



Universitat
de les Illes Balears

Títol: ANÁLISIS FILOGENÓMICO DE LA ESPECIE *P. stutzeri*

NOM AUTOR: *GUILLEM SEGUÍ CRESPI*

Memòria del Treball de Fi de Màster

Màster Universitari de Microbiologia Avançada
(Especialitat/Itinerari Recerca en Microbiologia Ambiental i Biotecnologia)

de la

UNIVERSITAT DE LES ILLES BALEARS

Curs Acadèmic 2015-2016

Data 06/07/2016

Nom Tutor del Treball: Margarita Gomila Ribas

Resumen

Dentro del género *Pseudomonas*, la especie *P. stutzeri* está ampliamente distribuida, destacando por su alta diversidad génica y metabólica. En este trabajo se realizará un estudio filogenómico combinando el análisis multigénico (MLSA) basado en el análisis de los genes 16S rRNA, *gyrB*, *rpoD* y los 40 genes ortólogos en copia única definidos en la plataforma online SpecI, con el análisis genómico basado en los índices de comparación genómica *in silico* (ANIb, ANIm, TETRA) de una colección de 77 genomas afiliados a las distintas genomovares de la especie *P. stutzeri* o a las distintas especies que conforman el grupo de *P. stutzeri*. Los objetivos del presente trabajo son: (1) Deducir las afiliaciones filogenéticas y taxonómicas de todos los genomas del grupo *P. stutzeri* disponibles hasta Febrero del 2016, (2) Realizar una comparación entre el análisis multigénico y los métodos basados en el análisis de genomas para esclarecer la afiliación filogenética de especies y/o genomovares del grupo; y (3) Comparar los métodos genómicos entre ellos para su uso en la delineación de especies y/o genomovares. Se observó que los valores de ANIb y ANIm del 95 % y de MLSA del 97 % actuaban como puntos de corte permitiendo discernir las diferentes genomovares, mientras que valores superiores agrupaban los miembros de una misma genomovar. Los resultados obtenidos permitieron confirmar las afiliaciones anteriormente obtenidas, asignar cepas a nuevas genomovares y otras a genomovares previamente descritas y por último, detectar asignaciones taxonómicas erróneas, como por ejemplo, la errónea asignación del genoma *P. stutzeri* YC-YH1 a la especie *P. stutzeri* en lugar de a *P. balearica*. La clasificación en 22 genomovares actual de la especie *P. stutzeri*, podría llevar en un futuro, a la asignación de éstas en diferentes especies con estatus taxonómico propio, pendientes de análisis quimiotaxonómicos y fenotípicos más detallados.

Abstract

Inside of *Pseudomonas* genus, the specie *Pseudomonas stutzeri* is widely distributed, sticking out by its highly genetic and metabolic diversity. In this current work will be developed a phylogenomic study combining a multigenic analysis (MLSA) based on analysis of the genes 16S rRNA, *gyrB*, *rpoD* and the 40 universal, single-copy phylogenetic marker genes described in the online platform SpecI, with the *in silico* genomic analysis focused on the genomic comparative ratings (ANIb, ANIm, TETRA) from the collection of 77 genomes affiliated to different genomovars of *P. stutzeri* species or to a distinct species belonging to the *P. stutzeri* group. The main goals of this present study are: (1) to deduct the phylogenetic and taxonomic affiliations of all the genomes of the group available before February, 2016, (2) to perform a comparative between the multigenic analysis and the methods based on the genome analysis to clarify the phylogenetic affiliation of species and/or genomovars group; and (3) to compare the genomic methods among them for be used in the delineation of the species and/or genomovars. It is noticed that values of 95 % of ANIb and ANIm, and 97 % of MLSA works as a cut-off allowing us to discern the different genomovars, meanwhile at higher values the members of the same genomovar were grouped. The results confirm the previously obtained affiliation, to assign strains to news genomovars and to the current described genomovars, and lastly, to detect wrong affiliations, such us, the misallocation of *P. stutzeri* YC-YH1 genome, assigned to *P. stutzeri* instead of *P. balearica* species. The current classification in 22 genomovars of *P. stutzeri*, could lead in the future allocation of these at distinct species with their own taxonomic status, waiting for thorough chemotaxonomic and phenotypic analysis in the near future.

Introducción

El género *Pseudomonas* actualmente está constituido por más de 144 especies válidamente descritas, incluyendo 10 subespecies, según la *List of Prokaryotic Names with Standing in Nomenclature* (Parte *et al.*, 2014). Este género perteneciente a las bacterias Gram negativas es el más numeroso en especies. Aumentando el número de éstas cada año. Desde la primera descripción de las bacterias del género *Pseudomonas* la taxonomía evolucionó a la par que las metodologías disponibles hasta el momento. Migula en el año 1894 describió el género *Pseudomonas* en base a las características morfológicas de los miembros de éste (Migula *et al.*, 1990). Durante muchos años este género incluyó especies no bien caracterizadas, la identificación de especies mantuvo errores hasta que las bases taxonómicas fueron claramente delimitadas, los fundamentos de estas bases eran las propiedades fisiológicas y bioquímicas (Stanier *et al.*, 1996). El género se revisó en 1894, y se estableció la subdivisión en 5 grupos justificándose en los resultados de hibridaciones DNA-DNA y rRNA-DNA (Palleroni, 1984). Más tarde, los cinco grupos fueron asociados a la clase *Proteobacteria* (De Vos and De Ley, 1983; De Vos *et al.*, 1985, 1989; De Ley, 1992); los miembros del género *Pseudomonas* “sensu stricto” pertenecían al grupo I RNA-DNA de la subclase *Gammaproteobacteria*. Desde entonces, varios autores han revisado el estado taxonómico del género (Moore *et al.*, 1996; Anzai *et al.*, 2000; Peix *et al.*, 2009; Mulet *et al.*, 2010, 2012a).

Dentro de este género destaca la especie, *Pseudomonas stutzeri* descrita por Burry y Stutzer en 1895, como un miembro del género *Pseudomonas* “sensu stricto”, perteneciente al filo *Proteobacteria* de la clase *Gammaproteobacteria* (Bennasar *et al.*, 1996). Estas bacterias Gram negativas constituyen el grupo más diverso y con un significado ecológico remarcable, en comparación con los grupos bacterianos restantes. Se caracterizan por tener forma bacilar y un flagelo polar, aunque excepcionalmente y bajo ciertas condiciones se podrían observar uno o dos flagelos laterales. Son bacterias catalasa y oxidasa positivas, con un metabolismo estrictamente respiratorio (Lalucat *et al.*, 2006). Además, juegan un rol muy importante en los ciclos biogeoquímicos del carbono y del nitrógeno, dadas sus particulares características metabólicas como desnitrificantes o la capacidad de utilizar almidón o maltosa como única fuente de carbono (Cladera *et al.*, 2004). No obstante, la hidrólisis de glucógeno o la arginina deshidrolasa dan negativas. Los diferentes microorganismos que forman el grupo de *P. stutzeri* se diferencian de las otras especies por no producir pigmentos fluorescentes. Los miembros de este grupo muestran una gran heterogeneidad fenotípica (Lalucat *et al.*, 2006), y

coinciden en la habilidad de crecer en medios con amonio o nitrato, no necesitan de factores de crecimiento y el oxígeno es el aceptor final de electrones en su metabolismo respiratorio. Aun así, todas las cepas pueden usar el nitrato como alternativa al aceptor de electrones. Además, tienen la capacidad de degradación de componentes aromáticos mediante mono- y di-oxigenasas. Al limitarnos a una identificación fenotípica de este grupo, se facilita la confusión con otras especies del género *Pseudomonas* (*P. putida*, *P. mendocina*) o incluso, con otros géneros (*Alcaligenes*, *Achromobacter*).

Estos microorganismos han sido aislados en diferentes hábitats desde suelos hasta sedimentos marinos, incluyendo muestras de la columna de agua. También se han encontrado algunas cepas en muestras clínicas, heces o aguas embotelladas (Rosello *et al.*, 1991). Con la habilidad de degradar xenobióticos o el potencial desnitrificante, como la fijación de nitrógeno, confieren a las *P. stutzeri* una importancia ecológica. Destaca por una elevada distribución geográfica o hábitats ocupados, que garantizan una gran versatilidad metabólica adaptativa (Scotta *et al.*, 2013).

La especie más cercana a *P. stutzeri* es *P. balearica* (antiguo miembro de la genomovar 6), comparte muchos rasgos fenotípicos básicos y pertenece a la misma rama filogenética en base al gen 16S rRNA. Se diferencia quimiotaxonómicamente por la habilidad de crecer a 42°C, aparte de otras pruebas bioquímicas.

Todas las cepas adscritas a la especie *P. stutzeri* forman un grupo con un origen monofilético, compartiendo ciertas características fenotípicas entre ellas que difieren del resto de especies. Aun así, varios grupos intraespecíficos pueden ser delineados genómicamente y filogenéticamente. En otros estudios taxonómicos polifásicos los grupos fenotípicamente similares, pero genotípicamente diferentes se han definido como genoespecies, genomoespecies o especies genómicas. Entendiéndose como tales especies que se pueden diferenciar mediante comparaciones de ácidos nucleicos (Lalucat *et al.*, 2006). El concepto de genomovar apareció por primera vez en 1991 por Roselló y colaboradores, para clarificar el estatus taxonómico de los subgrupos que se observaban dentro de la especie *P. stutzeri*. Es una aproximación para clasificar cepas individuales que destacan por ser genómicamente diferentes y tienen características fenotípicas relacionadas con las cepas más próximas (Roselló *et al.*, 1991). En el artículo de Roselló y colaboradores en el año 1991 se describieron grupos genómicos a partir de los resultados obtenidos de la hibridación DNA-DNA. Algunos de estos grupos descritos sólo se diferenciaban por el contenido en G+C.

En este artículo sugirieron y acuñaron la terminología de genomovar (gv). El término genomovar incluiría grupos genómicos “nomenespecies” a la espera de obtener un estatus propio. Estos autores sugirieron que la cepa de referencia para cada genomovar debería estar depositada en una colección de cultivo, y que la cepa de referencia de la primera genomovar era la cepa tipo de la especie *P. stutzeri*.

Atendiendo a este criterio, inicialmente se describieron 7 genomovares (gv) basándose en la hibridación DNA-DNA (gv 1, CCUG 11256^T; gv 2, ATCC 17588; gv 3, ATCC 11607; gv 4, 19SMN4; gv 5, DNSP21; gv 6, DSM 6083; y gv7, ATCC 17832) (Roselló *et al.*, 1991). Una de estas genomovares, la gv 6, con capacidad de degradar 2-metilnaftaleno, posteriormente se propuso como una especie diferente con entidad propia, a la que se denominó *P. balearica* (Bennasar *et al.*, 1996). En el 2003 García-Valdés y colaboradores, describieron 3 genomovares más (gv 8, JM300; gv 9, KC; gv 10, CLN100) combinando análisis filogenéticos y el análisis multigénico. El análisis multigénico (*Multilocus Sequence Anaylsis*, MLSA), basado en el análisis de siete genes esenciales (DNA ribosómico 16S, la región interespaciadora 1 (ITS-1), y los genes *gyrB*, codifica para la topoisomerasa II; *rpoD*, codifica para el factor sigma 70 de la RNA polimerasa; *nosZ*, gen codificante para una enzima oxido-reductasa de nitrógeno; *catA*, codifica para la catecol-1,2-dioxigenasa; y *nahH*, que codifica para la catecol-2,3-dioxigenasa), también confirmó el origen monofilético del grupo. En el estudio de Sikorski *et al.*, del año 2005 se definieron las genomovares 11-18 (gv 11, 28a50; gv 12, 28a39; gv 13, 28a22; gv 14, 28a3; gv 15, 4c29; gv 16, 24a13; gv 17, 24a75; gv 18, MT-1) formadas por aislados del suelo y sedimento marino, mediante la utilización de experimentos de hibridación DNA-DNA (con resultados inferiores al 70%), y la secuenciación de los genes ITS-1 y 16S rRNA. La genomovar 19 con la cepa de referencia CCUG 46542 fue determinada en base a los análisis filogenéticos de los genes 16S rRNA, *gyrB*, *rpoD* y el siderotipado (Mulet *et al.*, 2008). Las últimas genomovares asignadas al grupo de *P. stutzeri* (gv 20-22 asignadas con un solo representante, PE, A563/77, V81, respectivamente), determinación mediante MLSA y por espectroscopia de masas de células enteras por MALDI-TOF (*Whole-Cell Matrix- Assisted Laser-Desorption/ionisation Time-of-Flight Mass Spectrometry*, WC-MALDI-TOF MS) (Scotta *et al.*, 2013). Este último estudio mostró la eficacia del MLSA para discernir entre las genomovares y subrayó su papel como herramienta de estudios filogenéticos. También se obtuvo una buena correlación entre WC-MALDI-TOF MS y MLSA, demostrando ser ambas buenas herramientas para las aproximaciones filogenéticas.

