

Universidad del País Vasco / Euskal Herriko Unibertsitatea

Grado en Química

TRABAJO FIN DE GRADO

Anotación y análisis de genomas: bacterias ácido lácticas como fuente de compuestos de valor añadido

Autora: Annette Marrero Fernández

Dirigido por: Oier Etxebeste Juárez
María Teresa Dueñas Chasco

Donostia-San Sebastián, 28 de junio de 2021

ABREVIATURAS

| | |
|------------|---|
| BAL | bacterias ácido-lácticas |
| BLAST | Basic Local Alignment Search Tool |
| CDS | Secuencia codificante o <i>Coding Sequence</i> |
| CD-Search | Search for Conserved Domains |
| DNA | Ácido desoxirribonucleico o Deoxyribonucleic acid |
| dNDP | Difosfato de desoxirribonucleósido |
| dNTP | Trifosfato de desoxirribonucleósido |
| dsr | dextran sacarasa |
| <i>dsr</i> | genes que codificarían proteínas con actividad dextran sacarasa |
| EPS | exopolisacárido |
| <i>eps</i> | genes que codificarían proteínas implicadas en las síntesis de EPS |
| GH | glicosil hidrolasas |
| GS | glucan sacarasas |
| GT | glicosil transferasas |
| HePS | heteropolisacárido |
| HoPS | homopolisacárido |
| Kb | kilo base (1.000 nucleótidos o pares de bases) |
| NCBI | National Center for Biotechnology Information |
| NGS | Secuenciación de nueva generación o <i>Next-generation sequencing</i> |
| pb | pares de base |
| PCR | Reacción en Cadena de la Polimerasa o Polymerase Chain Reaction |
| RAST | Rapid Annotation using Subsystem Technology |

ÍNDICE:

| | |
|---|----|
| RESUMEN | 7 |
| ABSTRACT | 9 |
| 1. Introducción | 11 |
| 1.1. Bacterias ácido lácticas | 11 |
| 1.2. Estructura de la superficie celular de BAL..... | 11 |
| 1.3. Clasificación de los EPS producidos por las BAL..... | 13 |
| 1.3.1. Homopolisacáridos | 13 |
| 1.3.1.1. Síntesis de α -glucanos y β -glucanos..... | 13 |
| 1.3.2. Heteropolisacáridos | 14 |
| 1.3.2.1. Síntesis de HePS | 14 |
| 1.4. Aplicaciones de EPS como postbióticos. | 15 |
| 2. Objetivos | 17 |
| 3. Materiales y métodos | 19 |
| 3.1. Cepas bacterianas..... | 19 |
| 3.2. Ensamblaje, anotación y análisis de las secuencias de los genomas bacterianos..... | 19 |
| 3.3. Detección de los genes que sintetizarían proteínas con actividad dsr | 21 |
| 4. Resultados | 25 |
| 4.1. Tamaños de los <i>contigs</i> obtenidos y calidad del ensamblaje..... | 25 |
| 4.2. Diferencias entre las anotaciones realizadas con Prokka y RAST, y su comparación con la proporcionada por la empresa MicrobesNG..... | 26 |
| 4.3. Análisis de genes predictivamente relacionados con la síntesis EPS en el genoma de las cepas BAL de interés | 28 |
| 4.4. Identificación de genes que codificarían proteínas con actividad dextransacarasa | 38 |
| 4.5. PCR para la amplificación de los genes <i>dsr</i> en la cepa Ast1 | 40 |
| 5. Discusión | 41 |
| 6. Conclusiones & Conclusions | 45 |
| BIBLIOGRAFÍA | 49 |
| ANEXO | 57 |

RESUMEN

El presente trabajo tiene como finalidad realizar el ensamblaje a partir de la secuenciación de los genomas de cuatro cepas de bacterias ácido lácticas (BAL), para posteriormente realizar la anotación (predicción) de sus genes y analizar la presencia de genes y agrupaciones de genes que codifiquen funciones de interés. Dichas cepas están siendo estudiadas en el laboratorio de Biología de la Facultad de Químicas de la UPV/EHU debido a su posible uso como fuente de compuestos de valor añadido, como aditivos alimentarios.

