matrix, *Pseudomonas putida* Idaho displays the most red and has the lowest percentage of orthologous gene shared with other strains because *P. putida* Idaho has the biggest genome (almost 0.3 Mb larger than the average). *P. putida* W619 has the smallest genome of those studied and on average has the lowest number of genes shared with other strains. Green bars indicate that a higher percentage of orthologous genes are shared. Thus, a greener colour indicates that the strain is more closely related to those within the first monophyletic group (in Figure 4.1), which are solvent tolerant strains

separately from all other strains analysed (Figure 4.1). The results of the BLAST all-against-all comparison between the proteomes of the nine strains are shown in Figure 4.2. The given values are percentages of orthologous groups (OGs) shared between two proteomes. The BLAST matrix revealed a high similarity among the proteomes of all strains, which range between 74.4 and 89%. The heat map used in Figure 4.2 shows a correlation between the classification of the strains based on the phylogenetic tree carried out with the multilocus sequence analysis of four housekeeping genes, and the overall homology at the global proteome. These results suggest that the rate of accumulation of mutations in housekeeping genes is homogeneous with changes in the whole genome.

Compositional variation of DNA

In this work, we have defined the pangenome of the species *P. putida*, which is composed of three groups of genes: (1) a core genome, which contains genes present

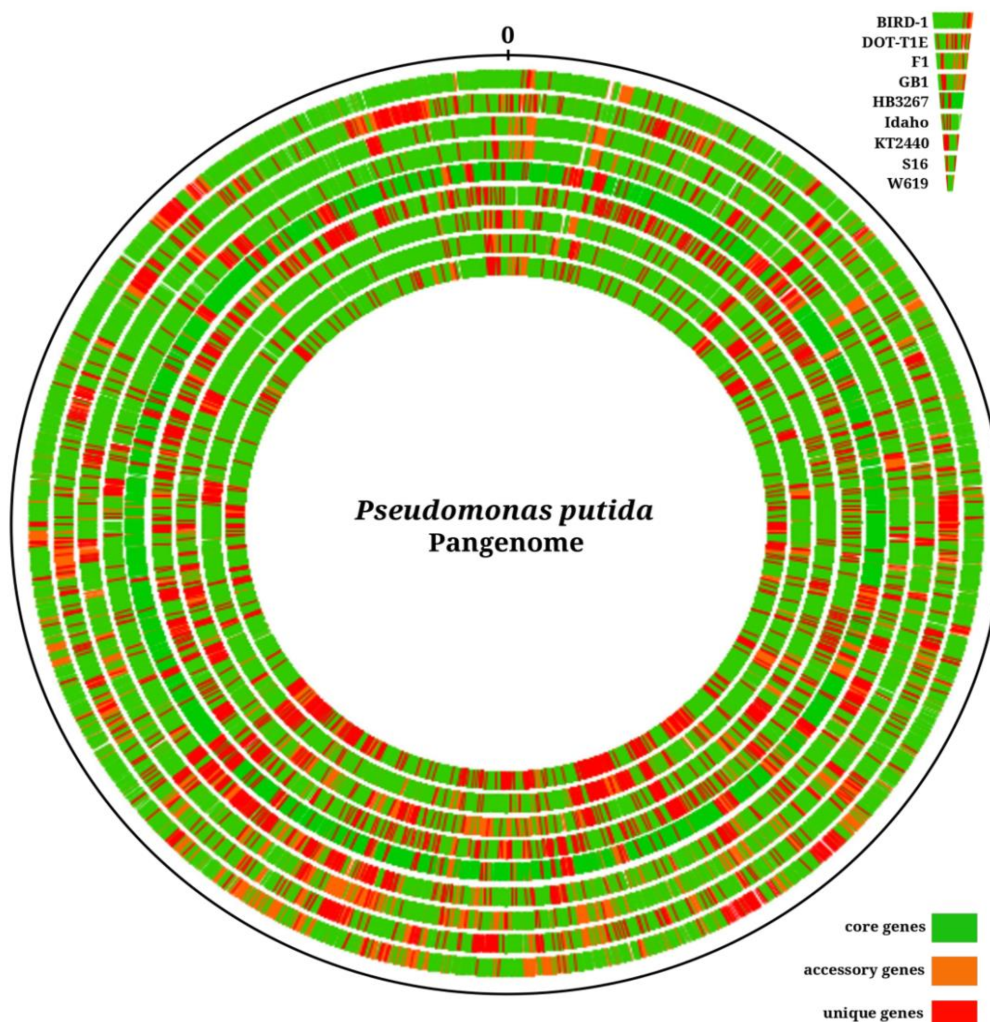


Figure 4. 3 Pie chart showing the distribution of the three sets of genes present in the pangenome of nine *Pseudomonas putida* strains. The diagram shows a higher concentration of unique and accessory genes in the region opposite to *ori*C located at the 0 minute mark.

in all strains; (2) an accessory genome, which contains genes present in two or more strains; and (3) a set called “unique genes”, which do not share sequence similarity with any other gene of the species. The vast majority of prokaryotic species have non-random codon usage. Codon usage, as with other structural DNA factors, is correlated with the potential expression level of a gene and its mutational pressure (Roymondal *et al*., 2009; Ran *et al*., 2014).

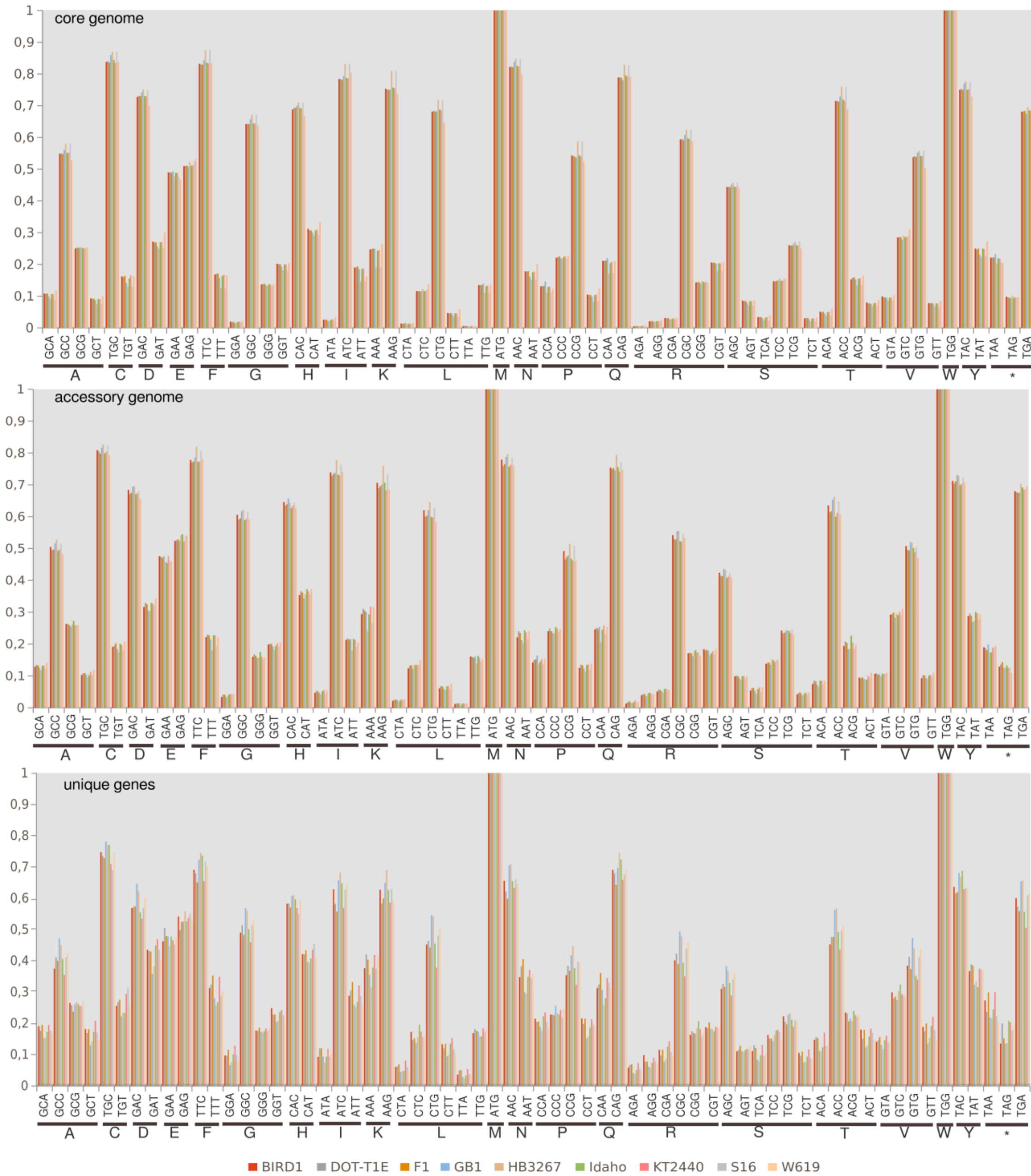


Figure 4. 4. Codon usage of the nine *Pseudomonas putida* strains as calculated using the cusp program from the EMBOSS package. Codon usage was calculated using gene sequences from the three groups of pangenome genes separately. The data are expressed as percentages and reflect the contribution of each codon. The codon usage percentages of the core and accessory genes show slight differences while the unique genes have a lower codon usage bias

To determine the potential existence of a relationship between the location of genes belonging to the three groups of genes that make the pangenome and their

involvement with compositional variations of DNA, genes of the three groups were analyzed according to their GC content and CAI and mapped in the circular chromosome of each strain. The overall GC content of the nine genomes is similar (61% - 63%) but it is not uniformly distributed across the chromosomes. The pie chart in Figure 4.3 shows that although accessory and unique genes are distributed throughout the chromosome, there is a greater concentration of accessory and unique genes in the region opposite to *ori*C located at the 0 minute mark. Statistical tests using ANOVA demonstrated that differences in GC content in the three groups of pangenome genes are significant ($p<0.001$). The GC content of unique genes was the lowest with an average value of 58.2%, followed by 61.6% for accessory genes and 62.9% for core genes. Subsequent analyses using CAI and the TukeyHSD test confirmed that the GC content averages between these groups are significantly different from each other ($p<0.001$ for CAI). Comparison of the codon usage percentages of the three groups of genes (Figure 4.4) revealed reduced codon usage bias in the set of unique genes, indicating that this cluster evolves at a different rate and may be expressed at different levels than the core and accessory genes (Ran *et al*., 2014). Furthermore, the codon usage divergence seen in the unique genes suggests that they were recently acquired through horizontal gene transfer.

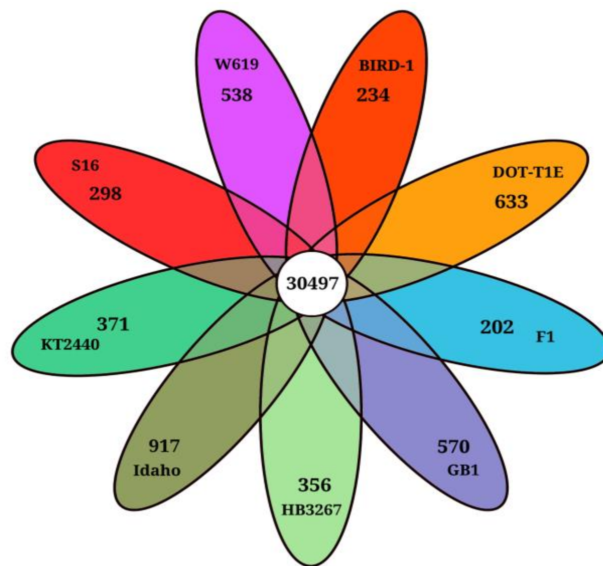


Figure 4. 5. Number of unique genes present in each strain that contribute to the pangenome of *Pseudomonas putida*. The central number, 30497 correspond to the entire number of genes that comprises the core genome of the *Pseudomonas putida* species based on the analysis of the nine *Pseudomonas putida* strains. Number represented to the extremes of the petals are the unique genes belonging to each strains that are no shared with any other of the strains in this study.

Table 4. 1. General features of the nine *Pseudomonas putida* genomes used in the pangenome study

P. putida	Genome size (MB)	Num. proteins	G+C content %	Coding density	Num. tRNA	Num. rRNA	Characteristics	NCBI or Genbank reference sequence
BIRD1	6.18	4,960	61.7	86.9	64	22	Plant growth promoting rhizobacteria	NC_017530.1
DOT-T1E	6.26	5,721	61.4	87.3	58	24	Solvent tolerant	NC_018220.1
F1	5.96	5,252	61.9	88.7	76	19	Solvent tolerant	NC_009512.1
GB1	6.08	5,409	61.9	89.4	74	22	Manganese oxidizer	NC_010322.1
HB3267	5.96	5,195	63.0	86.1	61	5	Hospital isolate	NC_019905.1
Idaho	6.36	5,711	61.8	81.2	41	5	Solvent tolerant	AGFJ01000000
KT2440	6.18	5,35	61.6	86.7	73	22	Soil microorganism	NC_002947.3
S16	5.98	5,218	62.3	84.9	70	19	Nicotine degrader	NC_015733.1
W619	5.77	5,182	61.4	88.9	75	22	Endophytic rhizosphere bacteria	NC_010501.1
Average	6.08	5,333	61.9	86.7	66	18	-	-

Descriptive statistics of OGs

The chromosomal protein coding sequences of all fully sequenced *Pseudomonas putida* strains were compared and analysed. The genomes of the strains of the species *P. putida* are about 6Mb in size and have a similar compact coding density. On average the genome of *Pseudomonas putida* encodes 5,333 proteins (Table 4.1) and the genomes showed a high level of sequence conservation (>74 %) (Figure 4.2). Using the protocol described in Materials and Methods, a total of 48,000 coding DNA sequences (CDS) were identified in the nine *Pseudomonas putida* strains, which were clustered into 6,078 OGs. Of the total CDS, 30,497 were highly conserved across all the nine genomes and accordingly comprised the *P. putida* core genome. We found that the core genes are clustered into 3,326 OGs, the additional genes are clustered into 2,752 OGs and the unique genes of the nine strains comprised 4,119 CDS. Of all strains analysed, *P. putida* Idaho had the most unique genes, with a total of 917 exclusive sequences (16.1 % of its total genome) (Figure 4.5).

The size of Pseudomonas putida's core genome and pangenome

To estimate the size of the core genome, inclusion of all possible combinations of sequences from the nine strains were simulated. The results of all permutations of the nine genomes are shown in Figure 4.6. Regression analysis revealed that as the number of genomes increases, the amount of conserved genes identified fits an exponential decaying function. Figure 4.6 shows that the number of genes in the core genome decreases at a slower rate with the addition of each new sequence until reaching a minimum of 3,385 genes.

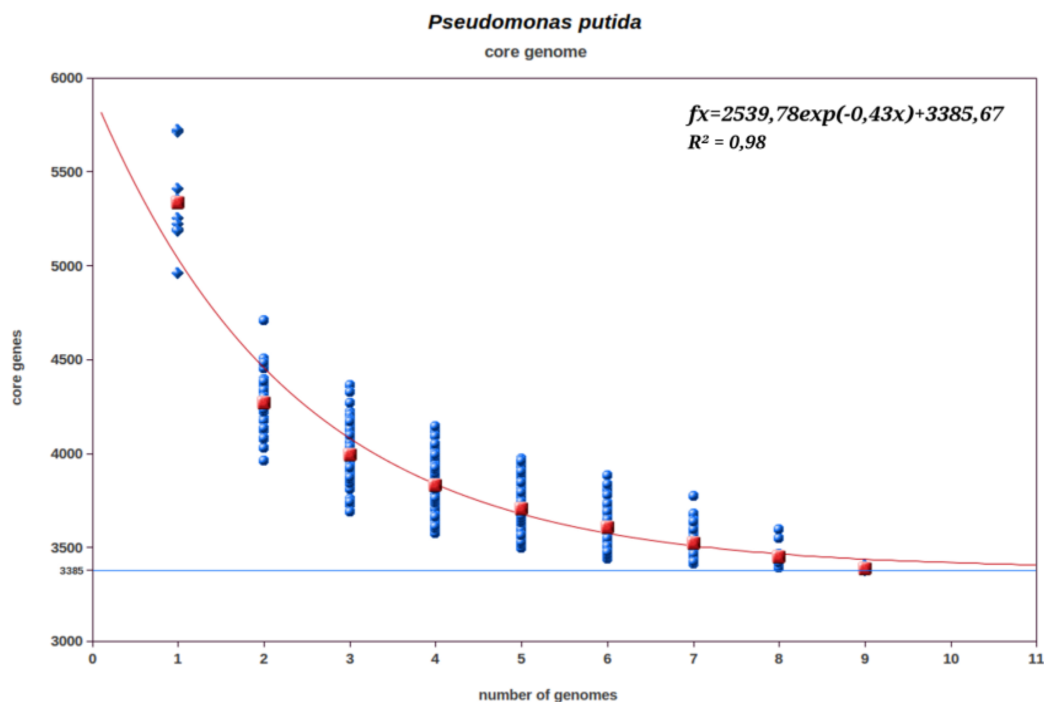


Figure 4. 6. Size of the core genome of *Pseudomonas putida*. The number of shared genes is plotted as a function of the number of n strains sequentially added. Red dots are the averages for each n. The continuous red curve represents the least-squares fit of an exponential decay function.

The continuous curve represents the least-squares fit to the function Fx=Kc*exp*[-n/τc] + Ω. The value of the Fx function is the number of shared genes for n strains and the extrapolated *P. putida* core genome size (Ω) is shown as a dashed line. Once this value is known for the function, we can infer the number of strains needed to decrease the number of core genes by half. We found that this value is 1.61, which suggests that the core genome of this strain decreased suddenly with the first value of n, but then remains relatively constant even as many more genomes are added. This is relevant because the analysis of the core genome can define the metabolic

and functional scope of *P. putida.* It also is worth noting that there are very low levels of paralogs present in the core genome, which indicates that this species has a low tendency for gene duplication (Karberg *et al*., 2011).