El estudio taxonómico, ha evolucionado en estos últimos años simultáneamente con las metodologías desde su primera descripción, mejorando las aproximaciones, eficiencia y abaratando costes y tiempo. Empezó con la hibridación DNA-DNA (DDH) y la cuantificación del contenido GC, siguiéndole el análisis de secuencias del gen 16S rRNA, y en los últimos años, el análisis multigénico basado en otros genes como marcadores moleculares como *gyrB*, *rpoB*, *recA* o *rpoD*. En el estudio de Mulet y colaboradores se demostró la facilidad en determinar la filogenia, clarificación del género e identificación de nuevas cepas basado en el análisis de las secuencias de cuatro genes esenciales (16S rRNA, *gyrB*, *rpoB*, and *rpoD*) de todas las especies (Mulet *et al.*, 2012a). En posteriores estudios se realizaron aproximaciones basándose en espectroscopia de masas por WC-MALDI-TOF, mostrándose la facilidad en la identificación de especie o incluso genomovar. En los últimos años la aparición de las tecnologías de secuenciación masiva ha revolucionado los estudios taxonómicos. La secuenciación de los genomas proporcionaría una herramienta importante en la taxonomía bacteriana como un método de comparación.

Las herramientas de análisis *in silico*, estrechamente relacionadas con la información de los genomas, ya sean completos o *drafts*, puede proporcionar información sumamente valiosa para establecer relaciones evolutivas y taxonómicas en la bacteriología. En 2005 Coenye *et al.*, publicaron el artículo “*Toward a prokaryotic genomic taxonomy*” donde mostraron los enfoques disponibles para llevar a cabo estudios taxonómicos entre especies de procariontas, basándose en la secuenciación de genomas. Estos métodos genómicos diseñados para sustituir la hibridación DNA-DNA experimental, han sido altamente discutidos y probados por la sociedad científica. De hecho, Gomila y colaboradores, probaron estos métodos, la frecuencia tetranucleotídica (TETRA), la identidad nucleotídica promedio (*Average Nucleotide Identity*, ANI) y el cálculo del *Genome-to-Genome distance* (GGDC) en el género *Pseudomonas*, mostrando que algunas de las clasificaciones realizadas eran erróneas y su correcta clasificación dudosa, siendo necesario en algunos casos una revisión de la afiliación de las especies al género (Gomila *et al.*, 2015).

El objetivo de este trabajo ha sido validar el uso de estas herramientas de análisis *in silico* en la especie *P. stutzeri*. Específicamente se pretende: (1) Inferir la afiliación filogenética y taxonómica de los genomas secuenciados de la especie *P. stutzeri*, (2) Comparar el análisis multigénico con los métodos basados en el análisis del genoma para la delineación de especies y genomovares en el grupo de *P. stutzeri*, y (3) Comparar los métodos genómicos

entre ellos para la delimitación de especies y/o genomovares. Para ello se analizaron todos los genomas de cepas de *P. stutzeri* disponibles en nuestro laboratorio, así como todos los genomas afiliados a la especie *P. stutzeri*, disponibles en la base de datos NCBI hasta Febrero del 2016.

En el estudio se ha incluido al menos un miembro de todas las genomovares de *P. stutzeri*, siendo en la mayoría de los casos la cepa de referencia de la genomovar, así como distintas cepas de la especie *P. balearica* (antigua gv 6). También se ha incluido en el análisis todos los genomas disponibles de las especies que forman el grupo de *P. stutzeri*, *P. azotifigens*, *P. chloritidismutans*, *P. kunmingensis*, *P. kuykendallii*, *P. songnenensis*, *P. zhaodongensis*, *P. xanthomarina*. (Mulet *et al.*, 2010, 2012, Gomila *et al.*, 2015). Los genomas correspondientes a las cepas de las especies *Azotobacter chroocum*, *A. vinelandii* y *Pseudomonas aeruginosa* se han incluido como *outgroup* del estudio.

Materiales y métodos

Genomas analizados

Se han analizado un total de 77 genomas, 5 de ellos considerados *outgroups* del estudio (4 cepas de *Azotobacter* y el genoma de la especie tipo del género, *Pseudomonas aeruginosa* DSM 50071^T). 51 genomas correspondían a miembros de las 22 genomovares de *P. stutzeri*. De los genomas restantes, 13 en total, corresponden a cepas de especies distintas, próximas al grupo *P. stutzeri*: *P. azotifigens*, *P. balearica*, *P. chloritidismutans*, *P. kunmingensis*, *P. kuykendallii*, *P. songnenensis*, *P. xanthomarina* y *P. zhaodongensis*. En el estudio contamos con 8 genomas repetidos con el fin de llevar a cabo un control interno de reproducibilidad de las técnicas utilizadas ya sea en el trimado (*P. stutzeri* 269) i bien diferentes carreras obtenidas de lecturas de Illumina (*P. stutzeri* CLN100), entre otros. De los 77 genomas, 37 corresponden a genomas de las bases de datos, mientras que el resto, 40 genomas están disponibles en nuestro laboratorio y no han sido publicados. De estos 40 genomas, a lo largo de este trabajo de fin de máster se ha llevado a cabo el ensamblaje de 13 de ellos, mediante el programa Newbler *De Novo Assembler* vs 2.7 (Roche). Todas las lecturas eran Illumina y antes de realizar el ensamblaje se analizó la calidad de éstas utilizando el programa FastQC (Andrews *et al.*, 2010). Para el ensamblaje de las lecturas se consideró una cobertura mínima del 60% de cobertura y un 98% de similitud entre ellas. En la Tabla 1 se detalla la información correspondiente a los genomas analizados en este estudio.

Tabla 1. Genomas analizados en este estudio. Se indican tanto los genomas pertenecientes a la base de datos del NCBI como los propios del grupo de investigación del laboratorio de microbiología. Los genomas de los microorganismos destacados con las letras superíndice A, B, C, D se presentaron por duplicado atendiendo a: A, están los genomas duplicados, diferenciándose en sí en que las lecturas del ensamblaje se trimaron o no. B, un genoma proviene de la base de datos NCBI, y el otro de la base de datos del grupo de investigación. C, ambos genomas corresponden a dos *runs* Illumina diferentes y a dos ensamblajes distintos, dado el elevado número de contigs que el grupo de investigación había obtenido. D, Ambos genomas difieren por el grupo de investigación que realizó la secuenciación y ensamblaje. En negrita se destacan las cepas de referencia para cada genomovar. Los genomas publicados por el grupo de investigación se destacaron subrayándose. Con el fondo gris se marcan los genomas ensamblados mediante el programa Newbler De Novo Assembler vs 2.7 (Roche) en este estudio. Genomovar (gv).

Organismo	Cepa	Cepas nombre trabajo	Genomovar	BioProject	Assembly	Contigs / Scaffolds	Tamaño (pb)	GC %	N50	Número de Acceso	Referencia genoma
<i>P. stutzeri</i>	ATCC 17588 ^A	01PstATCC17588T	gv1	PRJNA68131	GCA_000219605.1	1	4547930	63.9	-	CP002881.1	Chen et al., 2011
<i>P. stutzeri</i>	A1501	02PstA1501	gv1	PRJNA16817	GCA_000013785.1	1	4567418	63.9	-	CP000304.1	Yan et al., 2008
<i>P. stutzeri</i>	B1SMN1	<u>03PstB1SMN1AMVM01</u>	gv1	PRJNA170978	<u>GCA_000416345.1</u>	78	5318847	63.351	247369	<u>AMVM01</u>	<u>Busquets et al., 2013A</u>
<i>P. stutzeri</i>	NT0124	04PstNT0124JXTL01	gv1	PRJNA273703	GCA_000952205.1	35	4597183	64.226	387811	JXTL01	Unpublished
<i>P. stutzeri</i>	DSM 4166	05PstDSM4166	gv1	PRJNA63543	GCA_000195105.1	1	4689946	64	-	CP002622.1	Yu et al., 2011
<i>P. stutzeri</i>	T13	06PstT13_ALJB01	gv1	PRJNA170692	GCA_000282955.1	71	4648939	63.937	164111	ALJB01	Li et al., 2012
<i>P. stutzeri</i>	XLDN-R	07PstXLDNR_AKYE01	gv1	PRJNA168597	GCA_000280555.1	167	4695416	63.889	105529	AKYE01	Liu et al., 2012
<i>P. stutzeri</i>	11A	08Pst11A	gv1	-	-	90	4688906	63.680	224114	-	Unpublished
<i>P. stutzeri</i>	SD93936	09PstSD93936	gv1	-	-	214	4803292	63.950	41813	-	Unpublished
<i>P. stutzeri</i>	A160/74	10PstA160_74	gv1	-	-	112	4413735	64.090	77398	-	Unpublished
<i>P. stutzeri</i>	A665/78	11PstA665_78	gv1	-	-	121	4156905	64.069	58577	-	Unpublished
<i>P. stutzeri</i>	267 PSTU	53Pst267PSTUJVNQ01	? - gv1	PRJNA267549	GCA_001064225.1	144	4638698	63.808	6187	JVNQ01	Unpublished
<i>P. stutzeri</i>	C2	55PstC2LDWB01	? - gv1	PRJNA285410	GCA_001276475.1	126	4341300	64.114	70203	LDWB01	Unpublished
<i>P. stutzeri</i>	SLG510A3-8	63PstSLG510A38	? - gv1	PRJNA286160	GCA_001038645.1	1	4650155	64	-	CP011854.1	Unpublished
<i>P. stutzeri</i>	ATCC 14405	<u>12PstATCC14405ZobellAGSLC</u>	gv2	<u>PRJNA74687</u>	<u>GCA_000237885.2</u>	130	4525589	61.383	8883	<u>AGSL01</u>	<u>Pena et al., 2012</u>
<i>P. stutzeri</i>	DSM 50227	13PstDSM50227	gv3	-	-	45	4396089	63.073	407752	-	Unpublished
<i>P. kunmingensis</i> ^A	DSM 25974 ^T	14Pst269P	gv3	-	-	115	4693365	62.318	249047	-	Unpublished
<i>P. kunmingensis</i> ^A	DSM 25974 ^T	15Pst269P_trimat	gv3	-	-	78	4647845	62.604	291287	-	Unpublished
<i>P. stutzeri</i> ^B	ST-9	16PstST9_	gv3	-	-	116	4643775	62.724	162221	-	Unpublished
<i>P. stutzeri</i> ^B	ST-9	<u>17PstST9JXJL01</u>	gv3	<u>PRJNA266857</u>	<u>GCA_000982865.1</u>	116	4643775	62.724	162221	<u>JXJL01</u>	<u>Gomila et al., 2015</u>
<i>P. stutzeri</i>	AER51	18PstAER51	gv3	-	-	38	4686878	62.691	355775	-	Unpublished
<i>P. stutzeri</i>	CCUG 29243	<u>19PstCCUG29243 AN10</u>	gv3	<u>PRJNA167996</u>	<u>GCA_000267545.1</u>	1	4709064	62.7	-	<u>CP003677.1</u>	<u>unet-Galmés et al., 2012</u>
<i>P. stutzeri</i>	AN11	20PstAN11	gv3	-	-	29	4710198	62.899	355775	-	Unpublished
<i>P. xanthomarina</i>	S11	21PxanthoS11CCYE01	gv3	PRJEB7265	GCA_000825645.1	73	5051972	62.346	222853	CCYE01	Koechler et al., 2014
<i>P. stutzeri</i>	BAL361	22PstBAL361JXXD01	gv3	PRJNA261322	GCA_000935215.1	398	4878963	62.618	25831	JXXD01	entzon-Tilia et al., 2015
<i>P. chloritidismutans</i>	AW-1	23PchloriAW1AOFQ01	gv3	PRJNA182713	GCA_000495915.1	77	5056349	62.493	10397	AOFQ01	Unpublished
<i>P. stutzeri</i>	CCUG 46542	47Pstccug46542_146st	gv19	-	-	83	4733175	63.137	4730979	-	Unpublished
<i>P. stutzeri</i>	NF13	<u>48PstNF13AOBS01_9</u>	gv19	<u>PRJNA170977</u>	<u>GCA_000341615.1</u>	82	4674667	63.039	128824	<u>AOBS01</u>	<u>Pena et al., 2013</u>
<i>P. stutzeri</i>	19SMN4	<u>24Pstutzeri19SMN4</u>	gv4	PRJNA242326	GCA_000661915.1	2	4725662	62.21	-	CP007509.1	Unpublished
<i>P. stutzeri</i>	ST27MN3	26ST27MN3_123st	gv4	-	-	38	4654844	62.425	4653065	-	Unpublished
<i>P. stutzeri</i>	DNSP21	27DNSP21_125st	gv5	-	-	32	4379993	63.585	4378552	-	Unpublished
<i>P. stutzeri</i>	DSM 50238	28PstDSM50238_130st	gv7	-	-	32	4006942	64.463	4005282	-	Unpublished
<i>P. stutzeri</i>	AER2.7	29PstAER27	gv7	-	-	89	4072800	64.332	4070800	-	Unpublished
<i>P. stutzeri</i>	4FB3	30Pst4FB3	gv7	-	-	79	4061201	64.591	4054096	-	Unpublished
<i>Pseudomonas sp. Chol1</i>	Chol1	31PseudoChol1AMSL01	gv7	PRJNA174132	GCA_000306015.1	199	4869033	64.019	79285	AMSL01	Holert et al., 2013
<i>P. stutzeri</i>	KF716	57PstKF716BBBQQ01	? - gv7	PRJDB3357	GCA_001294145.1	30	4188013	64.223	370448	BBQQ01	Hirose et al., 2015
<i>P. stutzeri</i>	TS44	64PstTS44AJXE01	? - Nueva 2	PRJNA162447	GCA_000263395.1	78	4278818	64.352	130852	AJXE01	Li et al., 2012
<i>P. stutzeri</i>	DSM 10701	32PstutzeriDSM10701_JM300	gv8	<u>PRJNA78211</u>	<u>GCA_000279165.1</u>	1	4174118	63.2	-	<u>CP003725.1</u>	<u>Busquets et al., 2012</u>
<i>P. stutzeri</i>	KC	33PstKC_136st	gv9	-	-	28	4785124	61.735	4784515	-	Unpublished
<i>P. stutzeri</i> ^C	CLN100	34PstCLN100_137st	gv10	-	-	284	5138701	61.076	5112534	-	Unpublished
<i>P. stutzeri</i> ^C	CLN100	35PstCLN100	gv10	-	-	325	5163562	61.064	5136326	-	Unpublished