Las tareas de ensamblaje, anotación y análisis serán realizadas con técnicas básicas de análisis bioinformático masivo, un campo en constante expansión debido a la disponibilidad de un número cada vez mayor de genomas secuenciados. Dichas técnicas están ligadas al desarrollo de ordenadores capaces de procesar cada vez mayores volúmenes de información, por lo que hoy en día, un ordenador personal es suficiente para poder realizarlas de forma independiente y autónoma.

Algunas BAL sintetizan polisacáridos que son secretados al medio de cultivo (exopolisacáridos, EPS), mejorando las características sensoriales y tecnológicas de varios productos fermentados, fundamentalmente productos lácteos y los basados en cereales. Dado el interés biotecnológico de los EPS producidos por las BAL, este trabajo se centra en la búsqueda de agrupaciones de genes que pudieran intervenir en la síntesis de heteropolisacáridos (HePS, polisacáridos formados por más de un tipo de monosacárido) y en genes que sintetizarían proteínas con actividad dextransacarasa, que es fundamental en la síntesis de Homopolisacáridos (HoPS, polisacáridos formados por un tipo de monosacárido) de tipo dextrano.

ABSTRACT

The aim of the following work is to carry out the assembly of the genomes of four lactic acid bacteria (LAB), in order to annotate their genes and analyze the presence of genes and gene clusters encoding specific functions. These strains are being studied at the laboratory of Biology of the Faculty of Chemistry at the University of the Basque Country due to their potential as a source of value-added compounds such as food additives.

The assembly, annotation and analysis of the genomes will be carried out using basic bioinformatics techniques, a field in constant expansion due to the availability of an increasing number of sequenced genomes. These bioinformatics techniques are linked to the development of computers capable of processing even greater volumes of information. Thus, at the moment, a single personal computer can perform these analyses independently and autonomously, and in a reasonable amount of time.

Some LAB strains synthesize polysaccharides that are secreted to the culture medium (exopolysaccharides, EPS), improving the sensory and technological characteristics of various fermented products, mainly dairy products and those based on cereals. Given the biotechnological interest of the EPS produced by LAB, this work is focused on the search for gene *clusters* potentially involved in the synthesis of heteropolysaccharides (HePS; exopolysaccharides formed by at least two types of monosaccharides) and in genes that would encode proteins with dextransucrase activity, which is fundamental for the synthesis of dextran-type homopolysaccharides (HoPS, exopolysaccharides formed by only one type of monosaccharide).

1. Introducción

1.1. Bacterias ácido lácticas

Las bacterias ácido lácticas (BAL) se conocen como una familia heterogénea de microbios que son capaces de convertir, por fermentación, sustratos de carbohidratos en ácidos orgánicos, principalmente en ácido láctico. Se caracterizan por ser bacterias Gram-positivas, con forma de cocos y bacilos ^(1,2).

En general son microorganismos que están muy extendidos y pueden residir en diversos tipos de hábitats, desde cavidades humanas, hasta en alimentos como productos lácteos, cárnicos, vegetales procesados y bebidas alcohólicas tales como sidra, vino y cerveza. En consecuencia, están ganando atención hacia nuevos usos jugando un papel importante en muchos procesos de fermentación industrial y nutrición humana, pues su presencia en el tracto gastrointestinal confiere beneficios a la salud humana debido a las propiedades como probióticos de algunas BAL ^(3,4).

Hoy en día, con el aumento de la accesibilidad de la secuenciación de nueva generación (NGS), se dispone de una amplia biblioteca de genomas de cepas BAL secuenciados, lo que facilita una mejor comprensión de las características de estas especies, tales como sus procesos biológicos y su funcionalidad. Además, la disponibilidad de dichos genomas permite realizar análisis de genómica comparativa ^(1,5).

1.2. Estructura de la superficie celular de BAL

Las propiedades de la superficie celular de un microorganismo determinan en gran medida sus interacciones con el medio ambiente, entre las que se encuentra el intercambio de nutrientes y productos de desecho ⁶.

La membrana plasmática de las bacterias Gram-positivas está cubierta por una pared celular que está formada por una capa de peptidoglicano, sobre la cual se anclan, ácidos teicoicos (polímeros de un polialcohol unidos mediante enlaces fosfodiéster) y lipoteicoicos (unidades compuestas de ácido teicoicos y lípidos), que son fundamentales para mantener la forma e integridad de la célula. En muchas especies de bacterias la pared celular puede estar cubierta por una capa superficial adicional proteica, la capa S. La superficie bacteriana

puede estar también rodeada de otra capa más externa constituida por polisacáridos exocelulares o exopolisacáridos (EPS) (Figura 1) ^(7, 8).