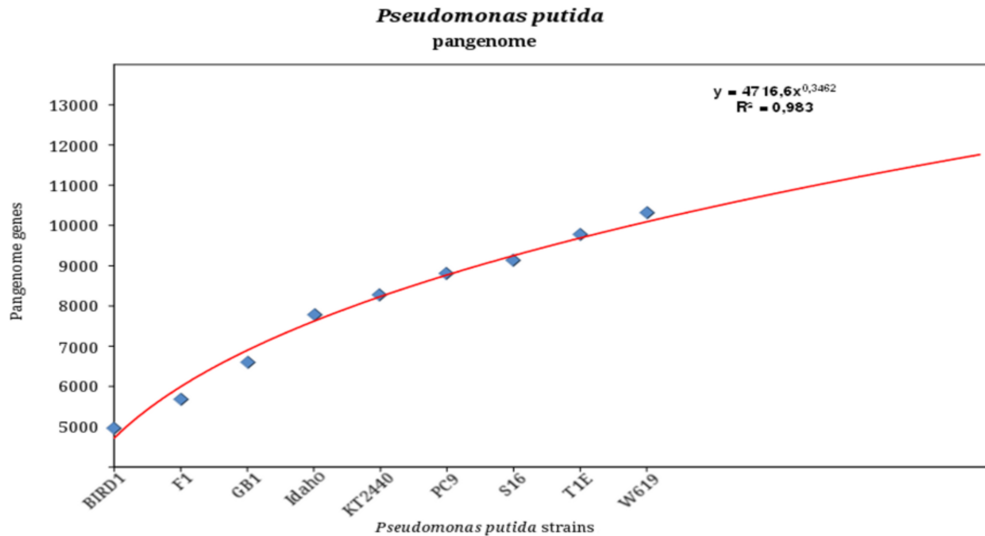


Figure 4. 7 Size of the *Pseudomonas putida* pangenome obtained upon adding a new genome. Blue diamonds are the averages for each n. The continuous red curve represents the least-squares fit of the function $y = 4{,}716.6x^{0{,}3462}$ (R=0.98), where “y” represents the pangenome size, which increases with the addition of each strain genome

To determine the global gene repertoire of the species, the number of new genes added by each genomic sequence was estimated (Figure 4.7). The plot of the number of genes was fitted by the function $Fx=kx^n$. In total, the number of gene families in the pangenome of *Pseudomonas putida*, identified by consecutive addition of the orthologous groups of core, accessory and unique genes of the strains in alphabetic order, is 10,297. The curve increases with the addition of each new strain and is far from saturation; thus, we can conclude that the pangenome of the *Pseudomonas putida* species is open. The high slope of the curve when the first strains of the analysis were added indicates that the genetic repertoire of the species is still growing in spite of a suitable adaptation to their ecological niche. The availability of a large genetic reservoir may be a key survival advantage for this species when faced with sudden environmental changes.

The core genes of the nine *Pseudomonas putida* strains were used for phylogenetic inference by a maximum likelihood tree construction (Suppl. Figure

4.1). It is remarkable the high fraction of confident nodes having a bootstrap value of 100 and the total correspondence with the results of the phylogenetic analysis performed using the 16S rRNA and the three housekeeping genes. The tree supports the claim that a close phylogenetic relationship exists between the strains that have similar ecological niche and our proposal that the ecological niche influences the content of conserved genes.

Functional analysis of core genes

In order to determine the functional profile of *Pseudomonas putida* core genome, each protein was assigned to a KEGG Orthology group (KO) and Gene Onthology (GO) functional category. The abundance of each KO category was plotted (Figure 4.8). For the core genes, a large number of genes (1,191 core genes) were classified as ABC transporters, with 70% of them identified as likely importer transport systems, which suggests that this species has a strong ability to take up a wide range

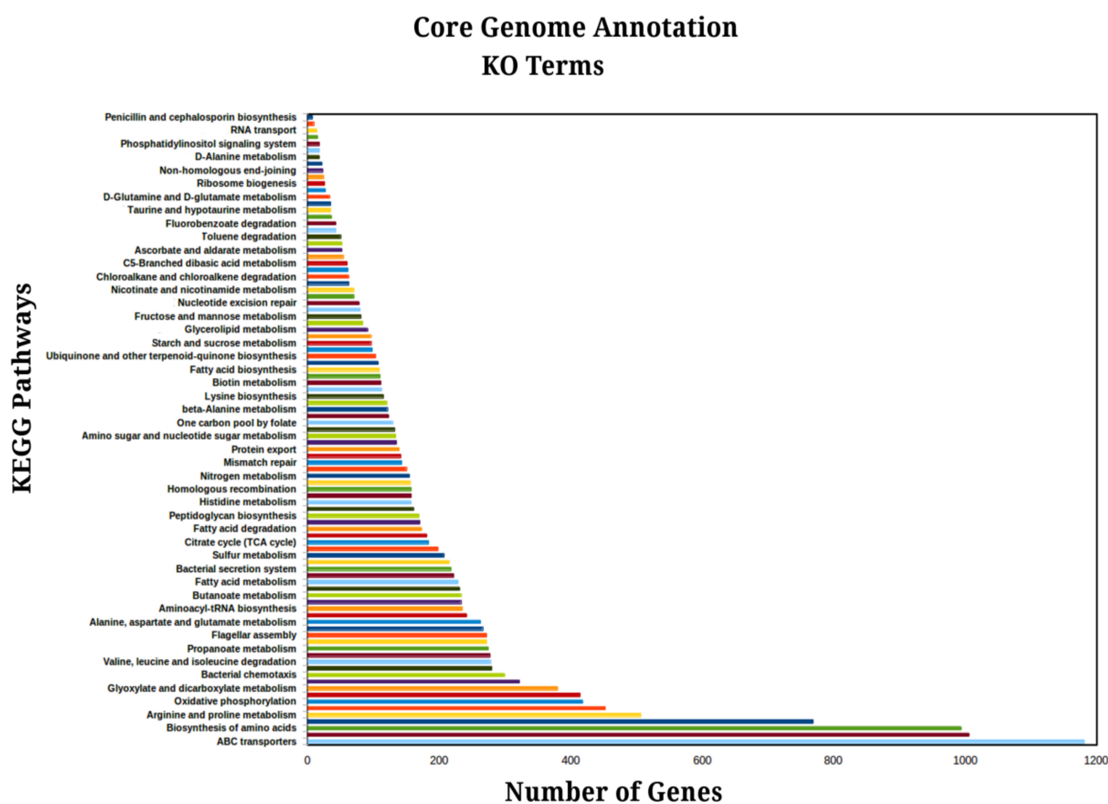


Figure 4. 8 Classification of core genes in *Pseudomonas putida*. Genes were classified by KEGG Orthology terms

of chemicals from the environment. In addition to being important to nutrient uptake, ABC transporters also play a key role in the extrusion of toxic compounds

and therefore are important in cellular homeostasis. As well as transporters, a large number of genes were found to be involved in the biosynthesis of amino acids and carbon metabolism. Molina-Henares *et al.* (2010) previously found that thirty of the genes involved in amino acid biosynthesis are essential for growth of *Pseudomonas putida* KT2440 in minimal medium because knock-outs in these genes led to auxotrophy (Molina-Henares *et al*., 2010). The same authors identified 17 additional genes as essential for growth in minimal-medium (for a total of 47 essential genes). All of these genes are part of the core genome of *Pseudomonas putida* and thirty-two of them are involved in amino acid metabolism. Of note is the remarkable number of genes for arginine and proline catabolism found in *P. putida* (Figure 4.9). There are several routes employed by bacteria to use arginine as a carbon and nitrogen source; in fact, *Pseudomonas putida* is able to catabolize arginine via four different pathways. Surprisingly, the genes for all of these pathways are present in the core genome a finding that emphasizes the importance of the metabolism of this amino acid in this species. The end-product of arginine metabolism via arginase or the AST pathway is glutamate, which is subsequently acetylated to produce N-acetylglutamate, leading to the formation of the intermediate ornithine. Ornithine is a pivotal compound in cell metabolism because it serves as a precursor for the biosynthesis as polyamines, siderophores and some antibiotics (Zúñiga *et al*., 2002; Lu, 2006). Genes for polyamine biosynthesis are well represented in the core genome, and all of the strains produce putrescine and have specific transporters for putrescine and spermidine. These compounds are essential metabolites required for cell growth and recent studies associated these compounds to processes such as the oxidative stress response, biofilm development and antibiotic resistance (Igarashi and Kashiwagi, 2010; Williams *et al*., 2010). Putrescine can also be synthesized from ornithine by ornithine decarboxylase (*speC*) or from arginine by the sequential actions of arginine decarboxylase (*speA*) and agmatinase (*speB*). These three genes are present in all *Pseudomonas putida* strains. Arginine is taken up by *P. putida* via the action of the integral membrane protein ArcD, an arginine/ornithine antiporter that mediates energy-independent exchange between ornithine in the cytosol and arginine in the medium. This protein is part of the *arcDABC* operon, which is found within the core genome. The other genes of the operon encode the arginine deiminase (ADI) pathway, through which L-arginine is transformed to ammonia and

carbon dioxide. When terminal electron acceptors such as oxygen are scarce, the ADI pathway also produces ATP (Itoh and Nakada, 2004).

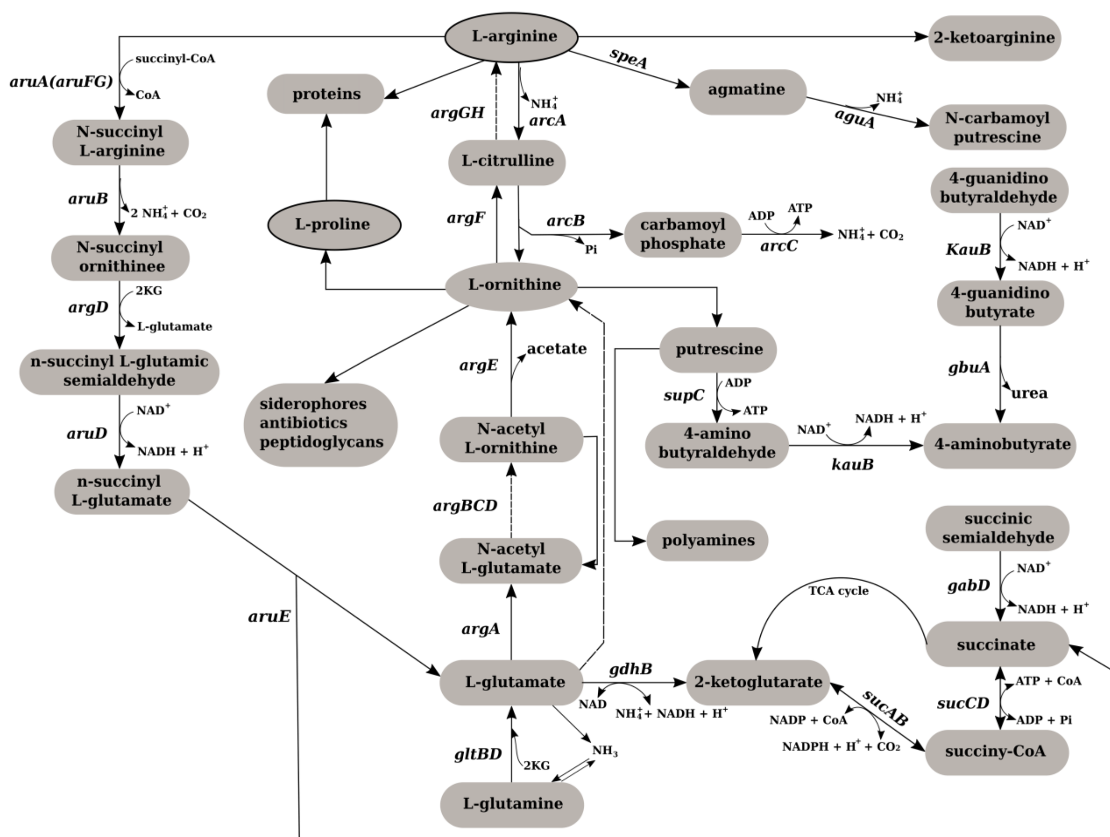


Figure 4. 9 Representation of the arginine and proline catabolism in *Pseudomonas putida* core genome. Only enzymes belonging to the core genes set were used to construct the flow diagram**.**

Pseudomonas putida is able to use a limited number of sugars as a carbon source, namely glucose, fructose and gluconate. Contrasting with the limited number of sugars used by *P. putida* is the richness of various glucose degradation pathways that are present (Suppl. Figure 4.2). These pathways converge with the production of 6-phosphogluconate, which is metabolized via the Entner-Doudoroff pathway.

All the genes for glucose metabolism and Entner-Doudoroff are present in the core genome of *Pseudomonas putida* except the *pgi* glucose epimerase gene, which is missing in the *P. putida* Idaho genome. The Entner-Doudoroff pathway enzymes Edd and Eda metabolize 6-phosphogluconate to glyceraldehyde-3-phosphate and pyruvate, which are Krebs cycle intermediates. Three homologous genes for glucose-6-phosphate dehydrogenase (*zwf-1*, *zwf-2*, *zwf-3)* are conserved in all nine

strains although according to previous studies, only *zwf-1* plays a major role in glucose metabolism and is also important for defence against oxidative stress (Kim *et al*., 2008). Glucose-6-phosphate dehydrogenase catalyzes the committed step of the pentose phosphate pathway and the enzyme is regulated by availability of the substrate $NADP^+$. As NADPH is utilized in reductive synthetic pathways, the increasing concentration of $NADP^+$ stimulates the pentose phosphate pathway to replenish NADPH. The pentose phosphate pathway, along with the Entner-Doudoroff pathway and the rest of the metabolic network, meet the high demand of NADPH needed to sustain the environmental lifestyle of this species (Chavarría *et al*., 2013). Pyruvate made from glucose is transformed into acetyl-coenzyme-A through the activity of pyruvate dehydrogenase and enters into the tricarboxylic acid cycle where it is subsequently decomposed to CO_2 and H_2O. As well as containing all the tricarboxylic acid cycle genes, the *Pseudomonas putida* core genome encodes all genes needed for glyoxylate cycle, through which organic material is recycled into the cell biomass rather than being released as carbon dioxide (Suppl. Figure 4.3).

Urea hydrolysis via urea amidohydrolase (urease), and the genes for the accessory proteins required for urease assembly and nickel insertion (encoded by the *ureDEFG* codon), are found in the core genome as well. At least two different types of urea transporter genes are present on the core genome, the *urtABCDE* operon which is induced under nitrogen-limiting conditions; and a low-affinity urea transporter that promotes the entry of urea into the cell when this compound is abundant in the environment. Both uptake systems provide a constant flow of ammonium to the cell, which serves as nitrogen source and also enables the colonization of acid environments. Nitrate and nitrite are abundant nitrogen sources in the environment and are taken up into the cell via specific transporters (nitrate/nitrite transport system components, including the NasS substrate-binding protein). Nitrate is reduced in the cytoplasm to nitrite by the action of a nitrate reductase and nitrite is then reduced to ammonium via a nitrite reductase. All the genes for nitrate and nitrite reduction and ammonia assimilation via L-glutamine and L-glutamate are part of the core genome of *P. putida.* The genome of *Pseudomonas putida* KT2440 possesses eight genes encoding glutamine synthetase, seven of which form part of the core genome. This emphasizes the importance of L-glutamine

metabolism in *Pseudomonas*. Recent studies in our group have shown that L-glutamine could play a key role also in stress response to organic solvents such as short chain alcohols (Cuenca *et al*., in preparation).

There are at least six core genome genes related to sulfate transport across the cell membrane (including the CysAWT sulfate transport system permease protein; the Sbp sulfate binding protein; the Pit-type sulfate permease; and the CysZ-type sulfate transporter). All the genes related to sulfate assimilation via hydrogen sulfide and cysteine biosynthesis (*cysDIJHNCM* genes) are also present in the *P. putida* core genome. The *ssuEADCBF* and *tauABCD* gene clusters required for the desulfonation of alkanesulfonates and taurine as sulfur sources are also present in the core genome and ensure that this species has an adequate sulfur supply for biosynthetic processes when sulfur is limiting in the environment.

In the same way, the low-affinity inorganic phosphate transporter system and the Pst (phosphate specific transport) system, ensure a constant flow of inorganic phosphate into the cell. In addition to these genes (encoded on *pstBACS)* is the PhoR-PhoB two component regulatory system, which detects and responds to the changes in environmental phosphate concentrations (Baek and Lee, 2007). Also present on the core genome is at least an alkaline phosphatase, which makes inorganic phosphate available by hydrolyzing organic molecules. Once the inorganic phosphate is inside the cell, the polyphosphate kinase (PPK) uses it for poly-phosphate biosynthesis, polymerizing the terminal phosphate of ATP into poly-phosphate in a freely reversible reaction (Hirota *et al*., 2010). The utilization and degradation of the poly-phosphate is catalyzed by an exopolyphosphatase, and several poly-phosphate specific kinases, including polyP-glucokinase and polyP-fructokinase, which are present in *Pseudomonas putida* core genome.

More than twenty core genes are involved in the response to oxidative stress. At least three catalases and two superoxide dismutase are found in the genome of all nine strains. Additionally, there are twelve glutathione-dependent enzymes present that catalyze the conjugation of electrophilic substrates to glutathione, as well as serving other functions.

Pseudomonas putida is a multi-flagellated species, having between five to seven flagella at one pole (Harwood *et al*., 1989; Theves *et al*., 2013; Martinez-García *et*

al., 2014) which enables it to explore the environment for nutrients and escape possible threat. In addition, flagella play a central role in adhesion, biofilm formation and chemotaxis (Kirov, 2003; Sampedro *et al*., 2015). In the *P. putida* core genome are more than thirty genes that make up structural components of the flagellum and serve as regulators, including *fliD*, *fliE*, *fliH*, *fliK* and *fliO* genes that are not present in other groups, such as *Alphaproteobacteria* (Liu and Ochman, 2007). In total, 55 genes in the core genome were identified as potentially involved in adhesion, biofilm formation, motility or attachment according to the early study by Duque *et al*. 2013.

