



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**Maestría y Doctorado en Ciencias Bioquímicas**

**ANÁLISIS TAXONÓMICO Y DEL POTENCIAL METABÓLICO DEL  
METAGENOMA DE QUESO COTIJA**

**TESIS**

**QUE PARA OPTAR POR EL GRADO DE:  
Doctor en Ciencias**

**PRESENTA:**

**Grisel Alejandra Escobar Zepeda**

**TUTOR PRINCIPAL**

**Dra. Maricarmen Quirasco Baruch  
Facultad de Química, UNAM**

**MIEMBROS DEL COMITÉ TUTOR**

**Dra. Carmen Wachter Rodarte  
Facultad de Química, UNAM**

**Dr. Miguel Ángel Cevallos Gaos  
Centro de Ciencias Genómicas, UNAM**

**CIUDAD DE MÉXICO, Abril de 2016**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# ANÁLISIS TAXONÓMICO Y DEL POTENCIAL METABÓLICO DEL METAGENOMA DE QUESO COTIJA

## RECONOCIMIENTOS

Esta tesis de doctorado se realizó bajo la tutoría de la Dra. Maricarmen Quirasco Baruch en el Laboratorio 312 del Conjunto E, Departamento de Alimentos y Biotecnología de la Facultad de Química en la Universidad Nacional Autónoma de México, y el análisis bioinformático se llevó a cabo bajo la dirección del Dr. Alejandro Sánchez Flores, jefe de la Unidad de Secuenciación Masiva y Bioinformática del Instituto de Biotecnología de la Universidad Nacional Autónoma de México, campus Cuernavaca, Mor.

Se reconoce el apoyo técnico de Jérôme Verleyen, Karel Estrada Guerra y Verónica Jiménez Jacinto de la Unidad de Secuenciación Masiva y Bioinformática del Instituto de Biotecnología de la UNAM.

El proyecto fue apoyado parcialmente por el PAPIIT IN222115. Durante mis estudios de doctorado gocé de una beca otorgada por CONACYT y apoyo para actividades extracurriculares de parte del “Programa de Apoyo a los Estudios del Posgrado” (PAEP).

### Miembros del Comité Tutor:

Dra. Carmen Wachter Rodarte	Facultad de Química, UNAM
Dr. Miguel Ángel Cevallos Gaos	Centro de Ciencias Genómicas, UNAM

### Jurado asignado:

PRESIDENTE	Dr. Francisco Ruíz Terán	Facultad de Química, UNAM
VOCAL	Dra. Rosa Ma. Gutiérrez Ríos	Instituto de Biotecnología, UNAM
VOCAL	Dr. Luis David Alcaráz Peraza	Instituto de Ecología, UNAM
VOCAL	Dra. Esperanza Martínez Romero	Centro de Ciencias Genómicas, UNAM
SECRETARIO	Dr. León Patricio Martínez Castilla	Facultad de Química, UNAM

## CONTENIDO

	Pág.
1. Resumen	1
2. Marco teórico	2
2.1. El queso Cotija artesanal madurado Región de Origen	2
2.2. La microbiota bacteriana del queso Cotija	3
2.3. Metabolismo de LAB en alimentos lácteos	6
Fermentación de carbohidratos	7
Catabolismo de aminoácidos	9
Catabolismo de ácidos grasos libres	11
El rol de las LAB en la inocuidad del queso	11
2.4. Nuevos enfoques en el análisis de comunidades microbianas	13
2.5. Tecnologías de secuenciación	14
2.6. Análisis bioinformático de los datos	17
Obtención de información taxonómica a partir de secuencias crudas	17
Anotación funcional: ensamble, predicción de genes y asignación de función	19
3. Justificación	22
4. Hipótesis	22
5. Objetivos	22
5.1. Objetivo principal	22
5.2. Objetivos particulares	22
6. Diagrama Experimental	23
7. Metodologías	23
7.1. Muestras de queso Cotija Región de Origen	23
7.2. Extracción de paquete celular y de ADN total	24
7.3. Análisis bioinformático de secuencias	25
Anotación taxonómica	25

Anotación funcional	26
Reconstrucción de genomas	28
8. Resultados y discusión	28
8.1. Reporte de calidad	29
8.2. Análisis taxonómico	31
Cálculo de índices de diversidad y esfuerzo de muestreo	34
Población dominante en el queso Cotija	36
Población subdominante y fracción de no dominantes	37
8.3. Anotación funcional	40
Búsqueda de compuestos de olor y sabor	43
Búsqueda de bacteriocinas y genes relacionados	47
8.4. Reconstrucción de genomas de las especies dominantes	51
9. Conclusiones	52
10. Perspectivas	53
Referencias	54
Anexo 1. Detección y eliminación de ADN contaminante	63
Detección de ADN bovino por q-PCR	63
Optimización de condiciones para eliminar ADN contaminante	64
Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S	66
Anexo 3. Tabla de anotación taxonómica por genes marcadores de copia única	77
Anexo 4. Resultados de la anotación taxonómica de eucariontes con el gen ribosomal 18S	79
Anexo 5. Artículo publicado producto de este trabajo de investigación:	
Metagenomic analysis of a Mexican ripened cheese reveals a unique complex microbiota	81
Glosario	93

## 1. Resumen

El queso Cotija Región de Origen es un alimento lácteo de producción artesanal originario de la sierra que se encuentra entre Jalisco y Michoacán, México. Al ser elaborado a partir de leche cruda de vaca sin adición de cultivo iniciador, su fermentación y maduración ocurren mediante la interacción de la microbiota inoculada durante su manufactura con los componentes del alimento. Dichos microorganismos pueden provenir de la leche, de la sal adicionada, de la atmósfera donde ocurre la maduración y del contacto con el quesero, entre otros.

El proceso de maduración, que puede durar desde 3 meses hasta años, se debe a la actividad de las enzimas presentes en el medio, que en su mayoría son de origen microbiano. Por ello, los factores que afectan el desarrollo de los microorganismos, la producción de enzimas y su actividad, influirán de forma determinante en este proceso. Entre estos factores, sobresalen la disminución del pH, la actividad acuosa y el potencial redox, que llevan a la selección de una microbiota capaz de resistir condiciones ácidas, pobres de oxígeno y estrés osmótico.

Se han realizado estudios de microbiología clásica y basados en métodos moleculares dependientes de PCR que han permitido la identificación de *Bacillus*, *Staphylococcus*, *Enterococcus*, *Lactococcus*, *Lactobacillus*, *Streptococcus*, *Marinilactibacillus*, *Vagococcus*, y *Virgibacillus* en queso Cotija Región de Origen. Además, en un ensayo de hibridación *in situ* se concluyó que los tres primeros géneros de la lista representan únicamente el 11% de la carga bacteriana total en términos de conteos de células marcadas, sin dominancia de ninguno de ellos en una pieza de queso Cotija. Los métodos dependientes de cultivo pueden favorecer la recuperación de especies que no son necesariamente las más abundantes, mientras que las técnicas independientes de cultivo basadas en la amplificación de regiones variables del gen ribosomal 16S, permiten la detección de microorganismos cultivables o no, sin embargo, su utilización implica sesgos relacionados con amplificación preferencial y problemas debidos a la heterogeneidad de dichos genes en multicopia.

Con el desarrollo de las tecnologías de secuenciación masiva de nueva generación, ha sido posible elucidar la información genética de cualquier sistema biológico con rendimientos muy altos y cada vez a menor costo. La secuenciación de un metagenoma permite obtener información sobre la diversidad especies así como del potencial metabólico del consorcio sin necesidad de cultivar microorganismos o clonar fragmentos de ADN.

El objetivo de este trabajo es explorar la diversidad bacteriana en una muestra representativa de queso Cotija Región de Origen, así como elucidar su potencial metabólico relacionado con la producción de compuestos responsables del perfil sensorial del alimento y la producción de bacteriocinas. El análisis del gen ribosomal 16S permitió identificar a *Lactobacillus*, *Leuconostoc* y *Weissella* como géneros dominantes y a más de 500 géneros no dominantes pertenecientes a 31 *Phyla* de bacterias y *Archaeas*. El análisis de marcadores filogenéticos de copia única permitió la reconstrucción de los genomas de dichas especies dominantes. La anotación funcional de genes permitió identificar que el consorcio tiene la capacidad de sintetizar una amplia gama de compuestos producto del metabolismo de ácidos grasos libres y aminoácidos de cadena ramificada, además, se encontraron genes de síntesis de bacteriocinas y una gran cantidad de genes de inmunidad.

## 2. Marco teórico

### 2.1 El queso Cotija artesanal madurado Región de Origen

El queso Cotija es un alimento mexicano que se produce de forma artesanal desde hace más de 400 años. Es originario de la sierra que se encuentra entre Jalisco y Michoacán (sierra de JalMich), de gran importancia en esta región y que ha ganado fama debido a que es reconocido internacionalmente por su sabor y calidad. Cada una de las piezas de queso pesa entre 20 y 30 kg, es madurado, de pasta friable y sabor intenso (Álvarez et al., 2005).

La zona productora es un área continua con forma de herradura que incluye los municipios de Santa María del Oro (Jalisco) y la parte sur de los municipios de Cotija y Tocumbo (Michoacán). Además se extiende a territorio de los municipios del norte de Jilotlán de los Dolores, oriente de Tamazula, sur de Valle de Juárez y de Quitupan (Jalisco); suroeste de los Reyes, Peribán y Tancitaro, y el norte de Buena Vista Tomatlán (Michoacán) (Figura 1). Los municipios que cuentan con productores que participan en el proceso de calificación del queso Cotija en la búsqueda de la denominación de origen son seis, Santa María del Oro, Jilotlán de los Dolores, Quitupan, Cotija, Tocumbo y Buena Vista Tomatlán (Pomeon et al., 2009).



Figura 1. Zona de producción del queso Cotija definida por las reglas de uso de la marca colectiva Queso Cotija Región de Origen, en verde y marrón se indica la zona reconocida (Álvarez et al., 2005).

La región donde se produce el queso es una ladera templada, muy plegada y con escalonamiento altitudinal que va desde los pies de monte que se elevan desde el valle de Tierra Caliente y que llega un poco antes de las cumbres frías del eje neovolcánico. Su altitud está comprendida entre los 700 y 1700 metros sobre el nivel del mar aprox. y las precipitaciones medias anuales son de alrededor de 900 mm al sur y 1200 mm en los otros puntos cardinales. La zona se caracteriza por una cubierta vegetal tipo selva baja caducifolia con vegetación secundaria irregular. Queda excluida la zona de bosque mixto (encino-pino), ubicada en las partes más altas de la región.

En este medio específico pastorea el ganado bovino productor de la leche con la cual se elabora el queso Cotija, y

son las características de clima, altura y suelo las que originan una vegetación típica del lugar, misma que se refleja en la composición de la leche producida. Por otra parte, la humedad relativa, vinculada a la temperatura, las lluvias y la altura, se relaciona con las características del queso elaborado y madurado en esta región.

La ocupación humana es de poblamiento escaso y disperso en ranchos aislados dedicados a la explotación ganadera con producción estacional de queso durante los meses de lluvia, generalmente de julio a octubre.

Con el paso de los años, la competencia desleal y la falta de apoyo para los productores de queso artesanal, muchos ranchos desaparecieron o abandonaron la producción de leche y queso. Para el año 2007, se estimaba que en la sierra de JalMich se encontraban alrededor de 200 ranchos, con una producción cada uno de 400 a 1500 kg de queso al año, dando aproximadamente un total de 400 Ton/año en toda la sierra (Pomeon et al., 2009). Actualmente es difícil hacer una estimación similar ya que tanto la producción como el número de ranchos es muy variable debido al incremento en la migración a los EE.UU. así como la situación de violencia vinculada al narcotráfico y el difícil acceso a algunos de los ranchos más lejanos.

El queso Cotija elaborado en la región delimitada por las reglas de uso (Figura 1) se obtiene a partir de leche cruda de vaca y sal de mar de grano de Cuyutlán, Col., sin adición de cultivo iniciador, ni tratamiento térmico, por lo que su maduración ocurre mediante la interacción de la microbiota inoculada de forma natural durante su manufactura y cuya dinámica poblacional cambia libremente durante un tiempo mínimo de tres meses, pero que puede prolongarse hasta años para su comercialización. En este proceso se modifican los valores de actividad acuosa ( $a_w$ ), potencial redox (Eh) y pH que llevan a la selección de ciertas especies que son capaces de resistir condiciones ácidas, de baja disponibilidad de agua y pobres de oxígeno (Escobar, 2012), lo que permite el desarrollo principalmente de bacterias ácido lácticas (LAB por sus siglas en inglés) y la exclusión de bacterias patógenas y coliformes, éstas últimas presentes en las primeras etapas de elaboración (Bravo, 2008).

Es ampliamente reconocido que la composición de los consorcios microbianos asociados a una matriz alimentaria dada, influye en la calidad y singularidad del producto final. Esto es claro en alimentos fermentados tradicionalmente, donde la biodiversidad microbiana se debe a los atributos geográficos de la zona en la que el alimento se elabora, y que tiene un fuerte impacto en el concepto de la Denominación de Origen (Montel et al., 2014).

## 2.2 La microbiota bacteriana del queso Cotija

Se han realizado estudios enfocados en la caracterización de la comunidad microbiana del queso Cotija basados en técnicas de amplificación e identificación de marcadores filogenéticos así como de microbiología clásica. Los primeros esfuerzos se enfocaron en el aislamiento e identificación de bacterias y levaduras lipolíticas por medio de la secuenciación de regiones variables de genes ribosomales y por pruebas bioquímicas en muestras con tiempos de maduración superior a 6 meses. En dicho trabajo se identificaron cepas bacterianas de los géneros



*Bacillus* y *Staphylococcus* (García, 2006). Siguiendo una estrategia similar, se aislaron y caracterizaron bacterias proteolíticas por secuenciación de la región V3 del gen ribosomal 16S en muestras de queso Cotija con distintos tiempos de maduración, de las que se aislaron colonias identificadas por dicho marcador como *Staphylococcus*, *Bacillus* y *Enterococcus* (Hernández, 2007).

Más tarde se monitorearon cambios en el desarrollo de BAL durante la maduración, donde se aislaron y caracterizaron por pruebas bioquímicas y reacción en cadena de la polimerasa (PCR por sus siglas en inglés) secuencias del ADNr 16S de *Lactobacillus*, *Lactococcus* y *Enterococcus* (Bravo, 2008; Cortés, 2009).

Un análisis más completo sobre riqueza de especies, dominancia relativa y dinámica poblacional de la microbiota del queso se llevó a cabo amplificando regiones variables de ADNr 16S con cebadores universales y su análisis por electroforesis en gel con gradiente desnaturizante (DGGE por sus siglas en inglés). La técnica permitió la identificación de secuencias pertenecientes a los géneros *Enterococcus*, *Marinilactibacillus* y *Vagococcus*. En dicho trabajo se concluyó que *Enterococcus* está presente durante toda la maduración sin ser dominante, mientras que el género *Marinilactibacillus* sí fue dominante, es decir, tuvo una abundancia superior con respecto al resto de los géneros encontrados (Zúñiga, 2009).

Con el objetivo de dilucidar el origen de algunos de los géneros encontrados en estos trabajos, se hizo el escrutinio por aislamiento e identificación de cepas por secuenciación del gen ribosomal 16S completo y por PCR-DGGE de las regiones variables V1 y V3 del ADNr 16S de las posibles fuentes de inoculación. Una de las más interesantes fue la sal de mar, donde se reportó la presencia de los géneros *Enterococcus* (Gómez, 2010), *Bacillus*, *Aerococcus* y *Staphylococcus*. El muestreo de otros puntos clave del proceso de elaboración llevó a la identificación en la superficie de la ubre de la vaca de *Klebsiella*, *Kocuria* y *Staphylococcus*; en la leche bronca de *Lactococcus*, *Staphylococcus*, *Rahnella*, *Kocuria*, *Acinetobacter* y *Klebsiella*; mientras que en la cuajada sin salar se detectó un grupo de bacterias no cultivables provenientes de piel humana o plantas, *Rahnella*, *Klebsiella*, *Lactococcus* y *Staphylococcus*; en la superficie de la mesa de amasado a *Staphylococcus*, *Brachybacterium*, *Kocuria*, *Enterobacter* y *Brevibacterium*; y por último en la cuajada salada a *Marinilactibacillus*, *Psychrobacter*, *Bacillus*, *Lactococcus*, *Staphylococcus*, *Klebsiella* y bacterias no cultivables de piel humana y plantas (García, 2011; Robles, 2014).

Con el fin de establecer diferencias entre la microbiota del producto de la marca colectiva Región de Origen y la de quesos “tipo Cotija” elaborados fuera de la región, se compararon muestras provenientes de la sierra de JalMich, con otras provenientes de Chiapas. Se concluyó que ambos quesos son diferentes en cuanto a características fisicoquímicas y de composición, y en muestras del queso Cotija de JalMich se identificaron por PCR-DGGE a bacterias de los géneros: *Bacillus*, *Virgibacillus* y *Vagococcus* con la región variable V3 del ADNr 16S como blanco, y a *Enterococcus faecium*, *Streptococcus infantarius* y *Staphylococcus equorum* usando ~340 bp el gen *rpoB* como marcador molecular (Casillas, 2013).

En resumen, cuando se han realizado estudios de caracterización taxonómica de bacterias asociadas al queso

Cotija por aislamiento o identificación directa de marcadores filogenéticos en muestras que cumplen con el tiempo mínimo de maduración de tres meses, se ha obtenido en total 9 géneros bacterianos (Tabla 1). Todos ellos pertenecen al *Phylum* de los *Firmicutes* y son LAB en su mayoría.

Tabla 1. Géneros bacterianos identificados por métodos moleculares en queso Cotija con 3 meses de maduración

Género bacteriano	Técnica de identificación	Referencia
<i>Enterococcus</i>	ARDRA y PCR de ADNr 16S	Hernández, 2007
	PCR y DGGE de ADNr 16S	Zúñiga, 2009
	PCR de ADNr 16S	Cortés, 2009
	PCR y DGGE de ADNr 16S	Gómez, 2010
	PCR y DGGE de <i>rpoB</i>	Casillas, 2013
<i>Bacillus</i>	PCR de ADNr 16S	García, 2006
	ARDRA y PCR de ADNr 16S	Hernández, 2007
	PCR y DGGE de ADNr 16S	Casillas, 2013
<i>Staphylococcus</i>	PCR de ADNr 16S	García, 2006
	ARDRA y PCR de ADNr 16S	Hernández, 2007
	PCR y DGGE de <i>rpoB</i>	Casillas, 2013
<i>Vagococcus</i>	PCR y DGGE de ADNr 16S	Zúñiga, 2009
	PCR y DGGE de ADNr 16S	Gómez, 2010
<i>Lactobacillus</i>	PCR y DGGE de ADNr 16S	Casillas, 2013
	PCR de ADNr 16S	Cortés, 2009
<i>Lactococcus</i>	PCR y DGGE de ADNr 16S	Gómez, 2010
<i>Streptococcus</i>	PCR de ADNr 16S	Cortés, 2009
<i>Virgibacillus</i>	PCR y DGGE de <i>rpoB</i>	Casillas, 2013
<i>Marinilactibacillus</i>	PCR y DGGE de ADNr 16S	Casillas, 2013
	PCR y DGGE de ADNr 16S	Zúñiga, 2009

Estos estudios abrieron el panorama sobre la riqueza de géneros bacterianos en el queso Cotija Región de Origen y sus posibles fuentes de inoculación, aunque no dejaron en claro su abundancia en la comunidad durante y al tiempo mínimo de maduración. Los tres géneros que encabezan la lista son los que se habían encontrado con mayor frecuencia y se caracterizan por poseer actividades lipolíticas y proteolíticas importantes en la generación del perfil de sabor y olor del queso. Sin embargo, la abundancia de dichos generos en conjunto, no rebasa el 11% de abundancia total relativa en el queso Cotija, calculada a partir de conteos de células marcadas con sondas fluorescentes género-específicas con respecto a una sonda universal de acuerdo a resultados obtenidos por hibridación *in situ* (FISH) (Escobar, 2012). De dicho trabajo se desprende la incógnita sobre si el 89% restante podría estar conformado por bacterias de los géneros que se enumeran en la Tabla 1 y/o por microorganismos que

no se han podido identificar por los métodos utilizados hasta el momento.

### 2.3 Metabolismo de LAB en alimentos lácteos

La mezcla de compuestos volátiles responsables del perfil sensorial de los alimentos lácteos proviene del metabolismo de carbohidratos, degradación y liberación de ácidos grasos libres por actividad de lipasas y estererasas, así como la degradación de aminoácidos libres producto de proteasas y peptidasas. La fuente principal de las enzimas involucradas en dicho metabolismo son las LAB (Smit et al., 2005).

Las bacterias ácido lácticas son un grupo que en su mayoría está relacionado filogenéticamente con el *Phylum* de los *Firmicutes* aunque también hay representantes dentro del *Phylum* de las actinobacterias como *Propionibacterium*, *Bifidobacterium* y *Corynebacterium* (Ouwehand y Vaughan, 2006). La descripción general de este grupo es que son cocos o bacilos Gram positivos, no esporulados, catalasa negativos, aerotolerantes, ácido tolerantes, nutricionalmente exigentes y fermentadores estrictos cuyo producto metabólico principal de la fermentación de carbohidratos es el ácido láctico. Hay muchos ejemplos de LAB que no cumplen con todas las características aquí mencionadas pero la única que se considera como indispensable es que deben ser bacterias Gram positivas (Axelsson, 2004).

Las LAB pueden clasificarse en dos grandes grupos de acuerdo a su capacidad para modificar las propiedades fisicoquímicas y sensoriales del alimento en el que se encuentran, misma que está relacionada con sus características metabólicas. Al primero de ellos se le reconoce como SLAB (bacterias ácido lácticas iniciadoras por sus siglas en inglés) y corresponde a aquellas que son productoras importantes de ácido láctico, responsables de la disminución drástica de pH en la primera etapa de la fermentación y básicamente se refiere a LAB con metabolismo de homofermentadoras obligadas. El segundo grupo son las NSLAB (no iniciadoras), que tienen importancia en el proceso de maduración, donde son responsables de la producción de una gama de metabolitos secundarios relacionados con olor y sabor, y cuyo metabolismo es principalmente heterofermentativo ya sea facultativo u obligado (Beresford y Williams, 2004; Felis y Dellaglio, 2007).

Algunas especies que no son productores importantes de lactato pero que son consideradas como SLAB debido a que contribuyen significativamente a la producción de compuestos de olor y sabor desde el inicio de la fermentación gracias a su metabolismo heterofermentativo son *Leuconostoc* y *Weissella* (Björkroth y Holzapfel, 2006).

De acuerdo con la teoría del balance de componentes, en la percepción del consumidor sobre el sabor y olor del queso, no sólo la concentración de cada uno de los componentes individuales es importante, sino el balance entre ellos. Es decir que aunque la composición de compuestos volátiles en dos alimentos sea la misma, si las proporciones son distintas, entonces el perfil de sabor global en ambos al momento de compararlos, cambiará. Si se enriquece preferentemente algunos de los componentes, por ejemplo, en una maduración acelerada, el perfil

sensorial del queso también será distinta. Una percepción desbalanceada de sabor en quesos se percibe como un defecto en el producto final (Smit et al, 2005). Los compuestos de sabor y olor más potentes son los aldehídos, alcoholes, ácidos carboxílicos y ésteres. Ejemplos de otros compuestos de los que se sabe tienen un papel importante en el perfil de olor y sabor en quesos son: cetonas, lactonas, ésteres, pirazinas, compuestos azufrados, compuestos carbonílicos, ácidos grasos libres de cadena corta (como acético, propiónico y butírico principalmente), aminoácidos libres y sales (Broadbent y Steele, 2005; Curioni y Bosset, 2002; d'Acampora Zellner et al., 2008).

### Fermentación de carbohidratos

La fermentación de carbohidratos por LAB ocurre por fosforilación a nivel de sustrato. En el caso del metabolismo de la glucosa, ésta es internalizada mediante el sistema PTS (sistema fosfotransferasa, por sus siglas en inglés), y es fosforilada para su activación dentro de la célula. La célula arranca la vía de la glucólisis (o Embden-Meyerhof-Parnas) hasta llegar a piruvato, que en condiciones pobres de oxígeno y exceso del azúcar, será reducida a lactato (Figura 2A). Este mecanismo ocurre durante la fermentación homoláctica en todas las LAB excepto en las heterofermentadoras obligadas.

El otro camino es la fermentación heteroláctica, también conocida como la ruta de las pentosas fosfato, donde el paso clave es la ruptura con la fosfocetolasa en gliceraldehído-3-fosfato (GAP) y acetil-fosfato. El primero se va a formación de lactato, mientras que cuando no haya otro aceptor de electrones, el acetil-fosfato es reducido a etanol vía acetyl-CoA y acetaldehído. Esta vía genera CO<sub>2</sub> como metabolito además de lactato y etanol (Figura 2B).

Las LAB en general son capaces de metabolizar una amplia gama de carbohidratos, como manosa y fructosa, que se incorporan en las rutas ya descritas a nivel de glucosa o fructosa fosfatadas después de ser isomerizadas y/o fosforiladas. Una excepción importante es la galactosa, que puede ser metabolizada por dos vías, si se internaliza mediante el sistema PTS, la galactosa es fosforilada al momento de transportarla y una vez dentro, es metabolizada mediante la ruta de la tagatosa-6-fosfato hasta GAP, que entra a glicólisis (Bissett y Anderson, 1974). La segunda opción es que la galactosa sea internalizada mediante permeasas en cuyo caso es fosforilada y metabolizada vía la ruta de Leloir hasta glucosa-6-fosfato que se incorpora a glicólisis. Estas rutas son la base en el metabolismo de lactosa que se internaliza como disacárido por medio de permeasas o mediante sistema PTS (Thomas et al., 1980).

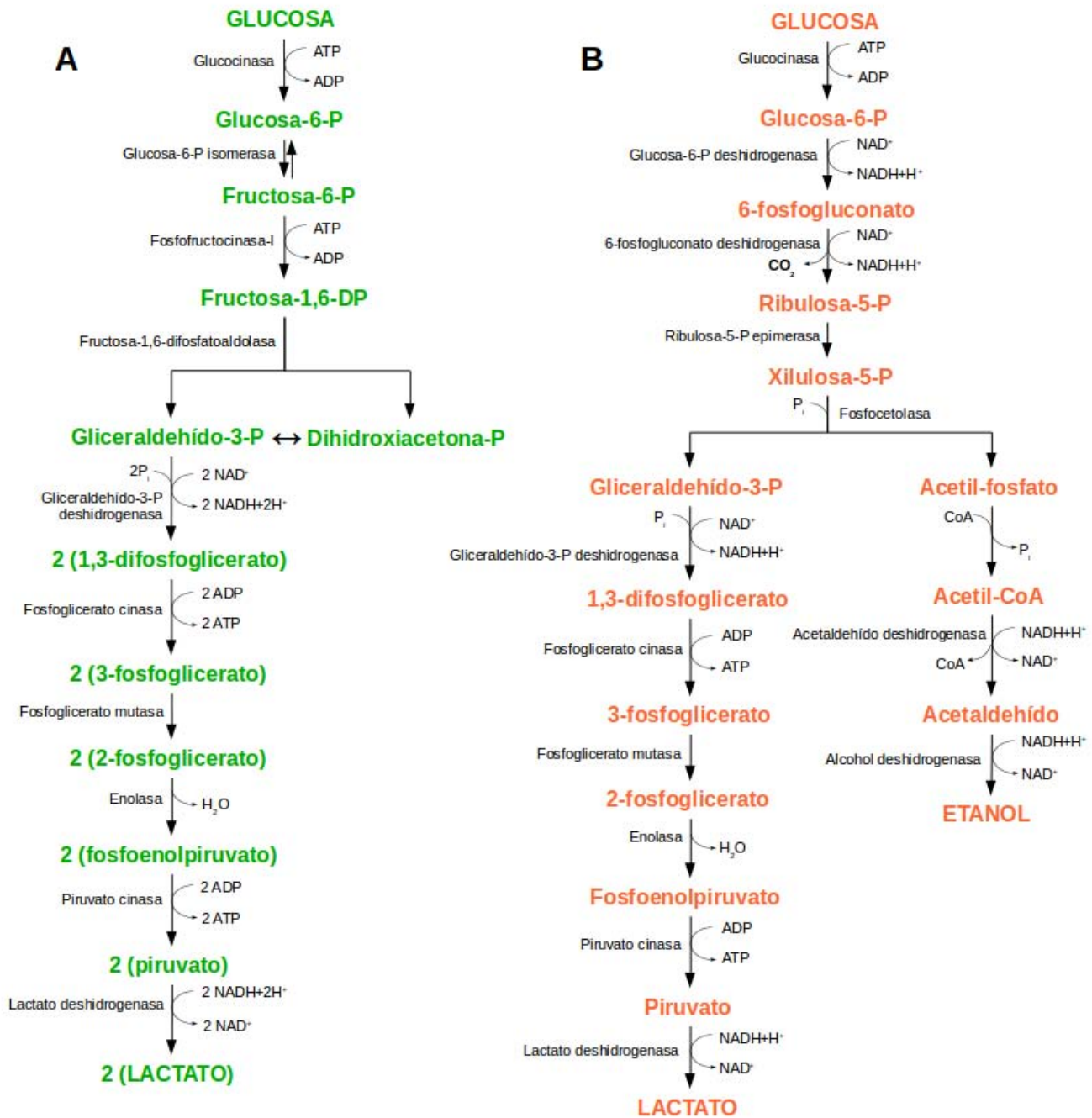


Figura 2. Rutas metabólicas de fermentación de la glucosa en LAB. A) Glucólisis y ruta homoláctica. B) Ruta de las pentosas fosfato o heteroláctica. Modificado de Axelsson, 2004.

De acuerdo al tipo de fermentación, las LAB pueden catalogarse en tres grandes grupos: Homofermentadoras obligadas, donde se incluye a los lactobacilos tipo I (*L. acidophilus*, *L. delbrueckii*, *L. farciminis* y *L. helveticus* entre otros) y algunas especies de otros géneros como *Lactococcus lactis* y *Pediococcus acidilactici*; heterofermentadoras obligadas donde se incluye a los lactobacilos del grupo III (por ejemplo *L. brevis*, *L. fermentum* y *L. buchneri*), leuconostocs, oenococos y weiselas; y heterofermentadoras facultativas donde se

encuentran los lactobacilos de tipo II (como *L. plantarum*, *L. rhamnosus*, *L. paracasei* y *L. casei*) y a la mayoría de las especies de enterococos, lactococos, pediococos, estreptococos, tetragenococos, carnobacterias y vagococos. Las LAB heterofermentadoras facultativas metabolizan las hexosas por vía homofermentativa y las pentosas y otros sustratos por vía heterofermentativa (Axelsson, 2004).

La cantidad de ácido láctico producido durante la maduración de un queso determinará el grado de pérdida de humedad, el pH final y la cantidad de lactosa residual, además de que éste merma en el desarrollo de microorganismos patógenos y de descomposición (Fox et al., 2000).

En general las LAB poseen un metabolismo plástico o adaptable a las condiciones en las que la célula se encuentra en función de las capacidades enzimáticas de la cepa. En las vías de fermentación principales usadas por LAB (Figura 2), el aceptor de electrones o hidrógeno es el piruvato. El cambio en el perfil de metabolitos puede deberse principalmente a alteraciones en el metabolismo de dicha molécula, el uso de aceptores de electrones alternativos o ambos (Axelsson, 2004).

Una de las rutas alternativas del piruvato de mayor importancia para la formación de compuestos relacionados con el perfil sensorial de los alimentos lácteos fermentados, es la producción de diacetilo y acetoína. Dichos compuestos se sintetizan a partir de un exceso de piruvato proveniente del catabolismo del citrato. El citrato internalizado en la célula mediante una permeasa es sustrato para una liasa que genera acetato y oxaloacetato, mismo que es descarboxilado a piruvato. El exceso de piruvato es entonces convertido a  $\alpha$ -acetolactato por una acetolactato sintasa y es a partir de éste que se obtiene principalmente acetoína y en menor proporción  $\text{CO}_2$  y diacetilo (McSweeney, 2004). Especies de *Leuconostoc*, *Pediococcus*, *Lactobacillus* y *Streptococcus* son productores importantes de diacetilo a partir de citrato siguiendo esta vía (Axelsson, 2004).

#### Catabolismo de aminoácidos

Además de los carbohidratos residuales que quedan después del desuerado, los componentes principales dentro de la matriz de cualquier queso, y que son sustrato para el metabolismo de LAB, son los lípidos y las proteínas, particularmente la caseína (Figura 3). La acción de proteasas y peptidasas dentro y fuera de la célula, lleva a la liberación de aminoácidos que pueden contribuir con notas específicas en el alimento. Por ejemplo, aminoácidos libres que confieren sabor dulce en orden decreciente son: D-triptófano, D-histidina, D-fenilalanina, D-tirosina, D-leucina, L-alanina y glicina; por otro lado, L-triptófano, L-fenilalanina, L-tirosina y L-leucina tienen sabor amargo; mientras que D- y L-cisteína y D- y L-metionina tienen notas sulfurosas; y finalmente, el ácido L-glutámico que tiene la propiedad única de potenciar otros sabores (Solms, 1969). Además, el catabolismo de los aminoácidos es un proceso importante para la formación de sabor en el queso, siendo los aminoácidos aromáticos, los de cadena ramificada y la metionina, los principales precursores de compuestos con olor y sabor (Rijnen et al., 2003).

La ruta catabólica de aminoácidos libres más importante en la obtención de compuestos de olor y sabor es la vía de las transaminasas, que implica la obtención de un  $\alpha$ -ceto ácido a partir de un aminoácido por acción de su respectiva aminotransferasa. El  $\alpha$ -cetoácido puede convertirse en aldehído, hidroxiácido o CoA-éster dependiendo de la enzima que actúe sobre el sustrato. Cuando se obtienen aldehídos por descarboxilación, éstos pueden ser deshidrogenados para formar alcoholes o hidrogenados para formar ácidos carboxílicos. Alcoholes y carboxilatos aportan olores y sabores menos intensos que los aldehídos y son sustratos para esterasas y aciltransferasas respectivamente en la formación de ésteres cuyo aporte al perfil de olor y sabor es menos importante.

Por otro lado, los hidroxiácidos que se obtienen directamente del  $\alpha$ -cetoácido por hidrogenación gracias a la acción de su deshidrogenasa, son compuestos que no aportan notas de olor y sabor al alimento ni son precursores de éstos, por lo que dicha reacción genera únicamente una disminución en la concentración del  $\alpha$ -cetoácido.

Una vía más corta para la formación de alcoholes y tioles a partir de aminoácidos libres es la de las liasas. Una de las reacciones más representativas de esta ruta es la obtención de metanotiol a partir de metionina así como la conversión de treonina a glicina y acetaldehído.

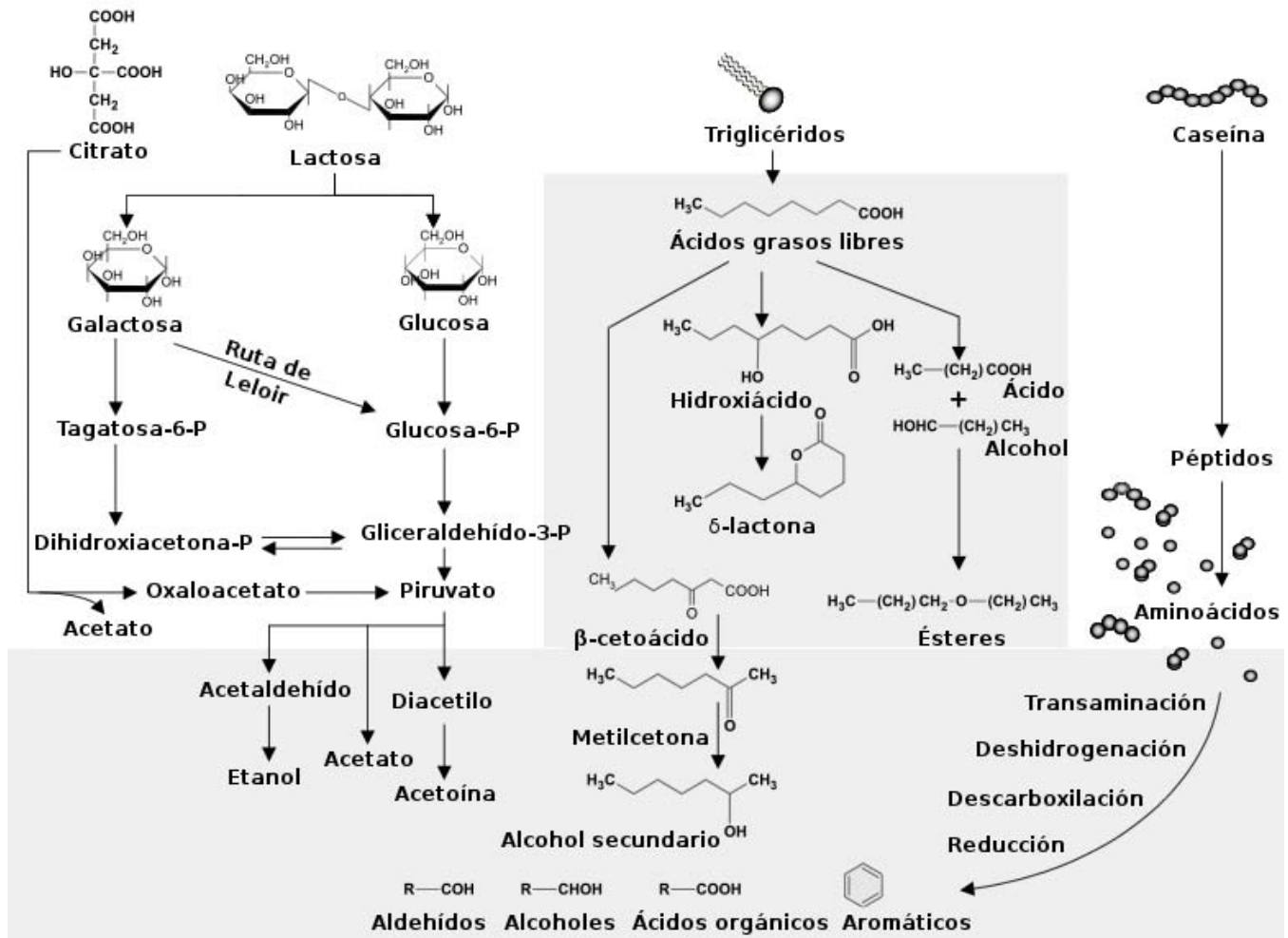


Figura 3. Resumen del catabolismo microbiano sobre los componentes del queso que llevan a la formación de compuestos de olor y sabor. Modificado de Marilley y Casey, 2004.

Finalmente, los aminoácidos pueden ser directamente descarboxilados para formar aminas con la liberación de CO<sub>2</sub> o desaminados para formar ácidos carboxílicos con la liberación de amoníaco (Smit et al., 2005).

#### Catabolismo de ácidos grasos libres

Aunque las LAB contribuyen poco a la lipólisis en quesos en comparación con la actividad que pueden tener hongos y levaduras, los ácidos grasos libres que se generan por acción de lipasas y esterases son fuente de sustratos para la formación de compuestos de olor y sabor (Figura 3) tales como metilcetonas, alcoholes secundarios, ésteres, alcanos y lactonas (McSweeney, 2004).

Al igual que los aminoácidos, algunos ácidos grasos libres pueden contribuir directamente con el perfil sensorial del queso, sobre todo los de cadena corta y media como el ácido butírico que se asocia al olor a rancio y sabor a grasa de leche (información obtenida de: <https://pubchem.ncbi.nlm.nih.gov/compound/264#section=Odor>); mientras que los ácidos grasos de cadena larga aportan más como estímulos texturales, cuyo descriptor podría ser oleoso o aceitoso, que con un sabor real (Mattes, 2009).

La reacción más sencilla que ocurre con los ácidos grasos libres es la del grupo carboxilo con etanol para formar etil ésteres y con tioles para formar tioésteres.

En un proceso más complejo, los ácidos grasos libres son sustrato para la  $\beta$ -oxidación, ruta mediante la cuál se reducen a  $\beta$ -cetoácidos que mediante descarboxilación pasan a metilcetonas (alcan-2-onas) con un átomo de carbono menos. Las síntesis de metilcetonas es importante en quesos madurados por hongos filamentosos como *Penicillium* y *Geotrichum*. Las metilcetonas son sustrato de reductasas en la formación de alcoholes secundarios, paso que es reversible en condiciones de aerobiosis (Molimard y Spinnler, 1996).

Otro tipo de compuestos importantes en el perfil de sabor y aroma en quesos son las lactonas, que son compuestos cíclicos que se forman a partir de la esterificación intramolecular de ácidos grasos hidroxilados. En queso las más importantes son las  $\gamma$ - y  $\beta$ -lactonas que se forman a partir de sus respectivos ácidos grasos  $\gamma$ - y  $\beta$ -hidroxilados (Hassan et al., 2013).

#### El rol de las LAB en la inocuidad del queso

Además de ser responsables de conferir características sensoriales al queso, las LAB juegan un papel importante en la inocuidad del mismo mediante la producción de ácidos orgánicos, peróxido de hidrógeno, diacetilo, CO<sub>2</sub>, moléculas de bajo peso molecular como heterociclos (Messens y De, 2002), compuestos con actividad antifúngica como ciertos ácidos grasos (Corsetti et al., 1996), ácido fenilactico (Lavermicocca et al., 2000) y/o bacteriocinas (Vuyst y Vandamme, 1994).

Particularmente las bacteriocinas han sido de especial interés en la búsqueda de compuestos antimicrobianos en seguridad alimentaria ya que entre otras razones, pueden ser sintetizadas por organismos considerados como



seguros para su uso en alimentos. Se estima que cerca del 99% de las bacterias conocidas son capaces de producir al menos una bacteriocina cuyo mecanismo de acción es comúnmente la formación de poros en la membrana o la inhibición de la síntesis de pared celular en especies cercanas a la de la bacteria productora aunque también hay ejemplos de bacteriocinas de amplio espectro (Snyder y Worobo, 2014).

En bacterias Gram negativas, las bacteriocinas se clasifican de acuerdo a su peso molecular o masa, las colisinas tienen una masa de entre 28 y 50 kDa mientras que la de las microcinas es de menos de 10 kDa. Éstas tienen diversos mecanismos de acción, pueden formar poros, degradar el peptidoglicano o ya dentro de la célula, digerir el ADN o ARN. Las microcinas son producidas principalmente por bacterias de la familia de las *Enterobacteriaceae*, son tolerantes a altas temperaturas, pHs extremos y a las proteasas.

En Gram positivas las LAB son productoras típicas de bacteriocinas, que se definen como péptidos antimicrobianos de bajo peso molecular con menos de 60 aminoácidos. Su clasificación ha sido problemática debido a la gran variedad de estructuras y funciones que pueden tener, pero las más aceptadas son las de Klaenhammer, (1993) y Cotter et al., (2005) que a continuación se resumen.

Bacteriocinas clase I ó lantibióticos. Son péptidos de menos de 28 aminoácidos (<5 kDa) modificados post-traduccionalmente, lineales o globulares que contienen lantionina,  $\beta$ -metil lantionina, y/o aminoácidos deshidratados (Yang et al., 2014). Dichas modificaciones resultan de la condensación de residuos de serina o treonina (Snyder y Worobo, 2014). Un ejemplo importante de este tipo de bacteriocinas es la nisina.

Bacteriocinas clase II o no lantibióticos tipo pediocinas. Péptidos de entre 30–50 aminoácidos (<10 kDa), termoestables y con carga neta positiva. Se reconocen varias subclases: clase IIa, péptidos con actividad antilisterial que conservan la secuencia consenso YGNGV/LxxC en el N-terminal y dos residuos de cisteína unidos por un puente disulfuro. Son de importancia para su uso en alimentos debido a su reconocida actividad antilisterial y a que son de amplio espectro, la pediocina PA-1 es un ejemplo de bacteriocinas de este tipo. Clase IIb, son bacteriocinas conformadas por dos péptidos que juntos forman un complejo de poración para iones monovalentes como  $\text{Na}^+$  y  $\text{K}^+$  en la membrana de la célula blanco, suelen tener el motivo GxxxG, en este grupo se incluyen a las  $\alpha$ -enterocinas y a lactococcina G. Clase IIc, bacteriocinas circulares que tienen un amplio rango de efectos en la permeabilidad de la membrana, la AS-48 es un ejemplo de este grupo. Clase IID, bacteriocinas lineales de un sólo péptido que no corresponden al tipo pediocinas, de amplio espectro antimicrobiano contra bacterias Gram positivas y negativas, ejemplos son las lactococcinas A y B. Adicionalmente, las lactococcinas forman parte de una familia pequeña de bacteriocinas que llevan un motivo distintivo de corte GG en el carboxilo terminal.

Bacteriocinas clase III o inestables al calor. Péptidos grandes (>30 kDa) que pueden actuar lisando la pared celular como la lisostafina o perturbando el potencial de membrana.

Bacteriocinas clase IV o complejas. Son péptidos que requieren modificaciones de naturaleza no proteínica para su

actividad. El péptido puede estar unido a carbohidratos o lípidos (Cotter et al., 2005). Dos ejemplos de esta clase son la sublancina (Oman et al., 2011) y la glicocina F (Brimble et al., 2015).

Por lo general, las bacteriocinas y los genes relacionados con la maquinaria de expresión y mecanismo de acción, se encuentran codificados en clusters policistrónicos que incluyen al gen estructural y alguna combinación de genes de inmunidad, secreción, regulación, factores de inducción y enzimas de modificación (Snyder y Worobo, 2014). La estructura del cluster es compleja y no corresponde técnicamente a la de un operón ya que los genes auxiliares no necesariamente se cotranscriben e incluso pueden estar codificados en cadenas complementarias y en distintos loci con respecto al gen estructural. Dichos genes pueden o no estar bajo la regulación transcripcional de un regulón.

La producción de bacteriocinas en ambientes nativos generalmente ocurre en niveles bajos e involucra interacciones con otras bacterias cercanas, es decir, está sujeta al fenómeno de *quorum sensing* y la expresión ocurre sólo cuando hay cierta densidad de células en el ambiente, durante la fase lag tardía de crecimiento (Snyder y Worobo, 2014; Yang et al., 2014).