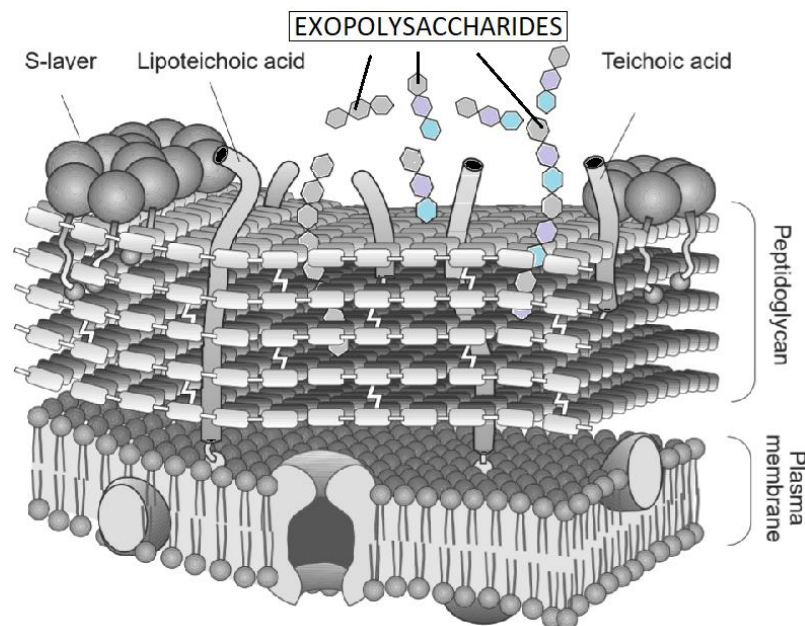


Figura 1. Esquema de la envoltura celular de las bacterias ácido lácticas. Figura modificada a partir de Delcour et al., 1999 ⁸. Capa de peptidoglicanos decorada con ácidos lipoteicoicos y teicoicos rodeada por la Capa S, y polisacáridos exocelulares o exopolisacáridos (EPS).

Los EPS microbianos que envuelven a la célula y están unidos covalentemente al ácido N-acetilmurámico del peptidoglicano son denominados polisacáridos capsulares (CPS). Otros polisacáridos extracelulares se mantienen unidos de forma débil a la superficie celular y/o son liberados al medio extracelular formando una matriz de EPS⁹.

La síntesis de estos polisacáridos en las BAL está mediada por tres mecanismos que pueden coincidir o no para una misma bacteria. La Vía Wzy-dependiente implica la participación de diferentes enzimas que son codificadas por un conjunto de genes que están agrupados en el genoma en forma de agrupaciones o *clusters* de genes. En segundo lugar, mediante una proteína sintasa que lleva a cabo el proceso de polimerización y transporte, utilizando como sustrato azúcares activados en el interior celular. Y, por último, mediante una enzima glicosilhidrolasa (GH), que lleva a cabo la síntesis extracelular utilizando principalmente sacarosa como sustrato ¹⁰.

1.3. Clasificación de los EPS producidos por las BAL

Los EPS se clasifican en dos grupos según su composición:

Homopolisacáridos (HoPS): polisacáridos compuestos a partir de un único tipo de monosacárido, que se enlazan entre sí por medio de uno o varios tipos de enlace.

Heteropolisacáridos (HePS): polisacáridos compuestos por unidades repetitivas que están formadas por más de un tipo de monosacárido, además de poder contener otros elementos de naturaleza diferente.

1.3.1. Homopolisacáridos

Los HoPS pueden ser glucanos, que son sintetizados por glucansacararas o glicosiltransferasas; o fructanos, por fructansacararas. Los glucanos, dependiendo del tipo de enlace pueden ser clasificados en α -glucanos y β -glucanos. Los dextranos se sintetizan a partir de sacarosa, y son α -glucanos, HoPS formados por moléculas de glucosa unidas mediante enlaces glicosídicos α -1,6 en una cadena principal, y con ramificaciones de α -glucopiranososa en las posiciones O-2, O-3 y O-4¹¹.

