Universitat de les Illes Balears

ESTUDIO MOLECULAR DE POBLACIONES DE *PSEUDOMONAS* AMBIENTALES

TESIS DOCTORAL

DAVID SÁNCHEZ BERMÚDEZ

DIRECTORA:

ELENA GARCÍA-VALDÉS PUKKITS

Departamento de Biología

Universitat de les Illes Balears

Palma de Mallorca, Septiembre 2013

Universitat de les Illes Balears

ESTUDIO MOLECULAR DE POBLACIONES DE *PSEUDOMONAS* AMBIENTALES

Tesis Doctoral presentada por David Sánchez Bermúdez para optar al título de Doctor en el programa Microbiología Ambiental y Biotecnología, de la Universitat de les Illes Balears, bajo la dirección de la Dra. Elena García-Valdés Pukkits.

V° B° Director de la Tesis

DRA. ELENA GARCÍA-VALDÉS PUKKITS

Catedrática de Universidad

Universitat de les Illes Balears

El doctorando

DAVID SÁNCHEZ BERMÚDEZ

PALMA DE MALLORCA, SEPTIEMBRE 2013

Index

Agradecimientos

Me siento aquí frente a una hoja en blanco a escribir la parte más fácil, pues tengo muy claro lo que quiero decir, y a la vez la más difícil de mi Tesis, pues no quiero dejarme a nadie en el tintero y deseo que mis agradecimientos sean justos y como se merecen. Y es que tengo tanto que agradecer… Hace ya cuatro años que empecé la Tesis, siete desde que pisé este laboratorio por primera vez, y por aquel entonces no pensé que llegaría este momento o lo veía tan lejano… Y aquí estoy, miles de experiencias después, buenas y malas, pero todas ellas trayéndome hasta este momento de final de Tesis. Momento que por una parte estaba deseando que llegase, para concluir una etapa, pero que por otra me da pena que se acabe, porque en su mayor parte lo he pasado muy bien y ha sido muy buena experiencia, no solo a nivel profesional si no también a nivel personal. Dicho esto, empecemos los agradecimientos, que la lista es larga.

En primer lugar, quiero dar las gracias a mi directora de Tesis, Elena García-Valdés, así como a Jorge Lalucat, ambos "jefes" a los que he de agradecer la oportunidad que me dieron en su momento abriéndome las puertas a este laboratorio, desde el momento que entré como alumno colaborador de mano de Jorge hasta el momento que Elena aceptó ser mi directora de Tesis. Gracias Elena por tu tiempo, pero sobre todo por haberte preocupado por mí más allá de lo profesional, lo cual ha sido muy importante para mí. Gracias Jorge por haber tenido siempre un momento para resolver mis dudas y tener siempre unas ganas contagiosas de ir más allá en el saber, así como una visión tan amplia de tantos temas. Gracias a ambos por el esfuerzo y trabajo puesto en esta Tesis y más en el sprint final.

Gracias también a los otros "jefes" del laboratorio, Rafa y Balbina, por haber tenido siempre vuestra puerta abierta para resolver cualquier duda y ayudarme en lo que os necesitase. Por supuesto, merece mención aparte otro "jefe", Toni, por su ayuda en todos los temas "bioinformáticos", gracias por haberme ayudado siempre de buen grado en todo y por tus discusiones científicas.

Also thanks to PhD. Sandra Matthijs and Mrs. Catherine Tricot from the Institut de Recherches Microbiologiques Jean-Marie Wiame (Bruselas) for agreeing to take me to their lab and enable my stay in Brussels during my Thesis. My stay in Brussels was a great experience not only professionally but also personally. I am convinced that the experience, being away from home, made me grow as a person. Thanks for making my stay easier and give me the opportunity to meet other beautiful country.

Me gustaría continuar mis agradecimientos a todas aquellas personas que durante estos años me han estando “aguantado” día tras días en el labo. En primer lugar a Marga, muchas gracias por haberme acogido y llevado de la mano desde mi época de alumno colaborador hasta el día de hoy haciéndome sentir parte de una gran familia; gracias por escucharme, preocuparte por mí, darme ánimos, ayudarme siempre con la mejor predisposición y paciencia y descubrirme “el mundo de la investigación”, contagiándome tus ganas y pasión por este trabajo, si no fuera por ti, no estaría aquí hoy. En segundo lugar a Magda, gracias por enseñarme tantas cosas durante todos estos años, por estar siempre dispuesta a escucharme, por ser mi guía con mucha paciencia y comprensión, por tus buenos consejos y ser otro hombro en el que llorar. Sé que debería extenderme más con vuestros agradecimientos, lo sé, pero también sé que no hay suficiente espacio en estas páginas para daros las suficientes gracias a ambas. Puedo decir sin lugar a dudas que sin vosotras no habría llegado hasta aquí y esta Tesis no habría sido posible, no sólo por lo que me habéis aportado profesionalmente sino también personalmente, muchísimas gracias.

Por otra parte, me gustaría agradecer sin que se enfaden por el orden, a Juana, por ser mi compi de experiencias, juntos hemos vivido el mismo camino prácticamente a la par, apoyándonos, ayudándonos, convirtiéndonos en confidentes, y lo más importante, en buenos amigos más allá de estas cuatro paredes. A Cris, mi compi de poyata, gracias por ser un hombro en el que llorar siempre que lo he necesitado, por escucharme y “adoptarme”, por ayudarme siempre, por tu paciencia, por todas esas conversaciones sobre cualquier cosa menos trabajo y todos esos momentos de risas haciendo mi día a día mucho más fácil, siempre te llevaré en mi “corazón”. A Claudia P., por enriquecerme personalmente enseñándome cosas de Colombia, por todos los momentos de risas y debates, en definitiva, por ser muy buena gente.

Me gustaría mencionar también a Arantxa, gracias, porque dos personas tan ingeniosas son mejor que una sola y así sus locuras se comprenden mutuamente y mejor, que suerte tienen en este laboratorio de habernos tenido a la vez. Gracias por todos los momentos de risas y desconexión. A Toni B. por ayudarme siempre que he necesitado que me echaras un cable, aunque me invadas mi sitio en cuanto me despisto, gracias por aportar otro cromosoma Y en este laboratorio, que si no, se me revelan todas en mi contra. He de agradecer también al “sector argentino” que aunque ya no están han sido importantes para mí: Mariana y Claudia S. Gracias a ambas también por enriquecerme enseñándome cosas de Argentina, y también por hacerme reír, se os echa de menos. A las nuevas incorporaciones, bueno ya no tan nuevas, que luego se me enfadan, pero tienen razón, pasa el tiempo y no llegaron ayer. A Bel, María y Cati, tres nuevas becarias de armas tomar con las que he podido reírme mucho, gracias por estar siempre dispuestas a aguantarme. Mucha suerte a las tres en vuestras respectivas tesis.

Puedo decir, a modo general a todos mis compañeros de trabajo, que en realidad he hecho grandes amigos, a los cuales agradezco el haber estado ahí, escuchándome, riéndonos con mis tonterías y comentarios, mostrando mucha paciencia cuando me he metido con ellos. Espero haberos hecho reír y que lo hayáis pasado tan bien conmigo como yo con vosotros, sois un equipo increíble y me siento muy afortunado de haber podido trabajar con todos vosotros, un equipo por el que a uno se le hace menos cuesta arriba ir a trabajar cada día, muchas gracias a todos, porque sin vosotros no lo habría conseguido.

Además, me gustaría agradecer a todas aquellas personas que han recorrido en algún momento mi paso por el labo; doctores: Joseph, Lady; practicums: Ana P., Andrés, Ana Cristina; post-docs: Mohamed, Farid. Todos ellos me han aportado algo en su momento, y se lo agradezco.

Y ya para terminar con la gente de mi día a día en el trabajo, me gustaría agradecer a todas aquellas personas que han colaborado de algún modo en esta Tesis. Los oficiales de laboratorio: Ángel, Xisco y Guillem, parte fundamental del funcionamiento de este laboratorio. De los "Serveis Cientificotécnics", a Trinidad y a Rosa, gracias a las dos por vuestra paciencia y buena disposición para ayudar a organizarme y pasar las muestras, facilitando mi trabajo y haciendo que llegase a terminarlo cuando era necesario. Gracias también a Toni Ramírez por su ayuda y buena disposición cuando le he necesitado como enlace con Son Dureta, y ahora Son Espases.

Por supuesto, merecen mención aparte todos mis amigos que han estado ahí, apoyándome, interesándose por mi tesis, escuchándome aunque no se enterasen de nada y estuviera siendo un pesado con mis quejas, lamentos, ideas. Gracias por levantarme el ánimo cuando lo he necesitado. A mis amigas de toda la vida del "cole": Cris, Gloria, Fanny que después de tanto tiempo, y aunque nos veamos poco, aún siguen ahí, y es que hay un vínculo muy fuerte entre nosotros a pesar de los diferentes caminos que hemos tomado. A mis amigas del "insti": Pol y Miriam, da igual el tiempo que pase o donde estemos, nuestra amistad también se mantiene en pie a lo largo del tiempo, gracias. En especial me gustaría agradecer a mis amigos de la "uni", que me han "sufrido" más, que me han apoyado y acompañado en esta aventura, estando a mi lado, impidiendo que me cayera en momentos difíciles, gracias a: Inma, Raquel, Jaime, Antonia, Cris, Aina y Sebas. Gracias chicos, sois todos los mejores, quién nos iba a decir cuando empezamos la carrera que llegaríamos hasta aquí, y cuanto me alegro de que sea así.

Por último, pero no menos importante, quiero agradecer a mi familia, a los que verdaderamente les debo el haber llegado hasta aquí. Ellos, que han estado cada día, aguantando mis cambios de humor, celebrando mis logros, perdonando mis ausencias, apoyándome, los que lo han dado todo y más para que pudiera conseguir mis objetivos. A los que les cuesta y no les importa no entender este estilo de vida, horarios, y en qué consiste mi trabajo, porque para ellos, todo eso da igual, lo primero ha sido siempre estar ahí para mí. Gracias a mis padres, mi hermano, mis abuelos, mi madrina, mi tata y mi primo favorito, gracias.

¡Muchas gracias a todos!

La elaboración de esta Tesis ha sido posible gracias a una beca predoctoral para la formación de personal investigador de la Conselleria d'Interior, Direcció General de Recerca, Desenvolupament Tecnològic i Innovació del Govern de les Illes Balears (FPI09) y del Fondo Social Europeo (ESF).

Resumen

Título de la Tesis: “Estudio molecular de poblaciones de *Pseudomonas* ambientales”

El principal objetivo de la Tesis Doctoral es el estudio en profundidad de poblaciones de *Pseudomonas* presentes en el ambiente. Para ello se ha realizado una prospección de los aislamientos de *Pseudomonas* procedentes de diversos hábitats.

Se han analizado muestras en suelos agrícolas, arenas de la zona intermareal, aguas freáticas y aguas de ríos. Los aislados de *Pseudomonas aeruginosa* procedentes de muestras ambientales, se han comparado con los aislados clínicos procedentes del Hospital Universitario Son Dureta.

Se han aplicado distintas metodologías en el análisis de estos estudios. Métodos dependientes de cultivo, basados en metodología tradicional, y métodos independientes de cultivo basados en nuevos métodos moleculares. Estos últimos incluyen: tipado de secuencias multilocus, secuenciación con cebadores específicos y análisis de DNA total procedente de muestras ambientales por clonación y pirosecuenciación con cebadores específicos. Esta última metodología se ha aplicado por primera vez en muestras ambientales de la manera realizada en esta Tesis. La genómica comparativa se ha aplicado para evaluar la diversidad intraclonal.

Se ha estimado el potencial de nuevas metodologías moleculares como la clonación, pirosecuenciación, y estudio de genomas, aplicándolo al análisis de la diversidad de las especies de *Pseudomonas*. Los datos obtenidos nos permiten tener una amplia perspectiva de estos métodos aplicados a la ecología y taxonomía del género *Pseudomonas.*

En el primer capítulo, se ha analizado la estructura y microdiversidad de 53 aislamientos de muestras ambientales y clínicas de Mallorca (España), mediante el análisis de secuencias multilocus (MLST). Patrones de multiresistencia a distintos antibióticos, solo se hallaron en aislamientos de origen clínico. El elevado número de nuevos alelos y secuencias tipo, halladas en una misma área, refleja la gran diversidad de las poblaciones de *P. aeruginosa*. A todo ello deben sumarse los índices de diversidad que también indicaron la alta diversidad de la población estudiada. Los tests de clonalidad demostraron que la recombinación juega un papel esencial en la distribución de los alelos. La secuencia tipo ST-1146 fué la única secuencia hallada en los dos tipos de muestras, tres aislamientos en muestras ambientales y un aislado clínico con distinto perfil de resistencia a antibióticos.

En el segundo capítulo, los cuatro genomas de los aislados ST-1146 fueron secuenciados y ensamblados *de novo.* Los resultados indicaron que el número de genes propios del aislado clínico (SD9) era superior al de los aislados de origen ambiental (P37, P47 y P49). Genes

relacionados con el bacteriófago Pf1 y otros relacionados con los bacteriófagos F116 y H66 solo se hallaron en SD9 pero no en las otras cepas del ST-1146 de origen ambiental. Los genes relacionados con el bacteriófago Pf1 de SD9 presentaron un elevado número de mutaciones respecto a los aislados de origen ambiental. La comparación genómica demostró que los aislados ST-1146 están estrechamente relacionados y los genes relacionados con la patogenicidad estudiados estaban conservados. El número de alelos exclusivos de SD9 fue 2,5 y 3,6 veces superior a los aislados de origen ambiental, al compararse todos ellos con los genomas de referencia de las cepas *P. aeruginosa* PAO1-UW y UCBPP-PA14.

En el tercer capítulo, el río Woluwe se tomó como hábitat modelo para el estudio de la diversidad de las especies del género *Pseudomonas*. Una muestra de agua no contaminada se analizó por métodos dependientes e independientes de cultivo. La identificación de los aislados de *Pseudomonas* se analizó por secuenciación y análisis de los cebadores del gen *rpoD.* Los métodos independientes de cultivo se basaron en la clonación y pirosecuenciación del amplicón del gen *rpoD* obtenido con los cebadores selectivos para dicho gen en *Pseudomonas*: PsEG30F-PsEG790R. Cabe destacar el elevado número de cepas de *Pseudomonas* obtenidas en las muestras por los tres métodos de análisis: 26 especies distribuidas en 13 grupos o subgrupos filogenéticos. La pirosecuenciación ha sido el mejor método de los utilizados; las secuencias obtenidas correspondieron a 24 de las especies totales observadas, con la excepción de *P. stutzeri* y *P. simiae*. El grupo filogenético predominante fue *Pseudomonas fluorescens.* En todos los métodos de análisis se halló un gran número de posibles nuevas especies indicando una enorme diversidad del género, no descrita hasta el momento.

En el cuarto capítulo, se aislaron cepas de *Pseudomonas* de muestras ambientales procedentes de suelos y zonas intermareales. En el proceso de identificación algunos de estos aislamientos no han podido ser asignados a especies conocidas de *Pseudomonas* considerándose como posibles nuevas especies. En el análisis de secuencias multilocus se incluyeron cepas procedentes de nuestra colección del laboratorio de Microbiología de la "Universitat de les Illes Baleares". El análisis de secuencias multilocus demostró que varios aislados podrían corresponder a 3 nuevas especies (6, 5 y 1 aislados de cada especie). La confirmación de estos resultados requerirá posteriores análisis. Otros cuatro aislados fueron estudiados mediante su caracterización morfológica, fisiológica, bioquímica, quimiotaxonómica y genómica. Estos estudios demostraron que los aislados no podían ser asignados a ninguna especie conocida de *Pseudomonas* por lo que se han propuesto dos cepas nuevas: *Pseudomonas aestusnigri* (cepa VGXO14^{T} = CECT 8317 ^{T} = CCUG 64165 ^{T}) y *Pseudomonas terricola* (cepa S58 ^{T} = CECT 8389 ^{T} = CCUG 64415 ^{T}).

Abstract

Thesis title: “Molecular analysis of environmental *Pseudomonas* populations”

The main goal of the present work is a thorough study of the diversity of *Pseudomonas* populations present in several habitats.

In chapter one, the population structure and microdiversity of 53 *Pseudomonas aeruginosa* isolates from environmental samples and clinical specimens obtained in Mallorca (Spain) has been analyzed by a multilocus sequence typing approach (MLST). Antibiotic multiresistance to several antibiotics was only found in isolates of clinical origin. The high number of new alleles and new sequence types found in a limited area reflects the great diversity of *P. aeruginosa* populations and the high plasticity of a paradoxically phylogenetic conserved genome of *P. aeruginosa*. The calculated genetic diversity index also demonstrated the high diversity of the population under study. Clonality tests demonstrated that recombination plays a key role in the distribution of alleles. The ST-1146 was the only one found in both kind of samples, 3 environmental isolates (from the same site isolated at 2 different dates) and 1 clinical isolate, with differences in its antibiotic susceptibility profile. For this reason, the 4 genomes of newly described sequence type ST-1146 have been sequenced and analyzed.

In the second chapter, the four genomes of ST-1146 were obtained and the sequences assembled *de novo* and compared with the CD-HIT program. Results showed that the number of isolate-specific genes was higher in the clinical isolate (SD9) than in environmental isolates (P37, P47 and P49). Some genes related to phage Pf1 and to other phages similar to bacteriophages F116 and H66 were found in isolate SD9 but not in the other isolates of ST-1146. The bacteriophage Pf1 region in isolate SD9 accumulated the highest number of mutations in comparison with the environmental isolates. Comparative genomic methods indicated that the isolates of ST-1146 are closely related, and most genes implicated in pathogenicity are highly conserved in the environmental isolates, suggesting the genetic potential for infectivity similar to that of the clinical isolate. Moreover, the four genomes were mapped against the reference genomes of *P. aeruginosa* PAO1-UW and UCBPP-PA14. A mutational profile was performed as a result of each comparison. The clinical isolate showed in both comparisons a number of exclusive alleles 2.5 and 3.6 times greater than the environmental isolates. These results suggest that the mutation pressure is not the same in the environmental isolates than in the clinical one.

In the third chapter, the River Woluwe has been taken as a model habitat for the study of the diversity of species in the genus *Pseudomonas*. A water sample from a non-contaminated

site at the source of the river was analyzed by culture-dependent and –independent methods. Identification of the *Pseudomonas* isolates was performed by sequencing and analysis of their *rpoD* sequence. Culture-independent methods consisted of a cloning and pyrosequencing of a specific *rpoD* amplicon obtained from total DNA extracted from the same sample and amplified by *Pseudomonas rpoD* primer set EGPsF340-EGPsR 780. It was remarkable the number of known species detected in the sample by the three different methods: 26 species distributed in 13 phylogenetic groups or subgroups within the genus. Pyrosequencing was the more powerful analysis; sequences obtained represented the 24 species with the exception of *P. stutzeri* and *P. simiae*. The predominant phylogenetic group within the *Pseudomonas* genus was *Pseudomonas fluorescens* group in the cultures and in the culture-independent analysis. In all analysis a high number of putative novel species were found indicating the enormous diversity not described yet.

In the fourth chapter, several *Pseudomonas* strains have been isolated from environmental samples, from soil and intertidal habitats. In the identification process, some of these strains have not been assigned to known *Pseudomonas* species and were considered members of putative novel species. In their phylogenetic characterization by MLSA we found that strains in the culture collection of our laboratory were close-related and therefore they were also included in the taxonomic characterization of these putative novel species. MLSA demonstrated that 3 putative novel species were represented by 6, 5 and 1 strains respectively, which will be the subject of additional studies. Four other strains were deeply studied by a taxonomic polyphasic approach, including morphological, physiological, biochemical, chemotaxonomic and genomic characterizations. These studies demonstrated that the four strains cannot be assigned to any of the known *Pseudomonas* species and we propose the creation of two novel species, *Pseudomonas aestusnigri* (strain VGXO14^{T} = CECT 8317 ^{T} = CCUG 64165 ^{T}) and *Pseudomonas terricola* (strain S58 ^{T} = CECT 8389 ^{T} = CCUG 64415 ^{T}).

Introduction

I.1. The genus *Pseudomonas*

I.1.1. Definition

The species in the genus *Pseudomonas* are defined as straight or slightly curved rods, 0.5-1.0 μm in diameter by 1.5-5.0 μm in length (Fig. I.1). Gram negative. They are motile by one or several polar flagella. No resting stages are known.

Pseudomonas species are aerobic, having a strictly respiratory type of metabolism with oxygen as the terminal electron acceptor; in some cases nitrate can be used as an alternative electron acceptor, allowing growth to occur anaerobically. All strains are chemoorganotrophic.

Xanthomonadins are not produced. Most of the species do not accumulate granules of poly-β-hydroxybutyrate (PHB), but accumulation of poli-hydroxialkanoates (PHA) of monomer length higher than C_4 may occur when growing on alkanes or gluconate.

Oxidase reaction is positive or negative and catalase is always positive. Strains include in their membranes hydroxylated fatty acids (3-OH 10:0 and 12:0; and 2-OH 12:0), and contain ubiquinone 9 as chemotaxonomic markers. The G+C content of the DNA is 58-69%.

They are not tolerant to acidic conditions and are not able to grow a pH lower than 4.5. They are widely distributed in nature. Some species are pathogens of animals or plants (Palleroni, 1984).

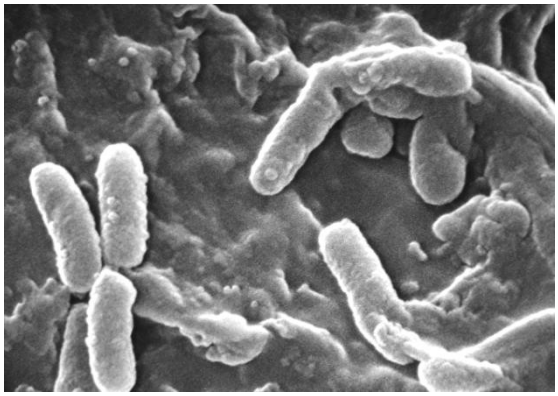


Figure I.1: Scanning electron microscopy (SEM) of *Pseudomonas aeruginosa* (picture from Centers for Disease Control and Prevention, United States Department of Health and Human Services; http://phil.cdc.gov/phil/home.asp).

I.1.2. *Pseudomonas* genus history

The genus *Pseudomonas* is one of the more diverse and its taxonomy has suffered many changes since its discovery (Palleroni, 1984). Taxonomy of *Pseudomonas*, as well as the

identification methods, has evolved together with the methodologies available. The genus *Pseudomonas* was described by Migula in 1895, according to their morphological characteristics (Migula, 1985).

For many years, the genus placed many species, not always well characterised, until the work of Stanier, Doudoroff and Palleroni (Stanier, 1966) in which the physiological and biochemical properties clearly established the taxonomical basis and identification of the species. In 1984 the genus was revised and a subdivision in 5 groups was made based on DNA-DNA hybridization and rRNA-DNA hybridization results (Palleroni, 1984). The actual members of the genus *Pseudomonas* (*sensu stricto*) belong to Palleroni's RNA group I, in the class *Gammaproteobacteria*.

From 1960, the number of new species grew up because the application of biochemical and physiological techniques in the study of bacteria taxonomy (Fig. I.2).

The introduction of molecular characterization techniques, based on genotypic characteristics changed the identification of new species in an important way (Fig. I.2). The application of techniques of DNA-DNA and DNA-rRNA hybridization, 16S rRNA or ITS gene sequencing allowed the description of new species and has redistributed some of the former *Pseudomonas* species into other genera, in the alpha, beta or gamma-classes of *Proteobacteria* (as *Acidovorax, Aminobacter, Brevundimonas, Burkholderia, Comamonas, Halomonas, Herbaspirillum, Hydrogenophaga, Marinobacter, Methylobacterium, Oligotropha, Pseudoalteromonas, Ralstonia, Sphingomonas, Stenotrophomonas, Telluria, Vogesella* and *Zavarzinia* genus) (Kersters *et al.*, 1996).

The genus *Pseudomonas* is one of the most complex bacterial genera. By now, the genus *Pseudomonas* is the genus of Gram negative bacteria with the highest number of species: 142 species and subspecies were recognized in (as stated in Euzeby's list of prokaryotic names and in the DSMZ web pages: http://www.bacterio.cict.fr and www.dsmz.de). Number of species described is growing continuously, as *Pseudomonas yangmingensis* y *Pseudomonas guariconensis* described this year (Wong *et al*., 2013; Toro *et al.*, 2013).

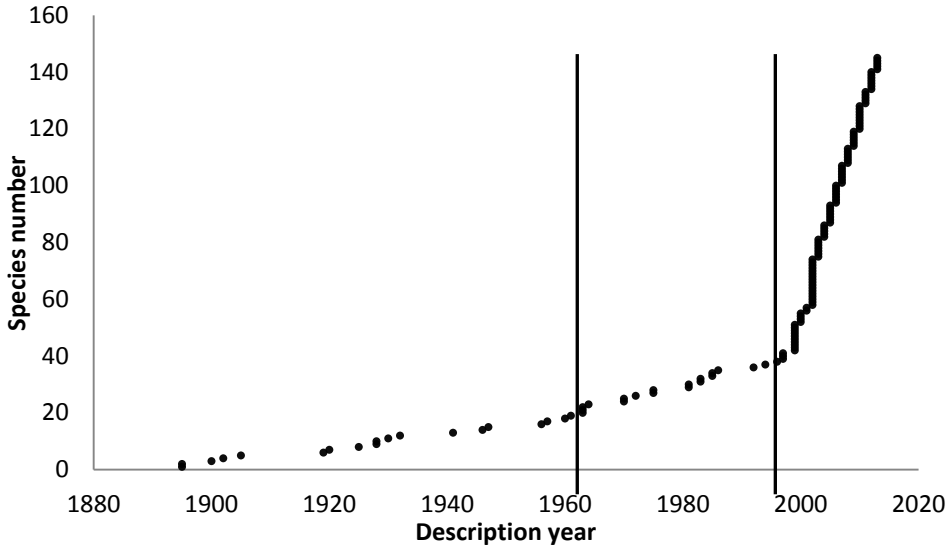


Figure I.2: Species number of the *Pseudomonas* genus along the years. The first line indicates the introduction of biochemical techniques and the second one the introduction of molecular techniques. It has been included 142 described species validated.

I.1.3. Phenotypic characteristics of *Pseudomonas*

The *Pseudomonas* genus is defined actually as straight or slightly curved rods, but the cells of *Pseudomonas* strains occasionally differ substantially in size and shape from the general definition. Thin sections of cells of *Pseudomonas* show cell walls and membrane characteristics of Gram negative bacteria (Palleroni, 1984).

Typically, *Pseudomonas* cells have polar flagella (Fig. I.3). Flagella insertion in some instances is not exactly polar but subpolar, and occasionally it may be difficult to differenciate from the so-called peritrichous type observed in members of other genera. In addition to the polar flagella, lateral flagella of short wavelength may be produced by strains of some species (*P. mendocina* and *P. stutzeri*) (Palleroni, 1984). Fimbriae (pili) of polar insertion have been reported for *P. aeruginosa* and *P. alcaligenes* in the early studies of Fuerst and Hayward (1969).

A morphological character that was used to differentiate members of the genus *Pseudomonas* from other aerobic pesudomonads was the inability of the former to accumulate endocellular granules of poly-β-hydroxybutyrate (PHB) when growing in media of low nitrogen content on various carbon sources.

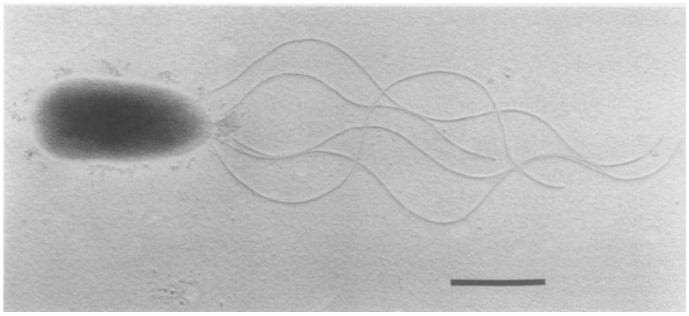


Figure I.3: Electron microscopy of *P. putida* PRS2000. Bar, 1 μm (Harwood *et al*., 1989).

Early taxonomic treatments of the genus *Pseudomonas* included pigmentation as a generic character, but this is no longer valid. In fact, the colonies and other cell masses always display some colors due to normal cellular components, but not due to the presence of pigments.

Six pigments are produced by *P. aeruginosa* and other fluorescent pseudomonads species: four phenazines (pyocyanine, pyorubin, chlororaphine, oxiphenazin), the *Pseudomonas* blue protein, and pyoverdine. Several phenazine pigments can be produced by a single strain.

Pyoverdines are also physiologically important because they function as efficient siderophores. Beside pyoverdines, the fluorescent pseudomonads also produce other strain-specific, but chemically related, siderophores called pseudobactin and pyochelin (not pigmented). Siderophores are iron-scavenging compounds. *Pseudomonas* spp. (as *P. fragi* and *P. mendocina*) do not produce siderophores in detectable amounts and are very sensitive to iron-limiting conditions.

Strains of *Pseudomonas* species can grow in minimal, chemically defined media with ammonium ions or nitrate as nitrogen source and a single organic compound as the sole carbon and energy source.

The best growth temperature of most strains is approximately 28 ^{o}C. Some species can grow at a substantial rate at 4 ^{o}C and thus can be considered psychrotrophic. For others the maximum temperature is about 45 ^{o}C, and therefore they are not true thermophiles. None of the members of the genus tolerates acidic conditions and growth is invariably negative a pH 4.5.

The metabolism of *Pseudomonas* is typically respiratory with oxygen as the terminal electron acceptor, but some species also can use nitrate as an alternate electron acceptor and can carry out oxygen-repressible denitrification (dissimilatory reduction of nitrate to N_2O or N_2). In most cases, denitrification is the property of all members of a given species, and only a few strains may be unable to denitrify.

Members of the genus *Pseudomonas* are notorious for their capacity of aerobic degradation of a number of hydrocarbons, aromatic compounds, and their derivatives, of which there are natural compounds and final products or intermediates from industrial activities.

One of the aspects that has attracted the attention of biochemist and bacteriologists for many years is the utilization of arginine by the pseudomonads. The arginine dihydrolase system has been used by taxonomist for differentiation of species, but not all strain of *Pseudomonas* has this system and is present in other genera (Palleroni, 1984).

I.1.4. Genotypic characteristics of *Pseudomonas*

The genome of *Pseudomonas* is a circle chromosome with a size of 4.5-7 Mbp. The mol% G+C content is 58-69. The larger genome is the *P. fluorescens* Pf-5 genome (7074893 bp) and the smaller the *P. stutzeri* A1501 genome (4567418 bp). *P. aeruginosa* is by far the best known from the genetic point of view. Strain PAO1 ("*Pseudomonas aeruginosa* One") is the one that has been most extensively studied.

Plasmids are important components of the genetic makeup of *Pseudomonas*. Some of them act as fertility factors, some may impart resistance to various agents (i.e. antibiotics), and others confer the capacity of degradation of unusual carbon sources, thus contributing to the nutritional versatility that is a striking feature of many members of the genus (Palleroni, 1984).

I.1.5. Ecology

The *Pseudomonas* genus is arguably one of the most fascinating and ecologically significant bacterial genera (Fig. I.4) (Guttman *et al*., 2008). *Pseudomonas* strains are environmental important bacteria in the recycling of nutrients, but also in medicine and biotechnology. Two interrelated characteristics are present in the genus *Pseudomonas*: its high number of species and its ubiquity. Consequently to this universal distribution a remarkable degree of physiological and genetic adaptability has to be expected. The genus is widely distributed in nature covering many different habitats, colonizing soils, waters plants and animals. A remarkable degree of physiological and genetic adaptability has to be expected due to such universal distribution (Mulet *et al*., 2012).

Healthy plant interactions

Many pseudomonads interact with plants and several species contribute to plant health

by antagonizing plant-pathogenic microorganisms (biocontrol) and directly influencing plant disease resistance and growth (plant growth promotion) – both as plant endophytes (Ryan *et al*., 2008) and as rhizosphere colonizers (Haas *et al*., 2005). Major secondary metabolites produced by *Pseudomonas* strains have been detected and investigated because of their antimicrobial activity (Leisinger *et al*., 1979).

Pseudomonads are also able to produce insecticides, including *P. fluorescens* Pf-5 Fit (for *P. fluorescens* insecticidal toxin) against tobacco hornworm (Pechy-Tarr *et al*., 2008); *P. entomophila* with currently unknown toxin(s) against Drosophila (Vodovar *et al*., 2005); and *P. syringae* with an unknown action against aphids (Stavrinides *et al*., 2009).

Bioremediation

Bioremediation uses microorganisms to degrade or detoxify hazardous environmental contaminants. The exceptional nutritional versatility of pseudomonads, coupled with the production of biosurfactants that can mobilize hydrocarbons and nonaqueous phase liquids into an aqueous phase (Desai *et al*., 1997), makes them excellent candidates for bioremediation. *P. aeruginosa* is frequently isolated in petroleum-contaminated soils and groundwater (Ridgway *et al*., 1990). *P. putida* is an important bio-dregadative species that is capable of eliminating some of the most deadly and challenging environmental toxins (Wackett, 2003). More strains of *P. putida* with significant degradative abilities have been isolated from the environment than any other bacterial species (Wackett, 2003). Other *Pseudomonas* species identified with bioremediation properties include *P. mendocina* (Whited *et al*., 1991) and *P. stutzeri* (Grimberg, 1996).

Pathogens

In the *Pseudomonas* genus, the relevant pathogenic species are:

P. aeruginosa, an opportunistic human pathogen which was identified by the Antimicrobial Availability Task Force of the Infectious Disease Society of America as one of the top six infectious disease threats (Tabot *et al*., 2006). It is the leading cause of mortality among Cystic Fibrosis patients, and one of the most common causes of hospital-associated pneumonia. This persistent pathogen attacks wounds, burns, eyes, the urinary tract, as well as causing general systemic infections.

P. syringae is a foliar plant pathogen that causes a variety of blights, speck, and spot diseases in many important agricultural crops, including tomato, soybeans, rice, and tobacco, to name just a few. Well over 50 different pathogenic varieties (pathovars) have been named within this complex. In addition to being a significant agricultural pathogen, this species together *P. viridiflava* are also one of the most important model systems for study of secreted virulence proteins and their role in pathogen-host interactions (Jakob *et al*., 2002).

P. entomophila was isolated from fruit flies and was subsequently found to be a pathogen of *Drosophila* (Vodovar *et al*., 2005).

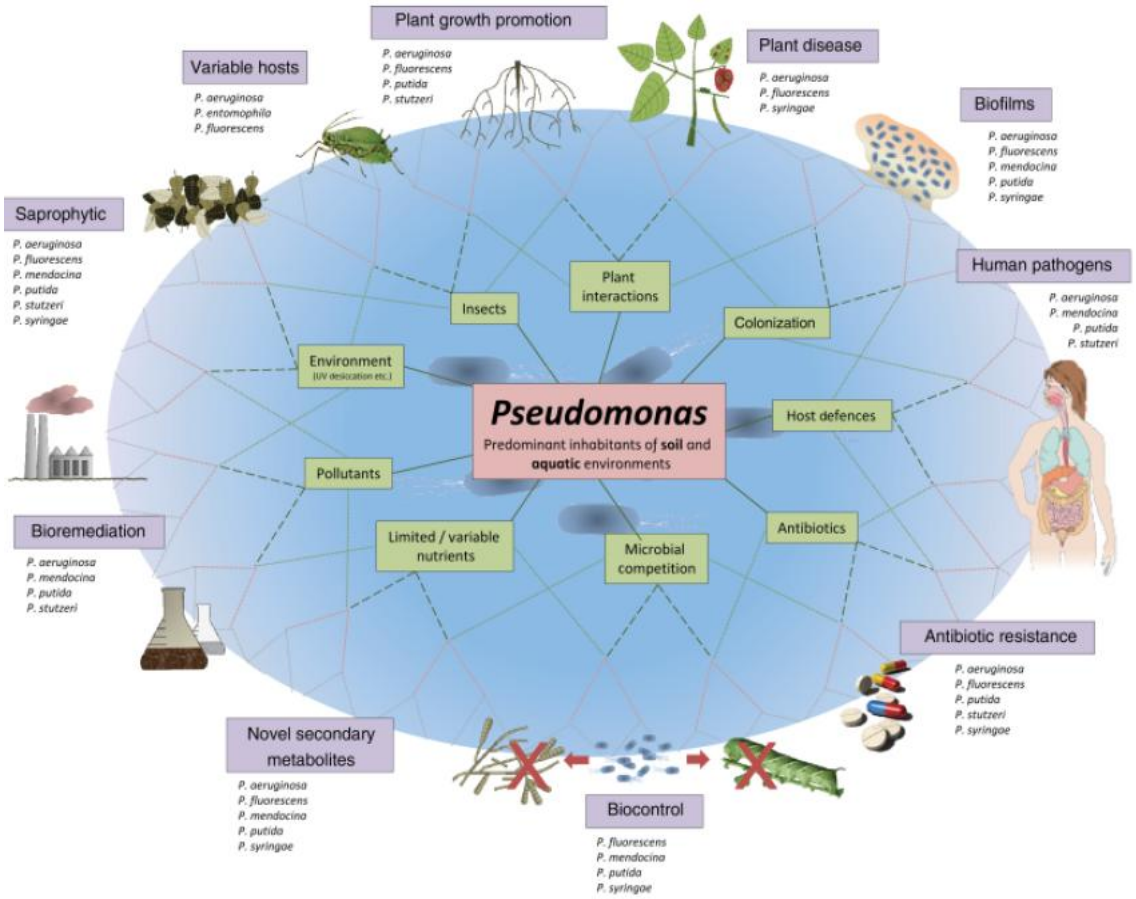


Figure I.4: General view of the ecological importance of the *Pseudomonas* genus (Silby *et al.,* 2011).

I.2. *Pseudomonas aeruginosa* species

Pseudomonas aeruginosa is a motile Gram negative rod-shaped bacterium. It differs from other members of the *Pseudomonas* genus because of its potential pathogenicity for human beings and other mammals. *P. aeruginosa* participates in infections in immunocompromised individuals, such as patients suffering from AIDS, cancer, burn wounds and cystic fibrosis (Lyczak *et al*., 2000; Willenbrock *et al*., 2006 and 2007). *P. aeruginosa* infections are normally difficult to eradicate.

P. aeruginosa is a generalist with a diverse metabolic competence. It can utilize simple small molecules as well as complex organic compounds as carbon sources and is capable of growing aerobically as well as anaerobically using NO_3 or arginine as respiratory electron acceptors.

P. aeruginosa grows over a wide range of temperatures, from common environmental temperatures up to 42 °C, with an optimum temperature at 37 °C.

Several schemes for the molecular typing of *P. aeruginosa* have been proposed to determine the relatedness of nosocomial pathogens. The MultiLocus Sequence Typing (MLST) previously described by Curran and collaborators (2004) has been used in this Thesis. Although other studies have compared numerous molecular typing schemes (Grundmann *et al*., 1995), the pulsed-field gel electrophoresis (PFGE) is still considered by most to be the "gold standard". MultiLocus Sequence typing (MLST) is a new typing technique that is becoming popular due to the ease of data analysis (Johnson *et al*., 2007). MLST is a highly discriminating, rapid, and portable DNA-based strain typing method in which regions from several housekeeping loci are sequenced from each strain (Guttman *et al*., 2008). Although some studies showed that PFGE had a higher discriminatory ability than MLST, MLST has the advantage that it gives information about the clonal relationships of isolates that PFGE does not (Johnson *et al*., 2007). The use of housekeping genes also focuses de dataset on the core genome, and consequently, is more likely to represent the clonal history of the species. MLST is rapidly becoming the gold standard for strain typing and it is the only method that permits precise phylogenetic analyses (Guttman *et al*., 2008). In this Thesis, seven housekeeping genes were analysed by MLST: *acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA*, and *trpE* gene (Table I.1) (Curran *et al*., 2004).

Table I.1. The seven housekeeping genes analysed in the MLST study.

Locus	Putative function of gene (strain)	Position in PAO1 genome (base pair)
acsA	Acetyl coenzyme A synthetase (PA0887)	969670
aroE	Shikimate dehydrogenase (PA0025)	26711
guaA	GMP synthase (PA3769)	4227237
mutL	DNA mismatch repair protein (PA4946)	5551681
nuoD	NADH dehydrogenase I chain C, D (PA2639)	2983963
ppsA	Phosphoenolpyruvate synthase (PA1770)	1914037
trpE	Anthralite synthetase component I (PA0609)	670980

(Curran *et al*., 2004).

The first completed genome of *P. aeruginosa* is the genome of PAO1, published in 2000 (Stover *et al*., 2000) (Figure I.5). It is 6.3 Mbp and contains 5570 open reading frames. A significant number (8.4%) of *P. aeruginosa* genes are predicted to be involved in regulation, which at the time of publication of the genome was the largest fraction of regulators among sequenced bacterial genomes. The large genome size and remarkable gene complexity of *P. aeruginosa* enable it to adapt and thrive in different environmental conditions and hosts. In addition to PAO1, UCBPP-PA14 as well as nine other clinical *P. aeruginosa* strains are currently available as completed genomes or as unfinished sequences.

Genome comparison analysis of PAO1 and five different clinical *P. aeruginosa* isolates (four of them are cystic fibrosis isolates) shows that about 80% of the PAO1 genome is conserved among other strains (Spencer *et al*., 2003). More than 30 relatively large regions of the PAO1 genome are found to be non conserved in the five clinical strains. Those regions include phages, the pyoverdine biosynthesis locus, genes encoding a putative type I secretion system and a putative restriction modification system.

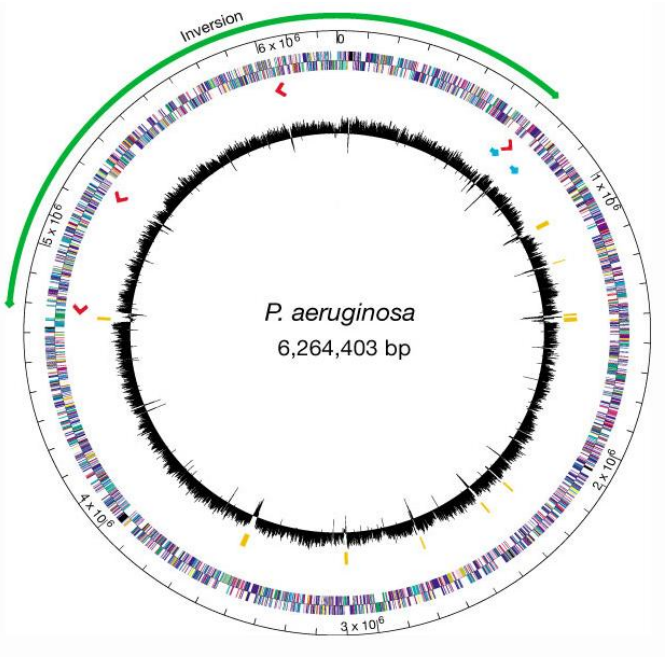


Figure I.5: Circular representation of the *P. aeruginosa* genome (Stover *et al*., 2000).

P. aeruginosa is an opportunistic pathogen capable of producing a wide variety of virulence factors, including lipopolysaccharides, flagellum, type IV pili, proteases, exotoxins, pyocyanin, exopolysaccharides, type III secretion, etc. Many of the extracellular virulence factors have been shown to be regulated by quorum-sensing signals (Brint *et al*., 1995; Pearson *et al*., 1997). Biofilm formation can also be regarded as a virulence factor (Van Alst *et al*., 2007). Acute infections with *P. aeruginosa* can be life threatening, resulting in severe tissue damage and septicemia (i.e. bloodstream invasion). Although the pathogenicity of *P. aeruginosa* is typically characterized by a high level of toxin production (Furukawa *et al*., 2006), several important exceptions to this are known, particularly in the cases of certain chronic infections of *P. aeruginosa*. These include infections of cystic fibrosis patients, where *P. aeruginosa* develops genetic adaptations during long-term persistence, in which virulence factors are normally selected against (Jelsbak *et al*., 2007; Lee *et al*., 2005; Smith *et al*., 2006). This shows *P. aeruginosa* is capable of choosing distinct strategies for different types of infections.

Another remarkable and unfortunate character of *P. aeruginosa* is its tolerance to many antimicrobial drugs. It has a number of inherent antibiotic-resistance mechanisms that include an AmpC β-lactamase that can be induced by β-lactams, which makes it inherently resistant to cephalothin and ampicillin (Livermore, 1995). *P. aeruginosa* also has efflux pumps such as MexAB–OprM, making it impermeable to many antibiotics. MexAB–OprM removes β-lactams, chloramphenicol, fluoroquinolones, novobiocin, as well as various dyes and detergents (Poole, 2001). Finally, it has been shown that through mutation, *P. aeruginosa* is capable of developing resistance to antibiotics that the strain is not inherently resistant to, such as aminoglycosides and colistin (Livermore, 2002).

I.3. Taxonomy of the genus *Pseudomonas*

Previous work in our laboratory with members of the species *P. stutzeri* (García-Valdés *et al.,* 2003; Cladera *et al.*, 2004, 2006a, 2006b; Mulet *et al.,* 2008) permitted us to generate the tools needed to extend the study to the other *Pseudomonas* species, as are the appropriate selection of genes (Cladera *et al.*, 2004; Mulet *et al.*, 2008), the improving of some PCR protocols and primers used (Mulet *et al.*, 2009), and the creation of a specific database, PseudoMLSA, now available, to compilate all these gene-sequences for the characterization and taxonomical identification of *Pseudomonas* strains (http://www.uib.es/microbiologiaBD/Welcome.php) (Bennasar *et al*., 2010).

Several genes have been used in this work to delineate the phylogenetic status of species in the genus *Pseudomonas*: the 16S rDNA was included, because as a universal marker, it permits the adscription of a strain in the genus and allows comparisons between very divergent bacteria (Santos et al., 2004) although it has been demonstrated that its resolution at intrageneric level was low (Moore *et al.*, 1996; Anzai *et al*., 2000; Yamamoto *et al.*, 2000).

The *gyrB* gene (Fig. I.6a), encodes the beta subunit of the gyrase (EC 5.99.1.3), responsible for the negative super coiling of the DNA, and *rpoD* (Fig. I.6b) is the gene encoding the sigma 70 subunit of the RNA polymerase (EC 2.7.7.6). Both genes have been used by Yamamoto and collaborators initially for the phylogenetic characterization of *Pseudomonas putida* strains, and later for 31 species of the *Pseudomonas* genus, establishing different complexes in it (Yamamoto et al., 1998 and 2000).

The *rpoB* gene (Fig. I.6c), encoding the beta-subunit of the RNA polymerase (EC 2.7.7.6),

has been postulated as a good candidate for phylogenetic analysis and identification of bacteria for clinical microbiologists (Adékambi *et al.*, 2009). In the *Pseudomonas* genus, this gene has been used by Tayeb and collaborators (2005), but also in some other organisms, like *Brevundimonas*, *Ralstonia*, *Comamonas*, or *Burkholderia* (Tayeb *et al.*, 2008), many of them former members of the genus *Pseudomonas* (*sensu lato*).

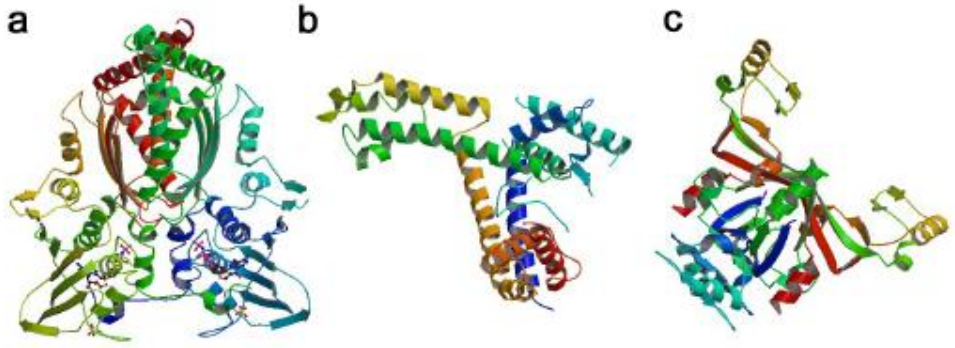


Figura I.6: a) gyrase B dimerization from *E. coli* (Brino *et al.*, 2000); b) Cristal structure of the *sigma* 70 subunit of the RNA polymerase from *E. coli* (Malhotra *et al.*, 1996); c) Cristal structure of the beta-subunit of the RNA polymerase from *Vibrio cholerae* (Zhang *et al.*, 2009). PDB (*Protein Data Bank*) ID: a) 1EI1 for the beta subunit of the gyrase, b) 1SIG for the *sigma* 70 subunit of the RNA polymerase, and c) 3E7H for the beta-subunit of the RNA polymerase (Kouranov *et al.*, 2006).

Moore and collaborators (1996) and Anzai and collaborators (2000), published their studies on the phylogeny of *Pseudomonas* based only on the analysis of the 16S rDNA. Later, Yamamoto and collaborators incorporated the use of the *gyrB* and *rpoD* genes, and 23 taxa were analysed phylogenetically (Yamamoto *et al*., 2000). Since the work of Hilario and collaborators (2004), in which *atpD*, *carA* and *recA* genes were incorporated into the analysis of 13 type strains of *Pseudomonas* (together with other reference strains), or the publication on the *rpoB* sequences of Tayeb and collaborators (2005) including 75 type strains, there has not been any review of the status of the *Pseudomonas* genus and on the phylogenetic relationships between their species based on DNA sequencing of representative genes, and only few considered the combined phylogenetic analysis of several genes in some species (Frapolli *et al.,* 2007; Kiewitz *et al*., 2000).

Partial sequences of the 16S rRNA, *gyrB*, *rpoB*, and *rpoD* genes of 107 *Pseudomonas* strains were analyzed by Mulet and collaborators (2010). This work demonstrated that the concatenated analysis of 3 genes (16S rRNA, *gyrB* and *rpoD*) was enough for the phylogenetic analysis of the genus. The inclusion of *rpoB* may be necessary in some cases, but it does not improve the resolution in discriminating the type strains.

Individual gene trees, as well as the concatenated sequences and a consensus analysis,

allowed the discrimination of two intrageneric groups (IG) or lineages, called *P. fluorescens* lineage and *P. aeruginosa* lineage (Fig. I.7).