#### 2.4 Nuevos enfoques en el análisis de comunidades microbianas

El análisis de comunidades microbianas presentes en alimentos con un enfoque ecológico, constituye la base tanto para el control de microorganismos indeseables como para el uso racional de los mismos en procesos de fermentación (Díaz y Wachter, 2003).

La necesidad de cultivar y aislar a los microorganismos para su identificación por métodos tradicionales ha limitado la comprensión de la diversidad microbiana, puesto que se estima que más del 90% de los que existen en ambientes naturales no pueden ser cultivados con los métodos estándar disponibles actualmente, además de la imposibilidad de obtener ciertas cepas puras ya que su crecimiento depende de las actividades metabólicas de otros o porque no se conocen sus condiciones de cultivo (Amann et al., 1995).

El estudio de la ecología de comunidades microbianas se remonta a los experimentos de Winogradsky en 1880 que fueron un parteaguas en el concepto de nicho ecológico y metabolismo de consorcios microbianos (Ackert, 2012). Otro punto crítico fue la implementación del uso de genes ribosomales como marcadores filogenéticos (Woese y Fox, 1977) y la secuenciación tipo Sanger (Sanger et al., 1977) lo que revolucionó el tema de la clasificación filogenética hasta entonces basada en características morfológicas y metabólicas de los microorganismos. A partir de estos avances y el consecuente diseño de cebadores universales contra regiones variables del ADNr 16S para ser usados en el estudio de mezclas de microorganismos (Pace et al., 1985), es que se abrió el panorama sobre una gran diversidad de secuencias pertenecientes a grupos taxonómicos nunca antes descritos (Giovannoni et al., 1990).

El desarrollo de técnicas moleculares basadas en la amplificación y/o análisis de secuencias que contienen

información filogenética tales como las ya mencionadas PCR-DGGE, FISH y RFLP (polimorfismo de longitud de fragmento de restricción) entre otras, permitieron la detección rápida y sensible de los microorganismos independientemente de que fueran o no cultivables, aunque más tarde fueron identificados ciertos sesgos intrínsecos como la amplificación preferencial de algunas secuencias sobre otras, la formación de productos quiméricos, problemas con impurezas en la muestra que merman en la eficiencia de las enzimas y visualización de los resultados, etc (Amann et al., 1995; Moter y Göbel, 2000).

Adicionalmente, la técnica de DGGE usada frecuentemente para separar amplicones con la misma longitud pero distinta secuencia, tiene sus propios sesgos y complicaciones, tales como la microheterogeneidad de secuencias de genes marcadores multicopia y la comigración de bandas (Ercolini, 2004). Además, no todas las bandas se cortan y purifican para mandar a secuenciar, lo que deja fuera del análisis a cierto sector de la población microbiana.

Una alternativa a la amplificación de marcadores moleculares es la extracción del ADN total de la muestra, su fraccionamiento y secuenciación, lo que se conoce como análisis metagenómico (Handelsman et al., 1998). Dicho enfoque se ha visto beneficiado por los avances en los métodos de secuenciación y con ello el crecimiento exponencial de las bases de datos y el desarrollo de herramientas computacionales.

En lo que a alimentos respecta, existen pocos reportes sobre análisis metagenómicos en quesos elaborados tradicionalmente. En uno de ellos se utilizó la metagenómica como herramienta en la búsqueda de funciones putativas relacionadas con la producción de compuestos volátiles en comunidades compuestas por hongos y bacterias en 137 muestras provenientes de las superficies de tres distintos tipos de quesos madurados naturalmente (Wolfe et al., 2014); mientras que en otro, se usó la metagenómica para monitorear los cambios de abundancia de especies y funciones durante la maduración de quesos inoculados superficialmente con nueve géneros de hongos y bacterias. Dicho estudio se complementó con el análisis de resultados de metatranscriptoma para identificar las actividades de la fracción metabólicamente activa (Dugat-Bony et al., 2015).

El éxito de estos y otros análisis metagenómicos en general, está relacionado con la disponibilidad de genomas de referencia y bases de datos de genes marcadores y funciones bien anotadas. En la actualidad hay disponibles más de 90,000 genomas secuenciados y ensamblados en más o menos fragmentos correspondientes a ~4,100 especies procariontes de acuerdo a la base de datos de genomas del NCBI (información obtenida de la tabla de genomas disponible en <http://www.ncbi.nlm.nih.gov/genome/browse/reference/>); mientras que la base de datos de GOLD (*Genomes OnLine Database*), reporta en su lista de organismos 55,908 genomas bacterianos y 1,231 para *Archaeas* tanto de organismos cultivables como no cultivables (información obtenida de: <https://gold.jgi.doe.gov/>).

## 2.5. Tecnologías de secuenciación

Como ya se mencionó, el desarrollo de la secuenciación de Sanger significó uno de los avances más importantes

en la investigación biológica y en particular, en la identificación de microorganismos. Los principios de la replicación del ADN fueron utilizados por Sanger et al. en 1977 en el desarrollo del proceso ahora conocido como la secuenciación de Sanger dideoxi. En este proceso la ADN polimerasa adiciona 2',3'-didesoxinucleótidos en la síntesis de una cadena de ADN nueva a partir del molde. Se llevan a cabo cuatro reacciones independientes donde cada una contiene cualquiera de los didesoxinucleótidos marcados ya sea radiactiva o fluorescentemente y una mezcla de los cuatro desoxinucleótidos sin modificar. Cuando un didesoxinucleótido (ddA, ddC, ddG, o ddT) se añade al extremo 3', la extensión de cadena termina, lo que permite obtener la formación de productos de extensión de todas las longitudes posibles terminados con dideoxinucleótidos en el extremo 3' (Figura 4).

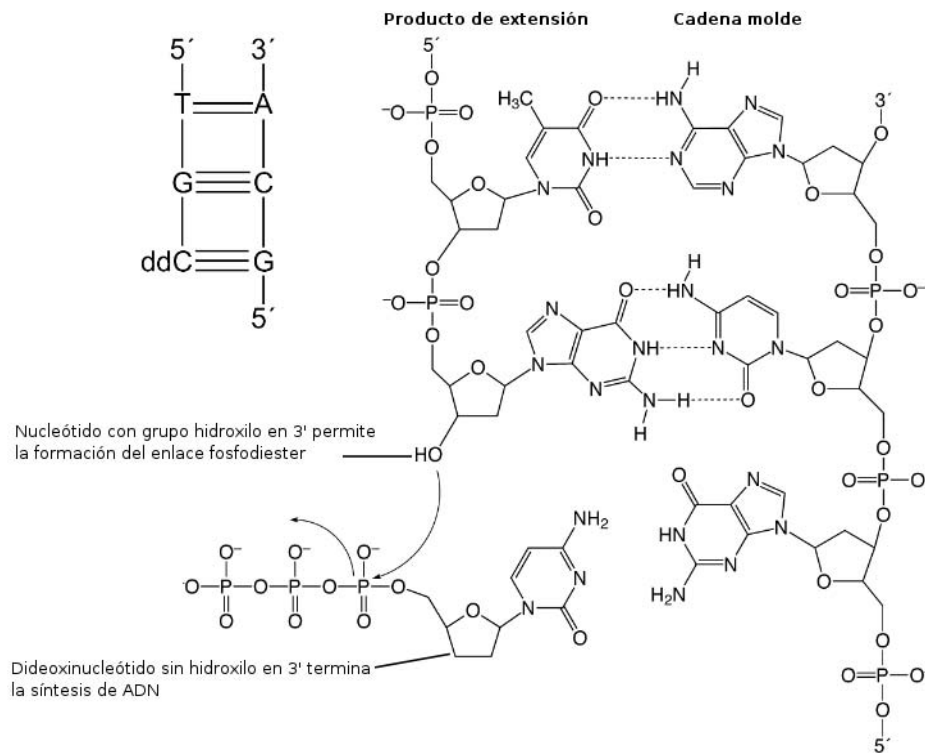


Figura 4. Fundamento de la química de secuenciación dideoxi de Sanger. Modificado de <http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/dna-sequencing-fragment-analysis/overview-of-dna-sequencing/faq.html>

Los productos de extensión se separan por electroforesis capilar con una resolución de una sola base. La técnica permite obtener lecturas de alrededor de 650 bp y es posible paralelizar las reacciones hasta 96 secuencias por corrida (Glenn, 2014).

A partir de este antecedente, se desarrollaron nuevas químicas de secuenciación basadas en el mismo principio de replicación del ADN, donde los principales esfuerzos se dirigieron a la miniaturización de los equipos con el fin de llevar a cabo la secuenciación de una gran cantidad de moléculas de ADN simultáneamente, de donde se desprende el concepto de secuenciación masiva.

La primera tecnología de este tipo la liberó Roche en el año 2005, y se conoce como 454 o pirosecuenciación. La

química se basa en el uso de una serie de enzimas que actúan de forma consecutiva para detectar el pirofosfato (PPi) liberado cada vez que la polimerasa incorpora un nucleótido a la cadena recién sintetizada de ADN. Dicho PPi es sustrato para la síntesis de un ATP que es utilizado por una luciferasa para generar una señal luminosa que el equipo es capaz de detectar. La intensidad de la señal registrada es proporcional al número de nucleótidos incorporados en cada ciclo (Margulies et al., 2005). Esta tecnología genera lecturas de hasta 650 bp y reporta hasta 1 millón de lecturas por corrida, pero tiene problemas de errores tipo inserciones y deleciones en una de cada 100 bases sobre todo cuando hay homopolímeros en la cadena molde (Glenn, 2014).

El homólogo de esta tecnología es IonTorrent, conocida como el potenciómetro más pequeño que existe, presenta rendimiento y tamaño de lectura similares a los de 454, pero se basa en la detección del cambio de pH debido a la liberación de un protón que ocurre cada vez que la polimerasa añade un nucleótido a la cadena recién sintetizada (Rothberg et al., 2011). Dicho método de detección es mucho menos costoso que el de la pirosecuenciación lo que la convierte en un claro prospecto para sustituirla una vez que Roche deje de darle soporte a partir del año 2016. La longitud de lectura es de hasta 175 bp y puede generar hasta 500 millones de lecturas por corrida con una tasa de error por inserciones o deleciones (indeles) en una de cada cien bases (Glenn, 2014).

Otra tecnología que ha ganado popularidad en análisis metagenómicos debido a su gran rendimiento y baja tasa de error es la de Illumina. La química de esta tecnología es la que más se parece a la de Sanger, la cadena nueva de ADN se sintetiza a partir de nucleótidos modificados que están marcados con un fluoróforo y bloqueados en el OH 3', aunque esta terminación es reversible de tal modo que en cada ciclo la etiqueta es eliminada y la adición de la siguiente base es posible. La ventaja de esta tecnología sobre las ya mencionadas es que la técnica de formación de clusters sobre el soporte sólido, hace posible la secuenciación desde ambos extremos de la cadena molde, lo que permite la obtención de lecturas pareadas donde, si se conoce la longitud del templado original, es posible calcular la distancia entre ambas lecturas, información que facilita su ensamble posteriormente (Bennett, 2004). La longitud de secuencia y el rendimiento dependen del equipo que se utilice pero va desde 22 millones de lecturas por corrida de 300 bp en el MiSeq v3, hasta 6000 millones de lecturas con una longitud de 150 bp en un equipo HiSeqX (con 2 celdas de flujo). Esta tecnología presenta errores de sustitución en una de cada 1000 bases secuenciadas (Glenn, 2014).

Además de las tecnologías descritas, hay otras que no se utilizan tanto en metagenómica, sino en resecuenciación de genomas o en la búsqueda de polimorfismos como en el caso de la química de SOLiD de Applied Biosystems (Shendure et al., 2005).

La búsqueda de la síntesis de cadenas cada vez más largas ha llevado al desarrollo de las llamadas tecnologías de tercera generación, tal es el caso de PacBio y Nanopore. La química de la primera de ellas se basa en la adición de nucleótidos marcados con diferentes fluorocromos, mismo que es escindido en tiempo real por la polimerasa que está fija en el fondo de un nanoposillo, desde donde se toman lecturas de fluorescencia por microscopía confocal (Fichot y Norman, 2013). La tecnología de PacBio permite obtener hasta 30 mil lecturas por corrida de 3 kb, pero

su tasa de error es de alrededor del 13%, más comunmente de indeles (Glenn, 2014). Por otro lado, Oxford Nanopore es la primera tecnología que no se basa en el proceso de replicación del ADN, sino en sensar el cambio de potencial de membrana que ocurre cuando una molécula de un tamaño dado pasa a través de un poro protéico (Kasianowicz et al., 1996). La tecnología permite obtener hasta 10 millones de lecturas con una longitud máxima de 10 kb y una tasa de error del 4% donde lo más común es encontrar deleciones (Glenn, 2014).

## 2.6. Análisis bioinformático de los datos

Dependiendo de cada tecnología, el formato del archivo de salida puede variar, pero siempre es posible convertirlo a formato fastq que es un tipo de texto plano que contiene información de cada una de las secuencias con su identificador y la calidad asociada a cada nucleótido.

La limpieza de adaptadores, secuencias contaminantes y el análisis de calidad es el primer paso a realizar ante cualquier proyecto de secuenciación. Además, es opcional hacer un paso de “trimming” o corte sobre las últimas bases de las lecturas cuando la calidad decae a niveles no aceptables, por ejemplo, a menos de 20 en score Phred definido como  $-10 \log(\text{probabilidad de error})$ . Es importante mencionar que un filtrado demasiado estricto podría llevar a la pérdida de información de secuencias con baja cobertura (Liu et al., 2012).

A partir de las lecturas que se obtienen de la secuenciación de ADN metagenómico, es posible la extracción de información tanto taxonómica mediante el análisis de genes que funcionan como marcadores filogenéticos, como del potencial metabólico por la predicción y anotación de secuencias codificantes. Los pasos del análisis pueden ser llevados a cabo en distinto orden al propuesto aquí e incluso algunos pasos podrían omitirse y/o añadirse otros.

### Obtención de información taxonómica a partir de secuencias crudas

Los métodos de anotación taxonómica pueden ser clasificados básicamente en dos grandes grupos: modos basados en similitud y métodos basados en composición.

Los métodos basados en similitud implican la extracción de secuencias de genes marcadores ya sea por mapeo directo de las lecturas contra bases de datos previamente construidas con identidad taxonómica conocida o mediante la predicción de las mismas usando principios matemáticos como los Modelos Ocultos de Markov (HMM por sus siglas en inglés).

Un modelo de Markov es un sistema que produce una cadena de Markov, y un modelo oculto de Markov es aquel donde las reglas para producir la cadena son desconocidas u ocultas. Dichas reglas incluyen dos probabilidades, la primera es que habrá cierta observación, y la segunda es que ocurrirá cierta transición entre estados dado el estado del modelo en un punto dado. Este método matemático permite solucionar cierto tipo de problemas: 1) dado el modelo, encontrar la probabilidad de las observaciones; 2) dado el modelo y las observaciones, encontrar la trayectoria de transición más probable entre estados; y 3) maximizar ya sea 1 ó 2 ajustando los parámetros del

modelo. Para la construcción de un HMM se requiere de los siguientes elementos: un alfabeto finito que define los estados observables, una colección finita de estados ocultos en el modelo, la probabilidad de observar el estado 2 dado el estado 1 y por último, las probabilidades de que ocurra la transición entre un estado oculto y otro en un solo paso (Rabiner y Juang, 1986).

Los HMM son efectivos modelando correlaciones entre símbolos, dominios o eventos por lo que han tenido éxito en la resolución de problemas biológicos como la predicción de genes, alineamiento de secuencias y predicción de estructura secundaria de proteínas (Yoon, 2009). Tomando como ejemplo la aplicación de los HMM en un alineamiento múltiple de secuencias, los perfiles del modelo convierten cada posición en un sistema de puntaje posición-específico adecuado para la búsqueda en bases de datos de homólogos. Para cada columna del consenso generado a partir del alineamiento múltiple de familias de secuencias, los estados del modelo son; coincidencias o *matches*, inserciones y deleciones. Un estado de coincidencia modela la distribución de residuos a través de esa posición, mientras que un estado de inserción o deleción, permite agregar o eliminar uno o más elementos en la columna dada y en la siguiente (Eddy, 1998).

Una vez identificadas y separadas las secuencias de interés en el metagenoma, éstas se agrupan en unidades taxonómicas operacionales (OTUs por sus siglas en inglés) y la taxonomía se asigna a cada grupo ya sea directamente por similitud con el mejor *hit* de la referencia o al ubicar su último ancestro común (LCA por sus siglas en inglés) cuando se toma en cuenta más de un resultado para cada OTU (Dröge y McHardy, 2012). En el método de LCA los resultados del mapeo se ponderan de acuerdo a un score dado, por ejemplo, por el porcentaje de similitud entre la secuencia problema y las referencias contra las que se haya encontrado un *match*. Los resultados se ubican dentro de un árbol filogenético de especies y en los casos donde los resultados son ambiguos y apuntan a distintos nodos del árbol, la identidad se asigna al nivel que resulte ser común para todas las coincidencias (Huson et al., 2007).

Se asume que la abundancia de secuencias de los genes marcadores se correlaciona directamente con la abundancia de los genomas de los que provienen, lo que se puede interpretar como abundancia de taxa en la comunidad ya sea de forma directa si no hay redundancia del marcador o pasando por una normalización cuando el mismo se encuentra en multicopia.

Por otro lado, dentro de los métodos basados en composición se incluyen aquellos que son independientes de la predicción de marcadores filogenéticos, por lo que se basan en características observables directamente sobre las secuencias, como por ejemplo, el contenido de GC. Los métodos más populares de asignación taxonómica basada en composición de secuencias, consisten en la clasificación de las lecturas con base en firmas de k-meros (secuencia con longitud “k” de bases, donde k es cualquier entero positivo de bases consecutivas) observables en genomas de referencia y comparando su frecuencia en las lecturas metagenómicas. Los métodos pueden tomar en cuenta ya sea sólo la presencia de k-meros o también su abundancia como evidencia de un origen taxonómico común entre la referencia y la secuencia problema (Kawulok and Deorowicz, 2015).

Muchos métodos de análisis bioinformático utilizan enfoques híbridos de las estrategias aquí descritas, por ejemplo en un primer paso pueden usar los espectros de k-meros para separar las lecturas en paquetes y después hacen la búsqueda de genes marcadores para darle identidad taxonómica a cada uno, esperando tener una sola especie en cada grupo de secuencias.

**Anotación funcional: ensamble, predicción de genes y asignación de función**

El ensamble se basa en el empalme de diferentes fragmentos de ADN que permitirán alinear los extremos similares y así reconstruir secuencias lineales completas conocidas como contigs.

Algunos problemas que podrían afectar la calidad del ensamble son regiones de baja cobertura, mala calidad de las secuencias, la longitud de las mismas en caso de que sean muy cortas y la existencia de secuencias repetidas en los genomas. Las lecturas que contienen el mismo ADN repetitivo ensamblaran entre si aunque provengan de regiones distintas del genoma, lo que ocurre por ejemplo, entre regiones conservadas de operones ribosomales.

En general, la cobertura del ensamble va de la mano con la profundidad de la secuenciación y la complejidad de la comunidad. Para el caso de las especies menos abundantes, menos copias de sus genomas habrán sido secuenciados y menos probable será ensamblar secuencias largas (Luo et al., 2012).

Una forma de abordar el problema de las regiones repetidas es el uso de lecturas pareadas o “mate pairs”, lo que facilita el ensamble cuando el tamaño del inserto es mayor que la region repetida, de modo que sabemos que esas dos lecturas van juntas a una distancia igual al tamaño del fragmento.

Actualmente se han desarrollado ensambladores específicos para la reconstrucción de fragmentos metagenómicos a partir de lecturas cortas. Éstos funcionan mediante la construcción de gráficas de Bruijn, cuyo fundamento se basa en encontrar la secuencia de letras que pudiera contener todas las posibles subcadenas de una longitud “k” dada (k-meros) a través de un alfabeto (Compeau et al., 2011). En la Figura 5 se muestra un ejemplo del método aplicado al ensamble de secuencias cortas.

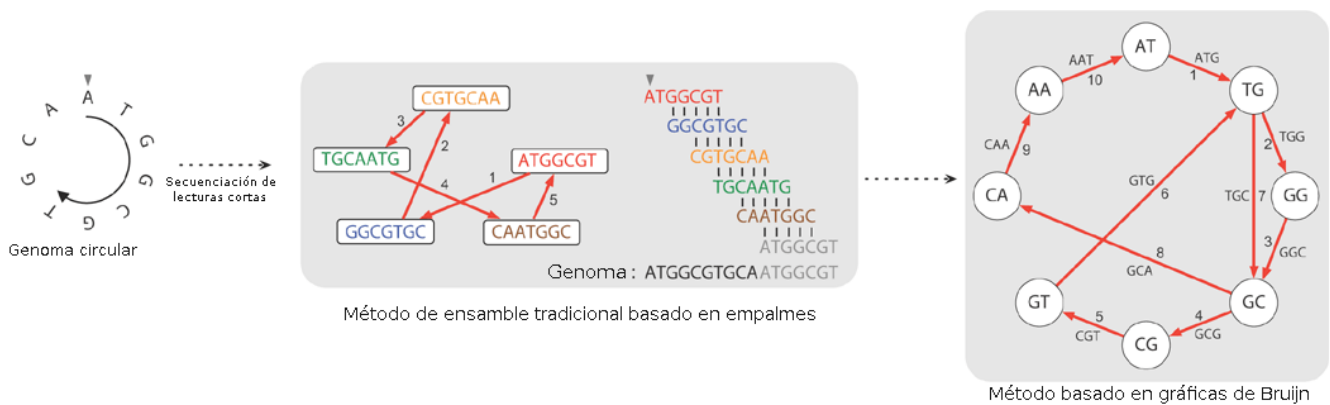


Figura 5. Representación de los algoritmos de ensamble de lecturas de ADN. Modificado de Compeau et al., 2011.



Intuitivamente, las lecturas con alta similitud de secuencia deben compartir k-meros en las regiones superpuestas. Dichos k-meros son generalmente más fáciles de encontrar que los solapamientos, por eso con esta estrategia se optimiza el tiempo y el costo computacional del ensamble. La elección del tamaño de k-mero es importante para generar el mejor ensamble posible. Éste debe ser lo suficientemente grande como para evitar solapamientos azarosos y lo suficientemente pequeño como para que ocurran la mayoría de los solapamientos verdaderos.

La calidad del ensamble se mide en función del tamaño y número de contigs obtenidos. Generalmente se toma en cuenta en las estadísticas la longitud máxima, la longitud promedio, el total de bases ensambladas, y el N50 o N90. El cálculo de N50 se obtiene ordenando los contigs de mayor a menor tamaño y se suman sus longitudes por ese orden hasta alcanzar el 50% del tamaño del genoma o del total de bases ensambladas en el metagenoma. La longitud del último contig añadido es el valor N50 (Miller et al., 2010).

Los programas diseñados para hacer la predicción de genes en secuencias metagenómicas, se basan en la aplicación de modelos heurísticos por métodos de cadenas de Markov generados a partir de parámetros observados en genomas de referencia. Estos algoritmos utilizan las características de las secuencias codificantes como uso preferencial de codones, frecuencia de aparición de dicodones, patrones en el uso de codones de inicio y paro, contenido de GC y tamaño de marcos de lectura abiertos (ORFs por sus siglas en inglés) así como patrones de secuencias identificadas como de unión al ribosoma para predecir posibles genes en la secuencia problema (Liu et al., 2013).

Una vez que se ha generado el archivo de posibles secuencias codificantes, el siguiente paso es la asignación de función. Esto se puede llevar a cabo en secuencias de nucleótidos, o preferentemente, en secuencias de aminoácidos.

Una forma de asignar función basada en homología, es por alineamientos locales usando casi siempre alguna versión de BLAST, contra bases de datos de secuencias de proteínas bien curadas como por ejemplo la de Swiss-Prot (Boeckmann et al., 2005, disponible por: [http://web.expasy.org/docs/swiss-prot\\_guideline.html](http://web.expasy.org/docs/swiss-prot_guideline.html)). Dicho método es computacionalmente costoso dependiendo el tamaño de la base de datos y el número de secuencias problema. Otro método cuya ventaja es que se puede aplicar cuando los porcentajes de identidad contra las secuencias en bases de datos de referencia como NR son bajos, es la predicción de dominios conservados de proteínas mediante la extrapolación de los ya mencionados HMM (Eddy, 2011) contra bases de datos como la de Pfam (Finn et al., 2016, disponible por: <http://pfam.xfam.org/>). Este enfoque permite la predicción de proteínas nuevas con dominios conocidos. Adicionalmente se pueden buscar características específicas sobre las secuencias predichas como péptido señal o dominios transmembranales, información que puede ser de utilidad en el momento de darle sentido biológico a los resultados. El análisis se puede complementar con el uso de herramientas para la predicción de genes de secuencias no codificantes como por ejemplo los CRISPRs o clustered regularly interspaced short palindromic repeats (Rho et al., 2012).

De cualquier forma, es importante resaltar que se ha observado en secuencias de ORF predichas a partir de datos

metagenómicos que sólo del 20 al 50% de las mismas pueden ser anotadas contra bases de datos como SEED, en sistemas taxonómicamente muy diversos tales como muestras de aguas marinas superficiales (Gilbert et al., 2010), dejando la pregunta inmediata de cuál es la importancia y la función de los genes sin anotación denominados como ORFans.

Dichos ORFans son responsables de la aparentemente interminable novedad genética en los metagenomas microbianos. Existen tres hipótesis sobre la existencia de esta fracción desconocida, una de ellas es que los ORFans podrían simplemente reflejar secuencias no codificantes predichas erróneamente debido a los algoritmos de detección imperfectos; otra explicación es que los ORFans son genes reales, pero codifican para funciones bioquímicas desconocidas; y la última es que aunque los genes sin anotación no tienen una identidad de secuencia lo suficientemente alta con genes conocidos para considerarlos homólogos, sí podrían tener similitud estructural con proteínas descritas previamente, por lo que representan familias de proteínas con función y estructura tridimensional conocida pero con secuencia de aminoácidos muy diferente a lo que hay en las bases de datos (Thomas et al., 2012).

### 3. Justificación

Los alimentos fermentados en general y los quesos elaborados artesanalmente en particular, son un reservorio de microorganismos no siempre estudiados que poseen actividades metabólicas interesantes desde el punto de vista alimentario. El análisis bioinformático del metagenoma del queso Cotija permitirá la caracterización del consorcio bacteriano presente en una muestra representativa con tres meses de maduración, tiempo que los productores consideran como mínimo para la venta del producto; así como la detección de funciones metabólicas relacionadas con la producción de compuestos de olor y sabor y otras relacionadas con su inocuidad.

En la actualidad existen pocos reportes de análisis metagenómico en alimentos de fermentación espontánea y muy pocos sobre quesos madurados, por lo que los resultados que se obtengan serán una aportación relevante en el campo de la biotecnología y microbiología de alimentos.

### 4. Hipótesis

Se espera que el queso Cotija artesanal madurado Región de Origen, por poseer características de potencial redox negativo, pH y actividad acuosa bajos, tenga una población bacteriana de baja complejidad taxonómica, misma que podrá ser descrita en su totalidad con la metodología propuesta y que estará conformada principalmente por miembros del *Phylum Firmicutes*.

A nivel de perfil metabólico, se espera que la batería de genes codificantes de la comunidad microbiana esté relacionada con las características del nicho del que proviene.

### 5. Objetivos

#### 5.1. Objetivo general

Llevar a cabo la caracterización taxonómica y del potencial metabólico de la microbiota bacteriana presente en una muestra representativa de queso Cotija que cumple con el tiempo mínimo de maduración (3 meses) mediante el análisis bioinformático del metagenoma.

#### 5.2. Objetivos particulares

Extraer del queso Cotija ADN metagenómico de buena calidad útil para secuenciación masiva.

Describir la diversidad bacteriana y evaluar la presencia de bacterias coliformes y patógenos a partir del resultado de riqueza y abundancia de especies bacterianas con base en diferentes marcadores filogenéticos.

Reconstruir los genomas de las especies dominantes a nivel de *draft*.

Buscar genes relacionados con la producción de compuestos de olor y sabor en quesos, así como de genes codificantes para bacteriocinas como parte de la descripción del perfil metabólico de la comunidad mediante la anotación funcional de marcos de lectura abiertos.

## 6. Diagrama Experimental

Para alcanzar los objetivos planteados, se propone la metodología que se ilustra en la Figura 6. En cada paso se utilizaron herramientas bioinformáticas específicas que se detallan en la descripción del análisis bioinformático de secuencias en la sección de metodologías.

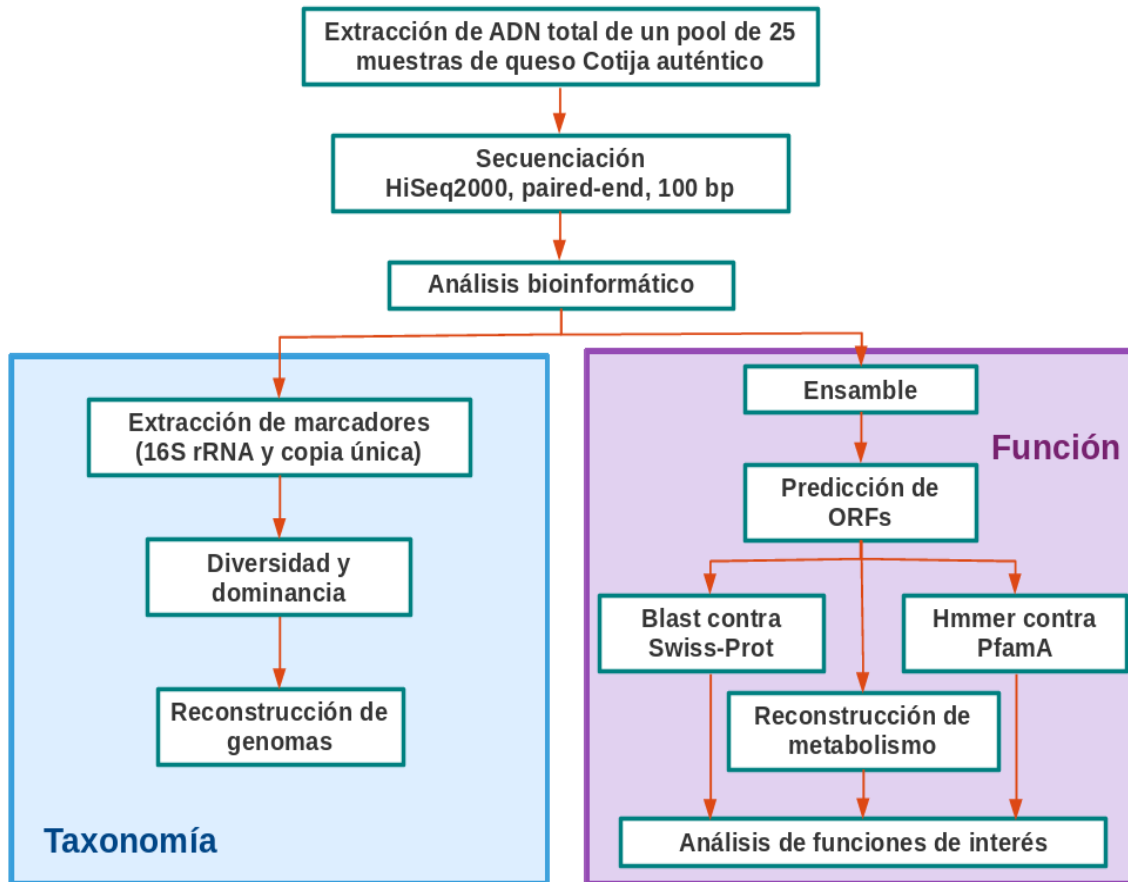


Figura 6. Diagrama general de la estrategia de análisis del metagenoma bacteriano de queso Cotija.

## 7. Metodologías

### 7.1. Muestras de queso Cotija Región de Origen

Se tomaron 25 muestras de queso Cotija provenientes de alguno de los municipios pertenecientes a la zona productora (Figura 1). La distribución geográfica de la procedencia de las muestras es la siguiente: 5 del municipio de Tocumbo, 8 de Cotija, 4 de Quitupan, 4 de Jilotlan y 4 de Santa María del Oro. Todas ellas cumplen con las especificaciones de la norma de queso Cotija (NMX-F-735-COFOCALEC, 2011) y con la cuenta de coliformes fecales establecida en la NOM-243-SSA1, 2010. Se pesaron 100 g de cada uno de los quesos homogeneizados a un tamaño de partícula de 2 mm y se hizo una mezcla que fue almacenada en una bolsa plástica de cierre hermético en refrigeración hasta su uso.

## 7.2. Extracción de paquete celular (Zúñiga, 2009) y de ADN total (Ampe et al., 1999; Yuan et al., 2012)

Se pesaron 15 g de la mezcla de quesos en la campana de flujo laminar dentro de bolsas para Stomacher (Seward) con 40 mL de citrato de sodio al 2% p/v, pH 8.0, previamente estéril. La mezcla se homogenizó en Stomacher 400 Circulator (Seward) durante 2 min a alta velocidad para llevar a cabo la ruptura mecánica de la matriz del queso. A continuación se adicionó 1 mL de solución de Neutrasas (Novo Nordisk) con el fin de romper la red proteínica y se mezcló a alta velocidad en Stomacher durante 5 min. La bolsa se incubó a 45°C durante 1 h en agitación a 100 rpm y al finalizar la incubación la mezcla se dividió en dos tubos Falcon estériles de 50 mL y se centrifugó a 3,500 rpm (2,160 X g) durante 7 min a 4°C para separar las fracciones de partículas mayores y grasa de las células de microorganismos que se encontraban suspendidas en la fase acuosa.

La fase intermedia (acuosa) se transfirió a un tubo de 50 mL estéril con ayuda de una micropipeta de 5 mL y se se centrifugó a 12,000 rpm (22,050 X g) durante 10 min a 4°C. El sobrenadante se desechó y el pellet se lavó 3 veces con solución salina al 0.85% p/v, pH 7.0, estéril. En el pellet se encontraba el paquete de microorganismos que se pudieron separar de los 15 g de queso.

El pellet lavado se resuspendió en 1 mL de solución salina en un tubo Eppendorf de 2 mL y se centrifugó a 12,000 rpm (13,300 X g) durante 5 min. El sobrenadante se eliminó y cada paquete celular se resuspendió en 1 mL de buffer para DNasa I (100 mM Tris pH 7.5, 25 mM de MgCl<sub>2</sub> y 5 mM de CaCl<sub>2</sub>). Se adicionó a cada tubo 1 U de Dnasa I con el fin de eliminar el ADN contaminante (Willner et al., 2009) de vaca. Las mezclas se incubaron durante 10 minutos en termomixer con agitación a 300 rpm de acuerdo a los resultados que se reportan en el Anexo 3. Al término de la incubación, la DNasa I se inactivó con un tratamiento de 10 minutos a 75°C y los tubos se centrifugaron a 14,000 rpm (18,200 X g) durante 10 minutos en microcentrífuga para después congelarlos a -20°C toda la noche.

Cada pellet se resuspendió en 450 µL de buffer TE y se adicionaron 50 µL de una solución de lisozima de 10 mg/mL (en buffer 100 mM Tris-HCl y 50 mM de EDTA, pH 8.0) y 10 µL de una solución de mutanolisina a una concentración de 1U/ µL (en buffer TE). Las muestras fueron agitadas en vortex durante 1 min e incubadas durante 1 h a 37°C en termomixer con agitación a 200 rpm. A cada tubo se adicionó la cantidad de RNasa A de páncreas de bovino necesaria para alcanzar una concentración final de 0.6 µg/mL y se incubó durante 30 min a 37°C. Terminado este tiempo, se adicionaron 50 µL de una solución 10 mg/mL de proteinasa K (en buffer 10 mM Tris-HCl, pH 7.5) a cada tubo y se incubaron durante 50 min a 50°C y luego por 10 min a 65°C.

La mezcla se dejó enfriar hasta 37°C y se adicionaron 50 µL de SDS al 10% p/v y se incubó a 37°C por 40 min más. La mezcla se dejó enfriar a temperatura ambiente y se agregaron 200 µL de una solución de NaCl 6M, mezclando lentamente por inversión. Se adicionaron 400 µL de una solución de fenol-cloroformo-isoamílico 25:24:1 en la campana de flujo laminar, la mezcla se agitó en vortex durante 30 s y se centrifugó a 14,000 rpm (18,200 X g) durante 10 min a 4°C para recuperar la fase acuosa que contenía el ADN. Ésta se transfirió a un tubo Eppendorf de 2 mL limpio y la extracción se repitió centrifugando hasta que la interfase orgánica fue nula.

Adicionalmente a la extracción se llevaron a cabo dos pasos de lavado con cloroformo para eliminar las trazas de fenol y por último se adicionó 1 mL de etanol al 100% v/v frío, se agitó por inversión de 2 a 5 veces y se incubó a -20°C durante mínimo una hora. La mezcla se centrifugó a 14,000 rpm (18,200 X g) durante 10 min en frío y el pellet se lavó con 300 µL de etanol al 75% v/v centrifugando 2 minutos en las mismas condiciones.

Se desechó el etanol y se evaporó el remanente en SpeedVac durante 15 minutos a 37°C. El pellet seco se resuspendió en 50 µL de agua grado biología molecular y los tubos se calentaron a 65°C durante 15 minutos para su total hidratación. Se llevó a cabo un paso extra de purificación pasando por columna del kit de purificación QIAquick de Qiagen siguiendo las especificaciones del fabricante.

Se midió concentración y pureza del ADN total por método espectrofotométrico (en equipo Epoch, BioTek) y en gel de agarosa al 1% y las extracciones se almacenaron en congelación a -20°C. La biblioteca se construyó en la unidad de secuenciación masiva y bioinformática del Instituto de Biotecnología (IBt) de la UNAM en Cuernavaca, Mor., con el kit TruSeq v2 (Illumina Inc.) y la secuenciación se llevó a cabo en el Laboratorio Nacional de Genómica y Diversidad (LANGEBIO) en Irapuato, Gto., en el equipo HiSeq2000 con configuración pareada de 100 ciclos.

### 7.3. Análisis bioinformático de secuencias

Las calidades de las secuencias crudas fueron visualizadas con el programa FastQC (disponible por: <http://www.bioinformatics.babraham.ac.uk>). Las secuencias crudas fueron mapeadas contra el genoma de *Bos taurus* con BWA v0.7.9 (Li y Durbin, 2009) con el fin de identificar contaminación por ADN de vaca en el resultado de la secuenciación.

#### Anotación taxonómica

La asignación de identidad taxonómica y el cálculo de abundancias se llevó a cabo mediante dos pipelines de acceso libre sobre las lecturas sin ensamblar. El primero de ellos es MOCAT v1.3 (Kultima et al., 2012), desarrollado en el European Molecular Biology Laboratory y el Beijing Genomics Institute. El software acepta como entrada un archivo en formato fastq con lecturas crudas. En un primer bloque, el pipeline hace el filtrado de las secuencias y las recorta por calidad con el software SolexaQA (score Phred<20) y usando código interno. Tiene la opción de hacer el escrutinio de secuencias contaminantes por ejemplo de humano o a partir de cualquier genoma de referencia correctamente indexado, así como la búsqueda y eliminación de adaptadores mediante alineamientos con SOAPaligner y USEARCH. En el segundo bloque, las lecturas filtradas son usadas en la búsqueda de secuencias de genes marcadores de copia única para lo que utiliza código interno y SOAPaligner con un valor de corte de 97% identidad y hasta 5 bases distintas contra la base de datos mOTU v1 que consiste en secuencias de genes marcadores extraídos de 3,496 genomas bacterianos completos procedentes del NCBI

(incluyendo RefMG v1 compuesta por genes extraídos de 1,753 genomas bacterianos de referencia) y 263 metagenomas disponibles públicamente de los proyectos de MetaHIT y HMP. Los perfiles obtenidos son abundancias de mOTUs y no todos pueden ser asignados taxonómicamente, sólo representan de forma robusta clusters de especies, que todavía no es información taxonómica. La abundancia de taxa y la asignación taxonómica a nivel de especie, se obtienen por alineamiento contra la base de datos RefMG. En un tercer bloque, el software tiene la opción de llevar a cabo el ensamble de lecturas con SOAPdenovo con un paso de revisión para eliminar indeles y quimeras con SOAPaligner y BWA, predecir secuencias codificantes ya sea con MetaGeneMark o Prodigal y extraer genes de copia única de marcadores individuales mediante fetchMG.

En el análisis del metagenoma de queso Cotija sólo utilizamos los dos primeros bloques de MOCAT, que incluyen el filtrado y recorte de las secuencias por calidad y la anotación taxonómica de las mismas.

La segunda herramienta es Parallel-Meta v2.4 (Su et al., 2014), que fue desarrollada por el Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences y fue diseñada para paralelizar el análisis masivo de datos metagenómicos en la obtención de información taxonómica y funcional. La anotación taxonómica se realiza por predicción de secuencias ribosomales mediante modelos ocultos de Markov e interrogando alguna de las bases de datos de genes ribosomales disponibles como GreenGenes (DeSantis et al., 2006), Ribosomal Database Project o RDP (Cole et al., 2014), SILVA (Quast et al., 2013) y una más destinada al análisis de metagenomas provenientes de cavidad oral, mediante el algoritmo BLAST con un valor esperado de corte predeterminado de  $1e-30$ . El programa incluye bases de datos. El pipeline tiene la opción de hacer el análisis taxonómico con el gen marcador 16S para procariontes o con el 18S para eucariontes; y la anotación funcional con los pasos de ensamble con Velvet, predicción de genes con FragGeneScan y la anotación mediante BLAST contra la base de datos de familias de proteínas organizadas en subsistemas del SEED. El software genera diagramas interactivos de abundancia y clasificación (conocidos como Kronas por el programa con el que se generan) de la información taxonómica y funcional, así como árboles filogenéticos. En el análisis de las secuencias del metagenoma de queso Cotija se utilizó únicamente el módulo de anotación taxonómica contra la base de datos del RDP y las búsquedas se realizaron con un valor de corte esperado (E-value) de  $1e-15$ .

A partir de las tablas de anotación obtenidas se construyó la curva de rarefacción usando el paquete vegan v2.3.0 (Oksanen et al., 2015) y se calcularon los estimadores de diversidad con el paquete phyloseq v1.12.2 (McMurdie y Holmes, 2013), disponibles por Bioconductor, de R.

Todos los programas fueron instalados en el servidor de la Unidad de Secuenciación Masiva y Bioinformática del Instituto de Biotecnología, UNAM, campus Cuernavaca, Mor.

#### Anotación funcional

El análisis funcional se realizó con una propuesta nueva que requiere de varios pasos independientes. El primero

de ellos fue el ensamble de las lecturas crudas sin filtrar con el software diseñado para metagenomas IDBA-UD v1.1.1 (Peng et al., 2012). Esta herramienta es un ensamblador iterativo *de novo* de secuencias cortas que se basa en la construcción de gráficas de Bruijn (Figura 5). En cada iteración, los contigs cortos y de baja cobertura se eliminan con umbrales de corte cada vez más estrictos. La información que brinda el uso de lecturas pareadas se utiliza para hacer ensambles locales que permiten generar k-meros no observados en regiones de baja profundidad. Con este enfoque IDBA-UD minimiza los gaps o huecos en el alineamiento y la formación de burbujas cuando hay más de un posible camino a seguir durante el ensamble, lo que permite formar contigs más largos que con otros ensambladores. Con el fin de obtener una estadística sobre el número de lecturas utilizadas en el ensamble, se mapearon todas las lecturas sobre los contigs usando BWA v0.7.9.

La predicción de secuencias codificantes se realizó sobre el ensamble usando MetaGeneMark v2.1.0 (Zhu et al., 2010), y con las secuencias traducidas a aminoácidos se hizo la anotación por homología contra las bases de datos de Swiss-Prot y Pfam-A.

UniProtKB/Swiss-Prot es el componente anotado manualmente de UniProtKB. UniProt es una colaboración entre el European Bioinformatics Institute, el Swiss Institute of Bioinformatics y el Protein Information Resource. Esta base de datos contiene registros con información de función sustentada experimentalmente extraída de la literatura y curados mediante análisis computacional. Por otro lado, la base de datos de Pfam está construída a base de una gran colección de familias de proteínas curadas manualmente, cada una representada por alineamientos múltiples y perfiles de HMM, dichos perfiles pueden a su vez estar agrupados en clanes donde se representa más de una familia de secuencias en una superfamilia (Finn et al., 2014).

Dado que la búsqueda de similitud contra secuencias de proteínas mediante alineamientos locales es computacionalmente muy costosa, la anotación contra Swiss-Prot se llevó a cabo sobre paquetes de secuencias problema de máximo 10,000 aminoácidos. La primera serie de alineamientos se realizó con umbral de E-value=1e-5 con el algoritmo ublastp (Edgar, 2010) del paquete de USEARCH v7.0.1090, que es más rápido que Blast. Con las secuencias que quedaron sin anotar de este paso, se volvieron a generar paquetes del mismo tamaño y se corrieron alineamientos con el algoritmo blastp de BLAST v2.2.28+ (Camacho et al., 2009) contra la misma base de datos nuevamente con un valor de corte de E-value=1e-5.

A la par, se generaron resultados de búsquedas de dominios conservados mediante el algoritmo hmmerScan de HMMER v3.1b1 (Eddy, 2011) contra los perfiles de la base de datos Pfam-A (Finn et al., 2016). Los resultados se integraron adaptando el software Trinotate v2.0.1 (Grabherr et al., 2011), lo que permitió agregar información sobre los términos de GO e identificadores de grupos ortólogos de COG asociados a cada gen con anotación.

De forma independiente a esta anotación, la reconstrucción de rutas metabólicas del metagenoma se realizó mediante GhostKOALA v2.0 (Kanehisa et al., 2016, disponible en línea por: <http://www.kegg.jp/ghostkoala/>), que es una herramienta de anotación interna de la Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa y Goto, 2000). Ésta asigna identificadores “K” a la secuencia problema mediante el buscador GHOSTX contra una



base de datos de genes no redundante del KEGG.

La búsqueda de funciones de interés relacionadas con la producción de compuestos de olor y sabor se llevó a cabo mediante el escrutinio de enzimas específicas a partir del resultado de la anotación con Swiss-Prot así como de rutas metabólicas completamente reconstruidas reportadas en la literatura como responsables de la síntesis de compuestos que aportan al perfil sensorial en quesos.