Availability of a mini-Tn5 mutant collection for P. putida KT2440

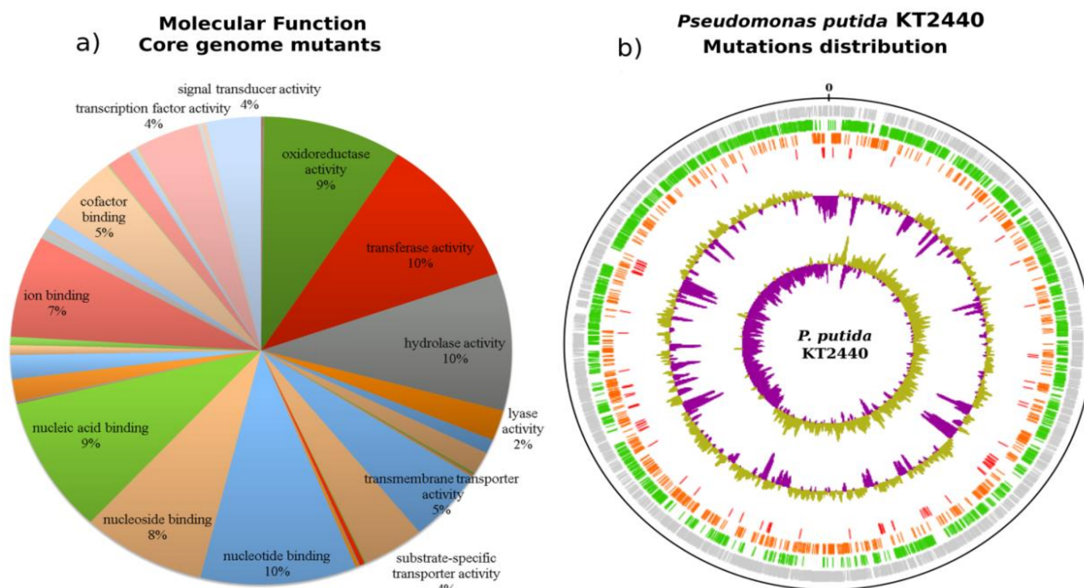


Figure 4. 10. Core genome mutants a) Representation of the GO "Molecular function" core genome mutants genes. The pie chart shows the percentage of activities presents in the core genome mutant genes. The non-redundant GO terms were summarized with REVIGO system (Supek et al., 2011); b) Pie chart showing the distribution of the mutations within the genome of *Pseudomonas putida* KT2440 strain. The mutated gene representations are separated according to their belonging to the three different groups of genes. Mutations of genes are represented in green colour, mutations belonging to accessory genes are represented in orange color and those belonging to unique genes are represented in red colour. The outermost grey ring represents the genes that are not mutated. Finally the two innermost rings, belong a GC plot, where areas with GC above range are in green colour, while GC below range is purple and a GC skew (G−C/G+C), in which regions above average are colored in green colour, while below average are colored in purple. Genes were mapped in the chromosome and GC plot and GC skew

We have generated a genome-wide mini-Tn*5* collection of mutants in the KT2440 strain (Duque *et al*., 2007), which serves as a powerful tool for functional analyses of *P. putida* genes. The mutants were generated through independent mutagenesis and some of were selected for deficiencies in the use of carbon or nitrogen sources, auxotrophic behaviour or enhanced sensitivity to environmental cues.

The mutants are stored individually and the mini-Tn5 insertion site has been sequenced for about 3000 independent clones. Our current analysis enabled us to identify 2620 mutations in open reading frames (ORFs) and 494 in intergenic regions. The functional information for the core genome mutants is provided in Figure 4.10a and in Supplementary Figures 4.4 and 4.5, which highlight mutations in genes that serve as cellular components or that function within biological processes. The complete list of mutated functions is available in Supplementary Table 4.1.

The mutants have been grouped according to whether they belong to the core genome, accessory genes or unique genes. We have identified 1821 mutants in the core genome (53% of core genes), 704 accessory genes (45% of accessory genes) and 95 unique genes (25% of unique genes) (Figure 4.10b). For the core mutants, we found a homogenous distribution of the mutations within the genome, indicating that no specific insertion sites or hotspots exist in the genome of this strain for mini-Tn-5 insertions. Regarding molecular functions, 43% of mutants were enzymes or potential enzymes, 11% were transporters, and 33% were regulator binding proteins. A number of activities were matched with phenotypes; for example, insertions in the *edd* (PP_1010) and *eda* (PP_1024) genes in the Entner-Doudoroff pathway were defective in glucose assimilation (del Castillo *et al*., 2007); also a mutant in the *nir* (PP_1705) gene was deficient in the utilization of nitrate and nitrite; and mutants in genes such as *argH, pheA* and *trpE* (PP_0184, PP_1769 and PP_0417 respectively) were auxotrophic (Molina-Henares *et al*., 2010). This collection serves as a useful tool for defining the core metabolism of the *Pseudomonas putida* species and all *P. putida* strains, as well as, in some cases, bacteria of the genus *Pseudomonas*. Importantly, the collection will enable accessory genes to be assigned functions, which will help to explain the unique characteristics of the various strains of the

genus *Pseudomonas* and in other genera that have acquired these genes via horizontal gene transfer.

Acknowledgments

We thank Ben Pakuts for critical reading of the manuscript.

References

Alexeyenko, A., Tamas, I., Liu, G., and Sonnhammer, E.L.L. (2006) Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* **22**: e9–e15.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., *et al*. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC genomics* **9**: 75.

Baek, J.H. and Lee, S.Y. (2007) Transcriptome analysis of phosphate starvation response in *Escherichia coli*. *Journal of microbiology and biotechnology* **17**: 244–252.

Bagdasarian, M., Lurz, R., Rückert, B., Franklin, F.C., Bagdasarian, M.M., Frey, J., and Timmis, K.N. (1981) Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*. *Gene* **16**: 237–247.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., *et al*. (1997) The complete genome sequence of *Escherichia coli K-12*. *Science* **277**: 1453–1462.

Calzada, J., Zamarro, M.T., Alcón, A., Santos, E., Díaz, E., García, J.L., *et al*. (2009) Analysis of dibenzothiophene desulfurization in a recombinant *Pseudomonas putida* strain. *Applied and environmental microbiology* **75**: 875–877.

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. (2009) DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**: 119–120.

del Castillo, T., Ramos, J.L., Rodríguez-Herva, J.J., Fuhrer, T., Sauer, U., and Duque, E. (2007) Convergent peripheral pathways catalyze initial glucose catabolism in *Pseudomonas putida* : genomic and flux analysis. *Journal of bacteriology* **189**: 5142–5152.

Castresana, J. (2000) Selection of Conserved Blocks from Multiple Alignments for their use in Phylogenetic Analysis. *Molecular biology and evolution* **17**: 540–552.

Chavarría, M., Nikel, P.I., Pérez-Pantoja, D., and de Lorenzo, V. (2013) The Entner-Doudoroff pathway empowers *Pseudomonas putida KT2440* with a high tolerance to oxidative stress. *Environmental microbiology* **15**: 1772–1785.

Darling, A.E., Mau, B., and Perna, N.T. (2010) progressiveMauve : multiple genome alignment with gene gain , loss and rearrangement. *PloS one* **5**: e11147.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012) jModelTest 2 : more models , new heuristics and parallel computing CircadiOmics : integrating circadian genomics , transcriptomics , proteomics. *Nature Methods* **9**: 772.

van Dillewijn, P., Nojiri, H., Van der Meer, J.R., and Wood, T.K. (2009) Bioremediation, a broad perspective. *Microbial biotechnology* **2**: 125–127.

Duque, E., Molina-Henares, A.J., de la Torre, J., Molina-Henares, M.A., del Castillo, T., Lam, J., and Ramos, J.L. (2007) Towards a genome-wide mutant library of *Pseudomonas putida* strain KT2440. In, Ramos,J.L. and Filloux,A. (eds), *Pseudomonas: A Model System in Biology*. The Netherlands: Springer, Dorchester, pp. 227–251.

Edgar, R. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Fernández, M., Niqui-Arroyo, J.L., Conde, S., Ramos, J.L., and Duque, E. (2012) Enhanced tolerance to naphthalene and enhanced rhizoremediation performance for *Pseudomonas putida* KT2440 via the NAH7 catabolic plasmid. *Applied and environmental microbiology* **78**: 5104–5110.

Frank, S., Klockgether, J., Hagendorf, P., Geffers, R., Schöck, U., Pohl, T., *et al.* (2011) *Pseudomonas putida* KT2440 genome update by cDNA sequencing and microarray transcriptomics. *Environmental microbiology* **13**: 1309–1326.

Harwood, C.S., Fosnaugh, K., and Dispensa, M. (1989) Flagellation of *Pseudomonas putida* and analysis of its motile behavior. *Journal of bacteriology* **171**: 4063–4066.

Hirota, R., Kuroda, A., Kato, J., and Ohtake, H. (2010) Bacterial phosphate metabolism and its application to phosphorus recovery and industrial bioprocesses. *Journal of bioscience and bioengineering* **109**: 423–432.

Igarashi, K. and Kashiwagi, K. (2010) Modulation of cellular function by polyamines. *The International Journal of Biochemistry & Cell Biology* **42**: 39–51.

Itoh, Y. and Nakada, Y. (2004) Arginine and Polyamine Metabolism. In, Ramos,J.L. (ed), *Pseudomonas volume 3: Biosynthesis of Macromolecules and Molecular Metabolism*. Kluwer Academic, New York, pp. 243–272.

Karberg, K.A., Olsen, G.J., and Davis, J.J. (2011) Similarity of genes horizontally acquired by *Escherichia coli* and *Salmonella enterica* is evidence of a supraspecies pangenome. *PNAS* **108**: 20154–20159.

Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., and Mclnerney, J.O. (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC evolutionary biology* **6**: 29.

Kim, J., Jeon, C.O., and Park, W. (2008) Dual regulation of zwf-1 by both 2-keto-3-deoxy-6-phosphogluconate and oxidative stress in *Pseudomonas putida*. *Microbiology* **154**: 3905–3916.

Kirov, S.M. (2003) Bacteria that express lateral flagella enable dissection of the multifunctional roles of flagella in pathogenesis. *FEMS microbiology letters* **224**: 151–159.

Li, J., McConkey, B.J., Cheng, Z., Guo, S., and Glick, B.R. (2013) Identification of plant growth-promoting bacteria-responsive proteins in cucumber roots under hypoxic stress using a proteomic approach. *Journal of Proteomics* **84**: 119–131.

Liu, R. and Ochman, H. (2007) Stepwise formation of the bacterial flagellar system. *PNAS* **104**: 7116–7121.

Lu, C. (2006) Pathways and regulation of bacterial arginine metabolism and perspectives for obtaining arginine overproducing strains. *Appl Microbiol Biotechnol* **70**: 261–272.

Martinez-García, E., Nikel, P.I., Chavarría, M., and de Lorenzo, V. (2014) The metabolic cost of flagellar motion in *Pseudomonas putida* KT2440. *Environmental microbiology* **16**: 291–303.

Matilla, M.A., Pizarro-Tobias, P., Roca, A., Fernández, M., Duque, E., Molina, L., *et al*. (2011) Complete genome of the plant growth-promoting rhizobacterium *Pseudomonas putida* BIRD-1. *Journal of bacteriology* **193**: 1290.

Molina, L., Ramos, C., Duque, E., Ronchel, M.C., García, J.M., Wyke, L., and Ramos, J.L. (2000) Survival of *Pseudomonas putida* KT2440 in soil and in the rhizosphere of plants under greenhouse and environmental conditions. **32**: 315–321.

Molina, L., Udaondo, Z., Duque, E., Fernández, M., Molina-Santiago, C., Roca, A., *et al*. (2014) Antibiotic resistance determinants in a *Pseudomonas putida* strain isolated from a hospital. *PloS one* **9**: e81604.

Molina-Henares, M.A., de la Torre, J., García-Salamanca, A., Molina-Henares, J.A., Herrera, M.C., Ramos, J.L., and Duque, E. (2010) Identification of conditionally essential genes for growth of *Pseudomonas putida* KT2440 on minimal medium through the screening of a genome-wide mutant library. *Environmental microbiology* **12**: 1468–1485.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35** : W182–W185.

Mulet, M., Lalucat, J., and García-Valdés, E. (2010) DNA sequence-based analysis of the *Pseudomonas* species. *Environmental microbiology* **12**: 1513–1530.

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Fouts, D.E., *et al*. (2002) Complete genome sequence and comparative analysis of the metabolically versatile. **4**: 799–808.

Okazaki, M., Sugita, T., Shimizu, M., Ohode, Y., Iwamoto, K., de Vrind-de Jong, E.W., *et al*. (1997) Partial purification and characterization of manganese-oxidizing factors of *Pseudomonas fluorescens* GB-1. *Applied and environmental microbiology* **63**: 4793–4799.

Palleroni, N.J. (1984) Gram negative aerobic rods and cocci, family I: Pseudomonadaceae. in Bergey's manual of systematic bacteriology, Williams a. N. R. Krieg and J. G. Holt (ed) Baltimore, Md.

Pérez-Yépez, J., Armas-capote, N., Velázquez, E., Pérez-galdona, R., Rivas, R., and León-barrios, M. (2014) Evaluation of seven housekeeping genes for multilocus sequence analysis of the genus *Mesorhizobium*: Resolving the taxonomic affiliation of the *Cicer canariense* rhizobia. *Systematic and Applied Microbiology* **37**: 553–559.

Phoenix, P., Keane, A., Patel, A., Bergeron, H., Ghoshal, S., and Lau, P.C.K. (2003) Characterization of a new solvent-responsive gene locus in *Pseudomonas putida* F1 and its functionalization as a versatile biosensor. *Environmental microbiology* **5**: 1309–1327.

Pizarro-Tobías, P., Fernández, M., Niqui, J.L., Solano, J., Duque, E., Ramos, J.L., and Roca, A. (2014) Restoration of a Mediterranean forest after a fire: bioremediation and rhizoremediation field-scale trial. *Microbial biotechnology* **8**: 77–92.

Prakash, D., Pandey, J., Tiwary, B.N., and Jain, R.K. (2010) A process optimization for bio-catalytic production of substituted catechols (3-nitrocatechol and 3-methylcatechol). *BMC biotechnology* **10**: 49.

Ramos, J.L., Mermod, N., and Timmis, K.N. (1987) Regulatory circuits controlling transcription of TOL plasmid operon encoding meta-cleavage pathway for degradation of alkylbenzoates by *Pseudomonas*. *Molecular Microbiology* **1**: 293–300.

Ramos, J.L., Stolz, A., Reineke, W., and Timmis, K.N. (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. *PNAS* **83**: 8467–8471.

Ramos-Díaz, M.A. and Ramos, J.L. (1998) Combined physical and genetic map of the *Pseudomonas putida* KT2440 chromosome. *Journal of bacteriology* **180**: 6352–63.

Ran, W., Kristensen, D.M., and Koonin, E. V (2014) Coupling between protein level selection and codon usage optimization in the evolution of bacteria and archaea. *MBio* **5**: 1–11.

Reineke, W., Jeenes, D.J., Williams, P.A., and Knackmuss, H.-J. (1982) TOL Plasmid pWWO in Constructed Halobenzoate-Degrading *Pseudomonas* Strains: Prevention of *meta* Pathway.

Remm, M., Storm, C.E. V, and Sonnhammer, E.L.L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**: 1041–1052.

Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics* **16**: 276–277.

Roca, A., Pizarro-Tobías, P., Udaondo, Z., Fernández, M., Matilla, M.A., Molina-Henares, M.A., *et al*. (2013) Analysis of the plant growth-promoting properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1. *Environmental microbiology* **15**: 780–794.

Roymondal, U., Das, S., and Sahoo, S. (2009) Predicting gene expression level from relative codon usage bias: an application to *Escherichia coli* genome. *DNA research* **16**: 13–30.

Sampedro, I., Parales, R.E., Krell, T., and Hill, J.E. (2015) Pseudomonas chemotaxis. *FEMS Microbiology Reviews* 17–46.

Santos, P.M., Benndorf, D., and Sá-Correia, I. (2004) Insights into *Pseudomonas putida* KT2440 response to phenol-induced stress by quantitative proteomics. *Proteomics* **4**: 2640–2652.

Sashidhar, B. and Podile, A.R. (2009) Transgenic expression of glucose dehydrogenase in *Azotobacter vinelandii* enhances mineral phosphate solubilization and growth of sorghum seedlings. *Microbial biotechnology* **2**: 521–529.

Segura, A. and Ramos, J.L. (2013) Plant-bacteria interactions in the removal of pollutants. *Current opinion in biotechnology* **24**: 467–473.

Stamatakis, A. (2014) RAxML Version 8 : A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one* **6**: e21800.

Taghavi, S., Barac, T., Greenberg, B., Borremans, B., Vangronsveld, J., and van der Lelie, D. (2005) Horizontal gene transfer to endogenous endophytic bacteria from poplar improves phytoremediation of toluene. *Applied and environmental microbiology* **71**: 8500–8505.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**: 2731–2739.

Tao, F., Tang, H., Gai, Z., Su, F., Wang, X., He, X., and Xu, P. (2011) Genome sequence of *Pseudomonas putida* Idaho, a unique organic-solvent-tolerant bacterium. *Journal of bacteriologyacteriology* **193**: 7011–7012.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., *et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *PNAS* **102**: 13950–13955.

Theves, M., Taktikos, J., Zaburdaev, V., Stark, H., and Beta, C. (2013) A Bacterial Swimmer with Two Alternating Speeds of Propagation. *Biophysj* **105**: 1915–1924.

Timmis, K.N. (2002) *Pseudomonas putida*: a cosmopolitan opportunist par excellence. *Environmental microbiologyicrobiology* **4**: 779–781.

Udaondo, Z., Molina, L., Daniels, C., Gómez, M.J., Molina-Henares, M.A., Matilla, M. a, *et al*. (2013) Metabolic potential of the organic-solvent tolerant *Pseudomonas putida* DOT-T1E deduced from its annotated genome. *Microbial biotechnology* **6**: 598–611.

Weinel, C., Nelson, K.E., and Tümmler, B. (2002) Global features of the *Pseudomonas putida* KT2440 genome sequence. **4**: 809–818.

Weyens, N., Truyens, S., Dupae, J., Newman, L., Taghavi, S., van der Lelie, D., *et al.* (2010) Potential of the TCE-degrading endophyte *Pseudomonas putida* W619-TCE to improve plant growth and reduce {TCE} phytotoxicity and evapotranspiration in poplar cuttings. *Environmental Pollution* **158**: 2915–2919.