Organismo	Cepa	Cepas nombre trabajo	Genomovar	BioProject	Assembly	Contigs / Scaffolds	Tamaño (pb)	GC %	N50	Número de Acceso	Referencia genoma
<i>P. stutzeri</i> ^f	CLN100	35PstCLN100	gv10	-	-	325	5163562	61.064	5136326	-	Unpublished
<i>P. stutzeri</i> ^p	MT1	45PstMT1	gv18	PRJDB2890	GCA_000828755.1	1	4861258	60.2	-	AP014655.1	Fujinami et al., 2014
<i>P. stutzeri</i> ^p	MT1	46PstMT1_145st	gv18	-	-	47	4794419	60.207	4793464	-	Unpublished
<i>P. stutzeri</i>	28a50	36Pst28a50	gv11	-	-	41	4717495	60.583	4711844	-	Unpublished
<i>P. stutzeri</i>	28a24	37Pst28a24	gv11	PRJNA240232	GCA_000590475.1	1	4731359	60.6	-	CP007441.1	Smith et al., 2014
<i>P. stutzeri</i>	28a39	38Pst28a39	gv12	-	-	34	4360045	60.703	4359095	-	Unpublished
<i>P. stutzeri</i>	SD25545	39PstSD25545	gv12	-	-	52	4228683	60.528	4228683	-	Unpublished
<i>P. stutzeri</i>	28a22	40Pst28a22	gv13	-	-	30	4622615	59.874	4620728	-	Unpublished
<i>P. stutzeri</i>	28a3	41Pst28a3	gv14	-	-	57	4968746	62.325	4967315	-	Unpublished
<i>P. stutzeri</i>	MF28	59PstMF28ATAR01	? - gv 14	PRJNA202932	GCA_000455665.1	91	4943564	62.289	12867	ATAR01	Chauhan et al., 2013
<i>P. stutzeri</i>	4c29	42Pst4c29	gv15	-	-	45	4444083	60.530	4441427	-	Unpublished
<i>P. stutzeri</i>	24a13	43Pst24a13	gv16	-	-	48	4795484	61.329	4790818	-	Unpublished
<i>P. stutzeri</i>	24a75	44Pst24a75	gv17	-	-	122	4726710	65.184	4720147	-	Unpublished
<i>P. stutzeri</i>	PE	49PstPE_161st	gv20	-	-	65	4813492	59.368	4809576	-	Unpublished
<i>P. zhaodongensis</i>	DSM 27559 ^T	50PzhaoDSM27559T	gv20	-	-	30	4662711	59.585	4662711	-	Unpublished
<i>P. stutzeri</i>	A563/77	51PstA563_77_147st	gv21	-	-	78	5393046	59.826	5388864	-	Unpublished
<i>P. stutzeri</i>	stvem81	52Pst_stvem81_163st	gv22	-	-	63	3698628	60.690	3694452	-	Unpublished
<i>P. stutzeri</i>	A31/70	54PstA31_70	? - Nueva 1	-	-	86	4755882	65.541	4750924	-	Unpublished
<i>P. stutzeri</i>	CCUG 58779	56PstCCUG58779	? - Nueva 1	-	-	97	4661691	65.685	4650523	-	Unpublished
<i>P. stutzeri</i>	SD129	61PstSD129	? - Nueva 1	-	-	81	4746475	65.549	4756345	-	Unpublished
<i>P. stutzeri</i>	KOS6	58PstKOS6AMCZ02	? - Nueva 3	PRJNA171881	GCA_000307775.2	5	4947214	62.7	2709949	AMCZ02	Grigoryeva et al., 2013
<i>P. stutzeri</i>	NT0128	60PstNT0128JYHV01	? - Nueva 4	PRJNA275697	GCA_000952685.1	47	4895864	61.041	250415	JYHV01	Unpublished
<i>P. stutzeri</i>	SDM-LAC	62PstSDMLAC_AGSX01	? - Nueva 5	PRJNA74741	GCA_000235745.2	199	4233213	60.560	39917	AGSX01	Jiang et al., 2012
<i>P. xanthomarina</i>	DSM 18231 ^T	72Pxantho_DSM18231T	? - Nueva 6	-	-	41	4302892	60.346	4224442	-	Unpublished
<i>P. stutzeri</i>	RCH2	65PstRCH2	? - Nueva 7	PRJNA60029	GCA_000327065.1	4	4575057	62.471	-	CP003071.1	Unpublished
<i>P. songnenensis</i>	DMS 27560 ^T	71PsongDMS27560T	? - Nueva 8	-	-	31	4325064	63.243	4325064	-	Unpublished
<i>P. azotifigens</i>	DSM 17556 ^T	67PazotifigensAUDU01	-	PRJNA188903	GCA_000425625.1	59	5017423	66.9	171594	AUDU01	Busquets et al., 2013B
<i>P. balearica</i>	DSM 6083 ^T	68Pbalearica_CP007511	<i>P. balearica</i>	PRJNA240670	GCA_000818015.1	1	4383480	64.655	-	CP007511.1	Bennasar et al., 2016
<i>P. balearica</i>	CCUG 66666	69PbaICCG66666	<i>P. balearica</i>	-	-	54	4226047	65.090	4223727	-	Unpublished
<i>P. balearica</i>	CCUG 66667	70PbaICCG66667	<i>P. balearica</i>	-	-	42	4335452	64.980	4333507	-	Unpublished
<i>P. stutzeri</i>	YC-YH1	66PstYCYH1JUDR01	? - <i>P. balearica</i>	PRJNA268715	GCA_000828205.1	56	4838453	64.070	568314	JUDR01	Shi et al., 2015
<i>P. kuykendallii</i>	LMG 22563 ^T	73PkuyLMG26364T	-	-	-	49	4681096	65.896	-	-	Unpublished
<i>A. chroococcum</i>	NCIMB 8003	74AchroococcumNCIMB8003	-	PRJNA224116/PRJNA200545	GCA_000817975.1	7	5192286	66.3	-	CP010415.1	Robson et al., 2015
<i>A. vinelandii</i>	CA	75AvinelandiiCA	-	PRJNA186698	GCA_000380335.1	1	5366370	65.7	-	CP005094.1	Noar and Bruno, 2013
<i>A. vinelandii</i>	CA6	76AvinelandiiCA6	-	PRJNA186699	GCA_000380365.1	1	5323210	65.6	-	CP005095.1	Noar and Bruno, 2013
<i>A. vinelandii</i>	DJ1	77AvinelandiiDJ1	-	PRJNA16	GCA_000021045.1	1	536532	65.7	-	CP001157.1	Setubal et al., 2009
<i>P. aeruginosa</i>	DSM 50071	78PaerDSM50071T	-	PRJNA285827	GCA_001045685.1	1	6317050	66.5	-	CP012001.1	Nakano et al., 2015

Comparación de genomas

Se ha calculado la correlación de frecuencia de tetranucleótidos (TETRA) y el promedio de identidad de nucleótidos (ANI) entre las distintas parejas de genomas del grupo *P. stutzeri*. Los cálculos estadísticos de las frecuencias de tetranucleótidos (TETRA) (Teeling *et al.*, 2004), el ANIb y ANIm se calcularon utilizando el programa JSpecies, disponible en la página web <http://www.imedea.uib.es/jspecies>. Para calcular el ANI, la secuencia genómica de una de las parejas a comparar se corta en fragmentos consecutivos de 1020 nucleótidos. Seguidamente estos fragmentos de 1020 nucleótidos se utilizan para buscar similitudes en la secuencia genómica del otro genoma a comparar, utilizando el algoritmo BLASTN (Altschul *et al.*, 1997) para el cálculo de ANIb (Altschul *et al.*, 1997; Goris *et al.*, 2007), o bien mediante la herramienta de alineamiento ultrarrápido MUMmer, ANIm (Kurtz *et al.*, 2004). El valor de ANI entre ambos genomas, se calcula a partir de la identidad media de todos los valores que muestran un promedio de identidad de secuencia del 30 % sobre una región alineada de al menos el 70 % de la longitud de secuencia. El punto de corte recomendado para delinear especies es de 95% para los parámetros ANIb y ANIm; y para el análisis TETRA se estipula un valor superior a 0.99 (Richter and Rosselló-Móra, 2009). Las matrices obtenidas para estos parámetros para cada pareja de valores se utilizaron para generar dendrogramas mediante el programa Permut Matrix aplicando el método de agrupamiento jerárquico (UPGMA) y la correlación de distancia de Pearson (Caraux and Pinloche, 2005).

Análisis multigénico

Para cada genoma se extrajeron las secuencias correspondientes a los genes 16S rRNA, *gyrB*, *rpoB* y *rpoD*, además de los 40 genes universales en copia única definidos como marcadores filogenéticos por la herramienta online SpecI (*species identification tool*; <http://www.bork.embl.de/software/specI/>), Tabla 2 (Mende *et al.*, 2013). Las secuencias del gen 16S rRNA, *gyrB* y *rpoD* se compararon con las secuencias correspondientes de todas las genomovares previamente publicadas por el grupo y disponibles en la base de datos del laboratorio de microbiología de la Universidad de las Islas Baleares (Cladera *et al.*, 2004; Sikorski *et al.*, 2005; Mulet *et al.*, 2008; Scotta *et al.*, 2012, 2013), para comprobar la autenticidad de las cepas. Para cada gen se realizaron los correspondientes árboles filogenéticos individuales, así como un análisis concatenado de los genes completos 16S rRNA, *gyrB*, *rpoD*, el concatenado utilizando fragmentos de los genes en base a la base de datos proporcionada por el grupo de investigación laboratorio de microbiología de la Universidad de las Islas Baleares, y en última instancia el análisis del concatenado de los

Tabla 2. Representación de los 40 genes universales en copia única definidos como marcadores filogenéticos por la herramienta online SpecI. Se puede observar en la tabla, el gen al que se refiere cada determinada COG, el *Locus tag*, el producto del gen y la longitud de la secuencia. ◆ Indica la secuencia génica descartada por su complejidad. ◇ Marca las secuencias génicas seleccionadas para llevar a cabo el análisis concatenado de los genes obtenidos mediante el programa SpecI. Con un * se indican los genes que no mantenían la topología filogenética descrita por Scotta y colaboradores (2013). nt, indica la longitud nucleotídica de cada gen.