En la búsqueda de bacteriocinas, se seleccionaron una serie de identificadores de dominios funcionales presentes en proteínas estructurales de bacteriocinas o relacionadas con su síntesis y sistema de inmunidad que protege a la célula productora. Dichos identificadores se utilizaron para hacer búsquedas en la tabla de anotación obtenida por HMMER. A partir de este resultado, se identificaron todos los contigs que contuvieran genes con la anotación de interés y se hizo la transferencia de anotación de los ORFs adyacentes considerando una ventana de diez genes río arriba y diez genes río abajo con el fin de identificar la mayor cantidad de secuencias codificantes para bacteriocinas aún cuando no se encontrara su anotación. Finalmente, cuando alguno de los genes dentro de la ventana o cluster no tuvo anotación de Pfam, se recurrió a la búsqueda de su resultado contra Swiss/Prot, y en caso de que no la hubiera se hizo una búsqueda nueva con el algoritmo BLAST contra la base de datos de NR (*non-redundant*) por ser una de las bases de datos más grandes de secuencias de genes. Este análisis se complementó con una búsqueda en las bases de datos de la herramienta BAGEL3 (van Heel et al., 2013, disponible en línea por: <http://bagel2.molgenrug.nl/index.php/bagel3>).

#### Reconstrucción de genomas

Se llevó a cabo la reconstrucción de los genomas de las especies más abundantes en el metagenoma con el pipeline ABACAS v1.3.1 (Assefa et al., 2009). Dicho software mapea los contigs metagenómicos contra cada genoma de referencia usando MUMmer y después les da orientación y genera una pseudomolécula que corresponde al genoma reconstruido. La visualización del genoma en una gráfica circular con respecto a su referencia se obtuvo con el software BRIG v0.95 (Alikhan et al., 2011).

Adicionalmente, se mapearon las lecturas crudas sobre cada referencia con el software BWA v0.7.9 con el fin de obtener un estimado de la cobertura para cada genoma. Los genomas de referencia que se usaron fueron los de las cepas de *Lactobacillus plantarum* WCFS1, *Leuconostoc mesenteroides* ATCC 8293 y *Weissella paramesenteroides* ATCC 33313, con números de acceso en el NCBI NC\_004567.2, NC\_008531.1 y NZ\_ACKU00000000.1, respectivamente.

### 8. Resultados y discusión

A partir del ADN extraído de la mezcla de 25 muestras de quesos Cotija, se construyó una sola biblioteca, misma que fue secuenciada por la tecnología de Illumina como se detalla en la parte final de la sección 7.2. Extracción de

paquete celular y de ADN total.

Como resultado de la secuenciación se obtuvo un total de 226,458,970 lecturas con longitud promedio de 100 bp a partir de fragmentos con una longitud de ~500bp. De éstas, alrededor de un 5% mapearon sobre el genoma de *Bos taurus* lo que indica que el tratamiento con DNasas antes de la ruptura de las células en la extracción de ADN fue adecuado (Anexo 1).

### 8.1. Reporte de calidad

En las Figuras 7A y B se muestra el diagrama de las calidades en score Phred para cada posición de las lecturas pareadas. Lo ideal es que todas las medias (línea azul) se mantengan por arriba del valor de 28, que corresponde a la zona verde en el fondo de la gráfica.

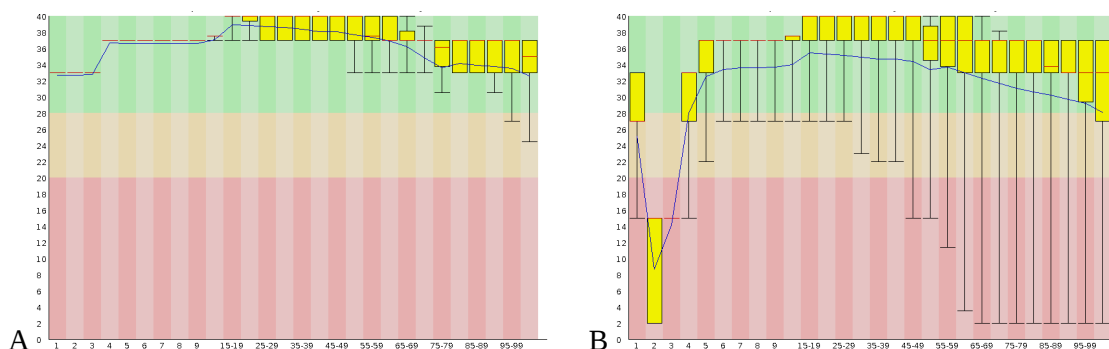


Figura 7. Box plot de calidad para las secuencias obtenidas por la secuenciación en sentido 5' (A) y en sentido 3' (B).

El decaimiento de la calidad en las últimas posiciones es un fenómeno que ya se ha reportado y que se relaciona con el daño que ocasiona al ADN la exposición al laser, aunado a que el llamado de bases es un consenso en un cluster de moldes y a que no todas las hebras se sintetizan a la misma velocidad. Dicha desincronización deriva en una acumulación en la tasa de error (Fuller et al., 2009).

Por otro lado, el decaimiento drástico que se observa en la segunda posición de la Figura 7B, es atípico, y pudo deberse a un problema relacionado con el desempeño del equipo durante el proceso de secuenciación. Cabe mencionar que a pesar de la caída de la calidad en las lecturas en sentido 3', la media de la calidad se encuentra siempre en la zona verde.

En cuanto a la frecuencia de nucleótidos en las secuencias, la cantidad relativa de cada base refleja su proporción en el metagenoma y en ningún caso se esperaría que fuera extremadamente desbalanceada. En el caso del metagenoma de queso Cotija (Figura 8A) se observa, en términos generales, una adición preferencial de timina y adenina en todas las posiciones lo cual se relaciona con un contenido de GC menor al 50% en los fragmentos genómicos que conforman la muestra.

En la distribución del contenido de GC en la población total de las lecturas se observan dos picos (Figura 8B, línea roja). El primero de ellos corresponde a la población más grande de secuencias con un máximo en 43% y el

segundo indica una población de menor abundancia con pico máximo en 71%.

Dados los contenidos de GC, las distribuciones de las lecturas podrían corresponder por un lado al *Phylum* de los *Firmicutes*, que es representativo de bacterias Gram positivas con bajo contenido de GC, y por otro, al de *Actinobacteria* en el que se clasifica a las bacterias Gram positivas generalmente con contenido de GC mayor al 55% (Doroghazi y Metcalf, 2013). De acuerdo a los antecedentes, se esperaba que la comunidad bacteriana del queso Cotija estuviera conformada principalmente por Firmicutes. En cuanto a las actinobacterias, no habían sido detectadas en queso Cotija anteriormente, aunque se encuentran con relativa frecuencia en análisis metagenómicos de alimentos fermentados naturalmente de origen vegetal (Jung et al., 2011; Lyu et al., 2013; Xie et al., 2013).

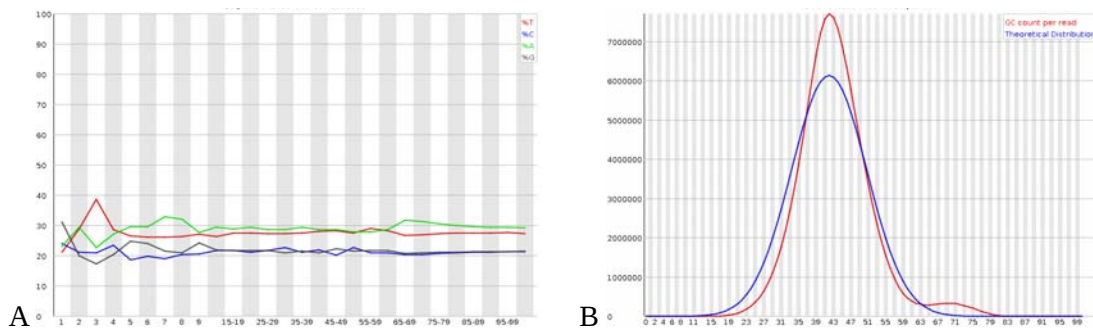


Figura 8. Gráfica de contenido de nucleótidos por posición (A) y distribución del %GC (B). Se muestran únicamente los resultados del análisis del conjunto de secuencias obtenidas en sentido 5'.

Un análisis del contenido de GC sobre las lecturas ensambladas confirmó que la fracción de lecturas minoritaria de alto contenido de GC no es producto de secuencias artefacto derivadas de errores de secuenciación.

Si la generación de secuencias de alto contenido de GC hubiera sido asarosa, habría sido imposible ensamblarlas en fragmentos de longitud más de 200 bases. La calidad y abundancia de dichas lecturas permitió hacer la reconstrucción de fragmentos genómicos, lo que respalda el hecho de que éstos representan una población real que forma parte de la estructura poblacional del metagenoma. (Figura 9).

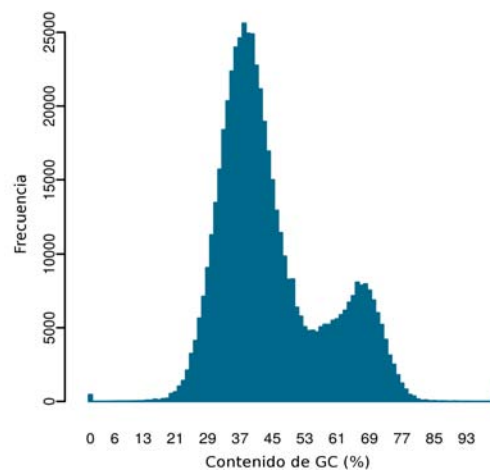


Figura 9. Histograma de distribución de contenido de GC en los contigs ensamblados.

## 8.2. Análisis taxonómico

La anotación de las secuencias metagenómicas de queso Cotija con genes marcadores de copia única permitió identificar una riqueza de 45 géneros bacterianos en los que se encuentran contenidas 86 especies diferentes, la mayoría de ellas Firmicutes (99.8%) y algunas de los *phila* Actinobacteria (~0.1%) y Proteobacteria (~0.01%). Estas actinobacterias forman parte de la subpoblación de alto contenido de GC observada en las Figuras 8B y 9.

Para fines prácticos de este análisis, los géneros bacterianos se clasificaron en tres grupos claramente definidos de acuerdo a su abundancia. Se consideró como dominantes a todos aquellos que presentaran más del 10% de abundancia; subdominantes si estaban entre el 1 y el 10% y como pobremente representados a los que tuvieran menos del 1% de abundancia relativa total. La población dominante estuvo compuesta por tres géneros que representan más del 80% de la riqueza total de especies, mientras que en la fracción subdominante se observaron cuatro géneros más (Tabla 2). Los otros 38 géneros detectados forman parte de la población no dominante.

Tabla 2. Anotación taxonómica de la fracción que representa más del 1% de la riqueza total relativa

Abundancia	Marcadores de copia única		Gen 16S ADNr	
	Género <sup>a</sup>	Especie <sup>b</sup>	Género <sup>a</sup>	
Dominante	<i>Lactobacillus</i> (33.4)	<i>plantarum</i> (95.9)	<i>Lactobacillus</i> spp. (20.7)	
		<i>farciiminis</i> (2.0)		
		<i>brevis</i> (0.7)		
		<i>rhamnosus</i> (0.7)		
		<i>casei</i> (0.5)		
		<i>paracasei</i> (0.05)		
		<i>coryniformis</i> (0.01)		
		<i>buchneri</i> (<0.01)		
	<i>Weissella</i> (26.7)	<i>paramesenteroides</i> (99.98 )	<i>Weissella</i> spp. (29.6)	
		<i>cibaria</i> (0.02)		
<i>Leuconostoc</i> (21.9)	<i>mesenteroides</i> (99.9)	<i>citreum</i> (0.03)	<i>Leuconostoc</i> spp. (11.3)	
				<i>kimchii</i> (<0.01)
				sp. C2 (<0.01)
Subdominante	<i>Aerococcus</i> (8.3)	<i>viridans</i> (100)	<i>Aerococcus</i> spp. (8.9)	
	<i>Enterococcus</i> (6.3)	<i>faecium</i> (51.0)	<i>Enterococcus</i> spp. (3.4)	
		<i>italicus</i> (29.7)		
		<i>faecalis</i> (16.8)		
		<i>casseliflavus</i> (1.5)		
		<i>gallinarum</i> (0.7)		
		sp. 7L76 (0.3)		

<i>Lactococcus</i> (2.0)	<i>garvieae</i> (71.2) <i>lactis</i> (28.8)	<i>Lactococcus</i> spp. (2.5)
<i>Staphylococcus</i> (1.1)	<i>aureus</i> (84.9) <i>saprophyticus</i> (13.7) <i>carneus</i> (0.7) <i>epidermidis</i> (0.5) <i>haemolyticus</i> (0.1) <i>hominis</i> (0.05) <i>lugdunensis</i> (0.01)	<i>Staphylococcus</i> spp. (3.0)
		<i>Bavariicoccus</i> spp. (2.2) <i>Marinilactibacillus</i> spp. (2.2) <i>Pediococcus</i> spp. (2.0) <i>Bacillus</i> spp. (1.9) <i>Fructobacillus</i> spp. (1.4) <i>Alkalibacterium</i> spp. (1.4) <i>Carnobacterium</i> spp. (1.0)

<sup>a</sup> En paréntesis la abundancia relativa de cada género con respecto a la abundancia total. <sup>b</sup> En paréntesis la abundancia relativa con respecto al 100% de su género respectivo.

Contrastantemente a lo obtenido en el análisis por genes marcadores de copia única, en la anotación con el gen ribosomal 16S se detectó una riqueza de 574 géneros bacterianos, un orden de magnitud más que con el enfoque anterior, de los cuáles el 98% fueron *Firmicutes* y el resto fueron identificados en los *Phyla Actinobacteria* y *Proteobacteria* de todas las clases, y otros minoritarios dentro de los *Phyla Acidobacteria* y *Bacteroidetes*, así como algunos géneros dentro del dominio de las *Archaeas* (Figura 10).

La fracción dominante estuvo integrada por los mismos tres géneros bacterianos identificados por marcadores de copia única, y en el caso de la población subdominante siete géneros más se integraron a la lista (Tabla 2). De los siete géneros bacterianos que no aparecen como subdominantes por copia única, *Pediococcus* (*acidilactici* y *pentosaseus*) y *Bacillus* (*cereus* y *subtilis*) fueron identificados como parte de la población que representa menos del 1%; mientras que *Bavariicoccus*, *Marinilactibacillus*, *Fructobacillus*, *Alkalibacterium* y *Carnobacterium* no fueron identificados mediante marcadores unicopia. Los más de 500 géneros bacterianos restantes tuvieron una abundancia relativa menor al 1%.

Lo anterior indica que la anotación taxonómica basada en el gen ribosomal 16S permite observar una mayor riqueza de géneros bacterianos debido a la disponibilidad de bases de datos más grandes. Las bases de datos incluídas en cada pipeline son de diferente tamaño, RDP tiene 8,127 entradas para eubacterias y 295 para *Archaeas*, mientras que RefMG tiene 3,319 y 130, respectivamente.

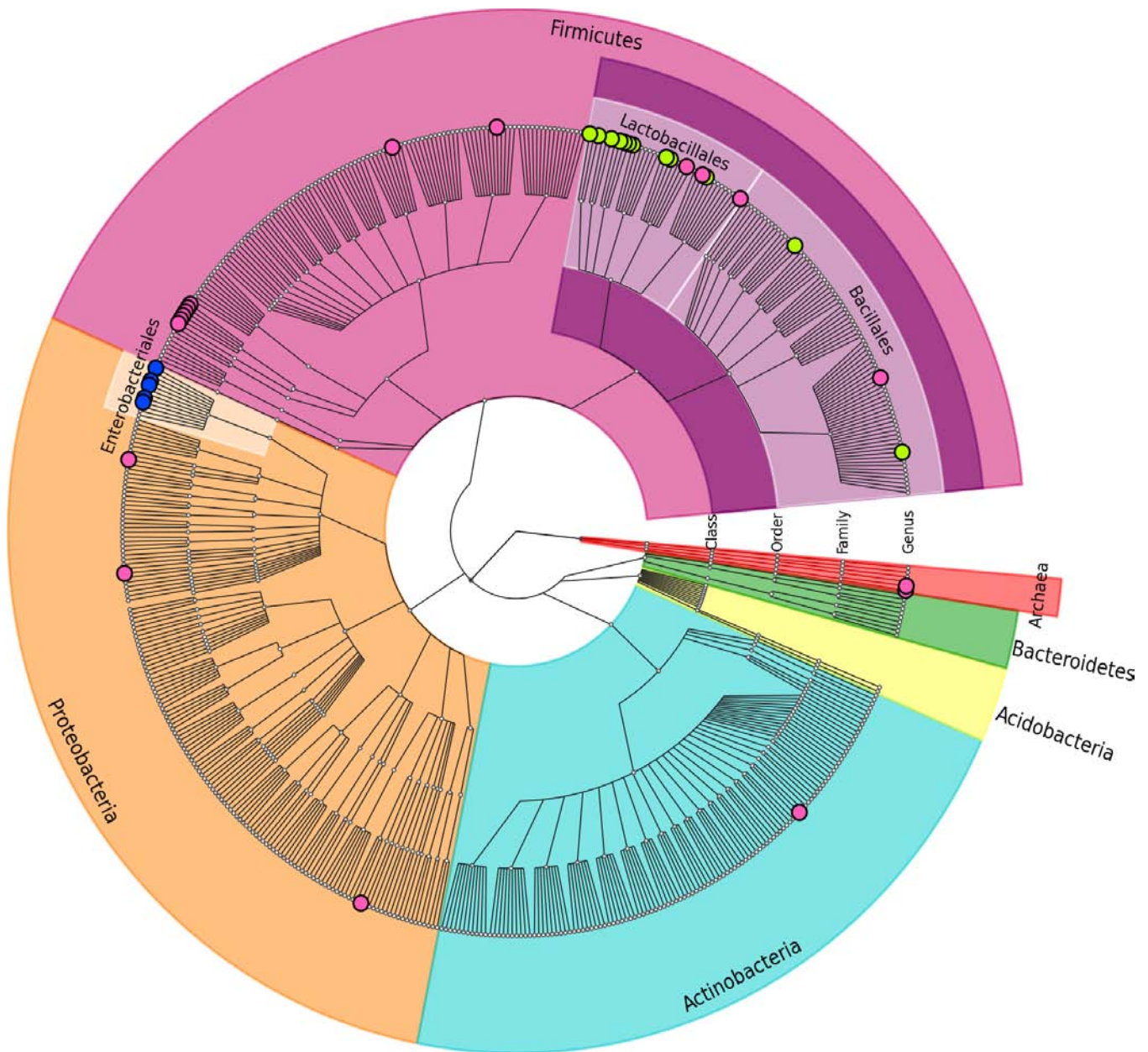


Figura 10. Cladograma de la anotación taxonómica de los principales *Phyla* detectados por el análisis del gen ribosomal 16S contra la base de datos RDP usando Parallel-meta v2.4. Algunos clados se resaltan en los nodos terminales: en rosa las halófilas (*Tetragenococcus*, *Marinilactibacillus*, *Halolactibacillus*, *Alkalibacterium*, *Dehalobacter*, *Thermohalobacter*, *Halanaerobiales*, *Haloglycomyces*, *Thiohalomonas*, *Desulfohalobium*, *Halomonas*, *Halobacterium*, *Haloquadratum*); en azul las coliformes (*Klebsiella*, *Enterobacter*, *Citrobacter*, *Hafnia*, *Escherichia*) y en verde aquellos géneros que representan más del 1% de abundancia total relativa (*Bacillus*, *Staphylococcus*, *Aerococcus*, *Enterococcus*, *Lactobacillus*, *Pediococcus*, *Leuconostoc*, *Weissella*, *Lactococcus*, *Alkalibacterium*, *Carnobacterium*, *Marinilactibacillus*, *Bavariicoccus*, *Fructobacillus*).  
 Figura generada con GraPhlAn v0.9.7 (Asnicar et al., 2015).

Por otro lado, la anotación con genes codificantes de copia única permite obtener resolución a nivel de especie. Ésto es debido a que la variabilidad en el gen ribosomal 16S está limitada por su funcionalidad en la maquinaria traduccional, lo que resulta en un alto grado de conservación en la secuencia entre especies lo que indica que una alta similitud de secuencias de genes ribosomales no implica necesariamente identidad a nivel de especie (Fox et

al., 1992; Klappenbach et al., 2000), adicional y contrastantemente se ha reportado que la heterogeneidad intragenómica de este gen multicopia puede llegar a ser de hasta 6.5% (Cilia et al., 1996; Wang et al., 1997). Por estas razones, en trabajos de reconstrucción de la estructura de comunidades microbianas y estimación de la diversidad, se recomienda complementar los resultados con el análisis de genes marcadores de copia única que podrían llegar a ser incluso más confiables que los obtenidos por 16S ADNr (Bonilla-Rosso et al., 2012).

De los 45 géneros detectados por marcadores de copia única, 38 se encuentran también en el resultado del análisis con el ribosomal 16S. Los siete restantes pertenecen a los géneros *Agrobacterium*, *Delftia*, *Frankia*, *Geodermatophilus*, *Kytococcus*, *Nakamurella* y *Pantoea*, que a nivel de familia sí fueron detectados en ambos casos. Cabe mencionar que ninguno de éstos géneros representa más del 0.0005% de la riqueza total.

Finalmente cabe señalar que ambos resultados de anotación taxonómica son congruentes y complementarios entre sí sobre todo en poblaciones dominantes.

#### Cálculo de índices de diversidad y esfuerzo de muestreo

De acuerdo a la anotación taxonómica obtenida con el marcador ribosomal 16S, que fue el método que permitió observar el mayor número de taxa anotados, se evaluó el esfuerzo de muestreo y se calcularon los índices de diversidad necesarios para hacer una descripción de la complejidad de la estructura de la comunidad bacteriana.

Una forma de estimar la eficiencia en el esfuerzo de muestreo es la construcción de curvas de acumulación de especies, que relacionan el tamaño de la muestra con el número de especies observado y su ajuste estadístico, permite la obtención de curvas de acumulación de especies en las que es posible interporlar (o extrapolar, aunque ésto no es tan recomendable pues el resultado depende del método de regresión (Thompson et al., 2003)), cualquier tamaño de muestra para estimar el número de especies en la misma, y donde un comportamiento asintótico indicaría que el muestreo es suficiente para observar el número máximo de especies en la comunidad (Krebs, 2014). En la Figura 12 se muestra la curva de acumulación de especies con comportamiento asintótico que representa el esfuerzo de muestreo, entendido como profundidad de la secuenciación, en el metagenoma del queso Cotija.

El número de especies en la muestra es el descriptor básico de la estructura de una comunidad en un área determinada, sin embargo, una medida de la diversidad de especies debe tomar en cuenta tanto la riqueza o número de especies, como sus abundancias, de tal modo que dadas dos comunidades con el mismo número de especies, aquella en la que la distribución de abundancias sea más equitativa, será la más diversa (Krebs, 2014). En la Tabla 3 se encuentran índices de diversidad calculados para la comunidad bacteriana del queso Cotija.

Los estimadores no paramétricos Chao1 y ACE conceptualizan la diversidad como riqueza total de especies. Cada uno agrega un factor de corrección distinto al número de especies observadas en la comunidad. Chao1 toma en cuenta las especies observadas una o dos veces, mientras que ACE incorpora datos de todas las especies con

menos de diez individuos. Ambos son sensibles al tamaño de la muestra y subestiman la riqueza real cuando el muestreo es pobre (Escalante, 2007). Para el caso del queso Cotija, ambos estimadores indican un número total de especies muy cercano al observado (Tabla 3) lo que soporta el comportamiento de la curva de rarefacción (Figura 12) que indica que la profundidad con la que se llevó a cabo la secuenciación del metagenoma fue la adecuada para observar la mayoría de las especies de muy baja abundancia.

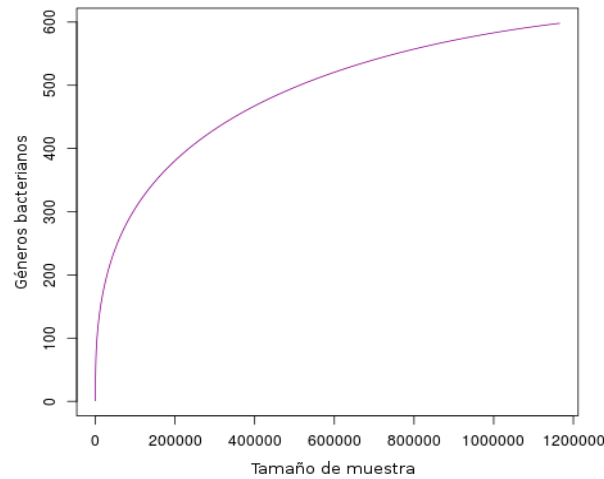


Figura 12. Curva de rarefacción del número de especies observadas en función al número de secuencias del marcador ribosomal 16S extraídas del metagenoma.

Tabla 3. Estimadores de diversidad calculados para la comunidad bacteriana descrita con el gen ribosomal 16S

Riqueza de especies	Chao1	ACE	Simpson (1-D)	Shannon (H)	Evenness (E)
598	629.63	660.48	0.84	2.52	0.39

Otro tipo de estimadores no paramétricos que son métricas de diversidad de especies son los índices de Simpson (D) y Shannon (H). El índice de Simpson (D), se puede entender como una medida de la heterogeneidad de abundancia en la muestra, y sugiere que la diversidad está inversamente relacionada con la probabilidad de que dos individuos en una comunidad elegidos al azar pertenezcan a la misma especie (Simpson, 1949). Una medida más fácil de interpretar es el complemento del índice de Simpson (1-D), es decir, la probabilidad de obtener dos organismos de la misma especie en un evento de muestreo. Su representación matemática toma valores que van desde cero para comunidades de baja diversidad, hasta valores cercanos a 1 (Krebs, 2014). Dicho índice tiene un valor mayor a 0.5 para la comunidad bacteriana del queso Cotija (Tabla 3), lo que indicaría que se trata de una comunidad con diversidad que va de baja a mediana.

Por otro lado, el índice de Shannon (H) es una medida de entropía donde el sistema es la comunidad. Se puede interpretar como la incertidumbre de predecir correctamente la especie del siguiente individuo colectado con base en la información observada. En una comunidad con una sola especie, H toma el valor de cero, pues la incertidumbre es nula, es decir que mientras mayor sea el valor de H, la incertidumbre aumenta (Shannon, 1948). En muestras biológicas el valor de H suele ser menor a 5.0 (Krebs, 2014), tal como ocurre en este caso (Tabla 3).



Al comparar valores de H calculados para muestras de estadios tempranos de la elaboración de quesos, se observa que el índice de Shannon para queso Cotija de encuentra dentro del rango de los valores reportados para inóculo de suero y cuajada madurada hasta por 24 horas, cuyos límites van desde 0.47 hasta 2.60 (Filippis et al., 2014). En otros procesos de fermentación de alimentos como el té de Pu-erh, se reporta una H=6.52 (Zhao et al., 2015), que ya esta por arriba de lo esperado para una muestra de diversidad media. Algunos metagenomas de ambientes que se consideran altamente diversos como suelos forestales (Kato et al., 2015) agua de mar y comensales de esponjas marinas (Kennedy et al., 2014) tienen valores de índice de Shannon de >8, 6.88 y 8.50 respectivamente; mientras que los de microbioma humano no superan valores de 2.65 (Li et al., 2012). Con base en la comparativa anterior podemos considerar a la microbiota del queso Cotija como de diversidad media-baja.

La equiparabilidad o evenness (E) es una métrica que intenta cuantificar la representación desigual en la abundancia de especies observadas con respecto a una comunidad hipotética donde todas las especies son igualmente abundantes (caso de máxima heterogeneidad). La forma más sencilla de estimarla es comparando algún índice (por ejemplo H o D) calculado para la muestra con respecto al mismo en condiciones de igualdad de abundancia para todas las especies. Valores cercanos a 1 indican que la comunidad tiene una estructura cercana a la máxima heterogeneidad posible (Krebs, 2014). Para la comunidad bacteriana del queso Cotija, la equiparabilidad en la abundancia es relativamente baja (Tabla 3), lo que se refleja en el hecho de que haya muy pocas especies dominantes y una cola muy larga de especies de baja abundancia. Dicha descripción corresponde a una estructura poblacional de mediana a baja diversidad de acuerdo a los valores de los índices obtenidos en la Tabla 3, aunque con alta riqueza de especies.

#### Población dominante en el queso Cotija

A los tres meses de maduración, los géneros y especies más exitosos hablando en términos de abundancia en el queso Cotija, son NSLAB, que tienen metabolismo heterofermentativo y son productores importantes de compuestos que confieren olor y sabor al alimento. Dicha comunidad bacteriana está representada principalmente por tres géneros cuya abundancia total relativa representa más del 60%. Estos son *Lactobacillus plantarum*, *Weissella paramesenteroides* y *Leuconostoc mesenteroides* (Tabla 2).

La presencia de bacterias del género *Lactobacillus* es común en quesos tanto frescos como madurados (Bautista-Gallego et al., 2014; Ercolini et al., 2012) y su abundancia esta directamente relacionada con el tiempo de maduración (Quigley et al., 2012). Otros generos que también han sido descritos frecuentemente como parte de la microbiota bacteriana de quesos madurados naturalmente elaborados con leche no pasteurizada y que fueron encontrados en el queso Cotija como parte de la población subdominante son *Streptococcus*, *Lactococcus* y *Enterococcus*, (Duthoit et al., 2003; Fuka et al., 2013; Masoud et al., 2012; Quigley et al., 2012). En contraste, los géneros *Leuconostoc* y *Weissella* no se asocian comunmente a quesos madurados y nunca han sido identificados como población dominante (Fuka et al., 2013; Masoud et al., 2012), por lo que este hecho puede considerarse

como una firma distintiva del queso Cotija Región de Origen.

La presencia de *Weissella* y *Leuconostoc* ha sido reportada en leche cruda de vaca (Masoud et al., 2012) así como en la superficie de contenedores de madera (Settanni et al., 2012) y en el aire de lugares donde se elaboran otro tipo de quesos (Litopoulou-Tzanetaki and Tzanetakis, 2014). Durante la elaboración del queso Cotija se añade sal de grano marina y la cuajada salada es amasada durante un tiempo considerable sobre una mesa, que suele ser de madera, hasta que la sal se incorpora completamente. Dicho proceso es característico de la elaboración de este queso, lo que podría explicar en parte lo peculiar de la estructura poblacional de su microbiota.

El hecho de no haber identificado por métodos tradicionales a dos de los géneros más importantes en dominancia puede deberse por un lado a la dificultad de cultivar en general a las LAB y en particular a estos géneros. Es posible que los medios de cultivo empleados tengan deficiencias en algunos factores de crecimiento específicos para las cepas provenientes del queso Cotija o que por el contrario, su composición con exceso de nutrientes favorezca el crecimiento de microorganismos no dominantes, aunado a que se ha reportado que en el caso de *Weissella*, ésta puede presentar crecimiento lento en los medios de cultivo convencionales (Björkroth y Holzapfel, 2006).

Por otro lado, ni *Weissella* ni *Leuconostoc* se han podido detectar por amplificación de las regiones variables V3-V4 en el análisis de comunidades bacterianas por PCR-DGGE realizados dentro de nuestro grupo de trabajo. Esto puede deberse a que la hibridación de los cebadores, aunque se consideren “universales”, no ocurre con la misma probabilidad sobre todas las secuencias de ADN ribosomal 16S en una mezcla (Hong et al., 2009; Roux et al., 2011), además, es importante resaltar que la mayoría de las bacterias identificadas en el metagenoma pertenecen al orden de los Lactobacillales (Figura 10), es decir que son filogenéticamente muy cercanas lo que puede generar una identificación errónea sobre todo cuando se trabaja con fragmentos cortos de las regiones variables de 16S (Nalbantoglu et al., 2014; Petrosino et al., 2009). Incluso algunos autores recomiendan utilizar cebadores específicos para grupos taxonómicos conocidos dada la amplia tasa de error que ocurre durante la detección e identificación cuando se utilizan estos métodos (Satokari et al., 2003). En un trabajo reciente, se reportó la presencia de *Weissella* y *Leuconostoc* mediante PCR-DGGE de las regiones variables V7-V8 del gen ribosomal 16S a partir de ADN extraído de muestras de queso Cotija Región de Origen (Chombo-Morales et al. 2016). Estos ejemplos ilustran los alcances del análisis metagenómico sobre aquellos basados en la amplificación de segmentos de marcadores moleculares.

#### Población subdominante y fracción de no dominantes

Todos los géneros bacterianos identificados por métodos dependientes e independientes de cultivo fueron identificados en el metagenoma mediante el análisis del gen ribosomal 16S. Además se confirmaron los resultados reportados por FISH para los generos *Bacillus* spp., *Staphylococcus* spp. y *Enterococcus* spp., cuya abundancia

fue en conjunto del 11% del total de la carga bacteriana total en el queso Cotija (Escobar, 2012), pues dichos resultados tienen concordancia con lo observado en el metagenoma, dado que los tres géneros mencionados pertenecen a la población de subdominantes y de acuerdo a las abundancias calculadas mediante la anotación taxonómica, la suma de estos tres géneros fue de 8.3%.

La fotografía de la estructura poblacional de la microbiota del queso Cotija a los tres meses de maduración, indica la presencia de una gran variedad de LAB tanto homo- como heterofermentadoras, lo que incrementa la diversidad de genes y rutas metabólicas posibles y por ende, la variedad de metabolitos que tienen repercusión en la inocuidad, atributos sensoriales y compuestos de importancia biotecnológica. Al inicio de cualquier fermentación láctica, las especies más abundantes o exitosas son las SLAB, que son productoras importantes de lactato con metabolismo homofermentativo tales como *Lactococcus lactis* y algunas especies de lactobacilos como *L. farciminis*. Conforme la maduración progresa, la población de SLAB disminuye y la de NSLAB se incrementa, dando paso al crecimiento de bacterias con metabolismo heterofermentativo obligado y facultativo tales como *Leuconostoc*, *Weissella* y otras especies de lactobacilos como *L. plantarum*, lo que permite una producción importante de compuestos de olor y sabor (Beresford y Williams, 2004; Björkroth y Holzapfel, 2006; McSweeney, 2004).

Además de los géneros reconocidos como SLAB y NSLAB, dentro de la fracción subdominante se encuentran bacterias de los géneros *Marinilactibacillus* y *Alkalibacterium*, reconocidos como bacterias halófilas alcalófilas lácticas (HALAB en inglés) y que han sido descritas como de origen marino (Ishikawa et al., 2013, 2007). El otro género que forma parte del grupo de las HALAB es *Halolactibacillus*, que también se encuentra en el queso Cotija formando parte de la población pobremente representada. *Tetragenococcus halophilus*, es otro ejemplo de halófilo que forma parte del consorcio bacteriano de queso Cotija y que ha sido identificado en queso tipo Cotija (Morales et al., 2011) (Tabla 4).

Tabla 4. Géneros de bacterias halófilas detectadas en queso Cotija y reportadas en otros quesos

Bacteria halófila identificada en queso Cotija	Otras fuentes donde se ha identificado	Referencias
<i>Tetragenococcus halophilus</i>	Queso tipo Cotija	(Morales et al., 2011)
<i>Marinilactibacillus</i> spp.	Quesos semiduros, semisuaves y suaves	(Feurer et al., 2004; Ishikawa et al., 2007; Maoz et al., 2003)
<i>Halolactibacillus</i> spp.	Organismos marinos vivos y muertos	(Ishikawa et al., 2005)
<i>Alkalibacterium</i> spp.	Quesos suaves y semiduros	(Ishikawa et al., 2013)

La lista de halófilos se extiende a bacterias de varios *Phyla* (representados como círculos rosas en la Figura 10), la mayoría del dominio de las eubacterias como Firmicutes (*Halolactibacillus*, *Halobacillus*, *Dehalobacter*, *Thermohalobacter*, y *Halanaerobiales*); Actinobacteria (*Haloglycomyces*), Proteobacteria (*Thiohalomonas*, *Desulfohalobium*, y *Halomonas*), Tenericutes (*Haloplasma*); así como a algunas del dominio Archaea

(*Halobacterium* y *Haloquadratum*). Todas ellas representan menos del 1% de abundancia y seguramente fueron inoculadas durante el salado de la masa.

Por último, dentro de la fracción de los pobremente representados, fueron descritos 78 géneros por marcadores de copia única y 560 por anotación del gen ribosomal 16S. Los últimos se encuentran agrupados en 31 *Phyla* dentro de los dominios *Archaea* y bacteria (Figura 11).

Dentro del grupo de los pobremente representados se encuentran géneros reportados en un trabajo reciente sobre la identificación de bacterias aisladas de diferentes fuentes durante el proceso de elaboración del queso Cotija (Robles, 2014), sumando a la lista a *Kocuria*, *Acinetobacter*, *Brachybacterium*, *Klebsiella*, *Enterobacter*, *Brevibacterium* y *Psychrobacter*.

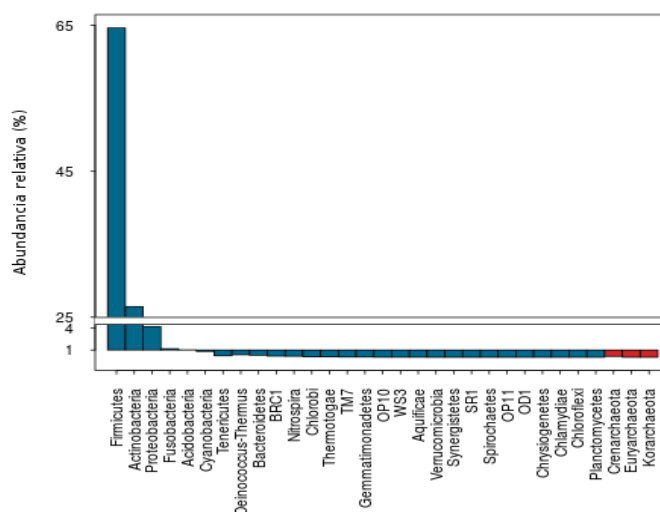


Figura 11. Diversidad a nivel de *Phylum* de la población pobremente representada identificada por 16S. Las barras azules son *Phyla* del dominio de las bacterias, y las rojas pertenecen al dominio de las *Archaeas*. El 100% corresponde a la cuenta total de los géneros incluidos en este grupo.

En esta lista hay algunos géneros catalogados como coliformes (representados como círculos azules en la Figura 10), aunque sus abundancias son muy bajas en el queso Cotija, para *Enterobacter* 0.0007%, *Klebsiella* 0.0004%, *Citrobacter* 0.0002% y *E. coli* 0.0007% con respecto a la abundancia total. Considerando que la carga bacteriana en el queso Cotija es de  $1 \times 10^5$  células/g (Escobar, 2012), la proporción de enterobacterias correspondería a 1-2 células/g lo que no representa un riesgo para el consumidor (NMX-F-735-COFOCALEC, 2011; NOM-243-SSA1, 2010). En el mismo sentido de calidad microbiológica e inocuidad, los patógenos que no aparecen en el queso por ninguno de los dos métodos de anotación taxonómica son *L. monocytogenes*, *P. aeuruginosa*, *Salmonella spp.*, *Yersinia spp.* y *Brucella spp.*

Con el fin de descartar que la *E. coli* encontrada en el queso pertenezca a alguna de las cepas identificadas como patógenas para humano, se llevó a cabo una búsqueda de genes asociados a virulencia en la tabla de anotación generada con Trinotate. Como referencia se tomó la lista de genes de virulencia para *E. coli* publicada en la base de datos del MOH Key Laboratory of Systems Biology of Pathogens, donde se reportan los genes asociados a

patogenicidad para ocho diferentes grupos de *E. coli*: enterotoxigénica (ETEC), enteroinvasiva (EIEC), enteropatógena (EPEC), enterohemorrágica (EHEC), enteroagregativa (EAEC), de adherencia difusa (DAEC), uropatógena (UPEC), y asociada a meningitis neonatal (NMEC). Aunque las cepas UPEC y NMEC no están reportadas como patógenos de alimentos, ambas cepas se incluyeron en la búsqueda de genes de virulencia. De la lista de genes asociados a las diferentes cepas patógenas, sólo se detectó el gen *aggR*, que codifica para un activador transcripcional en plásmido de EAEC. La *E. coli* enteroagregativa se reconoce como la causa más frecuente de diarrea en niños y adultos alrededor del mundo (Aslani et al., 2011). La presencia del gen que codifica para el activador transcripcional se ha utilizado como descriptor de cepas EAEC típicas (tEAEC), aunque esto no implica necesariamente que dicha *E. coli* pueda tener un fenotipo patogénico a menos que se identifiquen en la misma cepa otros genes de virulencia (Estrada-García et al., 2014); además, se han reportado cepas de tEAEC en muestras de heces de sujetos sin cuadro de diarrea (Aslani et al., 2011). Es decir que apesar de haber encontrado anotación para el activador transcripcional *AggR*, la presencia de *E. coli* en el queso no representa un riesgo para el consumidor dada a ausencia de genes de virulencia.

Por otro lado, dentro del grupo de organismos poco representados en el queso, se encontró una fracción de eucariones mediante la anotación del gen ribosomal 18S contra la base de datos de SILVA con un umbral E-value de  $1e-15$  usando el software Parallell-meta. Aún cuando este proyecto se enfocó en la descripción de la estructura poblacional bacteriana del queso Cotija y aunque el diseño experimental para la extracción de ADN fue optimizado para la lisis de bacterias, fue posible identificar a los géneros de eucariontes de la lista que se encuentra en el Anexo 4, donde los más abundantes son levaduras de las familias *Saccharomycetaceae* y *Debaryomycetaceae*, mismos que podrían contribuir con actividad lipolítica en la maduración del alimento, aunque esto tendría que ser comprobado experimentalmente.

### 8.3. Anotación funcional

Las lecturas crudas, sin filtrar por calidad, se ensamblaron (estadísticas en Tabla 5) y el ejercicio de mapeo de las mismas sobre los contigs demostró que el 96.75% de éstas fueron utilizadas en el ensamble.

La predicción de secuencias codificantes sobre el ensamble permitió recuperar ORFs con una longitud promedio de 630 nucleótidos y una distribución de longitud que se muestra en el histograma de la Figura 13, donde es claro que la mayoría de las secuencias codificantes tienen una longitud menor a 2,000 nucleótidos como es de esperarse en genomas bacterianos.

El método utilizado en este trabajo para la asignación de función a dichas secuencias tiene la versatilidad de contar con diferentes descriptores y clasificadores jerárquicos de función en una sola tabla de salida. Uno de ellos, los términos de GO, resultaron sumamente útiles en la visualización de las funciones codificadas más representadas en el metagenoma (Figura 14). Cabe mencionar que el análisis funcional debe entenderse como el potencial metabólico del consorcio bacteriano como un sistema y no tanto como una función asociada a un género

o especie bacteriana a menos que se lleve a cabo la anotación taxonómica del contig del que proviene el gen o genes de interés.

Tabla 5. Estadísticas del ensamble y anotación del metagenoma de queso Cotija

Ensamble	
Número total de bases ensambladas (Mb)	104.9
Número total de contigs	65,162
Tamaño promedio de los contigs (bp)	1,610
Tamaño del contig más largo (bp)	186,928
Tamaño del contig más corto (bp)	200
N50 / N90 (bp)	3,854 / 547
Anotación	
Total de secuencias codificantes predichas	143,736
ORFs con anotación de HMMER/Pfam	97,063
ORFs con anotación de BLAST/Swiss-Prot	75,989
ORFs con asignación de número K por GhostKoala	57,747
Secuencias del gen ribosomal 16S extraídas	1,165,460
Secuencias de genes marcadores de copia única extraídas	1,284,541

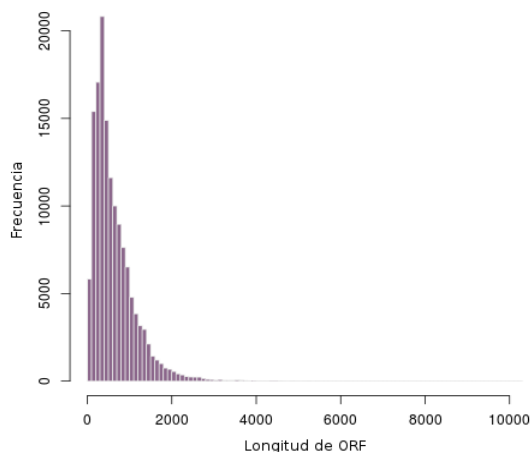


Figura 13. Histograma de la distribución de tamaños de ORFs predichos sobre el ensamble.

Los procesos más representados están relacionados con el metabolismo central de las bacterias, como el metabolismo de energía, del material genético y en general la actividad transferasa que está involucrada en una amplia gama de reacciones básicas dentro de la célula. Dejando de lado dichas funciones que son inherentes a todas las especies, llama la atención una cuenta alta en los términos de GO asociados con el metabolismo de lípidos, aminoácidos y compuestos aromáticos, así como con el transporte y metabolismo de ácidos orgánicos, cetonas, aminas y compuestos azufrados (Figura 14). Dichas actividades podrían estar implicadas en la

producción de compuestos que confieren olor y sabor al alimento o inocuidad, como en el caso del transporte de ácidos orgánicos. Además, hay anotación relacionada con funciones de secreción de proteínas, que tiene relevancia en la exportación de enzimas hidrolíticas y otros péptidos como por ejemplo bacteriocinas.

