UNIVERSIDAD COMPLUTENSE
MADRID

CARACTERIZACIÓN Y ANÁLISIS FUNCIONAL DEL GENOMA DE *Lactococcus garvieae*

TESIS DOCTORAL

Mónica Aguado Urda

2013

MEMORIA PARA OPTAR AL GRADO DE DOCTOR POR LA UNIVERSIDAD COMPLUTENSE DE MADRID

UNIVERSIDAD COMPLUTENSE
MADRID

FACULTAD DE VETERINARIA
DEPARTAMENTO DE SANIDAD ANIMAL

Dª. Mª del Mar Blanco Gutiérrez, Profesora Titular de Sanidad Animal de la Facultad de Veterinaria de la UCM, D. José F. Fernández-Garayzábal Fernández, Catedrático de Sanidad Animal de la Facultad de Veterinaria de la UCM y D. Guillermo H. López Campos, Profesor de Bioinformática Médica de la Facultad de Medicina de la Universidad de Melbourne (Australia),

CERTIFICAN:

Que la Tesis Doctoral titulada **"Caracterización y análisis funcional del genoma de *Lactococcus garvieae*"** ha sido realizada por la licenciada en Biología **Dª. Mónica Aguado Urda** en el Departamento de Sanidad Animal de la Facultad de Veterinaria de la Universidad Complutense de Madrid, bajo la dirección conjunta de los que suscriben, y estimamos que cumple las condiciones exigidas para su presentación como compendio de publicaciones para optar al Título de Doctor por la Universidad Complutense de Madrid.

El trabajo recogido en esta memoria ha sido financiado por el Ministerio de Economía y Competitividad a través de los proyectos AGL2005-04775, AGL2009-12447 y AGL2012-35419. Mónica Aguado Urda ha sido beneficiaria de una Beca Predoctoral asociada a un Contrato-artículo 83 de la LOU del Centro VISAVET de la UCM, y de una Beca Predoctoral UCM.

Madrid, noviembre de 2013

Los Directores de la Tesis Doctoral

Mª. del Mar Blanco Gutiérrez

José F. Fernández-Garayzábal

Guillermo H. López Campos

A mis padres

"Lo que sabemos es una gota, lo que ignoramos un inmenso océano"

(Isaac Newton)

"No paséis por alto a los genios"

(VV)

AGRADECIMIENTOS

Con la escritura de esta Tesis Doctoral se cierra una etapa de mi vida importante, una etapa llena de alegrías, también de dificultades, pero en definitiva una experiencia tremendamente positiva, que me hace sentir orgullosa del esfuerzo y trabajo realizados. **Llega el momento de agradeceros, a todos los que me habéis acompañado durante este tiempo**, todo lo que me habéis aportado, tanto a nivel profesional como personal, **haciendo posible que llegue este momento tan deseado.**

Quiero empezar dando las gracias a **mis tres directores de Tesis** (¡Sí, tres! ¡Qué locura!). ¡¡¡Gracias a los tres por darme esta oportunidad y por enseñarme tanto!!!

Gracias Mar, por ser no sólo mi directora y compañera durante estos años sino, y sobre todo, por ser amiga, a veces hasta segunda madre, por el apoyo y la confianza incondicionales, y por estar siempre presente en buenos y malos momentos. Siempre recordaré cómo, gracias a ti, todo ésto comenzó una tarde de otoño en la puerta de Biológicas... Fue amor ¡a primera vista! ¡Me encantas! ¡De mayor quiero ser como tú!

Gracias Jose, por todo lo que he aprendido contigo, por las interesantísimas discusiones científicas (siempre con un toque de humor), y por aportar esa calma y organización a todo lo que hacemos. Gracias por tenerme siempre en cuenta y confiar en mí. ¡Ah! ¡Y Cooper también te da las gracias por los trozos gigantes de pan duro!

Gracias Guille, por elegir ser mi director (¡Se siente! ¡Tú solito te metiste!), por todo lo que me has enseñado, por las risas, por los consejos, por descubrirme el maravilloso y prometedor mundo de la Bioinformática, y por dejarme ser "tu mochila" (todo el día pegada a ti haciéndote preguntas...). Tanto me pegué a ti que ¡te seguí hasta el otro lado del Mundo! Pero sobre todo, ¡gracias por ser mi amigo!

Gracias también, por supuesto, al resto del equipo: **Ali**, gracias por ser mi compañera "biológica-molecular" (benditas seamos entre todos los veterinarios...), por los ratos faenando en el laboratorio codo con codo, por contagiar ese entusiasmo científico que desprendes y que nos ha llevado a hacer tantas cosas en estos años, y ¡¡¡muchas gracias por esos debates científicos intensos en los que tanto aprendo!!! **Maite**, gracias por tu dulzura y cariño, por toda la ayuda que me has prestado, y por esas conversaciones tan agradables. ¡Es un placer trabajar contigo!

Gracias también a toda esa otra gente que forma o ha formado parte del **Departamento de Sanidad Animal** y que de un modo u otro habéis estado ahí durante estos años. Gracias a mis compis de labo y "de sitio": **Julio**, **Paloma,** y sobre todo a **mi Vere:** por las risas, la complicidad y la amistad. Gracias al Dr. **Lucas Domínguez** por el apoyo durante la primera fase de esta Tesis. También gracias a la gente de **VISAVET** que en algún momento me ha prestado su ayuda y su conversación. Gracias a **Bruno y "sus chicos"** por las conversaciones científicas y las "no científicas". Gracias a **Gustavo** y su equipo por su ayuda y su conversación. Gracias a **Espe, Mónica S., Alicia A., Ana D., Anabel, Flora...** por estar ahí, conversando e interesándoos por mí, y por el cariño del día a día. En definitiva, ¡gracias a toda esa gente de la Facul, que de alguna forma me ha acompañado durante estos años!

Quiero dar las gracias de manera especial a "mis niñas de Sanidad Animal" por ese gran equipo de "Lambruscas" que formábamos (y seguimos formando, aunque cada una por su camino...): **Noe, Leti, Natalia,** ¡inolvidables los días en el Departamento y sobre todo nuestras "noches de Ciencia" con comida, bebida y risas hasta reventar! ¡Lo vais a petar las tres en vuestras respectivas Tesis! ¡Os quiero!

Quiero dar las gracias también a toda esa gente de "mis otros lugares de trabajo":

Gracias a todos los que formáis la **Unidad de Bioinformática del ISCIII**, por tratarme tan bien siempre que he estado por allí, tanto a nivel personal como profesional. En especial a **Vicky** y a **Carmen**. Vicky, gracias por ser tan dulce y comprensiva, y por el cariño y el apoyo a todos los niveles que me has dado desde que aparecí por allí. Carmen, gracias por las risas y los buenos momentos como compañeras de mesa y de labo. ¡Gracias a las dos por vuestra amistad!

Gracias a toda la gente del grupo de **Enfermedades Infecciosas e Ictiopatología del IUSA de la ULPGC**. Gracias a **Félix** y a **Dani**, por permitirme formar parte del equipo durante unos meses, y por todo lo que pude aprender con vosotros. Por supuesto gracias a "las niñas de mi isla": **Fátima, Judit, Lore, Jimena, Valen** y **Lita**, por los buenos ratos en el laboratorio y fuera de él, por las risas, y sobre todo por esa gran amistad que me disteis y me seguís dando. También gracias a esa otra gente del IUSA que me ayudó, apoyó y me aportó muy buenos ratos por allí: Ana Sofía, Orestes, Begoña... ¡Gracias de corazón! ¡Entre todos hicísteis que mi estancia en Gran Canaria fuera maravillosa e inolvidable!

Gracias también a toda la gente de la "**Medical Informatics Research Unit**" de la **Universidad de Melbourne**, por hacerme más fácil mi genial estancia en Australia, y por todo lo aprendido durante mi tiempo en "el otro lado del Mundo". En especial mil gracias a **Fernando**, porque desde "el principio de los tiempos" (cuando me presenté hace años en el ISCIII pidiendo ayuda) me has ofrecido sin dudar apoyo y ayuda incondicionales.

No quiero olvidar a toda esa gente que forma parte de mi vida personal y que tan importante ha sido durante estos años, muchas veces sin comprender muy bien en qué consistía eso de "la Tesis", pero siempre

mostrando una confianza en mí y un apoyo inquebrantables ¡sin vosotros nunca lo habría conseguido!

Gracias a esas dos personas tan importantes en mi vida, sin las cuales nunca hubiera llegado hasta aquí, que confían siempre en todas mis decisiones y que me apoyan en todo lo que hago: **papá, mamá**... ¡No hay palabras! ¡Os lo debo todo! **¡Os quiero tanto!**

Gracias a mi compañero en la vida, mi otra mitad, la persona que siempre está ahí, me apoya y me mima. Hemos evolucionado con esta Tesis y por eso tenerte a mi lado al final de esta etapa es aún más especial. Gracias por esas clases magistrales de motivación en los momentos de desesperación, no sabes lo importantes que han sido para poder llegar hasta aquí... **¡Te quiero Dani!**

Gracias al resto de **mi súper familia** por el apoyo y el cariño que me ofrecéis siempre. Y gracias a esa otra gran familia que he elegido (y muy bien, por cierto) que sois **mis amigos**. A los de **Aranda** (especialmente ¡gracias **Isa!** porque tú sí que entiendes que el *Lactococcus* está de moda), a **los de Madrid** (especialmente ¡gracias **Sari, Carola, Silvi y Jose!** por preocuparos siempre por mí y por la evolución de esta Tesis), a **los de Burgos**, a **los de Medina**, y a los repartidos por el resto de España y del extranjero...

En definitiva, a toda esa gente que forma o ha formado parte de mi vida en algún momento... **¡¡¡GRACIAS!!! ¡Qué suerte tengo de teneros!**

Justificación

La presente Tesis Doctoral está constituida por un compendio de trabajos de investigación publicados durante el desarrollo de la misma en diversas revistas científicas o libros de carácter internacional. A continuación se presentan las referencias bibliográficas de cada una de las publicaciones que conforman esta Tesis:

1. **Aguado-Urda M**, López-Campos GH, Fernández-Garayzábal JF, Martín-Sánchez F, Gibello A, Domínguez L, Blanco MM (**2010**) Analysis of the genome content of *Lactococcus garvieae* by genomic interspecies microarray hybridization. **BMC Microbiol** 10: 79.

2. **Aguado-urda M**, Cutuli MT, Blanco MM, Aspiroz C, Tejedor JL, Fernández-Garayzábal JF, Gibello A (**2010**) Utilization of lactose and presence of the phospho-β -galactosidase (*lacG)* gene in *Lactococcus garvieae* isolates from different sources. **Int Microbiol**: 189–193.

3. **Aguado-Urda M**, López-Campos GH, Blanco MM, Fernández-Garayzábal JF, Cutuli MT, Aspiroz C, López-Alonso V, Gibello A (**2011**) Genome sequence of *Lactococcus garvieae* 21881, isolated in a case of human septicemia. **J Bacteriol** 193: 4033–4034.

4. **Aguado-Urda M**, López-Campos GH, Gibello A, Cutuli MT, López-Alonso V, Fernández-Garayzábal JF, Blanco MM (**2011**) Genome sequence of *Lactococcus garvieae* 8831, isolated from rainbow trout lactococcosis outbreaks in Spain. **J Bacteriol** 193: 4263–4264.

5. **Aguado-Urda M**, Gibello A, Blanco MM, López-Campos GH, Cutuli MT, Fernández-Garayzábal JF (**2012**) Characterization of plasmids in a human clinical strain of *Lactococcus garvieae*. **PLoS One** 7: e40119.

6. **Aguado-Urda M**. DNA Microarrays: Principles and Technologies. En: López-Campos G, Martínez-Suárez JV, **Aguado-Urda M**, López-Alonso V (**2012**) Microarray detection and characterization of bacterial foodborne pathogens. **Springer Briefs in Food, Health and Nutrition Series**. Springer. Nueva York.

7. **Aguado-Urda M**, Gibello A, Blanco MM, Fernández-Garayzábal JF, López-Alonso V, López-Campos GH (**2013**) Global transcriptome analysis of *Lactococcus garvieae* strains in response to temperature. **PLoS One** 8: e79692.

Resumen

Lactococcus garvieae es el agente etiológico de la lactococosis, una enfermedad con gran repercusión en acuicultura. Además, *L. garvieae* es una bacteria muy ubicua capaz de producir enfermedad en un amplio rango de hospedadores incluido el ser humano, considerándose actualmente como un agente potencialmente zoonótico. A pesar de la importancia creciente de este microorganismo, el conocimiento sobre las características y las propiedades de su genoma es limitado. Con el fin de aportar datos que contribuyan a un conocimiento más profundo sobre la biología y los mecanismos de patogenicidad de *L. garvieae,* se han planteado como objetivos generales de este trabajo de Tesis Doctoral, el estudio del contenido genético de esta bacteria y un análisis funcional del mismo.

Se realizó una primera aproximación al conocimiento del contenido genético de *L. garvieae* mediante hibridación genómica comparativa (CGH) inter-especie sobre microarrays de ADN, consiguiendo identificar por primera vez un gran número de genes en este patógeno. Posteriormente, la accesibilidad de las técnicas de ultrasecuenciación permitió la caracterización de los genomas de dos cepas clínicas de *L. garvieae* procedentes de humano (Lg21881) y de trucha arcoíris (Lg8831). La disponibilidad de los genomas de Lg21881 y Lg8831, así como de los de varias cepas más de *L. garvieae*, ha hecho posible realizar estudios de genómica comparativa intra-especie que han llevado a la identificación y caracterización molecular de elementos extracromosómicos únicos de la cepa Lg21881, que podrían estar implicados en su adaptación a un nicho ecológico específico. Además, los análisis de genómica comparativa han evidenciado una elevada variabilidad intra-especie en *L. garvieae* y han posibilitado una primera aproximación hacia la filogenia y evolución de este microorganismo. Por otra parte, y en relación con esta variabilidad intra-especie, se ha evaluado la utilidad de determinados caracteres propuestos como biomarcadores para la clasificación de aislados de *L. garvieae* en función de su origen.

El conocimiento del genoma de *L. garvieae* también ha permitido diseñar un microarray de ADN que demostró ser eficaz en el análisis de la expresión global en esta bacteria. Así, sobre dicho microarray se han realizado estudios funcionales de las cepas Lg21881 y Lg8831 en los que se han analizado los cambios de expresión global en función de la temperatura. Se ha observado que Lg21881 y Lg8831 presentan diferentes patrones de expresión, y se han detectado genes potencialmente implicados en su patogenicidad.

En conclusión, los resultados obtenidos en este trabajo aportan nuevos conocimientos sobre esta bacteria de importancia creciente en medicina veterinaria y humana, y abren camino hacia una comprensión más profunda de su biología y su evolución.

Summary

Lactococcus garvieae is the etiological agent of lactococcosis, a disease causing important economic losses in aquaculture. Moreover, *L. garvieae* is a ubiquitous bacterium able to infect a wide range of hosts, including humans, being considered a potential zoonotic agent. However, the knowledge about the characteristics and functionality of the genomic content of this important pathogen is still limited. Aiming to expand the existing knowledge about the biology and the pathogenicity mechanisms of *L. garvieae*, the main objectives of this Doctoral Thesis were the studio of the genomic content of this microorganism and its functional analysis.

An initial approach to analyze the genomic content of *L. garvieae* was carried out by means of genomic inter-species microarray hybridization. In these experiments, 267 genes were described for the first time in this bacterium. After this, next-generation sequencing techniques become accessible, enabling the genome sequencing of two clinical strains of *L. garvieae* isolated from human (Lg21881) and rainbow trout (Lg8831), respectively. Thanks to the availability of the Lg21881 and Lg8831 genome sequences altogheter with the publication of the genomes of other *L. garvieae* strains, it was possible to perform intra-species comparative genomic analyses. Such studies achieved the identification of extracromosomal material unique to Lg21881, which could be implicated in the adaptation to a specific ecological niche. Furthermore, comparative genomics evidenced high intra-species heterogeneity and allowed to obtain preliminary phylogenetic and evolutionary results on this species. Additionally, the usefulness of previously proposed biomarkers based on genotypic and phenotypic characteristics, which have been proposed as indicative of the origin of *L. garvieae* isolates, was assessed.

The knowledge of the genome sequence of *L. garvieae* also allowed the design and manufacturing of a specific DNA microarray, which performed successfully in the global gene expression analysis of this bacterium. Thus, transcriptomic studies were performed with the strains Lg21881 and Lg8831 grown at different temperatures. These global studies led to the identification of differential expression patterns, as well as genes potentially linked to the pathogenesis of *L. garvieae.*

In conclusion, the results of this work contribute to increase the knowledge on this bacterium, opening the possibility of a deep understanding of the biology and evolution of this pathogen with relevance in veterinary and human medicine.

Índice

CAPÍTULO I:

Antecedentes y Objetivos

1. *Lactococcus garvieae*

1.1. Generalidades

Lactococcus garvieae es una bacteria clasificada dentro de la familia *Streptococcaceae,* que fue aislada por primera vez en Reino Unido como agente causal de mastitis bovina (Garvie *et al*., 1981). Dos años más tarde, y de acuerdo con sus características morfológicas y bioquímicas, esta bacteria fue descrita como una nueva especie denominada *Streptococcus garvieae* (Collins *et al*., 1983). Tras el establecimiento del género *Lactococcus* en 1985, en el que se incluían los microorganismos inicialmente clasificados como estreptococos lácticos, la especie pasó a denominarse *Lactococcus garvieae* (Schleifer *et al*., 1985). Posteriormente se propuso una nueva especie denominada *Enterococcus seriolicida*, en la que se englobaban aislados de cocos Gram positivos responsables de importantes pérdidas económicas en la acuicultura japonesa (Kusuda *et al*., 1991). Más tarde, las similitudes encontradas tanto a nivel fenotípico como molecular entre esta nueva especie y aislados de *L. garvieae*, llevaron a concluir de manera independiente que *E. seriolicida* debía ser reclasificada como *L. garvieae* (Doménech *et al*., 1993, Eldar *et al*., 1996, Teixeira *et al*., 1996).

Las bacterias del género *Lactococcus* forman parte del grupo de microorganismos caracterizados como cocos Gram positivos, catalasa negativos y anaerobios facultativos. Son bacterias muy ubicuas, siendo sus hábitats más comunes los vegetales, la piel de los animales y la leche y los productos lácteos. Los lactococos se habían considerado únicamente como patógenos oportunistas hasta los años 90 (Aguirre y Collins, 1993). Sin embargo, su interés tanto en medicina humana como veterinaria, ha cobrado especial importancia en los últimos años, siendo actualmente considerados patógenos emergentes. De las siete especies reconocidas en el género *Lactococcus*, tres de ellas, *Lactococcus piscium, L. garvieae* y *Lactoccocus lactis* se consideran patógenas para el hombre y/o los animales (Aguirre y Collins, 1993; Facklam y Elliot, 1995; Bercovier *et al*., 1997; Goyache *et al*., 2001; Antolín *et al.,* 2004; Vendrell *et al*., 2006; Russo *et al*., 2012).

Desde un punto de vista morfológico, las células de *L. garvieae* son cocos Gram positivos que se disponen generalmente formando parejas o cadenas cortas (Figura 1). Es una bacteria anaerobia facultativa, no formadora de esporas e inmóvil. Puede crecer en un rango de temperaturas muy amplio, entre 10 y 45°C, siendo 37°C la temperatura óptima. Crece en un variado rango de medios de cultivo, tales como Agar Infusión Cerebro Corazón,

Agar Tripticasa Soja y Agar Sangre, entre otros. Es capaz de crecer en medios con pH entre 4,5 y 9,6 y que contengan concentraciones de NaCl de hasta un 6,5%. En medio sólido las colonias que forma son esféricas, de superficie lisa y coloración blanquecina, produciendo α-hemólisis en Agar Sangre (Vendrell *et al*., 2006) (Figura 1).

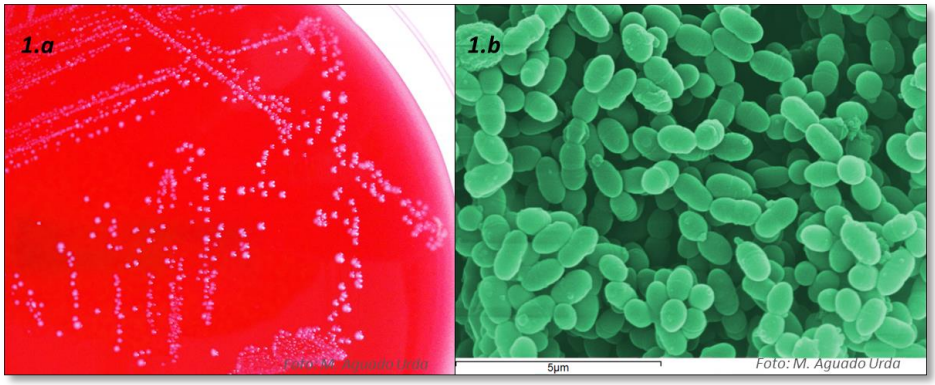


Figura 1. **1a**. Crecimiento de *L. garvieae* en Agar Sangre. **1b.** Fotografía de células de *L. garvieae* tomada con microscopio electrónico.

1.2. *L. garvieae* como agente patógeno

1.2.1. Patógeno de peces

L. garvieae es el agente etiológico de la lactococosis, un proceso septicémico que puede afectar a numerosas especies de peces, tanto de agua dulce como salada. La enfermedad es particularmente importante en el cultivo de la trucha arcoíris (*Oncorhynchus mykiss*), y con una elevada incidencia en países como Australia, Sudáfrica, Japón, Taiwán, Inglaterra y países del área mediterránea (Vendrell *et al*., 2006).

Los primeros brotes de lactococosis se describieron en Japón en el año 1974 (Kusuda *et al*., 1976), y desde entonces se han ido detectando paulatinamente en el resto del mundo. En España, los primeros brotes de lactococosis se detectaron en piscifactorías de trucha arcoíris a finales de los años 80 (Palacios *et al.,* 1988; Palacios *et al*., 1993). La lactococosis puede causar importantes pérdidas en el sector acuícola debido a la elevada mortalidad (superior al 50%), la disminución de la tasa de crecimiento y la imposibilidad de comercializar los peces afectados (Vendrell *et al*., 2006).

La lactococosis de la trucha es una enfermedad hemorrágica hiperaguda que suele cursar con cuadros septicémicos de elevada mortalidad. El cuadro patológico comienza con la rápida aparición de anorexia, letargia, pérdida de la orientación y natación errática. Los signos externos típicos de la enfermedad incluyen exoftalmia unilateral o bilateral, melanosis, presencia de hemorragias en las áreas periorbital e intraocular, bucal y en el opérculo, así como distensión abdominal y prolapso anal. A la necropsia se observan hemorragias en hígado y vejiga natatoria, edema meningoencefálico, pericarditis, esplenomegalia y una enteritis hemorrágica muy acusada (Eldar y Ghittino, 1999; Vendrell *et al*., 2006) (Figura 2).

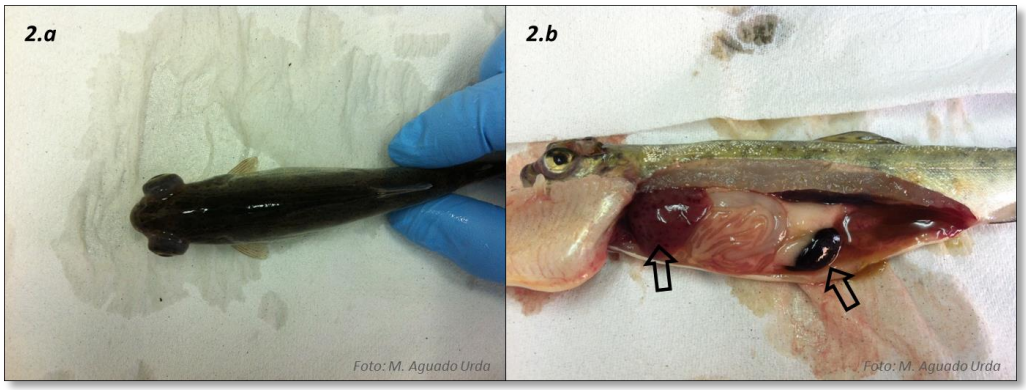


Figura 2. Lactococosis en la trucha arcoíris. **2a**. Exoftalmia bilateral. **2b.** Hemorragias en hígado y esplenomegalia.

La introducción de nuevos lotes de animales, huevos o gametos en una piscifactoría suele ser la vía de entrada más frecuente de *L. garvieae*. Los portadores asintomáticos son la principal fuente de infección debido a que, aunque no padecen la enfermedad, son portadores de la bacteria en su microbiota intestinal y la eliminan por las heces, infectando al resto de la población sana. Así, la infección se mantiene latente y el proceso se desencadena cuando las condiciones ambientales son óptimas para la proliferación de la bacteria (Ghittino y Múzquiz, 1998). La transmisión de la enfermedad se produce principalmente por mecanismos horizontales, siendo la vía fecal-oral la forma más frecuente de transmisión entre peces sanos y enfermos o portadores asintomáticos que eliminan el agente a través de las heces o del mucus (Vendrell *et al.*, 2006).

La temperatura y la calidad del agua son los principales factores condicionantes en la aparición de la enfermedad. La lactococosis presenta una clara estacionalidad, asociándose

a temperaturas del agua elevadas, y apareciendo los brotes más agudos, con elevadas tasas de mortalidad, cuando la temperatura del agua es superior a 18°C (Vendrell *et al*., 2006). Por otra parte, la evolución del proceso se ve favorecida por una mala calidad del agua asociada con alteraciones en las condiciones higiénico-sanitarias de la explotación y cuando el agua es deficiente en oxígeno (Ghittino y Múzquiz, 1998). Para evitar la aparición de brotes de lactococosis es importante un adecuado manejo de los animales, disminuyendo al mínimo la manipulación de los peces, eliminando los peces muertos o enfermos y manteniendo densidades bajas de cultivo. El control de la calidad microbiológica y las características físico-químicas del agua son factores esenciales para evitar la proliferación del microorganismo. Así, son fundamentales la limpieza y desinfección periódica de los estanques y de todos los utensilios de la piscifactoría, utilizando productos como sulfato de cobre, formol, amonios cuaternarios, cloramina T o agua oxigenada (Vendrell *et al*., 2006). Así mismo, es importante el control microbiológico de los productos destinados a la alimentación de los peces, ya que *L. garvieae* ha sido detectado en harinas de pescado destinadas a la producción de piensos para acuicultura (Tejedor, 2008).

El diagnóstico de la lactococosis se realiza en un principio mediante la observación de los signos clínicos típicos de la enfermedad. Posteriormente, el análisis microbiológico de los tejidos diana del pez (riñón, cerebro, hígado y bazo; y en menor medida ojos, intestino y sangre), permitirá confirmar el diagnóstico (Ghittino y Múzquiz, 1998). Para la detección de la bacteria en procesos subclínicos, el diagnóstico microbiológico clásico puede resultar insuficiente debido las bajas concentraciones de la bacteria en los tejidos, por lo que en la actualidad, los sistemas de diagnóstico moleculares basados en la reacción en cadena de la polimerasa (PCR) son los más empleados (Zlotkin *et al*., 1998). En los últimos años se han diseñado métodos de PCR múltiple que permiten la detección específica y simultánea de *L. garvieae* y otros patógenos responsables de enfermedades infecciosas en peces (Mata *et al*., 2004; Altinok, 2011), así como un sistema de PCR a tiempo real para la detección y cuantificación de *L. garvieae* con elevada sensibilidad y especificidad (Jung *et al*., 2010).

El tratamiento de la lactococosis se ha basado en la utilización de antibióticos como la oxitetraciclina, eritromicina, doxiciclina y amoxicilina (Munday, 1994). Sin embargo, a pesar de que numerosos antibióticos son efectivos contra la bacteria en condiciones de laboratorio, su uso en la práctica ha demostrado ser poco efectivo, probablemente debido a la rápida anorexia con la que cursa la enfermedad y a la aparición de resistencias en las bacterias (Bercovier *et al*., 1997; Vendrell *et al*., 2006). Por ello, la mejor forma de controlar la enfermedad es la profilaxis. Así, se han desarrollado algunas estrategias de vacunación para el control de la lactococosis. Los primeros ensayos consistieron en el desarrollo de

autovacunas que implicaban la inoculación intraperitoneal de células de cepas de *L. garvieae* (aisladas en la misma piscifactoría donde se desarrollaba el brote) inactivadas con formalina y calor, que conferían una protección variable dependiendo de la especie del pez, la cepa bacteriana, la formulación de la vacuna y la utilización o no de adyuvante (Prieta *et al*., 1993; Bercovier *et al*., 1997; Ravelo *et al*., 2006). Actualmente existen varias vacunas comerciales que representan una buena forma de prevención de la enfermedad, y que consisten en la aplicación de una bacterina por vía intraperitoneal a los peces con un peso mínimo de 50 gramos y a una temperatura del agua entre 12 y 14°C (Ghittino y Múzquiz, 1998, Brudeseth *et al*., 2013). La vacunación intraperitoneal supone un incremento en los gastos de producción, así como un estrés para los peces derivado de la manipulación que conlleva. Las vacunas orales representan una alternativa a la vacunación intraperitoneal, sin embargo, las vacunas orales desarrolladas hasta el momento frente a la lactococosis no garantizan una correcta inmunización (Romalde *et al*., 2004).

Otra alternativa que se está estudiando en los últimos años para la prevención de la lactococosis es la utilización de probióticos. La acción de los probióticos se fundamenta en la colonización rápida y focalizada del intestino del animal compitiendo con otras bacterias por nutrientes esenciales, en la producción de sustancias antimicrobianas como bacteriocinas y en el estímulo de la inmunidad innata, limitando así las capacidades de adherencia y colonización de los patógenos en el hospedador (Balcázar *et al*., 2007). Se ha demostrado que la administración como suplemento alimenticio de cepas de microorganismos como *Aeromonas sobria*, *Leuconostoc mesenteroides* y *Lactobacillus plantarum* aisladas del tracto digestivo de salmónidos sanos provocaba un descenso acusado de la mortalidad por lactococosis en los peces tratados (Brunt *et al*., 2005, Vendrell *et al*., 2008).

1.2.2. Patógeno humano y agente potencialmente zoonótico

Aunque la mayor importancia clínica de *L. garvieae* reside en su incidencia en la acuicultura, desde su descripción en 1983 como agente productor de mastitis bovina (Garvie *et al*., 1981; Collins *et al*., 1983), esta bacteria ha sido relacionada con varios casos de infecciones intra-mamarias en ganado bovino (Texeira *et al*., 1996; Carvalho *et al*., 1997; Devriese *et al*., 1999; Vela *et al*., 2000; Wyder *et al*., 2011; Tejedor *et al*., 2011). Además, se ha descrito el aislamiento de *L. garvieae* en diversas muestras clínicas de otros mamíferos: tonsilas de perros y vaca, tejido conjuntivo de tortuga (Pot *et al.*, 1996), músculo y riñón de delfín mular (Evans *et al.*, 2006), y de procesos neumónicos y pericarditis en cerdos (Tejedor *et al.*, 2011). La importancia de *L. garvieae* como agente patógeno humano está

aumentando en los últimos años debido al creciente número de casos en los que se aísla e identifica esta bacteria. Es muy probable que la incidencia real de infecciones por *L. garvieae* en humanos haya sido infravalorada, debido a las dificultades para distinguir los géneros *Lactococcus* y *Enterococcus* mediante las técnicas bioquímicas de identificación bacteriana de uso rutinario en hospitales, siendo necesarias técnicas más precisas para su correcta identificación (Aspiroz *et al*., 2007; Russo *et al.,* 2012). Algunos hospitales han incorporado ya en sus laboratorios nuevos paneles de identificación que incluyen a *L. garvieae*, como los nuevos paneles *Vitek2 System* (bioMérieux). No obstante, es conveniente que los resultados de identificación sean corroborados mediante técnicas moleculares como PCR y secuenciación del gen del ARN ribosómico 16S (Zlotkin *et al.,* 1998) o por MALDI-148-TOF-MS (Navas *et al*., 2013).

Los primeros casos descritos en la literatura de infecciones en humanos causadas por *L. garvieae* fueron referidos a principios de los años 90, describiendo a *L. garvieae* como agente causal de varios casos de endocarditis (Furutan *et al.*, 1991). Se han descrito diferentes tipos de infecciones causadas por *L. garvieae* en humanos hasta el momento, siendo la endocarditis la más frecuente con más de la mitad de los casos referidos en total (Tabla 1). *L. garvieae* también ha sido descrito como agente causal de casos de septicemia, peritonitis, absceso hepático, osteomielitis, espondilodiscitis, infección de prótesis de cadera y colecistitis acalculosa (Tabla 1). Actualmente, *L. garvieae* es considerado como un patógeno oportunista en humanos, definiéndose como principales factores de riesgo ante la infección: la edad avanzada, las enfermedades cardiovasculares y/o la colocación de válvulas prostéticas, las alteraciones en el tracto gastrointestinal, y los estados de inmunosupresión (Russo *et al*., 2012). Sin embargo, la capacidad de infectar y causar procesos clínicos en individuos inmunocompetentes ha quedado demostrada en algunos de los casos clínicos descritos (Li *et al*., 2008; Chan *et al*., 2011; Wang *et al*., 2007; Watanabe *et al*., 2011). El aumento de casos clínicos en el hombre asociados a infecciones por *L. garvieae* es especialmente significativo en los últimos años, como muestra el hecho de que el 50 % de los casos descritos se hayan publicado entre los años 2011-2013 (Figura 3).

Tabla 1. Casos de infección por *L. garvieae* en humanos descritos en la literatura desde 1991.

Tipo de infección	Referencias	Factores de riesgo/Comorbilidades	Fuente de infección hipotética
Endocarditis	Furutan *et al*., 1991	Información no disponible	
	Fefer *et al*., 1998	Edad avanzada, cardiopatías, marcapasos, dos prótesis de válvulas cardíacas, tratamiento con antiácidos, hipotiroidismo	
	Fihman *et al*., 2006	Edad avanzada, dos prótesis de válvulas cardíacas, úlcera duodenal	
	Vinh *et al*., 2006	Edad avanzada, enfermedad coronaria, cirugía de eliminación de pólipos malignos en el colon	
	Yiu *et al*., 2007	Enfermedad reumática del corazón, endocarditis infecciosa previa	Consumo de pescado
	Wang *et al*., 2007	Edad avanzada, prolapso de válvula mitral, úlcera gástrica, piedras en riñón	Consumo de pescado crudo
	Li *et al*., 2008	Ninguno	Contacto con pescado (chef)
	Zuily *et al*., 2011	Prótesis válvula cardíaca, marcapasos, pólipos en el colon, cirrosis	Consumo de marisco
	Wilbring *et al*., 2011	Prótesis válvula cardíaca, periodontitis crónica	Contacto con pescado (trabajador de piscifactoría)
	Watanabe *et al*., 2011	Ninguno	
	Hirakawa *et al*., 2011	Prótesis válvula cardíaca, hipertensión, diabetes, dislipemia	Contacto con pescado (perforación gingival con espina)
	Fleming *et al*., 2012	Linfoma, diabetes, prótesis válvula cardíaca	Consumo de pescado
	Russo *et al*., 2012	Tres prótesis de válvulas cardíacas, diverticulosis	Consumo de alimentos contaminados
	Navas *et al*., 2013	Prótesis válvula cardíaca, diabetes, EPOC	
Septicemia	Wang *et al.,* 2007	Estenosis esofágica, interposición del colon	Consumo de pescado
	Wang *et al.,* 2007	Diverticulosis, asma, hipotiroidismo, hipertensión arterial	
	Aspiroz *et al*., 2007	Edad avanzada, pancreatitis	
	Nadrah *et al.,* 2011	Edad avanzada, tres prótesis en válvulas cardíacas, marcapasos, enfermedad diverticular, anemia, esofagitis, hipotiroidismo	
	Fog-Moller y Andersen, 2012	Información no disponible	
Peritonitis	Wang *et al*., 2007	Perforación intestinal	Consumo de pescado crudo
	Chao *et al*., 2013	Nefropatía, diálisis peritoneal	Consumo de alimentos contaminados
Absceso hepático	Mofredj *et al*., 2000	Edad avanzada, carcinoma, prótesis biliar	
Osteomielitis	James *et al*., 2000	Dos prótesis de válvulas cardíacas	
Espondilodiscitis	Chan *et al*., 2011	Gastritis	
Infección en prótesis de cadera	Aubin *et al*., 2011	Obesidad, cardiomiopatía isquémica, hemocromatosis, hipertensión arterial, diabetes, alcoholismo	Contacto con pescado (pescadera)
Colecistitis acalculosa	Kim *et al*.,2013	Hígado graso, ulcera gástrica	Contacto con pescado (pescador)

En los últimos años se ha llegado a relacionar alguno de los citados casos con el contacto con pescado (Li *et al.*, 2008; Aubin *et al*., 2011; Hirakawa *et al*., 2011; Wilbring *et al*., 2011) o con el consumo de alimentos contaminados con *L. garvieae* (Wang *et al*., 2007; Russo *et al*., 2012). Se ha sugerido como probable vía de infección la ingestión de alimentos, especialmente pescado crudo o productos lácteos contaminados con *L. garvieae*, con el posterior paso al torrente sanguíneo de la bacteria debido a alteraciones anatómicas o fisiológicas del tracto gastrointestinal, y desde aquí, la extensión de la infección a órganos distantes donde existen patologías previas (Chan *et al*., 2011). Así, se ha descrito como uno de los factores de riesgo más importantes la coexistencia de alteraciones anatómicas o fisiológicas del tracto gastrointestinal con otros defectos previos locales como alteraciones de las válvulas cardíacas (Chan *et al*., 2011; Russo *et al*., 2012) (Tabla 1).

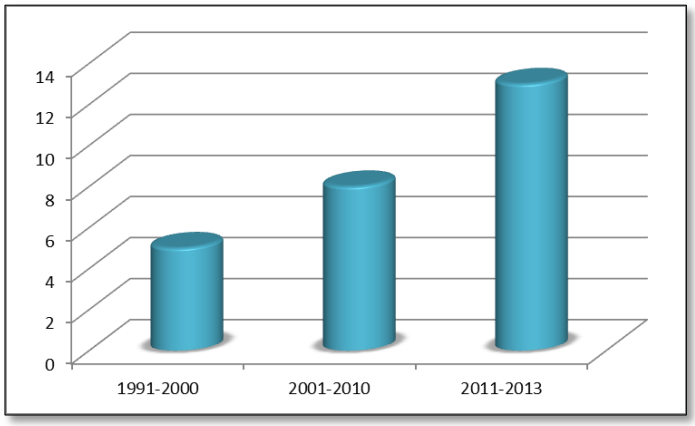


Figura 3. Evolución del número de casos de infección en el hombre por *L. garvieae* descritos en la literatura.

1.2.3. Mecanismos de patogenicidad

A pesar de la creciente importancia de *L. garvieae* tanto en medicina veterinaria como humana, el conocimiento sobre los mecanismos implicados en la patogenicidad de esta bacteria es aún limitado.

Los primeros trabajos dirigidos al estudio de los factores de virulencia de *L. garvieae* se realizaron en los años 80, describiéndose la implicación de posibles toxinas que poseían capacidad para reproducir la sintomatología clínica al inyectarlas en peces sanos (Kimura y Kusuda, 1982). A finales de los años 80 se consiguió identificar una toxina extracelular

similar a la estreptolisina O y que estaría implicada en la actividad hemolítica de la bacteria (Kusuda y Hamaguchi, 1988). Recientemente se han identificado en el genoma de varias cepas de *L. garvieae* genes que codifican hemolisinas que podrían estar implicadas en la septicemia hemorrágica que se produce durante la infección en peces (Reimundo Díaz-Fierros, 2011; Miyauchi *et al*., 2012).

Por otra parte, se ha descrito la presencia de sideróforos en *L. garvieae* que podrían estar implicados en la captación de hierro durante la infección (Schmidtke y Carson, 2003). Los sistemas de captación de hierro son importantes para las bacterias patógenas durante la infección, ya que el hierro es un elemento esencial para su supervivencia y los niveles disponibles de éste son muy imitados en el hospedador (Wandersman y Deleperaire, 2004). Así, se han descrito sistemas eficientes para la captación de hierro en bacterias patógenas de peces, con un papel importante en su virulencia (Magariños *et al*., 1994; Lemos y Osorio 2007; Zou *et al*., 2010).

La presencia de cápsula celular se ha asociado a la virulencia de *L. garvieae*, demostrándose que los serotipos capsulados son más virulentos que los no capsulados (Barnes *et al*., 2002, Ooyama *et al.*, 2002; Kang *et al*., 2004; Kawanishi, *et al*., 2007). La cápsula se considera un factor de virulencia esencial en bacterias, ya que proporciona protección frente a la opsonización y la fagocitosis, procesos implicados en la respuesta inmunitaria frente a las bacterias por parte del hospedador. Un análisis de serotipos capsulados y no capsulados de *L. garvieae* demostró la existencia de un mayor número de antígenos de superficie en los no capsulados, lo cual estaría directamente relacionado una mayor respuesta inmunitaria por parte del hospedador ante el serotipo no capsulado (Shin *et al*., 2007). Recientemente se ha conseguido identificar una isla genómica que contiene los genes necesarios para la síntesis de cápsula en un cepa de *L. garvieae* aislada de un caso clínico de peces, que no estaban presentes en los aislados no patógenos analizados (Miyauchi *et al.,* 2012).

También se ha descrito la presencia de apéndices similares a fimbrias en aislados de *L. garvieae*. Estas bacterias presentaban una cápsula fina, y eran más virulentas que los aislados no capsulados, pero menos que los aislados con una cápsula totalmente desarrollada (Hirokawa *et al*., 2004). Las fimbrias constituyen un factor de virulencia importante en estreptococos patógenos, ya que juegan un papel crucial en la adherencia a los tejidos del hospedador (Soriani y Telford, 2010).

La expresión de factores de adherencia constituye un proceso clave en el inicio de la infección, permitiendo que las bacterias se unan a receptores en los diferentes tejidos del hospedador y promoviendo la colonización de dichos tejidos (Nobbs *et al*., 2009). Se ha propuesto que la propagación y progresión de la infección por *L. garvieae* en trucha tiene como prerrequisito la colonización de la mucosa intestinal (Kusuda y Kawai, 1982; Taniguchi, 1982). En este sentido, algunos estudios realizados sugieren que los glicoesfingolípidos de la membrana plasmática de las células eucariotas actuarían como receptores para *L. garvieae* en los tejidos del hospedador (Shima *et al*., 2006). Además, se ha descrito que esta bacteria posee una elevada capacidad de adhesión a células intestinales, epiteliales y hepáticas de trucha (Vendrell *et al*., 2009). La secuenciación del genoma de varias cepas de *L. garvieae* ha permitido identificar varios genes que codifican proteínas potencialmente implicadas en la adhesión y/o internalización en tejidos del hospedador, como proteínas de unión a mucina, colágeno, fibrinógeno y fibronectina; así como genes que codifican productos proteicos homólogos a sortasas y a proteínas con dominios LPxTG de la familia internalina. Así pues, parece que existen genes que codifican proteínas de adhesión cuya presencia en el genoma, además, es específica en función del origen de la cepa (Reimundo *et al*., 2011a; Miyauchi *et al*., 2012).

Recientemente se han identificado algunos genes implicados en el establecimiento y desarrollo de la infección por *L. garvieae* en trucha mediante la técnica de mutagénesis de marcaje (Menéndez *et al*., 2007). Uno de los hallazgos más importantes de este trabajo fue la identificación del gen *dltA,* perteneciente al operón *dlt* y encargado de incorporar D-alanina a los ácidos teicoicos y lipoteicocicos de la pared celular bacteriana. En estudios posteriores se consiguió caracterizar estructural y funcionalmente el operón *dltABCD* completo y se concluyó el gen *dltA* tiene un papel determinante en la proliferación de *L. garvieae* dentro del hospedador (Reimundo *et al*., 2009).

1.3. Naturaleza ubicua de *L. garvieae*

Además de su implicación en procesos clínicos en diferentes especies animales y el hombre, *L. garvieae* se aísla frecuentemente a partir de numerosas muestras biológicas y ambientales. Así, se ha descrito su presencia en heces de gatos, caballos (Pot *et al.,* 1996) y perros (Ren *et al*., 2011), intestino de aves (Tejedor, 2008; Borrero *et al*., 2011), piel de rana toro (Mendoza *et al*., 2012), suelos (Klijn *et al*., 1995) y aguas de ríos y de depuradoras (Vela *et al*., 2000; Tejedor, 2008), considerándose a *L. garvieae* como una bacteria muy ubicua.

Actualmente, la mayor importancia de *L. garvieae* fuera del contexto clínico reside en su presencia en diversos tipos de alimentos destinados al consumo humano y animal. *L. garvieae* se aísla frecuentemente a partir de leche cruda y derivados lácteos, así como de carne cruda y productos cárnicos procesados de cerdo y ave. También ha sido aislado de otro tipo de alimentos destinados al consumo humano como verduras, y alimentos para animales como productos derivados del pescado, harinas de pescado y alimento fermentado para cerdos (Tabla 2).

Tabla 2. Presencia de *L. garvieae* en alimentos destinados al consumo humano y animal.

Grupo de alimentos	Alimento	Referencias
Leche	Leche cruda de vaca	Klijn *et al*., 1995; Villani *et al*., 2001; Lafarge *et al*., 2004; Fortina *et al.,* 2007; Tejedor, 2008; Delgado *et al*., 2013
	Leche cruda de cabra	Callon *et al.,* 2007
	Leche cruda de oveja	Prodromou *et al.,* 2001
	Leche cruda de yegua	An *et al.,* 2004
Derivados lácteos	Queso de cabra, Grecia	Hatzikamari *et al*., 1999
	Quesos de vaca, España	Flórez y Mayo, 2006; Alegría *et al.,* 2009
	Quesos de vaca, Italia	Fortina *et al.,* 2007; Foschino *et al*., 2008
	Queso de vaca, Egipto	El-Baradei *et al*., 2007
	Leche de búfalo fermentada, Egipto	El-Baradei *et al*., 2008
	Leche fermentada, Serbia	Jokovic *et al*., 2008
	Quesos de oveja, Irán	Edalatian *et al*., 2012
	Leches fermentadas de vaca, oveja y cabra, Jordania	Alrabadi, 2012
Carne y productos cárnicos	Carne de pollo	Barakat *et al*., 2000
	Morcilla de Burgos	Santos *et al*., 2005
	Salchichas de cerdo fermentadas	Rantsiou *et al*., 2005
	Carne de cerdo y pollo	Aquilanti *et al*., 2007
	Salchichas de cerdo fermentadas	Toshukowong *et al*., 2012
	Carne de pavo y salchicha de cerdo	Ricci *et al*., 2013
Vegetales	Rábano	Kawanishi *et al*., 2007
	Brócoli	Kawanishi *et al.,* 2007
Alimento para animales	Derivados de pescado para alimento de peces	Yasunaga *et al*., 1982
	Harinas de pescado	Tejedor, 2008
	Alimento fermentado para cerdos	Olstorpe *et al*., 2008

El interés de *L. garvieae* en el contexto de la industria alimentaria está aumentando en los últimos años al observarse su participación en las fermentaciones de productos lácteos y cárnicos, en los que contribuye al desarrollo de las características organolépticas de los productos finales (Rantsiou et al., 2005; Santos *et al*., 2005; Fortina *et al.*, 2007; Alegría *et al*., 2009). En el año 2010 se describió por primera vez a *L. garvieae* como componente de la microbiota intestinal en humanos (Kubota *et al*., 2010), sugiriéndose que este microorganismo podría formar parte de la microbiota intestinal, o de aquellas bacterias de tránsito que se ingieren con los alimentos. La elevada ubicuidad de *L. garvieae* y su frecuente aislamiento a partir de alimentos han llevado a algunos autores a plantearse los riesgos potenciales en materia de seguridad alimentaria que podrían derivarse de la presencia en alimentos de cepas patógenas de esta bacteria (Fortina *et al*., 2007; Foschino *et al*., 2008). Así, se han estudiado las diferencias genéticas entre cepas procedentes de diversos nichos ecológicos con el fin de detectar marcadores genéticos que sean capaces de distinguir entre cepas de diferentes orígenes, lo que podría resultar útil en materia de seguridad alimentaria. En este sentido se han propuesto la capacidad de metabolizar la lactosa y la presencia del gen de la fosfo-beta-galactosidasa (*lacG*) como marcadores para distinguir entre cepas no patógenas de origen láctico y cepas clínicas de peces (Fortina *et al*., 2009).

1.4. Variabilidad genética intra-especie en *L. garvieae*

El desarrollo de técnicas moleculares durante la última década ha permitido el estudio de la estructura de poblaciones microbianas incluidas en una misma especie. El estudio de la variabilidad intra-especie de una bacteria nos acerca a la comprensión de su fisiología y a la reconstrucción de su historia evolutiva, además de permitir profundizar en el conocimiento de los factores de virulencia en el caso de las bacterias patógenas.

Los primeros trabajos enfocados a la caracterización genética de aislados de *L. garvieae* de diferentes orígenes mostraron una elevada diversidad entre los aislados de ambientes terrestres, mientras que entre los aislados acuáticos se observó cierta homogeneidad genética (Vela *et al*., 2000; Kawanishi *et al*., 2006; Tejedor, 2008). En otros estudios se han incluido aislados procedentes de un amplio rango de orígenes y se ha evidenciado una elevada heterogenicidad genética intra-especie en *L. garvieae* (Fortina *et al.,* 2007; Reimundo Díaz-Fierros, 2011; Tejedor *et al*., 2011, Ferrario *et al*., 2012).

En general, se ha comprobado que los aislados clínicos procedentes de peces afectados por lactococosis forman grupos relacionados estrechamente a nivel genético en función de la zona geográfica de procedencia (Ravelo *et al*., 2003; Eyngor *et al*., 2004, Reimundo Díaz-Fierros, 2011; Tejedor, 2008). Sin embargo, fuera del ámbito de la lactococosis en peces (donde la homogeneidad genética estaría relacionada con una elevada adaptación al hospedador), los aislados de orígenes diversos presentan una elevada heterogenicidad genética, característica que se corresponde con la naturaleza ubicua de *L. garvieae* y con el carácter esporádico de las infecciones en otros hospedadores (Tejedor *et al*., 2011).

Por otra parte, las características genéticas únicas de aislados causantes de infecciones en el hombre han sido también objeto de estudio. En el año 2011, mediante la técnica de hibridación sustractiva supresora se consiguieron identificar genes presentes en un aislado humano ausentes en uno de trucha (Reimundo *et al*., 2011a); posteriormente se comprobó que la gran mayoría de estos genes, únicos del aislado humano, estaban también presentes en otra cepa de origen humano, sugiriéndose que los aislados humanos poseen características genéticas únicas que pueden estar relacionadas con la capacidad de causar infección el hombre (Miyauchi *et al*., 2012).

En los últimos trabajos publicados en los que se aborda el estudio de la diversidad genética intra-especie, se han llevado a cabo análisis filogenéticos que han conducido a proponer la existencia de dos líneas genómicas independientes de *L. garvieae* que se habrían separado tempranamente, y que después habrían dado lugar a la aparición de diversos genotipos como consecuencia de la presión selectiva ejercida por los estímulos ambientales relacionados con nichos ecológicos específicos (Ferrario *et al*., 2012).

2. Aproximaciones genómicas para el estudio de la biología bacteriana

Los enormes avances tecnológicos surgidos en el campo de la Biología Molecular durante las últimas décadas han provocado un gran cambio en la experimentación biológica, permitiendo el paso de las aproximaciones "gen a gen", en las que en un experimento se estudiaba un solo gen o un operón, hacia la posibilidad de realizar estudios globales, en los que se analizan miles de genes o incluso genomas completos en un solo experimento. La aplicación masiva de este tipo de tecnologías durante los últimos años hace que se defina la época científica actual como "la era genómica". La explosión en la cantidad de datos generados ha requerido el soporte de la Bioinformática para su correcta explotación, lo que ha llevado a que en la actualidad la Genómica y la Bioinformática estén profundamente entrelazadas.

2.1. Secuenciación de genomas bacterianos

La secuenciación de un genoma por completo es el primer paso en la aplicación de las aproximaciones globales o genómicas. Desde la liberación de la secuencia del genoma humano en el año 2000, la capacidad y la velocidad de generación de datos genómicos han experimentado un aumento exponencial. De forma simultánea, los costes de la tecnología de secuenciación genómica han ido disminuyendo, mientras que la eficacia de esta tecnología es mayor cada día. Las nuevas y altamente eficientes tecnologías de secuenciación, llamadas *Next-Generation Sequencing* (NGS) o secuenciación de última generación, se han convertido en técnicas accesibles y a precios competitivos en el mercado.

El primer genoma microbiano secuenciado fue el de *Haemophilus influenzae,* y en la publicación asociada a estos resultados se describía una nueva técnica de secuenciación genómica llamada *Whole Genome Shotgun* (WGS) *sequencing* (Fleischmann *et al*., 1995). Desde entonces, un gran número de genomas bacterianos han sido secuenciados mediante esta estrategia que, brevemente, consiste en generar lecturas a partir de localizaciones al azar a lo largo de genoma que después se ordenan usando técnicas bioinformáticas en función del solapamiento de fragmentos entre las lecturas. Actualmente, según la base de datos *Genomes Online Database* (GOLD), existen más de 30.000 proyectos de secuenciación genómica registrados, de los cuales el 80% corresponden a genomas bacterianos. Del total

de estos proyectos de secuenciación de bacterias, el 46% son genomas completos, es decir, que ya se han concluido y se dispone de los genomas completamente ensamblados; el 14% son secuencias genómicas incompletas y el 40% restante los representan proyectos que están en progreso de secuenciación (Figura 4).