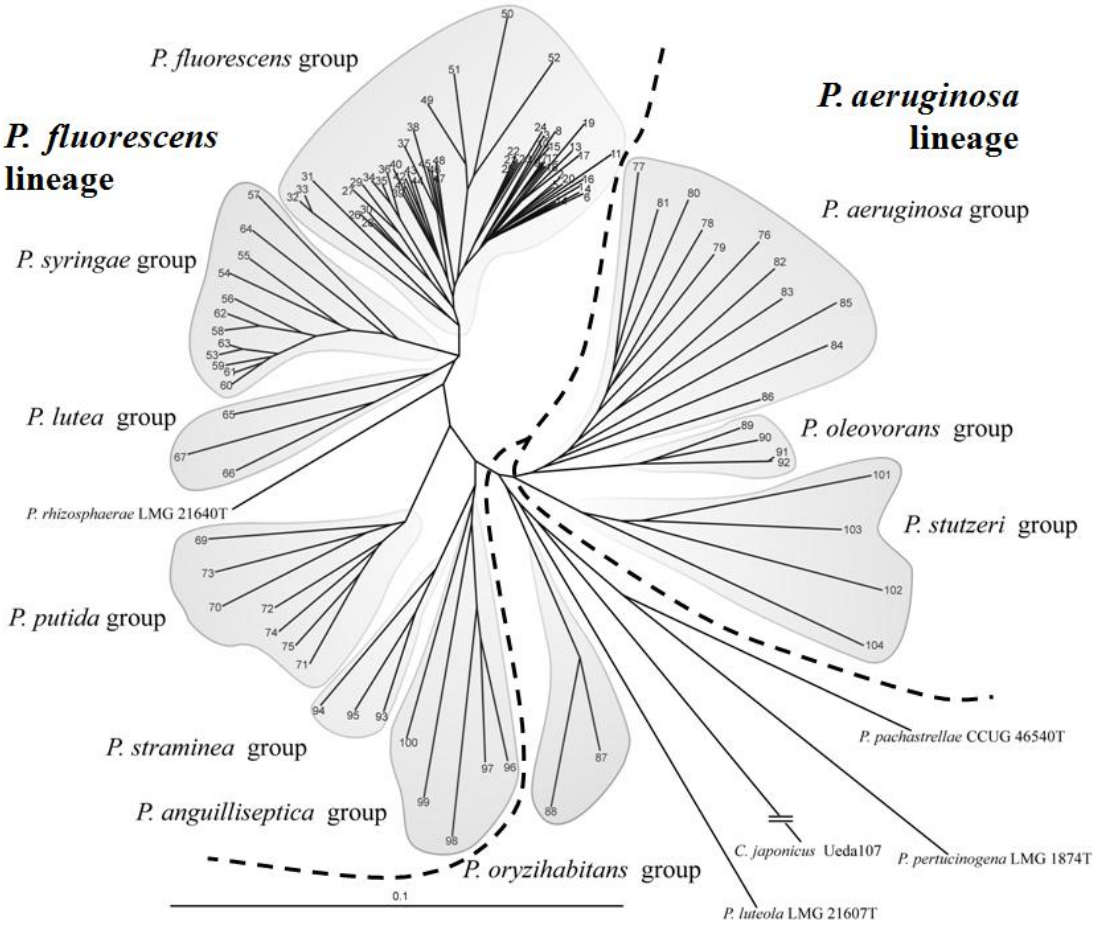


Figure I.7: Phylogenetic tree (unrooted) of 107 *Pseudomonas* type strains based on phylogenetic analysis of partial sequences of the 16S rRNA, *gyrB*, *rpoB* and *rpoD* genes. The bar indicates sequence divergence. Distance matrix was calculated by the Jukes-Cantor method. Dendrogram was generated by neighbor-joining. *Cellvibrio japonicus* Uada107 was used as outgroup. 1) *P. antarctica*^{T}, 2) *P. azotoformans*^{T}, 3) *P. cedrina*^{T}, 4) *P. costantinii*^{T}, 5) *P. extremorientalis*^{T}, 6) *P. fluorescens*^{T}, 7) *P. grimontii*^{T}, 8) *P. libaniensis*^{T}, 9) *P. marginalis*^{T}, 10) *P. orientalis*^{T}, 11) *P. palleroniana*^{T}, 12) *P. panacis*^{T}, 13) *P. poae*^{T}, 14) *P. salomonii*^{T}, 15) *P. synxantha*^{T}, 16) *P. tolaasii*^{T}, 17) *P. trivialis*^{T}, 18) *P. veronii*^{T}, 19) *P. rhodesiae*^{T}, 20) *P. simiae*^{T}, 21) *P. brenneri*^{T}, 22) *P. gessardi*^{T}, 23) *P. meridiana*^{T}, 24) *P. mucidolens*^{T}, 25) *P. proteolytica*^{T}, 26) *P. brassicacearum*^{T}, 27) *P. corrugata*^{T}, 28) *P. kilonensis*^{T}, 29) *P. mediterranea*^{T}, 30) *P. thivervalensis*^{T}, 31) *P. agarici*^{T}, 32) *P. asplenii*^{T}, 33) *P. fuscovaginae*^{T}, 34) *P. aurantiaca*^{T}, 35) *P. aureofaciens*^{T}, 36) *P. chlororaphis*^{T}, 37) *P. koreensis*^{T}, 38) *P. moraviensis*^{T}, 39) *P. jessenii*^{T}, 40)*P. vancouverensis*^{T}, 41) *P. umsongensis*^{T}, 42) *P. mohnii*^{T}, 43) *P. moorei*^{T}, 44) *P. reinekei*^{T}, 45) *P. frederiksbergesis*^{T}, 46) *P. mandelii*^{T}, 47) *P. lini*^{T}, 48) *P. migulae*^{T}, 49) *P. fragi*^{T}, 50) *P. lundensis*^{T}, 51) *P. psycrhophyla*^{T}, 52) *P. taetrolens*^{T}, 53) *P. amygdali*^{T}, 54) *P. avellanaea*^{T}, 55) *P. cannabina*^{T}, 56) *P. caricapapayae*^{T}, 57) *P. cichorii*^{T}, 58) *P. congelans*^{T}, 59) *P. ficuserectae*^{T}, 60) *P. meliae*^{T}, 61) *P. savastanoi*^{T}, 62) *P. syringae*^{T}, 63) *P. tremae*^{T}, 64) *P. viridiflava*^{T}, 65) *P. abietaniphila*^{T}, 66) *P. graminis*^{T}, 67) *P. lutea*^{T}, 69) *P. cremoricolorata*^{T}, 70) *P. fulva*^{T}, 71) *P. mosselii*^{T}, 72) *P. monteilii*^{T}, 73) *P. parafulva*^{T}, 74) *P. plecoglossicida*^{T}, 75) *P. putida*^{T}, 76) *P. aeruginosa*^{T}, 77) *P. citronellolis*^{T}, 78) *P. jinjuensis*^{T}, 79) *P. nitroreducens*^{T}, 80) *P. panipatensis*^{T}, 81) *P. knackmussii*^{T}, 82) *P. resinovorans*^{T}, 83) *P. otitidis*^{T}, 84) *P. indica*^{T}, 85) *P. thermotolerans*^{T}, 86) *P. alcaligenes*^{T}, 87) *P. oryzihabitans*^{T}, 88) *P. psycrhotolerans*^{T}, 89) *P. alcaliphila*^{T}, 90) *P. mendocina*^{T}, 91) *P. oleovorans*^{T}, 92) *P. pseudoalcaligenes*^{T}, 93) *P. argentinensis*^{T}, 94) *P. flavescens*^{T}, 95) *P.*

straminea^{T}, 96) *P. anguilliseptica*^{T}, 97) *P. peli*^{T}, 98) *P. guineae*^{T}, 99) *P. marincola*^{T}, 100) *P. borbori*^{T}, 101) *P. azotifigens*^{T}, 102) *P. balearica*^{T}, 103) *P. stutzeri*^{T}, 104) *P. xanthomarina*^{T}. Intrageneric groups (IG) or lineages, called Lineage *P. fluorescens* and Lineage *P. aeruginosa*, groups and subgroups habe been marked (Mulet *et al*., 2010).

In 2012, an update of that previous work was performed with 138 *Pseudomonas* strains (135 *Pseudomonas* type strains, *P. "alkylphenolia"*, and 2 *P. chlororaphis* subspecies), including recently described *Pseudomonas* species were analysed in order to reach a comprehensive view on the phylogenetic relationships of the species in the *Pseudomonas* genus (Mulet *et al*., 2012).

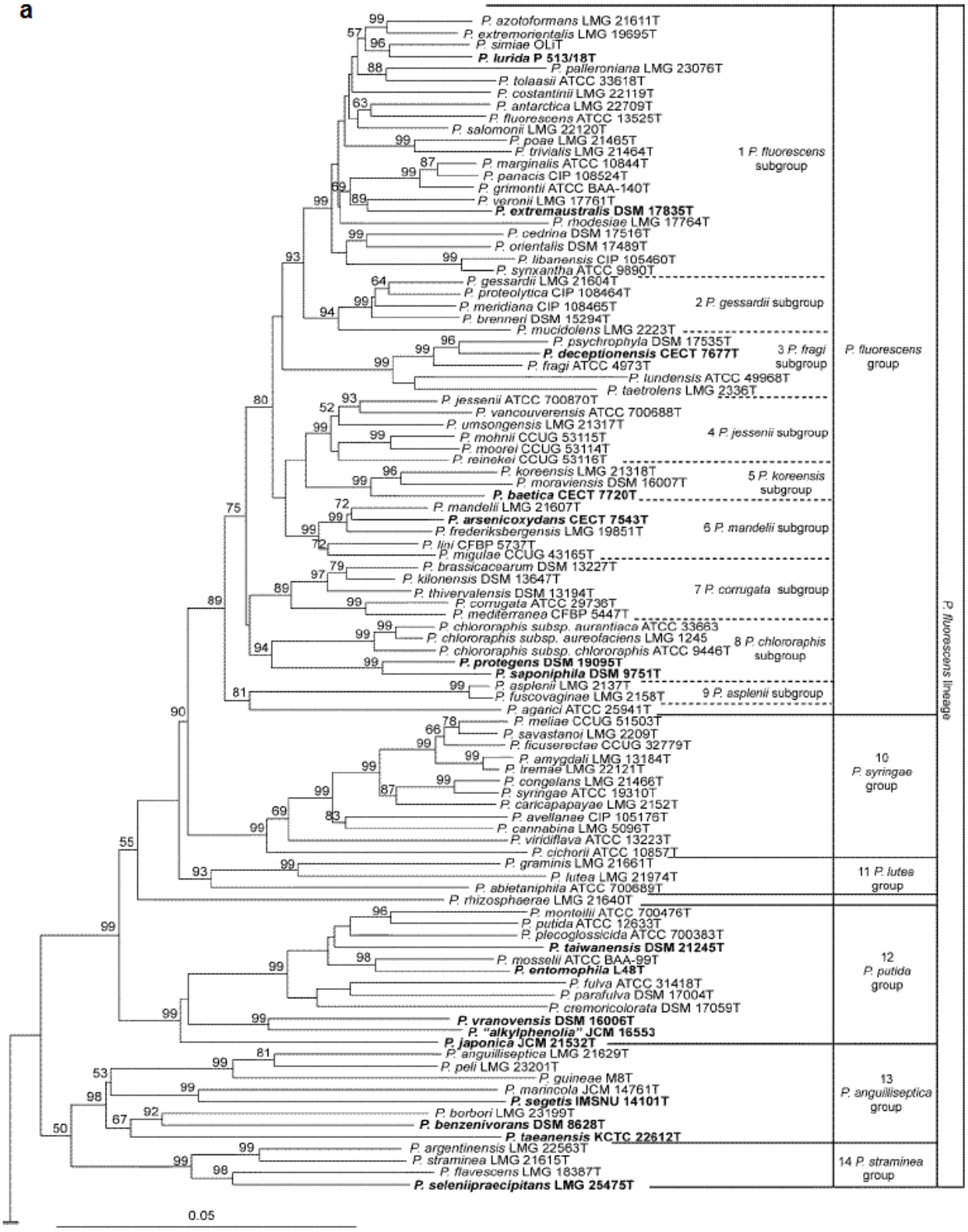
The first intrageneric group or lineage was divided into six groups (G), each one represented by the species *P. fluorescens*, *P. syringae*, *P. lutea*, *P. putida*, *P. anguilliseptica* and *P. straminea* (Fig. I.8a).

The *P. fluorescens* group was the most complex, and included 9 subgroups (SG), that were represented by the species *P. fluorescens*, *P. gessardi*, *P. fragi*, *P. mandelii*, *P. jesseni*, *P. koreensis*, *P. corrugata*, *P. chlororaphis*, and *P. asplenii*.

The second lineage, of *P. aeruginosa*, was divided into four main groups, represented by the species *P. aeruginosa, P. oleovorans, P. stutzeri and P. oryzihabitans* (Fig. I.8b).

P. agarici and *P. rhizospherae* affiliated in the phylogenetic analysis within the Lineage *P. fluorescens* and *P. indica* in the Lineage *P. aeruginosa*, were independent of any group. *P. pertucinogena group* were independent of any linage. *P. luteola, P. caeni* and *P. duriflava* should be considered outliers of the genus.

The bootstrap values of each complex branch of the individual, 3 or 4 concatenated genes analyzed showed the robustness of the analysis (Mulet *et al*., 2012).


a
P. azotoformans LMG 21611T
P. extremorientalis LMG 19695T
P. simiae OLiT
P. lurida P 513/18T
P. palleroniana LMG 23076T
P. tolaasii ATCC 33618T
P. costantinii LMG 22119T
P. antarctica LMG 22709T
P. fluorescens ATCC 13525T
P. salomonii LMG 22120T
P. poae LMG 21465T
P. trivialis LMG 21464T
1 P. fluorescens subgroup
P. marginalis ATCC 10844T
P. panacis CIP 108524T
P. grimontii ATCC BAA-140T
P. veronii LMG 17761T
P. extremaustralis DSM 17835T
P. rhodesiae LMG 17764T
P. cedrina DSM 17516T
P. orientalis DSM 17489T
P. libanensis CIP 105460T
P. synxantha ATCC 9890T
P. gessardii LMG 21604T
P. proteolytica CIP 108464T
P. meridiana CIP 108465T
2 P. gessardii subgroup
P. brenneri DSM 15294T
P. mucidolens LMG 2223T
P. psychrophyla DSM 17535T
P. deceptionensis CECT 7677T
3 P. fragi subgroup
P. fragi ATCC 4973T
P. lundensis ATCC 49968T
P. taetrolens LMG 2336T
P. fluorescens group
P. jessenii ATCC 700870T
P. vancouverensis ATCC 700688T
P. umsongensis LMG 21317T
P. mohnii CCUG 53115T
4 P. jessenii subgroup
P. moorei CCUG 53114T
P. reinekei CCUG 53116T
P. koreensis LMG 21318T
P. moraviensis DSM 16007T
5 P. koreensis subgroup
P. baetica CECT 7720T
P. mandelii LMG 21607T
P. arsenicoxydans CECT 7543T
P. frederiksbergensis LMG 19851T
6 P. mandelii subgroup
P. lini CFBP 5737T
P. migulae CCUG 43165T
P. brassicacearum DSM 13227T
P. kilonensis DSM 13647T
P. thivervalensis DSM 13194T
7 P. corrugata subgroup
P. corrugata ATCC 29736T
P. mediterranea CFBP 5447T
P. chlororaphis subsp. aurantiaca ATCC 33663
P. chlororaphis subsp. aureofaciens LMG 1245
P. chlororaphis subsp. chlororaphis ATCC 9446T
8 P. chlororaphis subgroup
P. protegens DSM 19095T
P. saponiphila DSM 9751T
P. asplenii LMG 2137T
P. fuscovaginae LMG 2158T
9 P. asplenii subgroup
P. agarici ATCC 25941T
P. meliae CCUG 51503T
P. savastanoi LMG 2209T
P. ficuserectae CCUG 32779T
P. amygdali LMG 13184T
P. tremae LMG 22121T
P. congelans LMG 21466T
P. syringae ATCC 19310T
P. caricapapayae LMG 2152T
P. avellanae CIP 105176T
P. cannabina LMG 5096T
P. viridiflava ATCC 13223T
P. cichorii ATCC 10857T
10 P. syringae group
P. graminis LMG 21661T
P. lutea LMG 21974T
P. abietaniphila ATCC 700689T
11 P. lutea group
P. rhizosphaerae LMG 21640T
P. monteilii ATCC 700476T
P. putida ATCC 12633T
P. plecoglossicida ATCC 700383T
P. taiwanensis DSM 21245T
P. mosselii ATCC BAA-99T
P. entomophila L48T
P. fulva ATCC 31418T
P. parafulva DSM 17004T
P. cremoricolorata DSM 17059T
P. vranovensis DSM 16006T
P. "alkylphenolia" JCM 16553
P. japonica JCM 21532T
12 P. putida group
P. anguilliseptica LMG 21629T
P. peli LMG 23201T
P. guineae M8T
P. marincola JCM 14761T
P. segetis IMSNU 14101T
P. borbori LMG 23199T
P. benzenivorans DSM 8628T
P. taeanensis KCTC 22612T
13 P. anguilliseptica group
P. argentinensis LMG 22563T
P. straminea LMG 21615T
P. flavescens LMG 18387T
P. seleniipraecipitans LMG 25475T
14 P. straminea group
P. fluorescens lineage
0.05

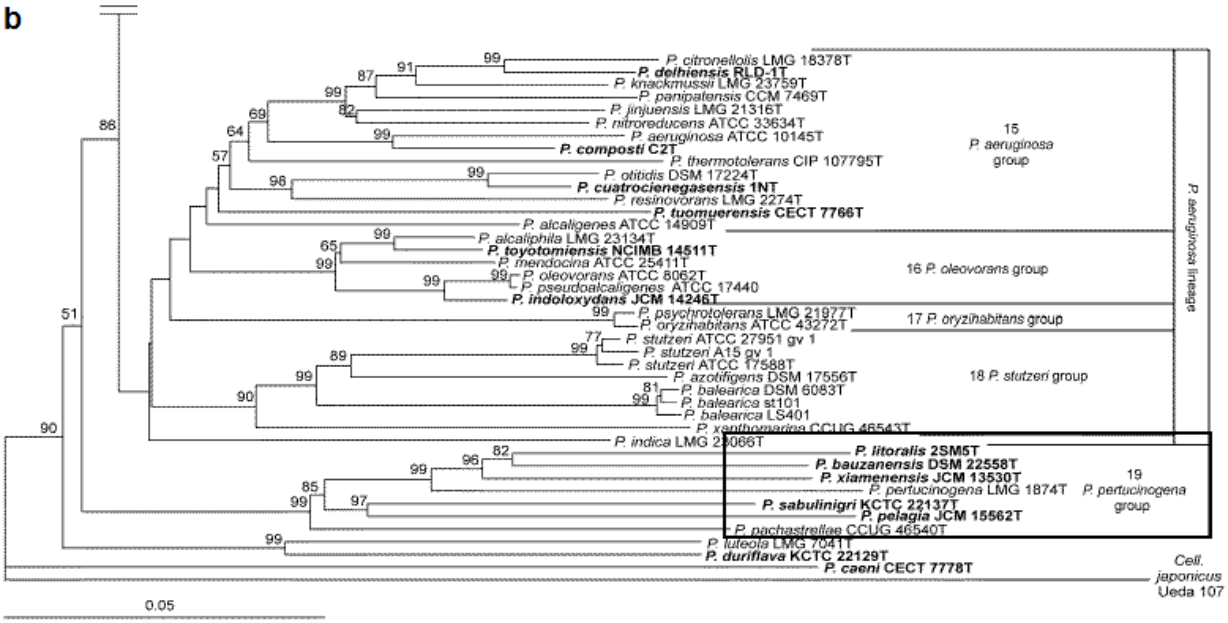


Figure I.8: Actual phylogenetic tree of the Pseudomonas genus. a) Phylogenetic branch of the *P. fluorescens* linage, b) Phylogenetic branch of the *P. aeruginosa* linage (Mulet *et al.*, 2012). Phylogenetic trees are based on the multigenic analysis of the concatenated RNAr 16S rRNA, *gyrB, rpoD* y *rpoB* genes. The species indicated in bold correspond to the most recent added species. The species in the box correspond to the new described group *P. pertucinogena*. The gene distances of the concatenated of the seven MLST genes were calculated using the Jukes-Cantor method, and phylogenetic tree was generated by neighbor-joining. Boostrap values of more than 500 (from 1000 replicates) are indicted at the nodes. *Cellvibrio japonicum* Ueda107 was used as outgroup (Mulet *et al.*, 2012).

I.4 The *rpoD* gene

I.4.1. Design and specifity test of the primer set PsEG30F/PsEG790R

The primer set 70F/70R for the *rpoD* gene was initially used by Yamamoto and collaborators for the study of *P. putida* strains (Yamamoto *et al*., 1998). Their combined analysis of the *gyrB* and *rpoD* sequences that was performed for 20 strains of *P. putida*, was compared with the resulting 16S rDNA phylogeny. These authors later extended the study to other members of the *Pseudomonas* genus, including 125 strains covering 31 species. The results that Yamamoto and collaborators obtained elucidated different complexes and phylogenetic lineages within the *Pseudomonas* group (Yamamoto *et al.,* 2000). The 70F and 70R primers that had five and six degenerations respectively were designed on sequences of *Buchnera aphidicola*, *E. coli* K-12, *Salmonella tiphymurium*, and *P. aeruginosa*. These primers were tested in our collection of *Pseudomonas* strains and a significant number of DNA from these strains could not be amplified.

The primer set PsEG30F/PsEG790R designed in our laboratory was based on all of the *Pseudomonas rpoD* gene sequences available in databases that represented 35 species from all *Pseudomonas* intrageneric phylogenetic clusters. The primers PsEG30F/PsEG790R show only a few degenerations, precisely two for the forward and one for the reverse primer, thus increasing their specificity (Table I.2). The only non-*Pseudomonas* bacterial genus showing significant similarity to both primers when these sequences were checked against the databases belonged to the genus *Alcanivorax* but not to any other close phylogenetically-related genus (Mulet *et al*., 2009).

Table I.2: Design of forward and reverse primers for the *Pseudomonas rpoD* gene (PsEG30F/PsEG790R).

																		Number of *Pseudomonas* type strains with each sequence
PsEG30F																		
5'	A	T	Y	G	A	A	A	T	C	G	C	C	A	A	R	C	G	3'
	·	·	C	·	·	A	·	·	·	·	·	·	·	·	G	·	·	18
	·	·	T	·	·	A	·	·	·	·	·	·	·	·	G	·	·	4
	·	·	C	·	·	A	·	·	·	·	·	·	·	·	A	·	·	10
	·	·	C	·	·	G	·	·	·	·	·	·	·	·	A	·	·	1(*P. balearica*)
PsEG790R																		
5'	T	C	A	A	G	G	A	M	A	T	C	A	A	C	C	G	3'	
	·	·	·	·	·	·	·	C	A	·	·	·	·	C	·	·		28
	·	·	·	·	·	·	·	C	G	·	·	·	·	C	·	·		1 (*P. citronellolis*)
	·	·	·	·	·	·	·	C	A	·	·	·	·	T	·	·		1 (*P. stutzeri*)
	·	·	·	·	·	·	·	A	A	·	·	·	·	C	·	·		2
	·	·	·	·	A	·	·	C	A	·	·	·	·	C	·	·		1 (*P.mediterranea*)

Y= C or T, R= A or G, M = C or A. (Mulet *et al*., 2009)

The PsEG30F/PsEG790R primer set amplified the expected *rpoD* internal fragment (Fig. I.9) of the 96 *Pseudomonas* type strains known at the time of the experiments of Mulet and collaborators (2009). They also successfully amplified a well characterized *Pseudomonas* collection consisting of more than 100 strains. Several strains isolated from environmental samples that were identified as non-*Pseudomonas* strains were unable to be amplified with these novel primers, with the exception of the *Alcanivorax* strains. The specificity of the primers was also verified by attempting the amplification of DNA from 10 clones that were not affiliated to the *Pseudomonas* phylogenetic branch but were part of the *rpoD* gene libraries 70F/70R. Importantly, none of these control experiments resulted in the production of the amplicon (Mulet *et al*., 2009).

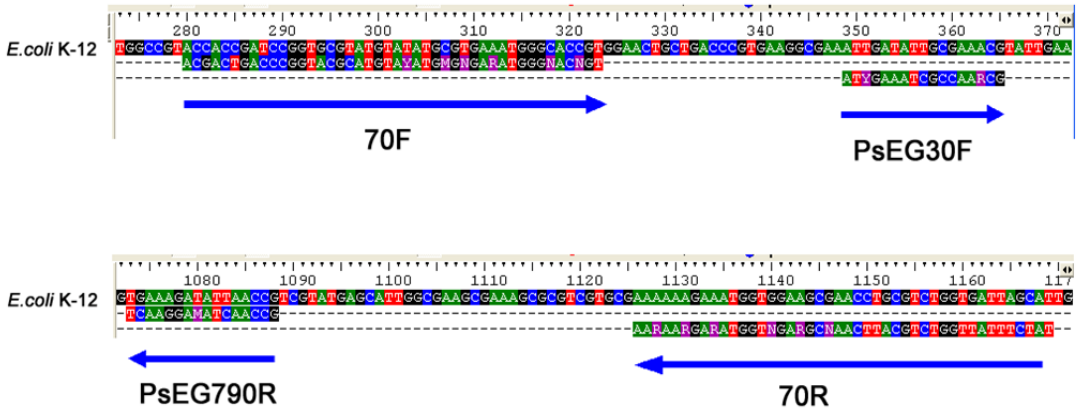


Figure I.9: Primers for the *Pseudomonas rpoD* gene located along a *rpoD* sequence (Mulet *et al*., 2009).

I.4.2. Detection of *Pseudomonas* in environmental DNA

Widmer and collaborators (1998) designed a set of primers (Ps-for/Ps-rev) based on 16S rDNA. The objective of the design of the Ps-for and Ps-rev primers was to develop a PCR approach that would allow for the selective detection of *Pseudomonas* (*sensu stricto*) in environmental samples (Widmer *et al.*, 1998). As far as we know, the 70F/70R set of primers has not been used previously in the analysis of environmental samples until Mulet and collaborators (2009). When these primers were used in the study of environmental samples by cloning the amplified DNA, more than 50 % of the genera detected were distant to the phylogenetic *Pseudomonas* branch. In some cases the percent identity assigned to the same clone showed values ranging from 58 % to 60 % to the genera *Pseudomonas*, *Chromobacterium* and *Stenotrophomonas.* In other cases the percent identity was 61 and 62 % for *P. aeruginosa* or *P. putida*, respectively. No percentage higher than 80 % was found for any of the 45 clones analysed. A possible explanation for this observation was that the *rpoD* gene database was not sufficiently exhaustive and has to be completed in order to be used for more detailed ecological studies. Another possibility was that novel *rpoD* phylogenetic branches have been detected in the environmental DNA samples, and they could represent groups of strains that have not yet been obtained in pure culture. Two samples were analysed in the study and, in both cases, similar results were obtained. A wide range of genera could be amplified from environmental DNA with these primers if they were not specific for *Pseudomonas*.

When it was applied to the sand samples, the Ps-for/Ps-rev primer set detected a few *Pseudomonas* species, and the 70F/70R primer set showed a low range of specificity. When the PsEG30F/PsEG790R primers were used directly with environmental DNA, no amplification products were obtained, indicating a scarce presence of *Pseudomonas* in the samples tested.

However, a higher diversity of *Pseudomonas* could be detected when primers PsEG30F/ PsEG790R were employed in the context of the nested procedure. In fact, 46 of the 84 clones analysed (55 %) belonged to the genus *Pseudomonas*. Clones belong to the genus Alcanivorax were also detected. However, both of the *Pseudomonas* and *Alcanivorax* groups were clearly separated in different *rpoD* phylogenetic branches such that the clones could be easily differentiated.

In general, it was concluded that the primers designed and tested in this study are sufficiently selective for the detection of *Pseudomonas*. In addition, these primers can be combined in a nested PCR procedure in order to increase the sensitivity for the detection of *Pseudomonas* in environments where they are in low numbers (Mulet *et al*., 2009).

In the present Thesis the primers PsEG30F/PsEG790R were used with no need of a nested PCR amplification.

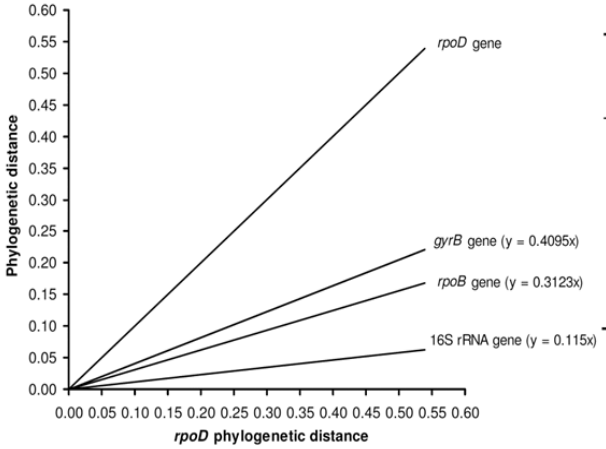
I.4.3. Analysis, comparison and selection of genes for phylogenic studies of the genus *Pseudomonas*

Four housekeeping genes were selected for a multigenic phylogenetic analysis of 107 type strains of the *Pseudomonas* genus: 16S rRNA, *gyrB, rpoB* and *rpoD* genes (Mulet *et al*., 2010). The four genes were compared in order to select the most discriminating gene and were used in a combined analysis to infer the phylogeny of the genus (Mulet *et al*., 2010).

For each single gene, a matrix of the phylogenetic distances between the 107 type strains was constructed, and the distances of pairs of strains (5671 values) were plotted. When were compared with *rpoD*, the distances of the four genes in the pair-wise comparisons correlated with R^2 values of 0.64, 0.75, and 0.69 for the 16S rRNA, *gyrB* and *rpoB* genes respectively. The discriminatory power of each gene was calculated as the ratio between the *rpoD* slope and the slopes of the other genes: *rpoD*/16S rDNA (8 times), *rpoD/rpoB* (3 times) and *rpoD/gyrB* (2 times). The more discriminating gene analysed was *rpoD*, followed by *gyrB*, *rpoB* and the 16S rRNA gene. The range and average distances for each gene are shown in (Fig. I.10).

In a similar way, the matrices constructed for the concatenated sequences of three genes (16S rRNA gene, *gyrB, rpoD*; 2870 nt) and four genes (16S rRNA gene, *gyrB, rpoB* and *rpoD*; 3726 nt) were compared in a pair-wise manner to assess the correlation between them and the relative discriminatory power of both sets of genes. They were well correlated and almost equally discriminating (3 genes vs. 4 genes: $y=1.0252x$, $R^2=0.987$). *rpoD* was well

correlated with all of them (3 genes: y=0.3741x/R^2=0.9077; 4 genes: y=0.3586x/ R^2=0.8881) and yielded the best resolution (Mulet *et al*., 2010)(Fig. I.10). Similar results were obtained in a study of *P. stutzeri* strains to determinate their genomovars (Scotta *et al*., 2013).



Gen	Mean similarity (%) ± standard deviation	% Maximal similarity	% Minimum similarity
16S rRNA	96.78±1.58	100.00	91.84
gyrB	82.54±5.20	99.62	68.17
rpoD	72.50±10.99	99.86	46.04
rpoB	88.44±4.14	99.78	73.75
3 concatenated genes	87.32±4.32	99.86	76.24
4 concatenated genes	87.60±4.18	99.76	77.20

Figure I.10: Least square tendency lines obtained for the phylogenetic distance between 107 type strain *Pseudomonas* for the different genes (16S rDNA, *gyrB* and *rpoB*) with respect to the *rpoD* gene. The slope is indicated in each case. The lines have been vertically shifted for the sake of clarity. The correlation coefficient R^2 is 0.6401 to 16S rDNA, 0.7501 to *gyrB* and 0.686 to *rpoB*. Mean, maximal and minimal similarities for each gene individually analyzed or in the concatenated analysis of 107 *Pseudomonas* type strains (Mulet *et al*., 2010).

Similarly, three other genes (*atpD, carA, recA)* from 13 type strains (Hilario *et al.*, 2004) were compared and the *rpoD* was also the most discriminating gene (Mulet *et al*., 2010). The genetic diversity of *oprI* and *oprL* sequences was also compared with *rpoD* sequences (Matthijs *et al*., 2012). The discriminatory of the *rpoD* gene was 3 times higher than *oprI* gene, and 2 times higher than *oprL* gene (Fig. I.11).

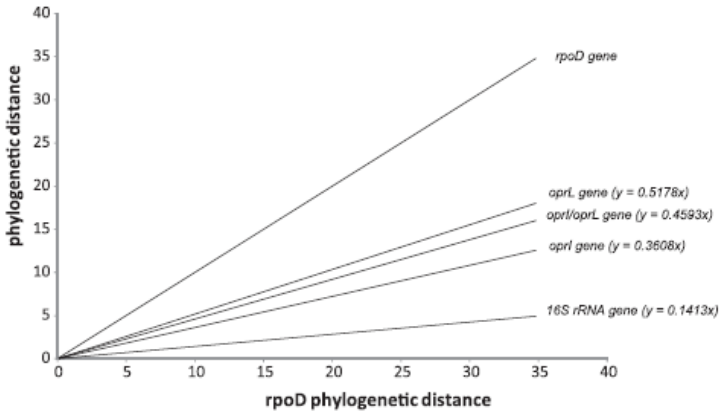


Figure I.11: Least square tendency lines obtained for correlation of phylogenetic distances between 75 Pseudomonas type strains. The slope is indicated for each marker. Distances are x100. When compared with *rpoD*, distances in pairwise comparisons showed correlation coefficients (R2) of 59.51%, 62.67%, 67.47% and 56.37% for oprI, oprL, concatenated oprI/oprL and 16S rRNA gene sequences, respectively (Matthijs *et al*., 2012).

I.5. Next-Generation Sequencing Technology

I.5.1. Illumina HiSeq 2000 technology

This technology has been used in this Thesis to obtain the four genomes sequenced and analyzed in the Chapter 2.

The first step is the library preparation; **s**amples consisting of longer fragments are first sheared into a random library of 100-300 base-pair long fragments. After fragmentation the ends of the obtained DNA-fragments are repaired and an A-overhang is added at the 3'-end of each strand. Afterwards, adaptors which are necessary for amplification and sequencing are ligated to both ends of the DNA-fragments. These fragments are then size selected and purified (Fig. I.12a).

The Cluster Generation is performed on the Illumina cBot. Single DNA-fragments are attached to the flow cell by hybridizing to oligos on its surface that are complementary to the ligated adaptors. The DNA-molecules are then amplified by a so called *bridge amplification* which results in a hundred of millions of unique clusters. Finally, the reverse strands are cleaved and washed away and the sequencing primer is hybridized to the DNA-templates (Fig. I. 12b).

During sequencing the huge amount of generated clusters are sequenced simultaneously. The DNA-templates are copied base by base using the four nucleotides (ACGT) which are fluorescently-labeled and reversibly terminated. After each synthesis step, the clusters are excited by a laser which causes fluorescence of the last incorporated base. After that, the fluorescence label and the blocking group are removed allowing the addition of the next base. The fluorescence signal after each incorporation step is captured by a built-in camera, producing images of the flow cell (Fig. I. 12c) (http://www.illumina.com).

I.5.2. Roche 454/GS FLX Sequencing technology

This technology has been used in this Thesis to obtain by pyrosequencing of an *rpoD* amplicon from an environmental sample in the Chapter 3.

The GS FLX sequencer supports sequencing of various different nucleic acid starting materials such as genomic DNA, PCR products, BACs and cDNA. Samples consisting of longer sequences are first sheared into a random library of 300-800 base-pair long fragments (Fig. I.13a).

Adaptors essential for purification, amplification and sequencing are added to both ends

of the fragments. If the sample is double stranded one strand is removed and the remaining single strands are used in the following steps (Fig. I.13b).

Aided by the adaptors individual fragments are captured on their own unique beads. Each bead carries a unique single-stranded library fragment. A bead and the bound fragment together, with a water-in-oil emulsion (with the amplifications reagents), form a microreactor (Fig. I.13c), so each fragment can be amplified without contamination via the so called emulsion PCR (emPCR). The entire emulsions are amplified in parallel to create millions of clonally copies of each library fragment on each bead. After amplification the emulsion shell is broken and the clonally amplified beads are ready for loading onto the fibre-optic PicoTiterDevice (PTP) for sequencing (Fig. I.13d).

The PicoTiterPlate, where the surface design allows for only one bead per well, is loaded with the amplified beads (Fig. I.13e) and smaller beads with the enzymes necessary for sequencing (Fig. I.13f).

Sequencing is accomplished by synthesizing the complementary strands of the bead attached templates. In a number of cycles the four bases (ATGC) are sequentially, individual nucleotides are flowed in sequence across the wells. Each incorporation of a nucleotide complementary to the template strand results in a chemiluminescent light signal recorded by the camera (Fig. I.13g). 454 Sequencing Data Analysis software uses the signal intensity of each incorporation event at each well position to determine the sequence of all reads in parallel (Fig. I.13h) (http://454.com/index.asp).

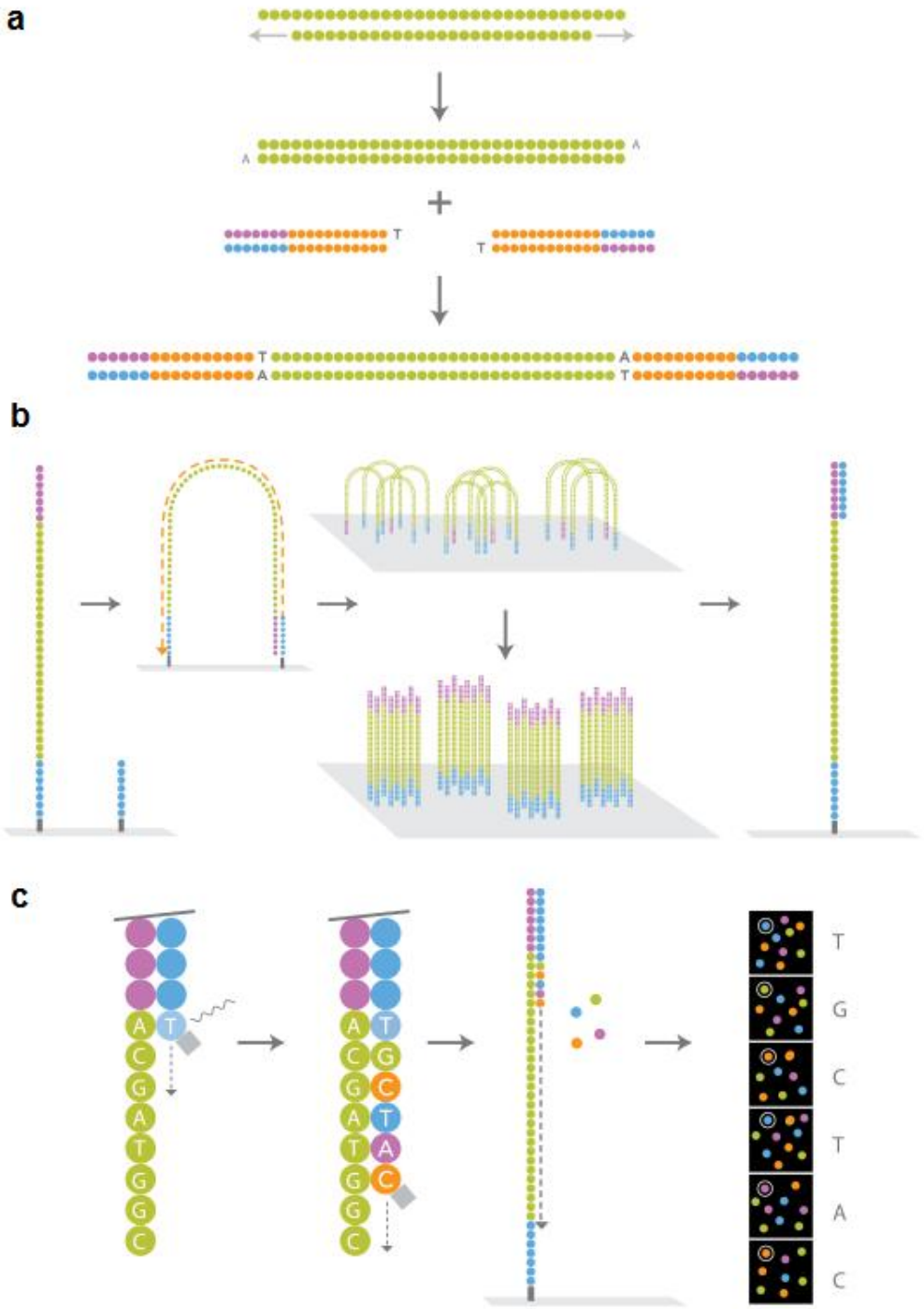


Figure I.12: Illumina HiSeq2000 sequencing diagram (Illumina Inc.).

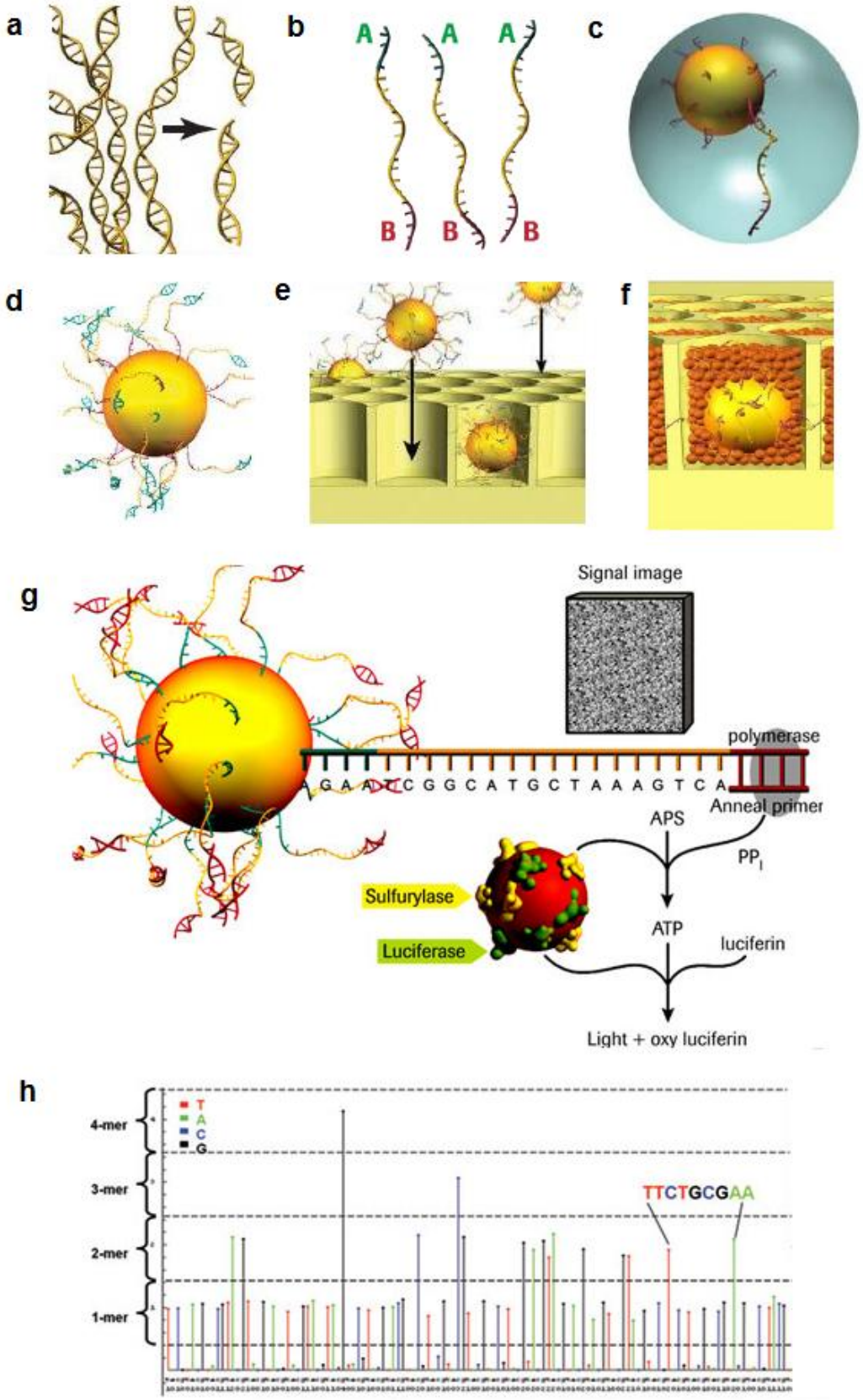


Figure I.13: Roche 454 pirosequencing diagram (Roche Diagnostics).

Objectives

One of the characteristics of the genus *Pseudomonas* is the ability of their species to survive in many habitats, from the Arctic to the Antarctic and in very different reservoirs: in water, soil, insects, nematodes, plants and humans. The main objective of the present work is a deep study of the diversity of *Pseudomonas* populations present in the environment.

Objectives:

1. Screening of *Pseudomonas* isolates from several and different habitats. Clinical, agricultural soil, intertidal sediments, subsurface water and river water samples will be prospected.

2. Application of different methodologies. Culture-dependent and independent methods, based on traditional microbiological methods and on newly developed molecular methodologies, including multilocus sequence typing, analysis of DNA obtained from environmental samples by cloning, sequencing and pyrosequencing will be tested. Comparative genomics of *Pseudomonas* isolates to understand the microdiversity between close related isolates from the same species will be performed.

3. Assessment of the potential of new methodologies in the description of the species diversity of the genus *Pseudomonas*. The data obtained will allow a better insight in the application of these methodologies in the ecology and taxonomy of the genus.

4. The description of new *Pseudomonas* species from different environmental samples. The *Pseudomonas* data compilation of these studies will be used to assess to which extent the diversity of the genus is known, and therefore the possibility to describe new species in the genus *Pseudomonas*.

Chapter 1

Population structure of environmental and clinical *Pseudomonas aeruginosa* isolates from the same geographical area

Abstract

The population structure and microdiversity of 53 *Pseudomonas aeruginosa* isolates from environmental samples and clinical specimens obtained in Mallorca (Spain) has been analyzed by a multilocus sequence typing approach (MLST). Thirty-seven sequence types (ST) were detected, 22 of which represent new STs: 8 from clinical isolates and 14 from environmental isolates. New alleles were found only in environmental isolates. The ST-1146 was the only one found in both kind of samples, 3 environmental isolates (from the same site isolated at 2 different dates) and 1 clinical isolate, with differences in its antibiotic susceptibility profile. Antibiotic multiresistance to several antibiotics was only found in isolates of clinical origin. Clonality tests demonstrated that recombination plays a key role in the distribution of alleles.

1.1. Introduction

The genetic repertoire of *Pseudomonas aeruginosa* reflects the life style of this bacterial species (Pirnay *et al*., 2009). It colonizes a broad spectrum of habitats reflecting its ability to exploit many different nutrition sources and its high potential for adaptation to new or changing environmental conditions (Ramos, 2004). Its metabolic versatility is derived from genes encoding not only enzymes participating in metabolic pathways but also a very high number of transcriptional regulators and two-component regulatory systems (Stover *et al*., 2000). *P. aeruginosa* clinical isolates are considered genotypically, chemotaxonomically, and functionally indistinguishable from environmental isolates (Pirnay *et al*., 2009). More than 500 regulatory genes have been identified in the genome of the *P. aeruginosa* strain PAO1 (Stover *et al*., 2000), and this regulation may explain its adaptive potential.

A multilocus sequence typing (MLST) scheme was developed to characterize unambiguously isolates of *P. aeruginosa*. MLST is a strain-typing system that focuses strictly on 7 conserved housekeeping genes (*acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA*, and *trpE*) (Curran *et al*., 2004). MLST is a powerful approach for understanding the evolution of the core genome and the processes that shape strain diversity (Kidd *et al*., 2012). It is very useful for detecting genetic relatedness, providing information about the clonal relationships between isolates (Johnson *et al*., 2007). Because of these advantages, we used an MLST scheme to characterize strains of *P. aeruginosa* isolated from the environment and from hospital specimens, which enabled a comparison with those strains stored in a database generated from various clinical and environmental *P. aeruginosa* strains. In addition, we attempted to clarify the phylogenetic relationships of these isolates by analyzing the concatenated sequences of the 7 genes used for MLST.

Some authors have argued that the species has a panmictic population structure (Denamur *et al*., 1993) and have proposed a net-like structure, characterized by high frequencies of recombination (Kiewitz *et al*., 2000). However, an epidemic structure has been proposed (Lomholt *et al*., 2001).

Different studies of *P. aeruginosa* populations and their distribution have been made. Most have focused on clinical strains with relatively few concentrating on environmental isolates (Khan *et al*., 2008; Kidd *et al*., 2012; Maatallah *et al*., 2011). In this study, *P. aeruginosa* isolates from both environmental (26 water samples taken from 7 wells and 1 pond) and clinical (27 specimens from the Son Dureta University Hospital Mallorca, Spain) sources were obtained and analyzed by MLST.

1.2. Material and Methods

1.2.1. *P. aeruginosa* isolates

Twenty-seven clinical and 26 environmental *P. aeruginosa* isolates were analyzed. Sources, date of isolation and antibiotic susceptibility profiles are indicated in Table 1.1a and 1.1b. The clinical strains were isolated on MacConkey agar plates (Biomerieux) at 37°C from clinical specimens at Son Dureta University Hospital from the 17 to 24 February 2009. Isolates were selected randomly, and none of them were derived from cystic fibrosis patients. The environmental strains were isolated from May 2009 to February 2011 from samples taken from 7 wells of subsurface waters and 1 pond, in Mallorca (Spain), located close to an urban solid waste treatment plant or to a waste construction deposit, at 10.3 and 43.3 km away from the hospital, respectively. Water samples (1-100 ml) were filtered through sterile Millipore membrane filters (0.45 µm pore size), placed on Cetrimide Agar (Merck) plates and incubated at 37°C. *P. aeruginosa* isolates were routinely cultivated at 30°C in Luria Broth (LB) or LB agar.