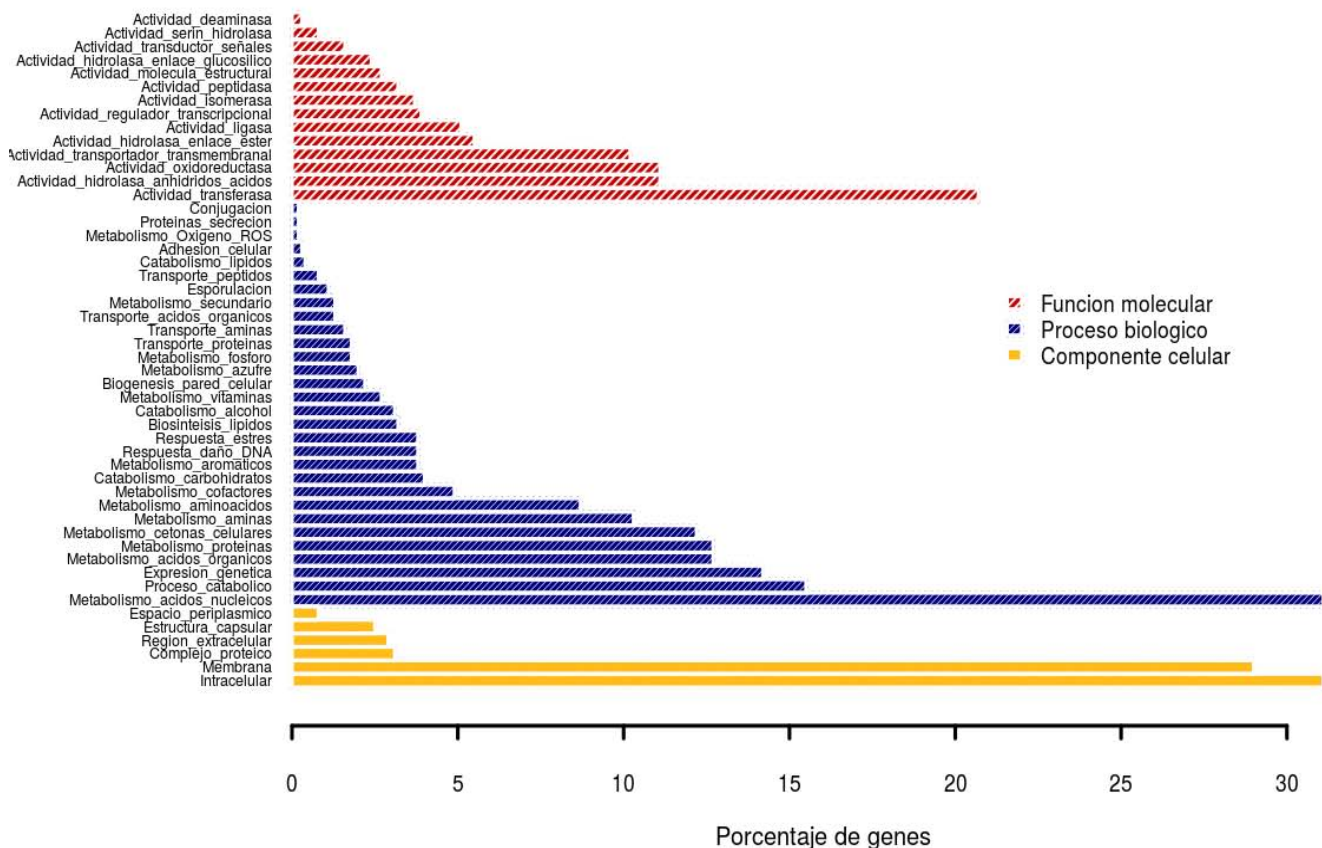


Figura 14. Términos de GO más representados en la anotación funcional del metagenoma de queso Cotija.

Otros procesos que llaman la atención es el metabolismo de oxígeno y especies reactivas de oxígeno, esporulación y la respuesta a estrés, que se explican como parte del arsenal del que disponen las bacterias exitosas en este ambiente que si bien es rico en nutrientes también es fisicoquímicamente hostil. Dichos genes y otros más como los relacionados con la producción de compuestos antimicrobianos tipo bacteriocinas podría estar relacionado con la transferencia horizontal de material genético, misma que está representada en el proceso de conjugación.

Por otro lado, la actividad peptidasa representa la utilización de nitrógeno por LAB, misma que genera como producto aminoácidos libres, cuyo catabolismo contribuye al perfil de sabor en quesos (Rijnen et al., 2003).

Otra actividad enzimática de importancia que aparece enriquecida en la anotación del metagenoma es la de hidrólisis sobre enlaces éster, lo que se traduce en la liberación de ácidos grasos, algunos de los cuáles confieren olor y sabor tales como el acético, butírico, propiónico, hexanoico, octanoico y los ramificados 4-metiloctanóico, 4-etiloctanóico y 3-metilvalérico (una discusión exhaustiva sobre el gusto de los ácido grasos libres fue escrita por Mattes, en el 2009). Los ácidos grasos libres que predominan en el queso Cotija son los insaturados 18:1, 18:2 y 18:3 y los saturados de 12, 13, 14 y 16 carbonos (Escobar, 2012), que podrían actuar como sustratos para la

producción de otros compuestos.

### Búsqueda de compuestos de olor y sabor

En el queso Cotija hay una gran variedad de bacterias ácido lácticas, lo que incrementa el número de genes disponibles y las posibles combinaciones de producción de metabolitos. Esto repercute en la complejidad del perfil de olor y sabor del alimento.

Se hizo un ecrutinio dirigido a la reconstrucción de rutas metabólicas relacionadas con la producción de metabolitos reportados en la literatura como responsables de conferir olor y sabor en quesos, mismos que se reportan en la Tabla 6. Lo anterior fue posible gracias al análisis realizado mediante GosthKOALA, que es una implementación de KEGG especializada en la reconstrucción de rutas metabólicas de metagenomas a partir de una lista de secuencias de proteínas.

Tabla 6. Rutas metabólicas involucradas en la síntesis de compuestos de olor y sabor en quesos

Ruta metabólica (KEGG id)	Compuesto <sup>a</sup>	Aroma / sabor <sup>b</sup>
Metabolismo de propanoato (00640)	Ácido propiónico	Pungente, rancio / agrio, ligeramente a queso
	Propionaldehído	Frutal, pungente / no reportado
	1-propanol	Mohoso, alcoholico / maduro, frutal
	2,3-butanodiona (diacetilo)	Mantequilla rancia / mantequilla
	Acetoína	Mantequilla, amaderado, yogurth / graso, cremoso, mantequilla
Metabolism de butanoato (00650)	Butirato	Rancio / grasa butírica
	Butanal	Pungente, olor aldéhico / no reportado
	1-butanol	Rancio, dulce, banana, fusel / banana, seco, fusel
Degradación de compuestos aromáticos (01220)	Benzaldehído	Almendras amargas / almendras amargas
Glicolisis/gliconeogénesis (00010)	Ácido benzóico	Inoloro o característico muy débil / amargo
	Etanol	Etéreo / ardiente
Metabolismo de piruvato (00620)	Ácido láctico	Inoloro / acre
	Ácido acético	Pungente, agrio, vinagre / ardiente
Metabolismo de nitrógeno (00910)	Acetaldehído	Pungente, afrutado / no reportado
	Amonio (NH <sub>4</sub> <sup>+</sup> )	Empalagoso, pungente, irritante, orina seca / no reportado



Degradación de ácidos grasos (00071)	Formación del respectivo aldehído y 1-alcohol	Frutal, floral, notas verdes / frutal (aldehídos) Dulce, fresco, floral / frutal, dulce (alcoholes derivados de 4-10 carbonos) Graso / grasoso, ceroso (alcoholes derivados de ácidos grasos de cadena larga)
Metabolismo de cisteína y metionina (00270)	Metanotiol	Pungente, col podrida, sulfuroso, ajo / desagradable a col
Metabolismo de fenilalanina (00360)	2-feniletanol	Como a rosas / amargo, dulce que recuerda a melocotón
	Fenilacetaldehído	Floral como a rosas <sup>a</sup> / no reportado
	Fenilacetato	No reportado / no reportado

<sup>a</sup> Compuestos identificados más frecuentemente en quesos (Smit et al., 2005). <sup>b</sup> Información disponible para cada compuesto en el sitio de PubChem del NCBI.

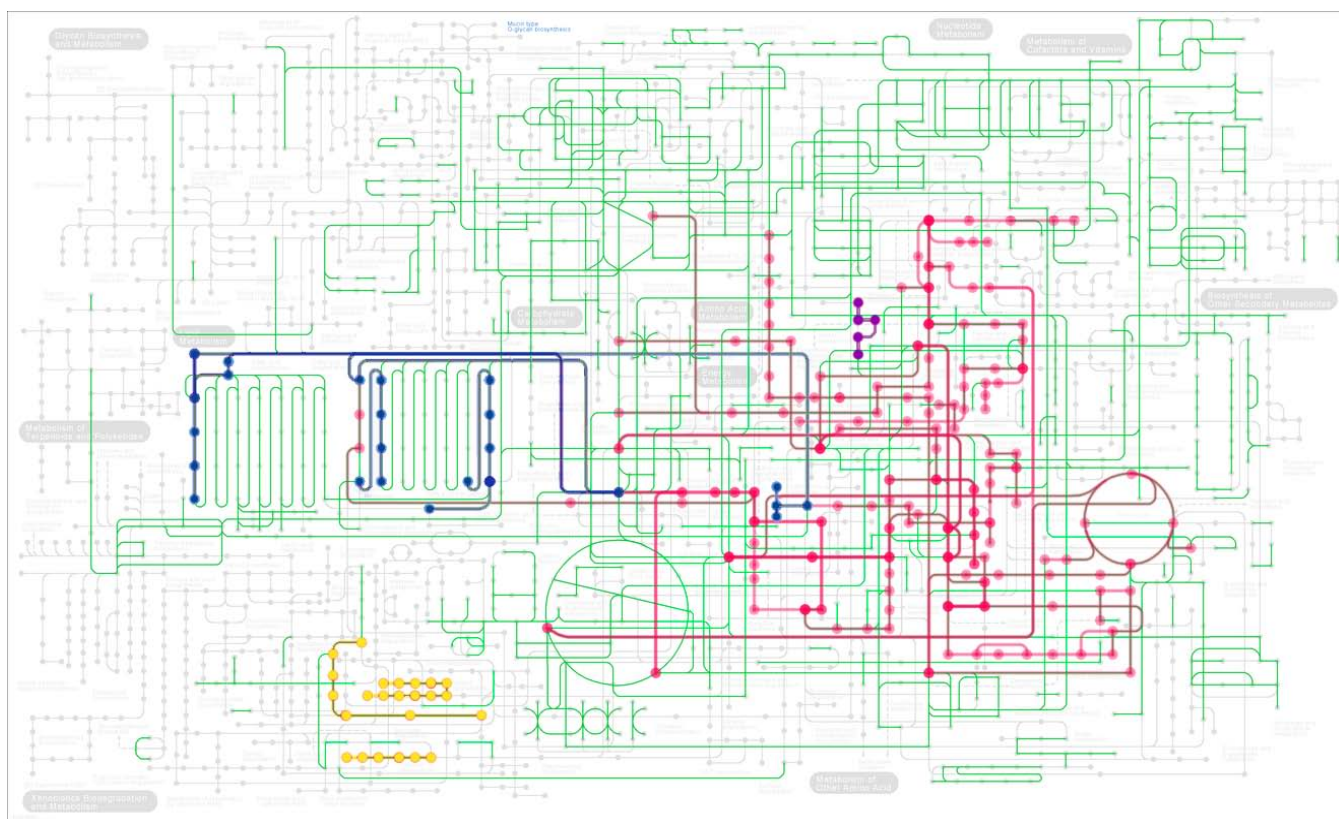


Figura 15. Mapa metabólico universal del KEGG. Las rutas metabólicas identificadas en el metagenoma del queso Cotija están representadas en verde. Se resaltan los módulos de interés en la producción de compuestos de olor y sabor: Amarillo, degradación de compuestos aromáticos; morado, metabolismo de azufre; azul, metabolismo de ácidos grasos; rosa, metabolismo de aminoácidos.

En el esquema metabólico del KEGG de la Figura 15 se resaltan las vías relacionadas con la producción de compuestos de olor y sabor. Todas las rutas metabólicas del metabolismo de aminoácidos están completamente representadas en el metagenoma del queso Cotija, así como las de catabolismo de ácidos grasos y compuestos

aromáticos, de acuerdo a lo observado en la anotación de términos de GO (Figura 14).

El consorcio bacteriano del queso Cotija tiene potencial para producir todos los compuestos de la Tabla 6, en cambio fue imposible la reconstrucción completa para la producción de otros compuestos reportados en quesos tales como propanona, indol, ácido isobutírico, ácido hidroxibenzóico, p-cresol y limoneno.

La reconstrucción de vías metabólicas se complementó con una búsqueda de enzimas relacionadas específicamente con el catabolismo de ácidos grasos, aminoácidos libres y otros metabolitos (Tabla 7). La búsqueda se basa en el escrutinio de identificadores de Swiss-Prot cuyo método de obtención se detalla en la sección 7.3 de Anotación funcional y que incluye la anotación de ORFs mediante búsquedas de BLAST contra esta base de datos. De esta forma se encontró anotación para una gran cantidad de aminotransferasas de aminoácidos de cadena ramificada pero ninguna para aminotrasferasas de los aminoácidos aromáticos tirosina y triptófano, lo cual resulta conveniente dado que algunos de sus compuestos de degradación, como el escatol, tienen perfiles de olor desagradables (Yvon y Rijnen, 2001). Adicionalmente, se identificaron genes con anotación para una amplia variedad de liasas de aminoácidos así como hidrolasas de ésteres carboxílicos.

Tabla 7. Reacciones específicas de catabolismos de ácidos grasos libres y aminoácidos

Reacción	Enzima	Número EC	Coincidencias en anotación*
L-leucina + 2-oxoglutarato $\rightleftharpoons$ 4-metil-2-oxopentanoato + L-glutamato	Transaminasa de aminoácidos de cadena ramificada <sup>c</sup>	2.6.1.42	50
L-iso-leucina + 2-oxoglutarato $\rightleftharpoons$ (S)-3-metil-2-oxopentanoato + L-glutamato			
L-valina + 2-oxoglutarato $\rightleftharpoons$ 3-metil-2-oxobutanoato + L-glutamato			
L-fenilalanina + piruvato $\rightleftharpoons$ fenilpiruvato + L-alanina	Transaminasa de fenilalanina (histidina) <sup>c</sup>	2.6.1.58	11
2-oxo ácido $\rightleftharpoons$ aldehído + CO <sub>2</sub>	Alfa cetoácido carboxilasa <sup>a</sup>	4.1.1.1	6
L-treonina $\rightleftharpoons$ glicina+ acetaldehído	L-treonina aldolasa <sup>c</sup>	4.1.2.48	3
Tirosina $\rightarrow$ tiramina +CO <sub>2</sub>	Tirosina descarboxilasa <sup>a</sup>	4.1.1.25	3
Metil cetonas (alcan-2-onas) $\rightleftharpoons$ alcan-2-ol	Carbonil reductasa I <sup>a</sup>	1.1.1.184	6
2-xileno $\rightarrow$ 2,3-dimetilfenol + 3,4-dimetilfenol	Fenol 2-mono oxigenasa <sup>b</sup>	1.14.13.7	10

Aldehído + NAD(+) → carboxilato + NADH	Aldehído deshidrogenasa <sup>a</sup>	1.2.1.3	79
Alcohol + NAD(+) <=> aldehído o cetona + NADH	Alcohol deshidrogenasa <sup>c</sup>	1.1.1.1	122
Alcohol secundario + NAD(+) <=> cetona + NADH	Alcohol deshidrogenasa <sup>c</sup>	1.1.1.1	122
(S)-metilmalonil-CoA + piruvato <=> propanoil-CoA + oxaloacetato	Metilmalonil-CoA carboxitransferasa <sup>a</sup>	2.1.3.1	5

\* Indica número de ORFs con anotación para dicha enzima en la tabla de anotación global.

Las macromoléculas mayoritarias en el queso, que son fuente de sustratos para la obtención de compuestos de olor y sabor, son las proteínas y los lípidos. A partir de la proteólisis de las caseínas, que son las proteínas mayoritarias en el queso, se generan aminoácidos libres, que a su vez son el sustrato de las vías catabólicas que permiten la obtención de una amplia variedad de compuestos de olor y sabor. Se ha reportado que *Lactobacillus platarum* es responsable de un incremento en la proteólisis secundaria en quesos madurados, promoviendo la producción de aminoácidos libres (Milesi et al., 2008). Particularmente en el queso Cotija se han aislado cepas proteolíticas de los géneros *Bacillus*, *Staphylococcus* y *Enterococcus* (Hernández, 2007).

El consorcio bacteriano del queso Cotija tiene potencial metabólico para transformar aminoácidos libres en sus respectivos  $\alpha$ -ceto ácidos por la vía de la transaminasa a partir de sustratos de cadena ramificada y fenilalanina. Dichos  $\alpha$ -ceto ácidos son degradados a aldehídos por las  $\alpha$ -ceto ácido descarboxilasas (Tabla 7). Es probable entonces que la fracción aldéhica en la mezcla de compuestos esté enriquecida en metabolitos producidos por el catabolismo de aminoácidos de cadena ramificada y fenilalanina así como de acetaldehído derivado del metabolismo de treonina (Tabla 7).

El otro sustrato importante en la producción de compuestos de olor y sabor en el queso son los ácidos grasos, mismos que en el queso Cotija se liberan por la actividad de bacterias lipolíticas de los géneros *Staphylococcus* y *Bacillus* (García, 2011, 2006).

En el metagenoma del queso Cotija hay una gran cantidad de genes con anotación para carboxil esterases que rompen enlaces éster para generar como productos un alcohol y un ácido carboxílico. La hidrólisis o la síntesis de ésteres son dependientes de la actividad de agua en el queso así como de la presencia de otros ácidos grasos libres y alcoholes (Broadbent y Steele, 2005). Las rutas metabólicas del catabolismo de ácidos grasos fueron reconstruídas en su totalidad mediante la anotación de GhostKOALA (Figura 15).

De acuerdo a lo observado en la anotación funcional del metagenoma del queso Cotija, los ácidos grasos libres son sustrato para reacciones de esterificación con alcoholes y tioles en la producción de ésteres y tioésteres. Los alcoholes pueden ser metabolitos de la degradación de carbohidratos y de los mismos ácidos grasos mientras que los tioles provienen del metabolismo de aminoácidos azufrados. Además, la degradación de ácidos grasos por  $\beta$ -oxidación con la generación concomitante de metil cetonas (Collins et al., 2003), lleva a la degradación de éstas últimas a su respectivo alcohol secundario por acción de la carbonil reductasa I (Tabla 7).

Además del catabolismo de aminoácidos y ácidos grasos, hay otros sustratos en el queso cuya degradación genera compuestos identificados como parte del perfil sensorial en quesos madurados. Tal es el caso de compuestos derivados del benceno como el tolueno y el xileno, pues se ha demostrado que su concentración en el queso está relacionada con la composición del cultivo iniciador (Centeno et al., 2002; Wang et al., 2012), y su degradación conduce a la formación de alcoholes como el 2,3-dimetilfenol y el 3,4-dimetilfenol. El primer compuesto se ha descrito con olor característico a humo y tinta y de sabor dulce, mientras que el segundo tiene olor a establo con notas fecales (Czerny et al., 2011). Éste se ha identificado en la superficie de quesos madurados (Urbach, 1997) y particularmente en el queso Cotija podría explicar ciertas notas identificadas empíricamente en el perfil sensorial. Cabe mencionar que dado que el potencial metabólico del consorcio para degradar aminoácidos, ácidos grasos libres y otros productos intermedios, la gama de compuestos que participan en el perfil de olor y sabor del queso es mucho más amplia. Particularmente la gran cantidad de aldehído y alcohol deshidrogenasas encontradas en la Tabla de anotación es un indicio de que el consorcio bacteriano tiene la capacidad para generar una mezcla compleja de compuestos con una amplia variedad de cetonas, aldehídos y ácidos orgánicos.

#### Búsqueda de bacteriocinas y genes relacionados

El desarrollo de BAL en sí mismo implica la inhibición de microorganismos indeseables y su habilidad para producir bacteriocinas es de importancia básica en estrategias de biopreservación en alimentos extendiendo la vida de anaquel y mejorando la seguridad del mismo (Ross et al., 2002).

El método de identificación de clusters de genes relacionados con bacteriocinas permitió la identificación de un total de 126 contigs de diferentes tamaños que contenían al menos un gen con la anotación de interés (Tabla 8), de los cuales, 43 tuvieron al menos un gen codificante para bacteriocina, 74 tuvieron sólo genes con anotación de proteínas de inmunidad pero sin el gen de la bacteriocina y sólo en 17 fue posible identificar tanto el gen de la bacteriocina como para su respectiva inmunidad (algunos ejemplos en la Figura 16). Por otro lado, nueve contigs tuvieron genes relacionados con la síntesis de lantibióticos pero el gen codificante de la bacteriocina no fue encontrado.

El resto de los ORFs tuvieron anotación para bacteriocinas de la clase II, es decir no lantibióticos, de los cuáles, la mayoría pertenecen a la subclase IIa de tipo pediocinas, o a la subclase IIc de bacteriocinas circulares (Cotter et al., 2005), en la que se incluye a la subtolisina A. Los genes con anotación de bacteriocina tipo lactococcina que incluye proteínas lineales de un sólo péptido no del tipo de las pediocinas, como las lactococcinas A y B (Nes et al., 2006) se consideran dentro de la clase IID (Cotter et al., 2005). También 9 de las 10 bacteriocinas detectadas con BAGEL3 fueron de clase II: cuatro fueron clasificadas como IIb, conformadas por dos péptidos; cuatro como de clase IID constituidas por un sólo péptido que no es del tipo pediocina; una fue identificada sólo como de clase II y otra más como putativa.

Tabla 8. Identificadores de Pfam usados en la búsqueda de bacteriocinas

Pfam	Descripción del dominio	Coincidencias en tabla*
PF02052	Gallidermina	1
PF01721	Bacteriocina de clase II	12
PF10439	Bacteriocina de clase II con péptido líder de doble glicina	15
PF08951	Inmunidad de enterocina A	88
PF05147	Lantionina sintetasa tipo C	3
PF04369	Familia de las de tipo lactococcinas	7
PF04738	Amino terminal de lantibiótico deshidratasa	3
PF04737	Amino terminal de lantibiótico deshidratasa	3
PF13376	YdeI o OmpD, asociado a protección a bacteriocinas	13
PF12173	Bacteriocina cíclica de clase IIc tipo gasicina A	2
PF11420	Bacteriocina subtilosina A	2
PF09683	Bacteriocina lactococcina 972	3

\* Indica número de ORFs con anotación para dicha enzima en la tabla de anotación global.

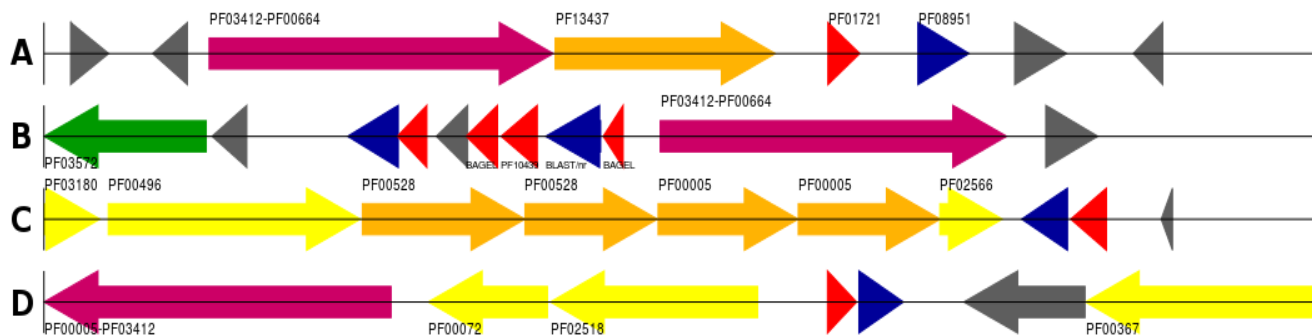


Figura 16. Ejemplos del contexto genómico de algunos clusters con gen estructural y de inmunidad. A-C) clusters identificados en el metagenoma del queso Cotija; D) cluster de bacteriocina en el genoma de *Pediococcus pentosaceus* ATCC 25745. Código de color: rojo, gen estructural de la bacteriocina (PF01721, a menos que se especifique otro); azul, gen de inmunidad (PF08951, a menos que se especifique otro); naranja, gen con dominio de transportador; verde, gen con dominio de peptidasa; rosa, gen con dominios de transporte y peptidasa; amarillo, otra anotación; gris, gen hipotético. Figura generada con el módulo GenomeDiagram de Python v2.7.

De los 52 genes estructurales para bacteriocinas identificados en el metagenoma del queso Cotija, sólo uno fue de clase I, es decir lantibiótico (Tabla 8). El contig en el que se encontró tuvo en la vecindad dos genes relacionados con la biosíntesis de lantibióticos, una proteína de la familia RepB y otros tres genes sin anotación clara que podrían estar asociados con el sistema de secreción de lantibióticos que suelen transcribirse en unidades múltiples de tres o cuatro genes (Kotelnikova y Gelfand, 2002).

Un tipo de bacteriocina encontrada en el metagenoma del queso que no entra en ninguna de las clasificaciones aceptadas actualmente son las lactococcinas 972, descritas como un homodímero que actúa inhibiendo la síntesis del septum durante la división celular en lugar de formar poros en la membrana como es el mecanismo clásico de acción de las bacteriocinas de clase II (Martínez et al., 1996).

La ventaja de conocer la anotación funcional de los genes contiguos a las secuencias de interés, es la posibilidad de identificar ORFs candidatos a nuevas bacteriocinas y/o genes de inmunidad de acuerdo a su contexto genómico.

En general, la anotación de los genes que forman parte de los clusters y que tienen anotación diferente a genes de inmunidad, bacteriocinas o síntesis, entran en alguna de las siguientes descripciones de acuerdo a las etiquetas de Pfam de los dominios que contienen: 1) Proteínas de transporte: transportador ABC o proteína de secreción de la familia HylD; 2) Peptidasas: C39 o S49; 3) Sistema de regulación de dos componentes: receptor de regulación de la respuesta, histidina cinasa, regulador de la transcripción o proteínas de transducción de señales; 4) Genes asociados a transferencia horizontal: conjugación, integrasas de fagos, resolvasas o transposasas; 5) Antídotos a toxinas e inhibidores del crecimiento; 6) Proteínas de respuesta a estrés: proteína universal de respuesta a estrés, respuesta a estrés osmótico o respuesta a estrés alcalino; 7) Replicación de ADN; 8) Genes putativos o sin anotación.

La presencia de estos genes se puede explicar desde el punto de vista del bioprocesamiento de la bacteriocina. Las peptidasas y transportadores de membrana se requieren para la escisión del péptido señal con el fin de exportar y activar a la bacteriocina en un sólo paso, por esta razón es común encontrar en la vecindad de la bacteriocina genes que codifican para proteínas que tienen ambos dominios (Ennahar et al., 2000). También se han reportado bacteriocinas de clase II sin péptido señal que no necesitan dicho procesamiento (Nes et al., 2006).

La presencia de genes de regulación también se espera en el cluster ya que la expresión de los genes codificantes de bacteriocinas está regulada y es dependiente de la presencia de señales de *quorum sensing*. Dicha regulación se basa en sistemas de dos componentes de transducción de señales que consiste en una cinasa de histidina unida a membrana y un regulador de respuesta que se activa gracias a la presencia de factores de inducción de prepéptido, también llamados feromonas, que son a nivel de secuencia parecidos a las bacteriocinas pero con mucha menor o nula actividad que forman parte del cluster (Ennahar et al., 2000; Snyder y Worobo, 2014).

Además de peptidasas, transportadores, genes de regulación y los genes estructurales y de inmunidad, el resto de las anotaciones son dispensables del cluster aunque aparezcan con relativa frecuencia. Tal es el caso de los genes relacionados con transferencia horizontal cuya presencia en el cluster tiene lógica dado que los genes de bacteriocinas y su procesamiento suelen estar codificados en plásmidos, cromosomas y/o transposones (Klaenhammer, 1993), y su susceptibilidad a ser transferidos horizontalmente es muy alta entre las bacterias lácticas presentes en alimentos fermentados (Rossi et al., 2014). Es probable que en cada evento de transferencia horizontal, la célula donadora se deshaga de un set de genes que confiere más de una ventaja a la receptora si

estos se encuentran físicamente próximos en el genoma ya sea formando parte de un operón o de un cluster (Lawrence y Roth, 1996). Esto podría explicar la presencia de genes relacionados con respuesta a estrés, antídotos a toxinas y en general aquellas que inhiben el crecimiento de otras bacterias.

Por otro lado, se esperaba observar el gen de la proteína de inmunidad río abajo del gen estructural de la bacteriocina ya que ambos genes son comúnmente cotranscritos y por lo general se reportan en este orden dentro del cluster (Kjos et al., 2011), pero esto no ocurrió en la mayoría de los casos. Sin embargo, así como se encontró una amplia variedad de genes codificantes con anotaciones diversas en la vecindad de los genes de bacteriocinas y de inmunidad sin un patrón definido de sintenia ni presencia de genes, también hubo muchos casos de genes sin anotación o hipotéticos (Figura 16). Éstos son de especial interés sobre todo en los casos de clusters donde sólo se encontró el gen estructural de la bacteriocina sin el gen de inmunidad o viceversa, situación que encontramos frecuentemente en este análisis, y más aún si la proteína hipotética se encontraba flanqueando dicho gen lo que podría indicar que dichas secuencias corresponden a genes que codifican para proteínas, ya sea de inmunidad o bacteriocinas con estructura primaria nueva, que fue imposible anotar por similitud con el método utilizado.

Adicionalmente, se observó una cantidad desproporcionada de genes de inmunidad (104 en total) con respecto al número de genes con anotación de bacteriocinas (52 incluyendo las de BAGEL3) y en muchos casos éstos no estaban flanqueados por secuencias de genes sin anotación lo que implica que posiblemente las proteínas de inmunidad son tan útiles o más que la propia bacteriocina, es decir, el hecho de poseer una batería de genes de inmunidad en el genoma implica una ventaja real para la cepa que la posee aún cuando no sea productora de bacteriocinas. Se observó de hecho un caso particular donde el contig contenía cuatro copias del gen de inmunidad y ningún gen codificante para la bacteriocina.

El fenómeno observado en este metagenoma se relaciona con lo que se conoce como inmunidad cruzada, es decir que los microorganismos productores de bacteriocinas son también propietarios de genes que podrían conferirles inmunidad contra otras bacteriocinas (Kjos et al., 2010). Además, se ha reportado que cepas no productoras pueden poseer genes de inmunidad que son probablemente remanentes de clusters de genes de bacteriocinas y que podrían hacer a las cepas resistentes si se expresan adecuadamente (Fimland et al., 2002; Mørretrø et al., 2005). Dicho fenómeno ha sido descrito como mimetismo de inmunidad, y se refiere a la resistencia observada en las cepas que no son productoras de bacteriocinas que poseen genes de inmunidad o a la inmunidad obtenida como consecuencia de la producción de una bacteriocina estrechamente relacionada con otra, es decir que el mimetismo puede ser resultado de la posesión o adquisición de genes de inmunidad homólogos (Draper et al., 2009).

Dada la alta frecuencia de genes de inmunidad no asociados a genes estructurales de bacteriocinas y sin proteínas hipotéticas adyacentes, el mimetismo de inmunidad podría jugar un papel importante como estrategia de sobrevivencia para las cepas bacterianas exitosas presentes al tiempo final de la maduración del queso Cotija.

#### 8.4. Reconstrucción de genomas de las especies dominantes

Gracias a que el metagenoma fue secuenciado con alta profundidad, fue posible reconstruir los genomas de las tres especies dominantes, cada uno en diferente número de fragmentos. La cobertura estimada en todos los casos es de más de 1400 veces y en ningún caso el número de gaps reales o de cobertura cero es de más de 78 (Tabla 9) y como se tienen cubiertas casi por completo las referencias y los contigs fueron orientados sobre las mismas (Figura 17), se considera que todas las reconstrucciones se encuentran a nivel de *draft* de alta calidad (Chain et al., 2009). Los porcentajes de lecturas mapeadas en todos los casos son consistentes con la abundancia relativa de dichas especies estimada mediante la proporción de marcadores filogenéticos (Tabla 2).

Tabla 9. Estadísticas de la reconstrucción de genomas

	<i>L. plantarum</i>	<i>L. mesenteroides</i>	<i>W. paramesenteroides</i>
	WCFS1	ATCC 8293	ATCC 33313
Lecturas mapeadas (%)	38.74	11.27	26.85
Cobertura estimada	1,863 x	1,473 x	2,135 x
Número total de contigs	1,560	297	992
Número de gaps no empalmantes	78	40	42
N50 / N90 (bp)	4,985 / 825	36,455 / 2,977	3,148 / 1,002
Tamaño promedio de contig	2,201	6,744	2,196
Contenido de GC (%)	44.37	37.90	38.27
Número de genes	3,998	2,141	2,755

Con el fin de establecer si las especies dominantes en el metagenoma son potenciales productores importantes de bacteriocinas, se hizo una búsqueda de los clusters de genes con dicha anotación en los contigs utilizados en la reconstrucción de los genomas. De esta búsqueda resultó que el genoma reconstruido de *W. paramesenteroides* no contiene regiones genómicas codificantes para bacteriocinas ni genes de inmunidad o algún otro relacionado. Por otro lado, en el genoma de *L. mesenteroides* se identificó un cluster con una bacteriocina putativa acompañada de siete ORFs de proteínas hipotéticas adyacentes, además de dos clusters más con genes de inmunidad pero sin gen codificante para la bacteriocina. Finalmente, en el genoma de *L. plantarum* no se encontraron genes con anotación para bacteriocinas, pero sí se identificaron 11 genes en diferentes contigs con anotación para proteínas de inmunidad donde sólo uno de ellos tuvo proteínas hipotéticas adyacentes. De cualquier modo, sería necesario realizar trabajo experimental para corroborar que alguno de estos genes corresponde a una bacteriocina nueva.



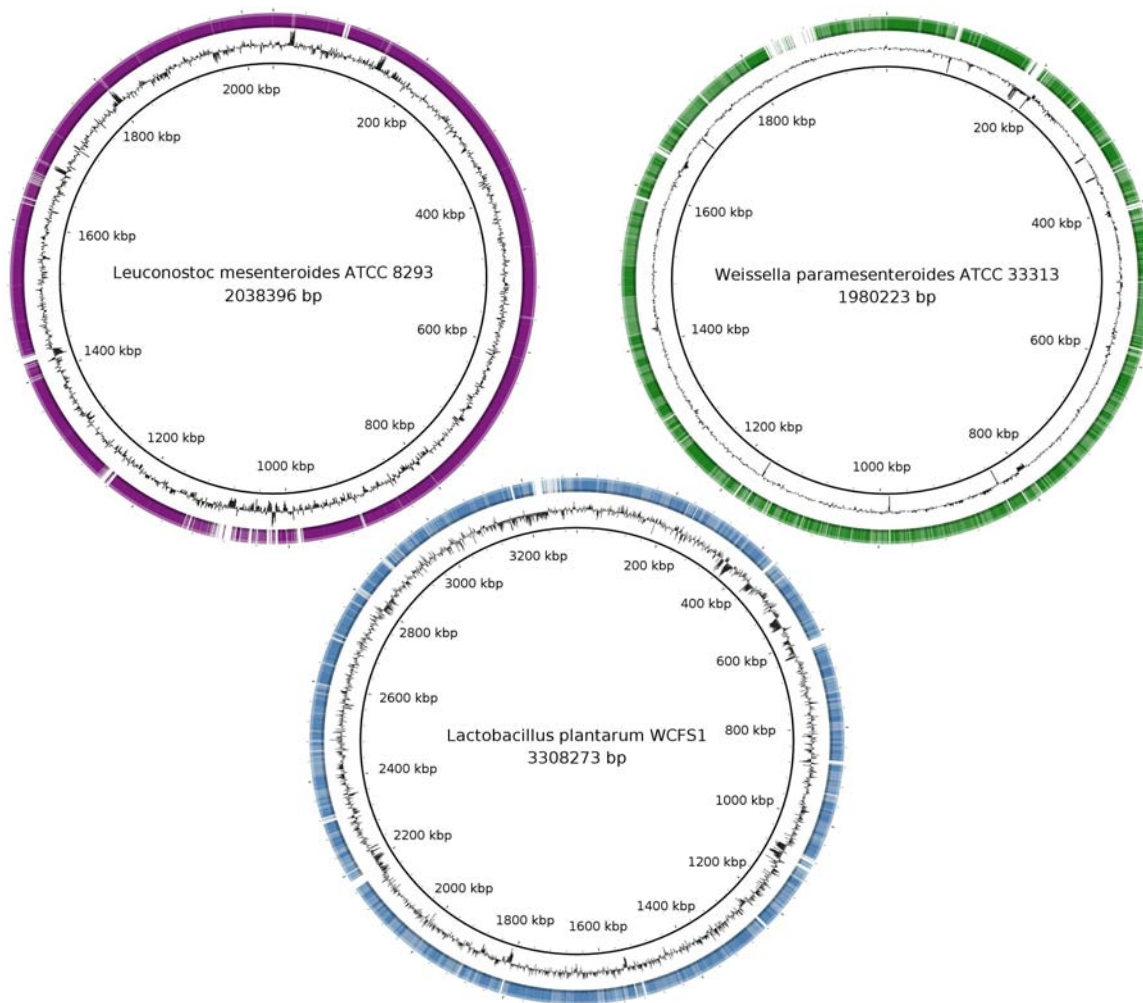


Figura 17. Comparación del genoma reconstruido de cada especie contra su respectiva referencia. El anillo central corresponde al genoma de referencia, el de enmedio representa su contenido de GC y el exterior es el alineamiento del genoma reconstruido con respecto a su referencia, mayor intensidad de color implica mayor similitud. Figura generada con el programa BRIG v0.95 (Alikhan et al., 2011).

## 9. Conclusiones

El enfoque metagenómico y la aplicación de las herramientas bioinformáticas propuestas en este trabajo permitieron llevar a cabo la caracterización tanto taxonómica como del potencial metabólico del consorcio bacteriano del queso Cotija con gran detalle. Esta estrategia es aplicable al análisis de cualquier metagenoma.

La microbiota bacteriana del queso Cotija a los tres meses de maduración está conformada casi en su totalidad por *Firmicutes* y tres géneros son los dominantes: *Lactobacillus*, *Leuconostoc* y *Weissella*, cuyos genomas fueron reconstruidos a nivel de *draft* de alta calidad. *Leuconostoc* y *Weissella* son detectados por primera vez en este trabajo y su dominancia es atípica en quesos madurados por lo que su presencia y abundancia pueden ser considerados como una firma distintiva del queso Cotija Región de Origen.

Se observó que el consorcio bacteriano de este queso está compuesto por más de 500 taxa no dominantes (<1% de

la abundancia total relativa) en los que se incluyen bacterias y *Achaegas* halófilas, lo que refleja la complejidad de la comunidad y lo infructuoso que resultaría intentar diseñar un cultivo iniciador con el fin de elaborar un queso idéntico fuera de la región de origen.

La alta profundidad de secuenciación permitió corroborar que el alimento está libre de bacterias patógenas y que las cuentas de coliformes están dentro de lo establecido por las normas, lo que refleja las buenas prácticas de higiene y la calidad de las materias primas con las que se elabora el queso Cotija, y que no representa un riesgo para la salud del consumidor aún cuando es elaborado con leche no pasteurizada.

El consorcio posee el potencial metabólico para generar una mezcla de compuestos relacionados con el perfil de olor y sabor a partir principalmente del catabolismo de aminoácidos y ácidos grasos libres, y que podría estar enriquecida en metabolitos derivados de aminoácidos de cadena ramificada así como de una gran variedad de aldehídos y alcoholes.

Los resultados indican que ninguna de las especies dominantes posee un arsenal importante de genes codificantes para bacteriocinas, sin embargo, dentro del consorcio existe una amplia variedad de genes que codifican para proteínas de inmunidad lo que sugiere que el mimetismo de inmunidad es una estrategia de sobrevivencia para las bacterias que forman parte de la comunidad en el producto madurado por tres meses.

## **10. Perspectivas**

Dentro de la lista de tareas derivadas de este análisis que quedaron fuera de los alcances de este trabajo pero que resultan de interés para completar el entendimiento de la estructura poblacional bacteriana en el queso Cotija y sus funciones están:

Enfocar esfuerzos en el aislamiento de los géneros identificados en el metagenoma, especialmente los dominantes.

Analizar la mezcla de compuesto volátiles que confiere el perfil sensorial del queso Cotija mediante técnicas como cromatografía de líquidos o gases con el fin de establecer qué compuestos forman parte del perfil de olor y sabor del alimento.

Usar la reconstrucción de los genomas de las especies dominantes para hacer la genómica comparativa contra especies provenientes de otros alimentos o ambientes.

Llevar a cabo la reconstrucción de los genomas de las especies subdominantes.

Evaluar la actividad antimicrobiana de los genes propuestos como candidatos a nuevas bacteriocinas mediante técnicas moleculares de clonación y expresión.