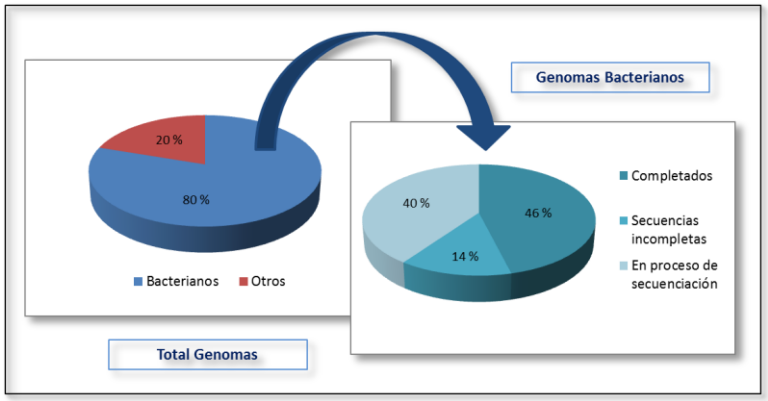


Figura 4. Proyectos de secuenciación genómica registrados hasta noviembre de 2013 según la base de datos GOLD.

La secuenciación del genoma de un microorganismo resulta una herramienta valiosa para la obtención de información fundamental acerca su biología, así como para una comprensión detallada de la evolución microbiana. Además, en el caso de los microorganismos patógenos, puede proporcionar información sobre sus mecanismos de patogenicidad, ya que permite identificar genes cuyos productos tienen un papel en el desarrollo de la infección, lo cual puede dirigir el diseño de estrategias para el control, el diagnóstico y la prevención de las enfermedades que producen. La secuenciación genómica de varias cepas de una misma especie o de especies filogenéticamente cercanas también proporciona información sobre las relaciones evolutivas entre ellas. La disponibilidad de genomas completos y el desarrollo de grandes bases de datos de secuencias, así como de herramientas bioinformáticas para su análisis, han facilitado el desarrollo de la genómica comparativa. La genómica comparativa aplicada al estudio de genomas microbianos, por ejemplo en el caso de una cepa patógena y una no patógena del mismo microorganismo, puede proporcionar datos importantes sobre los elementos genéticos responsables de la infección, así como las posibles presiones evolutivas que pueden haber influido en las diferencias entre dichas cepas (Guzmán *et al*., 2008).

La secuenciación de un genoma bacteriano completo es una tarea multidisciplinar que incluye, en general, los siguientes pasos: clonaje del genoma en fragmentos, obtención de la secuencia de dichos fragmentos, ensamblado de las secuencias obtenidas y anotación y análisis de las mismas mediante el uso de herramientas bioinformáticas.

2.1.1. Obtención de la secuencia genómica: tecnologías de secuenciación de última generación

Los métodos de secuenciación de última generación o de ultrasecuenciación se basan en una "secuenciación masiva paralela", llamada así porque permite secuenciar millones de fragmentos de ADN de forma simultánea. Las tecnologías de ultrasecuenciación que dominan el mercado actualmente se denominan "de segunda generación", para diferenciarlas de la primera generación de métodos de secuenciación basados en el método de Sanger (Sanger *et al*., 1977). Esta tecnología de última generación incluye básicamente los procesos de preparación de la muestra, secuenciado, y obtención y análisis de datos. La preparación de la muestra incluye la ruptura al azar del ADN de interés en fragmentos pequeños, que después son unidos a un soporte como nanoesferas en suspensión o una interfaz sólida, dependiendo de la tecnología utilizada. La inmovilización en espacios separados de los fragmentos molde de ADN permite que se produzcan de manera simultánea entre miles y millones de reacciones de secuenciación mediante un proceso de sucesivas incorporaciones de nucleótidos, lavados y escaneados que recogen los datos de la secuencia. Las diferentes tecnologías disponibles se diferencian según la longitud de las lecturas que producen (Tabla 3) y la tecnología química específica que usan para secuenciar y leer las moléculas de ADN generadas (Luciani *et al*., 2012).

La secuenciación de ADN es un área sometida a rápidos cambios de tecnología y plataformas, las cuales son continuamente actualizadas. Los grandes avances tecnológicos han permitido un gran aumento de la eficacia y velocidad de secuenciación, al mismo tiempo que ha disminuido su coste. Las desventajas de los métodos de secuenciación de última generación se deben a las cortas lecturas que generan y a una elevada tasa de error en determinados casos. Las llamadas tecnologías de tercera generación están ya irrumpiendo en el mercado y prometen una eficacia similar a las actuales, ofreciendo menor coste, lecturas más largas y novedosos métodos de lectura, lo que apunta a una nueva revolución en el campo de la secuenciación genómica (Henson *et al*., 2012).

Tabla 3. Tecnologías de ultrasecuenciación disponibles en la actualidad y comparación de sus principales características.

Plataforma	Generación	Tiempo aprox. por tanda	Longitud de las lecturas (pb)	Cobertura por tanda	Errores típicos
Roche 454 http://www.my454.com/	Segunda	23 h	700 de media (hasta 1000)	700 Mb	Inserciones/ Deleciones (*indels*) en regiones homopoliméricas
Illumina http://www.illumina.com/systems.ilmn	Segunda	9 días	100	600 Gb	Errores de sustitución al final de lecturas
SOLiD http://www.appliedbiosystems.com/absite/us/en/home.html	Segunda	9 días	75	77 Gb	Errores de sustitución
Ion torrent http://www.iontorrent.com/	Tercera	2 h	200	1 Gb	*Indels* en regiones homopoliméricas
Pacific Biosciences http://www.pacificbiosciences.com/	Tercera	1 día	2700 de media (hasta 5000)	90 Mb	*Indels* al azar

A continuación se describen brevemente las tecnologías de secuenciación más utilizadas en la actualidad.

2.1.1.1. Pirosecuenciación 454

El primer equipo de secuenciación de última generación que apareció en el mercado fue el GS20, de 454 Life Sciences (Basel, Suiza) en el año 2005. Esta tecnología se basa en un sistema de pirosecuenciación paralela a gran escala, que consiste en la fijación de fragmentos de ADN, previamente nebulizados y ligados a secuencias adaptadoras, a nanoesferas de unión a ADN que se encuentran inmersas en una emulsión agua-aceite, de manera que cada uno de los fragmentos de ADN fijado a las nanoesferas se amplifica mediante PCR (*emPCR* o PCR en emulsión). La pirosecuenciación se produce después sobre un soporte sólido donde cada nanoesfera queda aislada en un poro, de manera que cada una de ellas genera la lectura de un fragmento (Figura 5). La última versión de este tipo de secuenciación (454 GS FLX Titanium XL+) es capaz de producir lecturas de una longitud media de 700 pares de bases, que pueden llegar hasta 1000. La producción de lecturas relativamente largas tiene grandes ventajas que permiten una gran variedad de aplicaciones, como la secuenciación de genomas completos *de novo*, resecuenciación de genomas completos o de regiones de ADN de interés, metagenómica y análisis de ARN

(Henson *et al.*, 2012). El error característico de esta tecnología es la generación de inserciones o deleciones en las zonas homopoliméricas (Luciani *et al.*, 2012).

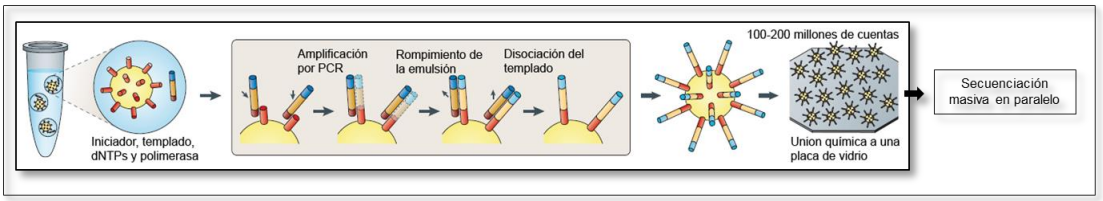


Figura 5. Esquema de funcionamiento de la pirosecuenciación 454. Modificado de Metzker (2010).

2.1.1.2. Tecnología Illumina

La liberación del Analizador Genómico de Illumina (California, USA) en 2007 marcó una verdadera revolución en la secuenciación genómica, permitiendo que la generación de lecturas cortas llegara a tener aplicaciones significativas en este campo. Esta tecnología se basa en la secuenciación de terminación reversible de ciclo. Tras el fraccionamiento del genoma se realiza una amplificación clonal de los fragmentos sobre una superficie sólida, creando grupos de fragmentos clonados. La secuenciación de los millones de racimos de fragmentos de ADN se realiza mediante síntesis empleando nucleótidos marcados que contienen terminaciones reversibles, permitiendo que cada ciclo del proceso de secuenciación ocurra en presencia de los cuatro nucleótidos simultáneamente (Nieva-Velasco, 2012). A diferencia de la pirosecuenciación, la extensión de la molécula de ADN se produce de nucleótido en nucleótido, de manera similar a la secuenciación clásica por el método de Sanger. Es un procedimiento más económico y potente que el 454, pero al generar lecturas de pequeño tamaño es incapaz de resolver repeticiones cortas. Las sustituciones de bases son los errores más típicos de esta tecnología. Debido a su elevada precisión y su bajo coste, esta tecnología se aplica con éxito en resecuenciación, ensamblado *de novo* y análisis y secuenciación de ARN.

2.1.1.3. Tecnología SOLiD

La tecnología SOLiD de Life Technologies (California, USA) realiza la secuenciación mediante ligación. Mediante este proceso una serie de oligonucleótidos de una longitud determinada se marcan de acuerdo a la posición a secuenciar. Los oligonucleótidos son después alineados y ligados, de manera que la preferencia de la enzima ADN ligasa por la

ligación de secuencias idénticas genera una señal que informa sobre la posición de cada nucleótido. La tecnología SOLiD genera más lecturas que sus competidoras, pero con una longitud de lectura menor. Su aplicación más eficiente es la del ensamblado *de novo* (Luciani *et al*., 2012).

2.1.1.4. Tecnologías de secuenciación de tercera generación

Mientras que las tecnologías de segunda generación están basadas típicamente en la obtención de secuencias a partir de la amplificación clonal de fragmentos, los métodos de tercera generación se basan en la lectura de una molécula única de ADN o ARN molde eliminando los pasos intermedios y generando secuencias de mayor tamaño que las anteriores (Henson *et al*., 2012). Algunos ejemplos son: *Ion Torrent* de Life Technologies (http://www.iontorrent.com/) y la plataforma de secuenciación a tiempo real de Pacific Biosciences (http://www.pacificbiosciences.com/). Estas tecnologías están experimentando un gran desarrollo en los últimos años y prometen una secuenciación mucho más rápida y a menor coste.

2.1.2. Ensamblado y anotación de genomas bacterianos

El montaje o ensamblado de secuencias se refiere al alineamiento y mezcla de múltiples fragmentos de ADN para obtener una secuencia mucho mayor que reconstruya la original. El ensamblado de las secuencias continúa siendo uno de los problemas principales de la Bioinformática. Esto se debe, en gran parte, a que los métodos de secuenciación masiva producen en general lecturas de pequeño tamaño, que deben ser ordenadas y unidas para conformar la secuencia del genoma completo (Henson *et al*., 2012).

Como se ha mencionado anteriormente, la estrategia más común para la secuenciación de un genoma completo es WGS (*Whole Genome Shotgun*). Esta aproximación implica la fragmentación al azar del genoma y su clonación para obtener una librería genómica que contiene los fragmentos que representan el genoma completo y que, después de secuenciados, han de ser ensamblados. Cada secuencia corta generada se denomina "lectura" o *read*. La secuencia original es reconstruida a partir de las lecturas mediante el empleo de herramientas bioinformáticas que se denominan "ensambladores", los cuales son capaces de ensamblar las lecturas en función del solapamiento de sus secuencias, formando así los llamados "fragmentos contiguos" o *contigs*. Las secuencias no resueltas que se generan entre *contigs* se denominan "huecos" o *gaps,* y suelen generarse en zonas del genoma que contienen secuencias repetitivas. La existencia previa de un

genoma de referencia que representa a la especie a la que pertenece el organismo problema facilita el proceso de ensamblado. Por el contrario, ante una secuenciación *de novo*, el proceso de ensamblado resulta más complicado. Actualmente existen estrategias aplicables desde el diseño experimental del proceso de secuenciación que mejoran la habilidad de los ensambladores para reconstruir posteriormente la secuencia de ADN original correctamente. Dichas estrategias generan lecturas a partir de fragmentos de ADN que contienen secuencias adaptadoras en los extremos, dando lugar a lecturas apareadas que facilitan el ensamblado y que resultan muy útiles ante una secuenciación *de novo*. En función de la metodología existen dos tipos de estrategias: lecturas de tipo *mate-pairs* y lecturas de tipo *paired-ends.* Gracias a los datos generados mediante ambos tipos se hace posible el ordenamiento de los *contigs,* así como conocer la distancia genómica aproximada entre ellos. Una colección ordenada de *contigs* entre los que existen *gaps* de longitud conocida pero de secuencia desconocida constituye un *scaffold* (Nagarajan y Pop, 2013). Así, se genera un mapa genómico compuesto por una serie de *contigs* colocados en el orden correcto entre los que quedarán *gaps* no resueltos pero de longitud conocida (Figura 6). El último paso para completar la secuencia genómica es el cierre de los *gaps*. Dependiendo del tamaño de los mismos, se pueden usar diferentes técnicas. Si el *gap* es pequeño (5-20 kb), lo más común es el uso de la PCR para amplificar y secuenciar la región en cuestión. Si el fragmento es mayor de 20 kb entonces será necesario clonarlo en un vector y secuenciar posteriormente el vector.

Otro aspecto importante para asegurar un buen proceso de ensamblado es la calidad de las secuencias generadas. Cada tecnología de secuenciación tiene su tasa de error característica, y para asegurar la fiabilidad de cada nucleótido asignado a una posición determinada del genoma es necesario obtener varias veces la información sobre dicha posición. El valor del número medio de veces que cada base del genoma está presente en las lecturas de secuenciación producidas se denomina “profundidad de cobertura” (*depth of coverage*, o simplemente, *coverage*) y es uno de los factores determinantes para evaluar la fiabilidad del ensamblado (Nagarajan y Pop, 2013).

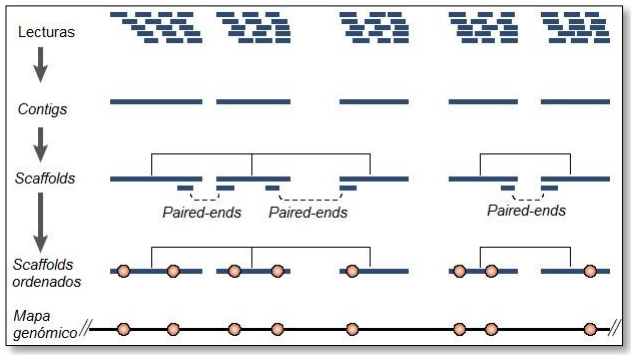


Figura 6. Fases en el proceso de ensamblado en una secuenciación genómica por *shotgun* unido al sistema de *paired-ends*. Modificado de Green (2001).

Finalmente, una vez el genoma de un organismo ha sido secuenciado y ensamblado, la siguiente fase es la anotación del mismo. La anotación se define como el proceso mediante el cual la información estructural y funcional de un genoma es atribuida mediante la identificación de los de genes que contiene, y que se realiza en función de la similitud con secuencias previamente caracterizadas y liberadas en bases de datos públicas. Este es un proceso complejo que requiere la utilización de métodos bioinformáticos basados en diferentes aproximaciones y el acceso a bases de datos, lo que ha provocado el desarrollo de herramientas bioinformáticas que han de ser ejecutadas para poder obtener la máxima información de una secuencia genómica. Sin el desarrollo de la Bioinformática, la secuencia genómica de un organismo sería indescifrable.

El primer paso en la anotación de un genoma es la búsqueda de genes, lo que resulta relativamente sencillo en el caso de un genoma procariota, debido a la elevada densidad de genes y a la ausencia de intrones. Las modernas herramientas bioinformáticas son capaces de realizar automáticamente la búsqueda de genes y a la vez apuntar otros parámetros estructurales del genoma, como la posición de ARNs de transferencia y ARNs ribosomales, así como el origen y el final de la replicación. La predicción de los posibles elementos que conforman el genoma sienta las bases para análisis posteriores, tales como la predicción de la función de cada gen. La predicción funcional se realiza generalmente mediante comparación frente a las bases de datos y búsqueda de genes y proteínas similares de función conocida o predicha. Sin embargo, este proceso puede llevar a errores en la anotación que son muy difíciles de detectar. Además, hay que tener en cuenta que la función de un gran número de genes, que puede llegar hasta una tercera parte en el caso de algunos organismos, es desconocida. Aunque algunos proyectos genómicos microbianos están bajo constante revisión, paralela a la aparición de nuevas herramientas

bioinformáticas de anotación, muchas de las anotaciones genómicas liberadas en las bases de datos nunca han sido revisadas (Guzmán *et al*., 2008).

Una vez conocida la secuencia y la anotación de un genoma, es posible plantear estudios funcionales globales que analicen la expresión de los genes identificados. Una de las técnicas más aplicadas para los estudios funcionales globales durante los últimos años ha sido la tecnología de "micromatrices" de ADN, más conocidas como microarrays de ADN. El diseño y construcción de un microarray de ADN en el que se represente el genoma completo del organismo en cuestión permite el estudio de los elementos constituyentes de un genoma de manera global y paralela, tanto para su caracterización estructural mediante aproximaciones basadas en genómica comparativa, como para el estudio y comprensión funcional del mismo mediante la caracterización de los transcritos generados a partir de dicho genoma. En el siguiente apartado de esta Tesis se explica en qué consiste la tecnología de microarrays de ADN y cuáles son sus aplicaciones.

2.2. Microarrays de ADN: principios y tecnologías

Un microarray o *chip* de ADN se define como una colección de sondas de ADN organizadas en la superficie de un soporte sólido y que pueden ser enfrentadas a ácidos nucleicos diana previamente marcados para generar datos cualitativos (por ejemplo, presencia/ausencia de un gen) o cuantitativos (por ejemplo, niveles de expresión génica). La base de esta tecnología es la hibridación de ácidos nucleicos (ADN o ARN) a gran escala, permitiendo la realización de experimentos a escala genómica en los que se analiza la homología del ADN o la expresión de miles de genes en un solo experimento. Este tipo de abordaje ha provocado profundos cambios durante los últimos años en el campo de la investigación biológica, promoviendo, como ya se ha comentado anteriormente, el paso de las aproximaciones clásicas en las que se estudiaba un solo gen o un operón, hacia estudios globales en los que es posible el estudio del contenido y/o el análisis funcional de genomas completos. El desarrollo de esta tecnología ha sido paralelo al incremento en el número de genomas secuenciados, por lo que los microarrays de ADN se han convertido en una herramienta fundamental en las principales áreas de investigación en Microbiología, como son la fisiología, epidemiología, ecología, patogénesis y filogenia microbianas (Aguado y Blanco, 2007).

En la publicación siguiente se describen los principios y las plataformas disponibles actualmente así como las aplicaciones y la metodología de la tecnología de microarrays de ADN.

Guillermo López-Campos
Joaquín V. Martínez-Suárez
Mónica Aguado-Urda • Victoria López-Alonso

Microarray Detection and Characterization of Bacterial Foodborne Pathogens

Springer

Guillermo López-Campos
Department of Bioinformatics
and Public Health
Spanish National Institute
of Health 'Carlos III'
Majadahonda, Madrid, Spain

Mónica Aguado-Urda
Department of Animal Health
Complutense University
Madrid, Spain

Joaquín V. Martínez-Suárez
Department of Food Technology
Spanish National Institute for Agricultural
and Food Research and Technology (INIA)
Madrid, Spain

Victoria López-Alonso
Department of Bioinformatics
and Public Health
Spanish National Institute
of Health 'Carlos III'
Majadahonda, Madrid, Spain

ISBN 978-1-4614-3249-4 e-ISBN 978-1-4614-3250-0
DOI 10.1007/978-1-4614-3250-0
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2012932853

Chapter 3
DNA Microarrays: Principles and Technologies

Abstract A microarray is defined as a collection of microscopic features densely organized onto the surface of a solid support that can be probed with target molecules chemically labeled to produce either quantitative or qualitative data. For biological applications, the features that make up the array can be DNA, RNA, proteins, polysaccharides, lipids, small organic compounds, or even whole cells. DNA microarray technology is the most popular and well-developed usage of microarrays. It has allowed researchers to perform large-scale quantitative experiments and has contributed to fundamental changes in the way to perform biological research, moving from a gene-by-gene approach to global or genome-wide systemic studies.

Keywords DNA microarray technologies • Microarray methodologies • Microarray applications

3.1 Introduction

The idea of performing biological reactions with one spatially immobilized reagent is not new. In 1975, Edwin Southern described a technique that transformed molecular biology. Later during the 1980s, researchers began to work on a technique in which molecules of a known identity were immobilized on a membrane or slide and the solution to be tested was labeled and hybridized to the surface. Such arrays were used in DNA mapping and sequencing. This work concluded with a publication by Fodor et al. in 1991 about a new technology called Affymax (later Affymetrix). The paper described protein and nucleotide arrays, their production using photolithography, and their applications.

In 1995, Schena et al. introduced the word “microarray” for the first time. Initially, the technique was greatly limited, but the amount of sequencing data began to grow rapidly and the power of microarrays increased tremendously. By the mid-1990s, scientists had extended microarray capabilities by using them to study gene expression levels of hundreds of genes simultaneously, and in the process established microarrays as a viable and flexible molecular biology tool. The twenty first century has witnessed an explosion in microarray-based publications, as researchers find more and more uses for the technology. It is difficult to find an issue of a contemporary biology journal that does not mention microarray technology (Wheelan et al. 2008).

3.2 DNA Microarray Principles and Technologies

The fundamental basis of DNA microarray technology involves the parallel hybridization of a mixture of nucleic acids (targets) with thousands of individual nucleic acids species (probes). Probes have a specific location on the array (spot or feature) so they can be identified by their spatial position in a single experiment. The probe sequences are immobilized in a tightly packed manner so that it is possible to place many different probes on a single small surface. While the probes are immobilized, the targets are deposited as a solution onto the array in order to hybridize. The target

sample is usually labeled with a fluorescent dye that can be detected by a light scanner that scans the surface of the microarray. The key to microarray technology is that a probe is detected at a level proportional to the amount of its target present in the labeled extract.

DNA microarrays have been widely used for gene expression profiling and genotype analysis. Observing all the microarray spots at the same time gives the profile of a sample; in the most common application of DNA microarrays, the target sample is mRNA and the total microarray image represents the transcriptional profile of the sample (de Rinaldis and Lahm 2006). Although mRNA expression profiling is the dominant application for DNA microarrays, they are also used in genomic DNA analyses such as the detection of alternatively spliced variants, the epigenetic status of the genome, DNA copy number changes and sequence polymorphisms, or to detect DNA–protein interactions. The potential applications include the analysis and characterization of any reaction product composed of nucleotide sequences (Shiu and Borevitz 2008). Due to advances in manufacture, robotics, and bioinformatics, microarray technology has continued to improve in terms of efficiency, reproducibility, sensitivity, and specificity. These improvements have allowed microarrays to transition from strictly the research setting to clinical diagnostic applications (Miller and Tang 2009). In addition to the ability to examine a large number of genes in parallel, the success of microarrays can be attributed to the versatility and flexibility of array designs. Currently, many microarray platforms are available, and custom array designs are possible and relatively cost-efficient. DNA microarray technology can be classified according to the method used for the deposition of the probe sequences onto the solid surface. This can be done by presynthesis (printed microarrays) or in situ synthesis (in situ-synthesized oligonucleotide microarrays). The solid surface used varies depending on the type of microarray employed.

There are downsides to microarray technology. Microarrays are still expensive, and more importantly, microarray experiments generate enormous amounts of complex data that are not analyzed easily. There is currently no consensus about how many repetitions need to be done, although the number fortunately appears to be low. The cost of arrays is independent of the cost of the reagents (Cy3 and Cy5 are expensive). The array scanner and workstation, the time and effort of a technician
and an analyst, and of course the cost associated with the time spent doing the experiment in the first place must all be added to the cost of microarray experimentation (Wheelan et al. 2008).

3.2.1 Printed DNA Microarrays

Arrays based on presynthesis of the probes are known as spotted or printed arrays because the probes are “printed” or spotted onto the microarray surface. In this technology, a robot spotter is used to place small quantities of a probe in solution on the microarray surface (commonly a glass microscope slide). Glass slides are an attractive medium for microarrays because they are economical, stable during high temperatures and stringent washes, nonporous (allowing for efficient kinetics during hybridization), and have minimal background fluorescence (de Rinaldis and Lahm 2006).

The probe spots can be applied by either noncontact or contact printing. A noncontact printer shoots small droplets of probe solution onto the glass slide. In contact printing, each print pin directly applies the probe solution onto the microarray surface. The result is the application of probe solution to create spots of 100–150 mm. Control for cross-contamination and consistency during the printing process is crucial to preserve the reliability of the microarray and consequent hybridization

data. Due to the relatively large size of the features, printed microarrays are of lower density (10,000–30,000 spots) than in situ-synthesized microarrays or high-density bead arrays, but offer considerably more features than suspension bead arrays (Miller and Tang 2009). The main advantage of the presynthesis of oligos is the ability to generate arrays with specifically desired sequences. The presynthesis technology is often used in small laboratories that want to have the freedom to target specific sequences, which are often not standard, and therefore need to design their own chips. This flexibility allows even the spotting of unknown DNA sequences, which can be essential for particular experimental designs.

According to the nature of the probe, printed arrays can be classified as doublestranded DNA microarrays or oligonucleotide microarrays. For double-stranded (ds) DNA microarrays, the probes consist of amplification products (amplicons) obtained by PCR using primers designed from a known genomic sequence, shotgun library clones, or cDNA. The double-stranded amplicons are denatured to allow the probes to be available for hybridization. Amplicons can be attached to the glass slide surface either by the electrostatic interaction between the negative charge of the phosphate backbone of the DNA and a positively charged coating on the slide surface or by UV-cross-linked covalent bonds between the thymidine bases in the DNA and amine groups on treated slides. A 200–800-bp dsDNA probe is recommended, but larger fragments of up to 1.3 kb in length also work. In a typical microarray design, each probe corresponds to one gene. PCR amplicons for microarrays should produce a high yield and specificity and no contamination (nonspecific amplification or contaminants that autofluoresce themselves or affect the attachment of targets to the microarray surface). Nevertheless, the generation of whole-genome DNA microarrays by high-throughput PCR amplification is a very laborious process, and dsDNA probes generally have high sensitivity but suffer in specificity. Extensive quality control by gel electrophoresis, product purification, and repetition of dropout reactions is necessary (Ehrenreich 2006). The ultimate assessment of probe specificity requires sequencing the products. Decreased specificity can be beneficial when analyzing a genomic sequence rich in natural polymorphisms, but it is disadvantageous when trying to discriminate between highly similar target sequences. Moreover, high specifi city is necessary for clinical diagnostic applications (Miller and Tang 2009). Two advantages that spotted dsDNA microarrays have are their higher hybridization sensitivity and their lower cost.

Oligonucleotide spotted microarray probes consists of short chemically synthesized sequences. Usually, the length of the probes ranges from 25 to 80 bp but may be as long as 150 bp for gene expression microarrays. Using oligonucleotides as probes is an alternative to double-stranded DNA because fewer errors are introduced during probe synthesis and it allows for the testing of small genomic regions, including polymorphisms. A decreased probe length may adversely affect the sensitivity compared to dsDNA probes, but the specificity is greater when short specific genomic regions are interrogated. Moreover, longer probes have higher melting temperatures and greater mismatch tolerance, leading to decreased specificity. Generally, the larger the length of the probe, the stronger the hybridization signal and the higher the sensitivity. Because of the small size of the probes in general, oligonucleotides are commonly attached to the slide surface by covalent coupling. Otherwise, a significant amount of probes would be lost during hybridization and washing. To achieve covalent linkage, slides are commonly coated with compounds with aldehyde or epoxy functional groups. Oligonucleotides are attached to the microarray surface via a modified 5 ¢ or 3 ¢ end. This increases the availability of the probe sequences for hybridization with potential targets, because they are not fixed to the surface by the nucleic acid backbone or by individual bases. A further improvement in sensitivity can be achieved by inserting spacer molecules between the oligonucleotides and the slide surface (most commonly a 5’ amino group) (Ehrenreich 2006). Despite

being easier to manufacture than dsDNA probes, oligonucleotide probes need to be designed carefully so that all the probes have similar melting temperatures (within a range of 5°C) and lack palindromic sequences.

Printed microarrays are relatively simple and inexpensive compared to in situsynthesized microarrays. Setting up microarray facilities is costly and requires a dedicated space in which the environment is well controlled (dust, humidity, temperature, ozone levels, etc.). However, if microarray core facilities are available at research centers and/or universities, these challenges don't affect the individual researchers as much. The huge scale of amplicon production along with the issues of quality control, information management, efficiency, and accuracy are the main difficulties in the manufacturing of printed dsDNA microarrays. Moreover, the design of oligonucleotide probes is laborious, and errors introduced during probe synthesis are a problem. There are a great number of commercially available whole-genome printed microarrays for select organisms. Additionally, printed dsDNA microarrays are important for the study of organisms that have not been fully sequenced.

3.2.2 In Situ-Synthesized Oligonucleotide Microarrays

For microarrays synthesized in situ, the manufacturing steps of cloning, amplification by PCR, and probe spotting are not necessary. This offers an important advantage by reducing the noise and variability of the system. Most array manufacturers offer standard ready-to-use chips designed for monitoring the expression profiles or the genotype of many common organisms such as human, mouse, rat, or yeast. In this case, the customer is not responsible for the complex problems of probe design or attaching the probes to the chip.

In situ-synthesized arrays are high-density microarrays that, like oligonucleotide printed arrays, use oligonucleotide probes. In contrast to printed oligonucleotide arrays, however, the oligonucleotide probes are synthesized directly on the surface of the microarray.

Arrays generated by this approach achieve DNA oligonucleotide synthesis by successive rounds of deprotection from UV light with many photolithographic masks. By combining solid-phase chemistry and photolithography, extremely highdensity oligonucleotide arrays containing more than 10 6 features can be manufactured. GeneChips (Affymetrix) are the most commonly known. In situ-synthesized probes are usually short (20–25 bp), and typically, multiple probes per target are included to improve the sensitivity, specifi city, and statistical accuracy. The use of probe sets is common to increase the specificity. A probe set includes one perfect match probe and one mismatch probe that contains a 1-bp difference in the middle position of the probe. Results from the exact probe versus the mismatched probe can therefore be used to detect and eliminate cross-hybridization.

GeneChips oligonucleotide probes (Affymetrix) are synthesized using semiconductor-based photochemical synthesis. There are synthetic linkers on the quartz surface modified with light-sensitive protecting groups. Consequently, the microarray surface is chemically protected from a nucleotide attachment until deprotected by light. Once the array surface is exposed to UV light, reactive nucleotides modified with a photolabile protecting group can be added to growing oligonucleotide chains. Photolithographic masks are used to target specific nucleotides to exact probe sites. Each photolithographic mask has a defi ned pattern of windows, which act as a filter to either transmit or block UV light. Areas of the microarray surface in which UV light has been blocked

will remain protected from the addition of nucleotides, whereas areas exposed to light will be deprotected, and specific nucleotides can be added. The pattern of windows in each mask directs the order of nucleotide addition. In situ probe synthesis is therefore achieved through the repeated cycling of masking, light exposure, and the addition of either A, C, T, or G bases to the growing oligonucleotides (Miller and Tang 2009).

Roche NimbleGen and Agilent Technologies are other important manufactures of high-density oligonucleotide microarrays. In both platforms, the oligonucleotide probes used are longer (60–100 bp). NimbleGen microarrays can contain more than 10 6 features, and different formats per slide are available, such as 1 × 2.1 million features, 3 × 720,000 features, 1 × 385,000 features, 4 × 72,000 features, and 12 × 135,000 features. Agilent microarrays are available in the following formats: SurePrint G3 formats (8 × 60,000 features, 4 × 180,000 features, 2 × 400,000 features, and 1 × 1,000,000 features), and SurePrint HD formats (8 × 15,000 features, 4 × 44,000 features, 2 × 105,000 features, and 1 × 244,000 features).

While NimbleGen manufactures probes by maskless photo-mediated synthesis similar to the process used for Genechips, Agilent employs inkjet technology for in situ synthesis. Affymetrix and NimbleGen microarray technologies use quartz wafers, while Agilent microarrays use glass slides and inkjet printing, which eliminates the need for either lithographic or digital masks. Unlike Affymetrix chips, which are hybridized with only one labeled target, the NimbleGen and Agilent platforms allow multicolor hybridizations and the use of longer oligonucleotides, thus increasing sensitivity.

Synthesized microarrays depend on commercial manufacturing because of the complexity behind chemical synthesis and the high cost involved in their production. The number of microbial genome microarrays that are commercially available for gene expression studies is continually growing. For specific applications, synthesized oligonucleotide arrays can be designed and ordered as custom microarrays. A custom Affymetrix microarray can be quite expensive, and the inflexibility of its custom mask makes the use of an Affymetrix-synthesized array impractical for some applications. In contrast, the NimbleGen and Agilent platforms are easily customized with an unique oligonucleotide sequence content. In addition, a web-based tool provided by Agilent called eArray allows users to design custom microarrays with no minimum manufacturing batch size requirements, making Agilent microarrays a primary choice for homebrew and pilot applications. The major advantages to these systems are the reproducibility of the manufacturing process and the standardization of reagents, instrumentation, and data analysis. Oligonucleotide microarrays generally allow much cleaner downstream hybridization data than amplicon-based microarrays. With oligonucleotide arrays, the ability to standardize probe concentrations and hybridization temperatures has resulted in considerable improvements in the accuracy and reproducibility of microarray data (Miller and Tang 2009).

3.2.3 Suspension Bead Arrays

Suspension bead arrays are three-dimensional arrays based on the use of microscopic polystyrene spheres as a solid support and flow cytometry for bead and target detection. Multiplexing is accomplished by using different microsphere sets based on color (Miller and Tang 2009).

The current generation of commercially available suspension bead arrays (LabMap System, Luminex) offers a 100-element array that uses a flow cytometer for analysis of the microspheres in suspension. The flow cytometer provides a method to simultaneously detect DNA-binding events on

each microsphere. Two spectrally distinct fluorophores are incorporated on polystyrene microspheres (FlowMetrix Beads). Using defined ratios of these fluorophores, a suspension microarray is created consisting of up to 100 different microsphere sets. The flow cytometer provides two lasers for simultaneous excitation of both fluorophores. A third fluorophore coupled to the reporter molecule quantifies the biomolecular interaction that has occurred on the surface of each microsphere. A suspension of microspheres is individualized by the hydrodynamic focusing effect of a flow cell. High-speed digital signal processing classifies the microspheres according to their spectral properties and thus is able to quantify the reaction on the surface. For each analyte, a defined quantity of microspheres is added to the sample. After mixing and incubating the analytes and microspheres, the detector molecules are added. A centrifugation or filtration step is needed after incubation to separate the unbound components. The washed bead suspension is directly read with the flow cytometer (Seidel and Niessner 2008).

Although the feature density of suspension bead arrays is the lowest of the platforms that have been reviewed here, the relative simplicity, powerful multiplexing capabilities, and relatively low cost make this platform the most practical system for high-throughput nucleic acid detection in applications such as clinical diagnosis of infectious diseases (Miller and Tang 2009).

3.2.4 High-Density Bead Arrays

BeadArrays (Illumina) offer an attractive substrate for the high-density detection of target nucleic acids. BeadArrays are based on 3- mm silica beads that randomly self-assemble onto one of two available substrates: the Sentrix Array Matrix (SAM) or the Sentrix BeadChip. The SAM contains 96 1.4-mm fiber-optic bundles. Each bundle is an individual array consisting of 50,000 5- mm light-conducting fibers, each of which is chemically fixed to create a microwell for a single bead. Each SAM allows the analysis of 96 independent samples. The Sentrix BeadChip is used to assay 1–16 samples at the same time on a silicon slide that has been processed by microelectromechanical systems technology to provide microwells for individual beads. BeadChips are more appropriate for very high-density applications such as whole-genome genotyping and for determining genome-wide single-nucleotide polymorphisms. In BeadArrays technology, the beads randomly separate to their final location on the array so the bead location must be mapped, which is in contrast to the known locations of printed or in situ-synthesized microarrays. The mapping is accomplished by a decoding process. Each bead has approximately 700,000 copies of a unique capture oligonucleotide covalently attached to it, which acts as the bead's identifier. After a series of hybridization and rinse steps, fluorescently labeled complementary oligonucleotides bind to their specific bead sequence, making the identification of the bead location possible (Seidel and Niessner 2008; Miller and Tang 2009). This decoding process offers the additional advantage of the quality control provided for each feature of the microarray. Since each manufactured microarray will not be identical, BeadArrays have built-in redundancy, which provides a crucial experimental control for inter-microarray comparative data (Miller and Tang 2009). BeadArrays have been successfully used for DNA methylation studies, gene expression profiling, and SNP genotyping.

3.3 Microarray Approaches and Applications

Regardless of the technology used, the final aim of any microarray assay is to provide a measure for each probe of the relative abundance of the complementary target in the sample of

interest. When well designed, microarrays can answer a huge number of questions with a single experiment or can increase the scope of an experiment to previously inaccessible levels (Wheelan et al. 2008). Thus, microarrays have been successfully used in a variety of applications, including sequencing, single-nucleotide polymorphism (SNP) detection, characterization of protein–DNA interactions, DNA computing, and others. However, mRNA profiling applications currently dominate microarray usage because of the amount of information that can be obtained about the functions of genes in cells and tissues. Applications can be divided into two main classes: mRNA gene expression profiling and analysis of genomic DNA.

3.3.1 mRNA Gene Expression Profiling

Undoubtedly, the most often used application of DNA microarray technology is transcription analysis. The transcriptome, or the entire set of genes transcribed in a specific cell at a given time under defined conditions, can effectively be analyzed using DNA microarrays. However, it is essential to spend some time thinking about exactly what type of gene expression changes are expected and in what type of cells those changes would be best detected. Microarray experiments are relatively easy to perform, but poor experimental design may yield results that are difficult or impossible to interpret (Ness 2006).

- ***Case versus control studies*.** The most common and basic question in DNA microarray experiments is whether genes appear to be upregulated or downregulated between two or more groups of samples. This type of analysis is essential because it provides the simplest characterization of the specific molecular differences associated with a specific biological effect. These signatures can be used to generate new hypotheses and guide the design of further experiments (Trevino et al. 2007). Treating a cell line or microorganism with a specific treatment condition generates immediate and rapid changes in gene expression that can be detected with microarray assays (Ness 2006). When comparing two biological conditions such as disease state vs. normal state, genes that are differentially expressed in the disease state can be identified and hypotheses can be made to identify the genes that play a causal role in the development of the disease. If the transcriptional data are also confirmed at the protein level, these genes can potentially pose an interest as candidates for drug targets in pharmaceutical research. Ideally, drugs could be designed to specifically inhibit any particular gene, protein, or signaling cascade, and if the target is specifically expressed in the diseased tissue, there is less chance of causing undesirable effects (de Rinaldis 2006). Simple case vs. control studies have given way to more powerful experimental designs to suggest targets and illuminate disease mechanisms.

- ***Comparison of samples*.** Microarray technology offers a rapid and sensitive way to compare gene expression profiles in tumors from different individuals as a potential clinical tool to identify which types of tumors might respond better to a particular treatment or for identifying patients with better or worse prognoses. Such information could be particularly helpful to make decisions about which therapeutic options are most appropriate. Nevertheless, these studies involve quite complex data and statistical analyses. Additionally, successful clinical studies require balanced cohorts designed by qualified biostatisticians to avoid common pitfalls and artifacts (Ness 2006).

- ***Functional response patterns*.** The power of expression profiling is most evident in experiments that explore a systematically varied set of conditions. Data redundancy is provided by sampling a smoothly varying process. Co-regulation of genes across a set of biological conditions can reveal functional gene groups (Stoughton 2005). These co-regulation–based groupings are a fairly accepted way to gain functional information using the "guilt by association" inference. In other words, a gene of unknown function is predicted to be associated with a functional role if its expression pattern is similar to that of a gene of known function (Quackenbush 2003).

- ***Studying pathways and biological gene networks*.** While standard gene expression analysis looks at each gene as an independent entity, pathways analysis is designed for the identification of coordinated changes in expression, affecting many genes at the same time. The idea at the basis of the various approaches and tools for pathway analysis is that of characterizing the behavior of groups of genes that act in concert to carry out a specific function. Defined gene sets can include annotation-based groups of genes belonging to the same functional category/pathway or can derive from the analysis of independent expression data, like clusters of co-expressed genes, genes expressed in a particular tissue type or in particular conditions (de Rinaldis and Lahm 2006).

3.3.2 Analysis of Genomic DNA

Genomes constantly experience expansion, contraction, and other changes due to events such as deletion, insertion, translocation, and inversion. These changes can be identified through comparative genomic hybridization (CGH), where DNA sequence differences throughout the entire genome are monitored by comparing differentially labeled test DNA and reference DNA (Shiu and Borevitz 2008).

- ***Genotyping*.** The rapid acquisition of genetic information was one of the original objectives of the development of Affymetrix microarray technology. In 1996, Chee et al. demonstrated a method of resequencing for point mutations using microarrays, which currently has become an established technique. The baseline method involves short probes complementary to every N-mer of the baseline target sequence along with additional probes with each of the four nucleotides at the putative mutation position. Each of these also can be paired with a "mismatch" probe to control for nonspecific hybridization. Chips have been designed for mutation detection in genes of particular interest to human health, showing the promise of these methods and also the difficulties associated with false detections when the underlying mutation rates are low (Stoughton 2005).

- ***Detailed characterization of microbial pathogens*.** By enabling parallel interrogation of pathogen genomes, microarrays offer promising improvements in the diagnosis of infectious diseases, monitoring of emerging infections, and examining the safety of food, water, and air (Stoughton 2005). Probes targeted to specific genes are used to detect the presence of virulence factors, antigenic determinants, and drug resistance determinants as well as to closely resolve related species of bacteria. Probes designed for genes of a baseline strain can be used to characterize and compare the genomes of test strains via competitive hybridization. CGH can be conducted both within species and between species to discover

instances of gene gain or loss. CGH is not restricted to genes, since arrays with nongenic sequences or even whole genomes are commercially available for several model organisms or can be fabricated on demand (Shiu and Borevitz 2008). Host–microbe interactions also could be studied in detail using a combination of genomic analysis of the pathogen and expression profiling of host immune cells (Stoughton 2005).

3.4 Microarray Methodologies

Microarray technology allows simultaneous massive parallel determination and multiple measurements of a variety of binding events. Moreover, it has the advantage of requiring a small amount of material, and it may easily be automated with the possibility of saving a great deal of time.

In a typical expression microarray experiment, the system relies on measuring the absolute intensity of the labeled target from each sample. In a one-color strategy, mRNA from one type of cells or tissues is isolated and reverse-transcribed to cDNA to avoid degradation. Usually, fluorescently labeled nucleotides are incorporated during the reverse transcription. After the labeling, cDNAs are hybridized on the microarray, which is then analyzed by a laser scanner, and the intensity of each feature is measured.

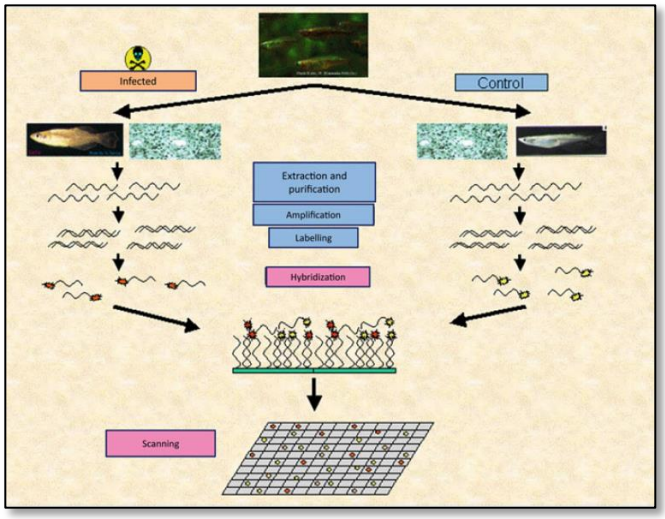


Fig. 3.1 Steps of a two-color approach strategy to do a comparative genomic hybridization experiment with DNA microarrays.

A different commonly used technique to measure gene expression involves RNA isolation from two separate samples. This is the two-color approach strategy, a comparative hybridization experiment between two samples. In this technique, the two different samples are each labeled with two different dyes that fluoresce at nonoverlapping wavelengths (typically, Cy3 and Cy5 are used). The two pools of labeled cDNAs are mixed and then hybridized on the array. If expression levels are

similar for both samples, then both Cy3- and Cy5-labeled cDNA are present, resulting in a yellow-colored spot. If expression levels from the Cy5-labeled cDNA are higher, then a red spot will appear. On the contrary, if expression levels from the Cy3- labeled cDNA are higher, then a green spot will be observed (Fig. 3.1).

3.4.1 Sample Preparation

There are many variations in the protocol for isolating and amplifying target nucleic acid. In the case of mRNA profiling, either mRNA or total RNA can be the starting material for amplification. In prokaryotes, mRNA purification is problematic because the nucleic acid lacks 3' polyadenylation, and there are no widely adopted protocols for selectively labeling the mRNA. Due to the lack of polyadenylation, random priming has to be used with either hexamers or nonamers. Thus, only total RNA can be labeled, resulting in a higher background and requiring a substantially higher amount of total RNA to be added to the labeling reaction. New protocols trying to solve this problem have been developed, such as preparing polyadenylated mRNA from prokaryotes; however, they have not been widely adopted. An additional problem in working with prokaryotic mRNA is its higher instability compared with eukaryotic mRNA. This instability demands special care during the preparation of prokaryotic RNA to avoid producing artifacts. It is possible to miss the expression of certain genes simply because of degradation of the corresponding mRNA during the experiment (Ehrenreich 2006). Good-quality total bacterial RNA can be prepared with commercial kits such as RNeasy (Qiagen), but the quality should be checked with a Bioanalyzer (Agilent Technologies) that expresses the RNA quality in the form of an RNA integrity number (RIN). RNA with a value higher than 8 on the RIN scale of 0–10 is considered suitable for hybridization.

Labeling molecules or modified nucleotides capable of being tagged with a labeling molecule can be incorporated during the synthesis of amplification products. In the Affymetrix protocol, the labeling takes place after hybridization. When the amount of sample material is scarce, an amplification step is needed. Nucleic acid amplification can be accomplished through reverse transcription of RNA followed by linear amplification via in vitro transcription (IVT) and/or polymerase chain reaction (PCR). Amplification can be 3' -biased or full-length, and this decision influences the process of probe design. 3' -Biased amplification methods take advantage of the ability to prime from polyadenylation sequences found in eukaryotic transcripts. Full-length amplification tends to employ random priming of the target molecules, either because poly(A) sequences do not exist (prokaryotic organisms) or because of a desire to obtain amplification product lengths that are more representative of the complete target sequences. The final product to be hybridized to the array can be either cDNA or cRNA. Linear and modest amplification, as well as post-synthetic incorporation of labels, are usually associated with more reproducible data. cDNA–DNA hybridizations are likely to suffer less from cross-hybridization, although the binding energies tend to be lower than those of cRNA–DNA duplexes (Stoughton 2005).

3.4.2 Hybridization

During hybridization, complementary sequences should be able to find each other. The fundamental parameters are time, stringency, concentration, and complexity of the sample and density of available binding sites. Other factors include the distribution of fragment lengths, steric effects of dye molecules, and surface chemistry. The optimization of stringency involves choosing

conditions in which, for most probes, perfect match duplexes have a high occupancy compared to mismatch duplexes. Overall, any given target sequence will pair and dissociate many times during the hybridization reaction, staying longer at high-binding-energy well-matched duplexes than in poorly matched duplexes. During this annealing, specificity increases over time as the reaction approaches equilibrium. In fact, the progression with time can be used to distinguish specific from nonspecific binding. Generally, long hybridization times at a relatively high stringency are required to obtain the best specificity (Stoughton 2005).

The hybridization of DNA microarrays can be done in two different ways. The classical approach includes placing the labeled target on a slide and carefully covering it. This requires some degree of skill to prevent gradients in hybridization and to avoid air bubbles. A hybridization chamber is usually needed to keep the temperature and humidity constant. Hybridization temperatures range from 40°C to 65°C depending on several factors, including the organism studied, the hybridization buffer composition, and the technology used. The hybridization temperature is critical for oligonucleotide microarrays and has to be carefully optimized. Hybridization solutions contain saline sodium citrate (SSC), sodium dodecyl sulfate (SDS) as detergent, nonspecific DNA such as salmon sperm DNA, blocking reagents to reduce the background like bovine serum albumin (BSA) or Denhardt's reagent, and labeled cDNA from the samples. Other factors such as agitation, microfluidic circulation (Affymetrix system), sonication, and the use of surfactants or buffers have the potential to speed up and improve hybridization. Washing off the unbound sample after hybridization is a crucial step. Stringency must be optimized here as well. More stringent washing steps are performed at the end of the washing procedure, which can be achieved by either decreasing the ionic strength or increasing the washing temperature (Ehrenreich 2006). Finally, the slides are dried and scanned within several hours after hybridization, because the fluorescence signal deteriorates with time. It is important to protect the slides from high levels of atmospheric ozone due to the sensitivity of some dyes (particularly during the drying step when the array surface is exposed to the air).

Automatic array hybridization stations are an alternative to the classical approach. The hybridization and washing steps are achieved by running programmed protocols. The results do not depend on the skill of the researcher and are very reproducible. Since the hybridization and washing conditions have to be fi ne-tuned according to the chemistry of the slide, the conditions are most easily optimized when a large number of arrays based on the same chemistry have to be handled identically. This is the case of the Affymetrix system.

3.4.3 Image Capture

After hybridization, the slide is read by a scanner, which consists of a device similar to a fluorescence microscope coupled with a laser, robotics, and a digital camera to record the fluorescent emission. Fluorescent labeling and detection on microarrays have replaced radioactive labeling because of the higher sensitivity and the fact that nonradioactive labeling is much easier and safer to handle. The amount of signal detected is presumed to be proportional to the amount of dye at each feature in the microarray and hence proportional to the target nucleic acid concentration of the complementary sequence in the sample. The output is a monochromatic digital image file typically in TIFF format for each fluorescent dye. False-color images containing red, green, and yellow are reconstructed by specialized software for visualization purposes only (Trevino et al. 2007).

Scanning of the fluorescent hybridization signal can be done with CCD imaging using filtered white light illumination. However, nowadays, it is common to use laser confocal scanners. The laser confocal approach has fundamental advantages, such as better signal-to-background ratios and less photobleaching of the labels (Stoughton 2005). In addition to scanners specific for Affymetrix technology, leading manufacturers of scanners include Agilent Technologies, Axon Instruments, and Genomic Solutions. Most devices have lasers and filter sets that are compatible with common fluorescent label pairs such as Cy3 and Cy5. Usually, scanners possess their own image processing software to reduce the raw images to spot intensities.

New labeling options, such as quantum dots and plasmon resonance particles, may finally allow efficiency as good as single-molecule detection and reduce requirements in amplification and quantity of biological sample input. Plasmon surface resonance (PSR) detection of molecular binding is potentially a way to scan microarrays without using any label, although this has yet to produce any commercial systems. PSR is an optical technique that investigates what happens at the interface of a thin metal-coated prism in contact with a solution, which is used to determine refractive index changes at the surface. When light is incident on the prism side at a particular angle (resonance angle), the intensity of the reflected light is at its minimum. In the presence of biomolecules on the metal surface, this angle variation is very sensitive. Changes in reflectivity give a signal that is proportional to the mass of the biomolecules bound to the surface. To detect one molecule, such as a DNA target, the ligand (probe) is immobilized onto the surface. As the target binds to the ligand, the mass and the refractive index increase; thus, detection of binding can be achieved without any label (Sassolas et al. 2008).

References

Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP (1996) Accessing genetic information with high-density DNA arrays. Science 274:610–614.

de Rinaldis E, Lahm A (2006) DNA microarrays: Current Applications. Horizon Bioscience, Norfolk, UK.

Ehrenreich A (2006) DNA microarray technology for the microbiologist: an overview. Appl Microbiol Biotechnol 73:255–273.

Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D (1991) Light-directed, spatially addressable parallel chemical synthesis. Science 251:767–773.

Miller MB, Tang YW (2009) Basic concepts of microarrays and potential applications in clinical microbiology. Clin Microbiol Rev 22:611–633.

Ness SA (2006) Basic Microarray Analysis. In: Larson RS (ed) Methods in Molecular Biology, vol. 316: Bioinformatics and Drug Discovery. Humana Press Inc. Totowa, New Jersey.

Quackenbush J (2003) Genomics. Microarrays – guilt by association. Science 302:240–241.

Sassolas A, Leca-Bouvier BD, Blum LJ (2008) DNA Biosensors and Microarrays. Chem Rev 108:109–139.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470.

Seidel M, Niessner R (2008) Automated analytical microarrays: a critical review. Anal Bioanal Chem 391:1521–1544.

Shiu SH, Borevitz JO (2008) The next generation of microarray research: applications in evolutionary and ecological genomics. Heredity 100:141–149.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517.

Stoughton RB (2005) Applications of DNA microarrays in Biology. Ann Rev Biochem 74:53–82.

Trevino V, Falciani F, Barrera-Saldaña HA (2007) DNA Microarrays: a Powerful Genomic Tool for Biomedical and Clinical Research. Mol Med 13:527–541.

Wheelan SJ, Murillo FM, Boeke JD (2008) The incredible shrinking world of DNA microarrays. Mol Biosyst 4:726–732.

2.3. La Bioinformática y los datos genómicos y transcriptómicos

Durante los últimos años, la generación de gran cantidad de datos derivados del desarrollo de las técnicas de Biología Molecular ha provocado la aparición de nuevos desafíos para los investigadores, que han de ser capaces de procesar eficientemente dichos datos. En este contexto es necesario contar con herramientas que permitan la gestión, el análisis y la interpretación de los datos generados, lo que ha propiciado un gran desarrollo en el campo de la Bioinformática. La Bioinformática se define como la aplicación de la tecnología de computación en el manejo y análisis de datos biológicos. Es por tanto, un área multidisciplinar cuyo objetivo último es descifrar la valiosa información contenida en las grandes masas de datos procedentes de la investigación biológica, para poder así obtener una comprensión más clara de la biología de los organismos. El desarrollo de la Bioinformática ha supuesto un profundo impacto en campos tan importantes como el de la Microbiología y la salud humana y animal, entre otros.