1.3.1.1. Síntesis de α -glucanos y β -glucanos

La síntesis de α -glucanos es llevada a cabo por glucansacararas (GS), que a menudo también se nombran según el producto que sintetizan, por ejemplo, dextransacararas (dsr), si producen dextranos. Las GS se clasifican como glicosil hidrolasas (GH) que pertenecen a la familia 70 por presentar grandes similitudes entre sus secuencias de aminoácidos. Estas enzimas extracelulares catalizan la escisión del enlace glicosídico entre los monosacáridos glucosa y fructosa, que componen el disacárido sacarosa, y con la energía liberada pueden transferir un resto de glucosa al polímero en formación. Hasta el momento se ha descrito la producción de α -glucanos en los siguientes géneros de BAL: *Leuconostoc*, *Lactobacillus*, *Streptococcus*, *Pediococcus*, *Weisella* y *Oenococcus* ¹².

La producción de β -glucanos se realiza a partir de la glucosa por medio de una glicosiltransferasa (GT), que es una proteína transmembrana y a diferencia de otras sintasas productoras de glucano, no tiene función GH ¹³.

1.3.2. Heteropolisacáridos

Los HePS producidos por las BAL muestran gran diversidad en su composición y estructura. La producción y secreción de estos requieren de la acción sincronizada de varias enzimas y proteínas codificadas en *clusters* de genes altamente conservados y en los que los genes correspondientes presentan normalmente la misma orientación ¹⁴.

1.3.2.1. Síntesis de HePS

La síntesis de HePS está mediada por la vía Wzy-dependiente, la cual se divide en diferentes pasos: la formación de precursores de azúcares activados, un ensamblaje asociado a la membrana celular y por último la polimerización y secreción del polisacárido. Estas tareas son realizadas por proteínas y enzimas codificadas por los genes que constituyen *clusters eps* (Figura 2) ¹⁰.

Las proteínas iniciadoras son GT de membrana (EpsE, codificadas por el gen *welE*) que se encargan de añadir el primer azúcar activado a un lípido de membrana. Para formar la unidad repetitiva del polisacárido, se coordinan diferentes enzimas GT (EpsF, -G, -H, etc., codificadas por genes *welF*, -G, -H, respectivamente), y serán tantas como cantidad de monosacáridos presentes haya en la unidad repetitiva del polisacárido. Una vez formada la unidad repetitiva, esta es transportada a través de la membrana plasmática por una proteína flipasa (Wzx) y, la polimerasa (Wzy) cataliza la unión de cada unidad estructural al polisacárido exocelular, quedando ya separado del lípido de membrana ^(10,15).

Existen diferentes genes moduladores a lo largo del proceso (*wzd*, *wzb*, *wze* y *wzr*) que codifican las proteínas (EpsC, -B, -D y -A) respectivamente, las cuales se encargan de determinar la longitud de la cadena del polisacárido que se está formando y hacer que la síntesis y exportación del polisacárido se haga satisfactoriamente. Estas proteínas reguladoras suelen ser proteínas tirosina kinasas fosforiladas y no fosforiladas. Wzr, por ejemplo, permite que el polisacárido se una a los peptidoglicanos de la pared celular.