Williams, B.J., Du, R.-H., Calcutt, M.W., Abdolrasulnia, R., Christman, B.W., and Blackwell, T.S. (2010) Discovery of an operon that participates in agmatine metabolism and regulates biofilm formation in *Pseudomonas aeruginosa*. *Molecular Microbiology* **76**: 104–119.

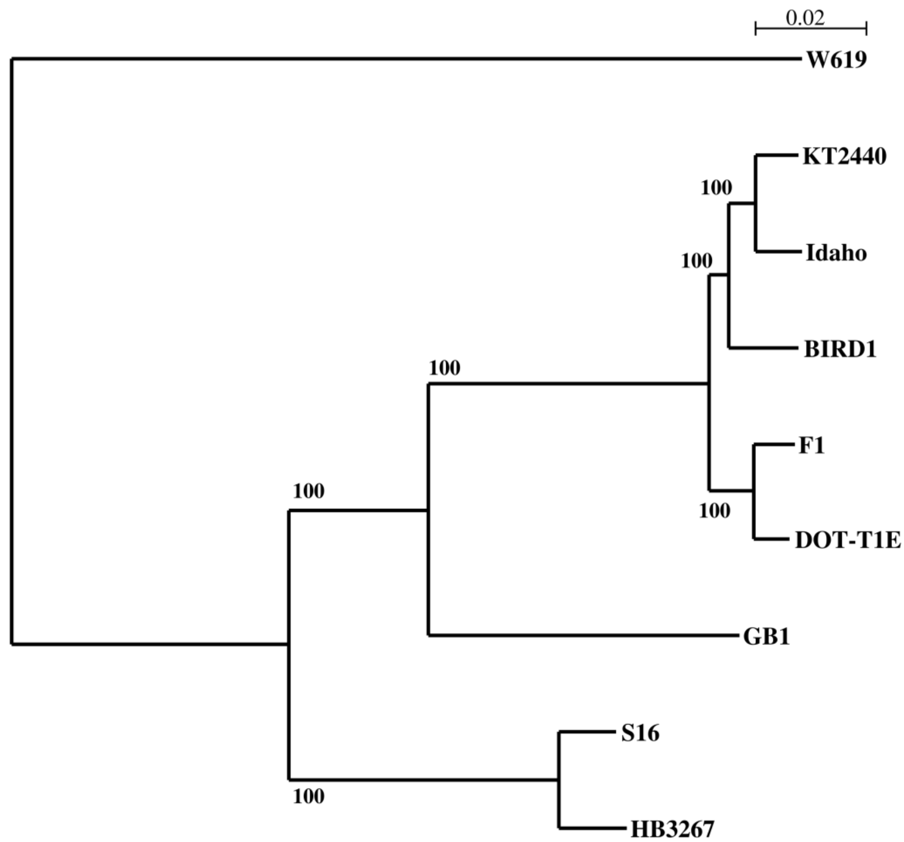
Yamamoto, S. and Harayama, S. (1998) Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes. *International journal of systematic bacteriology* **48**: 813–819.

Yamamoto, S., Kasai, H., Arnold, D.L., Jackson, R.W., Vivian, A., and Harayama, S. (2000) Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiology* **146**: 2385–2394.

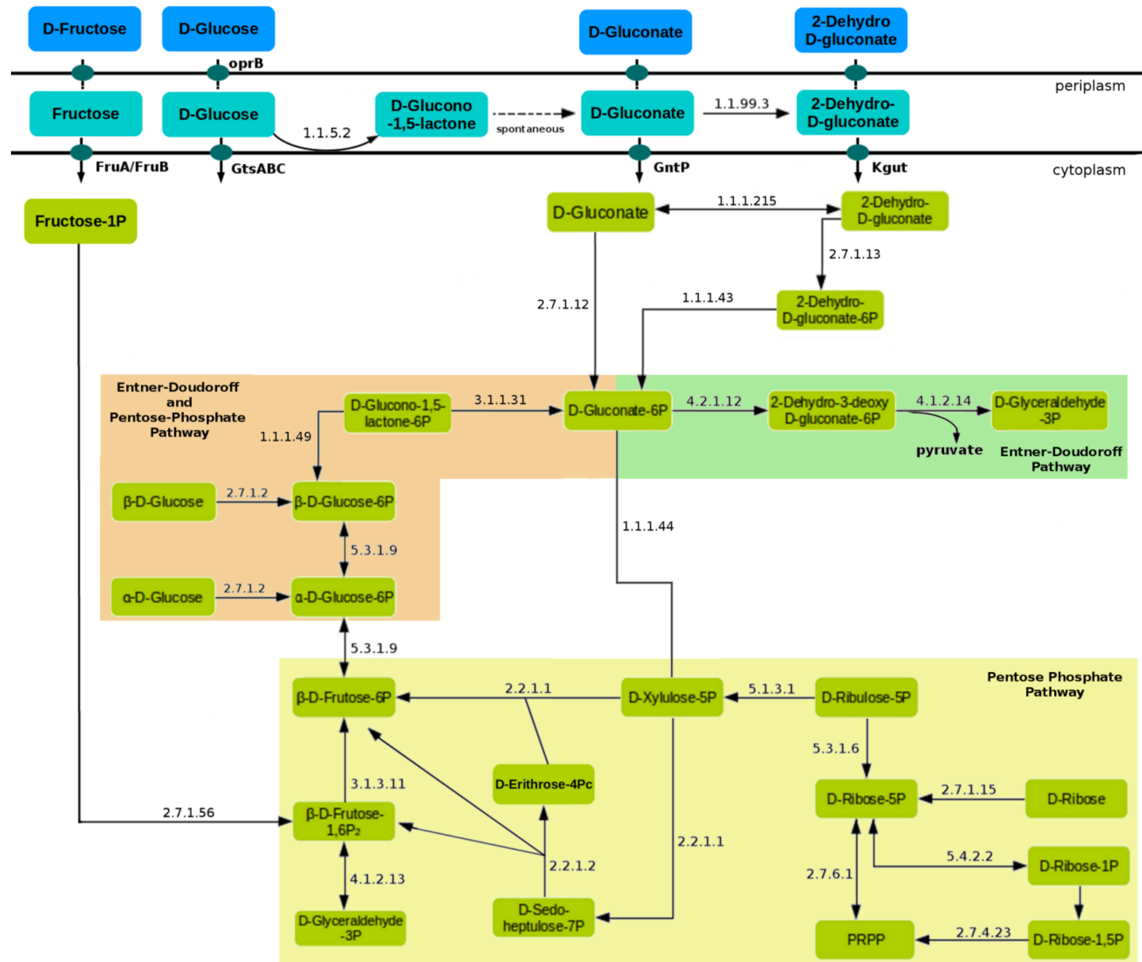
Yu, H., Tang, H., Wang, L., Yao, Y., Wu, G., and Xu, P. (2011) Complete genome sequence of the nicotine-degrading *Pseudomonas putida* strain S16. *Journal of bacteriology* **193**: 5541–5542.

Zúñiga, M., Miralles, M. del C., and Pérez-Martínez, G. (2002) The product of *arcR*, the sixth gene of the arc operon of *Lactobacillus sakei*, is essential for expression of the arginine deiminase pathway. *Applied and environmental microbiology* **68**: 6051–6058.

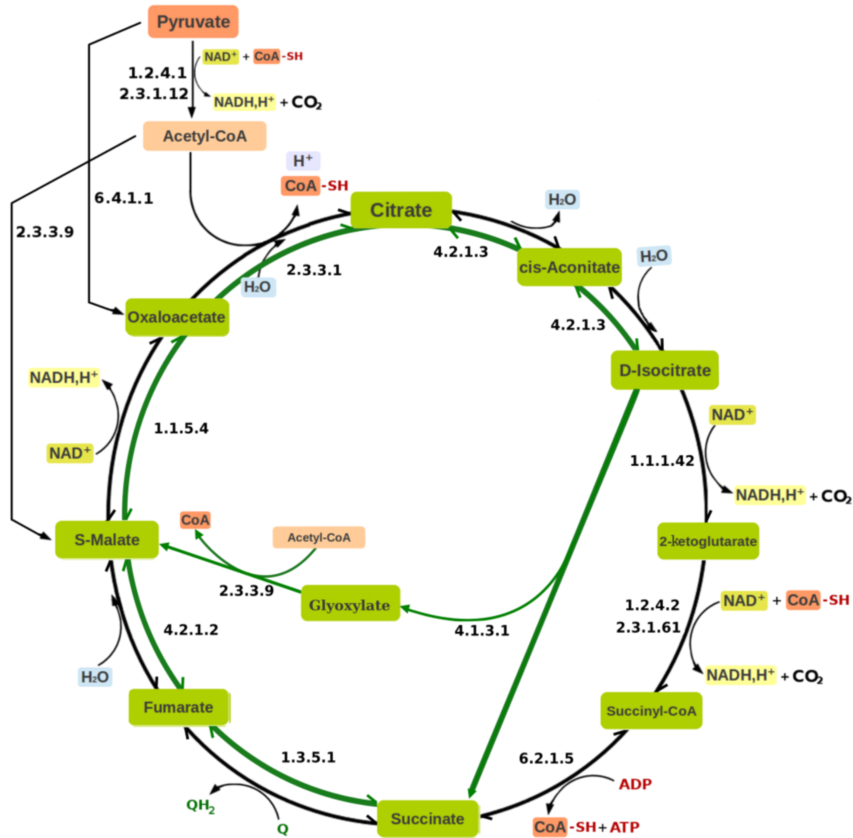
Supporting information



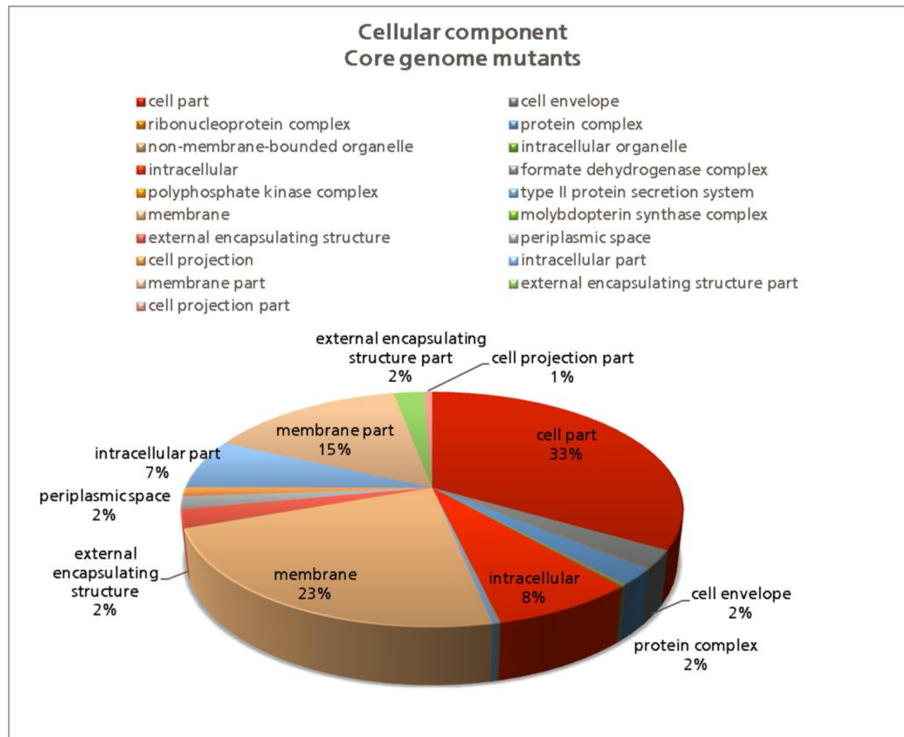
Suppl. Figure 4. 1 Maximum likelihood phylogenetic tree of the core genes from nine *Pseudomonas putida* strains based on a multilocus sequence alignment of 3,326 orthologous group of genes. Bootstrap values of 500 replication are shown at each node.



Suppl. Figure 4. 2. . Flow diagram for glucose metabolism in *Pseudomonas putida*. Only enzymes belonging to the core gene set were used to prepare the diagram.



Suppl. Figure 4. 3. Representation of the tricarboxylic acid cycle (black line) and glyoxylate cycle (green line) in *Pseudomonas putida* core genome. Only enzymes belonging to the core genes set were used to prepare the flow diagram



Suppl. Figure 4. 4. Representation of the GO "Cellular component" core genome mutants. The pie chart shows the percentage of activities presents in the core genome mutant genes. The nonredundant GO terms were summarized with REVIGO system (Supek *et al*., 2011).

Suppl. Figure 4. 5. Representation of the GO "Biological process" core genome mutants. The pie chart shows the percentage of activities presents in the core genome mutant genes. The nonredundant GO terms were summarized with REVIGO system (Supek *et al*., 2011).

Suppl. Table 4. 1. . Complete list of mutated functions in *P. putida* KT2440. The table is divided in three sections: Mutants in the set of accessory genes of *Pseudomonas putid*a KT2440; Mutants in the set of core genes of *Pseudomonas putida* KT2440; Mutants in the set of unique genes of *Pseudomonas putida* KT2440.

(Available as Appendix Online)

Chapter V

Accessory and unique genes of *Pseudomonas putida* pangenome

Zulema Udaondo and Juan Luis Ramos (2016)

Summary

The sum of all the genes present in the genomes of different strains of a bacterial species is known as the species pangenome. *Pseudomonas putida* pangenome consists of a set of genes present in all strains of the species and known as the core genome. Accessory and unique genes, in the pangenome are those genes shared by at least two strains (but not all) and the genes that are found only in an individual strain, respectively. In this chapter we show that accessory and unique genes are an important source of genetic variability in bacterial populations, allowing strains of the same species to better adapt to specific niches.

Introduction

The increasing number of complete genomes sequences available in public databases gives us the opportunity to compare genomic variations within genus and species as well as between strains of the same species, providing us with a clearer understanding of bacterial evolution. Nonetheless, it seems that a bacterial species will never be fully described because each strain has a specific set of genes what makes a gene pool much larger every time a new strain is sequenced (Medini *et al*., 2005; Mira *et al*., 2010; Segerman, 2012). The pangenome concept emerged as a new approach to describe bacterial species (Tettelin *et al*., 2005) and refers to the complete set of gene of a species at the time the sequence of a number of strains is available. Each species has a conserved set of genes present in all strains and known as core genes, additionally each strain has a number of accessory genes and unique genes. The existence of a core set of genes present in all strains of an species is a witness of the preservative natural evolution (Gil *et al*., 2004; Glass *et al*., 2006; Lapierre and Gogarten, 2009; Juhas *et al*., 2011). This group of core genes is often related to crucial cellular processes and it does not correspond to the minimal set of genes necessary for an organism to survive and thrive in nature (Koonin, 2003; Lapierre and Gogarten, 2009). Core genes, however, define the set of essential bricks on which the rest of a genome is built. Linked to the core genes are those traits derived from a common ancestor (Mann *et al*., 2013) that maintain the genetic integrity of the species over time (Lan and Reeves, 2000). The sets of genes in a bacterial genome that can be used to distinguish strains and serotypes and that represent key genes for survival in a particular environmental niche are known as accessory and unique genes (Young *et al*., 2006; Lapierre and Gogarten, 2009; Mira *et al*., 2010; Silby *et al*., 2011). The accessory genes are genes present in two or more but not all strains, while unique genes are exclusive to each strain in the analysis (Medini *et al*., 2005). Much of these “unique” genes come from other species or even genus, and they seem to have been acquired through horizontal gene transfer (HGT) with the help of mobile genetic elements during the evolutionary history of the strain. Because these genes are acquired from different origins and have a weaker sequence conservation between taxa, computational annotations of the set of unique genes usually result in a large percentage of the unique genes

annotated as that they encode hypothetical proteins, many of which related to phage or prophage proteins (Hiller *et al*., 2007; Walker *et al*., 2011).

Pseudomonas putida is a Gram-negative road-shaped bacterium species typically found in soil, water and moist environments. Members of this species have a broad metabolic versatility, which allows them to adapt to different habitats and nutritional environments including the rhizosphere of plants (Nelson *et al*., 2002; Wu *et al*., 2010; Roca *et al*., 2013). *Pseudomonas putida* shows a high robustness against extreme environmental conditions such as high temperature, extreme pH, or the presence of toxins or inhibiting solvents (Ramos-González *et al*., 2001; Krell *et al*., 2012; Poblete-Castro *et al*., 2012). Strains of this species have occasionally been isolated from patients in hospitals in Japan, USA, Italy and France (Weinstein and Von Graevenitz, 1971; Yoshino *et al*., 2011; Molina *et al*., 2014). Based on the previously observations, we propose that the species *P. putida* represents an outstanding model system for the study of the transition between pathogenic and saprotrophic lifestyles using a comparative genomics approach. We have previously described the set of core genes of the pangenome of *P. putida* (Udaondo *et al*., 2015) that characterize the essence of the species, defining the functional features derived from a common ancestor. Previous analysis of the strains specific genes in *P. putida* showed that this set of genes tends to have a distinctive base composition (G+C content) and a less biased codon usage (Médigue *et al*., 1991; Daubin and Ochman, 2004; Young *et al*., 2006; Karberg *et al*., 2011; Udaondo *et al*., 2015) than the vertically inherited genes. Other explanation about the unique genes is that they represent very fast evolving genes, or possibly pseudogenes, with rates of substitution and rearrangements that obscure their similarity to known proteins (Daubin and Ochman, 2004). The presence of new unique genes identified even after the addition of several genomes in the pangenome study suggest an “open pangenome” in which any new genome added to the analysis will increase the pangenome size (Medini *et al*., 2005; Mira *et al*., 2010; Mongodin *et al*., 2013). Open pangenomes have been observed in species that can inhabit a wide range of environments and/or have diverse lifestyles and possess efficient means of lateral gene transfer (Tettelin *et al*., 2005; Rasko *et al*., 2008; Donati *et al*., 2010; Mira *et al*., 2010; De Maayer *et al*., 2014). In this study we have identified the genetic traits linked to variation in niche adaptation and antibiotic resistance that are represented

in the sets of accessory and unique genes, to better understand the diversity within species and to present a more consistent definition of the species itself.