Uno de los primeros campos de avance de la Bioinformática fue el desarrollo de repositorios o bases de datos biológicos en los que almacenar de una forma estructurada los datos procedentes de diferentes áreas dentro del campo de la Biología Molecular. Las principales publicaciones científicas requieren el depósito de los resultados asociados a cualquier trabajo de investigación en una base de datos, así como la utilización de estándares para su depósito, en caso de que existan. En los últimos años existe una necesidad creciente de estandarización, que ha surgido de la mano del gran aumento en la cantidad de datos que generan las técnicas genómicas. En este sentido, el proyecto MIBBI (*Minimum Information for Biological and Biomedical Investigations*) de Biosharing (Taylor *et al*., 2008) fue creado como una guía para el establecimiento de estándares que garanticen la verificación, el análisis y la interpretación correcta de la información procedente de la experimentación biológica del conjunto de la comunidad científica. Existen tres principales bases de datos públicas de secuencias de ADN: la europea ENA (*European Nucleotide Archive*) del EMBL-EBI (*European Molecular Biology Laboratory – European Bioinformatics Institute*) (Leinonen *et al*., 2011), la americana GenBank del NCBI (*National Center for Biotechnology Information*) (Benson *et al*., 2013) y la japonesa DDBJ (*DNA Databank of Japan*) (Kosuge *et al.,* 2013). En todas ellas se puede encontrar la misma información sobre secuencias de ADN de cualquier organismo, que pueden ser desde secuencias nucleotídicas de fragmentos concretos de pequeño tamaño, hasta genomas completamente secuenciados, si bien el formato de los registros varía entre ellas. UniProt (*Universal Protein*) (UniProt Consortium, 2013) es un repositorio central de secuencias e información funcional de proteínas que aúna la información de las bases de datos Swiss-Prot (del Instituto Suizo de Bioinformática), TrEMBL (del EBI) y PIR (*Protein Information Resource*) de

la Universidad de Georgetown, lo que la ha convertido en el recurso líder mundial de almacenamiento de información sobre proteínas. La existencia de estas bases de datos ha supuesto un importante impulso para el desarrollo de herramientas bioinformáticas. En la actualidad, el análisis de secuencias incluye aspectos tales como la identificación automática de genes y sus estructuras (conocidos como *gene finding* y *gene prediction*), la identificación de polimorfismos y mutaciones genéticas o el alineamiento y comparación de secuencias (tanto de genes como de proteínas), siendo este último tipo de análisis uno de los más comúnmente empleados para distintos fines. Entre las herramientas de alineamiento y comparación de secuencias, BLAST (*Basic Local Alignment Search Tool*) del NCBI (Altschul *et al.*, 1990) es una de las más utilizadas.

2.3.1. Bases de datos y análisis de datos genómicos y de microarrays

La aplicación durante las últimas décadas de tecnologías moleculares globales como la secuenciación de última generación y los microarrays de ADN, ha supuesto la generación y el análisis de una gran cantidad de información, propiciando que la época que vivimos se haya definido como "la era genómica". Algunos científicos apuntan incluso a que actualmente estamos viviendo "la era posgenómica", en la que la comunidad científica trata de extraer toda la información posible de los datos generados durante "la era genómica". El análisis bioinformático de datos genómicos representa en la actualidad un área amplísima de aplicación de la Bioinformática, cubriendo aspectos fundamentales en el desarrollo de la presente Tesis Doctoral, tales como el ensamblado y la anotación durante el proceso de secuenciación genómica, los análisis cualitativos de genómica comparativa, y el análisis cuantitativo de los análisis de genómica funcional. También aborda otras áreas, como la integración y modelización a diferentes niveles de los diversos datos "-ómicos" mediante aproximaciones de biología de sistemas.

En la sección 2.1, dedicada a la secuenciación de genomas, ya se han mencionado algunos de los aspectos bioinformáticos asociados con los procesos de secuenciación y anotación de las secuencias genómicas. Existen numerosas herramientas que se usan en las diferentes fases de análisis de datos generados en la secuenciación de un genoma, como las que permiten el ensamblado de las secuencias, la búsqueda de genes y la anotación automática de los genomas una vez ensamblados, así como las dirigidas a la comparación de diferentes genomas y que resultan muy útiles en el campo de la genómica comparativa y la filogenia. Además de estas herramientas, existen bases de datos genómicos que además de contener la información sobre las secuencias de los genomas y su anotación, integran otros aspectos: regulación de la transcripción genómica; niveles de expresión de RNAs,

proteínas o metabolitos; interacciones entre moléculas; y rutas metabólicas y de señalización y sus interacciones. Una de las bases de datos genómicos más usadas en la interpretación de datos genómicos es la del KEGG (*Kyoto Encyclopedia of Genes and Genomes*), ya que integra gran cantidad de información a diferentes niveles biológicos (Kanehisa *et al.*, 2012). La integración de toda esta información hace posible el estudio de la Biología y de los procesos biológicos mediante lo que se ha definido como Biología de Sistemas, en la que se emplean los datos obtenidos para generar modelos integradores que expliquen los procesos biológicos como conjunto en lugar de como la suma de sus partes estudiadas individualmente.

Los datos capturados mediante microarrays de ADN se han convertido en un caso excepcional debido a la gran cantidad de información sobre diferentes procesos biológicos que éstos son capaces de generar. Los microarrays han supuesto un ejemplo del impacto de las nuevas tecnologías en la necesidad de crear estándares que permitan compartir los datos experimentales de una manera homogénea y efectiva para su publicación, tanto en la literatura como en las bases de datos públicas. En este sentido, MIAME (*Minimum Information About a Microarray Experiment*) es un estándar creado por la Sociedad FGED (*Functional Genomics Data*) (Brazma *et al.,* 2001) para la reposición de datos provenientes de experimentos con microarrays, que define la información necesaria para una correcta interpretación de los resultados, así como para asegurar su reproducibilidad, pero sin especificar el formato. Normalmente el formato dependerá de la base de datos donde vayan a ser depositados los resultados, siendo las más utilizadas: *ArrayExpress* del EMBL-EBI (Rustici *et al*., 2013) y GEO (*Gene Expression Omnibus*) del NCBI (Barrett *et al*., 2013).

La Bioinformática resulta esencial en cada una de las fases de la experimentación con microarrays de ADN. Para la fase de diseño existen aplicaciones informáticas específicas necesarias para un adecuado diseño de las sondas del microarray. El análisis de los datos generados por el escaneado del microarray tras la hibridación resulta una fase compleja que requiere el uso de diferentes herramientas para la obtención de imágenes, su transformación en datos numéricos, y finalmente el análisis y la gestión de dichos datos. En este caso, las herramientas deben permitir un correcto procesado que prepare los datos cuantitativos para un análisis estadístico que permita una interpretación biológica correcta. Existen numerosas aplicaciones bioinformáticas para el análisis de datos de microarrays, y el uso de una u otra normalmente depende del tipo de aplicación (análisis de expresión o genotipado) (López-Campos *et al*., 2012). Por último, es importante destacar que para un buen aprovechamiento de los datos generados mediante el uso de microarrays es necesario un nivel adecuado de formación y experiencia de los investigadores que permitan el correcto manejo e interpretación de la información.

3. Objetivos

Lactococcus garvieae es un microorganismo bien conocido por su importancia como patógeno de peces, y causante también de infecciones en otras especies, incluida la humana. Su naturaleza ubicua, y principalmente su potencial carácter zoonótico, hacen de esta bacteria un patógeno de interés creciente tanto en medicina veterinaria como humana. A pesar de ello, el conocimiento sobre el contenido genético de este patógeno resultaba muy limitado hasta hace pocos años, por lo que se planteó como objetivo general de la presente Tesis Doctoral el estudio del genoma de *L. garvieae* y su análisis funcional. Para ello, se han desarrollado los siguientes objetivos concretos:

- **Análisis del contenido genómico de *L. garvieae* mediante hibridación genómica comparativa inter-especie utilizando microarrays de ADN.**

- **Evaluación de la utilidad de biomarcadores descritos (gen de la foso-beta-galactosidasa-*lacG* y utilización de la lactosa) para la clasificación de aislados de *L. garvieae* en función de su origen.**

- **Determinación del contenido genómico de *L. garvieae* mediante secuenciación de dos cepas de origen clínico aisladas de lactococosis de trucha arcoíris (Lg8831) y de septicemia humana (Lg21881).**

- **Caracterización molecular y funcional de los plásmidos presentes en la cepa Lg21881.**

- **Análisis funcional del genoma de Lg8831 y Lg21881 mediante el estudio de sus perfiles de expresión en respuesta a la temperatura utilizando microarrays de ADN.**

CAPÍTULO II:

Análisis del genoma de *Lactococcus garvieae* mediante hibridación genómica comparativa inter-especie sobre microarrays de ADN

Resumen

Lactococcus garvieae es un patógeno bacteriano que afecta a diferentes especies de animales y al hombre. A pesar de su amplia distribución y su significación clínica, hasta hace pocos años existía una falta de conocimiento casi total sobre el contenido genético de esta bacteria. Si bien la secuenciación del genoma representa una buena aproximación para el conocimiento del contenido genético de cualquier microorganismo, en el año 2007 la secuenciación genómica resultaba aún una cara y relativamente lenta. Por ello, la tecnología de microarrays de ADN representaba una alternativa a la secuenciación completa del genoma. En los últimos años varios estudios han demostrado que la Hibridación Genómica Comparativa (CGH: *Comparative Genomic Hybridization*) mediante microarrays de ADN puede ser una buena aproximación para la búsqueda de nuevos genes en un organismo no secuenciado, mediante una hibridación heteróloga de su ADN sobre un microarray en el que se representan todos los ORFs de otro microorganismo secuenciado de distinta especie, pero filogenéticamente próximo (Dong *et al*., 2001; Hakenbeck *et al*., 2001; Fukiya *et al*., 2004). Gracias a esta aproximación se pueden identificar aquellos genes que son comunes entre el organismo secuenciado y el no secuenciado, lo que permite descubrir un gran número de genes en el organismo de interés en un solo experimento.

En el presente trabajo se analizó el contenido genómico de *L. garvieae* CECT4531 utilizando conjuntamente tanto un análisis *in silico* mediante herramientas bioinformáticas, como experimentos de CGH utilizando microarrays de ADN de dos microorganismos filogenéticamente cercanos (*Streptococcus pneumoniae* TIGR4 y *Lactococcus lactis* subsp. *lactis* IL1403) como microorganismos de referencia. Así, se consiguieron identificar 267 genes presentes en *L. garvieae*, la mayoría de los cuales (n= 258) se detectaron por primera vez en este patógeno. Gran parte de estos genes se asociaron con funciones ribosómicas, sistemas de conversión de energía y metabolismo de los azúcares. Algunos de los genes identificados podrían estar implicados en la patogénesis de las infecciones causadas por *L. garvieae*.

Este estudio proporcionó una primera aproximación al análisis del contenido del genoma de *L. garvieae*, permitiendo tener un conocimiento del mismo mucho más amplio del que se tenía hasta ese momento. Igualmente se demostró que la CGH puede ser un método útil para estudiar el contenido genético de microorganismos no secuenciados.

Aguado-Urda *et al. BMC Microbiology* 2010, **10**:79
http://www.biomedcentral.com/1471-2180/10/79

BMC Microbiology

RESEARCH ARTICLE Open Access

Analysis of the genome content of *Lactococcus garvieae* by genomic interspecies microarray hybridization

Mónica Aguado-Urda[1], Guillermo H López-Campos[2], José F Fernández-Garayzábal[1,3], Fernando Martín-Sánchez[2], Alicia Gibello[1], Lucas Domínguez[3], María M Blanco[1*]

Abstract

Background: *Lactococcus garvieae* is a bacterial pathogen that affects different animal species in addition to humans. Despite the widespread distribution and emerging clinical significance of *L. garvieae* in both veterinary and human medicine, there is almost a complete lack of knowledge about the genetic content of this microorganism. In the present study, the genomic content of *L. garvieae* CECT 4531 was analysed using bioinformatics tools and microarray-based comparative genomic hybridization (CGH) experiments. *Lactococcus lactis* subsp. *lactis* IL1403 and *Streptococcus pneumoniae* TIGR4 were used as reference microorganisms.

Results: The combination and integration of *in silico* analyses and *in vitro* CGH experiments, performed in comparison with the reference microorganisms, allowed establishment of an inter-species hybridization framework with a detection threshold based on a sequence similarity of ≥ 70%. With this threshold value, 267 genes were identified as having an analogue in *L. garvieae*, most of which (n = 258) have been documented for the first time in this pathogen. Most of the genes are related to ribosomal, sugar metabolism or energy conversion systems. Some of the identified genes, such as *als* and *mycA*, could be involved in the pathogenesis of *L. garvieae* infections.

Conclusions: In this study, we identified 267 genes that were potentially present in *L. garvieae* CECT 4531. Some of the identified genes could be involved in the pathogenesis of *L. garvieae* infections. These results provide the first insight into the genome content of *L. garvieae*.

Background

Lactococcus garvieae is one of the most important bacterial pathogens that affect different farmed fish species in many countries, although its major impact is on the trout farm industry [1,2]. In addition to farmed fish, this microorganism has also been isolated from a wide range of wild fish species, from both fresh and marine water, as well as from giant fresh water prawns [3] and from wild marine mammals [4]. The host range of *L. garvieae* is not limited to aquatic species. This agent has also been identified in cows and water buffalos with subclinical mastitis [5,6] and from cat and dog tonsils [7]. In humans it has been isolated from the urinary tract, blood, and skin and from patients with pneumonia, endocarditis or septicaemia [8-11]. Recently, intestinal disorders in humans have been associated with the consumption of raw fish contaminated with this pathogen [12], which suggests that *L. garvieae* could be considered as a potentially zoonotic bacterium [3,12]. Despite the widespread distribution and emerging clinical significance of *L. garvieae* in both veterinary and human medicine, there is almost a complete lack of knowledge about the genetic content of this microorganism.

In the last few years, research in microbial genetics has changed fundamentally, from an approach involving the characterization of individual genes to a global analysis of microbial genomes. The availability of complete genome sequences has enabled the development of high-throughput nucleic acid hybridization technologies including macro- and microarrays. Microarrays have the

* Correspondence: mmblanco@vet.ucm.es
[1]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad Complutense, 28040 Madrid, Spain

Aguado-Urda et al. BMC Microbiology 2010, 10:79
http://www.biomedcentral.com/1471-2180/10/79

capacity to monitor the genome content of bacterial strains or species very rapidly. Although whole-genome sequencing is definitely a powerful method for genetics, it is still expensive and time consuming. As an alternative, comparative genomic hybridization (CGH) experiments based on microarrays have been used to facilitate comparisons of unsequenced bacterial genomes. Array-based CGH using genome-wide DNA microarrays is used commonly to determine the genomic content of bacterial strains [13,14], but also for inter-species comparisons [14-16]. In this case, microarrays of closely related microorganisms that have been fully sequenced must be available. The primary advantage of this microarray approach is that it allows the identification of a large number of genes that are potentially present in an organism without the need for sequencing genomes. The disadvantage of this approach is that it indicates only the genes that are common between the fully sequenced relative and the strain of interest; genes unique to the strain of interest remain unknown [15,17]. In the present work the genetic content of *L. garvieae* CECT 4531 was studied by a combination of *in silico* analysis and *in vitro* microarray CGH experiments, using open reading frame (ORF) microarrays of two bacteria closely related to *L. garvieae*, namely *Lactococcus lactis* subsp. *lactis* IL1403 and *Streptococcus pneumoniae* TIGR4 [18,19].

Methods

Bacterial strains, culture conditions and isolation of genomic DNA

Lactococcus lactis subsp. *lactis* IL1403 (kindly provided by M.P. Gaya, INIA, Madrid, Spain) and *Streptococcus pneumoniae* TIGR4 (purchased form the American Type Culture Collection) were used as the reference sequenced microorganisms. The test strain of *Lactococcus garvieae* used for the experiments was CECT 4531 (purchased from the Spanish Type Culture Collection). The *L. lactis* subsp. *lactis* IL1403 and *L. garvieae* CECT 4531 were grown statically at 28°C in BHI broth (bioMérieux, Marcy l'Etoile, France). The *S. pneumoniae* TIGR4 was grown statically at 37°C in Todd Hewitt broth (Oxoid, Basingstoke, Hampshire, England). Cells were grown until the late-exponential phase of growth (OD600~1.5-2) and harvested for isolation and purification of genomic DNA using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's specifications. The DNA concentrations were determined spectrophotometrically.

DNA labelling

Aliquots (1-2 μg) of genomic DNA from the three strains were labelled fluorescently with Cy3-dUTP or Cy5-dUTP (Perkin-Elmer, Foster City, CA, USA), depending on whether the strain was used as a test or reference microorganism in the CGH experiments, respectively. Each DNA aliquot was fragmented by sonication to obtain fragments from 400 to 1000 bp. Fragmented DNA was mixed with 5 μL 10× NEBlot labelling buffer containing random sequence octamer oligonucleotides (New England Biolabs, Ipswich, MA, USA) and water to a final volume of 43.5 μL. This mixture was denatured by heating at 95°C for 5 min and then cooled for 5 min at 4°C. After this denaturing step, the remaining components of the labelling reaction were added: 5 μL of 10 × dNTP labelling mix (1.2 mM each dATP, dGTP and dCTP in 10 mM Tris pH 8.0, 1 mM EDTA) (New England Biolabs, Ipswich, MA, USA), 1.5 μL of 1 mM Cy3-dUTP or Cy5-dUTP and 1.5 μL of 10 U/μL Klenow fragment (Fermentas Life Sciences, Glen Burnie, MD, USA). The labelling reactions were incubated overnight at 37°C and then stopped by adding 2.5 μL of 0.5 M EDTA. Labelled DNA was purified from unincorporated label using a Qiaquick PCR Cleanup kit (Qiagen, Hilden, Germany) and dried under vacuum. The final DNA concentration and quality, as well as the labelling quality, were determined using a NanoDrop (NanoDrop Techonologies, Wilmington, DE, USA).

Array-based comparative genome hybridization (CGH)

The *L. lactis* subsp. *lactis* IL1403 and *S. pneumoniae* TIGR4 microarrays used for the CGH analysis were purchased from Eurogentec (Serain, Belgium). The *L. lactis* microarray contains 4608 spots: 2126 duplicated ORFs, 32 negative controls and 324 empty spots. The *S. pneumoniae* microarray contains 4608 spots: 2087 duplicated ORFs, 224 negative controls and 210 empty spots.

The CGH experiments were performed by means of competitive hybridizations using DNA of *L. lactis* subsp. *lactis* IL1403 or *S. pneumoniae* TIGR4, depending on the array, as positive controls. The DNAs to be hybridized on the same array were labelled with Cy3-dUTP and Cy5-dUTP, respectively. For each microarray hybridization reaction, aliquots (1-2 μg) of labelled genomic DNAs of the reference (labelled with Cy3) and test (labelled with Cy5) strains, were mixed in 45 μL EGT hybridization solution (Eurogentec, Serain, Belgium) and denatured at 65°C for 2 min. The hybridization mixture was then loaded onto a microarray slide, covered with a coverslip and incubated at 38°C overnight. Following hybridization, the slides were washed in 2 × SSC, 0.5% SDS for 5 min followed by a second wash step in 1 × SSC, 0.25% SDS for 5 min. Finally, slides were rinsed in 0.2 × SSC and dried by centrifugation.

The results presented herein represent a compilation of sixteen separate CGH experiments: *L. lactis* subsp. *lactis* IL1403 arrays (reference microorganism) were hybridized with *S. pneumoniae* TIGR4 (test microorganism) (n = 2);

Aguado-Urda *et al. BMC Microbiology* 2010, **10**:79
http://www.biomedcentral.com/1471-2180/10/79

S. pneumoniae TIGR4 arrays (reference microorganism) were hybridized with *L. lactis* subsp. *lactis* IL1403 (test microorganism) (n = 2); *L. lactis* subsp. *lactis* IL1403 arrays (reference microorganism) were hybridized with *L. garvieae* CECT 4531 (test microorganism) (n = 8); *S. pneumoniae* TIGR4 arrays (reference microorganism) were hybridized with *L. garvieae* CECT 4531 (test microorganism) (n = 4). The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [20] and are accessible through GEO Series accession number GSE19005. http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19005.

Data acquisition and analysis

The microarray was scanned after hybridization using a Scanarray HT microarray scanner (Perkin-Elmer). The signal intensity of the two fluors was determined using ImaGene software (BioDiscovery, El Segundo, CA, USA). Microarray data were analysed using ImaGene software, Microsoft Excel and an in-house designed and built Microsoft Access database [21]. Gene calling was based on a signal-to-noise ratio (SNR) >3 for each spot. After the CGH experiments, a gene was considered to show a positive result when it was present in at least three of the four CGH assays. In the case of the *L. garvieae* CECT 4531 hybridizations with the *L. lactis* subsp. *lactis* IL1403 arrays, it was necessary to perform a larger number of assays (n = 8), owing to the poor quality of one of the batches of arrays used. Thus, the criterion chosen to determine a positive result in this case was when the gene was present in at least five of the eight CGH assays.

In silico sequence analysis

Sequence analyses were carried out to assess the performance of the inter-species CGH protocol. Using the BLAT [22] and BLAST [23] programs, the sequences of the *L. lactis* microarray probes were aligned with the *S. pneumoniae* genome sequence, and *vice-versa*. The BLAT search parameters were 90%, 80% and 70% sequence identity (BLAT90, BLAT80 and BLAT70) and a 100 bp minimum alignment length (owing to the fact that the length of the array probe was between 100 and 400 bp). Available *L. garvieae* sequences of the nine previously identified genes that were positive in the CGH were aligned with the *L. lactis* subsp. *lactis* IL1403 or *S. pneumoniae* TIGR4 genomes and with the sequences of the immobilized probes of these genes in the corresponding microarray using BLAST [23] and BLAST 2 sequences [24] programs.

Results

Inter-species comparison framework

In silico analyses were performed to compare the sequences of the immobilized probes in the microarray of each reference organism with the sequences of their complete genomes available in GenBank (*L. lactis* subsp. *lactis* IL1403: NC_002662 and *S. pneumoniae* TIGR4: NC_003028). The BLAT alignment of the *L. lactis* IL1403 probes on the *S. pneumoniae* TIGR4 genome allowed the identification of 1 ORF with BLAT90, 65 ORFs with BLAT80 and 159 ORFs with BLAT70. Moreover, the BLAT alignment of the probes represented on the *S. pneumoniae* microarray on the *L. lactis* genome demonstrated 1 ORF, 63 ORFs and 165 ORFs for BLAT90, BLAT80 and BLAT70, respectively.

The CGH experiments based on swapping off the microarrays between *S. pneumoniae* and *L. lactis* identified 65 common ORFs. To evaluate the accuracy of the microarray CGH experiments, we compared these results with those of the *in silico* analysis. Out of the 65 genes, 47 (72%) showed similarities greater than 80%, 16 genes (25%) exhibited a similarity between 70% and 80%, and only 2 genes (3%) showed a similarity slightly lower than 70% (66-68%) (Table 1). In summary, 97% of the genes detected by CGH showed similarities greater than 70% at the nucleotide level.

After combined analysis of the results obtained *in silico* and *in vitro*, we established, under the hybridization conditions used in this study, a detection threshold based on a sequence similarity of ≥ 70% for alignments longer than 100 bp. This was established as the reference framework for the inter-species CGH assays.

In vitro microarray CGH experiments with *L. garvieae* CECT 4531 vs reference microorganisms *L. lactis* subsp. *lactis* IL1403 and *S. pneumoniae* TIGR4, and *in silico* analysis of available sequences from *L. garvieae*

The microarray CGH experiments identified 267 genes in *L. garvieae* that had analogues in *L. lactis* and/or *S. pneumoniae* (Additional file 1). Of these, 111 genes (41.6%) were identified only with the *L. lactis* microarray, 70 genes (26.2%) only with the microarray of *S. pneumoniae*, and 86 genes (32.2%) were identified with both microarrays. These genes belong to diverse functional groups (Table 2). Most of the genes (96.6%) have been documented for the first time in *L. garvieae*. Only nine genes (four present in both reference microorganisms: *atpD/SP1508, pfk/SP0896, tig/SP0400, tuf/SP1489*; three present in *L. lactis*: *als, ddl, galK*; two present in *S. pneumoniae: SP0766, SP1219*) out of the 267 genes detected have been either identified or sequenced before in diverse strains of *L. garvieae* (Tables 3 and 4). *In silico* analysis of these previously sequenced genes (n = 9) of *L. garvieae* were performed to assess the efficacy of the methodology. Alignments of these available sequences with the genomes of the corresponding reference microorganism and their respective array probes showed nucleotide identities ranging

Aguado-Urda *et al. BMC Microbiology* 2010, **10**:79
http://www.biomedcentral.com/1471-2180/10/79

Table 1 *In silico* analysis of the common genes detected by CGH in the reference microorganisms

CGH positive gene ID[a]	BLAT80[b]	BLAT70[b, c]	% Identity in BLAST[d]
atpD/SP1508	++		
dnaG/SP1072	-	-	66%
dnaJ/SP0519	+		
dnaK/SP0517	++		
enoA/SP1128	++		
fbaA/SP0605	++		
ftsZ/SP1666	++		
fusA/SP0273	++		
gapB/SP2012	++		
gltX/SP2069	-	++	
ldh/SP1220	++		
lepA/SP1200	-	++	
leuS/SP0254	++		
lysS/SP0713	-	++	
msmk/SP1580	++		
murA1/SP1081	-	+	
pepC/SP0281	+		
pfk/SP0896	-	+	
pgiA/SP2070	+		
pgk/SP0499	++		
pmg/SP1655	++		
ptnAB/SP0284	+		
ptnD/SP0282	-	+	
ptsI/SP1176	-	++	
purA/SP0019	-	-	76%
pyk/SP0897	++		
rplA/SP0631	++		
rplB/SP0212	++		
rplC/SP0209	++		
rplF/SP0225	+		
rplJ/SP1355	+		
rplK/SP0630	++		
rplL/SP1354	++		
rplM/SP0294	++		
rplN/SP0219	++		
rplO/SP0229	++		
rplQ/SP0237	+		
rplS/SP1293	++		
rplV/SP0214	+		
rpmI/SP0960	+		
rpmJ/SP0233	++		
rpoB/SP1961	-	+	
rpoC/SP1960	+	++	
rpoD/SP1073	++		
rpsA/SP0862	-	+	
rpsB/SP2215	-	++	
rpsC/SP0215	++		
rpsD/SP0085	++		
rpsE/SP0227	++		
rpsG/SP0272	++		
rpsJ/SP0208	++		
rpsK/SP0235	++		
rpsL/SP0271	++		
rpsM/SP0234	+		
rpsP/SP0775	+		
rpsS/SP0213	+		
topA/SP1263	-	+	
tuf/SP1489	++		
typA/SP0681	++		
upp/SP0745	-	-	77%
ychH/SP2097	+		
yciA/SP2096	-	-	68%
yjiF/SP1565	-	+	
yjkI/SP2209	-	+	
yyaL/SP0004	++		

[a] Reference organisms gene ID: *Lactococcus lactis* subsp. *lactis* IL1403/ *Streptococcus pneumoniae* TIGR4

[b] ++ Genes detected in both alignments, *L. lactis* subsp. *lactis* IL1403 array probes vs *S. pneumoniae* TIGR4 genome, and *S. pneumoniae* TIGR4 array probes vs *L. lactis* subsp. *lactis* IL1403 genome; + positive in one of the two cases.

[c] Only the results for the negative genes in BLAT80 are shown.

[d] Only the results for the negative genes in both BLAT80 and BLAT70 are shown.

between 70% and 86% (Tables 3 and 4). Most of the available sequences (80%) showed similarities greater than 75%.

Discussion

In the present study, commercial microarrays of *L. lactis* subsp. *lactis* IL1403 and *S. pneumoniae* TIGR4 were used to determine the presence of homologous genes in *L. garvieae*. Both *L. lactis* and *S. pneumoniae* were chosen as reference organisms because they are closely related to *L. garvieae* [18,19] and their genomes have been fully sequenced. Although these CGH experiments cannot detect and identify genes that are likely to exist only in the target microorganism, this approach reveals genes that are common to both the reference and the target organisms, allowing the identification of a large number of genes potentially present in an organism without the need for sequencing genomes [17,25].

In experiments that involve inter-species comparison it is necessary to establish a framework that allows accurate comparison and interpretation of the results. Thus, the first efforts were focused on establishing that framework by the combination and integration of *in silico* analyses and *in vitro* microarray CGH experiments to compare the reference organisms *L. lactis* subsp. *lactis* IL1403 and *S. pneumoniae* TIGR4. Signal intensity has been used to assess the level of similarity between two genes in inter-species CGH experiments [15]. However,

Aguado-Urda *et al. BMC Microbiology* 2010, **10**:79
http://www.biomedcentral.com/1471-2180/10/79

Table 2 Functional groups of genes identified in *L. garvieae* CECT 4531 according to the COG database

Functional Group	Homologous in *L. lactis* subsp. *lactis* IL1403	Homologous in *S. pneumoniae* TIGR4
Amino acid transport and metabolism	14	10
Carbohidrate transport and metabolism	24	15
Cell cycle control, cell division, cromosome partitioning	4	2
Cell wall/membrane/envelope biogenesis	5	4
Coenzime transport and metabolism	1	1
DNA replication, recombination and repair	8	12
Energy production and conversion	11	6
Inorganic ion transport and metabolism	4	5
Intracellular trafficking, secretion and vesicular transport	4	2
Lipid transport and metabolism	2	0
Nucleotide transport and metabolism	15	11
Phage capside proteins	1	0
Post translational modification, protein turnover, chaperones	8	8
Signal transduction mechanisms	2	3
Transcription	7	6
Translation, ribosomal structure and biogenesis	64	60
Unknown function	23	11
Total	197	156

Table 3 *In silico* analysis of the available sequences of the genes detected in *L. garvieae* by CGH

Gene ID	GenBank accession number of *L. garvieae* sequence	*L. garvieae* strain	Similarity with *L. lactis* subsp. *lactis* IL1403 gene (%)	Similarity with array probe (%)
als	EF450031	UNIUD074	77	76
atpD	AX111128	from patent WO0123604	86	86
ddl	AF170808	*E. serolicida*	72	75
galK	EU153555	DSM 20684	78	79
pfk	AB024532	SA8201	85	84
tig	AB024531	SA8201	82	-
tuf	AX109994	from patent WO0123604	80	77

Results for the *L. lactis* subsp. *lactis* IL1403 array based-CGH

this approach may be influenced, and therefore biased, by different factors, such as regional sample labelling effects, probe accessibility or local hybridization issues [13]. For these reasons, in the present study signal intensity was not considered for determining whether a gene was positive or not in the inter-species CGH experiments.

These analyses revealed that nearly all the genes common to *L. lactis* and *S. pneumoniae* that were detected by swap microarray CGH experiments (97%) exhibited a sequence similarity of at least 70% (Table 1). Only two genes (*dnaG* and *yciA*) detected in the microarray CGH experiments showed a sequence similarity slightly lower than 70% (66 and 68%, respectively; Table 1). Variability in the factors that influence the CGH signals, such as systematic errors (e.g. dye effects), copy number variation, and sequence divergence between the analysed samples [13], may explain these results. The comparison of the results of both analyses, *in silico* and *in vitro*, for the reference microorganisms (Table 1) allowed us to establish that, under our experimental conditions, it was possible to detect and identify inter-species hybridization with a detection threshold based on a sequence similarity of ≥70%.

Therefore, our threshold value of sequence similarity ≥70% was set up directly from the comparison of the results of the *in silico* and *in vitro* analyses of the present study. This threshold value was used subsequently to interpret the results of the microarray-based CGH experiments comparing *L. garvieae* and the reference microorganisms. Less stringent hybridization conditions would probably have allowed the identification of a larger number of genes, but this would have also resulted in lower specificity. Given that the final aim of the

Aguado-Urda *et al. BMC Microbiology* 2010, **10**:79
http://www.biomedcentral.com/1471-2180/10/79

Table 4 *In silico* analysis of the available sequences of the genes detected in *L. garvieae* by CGH

Gene ID	GenBank accession number of *L. garvieae* sequence	*L. garvieae* strain	Similarity with *S. pneumoniae* TIGR4 gene (%)	Similarity with array probe (%)
SP1508	AX111128	from patent WO0123604	82	82
SP0896	AB024532	SA8201	80	79
SP0766	AM490328	JIP 31-90 (2)	71	79
	AJ387925	CIP 102507 T	70	70
	AJ387923	*E. serolicida* ATCC49156	70	70
SP04000	AB024531	SA8201	74	-
SP1489	AX109994	from patent WO0123604	80	79
SP1219	AB364641	20-92	84	86
	AB364640	Lc.1236	84	85
	AB364639	Lc. 925	85	85
	AB364638	Lc. 881	84	84
	AB364637	Lc. 337	85	85
	AB364633	LMG9472	85	85
	AB364632	ATCC43921	84	84
	AB364627	G50202	84	84
	AB364626	KGLA5224	84	84
	AB364625	EH5803	83	83
	AB364624	KG9408	84	84

Results for the *S. pneumoniae* TIGR4 array based-CGH

experiment was the identification of genes potentially present in *L garvieae*, it was preferred to maintain stringent hybridization conditions, therefore increasing the specificity and the reliability of the results. Hence, the genes detected in the CGH experiments should have an analogue in *L. garvieae* with a nucleotide sequence identity greater than 70% with the respective gene in the reference organism.

The CGH hybridizations using *L. lactis* subsp. *lactis* IL1403 and *S. pneumoniae* TIGR4 microarrays identified 267 analogous genes in *L. garvieae* (Additional file 1). Only 3.4% of these genes (nine out of 267) have been characterized or sequenced previously by other groups in different strains of *L. garvieae* [[18,26-29], and GenBank sequences: AX109994, AB364624, AB364625, AB364626, AB364627, AB364632, AB364633, AB364637, AB364638, AB364639, AB364640, AB364641, EU153555]. The alignments of the available sequences of these nine previously of these nine previously identified genes in *L. garvieae* with both the sequences of these genes from the reference microorganisms and those from the array probe showed nucleotide similarities greater than 70% (70-86%) between them (Tables 3 and 4). These data are consistent with the detection threshold value discussed previously. Therefore it is reasonable to assume that the other genes detected in *L. garvieae* CECT 4531 by CGH experiments will also have at least 70% sequence similarity with the respective genes in the reference microorganisms. The positive result obtained in both CGH experiments for the *tig/SP0400* gene (Tables 3 and 4), was unexpected given the absence of similarity between the available sequence and the probes on both microarrays. This result could be explained by the fact that the available sequence for *L. garvieae* is partial, and it represents a part of the gene that does not correspond with the probe.

We classified the ORFs into clusters of orthologous genes (COGs) [30]. The 267 genes identified in *L. garvieae* CECT 4531 (Additional file 1) belong to diverse biological functional groups (Table 2). Most of the genes detected in *L. garvieae* (about 66%) were related to meaningful biological functions such as those related to ribosomal functions, sugar metabolism or energy conversion systems, which are usually represented in *Lactobacillales* [31]. The remaining genes identified included "housekeeping genes", such as *gyrB*, *sodA*, *recA*, *ileS*, *rpoD*, *dnaK* and *ddl* [19], genes of diverse functional groups and genes with unknown functions. Some of them are of interest because they could be involved in the pathogenesis of *L. garvieae* infections. For example, the gene *als*, which has been described as an important factor for host colonization by El Tor biotypes of *Vibrio cholerae* [32], has also been suggested to be one of the genes required for survival of *L. garvieae* in fish [27]. In addition, the gene *mycA*, which was detected for the first time in *L. garvieae* in the present study, encodes an antigen that cross-reacts with myosin, and members of

this family of proteins have been suggested to play an important role in the pathogenesis of streptococcal infections [33].

Sequencing of the genes identified in this work is beyond the scope of this initial study, but the data provided can be the starting point for future genetic analysis of *L. garvieae* strains from different ecological niches or adapted to different host species.

This study provides the first insight into the genome content of *L. garvieae* and suggests that CHG could be a useful approach for studying the genetic content of other Gram-positive catalase-negative cocci of human and veterinary relevance.

Conclusions

In the present work, a comparative analysis based on microarray interspecies hybridization and on the use of bioinformatic tools was used for the first time to study the genetic content of *L. garvieae* CECT 4531. It is important to remark that the integration of results from bioinformatics and microarray-based CGH requires the definition of a framework that allows an accurate comparison and interpretation of the results obtained. Once this framework was established, it was possible to identify 267 genes potentially present in *L. garvieae* CECT 4531. Some of the identified genes, such as the *als* and *mycA* genes, could be involved in the pathogenesis of *L. garvieae* infections.

In summary, these results provide the first insight into the genome content of *L. garvieae* and could be useful for future understanding of the genetics of this pathogenic microorganism.

Additional file 1: Genes potentially identified in *L. garvieae* CECT 4531 and their homologues in *L. lactis* subsp. *lactis* IL1403 and *S. pneumoniae* TIGR4.

Acknowledgements

This work was supported partially by projects AGL2005-04775 and AGL2009-12447 of the Ministerio Español de Ciencia e Innovación. M. Aguado-Urda was a recipient of a grant from Centro de Vigilancia Sanitaria Veterinaria (VISAVET), and a PhD grant from the Universidad Complutense de Madrid. The work of Dr. López-Campos and Dr. Martín-Sanchez was partially funded by the COMBIOMED Network and ONTOMINEBASE reseach project (Ministerio Español de Ciencia e Innovación). The authors thank M.P. Gaya for providing the *Lactococcus lactis* subsp *lactis* IL1403 strain.

Author details

[1]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad Complutense, 28040 Madrid, Spain. [2]Área de Bioinformática y Salud Pública, Instituto de Salud Carlos III, 28220 Madrid, Spain. [3]Centro de Vigilancia Sanitaria Veterinaria (VISAVET), Facultad de Veterinaria, Universidad Complutense, 28040 Madrid, Spain.

Authors' contributions

MAU carried out the microarray experiments and the bioinformatics analyses, and participated in the analysis of the data and drafting of the manuscript. GHLC designed the microarray experiments, and participated in analysis of the data and drafting of the manuscript. JFFG participated in and supervised drafting of the manuscript. FMS participated in and supervised the design of the microarray experiments and the analysis of the microarray data. AG participated in the design of the study and drafting of the manuscript. LD participated in the design of the study. MB conceived the study, and participated in its design and coordination. All the authors read and approved the final manuscript.

Received: 3 December 2009 Accepted: 16 March 2010
Published: 16 March 2010

References

1. Chen SC, Liaw LL, Su HY, Ko SC, Wu CY, Chaung HC, Tsai YH, Yang KL, Chen YC, Chen TH, Lin GR, Cheng SY, Lin YD, Lee JL, Lai CC, Weng YJ, Chu SY: **Lactococcus garvieae, a cause of disease in grey mullet, Mugil cephalus, in Taiwan.** *J Fish Dis* 2002, **25**:727-732.
2. Vela AI, Vázquez J, Gibello A, Blanco MM, Moreno MA, Liébana P, Albendea C, Alcalá B, Méndez A, Domínguez L, Fernández-Garayzábal JF: **Phenotypic and genetic characterization of *Lactococcus garvieae* isolated in Spain from lactococcosis outbreaks and comparison with isolates of other countries and sources.** *J Clin Microbiol* 2000, **38**:3791-3795.
3. Vendrell D, Balcázar JL, Ruiz-Zarzuela I, de Blas I, Gironés O, Múzquiz JL: ***Lactococcus garvieae* in fish: a review.** *Comp Immunol Microbiol Infect Dis* 2006, **29**:177-198.
4. Evans JJ, Pasnik DJ, Klesius PH, Al-Ablani S: **First report of *Streptococcus agalactiae* and *Lactococcus garvieae* from a wild bottlenose dolphin (*Tursiops truncatus*).** *J Wild Dis* 2006, **42**:561-569.
5. Collins MD, Farrow JAE, Phillips BA, Kandler O: ***Streptococcus garvieae* sp. nov. and *Streptococcus plantarum* sp.** *J Gen Microbiol* 1983, **129**:3427-3431.
6. Teixeira LM, Merquior VLC, Vianni MCV, Carvalho MGS, Fracalanzza SEL, Steigerwalt AG, Brenner DJ, Facklam RR: **Phenotypic and genotypic characterization of atypical *Lactococcus garvieae* strains isolated from water buffalos with subclinical mastitis and confirmation of *L garvieae* as a senior subjetive synonym of *Enterococcus seriolicida*.** *Int J Syst Bacteriol* 1996, **46**:664-668.
7. Pot B, Devriese LA, Ursi D, Vandamme P, Haesebrouck F, Kersters K: **Phenotypic identification and differentiation of *Lactococcus* strains isolated from animals.** *Syst Appl Microbiol* 1996, **19**:213-222.
8. Elliot JA, Collins MD, Pigott NE, Facklam RR: **Differentiation of *Lactococcus lactis* and *Lactococcus garvieae* from humans by comparison of whole-cell protein patterns.** *J Clin Microbiol* 1991, **20**:2731-2734.
9. Facklam RR, Elliot JA: **Identification, classification, and clinical relevance of catalase-negative, Gram-positive cocci, excluding the streptococci and enterococci.** *Clin Microbiol Rev* 1995, **8**:479-495.
10. Fefer JJ, Ratzan KR, Sharp SE, Saiz E: ***Lactococcus garvieae* endocarditis: report of a case and review of the literature.** *Diagn Microbiol Infect Dis* 1998, **32**:127-130.
11. Li WK, Chen YS, Wann SW, Liu YC, Tsai HT: ***Lactococcus garvieae* endocarditis with initial presentation of acute cerebral infarction in a healthy immunocompetent man.** *Inter Med* 2008, **47**:1143-1146.
12. Wang CY, Shie HS, Chen SC, Huang JP, Hsieh IC, Wen MS, Lin FC, Wu D: ***Lactococcus garvieae* infections in humans: possible association with aquaculture outbreaks.** *Int J Clin Pract* 2007, **61**:68-73.
13. van Hijum SAFT, Baerends RJS, Zomer AL, Karsens HA, Martín-Requena V, Trelles O, Kok J, Kuipers OP: **Supervised Lowess normalization of comparative genome hybridization data-application to lactococcal strain comparisons.** *BMC Bioinf* 2008, **9**:93.
14. Hakenbeck R, Balmelle N, Weber B, Gardes C, Keck W, de Saizieu A: **Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*.** *Infect Immun* 2001, **69**:2477-2486.
15. Dong Y, Glasner JD, Blattner FR, Triplett EW: **Genomic interspecies microarray hybridization: rapid discovery of three thousand genes in the maize endophyte, *Klebsiella pneumoniae* 342, by microarray hybridization with *Escherichia coli* K-12 Open Reading Frames.** *Appl Environ Microbiol* 2001, **67**:1911-1921.
16. Fukiya S, Mizoguchi H, Tobe T, Mori H: **Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray.** *J Bacteriol* 2004, **186**:3911-3921.
17. Zhang L, Reddi U, Srinivasan U, Li S, Borchardt SM, Pillai P, Mehta P, Styka AN, DeBusscher J, Marrs CF, Foxman B: **Combining microarray technology and molecular epidemiology to identify genes associated**

with invasive group B *Streptococcus*. *Interdiscip Perspect Infect Dis* 2008, **2008**:314762.
18. Michel C, Pelletier C, Boussaha M, Douet DG, Lautraite A, Tailliez P: **Diversity of lactic acid bacteria associated with fish and the fish farm environment, established by amplified rRNA gene restriction analysis.** *Appl Environ Microbiol* 2007, **73**:2947-2955.
19. Coenye T, Vandamme P: **Extracting phylogenetic information from whole-genome sequencing projects: the lactic acid bacteria as a test case.** *Microbiol* 2003, **149**:3507-3517.
20. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30**:207-210.
21. López-Campos GH, García-Albert L, Martín-Sánchez F, García-Sáez A: **Analysis and management of HIV peptide microarray experiments.** *Methods Inf Med* 2006, **45**:158-162.
22. Kent WJ: **BLAT-the BLAST-like alignment tool.** *Genome Res* 2002, **12**:656-664.
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
24. Tatusova TA, Maden TL: **BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences.** *FEMS Microbiol Lett* 1999, **2**:247-250.
25. Zhang R, Zhang CT: **The impact of comparative genomics on infectious disease research.** *Microbes and Infect* 2006, **8**:1613-1622.
26. Hirono I, Yamashita H, Park CI, Yoshida T, Aoki T: **Identification of genes in a KG - phenotype of *Lactococcus garvieae*, a fish pathogenic bacterium, whose proteins react with antiKG-rabbit serum.** *Microb Pathog* 1999, **27**:407-417.
27. Menéndez A, Fernández L, Reimundo P, Guijarro JA: **Genes required for *Lactococcus garvieae* survival in a fish host.** *Microbiology* 2007, **153**:3286-3294.
28. Ozawa Y, Courvalin P, Gaiimand M: **Identification of enterococci at the species level by sequencing of the genes for D-alanine: D-alanine ligases.** *Syst Appl Microbiol* 2000, **23**:230-237.
29. Poyart C, Quesnes G, Trieu-Cuot P: **Sequencing the gene encoding manganese dependent superoxide dismutase for rapid species identification of enterococci.** *J Clin Microbiol* 2000, **38**:415-418.
30. Tatusov RL, Galperin MY, Natale DA, Koonin EV: **The COG database: a tool for genome-scale analysis of protein functions and evolution.** *Nucleic Acids Res* 2000, **1**:33-36.
31. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welkeri D, Huges J, Goh Y, Benson A, Baldwin K, Lee JH, Díaz-Muñiz I, Dosti B, Smeianov V, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Parker C, Breidt F, Broadbent J, Hutkins R, O'Sullivan D, Steele J, Unlu G, Saier M, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D: **Comparative genomics of the lactic acid bacteria.** *PNAS* 2006, **103**:15611-15616.
32. Yoon SS, Mekalanos JJ: **2,3-butanediol synthesis and the emergence of the *Vibrio cholerae* El Tor biotype.** *Infect Immun* 2006, **74**:6547-6556.
33. Kil KS, Cunningham MW, Barnett LA: **Cloning and sequence analysis of a gene encoding a 67-kilodalton myosin-cross-reactive antigen of *Streptococcus pyogenes* reveals its similarity with class II major histocompatibility antigens.** *Infect Immun* 1994, **62**:2440-2449.

El material adicional asociado a esta publicación (*Additional file 1*) se encuentra disponible a través de la URL:

http://www.biomedcentral.com/content/supplementary/1471-2180-10-79-s1.doc

CAPÍTULO III:

Utilización de lactosa y presencia del gen *lacG* (fosfo-beta-galactosidasa) en aislados de *Lactococcus garvieae* de diferentes orígenes

Resumen

Además de su importancia como patógeno tanto en medicina veterinaria como humana, *Lactococcus garvieae* es una bacteria de gran ubicuidad que se aísla frecuentemente a partir de muestras ambientales y alimentarias. En la industria alimentaria además, presenta interés como bacteria láctica participante en fermentaciones de productos lácteos y cárnicos. Por todo ello, es importante conocer la diversidad intra-especie de esta bacteria y establecer posibles relaciones entre dicha diversidad y el origen de los aislados. Algunos autores han logrado establecer relaciones entre el perfil genético y el origen de algunas cepas, observando una clara diferenciación entre los aislados de peces y los de otros orígenes (Kawanishi *et al*., 2006; Foschino *et al.,* 2008).

Con el fin de definir marcadores genéticos capaces de distinguir las cepas de *L. garvieae* en función de su origen, Fortina *et al*. (2009) propusieron la capacidad de metabolizar lactosa como biomarcador diferencial entre las cepas de *L. garvieae* de origen lácteo y las aisladas de casos clínicos de lactococosis en peces. Sin embargo, dichos autores no estudiaron este carácter en cepas de otros orígenes. Con el fin de determinar la utilidad de este biomarcador en aislados de un mayor rango de orígenes, se realizó un estudio sobre la utilización de lactosa y la presencia del gen de la fosfo-beta-galactosidasa (*lacG*) en 57 aislados de *L. garvieae* de distinta procedencia (peces, rumiantes, cerdos, pájaros, aguas, alimentos y humanos).

Los datos del presente estudio demostraron que la capacidad o incapacidad de utilización de lactosa, así como la presencia o ausencia del gen *lacG* son características distribuidas heterogéneamente entre los aislados de *L. garvieae* estudiados y que, por tanto, su utilización como marcador para determinar el origen de cepas de *L. garvieae* debería ser reconsiderada.

RESEARCH ARTICLE

INTERNATIONAL MICROBIOLOGY (2010) 13:189-193
DOI: 10.2436/20.1501.01.125 ISSN: 1139-6709 **www.im.microbios.org**

INTERNATIONAL MICROBIOLOGY

Utilization of lactose and presence of the phospho-β-galactosidase (*lacG*) gene in *Lactococcus garvieae* isolates from different sources

Mónica Aguado-Urda,[1] M. Teresa Cutuli,[1] M. Mar Blanco,[1] Carmen Aspiroz,[2] José L. Tejedor,[3] José F. Fernández-Garayzábal,[1,4] Alicia Gibello[1]*

[1]Department of Animal Health, Faculty of Veterinary Science, Complutense University, 28040 Madrid, Spain. [2]Microbiology Unit, Royo Villanova Hospital, Zaragoza, Spain. [3]Dibaq Group, Fuentepelayo, Segovia, Spain. [4]Animal Health Surveillance Center (VISAVET), Complutense University of Madrid, Madrid, Spain

Received 12 November 2010 · Accepted 29 December

Summary. This study evaluates the utilization of lactose (Lac) and the presence of the phospho-β-galactosidase (*lacG*) gene as markers for distinguishing between fish (Lac$^-$/*lacG*$^-$) and dairy isolates (Lac$^+$/*lacG*$^+$) of *Lactococcus garvieae*, using a panel of *L. garvieae* isolates from different sources. None of the fish isolates produced acid from lactose (Lac$^-$), however Lac$^-$/*lacG*$^-$ isolates were observed in pigs, cows, birds and humans. Most of the dairy isolates (77.8%) were Lac$^+$/*lacG*$^+$, but some dairy isolates did not produce acid from this sugar. Data in the present study show that the ability to metabolize lactose and the presence of the *lacG* gene are heterogeneously scattered among *L. garvieae* isolates of different sources. Therefore, the use of these criteria as markers to differentiate between *L. garvieae* isolates of dairy and fish origin should be considered with caution. [**Int Microbiol** 2010; 13(4):189-193]

Keywords: *Lactococcus garvieae* · lactose metabolism · *lacG* gene

Introduction

Lactococcus garvieae is one of the most important bacterial pathogens that affect different farmed fish species in many countries. Its major impact, though, is on the trout farm industry [23]. In addition to farmed fish, this microorganism has been isolated also from a wide range of wild fish species, found both in fresh and marine water, as well as from giant fresh water prawns [24] and wild marine mammals [7]. The host range of *L. garvieae* is not limited to aquatic species, but is also associated to subclinical mastitis in cows and water buffalos, and pneumonia in pigs [4,5,22; Tejedor, PhD Thesis 2008]. In humans, *L. garvieae* has been isolated from patients affected by different clinical cases like urinary tract infections, pneumonia, endocarditis or septicaemia [15,17,25]. Like other animal pathogens of clinical relevance [3], *L. garvieae* is considered a potentially zoonosic bacteria [26]. In addition to its relevance as a pathogen, *L. garvieae* is widespread in nature, having been isolated from wild birds, rivers and sewage waters [Tejedor, PhD Thesis].

In the last years, *L. garvieae* has gained interest in the food industry because it has been isolated from different types of food, including vegetables [12], meat and meat products [19,21]. In particular, *L. garvieae* is considered to be a component of the natural occurring microbiota in dairy

***Corresponding author:** A. Gibello
Department of Animal Health
Faculty of Veterinary Science
Complutense University of Madrid
28040 Madrid, Spain
Tel. +34-913943900. Fax +34-913943908
E-mail: gibelloa@vet.ucm.es

products from raw milk of different ruminants [8, 10] and is also associated with the naturally fermentation of different artisan cheeses [1]. Recently, *L. garvieae* has been isolated also from faecal samples of healthy people, suggesting this microorganism could be part of the human commensal microbiota or transient bacteria ingested with food [14].

Lactococcus garvieae strains have a genetic diversity related to the specific host they colonize [11] and molecular genetic studies have revealed a low genetic relatedness between fish and dairy strains [8]. According to Fortina et al. [9], the *L. garvieae* isolates of both origins can be differentiated by the ability of dairy isolates to utilize lactose (Lac^+) and the presence of the phospho-β-galactosidase (*lacG*) gene, detectable in all isolates of dairy origin but lacking in fish isolates. However, *L. garvieae* from sources other than fish and dairy products were not analyzed by Fortina et al. Due to the widespread nature of *L. garvieae*, the aim of the present study is to evaluate the usefulness of the acidification of lactose and the presence of the *lacG* gen as indicators of the origin of *L. garvieae* in a set of isolates from different sources.

Materials and methods

Bacterial isolates. A total of 57 isolates of *Lactococcus garvieae* from food (n = 10), diseased trout (n = 10), ruminants (n = 7), pigs (n = 10), birds (n = 2), water (n = 8) and humans (n = 10) were studied. Isolates were considered clinical when isolated from clinical specimens (trout, ruminants, pigs and human). Otherwise they were considered as non-clinical (food, birds and water). Additional information regarding the studied isolates is given in Table 1. Bacteria were grown on Columbia blood agar plates (bioMérieux España) incubated at 30ºC for 24–48 h. Identification of *L. garvieae* was carried out by using the commercial identification system Rapid ID32 Strept (bioMérieux España) and by PCR [27].

Lactose utilization. Acid production from lactose was determined using the Rapid ID32 Strept strips and also confirmed by using phenol red broth base medium (Difco) supplemented with 1% (w/v) sugar (Sigma-Aldrich), after 48 h of incubation at 30°C.

β-Galactosidase activity. The production of the enzyme β–galactosidase was determined using the Rapid ID32 Strept strips and also using the commercial diagnostic ONPG assay (Difco).