1.2.2. Antibiotic resistance

Clinical Isolates were characterized phenotypically by routine identification tests and also with the MicroScan W/A system (Dade MicroScan Inc.). Antibiotic disk diffusion susceptibility tests were performed for the environmental and clinical isolates and interpreted according to the guidelines of the Clinical Laboratory Standards Institute (CLSI, 2009 and 2010). Ten antibiotics of six different antimicrobial categories were tested (Table 1.S1a and 1.S1b): aminoglycosides (amikacin, gentamicin and tobramycin), carbapenems (imipenem and meropenem), cephalosporins (ceftazidime and cefepime), fluoroquinolones (ciprofloxacin), penicilins and β–lactamases inhibitor (piperacillin-tazobactam), and monobactams (aztreonam). Antibiotics belonging to two required antimicrobial categories to define the resistance pattern, phosponic acids and polymyxins, were not tested. Multiresistance classification was performed according to Magiorakos and colleagues (2012) as indicated in Table 1.S1a and 1.S1b. MDR is defined as non-suceptibility to at least one agent in three or more antimicrobial categories. XDR is defined as non-suceptibility to at least one agent in all but two or fewer antimicrobial categories. PDR is defined as as non-suceptibility to all agents in all microbial categories.

1.2.3. DNA extraction, PCR amplification and DNA sequencing conditions

DNA extraction procedures were previously described by Ausubel and colleagues (1997). The seven housekeeping genes: *acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA* and *trpE* were selected for amplification and sequencing, according to the MLST scheme of *P. aeruginosa* website (http://pubmlst.org/paeruginosa/). PCR amplification reactions (volumes, nucleotides and primer concentrations) were performed as previously described (Mulet *et al*., 2009) in an Eppendorf thermocycler. The PCR conditions for the primers were slightly modified from the previous published conditions (Curran *et al*., 2004). The reactions were performed as previously described (Gomila *et al*., 2013). The amplified products were purified with Multiscreen PCR 96-well plates (Millipore). Sequence reactions were carried out using the ABI Prism BigDye Terminator version 3.1 and the sequences were read with an automatic sequence analyser (3130 genetic analyzer; Applied Biosystems).

1.2.4. MultiLocus Sequence Typing

A total of 2882 nucleotides for each strain was analyzed: 390 bp for *acsA*, 498 for *aroE*, 373 bp for *guaA*, 442 bp for *mutL*, 366 bp for *nuoD*, 370 bp for *ppsA* and 443 bp for *trpE* genes. Sequences were edited using the BioEdit program (Hall, 1999). A number was assigned to each distinct allele within a locus according to the number available in the *P. aeruginosa* MLST database website. For each isolate, the combination of alleles obtained at each locus defined its allelic profile or sequence type. Any allele that did not match with an existing one was designated as a “new” allele. Any sequence type that did not match with the existing database was numbered as a “new” sequence type. The new nucleotide sequences determined in this study, for each different allele of each locus, and new STs have been submitted to the *P. aeruginosa* MLST database.

1.2.5. Phylogenetic reconstruction

The concatenated sequences of the seven housekeeping genes were used to establish the phylogenetic relationships between the clinical and environmental isolates, including the sequences of the same genes from whole genomes sequenced of *P. aeruginosa* strains available at the National Center for Biotechnology Information (NCBI): LESB58 (NC_011770.1), M18 (NC_017548.1), NCGM2.S1 (NC_017549.1), PA7 (NC_009656.1), PAO1 (NC_002516.2), UCBPP-PA14 (NC_008463.1), 39016 (NZ_CM001020.1) and PACS2 (NZ_AAQW0100001.1). Phylogenetic reconstructions were performed as previously described (Mulet *et al*., 2009).

1.2.6. Population structure

Splits Tree

Clustering of STs was performed with the Sequence Type Analysis and Recombinational Tests, START v.2 program (Jolley *et al*., 2001) and split decomposition was analyzed with the Splits Tree software program v.4 (Huson *et al*., 2013).

eBURST and minimum spanning tree construction

The eBURST v3 algorithm was used to determine the evolutionary relationships between all STs, for phylogenetic analysis and for a stringent group definition of isolates (Feil *et al*., 2004). Two isolates that were identical at six or more alleles were considered to be part of the same clonal complex or Burst Group (BG) (Kidd *et al*., 2012). The minimum spanning tree was constructed with PHYLOViZ vs.1.0 software, which uses goeBURST, a recently refined version of the eBURST algorithm (Francisco *et al*., 2012). The ST node dimensions were plotted relative to the number of isolates within this ST.

1.2.7. Genetic Diversity index

Based on MLST data, the discriminatory index (D) of *P. aeruginosa* was calculated using Ridom Epicompare software (http://www.ridom.de/download.shtml). It calculates the average probability that the typing system will assign a different type to two unrelated strains randomly sampled in the microbial population of a given taxon.

1.2.8. Clonality and recombination tests

ClonalFrame

ClonalFrame, a computer package for the inference of bacterial microevolution, was used to investigate population structure by inferring relationships among STs. It estimates the relative probabilities that a nucleotide is changed as a result of recombination relative to point mutation (r/m ratio) (Didelot *et al*., 2009).

The standardized Index of association (I_A^S)

I_A^S between all STs was calculated by the START V.2 program to measure the clonality of the population (Jolley *et al*., 2001). Values of I_A^S that differ significantly from zero indicate that recombination has been rare or absent.

1.3. Results

1.3.1. *P. aeruginosa* isolates

Twenty-six environmental isolates were obtained from 8 water samples from different origins as indicated in Tables 1.1a and 1.1b. In routine analysis of 24 water samples during the sampling period of this study (June 2009 to February 2011) of wells 1, 10 and 12 none *P. aeruginosa* was detected in 33 % of the samples (2 samples in well 1, 2 samples in well 10 and 4 samples in well 12). The ranges of *P. aeruginosa* cfu/100 ml in the other samples were: 1 to more than 100 in well 1, 1 to 980 in well 10 and 2 to 8 in well 12.

Additionally, 27 *P. aeruginosa* strains were isolated from 27 clinical specimens obtained from 25 patients at the Son Dureta University Hospital (Palma de Mallorca, Spain). Eighteen isolates (66.6%) were obtained from respiratory samples and 9 (33.3%) from wounds. Dates of isolation and samples origin are given in Tables 1.1a and 1.1b.

1.3.2. Antibiotic resistance pattern

The resistance pattern to the different antibiotics tested for the 27 clinical isolates was as follows: 29.63% non-MDR (8 isolates), 37.04% possible MDR (10 isolates), 11.11% MDR (3 isolates), 7.41% MDR possible XDR (2 isolates) and 14.81% XDR possible PDR (4 isolates). The resistance pattern for the environmental isolates was: 96.15% non-MDR (25 isolates) and 3.85% possible MDR (1 isolate, P38). The antibiotic profiles are summarized in Tables 1.S1a and 1.S1b**.**

1.3.3. MultiLocus Sequence typing

All isolates were studied following the MLST methodology and were assigned to a ST. As indicated in Tables 1.1a and 1.1b, most clinical isolates were obtained from specimens from different patients, with only 2 exceptions: 4 isolates were obtained from the sputum specimens of 2 patients (patients 7 and 10; 2 isolates each) during the week studied (SD10, SD14 and SD13, SD15, respectively). In both cases, the isolates of the same patient were assigned to the same sequence type (ST-175 patient 7, ST-217 patient 10). ST-175 was the most frequently isolated ST in the clinical samples analyzed (11%). Two isolates were detected in the following STs: ST-217, ST-235 (represented by strain *P. aeruginosa* NCGM2.S1) and ST-253 (represented by strain PA14) together with the new ST-1246 and ST-1250. Among the environmental

samples studied, ST-1141 was the most frequent ST (6 isolates from the well 1) and ST-1146 the second most frequent ST (3 isolates from well 10 obtained at 2 different dates) (Tables 1.1a, 1.1b, 1.2a and 1.2b).

Table 1.1a: Clinical *P. aeruginosa* isolates analyzed in this study.

Clinical isolate	Patient	Sex[a]	Date (day/month/year)	Sample	Resistance pattern[b]
SD4	1	F	17/02/09	Non surgical injury (exuded)	Non MDR
SD5	2	M	24/02/09	Nasopharyngeal aspirate	Non MDR
SD6	3	M	24/02/09	Surgical injury (exuded)	Non MDR
SD7	4	M	24/02/09	Sputum	Non MDR
SD8	5	M	23/02/09	Ulcer	Non MDR, posible MDR
SD9	6	F	23/02/09	Ulcer	MDR, possible XDR
SD10	7	M	23/02/09	Sputum	XDR, possible PDR
SD11	8	M	23/02/09	Sputum	MDR
SD12	9	M	23/02/09	Bronchoaspirate	Non MDR, possible MDR
SD13	10	F	23/02/09	Sputum	Non MDR
SD14	7	M	22/02/09	Sputum	XDR, possible PDR
SD15	10	F	22/02/09	Sputum	Non MDR
SD16	11	F	22/02/09	Bronchoaspirate	XDR, possible PDR
SD17	12	F	22/02/09	Sputum	Non MDR, possible MDR
SD18	13	F	20/02/09	Decubitus ulcer	XDR, possible PDR
SD19	14	F	20/02/09	Decubitus ulcer	Non MDR, possible MDR
SD20	15	F	20/02/09	Ulcer	Non MDR, possible MDR
SD21	16	M	20/02/09	Sputum	MDR
SD22	17	F	19/02/09	Non surgical injury (exuded)	Non MDR
SD23	18	M	19/02/09	Tracheal aspirate	MDR, possible XDR
SD24	19	M	19/02/09	Bronchoaspirate	Non MDR, possible MDR
SD26	20	M	18/02/09	Bronchoaspirate	Non MDR, possible MDR
SD27	21	M	18/02/09	Tracheal aspirate	Non MDR, possible MDR
SD28	22	M	18/02/09	Bronchoaspirate	Non MDR, possible MDR
SD29	23	F	18/02/09	Surgical injury (exuded)	Non MDR
SD30	24	M	18/02/09	Sputum	MDR
SD31	25	F	17/02/09	Sputum	Non MDR, possible MDR

Table 1.1b: Environmental *P. aeruginosa* isolates analyzed in this study.

Environmental isolate	Date (month/year)	Sampling point	Situation	Resistance pattern[b]
P1	05/09	Well 1	Sta. Margalida	Non MDR
P4	05/09	Well 1	Sta. Margalida	Non MDR
P5	05/09	Well 1	Sta. Margalida	Non MDR
P9	05/09	Well 1	Sta. Margalida	Non MDR
P10	05/09	Well 1	Sta. Margalida	Non MDR
P12	05/09	Well 1	Sta. Margalida	Non MDR
P15	05/09	Well 1	Sta. Margalida	Non MDR
P17	05/09	Well 1	Sta. Margalida	Non MDR
P20	06/10	Well 12	Sta. Margalida	Non MDR
P23	06/10	Well 12	Sta. Margalida	Non MDR
P26	06/10	Well 12	Sta. Margalida	Non MDR
P27	06/10	Well 10	Sta. Margalida	Non MDR
P28	06/10	Well 10	Sta. Margalida	Non MDR
P30	08/10	Well 12	Sta. Margalida	Non MDR
P33	08/10	Well 12	Sta. Margalida	Non MDR
P36	10/10	Well 5	Son Reus	Non MDR
P37	10/10	Well 10	Sta. Margalida	Non MDR
P38	10/10	Well 12	Sta. Margalida	Non MDR, Possible MDR
P39	10/10	Well 39	Son Reus	Non MDR
P40	10/10	Well 53	Son Reus	Non MDR
P41	10/10	Pond	Son Reus	Non MDR
P42	10/10	Pond	Son Reus	Non MDR
P43	10/10	Pond	Son Reus	Non MDR
P46	02/11	Well 23	Son Reus	Non MDR
P47	02/11	Well 10	Sta. Margalida	Non MDR
P49	02/11	Well 10	Sta. Margalida	Non MDR

a) M: Male; F: Female.

b) Non MDR: non multidrug-resistant; MDR: multidrug-resistant; XDR: extensively drug-resistant; PDR: pandrug resistant; possible MDR: possible multidrug-resistant; possible XDR: possible extensively drug-resistant; possible PDR: possible pandrug resistant. (Magiorakos *et al.*. 2011).

The MLST analysis of the 53 isolates showed 37 different STs: 18 STs corresponded to 26 environmental strains, 20 STs corresponded to 27 clinical strains, and only ST-1146 was present in both types of samples analyzed. From the 37 different STs detected, 22 represented new STs: 8 from clinical strains (10 isolates) and 14 from environmental strains (15 isolates). Fourteen of the 22 new STs were detected as the result of new allele combinations (new allelic profile). Six of these STs originated from environmental samples and 8 from clinical specimens (Table 1.2a and 1.2b). Eight new STs were the result of the detection of 6 new alleles, all present in environmental isolates: 2 in *aroE* and *trpE* each, and 1 new allele in *guaA* and *mutL* each. All of the clinical MLST profiles were different from the environmental ones, with just one exception (ST-1146), which was detected in 1 clinical isolate (SD9) and 3 environmental isolates obtained from samples taken from the same well at different dates: one in October 2010 (P37) and 2 in February 2011 (P47 and P49). The ST-1146 contained the four most frequent alleles present in environmental isolates (*acsA* 5; *aroE* 11; *guaA* 57 and *nuoD* 1) and the two most frequent

alleles of clinical isolates (*ppsA* 6 and *trpE* 3). The allele *mutL* 33 was predominately detected in environmental isolates, and among all clinical isolates, it was only present in isolate SD9 of ST-1146 (Table 1.S2). The total number of polymorphic sites observed in the MLST genes was higher in the environmental isolates (98) than in the clinical ones (84).

Table 1.2a: Alleles and sequence type by multilocus sequence typing of clinical *P. aeruginosa* isolates analyzed in this study.

Clinical isolate	*acsA*	*aroE*	*guaA*	*mutL*	*nuoD*	*ppsA*	*trpE*	ST[a]
SD4	11	5	12	3	4	14	18	**1246**
SD5	13	4	5	5	12	7	15	308
SD6	39	5	9	11	27	5	19	**1252**
SD7	4	4	16	12	1	6	3	253
SD8	6	5	6	5	19	4	7	**1245**
SD9	5	11	57	33	1	6	3	1146
SD10	28	22	5	3	3	14	19	175
SD11	29	1	97	99	24	20	87	926
SD12	4	4	16	12	1	6	3	253
SD13	28	5	11	18	4	13	3	217
SD14	28	22	5	3	3	14	19	175
SD15	28	5	11	18	4	13	3	217
SD16	28	10	5	3	58	7	3	**1250**
SD17	28	10	5	3	58	7	3	**1250**
SD18	1	5	11	3	4	10	3	234
SD19	15	5	77	72	3	6	68	**1247**
SD20	11	5	12	3	4	14	18	**1246**
SD21	18	11	57	5	1	20	139	**1248**
SD22	38	11	3	13	1	2	4	235
SD23	36	27	28	3	4	13	7	179
SD24	17	5	12	3	14	4	7	244
SD26	35	11	25	6	13	6	84	**1251**
SD27	38	11	3	13	1	2	4	235
SD28	5	4	5	5	5	20	4	532
SD29	28	5	5	11	3	15	44	242
SD30	28	22	5	3	3	14	19	175
SD31	23	13	109	5	1	1	47	**1249**

a) New alleles and new sequence types described are indicated in bold.

Table 1.2b: Alleles and sequence type by multilocus sequence typing of environmental *P. aeruginosa* isolates analyzed in this study.

Environmental isolate	*acsA*	*aroE*	*guaA*	*mutL*	*nuoD*	*ppsA*	*trpE*	ST[a]
P1	113	4	5	67	1	17	26	1141
P4	113	4	5	67	1	17	26	1141
P5	113	4	5	67	1	17	26	1141
P9	113	4	5	67	1	17	26	1141
P10	113	4	5	67	1	17	3	**1257**
P12	113	4	5	67	1	17	26	1141
P15	113	4	5	67	1	17	10	**1258**
P17	113	4	5	67	1	17	26	1141
P20	5	**152**	25	5	1	20	10	**1259**
P23	5	**152**	25	5	1	20	**156**	**1261**
P26	5	**152**	25	5	1	20	**157**	**1262**
P27	5	11	57	33	1	6	10	**1254**
P28	5	11	57	33	1	6	26	1093
P30	5	**152**	25	5	1	20	26	**1260**
P33	5	**152**	25	5	1	20	**157**	**1262**
P36	6	5	36	63	4	30	**157**	**1264**
P37	5	11	57	33	1	6	3	1146
P38	5	8	57	6	1	33	3	**1253**
P39	5	5	57	13	2	40	**157**	**1263**
P40	16	22	5	3	2	15	**157**	**1265**
P41	11	20	1	65	4	4	2	**1255**
P42	89	**153**	**123**	**136**	30	24	3	**1266**
P43	11	20	1	65	4	4	47	**1256**
P46	18	11	57	5	1	20	26	1149
P47	5	11	57	33	1	6	3	1146
P49	5	11	57	33	1	6	3	1146

a) New alleles and new sequence types described are indicated in bold.

1.3.4. MLST genes phylogeny

The sequences of the 7 MLST genes were concatenated and the corresponding phylogenetic tree was constructed (Fig. 1.1). Different phylogenetic clusters were detected, corresponding to the MLST groupings. The *P. aeruginosa* isolates which genomes have been sequenced were also included in the analysis. The closest related sequenced isolate to the four isolates of ST-1146 (SD9, P37, P47 and P49) was the strain UCBPP-PA14 (99.65% of similarity), followed by strains NCGM2.S1 and 39016 (99.44%), LESB58 (98.91%), and PAO1, PACS2 and M18 (98.77%).

1.3.5. Population structure

The Splits Tree representation of the MLST data demonstrated a complex network distribution of the clinical and environmental isolates. Applying an eBURST analysis, 5 clonal complexes or BG groupings were detected, together with a high predominance of singletons (23 singletons of 53 isolates). Singletons were defined as strains that did not share 6 alleles with any other STs in the dataset of the 7 *loci* studied (Fig. 1.2). The ST-1146 isolates clustered in a BG group (BG-2) together with isolates P27 (ST-1254) and P28 (ST-1093); the 5 environmental isolates of BG-2 were isolated from samples taken from well 10. The interrelationship among the STs is shown in the minimum spanning tree representation (Fig. 1.3).

Other environmental isolates classified in STs previously published by other authors were also included in the analysis: 11 isolates from Mediterranean samples (Maatalah *et al*., 2011), 45 from Australia (Queensland) (Kidd *et al*., 2012) and 7 from the Pacific Ocean (Khan *et al*., 2008). All of these 63 STs were different to the 18 STs detected in the environmental isolates of the present study.

A graphical representation of the whole MLST database, with 1339 STs, including our isolates, also demonstrated a high dispersion of the STs found in this study (Fig. 1.S1). No other ST in the database was in the BG-2 (Fig. 1.S1). Only 8 of the 23 singletons previously found remained singletons when our 37 STs were analyzed with the whole MLST database.

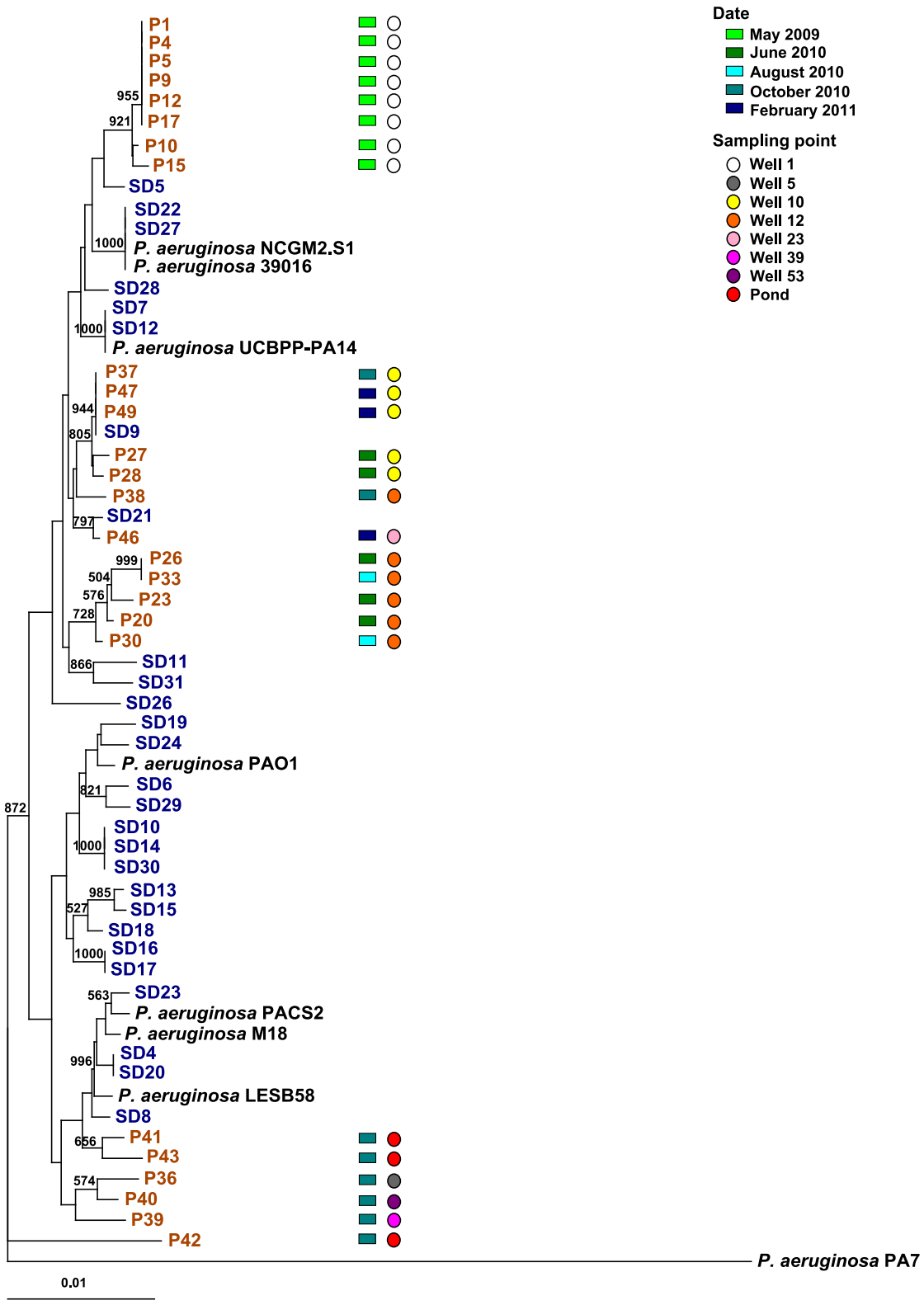


Figure 1.1: Phylogenetic relationship between the clinical and environmental isolates of this study and 8 *P. aeruginosa* retrieved from the National Center of Biotechnology Information. The gene distances of the seven MLST concatenated genes were calculated using the Jukes-Cantor method, and phylogenetic tree was generated by neighbor-joining. Bootstrap values of more than 500 (from 1000 replicates) are indicated at the nodes. Blue letters indicates clinical isolates and brown letters environmental isolates.

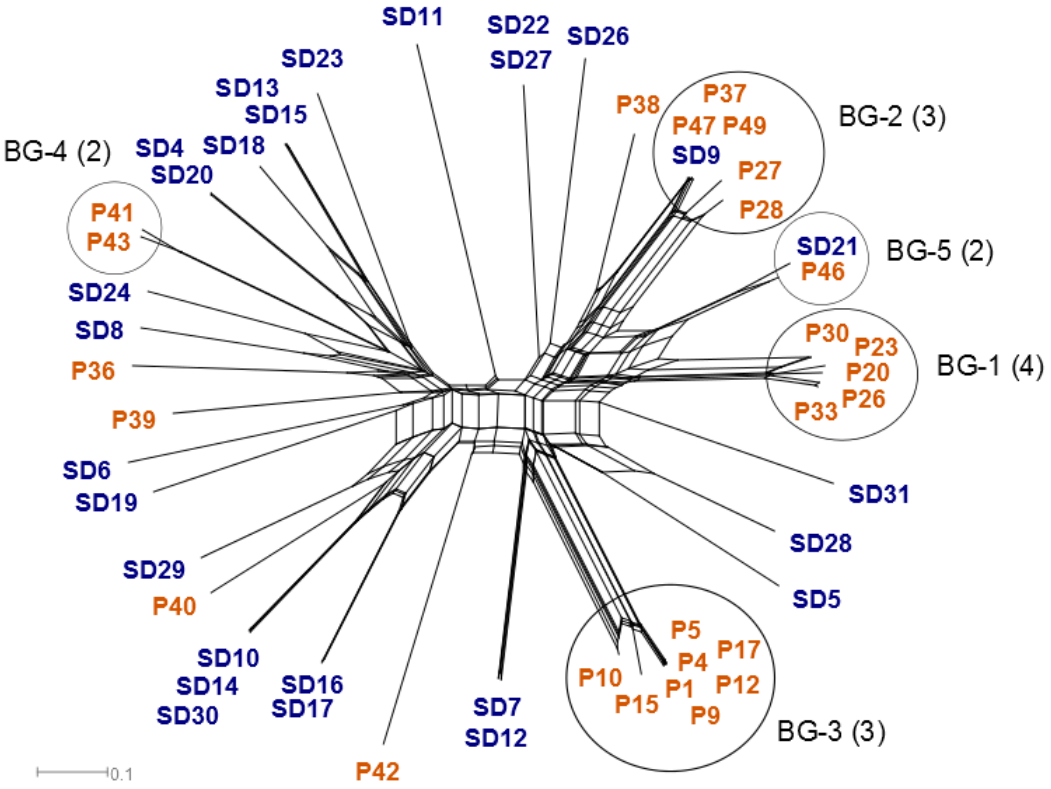


Figure 1.2: SplitsTree showing the distribution of all the sequence types obtained for the 53 *P. aeruginosa* isolates studied. Groups defined with eBURST algorithm, Burst Groups (BG), are indicated with a circle. In brackets are indicated the number of different sequence type in each BG. Blue letters indicates clinical isolates and orange letters environmental isolates.

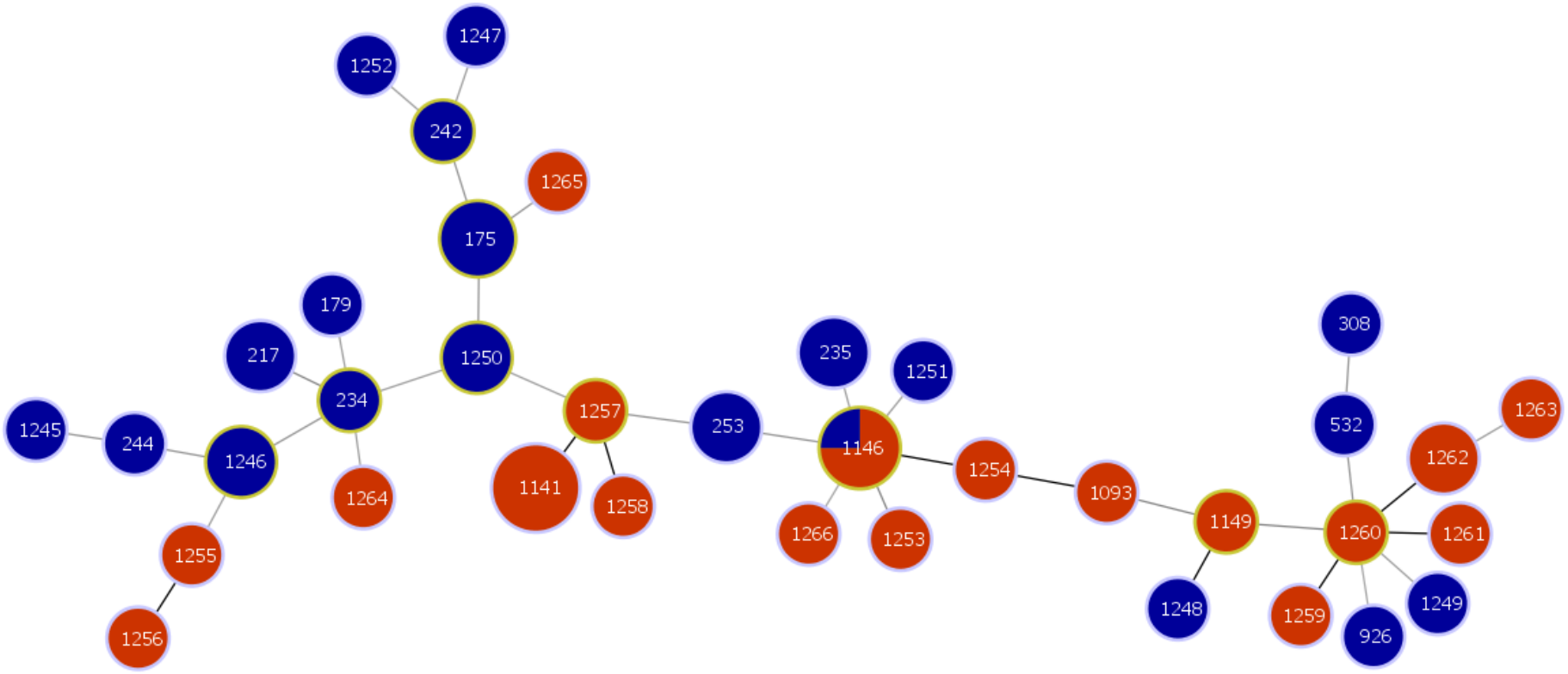


Figure 1.3: goeBURST minimum spannig tree representation of environmental (brown) and clinical (blue) *P. aeruginosa* MLST of this study.

1.3.6. Genetic diversity index

The genetic diversity values showed a high discriminative index for the 53 strains of this study: 0.977 (with a confidence interval 95% CI of 0.958-0.997) for clinical isolates, 0.942 (with a confidence interval 95% CI of 0.88-1.0) for the environmental isolates and 0.978 (with a confidence interval 95% CI of 0.961-0.995) for the whole set of STs. These data indicate a very high probability of obtaining *P. aeruginosa* isolates of different STs if 2 random isolates are collected from clinical or environmental samples (97.7% for clinical, 94.2% for environmental and 97.8% for both).

1.3.7. Clonality and recombination tests

Several indices have been used to analyze the predominance of recombination over mutation. The indices of standardized association, I_A^S, is a test of clonality, and showed the extent of a linkage equilibrium within a population by quantifying the amount of recombination among a set of sequences and detecting association between alleles at different loci. In the populations analyzed, linkage disequilibrium was detected: 0.47 for environmental strains, 0.19 for clinical strains and 0.26 for environmental and clinical strains together. I_A^S data demonstrated that recombination played a key role in the distribution of alleles.

The value obtained by the ClonalFrame method (r/m) for the environmental strains was 0.36 (95% CI: 0.052-0.89), whereas it was 0.002 (95% CI: 0.000159-0.009263) in clinical strains. For the 53 strains together, the value was 1.25 (CI 95% 0.50-2.24). These results reflected the fact that nucleotide substitutions in the 7 housekeeping genes studied are less frequent as a result of de novo mutation rather than recombination.

1.4. Discussion

In the publicly available *Pseudomomas* MLST database the number of STs in which water isolates are described is very scarce, only 99 from a total of 1595. Approximately half of the 99 STs (47) contained isolates from water and clinical samples.

In the present study, the population structure of *P. aeruginosa* environmental and clinical strains isolated from the same geographical region has been studied. The population structure of 53 isolates was analyzed by the MLST method, and the clonality and recombination of the isolates under study was assessed.

MLST is a powerful approach for understanding the evolution of the core genome and the processes that shape strains diversity. Therefore, it was taken as a first approach in the analysis. The high number of new alleles and new sequence types found (36% new STs in clinical and 63% new STs in environmental isolates) reflects the great diversity of *P. aeruginosa* populations. The calculated genetic diversity index also demonstrated the high diversity of the population under study. It is worth pointing out that the 8 singletons detected remain singletons when analyzed together with all the STs available at the MLST database (1339 STs) Fig. S1. The detection of a high number of new alleles not previously described in the MLST database for *P. aeruginosa* has also been reported in previous studies on environmental and clinical isolates from bigger geographical regions, Queensland, Australia (69%) and from Maryland, USA (60%) (Johnson *et al*., 2007; Kidd *et al*., 2012). None of the new environmental STs reported by Kidd, Johnson and colleagues were detected in our environmental isolates. However, several clinical isolates belonged to STs previously known to be present in clinical settings. These STs were: ST-235 (NCGM2.S1) (Viedma *et al*., 2009), ST-253 (PA14) (García-Castillo *et al*., 2011; Kidd *et al*., 2012; Wiehlmann *et al*., 2007) and ST-175, a widely dispersed clone in Mediterranean countries (Maatallah *et al*., 2011) and specifically reported in Spanish hospitals (García-Castillo *et al*., 2011). Other widely spread STs not detected in our study were ST-17 (ClonC) or ST-27, which are not present in Mediterranean countries or in Spanish hospitals (García-Castillo *et al*., 2011). None of the prevalent STs described in clinical settings have been found in our study of environmental samples.

ST-175 is widely distributed over the world and was the most frequently detected ST among the clinical specimens in this study: 3 isolates obtained from 2 patients. This ST was also the most prevalent in the studies of Cholley and colleagues (2011), García-Castillo and colleagues (2011) and was also detected by Johnson and colleagues (2007). One of the isolates of our study was MDR and the other 2 from the same patient were XDR and possibly PDR. ST-175 has been reported as a contaminant of hospital environments, a colonizer of the respiratory secretions of cystic fibrosis patients, and a strain associated with other multiresistant isolates of *P. aeruginosa*.

ST-1146 was the only ST found both in clinical and environmental samples in our study. A clear phenotypic trait that differentiated our clinical and environmental isolates was the antibiotic resistance pattern: the 3 environmental isolates are non-MDR (P37, P47, P49 isolates), while SD9 isolate presented a MDR-resistance pattern. Another clinical isolate of ST-1146 isolated recently from a clinical specimen in a different Majorcan hospital (PaC50) was found to be non-MDR (Gomila *et al*., 2013).

The MLST phylogenetic tree (Fig. 4) showed that the clinical isolates are located in

different branches, separated from the environmental ones, with only one exception: members of ST-1146 which included isolates P37, P47, P49 and SD9. The Splits Tree representation showed clinical and environmental strains in a complex net distribution (Figure 1), in clearly separated independent branches. Again, the only exception was ST-1146. The MLST phylogeny, based in only 7 genes, accurately reflected the phylogeny of the whole genomes.

Twenty-three of the 53 isolates were grouped in 5 clonal complexes, 3 of them (BG-1, BG-3 and BG-4) consisting only of environmental strains, and 2 of them (BG-2 and BG-5) of environmental and clinical isolates. Only 5 environmental isolates were singletons, while most clinical isolates were singletons (18 STs) with the only exception being members of ST-1146 and ST-1248. This is an indication of the high diversity of environmental STs compared with their clinical counterparts. In the 7 housekeeping genes analyzed, the most abundant allele in the clinical isolates was different from the most abundant allele in environmental isolates, and with a few exceptions, environmental and clinical STs did not share more than 2 alleles, indicating some type of isolation in their genealogies.

Recombination tests and clonality studies were performed with the complete set of the 53 isolates. The index of association I_A^S and the clonalFrame test (r/m), together with the structure of the Splits Tree and e-Burst representation, suggest that in *P. aeruginosa* populations, recombination is the prevalent evolutionary driving force, as it has been reported in other more extensive environmental studies (Kidd *et al*., 2012; Maatallah *et al*., 2011). The high values obtained in those studies could be due to the higher number of samples analyzed. Nevertheless, their data are concordant in all cases with the high diversity of new STs.

The presence of high number of new STs in a limited area reflected the high plasticity of a paradoxically phylogenetic conserved genome of *P. aeruginosa.* The genetic diversity found did not depend on the extension of the geographical area sampled, demonstrating that local populations are highly diverse. However, the antibiotic resistance pattern is a phenotypic trait highly correlated with isolates of clinical origin and absent in most of the environmental strains.

1.5. Supplementary data

Supplementary Table 1.S1a: Antibiotic profile and resistance pattern from the clinical *P. aeruginosa* strains analyzed.

	ANTIBIOTICS[a]										
Strain	**AMK**	**GEN**	**TOB**	**IMP**	**MEM**	**CAZ**	**FEP**	**CIP**	**ATM**	**TZP**	**Resistance pattern[b]**
SD4	S	S	S	S	S	S	S	S	S	S	**Non MDR**
SD5	S	S	S	S	S	S	S	S	S	S	**Non MDR**
SD6	S	S	S	S	S	S	S	S	S	S	**Non MDR**
SD7	S	S	S	S	S	S	S	S	S	S	**Non MDR**
SD8	S	I	S	S	S	S	S	S	S	S	**Possible MDR**
SD9	S	S	S	R	I	R	I	I	R	R	**MDR, possible XDR**
SD10	S	R	R	R	R	I	I	R	I	I	**XDR, possible PDR**
SD11	S	S	S	R	R	R	R	R	S	S	**MDR**
SD12	S	S	S	R	S	S	S	S	S	S	**Possible MDR**
SD13	S	S	S	S	S	S	S	S	S	S	**Non MDR**
SD14	S	R	R	R	R	R	I	R	I	I	**XDR, possible PDR**
SD15	S	S	S	S	S	S	S	S	S	S	**Non MDR**
SD16	I	R	S	R	R	R	R	R	R	R	**XDR, possible PDR**
SD17	S	S	S	S	S	S	I	R	S	S	**Possible MDR**
SD18	S	S	S	S	S	S	S	S	S	S	**XDR, possible PDR**
SD19	S	S	S	S	S	S	S	S	R	S	**Possible MDR**
SD20	S	S	S	S	S	S	S	S	I	S	**Possible MDR**
SD21	S	S	S	S	S	S	I	R	I	S	**MDR**
SD22	S	S	S	S	S	S	S	S	S	S	**Non MDR**
SD23	I	R	R	R	R	R	R	R	S	R	**MDR, possible XDR**
SD24	S	S	S	R	R	S	S	S	I	S	**Possible MDR**
SD26	S	S	S	S	S	S	S	R	S	S	**Possible MDR**
SD27	S	S	S	I	S	S	S	I	S	S	**Possible MDR**
SD28	S	S	S	S	S	S	I	S	R	S	**Possible MDR**
SD29	S	S	S	S	S	S	S	S	S	S	**Non MDR**
SD30	S	R	R	R	S	S	S	R	S	S	**MDR**
SD31	S	S	S	S	S	S	S	R	S	S	**Possible MDR**

a) Antibiotics tested: Amikacin (AMK), Gentamicin (GEN), Tobramycin (TOB), Imipenem (IMP), Meropenem (MEM), Ceftazidime (CAZ), Cefepime (FEP), Ciprofloxacin (CIP), Aztreonam (ATM) and Piperacilin/Tazobactam (TZP). Results are indicated as S: sensitive; I: intermediate, R: non-susceptibility. To establish the resistance pattern, Intermidiate resistance (I) was assumed as non-susceptibility (R).
b) Non MDR: non multidrug-resistant; MDR: multidrug-resistant; XDR: extensively drug-resistant; PDR: pandrug resistant; possible MDR: possible multidrug-resistant; possible XDR: possible extensively drug-resistant; possible PDR: possible pandrug resistant. (Magiorakos *et al.* 2011).

Supplementary Table 1.S1b: Antibiotic profile and resistance pattern from the environmental *P. aeruginosa* strains analyzed.

	ANTIBIOTICS[a]										
Strain	**AMK**	**GEN**	**TOB**	**IMP**	**MEM**	**CAZ**	**FEP**	**CIP**	**ATM**	**TZP**	**Resistance pattern[b]**
P1	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P4	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P5	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P9	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P10	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P12	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P15	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P17	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P20	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P23	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P26	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P27	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P28	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P30	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P33	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P36	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P37	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P38	S	S	S	S	S	S	S	S	I	S	**Possible MDR**
P39	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P40	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P41	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P42	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P43	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P46	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P47	S	S	S	S	S	S	S	S	S	S	**Non MDR**
P49	S	S	S	S	S	S	S	S	S	S	**Non MDR**

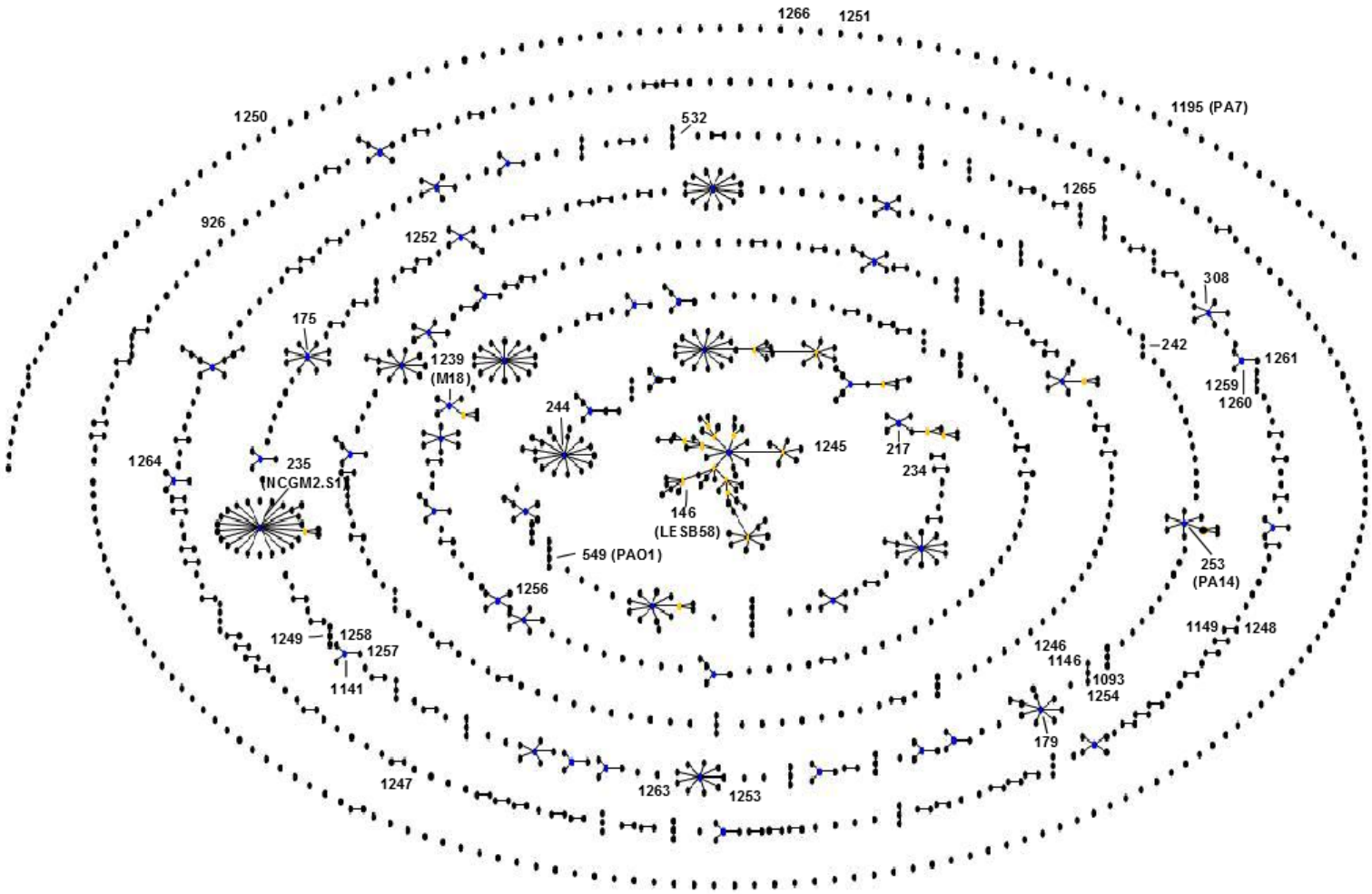
a) Antibiotics tested: Amikacin (AMK), Gentamicin (GEN), Tobramycin (TOB), Imipenem (IMP), Meropenem (MEM), Ceftazidime (CAZ), Cefepime (FEP), Ciprofloxacin (CIP), Aztreonam (ATM) and Piperacilin/Tazobactam (TZP). Results are indicated as S: sensitive; I: intermediate, R: non-susceptibility. To establish the resistance pattern, Intermidiate resistance (I) was assumed as non-susceptibility (R).

b) Non MDR: non multidrug-resistant; MDR: multidrug-resistant; XDR: extensively drug-resistant; PDR: pandrug resistant; possible MDR: possible multidrug-resistant; possible XDR: possible extensively drug-resistant; possible PDR: possible pandrug resistant. (Magiorakos *et al.* 2011).

Supplementary Table 1.S2. Genetic diversity of the selected *loci* among the 53 *P. aeruginosa* isolates analyzed in this study.

		Number of alleles			Allele more frequent			Polymorphic sites			Discriminatory index [Confidence interval (95% CI)]					
gene	**bp**[a]	**ES**[b]	**CS**[c]	**ALL**[d]	**ES**	**CS**	**ALL**	**ES**	**CS**	**ALL**	**ES**		**CS**		**ALL**	
acsA	390	7	16	19	5	28	5	10	14	17	0.708	[0.581 - 0.834]	0.909	[0.822 - 0.995]	0.886	[0.834 - 0.938]
aroE	498	8	8	12	152/11	5	4/5	24	16	26	0.831	[0.756 - 0.906]	0.815	[0.718 - 0.912]	0.851	[0.807 - 0.895]
guaA	373	6	13	16	57	5	5	9	14	15	0.769	[0.689 - 0.85]	0.895	[0.813 - 0.976]	0.851	[0.785 - 0.918]
mutL	442	9	10	14	67	3	3/5	15	8	17	0.834	[0.755 - 0.912]	0.832	[0.725 - 0.939]	0.883	[0.843 - 0.924]
nuoD	366	4	11	13	1	4	1	6	11	15	0.403	[0.178 - 0.628]	0.866	[0.797 - 0.935]	0.71	[0.592 - 0.829]
ppsA	370	9	11	16	17	14/6	6	15	9	17	0.834	[0.755 - 0.912]	0.912	[0.87 - 0.953]	0.903	[0.869 - 0.938]
trpE	443	7	12	17	26	3	3	19	22	25	0.803	[0.722 - 0.884]	0.883	[0.805 - 0.962]	0.888	[0.837 - 0.94]
MLST[e]	2882										0.942	[0.88 - 1.0]	0.977	[0.958 - 0.997]	0.978	[0.961 - 0.995]

a) base pairs; **b)** ES: Environmental strains; **c)** CS: Clinical strains; **d)** ALL: All strains; **e)** MLST: the 7 genes concatenated.



Supplementary Figure 1.S1: eBurst representation of all 1339 sequence types from the MLST *Pseudomonas aeruginosa* database. Sequence types of this study and of the *P. aeruginosa* genomes retrieved from the National Center of Biotechnology Information were indicated in the representation.

Chapter 2

Genome analysis of environmental and clinical *P. aeruginosa* isolates from Sequence Type ST-1146

Abstract

The genomes of four *Pseudomonas aeruginosa* isolates of the newly described sequence type ST-1146, three environmental isolates (P37, P47 and P49) and one clinical (SD9) isolate, with differences in their antibiotic susceptibility profiles have been sequenced and analysed. The genomes were mapped against the reference genomes of *P. aeruginosa* PAO1-UW and UCBPP-PA14. The allelic profiles have been grouped into functional categories. The highest number of differences belonged to the categories “Related to phage, transposon or plasmid” and “Secreted factors”. The clinical isolate showed a number of exclusive alleles 2.5 times greater than that for the environmental isolates.

The ORF of the four genomes were also assembled *de novo* and compared with the CD-HIT program. Results showed that the number of isolate-specific genes was higher in isolate D9 (195 genes) than in isolates P37, P47 and P49 (64, 38, and 61, respectively). Isolate P37 presented three exclusive genes related to the insertion elements of the IS1, IS3 and IS5 families. Some genes related to phage Pf1 and to other phages similar to bacteriophages F116 and H66 were found in isolate SD9 but not in the other isolates of ST-1146. The Pf1 region in isolate SD9 accumulated the highest number of mutations in comparison with PAO1. Comparative genomic methods indicated that the isolates of ST-1146 are closely related, and most genes implicated in pathogenicity are highly conserved in the environmental isolates, suggesting the genetic potential for infectivity similar to that of the clinical isolate.

2.1. Introduction

Pseudomonas aeruginosa is a Gram negative, aerobic, rod-shaped G*ammaproteobacteria* with polar inserted flagella. Environmental isolates of this ubiquitous bacterium are highly versatile and adapt easily to a large variety of natural ecosystems, although water is considered to be the primary habitat of this microorganism (Mavrodi *et al*., 2006). *P. aeruginosa* can cause a wide range of opportunistic infections in animals and humans (Pirnay *et al*., 2009). Some of these infections are associated with high rates of mortality in immunocompromised patients (Speert, 2002), and *P. aeruginosa* is a frequent cause of infections acquired by patients during hospitalisation (Cross *et al*., 1983). The colonisation of this broad spectrum of habitats is a result of the ability to exploit many different nutrition sources and the high potential to adapt to new (or changing) environmental conditions (Ramos, 2004).

The genomes of *P. aeruginosa* strains are larger than those of most sequenced bacteria, varying from 5.2 to 7.1 Mbp (Schmidt *et al*., 1996). This degree of variation has important implications for the methods used to study the evolution and epidemiology of this organism. Recent studies have suggested that more than 80% of the genome of the sequenced strain PAO1 is shared (with only 0.5% nucleotide divergence) by the cystic fibrosis (CF) and environmental strains (Spencer *et al*., 2003), indicating that the genome of this bacterium is characterised by a combination of relatively conserved core genes and variable accessory genes.

The divergence in genome size is caused by the so-called accessory genome. The major part of the genome, the core genome, is found in all *P. aeruginosa* strains with the respective DNA generally collinearly arranged (Römling *et al*., 1995). The core genome, with a few exceptions of loci that are subject to diversifying selection, is highly conserved among clonal complexes and shows sequence diversities of 0.5–0.7% (Spencer *et al*., 2003; Lee *et al*., 2006; Cramer *et al*., 2011). The elements of the accessory genome can be present in subgroups of the *P. aeruginosa* population but might also occur only in single strains (Klockgether *et al*., 2007; Wiehlmann *et al*., 2007). The individual composition of the accessory genome accounts for most of the intra- and inter-clonal genome diversity in *P. aeruginosa.* The elements of the accessory genome have apparently been acquired by horizontal gene transfer from different sources, including other species or genera. Upon integration into the host chromosome these elements appear as “foreign” blocks in the core genome. Therefore, a *P. aeruginosa* chromosome is often described as a mosaic structure of a conserved core genome frequently interrupted by the inserted portions of the accessory genome. The individual mosaics also

show remarkable plasticity (Klockgether *et al*., 2011). The ongoing acquisition of new foreign DNA, larger or smaller deletion events, mutations of single nucleotides and even chromosomal inversions (Römling *et al*., 1997; Ernst *et al*., 2003; Kresse *et al*., 2003; Smith *et al*., 2006; Klockgether *et al*., 2010; Cramer *et al*., 2011) are potentially affecting portions of the core and the accessory genome, and these processes continuously modify the genome and modulate the phenotype of a *P. aeruginosa* strain, thus differentiating the strains from each other.