Materials and Methods

Strain genome sequences. The genomes of *P. putida* strains BIRD1, DOT-T1E and HB3267 were determined previously by our group (Roca *et al*., 2013; Udaondo *et al*., 2013; Molina *et al*., 2014). When this project started, the genomes of other strains (*P. putida* F1 (Phoenix *et al*., 2003), *P. putida* GB1 (Okazaki *et al*., 1997), *P. putida* KT2440 (Nelson *et al*., 2002), *P. putida* S16 (Yu *et al*., 2011) and *P. putida* W619 (Taghavi *et al*., 2005)) were available at ftp://ftp.ncbi.nih.gov/genomes/Bacteria/ and used in the current study. The draft genome sequence of *Pseudomonas putida* Idaho (Tao *et al*., 2011) was downloaded from NCBI and annotated by our group using the RAST annotation server (Aziz *et al*., 2008).

Construction of ortholog groups (OGs). The proteomes of the nine strains used for pangenomic studies were clustered into OGs using a BBH approach using InParanoid 4.1 (Remm *et al*., 2001) and MultiParanoid (Alexeyenko *et al*., 2006). The BLOSUM 80 matrix was used during the InParanoid run. As the nine strains used in this study are evolutionarily closely related, MultiParanoid was reliably used to merge multiple pairwise orthologs from InParanoid into multi-strain ortholog groups. Genes were defined as unique if they did not have a bidirectional best hit in any of the eight other genomes.

KEGG ontology (KO) and Gene ontology (GO) annotation and analysis. For functional annotation of the three set of genes, we used the KAAS automatic genome annotation server (Moriya *et al*., 2007) to assign KO terms. GO categories were inferred for each set of genes through the alignment of protein sequences against Uniprot meeting the similarity criterion of at least 50% identity covering at least 50 % of both sequences. Functional profiles were summarized from BLAST results by mapping Uniprot IDs to Gene Ontology terms. For each GO term assigned to a given protein, all parents' terms were considered as well. GO

assignments were then grouped together into functional categories based on similarity and overlap. For functional analysis visualization, the lists of GO terms, were thereafter summarized with REVIGO (Supek *et al*., 2011) by clustering semantically close GO terms and with BGI WEGO (Ye *et al*., 2006) web program.

Results and Discussion

Descriptive analysis of the pangenome of P. putida

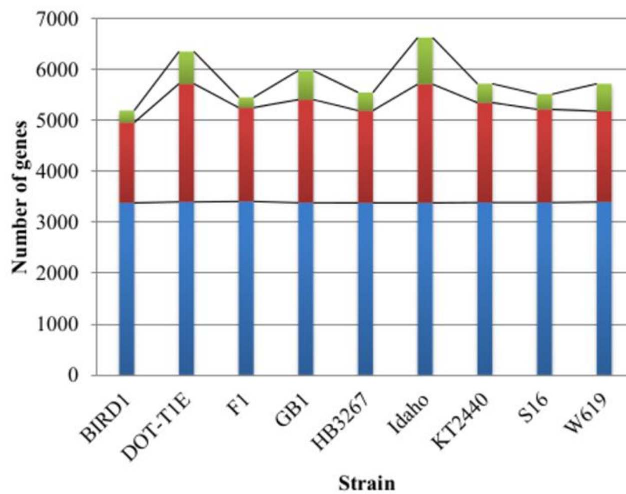


Figure 5. 1. Pangenome gene distribution of the number of genes across the nine *P. putida* strains of the study. The sizes of blue, red, and green areas in a stack indicate the number of genes that belong to the core, accessory and unique genes respectively

The pangenome of *P. putida* was constructed using the genome of nine strains (Udaondo *et al*., 2015). The chromosomes of the nine genomes of *P. putida* are on average 6.08 Mb, and have an average G+C content of 62 %. The annotated sequences support that 87% of the genome of these microorganisms code for CDS. The pangenome of the species revealed a total of 30,497 sequences as core genes with an average of 3,389 core sequences per genome that were clustered in 3,326 orthologous groups of genes. In comparison to other microbial pangenomic studies (Rasko *et al*., 2008; Blom *et al*., 2009; Donati *et al*., 2010; Mann *et al*., 2013) *P. putida* core genome size is in the middle of the range with an average of 63 % of the genes being part of the core genome. The clustering criteria used to establish the different set of genes was 80 % of sequence identity over 50 % length, which is a restrictive criterion in comparison to other pangenomic studies at the species level (Tettelin *et al*., 2005; Baltrus *et al*., 2011). This approach allows us to

have a robust set of core genes shared by all the strains. As observed for other bacterial species, the size of the conserved core of proteins from *P. putida* decreases as a function of the number of strains compared, while the size of the pangenome increases (Udaondo *et al*., 2015). The number of unique genes does not exceeded 16 % in any of the strains of the study. In total, there were 4,119 strain-specific genes identified in the pangenome of this species. The *P. putida* Idaho (Tao *et al*., 2011) was the strain with the largest number of unique genes, with 917 strain specific genes. We think this is the consequence of several factors as the low level of curation of Idaho genome, which was the only strain used in the study with an assembly at the level of contig. Besides, the genome of *P. putida* Idaho has one of the largest genomes of the study with 6,363,067 nucleotides as the sum of its 839 contigs which resulted in 5,713 proteins predicted and annotated using the RAST annotation server (Aziz *et al*., 2008). In the pangenome of *P. putida* the distribution of the core, accessory and unique genes in each genome is illustrated in Figure 5.1.

Functional analysis of the accessory genes

KEGG-based metabolic pathway reconstruction was used to compare the functional repertoire of *P. putida* accessory genes, assigning each protein to a KEGG Orthology term (KO) (Mao *et al*., 2005). Annotation was performed using the KAAS server from KEGG (Moriya *et al*., 2007), which provides functional

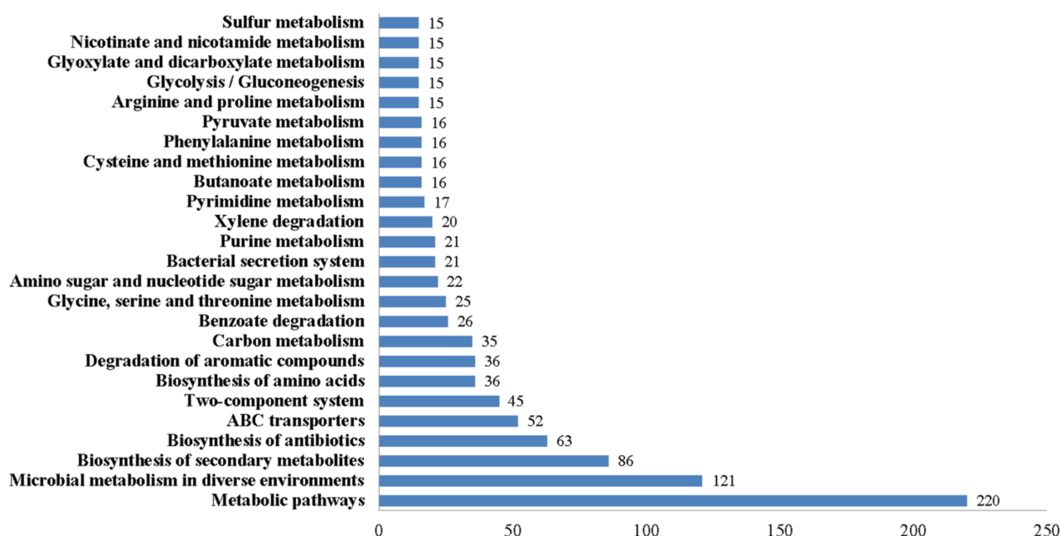


Figure 5. 2. Functional annotation of accessory genes using KEGG Automatic Annotation Server (KAAS). The chart shows the metabolic pathways and the number of KO terms associated to each pathway

annotation comparing the amino acids sequences uploaded to the server against a manually curated set of ortholog groups. The 40% percent of the total of 13,384 accessory genes were annotated with KO terms; the abundance of KO terms was plotted (Figure 5.2), showing that the largest number of KO terms annotated belongs to the generic set of “Metabolic pathways”, followed by functions related to the “Microbial metabolism in diverse environments”, “Biosynthesis of secondary metabolites” and “Biosynthesis of antibiotics”.

In a closer examination within the generic group “Microbial metabolism in diverse environments”, the KO terms most represented in the study belonged to genes related to degradation of aromatic compounds such as cymene, benzene, toluene, cumate and catechol. In addition there were abundant genes related to transport systems for sulfate, sulfonate, phosphate, nickel, iron, molybdate, putrescine, putrescine/spermidine, glycine-betaine/proline, the ribose transport system and the nitrate/taurine (NitT/TauT) family transport system. There are also numerous transport systems for branched-chain and polar amino acids. Other categories well represented were those related to bacterial secretion systems. Transport systems for alpha-hemolysin/cyclolysin, the hemophore/metalloprotease transport system, the adhesion protein transport system, and a large number of KO genes related to type II and VI general secretion system are present in the set of accessory genes. Regarding the two-component regulatory system, the CusS-CusR (cooper tolerance) type constitutes the largest group of KO genes. Finally, the multidrug resistence efflux pumps AcrAD-TolC, AcrAb-TolC, MdtABC, AmeABC, EmrAB and MeHI-OprD, complete the largest set of transporter hosted in the set of accessory genes of the *P. putida* species.

To compare functional roles of conserved genes in all strains and accessory genes, Gene Ontology (GO) terms at the second level of each functional category (biological process, molecular function, and cellular component) were assigned to the core and accessory genes. Here, we obtained the GO functional annotations for the nine *P. putida* genomes by using Blastp against Uniprot DB (The Uniprot Consortium, 2014). We found that 45% of the accessory genes were assigned with a GO term annotation. The WEGO software (Ye *et al*., 2006) was used to establish GO functional classifications for each set of genes in the pangenome and to examine the distribution of gene functions. The abundance of functional terms was

standarized by the total number of functional terms considering functional categories separately. A high-percentage of the genes in the cellular component category was classified as “cell part”, and “macromolecular complex”. The dominant terms in the molecular function category were catalytic activity, transporter activity, binding, and transcriptional regulator activity. In the biological processes category, most of the terms were classified under metabolic processes and cellular process. Figure 5.3 shows those functional categories that showed significant abundance of core genes compared to the accessory genes in the pangenome of *P. putida*. Categories as cofactor transporter, deaminase, structural molecule, cell motility, cellular component dissasembly and reproduction of a single-cell microorganism, were present mainly or exclusively in the core genome.

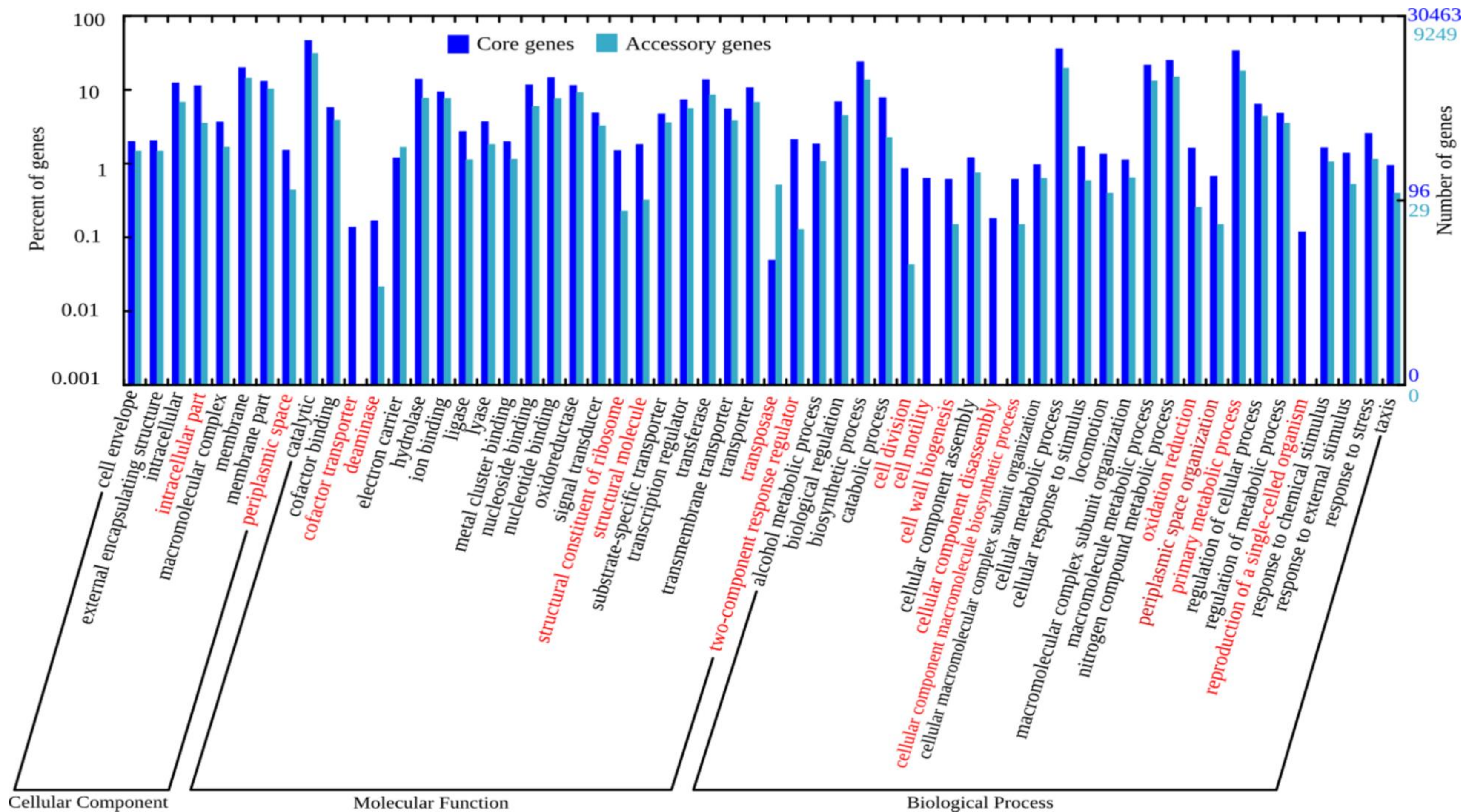


Figure 5. 3. Functional annotation with core and accessory genes GO terms. The left axis represents the percentage of associated gene to the total provided genes. The right axis shows the number of genes associated to the GO term. The p-value of Chi-Square test between the genes number of the two set of genes was calculated. Only ontologies with remarkable relationships in this representation (p-value below 0.05) were used. Terms with large differences between sets are written in red. Analysis made with BGI WEGO. (Ye et al., 2006)

Functional analysis of the unique genes

The structural analysis of the DNA composition in the three set of genes that constitute the pangenome, showed that unique genes have a distinct base composition (G+C) and a non-biased codon usage pattern (Udaondo *et al*., 2015), giving rise to the assumption that this set of genes were acquired from phylogenetically distant and disparate sources (Karberg *et al*., 2011). To clarify the exogenous acquisition of the unique genes, we used IslandViewer3 software (Dhillon *et al*., 2015) for computational identification, comparison, and visualization of genomic islands from the nine *P. putida* strains used in this analysis. This computational tool integrates three different genomic island prediction methods, IslandPick (Langille *et al*., 2008), SIGI-HMM (Stanke and Waack, 2003), and IslandPath-DIMOB (Langille *et al*., 2008). Once the genomic islands of the nine strains were predicted, the coordinates of the unique genes were mapped on the chromosome of the strains to figure out the coincidences between genes belonging to unique genes and genes harbored in genomic islands.

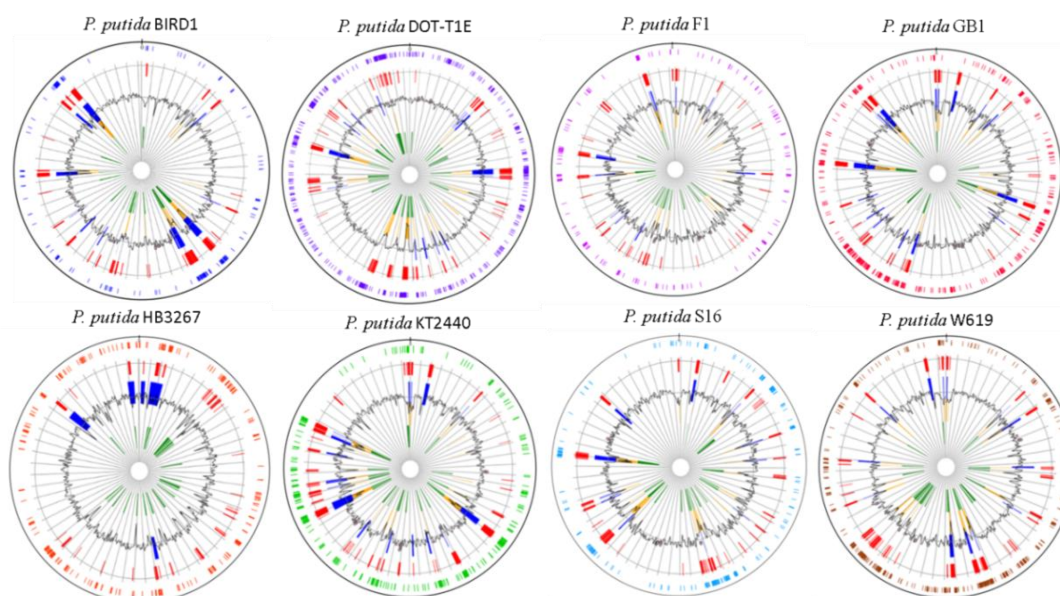


Figure 5. 4. Circular representation of the genome of eight *P. putida* strains. Circles display the position in the chromosome of the unique genes. In the inner rings the prediction of Islandviewer3 software for computational identification of genomic islands (Dhillon et al., 2015) is shown. The unique genes mapping was performed only with strains whose chromosome is closed, therefore the strain Idaho was not taken into account for this representation

As expected, several genomic islands were detected in all genomes and regions, a high concentration of unique genes matched the position of genomic island (Figure 5.4).

Table 5. 1. Summary of the number of unique genes harbored by each strain.