COG	Gen	Locus_tag	Producto	nt	COG	Gen	Locus_tag	Producto	nt
COG0012* ◇	<i>engD</i>	PSTAB_3240	translation-associated GTPase	1097	COG0100 ◇	<i>rpsK</i>	PSTAB_0693	30S ribosomal protein S11	390
COG0016* ◇	<i>pheS</i>	PSTAB_2266	phenylalanyl-tRNA synthetase subunit alpha	1016	COG0102 ◇	<i>rplM</i>	PSTAB_0963	50S ribosomal protein L13	429
COG0018	<i>argS</i>	PSTAB_0584	arginyl-tRNA synthetase	1768	COG0103	<i>rpsI</i>	PSTAB_0964	30S ribosomal protein S9	391
COG0048 ◇	<i>strA</i>	PSTAB_0665	ribosomal protein S12	372	COG0124*	<i>hisS</i>	PSTAB_3076	histidyl-tRNA synthetase	1273
COG0049* ◇	<i>rpsG</i>	PSTAB_0666	30S ribosomal protein S7	471	COG0172*	<i>serS</i>	PSTAB_2173	seryl-tRNA synthetase	1278
COG0052* ◇	<i>rpsB</i>	PSTAB_1441	30S ribosomal protein S2	713	COG0184	<i>rpsO</i>	PSTAB_3354	30S ribosomal protein S15	270
COG0080*	<i>rplK</i>	PSTAB_0659	50S ribosomal protein L11	432	COG0185	<i>rpsS</i>	PSTAB_0674	30S ribosomal protein S19	272
COG0081*	<i>rplA</i>	PSTAB_0660	50S ribosomal protein L1	694	COG0186	<i>rpsQ</i>	PSTAB_0679	30S ribosomal protein S17	267
COG0085 ◇	<i>rpoB</i>	PSTAB_0663	DNA-directed RNA polymerase subunit beta	4114	COG0197 ◇	<i>rplP</i>	PSTAB_0677	50S ribosomal protein L16	414
COG0087 ◇	<i>rplC</i>	PSTAB_0670	50S ribosomal protein L3	603	COG0200	<i>rplO</i>	PSTAB_0689	50S ribosomal protein L15	435
COG0088*	<i>rplD</i>	PSTAB_0671	50S ribosomal protein L4	603	COG0201 ◇	<i>secY</i>	PSTAB_0690	preprotein translocase subunit SecY	1330
COG0090	<i>rplB</i>	PSTAB_0673	50S ribosomal protein L2	825	COG0202 ◇	<i>rpoA</i>	PSTAB_0695	DNA-directed RNA polymerase subunit alpha	1002
COG0091 ◇	<i>rplV</i>	PSTAB_0675	50S ribosomal protein L22	333	COG0215	<i>cysS</i>	PSTAB_1954	cysteinylyl-tRNA synthetase	1361
COG0092 ◇	<i>rpsC</i>	PSTAB_0676	30S ribosomal protein S3	681	COG0256	<i>rplR</i>	PSTAB_0686	50S ribosomal protein L18	342
COG0093 ◇	<i>rplN</i>	PSTAB_0680	50S ribosomal protein L14	369	COG0495 ◇	<i>leuS</i>	PSTAB_3722	Leucyl-tRNA synthetase	2535
COG0094 ◇	<i>rplE</i>	PSTAB_0682	50S ribosomal protein L5	538	COG0522 ◇	<i>rpsD</i>	PSTAB_0694	30S ribosomal protein S4	621
COG0096	<i>rpsH</i>	PSTAB_0684	30S ribosomal protein S8	393	COG0525* ◇	<i>valS</i>	PSTAB_3107	valyl-tRNA synthetase	2958
COG0097	<i>rplF</i>	PSTAB_0685	50S ribosomal protein L6	456	COG0533	<i>yjgD</i>	PSTAB_0610	Probable O-sialoglycoprotein endopeptidase	969
COG0098 ◇	<i>rpsE</i>	PSTAB_0687	30S ribosomal protein S5	473	COG0541	<i>ffh</i>	PSTAB_109	signal recognition particle protein Ffh	1044
COG0099 ◇	<i>rpsM</i>	PSTAB_0692	30S ribosomal protein S13	357	COG0552 ◆	<i>ftsY</i>	PSTAB_3948	signal recognition particle receptor FtsY	1351

genes definidos en la plataforma online SpecI (Tabla 2 y Tabla suplementaria S1).

Las secuencias para cada gen se alinearon utilizando el método jerárquico para multialineamiento CLUSTAL X (Thompson *et al.*, 1997), implementado en el paquete del programa MEGA6 (Tamura *et al.*, 2013). Las secuencias se alinearon automáticamente y se confirmaron manualmente. Las distancias evolutivas y la similitud entre las parejas de valores se calcularon mediante el programa MEGA6 y el programa DNADIST incluido en el paquete de PHYLIP versión 3.5c (Felsenstein, 1989), respectivamente. Las distancias génicas se calcularon a partir de las secuencias nucleotídicas con el método de Jukes-Cantor (Jukes and Cantor, 1969), y los dendrogramas se crearon utilizando el método Neighbour-joining, implementados en el programa MEGA6. Los valores de Bootstrap se calcularon 1,000 veces. La visualización de los árboles filogenéticos se realizó también con el software MEGA6.

Delineación de especies y/o genomovares basados en el análisis multigénico y los índices genómicos

Se han representado gráficamente los resultados obtenidos en la comparación genómica para cada pareja de valores (TETRA, ANIb y ANIm), y el análisis multigénico concatenado, basado en el análisis de los genes 16S rRNA, *gyrB* y *rpoD*, así como el análisis concatenado de 20 genes de los 40 definidos en la herramienta digital SpecI, Tabla 2 y Tabla suplementaria S1.

Resultados y Discusión

Genomas analizados

Se analizaron 77 genomas, 5 de ellos considerados *outgroups* del estudio (4 cepas de *Azotobacter* y el genoma de la especie tipo del género, *Pseudomonas aeruginosa* DSM 50071^T). 51 genomas correspondían a miembros de las 22 genomovares de *P. stutzeri*. 13 genomas corresponden a cepas de especies distintas: *P. azotifigens*, *P. balearica*, *P. chloritidismutans*, *P. kunmingensis*, *P. kuykendallii*, *P. songnenensis*, *P. xanthomarina* y *P. zhaodongensis*. Los 77 genomas se compusieron por 37 genomas de las bases de datos, mientras que los 40 genomas restantes estaban disponibles en nuestro laboratorio y aún no han sido publicados. De estos 40 genomas, a lo largo de este trabajo de fin de master se ha llevado a cabo el ensamblaje de 13 de ellos, mediante el programa Newbler *De Novo* Assembler vs 2.7 (Roche). En la Tabla 1 está representada la información característica

correspondiente a los genomas analizados en el estudio como el tamaño del genoma en pares de bases, el contenido GC, el número de *contigs/scaffold*, la genomovar, o el número de acceso a la bases de datos. Se han analizado genomas completos y genomas en *draft*, cuyo contenido GC oscila entre 59-66 %.

Ensamblaje de genomas

A partir de las lecturas Illumina obtenidas con la plataforma MiSeq, se realizó el ensamblaje mediante el programa NewBler *De Novo* “Assembler” de 13 genomas, considerando una cobertura del 60 % y una identidad del 98% entre ellas. Los genomas ensamblados corresponden a los microorganismos: *P. songnenensis* DMS 27560^T, *P. stutzeri* AER 2.7, *P. stutzeri* A160/74, *P. stutzeri* A655/78, *P. stutzeri* A31/70, *P. stutzeri* CCUG 58779, *P. stutzeri* CLN100, *P. stutzeri* SD129, *P. stutzeri* SD25545, *P. stutzeri* SD93936, *P. stutzeri* 4FB3, *P. xanthomarina* DSM 18231^T, *P. zhaodongensis* DSM 27559^T cuyas características obtenidas están presentes en la Tabla 1. Los ensamblajes realizados variaron notablemente, entre 30 a 325 *contigs*. Previamente al ensamblaje, se revisó la calidad del set de lecturas Illumina, utilizando el programa FASTQC para confirmar la ausencia de los adaptadores de la plataforma de secuenciación Illumina 1.9 y, observar la calidad de las lecturas. Se comprobó en todos los casos la ausencia de adaptadores y una calidad superior a 30, tal y como se observa en la Figura 1.

✔ Basic Statistics

Measure	Value
Filename	7_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1226740
Sequences flagged as poor quality	0
Sequence length	50-300
%GC	59

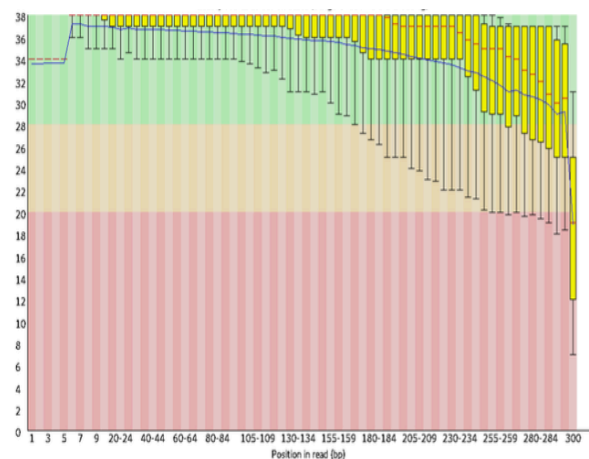


Figura 1. Ejemplo de la representación de la calidad del total de lecturas Illumina para la cepa *P. zhaodongensis* DSM 27559^T obtenidas en base al programa FastQC.

Análisis Multigénico

A partir de los genomas obtenidos de la base de datos y los disponibles en el laboratorio de Microbiología (Tabla 1), se realizó la extracción de las secuencias génicas mediante BLAST, basado en un cálculo computacional que extrae las secuencias similares a una secuencia referente. Las secuencias extraídas contienen la información del *contig* al que pertenecen, la longitud total alineada y la secuencia en sí misma. El BLAST realizado, correspondía tanto a los 40 genes ortólogos en copia única definidos por la herramienta online SpecI, así como *gyrB*, *rpoD*, *rpoB* y 16S rRNA establecidos Mulet y colaboradores en el 2008, para definir correctamente los distintos miembros de las genomovares en *P. stutzeri*.

En las 77 cepas estudiadas se encontraron secuencias génicas completas, secuencias génicas incompletas y en algunos casos, ausencia de la secuencia génica (Tabla suplementaria S1). Todos los genes presentes en el genoma están en monocopia, a excepción del gen 16S rRNA que presenta varias copias en el genoma, entre 1 a 7 copias encontradas en este estudio. En las secuencias génicas para *pheS* (COG0016), *rplK* (COG0080), *rplA* (VOG0081), *rpsI* (COG0103), *serS* (COG0172) o *vigD* (COG0533) varias cepas no presentaron secuencias génicas (entre 5 y 18 cepas). Adicionalmente, las cepas *Pseudomonas* sp. Chol1 y *P. chloritidismutans* AW-1, presentaban 2 secuencias idénticas en algunos de los genes analizados (*rpoD*, *engD* (COG0052), *argS* (COG0018), *rpsB* (COG0052), *hisS* (COG0124), *valS* (COG0525), *yigD* (COG0533); 16S rRNA, *rpoD*, *cysS* (COG0215); respectivamente).

En algunos casos, para una determinada COG y cepa se encontraron secuencias fragmentadas. En un principio, las secuencias fragmentadas aparentaban ser dos secuencias distintas. Sin embargo, se detectaron 2 posibilidades: (I) Ambos fragmentos pertenecían a la misma secuencia, y se designaban al mismo *contig*. Ambas secuencias eran contiguas. Sin embargo, existía una parte de la secuencia que discernía del BLAST. Por lo tanto, las secuencias problema fueron buscadas manualmente y se adquirieron las secuencias completas para los genes (16S rRNA, *valS* (COG0525)) que presentaban el problema. (II) En otro caso, las secuencias fragmentadas diferían en la asignación del *contig*. Mayoritariamente, estas secuencias se localizaron en posiciones terminales de los *contigs*. Estas secuencias, que presentaban el problema, se descartaron. El gen *ftsY*, correspondiente al COG0552 fue el único que presentó un elevado número de fragmentos de secuencias génicas por cepa presentando fragmentos dispuestos tanto en el mismo *contig* como presentes en diferentes *contigs*. No se encontraron secuencias del gen *ftsY* en las cepas de *P. stutzeri* PE 161st, CLN100 y *P. zhaodongensis* DSM 27559^T. Por todos estos motivos este gen se omitió de los estudios posteriores.

Autenticación de las cepas

Previamente a realizar ningún análisis se llevó a cabo la autenticación de todos los genomas analizados, con el objetivo de asegurar la correcta utilización de las cepas, y evitar posibles errores como el etiquetaje cruzado de genomas, la extracción de secuencias génicas erróneas y/o alteraciones puntuales. Se verificó la identidad de cada uno de los genomas mediante el análisis filogenético de los genes 16S rRNA, *rpoD* y *gyrB*. Se comparó la secuencia de estos genes con la base de datos proporcionada por el grupo de microbiología como referencia. Se comprobó la coherencia de todos los genomas utilizados. En todos los casos las secuencias extraídas en cada genoma concordaron con las secuencias presentes en la base de datos de microbiología. Agrupándose los mismos genomas en las mismas ramas filogenéticas sin ninguna excepción en los tres árboles filogenéticos.

Análisis filogenético. Árboles filogenéticos (individuales y concatenado)

A partir los 39 genes ortólogos en copia única definidos por la herramienta online SpecI y los genes 16S rRNA, *rpoD*, *rpoB* y *gyrB* definidos por Mulet y colaboradores (2008) se llevó a cabo el estudio filogenético, realizando los 39 árboles (SpecI) y 4 árboles (16S rRNA, *rpoD*, *rpoB* y *gyrB*) individualmente. El árbol filogenético para *ftsY*, COG0052 no se realizó debido a la complejidad de las secuencias obtenidas. Los 39 árboles obtenidos basados en el análisis de los genes de copia única del SpecI, mostraron, en general, una conservación en las agrupaciones por genomovar (árboles no mostrados). En algunos casos miembros de las genomovares 1 y 3 se distribuyeron en ramas diferentes del árbol filogenético, no mostrando la misma topología obtenida por Scotta y colaboradores (2013). Las COGs que muestran esta discrepancia están marcadas con un asterisco en la Tabla 2. Por último, cabe destacar que todos los miembros de la especie *P. balearica* y del género *Azotobacter*, se dispuso en la mayoría de las ocasiones entre las distintas genomovares definidas, dificultando el correcto discernimiento de las genomovares. El genoma de la cepa *P. stutzeri* YC-YH1 destacó por encontrarse muy próximo a la antigua genomovar 6, la especie *P. balearica*, sugiriendo una asignación filogenética errónea de este genoma.