P. aeruginosa strains have been preferentially studied in CF patients in the clinical context. Fewer environmental studies have been conducted than clinical studies (Khan *et al*., 2008; Kidd *et al*., 2012; Maatallah *et al*., 2011). The ability of *P. aeruginosa* to adapt to different habitats provides an excellent model for examining the mechanisms used by environmental strains of the ubiquitous *P. aeruginosa* species. The genomic structure of *P. aeruginosa* strains has also been analysed predominantly in the clinical isolates of patients with CF. Klockgether and collaborators (2011) have suggested that the sequencing of strains from environmental habitats should provide an unbiased overview of the genetic repertoire of the *P. aeruginosa* populations.

Few environmental strains have been sequenced thus far. Of the 62 *P. aeruginosa* genomes available on the NCBI website, only 1 strain was isolated from soil, and 1 strain was isolated from a water sample. In a screening of *P. aeruginosa* isolated from water and clinical specimens in Mallorca (Spain), sequence type ST-1146 was interesting because this sequence type was the only that included environmental and clinical isolates. In the present study, 4 genomes of intimately related isolates from this sequence type (ST-1146), 3 isolates from water samples (Mallorca, Spain) and 1 clinical respiratory, non-CF isolate, obtained at the Son Dureta University Hospital (Mallorca, Spain), were selected to be studied by comparative genomics. ST-1146 has the allelic profile 5-11-57-33-1-6-3 for the seven genes *acsA, aroE*, *guaA*, *mutL*, *nuoD*, *ppsA* and *trpE*, which was established by Curran and collaborators (Curran et al., 2004) and is available in the *P. aeruginosa* MLST database (http://www.pseudomonas.com/). The environmental isolates could be differentiated from the clinical isolate by the antibiotic susceptibility profile according to Magiorakos and collaborators (2012). The environmental isolates were non-multidrug resistant (non-MDR), and the clinical isolate was multidrug resistant and possibly extensively drug resistant (MDR, possible XDR) (resistant to aztreonam, ceftazidime, imipenem and piperacillin-tazobactam).

The genomic analysis of these four isolates was focused on three main aspects: a) the presence of exclusive and differentiating genes; b) the presence of mutations when compared with strains *P. aeruginosa* PAO1 and UCBPP-PA14 as references (the allele distribution in gene

categories, allelic profile comparisons) and the presence of nucleotide substitutions in ST-1146 isolates; and c) the analysis of genes that are involved in pathogenicity.

2.2. Materials and Methods

2.2.1. *P. aeruginosa* isolates

Environmental *P. aeruginosa* isolates were isolated on Cetrimide Agar (Merck) as the selective medium from 2 subsurface water samples taken from the same well in Santa Margalida (Mallorca, Spain). One sample was taken in October 2010 (isolate P37), and the second sample was taken in February 2011 (isolates P47 and P49). The clinical isolate (SD9) was isolated on MacConkey agar plates (bioMérieux) at 37°C from a patient´s ulcer. The four isolates were assigned to ST-1146 following previously described methods (Gomila *et al*., 2013).

2.2.2. *De novo* assembled genome analysis

The draft genome sequence of the P37, P47, P49 and SD9 isolates were obtained using the reads from Illumina HiSeq 2000 paired-end libraries. The reads were *de novo* assembled using the Newbler Assembler v2.7 program (Roche). The Whole Genome Shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accessions AMVN00000000 (SD9), AMVO00000000 (P37), AMVP00000000 (P39) and AMVQ00000000 (P49). The versions described are the first version.

2.2.3. Average Nucleotide Index (ANIb) and conserved DNA

ANIb and conserved DNA values from the 4 genomes in this study and other previously sequenced genomes of *P. aeruginosa* strains were calculated using the work package JSpecies (Ritcher *et al*., 2009).

2.2.4. Core and Pan-genome analysis

The Pan-genome analysis was performed as previously described (Özen *et al.*, 2012) using the Biotools for Comparative Microbial Genomics (CMG-Biotools). The BLAST Matrix and the Pan-Core genome plot construction were performed as suggested previously (Özen *et al.*, 2012; Tettelin *et al*., 2005) to display to what extent gene families were conserved within a set of 8 *P. aeruginosa* genomes available at NCBI and the 4 isolates (P37, P47, P49 and SD9) in our study. Two proteins with at least 50% sequence identity over at least 50% of their length were

considered as belonging in the same gene family.

2.2.5. Gene prediction and clustering analysis

The gene prediction was performed using the Metagenemark program (Zhu *et al*., 2010). Clustering analysis was initially performed with isolates P37, P47, P49 and SD9 and later included two *P. aeruginosa* reference genomes, PAO1-UW (NC_002516.2)(PAO1) and UCBPP-PA14 (CP000438) (PA14), using the CD-HIT program available at CD-HIT home page (Li *et al*., 2006; Huang *et al*., 2010; Fu *et al*., 2012). Two proteins with at least 50% sequence identity over at least 50% of the protein length were considered to belong to the same gene family. Proteins exclusive to each isolate and the proteins shared between two or more isolated were counted and represented in Venn diagrams.

2.2.6. Genome mutational profiles

The original lllumina reads of the 4 genomes sequenced were processed using the GS Reference Mapper software package version 2.6 (Roche Inc). High quality Illumina sequencing reads were aligned using both *P. aeruginosa* PAO1 and PA14 as reference genomes. Variants with respect to both reference sequences were identified with the GS Reference Mapper software (AllDiff and HCDiff reports). The presence of variant candidates (alleles) were detected using the high-confidence method (i.e., observed in >80% of the reads). The detected variants and positions were extracted for the purpose of comparison from the HCDiff output file. The GFF files, obtained from the *Pseudomonas* Database (Winsor *et al.,* 2011) from both reference genomes were used to annotate and count variants to generate the mutational gene profiles.

2.2.7. Polymorphic sites and nucleotide substitutions

The above original lllumina reads of the 4 genomes were aligned with strain PAO1 as reference with the Mauve program (Darling *et al*., 2010). Regions not present in the five genomes were discarded. Polymorphic sites between P37, P47, P49 and SD9 were calculated from this alignment with the DnaSP package, version 5.0 (Librado *et al*., 2009) and were localised along the chromosome.

2.2.8. Phylogenetic reconstructions

Phylogenetic reconstructions were performed as previously described (Mulet *et al*., 2009).

2.3. Results

2.3.1. Whole genome characteristics of isolates in ST-1146

The Illumina reads obtained for the four isolates in ST-1146 were *de novo* assembled and the draft genome annotated. The main features for each assembly and annotation prediction are provided in Table 2.1. The number of contigs ranged from 112 to 161. In all of the isolates, the percentages of bases with a consensus quality score of at least 40 (Q40), were higher than 99.9%. Based on the Metagenemark annotation prediction, between 5,841 and 5,873 ORFs were detected in the environmental isolates, and 5,972 ORFs were detected in the clinical isolate SD9.

2.3.2. Average Nucleotide Index (ANIb) and conserved DNA

The 4 isolates sequenced in this study showed a high similarity among them, with pair-wise ANIb values ranging from 99.95 to 99.99%. The three *P. aeruginosa* genomes closer to these four isolates were UCBPP-PA14, NCGM2.S1 and PAO1 with a range of 99.17-99.18 99.10-99.11 and 98.81-98.82%, respectively. The results of the ANIb analysis and the indices of conserved DNA are given in the Table 2.2. The indices of conserved DNA for the 4 isolates in ST-1146 with PAO1 ranged from 93.56 to 94.30 and were highly similar between them (98.49-99.55). No significant differences among the 3 environmental and clinical isolates were detected in these whole genome comparisons.

2.3.3. Core and Pan-genome analysis

The core/pan-genome analysis was calculated on 12 genomes (4 isolates of the ST-1146 and 8 *P. aeruginosa* genomes available in the GenBank). The results obtained from the 12 *P. aeruginosa* genomes analyzed (Supplementary Table 2.S1) were plotted from the first to the last column in an information accumulative manner (Fig. 2.1). Each column shows the BLAST hit results against all the previous ones in terms of new genes, new gene families, and pan/core genome sizes. The accumulative number of all gene families found increased as new genomes were added, leading to a pan-genome of 8093 genes in total, whereas the number of common gene families found in all genomes decreased with genome addition, leading to a core genome formed by 4556 genes after the addition of the last genome (*P. aeruginosa* PA7).

Table 2.1: Data of sequenced genomes of ST-1146 *P. aeruginosa* isolates of this study.

	P37	P47	P49	SD9
Ecological setting	Environmental	Environmental	Environmental	Clinical
Source	Subsurface water	Subsurface water	Subsurface water	Ulcer
Origin	Well 10	Well 10	Well 10	Patient 6
Date	October 2010	February 2011	February 2011	February 2009
Resistance pattern	Non MDR	Non MDR	Non MDR	MDR, possible XDR
Number of bases enssambled (bp)	6237921	6236369	6232858	6301986
N50 Contig Size	108312	140229	85677	76889
Largest Contig Size	298247	499099	310648	247959
Q40 Plus Bases (%)	99.93	99.96	99.94	99.93
Q39 Minus Bases (%)	0.07	0.04	0.06	0.07
GC-content (%)	66.61	66.63	66.63	66.56
Numbers of large contigs	148	112	145	161
Average contig size	42148	55681	42985	39142
Number of ORFs[a]	5873	5841	5871	5972
Genes[b]	5838	5834	5856	5914
Number of CDSs[b]	5798	5778	5806	5871
Total CDSs size (bp)	5579342	5556210	5577057	5603711
Coding %	89.4	89.1	89.5	88.9
Average CDS length (nt)	962	962	961	955
tRNAs[b]	37	53	47	40
rRNA (clusters)[b]	3	3	3	3
Number of genes with assigned function[b]	3887	3865	3890	3902
Number of genes without assigned function[b]	1911	1913	1916	1969

a) Data stablished with the MetaGene Mark program; b) Data stablished with the Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) from the National Center for Biotechnology Information (NCBI).

Table 2.2: Conserved DNA (upper right) and ANIb matrix data (lower left) of the genomes of the ST-1146 *P. aeruginosa* isolates of this study and from 8 *P. aeruginosa* genomes retrieved from the National Center of Biotechnology Information.

Conserved DNA /ANIb	ST	SD9	P37	P47	P49	PAO1	UCBPP-PA14	LESB58	M18	NCGM2.S1	39016	PACS2	PA7
SD9	1146	---	98.5	98.58	98.49	93.56	92.28	91.5	94.04	90.98	86.36	92.87	83.93
P37	1146	99.96	---	99.43	99.47	94.09	92.83	91.98	94.58	91.53	86.87	93.48	84.42
P47	1146	99.96	99.99	---	99.55	94.30	92.91	92.11	94.67	91.69	86.99	93.62	84.41
P49	1146	99.95	99.99	99.99	---	94.13	92.79	92.03	94.63	91.58	86.81	93.51	84.48
PAO1	549	98.81	98.82	98.81	98.82	---	93.80	93.84	95.47	91.80	86.74	94.40	85.34
UCBPP-PA14	253	99.17	99.17	99.18	99.18	98.80	---	91.22	92.27	92.12	87.34	92.67	85.00
LESB58	146	98.60	98.62	98.62	98.62	99.24	98.52	---	94.21	90.85	86.38	93.29	82.70
M18	1239	98.69	98.71	98.71	98.71	99.28	98.60	99.18	---	91.13	86.22	94.25	84.85
NCGM2.S1	235	99.11	99.11	99.11	99.10	98.67	98.91	98.43	98.81	---	91.82	91.55	83.30
39016	235	99.07	99.09	99.06	99.08	98.62	98.85	98.41	98.77	99.74	---	86.75	78.98
PACS2	1394	98.72	98.73	98.74	98.73	99.21	98.54	99.04	99.39	98.69	98.64	---	84.76
PA7	1195	93.24	93.24	93.27	93.24	93.19	93.12	93.06	93.31	93.20	93.20	93.27	---

Data between the 4 isolates sequenced in this study were marked in blue. The three closest genomes from the ANIb data were marked in orange. The three closest genomes from the conserved DNA were marked in yellow.

The gene content of the 4 isolates included in the ST-1146, as indicated in Table 4, ranged between 5789 (environmental isolate P47) and 5904 (clinical isolate SD9) genes. The whole genome cluster defined by the ST-1146 isolates shared 5405 genes (between 91.5% and 93.4% of the respective genomes), and the number of genes shared by the 3 environmental isolates was slightly higher (5434 genes), representing the 93.7 % (P37 genome), 93.9 % (P47 genome) and 93.6 % (P49 genome).

An analysis of the genetic novelty contributed by the genomes of each isolate of the ST-1146 through the progressive assessment of the pan-genome resulted in the observation that clinical isolate SD9 had the highest input among the 4 isolates, with 168 new genes. This value is slightly lower in comparison with the contribution of other *P. aeruginosa* genomes, which ranged between 179 new genes (*P. aeruginosa* M18) and 577 (*P. aeruginosa* PA7). The final pan-genome for ST-1146 is composed of 5758 genes.

Based on the criteria established for clustering of gene families in the pan-genome of *P. aeruginosa*, it was obtained a BLAST proteomic matrix that summarizes the results of the pair-wise comparison of genomes, including percentages and absolute numbers (Fig. 2.S1). The percentages in the BLASTMatrix were calculated by the number of gene families shared by 2 genomes divided by the gene families found in both. According to these data, homologies between the 4 proteomes derived from the genomes of ST-1146 isolates ranged from a minimal value of 95.0 % (SD9 *versus* P49) and a maximal value of 97.9 % (P47 versus P49). The proteome of isolate SD9 always showed less homology when compared against the proteomes of the 3 environmental isolates of ST-1146. When the ST-1146 environmental isolates were compared among themselves, the homology values were above 97.7%. When compared with the rest of the *P. aeruginosa* strains studied, the most closely related in terms of homologies was *P. aeruginosa* PAO1. Effectively, the 3 environmental isolates shared a similar fraction of protein family levels with this strain: 90.2% homology with the proteome of isolate P47, 89.7% with the P37 proteome and 89.6% for P49. The most closely related with SD9 is *P. aeruginosa* M18, which shares 88.4% of its proteome.

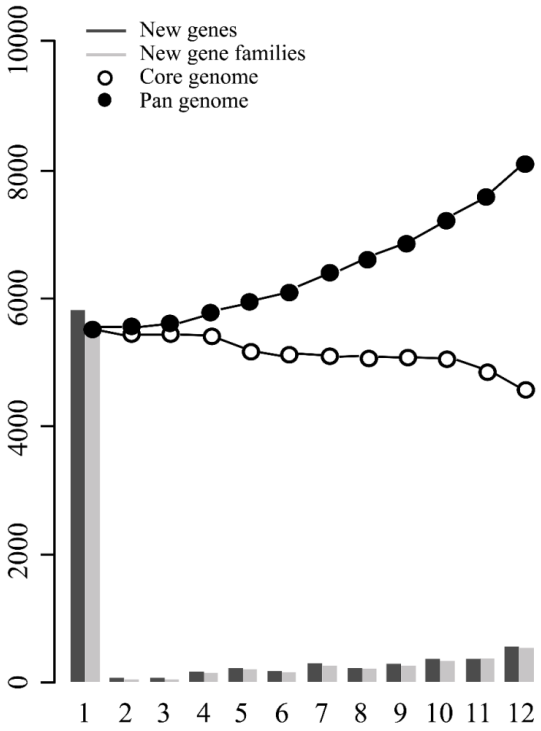


Figure 2.1: Pan/core-genome of ST-1146 *P. aeruginosa* isolates of this study and *P. aeruginosa* retrieved from the National Center of Biotechnology Information. 1: P37, 2: P47, 3: P49, 4: SD9, 5: PAO1, 6: M18, 7: PACS2, 8: UCBPP-PA14, 9: LESB58, 10: NCGM2.S1, 11: 39016, 12: PA7. The pan-genome is indicated with black circles and core-genome with white circles. The number of discovered novel genes (*dark bars*) and novel gene families (*light-grey bars*) are also shown for each added genome.

2.3.4. Gene comparisons with CD-HIT

Cluster comparisons of the four isolates in ST-1146 alone and with the reference genomes PAO1 and PA14 were performed using the program CD-HIT. Venn diagrams were generated to visualise the cluster distributions (Fig. 2.2). A total of 6,200 or 6,403 clusters were determined when PAO1 or PA14 were included in the analysis. As depicted in Figure 2.2, 5,038 genes were shared by the 4 isolates of ST-1146 and PAO1, and 5,150 genes were shared with PA14. Four-hundred and forty-eight protein clusters were present in the isolates of ST-1146 and not in PAO1, and 241 proteins were PAO1 strain specific (Fig. 2.2a). The numbers of strains- specific proteins in the ST-1146 isolates when PAO1 was included in the analysis were 241, 191, 64, 38 and 63 in strains PAO1, SD9, P37, P47 and P49, respectively. When strain PA14 was compared with the isolates of ST-1146, the numbers of strains-specific proteins were 441, 189, 64, 37 and 63 in strains PA14, SD9, P37, P47 and P49, respectively (Fig. 2.2b).

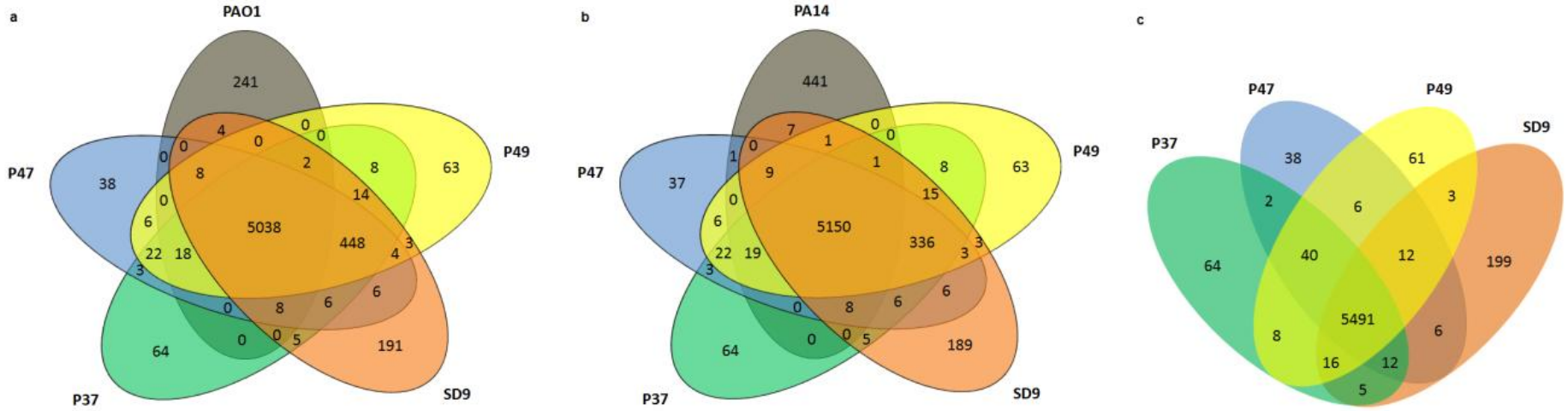


Figure 2.2: Venn diagram showing the number of shared and exclusive genes of isolates P37, P47, P49 and SD9 based on the Cd-hit results, referred to a) *P. aeruginosa* PAO1; b) *P. aeruginosa* PA14; c) ST-1146 isolates.

The four sequenced genomes of the ST-1146 isolates shared 5,491 protein clusters from the total of 5,959 analysed (Fig. 2.2c). The number of proteins shared by the three environmental isolates was slightly higher (5,573 proteins). The isolate-specific proteins were higher in SD9 (199) than in P37, P47 and P49 (64, 38, 61, respectively). The isolate-specific genes that were located at the beginning or at the end of a contig were 59 in SD9 (30%), 37 in P37 (57%), 19 in P47 (50%) and 36 in P49 (59%) (Table 2.3). It was assumed that these genes were incompletely sequenced, and they were not further analysed. The remaining isolate-specific genes were found in internal regions of the contigs (140, 27, 19 and 25 genes in isolates SD9, P37, P47 and P49, respectively). Genes considered relevant to the different characteristics between the four isolates (i.e., related to phages, antibiotic resistance, the ABC transport system, flagellation, or insertion elements) were studied exhaustively. Hypothetical proteins, those with unknown functions, and non-essential genes were not included. In total, 65, 6, 4 and 2 genes were studied in SD9, P37, P47 and P49, respectively (Table 2.3).

Table 2.3: Exclusive genes of ST-1146 isolates

Isolate	Exclusive genes	Truncated gene	Completed genes	Hypothetical proteins	Not hypothetical genes	Analysed genes
P37	64	37	27	9	18	6
P47	38	19	19	4	15	4
P49	61	36	25	5	20	2
SD9	199	59	140	27	113	69

Indels were detected in 11 sequences (3, 3, 2 and 3 sequences in P37, P47, P49 and SD9, respectively). More than one indel per gene were detected in some sequences. Three genes of isolate P37, related to insertion sequences from the IS1, IS3 and IS5 families, were not found in the other genomes. Two genes of SD9 were related to bacteriophage Pf1 (hypothetical proteins), and 59 other genes related to bacteriophages (most of them to bacteriophagues F116 and/or H66) were not found in the other genomes and will be discussed in a specific section. The isolate-specific proteins exhaustively analysed for the four isolates are shown in Supplementary Table 2.S2.

2.3.5. Genome mutational profiles and allele comparisons

The variant detection was performed using GS Reference Mapper and a conservative approach (high sequence read stringencies), resulting in the generation of high confidence variant detection (HCDiff) containing high confidence SNP sites and indels. The GFF files from the PAO1 and PA14 reference genomes were used to count and annotate the mutational profiles obtained for each comparison (Supplementary Table 2.S3 and 2.S4).

Nucleotide substitutions in the coding genes

According to mutational profiling, the nucleotide substitutions in the coding genes of isolates in the ST-1466 genomes compared with PAO1 were 49,931 in SD9, which was higher than in the environmental isolates P37, P47 and P49 (49,545; 49,444 and 49,578, respectively). Compared with PA14, the number of substitutions was lower: 32,387 in SD9, 31,982 in P37, 31,964 in P47, and 32,045 in P49. From a total of 5,565 genes present in strain PAO1, 5,126 genes with nucleotide substitutions were detected in ST-1146 isolates. ST-1146 isolates shared 4,506 (63.7%) identical genes (4067 genes identical between them but different to PAO1 and 439 genes identical between them and PAO1). From a total of 5,892 genes in PA14, 4,841 genes presented nucleotide substitutions in ST-1146 isolates.

Isolate-exclusive alleles

The alleles exclusive to each isolate and the alleles shared between two or more isolates obtained from the two mutational profiles of ST-1146 were counted and represented in Venn diagrams (Figure 2.3). The allele number compared to PAO1 and present in only one isolate and not in others was as follows: isolate P37, 205 alleles (4.0%); isolate P47, 196 alleles (3.8%); isolate P49, 211 alleles (4.1%); and finally, isolate SD9, 548 alleles (10.7%). The number of alleles exclusive to isolate SD9 (548) was approximately 2.5 times higher than the exclusive alleles in the environmental isolates, which shared 424 alleles not present in SD9. Compared to PA14, the allele numbers present in only one isolate were slightly lower, although isolate SD9 presented the highest number (476) of isolate-exclusive alleles (an average of 3.6 times higher than the environmental isolates)(Supplementary Table 2.S5 and 2.S6).

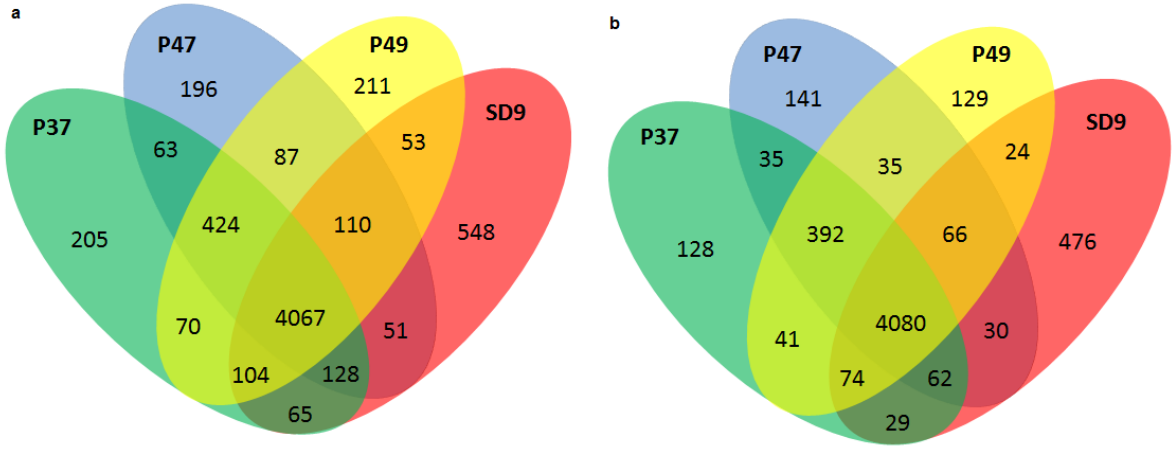


Figure 2.3: Venn diagram showing the number of shared and exclusive alleles of isolates P37, P47, P49 and SD9 referred to a) *P. aeruginosa* PAO1; b) *P. aeruginosa* PA14.

Functional categories

All of the genes of the ST-1146 isolates studied in the mutational profile were grouped by functional categories (Fig. 2.4, Supplementary Table 2.S5 and 2.S6). The exclusive alleles detected for each isolate were classified in 27 functional categories according to the PseudoCAP functional classification. All of these data are related to the total number of genes that are not identical to PAO1 and PA14 in each category, and the ratio among the alleles has been calculated (Fig. 2.4, Table 2.S5 and 2.S6). The genes grouped in the functional category "Membrane proteins" presented the highest absolute values of nucleotide substitutions and alleles in all ST-1146 isolates for the mutational profile compared with PAO1 and the functional category "Transport of small molecules" when compared with PA14. The categories corresponding to "Hypothetical, unclassified, unknown" and "Putative enzymes" were not considered. The percentage of exclusive alleles for each isolate (p, q, r, s) in each category (n) was calculated for the genes not identical to PAO1 or PA14. In the mutational profile of PAO1, the categories of genes "Related to phage, transposon or plasmid" and "Secreted factors" predominated in all isolates (range 10-22% in the environmental isolates and 29-33% in the clinical isolate SD9). In PA14, in all cases, except for P49, nucleotide substitutions in these two categories were also predominant. For P49, the predominant categories were "Related to phage, transposon or plasmid" (11%) and "Motility and attachment" (10%).

The percentage of the exclusive alleles classified by categories for the comparison with PAO1 ranged from 0-33% in isolate SD9, from 0-22% in isolate P37, from 0-20% in isolate P47 and from 0-19% in isolate P49 (Fig. 2.4 and Supplementary Table 2.S5).

The ratios of the different alleles between SD9 and each environmental isolate were calculated for each PseudoCAP functional class to determine if the nucleotide substitutions were balanced between the functional classes (Supplementary Table 2.S5 and 2.S6). For the comparison with PAO1, the highest ratio of different alleles between SD9 and P37 was 8 times higher in the category "Two-component regulatory systems"; the ratio was 8 times higher between SD9 and P47 in the categories "Amino acid biosynthesis and metabolism", "Nucleotide biosynthesis and metabolism" and "Two component regulatory systems"; and the ratio was 15 times higher between SD9 and P49 in the "Energy metabolism" category. In isolate SD9, the ratio of exclusive alleles in most of the categories was 2-15 times higher than in the environmental isolates. In the comparison with PA14, the highest ratio of different alleles was 7 times higher between SD9 and P37 in the categories "Two-component regulatory systems" and "Biosynthesis of cofactors"; the ratios were 8 times higher between SD9 and P47 in "Energy metabolism" and "Protein secretion/export apparatus"; and the ratio between SD9 and P49 was 10 times higher for "Two-component regulatory systems". In isolate SD9, the

ratio of exclusive alleles in most of the categories was 2-11 times higher than in the environmental isolates.

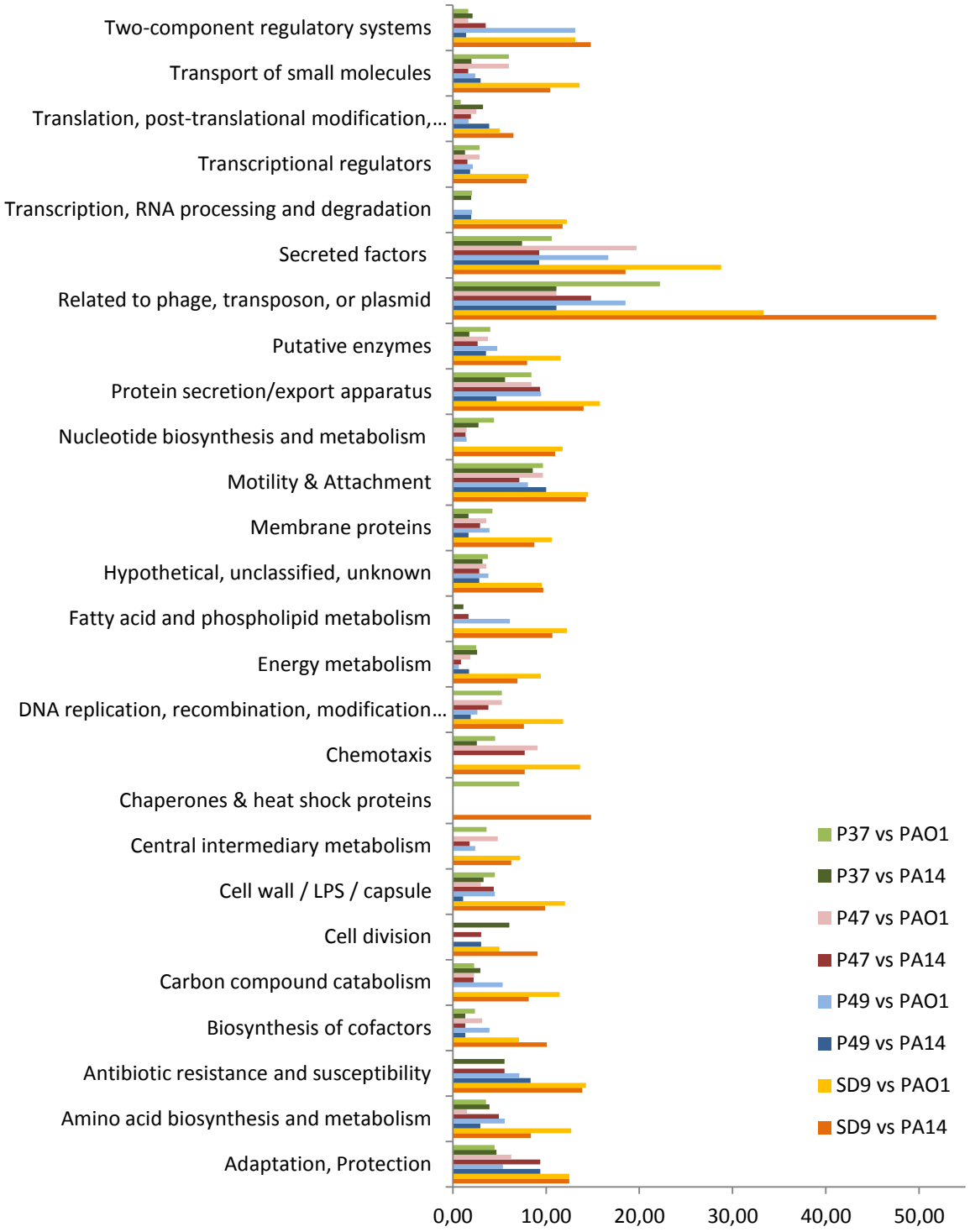


Figure 2.4: PseudoCAP Functional Classification of the exclusive alleles of isolates P37, P47, P49 and SD9 when compared with *P. aeruginosa* PAO1 and *P. aeruginosa* PA14.

The ratios of the different alleles between SD9 and each environmental isolate were calculated for each PseudoCAP functional class to determine if the nucleotide substitutions were balanced between the functional classes (Supplementary Table 2.S5 and 2.S6). For the

comparison with PAO1, the highest ratio of different alleles between SD9 and P37 was 8 times higher in the category “Two-component regulatory systems”; the ratio was 8 times higher between SD9 and P47 in the categories “Amino acid biosynthesis and metabolism”, “Nucleotide biosynthesis and metabolism” and “Two component regulatory systems“; and the ratio was 15 times higher between SD9 and P49 in the “Energy metabolism” category. In isolate SD9, the ratio of exclusive alleles in most of the categories was 2-15 times higher than in the environmental isolates. In the comparison with PA14, the highest ratio of different alleles was 7 times higher between SD9 and P37 in the categories “Two-component regulatory systems” and “Biosynthesis of cofactors”; the ratios were 8 times higher between SD9 and P47 in “Energy metabolism” and “Protein secretion/export apparatus”; and the ratio between SD9 and P49 was 10 times higher for “Two-component regulatory systems”. In isolate SD9, the ratio of exclusive alleles in most of the categories was 2-11 times higher than in the environmental isolates.

All 14 genes in the category “Antibiotic resistance and susceptibility” were identical in isolates P37 and P47, only one different allele was found in isolate P49, and different alleles were found in SD9 compared with PAO1. Some of these alleles were also different with PA14: 2 (P37, P47), 3 (P49) and 5 (SD9).

2.3.6. Nucleotide polymorphisms

The intraclonal diversity of the members of the clonal complex ST-1146 was also studied by comparing the nucleotide sequences of the four isolates mapped against the reference genome PAO1 to determine the nucleotide polymorphic site distributions in the isolates. For that purpose, the sequences located at the ends of the contigs and the sequences not present in all five genomes were discarded. The number of polymorphic sites for the 5 genomes was 56,626. The total number of nucleotide substitutions was 56,657; therefore, practically all of the polymorphic sites (99.95%) presented only 1 nucleotide substitution. The polymorphic sites for the 4 ST-1146 genomes were 1,056, and the total number of substitutions was 1,072. In the 1,072 polymorphic sites, 624 nucleotides were identical in the environmental isolates and different in SD9. The polymorphic sites resulting from a different nucleotide in P37, P47 and P49 were 147, 150 and 151, respectively. A description of the polymorphic sites is shown in Table 2.4.

Table 2.4: Distribution of nucleotide substitutions referred to the genome nucleotides of the isolates P37, P47, P49 and SD9.

SNPs	P37	P47	P49	SD9	Total
A	33	34	34	184	285
T	32	27	25	158	242
G	40	34	43	139	256
C	39	45	44	134	262
-	3	10	5	9	27
Total	**147**	**150**	**151**	**624**	**1072**

A representation of the nucleotide substitutions along the chromosome was created for each isolate (Fig. 2.5). The distribution along the chromosome was different for each studied isolate. The distribution of the polymorphic sites was not homogenous; some regions accumulated a higher number of nucleotide substitutions. The three environmental isolates showed a similar distribution of approximately 150 substitutions. A specific region of polymorphic sites was more evident in the SD9 isolate. In SD9, 139 substitutions (22.3%) were located in an 8,531 bp region. When this plateau was analysed in more detail, a total of 71 substitutions were found in intergenic regions, and 68 were located in genic regions, which is very similar to the genes of phage Pf1 of PAO1.

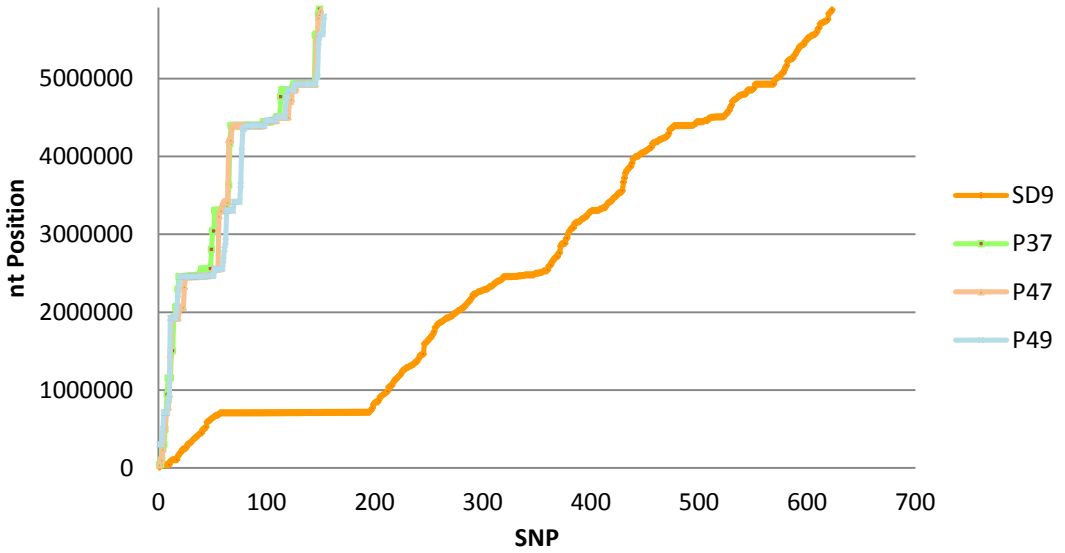


Figure 2.5: Distribution of nucleotide substitutions along the genome of the isolates P37, P47, P49 and SD9.

2.3.7. Pathogenic factors

Exoenzymes

The cytotoxin-encoding genes *exoS*, an indicator of an invasive phenotype, and *exoU*, a strain-specific gene associated with increased virulence, were included in the virulence factor

list of Wolfgang and collaborators (2003). The gene *exoS* was not found and *exoU* was detected in the 4 isolates and showed a 99.4% similarity with PA14_51530. The absence of *exoS* confirmed the previous reports that *exoS* and *exoU* are mutually exclusive (Table 2.5).

Table 2.5: Exoenzymes detected in the isolates P37, P47, P49 and SD9. The corresponding locus tag of strains *P. aeruginosa* PAO1 and PA14 are indicated.

Exoenzymes	PAO1	PA14	ST-1146	Function
exoT	PA0044	PA14_00560	PRESENT	Exoenzyme
exoY	PA2191	PA14_36345	PRESENT	Exoenzyme
exoU	ABSENT	PA14_51530	PRESENT	Exoenzyme
exoS	PA3841	ABSENT	NOT FOUND	Exoenzyme

PA14 Killing regions

Lee and collaborators (2006) defined 6 common and specific killing regions in strain *P. aeruginosa* PA14 that consist of 9 genes. All the PA14 killing regions were present in ST-1146 isolates, indicating that these regions possessed genes that are required for *Caenorhabditis elegans* killing. The killing genes were identical between the four ST-1146 isolates. The protein PilC had 15 and 56 different amino acids when compared with the corresponding PAO1 and PA14 proteins: PA04527 and PA14_58760 (90% and 86% homology, respectively). PilW could be differentiated by 1 amino acid (99% homology) from the corresponding protein from PAO1 (PA4552) and 132 amino acids (52% homology) from PA14 (PA14_60690). The other killing genes are identical to PA14 (Table 2.6).

Table 2.6: PA14 Killing regions detected in the isolates P37, P47, P49 and SD9. The corresponding locus tag of strains *P. aeruginosa* PAO1 and PA14 are indicated.

Killing regions	PAO1	ST-1146	Function
PA14_03370	ABSENT	PRESENT	Unknown
PA14_23420	ABSENT	PRESENT	O-antigen biosynthesis
PA14_23430	ABSENT	PRESENT	O-antigen biosynthesis
PA14_27680	ABSENT	PRESENT	Unknown
PA14_27700	ABSENT	PRESENT	Putative transcription regulator
PA14_58760	PA04527	PRESENT	pilC
PA14_59010	ABSENT	PRESENT	unknown
PA14_59070	ABSENT	PRESENT	unknown
PA14_60290	PA4552	PRESENT	pilW

Antibiotic resistance genes

Many CF isolates acquire hyper-mutator capabilities by mutations in *mutS* or *mutT*. Isolates of ST-1146 presented a *mutS* gene with 23 nucleotide substitutions compared with PAO1, and there were 10 substitutions for the *mutT* gene in these isolates (Table 2.7).

The major efflux pumps contributing to intrinsic and mutational antibiotic resistance are coded by the operon MexAB-OprM and by the ancillary system MexCD-OprJ. Both were present, and the MexEF-OprN and the MexGHI-OpmD efflux pump were also found. This pump confers resistance to aminoglycoside antibiotics, is required for biofilm formation, facilitates cell to cell communication and promotes virulence and growth in *P. aeruginosa.* In isolate SD9, *mexA* possessed 1 amino acid that was different from the environmental isolates, and P37 had another different amino acid in *mexF*. The comparison of efflux pump genes with PA14 showed a considerably higher number of nucleotide substitutions in *mexD*, with 62 different nucleotides resulting in 10 different amino acids. MexC and MexH protein were identical in ST-1146 isolates and showed 3 and 5 different amino acids from the corresponding protein in PA14 (PA14_09530 and PA14_60850).

Table 2.7a: Antibiotic resistance genes analyzed between ST-1146 isolates

	ST-1146				
	No. of different nucleotides				No. of different aminoacids
Gene	**P37**	**P47**	**P49**	**SD9**	
mexA	0	0	0	1	1
mexB	0	0	0	0	
mexC	0	0	0	0	
mexD	0	0	0	0	
mexE	0	0	0	0	
mexF	1	0	0	0	1
mexG	0	0	0	0	
mexH	0	0	0	0	
mexI	0	0	0	0	
mexT	0	0	0	0	
mexZ	0	0	0	0	
opmD	0	0	0	0	
oprD	0	0	0	1	1
oprJ	0	0	0	0	
oprM	0	0	0	0	
oprN	0	0	0	0	

The *oprD* gene codes for a specialised pore protein OprD, which allows for the selective permeation of basic amino acids and their structural analogues, such as the carbapenem antibiotics imipenem and meropenem. The *oprD* gene presented 134 mutations in the environmental isolates and 135 mutations in SD9, resulting in 28 different amino acids when compared to PAO1 (PA0958) and two deletions in positions 372 and 381. SD9 showed only 1 different amino acid (P, phenylalanine) when compared to the environmental isolates (S, serine) in position 278 (Table 2.7). The OprD protein was more similar to that present in *P. aeruginosa* PA7, an atypical, phylogenetically distant, non-respiratory *P. aeruginosa* strain. The

complete OprD proteins were aligned, and a corresponding UPGMA dendrogram was generated, confirming this result. The similarities among the ST-1146 isolates were 96%, 94% and 93 with respect to PA7, PA14 and PAO1 (Supplementary Figure 2.S2).

Table 2.7b: Antibiotic resistance genes analyzed of ST-1146 isolates compared with *P. aeruginosa* PAO1

Compared with PAO1						
	No. of differents nucleotides				No. of different aminoacids	Locus tag
Gene	P37	P47	P49	SD9		
mexA	16	16	16	17	2; 3	PA0425
mexB	17	17	17	17	1	PA0426
mexC	6	6	6	6	4	PA4599
mexD	12	12	12	12	2	PA4598
mexE	6	6	6	6	1	PA2493
mexF	11	10	10	10	0; 1	PA2494
mexG	2	2	2	2	1	PA4205
mexH	9	9	9	9	2	PA4206
mexI	3	3	3	3	1	PA4207
mexT	12	12	12	12	33	PA2492
mexZ	4	4	4	4	0	PA2020
opmD	15	15	15	15	2	PA4208
oprD	134	134	134	135	28; 29	PA0958
oprJ	8	8	8	8	1	PA4597
oprM	18	18	18	18	0	PA0427
oprN	12	12	12	12	1	PA2495

Table 2.7c: Antibiotic resistance genes analyzed of ST-1146 isolates compared with *P. aeruginosa* PA14

Compared with PA14						
	No. of differents nucleotides				No. of different aminoacids	Locus tag
Gene	P37	P47	P49	SD9		
mexA	11	11	11	12	2; 3	PA14_05530
mexB	14	14	14	14	1	PA14_05540
mexC	13	13	13	13	5	PA14_60850
mexD	62	62	62	62	10	PA14_60830
mexE	4	4	4	4	1	PA14_32400
mexF	7	6	6	6	1; 0	PA14_32390
mexG	2	2	2	2	1	PA14_09540
mexH	11	11	11	11	3	PA14_09530
mexI	12	12	12	12	0	PA14_09520
mexT	5	5	5	5	0	PA14_32410
mexZ	16	16	16	16	2	PA14_38380
opmD	12	12	12	12	3	PA14_09500
oprD	124	124	124	125	29; 30	PA14_51880
oprJ	12	12	12	12	3	PA14_60820
oprM	1	1	1	1	0	PA14_05550
oprN	13	13	13	13	0	PA14_32380

Pyocins

The isolates of ST-1146 presented 4 types of S-pyocins (Table 2.8): *pyoS5* was conserved in all isolates and was also identical with those of strains PAO1 and PA14. Pyocin S4 was present in the 6 strains with divergent nucleotides; *pyoS2* and *pyoS3A* were present and conserved in all of the isolates, but *pyoS2* was absent in strain PA14, and *pyoS3A* was absent in PAO1.

Table 2.8: Pyocins detected in the isolates P37, P47, P49 and SD9. The corresponding locus tag of strains *P. aeruginosa* PAO1 and PA14 are indicated.

Pyocins	PAO1	PA14	ST-1146	Function
pys2	PA1150	ABSENT	NOT FOUND	E2 colicins
imm2	PA1151	ABSENT	NOT FOUND	E2 colicins
pyoS3A	ABSENT	PA14_ 49520	PRESENT	S type pyocin protein
pyoS3I (imm)	ABSENT	PA14_ 49510	PRESENT	immunity protein
S4	PA3866	PA14_13940	PRESENT	S type pyocin protein
pyoS5	PA0985	PA14_ 59220	PRESENT	Ia and Ib
imm S5	PA0984	PA14_59230	PRESENT	immunity protein
R2	PA0622	PA14_08070	PRESENT	R2 pyocin

2.3.8. Phage related genes

The accessory genome is central to *P. aeruginosa* biology as a primary contributor to the genome evolution. The presence of phages and phage-like elements are considered to be reservoirs of genetic diversity. A high number of polimorphics sites were found between SD9 and the environmental strains in the Pf1 genes region (Fig. 2.6). Interestingly, these genes were found in a higher number of contigs (5-6 contigs) in the *de novo* assembled genomes in all ST-1146 isolates. All genes, with the exception of the first hypothetical protein, could be located in the same order as in PAO1. The first hypothetical protein of the Pf1 genes region in PAO1 was presented located in all ST-1146 isolates in a contig flanked by other genes not related to the Pf1. Pf1 is a single stranded DNA, filamentous phage that can be integrated as a prophage into the *P. aeruginosa* genome. Pf1 or Pf1-like phages are highly upregulated during *P. aeruginosa* biofilm development (Webb *et al*., 2004). A local Blast from Pf1 genes of SD9 against PA14 genomes showed that Pf1 genes are present, with the exception of the hypothetical protein that was not present and a low value of similarity of the integrase gene (52% similarity).

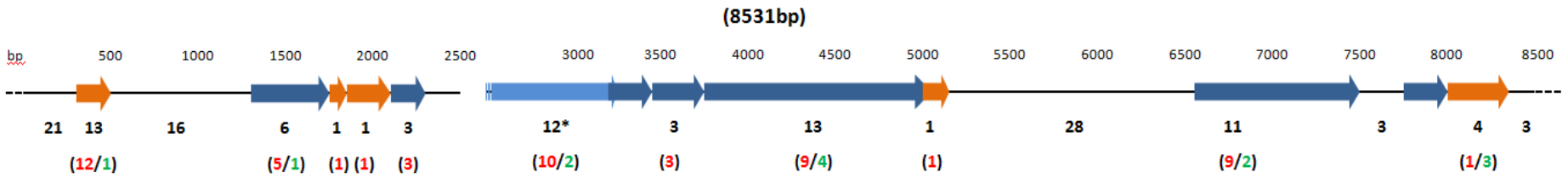


Figure 2.6: Nucleotide substitutions distribution in the hot spot region of original reads of SD9 mapped against PAO1. The genes related to bacteriophage Pf1 found in this region (8531 pb) with known function are indicated in blue; the hypothetical proteins are indicated in orange. The SNPs located in intragenic and genic regions are indicated below each region. Synonymous (red) and non synonymous (green) changes in the protein of the 3 environmental isolates are indicated in brackets**.** *shorter protein in SD9 as a result of one deletion which produces a stop codon.

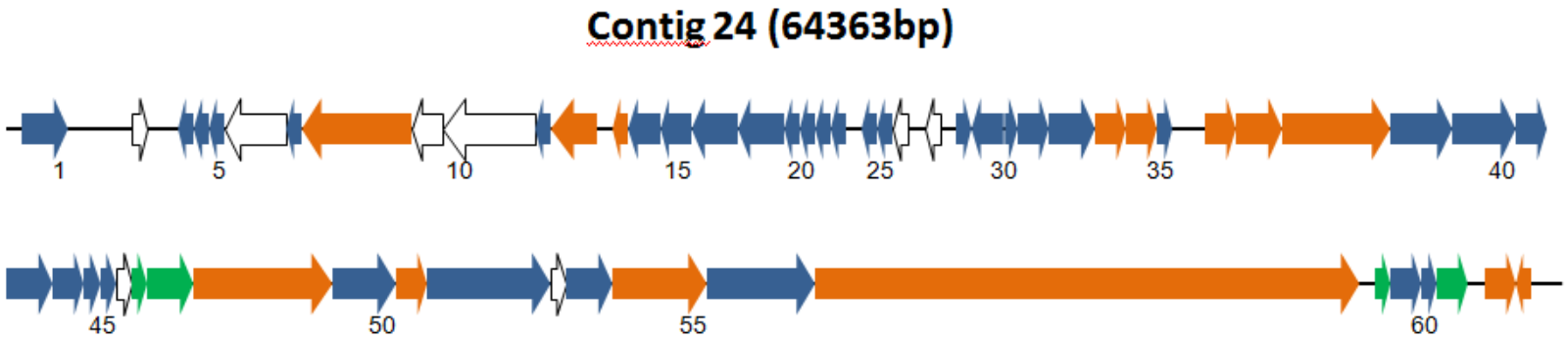


Figure 2.7: Gene map of phage present in SD9 and not present in the environmental isolates. Genes which first hit in BLAST was *P. aeruginosa* phage F116 are indicated in blue, genes which first hit was *Pseudomonas* phage H66 are indicated in orange. Those genes which its first hit could is the same are indicated in green. Genes not found in NCBI database are indicated in white. The numbers of genes are indicated below the genes.