## Referencias

- Ackert, L., 2012. Sergei Vinogradskii and the Cycle of Life: From the Thermodynamics of Life to Ecological Microbiology, 1850-1950. Springer Science & Business Media.
- Alais, C., Linden, G., 1991. Milk and dairy products, in: Food Biochemistry, Ellis Horwood Series in Food Science and Technology. Springer US, pp. 148–173.
- Alikhan, N.-F., Petty, N.K., Zakour, N.L.B., Beatson, S.A., 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12, 402. doi:10.1186/1471-2164-12-402
- Álvarez, B., Barragán, L., Chombo, M., 2005. Reglas de uso: marca colectiva : queso Cotija : región de origen. Asociación Regional de Productores.
- Amann, R.I., Ludwig, W., Schleifer, K.H., 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59, 143–169.
- Ampe, F., ben Omar, N., Moizan, C., Wacher, C., Guyot, J.-P., 1999. Polyphasic Study of the Spatial Distribution of Microorganisms in Mexican Pozol, a Fermented Maize Dough, Demonstrates the Need for Cultivation-Independent Methods To Investigate Traditional Fermentations. Appl Environ Microbiol 65, 5464–5473.
- Applied Biosystems, 2009. Chemistry Guide DNA Sequencing by Capillary Electrophoresis, Applied Biosystems Chemistry Guide, 2<sup>nd</sup> Ed. Disponible en: [https://www3.appliedbiosystems.com/cms/groups/mcb\\_support/documents/generaldocuments/cms\\_041003.pdf](https://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_041003.pdf) [Último acceso 1 Octubre 2015].
- Aslani, M.M., Alikhani, M.Y., Zavari, A., Yousefi, R., Zamani, A.R., 2011. Characterization of enteroaggregative Escherichia coli (EAEC) clinical isolates and their antibiotic resistance pattern. International Journal of Infectious Diseases 15, e136–e139. doi:10.1016/j.ijid.2010.10.002
- Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., Segata, N., 2015. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ 3. doi:10.7717/peerj.1029
- Assefa, S., Keane, T.M., Otto, T.D., Newbold, C., Berriman, M., 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics 25, 1968–1969. doi:10.1093/bioinformatics/btp347
- Axelsson, L., 2004. Lactic acid bacteria: Classification and physiology, in: Salminen, S., von Wright, A., Ouwehand, A. (Eds.), Lactic Acid Bacteria: Microbiology and Functional Aspects. Marcel Dekker, Inc., New York, pp. 1–66.
- Bautista-Gallego, J., Alessandria, V., Fontana, M., Bisotti, S., Taricco, S., Dolci, P., Cocolin, L., Rantsiou, K., 2014. Diversity and functional characterization of *Lactobacillus* spp. isolated throughout the ripening of a hard cheese. Int. J. Food Microbiol. 181, 60–66. doi:10.1016/j.ijfoodmicro.2014.04.020
- Bennett, S., 2004. Solexa Ltd. Pharmacogenomics 5, 433–438. doi:10.1517/14622416.5.4.433
- Beresford, T., Williams, A., 2004. The Microbiology of Cheese Ripening, in: Patrick F. Fox, P.L.H.M. Timothy M. Cogan and Timothy P. Guinee (Ed.), Cheese: Chemistry, Physics and Microbiology, Cheese Chemistry, Physics and Microbiology. Academic Press, pp. 287–317.
- Bissett, D., Anderson, R., 1974. Lactose and D -galactose metabolism in group N streptococci: presence of enzymes for both the d-galactose 1-phosphate and D -tagatose 6-phosphate pathways. J. Bacteriol. 117, 318–320.
- Björkroth, J., Holzappel, W., 2006. Genera *Leuconostoc*, *Oenococcus* and *Weissella*, in: Dr, M.D.P., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E. (Eds.), The Prokaryotes. Springer US, pp. 267–319.
- Boeckmann, B., Blatter, M.-C., Famiglietti, L., Hinz, U., Lane, L., Roehert, B., Bairoch, A., 2005. Protein variety and functional diversity: Swiss-Prot annotation in its biological context. Comptes Rendus Biologies 328, 882–899. doi:10.1016/j.crv.2005.06.001
- Bonilla-Rosso, G., Eguiarte, L.E., Romero, D., Travisano, M., Souza, V., 2012. Understanding microbial community diversity metrics derived from metagenomes: performance evaluation using simulated data sets. FEMS Microbiol Ecol 82, 37–49. doi:10.1111/j.1574-6941.2012.01405.x
- Bravo, M.A., 2008. Estudio de las poblaciones microbianas de interés biotecnológico aisladas de queso cotija (Tesis de licenciatura). Universidad Nacional Autónoma de México, Facultad de Química, México.
- Brimble, M.A., Edwards, P.J., Harris, P.W.R., Norris, G.E., Patchett, M.L., Wright, T.H., Yang, S.-H., Carley, S.E., 2015. Synthesis of the Antimicrobial S-Linked Glycopeptide, Glycocin F. Chem. Eur. J. 21, 3556–3561. doi:10.1002/chem.201405692

- Broadbent, J.R., Steele, J.L., 2005. Lactic acid bacteria genomics and cheese flavor. *ASM News* 71, 121–128.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421
- Casillas, H.J.L., 2013. Caracterización molecular de la microbiota bacteriana del queso Cotija elaborado en diversas regiones de México (Tesis de maestría). Universidad Nacional Autónoma de México, Facultad de Química, México.
- Centeno, J.A., Tomillo, F.J., Fernández-García, E., Gaya, P., Nuñez, M., 2002. Effect of Wild Strains of *Lactococcus lactis* on the Volatile Profile and the Sensory Characteristics of Ewes' Raw Milk Cheese. *Journal of Dairy Science* 85, 3164–3172. doi:10.3168/jds.S0022-0302(02)74404-4
- Chain, P.S.G., Grafham, D.V., Fulton, R.S., FitzGerald, M.G., Hostetler, J., Muzny, D., Ali, J., Birren, B., Bruce, D.C., Buhay, C., Cole, J.R., Ding, Y., Dugan, S., Field, D., Garrity, G.M., Gibbs, R., Graves, T., Han, C.S., Harrison, S.H., Highlander, S., Hugenholtz, P., Khouri, H.M., Kodira, C.D., Kolker, E., Kyrpides, N.C., Lang, D., Lapidus, A., Malfatti, S.A., Markowitz, V., Metha, T., Nelson, K.E., Parkhill, J., Pitluck, S., Qin, X., Read, T.D., Schmutz, J., Sozhamannan, S., Sterk, P., Strausberg, R.L., Sutton, G., Thomson, N.R., Tiedje, J.M., Weinstock, G., Wollam, A., Detter, J.C., 2009. Genome Project Standards in a New Era of Sequencing. *Science* 326. doi:10.1126/science.1180614
- Chombo-Morales, P., Kirchmayr, M., Gschaedler, A., Lugo-Cervantes, E., Villanueva-Rodríguez, S., 2016. Effects of controlling ripening conditions on the dynamics of the native microbial population of Mexican artisanal Cotija cheese assessed by PCR-DGGE. *LWT - Food Science and Technology* 65, 1153–1161. doi:10.1016/j.lwt.2015.09.044
- Cilia, V., Lafay, B., Christen, R., 1996. Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol Biol Evol* 13, 451–461.
- Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42, D633–D642. doi:10.1093/nar/gkt1244
- Collins, Y.F., McSweeney, P.L.H., Wilkinson, M.G., 2003. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. *International Dairy Journal* 13, 841–866. doi:10.1016/S0958-6946(03)00109-2
- Compeau, P.E.C., Pevzner, P.A., Tesler, G., 2011. How to apply de Bruijn graphs to genome assembly. *Nat Biotech* 29, 987–991. doi:10.1038/nbt.2023
- Corsetti, A., Gobbetti, M., Smacchi, E., 1996. Antibacterial activity of sourdough lactic acid bacteria: isolation of a bacteriocin-like inhibitory substance from *Lactobacillus sanfrancisco* C57. *Food Microbiology* 13, 447–456. doi:10.1006/fmic.1996.0051
- Cortés, D.L.A., 2009. Aplicación de la reacción en cadena de la polimerasa en la búsqueda de bacterias ácido lácticas en un queso artesanal mexicano (Tesis de licenciatura). Universidad Nacional Autónoma de México, Facultad de Química, México.
- Cotter, P.D., Hill, C., Ross, R.P., 2005. Bacteriocins: developing innate immunity for food. *Nat Rev Micro* 3, 777–788. doi:10.1038/nrmicro1273
- Curioni, P.M.G., Bosset, J.O., 2002. Key odorants in various cheese types as determined by gas chromatography-olfactometry. *International Dairy Journal* 12, 959–984. doi:10.1016/S0958-6946(02)00124-3
- Czerny, M., Brueckner, R., Kirchhoff, E., Schmitt, R., Buettner, A., 2011. The Influence of Molecular Structure on Odor Qualities and Odor Detection Thresholds of Volatile Alkylated Phenols. *Chem. Senses* bjr009. doi:10.1093/chemse/bjr009
- d'Acampora Zellner, B., Dugo, P., Dugo, G., Mondello, L., 2008. Gas chromatography-olfactometry in food flavour analysis. *J Chromatogr A* 1186, 123–143. doi:10.1016/j.chroma.2007.09.006
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi:10.1128/AEM.03006-05
- Díaz, R.G., Wachter, R.C., 2003. Métodos para el estudio de comunidades microbianas en alimentos fermentados. *Microbiología* 45, 30–40.
- Doroghazi, J.R., Metcalf, W.W., 2013. Comparative genomics of actinomycetes with a focus on natural product

- biosynthetic genes. *BMC Genomics* 14, 611. doi:10.1186/1471-2164-14-611
- Draper, L.A., Grainger, K., Deegan, L.H., Cotter, P.D., Hill, C., Ross, R.P., 2009. Cross-immunity and immune mimicry as mechanisms of resistance to the lantibiotic lactacin 3147. *Mol. Microbiol.* 71, 1043–1054. doi:10.1111/j.1365-2958.2008.06590.x
- Dröge, J., McHardy, A.C., 2012. Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. *Brief Bioinform* 13, 646–655. doi:10.1093/bib/bbs031
- Dugat-Bony, E., Straub, C., Teissandier, A., Onésime, D., Loux, V., Monnet, C., Irlinger, F., Landaud, S., Leclercq-Perlat, M.-N., Bento, P., Fraud, S., Gibrat, J.-F., Aubert, J., Fer, F., Guédon, E., Pons, N., Kennedy, S., Beckerich, J.-M., Swennen, D., Bonnarne, P., 2015. Overview of a surface-ripened cheese community functioning by meta-omics analyses. *PLoS ONE* 10, e0124360. doi:10.1371/journal.pone.0124360
- Duthoit, F., Godon, J., Montel, M., 2003. Bacterial Community Dynamics during Production of Registered Designation of Origin Salers Cheese as Evaluated by 16S rRNA Gene Single-Strand Conformation Polymorphism Analysis. *Appl. Environ. Microbiol.* 69, 3840–3848. doi:10.1128/AEM.69.7.3840-3848.2003
- Eddy, S.R., 1998. Profile hidden Markov models. *Bioinformatics* 14, 755–763.
- Eddy, S.R., 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol* 7, e1002195. doi:10.1371/journal.pcbi.1002195
- Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi:10.1093/bioinformatics/btq461
- Ennahar, S., Sashihara, T., Sonomoto, K., Ishizaki, A., 2000. Class IIa bacteriocins: biosynthesis, structure and activity. *FEMS Microbiology Reviews* 24, 85–106. doi:10.1111/j.1574-6976.2000.tb00534.x
- Ercolini, D., 2004. PCR-DGGE fingerprinting: novel strategies for detection of microbes in food. *Journal of Microbiological Methods* 56, 297–314. doi:10.1016/j.mimet.2003.11.006
- Ercolini, D., De Filippis, F., La Stora, A., Iacono, M., 2012. “Remake” by high-throughput sequencing of the microbiota involved in the production of water buffalo mozzarella cheese. *Appl. Environ. Microbiol.* 78, 8142–8145. doi:10.1128/AEM.02218-12
- Escalante, A., 2007. Capítulo 12. Ecología molecular en el estudio de comunidades bacterianas, in: *Ecología molecular*. Instituto Nacional de Ecología, México, D.F., p. 592.
- Escobar, Z.A., 2012. Estudio de la distribución espacial de la microbiota bacteriana en queso Cotija por la técnica de Fish (Tesis de maestría). Universidad Nacional Autónoma de México, Facultad de Química, México.
- Estrada-Garcia, T., Perez-Martinez, I., Bernal-Reynaga, R., Zaidi, M.B., 2014. Enterococci: A Pathogen Bridging the North and South. *Curr Trop Med Rep* 1, 88–96. doi:10.1007/s40475-014-0018-7
- Felis, G., Dellaglio, F., 2007. Taxonomy of *Lactobacilli* and *Bifidobacteria*. *Curr Issues Intest Microbiol* 8, 44–61.
- Feurer, C., Irlinger, F., Spinnler, H. e., Glaser, P., Vallaes, T., 2004. Assessment of the rind microbial diversity in a farmhouse-produced vs a pasteurized industrially produced soft red-smear cheese using both cultivation and rDNA-based methods. *Journal of Applied Microbiology* 97, 546–556. doi:10.1111/j.1365-2672.2004.02333.x
- Fichot, E.B., Norman, R.S., 2013. Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. *Microbiome* 1, 10. doi:10.1186/2049-2618-1-10
- Filippis, F.D., Stora, A.L., Stellato, G., Gatti, M., Ercolini, D., 2014. A Selected Core Microbiome Drives the Early Stages of Three Popular Italian Cheese Manufactures. *PLOS ONE* 9, e89680. doi:10.1371/journal.pone.0089680
- Fimland, G., Eijsink, V.G.H., Nissen-Meyer, J., 2002. Comparative studies of immunity proteins of pediocin-like bacteriocins. *Microbiology (Reading, Engl.)* 148, 3661–3670.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. *Nucl. Acids Res.* 42, D222–D230. doi:10.1093/nar/gkt1223
- Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. *Nucl. Acids Res.* 44, D279–D285. doi:10.1093/nar/gkv1344
- Fox, G.E., Wisotzkey, J.D., Jurtshuk, P., 1992. How close is close: 16S rRNA sequence identity may not be

- sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.* 42, 166–170.
- Fox, P.F., McSweeney, P.L.H., Cogan, T.M., 2000. *Fundamentals of Cheese Science*, 2000 edition. ed. Springer, Gaithersburg, MD.
- Fuka, M.M., Wallisch, S., Engel, M., Welzl, G., Havranek, J., Schloter, M., 2013. Dynamics of Bacterial Communities during the Ripening Process of Different Croatian Cheese Types Derived from Raw Ewe's Milk Cheeses. *PLoS One* 8. doi:10.1371/journal.pone.0080734
- Fuller, C.W., Middendorf, L.R., Benner, S.A., Church, G.M., Harris, T., Huang, X., Jovanovich, S.B., Nelson, J.R., Schloss, J.A., Schwartz, D.C., Vezenov, D.V., 2009. The challenges of sequencing by synthesis. *Nat Biotech* 27, 1013–1023. doi:10.1038/nbt.1585
- García, S.V., 2006. Aislamiento de microorganismos con mayor actividad lipolítica del queso Cotija (Tesis de licenciatura). Universidad Nacional Autónoma de México, Facultad de Química, México.
- García, S.V., 2011. Detección de microorganismos lipolíticos durante las etapas de elaboración de un queso artesanal, mediante técnicas moleculares (Tesis de maestría). Universidad Nacional Autónoma de México, Facultad de Química, México.
- Gilbert, J.A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B., Weynberg, K., Huse, S., Hughes, M., Joint, I., Somerfield, P.J., Mühling, M., 2010. The Taxonomic and Functional Diversity of Microbes at a Temperate Coastal Site: A “Multi-Omic” Study of Seasonal and Diel Temporal Variation. *PLOS ONE* 5, e15545. doi:10.1371/journal.pone.0015545
- Giovannoni, S.J., DeLong, E.F., Schmidt, T.M., Pace, N.R., 1990. Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. *Appl. Environ. Microbiol.* 56, 2572–2575.
- Glenn, T., 2014. 2014 NGS Field Guide: Overview | The Molecular Ecologist [WWW Document]. URL <http://www.molecularecologist.com/next-gen-fieldguide-2014/> (accessed 8.24.15).
- Gómez, C.B.E., 2010. Identificación de bacterias halófilas presentes en queso Cotija por métodos independientes de cultivo (Tesis de licenciatura). Universidad Nacional Autónoma de México, Facultad de Química, México.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi:10.1038/nbt.1883
- Guo, W., Jiang, L., Bhasin, S., Khan, S.M., Swerdlow, R.H., 2009. DNA extraction procedures meaningfully influence qPCR-based mtDNA copy number determination. *Mitochondrion* 9, 261–265. doi:10.1016/j.mito.2009.03.003
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., Goodman, R.M., 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5, R245–249.
- Harmon, R.J., 1994. Physiology of Mastitis and Factors Affecting Somatic Cell Counts<sup>1</sup>. *Journal of Dairy Science* 77, 2103–2112. doi:10.3168/jds.S0022-0302(94)77153-8
- Hassan, F.A.M., Gawad, M.A.M.A.E.-, Enab, A.K., 2013. Flavour Compounds in Cheese (Review). *Research on Precision Instrument and Machinery, Research on Precision Instrument and Machinery* 2.
- Hernández, M.N., 2007. Identificación de bacterias proteolíticas aisladas de queso Cotija, un estudio microbiológico y fisicoquímico (Tesis de licenciatura). Universidad Nacional Autónoma de México, Facultad de Química, México.
- Hurwitz, B.L., Deng, L., Poulos, B.T., Sullivan, M.B., 2013. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environmental Microbiology* 15, 1428–1440. doi:10.1111/j.1462-2920.2012.02836.x
- Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi:10.1101/gr.5969107
- Ishikawa, M., Kodama, K., Yasuda, H., Okamoto-Kainuma, A., Koizumi, K., Yamasato, K., 2007. Presence of halophilic and alkaliphilic lactic acid bacteria in various cheeses. *Lett. Appl. Microbiol.* 44, 308–313. doi:10.1111/j.1472-765X.2006.02073.x
- Ishikawa, M., Nakajima, K., Itamiya, Y., Furukawa, S., Yamamoto, Y., Yamasato, K., 2005. *Halolactibacillus*

- halophilus* gen. nov., sp. nov. and *Halolactibacillus miurensis* sp. nov., halophilic and alkaliphilic marine lactic acid bacteria constituting a phylogenetic lineage in *Bacillus* rRNA group 1. *Int. J. Syst. Evol. Microbiol.* 55, 2427–2439. doi:10.1099/ijs.0.63713-0
- Ishikawa, M., Yamasato, K., Kodama, K., Yasuda, H., Matsuyama, M., Okamoto-Kainuma, A., Koizumi, Y., 2013. *Alkalibacterium gilvum* sp. nov., slightly halophilic and alkaliphilic lactic acid bacterium isolated from soft and semi-hard cheeses. *Int. J. Syst. Evol. Microbiol.* 63, 1471–1478. doi:10.1099/ijs.0.042556-0
- Jung, J.Y., Lee, S.H., Kim, J.M., Park, M.S., Bae, J.-W., Hahn, Y., Madsen, E.L., Jeon, C.O., 2011. Metagenomic analysis of kimchi, a traditional Korean fermented food. *Appl. Environ. Microbiol.* 77, 2264–2274. doi:10.1128/AEM.02157-10
- Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27
- Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology, Computation Resources for Molecular Biology* 428, 726–731. doi:10.1016/j.jmb.2015.11.006
- Kasianowicz, J.J., Brandin, E., Branton, D., Deamer, D.W., 1996. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A* 93, 13770–13773.
- Kato, H., Mori, H., Maruyama, F., Toyoda, A., Oshima, K., Endo, R., Fuchu, G., Miyakoshi, M., Dozono, A., Ohtsubo, Y., Nagata, Y., Hattori, M., Fujiyama, A., Kurokawa, K., Tsuda, M., 2015. Time-series metagenomic analysis reveals robustness of soil microbiome against chemical disturbance. *DNA Res* dsv023. doi:10.1093/dnares/dsv023
- Kawulok, J., Deorowicz, S., 2015. CoMeta: Classification of Metagenomes Using k-mers. *PLOS ONE* 10, e0121453. doi:10.1371/journal.pone.0121453
- Kennedy, J., Flemer, B., Jackson, S.A., Morrissey, J.P., O’Gara, F., Dobson, A.D.W., 2014. Evidence of a Putative Deep Sea Specific Microbiome in Marine Sponges. *PLoS One* 9. doi:10.1371/journal.pone.0091092
- Kjos, M., Borrero, J., Opsata, M., Birri, D.J., Holo, H., Cintas, L.M., Snipen, L., Hernández, P.E., Nes, I.F., Diep, D.B., 2011. Target recognition, resistance, immunity and genome mining of class II bacteriocins from Gram-positive bacteria. *Microbiology* 157, 3256–3267. doi:10.1099/mic.0.052571-0
- Kjos, M., Snipen, L., Salehian, Z., Nes, I.F., Diep, D.B., 2010. The Abi Proteins and Their Involvement in Bacteriocin Self-Immunity. *Journal of Bacteriology* 192, 2068–2076. doi:10.1128/JB.01553-09
- Klaenhammer, T.R., 1993. Genetics of bacteriocins produced by lactic acid bacteria. *FEMS Microbiol. Rev.* 12, 39–85.
- Klappenbach, J.A., Dunbar, J.M., Schmidt, T.M., 2000. rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* 66, 1328–1333.
- Krebs, C., 2014. CHAPTER 12 - Species Diversity Measures, in: *Ecological Methodology*. Addison-Wesley Educational Publishers, Inc., Boston.
- Kultima, J.R., Sunagawa, S., Li, J., Chen, W., Chen, H., Mende, D.R., Arumugam, M., Pan, Q., Liu, B., Qin, J., Wang, J., Bork, P., 2012. MOCAT: A Metagenomics Assembly and Gene Prediction Toolkit. *PLoS ONE* 7, e47656. doi:10.1371/journal.pone.0047656
- Lavermicocca, P., Valerio, F., Evidente, A., Lazzaroni, S., Corsetti, A., Gobbetti, M., 2000. Purification and characterization of novel antifungal compounds from the sourdough *Lactobacillus plantarum* strain 21B. *Appl. Environ. Microbiol.* 66, 4084–4090.
- Lawrence, J.G., Roth, J.R., 1996. Selfish Operons: Horizontal Transfer May Drive the Evolution of Gene Clusters. *Genetics* 143, 1843–1860.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, K., Bihan, M., Yooseph, S., Methé, B.A., 2012. Analyses of the Microbial Diversity across the Human Microbiome. *PLoS One* 7. doi:10.1371/journal.pone.0032118
- Litopoulou-Tzanetaki, E., Tzanetakis, N., 2014. The Microfloras of Traditional Greek Cheeses, in: Donnelly, C. (Ed.), *Cheese and Microbes*. ASM Press, Washington, DC.
- Liu, Q., Guo, Y., Li, J., Long, J., Zhang, B., Shyr, Y., 2012. Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data. *BMC Genomics* 13, S8. doi:10.1186/1471-2164-13-S8-S8

- Liu, Y., Guo, J., Hu, G., Zhu, H., 2013. Gene prediction in metagenomic fragments based on the SVM algorithm. *BMC Bioinformatics* 14, S12. doi:10.1186/1471-2105-14-S5-S12
- Luo, C., Tsementzi, D., Kyrpides, N.C., Konstantinidis, K.T., 2012. Individual genome assembly from complex community short-read metagenomic datasets. *ISME J* 6, 898–901. doi:10.1038/ismej.2011.147
- Lyu, C., Chen, C., Ge, F., Liu, D., Zhao, S., Chen, D., 2013. A preliminary metagenomic study of puer tea during pile fermentation. *J. Sci. Food Agric.* 93, 3165–3174. doi:10.1002/jsfa.6149
- Maoz, A., Mayr, R., Scherer, S., 2003. Temporal stability and biodiversity of two complex antilisterial cheese-ripening microbial consortia. *Appl. Environ. Microbiol.* 69, 4012–4018.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Goodwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature* 437, 376–380. doi:10.1038/nature03959
- Martínez, B., Suárez, J.E., Rodríguez, A., 1996. Lactococcin 972 : a homodimeric lactococcal bacteriocin whose primary target is not the plasma membrane. *Microbiology*, 142, 2393–2398.
- Marilley, L., Casey, M.G., 2004. Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains. *International Journal of Food Microbiology* 90, 139–159. doi:10.1016/S0168-1605(03)00304-0
- Masoud, W., Vogensen, F.K., Lillevang, S., Abu Al-Soud, W., Sørensen, S.J., Jakobsen, M., 2012. The fate of indigenous microbiota, starter cultures, *Escherichia coli*, *Listeria innocua* and *Staphylococcus aureus* in Danish raw milk and cheeses determined by pyrosequencing and quantitative real time (qRT)-PCR. *International Journal of Food Microbiology* 153, 192–202. doi:10.1016/j.ijfoodmicro.2011.11.014
- Mattes, R.D., 2009. Is There a Fatty Acid Taste? *Annual Review of Nutrition* 29, 305–327. doi:10.1146/annurev-nutr-080508-141108
- McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* 8, e61217. doi:10.1371/journal.pone.0061217
- McSweeney, P.L.H., 2004. Biochemistry of cheese ripening. *International Journal of Dairy Technology* 57, 127–144. doi:10.1111/j.1471-0307.2004.00147.x
- Messens, W., De, V.L., 2002. Inhibitory substances produced by *Lactobacilli* isolated from sourdoughs--a review. *Int. J. Food Microbiol.* 72, 31–43.
- Milesi, M. m., McSweeney, P. l. h., Hynes, E. r., 2008. Viability and contribution to proteolysis of an adjunct culture of *Lactobacillus plantarum* in two model cheese systems: Cheddar cheese-type and soft-cheese type. *Journal of Applied Microbiology* 105, 884–892. doi:10.1111/j.1365-2672.2008.03813.x
- Miller, J.R., Koren, S., Sutton, G., 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95, 315–327. doi:10.1016/j.ygeno.2010.03.001
- MOH Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, 2014. VFDB - comparative pathogenomic composition of *Escherichia*. Virulence Factors of Pathogenic Bacteria. URL <http://www.mgc.ac.cn/cgi-bin/VFs/compvfs.cgi?Genus=Escherichia> [Último acceso 1 Diciembre 2015].
- Molimard, P., Spinnler, H.E., 1996. Review: Compounds Involved in the Flavor of Surface Mold-Ripened Cheeses: Origins and Properties. *Journal of Dairy Science* 79, 169–184. doi:10.3168/jds.S0022-0302(96)76348-8
- Montel, M.-C., Buchin, S., Mallet, A., Delbes-Paus, C., Vuitton, D.A., Desmasures, N., Berthier, F., 2014. Traditional cheeses: rich and diverse microbiota with associated benefits. *Int. J. Food Microbiol.* 177, 136–154. doi:10.1016/j.ijfoodmicro.2014.02.019
- Morales, F., Morales, J.I., Hernández, C.H., Hernández-Sánchez, H., 2011. Isolation and partial characterization of halotolerant lactic acid bacteria from two Mexican cheeses. *Appl. Biochem. Biotechnol.* 164, 889–905. doi:10.1007/s12010-011-9182-6



- Møretrø, T., Naterstad, K., Wang, E., Aasen, I.M., Chaillou, S., Zagorec, M., Axelsson, L., 2005. Sakacin P non-producing *Lactobacillus sakei* strains contain homologues of the sakacin P gene cluster. *Res. Microbiol.* 156, 949–960. doi:10.1016/j.resmic.2005.05.005
- Moter, A., Göbel, U.B., 2000. Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms. *Journal of Microbiological Methods* 41, 85–112. doi:10.1016/S0167-7012(00)00152-4
- Nalbantoglu, U., Cakar, A., Dogan, H., Abaci, N., Ustek, D., Sayood, K., Can, H., 2014. Metagenomic analysis of the microbial community in kefir grains. *Food Microbiol.* 41, 42–51. doi:10.1016/j.fm.2014.01.014
- Nes, I.F., Brede, D.A., Holo, H., 2006. Chapter 17 - The Nonantibiotic Heat-Stable Bacteriocins in Gram-Positive Bacteria, in: Kastin, A.J. (Ed.), *Handbook of Biologically Active Peptides*. Academic Press, Burlington, pp. 107–114.
- NMX-F-735-COFOCALEC, 2011. Norma Mexicana. Sistema producto Leche-Alimentos-Lácteos-Alimento Lácteo Regional-Queso Cotija Artesanal Madurado-Denominacion, Especificaciones y Métodos de Prueba.
- NOM-243-SSA1, 2010. Norma Oficial Mexicana. Productos y servicios. Leche, fórmula láctea, producto lácteo combinado y derivados lácteos. Disposiciones y especificaciones sanitarias. Métodos de prueba.
- Oman, T.J., Boettcher, J.M., Wang, H., Okalibe, X.N., van der Donk, W.A., 2011. Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat Chem Biol* 7, 78–80. doi:10.1038/nchembio.509
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2015. *vegan: Community Ecology Package*.
- Ouwehand, A.C., Vaughan, E.E., 2006. *Gastrointestinal Microbiology*. CRC Press.
- Pace, N.R., Stahl, D., Olsen, G.J., 1985. Analyzing natural microbial populations by rRNA sequences. *ASM News* 51.
- Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi:10.1093/bioinformatics/bts174
- Petrosino, J.F., Highlander, S., Luna, R.A., Gibbs, R.A., Versalovic, J., 2009. Metagenomic Pyrosequencing and Microbial Identification. *Clin Chem* 55, 856–866. doi:10.1373/clinchem.2008.107565
- Pomeon, T., Boucher, F., Escoto, F.C., 2009. Queso Cotija: Denominacion de origen o denominacion genérica. Universidad Autónoma Chapingo, Mexico.
- PROY-NMX-F-700-COFOCALEC, 2012. Norma Mexicana. Sistema Producto Leche-Alimento-Lácteo-Leche Cruda de Vaca-Especificaciones Fisicoquímicas, Sanitarias y Métodos de Prueba.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–596. doi:10.1093/nar/gks1219
- Quigley, L., O'Sullivan, O., Beresford, T.P., Ross, R.P., Fitzgerald, G.F., Cotter, P.D., 2012. High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses. *Appl. Environ. Microbiol.* 78, 5717–5723. doi:10.1128/AEM.00918-12
- Rabiner, L.R., Juang, B.H., 1986. An introduction to hidden Markov models. *ASSP Magazine, IEEE* 3, 4 – 16. doi:10.1109/MASSP.1986.1165342
- Reneau, J.K., 1986. Effective Use of Dairy Herd Improvement Somatic Cell Counts in Mastitis Control. *Journal of Dairy Science* 69, 1708–1720. doi:10.3168/jds.S0022-0302(86)80590-2
- Rho, M., Wu, Y.-W., Tang, H., Doak, T.G., Ye, Y., 2012. Diverse CRISPRs Evolving in Human Microbiomes. *PLoS Genet* 8, e1002441. doi:10.1371/journal.pgen.1002441
- Rijnen, L., Yvon, M., van Kranenburg, R., Courtin, P., Verheul, A., Chambellon, E., Smit, G., 2003. Lactococcal aminotransferases AraT and BcaT are key enzymes for the formation of aroma compounds from amino acids in cheese. *International Dairy Journal* 13, 805–812. doi:10.1016/S0958-6946(03)00102-X
- Robles, C.T., 2014. Identificación molecular de microorganismos aislados del queso Cotija y de sus materias primas (Tesis de licenciatura). Universidad Nacional Autónoma de México, Facultad de Química, México.
- Rossi, F., Rizzotti, L., Felis, G.E., Torriani, S., 2014. Horizontal gene transfer among microorganisms in food: current knowledge and future perspectives. *Food Microbiol.* 42, 232–243. doi:10.1016/j.fm.2014.04.004
- Ross, R.P., Morgan, S., Hill, C., 2002. Preservation and fermentation: past, present and future. *Int. J. Food Microbiol.* 79, 3–16.

- Rothberg, J.M., Hinze, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., Hoon, J., Simons, J.F., Marran, D., Myers, J.W., Davidson, J.F., Branting, A., Nobile, J.R., Puc, B.P., Light, D., Clark, T.A., Huber, M., Branciforte, J.T., Stoner, I.B., Cawley, S.E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J.A., Namsaraev, E., McKernan, K.J., Williams, A., Roth, G.T., Bustillo, J., 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352. doi:10.1038/nature10242
- Roux, S., Enault, F., le Bronner, G., Debroas, D., 2011. Comparison of 16S rRNA and protein-coding genes as molecular markers for assessing microbial diversity (Bacteria and Archaea) in ecosystems. *FEMS Microbiology Ecology* 78, 617–628. doi:10.1111/j.1574-6941.2011.01190.x
- Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74, 5463–5467.
- Satokari, R.M., Vaughan, E.E., Smidt, H., Saarela, M., Mättö, J., de Vos, W.M., 2003. Molecular Approaches for the Detection and Identification of *Bifidobacteria* and *Lactobacilli* in the Human Gastrointestinal Tract. *Systematic and Applied Microbiology* 26, 572–584. doi:10.1078/072320203770865882
- Settanni, L., Di Grigoli, A., Tornambé, G., Bellina, V., Francesca, N., Moschetti, G., Bonanno, A., 2012. Persistence of wild *Streptococcus thermophilus* strains on wooden vat and during the manufacture of a traditional Caciocavallo type cheese. *International Journal of Food Microbiology* 155, 73–81. doi:10.1016/j.ijfoodmicro.2012.01.022
- Shannon, C.E., 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x
- Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., Church, G.M., 2005. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science* 309, 1728–1732. doi:10.1126/science.1117389
- Simpson, E., 1949. Measurement of Diversity. *Nature* 688–688. doi:10.1038/163688a0
- Smit, G., Smit, B.A., Engels, W.J.M., 2005. Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.* 29, 591–610. doi:10.1016/j.femsre.2005.04.002
- Snyder, A.B., Worobo, R.W., 2014. Chemical and genetic characterization of bacteriocins: antimicrobial peptides for food safety. *J. Sci. Food Agric.* 94, 28–44. doi:10.1002/jsfa.6293
- Solms, J., 1969. Taste of amino acids, peptides, and proteins. *J. Agric. Food Chem.* 17, 686–688. doi:10.1021/jf60164a016
- Su, X., Pan, W., Song, B., Xu, J., Ning, K., 2014. Parallel-META 2.0: enhanced metagenomic data analysis with functional annotation, high performance computing and advanced visualization. *PLoS ONE* 9, e89323. doi:10.1371/journal.pone.0089323
- Thomas, T., Turner, K., Crow, V. 1980. Galactose fermentation by *Streptococcus lactis* and *Streptococcus cremoris*: pathways, products and regulation. *J. Bacteriol.* 144, 672–682.
- Thomas, T., Gilbert, J., Meyer, F., 2012. Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp* 2, 3. doi:10.1186/2042-5783-2-3
- Thompson, G.G., Withers, P.C., Pianka, E.R., Thompson, S.A., 2003. Assessing biodiversity with species accumulation curves; inventories of small reptiles by pit-trapping in Western Australia. *Austral Ecology* 28, 361–383. doi:10.1046/j.1442-9993.2003.01295.x
- Urbach, G., 1997. The flavour of milk and dairy products: II. Cheese: contribution of volatile compounds. *International Journal of Dairy Technology* 50, 79–89. doi:10.1111/j.1471-0307.1997.tb01743.x
- van Heel, A.J., de Jong, A., Montalbán-López, M., Kok, J., Kuipers, O.P., 2013. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res.* 41, W448–453. doi:10.1093/nar/gkt391
- Vuyst, L. de, Vandamme, E.J., 1994. *Bacteriocins of lactic acid bacteria: microbiology, genetics, and applications.* Blackie Academic & Professional.
- Wang, W., Zhang, L., Li, Y., 2012. Production of volatile compounds in reconstituted milk reduced-fat cheese and the physicochemical properties as affected by exopolysaccharide-producing strain. *Molecules* 17, 14393–14408. doi:10.3390/molecules171214393

- Wang, Y., Zhang, Z., Ramanan, N., 1997. The actinomycete *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes. *J Bacteriol* 179, 3270–3276.
- Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F.E., Silva, J., Tammadoni, S., Nosrat, B., Conrad, D., Rohwer, F., 2009. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS ONE* 4, e7370. doi:10.1371/journal.pone.0007370
- Woese, C.R., Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *PNAS* 74, 5088–5090. doi:10.1073/pnas.74.11.5088
- Wolfe, B.E., Button, J.E., Santarelli, M., Dutton, R.J., 2014. Cheese Rind Communities Provide Tractable Systems for In Situ and In Vitro Studies of Microbial Diversity. *Cell* 158, 422–433. doi:10.1016/j.cell.2014.05.041
- Xie, G., Wang, L., Gao, Q., Yu, W., Hong, X., Zhao, L., Zou, H., 2013. Microbial community structure in fermentation process of Shaoxing rice wine by Illumina-based metagenomic sequencing. *J. Sci. Food Agric.* 93, 3121–3125. doi:10.1002/jsfa.6058
- Yang, M.Y., Bowmaker, M., Reyes, A., Vergani, L., Angeli, P., Gringeri, E., Jacobs, H.T., Holt, I.J., 2002. Biased Incorporation of Ribonucleotides on the Mitochondrial L-Strand Accounts for Apparent Strand-Asymmetric DNA Replication. *Cell* 111, 495–505. doi:10.1016/S0092-8674(02)01075-9
- Yang, S.-C., Lin, C.-H., Sung, C.T., Fang, J.-Y., 2014. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Front Microbiol* 5, 241. doi:10.3389/fmicb.2014.00241
- Yoon, B.-J., 2009. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Curr Genomics* 10, 402–415. doi:10.2174/138920209789177575
- Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z., Forney, L.J., 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS ONE* 7, e33865. doi:10.1371/journal.pone.0033865
- Yvon, M., Rijnen, L., 2001. Cheese flavour formation by amino acid catabolism. *International Dairy Journal, Cheese Ripening and Technology* 11, 185–201. doi:10.1016/S0958-6946(01)00049-8
- Zhao, M., Zhang, D., Su, X., Duan, S., Wan, J., Yuan, W., Liu, B., Ma, Y., Pan, Y., 2015. An Integrated Metagenomics/Metaproteomics Investigation of the Microbial Communities and Enzymes in Solid-state Fermentation of Pu-erh tea. *Sci Rep* 5. doi:10.1038/srep10117
- Zhu, W., Lomsadze, A., Borodovsky, M., 2010. Ab initio gene identification in metagenomic sequences. *Nucl. Acids Res.* 38, e132–e132. doi:10.1093/nar/gkq275
- Zúñiga, A.B., 2009. Descripción e identificación de la comunidad bacteriana presente en el Cotija por métodos moleculares (Tesis de maestría). Universidad Nacional Autónoma de México, Facultad de Química, México.

Páginas de internet consultadas:

<http://www.ncbi.nlm.nih.gov/genome/browse/reference/> [Último acceso 1 Octubre 2015]

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> [Último acceso 1 Octubre 2015]

<http://pubchem.ncbi.nlm.nih.gov/compound/> [Último acceso 1 Abril 2016]

<https://gold.jgi.doe.gov/> [Último acceso 1 Abril 2016]

<http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/dna-sequencing-fragment-analysis/overview-of-dna-sequencing/faq.html> [Último acceso 1 Abril 2016]

<http://www.molecularecologist.com/next-gen-fieldguide-2014> [Último acceso 1 Octubre 2015]

## Anexo 1. Detección y eliminación de ADN contaminante

Se considera ADN contaminante todo aquel que es ajeno al metagenoma microbiano que se desea secuenciar. En alimentos lácteos la presencia de glóbulos blancos y células epiteliales provenientes de la ubre es normal y representa una fuente de ADN bovino en el queso.

El contenido celular somático en la leche de una ubre sana está compuesto por macrófagos (60%), linfocitos (25%) y leucocitos neutrófilos polimorfonucleares (15%). Cuando el conteo de células somáticas (CCS) resulta elevado, indica que hay un problema en la salud, generalmente mastitis, cuya causa principal es la infección bacteriana (Harmon, 1994). Otra causa podrían ser lesiones en la ubre ocasionadas en la mayoría de los casos por diseños deficientes en las instalaciones y equipo de ordeño o por accidentes entre una vaca y otra. Además las vacas viejas de más de 7 años de edad tienden a tener conteos más elevados que animales jóvenes debido al incrementando en la probabilidad de tener una infección crónica (Alais y Linden, 1991).

El umbral de conteo indicativo de inflamación o mastitis ha sido establecido históricamente entre 200,000 y 250,000 cel/mL de leche (Reneau, 1986), aunque en México de acuerdo a la norma para leche cruda de vaca (PROY-NMX-F-700-COFOCALEC-2012) el CCS va desde  $\leq 400,000$  hasta 1,000,000 de células/mL de leche de acuerdo al grado de calidad de la misma.

Aún cuando se espera la presencia de ADN bovino en el metagenoma del queso Cotija, éste se considera como material contaminante, pues su secuenciación no es de interés en este trabajo.

### Detección de ADN bovino por q-PCR

Se utilizó un juego de cebadores y sonda marcada con el fluoróforo FAM para PCR en tiempo real (q-PCR), dirigidos a ADN mitocondrial (ADNmt) de ganado bovino. Las mezclas de reacción se hicieron en un volumen de 25  $\mu$ L con 100 ng de ADN total en las condiciones estándar para tiempo real: 95°C, 15 s y 60°C, 60 s con las concentraciones de reactivos que se especifican en la Tabla A1.

Tabla A1. Volúmenes utilizados para la reacción de q-PCR

	Concentración inicial	Concentración final	Volúmen para 25 $\mu$ L
Taqman master mix	2X	1X	12.5
Mezcla de sonda y cebadores	---	---	2.5
ADN molde	50 ng/ $\mu$ L	100 ng	2.0
Agua	---	---	8.0

En todas las corridas se incluyeron controles sin ADN molde por triplicado. No se realizó curva estándar debido a que el objetivo era determinar presencia de ADN bovino por visualización de curva de amplificación por un métodos de alta sensibilidad.

Se llevó a cabo la detección en ADN total obtenido de tres paquetes celulares provenientes de muestras de queso Cotija extraídos independientemente. En la Figura A1 se muestra la gráfica de amplificación en todas las muestras a ciclos tempranos.

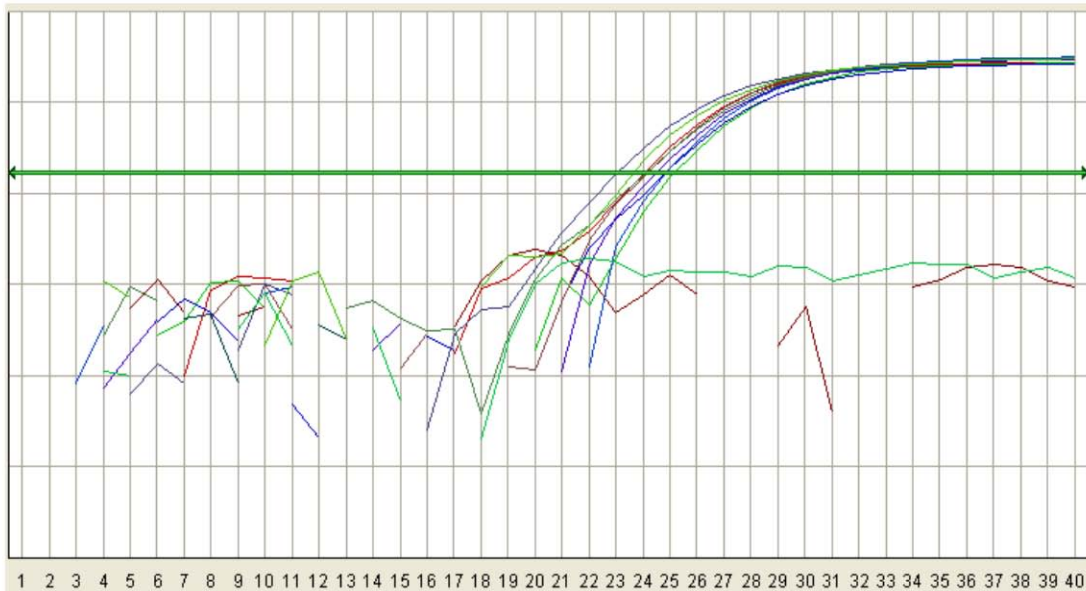


Figura A1. Amplificación por q-PCR de ADNmt vacuno control y en las muestras de ADN total de queso Cotija

#### Optimización de condiciones para eliminar ADN contaminante

Se probaron tres condiciones de tratamiento de los paquetes celulares resuspendidos en buffer de DNasa (100 mM Tris pH 7.5, 25 mM de  $MgCl_2$  y 5 mM de  $CaCl_2$ ) con 1 unidad de DNasa I a 37°C incubando durante 10, 30 y 60 minutos en termomixer con agitación a 200 rpm. La DNasa I se inactivó al término de este tiempo con un tratamiento de 10 minutos a 75°C y los tubos se centrifugaron a 14,000 (18,200 X g) rpm durante 10 minutos en microcentrífuga para después congelarlos a -20°C durante toda la noche y proseguir con la extracción y purificación del ADN total (Willner et al., 2009; Hurwitz et al., 2012). A la par se probó un paso de tratamiento con RNasas durante 30 min a 37°C en la extracción y purificación del ADN metagenómico para eliminar el ARN bacteriano.

Se elaboró un gel de agarosa al 1% para determinar el efecto de dichos tratamientos en la integridad del ADN extraído (Figura A2), cuyo resultado indicó que no hubo degradación en ninguna de ellas. Por otro lado, en la reacción de q-PCR contra ADN mitocondrial de vaca y se observó que después de los tratamientos con DNasa I, la señal de amplificación para todas las muestras desaparece (Figura A3), lo que indica que el tratamiento de 10 minutos es suficiente para eliminar el ADN contaminante sin afectar la integridad del material genético.

La señal también desaparece en la muestra tratada sólo con RNasa A debido a que durante la replicación el ADN mitocondrial suele presentar estructuras intermediarias híbridas de ADN y ARN que son producto de la

incorporación de oligos de ribonucleótidos dentro del bucle de ADN circular. Dichos híbridos pueden poseer más ARN que el ADNmt no replicante (Yang et al. , 2002). De hecho, el uso de RNasas no está recomendado en los protocolos de purificación de ADN mitocondrial (Guo et al. , 2009).

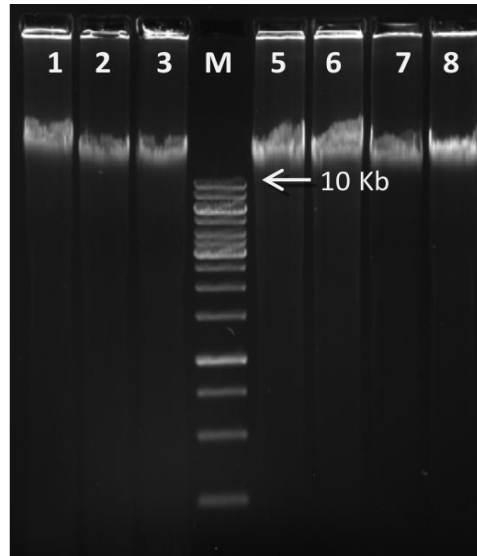


Figura A2. Gel de agarosa al 1% de las extracciones de ADN total después de cada tratamiento con DNasa y RNasa. Carriles 1, 2 y 3: Tratamiento con DNAsas, (1U incubado por 10, 30 y 60 min respectivamente); M: Marcador de peso molecular (1Kb, Fermentas); carriles 5 y 6: Tratamiento con RNAsas; carriles 7 y 8: Sin tratamiento de Dnasa ni RNasa.

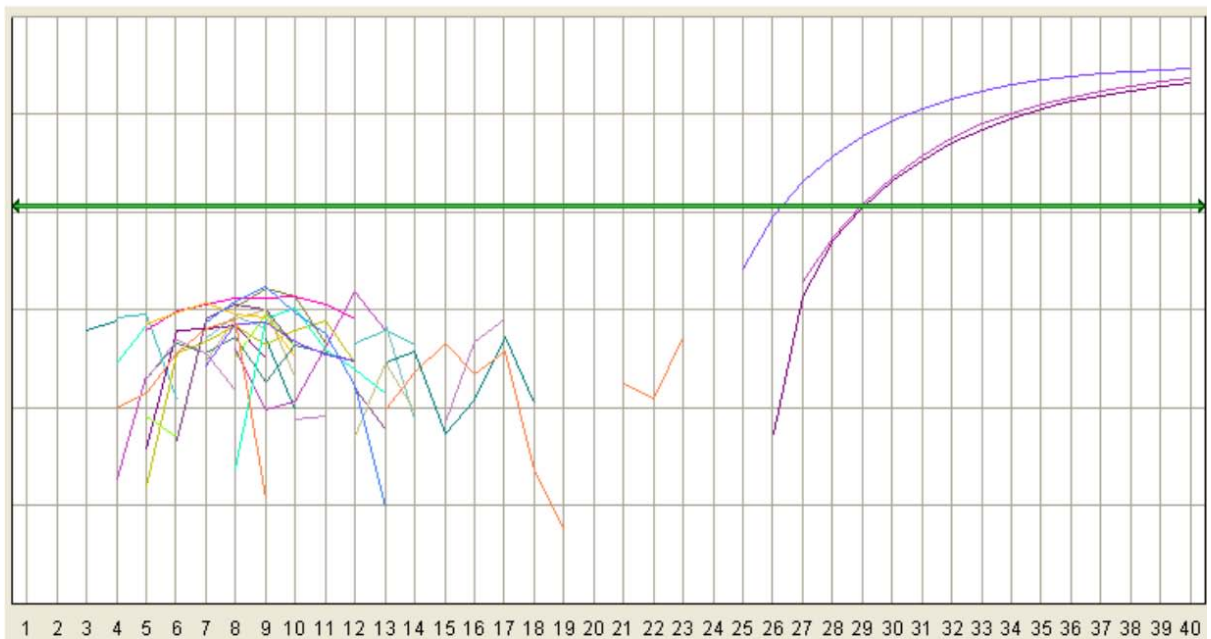


Figura A3. q-PCR de ADNmt vacuno después de tratamientos con DNAsas y RNAsas. Las señales de amplificación que se observan corresponden a las muestras de ADN extraído por método no modificado.

Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

Conteo	Porcentaje	Domino	Filum	Clase	Orden	Familia	Genero
344732	29.57905	Bacteria	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Weissella
241566	20.727095	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
131521	11.2849	Bacteria	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Leuconostoc
103326	8.865684	Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Aerococcus
39821	3.416762	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus
35054	3.007739	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
29628	2.542172	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Lactococcus
26109	2.240231	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Bavariococcus
25075	2.151511	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Marinilactibacillus
23274	1.99698	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Pediococcus
21655	1.858065	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
16363	1.403995	Bacteria	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Fructobacillus
16232	1.392755	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Alkalibacterium
11916	1.022429	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Carnobacterium
7160	0.61435	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Desemzia
4657	0.399585	Bacteria	Deinococcus-Thermus	Deinococci	Deinococcales	Trueperaceae	Truepera
4576	0.392635	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Kocuria
3927	0.336949	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium
3764	0.322963	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
3270	0.280576	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Brevibacillus
3148	0.270108	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XVIII	Symbiobacterium
3137	0.269164	Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Abiotrophia
2962	0.254149	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Vagococcus
2637	0.226263	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Granulicatella
2631	0.225748	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Dolosigranulum
2617	0.224547	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Oceanobacillus
2464	0.211419	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Brevibacteriaceae	Brevibacterium
2430	0.208501	Bacteria	Firmicutes	Bacilli	Bacillales	Thermoactinomycetacea	Seinonella
2335	0.20035	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Salinicoccus
2264	0.194258	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Halolactobacillus
1834	0.157363	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Halolactibacillus
1822	0.156333	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Isobaculum
1668	0.143119	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium
1583	0.135826	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Atopostipes
1541	0.132222	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Lacticingenium
1539	0.132051	Bacteria	Firmicutes	Bacilli	Bacillales	Listeriaceae	Brochothrix
1384	0.118751	Bacteria	Firmicutes	Bacilli	Bacillales	Sporolactobacillaceae	Sporolactobacillus
1297	0.111287	Bacteria	Acidobacteria	Acidobacteria_	Unclassified		
1192	0.102277	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Salirhabdus
1136	0.097472	Bacteria	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Oenococcus
1089	0.09344	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Propionibacterium
1088	0.093354	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Atopococcus
1075	0.092238	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Lactovum
1067	0.091552	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Streptomycetaceae	Streptomyces
1024	0.087862	Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Facklamia
1003	0.08606	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Catelicoccus
907	0.077823	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Tetragenococcus
852	0.073104	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	Kurthia
723	0.062036	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Viridibacillus
697	0.059805	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Trichococcus
686	0.058861	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Macroccoccus
599	0.051396	Bacteria	Firmicutes	Bacilli	Bacillales	Sporolactobacillaceae	Marinococcus
561	0.048136	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	Sporosarcina
555	0.047621	Bacteria	Firmicutes	Bacilli	Bacillales	Caryophanaceae	Caryophanon
551	0.047277	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Caldalkalibacillus
523	0.044875	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Virgibacillus
508	0.043588	Bacteria	Firmicutes	Bacilli	Bacillales	Sporolactobacillaceae	Sinobaca

Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

484	0.041529	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Citricoccus
477	0.040928	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Geobacillus
461	0.039555	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Marihabitans
459	0.039384	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus
442	0.037925	Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Globicatella
395	0.033892	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	Arcanobacterium
378	0.032434	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Sporotalea
367	0.03149	Bacteria	Fusobacteria	Fusobacteria	Fusobacteriales	Leptotrichiaceae	Leptotrichia
348	0.029859	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Dermabacteraceae	Dermabacter
336	0.02883	Bacteria	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium
332	0.028487	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Fodinibacter
324	0.0278	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Anoxybacillus
311	0.026685	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Dermabacteraceae	Brachybacterium
291	0.024969	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	Paraoerskovia
282	0.024196	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Sporomusa
281	0.024111	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XIV	Dethiosulfatibacter
274	0.02351	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Dermabacteraceae	Helcobacillus
271	0.023253	Bacteria	Firmicutes	Bacilli	Bacillales	Listeriaceae	Listeria
271	0.023253	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XVII	Sulfobacillus
265	0.022738	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Agrococcus
251	0.021537	Bacteria	Cyanobacteria	Cyanobacteria	Cyanobacteria_order	Chloroplast	Chlorophyta
226	0.019391	Bacteria	Acidobacteria	Acidobacteria_	Unclassified		
225	0.019306	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Jeotgalicoccus
221	0.018962	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Cellvibrio
205	0.01759	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Janibacter
196	0.016817	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Novosphingobium
189	0.016217	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Leifsonia
189	0.016217	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	Bhargavaea
181	0.01553	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Lentibacillus
167	0.014329	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Halobacillus
155	0.013299	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Exiguobacterium
153	0.013128	Bacteria	Firmicutes	Clostridia	Clostridiales	Heliobacteriaceae	Heliobacterium
150	0.01287	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Dermatophilaceae	Dermatophilus
141	0.012098	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Arthrobacter
139	0.011927	Bacteria	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Garciella
136	0.011669	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Rothia
129	0.011069	Bacteria	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Pseudoramibacter
126	0.010811	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Atopobacter
126	0.010811	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Caldicellulosiruptor
123	0.010554	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
110	0.009438	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Actinosynnemataceae	Lechevalieria
107	0.009181	Bacteria	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Acetobacterium
99	0.008495	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Actinosynnemataceae	Saccharothrix
97	0.008323	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Alkaliphilus
97	0.008323	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiales_incerta	Ideonella
96	0.008237	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	Pseudonocardia
92	0.007894	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Saccharibacillus
89	0.007636	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Gemella
76	0.006521	Bacteria	Deinococcus-Thermus	Deinococci	Deinococcales	Deinococcaceae	Deinococcus
76	0.006521	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Aneurinibacillus
75	0.006435	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Nosocomiicoccus
74	0.006349	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XI	Sedimentibacter
73	0.006264	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Rathayibacter
73	0.006264	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Cerasibacillus
71	0.006092	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	Actinomyces
71	0.006092	Bacteria	Firmicutes	Bacilli	Bacillales	Alicyclobacillaceae	Alicyclobacillus
70	0.006006	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Luteococcus
67	0.005749	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	Saccharopolyspora



Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

67	0.005749	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Veillonella
65	0.005577	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Desulfotomaculum
64	0.005491	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Kribbia
63	0.005406	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Melissococcus
62	0.00532	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Ornithinibacillus
62	0.00532	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Tepidanaerobacter
61	0.005234	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Natronobacillus
61	0.005234	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Alloiococcus
56	0.004805	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	Mycobacterium
54	0.004633	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillales_incertae_sedi	Solibacillus
54	0.004633	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Coprococcus
53	0.004548	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Aestuariimicrobium
52	0.004462	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Anaerovibrio
51	0.004376	Bacteria	BRC1	Unclassified			
49	0.004204	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Propionimicrobium
49	0.004204	Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Ignavigranum
48	0.004119	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Agromyces
48	0.004119	Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Doliosicoccus
48	0.004119	Bacteria	Proteobacteria	Deltaproteobact	Myxococcales	Cystobacteraceae	Anaeromyxobacter
48	0.004119	Bacteria	Tenericutes	Mollicutes	Entomoplasmatales	Entomoplasmataceae	Mesoplasma
46	0.003947	Bacteria	Nitrospira	Nitrospira	Nitrospirales	Nitrospiraceae	Nitrospira
45	0.003861	Bacteria	Firmicutes	Thermolithobac	Thermolithobacterial	Thermolithobacteraceae	Thermolithobacter
44	0.003775	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Sediminibacillus
44	0.003775	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Pilibacter
44	0.003775	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridiisalibacter
44	0.003775	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Selenomonas
43	0.00369	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Kineosporiaceae	Kineosporia
42	0.003604	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Terrabacter
42	0.003604	Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Eremococcus
41	0.003518	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	Kutzneria
40	0.003432	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Beutenbergiaceae	Serinibacter
38	0.003261	Bacteria	Acidobacteria	Acidobacteria_	Unclassified		
38	0.003261	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburia
37	0.003175	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Streptosporangiaceae	Acrocarpospora
37	0.003175	Bacteria	Firmicutes	Bacilli	Bacillales	Thermoactinomycetacea	Shimazuella
37	0.003175	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XIV	Howardella
37	0.003175	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Zymophilus
36	0.003089	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Jonesiaceae	Jonesia
36	0.003089	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Thermomonosporaceae	Actinoallomurus
35	0.003003	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Thermomonosporaceae	Actinomadura
34	0.002917	Bacteria	Tenericutes	Mollicutes	Mycoplasmatales	Mycoplasmataceae	Mycoplasma
33	0.002832	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	Actinobaculum
33	0.002832	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiodiaceae	Aeromicrobium
33	0.002832	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Pseudobutyrvibrio
32	0.002746	Bacteria	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Eubacterium
32	0.002746	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Robinsoniella
31	0.00266	Bacteria	Firmicutes	Clostridia	Natranaerobiales	Natranaerobiaceae	Dethiobacter
30	0.002574	Bacteria	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Cytophagaceae	Runella
30	0.002574	Bacteria	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Unclassified	
30	0.002574	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Dehalobacter
30	0.002574	Bacteria	Firmicutes	Clostridia	Halanaerobiales	Halanaerobiaceae	Halanaerobium
29	0.002488	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	Nocardia
28	0.002402	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Dermabacteraceae	Devriesea
28	0.002402	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Salinibacillus
27	0.002317	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Dermacoccaceae	Dermacoccus
27	0.002317	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Tessaracoccus
27	0.002317	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Gracilibacillus
27	0.002317	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	Planomicrobium

Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

27	0.002317	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Desulfonispora
26	0.002231	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Promicromonosporaceae	Myceligenens
26	0.002231	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Thermomonosporaceae	Thermomonospora
26	0.002231	Bacteria	Chlorobi	Chlorobia	Chlorobiales	Chlorobiaceae	Chlorobaculum
26	0.002231	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	Paenisporosarcina
25	0.002145	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Anaerospobacter
23	0.001973	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Actinoplanes
23	0.001973	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Salsuginibacillus
23	0.001973	Bacteria	Tenericutes	Mollicutes	Acholeplasmatales	Acholeplasmataceae	Acholeplasma
23	0.001973	Bacteria	Tenericutes	Mollicutes	Entomoplasmatales	Spiroplasmataceae	Spiroplasma
22	0.001888	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	Cellulomonas
22	0.001888	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Oxalophagus
21	0.001802	Bacteria	Proteobacteria	Betaproteobacte	Rhodocyclales	Rhodocyclaceae	Thauera
21	0.001802	Bacteria	Proteobacteria	Gammaproteob	Acidithiobacillales	Thermithiobacillaceae	Thermithiobacillus
20	0.001716	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Nesterenkonia
20	0.001716	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillales_incertae_sedi	Pullulanibacillus
20	0.001716	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Thermicanus
20	0.001716	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XI	Peptoniphilus
20	0.001716	Bacteria	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Turicibacter
19	0.00163	Archaea	Crenarchaeota	Thermoprotei	Desulfurococcales	Pyrodictiaceae	Pyrolobus
19	0.00163	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XIII	Mogibacterium
19	0.00163	Bacteria	Proteobacteria	Alphaproteobac	Caulobacterales	Caulobacteraceae	Phenylobacterium
18	0.001544	Archaea	Crenarchaeota	Thermoprotei	Desulfurococcales	Desulfurococcaceae	Ignicoccus
18	0.001544	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Salimicrobium
18	0.001544	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	Planococcus
18	0.001544	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Acetobacteraceae	Acidocella
17	0.001459	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	Millisia
17	0.001459	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Thermomonosporaceae	Actinocorallia
17	0.001459	Bacteria	Proteobacteria	Alphaproteobac	Caulobacterales	Hyphomonadaceae	Robiginitomaculum
16	0.001373	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Brooklawnia
16	0.001373	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Anaerოსinus
15	0.001287	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Arsenicoccus
15	0.001287	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	Rhodococcus
15	0.001287	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Rarobacteraceae	Rarobacter
15	0.001287	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Amphibacillus
15	0.001287	Bacteria	Proteobacteria	Deltaproteobact	Desulfobacterales	Desulfobacteraceae	Desulfobotulus
15	0.001287	Bacteria	Proteobacteria	Deltaproteobact	Desulfurellales	Desulfurellaceae	Desulfurella
14	0.001201	Bacteria	Acidobacteria	Acidobacteria_	Unclassified		
14	0.001201	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Tetrasphaera
14	0.001201	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Tsukamurellaceae	Tsukamurella
14	0.001201	Bacteria	Actinobacteria	Actinobacteria	Solirubrobacterales	Solirubrobacteraceae	Solirubrobacter
14	0.001201	Bacteria	Deinococcus-Thermus	Deinococci	Themales	Thermaceae	Unclassified
14	0.001201	Bacteria	Firmicutes	Bacilli	Bacillales	Sporolactobacillaceae	Tuberibacillus
14	0.001201	Bacteria	Proteobacteria	Epsilonproteob	Campylobacterales	Helicobacteraceae	Helicobacter
14	0.001201	Bacteria	Proteobacteria	Epsilonproteob	Nautiliales	Nautiliaceae	Nautilia
13	0.001115	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium
13	0.001115	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Paraliobacillus
13	0.001115	Bacteria	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Anaerofustis
13	0.001115	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XI	Anaerოსphaera
13	0.001115	Bacteria	Proteobacteria	Deltaproteobact	Desulfovibrionales	Desulfovibrionaceae	Desulfovibrio
13	0.001115	Bacteria	TM7	Unclassified			
12	0.00103	Bacteria	Acidobacteria	Acidobacteria_	Unclassified		
12	0.00103	Bacteria	Cyanobacteria	Cyanobacteria	Cyanobacteria_order	Family_II	GpIIb
12	0.00103	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Jeotgalibacillus
12	0.00103	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Pontibacillus
12	0.00103	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Tindallia
12	0.00103	Bacteria	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Erysipelothrix
12	0.00103	Bacteria	Gemmatimonadetes	Gemmatimonad	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas

Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

12	0.00103	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Acetobacteraceae	Acidicaldus
11	0.000944	Bacteria	Acidobacteria	Acidobacteria_	Unclassified		
11	0.000944	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Dietziaceae	Dietzia
11	0.000944	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Lysinibacillus
11	0.000944	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Thermobacillus
11	0.000944	Bacteria	Firmicutes	Bacilli	Bacillales	Thermoactinomycetacea	Planifilum
11	0.000944	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Thermoanaerobacter
11	0.000944	Bacteria	Proteobacteria	Gammaproteob:	Enterobacteriales	Enterobacteriaceae	Serratia
10	0.000858	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Glycomycetaceae	Haloglycomyces
10	0.000858	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiopsaceae	Marinactinospora
10	0.000858	Bacteria	Deinococcus-Thermus	Deinococci	Thermales	Thermaceae	Marinithermus
10	0.000858	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Caloramator
10	0.000858	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Johnsonella
10	0.000858	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Rhodospirillaceae	Caenispirillum
10	0.000858	Bacteria	Proteobacteria	Alphaproteobac	Sphingomonadales	Sphingomonadaceae	Sphingobium
10	0.000858	Bacteria	Proteobacteria	Gammaproteob:	Chromatiales	Chromatiaceae	Allochromatium
10	0.000858	Bacteria	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae	Kosmotoga
9	0.000772	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Anoxynatronum
9	0.000772	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Thermohalobacter
9	0.000772	Bacteria	Firmicutes	Clostridia	Halanaerobiales	Halobacteroidaceae	Orenia
9	0.000772	Bacteria	Firmicutes	Clostridia	Natranaerobiales	Natranaerobiaceae	Natranaerobius
9	0.000772	Bacteria	Proteobacteria	Deltaproteobact	Desulfobacteriales	Desulfobacteraceae	Desulfobacterium
9	0.000772	Bacteria	Proteobacteria	Deltaproteobact	Desulfuromonadales	Geobacteraceae	Geoalkalibacter
9	0.000772	Bacteria	Proteobacteria	Epsilonproteob:	Nautiliales	Nautiliaceae	Caminiibacter
9	0.000772	Bacteria	Proteobacteria	Gammaproteob:	Alteromonadales	Alteromonadaceae	Microbulbifer
8	0.000686	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Leucobacter
8	0.000686	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	Gordonia
8	0.000686	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillales_incertae_sedi	Rummeliibacillus
8	0.000686	Bacteria	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Alkalibacter
8	0.000686	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Desulfitobacterium
8	0.000686	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Desulfosporosinus
8	0.000686	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Mahella
8	0.000686	Bacteria	Proteobacteria	Deltaproteobact	Desulfarculales	Desulfarculaceae	Desulfarculus
8	0.000686	Bacteria	Proteobacteria	Gammaproteob:	Enterobacteriales	Enterobacteriaceae	Enterobacter
8	0.000686	Bacteria	Proteobacteria	Gammaproteob:	Enterobacteriales	Enterobacteriaceae	Escherichia/Shigella
7	0.000601	Bacteria	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flammeovirgaceae	Roseivirga
7	0.000601	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Ureibacillus
7	0.000601	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Vulcanibacillus
7	0.000601	Bacteria	Firmicutes	Bacilli	Bacillales	Thermoactinomycetacea	Mechercharimyces
7	0.000601	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XI	Tissierella
7	0.000601	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Sporotomaculum
7	0.000601	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptostreptococcaceae	Filifactor
7	0.000601	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Veillonella
7	0.000601	Bacteria	Firmicutes	Clostridia	Halanaerobiales	Halobacteroidaceae	Selenihalanaerobact
7	0.000601	Bacteria	Firmicutes	Clostridia	Natranaerobiales	Natranaerobiaceae	Natronovirga
7	0.000601	Bacteria	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Holdmania
7	0.000601	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Rhodobiaceae	Afifella
7	0.000601	Bacteria	Proteobacteria	Betaproteobacte	Methylophilales	Methylophilaceae	Methylotenera
7	0.000601	Bacteria	Proteobacteria	Deltaproteobact	Desulfuromonadales	Geobacteraceae	Geobacter
7	0.000601	Bacteria	Proteobacteria	Epsilonproteob:	Campylobacteriales	Campylobacteraceae	Campylobacter
7	0.000601	Bacteria	Proteobacteria	Gammaproteob:	Chromatiales	Ectothiorhodospiraceae	Ectothiorhodospira
7	0.000601	Bacteria	Proteobacteria	Gammaproteob:	Enterobacteriales	Enterobacteriaceae	Pectobacterium
7	0.000601	Bacteria	Proteobacteria	Gammaproteob:	Oceanospirillales	Hahellaceae	Hahella
7	0.000601	Bacteria	Tenericutes	Mollicutes	Anaeroplasmatales	Anaeroplasmataceae	Anaeroplasma
7	0.000601	Bacteria	Tenericutes	Mollicutes	Mycoplasmatales	Mycoplasmataceae	Ureaplasma
7	0.000601	Bacteria	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae	Thermosipho
6	0.000515	Bacteria	Acidobacteria	Holophagae	Holophagales	Holophagaceae	Holophaga
6	0.000515	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardaceae	Kibdelosporangium

Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

6	0.000515	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Yaniellaceae	Yaniella
6	0.000515	Bacteria	Cyanobacteria	Cyanobacteria	Cyanobacteria_order	Family_IV	GpIV
6	0.000515	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Cohnella
6	0.000515	Bacteria	Firmicutes	Clostridia	Clostridiales	Gracilibacteraceae	Gracilibacter
6	0.000515	Bacteria	Firmicutes	Clostridia	Halanaerobiales	Halobacteroidaceae	Halanaerobacter
6	0.000515	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Gelria
6	0.000515	Bacteria	OP10	Unclassified			
6	0.000515	Bacteria	Proteobacteria	Alphaproteobac	Caulobacterales	Hyphomonadaceae	Oceanicaulis
6	0.000515	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Phyllobacteriaceae	Hoeflea
6	0.000515	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Burkholderiaceae	Burkholderia
6	0.000515	Bacteria	Proteobacteria	Betaproteobacte	Neisseriales	Neisseriaceae	Silvimonas
6	0.000515	Bacteria	Proteobacteria	Deltaproteobact	Myxococcales	Cystobacteraceae	Cystobacter
6	0.000515	Bacteria	Proteobacteria	Gammaproteob:	Pseudomonadales	Moraxellaceae	Psychrobacter
5	0.000429	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	Oerskovia
5	0.000429	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Glycomycetaceae	Glycomyces
5	0.000429	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Kineosporiaceae	Kineococcus
5	0.000429	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Micromonospora
5	0.000429	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Rugosimonospora
5	0.000429	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiodaceae	Kribbella
5	0.000429	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Propionicicella
5	0.000429	Bacteria	Deinococcus-Thermus	Deinococci	Thermales	Thermaceae	Oceanithermus
5	0.000429	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Filobacillus
5	0.000429	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Saccharococcus
5	0.000429	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Caloranaerobacter
5	0.000429	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XIV	Anaerobranca
5	0.000429	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XIV	Proteocatella
5	0.000429	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Butyrivibrio
5	0.000429	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Peptococcus
5	0.000429	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Acetonema
5	0.000429	Bacteria	Firmicutes	Clostridia	Halanaerobiales	Halobacteroidaceae	Halonatronum
5	0.000429	Bacteria	Firmicutes	Clostridia	Halanaerobiales	Halobacteroidaceae	Nianiella
5	0.000429	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Moorella
5	0.000429	Bacteria	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Coprobacillus
5	0.000429	Bacteria	Proteobacteria	Alphaproteobac	Rhodobacterales	Rhodobacteraceae	Seohaecicola
5	0.000429	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Acetobacteraceae	Acetobacter
5	0.000429	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Acetobacteraceae	Asaia
5	0.000429	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Acetobacteraceae	Roseomonas
5	0.000429	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Rhodospirillaceae	Azospirillum
5	0.000429	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Comamonadaceae	Hydrogenophaga
5	0.000429	Bacteria	Proteobacteria	Betaproteobacte	Rhodocyclales	Rhodocyclaceae	Zoogloea
5	0.000429	Bacteria	Proteobacteria	Deltaproteobact	Syntrophobacterales	Syntrophobacteraceae	Desulfoglaeba
5	0.000429	Bacteria	Proteobacteria	Gammaproteob:	Enterobacteriales	Enterobacteriaceae	Klebsiella
5	0.000429	Bacteria	WS3	Unclassified			
4	0.000343	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	Mobiluncus
4	0.000343	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Sinomonas
4	0.000343	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Cryomorpaceae	Brumimicrobium
4	0.000343	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Terribacillus
4	0.000343	Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Ammoniphilus
4	0.000343	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XI	Anaerococcus
4	0.000343	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnobacterium
4	0.000343	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Hydrogenoanaeroba
4	0.000343	Bacteria	Firmicutes	Clostridia	Clostridiales	Syntrophomonadaceae	Syntrophomonas
4	0.000343	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Pectinatus
4	0.000343	Bacteria	OP11	Unclassified			
4	0.000343	Bacteria	Proteobacteria	Alphaproteobac	Parvularculales	Parvularculaceae	Parvularcula
4	0.000343	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Hyphomicrobiaceae	Zhangella
4	0.000343	Bacteria	Proteobacteria	Alphaproteobac	Rhodobacterales	Rhodobacteraceae	Paracoccus
4	0.000343	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Rhodospirillaceae	Defluviococcus

Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

4	0.000343	Bacteria	Proteobacteria	Alphaproteobac	Sphingomonadales	Sphingomonadaceae	Sphingomonas
4	0.000343	Bacteria	Proteobacteria	Deltaproteobact	Desulfovibrionales	Desulfovibrionaceae	Desulfocurvus
4	0.000343	Bacteria	Proteobacteria	Deltaproteobact	Desulfurellales	Desulfurellaceae	Hippea
4	0.000343	Bacteria	Proteobacteria	Gammaproteob:	Aeromonadales	Aeromonadaceae	Aeromonas
4	0.000343	Bacteria	Proteobacteria	Gammaproteob:	Chromatiales	Chromatiaceae	Thiocystis
4	0.000343	Bacteria	Proteobacteria	Gammaproteob:	Oceanospirillales	Oceanospirillaceae	Oceanobacter
4	0.000343	Bacteria	Proteobacteria	Gammaproteob:	Oceanospirillales	Oleiphilaceae	Oleiphilus
4	0.000343	Bacteria	SR1	Unclassified			
3	0.000257	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Catenulisporaceae	Catenulispora
3	0.000257	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Klugiella
3	0.000257	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Micrococcus
3	0.000257	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Renibacterium
3	0.000257	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioptaceae	Thermobifida
3	0.000257	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Promicromonosporaceae	Isosporicola
3	0.000257	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Microlunatus
3	0.000257	Bacteria	Chrysiogenetes	Chrysiogenetes	Chrysiogenales	Chrysiogenaceae	Chrysiogenes
3	0.000257	Bacteria	Firmicutes	Bacilli	Bacillales	Thermoactinomyetaceae	Thermoflavimicrobi
3	0.000257	Bacteria	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Allofustis
3	0.000257	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Caminicella
3	0.000257	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Catonella
3	0.000257	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Thermincola
3	0.000257	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Megamonas
3	0.000257	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Megasphaera
3	0.000257	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Succiniclasticum
3	0.000257	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Caldanaerovirga
3	0.000257	Bacteria	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Allobaculum
3	0.000257	Bacteria	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Catenibacterium
3	0.000257	Bacteria	OD1	Unclassified			
3	0.000257	Bacteria	Proteobacteria	Alphaproteobac	Caulobacterales	Hyphomonadaceae	Henriciella
3	0.000257	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Hyphomicrobiaceae	Prosthecomicrobiun
3	0.000257	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Methylcystaceae	Terasakiella
3	0.000257	Bacteria	Proteobacteria	Alphaproteobac	Rhodobacterales	Rhodobacteraceae	Amaricoccus
3	0.000257	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Acetobacteraceae	Roseococcus
3	0.000257	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Burkholderiales_incerta	Paucibacter
3	0.000257	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Comamonadaceae	Malikia
3	0.000257	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Comamonadaceae	Pseudorhodofera
3	0.000257	Bacteria	Proteobacteria	Betaproteobacte	Hydrogenophilales	Hydrogenophilaceae	Tepidiphilus
3	0.000257	Bacteria	Proteobacteria	Betaproteobacte	Neisseriales	Neisseriaceae	Kingella
3	0.000257	Bacteria	Proteobacteria	Gammaproteob:	Enterobacteriales	Enterobacteriaceae	Arsenophonus
3	0.000257	Bacteria	Proteobacteria	Gammaproteob:	Thiotrichales	Thiotrichales_incertae_s	Caedibacter
3	0.000257	Bacteria	Proteobacteria	Gammaproteob:	Vibrionales	Vibrionaceae	Vibrio
3	0.000257	Bacteria	Proteobacteria	Gammaproteob:	Xanthomonadales	Xanthomonadaceae	Xanthomonas
3	0.000257	Bacteria	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae	Thermotoga
2	0.000172	Bacteria	Acidobacteria	Acidobacteria_	Unclassified		
2	0.000172	Bacteria	Acidobacteria	Acidobacteria_	Unclassified		
2	0.000172	Bacteria	Acidobacteria	Holophagae	Acanthopleuribacteri	Acanthopleuribacteraceae	Acanthopleuribacter
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Actinosynnemataceae	Actinokineospora
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Bogoriellaceae	Bogoriella
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Frankineae_incertae_sec	Fodinicola
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Glycomycetaceae	Stackebrandtia
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Terracoccus
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Frigoribacterium
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Actinocatenispora
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Catellatospora
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Longispora
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioideae	Jiangella
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioideae	Nocardioides

Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Promicromonosporaceae	Promicromonospora
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Propioniferax
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardaceae	Saccharomonospora
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Sporichthyaceae	Sporichthya
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Streptomycetaceae	Kitasatospora
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Streptosporangiaceae	Nonomurea
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Thermomonosporaceae	Spirillospora
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae	Denitrobacterium
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Coriobacteriales	Coriobacteriaceae	Olsenella
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Solirubrobacterales	Conexibacteraceae	Conexibacter
2	0.000172	Bacteria	Actinobacteria	Actinobacteria	Solirubrobacterales	Patulibacteraceae	Patulibacter
2	0.000172	Bacteria	Aquificae	Aquificae	Aquificales	Desulfurobacteriaceae	Desulfurobacterium
2	0.000172	Bacteria	Aquificae	Aquificae	Aquificales	Unclassified	
2	0.000172	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
2	0.000172	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Cryomorpaceae	Cryomorpha
2	0.000172	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Arenibacter
2	0.000172	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Wautersiella
2	0.000172	Bacteria	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Cytophagaceae	Effluviibacter
2	0.000172	Bacteria	Chlamydiae	Chlamydiae	Chlamydiales	Chlamydiaceae	Chlamydia
2	0.000172	Bacteria	Chloroflexi	Anaerolineae	Anaerolineales	Anaerolineaceae	Anaerolinea
2	0.000172	Bacteria	Cyanobacteria	Cyanobacteria	Cyanobacteria_order	Chloroplast	Streptophyta
2	0.000172	Bacteria	Cyanobacteria	Cyanobacteria	Cyanobacteria_order	Family_X	GpX
2	0.000172	Bacteria	Deinococcus-Thermus	Deinococci	Thermales	Thermaceae	Vulcanithermus
2	0.000172	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Alkalibacillus
2	0.000172	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Halalkalibacillus
2	0.000172	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Paucisalibacillus
2	0.000172	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Piscibacillus
2	0.000172	Bacteria	Firmicutes	Bacilli	Bacillales	Thermoactinomycetacea	Desmospora
2	0.000172	Bacteria	Firmicutes	Bacilli	Bacillales	Thermoactinomycetacea	Thermoactinomyces
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Geosporobacter
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XI	Gallicola
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XI	Sporanaerobacter
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XII	Acidaminobacter
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XII	Fusibacter
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XII	Guggenheimella
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XIV	Proteiniborus
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptostreptococcaceae	Sporacetigenium
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ethanoligenens
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Sporobacter
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Subdoligranulum
2	0.000172	Bacteria	Firmicutes	Clostridia	Clostridiales	Syntrophomonadaceae	Thermosyntropha
2	0.000172	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Caldanaerobius
2	0.000172	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Unclassified
2	0.000172	Bacteria	Fusobacteria	Fusobacteria	Fusobacteriales	Leptotrichiaceae	Sebaldella
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Aurantimonadaceae	Aurantimonas
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Bradyrhizobiaceae	Afipia
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Methylobacteriaceae	Methylobacterium
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Rhizobiaceae	Ensifer
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Rhizobiaceae	Rhizobium
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Rhodobiaceae	Rhodobium
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhodobacterales	Rhodobacteraceae	Nereida
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhodobacterales	Rhodobacteraceae	Rhodobacter
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhodobacterales	Rhodobacteraceae	Roseibium
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhodobacterales	Rhodobacteraceae	Roseisalinus
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Acetobacteraceae	Acidiphilium
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Acetobacteraceae	Stella
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Rhodospirillaceae	Nisaea
2	0.000172	Bacteria	Proteobacteria	Alphaproteobac	Sneathiellales	Sneathiellaceae	Sneathiella

Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

2	0.000172	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Alcaligenaceae	Alcaligenes
2	0.000172	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Burkholderiales_incerta	Methylibium
2	0.000172	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Comamonadaceae	Verminephrobacter
2	0.000172	Bacteria	Proteobacteria	Betaproteobacte	Hydrogenophilales	Hydrogenophilaceae	Hydrogenophilus
2	0.000172	Bacteria	Proteobacteria	Betaproteobacte	Neisseriales	Neisseriaceae	Microvirgula
2	0.000172	Bacteria	Proteobacteria	Betaproteobacte	Neisseriales	Neisseriaceae	Paludibacterium
2	0.000172	Bacteria	Proteobacteria	Betaproteobacte	Rhodocyclales	Rhodocyclaceae	Rhodocyclus
2	0.000172	Bacteria	Proteobacteria	Betaproteobacte	Rhodocyclales	Rhodocyclaceae	Shinella
2	0.000172	Bacteria	Proteobacteria	Deltaproteobact	Desulfobacterales	Desulfobacteraceae	Desulfatibacillum
2	0.000172	Bacteria	Proteobacteria	Deltaproteobact	Desulfobacterales	Desulfobacteraceae	Desulfonema
2	0.000172	Bacteria	Proteobacteria	Deltaproteobact	Desulfobacterales	Desulfobacteraceae	Desulfosarcina
2	0.000172	Bacteria	Proteobacteria	Deltaproteobact	Desulfovibrionales	Desulfovibrionaceae	Lawsonia
2	0.000172	Bacteria	Proteobacteria	Deltaproteobact	Myxococcales	Phaselicystidaceae	Phaselicystis
2	0.000172	Bacteria	Proteobacteria	Epsilonproteob	Campylobacterales	Helicobacteraceae	Sulfurimonas
2	0.000172	Bacteria	Proteobacteria	Epsilonproteob	Campylobacterales	Helicobacteraceae	Wolinella
2	0.000172	Bacteria	Proteobacteria	Epsilonproteob	Nautiliales	Nautiliaceae	Nitratiruptor
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Alteromonadales	Alteromonadaceae	Glaciecola
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Alteromonadales	Alteromonadaceae	Saccharophagus
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Alteromonadales	Colwelliaceae	Colwellia
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Cardiobacteriales	Cardiobacteriaceae	Dichelobacter
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Chromatiales	Chromatiaceae	Thiorhodococcus
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Enterobacteriales	Enterobacteriaceae	Brenneria
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Enterobacteriales	Enterobacteriaceae	Citrobacter
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Enterobacteriales	Enterobacteriaceae	Kluyvera
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Enterobacteriales	Enterobacteriaceae	Trabulsilla
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Gammaproteobacteri	Gammaproteobacteria_i	Simidiuia
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Gammaproteobacteri	Gammaproteobacteria_i	Thiohalomonas
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Legionellales	Legionellaceae	Legionella
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Methylococcales	Methylococcaceae	Unclassified
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Oceanospirillales	Oceanospirillaceae	Bermanella
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Oceanospirillales	Oceanospirillaceae	Marinospirillum
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Oceanospirillales	Oceanospirillaceae	Neptuniibacter
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Oceanospirillales	Oceanospirillaceae	Nitricola
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Oceanospirillales	Oceanospirillaceae	Pseudospirillum
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Pseudomonadales	Moraxellaceae	Acinetobacter
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Pseudomonadales	Moraxellaceae	Enhydrobacter
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Xanthomonadales	Xanthomonadaceae	Lysobacter
2	0.000172	Bacteria	Proteobacteria	Gammaproteob	Xanthomonadales	Xanthomonadaceae	Pseudoxanthomonas
2	0.000172	Bacteria	Spirochaetes	Spirochaetes	Spirochaetales	Leptospiraceae	Leptonema
2	0.000172	Bacteria	Spirochaetes	Spirochaetes	Spirochaetales	Spirochaetaeae	Spirochaeta
2	0.000172	Bacteria	Synergistetes	Synergistia	Synergistales	Synergistaceae	Aminomonas
2	0.000172	Bacteria	Synergistetes	Synergistia	Synergistales	Synergistaceae	Dethiosulfovibrio
2	0.000172	Bacteria	Tenericutes	Mollicutes	Haloplasmatales	Haloplasmataceae	Haloplasma
2	0.000172	Bacteria	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae	Geotoga
2	0.000172	Bacteria	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae	Marinitoga
2	0.000172	Bacteria	Verrucomicrobia	Opitutae	Puniceicoccales	Puniceicoccaceae	Cerasicoccus
2	0.000172	Bacteria	Verrucomicrobia	Subdivision5	Unclassified		
1	0.000086	Archaea	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae	Halobacterium
1	0.000086	Archaea	Euryarchaeota	Halobacteria	Halobacteriales	Halobacteriaceae	Haloquadratum
1	0.000086	Archaea	Euryarchaeota	Methanobacteri	Methanobacteriales	Methanobacteriaceae	Methanosphaera
1	0.000086	Archaea	Korarchaeota	Korarchaeota_c	Korarchaeota_order_Korarchaeota_family_in		Candidatus_Korarchaeum
1	0.000086	Bacteria	Acidobacteria	Acidobacteria_	Unclassified		
1	0.000086	Bacteria	Acidobacteria	Acidobacteria_	Unclassified		
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Acidimicrobiales	Acidimicrobiaceae	Acidimicrobium
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Acidimicrobiales	Acidimicrobiaceae	Ferrihrix
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Dermacoccaceae	Demetria
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Dermatophilaceae	Kineosphaera
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Intrasporangium

Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Phycococcus
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Serinicoccus
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Clavibacter
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Yonghaparkia
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Pilimelia
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nakamurellaceae	Humicoccus
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiopsaceae	Nocardiopsis
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	Amycolatopsis
1	0.000086	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Streptosporangiaceae	Streptosporangium
1	0.000086	Bacteria	Aquificae	Aquificae	Aquificales	Aquificaceae	Thermocrinis
1	0.000086	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Coenonia
1	0.000086	Bacteria	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Cytophagaceae	Leadbetterella
1	0.000086	Bacteria	Chlamydiae	Chlamydiae	Chlamydiales	Waddliaceae	Waddlia
1	0.000086	Bacteria	Cyanobacteria	Cyanobacteria	Cyanobacteria_order	Chloroplast	Chlorarachniophyce
1	0.000086	Bacteria	Cyanobacteria	Cyanobacteria	Cyanobacteria_order	Family_VII	GpVII
1	0.000086	Bacteria	Deferribacteres	Deferribacteres	Deferribacterales	Deferribacteraceae	Denitrovibrio
1	0.000086	Bacteria	Deinococcus-Thermus	Deinococci	Thermales	Thermaceae	Thermus
1	0.000086	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Tenuibacillus
1	0.000086	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Thalassobacillus
1	0.000086	Bacteria	Firmicutes	Bacilli	Bacillales	Planococcaceae	Filibacter
1	0.000086	Bacteria	Firmicutes	Bacilli	Bacillales	Thermoactinomycetacea	Laceyella
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Natronincola
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Sarcina
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Thermotalea
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Heliobacteriaceae	Heliobacillus
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Heliobacteriaceae	Heliorestis
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XI	Tepidimicrobium
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XIII	Anaerovorax
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Incertae_Sedis_XVI	Carboxydocella
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Hespellia
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Moryella
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Oribacterium
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Syntrophococcus
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Pelotomaculum
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Acetivibrio
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Papillibacter
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Centipeda
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Sporolittus
1	0.000086	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Succinispira
1	0.000086	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Caldanaerobacter
1	0.000086	Bacteria	Firmicutes	Clostridia	Thermoanaerobacter	Thermoanaerobacteraceae	Thermacetogenium
1	0.000086	Bacteria	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae	Fusobacterium
1	0.000086	Bacteria	Planctomycetes	Planctomycetaci	Planctomycetales	Planctomycetaceae	Pirellula
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Alphaproteobacteria	Alphaproteobacteria_inc	Elioraea
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Kiloniellales	Kiloniellaceae	Kiloniella
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Bartonellaceae	Bartonella
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Bradyrhizobiaceae	Blastobacter
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Brucellaceae	Daeguia
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Hyphomicrobiaceae	Filomicrobium
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Rhizobiales	Methylobacteriaceae	Meganema
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Rhodobacterales	Rhodobacteraceae	Pannonibacter
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Rhodospirillales	Rhodospirillaceae	Tistrella
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Rickettsiales	Rickettsiaceae	Orientia
1	0.000086	Bacteria	Proteobacteria	Alphaproteobac	Sphingomonadales	Sphingomonadaceae	Sphingopyxis
1	0.000086	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Burkholderiaceae	Polynucleobacter
1	0.000086	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Burkholderiales_incerta	Aquabacterium
1	0.000086	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Comamonadaceae	Acidovorax
1	0.000086	Bacteria	Proteobacteria	Betaproteobacte	Burkholderiales	Comamonadaceae	Roseateles



Anexo 2. Tabla de anotación taxonómica por gen ribosomal 16S

1	0.000086	Bacteria	Proteobacteria	Betaproteobacte	Rhodocyclales	Rhodocyclaceae	Azonexus
1	0.000086	Bacteria	Proteobacteria	Betaproteobacte	Rhodocyclales	Rhodocyclaceae	Quatronicoccus
1	0.000086	Bacteria	Proteobacteria	Deltaproteobact	Desulfobacterales	Desulfobacteraceae	Desulfobacula
1	0.000086	Bacteria	Proteobacteria	Deltaproteobact	Desulfobacterales	Desulfobacteraceae	Desulfofaba
1	0.000086	Bacteria	Proteobacteria	Deltaproteobact	Desulfovibrionales	Desulfobacteraceae	Desulfobacterium
1	0.000086	Bacteria	Proteobacteria	Deltaproteobact	Desulfuromonadales	Desulfuromonadaceae	Pelobacter
1	0.000086	Bacteria	Proteobacteria	Epsilonproteob	Campylobacterales	Campylobacteraceae	Sulfurospirillum
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Acidithiobacillales	Acidithiobacillaceae	Acidithiobacillus
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Chromatiales	Chromatiaceae	Thiocapsa
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Chromatiales	Ectothiorhodospiraceae	Thiorhodospira
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Enterobacteriales	Enterobacteriaceae	Cedecea
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Enterobacteriales	Enterobacteriaceae	Hafnia
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Oceanospirillales	Hahellaceae	Endozoicomonas
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Oceanospirillales	Halomonadaceae	Halomonas
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Oceanospirillales	Oceanospirillaceae	Marinomonas
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Salinisphaerales	Salinisphaeraceae	Salinisphaera
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Thiotrichales	Piscirickettsiaceae	Thiomicrospira
1	0.000086	Bacteria	Proteobacteria	Gammaproteob	Xanthomonadales	Xanthomonadaceae	Xylella
1	0.000086	Bacteria	Tenericutes	Mollicutes	Entomoplasmatales	Entomoplasmataceae	Entomoplasma

1165460

Anexo 3. Tabla de anotación taxonómica por genes marcadores de copia única

Conteo	Porcentaje	Dominio	Filum	Clase	Orden	Familia	Genero_especie
411436	32.004766846	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus plantarum
343072	26.686870793	Bacteria	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Weissella paramesenteroides
281210	21.874752051	Bacteria	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Leuconostoc mesenteroides
106782	8.3063538761	Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Aerococcus viridans
41245	3.2083643837	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus faecium
24012	1.8678444801	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus italicus
18529	1.4413330989	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Lactococcus garvieae
13598	1.0577606713	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus faecalis
12369	0.9621592693	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus aureus
8777	0.6827449193	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus farciminis
7506	0.5838764229	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Lactococcus lactis
3149	0.244954284	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus brevis
2846	0.2213845323	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus rhamnosus
2291	0.1782122149	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus casei
1990	0.1547980391	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus saprophyticus
1310	0.1019022268	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Micrococcus caseolyticus
1245	0.0968460094	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus casseliflavus
1032	0.0802771741	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Corynebacteriaceae	Corynebacterium variabile
837	0.065108522	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Tetragenococcus halophilus
586	0.0455837442	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus gallinarum
265	0.0206138092	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Micrococcaceae	Kocuria rhizophila
233	0.0181245945	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus paracasei
220	0.0171133511	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus sp. 7L76
117	0.0091011912	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Pediococcus pentosaceus
105	0.0081677357	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus carnosus
96	0.0074676441	Bacteria	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Leuconostoc citreum
74	0.005756309	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus epidermidis
69	0.0053673692	Bacteria	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Weissella cibaria
49	0.00381161	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus coryniformis
47	0.0036560341	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Micrococcaceae	Micrococcus luteus
45	0.0035004582	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Propionibacteriaceae	Propionibacterium acnes
39	0.0030337304	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Klebsiella pneumoniae
36	0.0028003665	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Corynebacteriaceae	Corynebacterium ammoniagenes
36	0.0028003665	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Micrococcaceae	Arthrobacter arilaitensis
30	0.0023336388	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus buchneri
24	0.001866911	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Brevibacteriaceae	Brevibacterium linens
23	0.0017891231	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Vibrio caribbenthicus
19	0.0014779712	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus haemolyticus
18	0.0014001833	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Klebsiella oxytoca
15	0.0011668194	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Dermacoccaceae	Dermacoccus sp. Ellin185
15	0.0011668194	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Enterobacter cloacae
14	0.0010890314	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Dermabacteraceae	Brachybacterium faecium
10	0.0007778796	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus cereus
10	0.0007778796	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Micromonosporaceae	Micromonospora sp. ATCC 39149
7	0.0005445157	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus hominis
7	0.0005445157	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Agrobacterium tumefaciens
7	0.0005445157	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas mendocina
5	0.0003889398	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Pediococcus acidilactici
5	0.0003889398	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Nocardiodiaceae	Kribbella flavida
5	0.0003889398	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Streptomycetaceae	Streptomyces sp. SPB74
5	0.0003889398	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Vibrio mimicus
5	0.0003889398	Bacteria	Deinococcus-Thermu	Deinococci	Deinococcales	Deinococcaceae	Deinococcus radiodurans
4	0.0003111518	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Geodermatophilaceae	Geodermatophilus obscurus
4	0.0003111518	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Agrobacterium sp. ATCC31749
4	0.0003111518	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Enterobacter hormaechei
3	0.0002333639	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Nocardiaceae	Rhodococcus opacus
3	0.0002333639	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Gordoniaceae	Gordonia neofelifaecis
2	0.0001555759	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus vestibularis
2	0.0001555759	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus lugdunensis
2	0.0001555759	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus subtilis
2	0.0001555759	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Lysinibacillus sphaericus
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Corynebacteriaceae	Corynebacterium striatum

Anexo 3. Tabla de anotación taxonómica por genes marcadores de copia única

2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Corynebacteriaceae	Corynebacterium efficiens
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Corynebacteriaceae	Corynebacterium aurimucosum
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Corynebacteriaceae	Corynebacterium urealyticum
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Dermacoccaceae	Kytococcus sedentarius
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Nocardioideae	Aeromicrobium marinum
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Actinomycetaceae	Actinomyces sp. oral taxon 848
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Actinomycetaceae	Actinomyces viscosus
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Actinomycetaceae	Actinomyces sp. oral taxon 170
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Nocardiaceae	Rhodococcus erythropolis
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Cellulomonadaceae	Cellulomonas flavigena
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Cellulomonadaceae	Cellulomonas fimi
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Frankiaceae	Frankia sp. Eu11c
2	0.0001555759	Bacteria	Actinobacteria	Actinobacteria (class)	Actinomycetales	Nakamurellaceae	Nakamurella multipartita
2	0.0001555759	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter johnsonii
2	0.0001555759	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Enhydrobacter aerosaccus
2	0.0001555759	Bacteria	Proteobacteria	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	Aeromonas caviae
2	0.0001555759	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	Desulfovibrio vulgaris
1	0.000077788	Bacteria	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Leuconostoc kimchii
1	0.000077788	Bacteria	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Leuconostoc sp. C2
1	0.000077788	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus sp. C150
1	0.000077788	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Delftia sp. Cs1-4
1	0.000077788	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Delftia acidovorans
1	0.000077788	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Pantoea sp. aB
1	0.000077788	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Pantoea vagans

1285546

Anexo 4. Resultados de la anotación taxonómica de eucariontes con el gen ribosomal 18S