Detection and sequencing of the *lacG* gene. The presence of the *lacG* gene was determined by PCR using the primer set lac-F 5´-AGCTACTCGACGACCAACAC-3´ [9] and lacG-R 5´-AGCTACTCGACGACCAACAC-3´ [present study] designed to obtain an expected PCR product size of 969 bp (76.7% of the gene). *In vitro* amplification was carried out in a reaction mixture of 100 µl containing DNA template (10–20 ng), 1 µM of each primer, 100 µM of each dNTP (Biotools), 5 U of Ultratools DNA polymerase (Biotools) and its 1X amplification buffer. The amplifications were carried out in a Mastercycler gradient thermal cycler (Eppendorf) with the following parameters: an initial denaturation step of 94ºC for 2 min, 30 serial cycles of a denaturation step of 94ºC for 1 min, annealing at 50ºC for 2 min and extension at 72ºC for 2 min, followed by a final extension step of 72ºC for 10 min. Negative controls (no DNA template) were included in each batch of PCR reactions. PCR-generated products were detected by electrophoresis of 5 µl of each amplification mixture in 1% agarose gels supplemented with 1X SYBER safe (Invitrogen, Eugene, OR, USA). The *lacG* amplicons were purified (Geneclean Turbo Kit, MP Biomedicals LLC) and sequenced at the SECUGEN facilities (Biological Research Center, CIB-CSIC, Madrid, Spain) using the DyeDeoxy (dRhodamine) Terminator Cycle Sequencing kit in an automatic ABI Prism DNA sequencer (Applied Biosystems). Multiple-sequence alignments and similarities (%) of the *lacG* sequences in the *L. garvieae* isolates examined in this study, and the one available in GenBank (accession no. EU153556.1), were obtained with CLUSTALW2 software available from the EMBL-EBI web server [http://www.ebi.ac.uk/Tools/clustalw2/index.html].

Statistical analysis. The Fisher's exact test was used to determine the relationship between the utilization of lactose and presence of *lacG* gen and the clinical/non-clinical origin of the *L. garvieae* isolates, using SPSS 12.0 for Windows software (SPSS Inc., Chicago, IL). Differences were considered significant when probabilities were lower than 0.05.

Nucleotide sequence accession numbers. The *lacG* sequences of isolates T2-17, 1042, 306/79, MAM-77, DP1 and 21331-2, representatives of the different *L. garvieae* sources examined in this study, have been deposited in GenBank under the accession numbers FR687333-FR687338, respectively.

Results and Discussion

All *L. garvieae* isolates were properly identified with the Rapid ID32 Strept system. Biochemical identification was further confirmed by specific PCR, giving the expected amplification product of 1100 bp belonging to the 16S rRNA gene [27]. A good concordance was observed between the ability to acidify lactose and the presence of the *lacG* gene. All *L. garvieae* isolates that were able to utilize lactose (Lac^+) also carried the *lacG* gene (Table 1 and Fig. 1). The multiple-sequence alignment analysis revealed that *lacG* sequence is highly conserved among the different *L. garvieae* isolates analysed in this work and the previously available *lacG* sequence, showing similarities of between 98–99%.

In many bacteria, lactose catabolism is initiated by permease-promoted uptake of lactose and the generation of glucose and galactose by β-galactosidase. However, a small number of gram-positive bacteria follow a quite distinct pathway. This pathway begins with the phosphorylation of lactose upon uptake by the phosphoenolpyruvate-dependent phosphotransferase system (*lac*-PTS) and the catabolism of lactose phosphate into glucose and galactose 6-phosphate by the enzyme phospho-β-galactosidase, encoded by the *lacG* gene [2,16]. The ONPG assay was negative for all the isolates. These results indicate that *L. garvieae* does not catabolize lactose through the β-galactosidase enzyme and agree with the idea that in *L. garvieae* Lac^+ isolates, the lactose is

Table 1. Results of the lactose utilization (Lac) and presence of *lacG* gene for the *Lactococcus garvieae* isolates

Strain	Source	Pathological process	Lac and *lacG*
Food			
T1-1, T2-17, CAS-2	Casín cheese	–	+
4AB5, 3AA7	Cabrales cheese	–	+
1204	Salers cheese	–	–
N201	Saint-Nectarine cheese	–	+
1042	Raw milk	–	+
DK2-25	Kajmak[a]	–	–
LG-80	Morcilla Burgos[b]	–	–
Fish			
4876/006	Eel	Lactococcosis	–
8831/04, 98/4284, 5664, 1684, 8053/02, 5457, 5424, 5323, 8494/03	Trout	Lactococcosis	–
Human			
21881, BM06/00349	Human	Blood (bacteremia)	–
306/79	Human	Urine	+
240-88	Human	Urine	–
673-80	Human	Skin	–
2182-81, 1108-86, 364-88, 6690, 2486-87	Human	Blood	–
Ruminants			
1183, 1205	Buffalo	Subclinical mastitis	–
CECT[c] 4531[T], G-34, MAM-75	Cow	Subclinical mastitis	–
G-14, MAM-77	Cow	Subclinical mastitis	+
Pig			
1108/02, 1481/03, 1364/02, 1139/02	Pig	Pleuritis	–
205/03, 2497/03, 164/03	Pig	Pericarditis	–
391/03	Pig	Not known	–
396, 2487	Pig	Pneumonia	–
Wild bird			
H-14	Blue tit	–	–
30/02	Magpie	–	–
Water			
DP1, 2260-1	Sewage treatment plant	–	+
2260-2	Sewage treatment plant	–	–
2131-2, 2132-2	River	–	+
2544-1, 2134-1, 2133-1	River	–	–

[a]Serbian traditional fermented milk. [b]Spanish traditional black pudding. [c]Spanish Type Culture Collection.

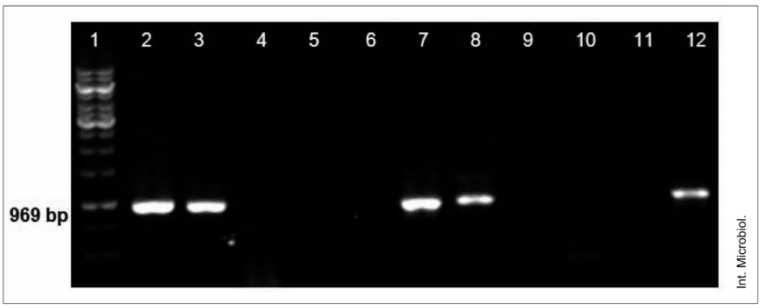


Fig. 1. Gel electrophoresis of *lacG* PCR amplicons from representative isolates of *L. garvieae*. Lane 1, molecular weight marker 1kb ladder (Biotools); lane 2 to 12, isolates T2-17, MAM-77, DK2-25, 8831/04, H14, 2132-2, DP1, 2133-1, 1108/02, 21881 and 306/79, respectively.

transported by the *lac*-PTS and hydrolyzed by the phospho-β-galactosidase enzyme. Unlike other lactic acid bacteria, in *L. garvieae* the *lacG* gene is chromosomally located [9].

None of the fish and pigs isolates of *L. garvieae* produced acid from lactose (Lac$^-$). Likewise, most of the *L. garvieae* isolates recovered from raw milk from cows with subclinical mastitis and most of the human isolates were Lac$^-$ (Table 1). In contrast, most of the dairy isolates (seven out of the nine isolates analyzed; 77.8%) were Lac$^+$; only two dairy isolates (DK2-25 and 1204) did not produce acid from this sugar. The isolate from black pudding, as well as both isolates from birds were Lac$^-$. The occurrence of Lac$^+$ and Lac$^-$ isolates of *L. garvieae* from the water samples of rivers and sewage treatment plants was comparable. In agreement with these results, Lac$^+$ isolates of *L. garvieae* have been detected also in soil and grass environmental samples [13].

The majority of the *L. garvieae* isolates from clinical specimens (91.9%) were Lac$^-$/*lacG*$^-$, a percentage that statistically was significantly higher ($p<0.05$) than that observed in non-clinical isolates (Table 1). This difference suggests that the inability to metabolize lactose and the absence of the *lacG* gene in *L. garvieae* could be related to the clinical origin of the isolate. Nevertheless, the high proportion of Lac$^-$/*lacG*$^-$ isolates observed also among the non-clinical isolates makes it difficult to use this phenotypic characteristic to anticipate the clinical origin of a *L. garvieae* isolate.

The ability to metabolize lactose and/or the presence of the *lacG* gene have been proposed as markers for distinguishing between dairy (Lac$^+$/*lacG*$^+$) and fish (Lac$^-$/*lacG*$^-$) *L. garvieae* isolates [8,9]. In the present study, in addition to fish, Lac$^-$/*lacG*$^-$ isolates of *L. garvieae* were also observed in pigs, cows, birds and humans (Table 1). Nevertheless, Lac$^+$ isolates of this microorganism had been isolated previously from fish [18,20].

Regarding the isolates of *L. garviae* of dairy origin, although most of them were Lac$^+$/*lacG*$^+$, some isolates (22.2%) were Lac$^-$/*lacG*$^-$. This percentage could be even higher considering that most of the *L. garvieae* isolates from raw milk from cows with subclinical mastitis were Lac$^-$/*lacG*$^-$ (71.4%; Table 1). *Lactococcus garvieae* is associated with a considerable number of subclinical mastitis in ruminants [5,6,22,23; Tejedor, PhD Thesis] and milk of these animals is commercialized to elaborate different dairy products. Consequently, it is likely to be expected that *L. garvieae* Lac$^-$/*lacG*$^-$ isolates present in the milk of animals with subclinical mastitis were further isolated in dairy products.

According to the results of the present study, as well as previous studies, the ability/inability to metabolize lactose and/or the presence/absence of the *lacG* gene are not exclusively related to dairy and fish isolates of *L. garvieae.* Therefore, the use of these criteria as markers to differentiate between *L. garvieae* isolates of dairy and fish origin should be considered with caution.

Acknowledgments. This study was supported by project AGL2009-12447 of the Spanish Ministry of Science and Innovation. M. Aguado-Urda was a recipient of a PhD grant from the Complutense University of Madrid. The authors thank Dr. B. Mayo from the Institute of Dairy Products of Asturias, Spain; Dr. J. Rovira from the University of Burgos, Spain; Dr. L. Topisirovic from the Institute of Molecular Genetics and Genetic Engineering, Belgrade, Serbia; and Dr. M.C. Montel from the Agricultural Research Institute, INRA, Aurillac, France, for kindly providing the *L. garvieae* food isolates. We also thank Juncal Fernández-Garayzábal for her assistance with the English revision of the manuscript.

References

1. Alegría A, Álvarez-Martín P, Sacristán N, Fernández E, Delgado S, Mayo B (2009) Diversity and evolution of the microbial populations during manufacture and ripening of Casín, a traditional Spanish, starter-free cheese made from cow's milk. Int J Food Microbiol 136:44-51

2. Aleksandrzak-Piekarczyk T, Kok J, Renault P, Bardowski J (2005) Alternative lactose catabolic pathway in *Lactococcus lactis* IL1403. Appl Environ Microbiol 71:6060-6069
3. Blume V, Luque I, Vela AI, Borge C, Maldonado A, Domínguez L, Tarradas C, Fernández-Garayzábal JF (2009) Genetic and virulence-phenotype characterization of serotypes 2 and 9 of *Streptococcus suis* swine isolates. Int Microbiol 12:161-166
4. Collins MD, Farrow JAE, Phillips BA, Kandler O (1983) *Streptococcus garvieae* and *Streptococcus plantarum* sp. nov. J Gen Microbiol 129:3427-3431
5. Devriese LA, Hommez J, Laevens H, Pot B, Vandamme P, Haesebrouck F (1999) Identification of aesculin-hydrolyzing streptococci, lactococci, aerococci and enterococci from subclinical intramammary infections in dairy cows. Vet Microbiol 70:87-94
6. Domenech A, Prieta J, Fernández-Garayzábal JF, Collins MD, Jones D, Domínguez L (1993) Phenotypic and phylogenetic evidence for a close relationship between *Lactoccoccus garvieae* and *Enterococcus seriolicida*. Microbiología SEM 9:63-68
7. Evans JJ, Pasnik DJ, Klesius PH, Al-Ablani S (2006) First report of *Streptococcus agalactiae* and *Lactococcus garvieae* from a wild bottlenose dolphin (*Tursiops truncatus*). J Wild Dis 42:561-569
8. Foschino R, Nucera D, Volponi G, Picozzi C, Ortoffi M, Bottero MT (2008) Comparison of *Lactococcus garvieae* strains isolated in northern Italy from dairy products and fishes through molecular typing. J Appl Microbiol 105:652-662
9. Fortina MG, Ricci G, Borgo F (2009) A study of lactose metabolism in *Lactococcus garvieae* reveals a genetic marker for distinguishing between dairy and fish biotypes. J Food Proc 72:1248-1254
10. Jokovic N, Nikolic M, Begovic J, Jovcic B, Savic D, Topisirovic L (2008) A survey of the lactic acid bacteria isolated from Serbian artisanal dairy product kajmak. Int J Food Microbiol 127:305-311
11. Kawanishi M, Yoshida T, Yagashiro S, et al. (2006) Differences between *Lactococcus garvieae* isolated from the genus *Seriola* in Japan and those isolated from other animals (trout, terrestrial animals from Europe) with regard to pathogenicity, phage susceptibility and genetic characterization. J Appl Microbiol 101:496–504
12. Kawanishi M, Yoshida T, Kijima M, et al. (2007) Characterization of *Lactococcus garvieae* isolated from radish and broccoli sprouts that exhibited a KG+ phenotype, lack of virulence and absence of a capsule. Lett Appl Microbiol 44:481-487
13. Klijn N, Weerkamp AH, de Vos WM (1995) Detection and characterization of lactose-utilizing *Lactococcus* spp. in natural ecosystems. Appl Environ Microbiol 61:788-792
14. Kubota H, Tsuji H, Matsuda K, Kurakawa T, Asahara T, Nomoto K (2010) Non-cultural detection of human intestinal catalase-negative, Gram-positive cocci by rRNA-targeted reverse transcription-PCR. Appl Environ Microbiol 76:5440-5451
15. Li WK, Chen YS, Wann SR, Liu YC, Tsai HC (2008) *Lactococcus garvieae* endocarditis with initial presentation of acute cerebral infarction in a healthy immunocompetent man. Intern Med 47:1143-1146
16. Maeda S, Gasson MJ (1986) Cloning, expression and location of the *Streptococcus lactis* gene for phospho–β-D-galactosidase. J Gen Microbiol 132:331-340
17. Mofredj A, Baraka D, Kloeti G, Dumont JL (2000) *Lactococcus garvieae* septicemia with liver abscess in an immunosuppressed patient. Am J Med 109:513-514
18. Pereira F, Ravelo C, Toranzo AE, Romalde JL (2004) *Lactococcus garvieae*, an emerging pathogen for the Portuguese trout culture. Bull Eur Ass Fish Pathol 24: 274-279
19. Rantsoiu K, Urso R, Iacumin L, Cantoni C, Cattaneo P, Comi G, Cocolin L (2005) Culture-dependent and -independent methods to investigate the microbial ecology of Italian fermented sausages. Appl Environ Microbiol 71:1977-1986
20. Ravelo C, Magariños B, Romalde JL, Toranzo AE (2001) Conventional versus miniaturized systems for the phenotypic characterization of *Lactococcus garvieae* strains. Bull Eur Ass Fish Pathol 21:136-144
21. Santos EM, Jaime I, Rovira J, Lyhs U, Korkeala H, Björkroth J (2005) Characterization and identification of lactic acid bacteria in "morcilla de Burgos". Int J Food Microbiol 97:285-296
22. Teixeira LM, Merquior VLC, Vianni MCE, et al. (1996) Phenotypic and genotypic characterization of atypical *Lactococcus garvieae* strains isolated from water buffalos with subclinical mastitis and confirmation of *L. garvieae* as a senior subjetive synonym of *Enterococcus seriolicida*. Int J Syst Bacteriol 46:664-668
23. Vela AI, Vázquez J, Gibello A, et al. (2000) Phenotypic and genetic characterization of *Lactococcus garvieae* isolated in Spain from lactococcosis outbreaks and comparison with isolates of other countries and sources. J Clin Microbiol 38:3791-3795
24. Vendrell D, Balcázar JL, Ruiz-Zarzuela I, de Blas I, Gironés O, Múzquiz JL (2006) *Lactococcus garvieae* in fish: a review. Comp Immunol Microbiol Infect Dis 29:177-198
25. Vinh DC, Nichol KA, Rand F, Embil JM (2006) Native-valve bacterial endocarditis caused by *Lactococcus garvieae*. Diagn Microbiol Infect Dis 56:91-94
26. Wang CY, Shie HS, Chen SC, et al. (2007) *Lactococcus garvieae* infections in humans: possible association with aquaculture outbreaks. Int J Clin Pract. 61:68-73
27. Zlotkin A, Eldar A, Ghittino C, Bercovier H (1998) Identification of *Lactococcus garvieae* by PCR. J Clin Microbiol 36:983-985

CAPÍTULO IV:

Secuenciación del genoma de los aislados clínicos *Lactococcus garvieae* 21881 (humano) y 8831 (trucha arcoíris)

Resumen

A pesar de los rápidos avances en los últimos años en el campo de la secuenciación genómica y de la gran importancia de *Lactococcus garvieae* como patógeno en la industria piscícola, así como causante de un número creciente de infecciones de en el hombre, el genoma de esta bacteria era aún desconocido a principios del año 2011.

La secuenciación de un genoma bacteriano resulta de gran utilidad para la obtención de información fundamental acerca de la biología de dicha bacteria. En el caso de bacterias patógenas, proporciona además información sobre sus mecanismos de patogenicidad, permitiendo identificar genes que podrían tener un papel importante en el desarrollo de la infección. La secuenciación genómica de varias cepas de una misma especie aporta información sobre las relaciones evolutivas entre las mismas. El elevado grado de diversidad intra-especie observada a menudo entre cepas secuenciadas de la misma especie sugiere que la información obtenida a partir de la secuenciación de un único genoma puede no ser totalmente representativa de la especie (Guzmán *et al*., 2008).

La tecnología de pirosecuenciación 454 de Roche se aplica desde principios de esta década en la secuenciación genómica de bacterias a un coste accesible. Se decidió utilizar la citada tecnología para la secuenciación de dos cepas de *L. garvieae* de importancia clínica: una cepa aislada de un caso de septicemia humana (Aspiroz *et al*., 2007) y un aislado representativo de la cepa epidémica de lactococosis en España (Vela *et al.,* 2000).

La secuenciación de la cepa 21881 supuso la primera liberación del genoma de una cepa de *L. garvieae*. Además, la posterior publicación del genoma de la cepa 8831 permitió la comparación de los genomas de dos cepas de *L. garvieae* que, mediante una aproximación preliminar, reveló una diferencia de tamaño de 0,1 Mb debida a la presencia de cinco plásmidos en la cepa 2881. La publicación de estos resultados abrió nuevas vías para el estudio de la biología de este importante patógeno animal y humano.

JOURNAL OF BACTERIOLOGY, Aug. 2011, p. 4033–4034 Vol. 193, No. 15
0021-9193/11/$12.00 doi:10.1128/JB.05090-11

Genome Sequence of *Lactococcus garvieae* 21881, Isolated in a Case of Human Septicemia▽

Mónica Aguado-Urda,[1] Guillermo H. López-Campos,[2] M. Mar Blanco,[1] José F. Fernández-Garayzábal,[1,3] M. Teresa Cutuli,[1] Carmen Aspiroz,[4] Victoria López-Alonso,[2] and Alicia Gibello[1]*

Department of Animal Health, Faculty of Veterinary Sciences, Complutense University, 28040 Madrid, Spain[1]; Bioinformatics and Public Health Department, Health Institute Carlos III, 28220 Madrid, Spain[2]; Animal Health Surveillance Center (VISAVET), Complutense University of Madrid, Madrid, Spain[3]; and Microbiology Unit, Royo Villanova Hospital, 50015 Zaragoza, Spain[4]

Received 14 April 2011/Accepted 9 May 2011

***Lactococcus garvieae* is a Gram-positive bacterium considered an important opportunistic emerging human pathogen and also a well-recognized fish pathogen. Here, we present the draft genome sequence of *Lactococcus garvieae* strain 21881 (2,164,557 bp, with a G+C content of 37.9%), which represents the first report of a genome sequence on *Lactococcus garvieae*.**

In recent years, *Lactococcus garvieae* has gained recognition as an opportunistic human pathogen, due to the increasing number of clinical cases in which has been involved. In humans, the most common manifestation is infective endocarditis, but it has also been associated with septicemia and urinary and skin infections (7, 14). In addition, *Lactococcus garvieae* is an important fish pathogen, mainly in the trout industry (13). The genetic content of *Lactococcus garvieae* has been studied previously by genomic interspecies microarray hybridization (1). Here, we present the first-draft genome sequence of the *Lactococcus garvieae* species.

Lactococcus garvieae strain 21881, isolated from human blood (from a 74-year-old male affected with septicemia) (4), was grown statically at 28°C in brain heart infusion (BHI) broth (bioMérieux, Marcy l'Etoile, France). Cells were grown until the late exponential phase (optical density at 600 nm [OD_{600}], ~1) and harvested for purification of genomic DNA using a DNeasy blood and tissue kit (Qiagen, Hilden, Germany). A whole-genome shotgun strategy using Roche 454 GS Titanium pyrosequencing was performed. The identified genome sequences were processed with Roche's software. Quality-filtered reads were assembled *in silico* using the 454 Newbler Assembler, version 2.3 (454 Life Sciences), which generated 91 contigs of sizes ranging between 257 and 169,878 bp. Open reading frames (ORFs) were predicted using Genemark.hmm (11). The annotation was done by merging the results obtained from the RAST (Rapid Annotation using Subsystem Technology) server (5), BLAST (3), tRNAscan-SE 1.21 (10), and RNAmmer 1.2 (9). In addition, annotation of the proposed genes was done by searching the contigs against the KEGG (8), UniProt (6), and COG (Clusters of Orthologous Groups) (12) databases to annotate the gene description.

The uncompleted draft genome includes 2,164,557 bases, with a G+C content of 37.9%, and is composed of 2,141 predicted coding sequences (CDSs), with an average size of 723 bp. There are single predicted copies of the 16S and 23S rRNA genes, two copies of the 5S rRNA, and 42 predicted tRNAs. The CDSs annotated by the COG database were classified into 21 functional COG groups. Of these, 1,938 (90%) were assigned a predicted function. Genes not functionally identified yet could be well identified upon closure of the genome. The draft genome contains 273 subsystems (sets of related functional roles) according to the RAST server. We used this information to reconstruct the metabolic network. There are many carbohydrate and protein metabolism features, as well as many genes related to cell wall and capsule biosynthesis. As previously shown, the phospho-beta-galactosidase (*lacG*) gene was absent in this strain (2), but it carries the genes involved in the galactose metabolism. There are also 25 virulence, disease, and defense features, including proteins involved in adhesion and resistance to antibiotics and toxic compounds. Several phage-related genes and some insertion sequences were identified.

The availability of the genome sequence of *Lactococcus garvieae* will provide a better background for future understanding of this organism's pathobiology.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under accession no. AFCC00000000. The version described in this paper is the first version, AFCC01000000.

This study was supported by project AGL2009-12447 of the Spanish Ministry of Science and Innovation. M. Aguado-Urda was a recipient of a Ph.D. grant from the Complutense University of Madrid. G. H. López-Campos's work was supported by the Ministry of Science and Innovation of Spain COMBIOMED RETICS and the European Commission INBIOMEDVISION.

We thank the Lifesequencing team for technical assistance, Daniel Ruiz-Bravo for his assistance in preparing the data for submission, and

* Corresponding author. Mailing address: Department of Animal Health, Faculty of Veterinary Science, Complutense University, Av. Puerta de Hierro s/n, 28040 Madrid, Spain. Phone: 34-91-3943900. Fax: 34-91-3943908. E-mail: gibelloa@vet.ucm.es.

▽ Published ahead of print on 27 May 2011.

Juncal Fernández-Garayzábal for her assistance with the English revision of the manuscript.

REFERENCES

1. **Aguado-Urda, M., et al.** 2010. Analysis of the genome content of *Lactococcus garvieae* by genomic interspecies microarray hybridization. BMC Microbiol. **10:**79.
2. **Aguado-Urda, M., et al.** 2010. Utilization of lactose and presence of the phospho-β-galactosidase (*lacG*) gene in *Lactococcus garvieae* isolates from different sources. Int. Microbiol. **13:**189–193.
3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
4. **Aspiroz, C., J. A. Saez-Nieto, I. Ruiz-Zarzuela, and A. I. Vela.** 2007. Bacteriemia por *Lactococcus garvieae* y *Citrobacter freundii.* Casos de Microbiolgía Clínica. Caso 387. http://www.f-soria.es/admfsoria/casos/img/387.pdf.
5. **Aziz, R. K., et al.** 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics **9:**75.
6. **Bairoch, A., et al.** 2005. The Universal Protein Resource (UniProt). Nucleic Acids Res. **33:**D154–D159.
7. **Chan, J. F. W., et al.** 20 March 2011. Primary infective spondylodiscitis caused by *Lactococcus garvieae* and a review of human *L. garvieae* infections. Infection. doi:10.1007/s15010-011-0094-8.
8. **Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori.** 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. **32:**D277–D280.
9. **Lagesen, K., et al.** 2007. RNammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. **35:**3100–3108.
10. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
11. **Lukashin, A., and M. Borodovsky.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26:**1107–1115.
12. **Tatusov, R. L., E. V. Koonin, and D. J. Lipman.** 1997. A genomic perspective on protein families. Science **278:**631–637.
13. **Vendrell, D., et al.** 2006. *Lactococcus garvieae* in fish: a review. Comp. Immunol. Microbiol. Infect. Dis. **29:**177–198.
14. **Wilbring, M., K. Alexiou, H. Reichenspurner, K. Matschke, and S. M. Tugtekin.** 20 February 2011. *Lactococcus garvieae* causing zoonotic prosthetic valve endocarditis. Clin. Res. Cardiol. doi:10.1007/s00392-011-0286-3.

JOURNAL OF BACTERIOLOGY, Aug. 2011, p. 4263–4264 Vol. 193, No. 16
0021-9193/11/$12.00 doi:10.1128/JB.05326-11

Genome Sequence of *Lactococcus garvieae* 8831, Isolated from Rainbow Trout Lactococcosis Outbreaks in Spain▽

Mónica Aguado-Urda,[1] Guillermo H. López-Campos,[2] Alicia Gibello,[1] M. Teresa Cutuli,[1] Victoria López-Alonso,[2] José F. Fernández-Garayzábal,[1,3] and M. Mar Blanco[1]*

Department of Animal Health, Faculty of Veterinary Sciences, Complutense University, 28040 Madrid, Spain[1]; Bioinformatics and Public Health Department, Health Institute Carlos III, 28220 Madrid, Spain[2]; and Animal Health Surveillance Center (VISAVET), Complutense University of Madrid, 28040 Madrid, Spain[3]

Received 17 May 2011/Accepted 6 June 2011

***Lactococcus garvieae* is the etiological agent of lactococcosis, one of the most important disease threats to the sustainability of the rainbow trout farming industry. Here, we present the draft genome sequence of *Lactococcus garvieae* strain 8831, isolated from diseased rainbow trout, which is composed of 2,087,276 bp with a G+C content of 38%.**

Lactococcus garvieae is the etiological agent of lactococcosis, a septicemic infection affecting different wild and farmed fish species, although its major clinical significance is in the trout industry, where it is responsible for important economic losses. Moreover, it has also been isolated from clinical specimens in other animal species, and in human medicine is considered a potentially opportunistic zoonotic pathogen (12). Recently the genome of a human clinical isolate of *Lactococcus garvieae* has been sequenced (1). Since intraspecific heterogeneity in the genetic content of this species related to host specificity has been suggested (8), we have sequenced the genome of *Lactococcus garvieae* strain 8831, isolated from diseased rainbow trout and responsible for most of the lactococcosis outbreaks in Spain.

Lactococcus garvieae strain 8831 was sequenced by a whole-genome shotgun strategy performed as described before for strain 21881 (1), which generated 87 contigs larger than 200 bases. Open reading frames (ORFs) were predicted using Glimmer 3.02 (6). Functional annotation was done by merging the results obtained from the RAST (Rapid Annotation using Subsystem Technology) server (3), BLAST (2), tRNAscan-SE 1.21 (10), and RNAmmer 1.2 (9) and by searching the contigs against the UniProt (4) and COG (Clusters of Orthologous Groups) (11) databases. The uncompleted draft genome includes 2,087,276 bases with a G+C content of 38% and is composed of 1,969 predicted coding sequences (CDSs). There are 48 predicted tRNAs and single predicted copies of the 16S and 23S rRNA genes, and two copies of the 5S rRNA. The ORFs annotated by the COG database were classified into 20 functional COG groups. The draft genome contains 270 subsystems (sets of related functional roles) according to the RAST server, and we used this information to reconstruct the metabolic network. There are 20 virulence, disease, and defense features, most of them involved in resistance to antibiotics and toxic compounds. Likewise, we identified the gene involved in resistance to clindamycin, which has been described to be characteristic of this species (7). As expected for a *Lactococcus garvieae* strain virulent for fish (5), genes related to capsule biosynthesis were found.

Preliminary comparison of the genome content of strains 21881 (1) and 8831 (this work) shows a difference of approximately 0.1 Mb in their genome sizes. This difference is due to the lack of plasmids in the 8831 strain. The availability of the genome sequence of *Lactococcus garvieae* 8831 will allow deeper comparative genomic studies for the analysis of intraspecific variations in this ubiquitous bacterium, especially focused on the identification of genes potentially involved in virulence.

Nucleotide sequence accession number. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under accession number AFCD00000000. The version described in this paper is the first version, AFCD01000000.

This study was supported by project AGL2009-12447 of the Spanish Ministry of Science and Innovation. M. Aguado-Urda was a recipient of a Ph.D. grant from the Complutense University of Madrid. G. H. López-Campos's work was supported by the Ministry of Science and Innovation of Spain COMBIOMED RETICS and the European Commission INBIOMEDVISION.

We thank Daniel Ruiz-Bravo for his assistance preparing the data for submission.

REFERENCES

1. **Aguado-Urda, M., et al.** 2011. Genome sequence of *Lactococcus garvieae* 21881, isolated from a case of human septicaemia. J. Bacteriol. **193:**4033–4034.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Aziz, R. K., et al.** 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics **9:**75.
4. **Bairoch, A., et al.** 2005. The Universal Protein Resource (UniProt). Nucleic Acids Res. **33:**D154–D159.
5. **Barnes, A. C., et al.** 2002. Resistance to serum killing may contribute to differences in the abilities of capsulate and non-capsulated isolates of *Lactococcus garvieae* to cause disease in rainbow trout (*Oncorhynchus mykiss* L.). Fish Shellfish Immunol. **12:**155–168.
6. **Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673–679.

* Corresponding author. Mailing address: Department of Animal Health, Faculty of Veterinary Science, Complutense University, Av. Puerta de Hierro s/n, 28040 Madrid, Spain. Phone: 34-91-3943718. Fax: 34-91-3943908. E-mail: mmblanco@vet.ucm.es.

▽ Published ahead of print on 17 June 2011.

7. **Elliott, J. A., and R. R. Facklam.** 1996. Antimicrobial susceptibilities of *Lactococcus lactis* and *Lactococcus garvieae* and a proposed method to discriminate between them. J. Clin. Microbiol. **34:**1296–1298.
8. **Kawanishi, M., et al.** 2006. Differences between *Lactococcus garvieae* isolated from the genus *Seriola* in Japan and those isolated from other animals (trout, terrestrial animals from Europe) with regard to pathogenicity, phage susceptibility and genetic characterization. J. Appl. Microbiol. **101:**496–504.
9. **Lagesen, K., et al.** 2007. RNammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. **35:**3100–3108.
10. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of tRNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
11. **Tatusov, R. L., E. V. Koonin, and D. J. Lipman.** 1997. A genomic perspective on protein families. Science **278:**631–637.
12. **Vendrell, D., et al.** 2006. *Lactococcus garvieae* in fish: a review. Comp. Immunol. Microbiol. Infect. Dis. **29:**177–198.

CAPÍTULO V:

Caracterización de los plásmidos presentes en *Lactococcus garvieae* 21881, aislado de un caso clínico de septicemia humana

Resumen

Lactococcus garvieae ha adquirido relevancia en medicina humana durante los últimos años debido al aumento en el número de casos de infecciones descritas en el hombre. Aunque la manifestación más común es la endocarditis, *L. garvieae* ha sido asociado a infecciones muy diversas (Russo *et al*., 2012). La liberación del primer genoma de este patógeno, perteneciente a una cepa aislada de un caso de septicemia humana (Aguado-Urda *et al*., 2011a) y las publicaciones de los genomas de otras cepas de distintos orígenes, abrió nuevas y numerosas posibilidades para el estudio del contenido genético individual, así como de la variación intra-especie en *L. garvieae.*

El estudio de las características únicas de una cepa bacteriana puede proporcionar información sobre la adaptación a un nicho ambiental específico de dicha cepa. En el caso de *L. garvieae*, el estudio de las características únicas de las cepas aisladas de procesos infecciosos en humanos puede acercarnos a la comprensión de los mecanismos que permiten a esta bacteria una colonización eficiente y el desarrollo de dichas infecciones (Reimundo *et al.,* 2011a).

La comparación global de los genomas de cepas aisladas de peces publicadas hasta el momento y la única cepa secuenciada de origen humano, reveló que *L. garvieae* 21881 posee cinco plásmidos que no están presentes en las cepas de peces. Estos plásmidos fueron denominados pGL1-pGL5, y presentaban unos pesos moleculares de 4.536 pb, 4.572 pb, 12.948 pb, 14.006 pb y 68.798 pb, respectivamente. El análisis detallado de las secuencias de estos plásmidos mostró que el material genético contenido en ellos proviene, en su mayoría, de procesos de transferencia genética horizontal con otras bacterias lácticas. Además, algunos de los genes plasmídicos caracterizados representan factores de virulencia potencialmente implicados en el desarrollo de la infección por *L. garvieae* 21881. Entre los resultados más significativos destacan la presencia de genes en los plásmidos pGL1, pGL2 y pGL5 que codifican distintas proteínas relacionadas con la síntesis y secreción de bacteriocinas. Asimismo, el plásmido pGL5 contiene diferentes genes que codifican proteínas que podrían estar implicadas en la adherencia de *L. garvieae* al mucus intestinal, así como en la interacción con las células epiteliales intestinales y otras células ricas en colágeno como las válvulas cardiacas.

OPEN ACCESS Freely available online

PLoS one

Characterization of Plasmids in a Human Clinical Strain of *Lactococcus garvieae*

Mónica Aguado-Urda[1☯], Alicia Gibello[1*☯], M. Mar Blanco[1], Guillermo H. López-Campos[2], M. Teresa Cutuli[1], José F. Fernández-Garayzábal[1,3]

1 Faculty of Veterinary Sciences, Department of Animal Health, Complutense University, Madrid, Spain, **2** Bioinformatics and Public Health Department, Health Institute Carlos III, Madrid, Spain, **3** Animal Health Surveillance Center (VISAVET), Complutense University of Madrid, Spain

Abstract

The present work describes the molecular characterization of five circular plasmids found in the human clinical strain *Lactococcus garvieae* 21881. The plasmids were designated pGL1-pGL5, with molecular sizes of 4,536 bp, 4,572 bp, 12,948 bp, 14,006 bp and 68,798 bp, respectively. Based on detailed sequence analysis, some of these plasmids appear to be mosaics composed of DNA obtained by modular exchange between different species of lactic acid bacteria. Based on sequence data and the derived presence of certain genes and proteins, the plasmid pGL2 appears to replicate via a rolling-circle mechanism, while the other four plasmids appear to belong to the group of lactococcal theta-type replicons. The plasmids pGL1, pGL2 and pGL5 encode putative proteins related with bacteriocin synthesis and bacteriocin secretion and immunity. The plasmid pGL5 harbors genes (*txn, orf5* and *orf25*) encoding proteins that could be considered putative virulence factors. The gene *txn* encodes a protein with an enzymatic domain corresponding to the family actin-ADP-ribosyltransferases toxins, which are known to play a key role in pathogenesis of a variety of bacterial pathogens. The genes *orf5* and *orf25* encode two putative surface proteins containing the cell wall-sorting motif LPXTG, with mucin-binding and collagen-binding protein domains, respectively. These proteins could be involved in the adherence of *L. garvieae* to mucus from the intestine, facilitating further interaction with intestinal epithelial cells and to collagenous tissues such as the collagen-rich heart valves. To our knowledge, this is the first report on the characterization of plasmids in a human clinical strain of this pathogen.

Citation: Aguado-Urda M, Gibello A, Blanco MM, López-Campos GH, Cutuli MT, et al. (2012) Characterization of Plasmids in a Human Clinical Strain of *Lactococcus garvieae*. PLoS ONE 7(6): e40119. doi:10.1371/journal.pone.0040119

Editor: Paul Jaak Janssen, Belgian Nuclear Research Centre SCK/CEN, Belgium

Received February 3, 2012; **Accepted** June 1, 2012; **Published** June 29, 2012

Funding: This study was supported by projects AGL2009-12447 and the Spanish RETIC COMBIOMED RD07/0067 funded by the Spanish Ministry of Science and Innovation and the INBIOMEDvision project funded by the European commission 7th framework programme grant agreement no: 270107. Dr. Aguado-Urda was a recipient of a PhD grant from the Complutense University of Madrid. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gibelloa@vet.ucm.es

☯ These authors contributed equally to this work.

Introduction

Lactococcus garvieae is a ubiquitous and widely distributed microorganism that has relevance in veterinary and human medicine. The increasing number of cases of human infections by *L. garvieae* reported in recent years have caused it to be considered an opportunistic emerging human pathogen. The most common manifestation is infective endocarditis involving either native or prosthetic valves [1–4], but it has also been associated with different clinical processes such as septicaemia, urinary infections and skin infections in healthy and immunocompromised patients [5–7]. *L. garvieae* is also an important bacterial fish pathogen responsible for lactococcosis, a septicemic infection affecting various wild and farmed fish species, particularly in trout [8]. It has also been isolated from clinical specimens in other animal species, such as cows and water buffalos with subclinical mastitis and pigs with pneumonia, and from cat and dog tonsils [8–10]. *L. garvieae* has also been found in vegetables, meat and sausages, but mainly in artisanal dairy products [11–13]. Because of this, *L. garvieae* is considered a potential emerging zoonotic pathogen [14,15].

Recently, the complete genome sequences of four clinical strains of *L. garvieae* isolated from yellowtail and trout [16–19], one human clinical strain [20] and one dairy strain [19] have been published. A preliminary genome comparison of the fish and human strains of *L. garvieae* 8831 and 22881, respectively, showed a difference in their genome sizes of approximately 0.1 Mb [16,20]. As shown in the present work, this difference is related to the presence of five plasmids found in the human strain. Plasmids are commonly found in many members of lactic acid bacteria [21–24], encoding relevant properties such as additional amino acid and carbohydrate metabolism, proteolysis activities, exopolysaccharide biosynthesis, bacteriophage resistance, bacteriocin production, drug resistance or virulence factors. Unlike other species of *Lactococcus*, such as *Lactococcus lactis* and *Lactococcus salivarius*, in which the genetic content of different plasmids have been studied and characterized [24,25], the data for *L. garvieae* are very limited. Fortina *et al.* [26] observed that *L. garvieae* strains of dairy origin often harbor plasmids, and Reimundo *et al.* [27] reported the presence of a 30-kb plasmid in a human clinical strain; however, neither study characterized the plasmids. There is only one study

that characterized a conjugative plasmid carrying multiple drug resistance genes in *L. garvieae* strains that were isolated from yellowtails [28]. In this work, we present the sequence analysis and characterization of five circular plasmids found in the human *L. garvieae* strain 21881.

Materials and Methods

Bacterial Strains and Growth Conditions

Lactococcus garvieae 21881 was isolated from the blood of a 74-year-old male patient affected by septicaemia [20]. The bacteria were grown in MRS broth (Cultimed, Panreac Laboratories) and incubated at 30°C for 24 h. The strain 8831, isolated in Spain in 2004 from diseased trout affected by lactococcosis [16], was used as a reference.

Plasmid Isolation and Plasmid Stability Tests

The isolation of plasmid DNA was performed using log-phase cultures (OD_{600}, ~1) using the method of Anderson and McKay [29]. The plasmid profile was observed by electrophoresis of 15 µL of each sample on a 0.6% (w/v) agarose gel supplemented with 1X Syber safe® (Invitrogen, Eugene, OR, USA).

The stability of the plasmids was determined after growing the cells in MRS liquid cultures for approximately 100 generations. Briefly, one colony of *L. garvieae* 21881 was grown overnight in MRS broth for 16 h (approximately 20 generations) for a total of 5 days. After this time without selection (approximately 100 generations), 50 colonies were picked and patched onto MRS agar plates (MRS broth supplemented with 1.5% agar). The patches were tested for plasmid isolation, and the plasmid stability was calculated as the percentage of patches or clones in the population that maintained the plasmid content.

DNA Manipulation

PCR primers for gap closure were designed based on DNA sequence information at the end of the corresponding contigs. In vitro amplification reactions of DNA were performed in a reaction mixture of 100 µL containing DNA template (10–20 ng), 1 µM of each primer, 100 µM of each dNTP (Biotools), 5 U of Ultratools DNA polymerase (Biotools) and its 1X amplification buffer. The amplifications were performed in a Mastercycler gradient thermal cycler (Eppendorf) following the optimal cycle profile: an initial denaturation step of 95°C for 5 min, 35 serial cycles of a denaturation step of 95°C for 45 seconds, annealing at the optimal annealing temperature corresponding to each primer set for 1 min and extension at 72°C for 2 min, followed by a final extension step of 72°C for 10 min. Negative controls (no DNA template) were included in each batch of PCR reactions. PCR-generated products were detected by electrophoresis of 5 µL of each amplification mixture on 1% agarose gels supplemented with 1X Syber safe®. The amplicons were purified (Geneclean Turbo Kit, MP Biomedicals LLC) and sequenced at SECUGEN facilities (Centro de Investigaciones Biológicas, CSIC, Spain) using the DyeDeoxy (dRhodamine) Terminator Cycle Sequencing kit in an automatic ABI Prism DNA sequencer.

Plasmid DNA Sequencing, Sequence Assembly and Annotation

The sequences of the five plasmids of *L. garvieae* 21881 were obtained from published whole genome sequencing data [20]. The whole sequences corresponding to each plasmid were completed by PCR reactions, which allowed us to join and fill the gaps of contigs: c42, c41 and c80 (pGL5); c20 (pGL4); c53 (pGL3); c102 (pGL2) and c101 (pGL1).

Open reading frames (Orfs) were identified using a combination of the GeneMark.hmm for prokaryotes (v2.8) program [30] with the *Lactococcus lactis* genome as a reference (http://exon.gatech.edu/gmhmm2_prok.cgi), and the ORF Finder tool (http://www.ncbi.nlm.nih.gov/projects/gorf/) from the National Center for Biotechnology Information (NCBI). Potential Orfs were subsequently manually filtered using the following criteria: i) only Orfs whose DNA sequences did not overlap with those of other Orfs or doing so in less than 21 nucleotides were considered, and ii) only Orfs larger than 50 amino acids (aa) were considered. The amino acid sequences of selected proteins were further analyzed using the BLASTp program and the NCBI's nonredundant protein database. Putative functions were assigned on the basis of the best BLASTp hit on an annotated protein. The predicted proteins were functionally categorized by using the clusters of orthologous groups (COG), conserved domain (CDD), the TIGR Gene Indices and the protein family (pfam) databases.

Figures of physical and genetic maps of plasmids pGL1-pGL5 and DNA sequence similarities with plasmids of other Gram-positive bacteria were generated by means of the BLAST Ring Image Generator (BRIG) using the default parameters [31].

Signal peptides and cleavage sites in Gram-positive bacteria amino acid sequences were predicted using the SignalP 4.0 Server [32] available at http://www.cbs.dtu.dk/services/SignalP/.

Horizontal Gene Transfer (HGT) Analysis

To identify putative horizontally transferred genes, sequence composition analysis was performed for some genes, including the calculation of GC composition and dinucleotide dissimilarity value (δ*) using the δρ-WEB tool (http://deltarho.amc.uva.nl/cgi-bin/bin/start.cgi). A high genomic dissimilarity between an input sequence and a representative genome sequence of the species from which the sequence was isolated suggests a heterologous origin of the input sequence and this difference can be expressed by the δ* value [33]. DNA fragments with different GC composition and/or a high dissimilarity value compared with those of the whole genome of *L. garvieae* 21881 were predicted to be HGT genes. The predicted horizontally transferred genes and gene clusters were checked for HGT mechanism-associated features such as neighboring mobile elements. Homologous sequences were aligned with MUSCLE [34], and phylogenetic trees were constructed using Phylogeny.fr [35]. The positions of orthologs from *L. garvieae* and other acid lactic bacteria in the phylogenetic trees were checked to confirm whether the predicted genes are horizontally transferred between genomes.

Susceptibility to Quinolones

The antimicrobial susceptibility to nalidixic acid, oxolinic acid, flumequine, ciprofloxacin, norfloxacin and moxifloxacin of the strain *L. garvieae* 21881 was determined by the disk diffusion method using commercially prepared antimicrobial disks (Oxoid, Ltd.). Inoculum was prepared from a 48 h Columbia blood agar plate by suspending four colonies in 5 mL of PBS and adjusted to a 0.5 McFarland standard. The disk diffusion test was performed as described by the Clinical and Laboratory Standards Institute [36]. The Mueller-Hinton-blood agar plates were examined after 24 h of incubation at 30°C. *Staphylococcus aureus* ATCC 25923 and *Escherichia coli* ATCC 25922 were included as quality control. As no specific inhibition zone diameter (IZD) breakpoints for *L.garvieae* are available, the IZD breakpoints were those recommended by the French Society of Microbiology [37] for testing Gram-positive microorganisms.

The amino acid sequences of GyrA, GyrB, ParC and ParE were obtained from their respective genes located in contigs 61, 39, and

61 [20]. The amino acid sequences of GyrA and ParC from *L. garvieae* 21881, including the quinolone resistance-determining regions (QRDR) were compared to those described for the quinolone sensitive *L. garvieae* strain KL99110 [38].

Nucleotide Sequence Accession Numbers

The DNA sequences and annotations corresponding to the five plasmids (pGL1 to pGL5) found in *L. garvieae* 21881 have been deposited in the EMBL database under the accession numbers HE650695, HE650696, HE650697, HE651325 and HE651326.

Results and Discussion

Plasmid Content and Sequence Analysis

The comparison between *L. garvieae* 21881 and the fish isolate *L. garvieae* 8831, used as reference, showed that *L. garvieae* 21881 contains five plasmids (Figure 1). We designated these plasmids pGL1, pGL2, pGL3, pGL4 and pGL5 in order of size from small to large; the nucleotide sequence of these plasmids determined molecular sizes of 4,536 bp, 4,572 bp, 12,948 bp, 14,006 bp and 68,798 bp, respectively. These plasmids explain the approximately 0.1 Mb difference in the genome of this strain compared with *L. garvieae* 8831 [16,20], which is lacking the plasmids (Figure 1). Alignments of pGL1-pGL5 and plasmid pKL0018 present in *L. garvieae* strains isolated from yellowtail (accession number AB290882) [28] were performed using the BLAST two sequences tool (www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). Significant DNA sequence similarity (66%) was only found between pGL2 and pKL0018 in the region containing the *repB* gene. The similarity of pGL1-pGL5 plasmids was also searched against the whole-genome shotgun contigs and nucleotide collection databases using the BLASTn program. Overall, the search results showed that the DNA sequences of pGL1-pGL5 plasmids were not present in any of the *L. garvieae* strains for which whole genome sequences are available. These results indicate that pGL1-pGL5 plasmids were solely present in our human strain 21881.

The GC contents of the plasmids were 35.3%, 37.4%, 36.3%, 32.17% and 34.3%, for pGL1 to pGL5, respectively. These values are slightly lower than the values described for *L. garvieae* chromosomal DNA (38–39%) [16–20] but are within the range (30–40%) exhibited by most of the lactococcal plasmids [39,40]. According to the criteria specified above, 91 Orfs were selected and are listed in Tables S1, S2, S3, S4, S5. All of the Orfs had an ATG start codon except for five starting with a TTG and one by a GTG codon. Putative biological functions were assigned to most of the Orfs. The genetic organization of each plasmid is depicted in Figures 2, 3, 4, 5, 6, and the details are presented in Tables S1, S2, S3, S4, S5.

The majority of the genes present on pGL1-pGL5 exhibited homology with genes located on lactococcal chromosomes (mainly *L. lactis*) or other lactococcal plasmids (Figures 2, 3, 4, 5, 6, Tables S1, S2, S3, S4, S5). However, some genes also shared a high percentage of DNA similarity (>90%) with genes present in other genera of lactic acid bacteria, mainly *Leuconostoc* but also *Lactobacillus*, *Enterococcus*, *Streptococcus* and *Weissella* (Tables S1, S2, S3, S4, S5). In the case of the pGL4 plasmid, the comparison of the nucleotide sequence of this plasmid with those in the databases revealed two different patterns of similarity (Figure 5). The region located between nucleotides 447 and 2,458 showed 99% DNA sequence similarity to a region of the plasmid pSRQ900 from *L. lactis* [41], coding for RepB and RepX proteins. In contrast, regions located between nucleotides 3,750 and 4,570 and between nucleotides 5,780 and

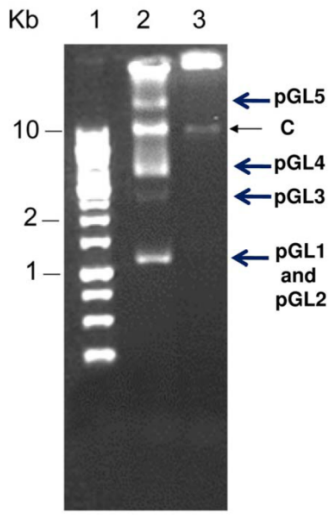


Figure 1. Gel electrophoresis of plasmid DNA from *L. garvieae* 21881 and *L. garvieae* 8831. Line 1, Biotools 1 kb DNA marker, line 2, *L. garvieae* 21881, line 3, *L. garvieae* 8831; C indicates the chromosomal DNA.
doi:10.1371/journal.pone.0040119.g001

8,166 showed 99% DNA sequence similarity to DNA regions on the *Leuconostoc citreum* plasmid pLCK4 [42]. Several genes on pGL4 (Table S4) exhibited DNA sequences identities higher than 90% with chromosomal or plasmidic genes of *Streptococcus parauberis* and different *Leuconostoc* species. Two IS elements were identified on the pGL4 plasmid (Table S4); one is similar to a transposase of an *IS946*-like element (an *ISS1*-family element) and the other to an inactive transposase of an *IS30* family element that is commonly found in the genome of *L. citreum* KM20 [42]. Moreover, *orf4* encodes an ATPase involved in DNA repair, which is 99% DNA identical to its homologous gene in *L. citreum* and *orf3* is 100% DNA identical to its homologous gene present in the plasmid pLCK4 (Figure 5) [42]. Likewise, on pGL3, *lai* showed 97% DNA sequence identity to that of *Weissella paramesenteroides* (Table S3) but only 63% DNA identity with the homologous chromosomal gene of *L. garvieae* (accession numbers CBK55550 and AFCC01000000). The genes *pox*, *orf2* and the sequence of insertion IS1297 exhibited DNA sequence identity percentages higher than 90% with genes found in *Lactobacillus buchneri*, *Enterococcus faecalis* and *Leuconostoc* sp., respectively (Table S3). IS elements are known to play an important role in lactococcal gene transfer. In contrast, site-specific recombinases have been studied less in *Lactococcus* and related bacteria. The *int* gene of pGL5 encodes a serine recombinase that is highly identical to the invertase/resolvase subfamily proteins (pfam 002399) from different lactic acid bacteria. This predicted protein shares 77% aa sequence identity to that of *Tetragenococcus halophilus* (Table S5). The presence of mobile elements such as transposon-related genes, IS elements, resolvases, integrases and relaxases, suggests that some of the unique Orfs found within the plasmids may have been acquired from other species via genetic exchange events. The high values of genomic dissimilarity, the differences in CG

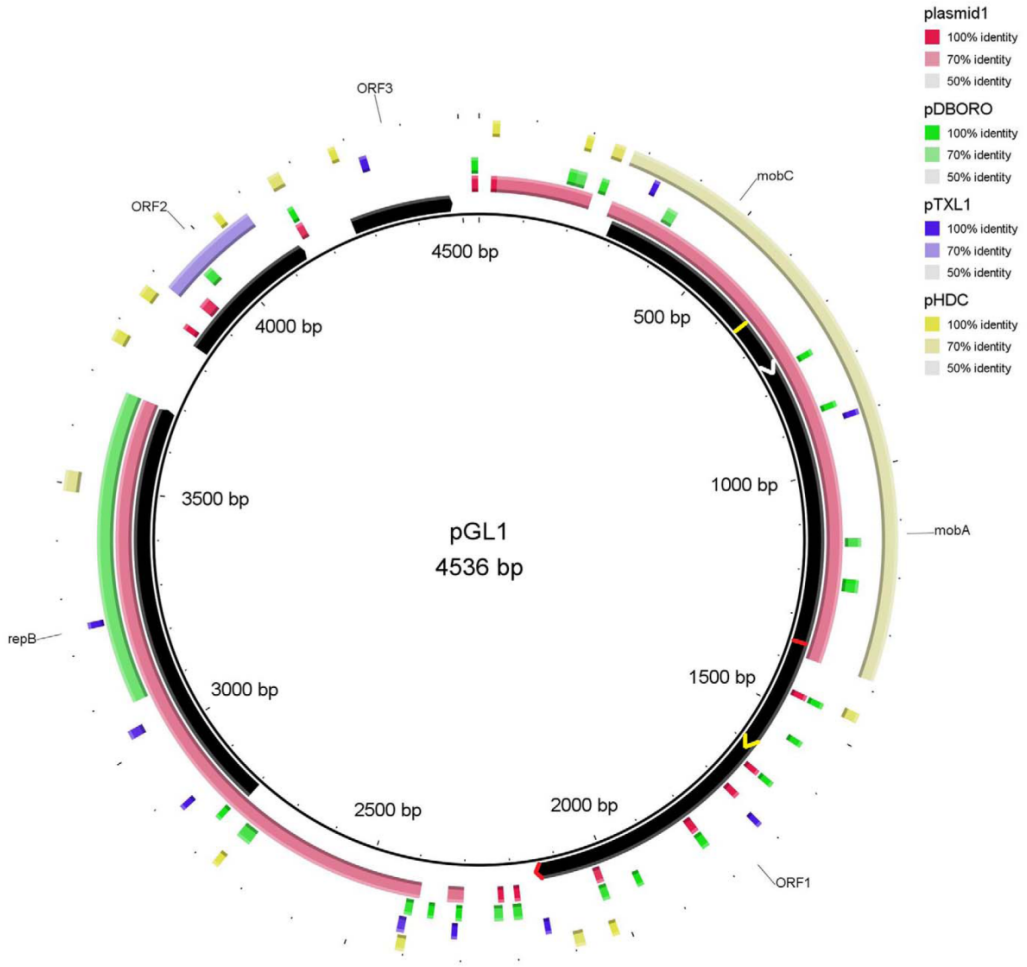


Figure 2. Plasmid map of pGL1 and DNA sequence similarity with other lactic bacteria plasmids. *Lactococcus lactis* subsp. *cremoris* SK11 plasmid1, *Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* plasmid pDBORO, *Leuconostoc mesenteroides* subsp. *mesenteroides* Y110 plasmid pTXL1 and *Tetragenococcus halophilus* plasmid pHDC were those exhibiting the highest coverage/similarity to pGL1. The Orfs are indicated by arrows showing the direction of transcription. The gene annotations are positioned at the midpoint of each gene. Overlapped genes *mobC, mobA* and *orf1* are indicated in white, yellow and red respectively.
doi:10.1371/journal.pone.0040119.g002

composition and the presence of transfer mechanisms-associated features for the analyzed genes (Table 1) together with the phylogenetic data (Figure 7), indicate that there have been HGT events between *L. garvieae* and other lactic acid bacteria as has been frequently observed among different species of this group of microorganisms [21,40,43]. The fact that most of these lactic acid bacteria (Figures 2, 3, 4, 5, 6, Tables S1, S2, S3, S4, S5) are usually present in dairy products is in line with the hypothesis that some dairy isolates of *L. garvieae* could be responsible of human infections [4].