When the exclusive cluster of proteins based on the CD-HIT analysis of each of the isolates was studied, the clinical isolate SD9 presented CDS in a contig of 64,363 bp related to previously described *P. aeruginosa* phages: F116 (65,195 Kb, 70 genes) and H66 (65,270 bp, 71 genes). A BLAST analysis indicated that phage F116 genes were 98% similar, with 76% coverage, and phage H66 was 98% similar with 78% coverage. In total, 53 of 70 genes presumptively belonging to phages similar to F116 or H66 were detected (98% similarity for 77% query coverage). A schematic representation of the phage is presented in Figure 2.7, and the relationship of the genes is shown in Supplementary Table 2.S2.

2.4. Discussion

For the first, time four *P. aeruginosa* genomes, three isolates from water, one of the main environmental *P. aeruginosa* reservoirs, and one clinical isolate, with the same sequence type have been studied by comparative genomics. The main objective of our study was to assess the genomic differences and similarities between closely related isolates of the same sequence type.

Whole genome comparisons based on the CD-HIT, method for clustering and comparing protein or nucleotide sequences, clearly demonstrated that members of ST-1146 were closely related genomically, sharing a high percentage of genes. Thus, the MLST profiling reflects the whole genome and the phylogeny of *P. aeruginosa* isolates. Moreover, the environmental isolates were more closely related among themselves than to the clinical isolate, which could be a consequence of the different habitat adaptation of both ecotypes. In a first approach, BOX and ERIC-PCR profiles of the 4 isolates of ST-1146 demonstrated already differences in their genomes (data not shown). These isolates can be differentiated not only by the gene content but also by the alleles of the shared genes. However, the differences among strains should be considered with respect to not only what genes or alleles are present but also how efficient these genes are regulated or expressed. With the CD-HIT method and the mutational profile analysis, strain PA14 is more closely related to the ST-1146 isolates than PAO1. This is in accord with previous analyses based on multilocus sequence typing, conserved DNA based and ANIb analysis (Table 2.2). Notably, PA14 is considered significantly more virulent than PAO1 (Li *et al.*, 2013; Mikkelsen et al., 2012).

We found a high number of transcriptional regulator genes (at least 437) in the isolates of ST-1146 and a high number of two-component regulatory systems (62), which reveals the

complexity of *P. aeruginosa* metabolic regulation. In general, 17% of *P. aeruginosa* genes are considered to be involved in regulation, but only 4% of the genes in *Escherichia coli* are involved in regulation. The differences in the allelic profile in these 2 gene categories between the 4 isolates might indicate adaptation to the habitat. The environmental isolates share more alleles in common in both categories than with SD9. The number of nucleotide substitutions detected in the clinical isolate is 2.5 and 3.5 times higher than those present in the environmental isolates when compared with PAO1 and PA14, respectively. These results suggest that the mutation pressure is not the same in the environmental isolates than in the clinical one.

Single nucleotide substitutions have been studied in closely related strains isolated from cystic fibrosis patients in Germany (RN3) and in California (PA14) (Klockgether *et al*., 2011). In RN3, 231 single nucleotide substitutions (SNPs) were reported to PA14. The authors suggested that the genes present in RN3 could provide a selective advantage to adapt and persist in CF, accumulating SNPs similar to those present in *retS* (major transcripts regulators), *mexH* (encoding efflux pumps), *pvdD* (siderophore), *cndS* (cyanide) or *phnA* (quinolone). These 5 genes are presented in all environmental or clinical isolates of ST-1146, although only *retS* and *pvdD* presented nucleotide substitutions in ST-1146 isolates. The amino acid change (alanine, A, by threonine, T) reported in RetS in clon C is considered an adaptive mutation during a chronic infection in CF airways (Cramer *et al*., 2011) and was also found in isolate SD9 (position 190) of ST-1146. The environmental isolates of ST-1146 and strains PAO1 and PA14 maintain an A in the same position. The other amino acid substitution detected by Cramer and collaborators (2011) in clon C (arginine, R by tryptophan, W) was not found in any of the ST-1146 isolates. MexH protein was found identical in ST-1146, but different from the corresponding PA14 protein. In our study, MexC and MexD proteins of other efflux pumps were also found identical in ST-1146 but different from the corresponding PA14 proteins. No correlation was found in the nucleotide substitutions in 4 ST-1146 isolates respect the nucleotide substitutions between PA14 and RN3. The intraclonal diversity did not evolve by random drift, but was driven by selective forces that do not affect the same genes in different strains of *P. aeruginosa*.

The category of genes “Related to phage, transposon or plasmid” presented a relevant number of nucleotide substitutions in all of the strains. All genes coding for phage Pf1 proteins were found in the ST-1146 isolates, and the corresponding alleles were identical in the environmental isolates but were different from those present in SD9. These data imply that the specific affected loci were subjected to the same diversifying selection pressure in the environmental isolate but not in the corresponding Pf1 genes of the clinical isolate, which

presents a different mutational rate. The prophage Pf1 genes were found in *P. aeruginosa* strains, as PA14 and RN3 (Klockgether *et al*., 2011) and the Pf1-like genes are considered to be the major mutation hotspot and the most rapidly evolving part of the genome. This is consistent with the view that phages span a high degree of genetic diversity and are prone to frequent lateral gene transfer; in addition, they are also the most activated genes in *P. aeruginosa* biofilms (Klockgether *et al*., 2011; Whiteley *et al*., 2001). The major mutational hotspot for PA14 and RN3 was the phage Pf1, with 87 SNSs (PA14_48890-PA14_49000). In our isolates, this hot spot was also present. Isolate SD9 showed a higher number of nucleotide substitutions respect the environmental isolates. The SNPs in the Pf1 genes of PA14 and RN3 were compared with isolate SD9. The nucleotide substitutions detected in the mutational hot spot region between strains PA14 and RN3 were not detected in isolate SD9.

The main difference in gene content between the environmental isolates and SD9 are the genes similar to the phage F116/H66 genes, which were present only in the clinical SD9. These genes are not present in the PA14 or PAO1 strains. The presence of genes related to the phage F116/H66 could represent a relevant advantage for genomic adaptation to a new host. The presence of phages reflects the genome plasticity of the studied isolates, which have adapted to different hosts. The presence of a Pf1 hot spot already observed in other clinical *P. aeruginosa* strains and the fact that the prophage islands are critical determinants of in vivo competitiveness (Winstansley *et al*., 2009) could justify the high number of nucleotide substitutions detected in the clinical isolate SD9 but not in the environmental isolates. As natural transformation is not encountered in *P. aeruginosa*, lateral gene transfer appears to play a major role for the observed dynamic genome (Maatallah *et al*., 2011)**.** Environmental bacteriophages are most likely responsible for a large amount of Horizontal Gene transfer (HGT) as these bacteriophages are formidable transducers of naturally occurring microbial communities of *P. aeruginosa* (Pirnay *et al*., 2009)**.** HGT may play a more important role than point mutations in the adaptation of *P. aeruginosa*. No two strains would be identical in terms of the gene content, and HGT continuously creates new strains with unique genetic characteristics (Kidd *et al*., 2012; Pirnay *et al*., 2009).

It has been argued that the ancestors of virulent bacteria and the origin of virulence determinants lie mostly in the environmental microbiota (Martínez, 2013). Previous experimental studies on the *P. aeruginosa* pathogenic factors (Lee *et al*., 2006; Wolfgang *et al*., 2003) allowed us to compare those data with the genes present in our isolates. Almost all of the pathogenic factors present and highly conserved in our isolates but are different from PAO1 and PA14. Genes associated with *P. aeruginosa* virulence are almost exclusively found in the category called “Secreted factors” (Wolfgang *et al*., 2003). Additional genes related to the

expression of virulence determinants are found in the categories “Motility and attachment”, “Protein secretion/export apparatus” and “Transcriptional regulators”. The high level of conservation in the 18 strains studied by Wolfgang and collaborators (2003) agrees with our results. The PAO1 genes detected by microarray by Wolfang and collaborators (2003) for the 4 environmental strains resulted in 97.2% of non-redundant known or predicted genes and 97.1% of all of the isolates studied. Our data indicate that 100% of these genes are present in the four genomes studied. The ST-1146 isolates also have *pilC* and *pilW*, which are killing genes, as defined by Lee and collaborators (2006), and are not present in the 4 environmental isolates studied by Wolfgang and collaborators (2003).

P. aeruginosa presents a type III secretion system (T3SS), which is not located in a pathogenicity island and is considered to be an old element in the evolution of this species (Martínez, 2013). T3SS is also not present in other environmental *Pseudomonas* species, such as *Pseudomonas putida* (Silby *et al*., 2011). The exoenzymes that belong to T3SS are *exoU*, a gene associated with increased virulence that makes the strains more cytotoxic to mammalian cells (Sato *et al*., 2004), and *exoS*, which is considered to be the major cytotoxin required for colonisation and dissemination during infection (Pirnay *et al*., 2009). A total of 65% of environmental strains and clinical non-CF strains have an *exoS* cytotoxin, whereas *exoU* is present only in 35% of these strains (Wolfgang *et al*., 2003). All of the ST-1146 isolates possess *exoU* and not *exoS*. The relative conservation of the genomes of these isolates implies that conserved selective pressures contributed to the evolution of these genomes in different environmental niches. Because the host cell contact–dependent type III secretion system secretes both ExoU and ExoS, the advantage afforded by the expression of either gene most likely involves the interaction with a target eukaryotic organism (Wolfgang *et al*., 2003). Our isolates have the PA14 specific killing genes defined by Lee and collaborators (2006), together with the *exo*U gene associated with increased virulence (Pirnay *et al*., 2009). These genes are considered prototypical elements that enhance the pathogenic characteristics in the strain harbouring these genes (Kung *et al*., 2010).

Replacement island genes, such as lipopolysaccharide (LPS) O antigens, pyoverdine, pili, and flagella, are critical determinants of *P. aeruginosa* fitness. The presence of a high number of mutations in these genes (*fglK, pilY1, pilB or pvdD* and *pvdL*) with respect to PAO1 shows that these surface exposed macromolecules are under intense selective pressures to resist phage killing or are in the pyoverdine locus to resist pyocin killing and to escape detection by the host immune system (Kung *et al*., 2010). The *pilC* and *pilW* genes have to be added to this group because these genes are considered to be killing factors. The ST-1146 isolates from water samples have the same high number of mutations with respect to PAO1 and PA14,

showing that this pressure also exists in nature. The same virulence factors required for infecting humans are also required for infecting plants, worms, or insects (Martínez, 2013). Some authors argue that the natural eukaryotic hosts (nematodes, insects, plants and amoebas) are the relevant natural hosts in which the selection and evolution of pathogenic traits occur, and the ability to infect humans is a secondary effect of this interaction. Another aspect could be that the pressure to survive phage attacks is what directs that pressure, giving an advantage in the colonisation of eukaryotes (Lee *et al*., 2006). Our results confirm the assumption of other authors that there are no specific clones selected for a specific disease (or habitat) (Pirnay *et al*., 2009) and that virulence is the result of a pool of pathogenicity-related genes that interact in various combinations in different genetic backgrounds (Lee *et al*., 2006).

Three phenotypic typing methods have been classically applied to discriminate *P. aeruginosa* strains: antibiotic profiles, pyocin typing and siderotyping. All 3 were analysed by comparing the responsible genes.

Isolate SD9 is multidrug resistant (MDR), and OprD is responsible for the selective permeation to carbapenem antibiotics. The *oprD* gene exhibits an unlimited number of alleles typical for a gene that is under strong selection for diversity. Pirnay and collaborators found 21 different defective *oprD* mutations conferring resistance to carbapenem antibiotics in clinical strains (CF and non-CF), and none these mutations were present in the environmental strains of that study. Members of narrow clonal complexes often show identical *opr*D sequences (Pirnay *et al*., 2009). In our study, the 3 environmental isolates showed identical nucleotide sequences for *oprD*, and SD9 differed by only 1 nucleotide. The differences with PAO1 were higher: 134 (P37), 134 (P47), 134 (P49) and 135 (SD9).

Interestingly, the environmental isolates were sensitive to imipenem while SD9 was resistant (Supplementary Table 1.S1a). None of the 21 mutations conferring resistance to carbapenem antibiotics described by Pirnay and collaborators (2009) were found in ST-1146 isolates. The *oprD* genes of the ST-1146 isolates were distant from the PAO1 gene and closer to strain PA7, while the *oprD* of PAO1 and PA14 were similar. It has been postulated that OprD mutations confer resistance to carbapenem antibiotics by not expressing porin, suggesting that OprD is a selective force in the early stages of CF colonisation, although the data obtained by Pirnay and collaborators cannot support this hypothesis. However, alterations in OprD are not the only mechanism for resistance against carbapenem antibiotics (Ocampo-Sosa et al., 2012).

Different types of bacteriocins have been described in *P. aeruginosa*, and bacteriocin typing (production or sensitivity) has been proposed for intraspecies differentiation. All ST-1146 isolates presented the same 4 conserved pyocins and could be distinguished from PAO1 and PA14 because 2 of the 4 pyocins were absent in these strains (1 pyocin in each strain). In

this way, ST-1146 isolates are combining several types and activities of colicins type E2 and E5 with DNAse activity or types Ia and Ib with the channel forming activity from pyoS5. The presence of a high number of pyocins could indicate the need for increased competition in a complex habitat.

P. aeruginosa does not require many *de novo* mutations to conquer a new habitat (Klockgether *et al*., 2011). *P. aeruginosa* can choose between distinct strategies in different types of infections (Kiil *et al*., 2008), which could have occurred in isolate SD9.

Our results are in accord with the argument that classical virulence determinants might be relevant for basic metabolic processes or in modulating predator-prey relationships (toxins) in natural, non-infective ecosystems (Martínez, 2013).

The extensive conservation of virulence genes in the genomes regardless of the clinical source suggests that the disease-causing ability of this opportunistic pathogen relies on a set of highly conserved pathogenic mechanisms. The conservation of virulence gene determinants also extends to the environmental reservoir, where a population is constantly changing but conserving the necessary tools to survive (Wolfgang *et al*., 2003). The clinical isolate SD9 has no specific gene that distinguishes this isolate from the environmental isolates, although a higher pressure for mutations could be detected, due most probably which is most likely due to the pressure of the eukaryotic host.

The genomic comparison of the isolates of this study with strains PAO1 and PA14 led us to conclude that not all of the genes of the genome are subject to the same evolutionary forces, as demonstrated by the following: a) The genomes are highly conserved between our isolates, including those genes studied classified as pathogenic factors, which shows that these genes are necessary to colonise environmental and clinical habitats. b) Although ST-1146 isolates have characteristics similar to strain PAO1, these isolates present well-known genes that enhance the pathogenic characteristics of powerful infective strain, such as PA14, which was the closest related strain in the phylogenetic analysis by MLST and ANIb. c) Several genes are not commonly present in all strains, such as pyocianins (from the type T3SS) or *exoU*, present in PA14 and ST-1146 but not present in PAO1.d) Some genes in ST-1146 are more similar to other *P. aeruginosa* strains (not PAO1 and PA14), such as *oprD* gene, more similar to the corresponding of *P. aeruginosa* PA7, a distant *P. aeruginosa* strain, an outlier from the species that is not considered *P. aeruginosa* by some researchers. e) Other differences in the accessory genome are represented by the presence of phages, such as F166/H66, or phages with high mutation rates in the clinical isolate, such as Pf1, which could be partially responsible for the acceleration of adaptation of SD9 when the isolate encounters a eukaryotic host. This result supports the hypothesis that phages are influencing factors that shape the genome of

the clinical strains. The environmental isolates, similar to the clinical isolates, have different potential degrees of infectiveness.

We consider the strains in the species *P. aeruginosa* to constitute a monophyletic phylogenetic branch in the genus (Mulet *et al*., 2012) with continuous changes in the populations and individuals. However, its integrity as a species is still maintained. Our study of the ST-1146 isolates showed that the genes implicated in pathogenicity are present and mostly conserved, and the environmental isolates must be considered potential pathogens.

2.5. Supplementary data

Supplementary Table 2.S1: Core and Pan-genome analysis results

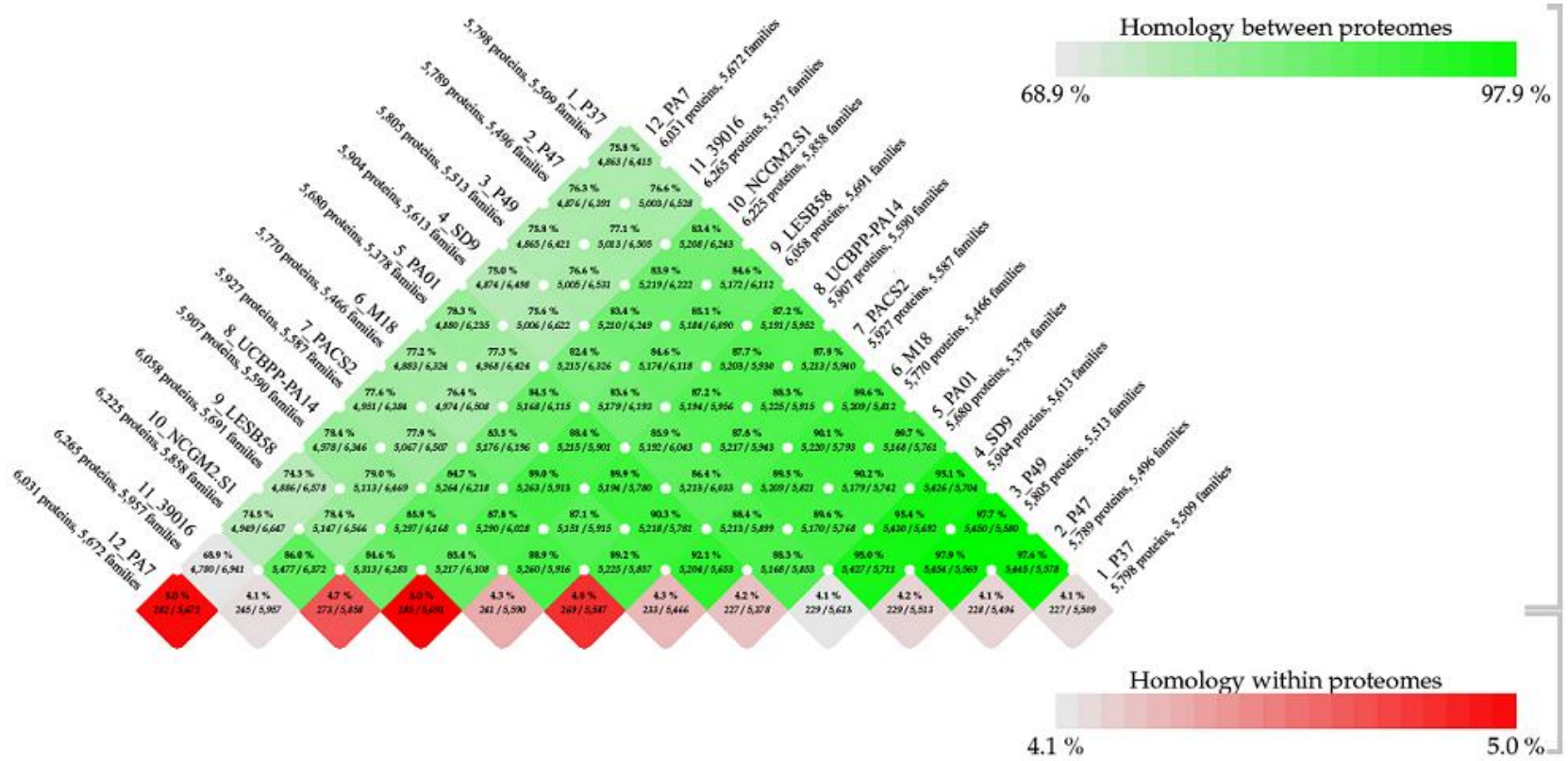
Strain	Total genes	New genes	New families	Pan genome	Core genome
P37	5798	5798	5509	5509	5509
P47	5789	60	49	5553	5443
P49	5805	46	43	5594	5434
SD9	5904	168	165	5758	5405
P. aeruginosa PAO1	5680	205	191	5941	5148
P. aeruginosa M18	5770	179	168	6106	5113
P. aeruginosa PACS2	5927	307	283	6386	5103
P. aeruginosa UCBPP-PA14	5907	223	223	6603	5074
P. aeruginosa LESB58	6058	288	264	6860	5055
P. aeruginosa NCGM2.S1	6225	368	359	7216	5039
P. aeruginosa Paer39016	6265	382	374	7583	4846
P. aeruginosa PA7	6031	577	532	8093	4556

Supplementary table 2.S3: Mutational profile of isolates P37, P47, P49 and SD compared with PAO1.

In extra CD.

Supplementary table 2.S4: Mutational profile of isolates P37, P47, P49 and SD compared with PA14.

In extra CD.



Supplementary Figure 2.S1: Genomic analysis of predicted proteome of ST-1146 *P. aeruginosa* isolates of this study and *P. aeruginosa* retrieved from the National Center of Biotechnology Information. The BLAST matrix, was based on pairwise whole-genome comparison. The percentage of similarity for each combination is presented by green color between genomes and red color shows similarity within the same genome.

Supplementary Table 2.S2: Description of the exclusive genes analysed in ST-1146 isolates.

Isolate	Gene_id[a]	Length (aminoacids)	Contig	Present in other ST-1146 isolates	Exclusivity reason	Description Blast Result	Query coverage (%)	Similarity (%)
P37	1985	164	41	YES	Deletion 1nt	probable transcriptional regulator	100	99
P37	2896	115	52	YES	Deletion 2nt	putative amino acid ABC transporter	100	100
P37	3647	766	73	YES	Deletion 2nt	putative ABC-type multidrug transport system	100	99
P37	5852	338	134	NO		insertion sequence (IS5 family)	100	100
P37	5856	167	137	NO		insertion element (IS1 family)	100	99
P37	5860	158	140	NO		insertion sequence (IS3 family)	100	100
P47	365	170	7	YES	Deletion 2nt	flagellar basal body rod protein FlgF	100	99
P47	366	114	7	YES	Part of gene 365	flagellar basal body rod protein	99	100
P47	3724	1170	50	YES	Deletion 2nt	PvdL; peptide synthase	99	99
P47	4899	166	66	YES	Deletion 1nt	putative outer membrane porin	100	100
P49	3450	224	70	YES	Deletion 2nt	putative secretion protein	100	99
P49	5863	135	138	YES	Deletion 1nt	putative ABC-type amino acid transporter	100	99
SD9	9	94	4	YES	Short fragment	*Pseudomonas* phage Pf1	100	98
SD9	34	149	7	YES	Deletion 1nt	putative clavaminic acid synthetase	99	97
SD9	632	410	24	NO		*P. aeruginosa* phague F116	99	90
SD9	633	115	24	NO		unknown		
SD9	634	53	24	NO		*P. aeruginosa* phague F116	100	98
SD9	635	140	24	NO		*P. aeruginosa* phague F116	100	97
SD9	636	40	24	NO		*P. aeruginosa* phague F116	93	100
SD9	637	483	24	NO		*Pseudomonas* phage PAJU2	27	98
SD9	638	143	24	NO		*Pseudomonas* phage phi297; *P. aeruginosa* phague F116	83	87
SD9	639	602	24	NO		*Pseudomonas* phage H66	60	93
SD9	640	206	24	NO		*Pseudomonas* phage YMC/01/01/P52_PAE_BP	100	99
SD9	641	703	24	NO		*Pseudomonas* phage phi297	100	99
SD9	642	62	24	NO		*P. aeruginosa* phague F116	100	98
SD9	643	407	24	NO		*Pseudomonas* phage H66	74	98

Isolate	Gene_id[a]	Length (aminoacids)	Contig	Present in other ST-1146 isolates	Exclusivity reason	Description Blast Result	Query coverage (%)	Similarity (%)
SD9	644	166	24	NO		*Pseudomonas* phage H66	100	98
SD9	645	201	24	NO		*P. aeruginosa* phague F116	100	99
SD9	646	184	24	NO		*P. aeruginosa* phague F116	100	99
SD9	647	338	24	NO		*P. aeruginosa* phague F116	100	97
SD9	648	251	24	NO		*P. aeruginosa* phague F116	100	99
SD9	649	50	24	NO		*P. aeruginosa* phague F116	100	100
SD9	650	39	24	NO		*P. aeruginosa* phague F116	100	99
SD9	651	123	24	NO		*P. aeruginosa* phague F116	100	99
SD9	652	73	24	NO		*P. aeruginosa* phague F116	100	100
SD9	653	48	24	NO		*P. aeruginosa* phague F116	100	99
SD9	654	167	24	NO		*P. aeruginosa* phague F116	100	99
SD9	655	123	24	NO		*P. aeruginosa* phague F116	100	99
SD9	656	70	24	NO		*P. aeruginosa* phague F116	100	93
SD9	657	52	24	NO		unknown		
SD9	658	68	24	NO		*P. aeruginosa* strain PACS171b clone fa1390, complete sequence	100	79
SD9	659	97	24	NO		*P. aeruginosa* phague F116	99	99
SD9	660	267	24	NO		*P. aeruginosa* phague F116	100	99
SD9	661	66	24	NO		*P. aeruginosa* phague F116	100	100
SD9	662	338	24	NO		*P. aeruginosa* phague F116	100	99
SD9	663	285	24	NO		*P. aeruginosa* phague F116	100	99
SD9	664	183	24	NO		*Pseudomonas* phage H66	100	96
SD9	665	169	24	NO		*Pseudomonas* phage H66	100	97
SD9	666	164	24	NO		*P. aeruginosa* phague F116	90	96
SD9	667	120	24	NO		*Pseudomonas* phage H66	100	97
SD9	668	427	24	NO		*Pseudomonas* phage H66	100	98
SD9	669	772	24	NO		*Pseudomonas* phage H66	99	97

Isolate	Gene_id[a]	Length (aminoacids)	Contig	Present in other ST-1146 isolates	Exclusivity reason	Description Blast Result	Query coverage (%)	Similarity (%)
SD9	670	376	24	NO		*P. aeruginosa* phague F116	100	99
SD9	671	430	24	NO		*P. aeruginosa* phague F116	100	99
SD9	672	159	24	NO		*P. aeruginosa* phague F116	100	100
SD9	673	290	24	NO		*P. aeruginosa* phague F116	100	99
SD9	674	225	24	NO		*P. aeruginosa* phague F116	100	96
SD9	675	68	24	NO		*P. aeruginosa* phague F116	100	99
SD9	676	135	24	NO		*P. aeruginosa* phague F116	100	97
SD9	677	131	24	NO		unknown		
SD9	678	111	24	NO		*Pseudomonas* phage H66/ P. aeruginosa phague F116	91	88
SD9	679	285	24	NO		*Pseudomonas* phage H66/ P. aeruginosa phague F116	100	99
SD9	680	987	24	NO		*Pseudomonas* phage H66	100	98
SD9	681	396	24	NO		*P. aeruginosa* phague F116	100	95
SD9	682	196	24	NO		*Pseudomonas* phage H66	100	96
SD9	683	865	24	NO		*P. aeruginosa* phague F116	100	87
SD9	684	109	24	NO		unknown		
SD9	685	283	24	NO		*P. aeruginosa* phague F116	99	78
SD9	686	630	24	NO		*Pseudomonas* phage H66	100	94
SD9	687	724	24	NO		*P. aeruginosa* phague F116	72	93
SD9	688	3782	24	NO		*Pseudomonas* phage H66	99	97
SD9	689	79	24	NO		*Pseudomonas* phage H66/ *P. aeruginosa* phague F116	100	99
SD9	690	177	24	NO		*P. aeruginosa* phague F116	100	97
SD9	691	172	24	NO		*P. aeruginosa* phague F116	100	97
SD9	692	197	24	NO		*Pseudomonas* phage H66/ *P. aeruginosa* phague F116	100	98
SD9	693	141	24	NO		*Pseudomonas* phage H66	98	93
SD9	694	170	24	NO		*Pseudomonas* phage H66	99	87
SD9	2506	71	56	NO		hypothetical protein of bacteriophage Pf1	100	98

Isolate	Gene_id[a]	Length (aminoacids)	Contig	Present in other ST-1146 isolates	Exclusivity reason	Description Blast Result	Query coverage (%)	Similarity (%)
SD9	2507	88	56	NO		hypothetical protein; Pseudomonas phage Pf1	100	96
SD9	4313	205	93	YES	Deletion 1nt	putative chemotaxis transducer	72	93
SD9	5640	480	127	YES	Deletion 1nt	hemagglutination activity domain protein	100	99

a) Gene identification establish by the MetaGene Mark program.

Supplementary Table 2.S5: Genes and alleles of ST-1146 compared with *P. aeruginosa* PAO1 classified by PseudoCAP Functional Categories.

	Genes				Alleles										
					P37		P47		P49		SD9		Ratio		
PseudoCAP Functional Class	**m**	**n**	**o**	**% o/n**	**p**	**% p/n**	**q**	**% q/n**	**r**	**% r/n**	**s**	**% r/n**	**s/p**	**s/q**	**s/r**
Adaptation, Protection	3	112	27	24.11	5	4.46	7	6.25	6	5.36	14	12.50	2.80	2.00	2.33
Amino acid biosynthesis and metabolism	6	197	47	23.86	7	3.55	3	1.52	11	5.58	25	12.69	3.57	8.33	2.27
Antibiotic resistance and susceptibility	0	14	3	21.43	0	0.00	0	0.00	1	7.14	2	14.29			2.00
Biosynthesis of cofactors	10	127	23	18.11	3	2.36	4	3.15	5	3.94	9	7.09	3.00	2.25	1.80
Carbon compound catabolism	5	131	30	22.90	3	2.29	3	2.29	7	5.34	15	11.45	5.00	5.00	2.14
Cell division	3	20	1	5.00	0	0.00	0	0.00	0	0.00	1	5.00			
Cell wall / LPS / capsule	7	133	26	19.55	6	4.51	4	3.01	6	4.51	16	12.03	2.67	4.00	2.67
Central intermediary metabolism	2	83	13	15.66	3	3.61	4	4.82	2	2.41	6	7.23	2.00	1.50	3.00
Chaperones & heat shock proteins	2	14	1	7.14	1	7.14	0	0.00	0	0.00	0	0.00	0.00		
Chemotaxis	1	22	6	27.27	1	4.55	2	9.09	0	0.00	3	13.64	3.00	1.50	
DNA replication, recombination, modification and repair	4	76	20	26.32	4	5.26	4	5.26	2	2.63	9	11.84	2.25	2.25	4.50
Energy metabolism	12	159	27	16.98	4	2.52	3	1.89	1	0.63	15	9.43	3.75	5.00	15.00
Fatty acid and phospholipid metabolism	5	49	11	22.45	0	0.00	0	0.00	3	6.12	6	12.24			2.00
Hypothetical, unclassified, unknown	209	1761	331	18.80	66	3.75	63	3.58	67	3.80	168	9.54	2.55	2.67	2.51
Membrane proteins	26	613	140	22.84	26	4.24	22	3.59	24	3.92	65	10.60	2.50	2.95	2.71
Motility & Attachment	4	62	19	30.65	6	9.68	6	9.68	5	8.06	9	14.52	1.50	1.50	1.80
Nucleotide biosynthesis and metabolism	4	68	11	16.18	3	4.41	1	1.47	1	1.47	8	11.76	2.67	8.00	8.00
Protein secretion/export apparatus	8	95	29	30.53	8	8.42	8	8.42	9	9.47	15	15.79	1.88	1.88	1.67
Putative enzymes	27	398	84	21.11	16	4.02	15	3.77	19	4.77	46	11.56	2.88	3.07	2.42
Related to phage, transposon, or plasmid	28	27	12	44.44	6	22.22	3	11.11	5	18.52	9	33.33	1.50	3.00	1.80
Secreted factors	5	66	31	46.97	7	10.61	13	19.70	11	16.67	19	28.79	2.71	1.46	1.73
Transcription, RNA processing and degradation	1	49	9	18.37	1	2.04	0	0.00	1	2.04	6	12.24	6.00		6.00
Transcriptional regulators	18	419	61	14.56	12	2.86	12	2.86	9	2.15	34	8.11	2.83	2.83	3.78
Translation, post-translational modification, degradation	37	119	15	12.61	1	0.84	3	2.52	2	1.68	6	5.04	6.00	2.00	3.00

	Genes				Alleles										
					P37		P47		P49		SD9		Ratio		
PseudoCAP Functional Class	**m**	**n**	**o**	**% o/n**	**p**	**% p/n**	**q**	**% q/n**	**r**	**% r/n**	**s**	**% r/n**	**s/p**	**s/q**	**s/r**
Transport of small molecules	11	250	62	24.80	15	6.00	15	6.00	6	2.40	34	13.60	2.27	2.27	5.67
Two-component regulatory systems	1	61	20	32.79	1	1.64	1	1.64	8	13.11	8	13.11	8.00	8.00	1.00
No classified	0	1	0	0.00											
Total	**439**	**5126**	**1059**	**20.66**	**205**	**4.00**	**196**	**3.82**	**211**	**4.12**	**548**	**10.69**	**2.67**	**2.80**	**2.60**

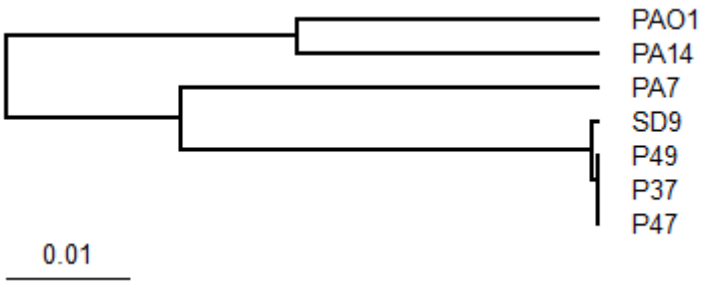
m, number of genes identical to PAO1**; n**, number of mutated genes compared with PAO1; **o** the number of mutated genes respect to PAO1, number of genes with some different mutation between the studied strains; **p, q, r y s,** unique alleles.

Supplementary Table 2.S6: Genes and alleles of ST-1146 compared with *P. aeruginosa* PAO1 classified by PseudoCAP Functional Categories.

	Genes				Alleles										
					P37		P47		P49		SD9		Ratio		
PseudoCAP Functional Class	**m**	**n**	**o**	**% o/n**	**p**	**% p/n**	**q**	**% q/n**	**r**	**% r/n**	**s**	**% r/n**	**s/p**	**s/q**	**s/r**
Adaptation, Protection	12	64	11	17.19	3	4.69	6	9.4	6	9.38	8	12.50	2.67	1.33	1.33
Amino acid biosynthesis and metabolism	18	203	34	16.75	8	3.94	10	4.9	6	2.96	17	8.37	2.13	1.70	2.83
Antibiotic resistance and susceptibility	7	36	11	30.56	2	5.56	2	5.6	3	8.33	5	13.89	2.50	2.50	1.67
Biosynthesis of cofactors	25	149	20	13.42	2	1.34	2	1.3	2	1.34	15	10.07	7.50	7.50	7.50
Carbon compound catabolism	9	135	17	12.59	4	2.96	3	2.2	0	0.00	11	8.15	2.75	3.67	
Cell division	3	33	5	15.15	2	6.06	1	3.0	1	3.03	3	9.09	1.50	3.00	3.00
Cell wall / LPS / capsule	21	91	14	15.38	3	3.30	4	4.4	1	1.10	9	9.89	3.00	2.25	9.00
Central intermediary metabolism	7	112	9	8.04	0	0.00	2	1.8	0	0.00	7	6.25		3.50	
Chaperones & heat shock proteins	7	27	4	14.81	0	0.00	0	0.0	0	0.00	4	14.81			
Chemotaxis	5	39	6	15.38	1	2.56	3	7.7	0	0.00	3	7.69	3.00	1.00	
DNA replication, recombination, modification and repair	20	105	16	15.24	0	0.00	4	3.8	2	1.90	8	7.62		2.00	4.00
Energy metabolism	36	231	26	11.26	6	2.60	2	0.9	4	1.73	16	6.93	2.67	8.00	4.00
Fatty acid and phospholipid metabolism	24	178	28	15.73	2	1.12	3	1.7	0	0.00	19	10.67	9.50	6.33	
Hypothetical, unclassified, unknown	487	1197	187	15.62	38	3.17	34	2.8	34	2.84	116	9.69	3.05	3.41	3.41
Membrane proteins	16	240	34	14.17	4	1.67	7	2.9	4	1.67	21	8.75	5.25	3.00	5.25
Motility & Attachment	27	70	22	31.43	6	8.57	5	7.1	7	10.00	10	14.29	1.67	2.00	1.43
Nucleotide biosynthesis and metabolism	12	73	12	16.44	2	2.74	1	1.4	0	0.00	8	10.96	4.00	8.00	
Protein secretion/export apparatus	19	107	24	22.43	6	5.61	10	9.3	5	4.67	15	14.02	2.50	1.50	3.00
Putative enzymes	52	339	51	15.04	6	1.77	9	2.7	12	3.54	27	7.96	4.50	3.00	2.25
Related to phage, transposon, or plasmid	54	27	18	66.67	3	11.11	4	14.8	3	11.11	14	51.85	4.67	3.50	4.67
Secreted factors	9	54	18	33.33	4	7.41	5	9.3	5	9.26	10	18.52	2.50	2.00	2.00
Transcription, RNA processing and degradation	7	51	8	15.69	1	1.96	0	0.0	1	1.96	6	11.76	6.00		6.00
Transcriptional regulators	50	380	48	12.63	5	1.32	6	1.6	7	1.84	30	7.89	6.00	5.00	4.29
Translation, post-translational modification, degradation	62	154	22	14.29	5	3.25	3	1.9	6	3.90	10	6.49	2.00	3.33	1.67

	Genes				Alleles										
					P37		P47		P49		SD9		Ratio		
PseudoCAP Functional Class	**m**	**n**	**o**	**% o/n**	**p**	**% p/n**	**q**	**% q/n**	**r**	**% r/n**	**s**	**% r/n**	**s/p**	**s/q**	**s/r**
Transport of small molecules	48	604	89	14.74	12	1.99	10	1.7	18	2.98	63	10.43	5.25	6.30	3.50
Two-component regulatory systems	14	142	27	19.01	3	2.11	5	3.5	2	1.41	21	14.79	7.00	4.20	10.50
No classified	0	0	0	0.00											
Total	**1051**	**4841**	**761**	**15.72**	**128**	**2.64**	**141**	**2.9**	**129**	**2.66**	**476**	**9.83**	**3.72**	**3.38**	**3.69**

m, number of genes identical to PA14**; n**, number of mutated genes compared with PA14; **o** the number of mutated genes respect to PAO1, number of genes with some different mutation between the studied strains; **p, q, r y s,** unique alleles.



Supplementary Figure 2.S2: UPGMA dendrogram of the imipenem outer membrane porin (OprD) of several *P. aeruginosa* strains and the four ST-1146 isolates.

Supplementary Table 2.S7: Synonymous and non synonymous nucleotide substitutions in the genes related to phage Pf1 in isolate SD9 compared with the environmental isolates.

Gene	lenght (bp)	No. of Nucleotide substitutions	No. of Synonymous subtitutions	No. of Non Synonymous subtitutions
hypothetical protein	213	13	12	1
helix destabilizing protein of bacteriophage Pf1	435	6	5	1
hypothetical protein	93	1	1	0
hypothetical protein	252	1	1	0
coat protein B of bacteriophage Pf1	240	3	3	0
coat protein A of bacteriophage Pf1	207	12	10	2
bacteriophage protein	357	3	3	0
zona occludens toxin	1275	13	9	4
hypothetical protein	102	1	1	0
bacteriophage integrase	984	11	9	2
prevent-host-death family protein	252	0		
hypothetical protein	348	4	1	3

bp) baise pairs

Chapter 3

Assessment of the species diversity of *Pseudomonas* by culture-dependent and -independent methods in a water sample from the River Woluve

Abstract

The River Woluwe has been taken as a model habitat for the study of the diversity of species in the genus *Pseudomonas*. A water sample from a non-contaminated site at the source of the river was analyzed by culture-dependent and –independent methods. Identification of the *Pseudomonas* isolates was performed by sequencing and analysis of their *rpoD* sequence. Culture-independent methods consisted of a cloning and pyrosequencing of a specific *rpoD* amplicon obtained from total DNA extracted from the same sample and amplified by *Pseudomonas rpoD* primer set PsEG30F-PsEGP790R. From a total of 14,540 pyrosequences, 6,228 corresponded to *Pseudomonas rpoD* sequences when compared in the NCBI database. The selection criteria for the pyrosequences were: sequences longer than 400 bp, with a Q40 average value higher than 25 and a percentage of similarity with a *Pseudomonas* species higher than 85%. From the 6228 *Pseudomonas rpoD* sequences, 5,345 sequences followed the established criteria of selection. It was remarkable the number of known species detected in the sample by the three different methods: 26 species distributed in 13 phylogenetic groups or subgroups within the genus. Pyrosequencing was a powerful tool, sequences obtained represented the 24 species with the exception of *P. stutzeri* and *P. simiae*. The predominant phylogenetic group within the *Pseudomonas* genus was *Pseudomonas fluorescens* group in the cultures and in the culture-independent analysis. In all analysis a high number of putative novel species were found: 11 in the cultured strains and 286 phylospecies were detected by pyrosequencing indicating the enormous diversity not described yet.

3.1. Introduction

Species of the genus *Pseudomonas* are ubiquitous bacteria in nature, occupying numerous and diverse ecological niches. The taxonomy of the genus is complex, comprising at least 142 recognized species at the time this manuscript was written (as stated in Euzeby's list of prokaryotic names and in the DSMZ web pages: http://www.bacterio.cict.fr and www.dsmz.de). These species are metabolically versatile, extremely heterogeneous from a nutritional perspective, and are of crucial importance in the carbon and nitrogen cycles. Some of its members are well known for their beneficial role to plants, others are used for bioremediation and as biocontrol agents, while yet other members are important as plant or animal pathogens (Mulet *et al*., 2009).

Previous work in our laboratory with members of the species *P. stutzeri* (García-Valdés *et al.*, 2003; Cladera *et al.*, 2004, 2006a, b; Mulet *et al.*, 2008) had permitted us to generate the tools needed to extend our study to other *Pseudomonas* species, such as the appropriate selection of genes (Cladera *et al.*, 2004; Mulet *et al.*, 2008), improvements in PCR protocols and primers used (Mulet *et al.*, 2009), and the creation of a specific database, PseudoMLSA, which is now available, to compile all of these gene sequences for the characterization and taxonomical identification of *Pseudomonas* strains (http://www.uib.es/microbiologiaBD/Welcome.php) (Bennasar et al., 2010). The *rpoD* is the gene encoding the sigma 70 subunit of RNA polymerase. The *rpoD* gene exhibited a high number of polymorphic sites (70.39%) and therefore is a good choice for phylogenetic and taxonomical analysis of species in the genus *Pseudomonas* (Mulet *et al*., 2010).

Species of the genus *Pseudomonas* are generally easy to isolate in standard growth media and several selective media have been tested for the isolation and enumeration in environmental samples. It is however still an open question how well bacterial isolates represent the total environmental population of *Pseudomonas* spp (Li *et al*., 2013) and therefore culture-independent methods are being developed. A highly selective *rpoD* pair of primers, PsEG30F/PsEG790R (amplicon of 736 bp), was designed for *Pseudomonas* genus and previously used to generate clone libraries from environmental DNA obtained from sand samples of the intertidal shore (Mulet *et al*., 2009). These primers present the advantages over other primers designed for similar purposes on the 16S rDNA (Li *et al.*, 2013; Widmer *et al.*, 1998) because they are longer and much better in discriminating *Pseudomonas* species. The limitatios of 16S rRNA sequence analysis and the advantages of new approaches based on other housekeeping genes are discussed by Pontes and collaborators (2007). These molecular methods used in microbial ecology studies are often limited to the analysis of a relatively small

number of clones and thus only a small fraction of the microbial diversity has been unraveled by the previous studies (Lenchi *et al*., 2013). In the last few years, pyrosequencing, allowing high throughput sequencing of DNA in a cost-effective manner, has been successfully applied to determine bacterial diversity within complex environmental ecosystems, such as marine deep water, hydrothermal vent samples and soil samples (Poulsen *et al*. 2013, Roesch *et al*., 2007; Sogin *et al*., 2006).

The River Woluwe is a typical lowland river with only 30 m altitude difference, 1–3 m wide and 5–50 cm deep, located in Brussels (Belgium). The sources of the river are ground water fed and are situated in the protected areas of a forest. The central part of the river is highly fragmented by diverse park and pond systems (flow-through and overflows) and by vaulted stretches. *Pseudomonas aeruginosa* population has been studied in these waters (Pirnay *et al*., 2005), as well as the *Pseudoma*s species: the *Pseudomonas aeruginosa* diversity by cultured strains (Pirnay *et al*., 2005) and the diversity of the genus *Pseudomonas*,using DNA-based molecular methodologies with the primers *oprI* and *oprL* for the identification of the cultured strains (Matthijs *et al*., 2012).

To assess the diversity of *Pseudomonas* species in a non-polluted sample obtained at the Woluwe River we used 2 approaches. A culture-dependent method was based on the culture of appropriate dilutions of the water sample in *Pseudomonas* selective medium and the isolates were identified by sequence analysis of the *rpoD* gene. For a culture-independent analysis, total DNA from the same sample was purified, amplified with selective *rpoD* gene primers and the amplicon was cloned in *E. coli* or was directly sequenced by pyrosequencing. Both methods were compared and the enormous diversity of phylotypes was analyzed.

3.2. Material and methods

3.2.1. Water Sampling

The water sample was obtained at sampling site W5 from the Woluwe River (Pirnay 2005), in Brussels (Belgium). W5 is located in the forest nearby the source. Surface water samples were collected (skimming the surface) in September 2011 with sterile 1 L bottles and immediately filtered, first through a 0.5μm filter (Millipore, USA, Type SVLP) and after through a 0.2 μm filter (Millipore, USA, Type GTTP).

3.2.2. Total cell count by DAPI

For cell enumeration purposes a water sample (1 ml) was filtered onto white polycarbonate filters (Millipore, USA, Type GTTP; pore size, 0.2 μm; diameter, 47 mm). Total cell numbers were determined by epifluorescence counting after staining with 4′, 6-diamidino-2-phenylindole (DAPI) as previously described (Glöckner *et al*., 1996).

3.2.3. Culture, isolation and identification of *Pseudomonas spp*.

The presence of *Pseudomonas* was determined by plate counts on a selective medium, CFC medium. CFC medium consisted of Pseudomonas agar F (Difco) containing cephaloridine (50 mg l^{-1}), fucidin (10 mg l^{-1}) and cetrimide (10 mg l^{-1}) (CFC supplement, Sigma). Overnight incubation was performed at 30°C. Strains were identified by sequencing the *rpoD* gene. Template DNA for the PCR reaction was obtained with the Wizard® Genomic DNA Purification Kit (Promega, Madison, USA) from an overnight culture of a fresh colony inoculated in 4 ml LB. Vials of DNA were subsequently stored at 20°C.

PCR amplification reactions of *rpoD* gene (volumes, nucleotides and primer concentrations) were performed as previously described (Mulet *et al*., 2009) in an Eppendorf thermocycler. Amplicons were purified with Multiscreen PCR 96-well plates (Millipore) and were then sequenced directly using the ABI PRISM BigDye vs. 3.1 Terminator Cycle sequencing kit (Applied Biosystems). Strains were identified by analysis of the sequences at the NCBI database and in an in-house *rpoD* database. Those colonies grown on CFC medium whose DNA could not be amplified by the selective rpoD primers were identified by sequencing and analyzing of their 16S rDNA.

3.2.4. Total DNA extraction and PCR conditions

The total DNA from 5 liters of freshwater sample was extracted using the MasterPure™ Complete DNA and RNA Purification Kit (Epicentre® Biotechnologies, Madison, USA). DNA purification was accomplished using the Wizard DNA Clean-Up System (Promega, Madison, USA). Purified environmental DNA was used to amplify the *rpoD* gene with the PsEG30F/PsEG790R pair of primers. PCR amplification was performed with a DNA thermocycler (Eppendorf). The PCR was carried out in a final volume of 50 μl using the KAPA2G Robust HotStart PCR kit (Kapa Biosystems) with 5U *Taq* DNA polymerase. The individual reaction mixtures contained 10μl of containing PCR KAPA2G buffer B, 0.4 μl of the deoxynucleotide mixture at 100 mM (Fermentas), 0.5 μl of each primer at a concentration of 100mM, 10 μl of 5X KAPA Enhancer 1, 0.4 μl of *Taq* DNA polymerase and 1 μl of template DNA. The PCR

program used was an initial denaturation of 4 min at 95°C, followed by 35 cycles of denaturation at 95°C for 30 s, annealing at 52°C for 20 s and extension at 72°C for 45 s, followed by incubation for 10 min at 72°C. The obtained PCR fragments were purified using the Qiagen PCR purification kit.

3.2.5. Clone library

The *rpoD* purified PCR product was used to generate the corresponding clone library using the CloneJET PCR cloning kit (Fermentas). The PCR product was ligated into the pJET1.2 blunt cloning vector and transformed into *Escherichia coli* XL1 Blue competent cells (Invitrogen) culture in LB agar plates, supplemented with 30 □g ampicillin per ml and incubated at 37°C for approximately 16-24 h. Plasmid DNA of the transformed cells was obtained by boiling the colonies for 5 min in 100 ml TE buffer (pH 8), followed by centrifugation for 5 min at 16,100 xg after which the supernatant was collected. The insertions of each clone were amplified using the same primers used for amplification. Five microlitres of the amplified PCR products were analysed by electrophoresis on a 1.5 % (w/v) agarose gel and stained with ethidium bromide. Amplicons were purified and sequenced as previously described above (PsEG30F/PsEG790R).

3.2.6. Sequence analysis of amplified genes

The partial *rpoD* sequences obtained from the cultured isolates and clones were compared with sequences from an in-house *rpoD* database. The partial 16S rDNA sequences obtained from the cultured isolates were compared with the NCBI gene database. All *rpoD* sequences were aligned and the phylogenetic analysis was done as previously described (Mulet *et al*., 2010). Similarities higher than 96 % with a known *Pseudomonas* type strain allowed the assignation of the sequence to a known *Pseudomonas* species. Values lower than 96 % ndicated that the sequence corresponded to a possible new species in the *Pseudomonas* genus.

3.2.7. *rpoD* 454 Pyrosequencing

The *rpoD* amplicon of the total DNA purified from the water sample was also analyzed by pyrosequencing with standard 454/Roche GS-FLX Titanium protocols. Sequences with a Q40 average lower than 25 and a length shorter than 400 nucleotides, as well as those considered chimaeras, were discarded from further analysis.