Conteo	Porcentaje	Dominio	Filum	Clase	Orden	Familia	Genero	Especie
330	16.500	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Kazachstaniaanispora
283	14.150	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomycetaceae	Debaryomycesudeni
200	10.000	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomycetaceae	Debaryomyces hansenii var.hansenii
196	9.800	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomycetaceae	Schwannomyces capriotti
155	7.750	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomycetaceae	Debaryomyces hansenii
145	7.250	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Naumovia castellii
65	3.250	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Williopsis salicomiae
65	3.250	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomycetaceae	Debaryomyces sp.MPSR02
56	2.800	Eukaryota	Fungi	Ascomycota	Dothideomycetes	Dothideomycetidae	Dothideaceae	Dothidea berberidis
48	2.400	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida dubliniensis
43	2.150	Eukaryota	Fungi	Ascomycota	Dothideomycetes	Dothideomycetidae	Teratosphaeriaceae	Teratosphaeria molleriana
27	1.350	Eukaryota	Fungi	Ascomycota	Dothideomycetes	Dothideomycetidae	Dothideaceae	Bagnisiella examinans
22	1.100	Eukaryota	Fungi	Ascomycota	Dothideomycetes	Dothideomycetidae	Mycosphaerellaceae	Mycosphaerella podagrariae
22	1.100	Eukaryota	Fungi	Ascomycota	Dothideomycetes	Dothideales	Mitosporic Dothioraceae	Kabatella microsticta
20	1.000	Eukaryota	Fungi	Fungiincertaesedis	Basalfungallineages	Unclassified zygomycete	Zygomycete sp.AM-2008a	Unclassified
19	0.950	Eukaryota	Environment	Uncultured eukary	Unclassified			
18	0.900	Eukaryota	Fungi	Dikarya	Ascomycota	mitosporicAscomycota	Bispora	Bispora christiansenii
18	0.900	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Saccharomyces cerevisiae
17	0.850	Eukaryota	Fungi	Dikarya	Ascomycota	mitosporicAscomycota	Moniliella	Moniliella suaveolens
17	0.850	Eukaryota	Fungi	Basidiomycota	Agaricomycotina	Sebacinales	Sebacinaceae	Sebacina vermifera
17	0.850	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Zygosaccharomyces lentus
16	0.800	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Williopsis pratensis
16	0.800	Eukaryota	Fungi	Ascomycota	Pezizomycotina	Pezizales	Sarcosomataceae	Urnula hiemalis
12	0.600	Eukaryota	Fungi	Basidiomycota	Agaricomycotina	Dacrymycetales	Dacrymycetaceae	Calocera viscosa
10	0.500	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomycetaceae	Lodderomyces elongisporus
8	0.400	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Saccharomyces spencerorum
8	0.400	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Lachancea kluyveri
7	0.350	Eukaryota	Fungi	Dikarya	Ascomycota	Mitosporic Ascomycota	Moniliella	Moniliella acetoabutans
7	0.350	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida lignicola
5	0.250	Eukaryota	Amoebozoa	Centramoebida	Acanthamoebidae	Protacanthamoeba	Protacanthamoeba bohemica	Unclassified
4	0.200	Eukaryota	Metazoa	Vertebrata	Eutheria	Sciurognathi	Murinae	Mus musculus
4	0.200	Eukaryota	Fungi	Environmental	unculturedfungus	Unclassified		
4	0.200	Eukaryota	Fungi	Basidiomycota	Exobasidiomycetes	Entylomatales	Mitosporic Entylomataceae	Tilletiopsis oryzicola
4	0.200	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Pichia sp.
4	0.200	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Dekkera anomala
4	0.200	Eukaryota	Fungi	Ascomycota	Eurotiomycetes	Eurotiomycetidae	Ajellomycetaceae	Ajellomyces capsulatus
3	0.150	Eukaryota	Fungi	Chytridiomycota	Chytridiomycetes	Lobulomycetales	Unclassified Lobulomycetales	Lobulomycetales sp.AF011
3	0.150	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Cyberlindnera saturnus
3	0.150	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida sp.SK75
3	0.150	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida lyxosophila
3	0.150	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Blastobotrys adenivorans
2	0.100	Eukaryota	Viridiplantae	Tracheophyta	eudicotyledons	asterids	Acanthaceae	Avicennia marina
2	0.100	Eukaryota	Fonticula	Fonticula	Fonticulaalba	Unclassified		
2	0.100	Eukaryota	Metazoa	Porifera	Ceractinomorpha	Poecilosclerida	Crambe	Crambe crambe
2	0.100	Eukaryota	Fungi	Fungi incertae sedi	Mucoromycotina	Mucorales	Mucor	Mucor mucedo
2	0.100	Eukaryota	Fungi	Dikarya	Ascomycota	Mitosporic Ascomycota	Phaeomoniella	Phaeomoniella prunicola
2	0.100	Eukaryota	Fungi	Dikarya	Ascomycota	Mitosporic Ascomycota	Graphium	Graphium rubrum
2	0.100	Eukaryota	Fungi	Basidiomycota	Agaricomycotina	Sebacinales	Sebacinaceae	Craterocolla cerasi
2	0.100	Eukaryota	Fungi	Basidiomycota	Agaricomycotina	Polyporales	Sparassidaceae	Sparassis sp.TENN50289/ss31
2	0.100	Eukaryota	Fungi	Basidiomycota	Agaricomycotina	Hymenochaetales	Hymenochaetales incertae sedi	Resinicium saccharicola
2	0.100	Eukaryota	Fungi	Basidiomycota	Agaricomycotina	Cystofilobasidiales	Cystofilobasidiaceae	Cystofilobasidium infirmominium
2	0.100	Eukaryota	Fungi	Ascomycota	Taphrinomycotina	Taphrinales	Taphrinaceae	Taphrina deformans
2	0.100	Eukaryota	Fungi	Ascomycota	Sordariomycetes	Sordariomycetidae	Lasiosphaeriaceae	Podospora asnerina
2	0.100	Eukaryota	Fungi	Ascomycota	Sordariomycetes	Microascales	Mitosporic Halosphaeriaceae	Monodictys castaneae
2	0.100	Eukaryota	Fungi	Ascomycota	Sordariomycetes	Hypocreomycetidae	Hypocreaceae	Hypocrea lixii
2	0.100	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Saccharomyces servazzii
2	0.100	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Saccharomyces bayanus
2	0.100	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida sp.BG02-7-20-001A-2-1
2	0.100	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida silvanorum
2	0.100	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Metschnikowiaceae	Metschnikowia sp.UWOP504-218.3
2	0.100	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomycetaceae	Meyerozyma guilliermondii ATCC6260
2	0.100	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomycetaceae	Meyerozyma guilliermondii
2	0.100	Eukaryota	Alveolata	Apicomplexa	Aconoidasida	Babesiidae	Babesia	Babesia odcoilei
1	0.050	Eukaryota	Metazoa	Malacostraca	Eucarida	Brachyura	Homoloidae	Latreilla valida
1	0.050	Eukaryota	Metazoa	Craniata	Euteleostomi	Teleostei	Ophichthidae	Echiophis punctifer
1	0.050	Eukaryota	Fungi	Glomeromycota	Glomeromycetes	Glomeraceae	Environmental samples	Uncultured Glomus
1	0.050	Eukaryota	Fungi	Fungi incertae sedi	Mucoromycotina	Mucorales	Zygorhynchus	Zygorhynchus heterogamus
1	0.050	Eukaryota	Fungi	Dikarya	Taphrinomycotina	Neolectales	Neolecta	Neolecta irregularis
1	0.050	Eukaryota	Fungi	Dikarya	environmentalsamples	unculturedDikarya	Unclassified	
1	0.050	Eukaryota	Fungi	Dikarya	Ascomycota	mitosporicAscomycota	Tetracladium	Tetracladium marchalianum
1	0.050	Eukaryota	Fungi	Dikarya	Ascomycota	mitosporicAscomycota	Spicellum	Spicellum roseum
1	0.050	Eukaryota	Fungi	Dikarya	Agaricomycotina	Agaricomycetes	Taiwanofungus	Taiwanofungus camphoratus
1	0.050	Eukaryota	Fungi	Chytridiomycota	environmentalsamples	Uncultured Chytridiomycota	Unclassified	
1	0.050	Eukaryota	Fungi	Basidiomycota	Pucciniomycotina	Agaricostilbomycetes	Mitosporic Agaricostilbomycetes	Bensingtonia musae
1	0.050	Eukaryota	Fungi	Basidiomycota	Agaricomycotina	Cystofilobasidiales	Cystofilobasidiaceae	Cystofilobasidium ferigula
1	0.050	Eukaryota	Fungi	Basidiomycota	Agaricomycotina	Agaricomycetidae	Agaricales	Uncultured Russulales
1	0.050	Eukaryota	Fungi	Ascomycota	Taphrinomycotina	Taphrinales	Taphrinaceae	Taphrina tosquinetii
1	0.050	Eukaryota	Fungi	Ascomycota	Taphrinomycotina	Taphrinales	Taphrinaceae	Taphrina farlowii
1	0.050	Eukaryota	Fungi	Ascomycota	Taphrinomycotina	Schizosaccharomycetales	Schizosaccharomycetaceae	Schizosaccharomyces japonicus
1	0.050	Eukaryota	Fungi	Ascomycota	Sordariomycetes	Hypocreomycetidae	Cordycipitaceae	Cordyceps bifusispora
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Wickerhamomycetaceae	Wickerhamomyces sydowiorum
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycopsidaceae	Saccharomycopsis fibuligera
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Saccharomyces cerevisiaeAWR1796
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Pichia triangularis

Anexo 4. Resultados de la anotación taxonómica de eucariontes con el gen ribosomal 18S

1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Pichia sp.SG3L01
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Pichia pijperi
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Pichia mexicana
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Kodamaea ohmeri
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Kluyveromyces wickerhamii
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Kazachstania sp.QMW-2009a
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Kazachstania exigua
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Saccharomycetaceae	Dekkera bruxellensis
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida sp.NRRLY-27170
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida psychrophila
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida palmioleophila
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida lassenensis
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Candida fennica
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Mitosporic Saccharomycetales	Brettanomyces naardenensis
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Endomycetaceae	Endomyces sp.
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomycetaceae	Debaryomyces maranus
1	0.050	Eukaryota	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Kazachstania	Kazachstania telluris
1	0.050	Eukaryota	Fungi	Ascomycota	Pezizomycotina	Pezizales	Sarcoscyphaceae	Sarcoscypha austriaca
1	0.050	Eukaryota	Fungi	Ascomycota	Pezizomycotina	Helotiales	Sclerotiniaceae	Botryotinia fuckelianaB05.10
1	0.050	Eukaryota	Fungi	Ascomycota	Dothideomycetes	Pleosporomycetidae	Delitschiaceae	Delitschia winteri
1	0.050	Eukaryota	Fungi	Ascomycota	Dothideomycetes	Dothideomycetidae	Teratosphaeriaceae	Teratosphaeriaceae sp. D007_09
1	0.050	Eukaryota	Alveolata	Dinophyceae	Syndiniales	environmentalsamples	Uncultured syndiniales	Unclassified
1	0.050	Eukaryota	Alveolata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Amphidinium	Amphidinium belauense
1	0.050	Eukaryota	Alveolata	Ciliophora	Nassophorea	Microthoracida	Leptopharynx	Leptopharynx costatus

2000



# Metagenomic analysis of a Mexican ripened cheese reveals a unique complex microbiota



Alejandra Escobar-Zepeda <sup>a</sup>, Alejandro Sanchez-Flores <sup>b</sup>, Maricarmen Quirasco Baruch <sup>a,\*</sup>

<sup>a</sup> Facultad de Química, Departamento de Alimentos y Biotecnología, Av. Universidad 3000, Universidad Nacional Autónoma de México, C. U., 04510, D. F., Mexico

<sup>b</sup> Instituto de Biotecnología, Unidad de Secuenciación Masiva y Bioinformática, Av. Universidad 2001, Chamilpa, 62210, Cuernavaca, Morelos, Mexico

## ARTICLE INFO

### Article history:

Received 6 August 2015

Received in revised form

9 February 2016

Accepted 9 February 2016

Available online 18 February 2016

### Keywords:

Raw-milk ripened cheese

Fermented food metagenomics

Bacteriocins

Lactic acid bacteria

High-throughput sequencing

## ABSTRACT

Cotija cheese is a Mexican handcrafted product made from raw cow milk whose ripening process occurs spontaneously and, presumably, it is influenced by environmental conditions. Its sensory characteristics and safety are probably the result of the balance between microbial populations and their metabolic capacity. In this work, we studied the dominance and richness of the bacteria in the Cotija cheese microbiome, as well as their metabolic potential by high-throughput sequencing. By the analysis of 16S ribosomal sequences, it was found that this metagenome is composed mainly of three dominant genera: *Lactobacillus*, *Leuconostoc* and *Weissella*, and more than 500 of non-dominant genera grouped in 31 phyla of both bacteria and archaea. The analysis of single-copy marker genes reported a similar result for dominant genera, although with greater resolution that reached the species level. Pathogenic bacteria such as *Salmonella*, *Listeria monocytogenes*, *Brucella* or *Mycobacterium* were not found. The Cotija cheese microbiome has the metabolic capacity for the synthesis of a wide range of flavor compounds, mainly involved with the metabolism of branched chain amino acids and free fatty acids. Genes associated with bacteriocin production and immunity were also found. Arguably, this is one of the most diverse metagenomes among the microbial communities related to fermented products.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

In Mexico, livestock breeding started approximately 500 years ago, with the arrival of the Spaniards, who introduced cows to the new region and, consequently, dairy products. Since then, ripened handmade cheeses of excellent quality have been produced in a traditional way. Some of these kinds of cheeses are Chihuahua, Ranchero and Cotija, which currently enjoy of great popularity and a significant economic impact throughout the country. Cotija cheese in particular has been manufactured with the same process over the last 400 years. The region where this cheese is produced is characterized by a deciduous forest vegetation that allows livestock to graze freely, resulting in the production of high-quality milk and cheese. This particular region comprises a surface of 2400 km<sup>2</sup> and an altitude of 700–1700 m above the sea level and its topography has a clear influence on the temperature (20–25 °C), rain

(900–1200 mm/year) and humidity (50–95% RH), parameters which are of utter important for ripening, as the ripening process is made at environmental conditions. The Cotija cheese production process starts with raw cow milk and with no addition of a starter culture. The manufacturing process is fairly established and no thermal treatments are used. Nevertheless, there are still some parameters that are variable among farmhouse producers, which is the case of clot cutting, quantity of grain sea salt added (~5% w/w), pressing and ripening. For instance, the ripening time must be of three months, but it can be extended for years (Álvarez et al., 2005). The Cotija cheese has been recognized internationally because of its sensory qualities with the award to the best foreign cheese at the 2006 World Championship of Quality Cheese in Cremona, Italy, however, the reputation of authentic Cotija cheese has been affected by the production and sale of lower-quality imitation and mislabeled cheeses, which sometimes show the presence of pathogens such as *Salmonella* (Centers for Disease Control and Prevention (CDC), 2008).

In cheese preparation using raw milk, microorganisms from different environmental sources, such as udder skin, grassland or stable soil, cow feces and even the farmer's skin, can be

\* Corresponding author.

E-mail addresses: [alexsf@ibt.unam.mx](mailto:alexsf@ibt.unam.mx) (A. Sanchez-Flores), [quirabma@unam.mx](mailto:quirabma@unam.mx) (M. Quirasco Baruch).

incorporated into the fermentation and ripening processes. The manufacture of Cotija cheese includes manual steps, such as whey draining and curd kneading for salting, where curd is in contact with wooden surfaces as tables or vats, which are also important microbial sources (Montel et al., 2014). Finally, another ingredient in the manufacture of this particular cheese is raw sea salt, which could be an additional source of microorganisms, especially halophilic bacteria (García, 2011). Previous studies on Cotija cheese bacterial diversity based on culture and PCR-DGGE of the V3 variable region of 16S rDNA, allowed the identification of ten bacterial genera: *Bacillus*, *Staphylococcus*, *Enterococcus*, *Lactococcus*, *Marinilactibacillus*, *Streptococcus*, *Vagococcus*, *Virgibacillus*, *Pediococcus* and *Lactobacillus* (Flores-Magallón et al., 2011; García, 2011; Zúñiga, 2009), however, the sensory characteristics of cheese made from raw milk are probably the result of a wider microbial population. The bacteria found in the final ripened dairy product are those which survived and that successfully adapted to the changing physicochemical environment, displacing other competitors during the ripening process. The role of each organism in the community is not yet fully understood in terms of the production of compounds related to flavor and aroma, but in particular to food safety, nonetheless, it is known that physicochemical changes and the production of antimicrobial compounds lead to a decrease of gram negative bacteria population, including pathogen and coliform bacteria at the end of the ripening time (Beresford and Williams, 2004).

Recently, the study of microbial consortia from different environments has been conducted through the analysis of the total DNA present in the sample, which is known as metagenomics (Handelsman et al., 1998). Metagenomics is a powerful methodology for the detection and identification of not only dominant genera, but also of those which are rare or have low prevalence (Sogin et al., 2006); another important feature of metagenomics is that it allows the characterization of the gene collection from a microbial community and not only its taxonomy. A metagenomics approach has been used successfully to inquire about functional and taxonomic features of microbial communities in fermented food studies (Jung et al., 2011; Nalbantoglu et al., 2014). Recently, it has been also used to monitor microbial abundance changes in a mock community of a surface-ripened cheese through time (Dugat-Bony et al., 2015) and to investigate metabolic potential in samples of three kinds of naturally aged cheeses (Wolfe et al., 2014).

The aim of this work is to explore the microbial richness of authentic ripened Cotija cheese, as well as to investigate the gene battery related with metabolic activities that could be involved in its sensory profile and those that could act as modulators of the composition of the bacterial consortium leading to obtain a microbiologically safe food product. The microbiota composition of Cotija cheese could give information about its microbiological quality which is relevant for several reasons, not only in order to know if pathogen bacteria are present, even in low quantities, but to establish the identity of dominant populations. The metabolism of the most abundant bacteria could modify the environment where microorganisms develop and lead to a selection of certain microbial groups during the ripening process, e.g. the production of organic acids could inhibit undesirable bacteria, such as coliforms and some pathogens.

## 2. Materials and methods

### 2.1. Cheese samples

Twenty-five 3-kg samples of authentic Cotija cheese from different farmhouses were collected and stored at  $-70^{\circ}\text{C}$  before analysis, which were manufactured by the ancient process,

following good hygiene practices. These complied the Mexican Official Standard for handcrafted ripened Cotija cheese and belonged to the following municipalities: Tocumbo (5 cheeses), Cotija (8), Quitupán (4), Jilotlán (4) and Santa María del Oro (4), (NMX-F-735-COFOCALEC, 2011). The authentic Cotija cheese is characterized by a water activity ( $a_w$ )  $< 0.91$ , a low pH value ( $\sim 5.1$ ) and redox potential around  $-116$  to  $-140$  mV (Escobar, 2012).

Each 3-kg sample was defrosted at  $4^{\circ}\text{C}$  for 18 h; then, its rind was removed using sterile instruments in a laminar flow cabinet and the rest of the piece was homogenized with a food processor (Oster, México City, México) in order to have 2 mm particles. The homogenized cheese samples were stored in labeled sterile resealable plastic bags and stored at  $4^{\circ}\text{C}$ . One-hundred-gram subsamples of each cheese were pooled for total DNA extraction.

### 2.2. Total DNA extraction

A 15-g sample of the Cotija cheese pool was homogenized in a Stomacher 400 Circulator (Seward Laboratory, London, UK) for 2 min at high paddle speed, with 40 mL of sodium citrate sterile buffer (2% w/v, pH 8.0); then 1-mL Neutrased solution (Novozymes Latin America, Paraná, Brazil) was added and it was mixed again at high paddle speed in the Stomacher for 5 min. The suspension was incubated at  $45^{\circ}\text{C}$  for 1 h under mild agitation to allow proteolysis. In order to separate fat and coarse particles, the suspension was centrifuged at  $3000\times g$  for 7 min at  $4^{\circ}\text{C}$ . The aqueous supernatant was recovered and centrifuged at  $22,000\times g$  for 10 min at  $4^{\circ}\text{C}$  and the pellet was washed three times with sterile saline solution (NaCl 0.85% w/v, pH 7.0). Before cellular rupture, each pellet was resuspended in 1 mL DNase I buffer (100 mM Tris pH 7.5, 25 mM  $\text{MgCl}_2$  and 5 mM  $\text{CaCl}_2$ ) with 1 U of DNase I (Thermo Scientific, Pittsburgh, KS, USA) and incubated for 10 min,  $37^{\circ}\text{C}$  and 300 rpm, in order to remove free DNA, mainly bovine. Subsequently, DNase I was heat inactivated at  $75^{\circ}\text{C}$  for 10 min. Bacterial lysis was performed with an enzymatic cocktail and incubation conditions were as follows: 10 mg/mL of lysozyme (Sigma–Aldrich, St. Luis, MO, USA) in buffer Tris–HCl 100 mM and 50 mM EDTA, pH 8.0 and 10 U of mutanolysin (Sigma–Aldrich, St. Luis, MO, USA) in TE buffer, incubated for 1 h at  $37^{\circ}\text{C}$ . Afterwards, 0.6 mg/mL RNase A from bovine pancreas (Thermo Scientific, Pittsburgh, KS, USA) was added and incubated for 30 min; after this incubation period, 0.5 mg/mL of proteinase K (Sigma–Aldrich, St. Luis, MO, USA) was added, and the mixture was incubated for 50 min at  $50^{\circ}\text{C}$  and then for 10 min at  $65^{\circ}\text{C}$ . Tubes were cooled down to  $37^{\circ}\text{C}$ , 50  $\mu\text{L}$  of SDS (10% w/v) were added and they were incubated for 40 min. Finally, 200  $\mu\text{L}$  of 6 M NaCl were added and slowly mixed by inversion at room temperature. The aqueous phase was extracted twice with phenol-chloroform-isoamyl alcohol (25:24:1) (pH 8) (Sigma–Aldrich, St. Luis, MO, USA) and washed twice with chloroform (Sigma–Aldrich, St. Luis, MO, USA) (Yuan et al., 2012). An additional purification step was carried out with the QIAquick purification kit (Qiagen, Inc. Germantown, MD, USA) according to manufacturer's instructions.

### 2.3. Library preparation and sequencing

An Illumina library was prepared from total DNA using the TruSeq kit v2 (Illumina, Inc., San Diego, CA, USA) following the manufacturer's specifications with an average fragment size of 500 bp. The sequencing was performed on the HiSeq2000/2500 (Illumina, Inc., San Diego, CA, USA) platform with a paired-end 100-cycle configuration at the National Laboratory of Genomics for Biodiversity (LANGEBIO), Irapuato, Guanajuato, Mexico.

Sequence data is available at SRA site from NCBI with SAMN03771674 bioproject identifier.

Read quality control was carried out using the FastQC software

(Andrews, 2015). Raw reads were mapped against the complete genome of *Bos taurus* (from NCBI genome database) with BWA software v0.7.9, (Li and Durbin, 2009) to remove cow's sequences.

#### 2.4. Taxonomic annotation

The Parallel-Meta pipeline v2.4 (Su et al., 2014) was used to assess the taxonomic annotation based on 16S rRNA marker genes in regards to the Ribosomal Database Project (RDP, Ribosomal Database Project, Michigan State University, USA) which is bundled within the software. The expectation value cut-off to assign taxonomy identity was  $1e-15$ . Additionally, the MOCAT v1.3 pipeline was used for the taxonomic assignment by comparison against their own single-copy gene databases (Kultima et al., 2012). Taxonomic assignment was performed considering an identity value of 95% and Krona charts were built with taxonomic data (Ondov et al., 2011).

#### 2.5. Metagenome assembly and dominant genomes reconstruction

*De novo* assembly was performed using IDBA-UD v1.1.1 (Peng et al., 2012) with default parameters. In order to assess the genomes of interest, total reads were mapped against reference genomes of the three dominant genera found. The most complete references in the NCBI genome database were chosen, namely *Lactobacillus plantarum* WCFS1, *Leuconostoc mesenteroides* ATCC 8293 and *Weissella paramesenteroides* ATCC 33313 (accession numbers NC\_004567.2, NC\_008531.1 and NZ\_ACKU00000000.1, respectively). To estimate coverage and abundance of each genome, the BWA software v0.7.9 was used to map reads over each mentioned reference; all assembled metagenomic contigs were also mapped on each reference using ABACAS v1.3.1 pipeline (Assefa et al., 2009) to order and orient contigs according to each reference genome. Finally, BRIG v0.95 (Alikhan et al., 2011) was used to generate ring images.

#### 2.6. Gene prediction and functional annotation

Open Reading Frame (ORF) prediction was performed by running MetaGeneMark v2.10 (Zhu et al., 2010) software on assembled sequences and product annotation of the predicted ORFs was carried out using Blast v2.2.28 + algorithm (Camacho et al., 2009) against UniProtKB/Swiss-Prot database. A functional domain annotation was performed using the hmmscan from HMMER v3.1b1 (Eddy, 2011) suite to search against Pfam-A database. All the annotation was integrated by adapting the Trinotate v2.0.1 (Grabherr et al., 2011) pipeline, and GhostKoala output annotator (GhostKOALA, 2015) was used for metabolic path visualization in KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa Laboratories, Japan).

The production of flavor and odor molecules was predicted by identification of genes corresponding to the complete metabolic pathways of interest, as well as searching for specific enzymatic annotation of reactions reported in literature related to the metabolism of free amino acids and fatty acids. Bacteriocin search was performed based on Pfam annotation of structural, synthesis and immunity genes. Neighbors of the annotated genes were of interest to describe complete clusters: if some genes without Pfam annotation in the cluster were found, a search was made into the BLAST/Swiss-Prot output; and in case of no match found, a BLAST/nr annotation was used in order to describe each cluster in the most complete way. Finally, bacteriocin analysis was enriched with a BAGEL3 (van Heel et al., 2013) search performed on all three databases available in such software.

### 3. Results

#### 3.1. Sequencing, taxonomic annotation and genome reconstruction

Reads showed an average Phred score above 28, reflecting a high quality sequencing procedure. Sequencing and assembly statistics are shown in Table 1. From the total reads, about 5% was mapped on *B. taurus* genome which indicates that the procedure to remove bovine DNA was appropriate (Materials and Methods). The distribution of %GC in raw reads allows the visual prediction of the presence of two different populations, one that is dominant, with a maximum peak of GC content around 42%, and another of high-GC content whose maximum peak is observed around 71% (Supplementary 1A).

The output of the sequence analysis, corresponding to Single-Copy Marker Genes (SCMG) found in the pool of reads, was a list of 45 bacterial genera, in which 86 different species were determined (SCMG Krona chart of Supplementary 2). Three species represented more than 80% of the total abundance, namely *L. plantarum*, *L. mesenteroides* and *W. paramesenteroides*. Four genera: *Aerococcus*, *Enterococcus*, *Lactococcus* and *Staphylococcus*, were considered as the subdominant population, which represented more than 1% and less than 10% of abundance each. Additionally, 78 other bacterial genera belonged to the non-dominant population (<1% abundance), which comprises *Firmicutes*, *Actinobacteria*, *Proteobacteria* and *Deinococcus-Thermus* phyla.

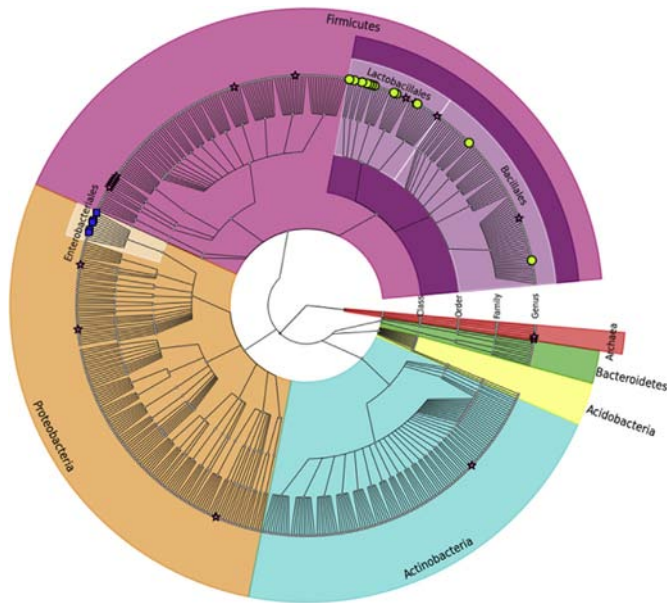
In contrast, the 16S rDNA sequence analysis, delivered a list of 574 bacterial genera (16S rDNA Krona chart of Supplementary 2). In this case, the resolution was only at genus level. The dominant population consisted of the same three genera determined by SCMG analysis, while the subdominant population was comprised of 11 genera. Non-dominant population, representing <1% of species richness, was grouped in 31 phyla of both bacteria and archaea, where *Firmicutes*, *Actinobacteria* and *Proteobacteria* were the most abundant ones (Fig. 1 and Supplementary 2).

Among the non-dominant population, *Propionibacterium*, *Bifidobacterium* and *Corynebacterium* are some of the *Actinobacteria* genera of biotechnological interest that were detected by means of both phylogenetic marker genes. Also, some halophilic and halotolerant bacteria, and marine Lactic Acid Bacteria (LAB) were found (Fig. 1). Within marine LAB are *Alkalibacterium* and *Marinilactibacillus*, both found as subdominant and described as part of soft and semi-hard cheese microbiota (Delcenserie et al., 2014; Ishikawa et al., 2007, 2013; Roth et al., 2010). The latter was previously detected in Cotija cheese in our group by PCR-dependent

**Table 1**  
Statistics of sequencing and bioinformatics analysis.

<b>Sequencing</b>	
Total number of reads	226,458,970
Total bases	22.6 Gb
Reads mapped on <i>Bos taurus</i> genome	11,300,573
<b>Assembly</b>	
Number of total bases in the assembly (Mbp)	104.9
Number of total contigs	65,162
Average size of contigs (bp)	1,610
Longest contig size (bp)	186,928
Shortest contig size (bp)	200
N50/N90 (bp)	3,854/547
<b>Annotation</b>	
Total of predicted Open Reading Frames (ORFs)	143,736
ORFs with Pfam annotation	97,063
ORFs with BLAST/Swiss-Prot annotation	75,989
ORFs with K number assigned by GhostKoala	57,747
16S rRNA reads extracted	1,165,460
Reads of Single Copy Marker Genes (SCMG) extracted	1,285,546





**Fig. 1.** Taxonomic annotation of the main bacterial phyla detected by 16S rRNA phylogenetic marker. Some clade markers are highlighted: halophilic bacteria or archaea in pink stars; coliform bacteria in blue squares; and dominant and subdominant population in green circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

methods (Zúñiga, 2009). Another marine halophilic LAB reported earlier in other cheeses and found in the Cotija as part of the non-dominant population is *Tetragenococcus halophilus* (Alegria et al., 2012; Morales et al., 2011). Among the non-dominant group, other halophilic bacteria were detected through metagenomic analysis, which have not been associated previously to cheese environments: *Dehalobacter*, *Thermohalobacter*, *Halanaerobiales*, *Haloglycomyces*, *Thiohalomonas*, *Desulfohalobium*, *Halomonas*, *Haloplasma* and two archaea, *Halobacterium* and *Haloquadratum* (Fig. 1 and Supplementary 2).

Also, within the non-dominant population determined by read abundance of 16S ribosomal gene marker, bacteria related to food sanitary quality of less than 0.001% of total richness, were identified: *Escherichia coli* (0.0007%), *Enterobacter* (0.0007%), *Campylobacter* (0.0006%), *Klebsiella* (0.0004%) and *Citrobacter* (0.0002%). There were no sequences with annotation related to pathogenic bacteria such as *Salmonella*, *Listeria monocytogenes*, *Brucella abortus*, *Mycobacterium bovis*, *M. avium*, *Yersinia* or *Pseudomonas aeruginosa* (Supplementary 2) (Montel et al., 2014).

### 3.2. Metagenomic assembly, gene prediction and functional annotation

The metagenome assembly size was about 105 Mb, as shown in Table 1. Mapping reads back to contigs, 96.75% of the total reads were recruited. Due to the high-deep sequencing it was possible to reconstruct the complete genomes of the most abundant species (Fig. 2). The percentage of reads mapped on each reference genome (Table 2) is consistent with the abundance that was estimated by phylogenetic marker counts. In all cases, the estimated coverage was over 1400× and all reconstructed genomes were larger than the references.

As for the functional annotation performed on all metagenomic contigs, the integration of BLAST and HMMER results with Trinotate allowed us to assign a function to 72% of sequences predicted as ORFs. On the other hand, the GhostKoala annotation software,

assigned a KEGG identifier (K number) to approximately 40% of total ORFs predicted (Table 1), and placed them on metabolic pathway maps, grouped in distinct functional categories, mainly involved in central cell metabolism. This information allowed us to reconstruct some metabolic pathways that are involved with the production of flavor and odor compounds. The analysis was complemented with the search of annotations of specific enzymes, because in some cases, the particular compound is not the final product of a pathway, and some additional enzymatic reactions might be needed in order to produce it (Table 3).

### 3.3. Screening of functional annotation related to cheese sensory properties

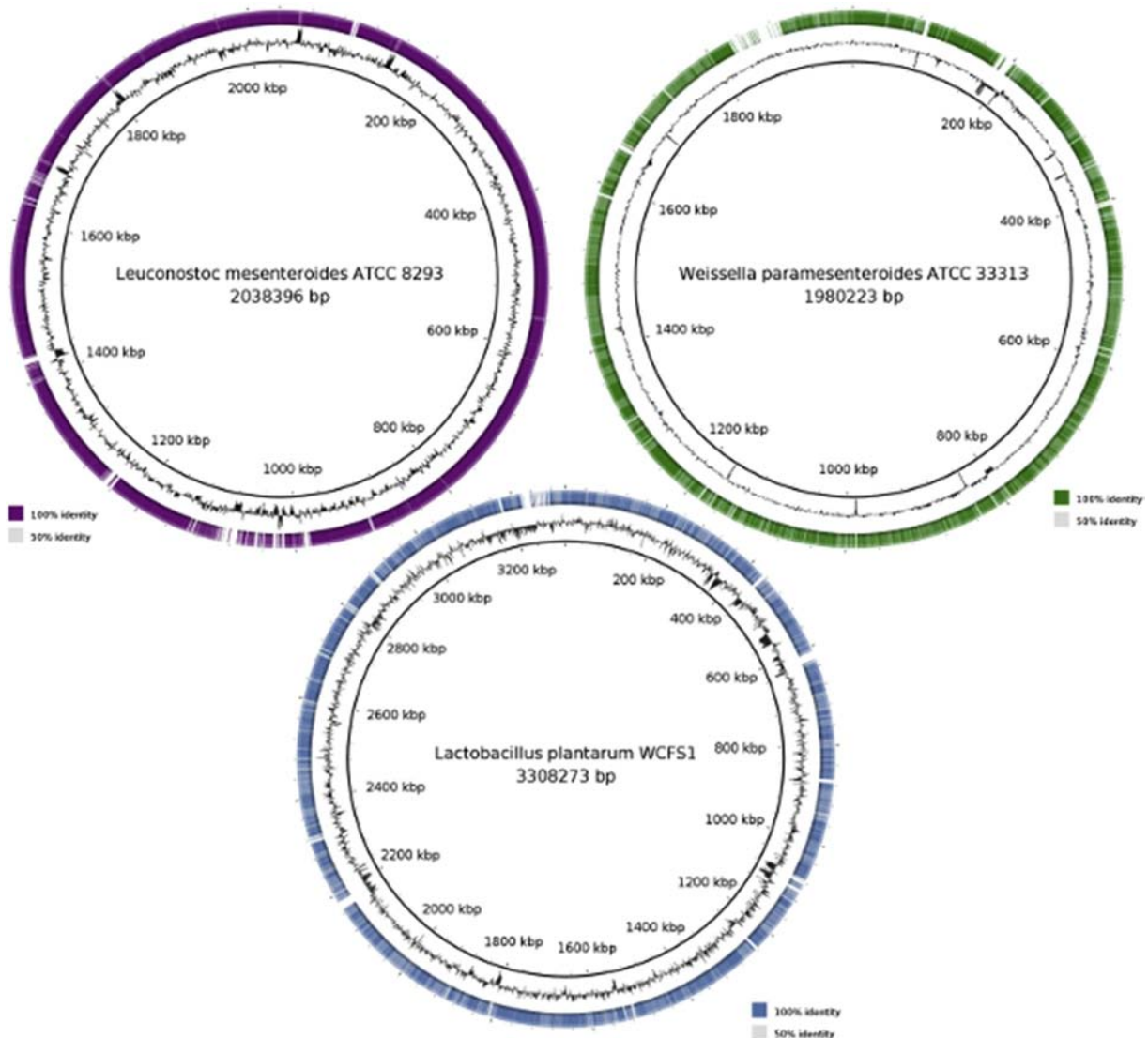
A list of the main compounds involved in flavor and/or aroma in cheese, and KEGG's metabolic pathways related to their production are shown in Supplementary material 3. Based on the GhostKoala reconstruction, the screening of such specific metabolic pathways showed that Cotija cheese metagenome has the metabolic potential to produce all the compounds mentioned in Supplementary material 3. Enzymes involved in the degradation of aromatic compounds, amino acids and fatty acids metabolism were detected (Fig. 3). A screening of Swiss-Prot identifiers of enzymes involved in fatty acid and free amino acid catabolism was performed (Table 3). Although many aminotransferases specific for phenylalanine and branched-chain amino acids were identified, no significant matches for specific tyrosine and tryptophan aminotransferases were found. In a broader search, a great variety of amino acid lyases and carboxyl ester hydrolases was also found.

### 3.4. Search for bacteriocin gene clusters

Bacteriocin genes are commonly grouped in clusters that include structural (bacteriocin-coding) and combinations of immunity, regulatory, secretion and modifying enzymes genes (Snyder and Worobo, 2014). Therefore, a strategy to find bacteriocins was to look not only for structural genes, but also for those sequences that could be part of the cluster context. The results from this analysis showed 126 contigs of different sizes containing at least one of the following gene annotations: immunity, structural or modifying enzymes. From those contigs, 9 had sequences annotated as class I modifying enzymes (lanthionine synthetase or lantibiotic dehydratase) but they did not contain any bacteriocin-coding sequence, and 74 contigs had only the immunity gene. In total, 43 contigs had at least one bacteriocin encoding gene and only 17 had structural and immunity genes together (some examples are shown in Fig. 4).

Besides, we identified 52 genes with bacteriocin annotation and 104 with bacteriocin immunity related annotation (Table 4). From the bacteriocin encoding genes, we found that only one, was a class I bacteriocin, annotated as gallidermin. The rest of bacteriocin-coding genes were annotated as class II bacteriocins. From those, most were pediocin-like or circular bacteriocins (Cotter et al., 2005), subtilisin A is included in the latter. There were some matches belonging to the lactococci-like group, which includes lactococci A and B (Nes et al., 2006). There were also three hits to lactococci 972 which acts by septum synthesis inhibition (Martínez et al., 1996).

Ten bacteriocins were identified by BAGEL3 software: four belonging to class IIb (two-peptide bacteriocins), four to class IIc (linear one-peptide, non-pediocin), one was putative and another one was reported as a class II bacteriocin.



**Fig. 2.** Reconstruction of genomes of the most abundant species in Cotija cheese metagenome. The genome of reference is in the inner ring. The middle ring represents the GC content of the reference. The comparison of the reconstructed genome with respect to the reference is in the outer ring, with color scales showing identity from BLAST analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 2**  
Genome reconstruction statistics.

	<i>L. plantarum</i> WCFS1	<i>L. mesenteroides</i> ATCC 8293	<i>W. paramesenteroides</i> ATCC 33313
Reads mapped (%)	38.74	11.27	26.85
Estimated coverage	1863×	1473×	2135×
Total of contigs	1,560	297	992
Number of non-overlapping (real) gaps	78	40	42
N50/N90 (bp)	4,985/825	36,455/2,977	3,148/1,002
Contig size average (bp)	2,201	6,744	2,196
GC content (%)	44.37	37.90	38.27
Number of genes	3,998	2,141	2,755

## 4. Discussion

### 4.1. Taxonomic characterization of the Cotija cheese bacterial consortium

Considering that Cotija cheese is made from raw milk, with no thermal treatment in the manufacturing process and that no starter

cultures are added, the initial bacterial microbiota is expected to be highly diverse. According to the FastQC analysis of raw reads and assembled metagenome sequences, our results show that such microbiome is comprised mainly by a population with low %GC content and a smaller proportion of microorganisms with high %GC content (Supplementary 1 A and B). Total reads were compared against two different databases (ribosomal and single copy marker

**Table 3**  
Specific enzymatic reactions that produce flavor and aroma compounds.

Reaction	Enzyme name	Enzyme EC number	Matches in the annotation
L-leucine + 2-oxoglutarate $\rightleftharpoons$ 4-methyl-2-oxopentanoate + L-glutamate	Branched-chain-amino-acid transaminase <sup>c</sup>	2.6.1.42	50
L-isoleucine + 2-oxoglutarate $\rightleftharpoons$ (S)-3-methyl-2-oxopentanoate + L-glutamate			
L-valine + 2-oxoglutarate $\rightleftharpoons$ 3-methyl-2-oxobutanoate + L-glutamate			
L-phenylalanine + pyruvate $\rightleftharpoons$ phenylpyruvate + L-alanine	Phenylalanine (histidine) transaminase <sup>c</sup>	2.6.1.58	11
2-oxo acid $\rightleftharpoons$ aldehyde + CO <sub>2</sub>	Alpha-ketoacid carboxylase <sup>a</sup>	4.1.1.1	6
L-threonine $\rightleftharpoons$ glycine + acetaldehyde	L-threonine aldolase <sup>c</sup>	4.1.2.48	3
Tyrosine $\rightarrow$ tyramine + CO <sub>2</sub>	Tyrosine decarboxylase <sup>a</sup>	4.1.1.25	3
Methyl ketones (alkan-2-ones) $\rightleftharpoons$ alkan-2-ol	Carbonyl reductase I <sup>a</sup>	1.1.1.184	6
2,2-xylene $\rightarrow$ 2,3-dimethylphenol + 3,4-dimethylphenol	Phenol 2-monoxygenase <sup>b</sup>	1.14.13.7	10
Aldehyde + NAD(+) $\rightarrow$ carboxylate + NADH	Aldehyde dehydrogenase <sup>a</sup>	1.2.1.3	79
An alcohol + NAD(+) $\rightleftharpoons$ an aldehyde or ketone + NADH	Alcohol dehydrogenase <sup>c</sup>	1.1.1.1	122
A secondary alcohol + NAD(+) $\rightleftharpoons$ a ketone + NADH			
(S)-methylmalonyl-CoA + pyruvate $\rightleftharpoons$ propanoyl-CoA + oxaloacetate	Methylmalonyl-CoA carboxytransferase <sup>a</sup>	2.1.3.1	5

<sup>a</sup> McSweeney, 2004.

<sup>b</sup> Urbach, 1997.

<sup>c</sup> Yvon and Rijnen, 2001.

genes) to characterize the bacterial population in Cotija cheese. The results obtained using both approaches were consistent for taxonomic identification and abundance, particularly for the most abundant organisms (Fig. 1 and Krona charts in Supplementary 2). Additionally, ribosomal gene analysis provided a higher sensitivity in terms of genera, due to the fact that the database is larger than the one used in the single-copy marker analysis.

Taxonomic annotation revealed that *Lactobacillus*, *Leuconostoc* and *Weissella* conform the dominant population; although the last two genera represent more than 40% of the relative abundance, they could not be detected by culture-dependent methods and by the V3 variable region of 16S rDNA amplification-DGGE (Flores-Magallón et al., 2011; Zúñiga, 2009), however, Chombo-Morales et al. (2016) were able to detect these genera in authentic Cotija cheese by the amplification of the V7–V8 regions of 16S rDNA.

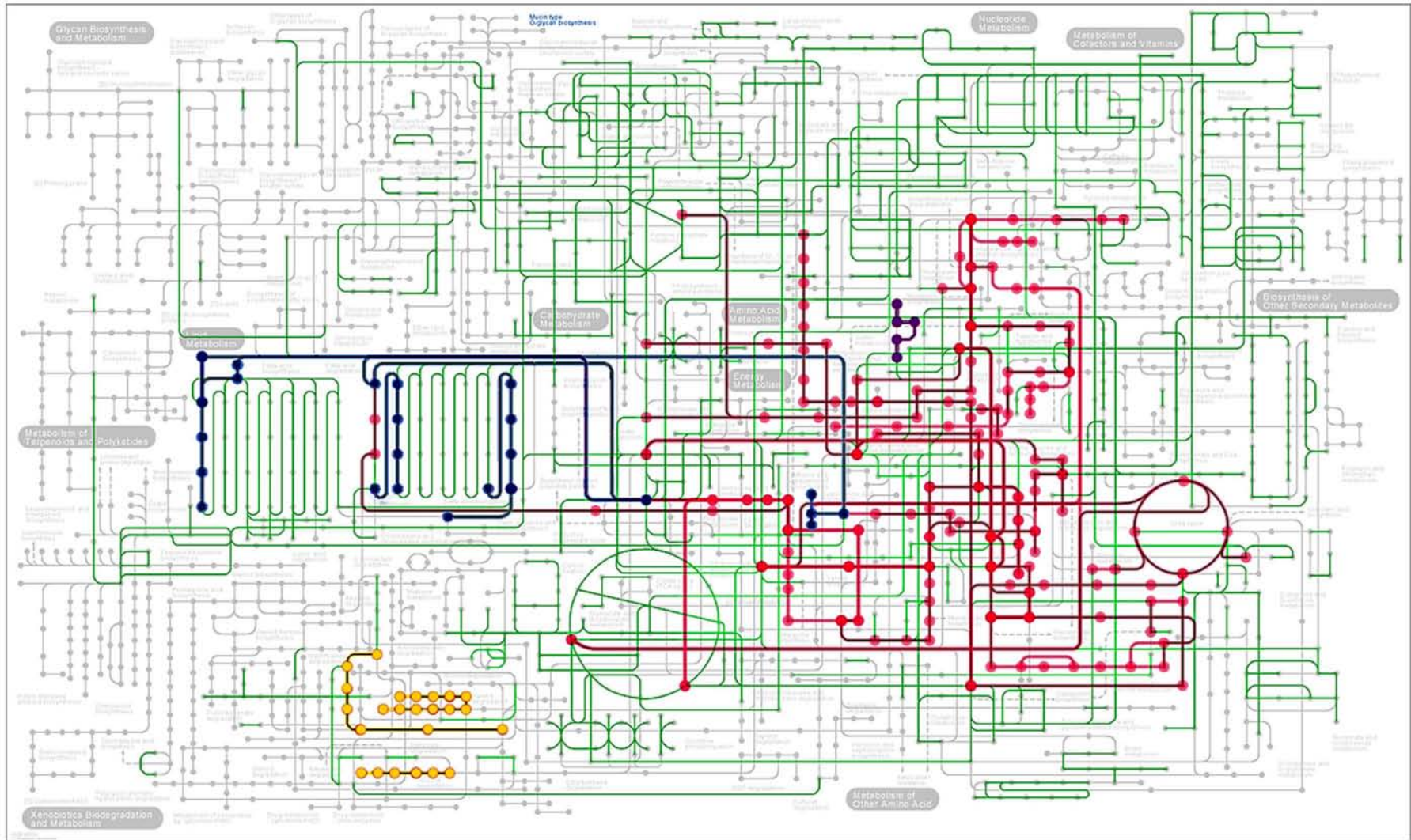
The genome reconstruction of the three dominant species *L. plantarum*, *L. mesenteroides* and *W. paramesenteroides* was possible due to a great coverage (between 1500 and 2000 $\times$ ) for each one of them (Table 2). It is noteworthy that the proportion of mapped reads for each reference genome agrees with the one calculated by the abundance of marker genes of the total metagenome reads (Supplementary material 2), suggesting that there is no bias in the abundance assessment.