Replication and Maintenance Systems

Lactococcal plasmids can replicate by two different mechanisms: theta and rolling circle replication (RCR).

Plasmid pGL2 appears to be a RCR plasmid, as suggested by the homology of its replication gene to those of other known RCR plasmids, such as pWCFS102 of *Lactobacillus plantarum*, pYSI8 of *Lactobacillus sakei* (Figure 3) [44,45] and pSSU1 of *Streptococcus suis* [46], all members of the rolling-circle replication pMV158 family. A putative double-stranded origin (*dso*) was identified at coordinates 531–676, which shares 95% identity with the *dso* of the plasmids pSMQ172 of *Streptococcus thermophilus* [47] and pSSU1 of

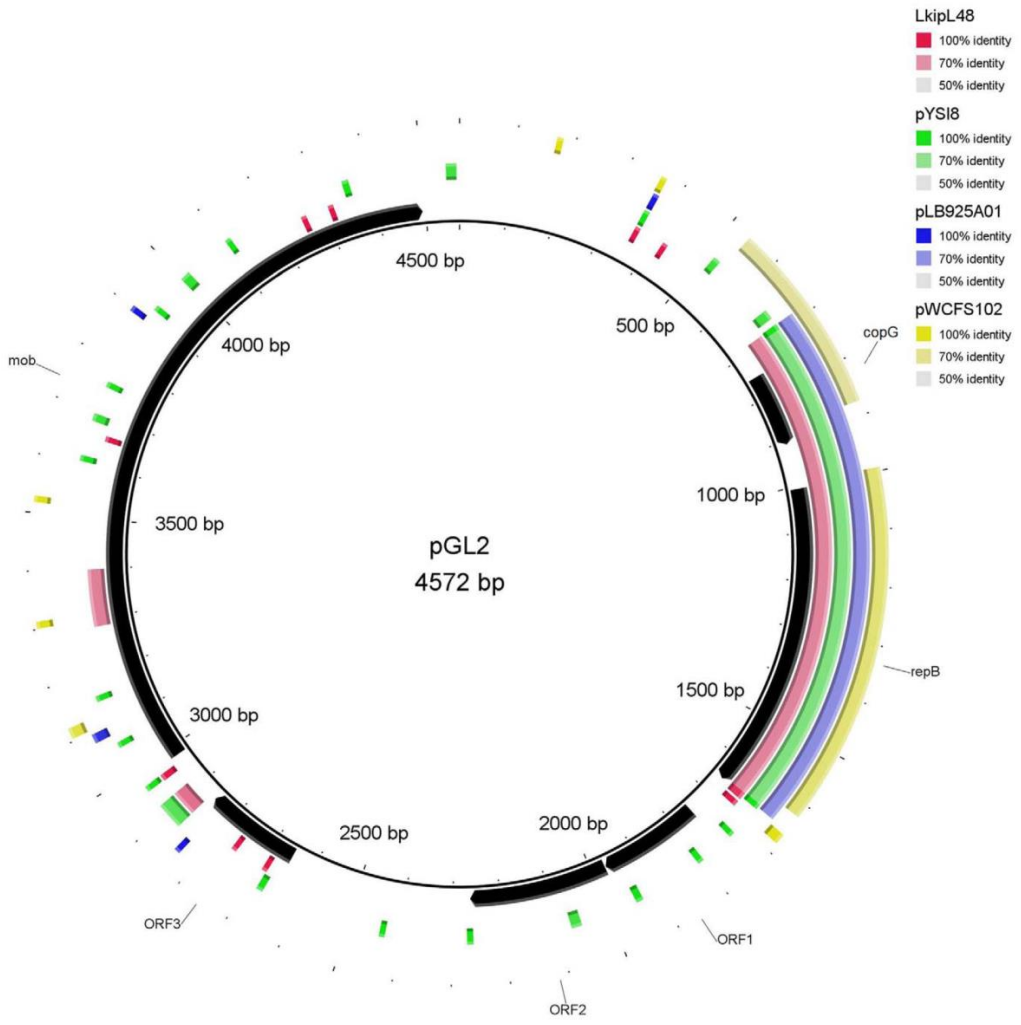


Figure 3. Plasmid map of pGL2 and DNA sequence similarity with other lactic bacteria plasmids. *Leuconostoc kimchii* IMSNU 11154 plasmid LkipL48, *Lactobacillus sakei* plasmid pYSI8, *Lactobacillus brevis* plasmid pLB925A01 and *Lactobacillus plantarum* WCFS1 plasmid pWCFS102 were those exhibiting the highest coverage/similarity to pGL2. The Orfs are indicated by arrows showing the direction of transcription. The gene annotations are positioned at the midpoint of each gene.
doi:10.1371/journal.pone.0040119.g003

S. suis [46]. In the case of the plasmids of the pMV158 family, the *dso* can be physically and functionally separated into two loci, termed *bind* (the binding region of the Rep protein) and *nic* (where the Rep protein cleaves specifically at the nick site, which is conserved among plasmids of the pMV158 family) [48]. Upstream of the *copG* gene, at position 542–550 bp, the conserved 9-mer nick sequence site TACTACGAC was identified. Next to this nick sequence, the inverted repeats IR-I elements that form the hairpin where the 5′-GpA dinucleotide is cleaved by RepB, and the two proximal direct repeats (PDRI and PDRII), were also detected (Figure 8). The locus *bind*, which in the case of pGL2 consists of four tandem 11-bp direct repeats (DR), was located 52 bp downstream of the locus *nic* (Figure 8). Another important element in the replicative process of RCR plasmids is the lagging-strand replication origin (*sso*), which is generally located at a short distance upstream of the *dso*. The *sso*s usually have extensive secondary structure, and unlike the *dso*s, their sequences are generally not homologous among plasmids belonging to the same family [25,49]. A putative *sso*-like region was found in pGL2 immediately upstream of the *dso*. This region contains several

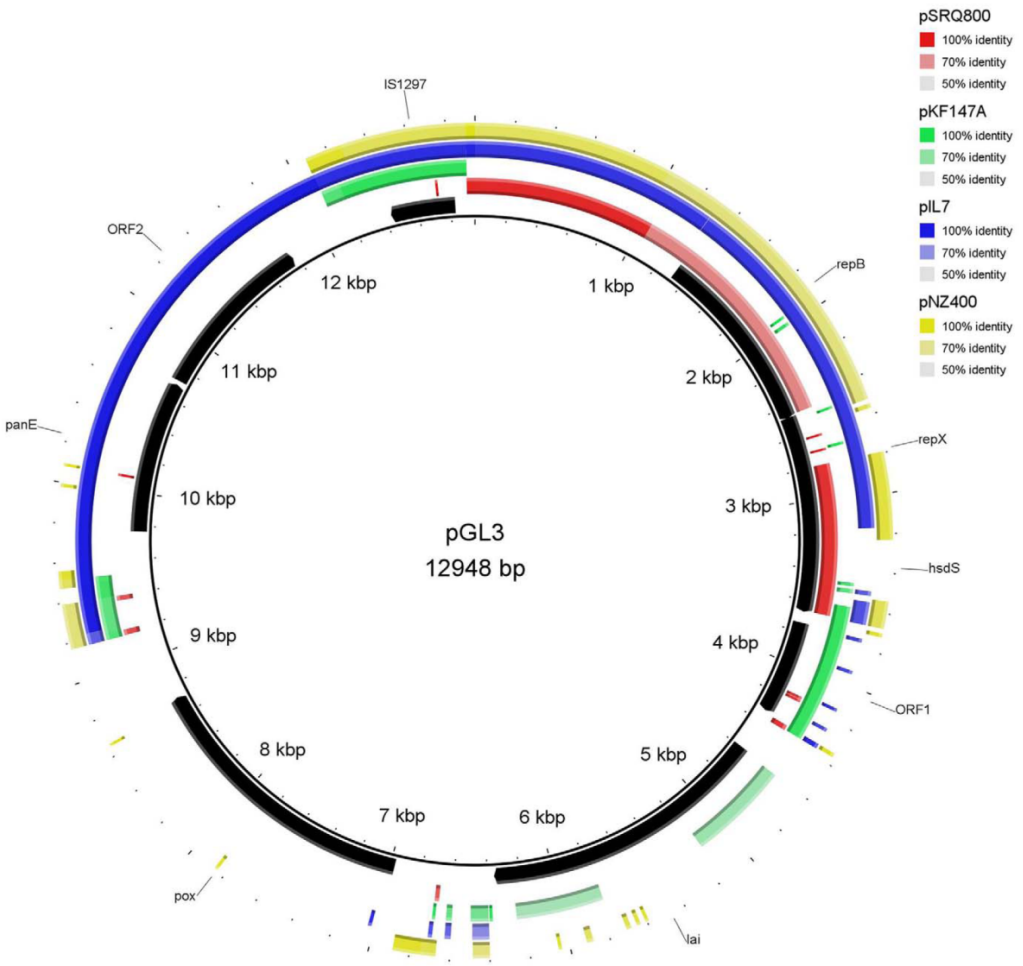


Figure 4. Plasmid map of pGL3 and DNA sequence similarity with other lactic bacteria plasmids. *Lactococcus lactis* plasmid pSRQ800, *Lactococcus lactis* subsp. *lactis* KF147 plasmid pKF147A, *Lactococcus lactis* subsp. *lactis* plasmid pIL7 and *Lactococcus lactis* subsp. *cremoris* plasmid pNZ4000 were those exhibiting the highest coverage/similarity to pGL3. The Orfs are indicated by arrows showing the direction of transcription. The gene annotations are positioned at the midpoint of each gene.
doi:10.1371/journal.pone.0040119.g004

inverted repeats sequences, which could generate stem-loop structures.

The plasmids pGL1, pGL3, pGL4 and pGL5 appear to be theta-replicating plasmids. The replication backbone of these plasmids contains an AT-rich region, the presumed origin of replication, which is located upstream the *rep* gene. This region is followed by three directly repeated sequences of 22 bp (DR, iterons) that are thought to interact directly with the Rep protein to initiate replication. Finally, following the iterons, there are two inverted repeats of 5 to 12 bp, designated as IRa and IRb, that usually overlap the putative −10 and −35 regions of the *rep* gene promoter [41]. The sequence characteristics of this region in plasmids pGL1, pGL3, pGL4 and pGL5 are shown in Figure 9. Plasmid pGL1 carries the *repB* gene (Figure 2) that encodes the replication initiation protein RepB (296 aa). RepB is very similar to the Rep proteins from the *L. lactis* subsp. *cremoris* plasmid 1 (Table S1) and *L. lactis* subsp. *lactis* plasmid pDBORO [22,50]. Downstream of the *repB* gene (position 3,858) lies *orf2* (Figure 2), which shared 59% aa sequence identity to a hypothetical protein of the plasmid pTXL1 from *Leuconostoc mesenteroides* subsp. *mesenteroides* that has a putative role in replication [51]. The pGL3 and pGL4 plasmids carry a replication gene *repB* that encodes a replication initiation protein RepB of 383 and 386 aa, respectively. Both RepB proteins contain the Rep3 superfamily

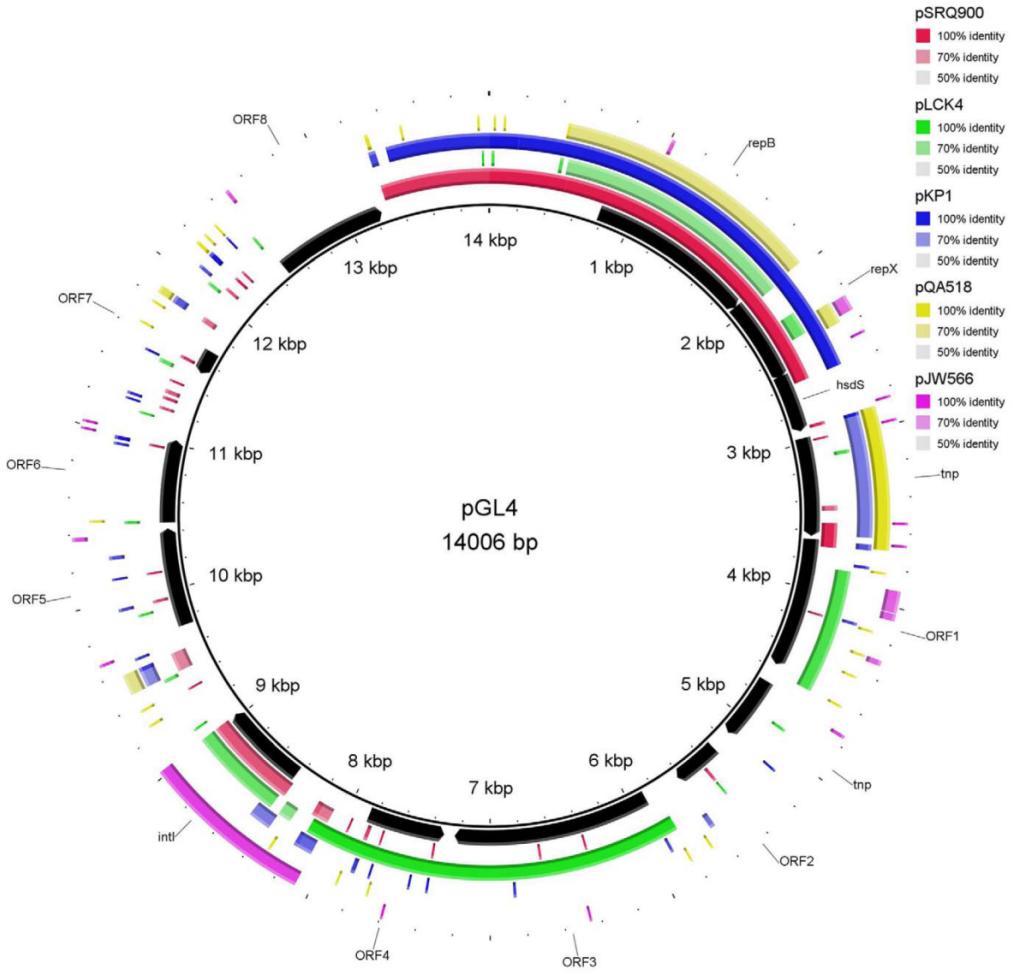


Figure 5. Plasmid map of pGL4 and DNA sequence similarity with other lactic bacteria plasmids. *Lactococcus lactis* plasmid pSRQ900, *Leuconostoc citreum* KM20 plasmid pLCK4, *Lactococcus lactis* subsp. *lactis* plasmid pKP1, *Lactococcus lactis* subsp. *cremoris* A76 plasmid pQA518 and *Lactococcus lactis* plasmid pJW566 were those exhibiting the highest coverage/similarity to pGL4. The Orfs are indicated by arrows showing the direction of transcription. The gene annotations are positioned at the midpoint of each gene.
doi:10.1371/journal.pone.0040119.g005

(pfam01051), and the *L. lactis* RepBC superfamily (pfam06430) conserved motifs and shared a high level of aa sequence identity to initiator proteins of lactococcal theta-replicating plasmids [41,52]. As often observed in theta-type replicons, a conserved region *repX-hsdS* was observed downstream of *repB* in pGL3 and pGL4. The *repX* gene (also known as *orfX*) is usually overlapped by one or two codons by the *repB* gene [24,41]. *repX*, very common in theta-replicating plasmids, is not essential for their replication, but in some plasmids, it participates in the control of plasmid copy number, plasmid stability or both [25,53]. The last gene of this transcription replication-module encodes HsdS, the specific S subunit of a type I restriction modification system [24]. pGL5 carries two replication genes that belong to the *L. lactis* RepBC-terminus superfamily. The *repA* and *repB* genes encode proteins that shared significant similarity (75–80%) with the replication initiator proteins from different lactococcal plasmids such as pSK11L [21], pIL4 [24], pCV56A, [54] and pS7a [55]. Both *rep* genes were found in the same orientation and had a 75% DNA similarity with each other.

Plasmid replication and plasmid stability are closely related processes. Plasmid stabilization requires accurate control mechanisms such as the proper plasmid copy number, plasmid multimer resolution, postsegregational killing and an active partitioning system [25]. In *L. garvieae* 21881, no plasmid loss was observed

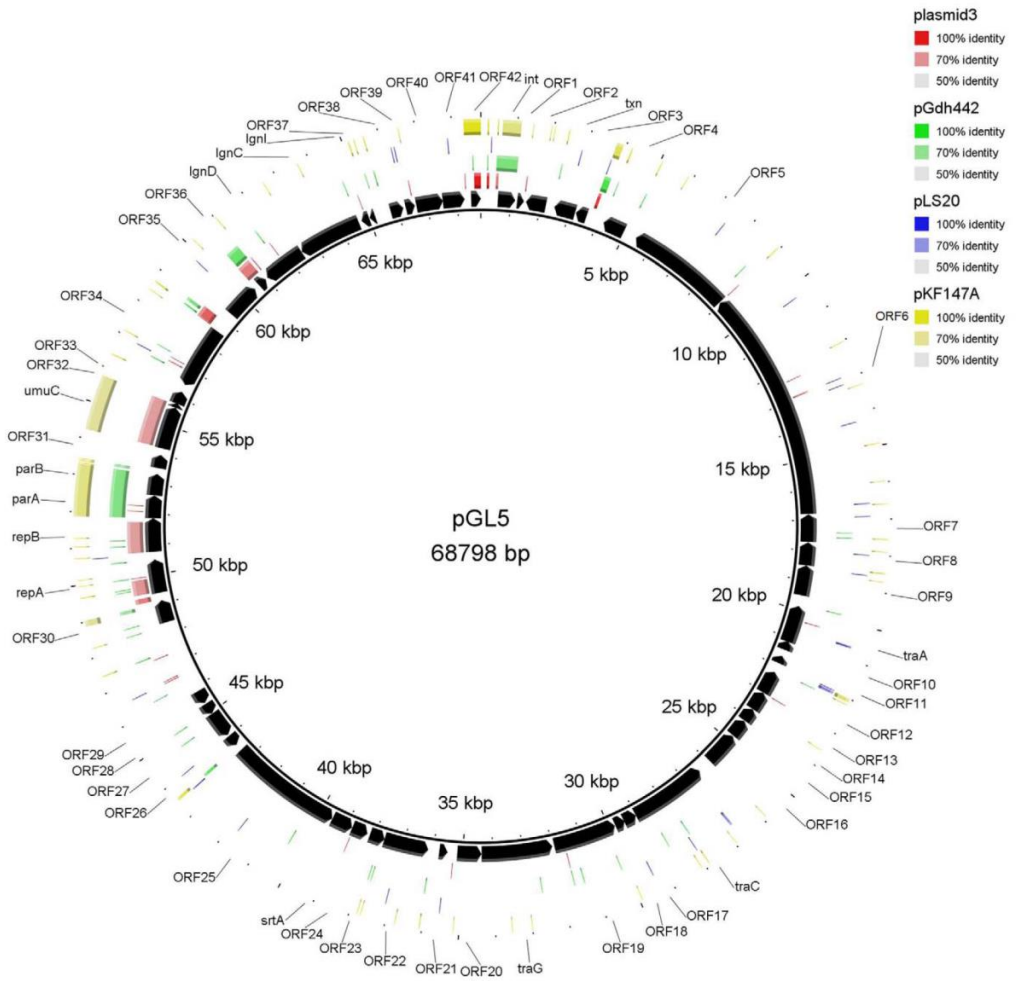


Figure 6. Plasmid map of pGL5 and DNA sequence similarity with other lactic bacteria plasmids. *Lactococcus lactis* subsp. *cremoris* SK11 plasmid 3, *Lactococcus lactis* plasmid pGdh442, *Bacillus subtilis* subsp. *natto* plasmid pLS20 and *Lactococcus lactis* subsp. *lactis* KF147 plasmid pKF147A were those exhibiting the highest coverage/similarity to pGL5. The Orfs are indicated by arrows showing the direction of transcription. The gene annotations are positioned at the midpoint of each gene.
doi:10.1371/journal.pone.0040119.g006

after approximately 100 generations of culturing in MRS medium without selection pressure, suggesting that the five plasmids possess stabilisation mechanisms. In pGL1, *orf2* encodes a protein with a putative role in replication control similar to that of the pTXL1 plasmid from *L. mesenteroides* [51] (Figure 2, Table S1). pGL2 also contains a plasmid replication copy-number control protein (CopG) similar to that found in *Lactobacillus brevis* (pLB925A01) and other plasmids of the pMV158 family (Figure 3). In these plasmids, two elements, the product of *copG* and a ctRNA (designated RNA II), are involved in the control of the synthesis of the Rep protein [56]. The gene that encodes the ctRNA (*maII*) was located between the genes *copG* and *repB* and it reads in the opposite direction. Figure 8 shows the putative *maII* gene, with its corresponding putative promoter and terminator regions. Both pGL3 and pGL4 carry the *repX* gene that, in *L. lactis*, participates in the control of plasmid copy number and/or plasmid stability [25,53]. The genes *parA* and *parB* on pGL5 encode putative partitioning determinants with a high aa identity (76% and 69%, respectively) to homologous proteins ParA and ParB from other *L. lactis* plasmids [24]. This partitioning system is likely to contribute to the segregational stability of pGL5, avoiding the risk of loss during cell division [24,40].

Table 1. Examples of horizontally transferred genes in *L. garvieae* 21881 plamids.

Orf (plasmid)	% GC	Genomic dissimilarity value (δ*) (10^3)	δ* plot position (1-100%)	HGT mechanism associated feature	Predicted function	Best BLASTP hit
orf2 (pGL1)	30.67	214.347	94.617	Mobile plasmid	Putative role in replication	*Leuconostoc mesenteroides*
orf3 (pGL1)	42.92	221.466	89.585	Mobile plasmid	Bacteriocin-like protein	*Enterococcus faecalis*
orf1 and *orf2* (pGL2)	31.84	173.000	93.41	Mobile plasmid	Bacteriocin-like protein and Enterocin A-like protein	*Streptococcus mitis*
lai (pGL3)	36.85	104.161	90.80	Transposase	Linoleate isomerase	*Weisella paramesenteroides*
pox (pGL3)	42.93	112.894	93.419	Transposase	Pyruvate oxidase	*Lactobacillus buchneri*
orf4 (pGL4)	26.26	164.409	92.172	Transposase	ATPase involved in DNA repair	*Leuconostoc citreum*
orf8 (pGL4)	32.95	140.024	90.215	Transposase	Methyl-transferase	*Streptococcus parauberis*
orf25 (pGL5)	39.49	94.570	96.064	Transposase	Collagen-binding LPXTG protein	*Enterococcus faecalis*
orf27 (pGL5)	40.22	163.538	97.523	Transposase	Conserved hypothetical protein	*Enterococcus faecalis*

doi:10.1371/journal.pone.0040119.t001

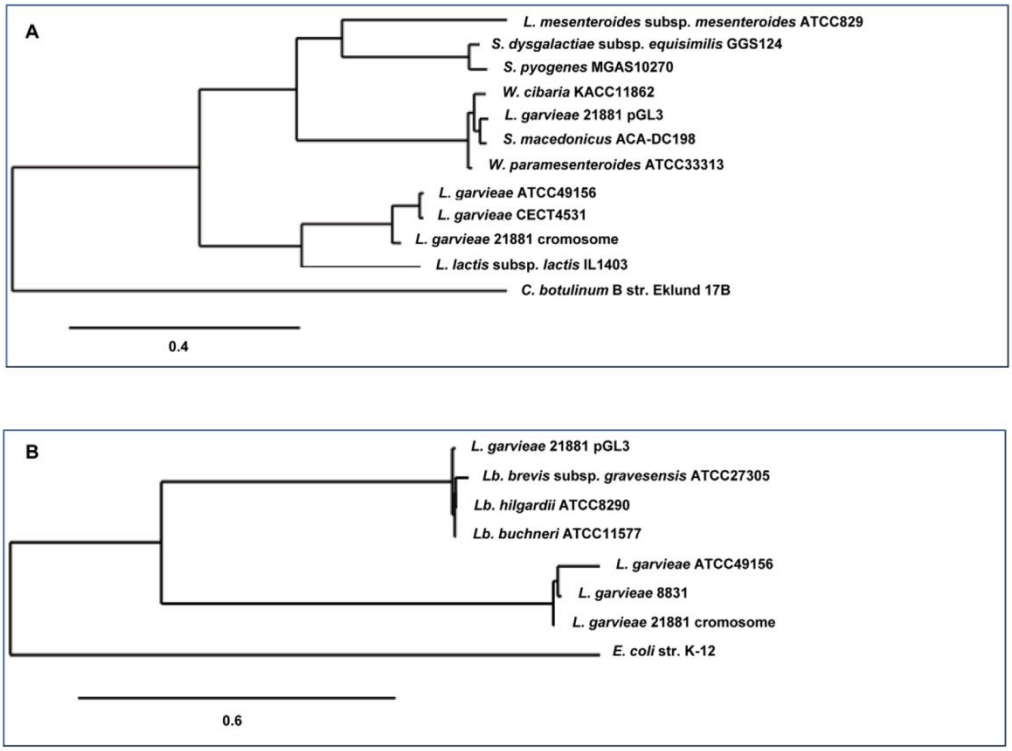


Figure 7. Phylogenetic trees of the *lai* and *pox* genes. The trees are constructed on the basis of the alignment of the DNA sequences of the *lai* (A) and *pox* (B) genes. Genes from *Clostridium botulinum* B str. Eklund 17B and *Escherichia coli* str. K-12 were used as the outgroups.
doi:10.1371/journal.pone.0040119.g007

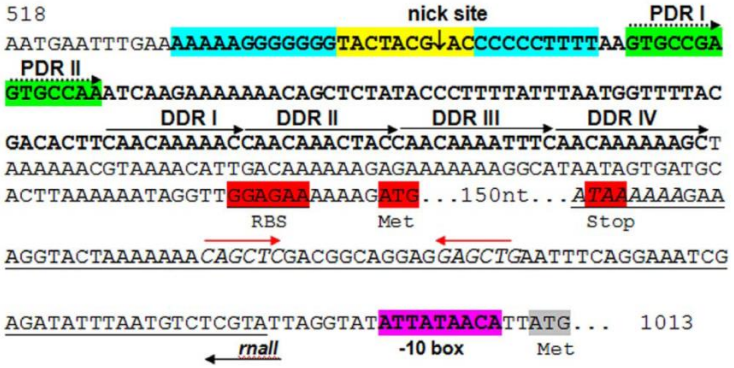


Figure 8. Proposed replication regions (*dso* and *rnaII*) of rolling circle pGL2 plasmid. The *dso* region is indicated in boldface. The *bind* locus contains four tandem distal direct repeats (DDR) which are marked with arrowheads indicating their orientation. The *nic* locus contains: the nick sequence (in yellow) flanked by the inverted repeat structures (in blue) and two proximal direct repeats (PDR) sequences in tandem (in green, with broken arrowheads indicating their orientation). The putative *rnaII* region is underlined. The predicted −10 consensus promoter sequence of *rnaII* is indicated in pink. The predicted transcription terminator region is indicated in italics and the inverted repeat sequences (involved in the RNA II hairpin formation) with red arrowheads indicating their orientation. The presumed ribosome-binding site (RBS) and start and stop codons of the *copG* gene are indicated in red. The start codon of *repB* is indicated in grey.
doi:10.1371/journal.pone.0040119.g008

Mobilization Genes

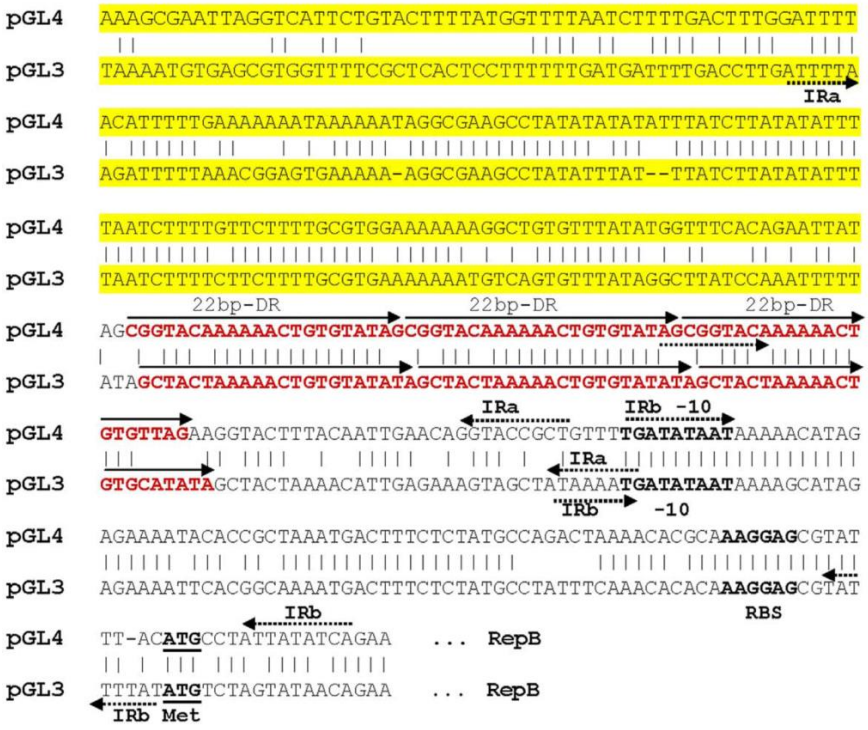
Mobility is an essential part of plasmid fitness. Plasmids can be classified into three categories according to mobility: conjugative, mobilizable and nonmobilizable. The protein component common to all transmissible (conjugative or mobilizable) plasmids is the relaxase, a key component in conjugation, because it recognizes the origin of transfer *oriT*. Additionally, conjugative plasmids carry the gene components involved in mating channel formation [57].

pGL1 was predicted to encode two mobilization proteins, MobC (124 aa) and MobA (320 aa), which showed 51% and 67% aa identity to their homologous proteins in the *T. halophilus* plasmid pHDC (Figure 2) [23]. Both the MobA and MobC proteins belong to the group of relaxases (pfam05713) involved in strand separation. A model has been suggested in which MobC acts as a molecular wedge for the relaxosome-induced melting of *oriT* DNA. The effect of MobC on strand separation may be partially complemented by the helical distortion induced by supercoiling. However, MobC extends the melted region through the nick site, thus providing the single-stranded substrate required for cleavage by MobA [58]. The pGL1 plasmid also contains a 234-bp region (bases 65 to 298) with AT-rich content (56.9%) overlapping the promoter sequences of MobC. This region contains the putative origin of transfer *(oriT)* with conservation of the postulated *nic* site (hexamer CTTGCA) just downstream of a conserved pair of inverted repeats (AAAAAAGGCT/TAGCCTTTTTT). The final segment of this region also contains another perfect inverted repeat sequence: TGTTTTTATTTTGTCA/TGACAAAA-TAAAAACA.

pGL2 was predicted to encode only one mobilization protein, Mob (504 aa), which was 54% aa identical to that of *Lactobacillus acidipiscis* plasmid pLAC1 (YP003650630) and 37% aa identical to the Mob protein encoded by the plasmid pVA380-1 found in *Streptococcus ferus* [59]. This protein is essential for site-specific cointegrative plasmid recombination, and its main biological function may be plasmid mobilization. The alignment of the pGL2 plasmid with *oriT* sequences of other pMV158 family plasmids allowed the identification of a putative *oriT* sequence upstream the *mob* gene. A putative nick site (AGTAAG ↓ TTA) was found at nucleotides 2903–2911 between the inverted repeat sequences (TAAAGT/ACTTTA) that may form the loop of hairpin [45,60].

Upstream of the origin of replication (*repB*) of pGL3 and pGL4 (Figures 4 and 5), there is a region of 231 bp that was 98% DNA identical to the transfer origin of the lactococcal plasmids pCD4 [52], pSRQ800, pSRQ900 [41] and pIL7 [24]. These putative *oriT* were located at nucleotides 65–297 on pGL3 and at nucleotides 13,490-13,721 on pGL4 and contain two inverted repeat sequences and a putative nick site CTTGCA. Neither pGL3 nor pGL4 carry any mobilization genes, suggesting that both plasmids would be non-mobilizable.

pGL5 carries three genes encoding proteins homologous to the TraA, TraG/TraD-TrwBVirD4 coupling-protein family (pfam 12696) and TraC-like proteins (also known as VirB4). The *traA* gene encodes a putative relaxase (PRK13878), with a domain that was 34% aa identical and 53% aa similar to a conjugative relaxase of the plasmid pMG2200 of *Enterococcus faecalis*. DNA relaxases are key enzymes in the initiation of conjugative transfer [57]. TraG is thought to be essential for DNA transfer in bacterial conjugation through the mating channel, and TraC is a protein involved in the translocation process [61]. Homologues of TraG and TraC have also been found in staphylococcal (pSK41 and pGO1) and lactococcal (pMRC01) plasmids. Genes encoding TraG and TraC appear to form an operon with other genes (*orf16-orf19*). The protein encoded by *orf16* displayed 42% aa identity to N-acetylmuramoyl-L-alanine amidase of *Streptococcus dysgalactiae* and contains two main enzymatic domains: a glucosaminidase domain (pfam01832) and a cysteine/histidine-dependant amidohydrolase/peptidase (designated CHAP; pfam05257) domain. Hence, Orf16 is predicted to be an exoenzyme able to hydrolyze the cell wall peptidoglycan and could therefore participate in facilitating the passage of DNA across the cell envelope by its peptidoglycan-degradation activity [39]. Similarly, Orf19 shared a 43% aa similarity to the putative membrane spanning protein of conjugative PFR55 plasmid from *Bacillus thurigiensis* [62] and could be involved in mating channel formation. Although no obvious


pGL4
pGL3
IRa
IRb
IRb -10
-10
RBS
RepB
Met
22bp-DR


pGL1
IRa
IRb
22bp-DR
RBS
Met
RepB
-10

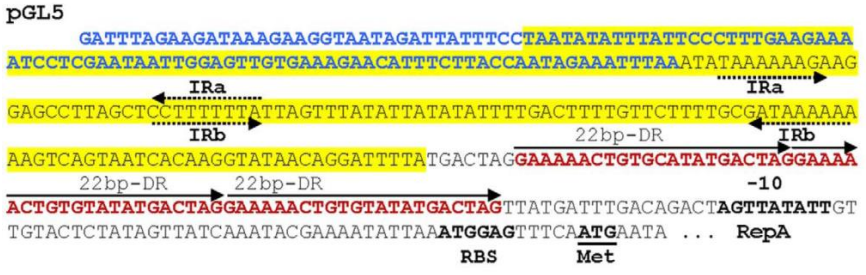

pGL5
IRa
IRb
22bp-DR
-10
RBS
Met
RepA

Figure 9. Proposed replication regions of theta-type plasmids pGL1, pGL3, pGL4 and pGL5. Sequence alignment of the upstream region of *repB* genes of only pGL3 and pGL4 showed significant identity (78%). Indicated in yellow are the AT- rich regions. Inverted repeat structures (IR) are represented by opposing broken arrows. The 22-bp directly repeated putative iteron sequences (DR) are indicated in red, and the arrowheads indicate their orientation. Predicted −10 consensus promoter sites and presumed ribosome-binding sites are in boldface. The start codons are underlined. The DNA sequence in blue corresponds to the last nucleotides of *orf30* on pGL5.
doi:10.1371/journal.pone.0040119.g009

candidate *oriT* region could be found for pGL5, its conjugation region appears to consist of modules that each display sequence similarity to the conjugal transfer determinants of pNP40 and other lactococcal conjugative plasmids [39].

According to these data, pGL1 and pGL2 could be mobilizable plasmids, and pGL5, which may encode the complete set of canonical proteins required for mobilization and transfer, could be a conjugative plasmid.

Plasmid Defense Mechanisms

Restriction-modification (R/M) systems (types I-IV) are the most common bacteriophage resistance mechanisms found in bacteria. The gene product of *hsdS* in pGL3 and pGL4 was identified as a S-subunit of type I R/M system (Tables S3 and S4), which is responsible for the specificity of endonuclease and methylase activities. The HsdS protein from pGL3 shared 99% aa identity with the homologous protein from pSRQ800 of *L. lactis* subsp. *lactis* (Figure 4), which conferrers resistance against phage P008 by changing the host type I R/M specificity [41]. In a similar fashion, it could be possible that HsdS encoded by pGL4 may be involved in the protection of *L. garvieae* 21881 from bacteriophage attack.

On the other hand, the protein encoded by *orf34* in pGL5 has a unique amino-terminal domain related to the KAP NTPase family proteins (pfam07693). Many of the prokaryotic KAP NTPases are encoded in plasmids. One of their possible functions might be the modification of the bacterial membrane that results in the exclusion of bacteriophages from the plasmid-carrying bacteria [63].

In addition to R/M systems, bacteriocin production is another common defense mechanism against competitive bacteria. The *orf3* and *orf1* genes located on pGL1 and pGL2, respectively, were predicted to encode two putative bacteriocin-like proteins (Table 2). Similarly, the protein products of *lgnD, lgnC, lgnI* and *orf37* on pGL5 appear to also be involved in the bacteriocin production, secretion and immunity (Table S5). These results are in accordance with current studies that confirm the presence of at least one bacteriocin in filtered supernatants from *L. garvieae* 21881 (data not shown).

The UmuC-like protein encoded by pGL5 appears to be involved in the replication of damaged DNA (UV protection and mutagenesis) and contains conserved domains corresponding to the Y-family of DNA polymerases, PolY/PolV/umuC subfamily (cd01700); IMS family (pfam00817: UV protection); and DinP (COG0389: nucleotidyltransferase/DNA polymerase involved in DNA repair). The *orf4* gene on pGL4 appears to encode an ATPase involved in DNA repair (Table S4). In addition, pGL3 contains the *lai* gene that encodes a potential linoleate isomerase/hydratase involved in stress tolerance in *Lactobacillus acidophilus* [64] and in the detoxification of linoleic acid effects in *Streptococcus pyogenes* [65]. These functions might help *L. garvieae* cells to adapt to specific environmental conditions.

Resistance to Drugs and Chemicals

The proteins encoded by *orf3* on pGL4 and *orf42* on pGL5 shared aa identity (100% and 36%, respectively) with efflux proteins associated with resistance to metal ions, such as copper or silver (Tables S4 and S5). The protein of 107 aa encoded by *orf2* on pGL4 showed a Small Multidrug-Resistance domain (pfam 00893) related with proteins involved in the export of a wide range of drugs, toxins and quaternary ammonium compounds [66]. Orf2 shared 49% aa identity (76% aa similarity) with the protein Smr encoded on the plasmid pSK41 from *S. aureus* (accession number AF051917) which is involved in ethidium bromide resistance.

On pGL5, *orf40* and *orf41*, which appear to be in the same operon, apparently encode a two-component drug resistance transport system similar to the ABC-type Multidrug-Resistance transporters. These are a large family of proteins involved in the transport of a wide variety of different compounds, such as sugars, peptides and more complex organic molecules (bacteriocins, antibiotics and chemicals). In bacteria, these transporters usually include an ATP-binding protein and one or two integral membrane proteins [66]. Orf40, a protein of 184 aa (Table 2), exhibited the COG1131 domain belonging to the ATPase component of ABC-type multidrug resistance system (ABC-type MDR); whereas Orf41, with a domain characteristic of the ABC2 membrane superfamily (COG0842; pfam 12698), could be the transporter membrane component. Orf40 and Orf41 shared 66% and 46% aa similarity, respectively, with the transporter system DrrA and DrrB from *Streptomyces peucetius* (accession number M73758) that confers resistance to daunorubicin and doxorubicin [67]. However, it was not possible to determine specifically which substrate would be transported by the Orf40 and Orf41 transporters proteins.

Interestingly, *orf35* on pGL5 encodes a protein of the Pentapeptide Repeat Protein (PRP) family (pfam00805, COG1357). The biological function of most PRP family members is unknown, but these proteins have been increasingly associated with quinolone resistance (Qnr). Qnr is encoded by plasmids in many Gram-negative bacteria and by chromosomal genes in Gram-positive bacteria [68,69]. *L. garvieae* 21881 was resistant to nalidixic acid (0 mm of diameter of inhibition zone), oxolinic acid (0 mm), flumequine (7 mm), ciprofloxacin (16 mm) and norfloxacin (14 mm) and was sensitive to moxifloxacin (22 mm), in an in vitro disk diffusion susceptibility assay. The main mechanism of quinolone resistance is the accumulation of mutations in the bacterial enzymes DNA gyrase and DNA topoisomerase IV [38,70]. No mutations were detected in the quinolone-resistance determining regions (QRDRs) of GyrA and ParC of *L. garvieae* 21881 (Figure 10) when compared with the GyrA and ParC sequences of the quinolone susceptible *L. garvieae* strain KL99110 [38]. Mutations in *gyrB* and *parE* are associated with quinolone resistance in other Gram-positive bacteria [70], but were not found in *L. garvieae* 21881 (data not shown). The analysis of the amino acid sequence of Orf35 revealed that critical residues at positions G56, C72, C92, G96, F114, and L159 and loops A and B, which are essential for the interaction with topoisomerases, were highly conserved in Orf35 with respect to the protein sequence of QnrA1, the variant associated with plasmid-mediated quinolone resistance [71]. To our knowledge, this is the first description of a plasmid-born PRP in Gram-positive bacteria, and further studies on the PRP Orf35 are required to elucidate its cellular function.

Table 2. Putative virulence factors and defense mechanisms in *L. garvieae* 21881 plasmids.

Plasmid	Orf	Size (aa)	Related protein	Organism	% Identity/ Similarity	Accession number
pGL1	*orf3*	72	Putative bacteriocin-like protein	*Streptococcus pyogenes* *Enterococcus faecalis*	57/66 46/58	ACT32367 EFM82101
pGL2	*orf1*	71	Putative bacteriocin-like protein	*Streptococcus mitis*	57/75	CBJ23192
pGL2	*orf2*	97	Bacteriocin- like immunity protein	*L. lactis* subsp. *lactis S. equi* subsp. *equi*	41/63 30/40	YP001032092 CAW94771
pGL4	*orf2*	107	Multidrug resistance protein	*Lactobacillus kisonensis* *Staphylococcus aureus*	76/89 49/76	EHO53128 NP863640
pGL5	*txn*	239	Actin-ADP-ribosylating protein. Putative toxin	*Clostridiun botulinum*	24/42	CAA35828
pGL5	*orf5*	119	Mucin-binding LPXTG protein	*Streptococcus anginosus*	34/51	ZP08525680
pGL5	*orf25*	1273	Collagen-binding LPXTG protein	*Enterococcus faecalis* *Enterococcus faecium*	36/51 32/47	ZP05574571 ZP06695453
pGL5	*orf35*	385	Pentapeptide repeats containing protein	Blood disease bacterium	33/49	CCA83758
pGL5	*lgnI*	88	Bacteriocin immunity protein	*Leuconostoc gelidum*	31/57	ZP08478983
pGL5	*orf37*	63	Putative bacteriocin			
pGL5	*orf40*	253	ABC-type multidrug transport protein	*L. lactis* subsp. *lactis*		AEU40217
pGL5	*orf41*	245	ABC-2 membrane transporter	*L. lactis* subsp. *lactis*	99/99	AEU40216

doi:10.1371/journal.pone.0040119.t002

Putative Virulence Factors

pGL5 harbors some genes (*txn*, *orf5* and *orf25*) that encode proteins that could be considered putative virulence factors (Table 2).

The gene *txn* encodes a protein of 239 residues that has the enzymatic domain corresponding to the family of actin-ADP-ribosyltransferases (pfam03496, cd00233). The bacterial ADP-ribosyltransferase toxins are a family of proteins that kill the target eukaryotic cells through the modification of proteins essential for the host organism, playing a key role in the pathogenesis of a variety of bacterial pathogens [72]. In particular, actin ADP-ribosylating bacterial toxins lead to a complete depolymerization or disaggregation of the actin cytoskeleton destroying the microfilament network that contributes to the cytopathic action of these toxins [72,73]. Txn of *L. garvieae* exhibits the catalytic Glu-X-Glu sequence (residues 199 to 201 in Txn), the NAD binding sites Ser-Thr-Ser sequence (amino acids 156–158 in Txn) and a conserved Arg residue (residing at position 129 in Txn), characteristic of the CT-group of many mono-ADP-ribosyltransferases [73]. Within this group, the toxins could be composed of by a single polypeptide, such as the clostridial C3 exoenzyme and C3-like toxins from *Bacillus cereus* and *S. aureus*, or by two components (binary AB-toxin), such as Iota toxin from *Clostridium perfringens*, the toxin C2 from *Clostridium botulinum*, the *Clostridium spiriforme* toxin,

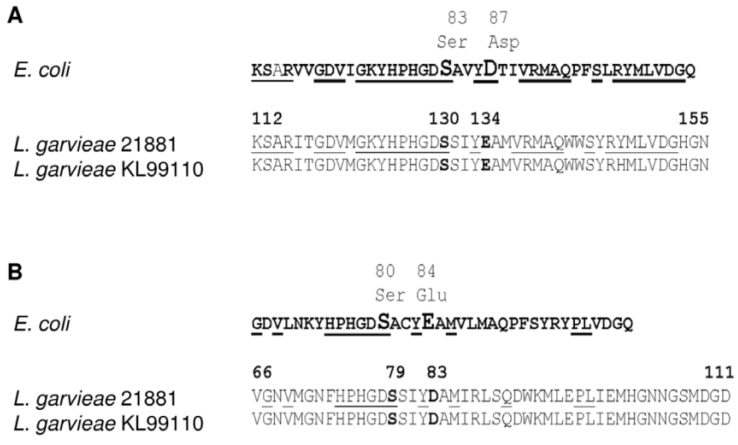


Figure 10. Amino acid sequence of the QRDRs from *L. garvieae* 21881. A) GyrA, B) ParC. Sequences from *L. garvieae* 21881 are compared with those described for the quinolone-sensitive *L. garvieae* strain KL99110. The position of amino acids critical for quinolone resistance (Ser, Asp and Glu) are indicated. Identical amino acids of *L. garvieae* 21881 to the standard QRDR regions for *E. coli* are underlined.
doi:10.1371/journal.pone.0040119.g010

or the vegetative insecticidal protein VIP2 from *B. cereus* [72,73]. The Txn protein of *L. garvieae* shared 24% aa identity and 42% aa similarity with the C3 exotoxin from *C. botulinum* (accession number CAA35828), indicating that it would belong to the single polypeptide class of C3-family ADP-ribosyltransferases toxins, whose amino acid sequences are considerably diverged [73]. The *orf3* gene, predicted to be located in the same operon as *txn*, encodes a thioredoxin protein, characterized by the TRX fold (amino acids 36–38). Many members of the thioredoxin (TRX)-like superfamily that do not contain the CXXC motif, such as Orf3, function as glutathione peroxidases, GSH transferases, arsenic reductases, transcriptional regulators or chaperones [74]. In contrast to the binary A-B toxins, the cell accessibility of C3-like toxins is unknown in many cases. Therefore, Orf3 might act as a transcriptional regulator of *txn*, as a facilitator of protein folding or in the secretion or cellular uptake of *L. garvieae* Txn toxin across the host membrane. Although ADP-ribosylating toxins have been identified in different species of Gram-positive and Gram-negative bacteria [73], to our knowledge, this is the first time these toxins have been detected in a species of *Lactococcus*.

The genes *orf5* and *orf25* encode two putative surface proteins containing the cell wall-sorting motif LPXTG (Tables 1 and S5) characteristic of Gram-positive cell-wall-anchored surface proteins. The gene *orf5* encodes a protein of 1,179 residues that contains on the carboxy-end a LPXTG-motif (LPQTG) and three mucin-binding protein domains (pfam06458) corresponding to positions 650–750, 850–950 and 1025–1125 of the protein. Bioinformatic analysis of the *orf5* gene product predicted the most likely cleavage site of the signal sequence peptide to be between positions 30 and 31 (ALA-DE). Most commensal and pathogenic bacteria attach to the intestinal mucosal cells through the interaction of adhesins to the mucosal receptors. Thus, the LPXTG-protein encoded by *orf5* could be involved in the binding of *L. garvieae* to mucus from the intestine facilitating further interaction with intestinal epithelial cells. This involvement is especially interesting considering that underlying gastro-intestinal disorders appear to be a factor that contributes to *L. garvieae* infection [14].

The gene *orf25* encodes another putative LPXTG surface protein of 1,271 aa that contains two conserved protein domains: a region on the N-terminal end (KSGKRW), characteristic of signal peptide of some serine-rich and heavily glycosylated proteins, and a collagen-binding protein domain (Cna protein B-type domain; pfam05738), which could facilitate the bacterial adherence to collagen. The Cna protein is an adhesin that plays an important role in the virulence of *S. aureus* [75,76]. Bioinformatic analysis of this protein predicted the most likely cleavage site of the signal sequence peptide, between positions 42 and 43: ALA-GG. Ten nucleotides upstream of *orf25* is the gene *srtA*, encoding a sortase A (Table S5), which appears to form part of the same operon. Sortase A covalently immobilizes the surface protein to the cell-wall peptidoglycan facilitating the bacterial adherence to the host cell [77]. Thus, the Orf25 protein might provide an advantage to *L. garvieae* for binding collagen substrates and collagenous tissues, such as the collagen-rich heart valves. It is interesting to note that endocarditis is the most common clinical manifestation of *L. garvieae* human infections [1].

LPXTG surface proteins are implicated in the pathogenesis of a number of bacteria, e.g., *S. aureus*, *Listeria monocytogenes*, enterococci or streptococci playing a functional role in the adherence to host cells [78–80]. Thus, the LPXTG surface proteins Orf5 and Orf25 could be implicated in the first interaction step of *L. garvieae* 21881 with human cells facilitating further infection. Genes *orf5* and *orf25* were not detected in any of the *L. garvieae* strains with whole genome sequences available [16–19]. *L. garvieae* 21881 also carries other predicted LPXTG-proteins in its chromosome [20]. One of these chromosomal LPXTG-proteins (located between contigs 25–26 from whole sequence of accession number AFCC01000000) showed high aa identity (98%) with the mucin-binding protein (accession number HM852546) present in another *L. garvieae* human clinical strain [27] but absent in fish strains [16,18,19,26]. These data suggest the existence of specific adhesins in human *L. garvieae* isolates and therefore the possibility of different strategies in *L. garvieae* for interacting with host cells. Further studies including more clinical strains will be necessary to further corroborate these results and elucidate the role of these virulence genes in the pathogenesis of *L. garvieae*.

This is the first report on the characterization of plasmids in a human clinical strain of *L. garvieae*. The similarity observed between the *orfs* found in the plasmid genome of this strain and the chromosomal or plasmidic genes of other lactic acid bacteria suggests the existence of horizontal gene transfer events among *L. garvieae* and these bacteria. These plasmids harbor genes related to drug resistance, bacteriocin production and genes that might help *L. garvieae* cells to adapt to specific environmental conditions. Most interesting was the detection of putative virulence genes, such as the gene *txn* that encodes a protein corresponding to the family actin-ADP-ribosyltransferases toxins and the genes *orf5* and *orf25* that encode surface LPXTG proteins with mucin and collagen-binding domains that may be implicated in the adherence to host cells. These results could be useful to understand the factors involved in host invasion and infection of *L. garvieae*.

Supporting Information

Table S1 Putative genes identified on pGL1. Mob proteins were classified into a relaxase (MOB) family according to Smillie et al. 2010.
(DOC)

Table S2 Putative genes identified on pGL2. Mob proteins were classified into a relaxase (MOB) family according to Smillie et al. 2010.
(DOC)

Table S3 Putative genes identified on pGL3.
(DOC)

Table S4 Putative genes identified on pGL4.
(DOC)

Table S5 Putative genes identified on pGL5. Mob proteins were classified into a relaxase (MOB) family according to Smillie et al. 2010.
(DOC)

Acknowledgments

The authors thank J.M. Rodríguez-Martínez for his help in the preliminary analysis of PRP, B. González-Zorn for his useful comments about *L. garvieae* 21881 quinolone resistance, M. W. J. van Passel for his help analyzing HGT events, M. Espinosa for his helpful comments on pGL2 and C. Aspiroz for kindly supplying the human clinical strain.

Author Contributions

Conceived and designed the experiments: AG MMB JFF-G. Performed the experiments: MA-U AG. Analyzed the data: MA-U AG GHL-C. Wrote the paper: MTC AG MMB JFF-G MA-U.