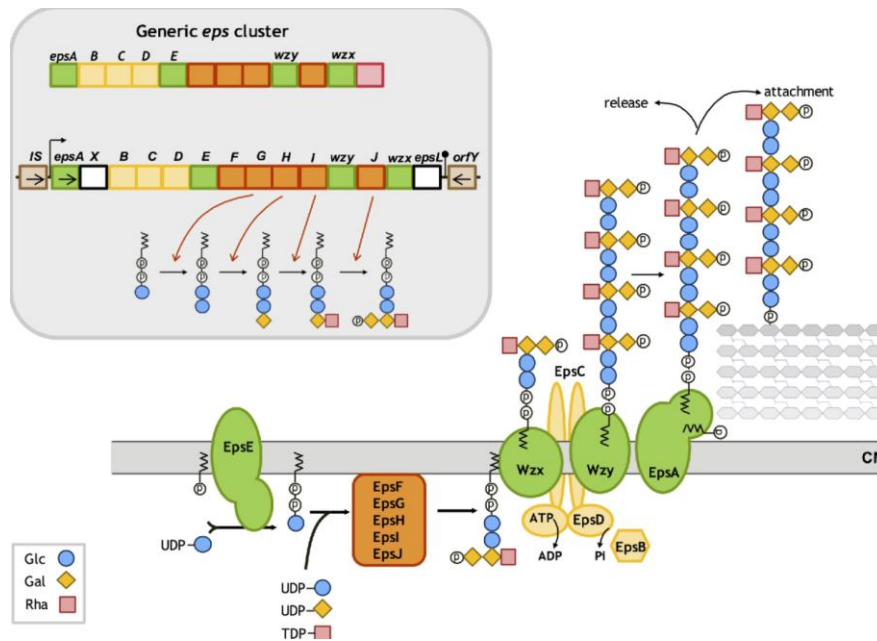


Figura 2. Esquema propuesto para la biosíntesis de HePS a través de la vía Wzy-dependiente. Figura tomada de Zeidan et al., 2017¹⁰. Se muestran los genes implicados en la síntesis y exportación de HePS en BAL. En la parte superior izquierda está representado un *cluster eps* genérico y debajo un *cluster eps* específico para la cepa *L. lactis* ssp. *cremoris* NIZO B40. Cada cuadro representa un gen, en verde genes polimerasa y flipasa; en amarillo genes moduladores; en naranja genes GT y en rojo, genes relacionados con la formación de azúcares precursores. En la parte inferior derecha queda representado el mecanismo de acción de las proteínas codificadas por los genes de la vía de Wzy-dependiente.

La formación de precursores es necesaria para comenzar con todo el proceso de síntesis. En las BAL, se estima que la mayoría de los azúcares y sus derivados detectados hasta el momento en estructuras de unidades repetitivas, tales como la galactosa, glucosa, N-acetilglucosamina, manosa y ribosa, son resultado de las rutas del metabolismo central. Sin embargo, los genes necesarios para sintetizar dTDP-ramnosa (*rmlA*, *rmlB*, *rmlC* y *rmlD*) pueden estar asociados a los *clusters eps*¹⁰.

1.4. Aplicaciones de EPS como postbióticos.

Algunas BAL presentan características beneficiosas para la salud debido a sus propiedades probióticas o capacidades metabólicas como son la producción de enzimas, péptidos, ácidos teicoicos, proteínas de la superficie celular, ácidos orgánicos y polisacáridos. Estos productos metabólicos son conocidos como postbióticos, los cuales han despertado gran interés en la industria agroalimentaria para la mejora de la funcionalidad de los alimentos, y consecuentemente promover eficazmente la salud de los consumidores¹⁶.

EPS de cepas BAL específicas se están utilizando actualmente como agentes estabilizantes, emulsionantes y aglutinantes¹⁷. Los EPS de determinadas cepas del género *Lactobacillus* funcionan como agentes antitumorales *in vitro*, y otros poseen propiedades antimicrobianas y antioxidantes¹⁸. Por todo lo mencionado, el posible uso de estos polisacáridos en productos farmacéuticos y alimentos funcionales ha generado un interés creciente para su investigación.

2. Objetivos

Este trabajo tiene como objetivos:

- 1) Realizar el ensamblaje, anotación y análisis de los genomas de cuatro cepas BAL, a partir de los archivos de secuenciación correspondientes.
- 2) Identificar posibles agrupaciones (*clusters*) y otros genes potencialmente necesarios para la síntesis de EPS, en función de si previamente se han descrito las cepas correspondientes como productoras de HePS o dextrano.
- 3) Identificar y amplificar mediante PCR genes que potencialmente codifican *dsr* como catalizadoras de la síntesis de dextrano.

3. Materiales y métodos

3.1. Cepas bacterianas

Para la realización de este trabajo se ha dispuesto de los archivos de secuenciación (.fastq) de los genomas de cuatro cepas BAL, denominadas E11, AC1, Ast1 y A19, según la nomenclatura del grupo de acogida GlycoBAL. Dichos archivos de secuenciación han sido proporcionados por este grupo, tras ser generados por la empresa MicrobesNG (Birmingham, Reino Unido), utilizando para ello la plataforma de secuenciación Illumina y con una cobertura de 30x (la longitud de la suma de todas las lecturas de secuenciación supone aproximadamente 30 veces los genomas bacterianos). GlycoBAL cuenta con una amplia colección de cepas BAL aisladas de diferentes matrices alimentarias. Las cuatro cepas indicadas previamente están siendo estudiadas debido a su posible uso como fuente de compuestos de valor añadido como aditivos alimentarios.