En el caso del árbol filogenético basado en el análisis del gen 16S rRNA, este presentó varias copias de este operón. Se analizaron todos los operones pero para llevar a cabo el análisis filogenético se escogió aquella secuencia genómica que fuese igual a la secuencia presente en la base de datos del área de microbiología, obtenida mediante secuenciación de Sanger. Tanto el árbol individual para el gen 16S rRNA como para los genes *rpoD*, *rpoB* y *gyrB* se observó

un mantenimiento de la topología y de las agrupaciones de genomovares en cierto grado, denotándose una falta de exactitud y precisión en el discernimiento de ellas utilizando solamente los árboles individuales.

Las diferentes topologías observadas en los 39 genes ortólogos definidos por la herramienta online SpeI se explicaría porque la mayoría de estas COGs correspondían a genes ribosómicos y por tanto con poco poder discriminativo por sí mismos. Se realizaron varios análisis de genes concatenados con el fin de aumentar el poder discriminativo de los árboles filogenéticos individuales. Se llevó a cabo el análisis concatenado de los genes 16S rRNA, *gyrB* y *rpoD* (en este orden), utilizando secuencias cortas según la base de datos del grupo de investigación, así como las secuencias génicas sin cortar. Por último, se llevó a cabo el análisis concatenado de 20 genes ortólogos de la herramienta SpeI siguiendo el criterio de no incluir los genes que presentaban alguna cepa ausente como se puede comprobar en la Figura 2, Tabla suplementaria S1. Los genomas correspondientes a las cepas de *P. stutzeri* MF28 y *P. stutzeri* YC-YH1 no presentaron la secuencia génica 16S rRNA, *rpoD* respectivamente por lo cual, ambas cepas se excluyeron del árbol filogenético concatenado. Las cepas CLN-100 y MT-1 también presentaron ausencia de la secuencia génica para el gen *gyrB*. No obstante, ambas cepas eran idénticas para los otros genes analizados con sus respectivas secuencias de la base de datos del grupo de investigación. Incorporándose en estos dos casos las secuencias del gen *gyrB* de la base de datos del grupo de investigación a nuestro análisis concatenado.

Según el procedimiento propuesto por Mulet y colaboradores (2010), se prosiguió el análisis filogenético creando el concatenado compuesto por las secuencias génicas del gen 16S rRNA, *rpoD*, y *gyrB*. Este análisis permitió una mejora en la eficiencia para discernir genomovares y especies en el género *Pseudomonas*. El concatenado de 3 genes, se desarrolló en el siguiente orden, el gen 16S rRNA (1033 pb), el gen *gyrB* (824 pb) y el gen *rpoD* (603 pb) (Figura 3). En los 77 microorganismos, *outgroup* incluido se estudió su diversidad genética. En el concatenado basado en el análisis de 3 genes, se mantuvo tanto el agrupamiento en genomovares como la topología, definida en el trabajo de Scotta y colaboradores (2013). En el árbol filogenético se pudo observar los genomas de *P. stutzeri* RCH2 y *P. songnenensis* DSM 27560^T intercalados entre las genomovares 4 y 9. La cepa *P. stutzeri* TS44 intercalada entre la genomovar 17 y 7. En general, las nuevas cepas incluidas, se adscribieron a antiguas genomovares: las cepas *P. stutzeri* DNSP21, *P. stutzeri* C2 y *P. stutzeri* SLG510A3-8 asociadas a la genomovar 1; en la genomovar 7 aparece la cepa *P. stutzeri* KF716. Por último, los resultados obtenidos para los genomas de las cepas *P. stutzeri*

Figura 2. Árbol filogenético basado en el análisis concatenado de los 20 genes ortólogos de la herramienta online SpeCI presentes en todos los genomas analizados (19748 nucleótidos). Marcado en colores las diferentes genomovares. La mayoría de cepas representantes pertenecen a la especie *P. stutzeri*. En caso contrario, la especie se nombró en el árbol.

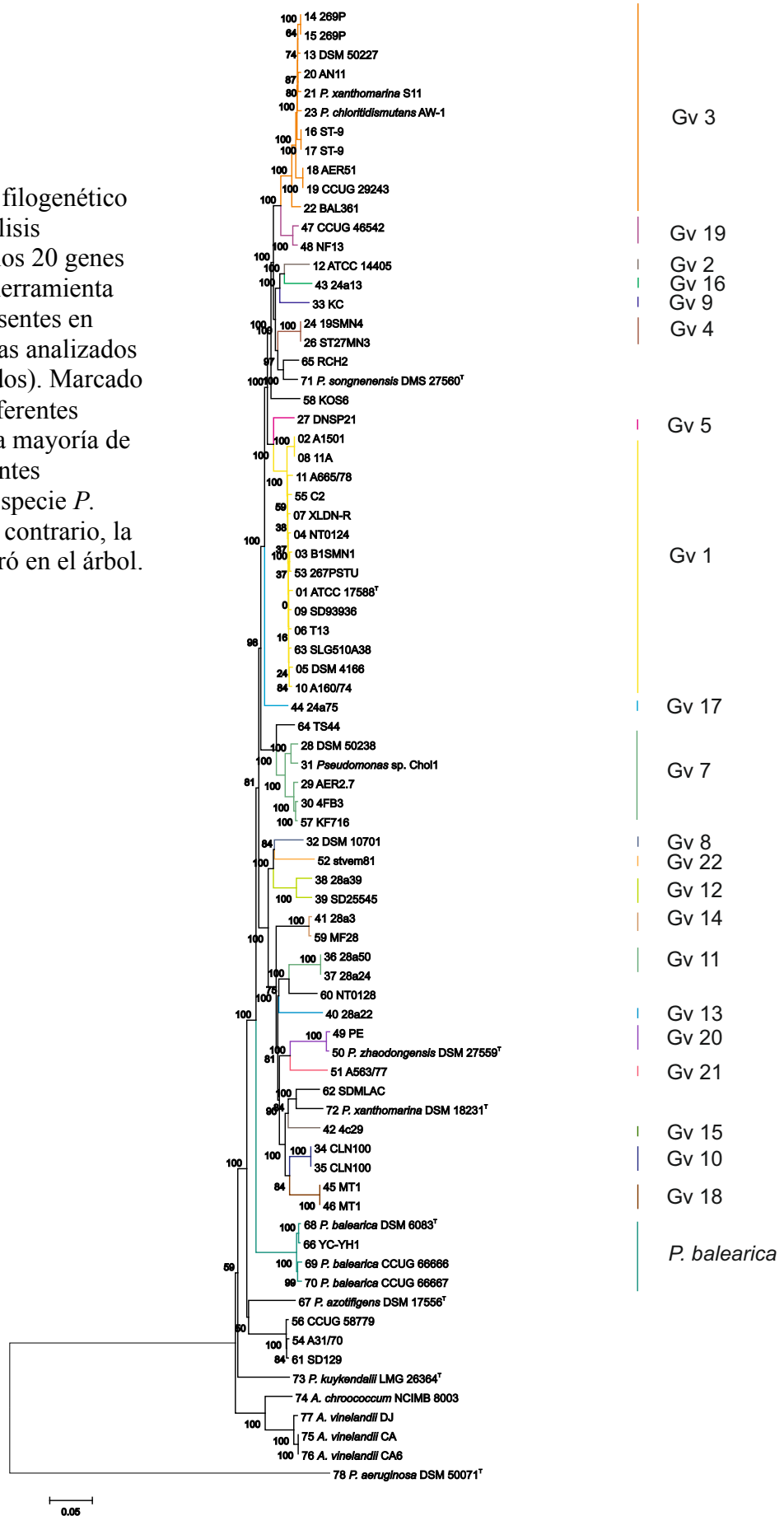
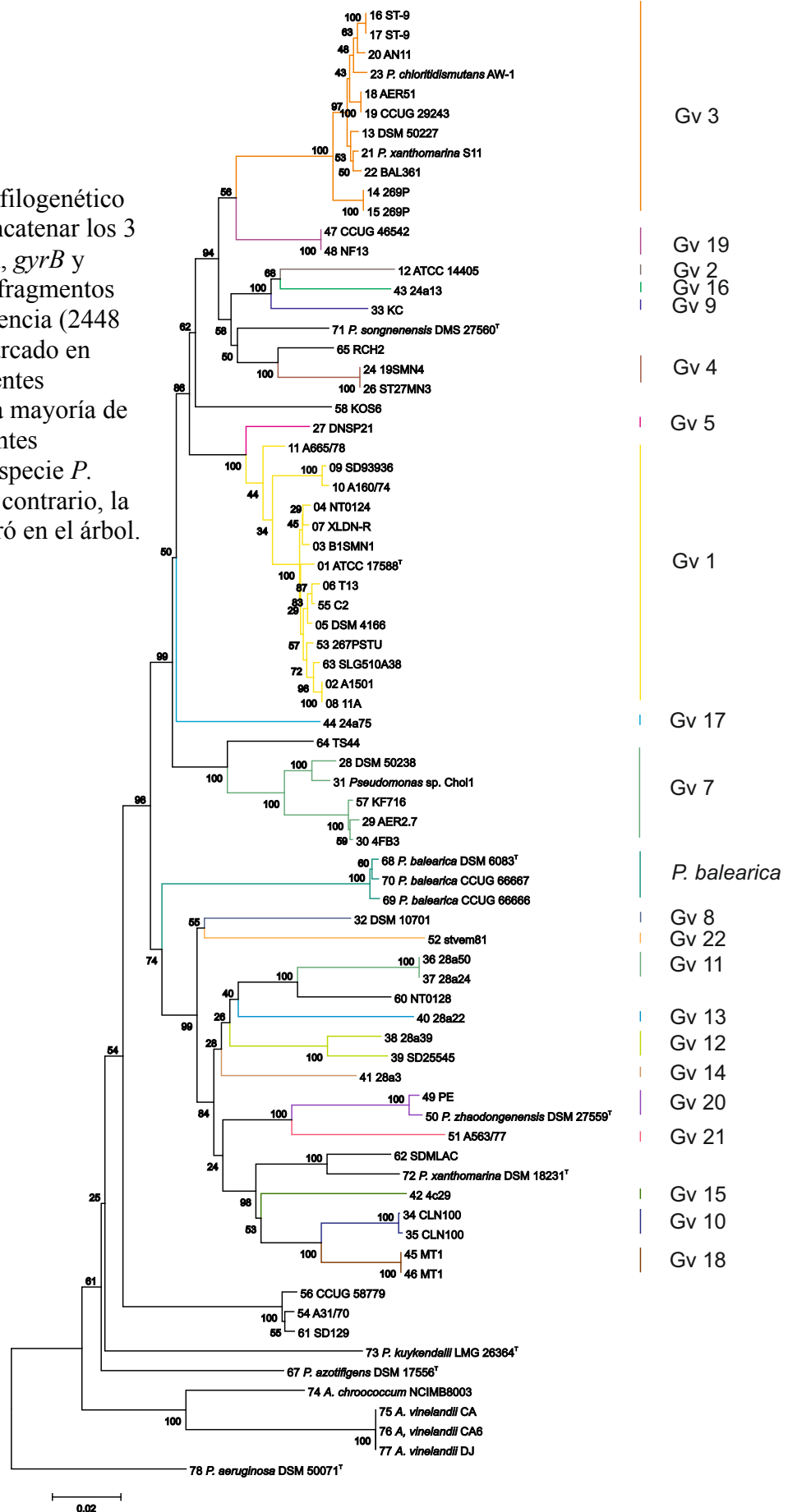


Figura 3. Árbol filogenético obtenido tras concatenar los 3 genes 16S rRNA, *gyrB* y *rpoD* utilizando fragmentos cortos de la secuencia (2448 nucleótidos). Marcado en colores las diferentes genomovares. La mayoría de cepas representantes pertenecen a la especie *P. stutzeri*. En caso contrario, la especie se nombró en el árbol.