Unique genes	**234**	**633**	**202**	**570**	**356**	**917**	**371**	**298**	**538**
Island genes	**394**	**573**	**356**	**422**	**245**	**488**	**553**	**370**	**537**
Unique genes in island	**140 60%**	**269 42%**	**93 46%**	**154 27%**	**122 34%**	**11 34%**	**211 57%**	**23 41%**	**35 44%**

The number of genes that are in genomic islands and the percentage of unique genes harbors in island are also shown.

We estimated that 43 % percent of the unique genes from the *P. putida* pangenome belong to genomic islands (Table 5.1), supporting the lateral acquisition of almost half of this set of genes. Although the large number of uncharacterized and hypothetical proteins hinder the functional classification of the unique set of genes, the KAAS system (Moriya *et al.*, 2007) and the Uniprot DB were used to annotate unique genes with KO and GO terms respectively. Enriched GO terms were clustered using REVIGO (Supek *et al.*, 2011) (Figure 5.5).

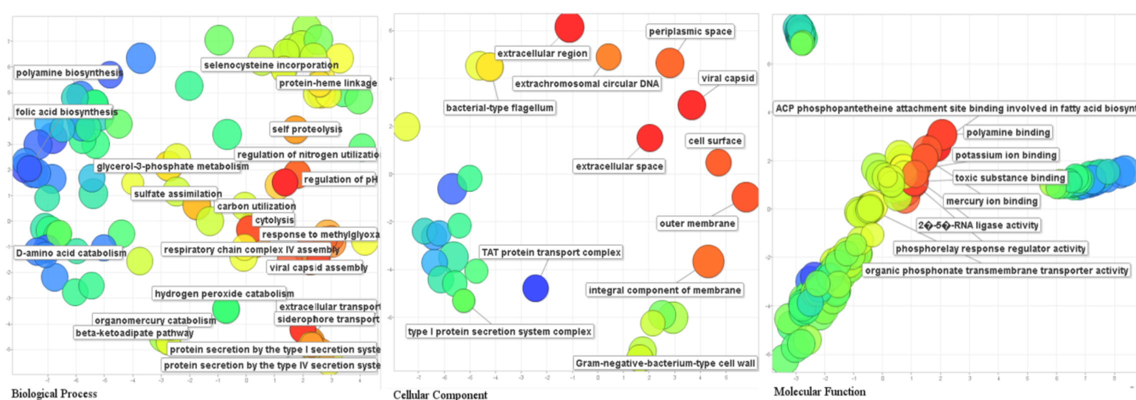


Figure 5. 5. Funtional analysis of the set of unique genes with GO terms. Enriched go terms were clustered with REVIGO (Supek et al., 2011). Each of the squared charts represents one of the Gene Ontology categories (Biological process, Cellular component and Molecular function)

The analysis showed that GO categories related with regulation of pH, nitrogen utilization, sulfate assimilation, siderophore transport and binding to mercury, toxic substances, potassium and polyamines are well represented in this set of genes. These functions in the set of unique genes are related with strain specifc capabilities.

That is the case of the PGPR strain *P. putida* BIRD1 in which Biological Process GO terms related with nitrogen compound metabolism and chemotaxis are the most abundant. In the case of the opportunistic pathogen strain HB3267, GO categories related with cytolysis, self proteolysis and detoxification were the most abundant terms; while in the solvent tolerant DOT-T1E strain GO terms related to cellular aromatic compound metabolism and polyamine transport were the most abundant. Therefore, this study provides support to the hypothesis that unique genes shape the environment in which certain strains can proliferate while others have serious impediments.

References

Alexeyenko, A., Tamas, I., Liu, G., and Sonnhammer, E.L.L. (2006) Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* **22**: e9–e15.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., *et al*. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC genomics* **9**: 75.

Baltrus, D. a, Nishimura, M.T., Romanchuk, A., Chang, J.H., Mukhtar, M.S., Cherkis, K., *et al*. (2011) Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 Pseudomonas syringae isolates. *PLoS pathogens* **7**: e1002132.

Blom, J., Albaum, S.P., Doppmeier, D., Puhler, A., Vorholter, F.-J., Zakrzewski, M., and Goesmann, A. (2009) EDGAR: A software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* **10**: 154.

Daubin, V. and Ochman, H. (2004) Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome research* **14**: 1036–42.

Dhillon, B.K., Laird, M.R., Shay, J.A., Winsor, G.L., Lo, R., Nizam, F., *et al*. (2015) IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic acids research* **43**: W104–8.

Donati, C., Hiller, N.L., Tettelin, H., Muzzi, A., Croucher, N.J., Angiuoli, S. V, *et al*. (2010) Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. *Genome biology* **11**: R107.

Gil, R., Silva, F.J., Peretó, J., and Pereto, J. (2004) Determination of the Core of a Minimal Bacterial Gene Set. *Microbiology and Molecular Biology Reviews* **68**: 518–537.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., *et al*. (2006) Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 425–430.

Hiller, N.L., Janto, B., Hogg, J.S., Boissy, R., Yu, S., Powell, E., *et al*. (2007) Comparative Genomic Analyses of Seventeen Streptococcus pneumoniae Strains: Insights into the Pneumococcal Supragenome. *Journal of Bacteriology* **189**: 8186–8195.

Juhas, M., Eberl, L., and Glass, J.I. (2011) Essence of life: Essential genes of minimal genomes. *Trends in Cell Biology* **21**: 562–568.

Karberg, K.A., Olsen, G.J., and Davis, J.J. (2011) Similarity of genes horizontally acquired by Escherichia coli and Salmonella enterica is evidence of a supraspecies pangenome. *Proceedings of the National Academy of Sciences* **108**: 20154–20159.

Koonin, E. V. (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Reviews Microbiology* **1**: 127–136.

Krell, T., Lacal, J., Guazzaroni, M.E., Busch, A., Silva-Jiménez, H., Fillet, S., *et al*. (2012) Responses of Pseudomonas putida to toxic aromatic carbon sources. *Journal of biotechnology* **160**: 25–32.

Lan, R. and Reeves, P.R. (2000) Intraspecies variation in bacterial genomes : the need for a species genome concept. *Trends in Microbiology* **8**: 396–401.

Langille, M.G.I., Hsiao, W.W.L., and Brinkman, F.S.L. (2008) Evaluation of genomic island predictors using a comparative genomics approach. *BMC bioinformatics* **9**: 329.

Lapierre, P. and Gogarten, J.P. (2009) Estimating the size of the bacterial pan-genome. *Trends in Genetics* **25**: 107–110.

De Maayer, P., Chan, W.Y., Rubagotti, E., Venter, S.N., Toth, I.K., Birch, P.R.J., and Coutinho, T.A. (2014) Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC genomics* **15**: 404.

Mann, R. a, Smits, T.H.M., Bühlmann, A., Blom, J., Goesmann, A., Frey, J.E., *et al*. (2013) Comparative genomics of 12 strains of Erwinia amylovora identifies a pan-genome with a large conserved core. *PloS one* **8**: e55644.

Mao, X., Cai, T., Olyarchuk, J.G., and Wei, L. (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**: 3787–3793.

Médigue, C., Rouxel, T., Vigier, P., Hénaut, a, and Danchin, a (1991) Evidence for horizontal gene transfer in Escherichia coli speciation. *Journal of molecular biology* **222**: 851–6.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005) The microbial pan-genome. *Current opinion in genetics & development* **15**: 589–94.

Mira, A., Martín-Cuadrado, A.B., D'Auria, G., and Rodríguez-Valera, F. (2010) The bacterial pan-genome: A new paradigm in microbiology. *International Microbiology* **13**: 45–57.

Molina, L., Udaondo, Z., Duque, E., Fernández, M., Molina-Santiago, C., Roca, A., *et al*. (2014) Antibiotic resistance determinants in a *Pseudomonas putida* strain isolated from a hospital. *PloS one* **9**: e81604.

Mongodin, E.F., Casjens, S.R., Bruno, J.F., Xu, Y., Drabek, E.F., Riley, D.R., *et al*. (2013) Inter- and intra-specific pan-genomes of Borrelia burgdorferi sensu lato: genome stability and adaptive radiation. *BMC genomics* **14**: 693.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35** : W182–W185.

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Fouts, D.E., *et al*. (2002) Complete genome sequence and comparative analysis of the metabolically versatile. **4**: 799–808.

Okazaki, M., Sugita, T., Shimizu, M., Ohode, Y., Iwamoto, K., de Vrind-de Jong, E.W., *et al*. (1997) Partial purification and characterization of manganese-oxidizing factors of *Pseudomonas fluorescens* GB-1. *Applied and environmental microbiology* **63**: 4793–4799.

Phoenix, P., Keane, A., Patel, A., Bergeron, H., Ghoshal, S., and Lau, P.C.K. (2003) Characterization of a new solvent-responsive gene locus in *Pseudomonas putida* F1 and its functionalization as a versatile biosensor. *Environmental microbiology* **5**: 1309–1327.

Poblete-Castro, I., Becker, J., Dohnt, K., dos Santos, V.M., and Wittmann, C. (2012) Industrial biotechnology of Pseudomonas putida and related species. *Applied microbiology and biotechnology* **93**: 2279–2290.

Ramos-González, M.I., Godoy, P., Alaminos, M., Ben-Bassat, A., and Ramos, J.L. (2001) Physiological characterization of Pseudomonas putida DOT-T1E tolerance to p-hydroxybenzoate. *Applied and environmental microbiology* **67**: 4338–4341.

Rasko, D.A., Rosovitz, M.J., Myers, G.S.A., Mongodin, E.F., Fricke, W.F., Gajer, P., *et al*. (2008) The Pangenome Structure of Escherichia coli: Comparative Genomic Analysis of E. coli Commensal and Pathogenic Isolates. *Journal of Bacteriology* **190**: 6881–6893.

Remm, M., Storm, C.E. V, and Sonnhammer, E.L.L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**: 1041–1052.

Roca, A., Pizarro-Tobías, P., Udaondo, Z., Fernández, M., Matilla, M. a, Molina-Henares, M.A., *et al*. (2013) Analysis of the plant growth-promoting properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1. *Environmental microbiology* **15**: 780–794.

Segerman, B. (2012) The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Frontiers in Cellular and Infection Microbiology* **2**: 1–8.

Silby, M.W., Winstanley, C., Godfrey, S. a C., Levy, S.B., and Jackson, R.W. (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS microbiology reviews* **35**: 652–80.

Stanke, M. and Waack, S. (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**: 215–225.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one* **6**: e21800.

Taghavi, S., Barac, T., Greenberg, B., Borremans, B., Vangronsveld, J., and van der Lelie, D. (2005) Horizontal gene transfer to endogenous endophytic bacteria from poplar improves phytoremediation of toluene. *Applied and environmental microbiology* **71**: 8500–8505.

Tao, F., Tang, H., Gai, Z., Su, F., Wang, X., He, X., and Xu, P. (2011a) Genome sequence of *Pseudomonas putida* Idaho, a unique organic-solvent-tolerant bacterium. *Journal of bacteriologyacteriology* **193**: 7011–7012.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., *et al*. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *PNAS* **102**: 13950–13955.

The Uniprot Consortium (2014) UniProt: a hub for protein information. *Nucleic Acids Research* **43**: D204–D212.

Udaondo, Z., Molina, L., Daniels, C., Gómez, M.J., Molina-Henares, M.A., Matilla, M. a, *et al*. (2013) Metabolic potential of the organic-solvent tolerant *Pseudomonas putida* DOT-T1E deduced from its annotated genome. *Microbial biotechnology* **6**: 598–611.

Udaondo, Z., Molina, L., Segura, A., Duque, E., and Ramos, J.L. (2015) Analysis of the core genome and pangenome of Pseudomonas putida. *Environmental Microbiology DOI: 10.1111/1462-2920.13015*.

Walker, P.J., Dietzgen, R.G., Joubert, D.A., and Blasdell, K.R. (2011) Rhabdovirus accessory genes. *Virus Res* **162**: 110–125.

Weinstein, J. and Von Graevenitz, A. (1971) Pathogenic significance of Pseuodomonas fluorescens and Pseudomonas putida. *Yale journal of biology and medicine* **44**: 265–273.

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

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., *et al*. (2006) WEGO: a web tool for plotting GO annotations. *Nucleic acids research* **34**: W293–7.

Yoshino, Y., Kitazawa, T., Kamimura, M., Tatsuno, K., Ota, Y., and Yotsuyanagi, H. (2011) *Pseudomonas putida* bacteremia in adult patients: five case reports and a review of the literature. *Journal of Infection and Chemotherapy* **17**: 278–282.

Young, J.P.W., Crossman, L.C., Johnston, A.W.B., Thomson, N.R., Ghazoui, Z.F., Hull, K.H., *et al*. (2006) The genome of Rhizobium leguminosarum has recognizable core and accessory components. *Genome biology* **7**: R34.

Yu, H., Tang, H., Wang, L., Yao, Y., Wu, G., and Xu, P. (2011) Complete genome sequence of the nicotine-degrading *Pseudomonas putida* strain S16. *Journal of bacteriology* **193**: 5541–5542.

GENERAL DISCUSSION

Bacteria represent the largest reservoir of gene functions (Alcaraz, 2014). Recent studies have revealed that in each individual bacterial genome there is a set of niche specific adaptive genes, that code for traits which endow each of the strains with their distinctive characteristics. In addition there are a set of genes that encode features that define their overall metabolism and characterize them as a member of the specific genus or species, this set of genes is known as a 'core genome'. This Ph.D has dealt with *Pseudomonas putida* strains, a group of cosmopolitan microbes that colonize freshwaters (Geszvain and Tebo, 2010), nutrient poor and rich soils (Nelson *et al*., 2002; Matilla *et al*., 2011; Wu *et al*., 2010; Pizarro-Tobías *et al*., 2014) and which are often associated with plant tissues (Wu *et al*., 2010). *Pseudomonas putida* has seldom been associated with humans, although they have been found on the skin of immunocompromised patients (Molina *et al*., 2013) and occasionally in the lungs of cystic fibrosis patients (Sauer and Camper, 2001; Søren Molin, personal communication). The main objective of this Ph.D research was to determine the *Pseudomonas putida* pangenome (Udaondo *et al*., 2015), and to analyze it to identify the genetic repertoire of this species. To this end, the genome sequences of nine *P. putida* strains isolated from different niches and countries and with different specific traits were compiled. Three of the strains were previously characterized in our lab in Granada, Spain; namely *P. putida* BIRD1 — a plant growth promoting rhizobacterium — (Matilla *et al*., 2011), *P. putida* DOT-T1E — a solvent tolerant strain considered an extremophile — (Mosqueda *et al*., 1999) and *P. putida* HB3267 — a strain isolated from an in-patient at the Hospital of Besançon, France (Molina *et al*., 2014). The genomes of these strains were automatically annotated and then these three genomes were manually re-annotated using the Pathway Tools software (Karp *et al*., 2002) and BLAST alignments, to ensure the quality of the pangenome project. As a first step, each of the genomes was analyzed to identify the set of genes that confer their main phenotypic traits; these results are presented in the first three chapters of this thesis. Then, the pangenome of the species *Pseudomonas putida* was defined using nine *P. putida* strains, by inferring the "core genes" — a set of genes shared by the whole species; and the accessory genes — a set of genes that are only present in at least two or more, but not all of the strains in the species (Udaondo *et al*., 2015). A set of genes that are only present in one strain were also identified and dubbed "unique genes".

Analysis of the *Pseudomonas putida* DOT-T1E (Udaondo *et al*., 2012, 2013) genome in chapter I, revealed the catabolic potential of this microorganism using the metabolic atlas based on international EC codes, these findings were then corroborated by Pathway Tools analysis. The possibility of using the detailed knowledge from the metabolic atlas combined with the ability of DOT-T1E strain to thrive in the presence of toxic solvents makes this strain a model microorganism in which to perform biotransformation reactions and produce value-added products at high concentration with minimal genetic manipulation (Udaondo *et al*., 2013).

Whole-cell transformation in biphasic reactors improves the productivity for reactions in which toxic substrates and/or toxic products are used/ produced (Déziel *et al*., 1999; Heipieper *et al*., 2007; Sardessai and Bhosle, 2008). Two-phase systems have several advantages, including easy recovery of the products (often in the organic phase), and the maintenance of a low concentration of toxic or inhibitory compounds in the aqueous phase (Sheldon, 2005; Tao *et al*., 2005; Prpich and Daugulis, 2007; Garikipati *et al*., 2009; Udaondo *et al*., 2012). Solvent tolerance and defense response mechanisms of cells to organic solvents have been reported and well characterized at the physiological, biochemical and genetic level (Segura *et al*., 1999; Ramos *et al*., 2002, 2015; Sardessai and Bhosle, 2002; Heipieper *et al*., 2007; Krell *et al*., 2012). The genome of DOT-T1E encodes at least 1,751 enzymatic reactions that account for the known pattern of C, N, P and S utilization by this strain. Using the information in the metabolic atlas a number of potential biotransformation reactions were deduced. Some of these reactions could be adapted to synthesize added-value products via direct transformation of substrates, either via genetic engineering blocking an existing pathway, reorganizing operons and genes, or by recruiting enzymes from other sources to achieve the desired biotransformation.