3.2.8. Post-run analysis of pyosequences

The sequences obtained from the pyrosequencing protocol were compared with sequences from the NCBI gene database. Sequences were filtered according to their first hits. In a first step only the *rpoD* sequences were selected (first filtration). In a second step those with a length higher than 400bp, with a similarity percentage with the closest species higher than 85% and Q_{40} average value higher than 25 were selected (second filtration). Afterwards, *rpoD* sequences similar to *Pseudomonas* species were retrieved from the *rpoD* sequences. The *rpoD* sequences assigned to *Pseudomonas* spp. were grouped according to their closest *rpoD* gene of a *Pseudomonas* type strain. For each group a deeper analysis was performed comparing the sequences with the closest *rpoD* sequences from an in-house database. Sequences with similarity higher than 96 % in the same phylogenetic branch were grouped. At least one representative sequence of each group was selected for further analysis, following the criteria of maximum length, coverage and high Q40 average. When it was possible two forward and two reverse sequences, assigned to the same group, were selected as representative sequences. Similarities higher than 96% with a known *Pseudomonas* type strain allowed the assignation of the representative sequences (and their represented group) to a known *Pseudomonas* species. Values lower than 96% indicated that the representative sequence (and their represented group) corresponded to a possible new species in the *Pseudomonas* genus. The coverage of sequences in a distant branch from all *rpoD* sequences was checked to discard chimeras. Sequences with coverage lower than 60% were discard (third filtration).

The *rpoD* sequences of other genera were grouped according to their genera and compared again with the NCBI database. Sequences with coverage lower than 60% were discarded.

3.2.9. Statistical analysis

Each different sequence was considered a single phylotype, and several phylotypes were grouped in a single phylospecies when they affiliated in the same phylogenetic branch and the similarity in the *rpoD* gene was higher than 96% (Mulet *et al*., 2010). Sequence information obtained from the pirosequences and isolates detected was used to calculate coverage, diversity indices, and phylotype and phylospecies richness with the PAST software package (Hammer, Ø. Et al 2001). The rarefaction curves were also obtained with the PAST software package.

3.3. Results

3.3.1. Total cell number count

Epifluorescence microscopy after DAPI staining of the sample filtered indicated a population of $1.02x10^{9}$ cells/ml.

3.3.2. Culture-dependent analysis

Hundred and thirty-six *Pseudomonas* type strains were first tested for growth on CFC medium. Hundred and twenty four type strains were able to grow on this selective medium at their optimal growth temperature after 24-72h incubation with the exception of twelve of them as indicated in Supplementary Table 3.S1.

188,4 cfu/ml were counted on CFC medium which represents a $1.85x10^{-5}$ % of the total cell count.

Hundred and sixty-two randomly selected colonies were identified by sequencing the *rpoD* gene (126 colonies), or the 16S rDNA when the *Pseudomonas* selective primers did not amplify (36). From the 126 *rpoD* sequences, 114 corresponded to *Pseudomonas rpoD* sequences and 12 to species in the genus *Aeromonas*. The remaining 36 colonies were identified by sequencing the 16S rDNA and were assigned to the following genera: *Aeromonas* (7), *Bacillus* (4), *Brevundimonas* (1), *Hafnia* (2), *Morganella* (3), *Staphylococcus* (1) and *Yersinia* (18). These results indicated that 30% of the colonies were not *Pseudomonas*.

Those *Pseudomonas* isolates with a similarity equal or higher than 96% with a type strain (the species threshold considered) were assigned to the corresponding species (7 *Pseudomonas* species, 24 isolates). The other *Pseudomonas* isolates were assigned to the phylogenetic group or subgroup which included the closest type strain to the new isolate (90 strains in 10 groups or subgroups). The *rpoD* similarity with the closest type strain was below the 96% species threshold and can be considered as representative of putative novel species (Table 3.1) (Supplementary Table 3.S4).

Table 3.1: *rpoD* sequences of the isolates

Closest *Pseudomonas* type strain	% similarity	No. of isolates	Species or group assignation	Probable new species
P. protegens	94.56-95.90	51	*P. chlororaphis* SG	I
P. brenneri	94.48-94.62	14	*P. gessardi* SG	II
P. saponiphila	93.93-93.76	11	*P. chlororaphis* SG	III
P. rhodesiae	98.03-98.45	8	*P. rhodesiae*	-
P. proteolytica	99.22-99.23	8	*P. proteolytica*	-
P. frederikbergensis	90.53-91.47	6	*P. mandelli* SG	IV
P. lurida	99.39-99.85	4	*P. lurida*	-
P. japonica	80.75	2	P. putida G	V
P.protegens	88.59-88.77	2	P. chlororaphis SG	VI
P. stutzeri (gv1)	99.24	1	*P. stutzeri*	-
P. rhodesiae	91.77	1	P. fluorescens SG	VII
P. simiae	99.84	1	*P. simiae*	-
P. lurida	93.95	1	P. fluorescens SG	VIII
P. grimonti/marginalis	95.59	1	P. fluorescens SG	IX
P. grimonti	99.85	1	*P. grimonti*	-
P. moraviensis	97.34	1	*P. moraviensis*	-
P. moraviensis	94.22	1	P. koreensis SG	X

3.3.3. Clone library analysis

The cloning of the PCR product resulted in a clone library consisting of 115 clones. Sixty-four clones showing the expected insert size were sequenced, while 51 clones did not generate an amplicon of the expected size or completely lacked an amplified insert. From the 64 clones analysed, 26 showed high similarity with *Pseudomonas* species (Table 3.2). Most of the clones belonged to the so-called *Pseudomonas syringae* group: 18 clones close to *Pseudomonas cichorii* (88.61-89.02 % of similarity) and 2 clones identified as *Pseudomonas viridiflava* (97.13 % of similarity). The other 6 clones belonged to the *Pseudomonas fluorescens* group. Only 2 of them were identified at the species level: *Pseudomonas grimonti*, with a 99.29-99.64 % of similarity. The phylogenetic distance of the remaining clones was high enough to represent a different *Pseudomonas* species: 1 close to *Pseudomonas frederikbergensis* (92.21 % of similarity), 1 close to *P. grimonti* (87.15 %), 1 close to *P. grimonti* and *Pseudomonas marginalis* (93.38 %) and 1 close to *Pseudomonas lurida* (95.46 %)(Supplementary Table 3.S4). The remaining 38 sequences were analyzed at the NCBI database and did not correspond to *Pseudomonas* species or were discarded (data not shown).

Table 3.2: Sequences assignation of the *rpoD* clone library

Closest *Pseudomonas* type species	% similarity	No. of clones	Specie or group assignation	Probable new especie
P. cichorii	88,61-89,02	18	*P. syringae* G	I
P. grimonti	99,29-99,64	2	*P. grimonti*	-
P. viridiflava	97,13	2	*P. viridiflava*	-
P. grimonti/marginalis	93,38	1	*P. fluorescens* SG	II
P. grimonti	87,15	1	*P. fluorescens* SG	III
P. frederikbergensis	92,21	1	*P. mandelli* SG	IV
P. lurida	95,46	1	*P. fluorescens* SG	V

3.3.4. Pyrosequencing of the environmental *rpoD* amplicon

A total of 14,540 sequences were obtained by pyrosequencing the *rpoD* amplicon of DNA extracted from the water sample (Supplementary table 3.S2). Eighty-seven per cent of the sequences were longer than 400 bp: 400-599 bp for 6,935 sequences (47.7%) and 600-799 bp for 5,750 sequences (39.55%) as indicated in Table 3.3.

Table 3.3: Sequences obtained by pyrosequencing of the *rpoD* amplicon (baise pairs)

All sequences			***Pseudomonas rpoD* sequences**		
Length (bp)	**No.**	**%**	**Length (bp)**	**No.**	**%**
<200	619	4,26	<200	169	2,71
200-399	1230	8,46	200-399	457	7,33
400-599	6935	47,70	400-599	2463	39,51
600-799	5750	39,55	600-799	3138	50,34
>800	6	0,04	>800	1	0,02
	14540			6228	

bp) baise pairs.

When all sequences were compared in the NCBI datase, 7,031 sequences (48.35%) were an *rpoD* gene as first hit and 6,228 corresponded to *Pseudomonas rpoD* sequences (88.58%). Phylogenetic assignation to genera of the 7,031 sequences and their abundance are shown in Table 3.4 and Supplementary Figure 3.S3.

Table 3.4: Phylogenetic assignation to genera of the *rpoD* sequences and number of sequences after each filtration step.

Phylogenetic assignation			Filtration step		
Genus	**Family**	**Class**	**1st**	**2on**	**3rd**
Pseudomonas	*Pseudomonadaceae*	*Gammaproteobacteria*	6228	5345	5165
Acidovorax	*Comamonadaceae*	*Betaproteobacteria*	351	176	170
Azoarcus	*Rhodocyclaceae*	*Betaproteobacteria*	143	102	0
Uncultured			123	82	8
Aeromonas	*Aeromonadaceae*	*Gammaproteobacteria*	83	62	56
Xanthomonas	*Xanthomonadaceae*	*Gammaproteobacteria*	24	12	0
Stenotrophomonas	*Xanthomonadaceae*	*Gammaproteobacteria*	15	11	0
Alcanivorax	*Alcanivoracaceae*	*Gammaproteobacteria*	13	11	0
Bordetella	*Alcaligenaceae*	*Betaproteobacteria*	9	5	0
Pseudoxanthomonas	*Xanthomonadaceae*	*Gammaproteobacteria*	8	5	1
Halomonas	*Halomonadaceae*	*Gammaproteobacteria*	6	3	0
Burkholderia	*Burkholderiaceae*	*Betaproteobacteria*	5	2	0
Chromohalobacter	*Halomonadaceae*	*Gammaproteobacteria*	4	1	0
Curvibacter	*Comamonadaceae*	*Betaproteobacteria*	3	0	0
Alteromonas	*Alteromonadaceae*	*Gammaproteobacteria*	2	0	0
Cupriavidus	*Burkholderiaceae*	*Betaproteobacteria*	2	1	0
Dechloromonas	*Rhodocyclaceae*	*Betaproteobacteria*	2	2	0
Rhodococcus	*Nocardiaceae*	*Actinobacteria*	2	2	0
Sphingomonas	*Sphingomonadaceae*	*Alphaproteobacteria*	2	2	0
Variovorax	*Comamonadaceae*	*Betaproteobacteria*	2	0	0
Achromobacter	*Alcaligenaceae*	*Betaproteobacteria*	1	0	0
Diaphorobacter	*Comamonadaceae*	*Betaproteobacteria*	1	1	0
Psychrobacter	*Moraxellaceae*	*Gammaproteobacteria*	1	0	0
Rhodobacter	*Rhodobacteraceae*	*Alphaproteobacteria*	1	1	0

From the 6228 *Pseudomonas rpoD* sequences, 5,345 sequences were longer than 400 bp, with a Q40 average value higher than 25 and a percentage of similarity with a *Pseudomonas* species higher than 85% (minimal distance between two type strains in the same *Pseudomonas* group or subgroup) and they were further analyzed. 2615 sequences corresponded to the forward primer and 2730 to the reverse primer. From the 5345 selected sequences, only 5165 were analyzed after the third filtration. From these sequences, only 559 (10.82%) were assigned to a 24 known *Pseudomonas* species. The remaining 4,606 were assigned to 280 phylospecies in a group or subgroup in the genus *Pseudomonas* (Supplementary Table 3.S4). Singletons were detected in high proportion in these phylospecies (210 sequences, 75%).

Ten known *Pseudomonas* species and 27 putative novel species were detected by forward sequences (528) and reverse sequences (4166) (were labelled “FR phylospecies”),

overlapping in at least 456 nucleotides. Bidirectional sequencing is normally recommended for all amplicon sequencing analysis because the combination of forward and reverse reads provides higher consensus sequence. In screening datasets for variants, one key criterion for determining whether a variant determination is valid is whether it is supported by both forward and reverse reads. Therefore, 4,694 forward and reverse sequences were assigned with high confidence to 37 species (putative or known) distributed in 11 of the 19 phylogenetic groups or subgroups defined in the genus *Pseudomonas* (data not shown).

3.3.5. Diversity indices

The number of clones in the library was low and for statistical analysis, only the cultured and the pyrosequencing results were considered (Table 3.5). The coverage index for the cultured isolates was estimated 0.53 and reached 0.97 in the analysis of the all phylospecies and the value was 1 when the conserved criterion of high confidence in species assignation was applied (FR phylospecies). The Chao1 index was used as an estimate of the total number of phylotypes or species present.

Table 3.5: Statistical indices.

	Cultured	**Pyrosequencing**	
		All phylospecies	**FR phylospecies[a]**
Total Sequences	114	5165	4694
Phylospecies	17	310	37
Singletons	8	217	0
Simpson index	0,239	0,530	0,641
Simpson diversity index	0,760	0,470	0,359
Coverage	0,529	0,977	1
Chao-1	26,33	1066	37
Shannon (H)	1,963	1,68	1,002

a) FR phylospecies, phylospecies with a forward and reverse representative sequences.

Rarefaction curves for the cultured strains and the *rpoD* sequences of the amplicon were constructed to determine the extent of the diversity of phylotypes and phylospecies detected (Figure 3.1). The total number of phylotypes detected was 5,165, distributed in 310 phylospecies (All phylospecies) and the rarefaction curve did not reach the saturation. However, the 37 phylospecies detected with the conservative criterion (FR phylospecies), represented by 4,696 sequences, reached a saturation curve, indicating that the total diversity in the sample was detected.

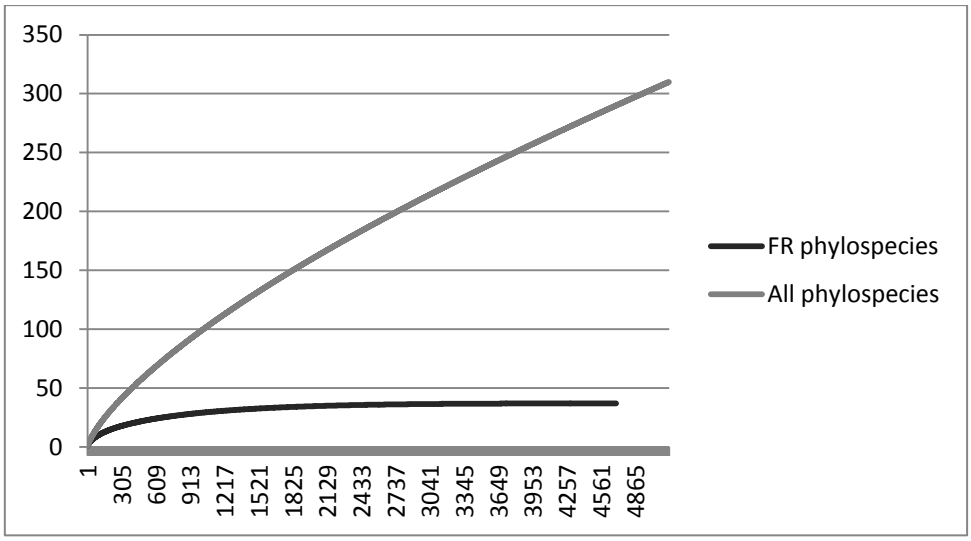


Figure 3.1: Rarefaction curves of pyrosequencing results. FR phylospecies were considerered that phylospecies with a forward and reverse representative sequences.

3.4. Discussion

Several selective media have been used for the isolation of *Pseudomonas* by other authors: Gould's S, King's B (Johnsen *et al*., 1999) or nutrient-poor NAA (Li *et al*., 2013). However, CFC media is considered as a high selective and is the mostly used selective medium for *Pseudomonas* (Matthijs *et al*., 2012) and has been also used previously in the *Pseudomonas* analysis of the River Woluwe. We tested 136 *Pseudomonas* species type strains for growth on CFC plates and only 12 type strains were not able to grow (9%). In the sample analyzed we could also detect colonies of other Gram negatives, as has been stated by other authors (Bridson, 2006; Matthijs *et al*., 2012), and also Gram positives in a lesser extent. The most frequent non-*Pseudomonas* species detected were *Aeromonas* spp. In a previous study (Matthijs *et al*., 2012) only 17% of the cultured *Pseudomonas* could be assigned to a known species using for identification the primers *oprI-oprL.* Similarly, in our study, 22% of the isolates could be identified at the species level.

The high selectivity of the *Pseudomonas rpoD* primers previously published (Mulet *et al*., 2009) was confirmed in the present study. Only few non-*Pseudomonas* cultured strain of the genus *Aeromonas* were amplified with this primers set. The analysis of the sequences of the *rpoD* amplicon also confirmed the selectivity of the primers, because only 4,35% of the *rpoD* sequences obtained were not assigned to the *Pseudomonas* genus after the third filtration.

In addition to sequencing errors and chimeras, PCR based methods introduce biases that can affect the results of microbial community structure analyses. Such PCR biases can be

particularly significant since the current procedure for amplicon-sequencing involves at least two different amplification steps, i.e., PCR amplification during initial sample preparation using template specific primers followed by emulsion PCR (emPCR) on Roche's 454 or bridge PCR on Illumina platforms, prior to sequencing (Pinto *et al*., 2012). Thus only the presence of species can be determined but not their abundance.

A comparison of the 3 methods indicated that all the sequences of the *Pseudomonas* isolates and those obtained in the clone library were assigned to species or phylogenetic groups or subgroups also detected in the pyrosequencing method, with the only exception of *P. stutzeri*, detected only in the cultured strains. It is clear the predominance of the members of the P. *fluorescens* G, being considerated as representantive of unpolluted waters and the absence of *Pseudomonas aeruginosa*. The W5 sampling point is located near of the source upstream of the sewage water entrances. These data is concordant with the data obtained by Mattijs and collaborators (2012) for the sampling point W2 also located near of the source.

The data obtained from the pyrosequences indicated the presence 18 species of *Pseudomonas* not found in culture, neither in the cloning sequences of this study. These sequences were assigned to: *P. abietaniphila*, *P. alcaligenes, P. aspleni, P. cichorii, P. flavescens, P. frederikbergensis, P jessenii, P. lini, , P. mandelii, , P. mohnii, P. peli, P. putida*, *P.salomonii, P. syringae*, *P. trivialis, P. veronii* and the newly described species *P. asturiensis* and *P. punonensis.* All the type strains of these species grew on CFC medium, with the exception of *P. asturiensis* and *P. punonensis* described after the analysis of the type strains in the CFC medium and no data are available. These data emphazises the high sensitivity of the pyrosequencing method.

It is remarkable the number of known species detected in the sample by the three different methods: 26 species distributed in 13 phylogenetic groups or subgroups within the genus. Only members of 4 groups were not detected: *P. luteola* G, *P. oryzihabitans* G, *P. pertucinogena* G and *P. oleovorans* G. The results obtained by pyrosequencing 24 species with the exception of *P. stutzeri* and *P. simiae*. *P. grimonti* was the only strain detected in all the sequences obtained in the three methods used. The pyrosequencing of the *rpoD* amplicon rendered a huge number of phylotypes. Only 11% could be assigned to a known species (24 species) with the accepted 96% similarity as species threshold. Additionally, 286 possible phylospecies not assigned to a known species have been detected, which could indicate that the real number of the existing *Pseudomonas* species could be at least 3 times the number of *Pseudomonas* species actually known. These results demonstrate the high potential of pyrosequencing to detect the enormous diversity of *Pseudomonas* which have not described yet, and simultaneously the need to stress the efforts to isolate the putative novel species for their physiological and biochemical characterization.

3.5. Supplementary data

Supplementary Table 4.S1: Growth Test of *Pseudomonas* type strain in CFC medium.

In extra CD

Supplementary Table 4.S2: First hit of the pyrosequences analyzed compared with the National Center Biotechnology for Information (NCBI) nucleotide database.

In extra CD

Supplementary Table 4.S3: *rpoD* sequences of *Pseudomonas* analyzed.

In extra CD

Supplementary Table 4.S4: Summarized results of *rpoD* sequences analyzed in Chapter 3.

	Isolates						rpoD sequences of the gene library						Sequences of the rpoD amplicon pyrosequenced					
	Assigned to species		Assigned to G or SG		Total		Assigned to species		Assigned to G or SG		Total		Assigned to species		Assigned to G or SG		Total	
	No. of sequences	No. of species	No. of sequences	No. of putative novel species	No. of sequences	No. of species	No. of sequences	No. of species	No. of sequences	No. of putative novel species	No. of sequences	No. of species	No. of sequences	No. of species	No. of sequences	No. of putative novel species	No. of sequences	No. of species
P. fluorescens SG	14	4	3	3	17	7	2	1	3	3	5	4	386	6	253	108	639	114
P. gessardi SG	8	1	14	1	22	2							1	1	1	1	2	2
P. fragi SG															15	3	15	3
P. jessenii SG													3	2	13	9	16	11
P. koreensis SG	1	1	1	1	2	2							1	1	92	14	93	15
P. mandelii SG			6	1	6	1			1	1	1	1	4	3	13	12	17	15
P. corrugata SG															7	4	7	4
P. chlororaphis SG			64	3	64	3									5	2	5	2
P. asplenii SG													2	1			2	1
P. fluorescens G															1	1	1	1
P. syringae G							2	1	18	1	20	2	126	4	3938	53	4064	57
P. lutea G													2	1	146	28	148	29
P. rhizospherae															12	7	12	7
P. putida G			2	1	2	1							6	1	60	11	66	12
P. anguilliseptica G													6	1	13	9	19	10
P. straminea G													15	2	14	12	29	14
P. aeruginosa G													7	1	23	12	30	13
P. stutzeri G			1	1	1	1												
Total	**23**	**6**	**91**	**11**	**114**	**17**	**4**	**2**	**22**	**5**	**26**	**7**	**559**	**24**	**4606**	**286**	**5165**	**310**

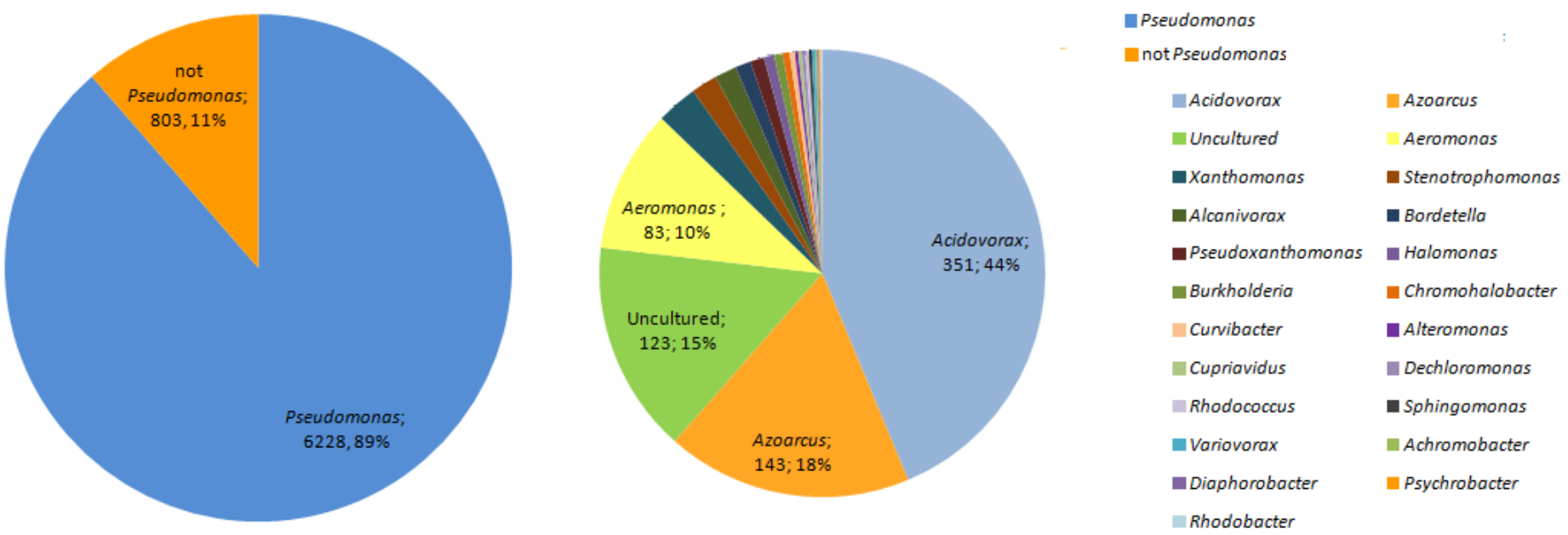


Figure 3.S1: Graphic representation of the assignation to genera of the *rpoD* sequences obtained by pyrosequencing of the amplicon of the environmental DNA after the first filtration (all sequences with an *rpoD* genes as fist hit when were compared with de National Center for Biotechnology Information).

Chapter 4

Isolation and characterization of *Pseudomonas* from soils and intertidal seashore habitats. Description of new species of *Pseudomonas*

Abstract

In the present chapter, several *Pseudomonas* strains have been isolated from environmental samples, from soil and intertidal habitats. In the identification process, some of these strains have not been assigned to known *Pseudomonas* species and were considered members of putative novel species. In their phylogenetic characterization by MLSA we found that strains in the culture collection of our laboratory were close-related and therefore they were also included in the taxonomic characterization of these putative novel species. MLSA demonstrated that 3 putative novel species were represented by 6, 5 and 1 strains respectively, which will be the subject of additional studies. Four other strains were deeply studied by a taxonomic polyphasic approach, including morphological, physiological, biochemical, chemotaxonomic and genomic characterizations. These studies demonstrated that the four strains cannot be assigned to any of the known *Pseudomonas* species and we propose the creation of two novel species, *Pseudomonas aestusnigri* (strain VGXO14T = CECT 8317 T = CCUG 64165 T) and *Pseudomonas terricola* (strain S58 T = CECT 8389 T = CCUG 64415 T) to allocate phylogeneticaly these four isolates in the *P. pertucinogena* group and in the *P. corrugata* subgroup within the genus *Pseudomonas*.

4.1. Introduction

The genus *Pseudomonas* was described by Migula in 1894. The genus *Pseudomonas* is one of the more diverse genera, and its taxonomy has undergone many changes since earlier descriptions. The genus *Pseudomonas* is ubiquitous, metabolically versatile, and important for the recycling of elements in the environment. *Pseudomonas* is the genus of Gram negative bacteria with the highest number of species, and new species are described continuously. In fact, 142 species names have been validly accepted within the genus *Pseudomonas* (http://www.bacterio.cict.fr). The genus is defined as a Gram negative, straight or slightly curved rod bacterium that is usually motile by polar flagella. The species are aerobic with a respiratory metabolism in which oxygen is the electron acceptor but can sometimes be substituted by an alternative terminal electron acceptor, such as nitrate. The oxidase test is positive or negative, and the catalase test is positive. The bacteria are chemoorganotrophs, and some of the species accumulate poly-beta-hydroxyalkanoates (PHA). The G+C content ranges from 58-69 mol% (Palleroni, 1984). *Pseudomonas* species appear across a wide distribution of environmental habitats and present a high level of phenotypic versatility (Moore *et al*., 1996).

The detection of *Pseudomonas* spp. and the analysis of the diversity of their consortia in terrestrial and aquatic environments have been typically based on cultivation approaches (Aagot *et al.* 2001; Sazakli *et al.* 2005). Several cultivation protocols, based on different nutritional sources and conditions, have been demonstrated to provide different results in terms of *Pseudomonas* diversity in soils (Sørheim *et al.* 1989; Aagot *et al.* 2001).

The characterisation of *Pseudomonas* type strains by multilocus sequence analysis (MLSA), concatenating the genes 16S rRNA, *gyrB*, *rpoD* and *rpoB,* permitted the establishment of two main lineages, *P. fluorescens* and *P. aeruginosa*, which were divided into several groups (G) and subgroups (SG). Some type strains did not belong to any of these lineages, being located in an independent phylogenetic branch, including those representing the species *P. oryzihabitans*, *P. psychrotolerans*, *P. luteola*, *P. pertucinogena* and *P. pachastrellae* (Mulet et al., 2010). In 2012, five type strains analysed by MLSA were phylogenetically close to *P. pertucinogena* and are now considered members of a new group: the *P. pertucinogena* group (Mulet et al., 2012).

In the present study, several strains have been isolated from agricultural soil which could not be assigned to known species in a preliminary identification. A deeper phylogenetic study by MLSA demonstrated that four strains could represent three putative novel species. Other culture collection strains available in our laboratory were close-related phylogenetically in the MLSA study and were also included in the taxonomic analysis.

In this chapter, we present the description of a new species that is related to the *P. corrugata* SG according to a polyphasic taxonomic approach, including phylogenetic, chemotaxonomic, phenotypic and genomic data of two strains: S58^T isolated from agricultural soil in Mallorca (Spain) and CFBP 4629 from the roots of *Helianthus annuus* in Charente, France in 1998., Other possible new species represented by strains S12 and S25, from the same agricultural soil, will be the subject of additional studies. In this Thesis were included some strains from culture collections related to some possible new species of Pseudomonas.

In a previous study (Mulet *et al*., 2011) on the diversity of autochthonous *Pseudomonas* populations in two contaminated beaches after the Prestige oil spill (Carnota municipality, Galicia, northwestern Spain) in November 2002, rendered 102 *Pseudomonas* isolates from intertidal sand contaminated with crude oil. From this analysis, several isolates were considered possible new species. These results were confirmed by the analysis of partial sequences of sigma 70 subunit of RNA polymerase, encoded by the *rpoD* gene. A deep taxonomic study of the characteristics of two of these strains is presented. A new *Pseudomonas* species in the *P. pertucinogena* group, with the strain VGXO14^T as the type strain, is proposed. Strain VGXO14^T was deposited in two culture collections under accession numbers CCUG 64165^T and CECT 8317^T.

4.2. Materials and Methods

4.2.1. Isolates

Agricultural soil isolates

Two soil samples were collected in September of 2009 (M0909) and September 2010 (M0910) from an agricultural soil sample taken at the University of the Balearic Islands Campus in Mallorca, Spain (39^o 38’ 11.8’’ N, 2^o 38’ 50.1’’ E). From a suspension of 2g of soil in 8 ml of Ringer, 100µl were plated on R2A agar (Scharlab), Gould S1 agar (Gould *et al*., 1985), Cetrimide agar (Merck) supplemented with nalidixic acid, Middlebrok agar supplemented with Middlebrook Oleic Albumin Dextrose Catalase Growth Supplement (MiddlebrooK OADC enrichment, Difco) in aerobic conditions. Mineral medium (Miller, 1972) supplemented with NO_3K 1% and etilenglicol (M9+E) or starch (M9+A) as carbon source were incubated in anaerobic conditions. All plates were incubated at 30^{o}C with the exception of R2A plates (room temperature). Different isolates were selected from different media according to their colony morphologies characteristics (Supplementary table 4.S1).

Culture collection strains

Previous studies in our laboratory of some culture collection strains were also included in this study, in order to evaluate their possible relatedness with the new isolates (Supplementary table 4.S1).

The strain CFBP 4629 was obtained from the CFBP culture collection. The strain CFBP 4629 was isolated from the roots of *Helianthus annuus* in Charente, France (1998). This strain was previously identified phenotypically as a member of *Pseudomonas putida* biovar B (Mulet *et al*., 2013). The strain IFO 3738 was also previously described as a member of *Pseudomonas putida* biovar unknown (Mulet *et al*., 2013).

Several Culture collection strains from the Plant Pathology Research Institute (Iran) were included in this study. Name, province, city, GPS position, host and isolation date are shown in table 4.1.

Table 4.1: Isolates from the Plant Pathology Research Institut (Iran) included in this study.

Isolate	Province	City	GPS position	Host	Isolation date
24F	Mazandaran	Babol	N36 22.749 E52 39.648	Washington Navel	2010
58F	Mazandaran	Chalus	N36 41.144 E51 20.702	Sour Orange	2009
102F	Mazandaran	Babol	N36 22.749 E52 39.648	Washington Navel	2010
103F	Golestan	Bandar gaz	N36 45.203 E53 55.359	Sour Orange	2010
122F	Mazandaran	Nashtaroud	N36 44.157 E51 01.020	Local Orange	2009
18F	Mazandaran	Chalus	N36 38.185 E51 28.737	Local Orange	2009
19F	Mazandaran	Nashtaroud	N36 44.009 E51 00.827	Local Orange	2009
35F	Gilan	Amlash	N37 02.374 E50 18.407	Local Orange	2010
44F	Mazandaran	Chamestan	N36 27.297 E52 14.180	Sour Orange	2010

Seashore isolates

Strains VGXO14^T and Vi1 have been isolated from intertidal sand contaminated by Prestige's crude oil from two different sites at Praia da Seda beach (municipality of Lariño; Lat. 42°46'-29.27''N, Long. 9°7'27.08''W). VGXO14^T was isolated in September 2004, after growth in marine mineral medium (MMM) (Mulet *et al*., 2011) with 1 g/litre potassium nitrate (0.11 M) under anaerobic conditions (Gaspack; Oxoid) and with 5 g/litre of glycerol as a carbon source. Vi1 was isolated in March 2004, selected after a previous enrichment in MMM with hexadecane as the only carbon and energy source, and selected after growth in glycerol in the same conditions as described for VGXO14^T (Mulet *et al*., 2011)(Supplementary table 4.S1).

4.2.2. DNA extraction, PCR amplification and DNA sequencing conditions

For biomass recovery, the strains were cultured routinely at 30 ^oC on Luria-Bertani medium (Miller, 1972). The DNA extraction, PCR amplification, primers used and DNA sequencing conditions were previously described by Mulet and collaborators (2010). The

amplified product was purified using a MultiScreen PCR 96-Well Plate (Millipore). The sequence analysis procedures have been previously described by Mulet and collaborators (2010).

4.2.3. Individual and concatenated trees

A series of individual trees based on the 16S rDNA, gyrase beta subunit (*gyrB)*, and RNA polymerase 70 sigma factor (*rpoD)* partial gene alignments were generated. A concatenated analysis of these three genes was also performed as previously described (Mulet *et al*., 2010). Individual tree based on beta subunit of the RNA polymerase (*rpoB*) and a concatenated analysis of four genes compared was performed when necessary. All trees were performed with the corresponding sequences of all species type strains described up to 2012.

4.2.4. DNA-DNA hybridisation

Genomic DNA was isolated using a previously described method (Marmur, 1961). DNA-DNA relatedness values were calculated in duplicate using a non-radioactive method, as previously described (Ziemke *et al*., 1998). DNA samples as probes against the other strains were double-labelled with DIG-11-dUTP and biotin-16dUTP using a nick-translation kit (Roche). Colour development was monitored at 405 nm using an iMark Microplate Absorbance Reader (BioRad).

4.2.5. Fatty acids analysis

FAME analysis was performed at the Spanish Type Culture Collection, CECT, in Valencia, Spain (http://cect.org/identificaciones). Strains were cultured for 24 hours in Trypticase Soy Agar at 28^{o}C. Fatty acids were extracted and prepared according to standard protocols, as described in the manufacturer instructions for the MIDI Microbial Identification System (Sasser, 1990). Cellular fatty acid content was analysed using GC with an Agilent 6850 chromatographic unit, with the MIDI Microbial Identification System using the RTSBA6 method (MIDI, 2008) and the Microbial Identification Sherlock software package version 6.1.

4.2.6. Phenotypic tests

The presence of fluorescent pigments was tested on King B medium (*Pseudomonas* agar F, Difco), and pyocyanin production was tested on King A medium (*Pseudomonas* agar P, Difco). The strains were characterised phenotypically using API 20 NE strips (bioMérieux) and Biolog GN2 MicroPlates (Biolog, Hayward, CA). The oxidase test was performed using

cytochrome oxidase test paper. The catalase activity was detected by adding 3% hydrogen peroxide solution to resting cells and examining the production of bubbles.

The cell size, morphology and flagella insertion were determined by transmission electron microscopy of cells from the exponential growth phase in LB. A Hitachi model H600 electron microscope was used at 75 kV. The samples were negatively stained with phosphotungstic acid (1%, pH 7.0), as described by Lalucat (1988).

Growth temperatures (4, 15, 20, 30, 37 and 42^{o}C) were determined in LB medium and growth in the presence of NaCl (0-10% w/v) and pH ranges (3-10) were observed in Nutrient Broth (Merck).

4.2.7. Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry

To determine the whole cell protein profile, a matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) analysis of strains VGXO14^{T} and Vi1 and their closely related type strains at the Scientific-Technical Services (University of Balearic Islands, Spain). Strains were cultured on LB plates at 30 ^{o}C for 24-48 h. One microliter of the extract (these extracts were obtained based on the ethanol/formic acid extraction procedure of Bruker Daltonics) was placed onto a spot of a ground steel plate. Each sample was overlaid with 1 µl of matrix solution (saturated solution of α-cyno-4-hydroxy-cinnamic acid in 50% acetonitrile-2.5% trifluoroacetic acid) and air dried at room temperature. Measurements were performed on an Autoflex III MALDI-TOF/TOF mass spectrometer (Bruker Daltonics, Leipzig, Germany). Spectra were recorded in the linear, positive mode in the mass range 2-20 kDa (IS1= 20 kV, IS2 = 18.7 kV, lens= 6.50 kV, and PIE= 200 ns).

For each spectrum 500 shots were collected approximately from different positions. The profile obtained for each species was analysed and compared, and the corresponding dendrogram was generated using the MALDI BioTyper software (version 1.0; Bruker Daltonics).

4.3. Results and Discussion

4.3.1. Phylogenetic characterization of isolates from soil

The 89 isolates obtained in the different media and growth conditions as well as their phylogentic affiliation are indicated in table XY Sixty-six isolates were Gram negative (75.28%) and 22 isolates were Gram positive (24.72%). The Gram negatives isolates: 62 (69.66%) belonged to the Proteobacteria class: 17 isolates were *Alphaproteobacteria* (19.10%), 2 *Betaproteobacteria* (2.25%) and 43 *Gammaproteobacteria* (48.31%). The

Gammaproteobacteria isolates were classified in the genera *Pantoea* (17 isolates), *Pseudomonas* (11 isolates), *Stenotrophomonas* (10 isolates), *Xanthomonas* (3 isolates), *Erwinia* and *Enterobacter* (1 isolate each) (Table 4.2 and Fig. 4.1).

The distribution of the isolates belonging to the Gram negatives groups reflected that the Proteobacteria is the predominant class (69.66%). The *Gammaproteobacteria* isolates were most abundant 48.31%. *Pantoea* represents the 19.10% of the isolates, followed by *Pseudomonas* (12.50%) and *Stenotrophomonas* (11.36%). In the *Alphaproteobacteria* group *Rhizobium* was another abundant genus representing 11.36% of the isolates. In Gram positives, *Actinobacteria* was the phylum with the highest number of representatives (16.85%), followed by *Bacillus* in the Firmicutes class (6.82%).

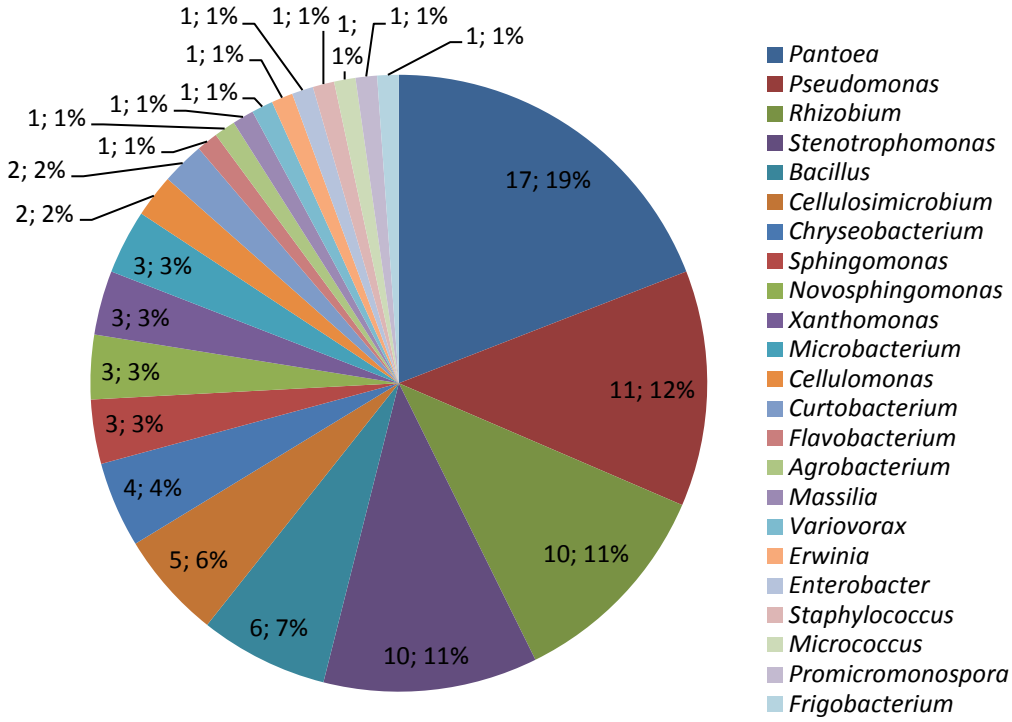


Figure 4.2: Distribution of isolates genera from samplings M0909 and M0910.

Table 4.2: Classification of the different isolates.

Gram stain	No. of isolates	%	Phylum	No. of isolates	%	Class	No. of isolates	%	Genus	M0909[a]	M0910[b]	No. of isolates	%
Gram negative	67	75,28	Bacteroidetes	5	5,62	*Flavobacteria*	5	5,62	*Chryseobacterium*	2	2	4	4,55
									Flavobacterium	1	0	1	1,14
			Proteobacteria	62	69,66	*Alphaproteobacteria*	17	19,10	*Rhizobium*	1	9	10	11,36
									Agrobacterium	0	1	1	1,14
									Sphingomonas	0	3	3	3,41
									Novosphingobium	0	3	3	3,41
						Betaproteobacteria	2	2,25	*Massilia*	1	0	1	1,14
									Variovorax	0	1	1	1,14
						Gammaproteobacteria	43	48,31	*Pantoea*	15	2	17	19,32
									Pseudomonas	4	7	11	12,50
									Stenotrophomonas	2	8	10	11,36
									Xanthomonas	0	3	3	3,41
									Erwinia	1	0	1	1,14
									Enterobacter	0	1	1	1,14
Gram positive	22	24,72	Firmicutes	7	7,87	*Bacilli*	7	7,87	*Staphylococcus*	0	1	1	1,14
									Bacillus	6	0	6	6,82
			Actinobacteria	15	16,85	*Actinobacteria*	1	1,12	*Micrococcus*	1	0	1	1,14
							2	2,25	*Cellulomonas*	2	0	2	2,27
							6	6,74	*Cellulosimicrobium*	0	5	5	5,68
									Promicromonospora	0	1	1	1,14
							5	5,62	*Microbacterium*	2	1	3	3,41
									Curtobacterium	0	2	2	2,27
							1	1,12	*Frigobacterium*	0	1	1	1,14
Total										**38**	**51**	**89**	

a) Sampling September 2009, **b)** Sampling September 2010

Eleven *Pseudomonas* isolates were obtained, 4 from M0909 sampling and 7 from M0910 sampling. According to the partial 16S rDNA sequences the closest phylogenetic type strains are indicated in Table 4.3.

Table 4.3: 16S rRNA gene analysis *of Pseudomonas* isolates from soil.

Isolate	New Name	Growth Media	Closest strain	% similarity
110909	S11	R2A	*Pseudomonas thivervalensis*	99.8
120909	S12	R2A	*Pseudomonas japonica*	98.8
150909	S15	R2A	*Pseudomonas mohnii*	100
170909	S17	R2A	*Pseudomonas umsongensiss*	99.5
10910	S1	R2A	*Pseudomonas koreensis*	99.8
250910	S25	Middlebrook	*Pseudomonas ficuserectae*	99.6
300910	S30	Middlebrook	*Pseudomonas viridiflava*	99.9
360910	S36	Middlebrook	*Pseudomonas umsongensis*	99.6
420910	S42	Gould S1	*Pseudomonas umsongensis*	99.6
430910	S43	Gould S1	*Pseudomonas koreensis*	99.8
580910	S58	M+A	*Pseudomonas thivervalensis*	99.8

4.3.2. Phylogenetic affiliation of the isolates

Following the *Pseudomonas* classification established in Mulet and collaborators (2012) a MLSA concatenated analysis of the 16Sr RNA, *gyrB* and *rpoD* genes allowed the following phylogenetic classification: 8 isolates belonged to the *P. fluorescens* group. S150909 (S15) was classified in the *P. jesseni* subgroup; S170909 (S17), S010910 (S1), S360910 (S36), S420910 (S42) and S430910 (S43) were classified in *P. koreensis* subgroup; and S110909 (S11) and S580910 (S58) were classified in the *P. corrugata* subgroup. S250910 (S25) and S300910 (S30) were located in the *P. syringae* group and S120909 (S12) in the *P. putida* group (Fig. 4.3).

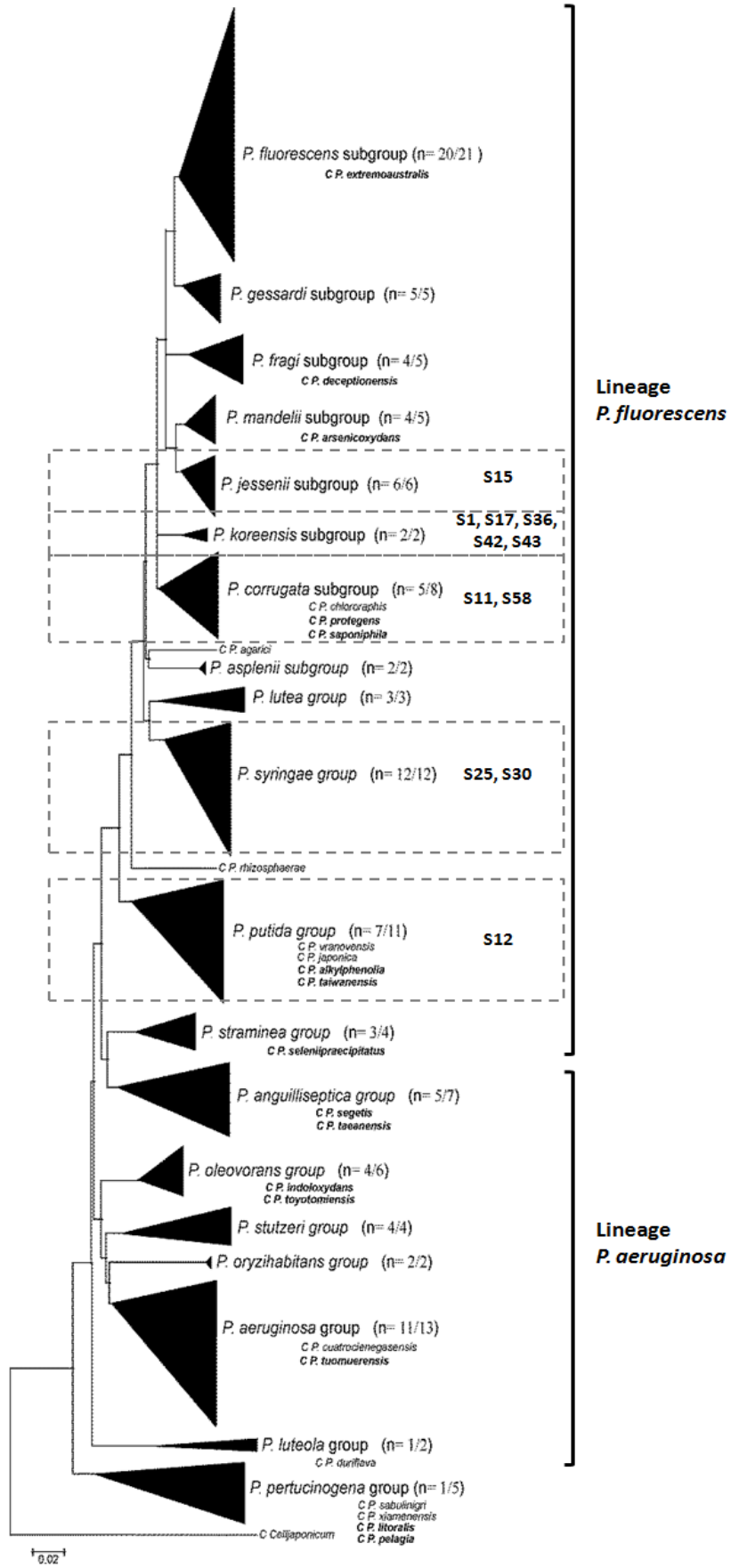


Figure 4.3: Phylogenetic tree with the ubication of soil isolates**.** Boxes indicate to which groups or subgroups was affiliated the *Pseudomonas* isolates.

Strain S30

P. viridiflava was the closest type strain to S30 with a 98.8% of similarity and this strain was assigned by the concatenated of 3 genes to this *Pseudomonas* species (Supplementary Table 4.S3). The 16S rRNA (1283 nucleotides), *gyrB* (804 nucleotides), and *rpoD* (701 nucleotides) genes have been included in the analysis, with a total of 2786 nucleotides being analysed.

Strains S1, S15, S17, S36, S42 and S46

An analysis of partial sequences of the 16S rRNA, *gyrB*, and *rpoD* genes among all of the *Pseudomonas* type strains revealed that strains S1, S15, S17, S36, S42 and S46 might represent a new species. The sequences of the 16S rRNA (1276 nucleotides), *gyrB* (807 nucleotides), and *rpoD* (696 nucleotides) genes have been included in the analysis, with a total of 2779 nucleotides being analysed.

P. umsongensis was the closest strain to S15 with a 97.5% of similarity in the concatenated tree (Fig. 4.5) (Supplementary Table 4.S2). Labelled DNA of S15 and *P. umsongensis* LMG 21317^T were hybridised separately with DNA from other type strains in the *P. jessenii* subgroup: *P. jessenii* ATCC 700870^T and *P. vancouverensis* ATCC 700688^T (data not shown). S15 showed DDH values higher than 70%, with the *P. umsongensis* LMG 21317^T accomplish the 70% threshold established for the species DNA-DNA discrimination. Similarities with labelled DNA of *P. umsongensis* LMG 21317^T were higher than 70% with the strains S15. These results confirmed that strain S15 can be affiliated to the species *P. umsongensis*.