References

1. Fefer JJ, Ratzan KR, Sharp SE, Saiz E (1998) *Lactococcus garvieae* endocarditis: report of a case and review of the literature. Diagn Microbiol Infect Dis. 32: 127–130.
2. Vinh DC, Nichol KA, Rand F, Embil JM (2006) Native-valve bacterial endocarditis caused by *Lactococcus garvieae*. Diagn Microbiol Infect Dis 56: 91–94.
3. Li WK, Chen YS, Wann SR, Liu YC, Tsai HC (2008) *Lactococcus garvieae* endocarditis with initial presentation of acute cerebral infarction in a healthy immunocompetent man. Intern Med 12: 1143–1146.
4. Watanabe Y, Naito T, Kikuchi K, Amari Y, Uehara Y, et al. (2011) Infective endocarditis with *Lactococcus garvieae* in Japan: a case report. J Med Case Reports 5: 356.
5. Mofredj A, Baraka D, Kloeti G, Dumont JL (2000) *Lactococcus garvieae* septicemia with liver abscess in an immunosuppressed patient. Am J Med 109: 513–514.
6. Chan JF, Woo PC, Teng JL, Lau SK, Leung SS, et al. (2011) Primary infective spondylodiscitis caused by *Lactococcus garvieae* and a review of human *L. garvieae* infections. Infection 39: 259–264.
7. Nadrah K, Cerar T, Papst L, Volkar-Meglič J, Matičič M, et al. (2011) *Lactococcus garvieae* septicaemia in a patient with artificial heart valves. Wien Klin Wochenschr 123: 677–679.
8. Vendrell D, Balcázar JL, Ruiz-Zarzuela I, de Blas I, Gironés O, et al. (2006) *Lactococcus garvieae* in fish: a review. Comp Immunol Microbiol Infect Dis 29: 177–198.
9. Teixeira LM, Merquior VL, Vianni MC, Carvalho MG, Fracalanzza SE, et al. (1996) Phenotypic and genotypic characterization of atypical *Lactococcus garvieae* strains isolated from water buffalos with subclinical mastitis and confirmation of *L. garvieae* as a senior subjective synonym of *Enterococcus seriolicida*. Int J Syst Bacteriol 46: 664–668.
10. Tejedor JL, Vela AI, Gibello A, Casamayor A, Domínguez L, et al. (2011) A genetic comparison of pig, cow and trout isolates of *Lactococcus garvieae* by PFGE analysis. Lett Appl Microbiol 53: 614–619.
11. Santos EM, Jaime I, Rovira J, Lyhs U, Korkeala H, et al. (2005) Characterization and identification of lactic acid bacteria in "morcilla de Burgos". Int J Food Microbiol 97: 285–296.
12. Kawanishi M, Yoshida T, Kijima M, Yagyu K, Nakai T, et al. (2007) Characterization of *Lactococcus garvieae* isolated from radish and broccoli sprouts that exhibited a KG+ phenotype, lack of virulence and absence of a capsule. Lett Appl Microbiol 44: 481–487.
13. Alegría A, Álvarez-Martín P, Sacristán N, Fernández E, Delgado S, et al. (2009) Diversity and evolution of the microbial populations during manufacture and ripening of Casín, a traditional Spanish, starter free cheese made from cow's milk. Int J Food Microbiol 136: 44–51.
14. Wang CY, Shie HS, Chen SC, Huang JP, Hsieh IC, et al. (2007) *Lactococcus garvieae* infections in humans: possible association with aquaculture outbreaks. Int J Clin Pract 61: 68–73.
15. Wilbring M, Alexiou K, Reichenspurner H, Matschke K, Tugtekin SM (2011) *Lactococcus garvieae* causing zoonotic prosthetic valve endocarditis. Clin Res Cardiol 100: 545–546.
16. Aguado-Urda M, López-Campos GH, Gibello A, Cutuli MT, López-Alonso V, et al. (2011) Genome sequence of *Lactococcus garvieae* 8831, isolated from rainbow trout lactococcosis outbreaks in Spain. J Bacteriol 193: 4263–4264.
17. Reimundo P, Pignatelli M, Alcaraz LD, D'Auria G, Moya A, et al. (2011) Genome sequence of *Lactococcus garvieae* UNIUD074, isolated in Italy from a lactococcosis outbreak. J Bacteriol 193: 3684–3685.
18. Morita H, Toh H, Oshima K, Yoshizaki M, Kawanishi M, et al. (2011) Complete genome sequence and comparative analysis of the fish pathogen *Lactococcus garvieae*. PLoS One 6: e23184.
19. Ricci G, Ferrario C, Borgo F, Rollando A, Fortina MG (2012) Genome sequences of *Lactococcus garvieae* TB25, isolated from Italian cheese, and *Lactococcus garvieae* LG9, isolated from Italian rainbow trout. J Bacteriol 194: 1249–1250.
20. Aguado-Urda M, López-Campos GH, Blanco MM, Fernández-Garayzábal JF, Cutuli MT, et al. (2011) Genome sequence of *Lactococcus garvieae* 21881, isolated from a case of human septicaemia. J Bacteriol 193: 4033–4034.
21. Siezen RJ, Renckens B, van Swam I, Peters S, van Kranenburg R, et al. (2005) Complete sequences of four plasmids of *Lactococcus lactis* subsp. *cremoris* SK11 reveal extensive adaptation to the dairy environment. Appl Environ Microbiol 71: 8371–8382.
22. Defoor E, Kryger MB, Martinussen J (2007) The orotate transporter encoded by oroP from *Lactococcus lactis* is required for orotate utilization and has utility as a food-grade selectable marker. Microbiology 153: 3645–3659.
23. Satomi M, Furushita M, Oikawa H, Yoshikawa-Takahashi M, Yano Y (2008) Analysis of a 30 kbp plasmid encoding histidine decarboxylase gene in *Tetragenococcus halophilus* isolated from fish sauce. Int J Food Microbiol 126: 202–209.
24. Górecki RK, Koryszewska-Bagińska A, Gołębiewski M, Żylińska J, Grynberg M, et al. (2011) Adaptative potential of the *Lactococcus lactis* IL594 strain encoded in its 7 plasmids. PLoS One 6: e22238.
25. Mills S, McAuliffe OA, Coffey A, Fitzgerald GF, Ross RP (2006) Plasmids of lactococci-genetic accessories or genetic necessities? FEMS Microbiol 30: 243–273.
26. Fortina MG, Ricci G, Borgo F (2009) A study of lactose metabolism in *Lactococcus garvieae* reveals a genetic marker for distinguishing between dairy and fish biotypes. J Food Proc 72: 1248–1254.
27. Reimundo P, Rivas AJ, Osorio CR, Méndez J, Pérez-Pascual D, et al. (2011) Application of suppressive subtractive hybridization to the identification of genetic differences between two *Lactococcus garvieae* strains showing distinct differences in virulence for rainbow trout and mouse. Microbiology 157: 2106–2119.
28. Maki T, Santos MD, Kondo H, Hirono I, Aoki T (2009) A transferable 20-kilobase multiple drug resistance-conferring R plasmid (pKL0018) from a fish pathogen (*Lactococcus garvieae*) is highly homologous to a conjugative multiple drug resistance-conferring enterococcal plasmid. Appl Environ Microbiol 75: 3370–3372.
29. Anderson DG, McKay L (1983) Simple and rapid method for isolating large plasmid DNA from lactic streptococci. Appl Environ Microbiol 46: 549–552.
30. Lukashin A, Borodovsky M (1998) GeneMark.hmm: new solutions for gene finding. NAR 26: 1107–1115.
31. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12: 402.
32. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785–786.
33. van Passel MW, Luyf AC, van Kampen AH, Bart A, van der Ende A (2005) Deltarho-web, an online tool to assess composition similarity of individual nucleic acid sequences. Bioinformatics 21: 3053–3055.
34. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.
35. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, et al. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36 (Web Server issue): W465-W469.
36. National Committee for Clinical Laboratory Standards (NCCLS) (2011) Performance standards for antimicrobial susceptibility testing; twenty-first informational supplement. NCCLS document M100-S21, vol.31, no.1. Wayne: NCCLS. 168 p.
37. Anonymous (2001) Comité de l'antibiogramme de la Societé Française de Microbiologie, report 2000–2001. Paris: Societé Française de Microbiologie. 42 p.
38. Maki T, Hirono I, Kondo H, Aoki T (2008) Drug resistance mechanism of the fish-patogenic bacterium *Lactococcus garvieae*. J Fish Dis 31: 461–468.
39. O'Driscoll J, Glynn F, Fitzgerald GF, van Sinderen D (2006) Sequence analysis of the lactococcal plasmid pNP40: a mobile replicon for coping with environmental hazards. J Bacteriol 188: 6629–6639.
40. Tanous C, Chambellon E, Yvon M (2007) Sequence analysis of the mobilizable lactococcal plasmid pGdh442 encoding glutamate dehydrogenase activity. Microbiology 153: 1664–1675.
41. Boucher I, Emond E, Parrot M, Moineau S (2001) DNA sequence analysis of three *Lactococcus lactis* plasmids encoding phage resistance mechanisms. J Dairy Sci 84: 1610–1620.
42. Kim JF, Jeong H, Lee JS, Choi SH, Ha M, et al. (2008) Complete genome sequence of *Leuconostoc citreum* KM20. J Bacteriol 190: 3093–3094.
43. Bosi E, Fani R, Fondi M (2011) The mosaicism of plasmids revealed by atypical genes detection and analysis. BMC Genomics 12: 403.
44. van Kranenburg R, Golic N, Bongers R, Leer RJ, de Vos WM, et al. (2005) Functional analysis of three plasmids from *Lactobacillus plantarum*. Appl Environ Microbiol 71: 1223–1230.
45. Zhai Z, Hao Y, Yin S, Luan C, Zhang L, et al. (2009) Characterization of a novel rolling-circle replication plasmid pYSI8 from *Lactobacillus sakei* YSI8. Plasmid 62: 30–34.
46. Takamatsu D, Osaki M, Sekizaki T (2000) Sequence analysis of a small cryptic plasmid isolated from *Streptococcus suis* serotype 2. Curr Microbiol 40: 61–66.
47. Turgeon N, Moineau S (2001) Isolation and characterization of a *Streptococcus thermophilus* plasmid closely related to the pMV158 family. Plasmid 45: 171–183.
48. Ruiz-Masó JA, Lurz R, Espinosa M, del Solar G (2007) Interactions between the RepB initiator protein of plasmid pMV158 and two distant DNA regions within the origin of replication. Nucleic Acid Res 35: 1230–1244.
49. Khan SA (1997) Rolling-circle replication of bacterial plasmids. Microbiol Mol Biol Rev 61: 442–455.
50. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, et al. (2006) Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA 103: 15611–15616.
51. Biet F, Cenatiempo Y, Fremaux C (2002) Identification of a replicon from pTXL1, a small cryptic plasmid from *Leuconostoc mesenteroides* subsp. *mesenteroides* Y110, and development of a food-grade vector. Appl Environ Microbiol 68: 6451–6456.
52. Emond E, Lavallée R, Drolet G, Moineau S, LaPointe G (2001) Molecular characterization of a theta replication plasmid and its use for development of a two-component food-grade cloning system for *Lactococcus lactis*. Appl Environ Microbiol 67: 1700–1709.

53. Sánchez C, Hernández de Rojas A, Martínez B, Argüelles ME, Suárez JE, et al. (2000) Nucleotide sequence and analysis of pBL1, a bacteriocin-producing plasmid from *Lactococcus lactis* IPLA 972. Plasmid 44: 239–249.
54. Gao Y, Lu Y, Teng KL, Chen ML, Zheng HJ, et al. (2011) Complete genome sequence of *Lactococcus lactis* subsp. *lactis* CV56, a probiotic strain isolated from the vaginas of healthy women. J Bacteriol 193: 2886–2887.
55. Strahinic I, Kojic M, Tolinacki M, Fira D, Topisirovic L (2009) Molecular characterization of plasmids pS7a and pS7b from *Lactococcus lactis* subsp. *lactis* bv. diacetylactis S50 as a base for the construction of mobilizable cloning vectors. J Appl Microbiol 106: 78–88.
56. del Solar G, Espinosa M (2001) In vitro analysis of the terminator T(II) of the inhibitor antisense rna II gene from plasmid pMV158. Plasmid 45: 75–87.
57. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F (2010) Mobility of plasmids. Microbiol Mol Biol Rev 74: 434–452.
58. Zhang S, Meyer R (1997) The relaxosome protein MobC promotes conjugal plasmid mobilization by extending DNA strand separation to the nick site at the origin of transfer. Mol Microbiol 25: 509–516.
59. LeBlanc DJ, Chen YY, Lee LN (1993) Identification and characterization of a mobilization gene in the streptococcal plasmid, pVA380–1. Plasmid 30: 296–302.
60. de las Rivas B, Marcobal A, Muñoz R (2004) Complete nucleotide sequence and structural organization of pPB1, a small *Lactobacillus plantarum* cryptic plasmid that originated by modular exchange. Plasmid 52: 203–211.
61. Grohmann E, Muth G, Espinosa M (2003) Conjugative plasmid transfer in gram-positive bacteria. Microbiol Mol Biol Rev 67: 277–301.
62. Amadio AF, Benintende GB, Zandomeni RO (2009) Complete sequence of three plasmids from *Bacillus thuringiensis* INTA-FR7–4 environmental isolate and comparison with related plasmids from the *Bacillus cereus* group. Plasmid 62: 172–182.
63. Aravind L, Iyer LM, Leipe DD, Koonin EV (2004) A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. Genome Biol 5: R30.
64. O'Flaherty SJ, Klaenhammer TR (2010) Functional and phenotypic characterization of a protein from *Lactobacillus acidophilus* involved in cell morphology, stress tolerance and adherence to intestinal cells. Microbiology 156: 3360–3367.
65. Volkov A, Liavonchanka A, Kamneva O, Fiedler T, Goebel C, et al. (2010) Myosin cross-reactive antigen of *Streptococcus pyogenes* M49 encodes a fatty acid double bond hydratase that plays a role in oleic acid detoxification and bacterial virulence. J Biol Chem 285: 10353–10361.
66. Putman M, van Veen HW, Konings WN (2000) Molecular properties of bacterial multidrug transporters. Microbiol Mol Biol Rev 64: 672–693.
67. Kaur P, Rao DK, Gandlur SM (2005) Biochemical characterization of domains in the membrane subunit DrrB that interact with the ABC subunit DrrA: identification of a conserved motif. Biochemistry 44: 2661–2670.
68. Vetting MW, Hegde SS, Fajardo JE, Fiser A, Roderick SL, et al. (2006) Pentapeptide repeat proteins. Biochemistry 45: 1–10.
69. Strahilevitz J, Jacoby GA, Hooper DC, Robicsek A (2009) Plasmid-mediated quinolone resistance: a multifaceted threat. Clin Microbiol Rev 22: 664–689.
70. Maeda Y, Murayama M, Goldsmith CE, Coulter WA, Mason C, et al. (2011) Molecular characterization and phylogenetic analysis of quinolone resistance-determining regions (QRDRs) of gyrA, gyrB, parC and parE gene loci in viridans group streptococci isolated from adult patients with cystic fibrosis. J Antimicrob Chemother 66: 476–486.
71. Rodríguez-Martínez JM, Briales A, Velasco C, Conejo MC, Martínez-Martínez L, et al. (2009) Mutational analysis of quinolone resistance in the plasmid-encoded pentapeptide repeat proteins QnrA, QnrB and QnrS. J Antimicrob Chemother 63: 1128–1134.
72. Han S, Tainer JA (2002) The ARTT motif and a unified structural understanding of substrate recognition in ADP-ribosylating bacterial toxins and eukaryotic ADP-ribosyltransferases. Int J Med Microbiol 291: 523–529.
73. Holbourn KP, Shone CC, Acharya KR (2006) A family of killer toxins. Exploring the mechanism of ADP-ribosylating toxins. FEBS J 273: 4579–4593.
74. Berndt C, Lillig CH, Holmgren A (2008) Thioredoxins and glutaredoxins as facilitators of protein folding. Biochim Biophys Acta 1783: 641–650.
75. Hartford O, McDevitt D, Foster TJ (1999) Matrix-binding proteins of *Staphylococcus aureus*: functional analysis of mutant andhybrid molecules. Microbiology 145: 2497–2505.
76. Nienaber JJC, Kuinkel BKS, Clarke-Pearson M, Lamlertthon S, Park L, et al. (2011) Methicillin-susceptible *Staphylococcus aureus* endocarditis isolates are associated with clonal complex 30 genotype and a distinct repertoire of enterotoxins and adhesins. J Infect Dis 204: 704–13.
77. Marraffini LA, Dedent AC, Schneewind O (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. Microbiol Mol Biol Rev 70: 192–221.
78. Davies JR, Svensäter G, Herzberg MC (2009) Identification of novel LPXTG-linked surface proteins from *Streptococcus gordonii*. Microbiology 155: 1977–1988.
79. Hendrickx AP, Willems RJ, Bonten MJ, van Schaik W (2009) LPxTG surface proteins of enterococci. Trends Microbiol 17: 423–430.
80. Reis O, Sousa S, Camejo A, Villiers V, Gouin E, et al. (2010) LapB, a novel *Listeria monocytogenes* LPXTG surface adhesion, required for entry into eukaryotic cells and virulence. J Infect Dis 202: 551–562.

El material adicional asociado a esta publicación se encuentra disponible a través de las siguientes URLs:

- ✓ Tabla Suplementaria 1 (*Table S1*): http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0040119.s001
- ✓ Tabla Suplementaria 2 (*Table S2*): http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0040119.s002
- ✓ Tabla Suplementaria 3 (*Table S3*): http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0040119.s003
- ✓ Tabla Suplementaria 4 (*Table S4*): http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0040119.s004
- ✓ Tabla Suplementaria 5 (*Table S5*): http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0040119.s005

CAPÍTULO VI:

Análisis de la influencia de la temperatura de crecimiento sobre el transcriptoma de *Lactococcus garvieae*

Resumen

La liberación de los genomas de varias cepas de *Lactococcus garvieae* en los últimos años ha abierto la posibilidad de la realización de estudios a escala genómica en este patógeno. Con el fin llevar cabo estudios de expresión génica global que permitieran la caracterización funcional del genoma de *L. garvieae*, se diseñó y construyó un microarray de ADN que contenía las secuencias genómicas de las dos cepas secuenciadas de origen clínico, Lg8831 (trucha) y Lg21881 (humano).

L. garvieae es una bacteria ubicua, capaz de adaptarse a múltiples nichos ecológicos y por lo tanto a diversas condiciones ambientales. La temperatura se ha definido como uno de los factores ambientales condicionantes para el desarrollo de la lactococosis (Vendrell *et al*., 2006). Así, el primer estudio funcional planteado sobre del microarray específico para *L. garvieae*, fue el análisis de la influencia de la temperatura sobre el transcriptoma de las cepas Lg8831 y Lg21881. Las temperaturas elegidas para el análisis fueron 18°C y 37°C, debido a que a 18°C se dan los brotes más agudos de lactococosis en peces, y 37°C es la temperatura corporal humana.

Los resultados de dichos experimentos evidenciaron una respuesta específica a la temperatura en función de la cepa. Entre los hallazgos más significativos, Lg8831 sobreexpresó a 18°C varios genes que codifican diferentes proteínas de respuesta específica a frío, implicados en una respuesta adaptativa eficiente de esta cepa a condiciones de baja temperatura. Otro resultado relevante fue la descripción por primera vez de un metabolismo respiratorio en *L. garvieae*, y la regulación de la expresión génica del mismo en Lg21881 resultó dependiente de la temperatura. La realización de este trabajo supone el primer estudio de expresión global en *L. garvieae* y sus resultados pueden ayudar a comprender los mecanismos que participan en la adaptación de este patógeno a diferentes condiciones ambientales.

OPEN ACCESS Freely available online

PLOS ONE

Global Transcriptome Analysis of *Lactococcus garvieae* Strains in Response to Temperature

Mónica Aguado-Urda[1], Alicia Gibello[1], M. del Mar Blanco[1], José F. Fernández-Garayzábal[1,2], Victoria López-Alonso[3], Guillermo H. López-Campos[4*]

1 Department of Animal Health, Faculty of Veterinary Sciences, Complutense University, Madrid, Spain, **2** Animal Health Surveillance Center (VISAVET), Complutense University, Madrid, Spain, **3** Bioinformatics and Public Health Department, Health Institute Carlos III, Madrid, Spain, **4** Health and Biomedical Informatics Research Unit, Medical School, University of Melbourne, Victoria, Australia

Abstract

Lactococcus garvieae is an important fish and an opportunistic human pathogen. The genomic sequences of several *L. garvieae* strains have been recently published, opening the possibility of global studies on the biology of this pathogen. In this study, a whole genome DNA microarray of two strains of *L. garvieae* was designed and validated. This DNA microarray was used to investigate the effects of growth temperature (18°C and 37°C) on the transcriptome of two clinical strains of *L. garvieae* that were isolated from fish (Lg8831) and from a human case of septicemia (Lg21881). The transcriptome profiles evidenced a strain-specific response to temperature, which was more evident at 18°C. Among the most significant findings, Lg8831 was found to up-regulate at 18°C several genes encoding different cold-shock and cold-induced proteins involved in an efficient adaptive response of this strain to low-temperature conditions. Another relevant result was the description, for the first time, of respiratory metabolism in *L. garvieae*, whose gene expression regulation was temperature-dependent in Lg21881. This study provides new insights about how environmental factors such as temperature can affect *L. garvieae* gene expression. These data could improve our understanding of the regulatory networks and adaptive biology of this important pathogen.

Citation: Aguado-Urda M, Gibello A, Blanco MdM, Fernández-Garayzábal JF, López-Alonso V, et al. (2013) Global Transcriptome Analysis of *Lactococcus garvieae* Strains in Response to Temperature . PLoS ONE 8(11): e79692. doi:10.1371/journal.pone.0079692

Editor: Willem van Schaik, University Medical Center Utrecht, Netherlands

Received June 21, 2013; **Accepted** September 24, 2013; **Published** November 4, 2013

Funding: This research was supported by grants AGL2009-12447 and AGL2012-35419 from the Spanish Ministry of Economy and Competiveness. MAU was recipient of a Ph.D. fellowship from the Complutense University of Madrid. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: guillermo.lopez@unimelb.edu.au

Introduction

Lactococcus garvieae is a ubiquitous and widely distributed microorganism that has relevance in veterinary and human medicine. Although this bacterium is one of the most important bacterial fish pathogens, affecting various wild and farmed fish species, particularly rainbow trout [1], it has also been isolated from other animal species, such as cows, buffalos, pigs, wild birds [2], cats, dogs, and horses [3]. *L. garvieae* has gained clinical relevance in human medicine during the last years, being considered an opportunistic and potentially zoonotic pathogen that causes a variety of infections [4]. In addition to its relevance as a pathogen, *L. garvieae* can also be isolated from rivers and sewage waters [2], and from different foods such as vegetables, meat and dairy products [5]. Recently, *L. garvieae* has also been isolated from fecal samples of healthy individuals, suggesting this microorganism could be either part of the human commensal microbiota or transient bacteria ingested with food [6]. This wide distribution of *L. garvieae* is likely related to its ability to adapt and survive in many environmental conditions including a wide range of pH (4.5 to 9.6), temperatures (from 10°C to 45°C), salinity concentrations (0 to 6.5%) and nutrient sources [7].

Bacteria usually respond to variations in environmental factors such as temperature with adaptive changes in their transcriptome [8-11]. Because *L. garvieae* is able to colonize multiple, diverse different environments, and because it causes infection in a broad range of different hosts, it must therefore be able to sense, adapt, and respond to these temperature fluctuations. Water temperature has been described as the most important environmental factor in the development of the *L. garvieae* infections in fish [1], but there is a complete lack of knowledge about the influence of temperature on *L. garvieae* gene expression.

Over the last few years, functional genomics approaches, including transcriptomics, have been increasingly used to obtain global gene expression profiles, thereby providing a comprehensive view of microorganism physiology [12,13].

Although the genome sequences of several *L. garvieae* strains from different origins have been published recently [14-21], such global approaches have not yet been used to study the transcriptome of this pathogen. In the present study, we used microarray expression analysis to evaluate global transcriptional changes occurring in two clinical *L. garvieae* strains, isolated from fish and a human, respectively, when incubated at 18°C and 37°C. These temperatures correspond to that at which fish lactococcosis outbreaks usually occur and the physiological temperature in humans, respectively. This first transcriptome analysis of two *L. garvieae* strains demonstrated that this bacterium responds globally to temperature.

Materials and Methods

Bacterial strains and culture conditions

Lactococus garvieae strain 8831 (Lg8831) was isolated from diseased rainbow trout suffering lactococcosis [16], and *Lactococcus garvieae* strain 21881 (Lg21881) was isolated from a case of human septicemia [15]. The growth kinetics of Lg8831 and Lg21881 were studied at 18°C and 37°C. To minimize variation in experimental culture conditions and ensure reproducibility, a standardized inoculum was prepared by adding 2 mL of an overnight culture of Lg8831 or Lg21881 incubated at 29°C in 150 ml BHI broth. Lag-phase and bacterial growth rate were determined in three independent experiments at both temperatures by monitoring OD_{600} each hour until OD_{600} ~ 1.5 was reached. Differences in lag-phase and bacterial growth rate were assessed using the Fisher exact test with the software SPSS 19.0 (IBM, New York, USA). Differences were considered significant when $p<0.05$.

RNA extraction and purification

Lg8831 and Lg21881 were grown aerobically in BHI broth (bioMérieux, Marcy l'Etoile, France) at several different temperatures (18°C, 29°C, and 37°C) and harvested at the mid-log phase (OD_{600} ~ 0.9) for RNA extraction. Total RNA from Lg8831 and Lg21881 was isolated from three independent samples (biological replicates) of each temperature condition by using RNeasy Protect Bacteria Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. Briefly, 1.5 mL of each culture sample were collected in 3 mL of RNAprotect and centrifuged. Pellets were resuspended in 200 µL of TE buffer containing 20 mg/mL of lysozyme (Sigma-Aldrich, Brondby, Denmark) and 20 µL of proteinase K solution (Qiagen), and incubated at RT for 60 minutes under agitation conditions. Total RNA extraction and purification was then performed using RNeasy kit columns including on-column DNA digestion. The quality and concentration of RNA was determined by using the RNA 6000 NanoKit on the Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA) at Genomics Unit facilities (Parque Científico de Madrid). Only high- quality RNA samples (RIN > 9) were used for the next step.

Microarray and experimental design

A DNA microarray was designed based on the published Lg8831 [16] and Lg21881 [15] genome sequences. The microarray, containing 60 mer oligonucleotide specific probes for the 1992 and 2275 ORFs of the Lg8831 and Lg21881 genomes respectively, was designed using Agilent's e-array software (https://earray.chem.agilent.com/earray/). The best probe method was used, and each probe was included in triplicate on the array. This custom oligonucleotide microarray was manufactured by Agilent Technologies on an 8x15K format.

A dual-color reference design experiment was used [22]. Interrogated samples of Lg8831and Lg21881 were grown in triplicate at 18°C and 37°C, and RNA was extracted. These RNAs were independently labeled with Cy-5. A reference sample made of an equimolar mix of Cy-3-labeled mixture of Lg8831 and Lg21881 RNAs was used for all hybridizations. These reference RNAs were obtain from both strains grown separately at 29°C, quantified and then mixed for labeling.

For the comparison of the Lg8831 and Lg21881 strain transcriptomic responses, a subset of common probes derived from those designed for Lg8831 was used. The criteria used for the selection of these subset of probes for the inter-strain comparisons was that probes designed for the strain Lg8831 should have a BLAT [23] similarity >98.4%, (no more than a single mismatch) with the Lg21881 genome. Thus, a total of 1302 probes (65% of probes for Lg8831) were selected and used.

cDNA synthesis, labeling of total RNA, hybridization and scanning

A total of 5 µg of total RNA of each sample replicate was reverse transcripted and labeled by using the FairPlay III Labeling Kit (Agilent Technologies) according to the protocol provided by the manufacturer except for the following: purification of the dye-coupled cDNA was performed by using QIAquick PCR purification kit (Qiagen). The yield and the specific activity of each labeling reaction were determined in a NanoDrop 1000 spectrophotometer (NanoDrop Technologies, Inc., Rockland, DE). The hybridization and washing steps were carried out according to the Agilent's "Two Color Microarray based Prokaryote Analysis" protocol. Image acquisition and scanning was performed on an Agilent G2565AA scanner according to the manufacturer's instructions (GE2_107_Sep09).

Data analysis

Image quantification was carried out using Agilent's Feature Extraction Software v 8.5.1.1. Background-subtracted data were log-transformed, LOWESS-normalized, and analyzed for the detection of differentially expressed genes using BRB-ArrayTools Software developed by Dr. Richard Simon and the BRB-ArrayTools Development Team [24]. Triplicate probes on the microarray were averaged for the analyses. The SAM algorithm was used to identify the differentially expressed genes using a 0.05 proportion of false discoveries (FDR) with a 90% confidence. A further filtering step was carried out by

Table 1. Primers used for RT-qPCR experiments.

Strain	Gene ID	Predicted protein function	Primer ID	Sequence 5'- 3'
Lg8831	8831_c60_g79	Glycine betaine ABC transporter permease/substrate binding protein	g79c60_Forward	GCCCGGTTTCGTTTACTTGA
			g79c60_Reverse	AGGGACCACACCGATACCAA
	8831_c18_g15	Fructose operon transcriptional regulator	g15c18_Forward	TTTGCTTGGTGGTCGAGTGA
			g15c18_Reverse	GATGGCCGAACTGCCAAT
	8831_c49_g33	Cold-shock protein A	g33c49_Forward	CATCACTGCTGAAGATGGTACTGA
			g33c49_Reverse	GAAGCCATCGCTTTGAATTTG
	8831_c4_g16	Crp family transcriptional regulator	g16c4_Forward	AAAAGCGGAAGGGATATTAATTGA
			g16c4_Reverse	GAGATTCCACAAAATCCTGCTATGT
	8831_c96_g10	Glutamate decarboxylase	g10c96_Forward	CCGAAATCGAAAACCGTTGT
			g10c96_Reverse	AACTGTTCGTTTTCACTTGCATTC
Lg21881	21881_c16_g17	Pyridine mercuric reductase	g17c16_Forward	GGTGTTTTTGACCGTGATGAAG
			g17c16_Reverse	CTTCTGCCTGACCGTCTCTTG
	21881_c16_g19	TetR-family transcriptional regulator	g19c16_Forward	TCCGCTATTGCGCCAGAT
			g19c16_Reverse	CCCAGTTTAGCATTGCAAGCA
	21881_c63_g5	Glycerol uptake facilitator protein	g5c63_Forward	TCTGAACAAGACGAAAGCAACAG
			g5c63_Reverse	CAAAGCCCCAACCGAGTGT
	21881_c30_g6	Alcohol dehydrogenase	g6c30_Forward	GCAGGTAATGGTGCCGTCTT
			g6c30_Reverse	GCAGGTCTTCAACTGCACGAT
	21881_c50_g82	Activator of 2-hydroxyglutaryl-CoA dehydratase	g82c50_Forward	GGGCTGCGCGTTGTTGTTAAT
			g82c50_Reverse	CACTACTTGCGCTGCAGCAT
Lg21881 & Lg8831	*gyrA* *	DNA gyrase subunit A	gyrA_Forward	ACGGAATGAACGAGCTTGGTA
			gyrA_Reverse	CCGGTAATACGGGCAGATTTT

*. Reference gene used in RT-qPCR experiments for both strains.

doi: 10.1371/journal.pone.0079692.t001

selecting among SAM results only those whose expression ratio between the analyzed conditions was ≥2.

Throughout this article we have used the term "up-regulated genes" to indicate those genes that were found differentially overexpressed at one temperature relative to the other.

Microarray accession numbers

The data presented in this publication have been deposited in the Gene Expression Omnibus (GEO) database at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE40318.

Validation of microarray data by reverse-transcription quantitative real-time PCR (RT-qPCR)

To validate the results obtained during the microarray analysis, RT-qPCR analysis was performed. A total of 10 genes were selected for the validation process. These genes were selected by taking the top three expressed and top two repressed genes at each comparison between the two temperatures for each strain. The *gyrA* gene was selected as an endogenous control because it did not present variation in its expression throughout the experiment. Triplicate assays utilizing the same three independent RNA samples used for the microarray hybridization were used for the RT-qPCR validation. The RT-qPCR primers were designed using Primer Express software version 2.0 (Applied Biosystems Technologies, Paisley, UK) on the basis of the consensus region of each of the sequences from the 10 genes of Lg8813 and Lg21881 strains (Table 1). cDNA was synthesized using the High Capacity RNA-to-cDNA Kit (Applied Biosystems). Amplification was performed in a 7900HT Fast Real-Time PCR Systems apparatus (Applied Biosystems) in triplicate for each sample using the FastStart Universal SYBR Green Master (Rox) (Roche Diagnostics, Basel, Switzerland) at Genomics Unit facilities (Parque Científico de Madrid). PCR amplification was initiated at 95°C for 10 min followed by 40 cycles at 95°C for 15 s and 60°C for 1 min. Fluorescence due to the binding of the SYBR Green to double-stranded DNA was measured at each cycle. The threshold cycle value (Ct) was obtained by automatic position of the threshold baseline at the mid-exponential phase of the curve. Data normalization and analysis were performed by means of RQ Manager Software Version 1.2. The 37°C condition was assigned as the reference sample for calculating relative gene expression.

Results and Discussion

Global expression patterns of Lg8831 and Lg21881 in response to growth temperature

We used whole-genome DNA microarrays to obtain a comprehensive overview of the molecular response of Lg8831 and Lg21881 strains grown at 18°C and 37°C. Differentially expressed genes detected with a 0.05% FDR and at least a two-fold change were selected. Under these conditions, 264 genes showed statistically significant differential expression for

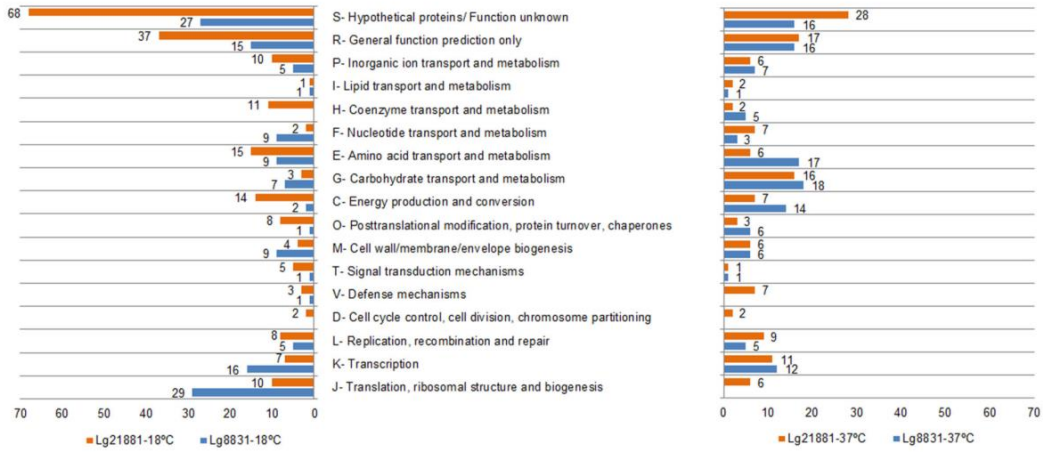


Figure 2. COG categories distribution of differentially expressed genes in Lg8831 and Lg21881 at 18°C and 37°C. Bars indicate the number of genes up-regulated in each functional group at each temperature in Lg8831 (blue) and Lg21881 (orange).
doi: 10.1371/journal.pone.0079692.g002

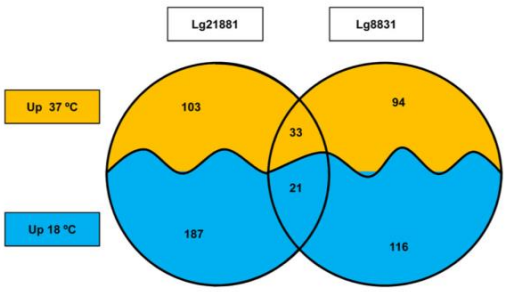


Figure 1. Comparison of the transcriptomic response of Lg8831 and Lg21881 grown at 18°C and 37°C. Up 37°C and Up 18°C represent up-regulated genes at each temperature.
doi: 10.1371/journal.pone.0079692.g001

Lg8831: 137 (6.9% of its genome) and 127 genes (6.4% of its genome) were up-regulated at 18°C and 37°C, respectively (Figure 1). For Lg21881, 344 differentially expressed genes were identified: 208 (9.2% of its genome) up-regulated at 18°C and 136 (6.0% of its genome) at 37°C (Figure 1). These results are in accordance with other studies that observed that bacteria can modify the expression of approximately 10% of their genes in response to an increase or decrease in growth temperature [10,11,25]. Genes displaying a significant differential expression are listed in Tables S1-S4.

The oligonucleotide array was designed to be optimized for the detection of each of the *L. garvieae* strains; therefore, for inter-strain comparisons, it was necessary to assess the similarity and specificity of the probes. Thus, this step allowed us to identify 1302 matching probes meeting the selection criteria used. Among these matching genes, we only found 21 genes up-regulated at 18°C in both strains (1.6% of the matching genes), and 33 genes up-regulated at 37°C in both strains (2.5% of the matching genes) (Figure 1). These results demonstrated that the core gene expression response to temperature is minimal, and evidenced a strain-specific variation in transcriptional responses to temperature of Lg8831 and Lg21881 strains.

We categorized the differentially regulated genes based on the Clusters of Orthologous Groups (COG). As shown in Figure 2, the pattern and distribution of differentially expressed genes among COG categories was different for Lg8831 and Lg21881 strains and depended on the temperature.

In relation to the functional gene categories, Lg8831 and Lg21881 up-regulated at 37°C mainly genes within categories related to bacterial growth and metabolism such as COG categories G (carbohydrate transport metabolism), E (amino acid transport and metabolism) and K (transcription) (Figure 2). This result was expected considering that 37°C is the optimal growth temperature for this microorganism [7]. However, at 18°C the distribution pattern between gene categories varied greatly between strains (Figure 2). Thus, COG categories J (translation, ribosomal structure and biogenesis) and K (transcription) were the most represented at 18°C for Lg8831, whereas Lg21881 primarily up-regulated genes within the COG categories C (energy production and conversion) and H (coenzyme transport and metabolism) (Figure 2). Overall, these results (Figures 1 and 2) indicate a different transcriptional response of Lg8831 and Lg21881 to temperature, which was more evident at 18°C. The most significant differences observed in both strains are discussed in detail below.

Table 2. Growth kinetics parameters for Lg8831 and Lg21881 at 18°C and 37°C.

	18°C		37°C	
Strain	**Lg8831**	**Lg21881**	**Lg8831**	**Lg21881**
Lag-phase (h)	7.65 ± 0.60	12.32 ± 0.60	2.05 ± 0.10	2.1 ± 0.20
Growth rate (h^{-1})	0.14 ± 0.03	0.18 ± 0.05	0.63 ± 0.06	0.77 ± 0.05

doi: 10.1371/journal.pone.0079692.t002

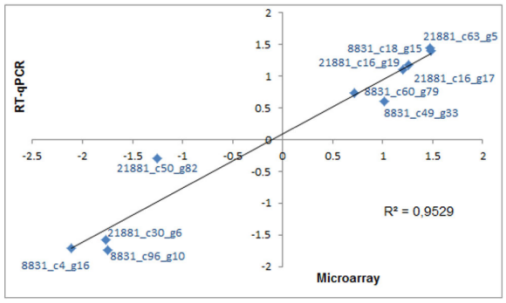


Figure 3. Correlation between RT-qPCR and microarray analysis data. The $\log_{10}$ values of the real-time RT-qPCR were plotted against the microarray analysis $\log_{10}$ values. The expression levels for 10 selected genes are shown.

doi: 10.1371/journal.pone.0079692.g003

To assess the influence of the temperature in the growth kinetics of Lg8831 and Lg21881 and look for a possible relation between differences in their transcriptomes and differences in growth, we performed growth kinetics studies at 18°C and 37°C for both strains. At the same temperature (18°C or 37 °C), Lg8831 and Lg21881 exhibited similar growth rates (Table 2), but the growth rate at 37°C were higher for both strains compared to 18°C ($p<0.05$). However, at 18°C Lg8831 showed a shorter lag-phase (7.65 h) than that exhibited by Lg21881 (12.32 h) (Table 2) and this difference was statistically significant ($p<0.05$). The shorter lag-phase of Lg8831 grown at 18°C suggests that this strain could be better adapted to grow at low temperatures.

DNA microarray analysis verification

To validate the microarray transcription profiling results, 10 differentially expressed genes were selected for quantitative RT-qPCR experiments with the same RNA samples used in the array hybridizations. Among them, six genes (three for each strain) were up-regulated at 18°C, and four genes (two for each strain) were repressed. There was a strong positive correlation ($r=0.953$) between the data obtained by both techniques (Figure 3). Thus, this strong correlation allowed us to validate the microarray results.

Transcriptional profiling of Lg8831 and Lg21881 at 18°C compared to 37°C

Induction of a specific cold stress response in Lg8831. The functional analysis of the differentially up-regulated genes in Lg8831 at 18°C showed that a large number of genes were involved in translation, ribosomal structure and biogenesis (COG category J; 21%) and in transcription (COG category K; 12%). The up-regulation of genes within functional COG category J at 18°C could be implicated in the adaptation of Lg8831 to low temperature. Most of the up-regulated genes among this category corresponded to ribosomal proteins (Table S1). It has been previously shown that ribosomal proteins, translation factors and other components that help in the function and/or biogenesis of ribosomes are induced after exposure to low temperatures [26]. Ribosomal proteins act as chaperones ensuring accurate translation at low temperatures [27]. Furthermore, it is not surprising that a large proportion of genes involved in COG category K were up-regulated at 18°C, because is necessary the expression of specific transcription factors for the initiation of the transcription of cold-inducible genes whose products would assist in the adaptation to temperatures below the optimum [28].

A great number of the genes among those found to be up-regulated at 18°C in Lg8831 encode proteins involved in a specific low-temperature response (Table S1, Figure 4). Thus, we observed the up-regulation of three *csp* homologues encoding cold-shock proteins (CSPs). The induction of CSPs has been described for a great number of bacteria exposed to low temperatures [26], mainly acting as RNA chaperones facilitating transcription and translation at low temperature in mesophilic bacteria [28,30]. We also observed the up-regulation of genes encoding previously described cold-induced proteins (CIPs), which are necessary for cellular adaptation to low temperature [26,31]. Some of these up-regulated genes include: genes involved in transcription such as *nusA* (transcription factor NusA) [30] and *greA* (transcription elongation factor GreA); and in translation, such as *infA* (translation initiation factor IF1) [32], *infB* (translation initiation factor IF2) [29], *infC* (translation initiation factor IF3) [33], and *rbfA* (ribosome binding factor A) [34,35] (Table S1).

Transport systems allow bacteria to cope with environmental changes [31]. Genes encoding transport systems represented 23% of the up-regulated genes at 18°C, suggesting that the transport and secretion of molecules through the membrane could play significant role in the cold-adaptive process of Lg8831. Within this group, ABC transporters represented more than 50% (Table S1). Although energetically costly, ABC transporters usually have very high affinities for their solutes and catalyze transport at high rates [36]. Hence, induction of ABC transporters may allow Lg8831 to efficiently scavenge essential solutes giving it a competitive advantage when grown at low temperatures and assisting in cold adaptation. Transcription of *busAA* and *busAB* homologues encoding a glycine betaine uptake ABC transporter system, as well as transcription of the *potABCD* operon encoding the spermidine putrescine uptake ABC transport system, were found to be up-regulated at 18°C (Table S1). Both transport systems are

Figure 4. Comparison of the expression levels of cold stress-related genes in Lg8831 and Lg21881 at 18°C and 37°C by microarray analysis.

doi: 10.1371/journal.pone.0079692.g004

important response mechanisms to cold conditions. Spermidine putrescine, and glycine betaine are compatible solutes that are known to accumulate during bacterial growth at low temperature, acting as cryoprotectants that prevent denaturing and the cold-induced aggregation of proteins [26,37,38].

Cryoprotectant substances are also believed to play a role in maintaining optimum membrane fluidity at low temperature [39]. Spermidine has been described as a bacteriolysis-inducing activator in staphylococci, bacilli, and in certain streptococcal species [40]; and pronounced cell wall breakdown has been associated to responses to stress conditions [41,42]. Interestingly, we found the up-regulation of three genes encoding autolytic cell-wall breaking enzymes (gene25_contig19, gene51_contig19 and gene20_contig36), which act as muramidases that hydrolyze the beta-1,4-linked polysaccharides of the peptidoglycan (Table S1). One of the regulation mechanisms of autolytic enzymes is the teichoic acids system [43,44]. In this sense, 4 genes encoding proteins implicated in teichoic acid biosynthesis were found to be up-regulated (*tagD, tagF, tagG, tagH*) (Table S1). Additionally, *epsL* (encoding exopolysaccharide biosynthesis protein) and *glmS* (encoding glucosamine-fructose-6-phosphate aminotransferase) homologues were also up-regulated. These two genes are related to biosynthetic processes of membrane components. Hence, genes related to membrane functions (COG category M) that represents 6.5% of the up-regulated genes at 18°C in Lg8831, appear to be linked to both biological processes: the degradation and the biosynthesis of membrane components. This observation could be related to the reorganization of membrane composition, which is associated with growth at low temperatures [26,28].

Another significant result was the pronounced up-regulation of homologues of fructose metabolism (*fru* operon: *fruRCA*) (Table S1). This operon is regulated by the repressor FruR in the absence of fructose in *Lactococcus lactis* and low-GC Gram-positive bacteria. The FruR effector is fructose-1-phosphate, which is necessary to relieve repression by FruR [45]. Therefore, the observed up-regulation of the *fru* operon in Lg8831 requires the presence of fructose-1-phosphate, which is likely generated as a consequence of the cell wall breakdown processes described above (KEGG pathway: lgr00520) that yields fructose-6-phosphate. This fructose-6-phosphate is then redirected to the glycolysis pathway and converted to fructose-1,6-biphosphate, which is further spontaneously dephosphorylated to fructose-1-phosphate [45].

Thus, a relevant finding of this work is that Lg8831 is able to sense and adapt to temperatures below its optimum by means of differential expression of genes whose products are involved in a specific cold response. Additionally, this specific and coordinated remodeling of the transcriptome could be related to the differential growth kinetics, associated with a shorter lag phase, demonstrated by this strain (Table 2).

Induction of a general stress response in Lg21881. The largest group among the up-regulated genes at 18°C (31%) was the phage, prophage, and chromosomal mobile elements. The up-regulation of this group of genes has been observed in several bacteria under different stress conditions [46-48] and phage-related proteins have been previously reported to be up-regulated in group A *Streptococcus* at 29°C compared to 37°C [8]. Phages, prophages, and mobile elements play a role in bacterial genome diversification, and there is evidence of the

induction of genetic exchange mechanisms as a general response to stress in Gram-positive bacteria [49]. An explanation to this fact is that, at low growth rates, cells enter into a state favorable for DNA rearrangements [48].

Furthermore, we observed the up-regulation of some genes related to the cold-stress response that were also found to be up-regulated in Lg8831 at 18°C, such as *nusA*, *cspA* and the *potABCD* operon (Table S2, Figure 4). The Lg21881 *opuB* homologue was up-regulated at 18°C. The Opu transport system is described as involved in cryoprotection in bacteria, serving for the acquisition of a substantial number of osmoprotectants. It has been demonstrated that the Opu transporters mediate the uptake of the cryoprotectant solute glycine betaine in *Bacillus subtilis* cold-stressed cells [37]. In addition to *nusA*, *cspA*, *potABCD* and *opuB*, we also found up-regulated genes that have been widely described as general stress responsive genes (Table S2) such as: (i) two transcripts encoding OsmC (osmotically inducible protein C)-like proteins (gene32_contig62 and gene67_contig49), whose expression is regulated by multiple stress conditions [50]; (ii) the gene66_contig50 transcript, which encodes a general stress family protein; and (iii) the gene15_contig69, which encodes a Clp protease that has been described as general biomarker of stress adaptive behavior in a wide range of bacteria [51].

Several genes encoding proteins presumed to be involved in iron homeostasis were up-regulated at 18°C in Lg21881. These genes include: the *fhu* operon (see below); genes of the iron ABC transport system (gene26_contig33, gene27_contig33, gene28_contig33 and gene29_contig33; see TS4); *sufB*, *sufC*, *sufD*, *sufS* homologues; and a gene encoding a NifU-like protein. Iron transporter proteins as well as other proteins involved in iron homeostasis are differentially expressed in response to growth temperatures in group A *Streptococcus* [8]. Suf and NifU proteins play a role in the iron-sulfur (Fe-S) cluster assembly machinery [52]. Bacterial Fe-S proteins are essential regulators of gene transcription under stress conditions, acting as sensors of the environment and enabling the organism to adapt to prevailing conditions [53]. These genes involved in iron homeostasis could assist Lg21881 to adapt to low temperature conditions contributing to the general stress response showed by this strain at 18°C.

We observed a strong induction of the *glpKDF* operon at 18°C (Table S2). Up-regulation of GlpD has also been reported in *Lactobacillus sakei* when exposed to low temperature [54]. It is known that glycerol metabolism is linked to membrane properties. Glycerol-3-phosphate can be converted to phosphatidic acid towards the activity of an acetyltransferase, which leads to membrane phospholipid synthesis that it is likely linked to changes in membrane composition associated to maintenance of membrane integrity when grown at low temperatures [28,54]. The observed up-regulation of an acetyltranferase transcript (gene8_contig37; see Table S3), which may be related to the synthesis of phosphatidic acid, makes sense in this light. Alternately, the activity of GlpD is also related to the induction of respiratory metabolism in Lg21881 when grown at 18°C (see below).

Aerobic respiration activation in Lg21881. Energy production and conversion, and coenzyme transport and metabolism (COG categories C and H) represented 13% of the up-regulated genes at 18°C in Lg21881. An important finding in this study is that 11 genes within these categories (45%) were related to an aerobic respiratory metabolism. LAB have usually been considered as fermentative bacteria, but in recent years some studies have noted the potential of the respiratory metabolism in different lactococcal strains, demonstrating that they undergo respiration and thrive when they are grown in the presence of oxygen and a heme source [55-57].

In *L. lactis* subsp. *lactis*, the membrane respiration chain requires dehydrogenases (membrane proteins NoxB and/or GlpD), menaquinones (encoded by *men* genes), and a terminal electron acceptor (the cytochrome oxidase encoded by *cyd* genes) [55,57]. At 18°C in Lg21881, we observed the up-regulation of homologues within the menaquinone biosynthesis pathway (*menB*, *menC*, *menD*, *menE*, *menH* and *ubiE*, see Table S2), and two homologues of the cytochrome oxidase synthesis operon (*cydC* and *cydD*). Additionally, the strong up-regulation of *glpD* (see above) is likely related to both independent biological processes occurring at 18°C in Lg21881: playing a role as a membrane dehydrogenase that is necessary for the aerobic respiration process, and in the biosynthesis of membrane phospholipids associated with membrane changes at low temperatures.

Homologues within the *fhu* operon (*fhuB fhuC* and *fhuD*) encoding subunits of the iron complex ABC transport system were also up-regulated at 18°C. The *fhu* operon is responsible for heme uptake in respiring *L. lactis* cells [57] and is likely to be responsible for heme import from the external environment in Lg21881. Heme is required to activate a respiration metabolic pathway in streptococci such us *L. lactis* and *Streptococcus agalactiae* [55,58]. Because Lg21881 lacks the appropriate genes for heme biosynthesis, heme must be present in the medium to activate respiration. Although BHI broth was not supplemented with heme, trace amounts of heme in BHI medium have been reported to activate respiration-related genes [58].

We also observed in Lg21881 the up-regulation of genes encoding enzymes related to mixed-acid fermentation that could be associated to the respiratory metabolism such as proteins within the pyruvate dehydrogenase complex (*pdhABD* and dihydrolipoamide acetyltransferase homologues), *als* (acetolactate synthase), and *pycA* (pyruvate carboxylase) (Table S2). When oxygen and heme are present, *L. lactis* shifts to a mixed-acid fermentation, more complete glucose utilization, and energy generation by NADH oxidation via an electron transfer chain [36]. LAB lack a complete Krebs cycle, and consequently, NADH, which is required for the respiratory chain, is produced by carbon catabolism. Once phosphorylated, sugar is catabolized to pyruvate via glycolysis with production of ATP and NADH. Thus, pyruvate dehydrogenase (Pdh) provides extra NADH from pyruvate when oxygen is present [57]. Moreover, we observed in Lg21881 the up-regulation of NADH oxidase, an enzyme that in *L. lactis* is responsible for the shift to the mixed acid fermentation pathway in aerobically grown cells [59].

The up-regulation of respiration-related genes at 18°C in Lg21881 may be linked to the increased solubility of oxygen at

lower temperatures, which may increase oxidative stress. Thus, the activation of respiration activity can protect cells against damage by consuming oxygen via respiratory metabolism [53]. We did not, however, observe a significant up-regulation of the respiratory pathways in strain Lg8831 at 18°C highlighting the possibility that this phenomenon is strain-dependent. The specific up-regulation of aerobic respiration-related genes in Lg21881 may be related to the general response to stress observed at 18°C for this strain. To the best of our knowledge, this is the first time respiratory metabolism has been described in *L. garvieae*.

Transcriptional profiling of Lg8831 and Lg21881 at 37°C compared to 18°C

Carbohydrate transport- and metabolism-related genes (COG category G) were the largest group among the up-regulated genes in both strains (Figure 2), which is in accordance with the high growth rate and metabolic activity at 37°C shown by both strains (Table 2).