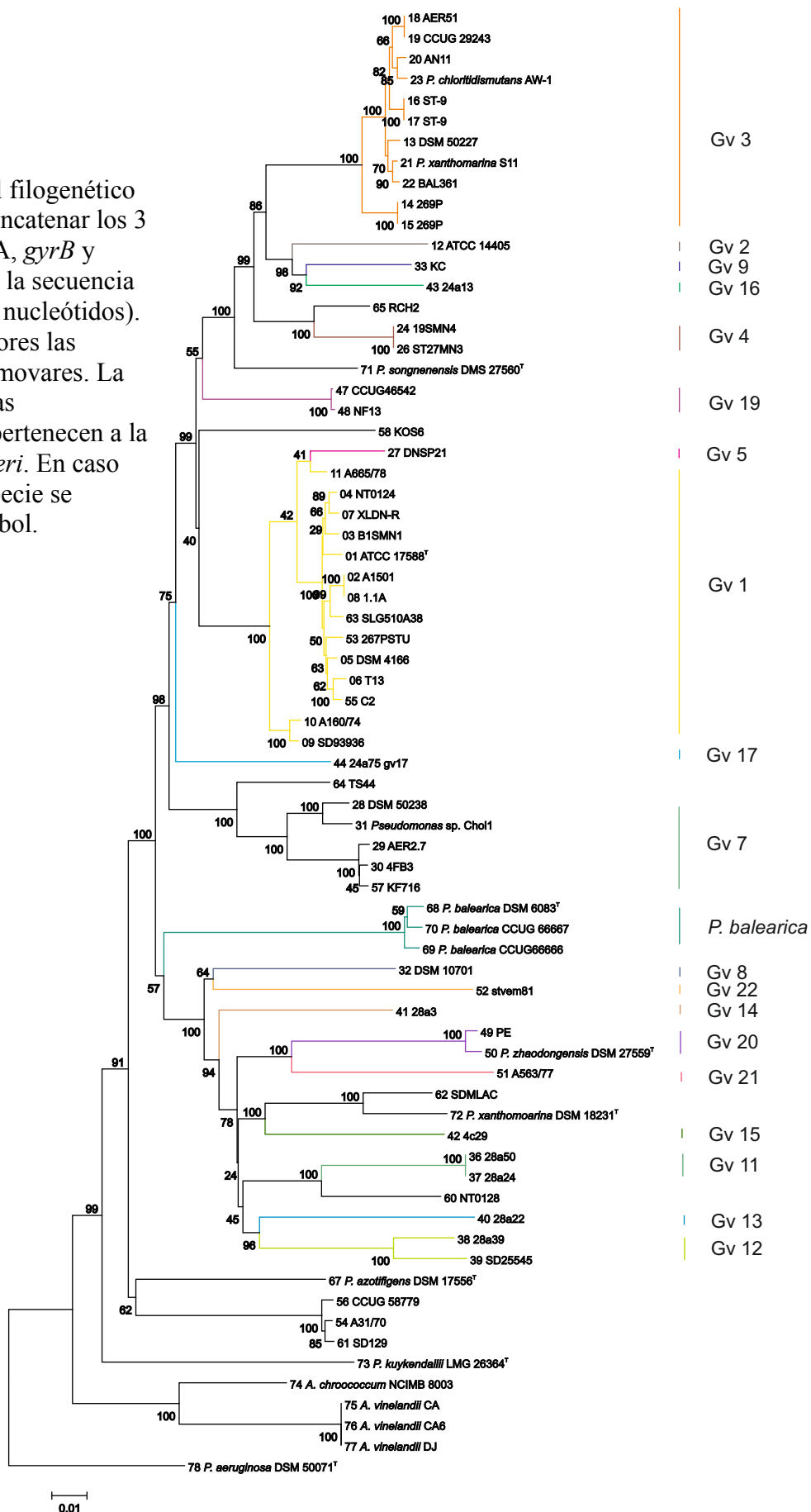


CCUG 58779, *P. stutzeri* A31/70 y *P. stutzeri* SD129 sugirieron que estas cepas podían considerarse una nueva genomovar (nueva 1), la cepa TS44 (nueva 2), KOS6 (nueva 3), NT0128 (nueva 4), SDMLAC (nueva 5), *P. xanthomarina* DSM 18231^T (nueva 6) en el grupo de *P. stutzeri* o una nueva especie si se encontrasen características fenotípicas suficientes. También, las cepas *P. songnenensis* DMS 27560^T o *P. stutzeri* TS44 conformarían dos nuevas genomovares (nueva 7 y nueva 8; entre ellas muestran un 91 % de ANIb). El concatenado completo se observaron resultados similares al concatenado de 3 genes cortado (Figura 4), con la ausencia *P. stutzeri* MF28 y *P. stutzeri* YC-YH1 como ya se explicó. Además, en este caso se eliminaron las cepas CLN-100 y MT-1 al no disponer de la secuencia génica *gyrB* completa. En el árbol del concatenado de las 3 secuencias génicas completas no se pudo discernir la gv 1 de la gv 5.

Comparación de genomas *in silico* (TETRA, ANIb, ANIm)

Se llevó a cabo una comparación *in silico* de los 77 genomas analizados. En total se trabajó con 3003 parejas de valores. En las diferentes estrategias comparativas, se observó como el análisis de frecuencia tetranucleotídica (TETRA) confirmó los resultados menos concluyentes. Los valores no eran muy discriminativos, variando solamente en el tercer decimal, resultando una diferenciación entre las genomovares poco fiable. En cuanto a la comparación genómica basada en los análisis ANIb, y ANIm estos valores mostraron resultados poco variables entre ambas estrategias. Con valores superiores al 95 % (ANIb, ANIm) se pudo discernir las agrupaciones por genomovar, y se corroboraron las 22 genomovares existentes. Los genomas pertenecientes a las cepas *P. stutzeri* DNSP21, *P. stutzeri* C2 y *P. stutzeri* SLG510A3-8 se asociaron a la genomovar 1 con valores del 97 %. La genomovar 7 incluyó un nuevo genoma, *P. stutzeri* KF716 entre el 95-97 %. Por último, los resultados obtenidos para los genomas de las cepas *P. stutzeri* CCUG 58779, *P. stutzeri* A31/70 y *P. stutzeri* SD129, con una similitud del 98 -99 % entre ellas, sugirieron a estas cepas como integrantes de una posible nueva genomovar. La genomovar 14, compuesta por la cepa 28a3, se le afilió la cepa MF28 con un valor de ANIb del 98 %. También destacó de nuevo, la similitud de *P. stutzeri* YC-YH1 con el grupo de *P. balearica* con un valor del 97 % en ambas estrategias (ANIb y ANIm). La afinidad entre *P. stutzeri* YCYH1 y *P. balearica* corroboraría los resultados obtenidos en las comparaciones de genomas (ANIb, ANIm, TETRA), estos resultados explicarían un error en la clasificación en términos de especie. Nos encontramos delante una posible nueva *P. balearica*. Sin embargo, los resultados de los arboles individuales, más los

Figura 4. Árbol filogenético obtenido tras concatenar los 3 genes 16S rRNA, *gyrB* y *rpoD* utilizando la secuencia completa (4897 nucleótidos). Marcado en colores las diferentes genomovares. La mayoría de cepas representantes pertenecen a la especie *P. stutzeri*. En caso contrario, la especie se nombró en el árbol.



resultados de los genomas proporcionaron finalmente información concluyente para ambas cepas. A partir de las 3003 comparaciones genómicas se llevaron a cabo dendrogramas utilizando los resultados obtenidos para el análisis de ANIm, ANIb o TETRA (datos no mostrados). Estos dendrogramas mostraron un buen agrupamiento de las genomovares, tanto en el caso de ANIb (Figura 5) como ANIm. Mostraron un mantenimiento de la topología, agrupando correctamente las diferentes genomovares.

Análisis comparativo

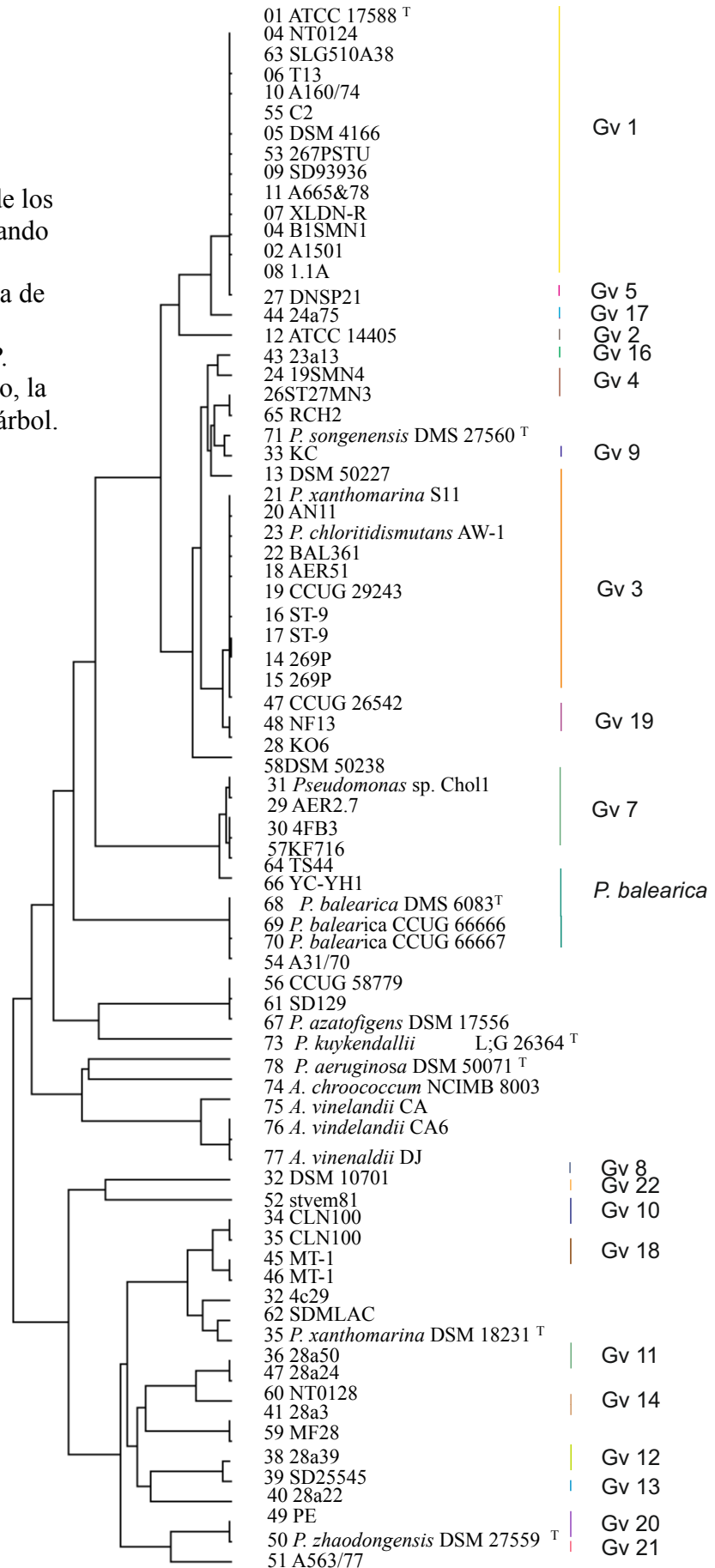
Tanto el análisis multigénico como el análisis *in silico* de los genomas (ANIb, ANI, TETRA) se compararon entre sí. Se calculó la tabla de similitud entre las diferentes cepas para el análisis concatenado de 3 genes cortos, de 3 genes completos y del concatenado de los 20 genes ortólogos del *SpeI*. Se utilizaron 3003 parejas de valores cuando los 77 genomas estaban presentes en el análisis multigénico y 2845 parejas de valores, en los concatenados basados en el análisis de tres genes, dada la ausencia de algunos genes en los genomas de *P. stutzeri* YC-YH1 y *P. stutzeri* MF28.

Los resultados observados en la Figura 6 nos permitieron comprobar que a valores superiores del 97 % en MLSA, y superiores al 95 % para ANIb y ANIm se agrupaban los miembros de una misma genomovar. A valores inferiores al 94% en MLSA y 90% en ANIb y ANIm se apreciaban claramente las diferencias intergenomovares.

No obstante, entre los valores superior al 93 % e inferior al 97 en MLSA y superior al 90% e inferior al 95 % para los valores de ANIb e ANIm existían resultados complicados. Estos valores correspondían a los valores más próximos entre diferentes genomovares, como los miembros de las genomovares 1 y 5, o entre la gv3 y gv19 y la gv10 y gv18. En estos casos sería necesario una mayor investigación con el fin de clarificar dicho rango. Los resultados comparados con la estrategia TETRA no fueron concluyentes, al presentar dificultades en inferir en las genomovares tanto intra- como inter-específicas.

Por último, en el MLSA del concatenado de 20 genes ortólogos del *SpeI* se observó una correlación con los resultados obtenidos con los otros análisis multigénicos basados en MLSA.

Figura 5. Dendograma de los resultados ANIb. Destacando en colores las diferentes genomovares. La mayoría de cepas representantes pertenecen a la especie *P. stutzeri*. En caso contrario, la especie se nombró en el árbol.