P. moraviensis was the closer type strain to S1, S17, S36, S42 and S43 with a 97.7, 97.7, 97.8, 97.8 and 97.6% of similarity respectively in the concatenated tree (Fig. 4.5) (Supplementary Table 4.S2). Labelled DNA of S17 and *P. moraviensis* DSM 16007^T were hybridised separately with DNA from S1, S36 and *P. koreensis* LMG 21318T (data not shown). S17 showed DDH values higher than 70%, with the *P. moraviensis* DSM 16007^T and accomplished the 70% threshold established for the species DNA-DNA discrimination. Similarities with labelled DNA of *P. moraviensis* DSM 16007^T were higher than 70% with the strains S17, S1, S36. These results confirmed that strains S17, S1, S36, S42 and S43 can be affiliated to *P. moraviensis*. The five strains are 97.7-99.2% similar according to the concatenated analysis.

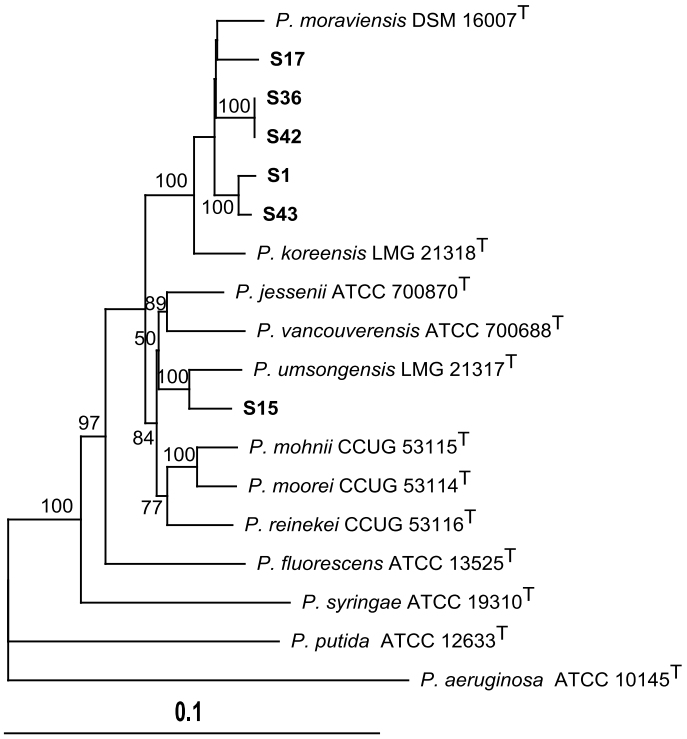


Fig. 4.5: Phylogenetic tree based on concatenated 16S rRNA, *gyrB* and *rpoD* genes of S1, S15, S17, S36, S42 and S43, and phylogenetic close members of *Pseudomonas*. Distance matrices were calculated by the Jukes-Cantor method. Dendrograms were generated by the neighbor-joining method. *Pseudomonas aeruginosa* ATCC 10145^T was used as outgroup. The bar indicates sequence divergence. Percentage bootstrap values of more than 50% (from 1000 replicates) are indicated at the nodes.

4.3.3. Taxonomic characterization of soil strain S12

An analysis of partial sequences of the 16S rRNA, *gyrB*, and *rpoD* genes among all of the *Pseudomonas* type and the culture collection strains revealed that strains S12, 18F, 19F, 35F, 42F and IFO 3738 might represent a new species. The sequences of the 16S rRNA (1274 nucleotides), *gyrB* (804 nucleotides), and *rpoD* (680 nucleotides) genes have been included in the analysis, with a total of 2773 nucleotides being analysed. In all individual and concatenated gene trees studied, S12, 18F, 19F, 35F, 42F and IFO 3738 were located in the *P. putida* G. In all gene trees all strains clustered together and formed a branch that was independent from the other type strains in the *P. putida* G. *P. japonica* JCM 21532^T was the closest strain (Fig. 4.6). High bootstrap values supported these four trees (concatenated 86%, 16S rRNA gene 55%, *rpoD* gene 64% and *gyrB* gene 72%). The concatenated tree indicated that the closest type strain to strain S12, 18F, 19F, 35F, 42F and IFO 3738 was *P. japonica* JCM 21532^T (91.9% for S12, 18F, 19F, 35F and 42F; and 91.3% for IFO 3738) (Supplementary Table 4.S3).

The lowest similarity in the concatenated analysis within the *P. putida* group was 87.5% between *P. fulva* ATCC 21418^T and *P. vranovensis* DSM 16006 ^{T}; the highest similarity was 96.1%

between *P. mosselii* ATCC BAA-99^T and *P. entomophila* L48^T. The intragroup average similarity value calculated for members of the *P. putida* G was 91.7%. The distance to *P. aeruginosa* ATCC 10145^T, which is the type species of the *Pseudomonas* genus, was 81.5 to 85.2% (Supplementary Table 4.S3). The similarities from 91.3 -91.9% of S12, 18F, 19F, 35F, 42F and IFO 3738 strains did not allow their affiliation whith any *Pseudomonas* species previously described and should be considered as hypothetical representatives of a new species. The six strains are 99.2-100% similar according to the concatenated analysis.

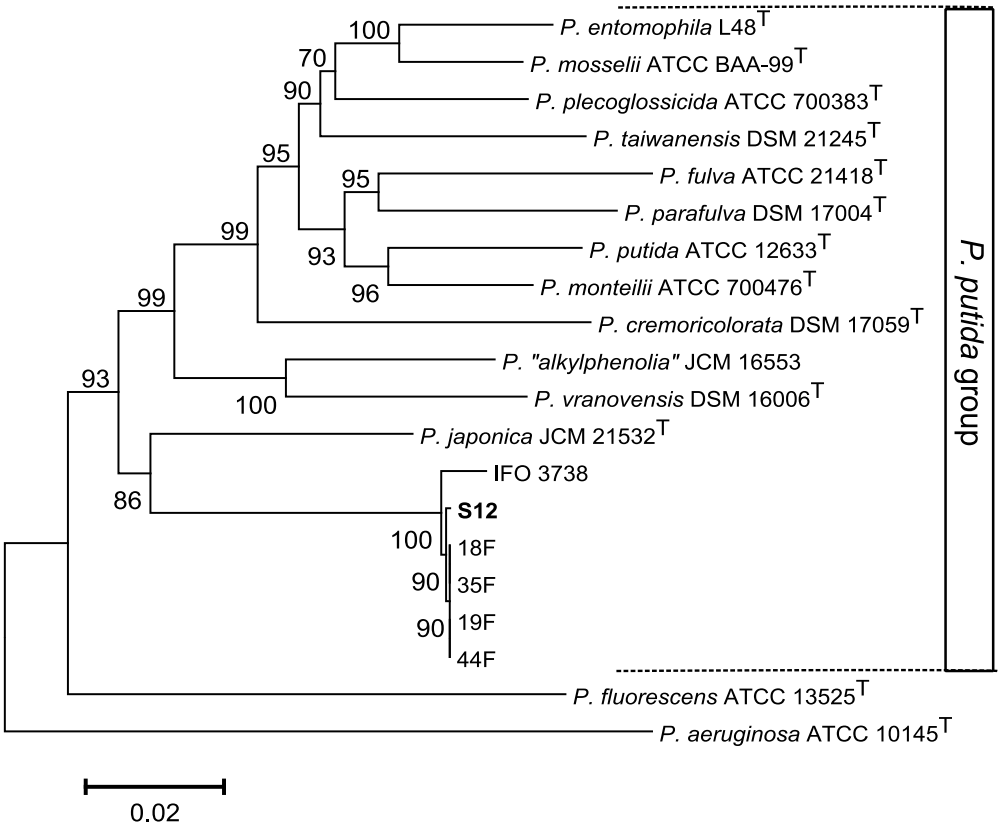


Fig. 4.6: Phylogenetic tree based on concatenated 16S rRNA, *gyrB* and *rpoD* genes of S12, 18F, 19F, 35F, 42F and IFO 3738 and phylogenetic close members of *Pseudomonas*. Distance matrices were calculated by the Jukes-Cantor method. Dendrograms were generated by the neighbor-joining method. *Pseudomonas aeruginosa* ATCC 10145^T was used as outgroup. The bar indicates sequence divergence. Percentage bootstrap values of more than 50% (from 1000 replicates) are indicated at the nodes.

Labelled DNA of S12, 18F and *P. japonica* JCM 21532^T were hybridised separately with DNA from 44F and IFO 3738. Labelled DNA of S12 and 18F strains showed DDH values below 50% with the *P. japonica* JCM 21532^T not accomplish the 70% threshold established for the species DNA-DNA discrimination. Similarities with labelled DNA of *P. japonica* JCM 21532^T were lower than 40% with the strains S12, 18F and IFO 3738. Labelled DNA of S12 and 18F showed DDH values of 100% between them and with IFO 3738. These results confirmed that strains S12, 18F, 19F, 35F and 44F belonged to the same *Pseudomonas* species but were not

members of any other previously described species (Table 4.1).

Strains S12, 18F, 19F, 35F, 44F and IFO 3738 are Gram negative, rod-shaped bacteria. The strain S12 was positive for catalase and oxidase activities. After incubation for 24 h at 30 ^{o}C on LB plates, the colonies were round (1-3 mm of diameter) slightly convex and coloured beige, regular margins. Strain S12 was able to grow in LB medium at 4-37 ^{o}C in 24-48h. However, no growth was detected at 42 ^{o}C. Growth was observed on nutrient broth in the presence of 0-8% NaCl (w/v) and tolerated pH levels ranging from 5-10.

Strains S12 did not exhibited fluorescent pigmentation on *Pseudomonas* agar F, and no pyocyanin production was observed on *Pseudomonas* agar P when was cultured for 24-48 h at 30 ^{o}C. 18F, 19F, 35F, 44F and IFO3738 exhibited fluorescent pigmentation on *Pseudomonas* agar F.

The preliminary phenotypical data together with the DDH and concatenated genes analysis indicates that S12, 18F, 19F, 35F, 44F and IFO 3738 strains could be candidates for a putative novel species. Additional physiological and biochemical characteristics should be included to describe properly these strains as new species.

4.3.4. Taxonomic characterization of soil strain S25

An analysis of partial sequences of the 16S rRNA, *gyrB,* and *rpoD* genes among all of the *Pseudomonas* type and culture collection strains revealed that strains S25, 24F, 58F, 102F, 103F and 122F might represent a new species. The sequences of the 16S rRNA (1283 nucleotides), *gyrB* (804 nucleotides), and *rpoD* (701 nucleotides) genes have been included in the analysis, with a total of 2786 nucleotides being analysed. In all individual and concatenated gene trees studied, S25 and 24F, 58F, 102F, 103F and 122F were located in the *P. syringae* G. In all gene trees all strains clustered together and formed a branch that was independent from the other type strains in the *P. syringae* G. *P. caricapapayae* LMG 2152^{T} was the closest strain (Fig. 4.7). High bootstrap values supported these tree branches (concatenated 70%, 16S rRNA gene 38%, *rpoD* gene 70% and *gyrB* gene 75%). The concatenated tree indicated that the closest type strain to strain S25, 24F, 58F, 102F, 103F and 122F was *P. caricapapayae* LMG 2152^{T} (91.8%, 91.6%, 91.7%, 91.7%, 91.6% and 91.6% respectively) (Supplementary Table 4.S4).

The lowest similarity in the concatenated analysis within the *P. syringae* group was 90.8% between *P. avellanae* CIP 105176^{T} and *P. cichorii* ATCC 10857^{T}; the highest similarity was 99.0% between *P. amydali* LMG 1384^{T} and *P. tremae* LMG 22121^{T}. The intragroup average similarity value calculated for members of the *P. syringae* G was 95.2%. The distance to *P. aeruginosa*

ATCC 10145^T, which is the type species of the *Pseudomonas* genus, was 81.60 to 83.14% (Supplementary Table 4.S4). The similarities from 91.6-91.8% of strain S25, 24F, 58F, 102F, 103F and 122F strains did not allow the affiliation with any other *Pseudomonas* species previously described and should be considered as hypothetical representatives of a new species. The strains 24F, 58F, 102F, 103F and 122F were 99.5-99.9% similar according to the concatenated analysis, but S25 was 95.9-96.0% with the strains 24F, 58F, 102F, 103F and 122F. They could represent two different new *Pseudomonas* species.

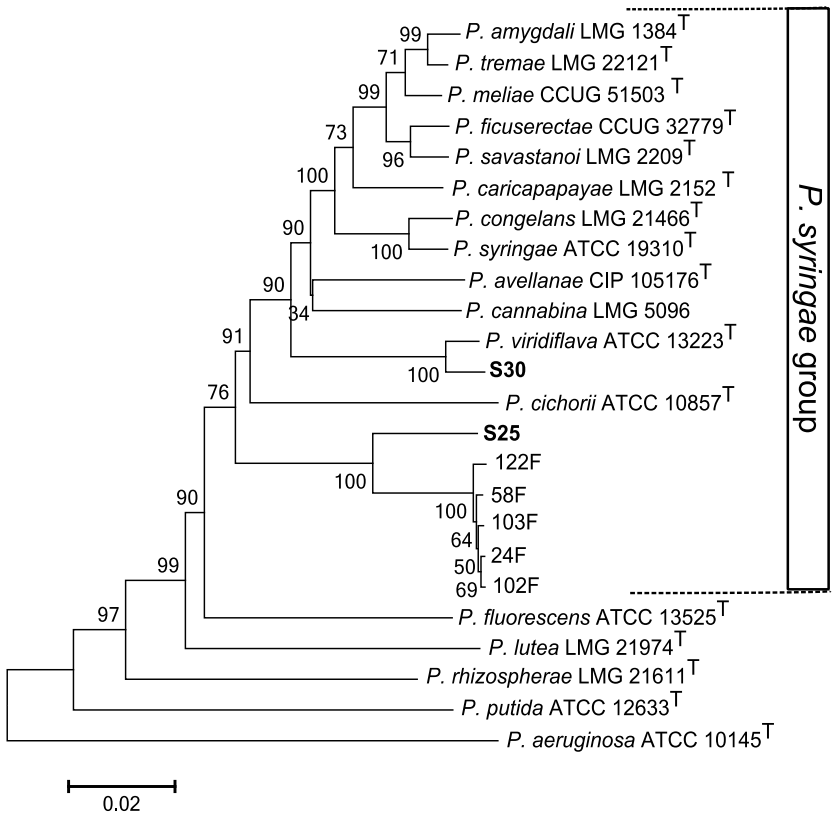


Fig. 4.7: Phylogenetic tree based on concatenated 16S rRNA, *gyrB* and *rpoD* genes of S25, 24F, 58F, 102F, 103F and 122F and phylogenetic close members of *Pseudomonas*. Distance matrices were calculated by the Jukes-Cantor method. Dendrograms were generated by the neighbor-joining method. *Pseudomonas aeruginosa* ATCC 10145^T was used as outgroup. The bar indicates sequence divergence. Percentage bootstrap values of more than 50% (from 1000 replicates) are indicated at the nodes.

Labelled DNA of S25 and 122F were hybridised separately with DNA from all type strains of P. *syringae* group. Labelled DNA of S25^T and 122F strains showed DDH values below 60%, with all type strains of P. *syringae* group did not accomplish the 70% threshold established for the species DNA-DNA discrimination. Labelled DNA of S25 and 122F showed DDH values below 50%. These results confirmed that strains S25, 24F, 58F, 102F, 103F and 122F are *Pseudomonas* species but not members of any other previously described species (Table 4.1), and they cannot be allocated in the same species; S25 has to be considered a different species than

strains 24F, 58F, 102F, 103F and 122F.

Strain S25 is a Gram negative, rod-shaped bacterium. The strain S25 was positive for catalase and oxidase activities. After incubation for 24 h at 30^0 C on LB plates, the colonies were round (1-1,5 mm of diameter) flat and coloured beige, regular margins. Strain S25 was able to grow in LB medium at 4-30 ^oC in 24-48h and showed a slow growth at 37 ^oC after one week of incubation. However, no growth was detected at 42 ^oC. Growth was observed on nutrient broth in the presence of 0-6% NaCl (w/v) and tolerated pH levels ranging from 5-9. Strain S25 did not exhibit fluorescent pigmentation on *Pseudomonas* agar F, and no pyocyanin production was observed on *Pseudomonas* agar P when cultured for 24-48 h at 30 ^oC.

The molecular and preliminary phenotypic data indicated that S25, 24F, 58F, 102F, 103F and 122F strains could be candidates for a putative novel species. Additional physiological and biochemical characteristics should be included to describe properly these strains as new species.

4.3.5. Taxonomic characterization of soil strains S11 and S58^T

Pseudomonas thivervalensis was the closest type strain to S11 and S58^T with a 97.70 and 97.42% of similarity respectively. An analysis of partial sequences of the 16S rRNA, *gyrB, rpoD* and *rpoB* genes among all of the *Pseudomonas* type strains revealed that strains S58^T and CFBP 4629 might represent a new species. The sequences of the 16S rRNA (1280 nucleotides), *gyrB* (803 nucleotides), *rpoD* (723 nucleotides) and *rpoB* (905 nucleotides) genes have been included in the analysis, with a total of 3713 nucleotides being analysed. In all individual and concatenated gene trees studied, S58^T and CFBP 4629 were located in the *P. corrugata* SG, within the *P. fluorescens* G. The strain CFBP 4629 had already been affiliated with the *P. corrugata* SG in a previous phylogenetic study (Mulet *et al*., 2013). In all gene trees, with the exception of the 16S rDNA, strains S58^T and CFBP 4629 clustered together and formed a branch that was independent from the other type strains in the *P. corrugata* SG. *P. thivervalensis* DSM 13194^T was the closest strain (Fig. 4.8.). High bootstrap values supported these tree branches (concatenated 82%, *rpoB* gene 79%, *rpoD* gene 72% and *gyrB* gene 59%). The concatenated tree indicated that the closest type strain to strain S58^T and CFBP 4629 was *P. thivervalensis* DSM 13194^T (97.4% and 97.6% similarity, respectively) (Supplementary Table 4.S5). Similar results with high bootstrap values were obtained when the maximum likelihood and maximum parsimony algorithms were applied (data not shown).

The lowest similarity in the concatenated analysis within the *P. corrugata* group was 95.0% between *P. brassicacearum* DSM 13227^T and *P. kilonensis* DSM 13647^T; the highest

similarity was 98.0% between *P. brassicacearum* DSM 13227^T and *P. corrugata* ATCC 29736^T. The intragroup average similarity value calculated for members of the *P. corrugata* SG was 96.8%. The distance to *P. aeruginosa* ATCC 10145^T, which is the type species of the *Pseudomonas* genus, was 85.2 to 84.4% (Supplementary Table 4.S5). The distance to the type strains of the other subgroups in the *P. fluorescens* G was 95.5 to 91.5% (data not shown). The similarities higher than 97% between strains S58^T and CFBP 4629 and between strains *P. thivervalensis* DSM 13194^T and *P. brassicacearum* DSM 13227^T affiliated both sets of strains to the *P. corrugata* SG; however, the similarity values were similar when comparing previously established type strains of the same *P. corrugata* SG. Strains S58^T and CFBP 4629 cannot be affiliated with any other *Pseudomonas* species previously described and should be considered as hypothetical representatives of a new species. Both strains were 99.5% similar according to the concatenated analysis.

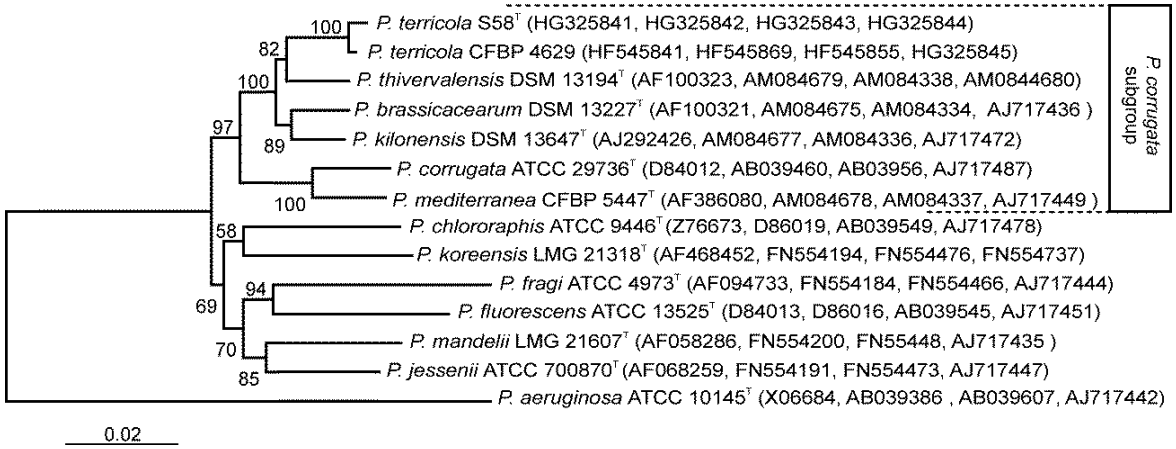


Fig. 4.8: Phylogenetic tree based on concatenated 16S rRNA, *gyrB*, *rpoD* and *rpoB* gene of *Pseudomonas terricola* and phylogenetic close members of *Pseudomonas*. Distance matrices were calculated by the Jukes-Cantor method. Dendrograms were generated by the neighbor-joining method. *Pseudomonas aeruginosa* ATCC 10145^T was used as outgroup. The bar indicates sequence divergence. Percentage bootstrap values of more than 50% (from 1000 replicates) are indicated at the nodes. GenBank accession numbers are given in parentheses in the following order: 16S rRNA, *gyrB*, *rpoD* and *rpoB* gene.

Labelled DNA of S58^T and *P. thivervalensis* DSM 13194^T were hybridised separately with DNA from strain CFBP 4629 and the other type strains in the *P. corrugata* group (Table 4.4). All type strains showed DDH values below 60%, with the S58^T labelled DNA falling clearly below the 70% threshold established for the species DNA-DNA discrimination. The strains S58^T and CFBP 4629 showed 84% total DNA-DNA similarity, indicating that both strains belong to the same species. Similarities with labelled DNA of *P. thivervalensis* were lower than 70% with the strains S58^T and CFBP 4629. These results confirmed that strains S58^T and CFBP 4629 were not members of any other previously described species.

However, DNA-DNA hybridisation data suggested a close affiliation between some of the type strains in the *P. corrugata* SG. *P. thivervalensis* and *P. kilonensis* showed approximately 70% DDH values, and the DDH value between *P. thivervalensis* and *P. mediterranea* was slightly higher than the 70% threshold (75%). Similar results have been obtained in previous studies (Supplementary Table S3). Sikorski and colleagues reported a DNA hybridisation similarity between *P. kilonensis* DSM 13647^T and *P. thivervalensis* DSM 13194^T that almost reached the 70% threshold value, and *P. kilonensis* DSM 13647^T and *P. brassicacearum* DSM 13227^T showed 75% DDH values. The two species *P. brassicacearum* and *P. thivervalensis* could be distinguished from *P. kilonensis* phenotypically and by riboprinting, but they were shown to be closely related phylogenetically (16S rDNA similarity) (Sikorski *et al*., 2001).

The cellular fatty acid methyl ester composition (CFA) was determined using gas chromatography. Fatty acid profiles were similar in strains S58^T and CFBP 4629 (Table 4.5). The major CFAs for the two strains included summed feature 3 ($C_{16:1}$ ω7*c* and $C_{16:1}$ ω6*c*; 33.2-31.1%), $C_{16:0}$ (25.9-25.5%), summed feature 8 ($C_{18:1}$ ω7*c* and $C_{18:1}$ ω6*c*; 24.0-23.2%), $C_{12:0}$ (4.5-5.0%), $C_{12:0}$ 3-OH (4.3-4.2%), $C_{10:0}$ 3-OH (3.5-3.6%) and $C_{12:0}$ 2-OH (3.0-2.7%); all the other detected CFA values were lower than 2%. The only differences detected in the fatty acid profiles between the two strains were anteiso- $C_{17:0}$ and $C_{18:1}$ ω9*c* (0.7% and 0.6%, respectively), which were present in low percentages in strain S58^T but not in strain CFBP 4629. Both strains could be differentiated from *P. thivervalensis* DSM 13194^T *and P. brassicacearum* DSM 13227^T by the fatty acids summed feature 8 ($C_{18:1}$ ω7*c* and $C_{18:1}$ ω6*c*) and $C_{17:0}$ cyclo, which exhibited different percentage values (24.0-23.2% to 18.7-14.0%).

Strains S58^T and CFBP 4629 are Gram negative, rod-shaped bacteria (0.5-0.8 µm wide and 2.2-3.0 µm long). Cells are motile, and single polar flagella were observed under transmission electron microscopy (Supplementary Fig. 4.S1). The strain S58^T was positive for catalase and oxidase activities. After incubation for 24 h at 30 ^{o}C on LB plates, the colonies were round (1-2 mm of diameter) and coloured beige and had flat, regular margins, were translucent and produced brown-black diffusible pigment when grown on LB plates after incubation for 48 h at 30 ^{o}C. Strain S58^T was able to grow in LB medium at 18-37 ^{o}C in 24 h and showed a slow growth at 4 ^{o}C after one week of incubation. However, no growth was detected at 42 ^{o}C. Growth was observed on nutrient broth in the presence of 0-6% NaCl (w/v) and tolerated pH levels ranging from 5-10. Strains S58^T and CFBP 4629 exhibited fluorescent pigmentation on *Pseudomonas* agar F, and no pyocyanin production was observed on *Pseudomonas* agar P when cultured for 24-48 h at 30 ^{o}C.

Table 4.4: Characteristics that differentiate *P. terricola* S58^T from related species of the genus *Pseudomonas*. Strains: 1. *P. terricola* S58^T, 2. *P. terricola* CFBP 4629, 3. *P. thivervalensis* DSM 13194^T, 4. *P. brassicacearum* DSM 13227^T, 5. *P. corrugata* ATCC 29736^T, 6. *P. kilonensis* DSM 13647^T, 7. *P. mediterranea* CFBP 5447^T.

Characteristics	1	2	3	4	5	6	7
16S rDNA similarity (%) with S58^T	100	99.8	99.8	99.4	99.4	99.7	98.7
MLSA 4 genes (%) with S58^T	100	99.5	97.4	97.2	94.9	97.0	95.2
DDH (%)[a]							
S58^T labelled	100	83.5	58.1	41.8	41.9	46.7	47.2
P. thivervalensis^T labelled	56.9	63.7	100	56.0	60.3	70.1	74.9
Activity of enzymes (API 20NE test)							
Reduction of nitrate a nitrite	-	-	-	-	-	-	+
Denitrification	+	+	-	-	+	+	-
Arginine dihydrolase	-	-	+	+	-	+	-
Esculin hydrolysis (β-glucosidase)	+	-	-	-	-	-	-
Gelatin hydrolysis (protease)	+	+	+	+	-	+	+
Growth on (API 20NE test)							
N-Acetil-glucosamine	+	+	+	+	+	-	+
Maltose	+	-	-	-	-	-	-
Carbon sources (Biolog GN2)							
α-cyclodextrin	+	-	-	-	-	-	-
Dextrin	+	w	-	-	-	-	-
Tween 40	+	+	+	+	w	-	+
N-Acetyl-D-glucosamine	+	+	+	+	-	-	+
D-Cellobiose	+	-	-	-	-	-	-
Maltose	+	-	-	-	-	-	-
D-Sorbitol	+	+	+	+	w	+	-
D-Galacturonic Acid	+	+	w	w	+	-	+
D-Glucuronic Acid	w	+	w	+	+	-	+
γ- Hydroxy Butyric Acid	+	-	+	+	w	+	-
α-keto Butyric Acid	+	-	-	-	-	-	-
α-keto Valeric Acid	w	-	w	-	w	w	-
Succinamic Acid	-	-	+	+	+	+	+
Glucuronamide	w	w	+	+	+	-	+
Glycyl-L-Aspartic Acid	+	-	-	-	-	-	-
L-Ornithine	-	+	-	-	w	w	+
2-Aminoethanol	-	+	+	+	w	+	-
2,3-Butanediol	w	w	+	+	-	-	+

Unless otherwise indicated all data were obtained in this study. In the API 20 NE test all strains were positive for the following characteristics: assimilation of glucose, arabinose, mannose, mannitol, potassium gluconate, capric acid, malic acid and trisodium citrate; but were negative: indole production, fermentation of glucose, urease, β-galactosidase, assimilation of adipic acid and phenylacetic acid. In the oxidase and catalase tests all the strains were positive. In the Biolog GN2 test all strains were positive for the following characteristics: Tween 80, L-arabinose, D-arabitol, D-fructose, D-galactose, α-D-glucose, m-inositol, D-mannitol, D-mannose, sucrose, D-trehalose, pyruvic acid methyl ester, mono-methyl-succinate, acetic acid, cis-aconitic acid, citric acid, formic acid, D-galactonic acid lactone, D-gluconic acid, D-glucosaminic acid, α-hydroxy butyric acid, β-hydroxy butyric acid, p-hydroxy phenylacetic acid, a-keto glutaric acid, D,L-lactic acid, malonic acid, propionic acid, quinic acid, D-saccharic acid, succinic acid, bromo succinic acid, L-alaninamide, D-alanine, L-alanine, L-alanyl-glycine, L-asparagine, L-aspartic acid, L-glutamic acid, glycyl-L-glutamic acid, L-histamine, hydroxy-L-proline, L-leucine, L-proline, L-pyroglutamic acid, L-serine, L-threonine, D,L-carnitine, γ-amino butytic acid, urocanic acid, inosine, uridine, putrescine, glycerol, D,L-α-glycerol phosphate, but were negative: glycogen, N-acetyl-D-galactosamine, adonitol, i-erythritol, L-fucose, gentiobiose, α-D-lactose, lactulose, D-melibiose, β-methyl-D-glucoside, D-psicose, D-raffinose, L-rhamnose, turanose, xylitol, itaconic acid, sebacic acid, L-phenylalanine, D-serine, thymidine, phenyethylamine, glucose-1-phosphate, glucosa-6-phosphate. +, positive; -, negative; w, weak; ND, not determined. [a]

Pooled standard deviations in the DDH experiments were 0.80 and 0.77 (with labelled DNA from *P. terricola* S58^T and *P. thivervalensis* DSM 13194^T, respectively).

Table 4.5: Cell fatty acid (CFA) profiles of *P. terricola* S58^T and CFBP 4629 and the type strains of related species of the genus *Pseudomonas*. Strains: 1. S58^T, 2. CFBP 4629, 3. *P. thivervalensis* DSM 13194^T, 4. *P. brassicacearum* DSM 13227^T and 5. *P. kilonensis* DSM 13647^T. Values are percentages of total CFAs. All strains were cultured on TSA medium at 28 ^oC.

Fatty acid (%)	1	2	3	4	5
$C_{10:0}$	0.2	0.2	-	0.3	0.4
$C_{12:0}$	4.5	5.0	3.7	4.0	6.6
$C_{14:0}$	0.5	0.3	0.3	0.6	0.3
$C_{16:0}$	25.5	25.9	27.2	31.8	28.0
$C_{17:0}$	0.1	0.1	-	-	-
$C_{18:0}$	1.3	0.5	0.6	1.5	0.8
$C_{18:1}$ $\omega 9c$	0.6	-	-	1.2	-
anteiso- $C_{17:0}$	0.7	-	-	-	-
$C_{12:0}$ 2-OH	3.0	2.7	4.0	3.9	1.8
$C_{10:0}$ 3-OH	3.5	3.6	4.1	3.5	4.1
$C_{12:1}$ 3-OH	0.4	0.2	0.6	-	0.6
$C_{12:0}$ 3-OH	4.2	4.3	5.2	4.5	5.0
$C_{17:0}$ cyclo	0.8	0.3	3.2	3.2	0.5
unknown 18.905	-	-	0.3	-	-
[a]Summed feature 2	0.3	-	-	-	-
[a]Summed feature 3	31.1	33.2	32.2	31.5	35.4
[a]Summed feature 5	0.1	-	-	-	-
[a]Summed feature 8	23.2	24.0	18.7	14.0	16.5

[a]Summed feature represents a set of more than one CFA that could not be resolved; Summed feature 2 contained $C_{12:0}$ aldehydes, $C_{14:0}$ 3-OH and iso-$C_{16:1}$, Summed feature 3 contained $C_{16:1}$ $\omega 7c$ and $C_{16:1}$ $\omega 6c$, Summed feature 5 contained $C_{18:2}$ $\omega 6$, $C_{18:2}$ $\omega 9c$ and anteiso-$C_{18:0}$ and Summed feature 8 contained $C_{18:1}$ $\omega 7c$ and $C_{18:1}$ $\omega 6c$; -, Not detectable.

Additional physiological and biochemical characteristics were tested using API 20NE and Biolog GN2 MicroPlates. Differential phenotypic characteristics are indicated in Table 4.4. The strain S58^T differed in eight biochemical tests from the other type strains. The strain S58^T was positive for esculin hydrolysis (β-glucosidase), assimilation of maltose and oxidation of α-cyclodextrin, dextrin, D-cellobiose, maltose, α-keto butyric acid and glycyl-L-aspartic acid. The strain S58^T differed from CFBP 4629 in 11 biochemical tests and from the other strains of the *P. corrugata* SG in more than 11. The strain S58^T differed from *P. thivervalensis* DSM 13194^T in 12 different phenotypic characteristics, from *P. brassicacearum* DSM 13227^T in 13, from *P. corrugata* ATCC 29736^T in 14, from *P. kilonensis* DSM 13647^T in 19 and from *P. mediterranea* CFBP 5447^T in 15. Strains S58^T and CFBP 4629 were able to use dextrin as carbon source, in contrast to the remainder of the type strains studied in the group.

Considering the phylogenetic, chemotaxonomic, genotypic and phenotypic characteristics presented, we propose a new species, *Pseudomonas terricola* sp. nov., with *P. terricola* S58^T as the type strain.

4.3.6. Taxonomic characterization of seashore strains VGXO14^T and Vi1

The *P. pertucinogena* group has experienced a significant increase in the number of newly described species compared with the other groups in recent years. In 2004, this group contained only *P. pertucinogena* (Kawai *et al*., 1975); one year later, *P. pachastrellae* (Romanenko *et al*., 2005) was described, and in the following years, several other species were added, including *P. xiamenensis* (Lai *et al*., 2008), *P pelagia* (Hwang *et al*., 2009), *P. sabulinigri* (Kim *et al*., 2009), *P. bauzanensis* (Zhang *et al*., 2011) and *P. litoralis* (Pascual *et al*., 2012), as shown in Figure. 4.9.

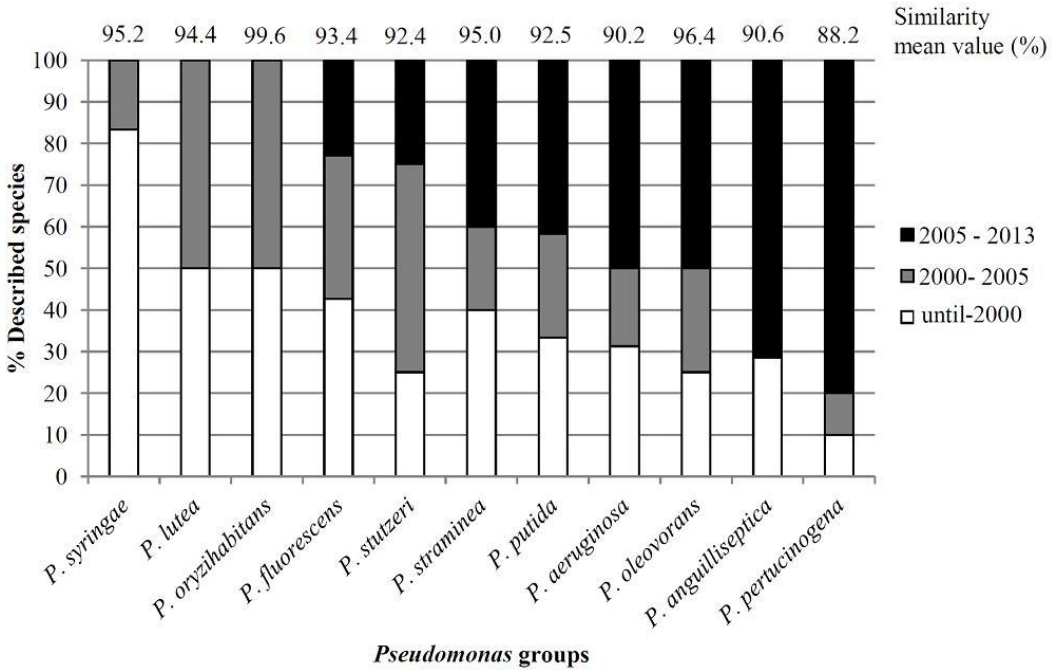


Figure 4.9: Distribution of 134 *Pseudomonas* described species in the different groups established (http://www.bacterio.cict.fr/). Similarity mean value (%) is the average value of similarities in the concatenated sequences between the strains studied within the group.

A preliminary analysis of the partial sequences of the *rpoD* gene compared among all of the *Pseudomonas* type strains revealed that strains VGXO14^T and Vi1 could be representatives of a new species (Mulet *et al*., 2011). A complete phylogenetic analysis has now been accomplished. The sequences of the 16S rRNA (1283 nucleotides) and *gyrB* (802 nucleotides) genes have been included in the analysis, together with an enlarged *rpoD* (741 nucleotides) gene sequence, with a total of 2826 nucleotides. These sequences have been aligned and

compared with the 133 *Pseudomonas* type strains of the data presented in Mulet and collaborators (2012). In all individual and concatenated gene trees studied**,** VGXO14^T and Vi1 were located in the same branch independently from the other type strains in the *P. pertucinogena* group, with *P. pachastrellae* CCUG 46540^T being the closest strain to the group (Fig. 4.9). High bootstrap values supported these tree branches (16S rRNA gene 100%, *gyrB* gene 66%, *rpoD* gene 99% and concatenated 100%). The concatenated tree showed that the closest type strain to strain VGXO14^T and Vi1 is *P. pachastrellae* CCUG 46540^T (90.4% similarity for both) (Supplementary Table 4.S6). Similar results with high bootstrap values were obtained when maximum likelihood and maximum parsimony algorithms were used (data not shown).

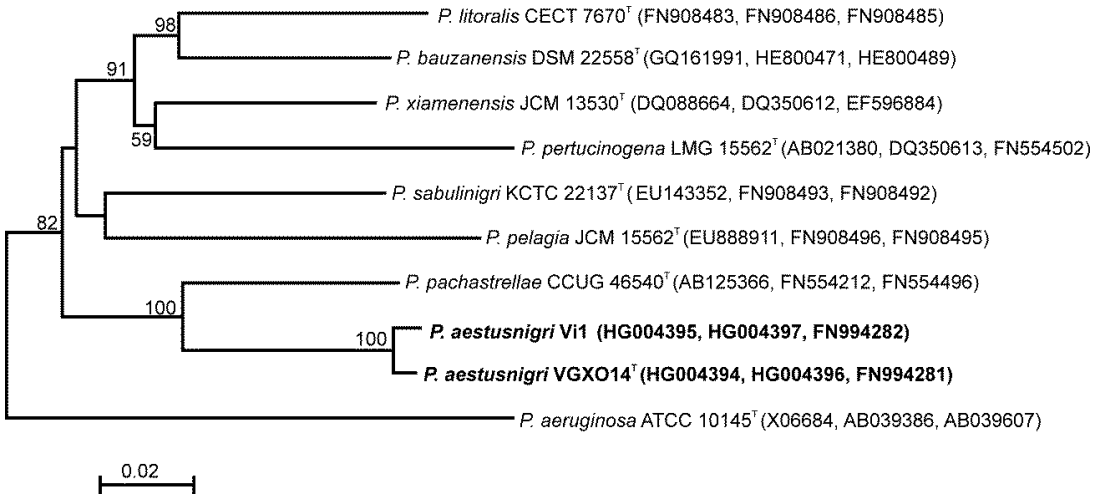


Figure 4.9: Phylogenetic tree based on concatenated 16S rRNA, *gyrB* and *rpoD* gene of *Pseudomonas aestusnigri* and phylogenetic close members of *Pseudomonas*. Distance matrices were calculated by the Jukes-Cantor method. Dendrograms were generated by the neighbor-joining method. *Pseudomonas aeruginosa* ATCC 10145^T was used as outgroup. The bar indicates sequence divergence. Percentage bootstrap values of more than 50% (from 1000 replicates) are indicated at the nodes. GenBank accession numbers are given in parentheses in the following order: 16S rRNA, *gyrB* and *rpoD* gene.

The lowest similarity in the concatenated analysis within the *P. pertucinogena* group was 81.8% between *P. pelagia* JCM 15562^T and *P. pertucinogena* LMG 1874^T; the highest similarity was 90.3% between *P. bauzanensis* DSM 22558^T and *P. xiamenensis* JCM 13530^T. The intragroup average similarity value calculated for members of the *P. pertucinogena* group was 88.2%. The distance to *P. aeruginosa* ATCC 10145^T, which is the type species of the *Pseudomonas* genus, was 81.6% (Supplementary Table 4.S6). The 90.4% similarities between strains VGXO14^T and Vi1 and *P. pachastrellae* affiliated both strains to the *P. pertucinogena* group; however, this similarity is lower than 97%, the threshold established to discriminate species in the genus *Pseudomonas* by MLSA of three concatenated genes (Mulet *et al*., 2010). Strains VGXO14^T and Vi1 cannot be affiliated with any *Pseudomonas* species previously

described and should be considered as representatives of a new species. Both strains are 99% similar in the concatenated analysis.

All type strains showed DDH values below 40% with VGXO14^T-labelled DNA, clearly below the 70% established for the species DNA-DNA discrimination. The strains VGXO14^T and Vi1 showed 80 % total DNA-DNA similarity among them, indicating that both strains belong to the same species. Similarities with labelled DNA of *P. pachastrellae* were lower than 70% with the other strains tested in the group. These results confirmed that strains VGXO14^T and Vi1 were not members of any other previously described species (Table 4.6).

Table 4.6: DNA-hybridization relatedness values and gene sequence similarities between *Pseudomonas aestusnigri* GXO14T and closely related *Pseudomonas* species.

Bacterial strains	**Reassociation (%)[a] with labeled DNA from**		**Gene sequence similarities (%) with respect to strain GXO14^T**	
	***P. aestusnigri* VGXO14^T**	***P. pachastrellae* CCUG 46540^T**	**16S rDNA**	**MLSA**
P. aestusnigri VGXO14^T	100.0	50.0[b]	100.0	100.0
P. aestusnigri Vi1	79.6	66.6	100.0	99.0
P. pachastrellae CCUG 46540^T	32.0[b]	100	99.1	90.4
P. bauzanensis DSM 22558^T	24.3	47.0	95.3	85.1
P. litoralis CECT 7670^T	25.2	35.7	95.3	85.0
P. pertucinogena LMG 1874^T	24.1	48.6	94.3	82.8
P. pelagia JCM 15562^T	26.6	47.8	95.1	83.2
P. sabulinigri KCTC 22137^T	29.4	27.7	95.9	85.9
P. xiamenensis JCM 13530^T	29.7	35.8	95.2	86.2

[a]Pooled standard deviations in experiments DDH were 0.38-0.40 and 0.69-0.82 (with labelled DNA from *P. aestusnigri* VGXO14^T and *P. pachastrellae* CCUG 46540^T, respectively). Reassociation values are means of two determinations.
[b]Reassociation values are means of four determinations, in two different experiments.

The dendrogram obtained by WC-MALDI-TOF MS analysis showed that the two strains VGXO14^T and Vi1 clustered together at 90% similarity and that the similarity to the other strains was less than 35%. The *P. pachastrellae* type strain was the closest strain. The distances to the other strains in the *P. pertucinogena* group were higher than 30%. In the MALDI-TOF MS analysis, VGXO14^T and Vi1 showed four common m/z peaks (supermass values) that were not present in *P. pachastrellae,* the closest type strain: 4808 m/z; 6189-6191 m/z, 7007-7008) m/z, 9622 m/z (Supplementary Figure 4.S2). These data further supported the conclusion that both strains represented a distinct species that was separated from all other species of the *Pseudomonas* genus, even at the level of expression of the most abundant cellular proteins.

Fatty acid profiles were similar in strains VGXO14^T and Vi1 (Table 4.7). Major CFAs for the two strains were summed feature 8 ($C_{18:1}$ ω*6c* and/or $C_{18:1}$ ω*7c*; 33.3-34.5%), summed feature 3 ($C_{16:1}$ ω*7c* and/or $C_{16:1}$ ω*6c*; 25.3-25.8%), $C_{16:0}$ (16.3-16.2%), $C_{12:0}$ (9.5-8.9%), $C_{10:0}$ 3-OH (5.6-5.3%)

and $C_{12:0}$ 3-OH (4.8-4.7%); all other detected CFA values were lower than 2%. The only difference detected in the fatty acid profiles between the two strains was iso-$C_{16:0}$ (0.4%), which was present in low percentage in strain VGXO14^T but not in strain Vi1. Both strains could be differentiated from *P. pachastrellae* CCUG 46540^T by the fatty acid $C_{17:1}$ cyclo and by traces of $C_{17:1}$ ω8*c* and $C_{11:0}$ 3-0H, and by summed feature 3 and 8.

Table 4.7: Cell fatty acid (CFA) profiles of *P. aestusnigri* VGXO14^T and Vi1 and the all type strains *P. pertucinogena* group. Strains: 1. VGX014^T; 2. Vi1; 3. *P. pachastrellae* CCUG 46540^T; 4. *P. bauzanensis* DSM 22558^T; 5. *P. litoralis* CECT 7670T; 6. *P. pelagia* JCM 15562^T; 7. *P. pertucinogena* LMG 1874^T; 8. *P. sabulinigri* KCTC 22137^T; 9. *P. xiamenensis* JCM 13530^T.

Fatty acid (%)	1	2	3	4 [b]	5 [c]	6 [c]	7 [c]	8 [c]	9 [c]
$C_{10:0}$	TR	TR	TR	TR	-	-	-	-	-
$C_{12:0}$	8.9	9.5	10.1	4.5	7.9	8.1	6.1	7.6	9.3
$C_{13:0}$	-	-	-	-	-	TR	1.3	-	-
$C_{14:0}$	TR	1.0	1.6	TR	TR	TR	1.5	1.0	TR
$C_{16:0}$	16.2	16.3	16.9	9.9	18.5	11.8	7.9	18.5	16.1
$C_{17:0}$	-	-	-	TR	TR	TR	3.3	TR	3.0
$C_{18:0}$	TR	TR	TR	TR	1.2	1.0	TR	1.5	1.1
$C_{15:1}$ ω6*c*	-	-	-	TR	TR	TR	1.5	TR	-
$C_{17:1}$ ω8*c*	TR	TR	-	TR	TR	1.1	5.3	TR	TR
$C_{18:1}$ ω9*c*	-	-	-	TR	-	-	-	-	-
iso-$C_{16:0}$	TR	-	-	-	-	-	-	-	-
iso-$C_{17:0}$	-	-	-	-	TR	TR	1.6	1.2	-
anteiso-$C_{17:0}$	-	-	-	-	-	-	TR	1.4	-
$C_{10:0}$ 3-OH	5.3	5.6	3.9	2.0	3.0	3.1	3.7	2.5	3.5
$C_{11:0}$ 3-OH	TR	TR	-	TR	TR	TR	1.7	TR	-
$C_{12:0}$ 3-OH	4.7	4.8	5.1	1.9	4.2	4.4	3.4	4.0	4.5
$C_{16:0}$ 3-OH	-	-	-	TR	-	-	-	-	-
$C_{18:0}$ 3-OH	-	-	-	-	-	-	2.0	-	-
$C_{17:0}$ cyclo	1.3	1.3	-	7.4	8.9	4.4	9.0	1.2	5.1
$C_{19:0}$ cyclo ω8*c*	-	-	-	2.9	1.8	1.3	3.8	TR	-
[a]Summed feature 3	25.3	25.8	34.5	10.3	18.1	24.6	14.3	20.4	12.8
[a]Summed feature 7	TR	TR	TR	-	-	-	-	-	-
[a]Summed feature 8	34.5	33.3	26.9	54.8	32.6	35.4	25.7	36.8	42.5

[a]Summed feature represents a set of more than one CFA that could not be resolved; Summed feature 3 contained $C_{16:1}$ ω7*c* and/or $C_{16:1}$ ω6*c*; Summed feature 7 contained unknown 18.846, $C_{19:1}$ ω6*c* and /or $C_{19:0}$ cyclo ω10*c*; Summed feature 8 contained $C_{18:1}$ ω6*c* and/or $C_{18:1}$ ω7*c*; -, Not detectable; TR, trace amount (<1 %).
[b]Data from Zhang *et al*. (2011).
[c]Data from Pascual *et al*. (2012).

Strains VGXO14^T and Vi1 are Gram negative, rod-shaped bacteria (0.6-0.7 µm wide and 1.6-2.0 µm long). Cells are motile, and single polar flagella were observed under transmission electron microscopy (Supplementary Figure 4.S3). The strain VGXO14^T was positive for catalase and oxidase activities. After incubation for 48 h at 30 ^oC on LB plates, colonies were round, convex, beige colour, bright and with entire margins (1-3 mm of diameter). Strain VGXO14^T was

able to grow in liquid LB at 18-37 ^{o}C in 24 h and showed a slow growth at 42 ^{o}C after one week of incubation. No growth was detected at 4^{o}C. Growth was observed on Nutrient Broth in the presence of 2-10% NaCl (w/v) and tolerated pH ranging from 6-10. Strains VGXO14^{T}, Vi1 and all of the other strains in the *P. pertucinogena* group failed to produce either fluorescent pigments or pyocyanin when cultured for 24-48 h at 30 ^{o}C on King B or King A medium.

Differential phenotypic characteristics are indicated in Table 4.8. Strains VGXO14^{T} and Vi1 showed identical patterns of carbon assimilation and enzymatic activity in the API 20NE strips. Although the sand samples were enriched in denitrifying bacteria in the early stages of isolation, neither strain in pure culture was able to reduce nitrate to nitrite or to denitrify. Strains VGXO14^{T} and Vi1 could be differentiated from *P. pachastrellae* by the assimilation of malic acid and trisodium citrate. All type strains in the *P. pertucinogena* group, including strains VGXO14^{T} and Vi1, were negative for indole production, glucose fermentation, arginine dihydrolase, urease, esculine hydrolysis, gelatin hydrolysis, β-galactosidase and assimilation of D-glucose, L-arabinose, D-mannose, D-mannitol, N-acetyl-glucosamine, D-maltose, potassium gluconate and phenylacetic acid. No phenotypic test in the API 20 NE strips was positive for all the strains tested.