In chapter II we analyzed *Pseudomonas putida* BIRD-1 which is a plant growth promoting rhizobacterium. The rhizosphere, or "the narrow zone of soil influenced by the root system" (Dobbelaere *et al*., 2003), is a complex environment where critical interactions between roots and microorganisms take place (Molina *et al*., 2000; Uroz *et al*., 2010; Blom *et al*., 2011; Beneduzi *et al*., 2012). Roots excrete a diverse array of primary metabolites, such as amino acids, organic acids, flavonoids, glucosinolates, fatty acids, polysaccharides, and proteins (Ahmad *et al*., 2011; Li *et*

al., 2013; De la Peña and Loyola-Vargas, 2014), this changes the surrounding composition of the root and thereby influences the edaphic and biological composition of the soil (Micallef *et al*., 2009; Chaparro *et al*., 2013). In this "nutrient rich environment" microbes find food on which to thrive. It should be noted that the microbial community composition of the rhizosphere differs between plant species (Batten *et al*., 2006; Chaparro *et al*., 2013; De la Peña and Loyola-Vargas, 2014), and is also affected by plant age and type (Micallef *et al*., 2009). The plant-bacteria relationships that occur at the root-soil interface are probably opportunistic interactions, which favor the growth of some microorganisms that benefit plant health and prevent growth of harmful microbes (Bertin *et al*., 2003; Ahmad *et al*., 2011; Chaparro *et al*., 2013; Ipek *et al*., 2014). Plant growth promoting rhizobacteria (PGPR) are free-living microorganisms with beneficial effects on crop production that colonize the rhizosphere of plants (Glick, 1995; Beneduzi *et al*., 2012). PGPR either provide plants with precursors of plant-hormone favoring development or solubilize nutrients, facilitating their uptake by the plant from the environment (Beneduzi *et al*., 2012). In addition, a number of PGPR induce systemic resistance (Maurhofer *et al*., 1998; Choudhary *et al*., 2007; Verhagen *et al*., 2009), making plants resistant to pathogens. Members of the *Pseudomonas* genus have been reported to be the most effective root-colonizing bacteria (Lugtenberg *et al*., 2001; Sørensen *et al*., 2001; Zamioudis *et al*., 2013).

Table 7. PGPR properties found in *Pseudomonas putida* BIRD1 genome

Plant growth promoting properties identified in *P. putida* BIRD-1	Beneficial action to the plant
Production of pyoverdine, an iron-siderophore	Provide iron to the plant and inhibits fungal pathogens by sequestering iron
Production of organic acids	Inorganic phosphate solubilization
Production of phosphatases/phytases	Mineralization of organic phosphorous
Production of phytohormones (IAA)	Increases root growth, increases number of secondary roots and root hairs
Production of ACC deaminase	Reduces ethylene concentration in the root thus increasing root elongation

In addition to the beneficial traits that microorganisms provide to the plant, *P. putida* BIRD-1 has many genes involved in adhesion to abiotic and biotic surfaces and involved in chemotaxis to respond to amino acids exudated by the roots. Genes involved in trehalose synthesis were also found in the genome of BIRD-1 which favours survival of this strain in soils with low-moisture levels. The study presented in chapter II is an example of the combined use of the "dry lab" (bioinformatics) and "wet lab" work, in which the bioinformatic findings were corroborated by experimental assays. In this process, all of the PGPR mechanisms described in the literature were identified by comparison with information deposited in public databases, and then all of the genes required for these functions were identified in the BIRD-1 genome using sequence homology techniques as well as synteny analysis. The presence of functional domains and motifs were confirmed in the target proteins.

The results presented in chapter III are related to the strain *P. putida* HB3267, an isolate from an in-patient from the Hospital of Besançon in France (Molina *et al.*, 2014). This strain kills insects and is resistant to the majority of antibiotics used in hospitals and laboratories. These traits are not usual in a member of this species. This chapter revealed that multidrug resistance in HB3267 is due to the presence of resistant genes located in the chromosome and in plasmid pPC9. In the chromosome were identified resistance determinants for the quinolones and fluoroquinolones, cationic antimicrobial peptides, aminoglycosides, tetracyclines, β-lactams and chloramphenicol. In the case of HB3267 a "hot spot" mutation site was identified in the Quinolone Resistance Determining Regions (QRDRs) particularly in genes encoding topoisomerases, namely *gyrA*, *gyrB*, *parC* and *parE*, where the change of three amino acids (Thr83Ile and Val894Gly in GyrA protein and Ser87Leu in ParC protein) likely conferred the resistance to quinolones. The change of Thr83Ile in GyrA in combination with other changes in the sequence of amino acids of ParC have been described in other studies of QRDRs (Hirose *et al.*, 2002; Liu *et al.*, 2012; Khalil *et al.*, 2015). The β-lactamase AmpC is the main resistance determinant for β-lactams in clinical isolates of *P. aeruginosa* (Berrazeg *et al.*, 2015). The AmpC protein of strain HB3267 varies in two amino acids with respect to the AmpC protein from the KT2440 strain (which is relatively sensitive to β-lactams). These two variations may explain the HB3267 resistance to these antibiotics.

A large number of plasmids of different sizes and incompatibility groups have been described in the genus *Pseudomonas*, many of which were discovered in strains of the species *putida*. Although it would be remiss to ignore the valuable genetic information contained on these plasmids, in chapter IV, we only considered chromosomal genes to define the *P. putida* pangenome. This is important in the analysis of the pangenome because genes involved in antibiotic resistance and the degradation of xenobiotic compounds are often encoded on plasmids. If these genes would have been taken into account this would have increased the number of accessory and unique genes. In terms of the pangenome of *P. putida* it would have added strength to our proposal that *P. putida* species has an open pangenome (Figure 4.7).

The compilation of the genetic information analyzed in this thesis is presented in chapter IV. In this chapter the core genome of the nine strains was shown to be comprised of 3,326 clusters of homologous genes, which correspond to 30,497 genes that represent 64% of all the genes considered in the study. The accessory genes and the unique genes represented 27% and 9% of total genes, respectively.

Table 8. Number of clusters of core genes

	Protein Number	Cluster Number	Core genes	Unique genes	Percent of core genes	Percent of accessory genes	Percent of unique genes
BIRD1	4,960	4,726	3,380	234	68%	27%	5%
DOT-T1E	5,721	5,088	3,398	633	59%	30%	11%
F1	5,252	5,050	3,407	202	65%	31%	4%
GB1	5,409	4,839	3,383	570	63%	26%	11%
HB3267	5,195	4,839	3,380	356	65%	28%	7%
Idaho	5,713	4,796	3,380	917	59%	25%	16%
KT2440	5,350	4,979	3,386	371	63%	30%	7%
S16	5,218	4,920	3,386	298	65%	24%	11%
W619	5,182	5,644	3,397	538	66%	24%	10%
	∑48,000	∑44,881	∑30,497	∑4,119	$\bar{x}$ 64%	$\bar{x}$ 27%	$\bar{x}$ 9%

The number of clusters of core genes is 3,326. Genomes with more than 3,326 genes in their core genome have the remaining genes as paralog genes. For example, *P. putida* BIRD1 has 3,380 core genes within the 3,326 core clusters, that means that BIRD1 strain has 53 paralog genes in its core genome ∑=summation; $\bar{x}$=average.

A critical step in the comparison of genomes is the definition of the homology relationship between genes belonging to different species or strains. To infer whether two genes descend from a common ancestral sequence we chose the bidirectional best hit approach, that is, the output hits are the sequences that score highest with each other in the alignment. The cutoff point was 80% sequence similarity and an overlap cutoff of 50% to avoid short domain-level matches. The bidirectional best hit (BBH) is a frequent approach for finding ortholog sequences between genomes; however it does not take into account the duplication events that can occur after the speciation event (these are defined as paralogs). To overcome this we used the Perl scripts *Inparanoid* (O'Brien, 2004) and *Multiparanoid* (Alexeyenko *et al*., 2006) which detect BBHs between a pair of organisms and then apply additional statistical rules to add in-paralog sequences that have arisen from duplication after speciation.

The average gene number of the analyzed *P. putida* genomes is 3,388 core genes, 1,487 accessory genes and 458 unique genes. The core genome of the species *P. putida* is, therefore, the largest set of genes. This is also the case for the pangenome of other species such as *Aggregatibacter actinomycetemcomitans* (Kittichotirat *et al*., 2011), *Procholorococcus* (Kettler *et al*., 2007) and *Streptococcus agalactiae* (Tettelin *et al*., 2005). The number of genes in the core genome depends on the number of genomes in the sample (Tettelin *et al*., 2005; Kittichotirat *et al*., 2011; Collins and Higgs, 2012), and it often decreases when new genomes are added (Kittichotirat *et al*., 2011; Collins and Higgs, 2012). In *P. putida* the comparison of two genomes brought a first relevant cut in the set of common genes (at n=1.61, figure 4.6) which then decreased slowly as the number of genomes is added. This is not unusual as the same occurs in other bacterial species, e.g. *Procholorococcus* (Kettler *et al*., 2007), *Haemophilus influenza* (Hogg *et al*., 2007), *Streptocuccus agalactiae* (Tettelin *et al*., 2005), and several other species of *Streptococcus* (Lefébure and Stanhope, 2007).

The extended core of shared genes in *P. putida* represents about 63 % of all the genes in the genome and includes the backbone of essential metabolic reactions. The most important implication of orthology is that ortholog genes have retained the same ancestral function (Wolf and Koonin, 2012), because they are all derived from a single ancestral gene in the last common ancestor of the compared species

(Koonin, 2005). Clusters of core genes consist of at least nine genes, one of each strain, but there are sets of ortholog genes in the core genome with more than one gene from a strain. Indeed core genes are clustered in 3,326 sets of genes, but the strains with the lowest number of core genes are BIRD1, HB3267 and Idaho with 3,380 core genes. Differences shown between the numbers of core genes of the nine strains and with the number of core clusters are due to the presence of paralog genes in a given genome. Families of paralog genes comprise a significant portion of prokaryotic gene sets (Borodovsky *et al*., 1995; Jordan *et al*., 2001; Wolf and Koonin, 2012) and reflect the functional diversification from duplication of genes of the microorganism. For example signaling protein families are often part of large paralog families that expand through duplication and divergence (Alm *et al*., 2006; Capra *et al*., 2012). As paralog genes diverge more quickly than ortholog genes, the emergence of a great number of paralog genes in a pangenome analysis could be an indicator of cell response to variable environmental conditions, contributing to fitness through dosage effects (Kondrashov *et al*., 2002). As such, these families of paralog genes seem likely to contribute substantially to the genomic determinants of phenotypic differences between bacterial lineages (Jordan *et al*., 2001). Against this background, paralog genes in a pangenome would be more highly represented in the accessory genome, where the selective pressure on genes is lower and where genes bear the phenotypic differences between bacterial lineages. That is the case for the *Pseudomonas putida* pangenome. The study of paralog genes therefore, could contribute to the quantification of genomic differentiation among strains and species.

Comparison of the distribution of the functional information associated with the accessory and the core genes was analyzed by assigning KO terms from KAAS (KEGG Automatic Annotation Server) (Moriya *et al*., 2007) using the BBH (Figure 5.2) and GO terms (The Gene Ontology Consortium, 2010). We found that 61% of the core genes were functionally annotated with KO terms and 78% were annotated with GO terms. We found that 40% of the accessory genes were also annotated with KO terms and 45% with GO terms. As expected, the results show that genes required for basic metabolic functions and housekeeping are overrepresented in the core genome in comparison to the accessory genome (Figure 5.2). GO terms associated with cell division, cell motility, cellular component disassembly, reproduction of single cell organism and cofactors transporter are significantly

overrepresented in core genome genes. On the other hand, genes involved in regulation of cellular and metabolic processes and response to chemical stimuli are well represented among core genes. The only GO terms better represented in the accessory genes, than in the core genes (in a statistically significant way), are those associated with transposases and electron carriers. KO terms associated with microbial metabolism in diverse environments and the biosynthesis of secondary metabolites are the most highly represented in the accessory genes (Figure 5.1).

In Chapter IV a distinctive base composition (i.e. G+C % and in codon usage) was notable when accessory and unique genes and core genes were compared globally (Figure 4.4). The structural changes in DNA composition are indicative of genes acquired from phylogenetically distant sources by horizontal gene transfer (HGT) (Médigue *et al.*, 1991; Daubin and Ochman, 2004; Ochman *et al.*, 2005). Further analysis in chapter V confirmed that unique genes are often grouped in gene islands (Figure 5.4). In spite of the fact that most of the unique genes in the *P. putida* pangenome are annotated as hypothetical or uncharacterized proteins (49 % of the unique sequences), the annotation with GO terms revealed unique genes harboring key functions that contribute to specific behavior of the individual microorganism as well as genes involved in chemotaxis and nitrogen compound metabolism in BIRD1 strain. This analysis also revealed genes involved in cellular aromatic compound metabolism, polyamine transport and the cellular response to starvation in the DOT-T1E strain and genes required for nitrogen compounds metabolism, siderophore transport, detoxification and cytolysis in HB3267 strain.

Taken together, the results in this thesis have contributed to a better understanding of the genomic composition and metabolism of the species *Pseudomonas putida*. Comparison analysis of the nine strains showed a high similarity of the proteomes, in which, 74-89 % of the proteins were shared between the strains. Pangenomic analysis revealed that the *P. putida* species has a large core set of genes which represent the metabolism of the species and a set of accessory and unique genes which provide genetic variability to the strains. From the functional annotation and the metabolic reconstruction of the core set of genes it can be concluded that all *P. putida* strains share an important number of efflux pumps, mainly ABC transporters; 1,191 genes were shown to be ABC transporter genes. Of those, 70 % were shown to encode import transport system, which implies a rich

versatility for the species to take up compounds from the environment. Besides being a central part of the nutrient uptake system, ABC transporters, also have a key role in the extrusion of toxic compounds. Another important category was the metabolism of arginine, proline and ornithine which are pivotal precursors of several important products, such as, polyamines, siderophores and some antibiotics, being also a source of energy in the urea cycle, in which arginine deiminase (ADI) plays a role in the absence of oxygen (Liu and Pilone, 1998; Tonon and Lonvaud-Funel, 2000). The analysis of the accessory and unique genes unveiled that in spite of the abundance of poorly annotated and hypothetical proteins, relevant phenotypic traits of the species are encoded in these sets of genes. *Pseudomonas putida* is a microorganism with a high biotechnological potential, its rich intrinsic pathway repertoire enables this species to degrade and metabolize a broad range of compounds (Poblete-Castro *et al*., 2012). The presence of multiple genes that encode proteins that catalyze the same reaction and the coexistence of alternative pathways for the utilization of some metabolites emphasize the ubiquitous nature of the bacterium and speak of its flexibility in adapting to the different niches in which it resides. The definition of the pangenome of *Pseudomonas putida* in this thesis work has provided a foundation for future studies. Not only gene diversity plays a role in the cosmopolitan life of *P. putida*, but also the control of the expression of core and accessory genes, as well as variations linked to gene duplications (gene dosage) (Gevers *et al*., 2004; Kondrashov, 2012). The gene expression layer adds complexity to the mechanisms that have to be taken into account to explain the specific phenotypes of the different strains.

References

Ahmad, F., Husain, F., and Ahmad, I. (2011) Rhizosphere and root colonization by bacterial inoculants and their monitoring methods: A critical area in PGPR research. In, Ahmad,I., Ahmad,F., and Pichtel,J. (eds), *Microbes and Microbial technology*. Springer New York, New York, NY, pp. 363–391.

Alcaraz, L.D. (2014) Pan-genomics: Unmasking the gene diversity hidden in the bacteria species. *PeerJ PrePrints* **2**: e113v2.

Alexeyenko, A., Tamas, I., Liu, G., and Sonnhammer, E.L.L. (2006) Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* **22**: e9–e15.

Alm, E., Huang, K., and Arkin, A. (2006) The Evolution of Two-Component Systems in Bacteria Reveals Different Strategies for Niche Adaptation. *PLoS Computational Biology* **2**: e143.

Batten, K.M., Scow, K.M., Davies, K.F., and Harrison, S.P. (2006) Two Invasive Plants Alter Soil Microbial Community Composition in Serpentine Grasslands. *Biological Invasions* **8**: 217–230.

Beneduzi, A., Ambrosini, A., and Passaglia, L.M.P. (2012) Plant growth-promoting rhizobacteria (PGPR): Their potential as antagonists and biocontrol agents. *Genetics and molecular biology* **35**: 1044–1051.

Berrazeg, M., Jeannot, K., Ntsogo Enguéné, V.Y., Broutin, I., Loeffert, S., Fournier, D., and Plésiat, P. (2015) Mutations in β-Lactamase AmpC Increase Resistance of *Pseudomonas aeruginosa* Isolates to Antipseudomonal Cephalosporins. *Antimicrobial Agents and Chemotherapy* **59**: 6248–6255.

Bertin, C., Yang, X., and Weston, L.A. (2003) The role of root exudates and allelochemicals in the rhizosphere. *Plant and Soil* **256**: 67–83.

Blom, D., Fabbri, C., Connor, E.C., Schiestl, F.P., Klauser, D.R., Boller, T., *et al*. (2011) Production of plant growth modulating volatiles is widespread among rhizosphere bacteria and strongly depends on culture conditions. *Environmental microbiology* **13**: 3047–58.

Borodovsky, M.Y., McIninch, J., Koonin, E. V, Rudd, K.E., Médigue, C., and Danchin, A. (1995) Detection of new genes in a bacterial genome using Markov models for three gene classes. *Nucleic Acids Research* **23**: 3554–3562.

Capra, E.J., Perchuk, B.S., Skerker, J.M., and Laub, M.T. (2012) Adaptive Mutations that Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families. *Cell* **150**: 222–232.

Chaparro, J.M., Badri, D. V., Bakker, M.G., Sugiyama, A., Manter, D.K., and Vivanco, J.M. (2013) Root Exudation of Phytochemicals in Arabidopsis Follows Specific Patterns That Are Developmentally Programmed and Correlate with Soil Microbial Functions. *PLoS ONE* **8**: e55731.

Choudhary, D.K., Prakash, A., and Johri, B.N. (2007) Induced systemic resistance (ISR) in plants: Mechanism of action. *Indian Journal of Microbiology* **47**: 289–297.

Collins, R.E. and Higgs, P.G. (2012) Testing the infinitely many genes model for the evolution of the bacterial core genome and pangenome. *Molecular biology and evolution* **29**: 3413–25.

Daubin, V. and Ochman, H. (2004) Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome research* **14**: 1036–42.

Déziel, E., Comeau, Y., and Villemur, R. (1999) Two-liquid-phase bioreactors for enhanced degradation of hydrophobic/toxic compounds. *Biodegradation* **10**: 219–233.

Dobbelaere, S., Vanderleyden, J., and Okon, Y. (2003) Critical Reviews in Plant Sciences Plant Growth-Promoting Effects of Diazotrophs in the Rhizosphere Plant Growth-Promoting Effects of Diazotrophs in the Rhizosphere. *Critical Reviews in Plant Sciences, 22:2, 107-149* **22**: 107–149.

Garikipati, S.V.B.J., McIver, A.M., and Peeples, T.L. (2009) Whole-cell biocatalysis for 1-naphthol production in liquid-liquid biphasic systems. *Applied and environmental microbiology* **75**: 6545–52.