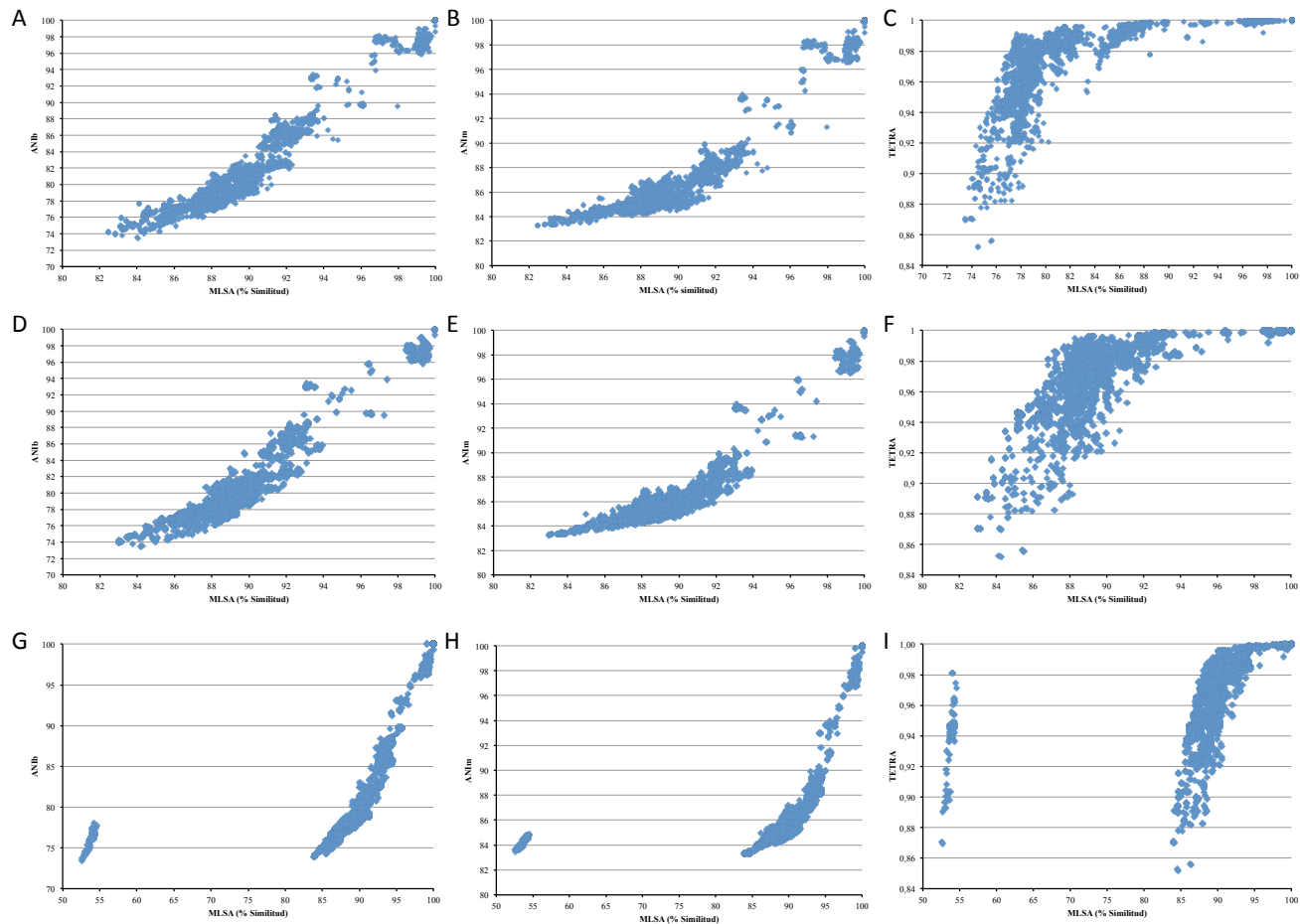


Figura 6. Representación gráfica entre los índices genómicos calculados y el porcentaje de similitud obtenido en el análisis MLSA. A, B y C son los gráficos obtenidos al comparar el porcentaje de similitud del concatenado de 3 genes cortados contra ANIb, ANIm y TETRA respectivamente. D, E, F corresponden al gráfico obtenido utilizando porcentaje de similitud del concatenado de 3 genes completos contra ANIb, ANIm y TETRA respectivamente. Por último, los gráficos G, H, I correspondieron al porcentaje de similitud del concatenado de los 20 genes universales en copia única definidos como marcadores filogenéticos por la herramienta online SpeI contra ANIb, ANIm y TETRA respectivamente.

Conclusiones

1. Los estudios de comparaciones genómicas ANIb, ANIm y MLSA son una excelente herramienta en estudios filogenéticos para la discriminación de especies. En este estudio, los resultados obtenidos corroboraron una discriminación de las diferentes genomovares del grupo *P. stutzeri*, utilizando valores superiores al 95 % de similitud para ANIb y ANIm, y valores superiores al 97 % en MLSA.
2. Estas aproximaciones han permitido inferir la afiliación filogenética y taxonómica de los 77 genomas analizados, alguno no correctamente asignado.
3. Los resultados obtenidos a partir de los genomas sugieren que las cepas CCUG 58779, SD129, A31/70 conformarían una nueva genomovar. Adicionalmente, las cepas C2, DSM4166, 267 PST, SLG510A38, serían miembros de la gv 1, y la cepa KF716 de la gv 7. Los genomas de las cepas TS44, KOS6, NT0128, RCH2 y SDM-LAC, corresponderían a nuevas genomovares. Por otro lado, *P. stutzeri* YC-YH1 se asignaría a la especie *P. balearica*.
4. Las especies *P. chloritidismutans* AW-1, *P. kunmingensis* DSM 25974^T, *P. songnenensis* DMS 27560^T, *Pseudomonas* sp. Ch011, *P. xanthomarina* DSM18231^T, *P. xanthomarina* S11, y *P. zhaodongensis* DSM 27559^T se asignaron a la especie *P. stutzeri*. Con la información adquirida en el estudio se propuso las afiliaciones de *P. chloritidismutans* AW-1, *P. kunmingensis* DSM 25974^T, *P. xanthomarina* S11 a la gv 3; *Pseudomonas* sp. Ch011 a la gv 7; *P. zhaodongensis* DSM 27559^T a la gv 20. Los genomas de las especies *P. xanthomarina* DSM 18231^T y *P. songnenensis* DSM 27560^T corresponderían a nuevas genomovares dentro de la especie *P. stutzeri*.
5. Las especies *P. kuykendallii* LMG 22563^T y *P. azotifigens* DSM 17556^T mantienen su estatus taxonómico propio.
6. Para proponer una especie nueva dentro del complejo de especies de *P. stutzeri* no es suficiente compararlo con la cepa tipo de *P. stutzeri*, sino que se deben incluir las cepas de referencia de cada una de las genomovares conocidas hasta el momento.
7. Todo el estudio presente exhibe una patente necesidad de profundizar en las investigaciones relacionadas con el grupo de *P. stutzeri*. Focalizándose en los detalles de cada genomovar descrita con el objetivo de clarificar las adscripciones taxonómicas correctamente. Los estudios genómicos, así como otros estudios posteriores como por ejemplo la espectrometría de masas MALDI-TOF permitirán otorgar un estatus taxonómico propio a las 22 genomovares descritas hasta el momento y para las futuras genomovares por describir.

Referencias

1. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, M., *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
2. Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Reference Source.
3. 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, 1563–1589. doi: 10.1099/00207713-50-4-1563
4. Bennasar, A., Salva-Serra, F., Jaen-Luchoro, D., Segui, C., Aliaga, F., Busquets, A., Gomila, M., Moore, E.R., Lalucat, J. (2016) Complete genome sequence of *Pseudomonas balearica* DSM 6083^T. *Genome Announc.* 4 (2)
5. Bennasar, A., Rosselo-Mora, R., Lalucat, J., Moore, E. R. B. (1996). 16S rRNA gene sequence analysis relative to genomovars of *Pseudomonas stutzeri* and proposal of *Pseudomonas balearica* sp. nov. *Int. J. Syst. Bacteriol.* 46 (1): 200–205. doi:10.1099/00207713-46-1-200.
6. Bentzon-Tilia, M., Severin, I., Hansen, L.H., Riemann, L. (2015). Genomics and ecophysiology of heterotrophic nitrogen-fixing bacteria isolated from estuarine surface water. *MBio* 6 (4), e00929-15
7. Brunet-Galmés, I., Busquets, A., Peña, A., Gomila, M., Nogales, B., García-Valdés, E., Lalucat, J., Bennasar, A., Bosch, R. (2012). Complete genome sequence of the naphthalene-degrading bacterium *Pseudomonas stutzeri* AN10 (CCUG 29243). *J. Bacteriol.* 194(23):6642-3. doi: 10.1128/JB.01753-12.
8. Busquets, A., Pena, A., Gomila, M., Bosch, R., Nogales, B., Garcia-Valdes, E., Lalucat, J., Bennasar, A. (2012). Genome sequence of *Pseudomonas stutzeri* strain JM300 (DSM 10701), a soil isolate and model organism for natural transformation. *J. Bacteriol.* 194 (19), 5477-5478
9. Busquets, A., Pena, A., Gomila, M., Mayol, J., Bosch, R., Nogales, B., Garcia-Valdes, E., Bennasar, A., Lalucat, J. (2013). Draft genome sequence of *Pseudomonas stutzeri* strain B1SMN1, a nitrogen-fixing and naphthalene-degrading strain isolated from wastewater. *Genome Announc.* 1 (4), e00584-13
10. Busquets, A., Pena, A., Gomila, M., Mulet, M., Mayol, J., Garcia-Valdes, E., Bennasar, A., Huntemann, M., Han, J., Chen, I.M., Mavromatis, K., Markowitz, V., Palaniappan, K., Ivanova, N., Schaumberg, A., Pati, A., Reddy, T.B., Nordberg, H., Woyke, T., Klenk, H.P., Kyripides, N., Lalucat, J. (2013) Draft genome sequence of *Pseudomonas azotifigens* strain

- DSM 17556^T(6H33bT), a nitrogen fixer strain isolated from a compost pile. *Genome Announc.* 1 (6), e00893-13
11. Caraux, G., Pinloche, S. (2005). Permutmatrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* 21, 1280–1281. doi: 10.1093/bioinformatics/bti141
 12. Chauhan, A., Green, S., Pathak, A., Thomas, J., Venkatramanan, R. (2013). Whole-genome sequences of five oyster-associated bacteria show potential for crude oil hydrocarbon degradation. *Genome Announc.* 1 (5), e00802-13
 13. Chen, M., Yan, Y., Zhang, W., Lu, W., Wang, J., Ping, S., Lin, M. (2011). Complete genome sequence of the type strain *Pseudomonas stutzeri* CGMCC 1.1803. *J. Bacteriol.* 193 (21), 6095
 14. Cladera A. M., Bennasar A., Barceló M., Lalucat J., García-Valdés E. (2004). Comparative genetic diversity of *Pseudomonas stutzeri* genomovars, *J. Bacteriol.* 186(16): 5239–5248. doi: 10.1128/JB.186.16.5239-5248.
 15. Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P., Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol. Rev.* 29, 147–167. doi: 10.1016/j.femsre.2004.11.004
 16. De Ley, J. (1992). The proteobacteria: ribosomal RNA cistron similarities and bacterial taxonomy, in the prokaryotes, a Handbook on the Biology of Bacteria, Ecophysiology, Isolation, Identification and Applications, 2nd Edn., Vol. 2, eds Balows A., Truper H. G., Dworkin M., Harder W., Schleifer K. H., editors. (Berlin: Springer;), 2111–2140.
 17. De Vos, P., De Ley, J. (1983). Intra- and intergeneric similarities of *Pseudomonas* and *Xanthomonas* ribosomal ribonucleic acid cistrons. *Int. J. Syst. Bacteriol.* 33, 487–509. 10.1099/00207713-33-3-487
 18. De Vos, P., Goor, M., Gillis, M., De Ley, J. (1985). Ribosomal ribonucleic acid cistron similarities of phytopathogenic *Pseudomonas* species. *Int. J. Syst. Bacteriol.* 35, 169–184. 10.1099/00207713-35-2-169
 19. De Vos, P., Van Landschoot, A., Segers, P., Tytgat, R., Gillis, M., Bauwens, M., *et al.* (1989). Genotypic relationships and taxonomic localization of unclassified *Pseudomonas* and *Pseudomonas*-like strains by deoxyribonucleic acid: ribosomal ribonucleic acid hybridizations. *Int. J. Syst. Bacteriol.* 39, 35–49. 10.1099/00207713-39-1-35
 20. Felsenstein, J. (1989) PHYLIP—phylogeny inference package (version 3.0). *Cladistics* 5:164–166
 21. Fujinami, S., Oikawa, Y., Araki, T., Shinmura, Y., Midorikawa, R., Ishizaka, H., Kato, C., Horikoshi, K., Ito, M., Tamegai, H. (2014). Genome sequence of the deep-sea denitrifier *Pseudomonas* sp. strain MT-1, isolated from the Mariana Trench. *Genome Announc.* 2(6),

e01313-14. doi: 10.1128/genomeA.01313-14.