3.2. Ensamblaje, anotación y análisis de las secuencias de los genomas bacterianos

En la secuenciación masiva a través de la plataforma Illumina se genera una gran cantidad de secuencias cortas o lecturas (*reads*), de no más de 400 pb (Figura 3). Muchos de estos *reads* solapan parcialmente, de modo que la mayoría de ellos (pero no todos) pueden ser ensamblados para así formar secuencias más largas, llamadas *contigs*, cuya suma representaría el genoma analizado. Posteriormente, el producto del ensamblaje se anota, es decir, se identifican los genes o secuencias probablemente codificantes de proteínas (CDS, de *coding sequences*) y se les asocia una función (Figura 3). En este último caso, una proporción de genes quedan anotados con funciones desconocidas, al no poder inferírseles por similitud función alguna. Finalmente, se han analizado las secuencias resultantes y se ha centrado la búsqueda en la identificación de posibles agrupaciones de genes *eps* y genes que puedan codificar la actividad *dsr*, como catalizadora en la síntesis de dextrano. Las tareas de ensamblaje, anotación y análisis han sido realizadas con plataformas de libre acceso disponibles en Internet (Figura 3).

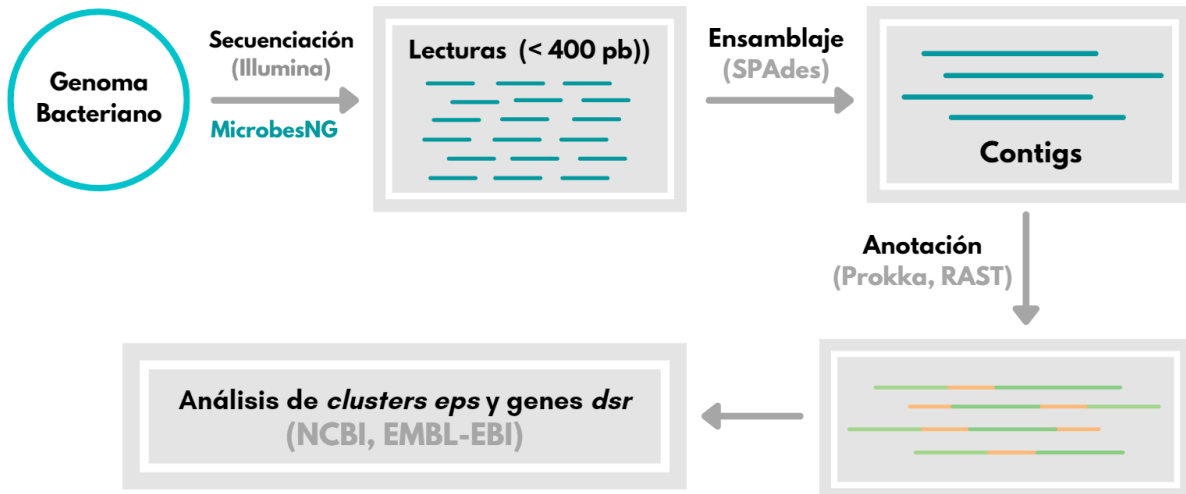


Figura 3. Flujo de trabajo seguido. Partiendo de la secuenciación de los genomas bacterianos, éstos se ensamblan y anotan con diferentes herramientas para analizar las CDS, y detectar así, posibles agrupaciones de genes *eps* y, la posible actividad *dsr*, como catalizadora en la síntesis de dextrano. El color verde tras la anotación con Prokka o RAST, indica las probables CDS anotadas.

Ensamblaje:

El ensamblaje de las secuencias se realizó utilizando el ensamblador SPAdes¹⁹, una herramienta de la plataforma Galaxy (<https://usegalaxy.org/>)²⁰ para analizar genomas pequeños. SPAdes parte de los archivos de secuenciación (*fastq*) de ambas hebras del DNA (*forward* y *reverse reads*). Se utilizaron los parámetros que el programa elige por defecto (ver Anexo I). A este tipo de secuenciación, donde se secuencian ambas hebras del DNA, se le denomina *paired-end*.