We observed the up-regulation of genes involved in an elevated glycolytic activity at 37°C in Lg8831, which is likely related to the higher growth rate at 37°C compared to 18°C (Table 2). This group includes: (i) genes encoding glycolytic enzymes such as *gapB* (glyceraldehyde-3-phosphate dehydrogenase), *fbaA* (fructose-biphosphate aldolase), *scrK* (fructokinase) and *pyk* (pyruvate kinase); (ii) enzymes of fermentation pathways such as ldh (lactate dehydrogenase) and *pdhAB* (pyruvate dehydrogenase complex); and (iii) PTS systems, necessary for the transport and phosphorylation of sugars before their utilization and genes within the energy production and conversion pathways (Table S3). The growth of lactic acid bacteria is characterized by the generation of acidic end products of fermentation (mainly lactic acid), which accumulate in the extracellular milieu leading to an acidification of the environment [60]. In this regard, the expression of acid resistance-related genes could help to cope with these environmental changes. Thus, 17 of the up-regulated genes at 37°C in Lg8831 (13%) were related to an acid resistance response, including: (i) genes within the three main systems involved in lactococcal pH homeostasis: the F_1F_0-ATPase proton pump, the glutamate decarboxylase (GAD) system and the arginine deiminase (ADI) pathway [61]; (ii) the *lctO* gene encoding L-lactate oxidase, an enzyme that has been implicated in the removal of excess lactate that accumulates after fermentation [62]; and (iii) other described biomarkers of the acid resistance response such as *dnaK* and *groEL* [60], and *clpE* and *clpB* [51] (Table S3).

In addition, 19 out of the 136 up-regulated transcripts at 37°C (14%) in Lg21881 corresponded to plasmid-encoded genes (Table S4). Lg21881 carries five circular plasmids encoding heterogeneous functions acquired by horizontal gene transfer from other lactic acid bacteria (LAB) [63]. The optimum growth temperature for *L. garvieae* is 37°C [7]. It is therefore expected that the expression of extra chromosomal material is favored at this temperature.

Responses potentially linked to the pathogenesis of *L. garvieae*

The influence of temperature in the expression of virulence factors has been extensively studied in different bacterial pathogens such as *Listeria monocytogenes* [64] and *Pseudomonas aeruginosa* [65]. In this study, several genes encoding proteins associated with the pathogenesis of different pathogens were up-regulated in Lg8831 at 18°C or in Lg21881 at 37°C.

The *rpoE* homologue encoding the delta subunit of RNA polymerase, the cold-responsive *potABCD* operon, and three genes (gene25_contig19, gene51_contig19, gene20_contig36) encoding autolytic enzymes were up-regulated in Lg8831 at 18°C (Table S1). RpoE has been described as an essential global modulator of environmental adaptation in Gram-positive bacteria such us *S. mutans* [66], and it has also been linked to the virulence of *S. agalactiae* [67,68] and the fish pathogen *Vibrio harveyi* [69]. The cold-responsive *potABCD* operon has been described as necessary for the pathogenesis of *S. pneumoniae* in experimentally infected mice [70]. In particular, it has been demonstrated that PotD it is intricately linked to the fitness, survival, and pathogenesis of pneumococci in host microenvironments [71]. Autolytic enzymes also play an important role in the infectious process in Gram-positive pathogens [40]. When bacteriolysis occurs during infection, cell wall- and membrane-associated lipopolysaccharides (endotoxin), lipoteichoic acids, and peptidoglycan are released. These compounds can act on macrophages to induce the release of reactive oxygen and nitrogen species, cytotoxic cytokines, hydrolases, and proteinases, and also activate the coagulation and complement cascades. All these agents and processes are involved in the pathophysiology of septic shock and multiple organ failure resulting from severe microbial infection [40]. Lg8831 was isolated from rainbow trout suffering lactococcosis, a disease greatly influenced by temperature [1]. Therefore, the up-regulation of these genes in Lg8331 at 18°C could be relevant in the pathogenesis of lactococcosis, a disease characterized by a severe generalized septicemic process in affected fish. To confirm this hypothesis, our group is currently performing transcriptomic studies in experimentally infected trout.

In Lg21881, genes that may be associated with pathogenesis were those related to manganese homeostasis up-regulated at 37°C: a *mtnR* homologue encoding a manganese-dependent transcriptional regulator of the DtxR family; *mtsABC* homologues encoding a manganese ABC transporter; and a *mntH* homologue encoding a Nramp family manganese transport protein (Table S4). MtnR was first described in *B. subtilis* as a transcriptional regulator, which responded to manganese. The manganese ABC transporter and MntH are the targets of MntR regulation [72]. Both, Nramp and ABC class manganese transporters have been extensively described as essential for the virulence of several Gram-positive pathogens [73]. In particular, MntH has been linked to the virulence of pathogens such as *Salmonella enterica* [74], *Escherichia coli* [75], *Brucella abortus* [76] and *Yersinia pestis* [77]. Thus, as in other pathogens, the regulation of the

expression of these manganese transporters could also play a role in the pathogenesis of infections caused by Lg21881.

Conclusion

The data coming from the whole genome DNA microarray of *L. garvieae* constructed in the present study demonstrated excellent agreement in the microarray validation experiments, indicating that it can be a useful tool for transcriptome studies in *L. garvieae*. In the present study, this microarray was used to perform the first transcriptome analysis of *L. garvieae*. This study of gene expression revealed that Lg8831 and Lg21881 differentially respond to temperature suggesting strain-specific adaptation mechanisms. In addition, it is the first time that a respiratory metabolism has been described in *L. garvieae*. These data extend our understanding of the regulatory networks and biology of this important pathogen.

Supporting Information

Table S1. Genes showing significant up-regulation by microarray hybridization in Lg8831 grown at 18°C compared to 37°C.
(DOC)

Table S2. Genes showing significant up-regulation by microarray hybridization in Lg21881 grown at 18°C compared to 37°C.
(DOC)

Table S3. Genes showing significant up-regulation by microarray hybridization in Lg8831 grown at 37°C compared to 18°C.
(DOC)

Table S4. Genes showing significant up-regulation by microarray hybridization in Lg21881 grown at 37°C compared to 18°C.
(DOC)

Acknowledgements

The authors would like to thank Dr. B. Perez-Villamil for the use of the microarray facilities for microarray hybridization and scanning, C. Aspiroz for kindly supplying the human clinical strain, and Dr. P. Gaudu for his helpful comments on the respiratory metabolism of *L. garvieae.*

Author Contributions

Conceived and designed the experiments: AG MMB GHLC. Performed the experiments: MAU. Analyzed the data: MAU VLA GHLC. Wrote the manuscript: MAU AG MMB JFFG GHLC.

References

1. Vendrell D, Balcázar JL, Ruiz-Zarzuela I, de Blas I, Gironés O et al. (2006) *Lactococcus garvieae* in fish: a review. Comp Immunol Microbiol Infect Dis 29: 177–198. doi:10.1016/j.cimid.2006.06.003. PubMed: 16935332.
2. Aguado-urda M, Cutuli MT, Blanco MM, Aspiroz C, Tejedor JL et al. (2010) Utilization of lactose and presence of the phospho-β-galactosidase (*lacG*) gene in *Lactococcus garvieae* isolates from different sources. Int Microbiol 13: 189–193. PubMed: 21404213.
3. Pot B, Devriese LA, Ursi D, Vandamme P, Haesebrouck F et al. (1996) Phenotypic identification and differentiation of *Lactococcus* strains isolated from animals. Syst Appl Microbiol 19: 213–222. doi:10.1016/S0723-2020(96)80047-6.
4. Russo G, Iannetta M, D'Abramo A, Mascellino MT, Pantosti A et al. (2012) *Lactococcus garvieae* endocarditis in a patient with colonic diverticulosis: first case report in Italy and review of the literature. New Microbiol 35: 495–501. PubMed: 23109019.
5. Ferrario C, Ricci G, Borgo F, Rollando A, Fortina MG (2012) Genetic investigation within *Lactococcus garvieae* revealed two genomic lineages. FEMS Microbiol Lett 332: 153–161. doi:10.1111/j.1574-6968.2012.02591.x. PubMed: 22568590.
6. Kubota H, Tsuji H, Matsuda K, Kurakawa T, Asahara T et al. (2010) Detection of human intestinal catalase-negative, Gram-positive cocci by rRNA-targeted reverse transcription-PCR. Appl Environ Microbiol 76: 5440–5451. doi:10.1128/AEM.03132-09. PubMed: 20581195.
7. Kusuda R, Kawai K, Salati F, Banner CR, Fryer JL (1991) *Enterococcus seriolicida* sp. nov., a fish pathogen. Int J Syst Bacteriol 41: 406–409. doi:10.1099/00207713-41-3-406. PubMed: 1883715.
8. Smoot LM, Smoot JC, Graham MR, Somerville GA, Sturdevant DE et al. (2001) Global differential gene expression in response to growth temperature alteration in group A *Streptococcus*. Proc Natl Acad Sci U S A 98: 10416–10421. doi:10.1073/pnas.191267598. PubMed: 11517341.
9. Stintzi A (2003) Gene expression profile of *Campylobacter jejuni* in response to growth temperature variation. J Bacteriol 185: 2009–2016. doi:10.1128/JB.185.6.2009-2016.2003. PubMed: 12618466.
10. Mereghetti L, Sitkiewicz I, Green NM, Musser JM (2008) Remodeling of the *Streptococcus agalactiae* transcriptome in response to growth temperature. PLOS ONE 3: e2785. doi:10.1371/journal.pone.0002785. PubMed: 18665215.
11. Wu DQ, Ye J, Ou HY, Wei X, Huang X et al. (2011) Genomic analysis and temperature-dependent transcriptome profiles of the rhizosphere originating strain *Pseudomonas aeruginosa* M18. BMC Genomics 12: 438. doi:10.1186/1471-2164-12-438. PubMed: 21884571.
12. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H et al. (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. Nature 459: 950–956. doi:10.1038/nature08080. PubMed: 19448609.
13. Kocabaş P, Calik P, Calik G, Ozdamar TH (2009) Microarray Studies in *Bacillus subtilis*. Biotechnol J 4: 1012-1027. doi:10.1002/biot.200800330. PubMed: 19618384.
14. Reimundo P, Pignatelli M, Alcaraz LD, D'Auria G, Moya A et al. (2011) Genome sequence of *Lactococcus garvieae* UNIUD074, isolated in Italy from a lactococcosis outbreak. J Bacteriol 193: 3684–3685. doi: 10.1128/JB.05210-11. PubMed: 21602331.
15. Aguado-Urda M, López-Campos GH, Blanco MM, Fernández-Garayzábal JF, Cutuli MT et al. (2011) Genome sequence of *Lactococcus garvieae* 21881, isolated in a case of human septicemia. J Bacteriol 193: 4033–4034. doi:10.1128/JB.05090-11. PubMed: 21622739.
16. Aguado-Urda M, López-Campos GH, Gibello A, Cutuli MT, López-Alonso V et al. (2011) Genome sequence of *Lactococcus garvieae* 8831, isolated from rainbow trout lactococcosis outbreaks in Spain. J Bacteriol 193: 4263–4264. doi:10.1128/JB.05326-11. PubMed: 21685280.
17. Morita H, Toh H, Oshima K, Yoshizaki M, Kawanishi M et al. (2011) Complete genome sequence and comparative analysis of the fish pathogen *Lactococcus garvieae*. PLOS ONE 6: e23184. doi:10.1371/journal.pone.0023184. PubMed: 21829716.
18. Ricci G, Ferrario C, Borgo F, Rollando A, Fortina MG (2012) Genome Sequences of *Lactococcus garvieae* TB25, isolated from Italian cheese, and *Lactococcus garvieae* LG9, isolated from Italian rainbow

trout. J Bacteriol 194: 1249–1250. doi:10.1128/JB.06655-11. PubMed: 22328751.
19. Flórez AB, Reimundo P, Delgado S, Fernández E, Alegría A et al. (2012) Genome sequence of *Lactococcus garvieae* IPLA 31405, a bacteriocin-producing, tetracycline-resistant strain isolated from a raw-milk cheese. J Bacteriol 194: 5118–5119. doi:10.1128/JB.00975-12. PubMed: 22933752.
20. Gabrielsen C, Brede Da, Hernández PE, Nes IF, Diep DB (2012) Genome sequence of the bacteriocin-producing strain *Lactococcus garvieae* DCC43. J Bacteriol 194: 6976–6977. doi:10.1128/JB.01864-12. PubMed: 23209230.
21. Ricci G, Ferrario C, Borgo F, Eraclio G, Fortina MG (2013) Genome sequences of two *Lactococcus garvieae* strains isolated from meat. Genome Announc 1: e00018-12. PubMed: 23405320.
22. Knapen D, Vergauwen L, Laukens K, Blust R (2009) Best practices for hybridization design in two-colour microarray analysis. Trends Biotechnol 27: 406–414. doi:10.1016/j.tibtech.2009.03.007. PubMed: 19481825.
23. Kent WJ (2002) BLAT — The BLAST-Like Alignment Tool. Genome Res 12: 656–664. doi:10.1101/gr.229202. Article published online before March 2002 PubMed: 11932250
24. Simon R, Lam A, Li MC, Ngan M, Menenzes S et al. (2007) Analysis of gene expression data using BRB-array tools. Cancer Inform 3: 11–17. PubMed: 19455231.
25. Li JS, Bi YT, Dong C, Yang JF, Liang WD (2011) Transcriptome analysis of adaptive heat shock response of *Streptococcus thermophilus*. PLOS ONE 6: e25777. doi:10.1371/journal.pone.0025777. PubMed: 22022447.
26. Shivaji S, Prakash JSS (2010) How do bacteria sense and respond to low temperature? Arch Microbiol 192: 85–95. doi:10.1007/s00203-009-0539-y. PubMed: 20049417.
27. Wick LM, Egli T (2004) Molecular components of physiological stress responses in *Escherichia coli*. Adv Biochem Eng/Biotechnol 89: 1-45. doi:10.1007/b93957. PubMed: 15217154.
28. Chattopadhyay MK (2006) Mechanism of bacterial adaptation to low temperature. J Biosci 31: 157–165. doi:10.1007/BF02705244. PubMed: 16595884.
29. Phadtare S (2004) Recent developments in bacterial cold-shock response. Curr Issues Mol Biol 6: 125–136. PubMed: 15119823.
30. Phadtare S, Severinov K (2010) RNA remodeling and gene regulation by cold shock proteins. RNA Biol 7: 788–795. doi:10.4161/rna.7.6.13482. PubMed: 21045540.
31. Ray MK (2006) Cold stress response of low temperature adapted bacteria. In: AS SreedharUK Srinivas. Stress Response: A Molecular Biology Approach. Kerala: Research Singpost. pp. 1-23.
32. Giangrossi M, Brandi A, Giuliodori AM, Gualerzi CO, Pon CL (2007) Cold-shock-induced de novo transcription and translation of *infA* and role of IF1 during cold adaptation. Mol Microbiol 64: 807-821. doi:10.1111/j.1365-2958.2007.05699.x. PubMed: 17462025.
33. Giuliodori AM, Brandi A, Giangrossi M, Gualerzi O, Pon CL (2007) Cold-stress-induced de novo expression of *infC* and role of IF3 in cold-shock translational bias. RNA 13: 1355–1365. doi:10.1261/rna.455607. PubMed: 17592046.
34. Datta PP, Wilson DN, Kawazoe M, Swami NK, Kaminishi T et al. (2007) Structural aspects of RbfA action during small ribosomal subunit assembly. Mol Cell 28: 434–445. doi:10.1016/j.molcel.2007.08.026. PubMed: 17996707.
35. Jones PG, Inouye M (1996) RbfA, a 30S ribosomal binding factor, is a cold-shock protein whose absence triggers the cold-shock response. Mol Microbiol 21: 1207-1218. doi:10.1111/j.1365-2958.1996.tb02582.x. PubMed: 8898389.
36. Price CE, Zeyniyev A, Kuipers OP, Kok J (2012) From meadows to milk to mucosa-adaptation of *Streptococcus* and *Lactococcus* species to their nutritional environments. FEMS Microbiol Rev 36: 949–997.
37. Hoffmann T, Bremer E (2011) Protection of *Bacillus subtilis* against cold-stress via compatible solute acquisition. J Bacteriol 193: 1552-1562. doi:10.1128/JB.01319-10. PubMed: 21296969.
38. Angelidis AS, Smith GM (2003) Role of the glycine betaine and carnitine transporters in adaptation of Listeria monocytogenes to chill stress in defined medium. Appl Environ Microbiol 69: 7492-7498. doi:10.1128/AEM.69.12.7492-7498.2003. PubMed: 14660402.
39. Harishchandra RK, Wulff S, Lentzen G, Neuhaus T, Galla HJ (2010) The effect of compatible solute ectoines on the structural organization of lipid monolayer and bilayer membranes. Biophys Chem 150: 37–46. doi:10.1016/j.bpc.2010.02.007. PubMed: 20206435.
40. Ginsburg I (2002) The role of bacteriolysis in the pathophysiology of inflammation, infection and post-infectious sequelae. APMIS 110: 753-770. doi:10.1034/j.1600-0463.2002.1101101.x. PubMed: 12588416.
41. Rice KC, Bayles KW (2003) Death's toolbox: examining the molecular components of bacterial programmed cell death. Mol Microbiol 50: 729–738. doi:10.1046/j.1365-2958.2003.t01-1-03720.x. PubMed: 14617136.
42. Liu Y, Burne RA (2011) The major autolysin of *Streptococcus gordonii* is subject to complex regulation and modulates stress tolerance, biofilm formation, and extracellular-DNA release. J Bacteriol 193: 2826–2837. doi:10.1128/JB.00056-11. PubMed: 21478346.
43. Garcia-Bustos JF, Tomasz A (1987) Teichoic acid-containing muropeptides from *Streptococcus pneumoniae* as substrates for the pneumococcal autolysin. J Bacteriol 169: 447-453. PubMed: 2879828.
44. Fedtke I, Mader D, Kohler T, Moll H, Nicholson G et al. (2007) A Staphylococcus *aureus ypfP* mutant with strongly reduced lipoteichoic acid (LTA) content: LTA governs bacterial surface properties and autolysin activity. Mol Microbiol 65: 1078–1091. doi:10.1111/j.1365-2958.2007.05854.x. PubMed: 17640274.
45. Barrière C, Veiga-da-Cunha M, Pons N, Guédon E, van Hijum SAFT et al. (2005) Fructose utilization in *Lactococcus lactis* as a model for low-GC Gram-positive bacteria: its regulator, signal, and DNA-binding site. J Bacteriol 187: 3752–3761. doi:10.1128/JB.187.11.3752-3761.2005. PubMed: 15901699.
46. Redon E, Loubiere P, Cocaign-Bousquet M (2005) Transcriptome analysis of the progressive adaptation of *Lactococcus lactis* to carbon starvation. J Bacteriol 187: 3589–3592. doi:10.1128/JB.187.10.3589-3592.2005. PubMed: 15866950.
47. Raynaud S, Perrin R, Cocaign-bousquet M, Loubiere P (2005) Metabolic and transcriptomic adaptation of *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis* in response to autoacidification and temperature downshift in skim milk. Appl Environ Microbiol, 71: 8016–8023. doi:10.1128/AEM.71.12.8016-8023.2005. PubMed: 16332781.
48. Dressaire C, Redon E, Milhem H, Besse P, Loubière P et al. (2008) Growth rate regulated genes and their wide involvement in the *Lactococcus lactis* stress responses. BMC Genomics 9: 343. doi:10.1186/1471-2164-9-343. PubMed: 18644113.
49. Claverys JP, Prudhomme M, Martin B (2006) Induction of competence regulons as a general response to stress in gram-positive bacteria. Annu Rev Microbiol 60: 451–475. doi:10.1146/annurev.micro.60.080805.142139. PubMed: 16771651.
50. Dubbs JM, Mongkolsuk S (2007) Peroxiredoxins in bacterial antioxidant defense. Subcell Biochem 44: 143-193. doi:10.1007/978-1-4020-6051-9_7. PubMed: 18084893.
51. den Besten HM, Arvind A, Gaballo HM, Moezelaar R, Zwietering MH et al. (2010) Short- and long-term biomarkers for bacterial robustness: a framework for quantifying correlations between cellular indicators and adaptive behavior. PLOS ONE 5: e13746. doi:10.1371/journal.pone.0013746. PubMed: 21060783.
52. Ayala-Castro C, Saini A, Outten FW (2008) Fe-S cluster assembly pathways in bacteria. Microbiol Mol Biol Rev 72: 110-125. doi:10.1128/MMBR.00034-07. PubMed: 18322036.
53. Crack JC, Green J, Thomson AJ, Le Brun NE (2012) Iron-sulfur cluster sensor-regulators. Curr Opin Chem Biol 16: 35-44. doi:10.1016/j.cbpa.2012.02.009. PubMed: 22387135.
54. Marceau A, Zagorec M (2004) Evidence for involvement of at least six proteins in adaptation of *Lactobacillus sakei* to cold temperatures and addition of NaCl. Appl Environ Microbiol 70: 7260–7268. doi:10.1128/AEM.70.12.7260-7268.2004. PubMed: 15574925.
55. Gaudu P, Vido K, Cesselin B, Kulakauskas S, Tremblay J et al. (2002) Respiration capacity and consequences in *Lactococcus lactis*. Antonie Van Leeuwenhoek 82: 263–269. doi:10.1023/A:1020635600343. PubMed: 12369192.
56. Rezaïki L, Cesselin B, Yamamoto Y, Vido K, van West E et al. (2004) Respiration metabolism reduces oxidative and acid stress to improve long-term survival of *Lactococcus lactis*. Mol Microbiol 53: 1331–1342. doi:10.1111/j.1365-2958.2004.04217.x. PubMed: 15387813.
57. Pedersen MB, Gaudu P, Lechardeur D, Petit M-A, Gruss A (2012) Aerobic respiration metabolism in lactic acid bacteria and uses in biotechnology. Annu Rev Foods Sci Technol 3: 37–58. doi:10.1146/annurev-food-022811-101255.
58. Fernandez A, Lechardeur D, Derré-Bobillot A, Couvé E, Gaudu P et al. (2010) Two coregulated efflux transporters modulate intracellular heme and protoporphyrin IX availability in *Streptococcus agalactiae*. PLOS Pathog 6: e1000860. PubMed: 20421944.
59. Gaudu P, Lamberet G, Poncet S, Gruss A (2003) CcpA regulation of aerobic and respiration growth in *Lactococcus lactis*. Mol Microbiol 50: 183–192. doi:10.1046/j.1365-2958.2003.03700.x. PubMed: 14507373.
60. van de Guchte M, Serror P, Chervaux C, Smokvina T, Ehrlich SD et al. (2002) Stress responses in lactic acid bacteria. Antonie Van Leeuwenhoek 82: 187-216. doi:10.1023/A:1020631532202. PubMed: 12369188.

61. Cotter PD, Hill C (2003) Surviving the acid test: responses of gram-positive bacteria to low pH. Microbiol. Mol Biol 67: 429-453.
62. Gibello A, Collins MD, Domínguez L, Fernández-Garayzábal JF, Richardson PT (1999) Cloning and analysis of the L-lactate utilization genes from *Streptococcus iniae*. Appl Environ Microbiol 65: 4346-4350. PubMed: 10508058.
63. Aguado-Urda M, Gibello A, Blanco MM, López-Campos GH, Cutuli MT et al. (2012) Characterization of Plasmids in a Human Clinical Strain of *Lactococcus garvieae*. PLOS ONE 7: e40119. doi:10.1371/journal.pone.0040119. PubMed: 22768237.
64. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H et al. (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. Nature 459: 950-956. doi:10.1038/nature08080. PubMed: 19448609.
65. Wurtzel O, Yoder-Himes DR, Han K, Dandekar AA, Edelheit S et al. (2012) The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. PLOS Pathog 8: e1002945. PubMed: 23028334.
66. Xue X, Tomasch J, Sztajer H, Wagner-Döbler I (2010) The delta subunit of RNA polymerase, RpoE, is a global modulator of *Streptococcus mutans* environmental adaptation. J Bacteriol 192: 5081-5092. doi:10.1128/JB.00653-10. PubMed: 20675470.
67. Jones AL, Needham RH, Rubens CE (2003) The Delta subunit of RNA polymerase is required for virulence of *Streptococcus agalactiae*. Infect Immun 71: 4011-4017. doi:10.1128/IAI.71.7.4011-4017.2003. PubMed: 12819089.
68. Seepersaud R, Needham RH, Kim CS, Jones AL (2006) Abundance of the delta subunit of RNA polymerase is linked to the virulence of *Streptococcus agalactiae*. J Bacteriol 188: 2096-2105. doi:10.1128/JB.188.6.2096-2105.2006. PubMed: 16513739.
69. Rattanama P, Thompson JR, Kongkerd N, Srinitiwarawong K, Vuddhakul V et al. (2012) Sigma E regulators control hemolytic activity and virulence in a shrimp pathogenic *Vibrio harveyi*. PLOS ONE 7: e32523. doi:10.1371/journal.pone.0032523. PubMed: 22384269.
70. Shah P, Nanduri B, Swiatlo E, Ma Y, Pendarvis K (2011) Polyamine biosynthesis and transport mechanisms are crucial for fitness and pathogenesis of *Streptococcus pneumoniae*. Microbiol 157: 504–515. doi:10.1099/mic.0.042564-0. PubMed: 20966092.
71. Ware D, Jiang Y, Lin W, Swiatlo E (2006) Involvement of potD in *Streptococcus pneumoniae* polyamine transport and pathogenesis. Infect Immun 74: 352-361. doi:10.1128/IAI.74.1.352-361.2006. PubMed: 16368990.
72. Que Q, Helmann JD (2000) Manganese homeostasis in *Bacillus subtilis* is regulated by MntR, a bifunctional regulator related to the diphtheria toxin repressor family of proteins. Mol Microbiol 35: 1454-1468. PubMed: 10760146.
73. Papp-Wallace KM, Maguire ME (2000) Manganese transport and the role of manganese in virulence. Annu Rev Microbiol 60: 187-209. PubMed: 16704341.
74. Boyer E, Bergevin I, Malo D, Gros P, Cellier MFM (2002) Acquisition of Mn(II) in addition to Fe(II) is required for full virulence of *Salmonella enterica* Serovar Typhimurium. Infect Immun 70: 6032–6042. doi: 10.1128/IAI.70.11.6032-6042.2002. PubMed: 12379679.
75. Sabri M, Caza M, Proulx J, Lymberopoulos MH, Brée A et al. (2008) Contribution of the SitABCD, MntH, and FeoB metal transporters to the virulence of avian pathogenic Escherichia coli O78 strain 7122. Infect Immun 76: 601–611.
76. Anderson ES, Paulley JT, Gaines JM, Valderas MW, Martin DW et al. (2009) The manganese transporter MntH is a critical virulence determinant for *Brucella abortus* 2308 in experimentally infected mice. Infect Immun 77: 3466–3474. doi:10.1128/IAI.00444-09. PubMed: 19487482.
77. Perry RD, Craig SK, Abney J, Bobrov AG, Kirillina O et al. (2012) Manganese transporters Yfe and MntH are Fur-regulated and important for the virulence of *Yersinia pestis*. Microbiol 158: 804-815. doi: 10.1099/mic.0.053710-0. PubMed: 22222497.

El material adicional asociado a esta publicación se encuentra disponible a través de las siguientes URLs:

- ✓ Tabla Suplementaria 1 (*Table S1*): http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0079692.s001
- ✓ Tabla Suplementaria 2 (*Table S2*): http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0079692.s002
- ✓ Tabla Suplementaria 3 (*Table S3*): http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0079692.s003
- ✓ Tabla Suplementaria 4 (*Table S4*): http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0079692.s004

CAPÍTULO VII:

Discusión

A pesar de la importancia de *Lactococcus garvieae* como microorganismo patógeno en medicina veterinaria y humana, además de su posible carácter zoonótico, cuando se inició esta Tesis Doctoral la información disponible sobre el genoma de esta bacteria era muy escasa. Por ello, se planteó el estudio del contenido genético de este microorganismo como principal objetivo de esta Tesis, el cual se ha abordado a través de las técnicas disponibles en cada momento. La elucidación del contenido genético de *L. garvieae* y el análisis funcional del mismo tienen como fin aportar conocimientos sobre los mecanismos evolutivos, adaptativos y de patogenicidad de esta bacteria de interés creciente, lo que puede abrir vías hacia el desarrollo de estrategias para su control.

1. Análisis del genoma de *L. garvieae* mediante Hibridación Genómica Comparativa (CGH)

Cuando iniciamos nuestro trabajo sobre el estudio del contenido genético de *L. garvieae*, existía en las bases de datos un número limitado de entradas para la búsqueda de secuencias conocidas de *L. garvieae*, perteneciendo la mayoría de ellas a secuencias de ARN ribosómico 16S. Así, el conocimiento sobre el genoma de este importante patógeno se limitaba a un escaso número de genes secuenciados. En ese mismo año, el uso de las tecnologías de secuenciación genómica de última generación comenzaba a extenderse, aunque aún resultaban técnicas relativamente inaccesibles por su complejidad y su alto precio. En el caso de los microorganismos, se secuenciaban únicamente los genomas de aquellos de gran interés clínico e industrial, incluyendo algunos pertenecientes a la familia *Streptococcaceae*. Una vez conocido el genoma de un organismo, es posible diseñar un microarray que represente el genoma completo del mismo, el cual puede utilizarse con distintos fines. Una de las aproximaciones en el uso de los microarrays en Microbiología es la comparación de un genoma microbiano problema y otro secuenciado mediante la técnica denominada Hibridación Genómica Comparativa (CGH: *Comparative Genomic Hybridization*). Estas comparaciones se pueden realizar entre cepas de la misma especie, o bien entre microorganismos filogenéticamente cercanos (Aguado y Blanco, 2007; Guzmán *et al*., 2008). Varios estudios han demostrado que la técnica de CGH sobre microarrays de ADN constituye una buena aproximación para la búsqueda de nuevos genes en una especie cuyo genoma se desconoce. En este caso la técnica consiste en realizar una CGH inter-especie del ADN genómico del microorganismo problema sobre un microarray en el que se representan todos los marcos abiertos de lectura (ORFs: *Open Reading Frames*) del genoma de un microorganismo secuenciado de distinta especie pero filogenéticamente próximo

(Dong *et al.,* 2001, Hakenbeck *et al*., 2001; Fukiya *et al*., 2004). Mediante esta aproximación es posible determinar los genes comunes a ambos microorganismos (referencia y problema), permitiendo la identificación de un gran número de genes en el microorganismo problema sin la necesidad de secuenciar su genoma. Además, esta técnica puede proporcionar valiosa información sobre la diversidad y la evolución de los organismos y se ha aplicado con éxito al estudio de microorganismos patógenos (Aguado y Blanco, 2007). La principal limitación de esta aproximación es que los genes únicos del microorganismo estudiado permanecen sin identificar (Dong *et al*., 2001; Fukiya *et al*., 2004). De esta manera, se abordó el estudio del contenido genómico de *L. garvieae* mediante CGH sobre microarrays comerciales de ORFs de los microorganismos filogenéticamente próximos *Lactococcus lactis* subsp. *lactis* y *Streptococcus pneunoniae* (Ver Capítulo II).

Cuando se realizan experimentos de CGH inter-especie es necesario establecer un marco de trabajo que permita una interpretación precisa de los resultados. Aunque en experimentos previos de CGH inter-especie se ha utilizado la intensidad de señal como indicador del nivel de similitud entre las secuencias de genes homólogos (Dong *et al*., 2001), este tipo interpretación puede verse influenciada por factores experimentales, como efectos de marcaje diferencial, accesibilidad de las sondas o problemas de hibridación en algunos puntos del microarray (van Hijum *et al*., 2008). Por estas razones no se utilizó este criterio en nuestro estudio de CGH inter-especie, sino que se estableció un marco de trabajo definido mediante la combinación de experimentos de CGH *in vitro* y el análisis de secuencias *in silico*. Este abordaje permitió establecer una relación entre los resultados obtenidos mediante los experimentos de CGH y la similitud de las secuencias de ADN detectadas, conformando el citado marco de trabajo que permitió una interpretación precisa de los resultados de CGH inter-especie para *L. garvieae*.

Tanto los estudios *in silico* como los estudios *in vitro* realizados con los organismos de referencia (*L. lactis* subsp. *lactis* IL1403 y *S. pneumoniae* TIGR4) se basaron en comparaciones del genoma de uno de los microorganismos de referencia con las secuencias de las sondas inmovilizadas en microarray de OFRs del otro microorganismo, y viceversa. Los resultados establecieron una correlación entre los resultados de CGH y la similitud de las secuencias detectadas, permitiendo evaluar el rendimiento de la CGH inter-especie bajo las condiciones experimentales fijadas. Así, se comprobó que el 97% de los genes comunes a los dos microorganismos de referencia detectados mediante CGH inter-especie tenían una similitud de secuencia del 70% o mayor (Capítulo II, Tabla 1); esto permitió establecer que en las condiciones experimentales fijadas para la CGH inter-especie, resultaba posible

identificar la presencia de genes homólogos en un microorganismo problema que presentasen un 70% o más de similitud de secuencia. Una vez establecido este umbral de detección, se procedió a la interpretación de los resultados de CGH de *L. garvieae* CECT4531 sobre microarrays de los dos microorganismos de referencia.

En el marco de trabajo establecido para la CGH se lograron identificar 267 genes en *L. garvieae* CECT4531 que presentaban, al menos, un 70% de similitud de secuencia con sus homólogos en los microorganismos de referencia (Capítulo II, Material Adicional). Si se hubieran aplicado unas condiciones experimentales de hibridación menos restrictivas, quizá habría sido posible la identificación de un mayor número de genes, pero con menor especificidad en la detección. Debido a que el principal objetivo de este estudio era la identificación de genes presentes en *L. garvieae*, se optó por mantener unas condiciones de hibridación restrictivas que permitieran una elevada especificidad y fiabilidad de los resultados.

El estudio de los 267 genes identificados en *L. garvieae* reflejó que éstos intervenían en funciones biológicas muy diversas (Capítulo II, Tabla 2). Algunos de los genes identificados resultaron de especial interés ya que podrían estar implicados en la patogénesis de las infecciones causadas por *L. garvieae*. Por ejemplo, el gen *als*, que codifica una acetolactato sintasa, ha sido descrito como necesario para la supervivencia de *L. garvieae* durante la infección en pez (Menéndez *et al*., 2007) y como un factor implicado en la colonización de los biotipos El Tor de *Vibrio cholerae* (Yoon *et al*., 2006). Por otra parte, el gen *mycA*, que codifica un antígeno de reacción cruzada con miosina, ha sido relacionado con la virulencia de estreptococos patógenos en humanos (Volkov *et al*., 2010).

Este estudio supuso el primer análisis del contenido genómico global de *L. garvieae*, ampliando el conocimiento que hasta ese momento se tenía sobre este patógeno. Asimismo, los resultados obtenidos confirmaron que la CGH inter-especie sobre microarrays de ADN es una buena aproximación para el estudio del contenido genómico de bacterias de relevancia que no han sido secuenciadas completamente. Esto también se refleja en otros trabajos publicados recientemente dirigidos tanto al análisis del contenido genómico, como en al estudio de los transcriptomas de microorganismos de los que aún no se conoce su genoma (Van Hijum *et al*., 2008; Park *et al*., 2012).

2. Secuenciación del genoma de *L. garvieae* y estudios de la variabilidad intra-especie

A principios de esta década la tecnología de secuenciación genómica alcanzó un gran nivel de desarrollo, lo que se tradujo en una reducción de los costes y por tanto, en un incremento en la accesibilidad de la misma por parte de los investigadores. La secuenciación de un genoma bacteriano abre las vías hacia la exploración de la biología bacteriana, y el conocimiento de los mecanismos de patogenicidad de las bacterias de interés clínico. Además, la secuenciación genómica de varias cepas de una misma especie permite establecer relaciones evolutivas entre ellas. La elevada diversidad observada a menudo entre genomas de la misma especie sugiere la conveniencia de secuenciar diferentes aislados con el fin de obtener una visión más exacta del contenido genómico de dicha especie (Guzmán *et al*., 2008). Por estas razones se decidió secuenciar dos cepas de *L. garvieae* con interés clínico: un aislado de trucha representativo de la cepa epidémica causante de los brotes de lactococosis en España (Lg8831) (Vela *et al*., 2000), y un aislado de un caso de septicemia humana (Lg21881) (Aspiroz *et al*., 2007) (Ver Capítulo IV). Para ello, se utilizó la pirosecuenciación 454 de Roche, tecnología que se aplica con éxito en la secuenciación de genomas bacterianos a un coste asequible.

Uno de los problemas en la secuenciación *de novo* de genomas pequeños es que las lecturas cortas generadas mediante la estrategia de secuenciación genómica *Whole Genome Shotgun Sequencing* (WGS), pueden conducir a un número más alto de regiones en las que no se alinean las lecturas (*gaps*), causando una fragmentación mayor y una menor calidad de los fragmentos alineados o *contigs*, lo que ocurre mayoritariamente en las regiones del genoma que contienen elementos repetitivos. Para solucionar este inconveniente, algunas plataformas de secuenciación de última generación ofrecen el sistema *paired-ends,* mediante el cual ambos finales de un fragmento de ADN son ordenados, a diferencia de la secuenciación de una sola lectura, en la que sólo se ordena uno de los finales. Las lecturas apareadas que se obtienen mediante este sistema permiten conocer la orientación y la posición relativa de los *contigs*, haciendo posible completar *gaps* en la secuencia consenso y permitiendo ordenar los *contigs* en fragmentos de mayor longitud o *scaffolds* (Nagarajan y Pop, 2013). Así, se aplicó la estrategia de secuenciación genómica mediante WGS unida a la metodología de *paired-ends* para la secuenciación de la cepa Lg21881, y únicamente WGS para la secuenciación de la cepa Lg8831. De esta forma, la cepa Lg21881 serviría como referencia para el ensamblado del genoma de la cepa Lg8831.

Tras la secuenciación, el ensamblado y la anotación automática de las secuencias de los genomas de Lg21881 y Lg8831 mediante herramientas bioinformáticas, se obtuvieron los mapas genómicos de ambas cepas, que fueron depositados en la base de datos genómicos del NCBI (Lg21881: AFCC00000000; Lg8831: AFCD00000000). Además, se analizaron ambos genomas mediante la herramienta RAST (*Rapid Annotation using Subsystem Technology*) (Aziz *et al*., 2008) utilizando la aplicación específica para procariotas, lo que permitió una clasificación de los genes identificados por grupos funcionales. Los resultados de este análisis se muestran en la Figura 7.

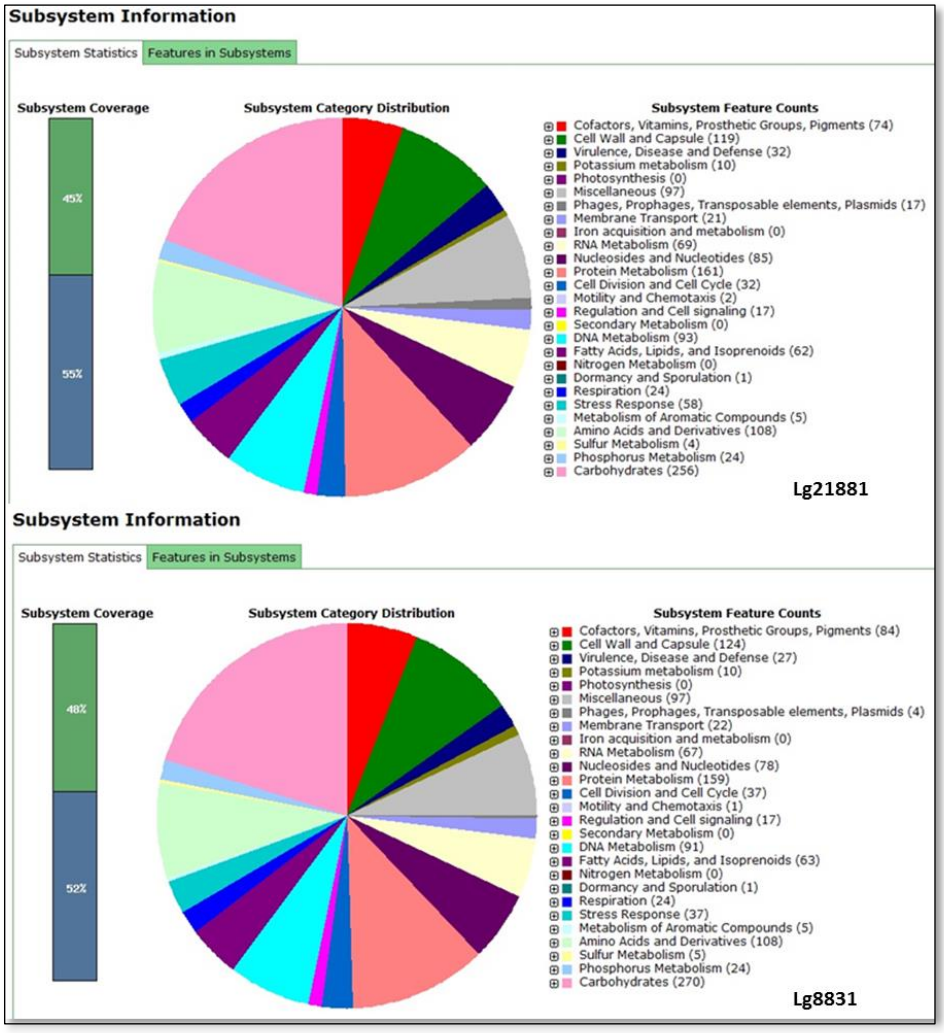


Figura 7. Resultados del análisis funcional mediante RAST para las cepas Lg21881 (superior) y Lg8831 (inferior).

Es importante señalar que además de la liberación del genoma de las cepas Lg8831 y Lg21881, hasta el momento se han liberado los genomas de otras nueve cepas de *L. garvieae* (Tabla 4; http://www.ncbi.nlm.nih.gov/genome/genomes/699), lo que pone de manifiesto el interés creciente en el estudio de este patógeno.

Tabla 4. Genomas de *L. garvieae* secuenciados hasta el momento.

Cepa	Procedencia del aislado	Publicación asociada	Tamaño (Mb)	% G+C	Nivel de ensamblado	Número de acceso (NCBI)
Lg21881	Septicemia humana	Aguado-Urda *et al*. (2011a)	2,16	37,9	91 *contigs*	AFCC00000000
Lg8831	Lactococosis trucha arcoíris (España)	Aguado-Urda *et al.* (2011b)	2,09	38	87 *contigs*	AFCD00000000
UNIUD074	Lactococosis trucha arcoíris (Italia)	Reimundo *et al.* (2011b)	2,17	38,7	25 *contigs*	AFHF01000000
ATCC49156	Lactococosis pez limón (Japón)	Morita *et al.* (2011)	1,95	38,8	Cromosoma completo	AP009332
Lg2	Lactococosis pez limón (Japón)	Morita *et al,* (2011)	1,96	38,8	Cromosoma completo	AP009333
DCC43	Intestino de pato	Gabrielsen *et al*. (2012)	2,24	37,7	69 *contigs*	AMQS01000000
IPLA31405	Queso (España)	Flórez *et al*. (2012)	2,05	38,5	23 *contigs*	AKFO01000000
LG9	Lactococosis trucha arcoíris (Italia)	Ricci *et al.* (2012)	2,09	38,5	140 *contigs*	AGQY01000000
TB25	Queso (Italia)	Ricci *et al*. (2012)	2,01	38,1	92 contigs	AGQX01000000
Tac2	Carne de pavo	Ricci *et al*. (2013)	2,24	38,2	97 *contigs*	AMFE01000000
I113	Carne de cerdo	Ricci *et al*., (2013)	2,18	37,9	49 *contigs*	AMFD01000000

La disponibilidad de estas secuencias genómicas de *L. garvieae* posibilitó la realización de análisis orientados a la validación de los resultados obtenidos en los estudios previos sobre el contenido genético de esta bacteria mediante CGH. Es importante apuntar que los experimentos de CGH fueron realizados con la cepa *L. garvieae* CECT4531, de la que no se conoce su genoma por el momento, y que puede presentar variaciones en su

contenido con respecto a las cepas secuenciadas. Para esta validación se utilizó la cepa *L. garvieae* ATCC49156, cuyo genoma público se encuentra completo y cerrado en la base de datos del NCBI (Morita *et al*., 2011). Este análisis, realizado mediante la herramienta de comparación de secuencias BLAT (Kent, 2002), reveló que el 82% (n=216) de los genes de *L. garvieae* CECT4531 identificados en los experimentos de CGH estaban también presentes en el genoma de la cepa ATCC49156. El 18% restante de los genes (n=53) no se identificaron en la cepa ATCC49156, lo que podría deberse a que dichos genes fueran específicos de *L. garvieae* CECT4531. Para comprobar esta hipótesis se realizaron análisis complementarios de secuencias mediante BLAT, en los que se incluyeron todas las secuencias genómicas de *L. garvieae* disponibles en las bases de datos del NCBI. Así, se consiguió elevar el número de genes validados (presentes en el genoma de una o varias de las cepas secuenciadas) hasta el 94,4% (n=252). Los genes restantes (n=15) que no pudieron ser validados automáticamente en las comparaciones mediante herramienta BLAT, fueron analizados uno a uno manualmente. El análisis individual de estos genes se realizó mediante la herramienta BLAST, obteniéndose resultados positivos para 13 de los 15 genes, que fueron detectados en alguna de las secuencias genómicas de *L. garvieae.* Estos resultados validan los resultados iniciales acerca del contenido genómico de *L. garvieae* obtenidos mediante CGH y ponen de manifiesto asimismo la variabilidad del contenido genómico que existe entre diferentes cepas de esta bacteria.

El paso siguiente en el análisis de los genomas de las cepas Lg21881 y Lg8831 consistió en estudios preliminares de genómica comparativa. La genómica comparativa se refiere al estudio de las relaciones entre genomas de diferentes especies o cepas de la misma especie, e implica el uso de herramientas bioinformáticas o procesos experimentales que identifiquen zonas de similitud entre los genomas, revelando las características únicas de cada genoma. Cuando se analizan genomas bacterianos cercanos, como en este caso, este análisis puede mejorar la anotación funcional de los mismos, detectar la transmisión de transposones, así como de genes de resistencia a antibióticos y elementos extracromosómicos, y aportar información sobre la adaptación de cada cepa a su nicho ecológico (Guzmán *et al.,* 2008). La comparación preliminar de las secuencias genómicas de Lg21881 y Lg8831 se realizó mediante la herramienta bioinformática MAUVE (Darling *et al*., 2004) usando el algoritmo *progressiveMauve* (Darling *et al*., 2010) y los resultados se muestran en la Figura 8. La diferencia más notable que se encontró fue el mayor tamaño del genoma de la cepa Lg21881 (aproximadamente 0,1 Mb), comprobándose que esta diferencia se debía a la presencia de 5 plásmidos en Lg21881 (Capítulo V, Figura 1). Las diferencias entre genomas de la misma especie normalmente radican en la adquisición de grandes islas genómicas y/o material extracromosómico, que ocurren principalmente a través de fenómenos de transferencia genética horizontal (Mira *et al*., 2010). Estas

variaciones ocurren en el genoma "accesorio", es decir, en aquellos genes que no pertenecen al genoma "núcleo" (parte compartida por todos los miembros de la misma especie) y que presumiblemente no son esenciales para las funciones celulares básicas, pero que aportan a cada cepa características determinadas, como por ejemplo, resistencia a antibióticos o factores de virulencia. Por ello, el estudio del genoma accesorio microbiano puede ayudar a elucidar los mecanismos potencialmente implicados en las diferencias fenotípicas que existen entre cepas de la misma especie bacteriana. Estos estudios se han visto facilitados considerablemente por la secuenciación genómica de última generación (Laing *et al*., 2011).

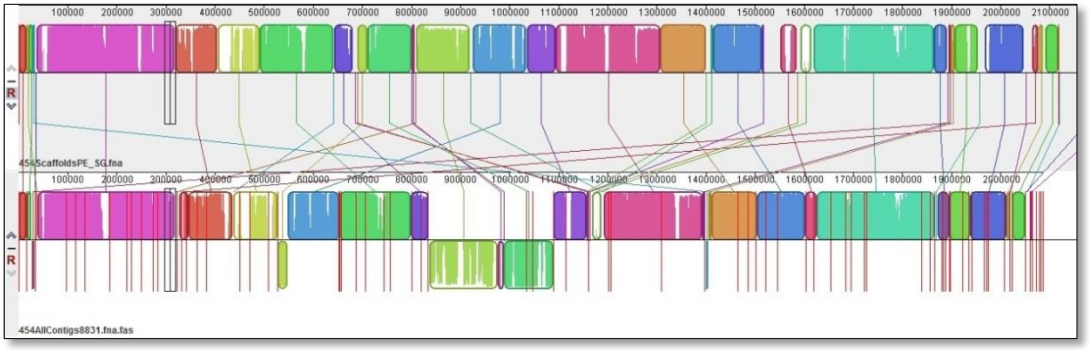


Figura 8. Alineamiento mediante MAUVE de los genomas de Lg21881 (superior) y Lg8831 (inferior).

Por este motivo se decidió profundizar en el estudio de los cinco plásmidos denominados pGL1-pGL5 presentes en la cepa de origen clínico humano Lg21881 (Ver Capítulo V). Las bacterias ácido-lácticas normalmente poseen plásmidos que les aportan diversas funciones adicionales como resistencia a bacteriófagos, producción de bacteriocinas, o resistencia a antibióticos entre otras (Mills *et al*., 2006). A diferencia de otros lactococos como *L. lactis*, bacteria de las que se ha estudiado y caracterizado el contenido de sus plásmidos (Siezen *et al*., 2005; Górecki *et al*., 2011), la información sobre el contenido plasmídico de *L. garvieae* resultaba muy limitada antes de la secuenciación de los genomas de varias cepas.

De acuerdo con sus características, pGL3 y pGL4 son plásmidos no movilizables. Por el contrario, el análisis de las secuencias de pGL1, pGL2 y pGL5, reveló que son plásmidos movilizables, característica confirmada por la presencia de estos plásmidos en otras cepas de *L. garvieae.* Gran parte de los genes identificados en los plásmidos pGL1-pGL5

presentaron elevada similitud con homólogos en otros géneros de bacterias acido-lácticas como *Leuconostoc*, *Lactobacillus*, *Enterococcus, Streptococcus* y *Weissella.* La presencia en estos plásmidos de elementos móviles y genes relacionados, como transposasas, secuencias de inserción, resolvasas, integrasas y relaxasas sugirieron que algunos de estos ORFs podrían haber sido adquiridos mediante eventos de transferencia genética horizontal (HGT: *Horizontal Gene Transfer*), hecho que quedó confirmado tras los análisis de HGT (Capítulo V, Tabla 1). El hecho de que hayan existido eventos de HGT entre Lg21881 y bacterias ácido-lácticas que forman parte del grupo de microorganismos que típicamente participan en las fermentaciones lácticas, sugiere la posibilidad de que Lg21881 podría proceder de un nicho ecológico alimentario. Esta hipótesis estaría en la misma línea de aquellas que proponen que la fuente de infección por *L. garvieae* en humanos podría ser el consumo de alimentos contaminados (Wang *et al*., 2007; Wilbring *et al*., 2011; Russo *et al*., 2012). En este sentido, es destacable el hecho de que las secuencias completas de pGL1 y pGL5 se encuentran presentes en el genoma la cepa de *L. garvieae* Tac2 aislada de carne de pavo, y que la secuencia de pGL2 está presente en el genoma de la cepa de *L. garvieae* I113 aislada de carne de cerdo (Ricci *et al*., 2013). La posibilidad de que *L. garvieae* sea considerado un patógeno alimentario abriría nuevas líneas de investigación enfocadas al esclarecimiento de aspectos de enorme interés desde el punto de vista de seguridad alimentaria, como el riesgo que podría suponer la presencia en alimentos de cepas de *L. garvieae* potencialmente patógenas para el hombre.

Otras características relevantes encontradas en el contenido plasmídico de Lg21881 son los sistemas de producción de bacteriocinas, un mecanismo de defensa ante bacterias competidoras. Así, se determinó que pGL1 y pGL2 contenían genes que codifican proteínas similares a bacteriocinas (Capítulo V, Tabla 2). En pGL5, además de un gen que codifica una potencial bacteriocina, se identificaron los genes necesarios para su producción, secreción e inmunidad (Capítulo V, Tabla S5). En estudios posteriores se logró purificar y caracterizar esta bacteriocina, denominada garvicina A, demostrándose que posee actividad específica frente a cepas clínicas de *L. garvieae* aisladas de peces y de ganado bovino, entre otras (Maldonado-Barragán *et al*., 2013). La potente inhibición ejercida por la garvicina A sobre cepas patógenas de *L. garvieae* podría ser de gran utilidad en el tratamiento de las infecciones producidas por este patógeno en el ámbito veterinario.

Además, en pGL5 se detectaron algunos genes potencialmente implicados en la virulencia de esta cepa, como el gen *txn* que codifica una proteína de la familia de las toxinas actina-ADP-ribosiltransferasas, así como dos genes que codifican sendas proteínas con dominios LPxTG de unión a mucina y colágeno respectivamente (Capítulo V, Tabla 2). Las proteínas con dominios LPxTG están implicadas en la adhesión al mucus intestinal y en

la internalización en las células del hospedador (Linden *et al*., 2008; Hendrickx *et al.,* 2009), y resultan importantes para la infección de patógenos como *Listeria monocytogenes,* que es capaz de atravesar la barrera intestinal, produciendo infecciones de tipo sistémico. Por lo tanto, las dos proteínas con dominios LPxTG identificadas podrían estar implicadas en la adherencia y colonización por parte de *L. garvieae* en las células del hospedador. En este sentido, se ha sugerido que la mayor tasa de adhesión e internalización de la cepa Lg21881 con respecto a la cepa Lg8831 puede estar relacionada con las proteínas LPxTG únicas de la cepa Lg21881 (Aguado-Urda *et al*., 2012).