22. Gomila, M., Peña, A., Mulet, M., Lalucat, J., García-Valdés, E. (2015). Phylogenomics and systematics in *Pseudomonas*. *Frontiers in Microbiology*. Vol. 6 (214), doi: 10.3389/fmicb.2015.00214
23. Gomila, M., Busquets, A., Garcia-Valdes, E., Michael, E., Cahan, R., Nitzan, Y., Lalucat, J. (2015). Draft genome sequence of the toluene-degrading *Pseudomonas stutzeri* strain ST-9. *Genome Announc.* 3 (3), e00567-15
24. Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M. (2007). DNA-DNA hybridization values and their relationship to whole genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ij.s.0.64483-0
25. Grigoryeva, T.V., Laikov, A.V., Naumova, R.P., Manolov, A.I., Larin, A.K., Karpova, I.Y., Semashko, T.A., Alexeev, D.G., Kostryukova, E.S., Muller, R., Govorun, V.M. (2013). Draft genome of the nitrogen-fixing bacterium *Pseudomonas stutzeri* Strain KOS6 isolated from Industrial hydrocarbon sludge. *Genome Announc.* 1 (1)
26. Hirose, J., Yamazoe, A., Hosoyama, A., Kimura, N., Suenaga, H., Watanabe, T., Fujihara, H., Futagami, T., Goto, M., Furukawa, K. (2015). Draft genome sequence of the polychlorinated biphenyl-degrading bacterium *Pseudomonas stutzeri* KF716 (NBRC 110668). *Genome Announc.* 3 (5)
27. Holert, J., Alam, I., Larsen, M., Antunes, A., Bajic, V.B., Stingl, U., Philipp, B. (2013). Genome sequence of *Pseudomonas* sp. strain Choll1, a model organism for the degradation of bile salts and other steroid compounds. *Genome Announc.* 1 (1), E00014-12
28. Jiang, T., Gao, C., Su, F., Zhang, W., Hu, C., Dou, P., Zheng, Z., Tao, F., Ma, C., Xu, P. (2012). Genome sequence of *Pseudomonas stutzeri* SDM-LAC, a typical strain for studying the molecular mechanism of lactate utilization. *J. Bacteriol.* 194 (4), 894-895
29. Jukes, T.H., Cantor, C.R. (1969) Evolution of protein molecules. In: Munro HN (ed) *Mammalian protein metabolism*. Academic, New York, pp 21–132
30. Koechler, S., Arsène-Ploetze, F., Brochier-Armanet, C., Goulhen-Chollet, F., Heinrich-Salmeron A., Jost, B., Lièvre-mont, D., Philipps, M., Plewniak, F., Bertin, P.N., Lett, M.C. (2015). Constitutive arsenite oxidase expression detected in arsenic-hypertolerant *Pseudomonas xanthomarina* S11. *Res. Microbiol.* 166, 205-214
31. Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., *et al.* (2004). Versatile and open software for comparing large genomes. *Genome. Biol.* 5, R12. doi: 10.1186/gb-2004-5-2-r12
32. Lalucat, J., Bennasar, A., Bosch, R., García-Valdés, E., Palleroni N. J. (2006). Biology of *Pseudomonas stutzeri*. *Microbiol. Mol. Biol. Rev.* 70(2):510-47

33. Li, A., Gai, Z., Cui, D., Ma, F., Yang, J., Zhang, X., Sun, Y., Ren, N. (2012). Genome sequence of a highly efficient aerobic denitrifying bacterium, *Pseudomonas stutzeri* T13. *J. Bacteriol.* 194 (20), 5720
34. Li, X., Gong, J., Hu, Y., Cai, L., Johnstone, L., Grass, G., Rensing, C., Wang, G. (2012). Genome sequence of the moderately halotolerant, arsenite-oxidizing bacterium *Pseudomonas stutzeri* TS44. *J. Bacteriol.* 194 (16), 4473-4474
35. Liu, X., Gai, Z., Tao, F., Yu, H., Tang, H., Xu, P. (2012). Genome sequences of *Pseudomonas luteola* XLDN4-9 and *Pseudomonas stutzeri* XLDN-R, two efficient carbazole-degrading strains. *J. Bacteriol.* 194 (20), 5701-5702
36. Mandel, M. (1966). Desoxiribonucleic acid base composition in the genus *Pseudomonas*. *J. Gen. Microbiol.* **43**:273–292.
37. Mende, D.R., Sunagawa, S., Zeller, G., Bork, P. (2013) Accurate and universal delineation of prokaryotic species. *Nature methods* 10 (9): 881-884. doi: 10.1038/NMETH.2575
38. Migula, W. (1900). *System der Bakterien. Handbuck der Morphologie, Entwicklung-Geschichte und Systematik der Bakterien*, Vol. 2. Jena: Verlag von Gustav Fischer. doi: 10.1099/00207713-37-4-463
39. Moore, E.R.B., Mua, M., Arnscheidt, A., Böttger, E.C., Hutson, R.A., Collins, M.D., *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, 476–492. doi: 10.1016/S0723-2020(96)80021-X
40. Mulet, M., Gomila, M., Gruffaz, C., Meyer, J-M., Palleroni, N.J., 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.* 2008 Oct;58(Pt 10):2309-15. doi: 10.1099/ijs.0.65797-0
41. Mulet, M., García-Valdés, E., Lalucat, J. (2013). Phylogenetic affiliation of *Pseudomonas putida* biovar A and B strains. *Res. Microbiol.* 164, 351–359. doi: 10.1016/j.resmic.2013.01.009
42. Mulet, M., Gomila, M., Lemaitre, B., Lalucat, J., García-Valdés, E. (2012b). Taxonomic characterization of *Pseudomonas* strain L48 and formal proposal of *Pseudomonas entomophila* sp. nov. *Syst. Appl. Microbiol.* 35, 145–149. doi: 10.1016/j.syapm.2011.12.003
43. Mulet, M., Gomila, M., Scotta, C., Sánchez, D., Lalucat, J., García-Valdés, E. (2012a). 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, 455–464. doi: 10.1016/j.syapm.2012. 08.007

44. Mulet, M., Lalucat, J., García-Valdés, E. (2010). DNA sequence-based analysis of the *Pseudomonas* species. *Environ. Microbiol.* 12:1513– 1530
45. Nakano, K., Terabayashi, Y., Shiroma, A., Shimoji, M., Tamotsu, H., Ashimine, N., Ohki, S., Shinzato, M., Teruya, K., Satou, K., Hirano, T. (2015) First complete genome sequence of *Pseudomonas aeruginosa* (Schroeter 1872) Migula 1900 (DSM 50071^T), determined using PacBio single-molecule real-time technology. *Genome Announc* 3 (4), e00932-15
46. Noar, J.D., Bruno-Barcena, J.M. (2013) Complete genome sequences of *Azotobacter vinelandii* wild-type strain CA and tungsten-tolerant mutant strain CA6. *Genome Announc* 1 (3)
47. Palleroni N.J. (1984). Genus I. *Pseudomonas* Migula 1894, in *Bergey's manual of systematic bacteriology*, Vol. 1, eds Krieg N. R., Holt J. G., editors. (Baltimore, MD: Williams & Wilkins;), 141–199.
48. Parte, A.(2014). LPSN-list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res.* 42, D613–D616. doi:10.1093/nar/gkt1111
49. Peix, A., Ramírez-Bahena, M.H., Velázquez, E. (2009). Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infec. Gen. Evol.* 9, 1132–1147. doi: 10.1016/j.meegid.2009.08.001
50. Pena, A., Busquets, A., Gomila, M., Bosch, R., Nogales, B., Garcia-Valdes, E., Lalucat, J., Bennasar, A. (2012). Draft genome of *Pseudomonas stutzeri* strain ZoBell (CCUG 16156), a marine isolate and model organism for denitrification studies. *J. Bacteriol.* 194 (5), 1277-1278
51. Pena, A., Busquets, A., Gomila, M., Mayol, J., Bosch, R., Nogales, B., Garcia-Valdes, E., Bennasar, A., Lalucat, J. (2013). Draft genome of *Pseudomonas stutzeri* strain NF13, a nitrogen fixer isolated from the Galapagos rift hydrothermal vent. *Genome Announc* 1 (2), E0011313
52. Richter, M., Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
53. Robson, R.L., Jones, R., Robson, R.M., Schwartz, A. , Richardson, T.H. (2015) *Azotobacter* genomes: The genome of *Azotobacter chroococcum* NCIMB 8003 (ATCC 4412). *PLoS ONE* 10 (6), E0127997
54. Rosselló, R., Garcia-Valdes E., Lalucat, J., Ursing, J. (1991). Genotypic and phenotypic diversity of *Pseudomonas stutzeri*. *Syst. Appl. Microbiol.* 14 150-157.
55. 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.* 522–32.

doi:10.1007/s00248-013-0246-8.

56. Scotta, C., Mulet, M., Sánchez, D., Gomila, M., Ramírez, A., Bennasar, A., García-Valdés, E., Holmes, B., Lalucat, J. (2012). Identification and genomovar assignment of clinical strains of *Pseudomonas stutzeri*. *Eur. J. Clin. Microbiol. Infect. Dis.* 31:2133–2139. DOI 10.1007/s10096-012-1547-4
57. Setubal, J.C., dos Santos, P., Goldman, B.S., Ertesvag, H., Espin, G., Rubio, L.M., Valla, S., Almeida, N.F., Balasubramanian, D., Cromes, L., Curatti, L., Du, Z., Godsy, E., Goodner, B., Hellner-Burris, K., Hernandez, J.A., Houmiel, K., Imperial, J., Kennedy, C., Larson, T.J., Latreille, P., Ligon, L.S., Lu, J., Maerk, M., Miller, N.M., Norton, S., O'Carroll, I.P., Paulsen, I., Raulfs, E.C., Roemer, R., Rosser, J., Segura, D., Slater, S., Stricklin, S.L., Studholme, D.J., Sun, J., Viana, C.J., Wallin, E., Wang, B., Wheeler, C., Zhu, H., Dean, D.R., Dixon, R., Wood, D. (2009) Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *J. Bacteriol.* 191 (14), 4534-4545
58. Shi, Y.H., Ren, L., Jia, Y., Yan, Y.C. (2015). Genome sequence of organophosphorus pesticide-degrading bacterium *Pseudomonas stutzeri* strain YC-YH1. *Genome Announc.* 3(2). pii: e00192-15. doi: 10.1128/genomeA.00192-15.
59. Sikorski, J., Lalucat, J., Wackernagel, W. (2005). Genomovars 11 to 18 of *Pseudomonas stutzeri*, identified among isolates from soil and marine sediment. *Int. J. Syst. Evol. Microbiol.*, 55, 1767–1770. DOI 10.1099/ijs.0.63535-0
60. Smith, B.A., Dougherty, K.M., Baltrus, D.A. (2014). Complete genome sequence of the highly transformable *Pseudomonas stutzeri* strain 28a24. *Genome Announc.* 2(3). pii: e00543-14. doi: 10.1128/genomeA.00543-14.
61. Stanier, R.Y., Palleroni, N.J., Doudoroff M. (1966). The aerobic *Pseudomonads*: a taxonomic study. *J. Gen. Microbiol.* 43, 159–271. 10.1099/00221287-43-2-159
62. Tamura, K., Stecher, G., Peterson, D., Filipinski, A., Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* (12):2725-9. doi: 10.1093/molbev/mst197
63. Teeling, H., Meyerdierks, A., Bauer, M., Amann, R., Glöckner, F. O. (2004). Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ. Microbiol.* 6, 938–947. doi: 10.1111/j.1462-2920.2004.00624.x
64. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G. (1997). The Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876–4882
65. Yan, Y., Yang, J., Dou, Y., Chen, M., Ping, S., Peng, J., Lu, W., Zhang, W., Yao, Z., Li, H., Liu, W., He, S., Geng, L., Zhang, X., Yang, F., Yu, H., Zhan, Y., Li, D., Lin, Z.,

- Wang, Y., Elmerich, C., Lin, M., Jin, Q. (2008). Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501. Proc. Natl. Acad. Sci. U S A. 105(21):7564-9. doi: 10.1073/pnas.0801093105.
66. Yu, H., Yuan, M., Lu, W., Yang, J., Dai, S., Li, Q., Yang, Z., Dong, J., Sun, L., Deng, Z., Zhang, W., Chen, M., Ping, S., Han, Y., Zhan, Y., Yan, Y., Jin, Q., Lin, M. (2011). Complete genome sequence of the nitrogen-fixing and rhizosphere-associated bacterium *Pseudomonas stutzeri* strain DSM4166. J. Bacteriol. 193(13):3422-3. doi: 10.1128/JB.05039-11.

Agradecimientos

Al Dr. Edward Moore de la colección de cultivos sueca de *Göteborgs Universitet* y al alumno de doctorado Francisco Salva de *Göteborgs Universitet*, al ceder los genomas *P. balearica* CCUG 66666 y CCUG 66667 relacionados con la especie *P. stutzeri*.

Al grupo de investigación de microbiología de la *Universitat de les Illes Balears*, por brindarme su conocimiento para alcanzar los objetivos trazados en este proyecto.

A la Dra. Balbina Nogales Fernández, Dra. Aranzazu Peña Pardo encargadas de la asignatura Seminarios de Especialización y al Dr. Jorge Lalucat Jo encargado de la asignatura Avances en Diversidad Microbiana por su asesoría y introducción en el mundo de la investigación.

Al Dr. Rafael Bosch Zaragoza tutor de mi trabajo de fin de grado y profesor de microbiología por introducirme en el mundo de las herramientas bioinformáticas de suma importancia para la realización de este proyecto.

Finalmente, agradecer a la Dra. Margarita Gomila Ribas, tutora del trabajo fin de máster y encargada de la asignatura Seminarios de Especialización por su asesoría y conocimientos brindados en el desarrollo de este proyecto.

ANEXO