In the Biolog GN2 plates, all the type strains, as well as strains VGXO14^{T} and Vi1 were able to oxidise the following substrates: L-alanine, sebacic acid, Tween 40 and Tween 80. A large number of the carbon sources tested were not used. All strains in this study showed a negative or weak assimilation of a high percentage of the substrates tested (75-80%). This limited carbon utilisation differentiates species in the *P. pertucinogena* group from other species in the genus *Pseudomonas* (Kim *et al*., 2009; Romanenko *et al*., 2005). Strains VGXO14^{T} and Vi1 were not able to oxidise L-arabinose, which represents a difference from the rest of strains studied in the group.

Considering the phylogenetic, chemotaxonomic, genotypic and phenotypic characteristics presented, we propose a new species, *Pseudomonas aestusnigri* sp. nov., with *P. aestusnigri* VGXO14^{T} as the type strain.

Table 4.8: Differential phenotypic characteristics of VGXO14^T from related species of the genus *Pseudomonas*. Strains: 1. VGXO14^T; 2. Vi1; 3. *P. pachastrellae* CCUG 46540^T; 4. *P. bauzanensis* DSM 22558^T; 5. *P. litoralis* CECT 7670^T; 6. *P. pertucinogena* LMG 1874^T; 7. *P. sabulinigri* KCTC 22137^T; 8. *P. xiamenensis* JCM 13530^T.

Characteristics	1	2	3	4	5	6	7	8
Range for growth at/on[a]								
Temperature (^oC)	18-37	18-37	7-41	5-30	15-37	ND	4-37	10-45
NaCl (%) (w/v)	2-10	0-8	0-10	0-7	0-15	ND	0-10	0-8
pH	6-10	6-10	ND	7-8	ND	ND	5.5-10	ND
Assimilation of (API 20NE test)								
Capric Acid	+	+	+	-	w	-	+	+
Adipic Acid	+	+	w	+	+	w	+	-
Malic Acid	-	-	+	+	w	w	-	+
Trisodium Citrate	-	-	+	+	-	-	-	+
Carbon sources (Biolog GN2)								
α-Cyclodextrin	-	-	+	-	-	-	-	-
Dextrin	+	-	w	-	+	+	+	+
Glycogen	-	-	-	w	-	w	+	w
L-Arabinose	-	-	+	+	+	+	+	+
α-D-Glucose	-	-	-	w	+	-	-	-
Lactulose	-	-	-	+	-	-	w	-
Methyl Pyruvate	+	+	+	-	+	+	+	+
Mono-Methyl-Succinate	+	-	+	+	+	+	-	+
Acetic Acid	+	-	-	+	+	+	w	+
Cis-Aconitic Acid	-	-	-	-	+	-	-	+
Citric Acid	-	-	-	-	w	+	-	+
D-Galacturonic Acid	-	-	-	-	+	w	-	-
D-Glucuronic Acid	-	-	-	-	+	-	-	-
α-Hydroxy Butyric Acid	-	-	-	-	-	-	-	+
β-Hydroxy Butyric Acid	+	-	+	+	-	+	-	+
Itaconic Acid	-	-	-	+	-	-	-	-
α-Keto Butyric Acid	-	-	-	-	+	-	-	+
α-Keto Glutaric Acid	-	-	+	+	+	-	-	w
α-Keto Valeric Acid	-	-	-	-	+	-	-	w
D,L-Lactic Acid	+	+	+	+	+	-	+	+
Propionic Acid	+	-	-	+	+	+	-	+
Quinic Acid	-	-	-	-	-	-	-	+
Succinic Acid	+	-	+	+	+	+	+	+
Bromo Succinic Acid	+	-	+	+	+	+	-	+
Succinamic Acid	w	-	+	w	w	w	w	w
L-Alaninamide	w	-	-	+	+	+	-	+
D-Alanine	w	-	-	+	+	+	w	+
L-Asparagine	+	-	+	-	-	-	+	-
L-Glutamic Adid	w	-	+	-	+	+	+	-
L-Leucine	-	-	-	-	+	-	-	-
L-Proline	+	-	w	-	+	-	+	+
L-Pyroglutamic Acid	-	-	-	-	-	-	+	-
D-Serine	-	-	-	w	-	+	-	+
L-Serine	-	-	-	-	-	+	-	w
g-Amino Butytic Acid	+	-	-	-	-	+	-	-
Thymidine	-	-	-	+	-	w	-	w
Glucose-6-phosphate	+	-	-	-	-	-	-	-

Unless otherwise indicated all data were obtained in this study. In the API 20 NE test all strains were negative for reduction of nitrate a nitrite and denitrification, indole production, glucose fermentation, arginine dihydrolase, urease, esculine hydrolysis, gelatin hydrolysis, β-galactosidase and assimilation of D-glucose, L-arabinose, D-mannose, D-mannitol, N-acetyl-glucosamine, D-maltose, potassium gluconate and phenylacetic acid. In the oxidase and catalase tests all the strains were positive. In the Biolog GN2 test all strains were positive for the

following characteristics: L-alanine, sebacic acid, Tween 40 and Tween 80; and were negative or weak for the utilization of the other substrates. Positive (+), negative (-), weak (w) and not determined (ND).
[a]Growth at/on different temperatures, % (w/v) [NaCl] range and pH range for other closely related *Pseudomonas* type strains were taken from Romanenko *et al*. (2005), Lai *et al*. (2008), Kim *et al*. (2009), Zhang *et al*. (2011) and Pascual *et al*. (2012).

4.4. Concluding remarks

4.4.1. Putative novel *Pseudomonas* species

The preliminary phenotypical data together with the DDH and concatenated genes analysis indicates that strain S12, 18F, 19F, 35F, 44F and IFO 3738 could be candidates for a new species and strains S25^T, 24F, 58F, 102F, 103F and 122F could be candidates for a two new species. Additional physiological and biochemical characteristics should be included to describe properly these three new species.

4.4.2. Description of *Pseudomonas terricola* sp. nov.

Pseudomonas terricola (ter.ri'co.la. L. n. terra, earth, soil; L. suff. cola, inhabitant, dweller; N.L. n. terricola, soil-dweller)

Cells are Gram negative, rod-shaped (2.2-3.0 μm long and 0.5-0.8 μm wide) and motile by means of one polar inserted flagellum. Cells are catalase-positive, oxidase-positive, oxidative metabolism (aerobic or denitrifiers) and non fermentative. The colonies are round (1-2 mm of diameter), color beige, flat, regular margins and translucent and produce brown-black diffusible pigment when grown on LB plates after incubation for 48h at 30 ^oC. A fluorescent pigment is produced on *Pseudomonas* agar F, but no pyocianin is produced on *Pseudomonas* agar P. Growth occurs at 4-37 ^oC (optimal temperature is 25-30 ^oC), at a pH of 5-10 and at 0-6% (w/v) NaCl.

In API 20NE tests, denitrification and gelatinase are positive. The following tests were negative: indole production, glucose fermentation, arginine dihydrolase, urease and β-galactosidase; and esculin hydrolysis was variable. Glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, potassium gluconate, malate, caprate and trisodium citrate are assimilated; adipate and phenylacetate are not assimilated; and assimilation maltose was variable. The results obtained with Biolog GN2 microplates indicate that the following substrates are utilised: dextrin, Tween 40 , Tween 80, N-acetyl-D-glucosamine, L-arabinose, D-arabitol, D-fructose, D-galactose , α-D-glucose, m-inositol ,D-mannitol ,D-mannose ,D-sorbitol,

sucrose, D-trehalose, methyl pyruvate , mono-methyl-succinate, acetic acid, cis-aconitic acid, citric acid, formic acid, D-galactonic acid lactone, D-galacturonic acid, D-gluconic acid, D-glucosaminic acid, D-glucuronic acid, α-hydroxy butyric acid, β-hydroxy butyric acid, p-hydroxy phenylacetic acid, α-keto glutaric acid, D,L-lactic acid, malonic acid, propionic acid, quinic acid, D-saccharic acid, succinic acid, bromo succinic acid, glucuronamide, L-alaninamide, D-alanine, L-alanine, L-alanyl-glycine, L-asparagine, L-aspartic acid, L-glutamic acid, glycyl-L-glutamic acid, L-histamine, hydroxy-L-proline, L-leucine, L-proline, L-pyroglutamic acid, L-serine, L-threonine, D,L-carnitine, γ-amino butytic acid, urocanic acid, inosine, uridine, putrescine, 2,3-butanediol, glycerol, D,L-α-glycerol phosphate; and variable for the utilization of α-cyclodextrin, D-cellobiose, maltose, γ-hydroxy butyric acid, α-keto butyric acid, α-keto valeric acid, glycyl-L-aspartic acid, L-ornitine and 2-aminoethanol. The other organic substrates included in the Biolog GN2 microplates are not utilised. The predominant fatty acids are summed feature 3 ($C_{16:1}$ *ω7c* and $C_{16:1}$ *ω6c*), $C_{16:0}$, summed feature 8 ($C_{18:1}$ *ω7c* and $C_{18:1}$ *ω6c*), $C_{12:0}$, $C_{12:0}$ 3-OH, $C_{10:0}$ 3-OH and $C_{12:0}$ 2-OH. The type strain S58^T (CECT 8389^T and CCUG 64415^T) was isolated from agricultural soil, Mallorca, Spain.

4.4.3. Description of *Pseudomonas aestusnigri* sp. nov.

Pseudomonas aestusnigri (aestusnigri: a.es.ti.ni'gri. L. n. aestus, tide; L. adj. niger-gra -grum, black; N.L. gen. n. aestusnigri, of black tide).

Cells are Gram negative, non-pigmented, rod-shaped (1.6-2.0 µm long and 0.6-0.7 µm wide) that are mobile by means of one polar flagellum. Cells are catalase-positive, oxidase-positive, strictly aerobic and non fermentative. The colonies are round, convex, color beige, bright and with entire margins (1-3 mm of diameter) on LB plates after incubation for 48h at 30 ^oC. Growth occurs at 18-42 ^oC (optimal temperature is 25-30 ^oC), at a pH of 6-10 and at 2-10% (w/v) NaCl. In API 20NE tests, strains are positive for assimilation of capric acid and adipic acid. In Biolog GN2 Microplates tests, are positive for utilization of Tween 40, Tween 80, pyruvic acid methyl ester, D,L-lactic acid, sebacic acid and L-alanine; and weak, negative or variable for the utilization of the other substrates. The predominant fatty acids are summed feature 8 ($C_{18:1}$*ω6c* and/or $C_{18:1}$ *ω7c*), summed feature 3 ($C_{16:1}$ *ω7c* and/or $C_{16:1}$ *ω6c*), $C_{16:0}$, $C_{12:0}$, $C_{10:0}$ 3-OH and $C_{12:0}$ 3-OH. The type strain VGXO14^T (CECT 8317 and CCUG 64165) was isolated from a crude-oil contaminated intertidal sand, Galicia, Spain.

4.5. Supplementary data

Supplementary Table 4.S1: Strains analyzed in this study.

Isolate	Origin	Isolation date	Collection
10909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
20909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
30909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
40909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
50909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
60909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
70909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
80909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
90909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
100909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
110909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
120909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
130909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
140909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
150909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
160909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
170909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
180909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
190909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
200909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
210909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
220909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
240909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
250909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
260909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
270909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
280909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
290909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
300909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
310909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
320909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
350909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
360909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
370909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
380909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
390909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
400909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
410909	agricultural soil, Mallorca (Spain)	September, 2009	Universitat de les Illes Balears (Spain)
10910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
20910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
30910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
40910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
50910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
60910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
70910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
90910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
100910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
110910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
120910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
130910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)

140910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
150910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
160910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
170910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
180910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
190910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
200910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
210910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
220910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
230910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
250910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
260910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
270910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
290910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
300910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
310910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
320910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
330910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
350910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
360910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
370910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
390910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
400910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
420910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
430910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
440910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
450910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
460910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
470910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
480910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
490910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
500910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
510910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
520910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
530910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
540910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
560910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
570910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
580910	agricultural soil, Mallorca (Spain)	September, 2010	Universitat de les Illes Balears (Spain)
CFBP 4629	*Helianthus annuus*, Charente (France)	1998	CFBP culture collection (France)
IFO3738[a]	-	-	-
18F	Local Orange, Mazandaran (Iran)	2009	Plant Pathology Research Institute (Iran)
19F	Local Orange, Mazandaran (Iran)	2009	Plant Pathology Research Institute (Iran)
24F	Washington Navel, Mazandaran (Iran)	2010	Plant Pathology Research Institute (Iran)
35F	Local Orange, Gilan (Iran)	2010	Plant Pathology Research Institute (Iran)
44F	Sour Orange, Mazandaran (Iran)	2010	Plant Pathology Research Institute (Iran)
58F	Sour Orange, Mazandaran (Iran)	2009	Plant Pathology Research Institute (Iran)
102F	Washington Navel, Mazandaran (Iran)	2010	Plant Pathology Research Institute (Iran)
103F	Sour Orange, Golestan (Iran)	2010	Plant Pathology Research Institute (Iran)
122F	Local Orange, Mazandaran (Iran)	2009	Plant Pathology Research Institute (Iran)
VGXO14	intertidal sand , Lariño (Spain)	2004	Universitat de les Illes Balears (Spain)
Vi1	intertidal sand , Lariño (Spain)	2004	Universitat de les Illes Balears (Spain)

a) Sequences were retrieved from Yamamoto and Harayama (1998).

Supplementary Table 4.S2. Concatenated, 16S rRNA, *gyrB* and *rpoD* genes pairwise sequence similarity between strains S1, S15, S17, S36, S42, S43 and the types strains used in this study.

In extra CD

Supplementary Table 4.S3. Concatenated, 16S rRNA, *gyrB* and *rpoD* genes pairwise sequence similarity between strain S12, 18F, 19F, 35F, 44F, IFO 3738 and the types strains used in this study.

In extra CD

Supplementary Table 4.S4. Concatenated, 16S rRNA, *gyrB* and *rpoD* genes pairwise sequence similarity between strains S25, S30, 24F, 58F, 102F, 103F, 122F and the types strains used in this study.

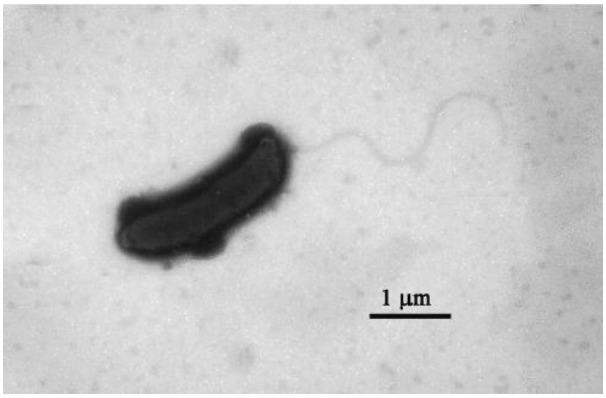
In extra CD

Supplementary Table 4.S5. Concatenated, 16S rRNA, *gyrB*, *rpoD,* and *rpoB* genes pairwise sequence similarity between strains S58, CFBP 4629 and the types strains used in this study.

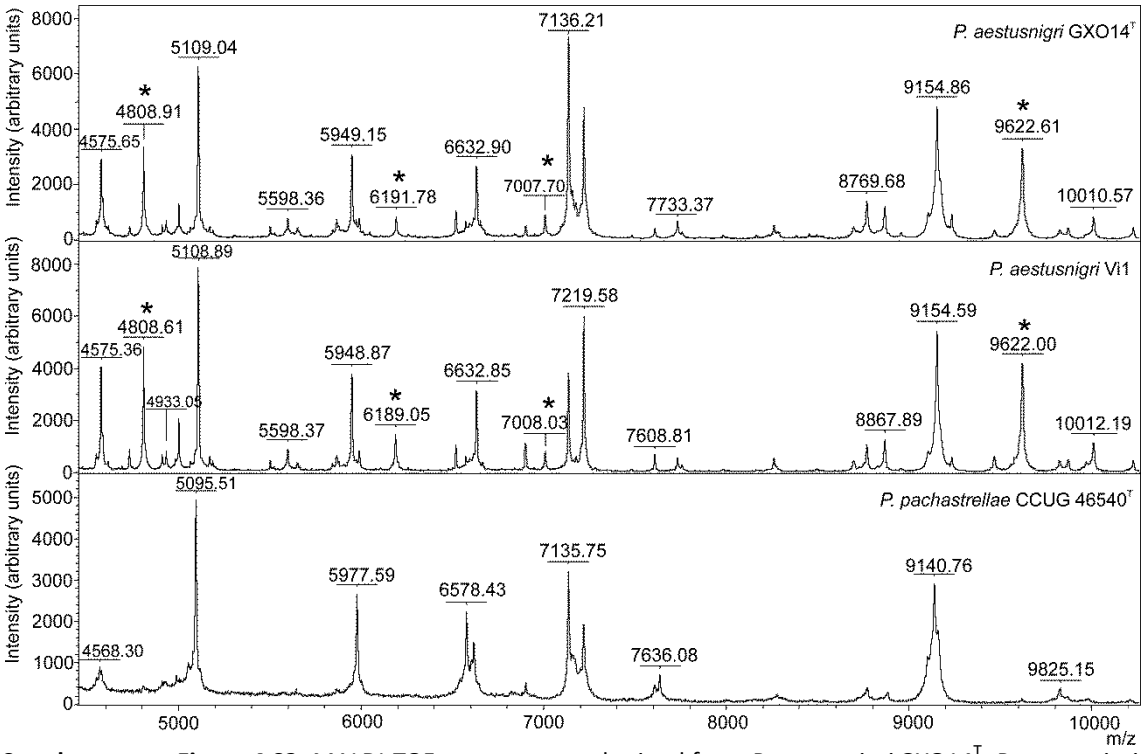
In extra CD

Supplementary Table 4.S6. Concatenated, 16S rRNA, *gyrB* and *rpoD* genes pairwise sequence similarity between strains VGXO14^{T}, Vi1 and the types strains used in this study.

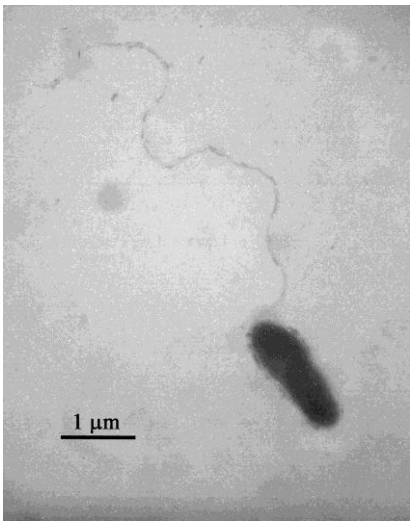
In extra CD



Supplementary Figure 4.S1: Transmission electron microscopy of a negatively stained cell of *P. terricola* S58T. showing polar flagellation as observed in an Hitachi model H600 electron microscope (bar, 1 μm).



Supplementary Figure 4.S2: MALDI-TOF mass spectra obtained from *P. aestusnigri* GXO14^T, *P. aestusnigri* Vi1 and *P. pachastrellae* CCUG 46540^T. *Peaks marked correspond to *P. aestusnigri* strains common peaks (m/z) not present in *P. pachastrellae* CCUG 46540T.



Supplementary Figure 4.S3: Transmission electron microscopy of a negatively stained cell of *P. aestusnigri* VGXO14^T showing polar flagellation as observed in an Hitachi model H600 electron microscope (bar, 1 µm).

Conclusions

Conclusions

1. The study of the population structure and genetic diversity of environmental and clinical *Pseudomonas aeruginosa* revealed a high number of new alleles and new sequence types in a limited area, reflecting the great diversity of *P. aeruginosa* populations and the high plasticity of a phylogenetic conserved species as *P. aeruginosa*.

2. The presence of *Pseudomonas aeruginosa* isolates of the same sequence type, ST-1146, has been demonstrated in clinical and environmental habitats.

3. The comparison of 4 genomes of isolates of ST-1146 showed that the isolate-specific genes in the clinical isolate SD9 were mainly related to phages. SD9 harbored exclusive genes similar to F116/H66 phage and presented the highest number of mutations accumulated in other phage region (phage Pf1-like) which could represent a relevant adaptation advantage.

4. The clinical isolate SD9 showed in both mutational profile analysis in the comparisons with *P. aeruginosa* PAO1 and UCBPP-PA14 strains a number of exclusive alleles higher than the environmental isolates, suggesting that the mutation pressure is not the same in the environmental isolates and in the clinical one.

5. Comparative genomic indicated that the isolates of ST-1146 are closely related, and genes implicated in pathogenicity studied are highly conserved in the environmental isolates.

6. In the assessment of the species diversity of *Pseudomonas* from the river Woluwe, the *Pseudomonas rpoD* selective primers PsEG30F/PsEG790R, were useful to differentiate a high number of species and phylospecies in culture-dependent and independent methods.

7. Pyrosequencing of the *rpoD* environmental amplicon was a powerful tool for DNA-molecular analysis in the study of *Pseudomonas* diversity.

8. In all analysis a high number of putative novel species were found, indicating the enormous diversity not described yet.

9. In the characterization of *Pseudomonas* from agricultural soil, the preliminary phenotypical data, together with the DNA-DNA hybridization data and the multilocus sequence analysis indicated that strains S12, 18F, 19F, 35F, 44F and IFO 3738 could be candidates for a new species. Strain S25, and strains 24F, 58F, 102F, 103F and 122F could be candidates for two more new species. Additional physiological and biochemical characteristics should be included to describe properly these strains as new species.

10. Considering the phylogenetic, chemotaxonomic, genotypic and phenotypic characteristics presented, we propose a new species, *Pseudomonas terricola* sp. nov., with *P. terricola* S58^{T} as the type strain.

11. Considering the phylogenetic, chemotaxonomic, genotypic and phenotypic characteristics presented, we propose a new species, *Pseudomonas aestusnigri* sp. nov., with *P. aestusnigri* VGXO14^{T} as the type strain.

References

References

Ausubel F, Brent R, Kingston R, Moore D, Seidma J, Smith JA *et al*. (1997). Current protocols in molecular biology. J. Wiley & Sons. Inc., New York, vol1.

Adékambi T, Drancourt M, Raoult D (2009). The *rpoB* gene as a tool for clinical microbiologists. *Trends Microbiol* **17**: 37-45.

Allen L, Dockrell DH, Pattery T, Lee DG, Cornelis P, Hellewell PG, Whyte MK (2005). Pyocyanin production by *Pseudomonas aeruginosa* induces neutrophil apoptosis and impairs neutrophil-mediated host defenses in vivo. *J Immunol* **174**: 3643-9.

Anzai Y, Kim H, Park J, Wakabayashi H, Oyaizu H (2000). Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *Int J Syst Evol Microbiol* **50:** 1563-89.

Bennasar A, Mulet M, Lalucat J, García-Valdés E (2010). PseudoMLSA: a database for multigenic sequence analysis of Pseudomonas species. *BMC Microbiol* **10:** 118.

Bridson, E.Y. (2006). *Pseudomonas* Agar Base. The OXOID Manual, pp. 2/2-287/289.

Brint JM, Ohman DE (1995). Synthesis of multiple exoproducts in *Pseudomonas aeruginosa* is under the control of RhlR-RhlI, another set of regulators in strain PAO1 with homology to the autoinducer-responsive LuxR-LuxI family. *J Bacteriol* **177**: 7155–7163.

Cholley P, Thouverez M, Hocquet D, van der Mee-Marquet N, Talon D, Bertrand X (2011). Most multigrug-resistance *Pseudomonas aeruginosa* isolates from hospitals in eastern France belong to a few clonal types. *J Clin Microbiol* **49**: 2578-2583.

Cladera A, Bennasar A, Barceló M, Lalucat J, García-Valdés E (2004). Comparative genetic diversity of *Pseudomonas stutzeri* genomovars, clonal structure, and phylogeny of the species. *J Bacteriol* **186**: 5239-48.

Cladera AM, García-Valdés E, Lalucat J (2006a). Genotype versus phenotype in the circumscription of bacterial species: the case of *Pseudomonas stutzeri* and *Pseudomonas chloritidismutans*. *Archiv Microbiol* **184**: 353-61.

Cladera AM, Sepúlveda-Torres LC, Valens-Vadell M, Meyer J, Lalucat J, García-Valdés E (2006b). A detailed phenotypic and genotypic description of *Pseudomonas* strain OX1. *Syst Appl Microbiol* **29**: 422-30.

Clinical Laboratory Standards Institute (2009). Performance standards for antimicrobial disk susceptibility tests; 10th edition. Approved standard M2-A10. Clinical and Laboratory Standards Institute, Wayne, PA.

Clinical Laboratory Standards Institute (2010). Performance standards for antimicrobial susceptibility testing; 19th informational supplement. CLSI document M100-S20. Clinical and Laboratory Standards Institute, Wayne, PA.

Cramer N, Klockgether J, Wrasman K, Schmidt M, Davenport C, Tümmler B (2011). Microevolution of the major common *Pseudomonas aeruginosa* clones C and PA14 in cystic fibrosis lungs. *Environ Microbiol* **13**: 1590–1604.

Cross A, Allen JR, Burke J, Ducel G, Harris A *et al*. (1983). Nosocomial infections due to *Pseudomonas aeruginosa*: review of recent trends. *Rev. Infect. Dis.* **5 (Suppl.5)**: S837–S845.

Curran B, Jonas D, Grundmann H, Pitt T, Dowson CG (2004). Development of a Multilocus Sequence Typing scheme for the opportunistic pathogen *Pseudomonas aeruginosa. J Clin Microbiol* **42**: 5644–5649.

Darling AE, Mau B, Perna NT (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss, and Rearrangement. *PLoS One*. **5**: e11147.

Denamur E, Picard B, Decoux G, Denis JB, Elion J (1993). The absence of correlation between allozyme and rrn RFLP analysis indicates a high gene flow rate within human clinical *Pseudomonas aeruginosa* isolates. *FEMS Microbiol Lett* **110**: 275–280.

Desai JD, Banat IM (1997). Microbial production of surfactants and their commercial potential. *Microbiol Mol Biol R* **61**: 47–64.

Didelot X, Falush D (2007). Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**: 1251–1266.

Ernst RK, D'Argenio DA, Ichikawa JK, Bangera MG, Selgrade S *et al*. (2003). Genome mosaicism is conserved but not unique in *Pseudomonas aeruginosa* isolates from the airways of young children with cystic fibrosis. *Environ Microbiol* **5**: 1341–1349.

Feil EJ, Li B, Aanensen DM, Hanage WP, Spratt BG (2004). eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* **186**: 1518-1530.

Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA (2012). PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* **13**: 87.

Frapolli M, Défago M, Moënne-Loccoz Y (**2007).** Multilocus sequence analysis of biocontrol fluorescent *Pseudomonas* sp. producing the antifungal compound 2,4-diacetylphloroglucinol. *Environ Microbiol* **9**: 1939-55.

Fu L, Niu B, Zhu Z, Wu S, Li W (2012). CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics* **28**: 3150-3152.

Fuerts JA, Hayward AC (1969). Surface appendages similar to fimbriae (pili) on *Pseudomonas* species. *J Gen Microbiol* **58**: 227-237.

Furukawa S, Kuchma SL, O'Toole GA (2006). Keeping their options open: acute versus persistent infections. *J Bacteriol* 188: 1211–1217.

García-Castillo M, del Campo R, Morosini MI, Riera E, Cabot G, Willems R *et al*. (2011). Wide Dispersion of ST175 Clone despite High Genetic Diversity of Carbapenem-Nonsusceptible *Pseudomonas aeruginosa* clinical strains in 16 spanish hospitals. *J Clin Microbiol* **49**: 2905–2910.

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

Grimberg SJ (1996). Quantifying the biodegradation of phenanthrene by *Pseudomonas stutzeri* P16 in the presence of a nonionic surfactant. *Appl Environ Microb* **62**: 2387–2392.

Gomila M, Gascó J, Busquets A, Gil J, Bernabéu R, Buades JM, Lalucat J (2005). Identification of culturable bacteria present in haemodialysis water and fluid. *FEMS Microbiol Ecol* **52**: 101–114

Gomila M, Gallegos MC, Fernández-Baca V, Pareja A, Pascual M, Díaz-Antolín P *et al*. (2013). Genetic diversity of clinical *Pseudomonas aeruginosa* isolates in a public hospital in Spain. BMC Microbiology 13:138.

Gould WD, Hagedorn C, Bardinelli TR, Zablotowicz RM (1985). New selective media for enumeration and recovery of fluorescent pseudomonads from various habitats. *Appl Environ Microbiol* **49**: 28–32.

Grundmann H, Schneider C, Hartung D, Daschner FD, Pitt TL (1995). Discriminatory power of three DNA-based typing techniques for *Pseudomonas aeruginosa*. *J Clin Microbiol* **33:** 528–534.

Guttman DS, Morgan RL, Wang PW (2008). The Evolution of the Pseudomonads. In: P*seudomonas syringae* Pathovars and Related Pathogens – Identification, Epidemiology and Genomics, pp 307-319.

Haas D, Defago G (2005). Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nat Rev Microbiol* **3**: 307–319.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95-98.

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

Harwood CS, Fosnaugh K, Dispensa M (1989). Flagellation of *Pseudomonas putida* and analysis of its motile behavior. *J Bacteriol* **171**: 4063-66.

Hilario E, Buckley TR, Young JM (2004). Improved resolution on the phylogenetic relationships among *Pseudomonas* by the combined analysis of *atpD*, *carA*, *recA* and 16S rDNA. *Antonie van Leeuwenhoek* **86**: 51-64.

Holland SM (2003). Analytic rarefaction user's manual, version 1.3 (available at http://www.uga.edu/_strata/software/). University of Georgia, GA.

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

Huson DH, Bryant D (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254-267.

Hwang CY, Zhang GI, Kang SH, Kim HJ, Cho BC (2009). *Pseudomonas pelagia* sp. nov., isolated from a culture of the Antarctic green alga *Pyramimonas gelidicola*. *Int J Syst Evol Microbiol* **59**: 3019–3024.

Jakob K, Goss EM, Araki H, Van T, Kreitman M, Bergelson J (2002). *Pseudomonas viridiflava* and *P. syringae* — Natural Pathogens of *Arabidopsis thaliana*. *Mol Plant Microbe Interact* **15**: 1195-203.

Jelsbak L, Johansen HK, Frost AL, Thogersen R, Thomsen LE, Ciofu O *et al*. (2007). Molecular epidemiology and dynamics of pseudomonas aeruginosa populations in lungs of cystic fibrosis patients. *Infect Immun* **75**: 2214–2224.

Johnsen K, Nielsen P (1999). Diversity of *Pseudomonas* strains isolated with King's B and Gould's S1 agar determined by repetitive extragenic palindromic-polymerase chain reaction, 16S rDNA sequencing and Fourier transform infrared spectroscopy characterisation. *FEMS Microbiol Lett* **173**: 155-62.

Johnson JK, Arduino SM, Stine OC, Johnson JA, Harris AD (2007). Multilocus Sequence Typing compared to pulsed-Field gel electrophoresis for molecular typing of *Pseudomonas aeruginosa.* *J Clin Microbiol* **45**: 3707–3712.

Jolley KA, Feil EJ, Chan MS, Maiden MC (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**: 1230-1231.

Kawai Y, Yabuuchi E (1975). *Pseudomonas pertucinogena* sp. nov., an organism previously misidentified as *Bordetella pertussis*. *Int J Syst Bacteriol* **25**: 317–323.

Kersters K, Ludwing W, Vancanneyt M, Devos P, Gillis M, Schleifer K (1996). Recent changes in the classification of the pseudomonads: An overview. *Syst Appl Microbiol* **19:** 465-47.

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

Kidd TJ, Ritchie SR, Ramsay KA, Grimwood K, Bell SC, Rainey PB (2012). *Pseudomonas aeruginosa* exhibits frequent recombination, but only a limited association between genotype and ecological setting. *PLoS ONE* **7**: e44199.

Kiewitz C, Tümmler B (2000). Sequence diversity of *Pseudomonas aeruginosa*: impact on population structure and genome evolution. *J Bacteriol* **182**: 3125–3135.

Kiil K, Binnewies TT, Willenbrock H, Kirkelund Hansen S, Yang L, Jelsbak L *et al*. (2008). Comparative Genomics of Pseudomonas. In: *Pseudomonas*. Model Organism, Pathogen, Cell Factory. Edited by Bernd H.A. Rehm.

Kim KH, Roh SW, Chang HW, Nam YD, Yoon JH, Jeon CO *et al*. (2009). *Pseudomonas sabulinigri* sp. nov., isolated from black beach sand. *Int J Syst Evol Microbiol* **59**: 38–41.

Klockgether J, Würdemann D, Reva O, Wiehlmann L, Tümmler B (2007). Diversity of the abundant pKLC102/PAGI-2 family of genomic islands in *Pseudomonas aeruginosa*. *J Bacteriol* **189**: 2443–2459.

Klockgether J, Munder A, Neugebauer J, Davenport CF, Stanke F, Larbig KD *et al*. (2010). Genome diversity of *Pseudomonas aeruginosa* PAO1 Laboratory Strains. *J Bacteriol* **192**: 1113–1121.

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

Kresse AU, Dinesh SD, Larbig K, Römling U (2003). Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs. *Mol. Microbiol.* **47**: 145–158.

Koeuth T, Versalovic J, Lupski JR (1995). Differential subsequence conservation supports the mosaic nature of interspersed repetitive BOX elements in bacteria. *Genome Res* **5**: 408–418.

Kung VL, Ozer EA, Hauser AR (2010). The Accessory Genome of *Pseudomonas aeruginosa.* Microbiol. Mol. Biol. Rev. p. 621–641.

Lai Q, Shao Z (2008). *Pseudomonas xiamenensis* sp. nov., a denitrifying bacterium isolated from activated sludge. *Int J Syst Evol Microbiol* **58**: 1911–1915.

Lalucat J (1988). Analysis of Refractile (R) Bodies. In: Mayer, E. (Eds.), Methods in Microbiology - Electron Microscopy in Microbiology (vol. 20). Academic Press, London, pp. 79–90.

Lee B, Haagensen JA, Ciofu O, Andersen JB, Hoiby N, Molin S (2005). Heterogeneity of biofilms formed by nonmucoid Pseudomonas aeruginosa isolates from patients with cystic fibrosis. *J Clin Microbiol* **43**: 5247–5255.

Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL, Miyata S *et al*. (2006). Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol* **7**: R90.

Leisinger T, Margraff R (1979). Secondary metabolites of the fluorescent pseudomonads. *Microbiol Mol Biol R* 43: 422–442.

Lenchi N, İnceoğlu Ö, Kebbouche-Gana S, Gana ML, Llirós M, Servais P, García-Armisen T (2013). Diversity of Microbial Communities in Production and Injection Waters of Algerian Oilfields Revealed by 16S rRNA Gene Amplicon 454 Pyrosequencing. *PLoS ONE* **8**: e66588.

Li L, Abu Al-Soud W, Bergmark L, Riber L, Hansen LH, Magid J, Sørensen SJ (2013). Investigating the diversity of *Pseudomonas* spp. in soil using culture dependent and independent techniques. *Curr Microbiol* **67:** 423-30.

Li W, Godzik A (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658-9.

Librado P, Rozas J (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451-1452.

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

Livermore DM (2002). Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa*: our worst nightmare? *Clin Infect Dis* **34**: 634–640.

Lomholt JA, Poulsen K, Kilian M (2001). Epidemic population structure of *Pseudomonas aeruginosa*: evidence for a clone that is pathogenic to the eye and that has a distinct combination of virulence factors. *Infect Immun* **69**: 6284–6295.

Lyczak JB, Cannon CL, Pier GB (2000). Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect* **2**: 1051–1060.

Maatallah M, Cheriaa J, Backhrouf A, Iversen A, Grundmann H, Do T *et al*. (2011). Population structure of *Pseudomonas aeruginosa* from five Mediterranean countries: evidence for frequent recombination and epidemic occurrence of CC235. *PLoS ONE* 6: e25617.

Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG *et al*. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect* **18**: 268–281.

Marmur J (1961). A procedure for the isolation of deoxyribonucleic acid from microorganisms. *J Mol Biol* **3**: 208–218.

Martínez JL (2013). Bacterial pathogens: from natural ecosystems to human hosts. *Environ Microbiol* **15**: 325–333.

Matthijs S, Coorevits A, Gebrekidan TT, Tricot C, Wauven CV, Pirnay JP, De Vos P, Cornelis P (2012). Evaluation of oprI and oprL genes as molecular markers for the genus Pseudomonas and their use in studying the biodiversity of a small Belgian River. *Res Microbiol* **164**: 254-61.

Mavrodi DV, Blankenfeldt W, Thomashow LS (2006). Phenazine compounds in fluorescent *Pseudomonas* spp. biosynthesis and regulation. *Annu Rev Phytopathol* **44**: 417-45.

MIDI (2008). Sherlock Microbial Identification System Operating Manual, version 6.1. Newark, DE: MIDI Inc.

Migula W (1895). *Bacteriaceae* (Stabchenbacterien). In: *Die Naturlichen Pfanzenfamilien*, pp. 20-30. Edited by A Engler and K Prantil. Leipzig, W Engelmann, Leipzig, Teil I, Abteilung Ia.

Mikkelsen H, McMullan R, Filloux A (2011). The *Pseudomonas aeruginosa* Reference Strain PA14 Displays Increased Virulence Due to a Mutation in ladS. *PLoS ONE* **6**: e29113.

Miller, JM (1972). Experiments in molecular genetics. New York: Cold spring Harbour.

Mira A, Martín-Cuadrado AB, D'Auria G, Rodríguez-Valera F (2010). The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol* **13**: 45-57.

Moore ERB, Mua M, Arnscheidt A, Böttger EC, Hutson RA, Collins MD et al. (1996). The determination and comparison of the 16S rDNA gene sequences of species of the genus *Pseudomonas* (sensu stricto) and estimation of the natural intrageneric relationships. *Syst Appl Microbiol* **19**: 476-92.

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

Mulet M, Bennasar A, Lalucat J, García-Valdés E (2009). An *rpoD*-based PCR procedure for the identification of *Pseudomonas* species and for their detection in environmental samples. *Mol Cell Probes* **23**: 140-147.

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

Mulet M, David Z, Nogales B, Bosch R, Lalucat J, García-Valdés E (2011). *Pseudomonas* diversity in crude-oil-contaminated intertidal sand samples obtained after the *Prestige* oil spill. *Appl Environ Microbiol* **77**: 1076–1085.

Mulet M, Gomila M, Scotta C, Sánchez D, Lalucat J, García-Valdés E (2012). Concordance between whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry and multilocus sequence analysis approaches in species discrimination within the genus *Pseudomonas*. *Syst Appl Microbiol* **35**: 455-64.

Mulet M, García-Valdés E, Lalucat J (2013). Phylogenetic affiliation of *Pseudomonas putida* biovar A and B strains. *Res Microbiol* **164**: 351-359.

Ocampo-Sosa AA, Cabot G, Rodríguez C, Roman E, Tubau F, Macia MD *et al*. (2012). Alterations of OprD in carbapenem-intermediate and -susceptible strains of *Pseudomonas aeruginosa* isolated from patients with bacteremia in a Spanish multicenter study. *Antimicrob Agents Chemother* **56**: 1703-13

Özen AI, Ussery DW (2012). Defining the *Pseudomonas* genus: where do we draw the line with *Azotobacter*? *Microb Ecol* **63**: 239-48.

Palleroni NJ (**1984).** Genus I *Pseudomonas* Migula 1894. In: *Bergey's Manual of Systematic Bacteriology*, vol. 2, pp. 323-79. Edited by NR Krieg and JG Holt. The Williams & Wilkins Co, Baltimore. Maryland, USA.

Pascual J, Lucena T, Ruvira MA, Giordano A, Gambacorta A, Garay E *et al*. (2012). *Pseudomonas litoralis* sp. nov., isolated from Mediterranean seawater. *Int J Syst Evol Microbiol* **62**: 438–444.

Pearson JP, Pesci EC, Iglewski BH (1997). Roles of Pseudomonas aeruginosa las and rhl quorum-sensing systems in control of elastase and rhamnolipid biosynthesis genes. *J Bacteriol* **179**: 5756–5767.

Péchy-Tarr M, Bruck DJ, Maurhofer M, Fischer E, Vogne C, Henkels MD *et al*. (2008). Molecular analysis of a novel gene cluster encoding an insect toxin in plant-associated strains of *Pseudomonas fluorescens*. *Environ Microbiol* **10**: 2368–2386.

Pinto AJ, Raskin L (2012). PCR Biases Distort Bacterial and Archaeal Community Structure in Pyrosequencing Datasets. *PLoS ONE* **7**: e43093.

Pirnay JP, Matthijs S, Colak H, Chablain P, Bilocq F, Van Eldere J *et al*. (2005). Global *Pseudomonas aeruginosa* biodiversity as reflected in a Belgian river. *Environ Microbiol* **7**: 969–980.

Pirnay JP, Bilocq F, Pot B, Cornelis P, Zizi M, van Eldere J *et al*. (2009). *Pseudomonas aeruginosa* population structure revisited. *PLoS ONE* **4**: e7740.

Pontes DS, Lima-Bittencourt CI, Chartone-Souza E, Amaral Nascimento AM (2007). Molecular approaches: advantages and artifacts in assessing bacterial diversity. *J Ind Microbiol Biotechnol* **34**: 463-473.

Poole K (2001). Multidrug efflux pumps and antimicrobial resistance in *Pseudomonas aeruginosa* and related organisms. *J Mol Microbiol Biotechnol* **3**: 255–264.

Porter KG, Feig YS (1980). The use of DAPI for identifying and counting aquatic microflora. *Limnol Oceanogr* **25**: 943–948.

Poulsen PHB, Al-Soud WA, Bergmark L, Magid J, Hansen LH, Sørensen SJ (2013). Effects of fertilization with urban and agricultural organic wastes in a field trial—Prokaryotic diversity investigated by pyrosequencing. *Soil Biol Biochem* **57**: 784–793.

Ramos JL (2004). Pseudomonas Volume 1: Genomics, Life Style and Molecular Architecture. New York: Kluwer Academics/Plenum Publishers.

Richter M, Rosselló-Móra R (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* **106**: 19126–19131.

Ridgway HF, Safarik J, Phipps D, Carl P & Clark D (1990). Identification and catabolic activity of well-derived gasoline degrading bacteria from a contaminated aquifer. *Appl Environ Microb* **56**: 3565–3575.

Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD *et al*. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1**: 283–290.

Romanenko LA, Uchino M, Falsen E, Frolova GM, Zhukova NV, Mikhailov VV (2005). *Pseudomonas pachastrellae* sp. nov., isolated from a marine sponge. *Int J Syst Evol Microbiol* **55**: 919–924.

Römling U, Greipel J, Tümmler B (1995). Gradient of genomic diversity in the *Pseudomonas aeruginosa* chromosome. *Mol Microbiol* **17**: 323–332.

Römling U, Schmidt KD, Tümmler B (1997). Large genome rearrangements discovered by the detailed analysis of 21 *Pseudomonas aeruginosa* clone C isolates found in environment and disease habitats. *J. Mol.Biol.* 271, 386–404.

Ryan RP, Germaine K, Franks A, Ryan DJ, Dowling DN (2008). Bacterial endophytes: recent developments and applications. *FEMS Microbiol Lett* **278**: 1–9.

Santos SR, Ochman H (2004). Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ Microbiol* **6**: 754-9.

Sasser M (1990). Identification of bacteria by gas chromatography of cellular fatty acids, MIDI Technical Note 101. Newark, DE: MIDI Inc.

Sato H, Frank DW (2004). ExoU is a potent intracellular phospholipase. *Mol Microbiol* **53:** 1279-1290.

Schmidt KD, Tummler B, Romling U (1996). Comparative genome mapping of *Pseudomonas aeruginosa* PAO with *P. aeruginosa* C, which belongs to a major clone in cystic fibrosis patients and aquatic habitats. *J. Bacteriol* **178**: 85–93.

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

Scotta C, Gomila M, Mulet M, Lalucat J, García-Valdés E (2013). Whole-Cell MALDI-TOF Mass Spectrometry and Multilocus Sequence Analysis in the Discrimination of *Pseudomonas stutzeri* Populations: Three Novel Genomovars. *Microb Ecol* In press.

Silby MW, Winstanley C, Godfrey SAC, Levy SB, Jackson RW (2011). Pseudomonas genomes: diverse and adaptable. *FEMS Microbiol Rev* **35**: 652–680

Sikorski J, Stackebrandt E, Wackernagel W (2001). *Pseudomonas kilonensis* sp. nov., a bacterium isolated from agricultural soil. *Int J Syst Evol Microbiol* **51**: 1549-1555.

Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, *et al*. (2006). Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci USA* 103: 8487–8492.

Smith EE, Sims EH, Spencer DH, Kaul R, Olson MV (2005). Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*. *J Bacteriol* **187**: 2138–2147.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006). Microbial diversity in the deep sea and the underexplored ‘‘rare biosphere’’. *Proc Natl Acad Sci USA* **103**: 12115–12120.

Speert DP (2002). Molecular epidemiology of *Pseudomonas aeruginosa*. *Frontiers Biosci* **7**: e354–361.

Spencer DH, Kas A, Smith EE, Raymond CK, Sims EH, Hastings M *et al*. (2003). Whole-genome sequence variation among multiple isolates of *Pseudomonas aeruginosa*. *J Bacteriol* **185**: 1316–1325.

Stanier RY (1966). In: Stanier RY, Palleroni NJ and Doudoroff M. The aerobic pseudomonads: a taxonomic study. *J Gen Microbiol* **43**: 159-271.

Stavrinides J, McCloskey JK, Ochman H (2009). Pea aphid as both host and vector for the phytopathogenic bacterium *Pseudomonas syringae*. *Appl Environ Microb* **75**: 2230–2235.

Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ *et al.* (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**: 959–964.

Talbot GH, Bradley J, Edwards JE Jr., Gilbert D, Scheld M, Bartlett JG (2006). Bad bugs need drugs: an update on the development pipeline from the Antimicrobial Availability Task Force of the Infectious Diseases Society of America. *Clin Infect Dis* **42**: 657–668.

Tayeb L, Ageron E, Grimont F, Grimont PA (2005). Molecular phylogeny of the genus *Pseudomonas* based on *rpoB* sequences and application for the identification of isolates. *Res Microbiol* **156**: 763-73.

Tayeb LA, Lefevre M, Passet V, Diancourt L, Brisse S, Grimont PAD (2008). Comparative phylogenies of *Burkholderia*, *Ralstonia*, *Comamonas*, *Brevundimonas* and related organisms derived from *rpoB*, *gyrB* and *rrs* gene sequences. *Res Microbiol* **159**, 169-77.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL *et al*. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci USA* **102:** 13950-13955.

Toro M, Ramírez-Bahena MH, Cuesta MJ, Velázquez E, Peix A (2013). *Pseudomonas guariconensis* sp. nov., isolated from rhizospheric soil in Guarico state, Venezuela. *Int J Syst Evol Microbiol* In press.

Van Alst NE, Picardo KF, Iglewski BH, Haidaris CG (2007). Nitrate sensing and metabolism modulate motility, biofilm formation, and virulence in Pseudomonas aeruginosa. Infect Immun, 75, 3780–3790.

Versalovic J, Koeuth T, Lupski JR (1991). Distribution of DNA repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res* **19**: 6823-6831.

Viedma E, Juan C, Acosta J, Zamorano L, Otero JR, Sanz F *et al*. (2009). Nosocomial Spread of Colistin-Only-Sensitive Sequence Type 235 *Pseudomonas aeruginosa* isolates producing the extended-spectrum β-lactamases GES-1 and GES-5 in Spain. *Antimicrob Agents Chemother* **53**: 4930–4933.

Vodovar N, Vinals M, Liehl P, Basset A, Degrouard J, Spellman P *et al*. (2005). Drosophila host defense after oral infection by an entomopathogenic *Pseudomonas* species. *P Natl Acad Sci USA* **102**: 11414–11419.

Wackett LP (2003). *Pseudomonas putida*—a versatile biocatalyst. *Nat Biotechnol* **21**: 136–138.

Webb JS, Lau M, Kjelleberg S (2004). Bacteriophage and Phenotypic Variation in *Pseudomonas aeruginosa* Biofilm Development. *J Bacteriol* **186**: 8066–8073.

Widmer F, Seidler RJ, Gillevet PM, Watrud LS, Di Giovanni GD (1998). A highly selective PCR protocol for detecting 16S rRNA genes of the genus *Pseudomonas* (*sensu stricto*) in environmental samples. *Appl Environ Microbiol* **64**: 2545-53.

Wiehlmann L, Wagner G, Cramer N, Siebert B, Gudowius P, Morales G *et al.* (2007). Population structure of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* **104**: 8101–8106.

Willenbrock H, Friis C, Friis AS, Ussery DW (2006). An environmental signature for 323 microbial genomes based on codon adaptation indices. *Genome Biol* **7**: R114.

Willenbrock H, Ussery DW (2007). Prediction of highly expressed genes in microbes based on chromatin accessibility. *BMC Mol Biol* **8**: 11.

Winsor GL, Lam DK, Fleming L, Lo R, Whiteside MD, Yu NY, Hancock RE, Brinkman FS (2011). *Pseudomonas* Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. *Nucleic Acids Res* **39**: D596-600.

Winstanley C, Langille MGI, Fothergill JL *et al*. (2009). Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. *Genome Research* **19**: 12-23.

Whited GM, Gibson DT (1991). Separation and partial characterization of the enzymes of the toluene-4-monooxygenase catabolic pathway in *Pseudomonas mendocina* KR1. *J Bacteriol* **173**: 3017–3020.

Whiteley M, Bangera MG, Bumgarner RE, Parsek MR, Teitzel GM, Lory S, Greenberg EP (2001). Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* **413:** 860-4**.**

Wolfgang MC, Kulasekara BR, Liang X, Boyd D, Wu K, Yang Q, *et al*. (2003). Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa. Proc Natl Acad Sci USA* **100**: 8484–8489.

Wong BT, Lee DJ (2003). *Pseudomonas yangmingensis* sp. nov., an alkaliphilic denitrifying species isolated from a hot spring. *J Biosci Bioeng* In press.

Yamamoto S, Harayama S (1998). Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes. *Int J Syst Bacteriol* **48**: 813-9.

Yamamoto S, Kasai H, Arnold DL, Jackson RW, Vivian A, Harayama S (2000). Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiology* **146**: 2385-2394.

Zhang DC, Liu HC, Zhou YG, Schinner F, Margesin R (2011). *Pseudomonas bauzanensis* sp. nov., isolated from soil. *Int J Syst Evol Microbiol* **61**: 2333–2337.

Ziemke F, Höfle MG, Lalucat J, Rosselló-Mora R (1998). Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica* sp. nov. *Int J Syst. Bacteriol* **48**: 179–186.

Zhu W, Lomsadze A, Borodovsky M (2010). *Ab initio* gene identification in metagenomic sequences. *Nucleic Acids Research* **38**: e132.