Los estudios de caracterización genética en *L. garvieae* previos a la secuenciación de su genoma mostraron que existía una elevada heterogeneidad genética intra-especie que en algunos casos estaba relacionada con el hospedador y/o la zona geográfica de origen (Vela *et al*., 2000; Ravelo *et al*., 2003; Kawanishi *et al*., 2006; Reimundo Díaz-Fierros, 2011). Esta elevada variabilidad genética se ha puesto de manifiesto en los estudios de genómica comparativa realizados tras la secuenciación de los genomas de varias cepas de *L. garvieae* (Tabla 4). El primer estudio de genómica comparativa publicado sobre las cinco cepas de *L. garvieae* que habían sido secuenciadas a finales de 2011, puso de manifiesto que existía un genoma núcleo compartido entre todas las cepas, y genes auxiliares únicos de cada una, o comunes a varias de las cepas analizadas. Además, se identificaron algunos genes potencialmente implicados en la patogénesis de la infección por *L. garvieae*, entre los cuales se describieron posibles factores de virulencia únicos de cada cepa que podrían relacionarse con la especificidad de hospedador (Miyauchi *et al*., 2012). Así, en este estudio se describieron 18 genes que únicamente estaban presentes en las cepas aisladas de casos clínicos humanos Lg21881 y LgHF. Con anterioridad, Reimundo *et al*. (2011a) habían comprobado que estos mismos genes se encontraban en la cepa LgHF pero no estaban presentes en una cepa clínica de pez. Estos resultados sugerían que algunos de estos 18 genes presentes en las cepas de origen clínico humano podrían estar implicados en la patogénesis de las infecciones por *L. garvieae* en el hombre (Miyauchi *et al*., 2012), y ponen de manifiesto la importancia de caracterizar aquellos elementos genómicos únicos de determinadas cepas de un microorganismo.

Los estudios de genómica comparativa permiten también establecer relaciones filogenéticas entre diferentes cepas de una especie bacteriana, así como clarificar aspectos sobre los posibles mecanismos evolutivos a los que han sido sometidos los microorganismos (Laing *et al*., 2011). Aunque los estudios de comparación de las once cepas de *L. garvieae* secuenciadas hasta la actualidad no están contemplados como uno de los objetivos definidos en la presente Tesis Doctoral, se decidió realizar una comparación genómica preliminar que permitiera establecer relaciones filogenéticas entre los genomas

de *L. garvieae* disponibles mediante una aproximación *in silico* empleando herramientas bioinformáticas basadas en lo que se conoce como *Genome Blast Distance Phylogeny* (GBDP) (Henz *et al*., 2005). Los resultados de este análisis identificaron dos líneas genéticas diferenciadas en *L. garvieae* (Figura 9). Estos resultados son coincidentes con los de los últimos estudios comparativos publicados, que ya han apuntado la idea de que existen en *L. garvieae*, líneas genómicas independientes que derivarían de un ancestro común, y que se separaron debido a estímulos ambientales derivados de nichos ecológicos específicos (Ferrario *et al*., 2012). Además, se observó que la cepa DCC43 está muy alejada filogenéticamente del resto de las cepas analizadas, lo que sugiere que podría tratarse de una especie diferente (Figura 9). Estos resultados son coincidentes con las conclusiones de los estudios más recientes acerca de las relaciones evolutivas entre aislados de *L. garvieae* (Ferrario *et al*. 2013).

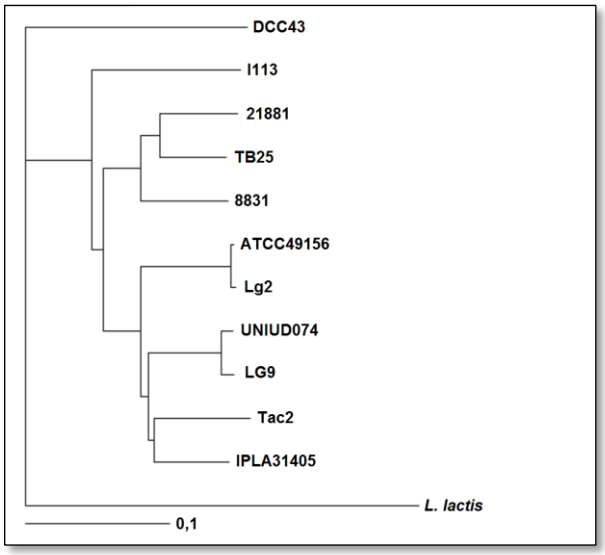


Figura 9. Árbol filogenético de las cepas secuenciadas de *L. garvieae* generado por el método CVTree (Qi *et al*., 2004; Xu y Hao, 2009) en función de los resultados de los análisis basados en GBDP. Se ha incluido a *L. lactis* como *outgroup*.

Durante los últimos años se ha observado un interés creciente por *L. garvieae* en el ámbito de la industria alimentaria, debido a su frecuente aislamiento a partir de diferentes tipos de alimentos (Santos *et al*., 2005; Kawanishi *et al*., 2007; Ricci *et al*., 2013). Esto es especialmente relevante en la industria láctea en particular, por su habitual presencia en diferentes productos lácteos (Foschino *et al*., 2008; Jokovic *et al*., 2008; Alegría *et al*., 2009). Como ya se ha mencionado, algunos autores han apuntado a la ingestión de alimentos

contaminados con *L. garvieae* como una posible ruta de infección en el hombre (Wang *et al*., 2007; Wilbring *et al*., 2011). En este contexto, resultaría de gran utilidad poder diferenciar entre los aislados clínicos y los de origen alimentario de *L. garvieae* (Foschino *et al*., 2008; Fortina *et al.,* 2009). Así, se propuso emplear la capacidad de utilizar lactosa como fuente de carbono (Lac+) y la presencia del gen de la fosfo-beta-galactosidasa (*lacG*), como marcadores para diferenciar los aislados de *L. garvieae* procedentes de productos lácteos y los aislados clínicos de lactococosis en peces (Fortina *et al*., 2009). Sin embargo, estos autores no analizaron aislados de *L. garvieae* de otros orígenes a pesar de la conocida ubicuidad de este micoorganismo. Por ello, se decidió analizar la utilidad de dichos marcadores en un conjunto de aislados de *L. garvieae* procedentes de un amplio rango de orígenes (Ver Capítulo III).

La mayoría de los aislados analizados procedentes de muestras clínicas (91,9%) resultaron ser Lac-/*lacG*-, independientemente de su origen, porcentaje significativamente mayor que el de los aislados no clínicos. Todos los aislados procedentes de truchas y cerdos, así como la mayoría de los de leche procedente de vacas con mastitis subclínica, y de los aislados de humanos, fueron incapaces de metabolizar la lactosa (Lac-/*lacG*-) (Capítulo III, Tabla 1). Además, en el estudio de los genomas secuenciados de Lg21881 y Lg8831 se comprobó la ausencia del gen *lacG*, validando los resultados obtenidos para estas cepas en este estudio. Estos resultados podrían sugerir una relación entre la incapacidad de metabolizar lactosa y el origen clínico de las muestras. Sin embargo, esta hipótesis se ve limitada por la existencia de aislados no clínicos de *L. garvieae* Lac-/*lacG*- (procedentes de morcilla de Burgos y de aves), así como la existencia de aislados clínicos de peces Lac+/*lacG*+ (Ravelo *et al.,* 2001; Pereira *et al.,* 2004). Por otra parte, aunque la mayoría de los aislados procedentes de productos lácteos (77,8%) resultaron capaces de utilizar la lactosa (Lac+/*lacG*+), algunos aislados lácteos resultaron Lac-/*lacG*- (Capítulo III, Tabla 1). Igualmente, entre los *L. garvieae* procedentes de aguas el número de aislados positivos y negativos fue comparable para ambos caracteres, resultado que coincide con estudios previos que han descrito aislados tanto positivos como negativos para la utilización de la lactosa procedentes de muestras ambientales (Klijn *et al.*, 1995).

En definitiva, de acuerdo con los resultados de este estudio, la capacidad o no de metabolizar la lactosa y la presencia o ausencia del gen *lacG* no son exclusivas de los aislados de *L. garvieae* procedentes de productos lácteos o de peces respectivamente. Por tanto, el uso de estos criterios como biomarcadores del origen de los aislados de *L. garvieae* debería ser reconsiderado.

3. Análisis funcional del genoma de *L. garvieae*

La disponibilidad de la información del contenido genómico de un microorganismo no puede resolver por sí misma preguntas sobre las características funcionales de su genoma. Por este motivo, una vez obtenida la secuencia del genoma de *L. garvieae*, decidimos realizar el análisis de expresión global en este patógeno mediante estudios transcriptómicos utilizando microarrays de ADN que representaban su genoma completo. Este tipo de estudios han sido extensamente utilizados para analizar los perfiles de expresión de diferentes microorganismos en un amplio rango de condiciones experimentales, y aportan una visión global y detallada de aspectos funcionales de la fisiología microbiana (Toledo-Arana *et al*., 2009; Kocabas *et al*., 2009).

La variación de factores ambientales, como la temperatura, provoca cambios adaptativos que se reflejan en el transcriptoma bacteriano (Stintzi, 2003; Mereghetti *et al.,* 2008). *L. garvieae* es una bacteria ubicua, capaz de colonizar diversos nichos ecológicos y de causar infección en un amplio rango de hospedadores; por lo tanto debe ser capaz de adaptarse a fluctuaciones ambientales. Además, la temperatura del agua ha se ha descrito como uno de los factores ambientales más influyentes en el desarrollo de la lactococosis en peces (Vendrell *et al*. 2006). Por todo ello, se diseñó un microarray de ADN a partir de los datos previamente obtenidos en la secuenciación genómica de las cepas Lg21881 y Lg8831, que permitió analizar el efecto de la temperatura en la expresión global del genoma en ambas cepas (Ver Capítulo VI). Así, se estudiaron los transcriptomas de las cepas Lg8831 y Lg21881 cultivadas a 18°C, temperatura a la usualmente que se producen los brotes agudos de lactococosis, y 37°C, temperatura fisiológica humana.

El diseño del microarray que contenía las secuencias genómicas completas de Lg21881 y Lg8831 fue validado, mostrando una excelente correlación en sus resultados con los de experimentos de PCR cuantitativa (Capítulo VI, Figura 3), y representa el primer microarray de ADN disponible para la realización de estudios de genómica funcional global en *L. garvieae*.

Los resultados del primer análisis funcional de los transcriptomas de las cepas Lg21881 y Lg8831 obtenidos a diferentes temperaturas de crecimiento mostraron que las dos cepas estudiadas responden de manera global y específica a la temperatura (Capítulo VI, Figura 1). En ambos casos, el número de genes expresados diferencialmente en función

de la temperatura se aproximó al 10% del total de genes presentes en cada genoma, resultado concordante con estudios similares en otras bacterias como *Streptococcus agalactiae* (Mereghetti *et al.,* 2008) o *Streptococcus thermophilus* (Li *et al*., 2011). Además, al comparar los resultados de ambas cepas se observó una respuesta transcripcional específica en función de la cepa, que resultó ser más evidente a 18°C, como mostró el análisis por categorías funcionales (Capítulo VI, Figura 2). Entre los genes sobreexpresados a 18°C, las categorías funcionales J (traducción, estructura y biogénesis ribosómicas) y K (transcripción) fueron las más representadas en Lg8831, mientras que las categorías C (producción y conversión de energía) y H (transporte y metabolismo de coenzimas) resultaron las mayoritarias en Lg21881. El análisis funcional de los genes sobreexpresados a 18°C por la cepa Lg8831, mostró que esta cepa es capaz de adaptarse de manera eficaz a temperaturas por debajo de la óptima mediante la expresión de genes implicados en una respuesta específica a estrés por frío (Capítulo V, Figura 4). Esta remodelación específica y coordinada de su transcriptoma a 18°C podría estar relacionada además con la fase de latencia más corta que presenta esta cepa durante su crecimiento a esta temperatura. Estos datos sugieren que la cepa clínica de pez podría estar mejor adaptada al crecimiento a temperaturas por debajo de la óptima. En el caso de Lg21881, a 18°C mostró la sobreexpresión de genes relacionados con una repuesta inespecífica a estrés. Además, a esta temperatura se detectó la sobreexpresión de genes relacionados con un metabolismo de tipo respiratorio. Es la primera vez que se describe este tipo de metabolismo en *L. garvieae*, que además parece ser temperatura-dependiente en Lg21881.

Por otra parte, se analizó en ambas cepas la influencia de la temperatura sobre posibles factores de virulencia. La temperatura ha sido descrita como un factor relevante en la expresión de factores de virulencia en patógenos oportunistas como *Pseudomonas aeruginosa* (Wurtzel *et al*., 2012) y agentes zoonóticos como *L. monocytogenes* (Toledo-Arana *et al.,* 2009). En nuestro estudio se encontraron varios genes que han sido relacionados con la patogénesis de infecciones causadas por otras bacterias. Así, el gen *rpoE,* que codifica la subunidad delta de la ARN polimerasa, se encontró sobreexpresado a 18°C en Lg8831. RpoE se ha descrito como un modulador esencial de la adaptación ambiental en bacterias Gram-positivas (Xue *et al.,* 2010), y se ha asociado con la virulencia de *S. agalactiae* (Jones *et al*., 2003; Seepersaud *et al*., 2006) y el patógeno de peces *Vibrio harveyi* (Rattanama *et al*., 2012). De la misma manera, el gen *potD,* que juega un papel esencial en la adaptación, la supervivencia y la patogénesis de *S. pneumoniae* en el hospedador (Ware *et al*., 2006), se encontró sobreexpresado en Lg8831 a 18°C, por lo que su expresión diferencial podría jugar un papel esencial en la supervivencia y la adaptación de esta cepa al ambiente dentro de la trucha, con su implicación en el desencadenamiento

de la lactococosis. Además, en la cepa Lg8831 también se encontraron sobreexpresados a 18°C diversos genes que codifican enzimas autolíticas. La autolisis tiene un papel esencial en el desarrollo de las infecciones producidas por microorganismos Gram-positivos, debido a que provoca la liberación de moléculas de la pared y la membrana bacterianas; esto origina la liberación por parte de los macrófagos del hospedador de productos reactivos de oxígeno y nitrógeno, citoquinas citotóxicas, hidrolasas, proteinasas, así como la activación de procesos de coagulación y la cascada del complemento. Todos estos procesos están implicados en el desarrollo del *shock* séptico y fallo multiorgánico producidos por una infección microbiana grave (Ginsburg, 2002). En este sentido, es importante remarcar que la lactococosis es una enfermedad sistémica que cursa con septicemia en los peces afectados y por tanto, estas enzimas autolíticas podrían estar implicadas en el desarrollo de la misma.

En Lg21881 se encontraron sobreexpresados a 37°C varios genes relacionados con la homeostasis del manganeso, entre ellos un transportador del tipo ABC específico para manganeso y una proteína transportadora de manganeso de la familia Nramp. Ambos tipos de transportadores, particularmente las proteínas Nramp, han sido extensamente relacionados con la patogénesis de las infecciones causadas en humanos por diferentes bacterias Gram-positivas (Papp-Wallace y Maguire, 2006). De la misma manera, la expresión diferencial de estos genes podría jugar un papel en la patogénesis de la infección por Lg21881 en humanos.

Los resultados del primer estudio de genómica funcional en *L. garvieae* mediante microarrays de ADN han contribuido a mejorar el conocimiento sobre las vías de regulación para la adaptación a cambios en condiciones ambientales como la temperatura y, en consecuencia, mejorar la comprensión de la biología en general, y de los mecanismos de adaptación a diferentes nichos ecológicos en particular, de este importante patógeno.

Los resultados de la presente Tesis Doctoral resultan pioneros en el campo de la genómica de *L. garvieae*, destacando los diferentes estudios abordados sobre el genoma de este patógeno, así como el primer análisis funcional global del mismo. Los resultados obtenidos aportan nuevos conocimientos sobre esta bacteria de importancia creciente en medicina veterinaria y humana, abren camino hacia otros estudios genómicos globales en *L. garvieae*, y sientan las bases para una comprensión más profunda de su evolución y biología.

CONCLUSIONES

A continuación se exponen las conclusiones extraídas a partir del trabajo desarrollado en la presente Tesis Doctoral:

1. **La hibridación genómica comparativa inter-especie mediante microarrays de ADN es una aproximación válida para el estudio del contenido genómico de bacterias cuyo genoma no ha sido completamente secuenciado.**

2. **La secuenciación del genoma de *L. garvieae* ha puesto de manifiesto la elevada variabilidad genética que existe en esta especie.**

3. **Los cinco plásmidos presentes en *L. garvieae* 21881 confieren a esta cepa características únicas, algunas de las cuales podrían estar implicadas en la adaptación a determinados nichos ecológicos.**

4. **La utilización de lactosa y el gen de la fosfo-beta-galactosidasa (*lacG*) no son biomarcadores adecuados para la clasificación de aislados de *L. garvieae* en función de su origen.**

5. **El microarray de ADN diseñado y validado en el presente trabajo representa el primer microarray disponible para *L. garvieae,* cuya utilización facilitará futuros estudios de genómica funcional en este patógeno.**

6. **Las cepas de *L. garvieae* 21881 y 8831 presentan distinta respuesta transcripcional global a la temperatura, mostrando *L. garvieae* 8831 una respuesta adaptativa específica y coordinada por debajo de su temperatura óptima de crecimiento.**

BIBLIOGRAFÍA

Aguado M, Blanco M (2007) Microarrays de ADN en Microbiología. RCCV. 1: 125-134.

Aguado-Urda M, Gibello A, Vega J, Acosta F, Padilla D, Blanco MM (2012) Adherencia e invasión de *Lactococcus garvieae* en células no fagocíticas de trucha. XI Congreso SEM Microbiología del Medio Acuático, Barcelona.

Aguirre M, Collins MD (1993) Lactic acid bacteria and human clinical infection. J Appl Bacteriol. 75: 95-107.

Alegría A, Alvarez-Martín P, Sacristán N, Fernández E, Delgado S, *et al.* (2009) Diversity and evolution of the microbial populations during manufacture and ripening of Casín, a traditional Spanish, starter-free cheese made from cow's milk. Int J Food Microbiol. 136: 44–51.

Alrabadi NI (2012) The effect of several antibiotics on *Lactococcus garvieae* isolated from jordanian dairy products. Am J Agr Biol Sci. 7: 468-472.

Altinok I (2011) Multiplex PCR assay for detection of four major bacterial pathogens causing rainbow trout disease. Dis Aquat Organ. 93: 199-206.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. 215: 403-410.

An Y, Adachi Y, Ogawa Y (2004) Classification of lactic acid bacteria isolated from chigee and mare milk collected in Inner Mongolia. Anim Sci. 75: 245-252.

Antolín J, Cigüenza R, Salueña I, Vázquez E, Hernández J, *et al.* (2004) Liver abscess caused by *Lactococcus lactis cremoris*: a new pathogen. Scand J Infect Dis. 36: 490-491.

Aquilanti L, Garofalo C, Osimani A, Silvestri G, Vignaroli C, *et al.* (2007) Isolation and molecular characterization of antibiotic-resistant lactic acid bacteria from poultry and swine meat products. J Food Prot. 70: 557-565.

Aspiroz C, Sáez-nieto JA, Vela AI (2007) Bacteriemia por *Lactococcus garvieae* y *Citrobacter freundii*. Casos de Microbiología Clínica. Caso nº 387.

Aubin GG, Bémer P, Guillouzouic A, Crémet L, Touchais S, *et al.* (2011) First report of a hip prosthetic and joint infection caused by *Lactococcus garvieae* in a woman fishmonger. J Clin Microbiol. 49: 2074-2076.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, *et al.* (2008) The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 9:75.

Balcázar JL, Vendrell D, de Blas I, Ruiz-Zarzuela I, Gironés O, *et al.* (2007) In vitro competitive adhesion and production of antagonistic compounds by lactic acid bacteria against fish pathogens. Vet Microbiol. 122: 373–380.

Barakat RK, Griffiths MW, Harris LJ (2000) Isolation and characterization of *Carnobacterium, Lactococcus*, and *Enterococcus* spp. from cooked, modified atmosphere packaged, refrigerated, poultry meat. Int J Food Microbiol. 62: 83-94.

Barnes AC, Guyot C, Hansen BG, Mackenzie K, Horne MT, *et al*. (2002) Resistance to serum killing may contribute to differences in the abilities of capsulate and non-capsulated isolates of *Lactococcus garvieae* to cause disease in rainbow trout (*Oncorhynchus mykiss* L.). Fish Shellfish Immunol. 12: 155–168.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, *et al.* (2013) NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 41 (Database issue): D991-D995.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, *et al.* (2013) GenBank. Nucleic Acids Res. 41 (Database issue): D36-D42.

Bercovier H, Ghittino C, Eldar A (1997) Immunization with bacterial antigens: infections with streptococci and related organisms. Dev Biol Stand. 90: 153-160.

Borrero J, Brede DA, Skaugen M, Diep DB, Herranz C, *et al.* (2011) Characterization of garvicin ML, a novel circular bacteriocin produced by *Lactococcus garvieae* DCC43, isolated from mallard ducks (*Anas platyrhynchos*). Appl Environ Microbiol. 77: 369-373.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, *et al.* (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet. 29: 365-371.

Brudeseth BE, Wiulsrød R, Fredriksen BN, Lindmo K, Løkling K-E, *et al*. (2013) Status and future perspectives of vaccines for industrialised fin-fish farming. Fish Shellfish Immunol. Jun 12. [Epub ahead of print]

Brunt J, Austin B (2005) Use of a probiotic to control lactococcosis and streptococcosis in rainbow trout, *Oncorhynchus mykiss* (Walbaum). J Fish Dis. 28: 693–701.

Callon C, Duthoit F, Delbès C, Ferrand M, Le Frileux Y, *et al*. (2007) Stability of microbial communities in goat milk during a lactation year: molecular approaches. Syst Appl Microbiol. 30: 547-560.

Carvalho MG, Vianni MC, Elliott JA, Reeves M, Facklam RR, *et al*. (1997) Molecular analysis of *Lactococcus garvieae* and *Enterococcus gallinarum* isolated from water buffalos with subclinical mastitis. Adv Exp Med Biol. 418: 401-404.

Chan JF, Woo PC, Teng JL, Lau SK, Leung SS, *et al.* (2011) Primary infective spondylodiscitis caused by *Lactococcus garvieae* and a review of human *L. garvieae* infections. Infection. 39: 259-264.

Chao CT, Lai CF, Huang JW (2013) *Lactococcus garvieae* peritoneal dialysis peritonitis. Perit Dial Int. 33: 100-101.

Collins MD, Farrow JA, Phillips BA, Kandler O (1983) *Streptococcus garvieae* sp. nov. and *Streptococcus plantarum* sp. nov. J Gen Microbiol. 129: 3427–3431.

Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14: 1394-1403.

Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 5: e11147.

Delgado S, Rachid CT, Fernández E, Rychlik T, Alegría A, *et al.* (2013) Diversity of thermophilic bacteria in raw, pasteurized and selectively-cultured milk, as assessed by culturing, PCR-DGGE and pyrosequencing. Food Microbiol. 36: 103-111.

Devriese LA, Hommez J, Laevens H, Pot B, Vandamme P, *et al.* (1999) Identification of aesculin-hydrolyzing streptococci, lactococci, aerococci and enterococci from subclinical intramammary infections in dairy cows. Vet Microbiol. 70: 87-94.

Doménech A, Prieta J, Fernández-Garayzábal JF, Collins MD, Jones D, *et al.* (1993) Phenotypic and phylogenetic evidence for a close relationship between *Lactococcus garvieae* and *Enterococcus seriolicida*. Microbiologia. 9: 63-68.

Dong Y, Glasner JD, Blattner FR, Triplett EW (2001) Genomic interspecies microarray hybridization: rapid discovery of three thousand genes in the maize endophyte, *Klebsiella pneumoniae* 342, by microarray hybridization with *Escherichia coli* K-12 open reading frames. Appl Environ Microbiol. 67: 1911-1921.

Edalatian MR, Habibi-Najafi MB, Mortazavi SA, Alegría A, Nassiri MR, *et al.* (2012) Microbial diversity of the traditional Iranian cheeses Lighvan and Koozeh, as revealed by polyphasic culturing and culture-independent approaches. Dairy Sci Technol. 92: 75-90.

El-Baradei G, Delacroix-Buchet A, Ogier JC (2007) Biodiversity of bacterial ecosystems in traditional Egyptian Domiati cheese. Appl Environ Microbiol. 73: 1248-1255.

El-Baradei G, Delacroix-Buchet A, Ogier JC (2008) Bacterial biodiversity of traditional Zabady fermented milk. Int J Food Microbiol. 121: 295-301.

Eldar A, Ghittino C, Asanta L, Bozzetta E, Goria M, *et al.* (1996) *Enterococcus seriolicida* is a junior synonym of *Lactococcus garvieae*, a causative agent of septicemia and meningoencephalitis in fish. Curr Microbiol. 32: 85-88.

Eldar A, Ghittino C (1999) *Lactococcus garvieae* and *Streptococcus iniae* infections in rainbow trout *Oncorhynchus mykiss*: similar, but different diseases. Dis Aquat Organ. 36: 227-231.

Evans JJ, Pasnik DJ, Klesius PH, Al-Ablani S (2006) First report of *Streptococcus agalactiae* and *Lactococcus garvieae* from a wild bottlenose dolphin (*Tursiops truncatus*). J Wildl Dis. 42: 561-569.

Eyngor M, Zlotkin A, Ghittino C, Prearo M, Douet DG, *et al.* (2004) Clonality and diversity of the fish pathogen *Lactococcus garvieae* in Mediterranean countries. Appl Environ Microbiol. 70: 5132-5137.

Facklam R, Elliott JA (1995) Identification, classification, and clinical relevance of catalase-negative, gram-positive cocci, excluding the streptococci and enterococci. Clin Microbiol Rev. 8: 479-495.

Fefer JJ, Ratzan KR, Sharp SE, Saiz E (1998) *Lactococcus garvieae* endocarditis: report of a case and review of the literature. Diagn Microbiol Infect Dis. 32: 127-130.

Ferrario C, Ricci G, Borgo F, Rollando A, Fortina MG (2012) Genetic investigation within *Lactococcus garvieae* revealed two genomic lineages. FEMS Microbiol Lett. 332: 153-161.

Ferrario C, Ricci G, Milani C, Lugli GA, Ventura M, *et al*. (2013) *Lactococcus garvieae:* Where is it from? A first approach to explore the evolutionary history of this emerging pathogen. PLoS One. Nov 19. Aceptado para su publicación.

Fihman V, Raskine L, Barrou Z, Kiffel C, Riahi J, *et al.* (2006) *Lactococcus garvieae* endocarditis: identification by 16S rRNA and sodA sequence analysis. J Infect. 52: e3-6.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, *et al*. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science. 269: 496-512.

Fleming H, Fowler SV, Nguyen L, Hofinger DM (2012) *Lactococcus garvieae* multi-valve infective endocarditis in a traveler returning from South Korea. Travel Med Infect Dis. 10: 101-104.

Flórez AB, Mayo B (2006) Microbial diversity and succession during the manufacture and ripening of traditional, Spanish, blue-veined Cabrales cheese, as determined by PCR-DGGE. Int J Food Microbiol. 110: 165-171.

Flórez AB, Reimundo P, Delgado S, Fernández E, Alegría A, *et al*. (2012) Genome sequence of *Lactococcus garvieae* IPLA 31405, a bacteriocin-producing, tetracycline-resistant strain isolated from a raw-milk cheese. J Bacteriol. 194: 5118–5119.

Fog-Møller T, Andersen J (2012) Serious infection with *Lactococcus garvieae*. Ugeskr Laeger. 174: 1096-1097.

Fortina MG, Ricci G, Foschino R, Picozzi C, Dolci P, *et al.* (2007) Phenotypic typing, technological properties and safety aspects of *Lactococcus garvieae* strains from dairy environments. J Appl Microbiol. 103: 445–453.

Fortina MG, Ricci G, Borgo F (2009) A study of lactose metabolism in *Lactococcus garvieae* reveals a genetic marker for distinguishing between dairy and fish biotypes. J Food Prot. 72: 1248-54.

Foschino R, Nucera D, Volponi G, Picozzi C, Ortoffi M, *et al.* (2008) Comparison of *Lactococcus garvieae* strains isolated in northern Italy from dairy products and fishes through molecular typing. J Appl Microbiol. 105: 652–662.

Fukiya S, Mizoguchi H, Tobe T, Mori H (2004) Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* Strains revealed by comparative genomic hybridization microarray. J Bacteriol. 186: 3911-3921.

Furutan NP, Breiman RF, Fischer MA, Facklam RR (1991) *Lactococcus garvieae* infections in humans: a cause of prosthetic valve endocarditis (abstract C-297). 91st General Meeting American Society for Microbiology. 109.

Gabrielsen C, Brede D, Hernández PE, Nes IF, Diep DB (2012) Genome sequence of the bacteriocin-producing strain *Lactococcus garvieae* DCC43. J Bacteriol 194: 6976–6977.

Garvie EI, Farrow JAE, Phillips BA (1981) A taxonomic study of some strains of streptococci which grow at 10 °C but not at 45 °C including *Streptococcus lactis* and *Streptococcus cremoris.* Zentralblatt für Bakteriologie Mikrobiologie und Hygiene: I. Abt. Originale C: Allgemeine, angewandte und ökologische Mikrobiologie 2: 151-165.

Ghittino C, Múzquiz JL (1998) La estreptococosis de la trucha arco iris en España. Reunión de Piscicultores. Zaragoza. Rev Aquatic. 2.

Ginsburg I (2002) The role of bacteriolysis in the pathophysiology of inflammation, infection and post-infectious sequelae. APMIS. 110: 753-70.

Górecki RK, Koryszewska-Bagińska A, Gołębiewski M, Żylińska J, Grynberg M, *et al.* (2011) Adaptative potential of the *Lactococcus lactis* IL594 strain encoded in its 7 plasmids. PLoS One. 6: e22238.

Goyache J, Vela AI, Gibello A, Blanco MM, Briones V, *et al.* (2001) *Lactococcus lactis* subsp. *lactis* infection in waterfowl: first confirmation in animals. Emerg Infect Dis. 7: 884-886.

Green ED (2001) Strategies for the systematic sequencing of complex genomes. Nat Rev Genet. 2: 573-583.

Guzmán E, Romeu A, Garcia-Vallve S (2008) Completely sequenced genomes of pathogenic bacteria: a review. Enferm Infecc Microbiol Clin. 26: 88-98.

Hakenbeck R, Balmelle N, Weber B, Gardès C, Keck W, *et al.* (2001) Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*. Infect Immun. 69: 2477-2486.

Hatzikamari M, Litopoulou-Tzanetaki E, Tzanetakis N (1999) Microbiological characteristics of anevato: a traditional greek cheese. J Appl Microbiol. 87: 595-601.

Hendrickx AP, Willems RJ, Bonten MJ, van Schaik W (2009) LPxTG surface proteins of enterococci. Trends Microbiol. 17: 423-430.

Henson J, Tischler G, Ning Z (2012) Next-generation sequencing and large genome assemblies. Pharmacogenomics. 13: 901-915.

Henz SR, Huson DH, Auch AF, Nieselt-Struwe K, Schuster SC (2005) Whole-genome prokaryotic phylogeny. Bioinformatics. 21: 2329-2335.

Hirakawa TF, Costa FA, Vilela MC, Rigon M, Abensur H, *et al.* (2011) *Lactococcus garvieae* endocarditis: first case report in Latin America. Arq Bras Cardiol. 97: e108-110.

Hirokawa Y, Irie T, Ooyama T, Yasuda H, Nakamura A, *et al.* (2004). *Lactococcus garvieae* strains from Yellowtail *Seriola quinqueradiata* carry different lengths of fimbriae on their cell surface. Fisheries Sci. 70: 521-523.

James PR, Hardman SM, Patterson DL (2000) Osteomyelitis and possible endocarditis secondary to *Lactococcus garvieae*: a first case report. Postgrad Med J. 76: 301-303.

Jokovic N, Nikolic M, Begovic J, Jovcic B, Savic D, *et al.* (2008) A survey of the lactic acid bacteria isolated from Serbian artisanal dairy product kajmak. Int J Food Microbiol. 127: 305-311.

Jones AL, Needham RH, Rubens CE (2003) The Delta subunit of RNA polymerase is required for virulence of *Streptococcus agalactiae*. Infect Immun. 71: 4011-4017.

Jung MY, Chang YH, Kim W (2010) A real-time PCR assay for detection and quantification of *Lactococcus garvieae*. J Appl Microbiol. 108: 1694-1701.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40(Database issue): D109-D114.

Kang SH, Shin GW, Shin YS, J PK, Kim YR, *et al.* (2004) Experimental evaluation of pathogenicity of *Lactococcus garvieae* in black rockfish (*Sebastes schlegeli*). J Vet Sci. 5: 387-390.

Kawanishi M, Yoshida T, Yagashiro S, Kijima M, Yagyu K, *et al*. (2006) Differences between *Lactococcus garvieae* isolated from the genus *Seriola* in Japan and those isolated from other animals (trout, terrestrial animals from Europe) with regard to pathogenicity, phage susceptibility and genetic characterization. J Appl Microbiol. 101: 496–504.

Kawanishi M, Yoshida T, Kijima M, Yagyu K, Nakai T, *et al.* (2007) Characterization of *Lactococcus garvieae* isolated from radish and broccoli sprouts that exhibited a KG+ phenotype, lack of virulence and absence of a capsule. Lett Appl Microbiol. 44: 481–487.

Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res. 12: 656-664.

Kim JH, Go J, Cho CR, Kim JI, Lee MS, *et al*. (2013) First report of human acute acalculous cholecystitis caused by the fish pathogen *Lactococcus garvieae*. J Clin Microbiol. 51: 712-714.

Kimura H, Kusuda R (1982) Studies on the pathogenesis of streptococcal infection in cultured yellowtails *Seriola* spp.: effect of crude exotoxin fraction from cell-free culture on experimental streptococcal infection. J Fish Dis. 5: 471-478.

Klijn N, Weerkamp AH, de Vos WM (1995) Detection and characterization of lactose-utilizing *Lactococcus* spp. in natural ecosystems. Appl Environ Microbiol. 61: 788-792.

Kocabas P, Calik P, Calik G, Ozdamar TH (2009) Microarray Studies in *Bacillus subtilis*. Biotechnol J. 4: 1012-1027.

Kosuge T, Mashima J, Kodama Y, Fujisawa T, Kaminuma E, *et al*. (2013) DDBJ progress report: a new submission system for leading to a correct annotation. Nucleic Acids Res. Nov 4. [Epub ahead of print]

Kubota H, Tsuji H, Matsuda K, Kurakawa T, Asahara T, *et al.* (2010) Detection of human intestinal catalase-negative, Gram-positive cocci by rRNA-targeted reverse transcription-PCR. Appl Environ Microbiol. 76: 5440–5451.

Kusuda R, Kawai K, Toyoshima T, Komatsu I (1976) A new pathogenic bacterium belonging to the genus *Streptococcus*, isolated from an epizootic of cultured yellowtail. Nippon Suisan Gakkaishi. 42: 1345–1352.

Kusuda R, Kawai K (1982) Characteristics of *Streptococcus* sp. pathogenic to yellowtail. Fish Pathol. 17: 11-16.

Kusuda R, Hamaguchi M (1988). Extracellular and intracellular toxins of *Streptococcus* sp. in the Yellowtail. Bull Eur Ass Fish Pathol. 8: 9-10.

Kusuda R, Kawa K, Salat F, Banner CR, Fryer JL (1991) *Enterococcus seriolicida* sp. nov., a fish pathogen. Int J Syst Bacteriol. 41: 406–409.

Lafarge V, Ogier JC, Girard V, Maladen V, Leveau JY, *et al.* (2004) Raw cow milk bacterial population shifts attributable to refrigeration. Appl Environ Microbiol. 70: 5644-5650.

Laing CR, Zhang Y, Thomas JE, Gannon VP (2011) Everything at once: comparative analysis of the genomes of bacterial pathogens. Vet Microbiol. 153: 13-26.

Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, *et al.* (2011) The European Nucleotide Archive. Nucleic Acids Res. 39 (Database issue): D28-D31.

Lemos ML, Osorio CR (2007) Heme, an iron supply for vibrios pathogenic for fish. Biometals 20: 615-626.

Li WK, Chen YS, Wann SR, Liu YC, Tsai HC (2008) *Lactococcus garvieae* endocarditis with initial presentation of acute cerebral infarction in a healthy immunocompetent man. Intern Med. 47: 1143-1146.

Li JS, Bi YT, Dong C, Yang JF, Liang WD (2011) Transcriptome analysis of adaptive heat shock response of *Streptococcus thermophilus*. PLoS One. 6: e25777.

Linden SK, Sutton P, Karlsson NG, Korolik V, McGuckin MA (2008) Mucins in the mucosal barrier to infection. Mucosal Immunol. 1: 183-197.

López-Campos G. Bioinformatics in Support of Microarray Experiements. En: López-Campos G, Martínez-Suárez JV, Aguado-Urda M, López-Alonso V (2012) Microarray detection and characterization of bacterial foodborne pathogens. Springer Briefs in Food, Health and Nutrition Series. Springer. Nueva York.

Luciani F, Bull RA, Lloyd AR (2012) Next generation deep sequencing and vaccine design: today and tomorrow. Trends Biotechnol. 30: 443-452.

Magariños B, Romalde JL, Lemos ML, Barja JL, Toranzo AE (1994) Iron uptake by *Pasteurella piscicida* and its role in pathogenicity for fish. Appl Environ Microbiol. 60: 2990-2998.

Maldonado-Barragán A, Cárdenas N, Martínez B, Ruiz-Barba JL, Fernández-Garayzábal JF, *et al.* (2013) Garvicin A, a novel class IId bacteriocin from *Lactococcus garvieae* that inhibits septum formation in *L. garvieae* strains. Appl Environ Microbiol. 79: 4336-4346.

Mata AI, Gibello A, Casamayor A, Blanco MM, Domínguez L, *et al*. (2004) Multiplex PCR assay for detection of bacterial pathogens associated with warm-water Streptococcosis in fish. Appl Environ Microbiol. 70: 3183-3187.

Mendoza GM, Pasteris SE, Ale CE, Otero MC, Bühler MI, *et al.* (2012) Cultivable microbiota of *Lithobates catesbeianus* and advances in the selection of lactic acid bacteria as biological control agents in raniculture. Res Vet Sci. 93: 1160-1167.

Menéndez A, Fernández L, Reimundo P, Guijarro JA (2007) Genes required for *Lactococcus garvieae* survival in a fish host. Microbiol. 153: 3286-3294.

Mereghetti L, Sitkiewicz I, Green NM, Musser JM (2008) Remodeling of the *Streptococcus agalactiae* transcriptome in response to growth temperature. PLoS One. 3: e2785.

Metzker ML (2010) Sequencing technologies - the next generation. Nat Rev Genet. 11: 31-46.

Mills S, McAuliffe OE, Coffey A, Fitzgerald GF, Ross RP (2006) Plasmids of lactococci - genetic accessories or genetic necessities? FEMS Microbiol Rev. 30: 243-273.

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.

Miyauchi E, Toh H, Nakano A, Tanabe S, Morita H (2012) Comparative genomic analysis of *Lactococcus garvieae* strains isolated from different sources reveals candidate virulence genes. Int J Microbiol. 2012: 728276.

Mofredj A, Baraka D, Kloeti G, Dumont JL (2000) *Lactococcus garvieae* septicemia with liver abscess in an immunosuppressed patient. Am J Med. 109: 513-514.

Morita H, Toh H, Oshima K, Yoshizaki M, Kawanishi M, *et al.* (2011) Complete genome sequence and comparative analysis of the fish pathogen *Lactococcus garvieae*. PLoS One. 6: e23184.

Munday BL (1994). Fish. En: Coopert SB, editor. Antimicrobial Prescribing Guidelines for Veterinarians: a Post Graduate FoundationPublication. University of Sydney, Australia, in association with the National Health and Medical Research Council. 305-325.

Nadrah K, Cerar T, Papst L, Volkar-Meglič J, Matičič M, *et al.* (2011) *Lactococcus garvieae* septicaemia in a patient with artificial heart valves. Wien Klin Wochenschr. 123: 677-679.

Nagarajan M, Pop M (2013) Sequence assembly demystified. Nat Rev Genet. 14: 157-167.

Navas ME, Hall G, El Bejjani D (2013) A case of endocarditis caused by *Lactococcus garvieae* and suggested methods for identification. J Clin Microbiol. 51: 1990-1992.

Nieva-Velasco CM (2012) la aplicación de NGS como técnicas para el estudio del microbioma humano. Trabajo de fin de Máster. Universidad Internacional de Andalucía.

Nobbs AH, Lamont RJ, Jenkinson HF (2009) *Streptococcus* adherence and colonization. Microbiol Mol Biol Rev. 73: 407-450.

Olstorpe M, Lyberg K, Lindberg JE, Schnürer J, Passoth V (2008) Population diversity of yeasts and lactic acid bacteria in pig feed fermented with whey, wet wheat distillers' grains, or water at different temperatures. Appl Environ Microbiol. 74: 1696-1703.

Ooyama T, Hirokawa Y, Minami T, Yasuda H, Nakai T, *et al.* (2002) Cell-surface properties of *Lactococcus garvieae* strains and their immunogenicity in the yellowtail *Seriola quinqueradiata*. Dis Aquat Organ. 51: 169-177.

Palacios MA, Benayas E, Solorzano L, Martínez M, Bautista, C (1988) Señalización del primer caso de Estreptococcosis en la truchicultura intensiva en España. En: Universidad Complutense de Madrid–Asociacion Española de Aquicultura: Simposio sobre el cultivo intensivo de peces continentales, Madrid.

Palacios MA, Zamora MJ, Velazquez J, Zamora E, Durán A (1993) Streptococcosis in rainbow trout (*Oncorhynchus mykiss*) in Spain. Bollettino della Società Italiana di Patologia Ittica. 13: 11–14.

Papp-Wallace KM, Maguire ME (2006) Manganese Transport and the Role of Manganese in Virulence. Annu Rev Microbiol. 60: 187-209.

Park HK, Myung SC, Kim W (2012) Comparative transcriptomic analysis of *Streptococcus pseudopneumoniae* with viridans group streptococci. BMC Microbiol. 12: 77.

Pereira F, Ravelo C, Toranzo AE, Romalde JL (2004) *Lactococcus garvieae*, an emerging pathogen for the Portuguese trout culture. Bull Eur Ass Fish Pathol. 24: 274-279.

Pot B, Devriese LA, Ursi D, Vandamme P, Haesebrouck F, *et al.* (1996) Phenotypic identification and differentiation of *Lactococcus* strains isolated from animals. System Appl Microbiol. 19: 213–222.

Prieta J, Doménech A, Fernández-Garayzábal JF, Collins MD, Rodríguez UM, *et al*. (1993) Lactococosis de la trucha arco iris. Medicina Veterinaria. 10: 367-373.

Prodromou K, Thasitou P, Haritonidou E, Tzanetakis N, Litopoulou-Tzanetaki E (2001) Microbiology of “Orinotyri”, a ewe's milk cheese from the Greek mountains. Food Microbiol. 18: 319-328.

Qi J, Luo H, Hao B (2004) CVTree: a phylogenetic tree reconstruction tool based on whole genomes. Nucleic Acids Res. 32(Web Server issue): W45-47.

Rantsiou K, Urso R, Iacumin L, Cantoni C, Cattaneo P, *et al.* (2005) Culture-dependent and -independent methods to investigate the microbial ecology of Italian fermented sausages. Appl Environ Microbiol. 71: 1977-1986.

Rattanama P, Thompson JR, Kongkerd N, Srinitiwarawong K, Vuddhakul V, *et al.* (2012) Sigma E regulators control hemolytic activity and virulence in a shrimp pathogenic *Vibrio harveyi*. PLoS One. 7: e32523.

Ravelo C, Magariños B, Romalde JL, Toranzo AE (2001) Conventional versus miniaturized systems for the phenotypic characterization of *Lactococcus garvieae* strains. Bull Eur Ass Fish Pathol. 21: 136-144.

Ravelo C, Magariños B, López-Romalde S, Toranzo AE, Romalde JL (2003) Molecular fingerprinting of fish-pathogenic *Lactococcus garvieae* strains by random amplified polymorphic DNA analysis. J Clin Microbiol. 41: 751-756.

Ravelo C, Magariños B, Herrero MC, Costa L, Toranzo AE, *et al.* (2006) Use of adjuvanted vaccines to lengthen the protection against lactococcosis in rainbow trout (*Oncorhynchus mykiss*). Aquaculture 251: 153–158.

Reimundo P, Menéndez A, Méndez J, Pérez-Pascual D, Navais R, *et al.* (2010) dltA gene mutation in the teichoic acids alanylation system of *Lactococcus garvieae* results in diminished proliferation in its natural host. Vet Microbiol. 143: 434–439.

Reimundo P, Rivas AJ, Osorio CR, Méndez J, Pérez-Pascual D, *et al.* (2011) Application of suppressive subtractive hybridization to the identification of genetic differences between two *Lactococcus garvieae* strains showing distinct differences in virulence for rainbow trout and mouse. Microbiology. 157: 2106-2119.

Reimundo P, Pignatelli M, Alcaraz LD, D'Auria G, Moya A, *et al.* (2011) Genome sequence of *Lactococcus garvieae* UNIUD074, isolated in Italy from a lactococcosis outbreak. J Bacteriol. 193: 3684-3685.

Reimundo Díaz-Fierros, P (2011) Variabilidad intraespecífica y características del genoma del patógeno *Lactococcus garvieae*. Tesis Doctoral, Departamento de Biología Funcional, Universidad de Oviedo.

Ren Z, Pan C, Jiang L, Wu C, Liu Y, *et al.* (2011) Oxalate-degrading capacities of lactic acid bacteria in canine feces. Vet Microbiol. 152: 368-373.

Ricci G, Ferrario C, Borgo F, Rollando A, Fortina MG (2012) Genome Sequences of *Lactococcus garvieae* TB25, Isolated from Italian Cheese, and *Lactococcus garvieae* LG9, Isolated from Italian Rainbow Trout. J Bacteriol. 194: 1249–1250.

Ricci G, Ferrario C, Borgo F, Eraclio G, Fortina MG (2013) Genome Sequences of Two *Lactococcus garvieae* Strains Isolated from Meat. Genome Announc. 1: 2011–2012.

Romalde JL, Luzardo-Alvárez A, Ravelo C, Toranzo AE, Blanco-Méndez J (2004) Oral immunization using alginate microparticles as a useful strategy for booster vaccination against fish lactoccocosis. Aquaculture. 236: 119–129.

Russo G, Iannetta M, D'Abramo A, Mascellino MT, Pantosti A, *et al.* (2012) *Lactococcus garvieae* endocarditis in a patient with colonic diverticulosis: first case report in Italy and review of the literature. New Microbiol. 35: 495-501.

Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, *et al.* (2013) ArrayExpress update--trends in database growth and links to data analysis tools. Nucleic Acids Res. 41(Database issue): D987-D990.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA. 74: 5463-5467.

Santos EM, Jaime I, Rovira J, Lyhs U, Korkeala H, *et al.* (2005) Characterization and identification of lactic acid bacteria in "morcilla de Burgos". Int J Food Microbiol. 97: 285–296.

Schleifer KH, Kraus J, Dvorak C, Kilpper-Bälz R, Collins MD, *et al.* (1985) Transfer of *Streptococcus lactis* and related *Streptococci* to the genus *Lactococcus* gen. nov. Syst Appl Microbiol. 6: 183-195.

Schmidtke LM, Carson J (2003) Antigen recognition by rainbow trout (*Oncorhynchus mykiss*) of whole cell proteins expressed by *Lactococcus garvieae* when obtained directly from fish and under iron limited culture conditions. Vet Microbiol. 9: 363–371.

Seepersaud R, Needham RH, Kim CS, Jones AL (2006) Abundance of the delta subunit of RNA polymerase is linked to the virulence of *Streptococcus agalactiae*. J Bacteriol. 188: 2096-2105.

Siezen RJ, Renckens B, van Swam I, Peters S, van Kranenburg R, *et al.* (2005) Complete sequences of four plasmids of *Lactococcus lactis* subsp. *cremoris* SK11 reveal extensive adaptation to the dairy environment. Appl Environ Microbiol. 71: 8371-8382.

Soriani M, Telford JL (2010) Relevance of pili in pathogenic streptococci pathogenesis and vaccine development. Future Microbiol. 5: 735-747.

Stintzi A (2003) Gene expression profile of *Campylobacter jejuni* in response to growth temperature variation. J Bacteriol. 185: 2009-2016.

Taniguchi M (1982) Experiment on peroral inoculation via food to induce yellowtail streptococcosis. Bull Japan Soc Sci Fish. 48: 1717-1720.

Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, et al. (2008) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nat Biotechnol. 26: 889-896.

Teixeira LM, Merquior VL, Vianni MC, Carvalho MG, Fracalanzza SE, *et al.* (1996) Phenotypic and genotypic characterization of atypical *Lactococcus garvieae* strains isolated from water buffalos with subclinical mastitis and confirmation of *L. garvieae* as a senior subjective synonym of *Enterococcus seriolicida*. Int J Syst Bacteriol. 46: 664-668.

Tejedor JL (2008) Puesta a punto de una metodología para la detección de cocos Gram positivos patógenos de peces en piensos compuestos y muestras ambientales: estudio de su presencia y caracterización de *Lactococcus garvieae* en estas matrices. Tesis Doctoral Departamento de Sanidad Animal, Universidad Complutense de Madrid.

Tejedor JL, Vela AI, Gibello A, Casamayor A, Domínguez L, *et al.* (2011) A genetic comparison of pig, cow and trout isolates of *Lactococcus garvieae* by PFGE analysis. Lett Appl Microbiol. 53: 614-619.

Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, *et al.* (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. Nature. 459: 950-956.

Tosukhowong A, Zendo T, Visessanguan W, Roytrakul S, Pumpuang L*, et al.* (2012) Garvieacin Q, a novel class II bacteriocin from *Lactococcus garvieae* BCC 43578. Appl Environ Microbiol. 78: 1619-1623.

UniProt Consortium (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res. 41: D43-D47.

van Hijum SA, Baerends RJ, Zomer AL, Karsens HA, Martin-Requena V, *et al*. (2008) Supervised Lowess normalization of comparative genome hybridization data--application to lactococcal strain comparisons. BMC Bioinformatics. 9: 93.

Vela AI, Vázquez J, Gibello A, Blanco MM, Moreno M, *et al*. (2000) Phenotypic and genetic characterization of *Lactococcus garvieae* isolated in Spain from lactococcosis outbreaks

and comparison with isolates of other countries and sources. J Clin Microbiol. 38: 3791–3795.

Vendrell D, Balcázar JL, Ruiz-Zarzuela I, de Blas I, Gironés O, *et al.* (2006) *Lactococcus garvieae* in fish: a review. Infect Dis. 29: 177-198.

Vendrell D, Balcázar JL, de Blas I, Ruiz-Zarzuela I, Gironés O, *et al.* (2008) Protection of rainbow trout (*Oncorhynchus mykiss*) from lactococcosis by probiotic bacteria. Comp Immunol Microbiol Infect Dis 31: 337–345.

Villani F, Aponte M, Blaiotta G, Mauriello G, Pepe O, *et al.* (2001) Detection and characterization of a bacteriocin, garviecin L1-5, produced by *Lactococcus garvieae* isolated from raw cow's milk. J Appl Microbiol. 90: 430-439.

Vinh DC, Nichol KA, Rand F, Embil JM (2006) Native-valve bacterial endocarditis caused by *Lactococcus garvieae*. Diagn Microbiol Infect Dis. 56: 91-94.

Volkov A, Liavonchanka A, Kamneva O, Fiedler T, Goebel C, *et al.* (2010) Myosin cross-reactive antigen of *Streptococcus pyogenes* M49 encodes a fatty acid double bond hydratase that plays a role in oleic acid detoxification and bacterial virulence. J Biol Chem. 285: 10353-10361.

Wandersman C, Delepelaire P (2004) Bacterial iron sources: from siderophores to hemophores. Annu Rev Microbiol. 58: 611-647.

Wang CY, Shie HS, Chen SC, Huang JP, Hsieh IC, *et al.* (2007) *Lactococcus garvieae* infections in humans: possible association with aquaculture outbreaks. Int J Clin Pract. 61: 68-73.

Ware D, Jiang Y, Lin W, Swiatlo E (2006) Involvement of potD in *Streptococcus pneumoniae* Polyamine Transport and Pathogenesis. Infect Immun. 74: 352–361.

Watanabe Y, Naito T, Kikuchi K, Amari Y, Uehara Y, *et al.* (2011) Infective endocarditis with *Lactococcus garvieae* in Japan: a case report. J Med Case Rep. 5: 356.

Wilbring M, Alexiou K, Reichenspurner H, Matschke K, Tugtekin SM (2011) *Lactococcus garvieae* causing zoonotic prosthetic valve endocarditis. Clin Res Cardiol. 100: 545-546.

Wurtzel O, Yoder-Himes DR, Han K, Dandekar AA, Edelheit S, Greenberg EP, *et al.* (2012) The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. PLoS Pathog. 8: e1002945.

Wyder AB, Boss R, Naskova J, Kaufmann T, Steiner A, *et al.* (2011) *Streptococcus* spp. and related bacteria: their identification and their pathogenic potential for chronic mastitis - a molecular approach. Res Vet Sci. 91: 349-57.

Xu Z, Hao B (2009) CVTree update: a newly designed phylogenetic study platform using composition vectors and whole genomes. Nucleic Acids Res. 37(Web Server issue): W174-178.

Xue X, Tomasch J, Sztajer H, Wagner-Döbler I (2010) The delta subunit of RNA polymerase, RpoE, is a global modulator of *Streptococcus mutans* environmental adaptation. J Bacteriol. 192: 5081-5092.

Yasunaga N (1982) Occurrence of *Streptococcus* sp., a pathogen of cultured yellowtail, in muscle of sardine for diets. Fish Pathol. 17: 195–198.

Yiu KH, Siu CW, To KK, Jim MH, Lee KL, *et al.* (2007) A rare cause of infective endocarditis; *Lactococcus garvieae*. Int J Cardiol. 114: 286-287.

Yoon SS, Mekalanos JJ (2006) 2,3-butanediol synthesis and the emergence of the *Vibrio cholerae* El Tor biotype. Infect Immun. 74: 6547-6556.

Zlotkin A, Eldar A, Ghittino C, Bercovier H (1998) Identification of *Lactococcus garvieae* by PCR. J Clin Microbiol. 36: 983-985.

Zou L, Wang J, Huang B, Xie M, Li A (2010) A solute-binding protein for iron transport in *Streptococcus iniae.* BMC Microbiol. 10: 309.

Zuily S, Mami Z, Meune C (2011) *Lactococcus garvieae* endocarditis. Arch Cardiovasc Dis. 104: 138-139.