The genus *Lactobacillus* is commonly found in both fresh and ripened cheeses (Bautista-Gallego et al., 2014; Ercolini et al., 2012) and its abundance has been correlated to maturation time (Quigley et al., 2012). Most of the genera detected in this work as dominant and subdominant (*Lactobacillus*, *Streptococcus*, *Lactococcus* and *Enterococcus*), have often been described as a part of the microbiota of spontaneously ripened cheese, made from non-pasteurized milk (Duthoit et al., 2003; Fuka et al., 2013; Masoud et al., 2012; Quigley et al., 2012). In contrast, *Leuconostoc* and *Weissella* are not commonly associated with cheese environment (Chombo-Morales et al., 2016; Fuka et al., 2013; Masoud et al., 2012). The presence of these particular bacteria as dominant genera could be considered as a distinctive signature of the authentic ripened Cotija cheese. The origin of *Weissella* and *Leuconostoc* species and their presence in Cotija cheese as dominant populations is uncertain, nevertheless, strains of both genus have been isolated from different sources in the manufacture of some cheeses, such as from MRS agar plates exposed to the air of cheese-making room used to make traditional Greek cheese (Litopoulou-Tzanetaki and Tzanetakis, 2014); from the wood surface of the containers used in cheese manufacture (Settanni et al., 2012), and even from cow raw milk (Masoud et al., 2012).

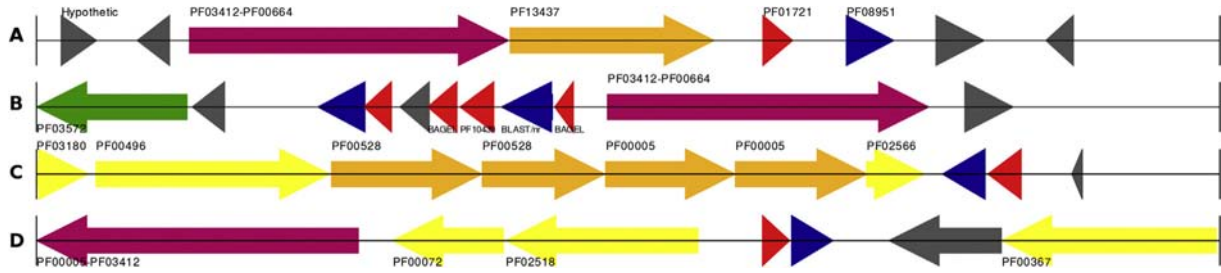
#### 4.2. Metagenomic approach applied to a microbiological safety analysis

The bacterial community of Cotija cheese is rich in LAB, practically free from coliform and it shows a total absence of pathogenic bacteria such as *L. monocytogenes*, *P. aeruginosa*, *Salmonella* spp., *Yersinia* spp. and *Brucella* spp. (Supplementary material 2), however, some bacteria from the family *Enterobacteriaceae* were detected in a very low abundance (Fig. 1 and Supplementary 2). Considering that this particular cheese has an estimated bacterial number of  $1 \times 10^5$  CFU/g, as determined by FISH (Escobar, 2012), this low proportion would correspond to a count of 1–2 CFU/g, which may not represent a risk for the consumer. Due to the fact that *E. coli* was found, a screening of genes reported as associated with virulence in eight different groups of pathogenic *E. coli*: enterotoxigenic (ETEC), enteroinvasive (EIEC), enteropathogenic (EPEC), enterohemorrhagic (EHEC), enteroaggregative (EAEC), diffusely adherent (DAEC), uropathogenic (UPEC), and neonatal meningitis-associated (NMEC), was performed (MOH Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, 2014). The UPEC and NMEC strains are not described as foodborne pathogens, nevertheless, virulence genes associated with these strains were searched for in the Cotija cheese annotation output, as well. No virulence genes belonging to any of the pathogenic groups mentioned before were found. Accordingly, we can infer that the *E. coli* strains present in Cotija cheese are non-pathogenic and thus, their presence does not represent a risk to the consumer's health.

It is worth mentioning that quality of raw materials is essential in obtaining safe dairy products and the desired sensory characteristics. Particularly in the case of the authentic farm-made Cotija cheese, producers participate into governmental programs in order to maintain their livestock free of brucellosis and tuberculosis. Besides, they follow hygienic production practices to attain the Collective Trademark and to accomplish the Mexican Standard, where it is established that a three month-ripening time should be carried out for the microbial safety of the product (Álvarez et al., 2005; NMX-F-735-COFOCALEC, 2011). The absence of enteric and pathogenic bacteria is explained by lactic and other organic acid production by LAB, pH decrease, as well as by antimicrobial peptide activity (Beresford and Williams, 2004) as depicted in Fig. 5.



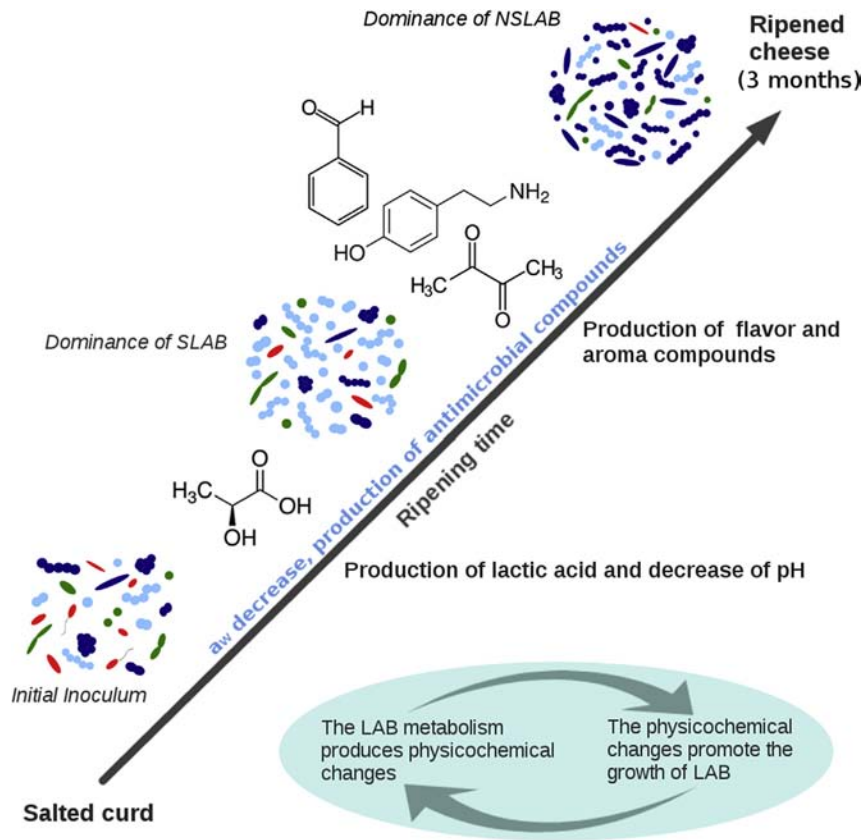
**Fig. 3.** KEGG map of metabolic pathways present in Cotija cheese metagenome. In green, all pathways represented in Cotija cheese metagenome. Modules of interest are highlighted: In yellow, aromatic compounds degradation; purple, sulfur metabolism; blue, fatty acid metabolism; pink, amino acid metabolism. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Examples of genomic context of bacteriocin clusters. A–C, clusters found in Cotija cheese metagenome. D, *Pediococcus pentosaceus* ATCC 25745 bacteriocin cluster, as a reference. Color code: red, bacteriocin structural gene (PF01721, unless other specified); blue, immunity gene (PF08951, unless other specified); orange, transport domain; green, peptidase domain; pink, transport and peptidase domains in a single gene; yellow, other annotation; gray, hypothetical protein. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 4**  
Bacteriocin and related genes found by Pfam analysis on metagenomic annotation.

Pfam ID	Description	Number of matches in the annotation output
PF02052	Gallidermin	1
PF01721	Class II bacteriocin	12
PF10439	Bacteriocin class II with double-glycine leader peptide	15
PF08951	Enterocin A immunity	88
PF05147	Lanthionine synthetase C-like protein	3
PF04369	Lactococcin-like family	7
PF04738	Lantibiotic dehydratase, N terminus	3
PF04737	Lantibiotic dehydratase, N terminus	3
PF13376	Bacteriocin-protection, YdeI or OmpD-associated	13
PF12173	Bacteriocin class IIc cyclic gassericin A-like	2
PF11420	Bacteriocin subtilosin A	2
PF09683	Bacteriocin (Lactococcin_972)	3



**Fig. 5.** Schematic representation of a ripening cheese process.

#### 4.3. Role of bacteria in Cotija cheese ripening

In Cotija cheese, there is a mixture of a variety of both homo and heterofermentative LAB species (Fig. 1 and Supplementary Material 2), enabling the production of a wide range of compounds important for food safety, sensory attributes and important biotechnological compounds, such as enzymes. At the beginning of any lactic fermentation, the most successful species are lactate producers, e.g. *Lactococcus lactis* and some species of *Lactobacillus*. In cheese, as maturation proceeds, this population is diminished, allowing the development of a secondary microbiota that comprises mainly facultative heterofermentative species, which synthesize aroma and flavor compounds due to their lipolytic and proteolytic activity during ripening, as illustrated in Fig. 5. Such is the case of *Leuconostoc* and *Weissella* genus and heterofermentative *L. plantarum* (Beresford and Williams, 2004; Björkroth and Holzapfel, 2006; McSweeney, 2004).

The enzymes necessary to produce flavor and aroma compounds in LAB are highly strain dependent, however some strains have been studied because of their activity in the production of cheese flavor compounds. *Leuconostoc* spp. is a recognized producer of diacetyl and acetoin. Also some citrate-positive *Lactococcus* strains and mesophilic lactobacilli, produce diacetyl and formate (Fox et al., 1995). Strains of *L. plantarum*, *casei* and *paracasei* produce esterases (intracellular mainly) and degrade sulfur, aromatic and branched chain amino acids to produce a variety of thiols, alcohols and aldehydes; some *L. lactis* strains produce intra- and extracellular lipases and proteases, and some other strains have been reported as capable to degrade methionine to methanethiol and methional. Additionally, some strains of *Propionibacterium*, *Carnobacterium*, and *Enterococcus faecalis* are capable of producing carboxylic acids form oxidative decarboxylation of  $\alpha$ -ketoacids; likewise, LAB including lactococci, lactobacilli, leuconostocs and pediococci are capable of generating esters from carboxylic acids and alcohols (Collins et al., 2003; Yvon and Rijnen, 2001).

In addition to LAB, a group of non-dominant halophilic microorganisms were found. These bacteria could have been inoculated with the salt used during Cotija cheese manufacture, due to the fact that unprocessed sea salt is added directly to the curd, some marine LAB were found as subdominant populations, such as *Marinilactibacillus* and *Alkalibacterium*, which might have adapted to the cheese environment because of their metabolic capacities, which can explain why they were more abundant than the other halophilic bacteria.

#### 4.4. Functional annotation and metagenome metabolic potential

It is well known that the sensory profile in a fermented food is the result of a highly complex mixture of compounds, which come from the microbial metabolism of carbohydrates, lipids and proteins, and from degradation of free amino acids and fatty acids. The most relevant flavor compounds produced by microbial activity in ripened cheese are the aldehydes, alcohols, carboxylic acids and esters. Other flavor compounds that could provide distinctive flavor notes are ketones, lactones, pyrazines, sulfurous compounds, free fatty acids (acetic and propionic mainly), free amino acids, and salts (Smit et al., 2005). Sometimes the flavor compound is not the final product of a metabolic pathway and independent enzymatic reactions could be needed in order to produce them (Table 3).

*L. plantarum*, which was detected as dominant in Cotija cheese, has been recognized as responsible for an increase of secondary proteolysis in cheese, promoting amino acid release (Milesi et al., 2008). Cotija cheese microbiome has the metabolic potential to transform free amino acids into their respective  $\alpha$ -keto acid by transamination of branched-chain and aromatic amino acids, such

as phenylalanine. Those  $\alpha$ -keto acids are degraded to aldehydes by  $\alpha$ -keto acid decarboxylases, which were also found in the annotation (Table 3). Accordingly, cheese flavor profile could be enriched by aldehydic compounds produced by the catabolism of branched chain amino acids and phenylalanine, and also with acetaldehyde derived from threonine (Table 3 and Supplementary material 3). It is interesting that no transaminases for tyrosine and tryptophan were found, due to the fact that some of their corresponding products, such as skatole, are responsible for unpleasant odors (Yvon and Rijnen, 2001).

Another important source of flavor and aroma compounds are lipids. Some lipolytic strains of *Bacillus* and *Staphylococcus* have been isolated from Cotija cheese (García, 2011), and fatty acid catabolism pathways were completely reconstructed by Ghost-Koala annotation in this metagenome (Fig. 3). The degradation of free fatty acids (FFA) is accomplished by  $\beta$ -oxidation with the concomitant generation of methyl-ketones (Table 3) (Collins et al., 2003) and the enzymes required to convert methyl-ketones into their respective secondary alcohol were found as well (Table 3). Likewise, FFA could produce esters or thioesters by an esterification reaction with an alcohol or a thiol. Alcohols can be produced from carbohydrate metabolism and thiols from sulfur amino acid metabolism. A great variety of aldehyde and alcohol dehydrogenases were found in the metagenome annotation output as well, whose products can contribute to flavor with a variety of ketones, aldehydes and carboxylates (Table 3).

Besides the degradation of proteins and lipids, the presence of benzene derivatives, e.g. toluene and xylene, is important for the flavor and aroma profiles in cheese. In lab-made products, the concentration of such derivatives was associated with the microbial composition of the starter culture (Centeno et al., 2002; Wang et al., 2012). Enzymes related to the synthesis of alcohols from xylene degradation were found in the Cotija cheese metagenome, whose presence could explain some odor notes detected in the traditional product (Table 3). Phenolic alcohols such as 2,3-dimethylphenol and 3,4-dimethylphenol are part of the sensory profile of cheese (Urbach, 1997); the former has smoky and ink-like odors and a sweet flavor, as the latter has predominantly horse stable-like odor notes (Czerny et al., 2011).

#### 4.5. Genomic context of bacteriocin genes

LAB play a main role in food safety because of the production of antimicrobial compounds such as bacteriocins, which are important in food biopreservation (Ross et al., 2002). As for the metagenome from Cotija cheese, there were 43 clusters that contained at least one coding-gene for bacteriocins, and some of them were large enough to analyze the genomic context. The only structural gene of class I bacteriocins, gallidermin, was in a cluster with two genes associated with a post translational modification for lantibiotic biosynthesis, a RepB family protein and other three genes without clear annotation. The remaining clusters contained class II bacteriocins, mainly of class IIa, described as antilisterial. In Fig. 4, some class II bacteriocin-clusters are shown. Bacteriocin and ancillary genes are some of the genetic traits most susceptible to be transferred horizontally among LAB in fermented foods, along with antibiotic resistance and immunity to bacteriophages (Rossi et al., 2014).

It is well known that bacteriocins are coded in gene clusters which might include some of the following genes: bacteriocin modifying enzymes, immunity, secretion or regulatory proteins, in any genomic context (Snyder and Worobo, 2014). In most contigs, downstream the coding bacteriocin gene there were proteins with hypothetical or no annotation (Fig. 4). Further characterization is needed in order to find out the function of those unknown

sequences. In some other cases, when the immunity protein was absent, there were two or more bacteriocin-like coding genes in tandem, which have been reported as substances with a regulatory role acting as bacterial pheromones (Snyder and Worobo, 2014).

A different scenario are contigs where only the immunity gene was found. It is not clear if the adjacent gene with no annotation corresponds to a new bacteriocin, or if the immunity gene is independent of a bacteriocin cluster. Immunity genes could be an advantage to the holder strain, although the acquisition mechanisms have to be characterized. Several non-producing bacteriocin strains have been shown to contain immunity genes, which are probably residual bacteriocin gene clusters and which could render resistant strains conferring cross-immunity (Fimland et al., 2002; Kjos et al., 2011; Mørseth et al., 2005). In the case of the reconstructed genomes from dominant species, ORFs with products annotation related to bacteriocin clusters were studied. Results from this analysis showed that there were no genomic regions with the genes of interest in *W. paramesenteroides*, however the genome of *L. mesenteroides* has a putative bacteriocin and seven contiguous ORFs annotated as hypothetical proteins. In addition, two regions with immunity genes, but no structural gene for the bacteriocin were detected, from which only one has a hypothetical protein immediately upstream. Further analysis, however, is needed to determine the presence of a new bacteriocin. As for the *L. plantarum* reconstructed genome, no coding genes for bacteriocins were found. Similarly, immunity genes were present, from which in only one region there were adjacent genes with hypothetical protein annotation. As reported for other bacterial species, an immune mimicry survival strategy in this strain can be speculated (Draper et al., 2009).

## 5. Conclusions

The use of a whole metagenome sequencing approach and bioinformatics were applied to perform a comprehensive analysis of the Cotija cheese microbiome, for taxonomical and functional characterization at a fine detail level. The bacterial microbiota in the three-month ripened Cotija cheese is conformed in ~98% of *Firmicutes*, with dominance of three genera: *Lactobacillus*, *Leuconostoc* and *Weissella*, whose genomes were entirely reconstructed at a high-quality draft level. Interestingly, this analysis allowed us to become aware of the presence of over 500 taxa, including halophilic bacteria and archaea, in the non-dominant population (<1%). This finding reflects the complexity and uniqueness of the ripening process of a handcrafted dairy product such as Cotija cheese. Considering the microbial complexity found, efforts to design a starter culture for a similar fermentation would not be worth the effort.

The use of this highly sensitive method allowed us to assess the microbial safety of a dairy product made from raw milk. These findings confirm that if good quality raw materials and hygienic practices are used, along with a three-month ripening process, this traditional practice allows farm-house producers to obtain a good microbiological quality cheese.

This consortium has the metabolic potential to generate a flavor and aroma profile mainly given by amino acid and free fatty acid catabolism, and that could be enriched with metabolites derived from branched-chain amino acids, as well as with secondary alcohols produced from short chain free fatty acids. However, an analysis of volatile compounds from Cotija cheese is needed to confirm the presence of such compounds.

Results suggest that the diversity of immunity genes against bacteriocins in Cotija cheese microbiome is an important strategy to maintain the population dynamics, however it remains unknown what the driving forces of such taxonomic diversity are. A

further molecular mining study, with other bioinformatic and molecular tools is needed in order to search for new bacteriocins and to analyze their genomic context.

Our findings can partially explain the role of a bacterial population in the sensory attribute balance in a fermented dairy product and its contribution to its safety for human consumption. Besides, the functional information generated can be helpful to search for new enzymes with biotechnological applications.

## Acknowledgments

We thank Jérôme Verleyen and Karel Estrada Guerra, at the Unidad de Secuenciación Masiva y Bioinformática-UNAM (Massive Sequencing and Bioinformatics Unit-UNAM), for their bioinformatics support and Verónica Jiménez Jacinto for uploading metagenomic data to NCBI repository database. We thank Dr. Carmen Wachter Rodarte and Dr. Miguel Ángel Cevallos Gaos for their comments and contributions to this project. Finally, we thank Arturo Vera-Ponce de Leon for his valuable contribution to the development of the metagenomic analysis method. This work was supported by PAPIIT IN222115 grant. Alejandra Escobar-Zepeda thanks CONACyT for her PhD scholarship.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.fm.2016.02.004>.

## References

- Alegria, A., Szczesny, P., Mayo, B., Bardowski, J., Kowalczyk, M., 2012. Biodiversity in Oscypek, a traditional Polish cheese, determined by culture-dependent and -independent approaches. *Appl. Environ. Microbiol.* 78, 1890–1898. <http://dx.doi.org/10.1128/AEM.06081-11>.
- Alikhan, N.-F., Petty, N.K., Zakour, N.L.B., Beatson, S.A., 2011. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genom.* 12, 402. <http://dx.doi.org/10.1186/1471-2164-12-402>.
- Álvarez, R., Barragán, E., Chombo, P., 2005. Reglas de uso. Marca colectiva. Queso Cotija Región de Origen. Asociación Regional de Productores. El Colegio de Michoacán, Mexico.
- Andrews, S., 2015. Babraham Bioinformatics – FastQC a Quality Control Tool for High Throughput Sequence Data. URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 7.17.15.).
- Assefa, S., Keane, T.M., Otto, T.D., Newbold, C., Berriman, M., 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* 25, 1968–1969. <http://dx.doi.org/10.1093/bioinformatics/btp347>.
- Bautista-Gallego, J., Alessandria, V., Fontana, M., Bisotti, S., Tarico, S., Dolci, P., Coccolin, L., Rantsiou, K., 2014. Diversity and functional characterization of *Lactobacillus* spp. isolated throughout the ripening of a hard cheese. *Int. J. Food Microbiol.* 181, 60–66. <http://dx.doi.org/10.1016/j.ijfoodmicro.2014.04.020>.
- Beresford, T., Williams, A., 2004. The microbiology of cheese ripening. In: Fox, Patrick F., P.L.H.M., Cogan, Timothy M., Guinee, Timothy P. (Eds.), *Cheese: Chemistry, Physics and Microbiology, Cheese Chemistry, Physics and Microbiology*. Academic Press, pp. 287–317.
- Björkroth, J., Holzapfel, W., 2006. Genera *Leuconostoc*, *Oenococcus* and *Weissella*. In: Dr, M.D.P., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E. (Eds.), *The Prokaryotes*. Springer US, pp. 267–319.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinform.* 10, 421. <http://dx.doi.org/10.1186/1471-2105-10-421>.
- Centers for Disease Control and Prevention (CDC), 2008. Outbreak of multidrug-resistant *Salmonella enterica* serotype Newport infections associated with consumption of unpasteurized Mexican-style aged cheese—Illinois, March 2006–April 2007. *MMWR Morb. Mortal. Wkly. Rep.* 57, 432–435.
- Centeno, J.A., Tomillo, F.J., Fernández-García, E., Gaya, P., Nuñez, M., 2002. Effect of wild strains of *Lactococcus lactis* on the volatile profile and the sensory characteristics of Ewes' raw milk cheese. *J. Dairy Sci.* 85, 3164–3172. [http://dx.doi.org/10.3168/jds.S0022-0302\(02\)74404-4](http://dx.doi.org/10.3168/jds.S0022-0302(02)74404-4).
- Chombo-Morales, P., Kirchmayr, M., Gschaedler, A., Lugo-Cervantes, E., Villanueva-Rodríguez, S., 2016. Effects of controlling ripening conditions on the dynamics of the native microbial population of Mexican artisanal Cotija cheese assessed by PCR-DGGE. *LWT – Food Sci. Technol.* 65, 1153–1161. <http://dx.doi.org/10.1016/j.lwt.2015.09.044>.
- Collins, Y.F., McSweeney, P.L.H., Wilkinson, M.G., 2003. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. *Int. Dairy J.* 13, 841–866.

- [http://dx.doi.org/10.1016/S0958-6946\(03\)00109-2](http://dx.doi.org/10.1016/S0958-6946(03)00109-2).
- Cotter, P.D., Hill, C., Ross, R.P., 2005. Bacteriocins: developing innate immunity for food. *Nat. Rev. Microbiol.* 3, 777–788. <http://dx.doi.org/10.1038/nrmicro1273>.
- Czerny, M., Brueckner, R., Kirchoff, E., Schmitt, R., Buettner, A., 2011. The influence of molecular structure on odor qualities and odor detection thresholds of volatile alkylated phenols. *Chem. Senses* 36 (6), 539–553. <http://dx.doi.org/10.1093/chemse/bjr009>.
- Delcenserie, V., Taminiau, B., Delhalle, L., Nezer, C., Doyen, P., Crevecoeur, S., Roussey, D., Korsak, N., Daube, G., 2014. Microbiota characterization of a Belgian protected designation of origin cheese, Herve cheese, using metagenomic analysis. *J. Dairy Sci.* 97, 6046–6056. <http://dx.doi.org/10.3168/jds.2014-8225>.
- Draper, L.A., Grainger, K., Deegan, L.H., Cotter, P.D., Hill, C., Ross, R.P., 2009. Cross-immunity and immune mimicry as mechanisms of resistance to the antibiotoxic lacticin 3147. *Mol. Microbiol.* 71, 1043–1054. <http://dx.doi.org/10.1111/j.1365-2958.2008.06590.x>.
- Dugat-Bony, E., Straub, C., Teissandier, A., Onésime, D., Loux, V., Monnet, C., Irlinger, F., Landau, S., Leclercq-Perlat, M.-N., Bento, P., Fraud, S., Gibrat, J.-F., Aubert, J., Fer, F., Guédon, E., Pons, N., Kennedy, S., Beckerich, J.-M., Swennen, D., Bonnarme, P., 2015. Overview of a surface-ripened cheese community functioning by meta-omics analyses. *PLoS One* 10, e0124360. <http://dx.doi.org/10.1371/journal.pone.0124360>.
- Duthoit, F., Godon, J.-J., Montel, M.-C., 2003. Bacterial community dynamics during production of registered designation of origin salers cheese as evaluated by 16S rRNA gene single-strand conformation polymorphism analysis. *Appl. Environ. Microbiol.* 69, 3840–3848. <http://dx.doi.org/10.1128/AEM.69.7.3840-3848.2003>.
- Eddy, S.R., 2011. Accelerated profile HMM searches. *PLoS Comput. Biol.* 7, e1002195. <http://dx.doi.org/10.1371/journal.pcbi.1002195>.
- Ercolini, D., De Filippis, F., La Storia, A., Iacono, M., 2012. “Remake” by high-throughput sequencing of the microbiota involved in the production of water Buffalo Mozzarella cheese. *Appl. Environ. Microbiol.* 78, 8142–8145. <http://dx.doi.org/10.1128/AEM.02218-12>.
- Escobar, G., 2012. Estudio de la distribución espacial de la microbiota bacteriana en queso Cotija por la técnica de FISH (Thesis). Universidad Nacional Autónoma de México, Facultad de Química, México. Available at: [www.dgbiblio.unam.mx](http://www.dgbiblio.unam.mx).
- Fimland, G., Eijsink, V.G.H., Nissen-Meyer, J., 2002. Comparative studies of immunity proteins of pediocin-like bacteriocins. *Microbiology* 148, 3661–3670.
- Flores-Magallón, R., Oliva-Hernández, A.A., Narváez-Zapata, J.A., 2011. Characterization of microbial traits involved with the elaboration of the Cotija cheese. *Food. Sci. Biotechnol.* 20, 997–1003. <http://dx.doi.org/10.1007/s10068-011-0137-z>.
- Fox, P.F., Singh, T.K., McSweeney, P.L.H., 1995. Biogenesis of flavour compounds in cheese. In: Malin, E.L., Tunick, M.H. (Eds.), *Chemistry of Structure-function Relationships in Cheese, Advances in Experimental Medicine and Biology*. Springer, US, pp. 59–98.
- Fuka, M.M., Wallisch, S., Engel, M., Welzl, G., Havranek, J., Schloter, M., 2013. Dynamics of bacterial communities during the ripening process of different croatian cheese types derived from raw Ewe's milk cheeses. *PLoS One* 8 (11). <http://dx.doi.org/10.1371/journal.pone.0080734>.
- García, V., 2011. Detección de microorganismos lipolíticos durante las etapas de elaboración de un queso artesanal, mediante técnicas moleculares (Thesis). Universidad Nacional Autónoma de México, Facultad de Química, México. Available at: [www.dgbiblio.unam.mx](http://www.dgbiblio.unam.mx).
- GhostKOALA, 2015. URL <http://www.kegg.jp/ghostkoala/> (accessed 12.11.15).
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. <http://dx.doi.org/10.1038/nbt.1883>.
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., Goodman, R.M., 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5, R245–R249. [http://dx.doi.org/10.1016/S1074-5521\(98\)90108-9](http://dx.doi.org/10.1016/S1074-5521(98)90108-9).
- Ishikawa, M., Kodama, K., Yasuda, H., Okamoto-Kainuma, A., Koizumi, K., Yamasato, K., 2007. Presence of halophilic and alkaliphilic lactic acid bacteria in various cheeses. *Lett. Appl. Microbiol.* 44, 308–313. <http://dx.doi.org/10.1111/j.1472-765X.2006.02073.x>.
- Ishikawa, M., Yamasato, K., Kodama, K., Yasuda, H., Matsuyama, M., Okamoto-Kainuma, A., Koizumi, Y., 2013. *Alkalibacterium gilvum* sp. nov., slightly halophilic and alkaliphilic lactic acid bacterium isolated from soft and semi-hard cheeses. *Int. J. Syst. Evol. Microbiol.* 63, 1471–1478. <http://dx.doi.org/10.1099/ijs.0.042556-0>.
- Jung, J.Y., Lee, S.H., Kim, J.M., Park, M.S., Bae, J.-W., Hahn, Y., Madsen, E.L., Jeon, C.O., 2011. Metagenomic analysis of kimchi, a traditional Korean fermented food. *Appl. Environ. Microbiol.* 77, 2264–2274. <http://dx.doi.org/10.1128/AEM.02157-10>.
- Kjos, M., Borrero, J., Opsata, M., Birri, D.J., Holo, H., Cintas, L.M., Snipen, L., Hernández, P.E., Nes, I.F., Diep, D.B., 2011. Target recognition, resistance, immunity and genome mining of class II bacteriocins from gram-positive bacteria. *Microbiology* 157, 3256–3267. <http://dx.doi.org/10.1099/mic.0.052571-0>.
- Kultima, J.R., Sunagawa, S., Li, J., Chen, W., Chen, H., Mende, D.R., Arumugam, M., Pan, Q., Liu, B., Qin, J., Wang, J., Bork, P., 2012. MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS One* 7, e47656. <http://dx.doi.org/10.1371/journal.pone.0047656>.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <http://dx.doi.org/10.1093/bioinformatics/btp324>.
- Litopoulou-Tzanetaki, E., Tzanetakis, N., 2014. The microfloras of traditional greek cheeses. In: Donnelly, C. (Ed.), *Cheese and Microbes*. ASM Press, Washington, DC, pp. 177–218.
- Martínez, B., Suárez, J.E., Rodríguez, A., 1996. Lactococcin 972: a homodimeric lactococcal bacteriocin whose primary target is not the plasma membrane. *Microbiology* 142 (Pt 9), 2393–2398. <http://dx.doi.org/10.1099/00221287-142-9-2393>.
- Masoud, W., Vogensen, F.K., Lillevang, S., Abu Al-Soud, W., Sørensen, S.J., Jakobsen, M., 2012. The fate of indigenous microbiota, starter cultures, *Escherichia coli*, *Listeria innocua* and *Staphylococcus aureus* in Danish raw milk and cheeses determined by pyrosequencing and quantitative real time (qRT)-PCR. *Int. J. Food Microbiol.* 153, 192–202. <http://dx.doi.org/10.1016/j.jfoodmicro.2011.11.014>.
- McSweeney, P.L.H., 2004. Biochemistry of cheese ripening. *Int. J. Dairy Technol.* 57, 127–144. <http://dx.doi.org/10.1111/j.1471-0307.2004.00147.x>.
- Milesi, M., McSweeney, P., Hynes, E., 2008. Viability and contribution to proteolysis of an adjunct culture of *Lactobacillus plantarum* in two model cheese systems: cheddar cheese-type and soft-cheese type. *J. Appl. Microbiol.* 105, 884–892. <http://dx.doi.org/10.1111/j.1365-2672.2008.03813.x>.
- MOH Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, 2014. VFDB - Comparative Pathogenomic Composition of *Escherichia*. Virulence Factors of Pathogenic Bacteria. URL: <http://www.mgc.ac.cn/cgi-bin/VFs/compvifs.cgi?Genus=Escherichia> (accessed 12.11.15).
- Montel, M.-C., Buchin, S., Mallet, A., Delbes-Paus, C., Vuitton, D.A., Desmases, N., Berthier, F., 2014. Traditional cheeses: rich and diverse microbiota with associated benefits. *Int. J. Food Microbiol.* 177, 136–154. <http://dx.doi.org/10.1016/j.jfoodmicro.2014.02.019>.
- Morales, F., Morales, J.L., Hernández, C.H., Hernández-Sánchez, H., 2011. Isolation and partial characterization of halotolerant lactic acid bacteria from two Mexican cheeses. *Appl. Biochem. Biotechnol.* 164, 889–905. <http://dx.doi.org/10.1007/s12010-011-9182-6>.
- Moretø, T., Naterstad, K., Wang, E., Aasen, I.M., Chaillou, S., Zagorec, M., Axelsson, L., 2005. Sakacin P non-producing *Lactobacillus sakei* strains contain homologues of the sakacin P gene cluster. *Res. Microbiol.* 156, 949–960. <http://dx.doi.org/10.1016/j.resmic.2005.05.005>.
- Nalbantoglu, U., Cakar, A., Dogan, H., Abaci, N., Ustek, D., Sayood, K., Can, H., 2014. Metagenomic analysis of the microbial community in kefir grains. *Food Microbiol.* 41, 42–51. <http://dx.doi.org/10.1016/j.fm.2014.01.014>.
- Nes, I.F., Brede, D.A., Holo, H., 2006. Chapter 17-The nonantibiotic heat-stable bacteriocins in gram-positive bacteria. In: Kastin, A.J. (Ed.), *Handbook of Biologically Active Peptides*. Academic Press, Burlington, pp. 107–114.
- NMX-F-735-COFOCALEC, 2011. Mexican Official Standard. Sistema Producto Leche-Alimentos-Lácteos-Alimento Lácteo Regional-Queso Cotija Artesanal Madurado-Denominación, Especificaciones y Métodos de Prueba.
- Ondov, B.D., Bergman, N.H., Phillippy, A.M., 2011. Interactive metagenomic visualization in a web browser. *BMC Bioinform.* 12, 385. <http://dx.doi.org/10.1186/1471-2105-12-385>.
- Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. <http://dx.doi.org/10.1093/bioinformatics/bts174>.
- Quigley, L., O'Sullivan, O., Beresford, T.P., Ross, R.P., Fitzgerald, G.F., Cotter, P.D., 2012. High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses. *Appl. Environ. Microbiol.* 78, 5717–5723. <http://dx.doi.org/10.1128/AEM.00918-12>.
- Ross, R.P., Morgan, S., Hill, C., 2002. Preservation and fermentation: past, present and future. *Int. J. Food Microbiol.* 79, 3–16.
- Rossi, F., Rizzotti, L., Felis, G.E., Torriani, S., 2014. Horizontal gene transfer among microorganisms in food: current knowledge and future perspectives. *Food Microbiol.* 42, 232–243. <http://dx.doi.org/10.1016/j.fm.2014.04.004>.
- Roth, E., Miescher-Schwenninger, S., Hasler, M., Eugster-Meier, E., Lacroix, C., 2010. Population dynamics of two antilisterial cheese surface consortia revealed by temporal temperature gradient gel electrophoresis. *BMC Microbiol.* 10, 74. <http://dx.doi.org/10.1186/1471-2180-10-74>.
- Settanni, L., Di Grigoli, A., Tornambé, G., Bellina, V., Francesca, N., Moschetti, G., Bonanno, A., 2012. Persistence of wild *Streptococcus thermophilus* strains on wooden vat and during the manufacture of a traditional Caciocavallo type cheese. *Int. J. Food Microbiol.* 155 (1–2), 73–81. <http://dx.doi.org/10.1016/j.jfoodmicro.2012.01.022>.
- Smit, G., Smit, B.A., Engels, W.J.M., 2005. Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.* 29, 591–610. <http://dx.doi.org/10.1016/j.femsre.2005.04.002>.
- Snyder, A.B., Worobo, R.W., 2014. Chemical and genetic characterization of bacteriocins: antimicrobial peptides for food safety. *J. Sci. Food Agric.* 94, 28–44. <http://dx.doi.org/10.1002/jsfa.6293>.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12115–12120. <http://dx.doi.org/10.1073/pnas.0605127103>.
- Su, X., Pan, W., Song, B., Xu, J., Ning, K., 2014. Parallel-META 2.0: enhanced metagenomic data analysis with functional annotation, high performance computing and advanced visualization. *PLoS One* 9, e89323. <http://dx.doi.org/10.1371/journal.pone.0089323>.



- Urbach, G., 1997. The flavour of milk and dairy products: II. Cheese: contribution of volatile compounds. *Int. J. Dairy Technol.* 50, 79–89. <http://dx.doi.org/10.1111/j.1471-0307.1997.tb01743.x>.
- van Heel, A.J., de Jong, A., Montalbán-López, M., Kok, J., Kuipers, O.P., 2013. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res.* 41, W448–W453. <http://dx.doi.org/10.1093/nar/gkt391>.
- Wang, W., Zhang, L., Li, Y., 2012. Production of volatile compounds in reconstituted milk reduced-fat cheese and the physicochemical properties as affected by exopolysaccharide-producing strain. *Molecules* 17, 14393–14408. <http://dx.doi.org/10.3390/molecules171214393>.
- Wolfe, B.E., Button, J.E., Santarelli, M., Dutton, R.J., 2014. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. *Cell* 158, 422–433. <http://dx.doi.org/10.1016/j.cell.2014.05.041>.
- Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z., Forney, L.J., 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* 7, e33865. <http://dx.doi.org/10.1371/journal.pone.0033865>.
- Yvon, M., Rijnen, L., 2001. Cheese flavour formation by amino acid catabolism. *Int. Dairy J.* 11, 185–201. [http://dx.doi.org/10.1016/S0958-6946\(01\)00049-8](http://dx.doi.org/10.1016/S0958-6946(01)00049-8).
- Zhu, W., Lomsadze, A., Borodovsky, M., 2010. *Ab initio* gene identification in metagenomic sequences. *Nucl. Acids Res.* 38 <http://dx.doi.org/10.1093/nar/gkq275> e132–e132.
- Zúñiga, A., 2009. Descripción e identificación de la comunidad bacteriana presente en el Cotija por métodos moleculares (Thesis). Universidad Nacional Autónoma de México, Facultad de Química, México. Available at: [www.dgbiblio.unam.mx](http://www.dgbiblio.unam.mx).

## GLOSARIO

**Actividad acuosa ( $a_w$ ):** Término implementado para describir la intensidad con que el agua se asocia a los componentes no acuosos de un alimento. Se define matemáticamente como la presión de vapor relativa ( $a_w \approx p/p_0$ ; donde “p” es la presión de vapor del agua en solución y “ $p_0$ ” es la presión de vapor del agua pura), y puede tomar el valor máximo de uno para el agua pura.

**Bacterias coliformes:** Las bacterias coliformes son bacilos Gram negativos no esporulados, fermentadores de lactosa con producción de gas y ácido cuando se incuban a 37°C. Se usan como indicadores de calidad microbiológica, específicamente de contaminación fecal debido a su origen como comensales intestinales de animales de sangre caliente, aunque se les encuentra frecuentemente en ambientes acuáticos, suelo y en la superficie de vegetales. Los géneros *Citrobacter*, *Enterobacter*, *Hafnia*, *Klebsiella* y *Escherichia* son considerados coliformes totales, mientras que *E. coli* es considerado como coliforme fecal y se le reconoce por su habilidad para fermentar lactosa a 44°C. En la NMX-F-735-COFOCALEC-2011, se especifica para queso Cotija Artesanal Madurado la cuenta de coliformes fecales <10 NMP(Número Más Probable)/g.

**Contig:** Es una secuencia genómica continua de ADN obtenida a partir del ensamble de secuencias más cortas, donde la posición de cada una de las bases se considera como de alta confiabilidad. Dentro de dicha secuencia no debe haber huecos (*gaps*) o bases ambíguas diferentes de A, C, G o T.

**Denominación de Origen:** La Denominación de Origen es una figura jurídica de protección en materia de Propiedad Industrial. Con ésta, se protege productos que cuentan con características propias de una región geográfica particular, como son los factores naturales y humanos que determinan su calidad o características peculiares. En México, las Denominaciones de Origen son propiedad del Estado.

**Diversidad de especies:** La diversidad de una comunidad depende de la riqueza (número de especies) y de la abundancia relativa de éstas (equitatividad o *evenness*). En un par de muestras con el mismo número de especies, aquella en la que la abundancia relativa de las especies sea más equitativa, se considerará como más diversa.

**Estadística N50:** Permite obtener una idea de la distribución del tamaño de los contigs en un ensamble, se obtiene ordenando los contigs de mayor a menor tamaño y sumando sus longitudes por ese orden hasta alcanzar el 50% del tamaño del genoma o del total de bases ensambladas en un metagenoma. La longitud del último contig sumado es el valor N50.

**Estimador de diversidad:** Determinación estadística de la diversidad a partir de la abundancia relativa de las especies presentes en una muestra del tamaño necesario para tener completamente representada a la comunidad. Existen varios métodos de estimar diversidad, los métodos paramétricos estiman el número de especies no observadas al ajustar los datos del muestreo a modelos de abundancia relativa de especies, ejemplos de estos modelos son lognormal y Poisson normal; los estimadores no paramétricos ponderan de diferente manera a las

especies abundantes en relación con las especies de baja abundancia, y se consideran más adecuados para su aplicación en ecología microbiana; ejemplos de estos estimadores son los de Simpson y Shannon.

**Homología:** En el contexto biológico, homología es la existencia de un ancestro común que se comparte para una estructura o gen en especies diferentes.

**Indeles:** Inserciones o deleciones de bases en una secuencia genómica.

**Inocuidad:** Que no hace daño. En el contexto alimentario, la inocuidad es un concepto que se refiere a las condiciones y prácticas que preservan la calidad de los alimentos para prevenir su contaminación y las enfermedades transmitidas por su consumo.

**k-mero:** Secuencia con longitud “k” de bases, donde “k” es cualquier entero positivo de bases consecutivas.

**Lantibiótico:** Bacteriocinas clase I, son péptidos con actividad antimicrobiana de menos de 28 aminoácidos (<5 kDa) modificados post-traduccionalmente, lineales o globulares que contienen lantionina, β-metil lantionina, y/o aminoácidos deshidratados. Dichas modificaciones resultan de la condensación de residuos de serina o treonina. Un ejemplo importante de este tipo de bacteriocinas es la nisina.

**Lecturas:** Secuencias de ADN que se obtienen como resultado de la secuenciación de un genoma o metagenoma *shotgun* o de amplicones.

**Lecturas pareadas:** Protocolo de secuenciación que implica la obtención de lecturas desde ambos extremos de cada fragmento, lo que provee información sobre localización relativa y orientación de cada par de lecturas.

**Marca colectiva:** El registro de una Marca Colectiva en México, tiene como fin distinguir en el mercado, los productos o servicios de los miembros de la asociación que lo solicita de productos o servicios de terceros. En la solicitud se deben incluir las reglas de uso del producto a registrar. Su uso se aplica en la protección de un producto con calidad vinculada a su origen. Las Marcas Colectivas son propiedad del grupo de productores que las registran.

**Marcador filogenético:** Es un fragmento (o *locus*) de una secuencia ya sea codificante o no codificante de ADN que se utiliza en la reconstrucción filogenética mediante la comparación de los homólogos de secuencias de nucleótidos o de proteínas. Para que una secuencia pueda fungir como marcador filogenético no debe ser variable entre especies o debe tener variación predecible e idealmente, las secuencias deben estar disponibles para la mayoría o todas las especies de un género. El grado de similitud entre las secuencias, indica divergencia genética como resultado de la evolución molecular en el tiempo.

**Potencial de hidrógeno (pH):** Se entiende como potencial de hidrógeno la concentración molar de iones hidrógeno o protones (H<sup>+</sup>) en una solución. Matemáticamente se expresa como  $\text{pH} = -\log_{10}[\text{H}^+]$ .

**Potencial metabólico:** Conjunto de predicciones de función obtenidas a partir de secuencias metagenómicas que no se asocian necesariamente a un género o especie en particular, sino al consorcio como un sistema completo.

**Potencial redox (Eh):** El potencial redox es la medida de la afinidad de una sustancia por electrones comparada con la del hidrógeno, cuyo valor de potencial (E) es de cero. El potencial redox de un sistema está determinado por la cantidad total de especies oxidantes y reductoras ([Ox]+[Red]) y depende en gran medida de la cantidad de oxígeno disuelto y del pH del sistema, como se describe en la ecuación:  $Eh(mV) = 1.234 - (0.058 * pH) + (0.0145) * (\log_{10} pO_2)$ ; donde  $pO_2$  es la presión parcial de oxígeno expresada en atmósferas. Dicho potencial es positivo cuando ocurre oxidación y negativo cuando se produce una reducción. Debido a la fermentación de lactosa a ácido láctico, así como a la reducción de pequeñas cantidades de  $O_2$  a  $H_2O$ , el valor típico de Eh en quesos madurados es de aproximadamente -240 a -250 mV.

**Quorum sensing:** Es la regulación de la expresión de genes como respuesta a fluctuaciones en la densidad poblacional a nivel celular. En bacterias, la comunicación celular ocurre mediante la liberación de moléculas llamadas autoinductores, cuya concentración incrementa en el medio con la densidad de células. Algunos ejemplos de procesos regulados por *quorum sensing* en bacterias son virulencia, conjugación, producción de antibióticos, esporulación y la formación de biopelículas.

**Regulón:** Grupo de genes que están bajo el control del mismo gen regulador que codifica para una proteína que actúa ya sea como activador o como represor de su transcripción. Este sistema es la unidad básica de respuesta celular en bacterias.

**Secuencias adaptadoras:** Secuencias cortas de ADN de doble cadena que se adicionan a los extremos de los fragmentos cuya secuencia se desea conocer. Dichos oligos son necesarios para llevar a cabo la secuenciación y su patrón de nucleótidos varía de acuerdo a la tecnología utilizada con este fin. Los adaptadores indexados contienen además información sobre el origen de las muestras cuando se lleva a cabo secuenciación de varias muestras en una misma corrida.

**Transferencia horizontal de genes:** Transmisión no genealógica de material genético de un organismo donador a otro receptor que pueden estar o no relacionados taxonómicamente. Dicho mecanismo puede ocurrir en bacterias por tres vías distintas, conjugación, transformación o transducción.