Geszvain, K. and Tebo, B.M. (2010) Identification of a two-component regulatory pathway essential for Mn(II) oxidation in *Pseudomonas putida* GB-1. *Applied and environmental microbiology* **76**: 1224–31.

Gevers, D., Vandepoele, K., Simillion, C., and Van de Peer, Y. (2004) Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends in Microbiology* **12**: 148–154.

Glick, B.R. (1995) The enhancement of plant growth by free-living bacteria. *Canadian Journal of Microbiology* **41**: 109–117.

Heipieper, H.J., Neumann, G., Cornelissen, S., and Meinhardt, F. (2007) Solvent-tolerant bacteria for biotransformations in two-phase fermentation systems. *Applied*

microbiology and biotechnology **74**: 961–73.

Hirose, K., Hashimoto, A., Tamura, K., Kawamura, Y., Ezaki, T., Sagara, H., and Watanabe, H. (2002) DNA Sequence Analysis of DNA Gyrase and DNA Topoisomerase IV Quinolone Resistance-Determining Regions of *Salmonella enterica* Serovar Typhi and Serovar Paratyphi A. *Antimicrobial Agents and Chemotherapy* **46**: 3249–3252.

Hogg, J.S., Hu, F.Z., Janto, B., Boissy, R., Hayes, J., Keefe, R., *et al*. (2007) Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome biology* **8**: R103.

Ipek, M., Pirlak, L., Esitken, A., Figen Dönmez, M., Turan, M., and Sahin, F. (2014) Plant Growth-Promoting Rhizobacteria (Pgpr) Increase Yield, Growth And Nutrition Of Strawberry Under High-Calcareous Soil Conditions. *Journal of Plant Nutrition* **37**: 990–1001.

Jordan, I.K., Makarova, K.S., Spouge, J.L., Wolf, Y.I., and Koonin, E. V. (2001) Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Research* **11**: 555–565.

Karp, P.D., Paley, S., and Romero, P. (2002) The Pathway Tools software. *Bioinformatics (Oxford, England)* **18 Suppl 1**: S225–S232.

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., *et al*. (2007) Patterns and Implications of Gene Gain and Loss in the Evolution of *Prochlorococcus*. *PLoS Genetics* **3**: e231.

Khalil, D., Becker, C.A.M., and Tardy, F. (2015) Alterations in the Quinolone Resistance Determining Regions and fluoroquinolone resistance in clinical isolates and laboratory-derived mutants of *Mycoplasma bovis* : not all genotypes may be equal. *Applied and Environmental Microbiology* AEM.03280–15.

Kittichotirat, W., Bumgarner, R.E., Asikainen, S., and Chen, C. (2011) Identification of the pangenome and its components in 14 distinct *Aggregatibacter actinomycetemcomitans* strains by comparative genomic analysis. *PloS one* **6**: e22420.

Kondrashov, F. A. (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences* **279**(1749):5048–5057.

Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I., and Koonin, E. V (2002) Selection in the

evolution of gene duplications. *Genome biology* **3(2)**: research0008.1-research0008.9.

Koonin, E. V (2005) Orthologs, paralogs, and evolutionary genomics. *Annual review of genetics* **39**: 309–338.

Krell, T., Lacal, J., Guazzaroni, M.E., Busch, A., Silva-Jiménez, H., Fillet, S., *et al*. (2012) Responses of *Pseudomonas putida* to toxic aromatic carbon sources. *Journal of biotechnology* **160**: 25–32.

De la Peña, C. and Loyola-Vargas, V.M. (2014) Biotic interactions in the rhizosphere: a diverse cooperative enterprise for plant productivity. *Plant physiology* **166**: 701–719.

Lefébure, T. and Stanhope, M.J. (2007) Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biology* **8**: R71.

Li, X., Zhang, T., Wang, X., Hua, K., Zhao, L., and Han, Z. (2013) The Composition of Root Exudates from Two Different Resistant Peanut Cultivars and Their Effects on the Growth of Soil-Borne Pathogen. *International Journal of Biological Sciences* **9**: 164–173.

Liu, S.-Q. and Pilone, G.J. (1998) A REVIEW : Arginine metabolism in wine lactic acid bacteria and its practical significance. *Journal of Applied Microbiology* **84**: 315–327.

Liu, Y.H., Kuo, S.C., Lee, Y.T., Chang, I.C.Y., Yang, S.P., Chen, T.L., and Fung, C.P. (2012) Amino acid substitutions of quinolone resistance determining regions in GyrA and ParC associated with quinolone resistance in *Acinetobacter baumannii* and *Acinetobacter* genomic species 13TU. *Journal of Microbiology, Immunology and Infection* **45**: 108–112.

Lugtenberg, B.J.J., Dekkers, L., and Bloemberg, G. V (2001) Molecular determinants of rhizosphere colonization by *Pseudomonas*. *Annual Review of Phytopathology*. **39**:461-490

Matilla, M.A., Pizarro-Tobias, P., Roca, A., Fernández, M., Duque, E., Molina, L., *et al*. (2011) Complete genome of the plant growth-promoting rhizobacterium *Pseudomonas putida* BIRD-1. *Journal of bacteriology* **193**: 1290.

Maurhofer, M., Reimmann, C., Schmidli-Sacherer, P., Heeb, S., Haas, D., and Défago, G. (1998) Salicylic Acid Biosynthetic Genes Expressed in *Pseudomonas* fluorescens Strain P3 Improve the Induction of Systemic Resistance in Tobacco Against Tobacco Necrosis Virus. *Phytopathology* **88**: 678–84.

Médigue, C., Rouxel, T., Vigier, P., Hénaut, a, and Danchin, a (1991) Evidence for

horizontal gene transfer in Escherichia coli speciation. *Journal of molecular biology* **222**: 851–6.

Micallef, S.A., Channer, S., Shiaris, M.P., and Colón-Carmona, A. (2009) Plant age and genotype impact the progression of bacterial community succession in the *Arabidopsis* rhizosphere. *Plant signaling & behavior* **4**: 777–80.

Molina, L., Bernal, P., Udaondo, Z., Segura, A., and Ramos, J.-L. (2013) Complete genome sequence of a *Pseudomonas putida* clinical isolate, strain H8234. *Genome Announcements* **1.**:

Molina, L., Ramos, C., Duque, E., Ronchel, M.C., García, J.M., Wyke, L., and Ramos, J.L. (2000) Survival of *Pseudomonas putida* KT2440 in soil and in the rhizosphere of plants under greenhouse and environmental conditions. **32**: 315–321.

Molina, L., Udaondo, Z., Duque, E., Fernández, M., Molina-Santiago, C., Roca, A., *et al*. (2014) Antibiotic resistance determinants in a *Pseudomonas putida* strain isolated from a hospital. *PloS one* **9**: e81604.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35** : W182–W185.

Mosqueda, G., Ramos-González, M.I., and Ramos, J.L. (1999) Toluene metabolism by the solvent-tolerant *Pseudomonas putida* DOT-T1 strain, and its role in solvent impermeabilization. *Gene* **232**: 69–76.

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Fouts, D.E., *et al*. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environmental Microbiology* **4**: 799–808.

O'Brien, K.P. (2004) Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Research* **33**: D476–D480.

Ochman, H., Lerat, E., and Daubin, V. (2005) Examining bacterial species under the specter of gene transfer and exchange. *Proceedings of the National Academy of Sciences of the United States of America* **102 Suppl** : 6595–9.

Pizarro-Tobías, P., Fernández, M., Niqui, J.L., Solano, J., Duque, E., Ramos, J.L., and Roca, A. (2014) Restoration of a Mediterranean forest after a fire: bioremediation and rhizoremediation field-scale trial. *Microbial biotechnology* **8**: 77–92.

Poblete-Castro, I., Becker, J., Dohnt, K., dos Santos, V.M., and Wittmann, C. (2012) Industrial biotechnology of *Pseudomonas putida* and related species. *Applied*

microbiology and biotechnology **93**: 2279–2290.

Prpich, G.P. and Daugulis, A.J. (2007) Solvent selection for enhanced bioproduction of 3-methylcatechol in a two-phase partitioning bioreactor. *Biotechnology and bioengineering* **97**: 536–43.

Ramos, J.L., Duque, E., Gallegos, M.-T., Godoy, P., Ramos-González, M.I., Rojas, A., *et al.* (2002) Mechanisms of Solvent Tolerance in Gram-Negative Bacteria. *Annual Review of Microbiology* **56**: 743–768.

Ramos, J.-L., Sol Cuenca, M., Molina-Santiago, C., Segura, A., Duque, E., Gómez-García, M.R., *et al.* (2015) Mechanisms of solvent resistance mediated by interplay of cellular factors in *Pseudomonas putida*. *FEMS microbiology reviews* **39**: 555–66.

Sardessai, Y. and Bhosle, S. (2002) Tolerance of bacteria to organic solvents. *Research in microbiology* **153**: 263–268.

Sardessai, Y.N. and Bhosle, S. (2008) Industrial potential of organic solvent tolerant bacteria. *Biotechnology progress* **20**: 655–660.

Sauer, K. and Camper, A.K. (2001) Characterization of Phenotypic Changes in *Pseudomonas putida* in Response to Surface-Associated Growth Characterization of Phenotypic Changes in *Pseudomonas putida* in Response to Surface-Associated Growth. *Society* **183**: 6579–6589.

Segura, A., Junker, F., Duque, E., Mosqueda, G., and Ramos, J.L. (1999) Multiple responses of Gram-negative bacteria to organic solvents. *Environmental Microbiology* **1**: 470–470.

Sheldon, R.A. (2005) Green solvents for sustainable organic synthesis: state of the art. *Green Chemistry* **7**: 267.

Sørensen, J., Jensen, L.E., and Nybroe, O. (2001) Soil and rhizosphere as habitats for *Pseudomonas* inoculants: new knowledge on distribution, activity and physiological state derived from micro-scale and single-cell studies. *Plant and Soil* **232**: 97–108.

Tao, Y., Bentley, W.E., and Wood, T.K. (2005) Phenol and 2-naphthol production by toluene 4-monooxygenases using an aqueous/dioctyl phthalate system. *Applied microbiology and biotechnology* **68**: 614–21.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., *et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *PNAS* **102**: 13950–13955.

The Gene Ontology Consortium (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic acids research* **38**: D331–335.

Tonon, T. and Lonvaud-Funel, a. (2000) Metabolism of arginine and its positive effect on growth and revival of *Oenococcus oeni*. *Journal of Applied Microbiology* **89**: 526–531.

Udaondo, Z., Duque, E., Fernández, M., Molina, L., de la Torre, J., Bernal, P., *et al*. (2012) Analysis of solvent tolerance in *Pseudomonas putida* DOT-T1E based on its genome sequence and a collection of mutants. *FEBS letters* **586**: 2932–8.

Udaondo, Z., Molina, L., Daniels, C., Gómez, M.J., Molina-Henares, M.A., Matilla, M. a, *et al*. (2013) Metabolic potential of the organic-solvent tolerant *Pseudomonas putida* DOT-T1E deduced from its annotated genome. *Microbial biotechnology* **6**: 598–611.

Udaondo, Z., Molina, L., Segura, A., Duque, E., and Ramos, J.L. (2015) Analysis of the core genome and pangenome of *Pseudomonas putida*. *Environmental Microbiology*. **DOI:** 10.1111/1462-2920.13015

Uroz, S., Buée, M., Murat, C., Frey-Klett, P., and Martin, F. (2010) Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environmental microbiology reports* **2**: 281–8.

Verhagen, B.W.M., Trotel-Aziz, P., Couderchet, M., Hofte, M., and Aziz, A. (2009) *Pseudomonas* spp.-induced systemic resistance to Botrytis cinerea is associated with induction and priming of defence responses in grapevine. *Journal of Experimental Botany* **61**: 249–260.

Wolf, Y.I. and Koonin, E. V. (2012) A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes. *Genome Biology and Evolution* **4**: 1286–1294.

Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J., and van der Lelie, D. (2010) Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS microbiology reviews* **35**: 299–323.

Zamioudis, C., Mastranesti, P., Dhonukshe, P., Blilou, I., and Pieterse, C.M.J. (2013) Unraveling root developmental programs initiated by beneficial *Pseudomonas* spp. bacteria. *Plant physiology* **162**: 304–318.

CONCLUSIONS/CONCLUSIONES

These are the conclusions drawn from the study of the genome sequence of several *Pseudomonas putida* strains and from the experience gained during completion of the pangenome of this species:

1. Knowledge derived from the metabolic atlas of DOT-T1E and the ability of the strain to thrive in the presence of toxic solvents, makes DOT-T1E a model microorganism to design biotransformation reactions to produce added-value chemicals at high concentrations with minimal genetic manipulations.

2. *Pseudomonas putida* BIRD1 is a Plant Growth Promoting Rhizobacterium capable of solubilizing organic phosphate through the production of phosphatases and inorganic acids, as well as to solubilize iron using pyoverdine as a siderophore. Moreover the strain is capable of synthesizing the precursor of the plant growth hormone, indolacetic acid. BIRD1 can also metabolize a wide range of compounds released by plant roots as either carbon or nitrogen sources. This set of properties enables the strain to survive and colonize the plant rhizosphere.

3. *Pseudomonas putida* HB3267 was isolated from an in-patient in a French hospital. The strain can be both an opportunistic pathogen and an insect killer. Analysis of the genome of HB3267 revealed that resistance determinants for multiple antibiotics are located chromosomally and on plasmid pPC9. *Pseudomonas putida* HB3267 resistance to antibiotics is linked to a wide set of efflux pumps (TtgABC and TtgGHI), hot spot mutations in key genes such as *gyr*A, *gyr*B, *par*C, *pho*Q, *pho*P, *amp*C, *pqq*C, and duplications that increase gene dosage (*fir*A, *pqq* genes).

4. *Pseudomonas putida* species has an open pangenome with a large core set of genes that represent the central metabolism of the species, as well as a set of accessory and unique genes that provide genetic diversity to the strains.

5. In the set of core genes of *Pseudomonas putida* those encoding ABC transporters are well represented, as are those related to the biosynthesis of amino acids and catabolism of arginine and proline, and central pathways for the catabolism of aromatic compounds.

6. Statistical tests using ANOVA demonstrated that differences in G+C content and CAI index in the three groups of the pangenome genes are significant ($p<0.001$). The different G+C content and the reduced bias of codon usage in the set of unique genes, shows that this cluster evolves at a different rate and may be expressed at different levels than core and accessory genes.

7. The set of accessory genes encodes functions related with microbial metabolism in diverse environments, the biosynthesis of secondary metabolites and the biosynthesis of antibiotics.

8. The set of unique genes, most of which are poorly annotated, include genes that encode relevant phenotypic traits; i.e., functions related with the regulation of nitrogen utilization, siderophore transport, polyamine biosynthesis and extrusion of toxic substances.

Estas son las conclusiones obtenidas del estudio de la secuencia del genoma de varias cepas de *Pseudomonas putida* y de la experiencia obtenida durante la finalización del pangenoma de esta especie:

1. El conocimiento extraído del atlas metabólico de la cepa DOT-T1E combinado con su capacidad para proliferar en la presencia de disolventes tóxicos, convierten a esta cepa en un microoganismo modelo para el diseño de reacciones de biotransformación para producir a alta concentración, compuestos de valor añadido con una manipulación genética mínima.
2. *Pseudomonas putida* BIRD-1 es una rizobacteria promotora del crecimiento de plantas capaz de solubilizar fosfato orgánico a través de la producción de fosfatasas y ácidos inorgánicos, al igual que solubilizando hierro utilizando el sideróforo pioverdina. Además esta cepa es capaz de sintetizar el precursor de la hormona vegetal, ácido indol acético. BIRD-1 es también capaz de metabolizar un amplio rango de compuestos liberados por la raíz de la planta tanto como fuente de carbono como de nitrógeno. Este conjunto de propiedades permite a esta cepa sobrevivir y colonizar la rizosfera.
3. *Pseudomonas putida* HB3267 fue aislada de un paciente en un hospital francés. Esta cepa es un patógeno oportunista que puede también matar insectos. Análisis del genoma de HB3276 revelaron que los determinantes de resistencia para múltiples antibióticos están localizados tanto en el cromosoma como en el plásmido pPC9 que alberga la cepa. La resistencia a antibióticos de *Pseudomonas putida* HB3267 está unida a un amplio conjunto de bombas de eflujo (TtgABC y TtgGHI), mutaciones en sitios "hot spot" en genes clave tales como *gyr*A, *gyr*B, *par*C, *pho*Q, *pho*P, *amp*C, *pqq*C, y duplicaciones que incrementan la dosis génica (genes *fir*A, *ppq*).
4. La especie *Pseudomonas putida* presenta un pangenoma abierto con un gran conjunto de genes del *core* que codifican el conjunto de reacciones del metabolismo central de la especie; al igual que un conjunto de genes accesorios y únicos que proporcionan diversidad genómica a las cepas.
5. En el conjunto de genes del *core* de *Pseudomonas putida,* los genes que codifican transportadores ABC están ampliamente representados, al igual que los genes relacionados con la biosíntesis de aminoácidos, genes para el

catabolismo de la arginina y la prolina y genes relacionados con el catabolismo central de compuestos aromáticos.

6. Test estadísticos de análisis de la varianza (ANOVA) demostraron que las diferencias en el contenido G+C y el índice CAI es distinto en los tres grupos de genes del pangenoma de forma estadísticamente significativa ($p<0.001$). El diferente contenido de G+C del genoma y el sesgo reducido del uso de codones en el conjunto de genes únicos, muestra que este grupo de genes evoluciona a una tasa distinta y pueden expresarse a niveles distintos con respecto a los genes accesorios y del *core*.
7. El grupo de genes accesorios codifican funciones relacionadas con el metabolismo microbiano en diferentes ambientes, la biosíntesis de metabolitos secundarios y la biosíntesis de antibióticos.
8. El grupo de genes únicos, que en su mayoría carecen de anotación funcional, incluye genes que codifican rasgos fenotípicos relevantes, como por ejemplo, funciones relacionadas con la regulación de la utilización de nitrógeno, el transporte de sideróforos, la biosíntesis de poliaminas y la extrusión de sustancias tóxicas.