Anotación:

Para la anotación se utilizaron dos herramientas. La primera es Prokka²¹, otra herramienta accesible en la plataforma Galaxy y que realiza la anotación de genomas procariontes, en nuestro caso a partir del archivo FASTA generado por SPAdes en el ensamblaje (ver Anexo I.I). A modo de comparación, se empleó también una segunda plataforma, RAST (*Rapid Annotation using Subsystem Technology*)²², un servicio totalmente automatizado para anotar genomas bacterianos. De igual manera, en RAST es también necesario cargar el archivo FASTA generado por SPAdes, y los parámetros de estudio seleccionados fueron los que el programa establece por defecto (Anexo I. II).

Búsqueda y análisis de posibles *clusters eps* y actividad *dsr*:

Para este estudio se tuvo en cuenta la información obtenida en la anotación realizada con RAST, para así predecir las funciones de los genes. El portal *National Center for Biotechnology Information* (NCBI)²³, contiene diversas herramientas que permiten la búsqueda de dominios conservados dentro de una proteína. BLAST (Basic Local Alignment Search Tool) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), es una de sus herramientas de búsqueda de alineación, que encuentra, en base a la similitud, regiones conservadas entre secuencias. Más específicamente, BLASTp (BLAST Protein), compara las secuencias de las proteínas de interés con las millones de secuencias peptídicas de la base de datos y calcula la significación estadística de las coincidencias, proporcionando diferentes parámetros que se tendrán en cuenta a la hora de realizar el análisis. CD-Search (Search for Conserved Domains)²⁴, es otra de las funciones disponibles en el NCBI, la cual permite la búsqueda de dominios conservados dentro de una proteína (también se puede partir de una secuencia de DNA codificante).

The European Bioinformatics Institute, que forma parte del *European Molecular Biology Laboratory* (EMBL-EBI)²⁵, ha desarrollado también bases de datos y diversas herramientas para el análisis bioinformático. InterPro²⁶ proporciona un análisis funcional de proteínas, las clasifica por familias y predice las coordenadas y extensión de sus dominios, y sitios que probablemente sean funcionalmente importantes, además de ofrecer información acerca de los mismos. HMMER²⁷, por otro lado, busca homologías, de manera rápida, entre las secuencias de aminoácidos, entre la proteína de interés y las de la base de datos, en base a los algoritmos *Hidden Markov Models* ²⁸.

3.3. Detección de los genes que sintetizarían proteínas con actividad *dsr*

La detección de los genes *dsr*, se llevó a cabo mediante PCR, las cuales van dirigidas a ampliar la zona del genoma en la que se encuentren dichos genes *dsr*. Se diseñaron, para la amplificación, oligonucleótidos (o cebadores) a partir de las secuencias de nucleótidos de los posibles genes *dsr*. La amplificación de los genes *dsr* identificados, se realizó de entre las cuatro cepas BAL, solo en Ast1 a partir de las secuencias de nucleótidos de los genes *dsr* (ver Anexo II).

La selección de los parámetros para el diseño de los oligonucleótidos (se requieren dos cebadores para cada amplificación) se realizó con la ayuda de la plataforma *Oligo Evaluator* de la compañía Sigma-Aldrich, la cual permite hacer una evaluación de la probable eficiencia de oligonucleótidos (Tabla 1). Se requieren temperaturas de fusión que no fuesen menores de 65°C, con una diferencia de no más que 5°C entre las de las parejas de oligonucleótidos a utilizar en una misma reacción de PCR. En segundo lugar, la posibilidad de formación de estructuras secundarias (plegamientos internos en la secuencia del cebador) debía ser nula, muy débil o débil. Es importante también que la probabilidad de formación de dímeros entre los cebadores sea baja o nula, para evitar que se hibriden los cebadores entre sí y que no lo hagan correctamente con el DNA molde. Por último, se buscaron cebadores con concentraciones de GC (Guanosina y Citosina) superiores al 50%, ya que estas bases nitrogenadas forman tres puentes de hidrógeno entre ellas, por lo que las uniones del cebador al DNA molde se realizarán con mayor eficiencia si hay concentraciones mayores de estas (Tabla 1).

Las reacciones de PCR se llevaron a cabo utilizando dos réplicas, dos muestras diferentes (a y b), del DNA molde (DNA de la cepa Ast1). Los probables genes *dsr* identificados en la cepa Ast1 fueron *peg.1686* y *peg.1246*, por lo que los cebadores diseñados se denominaron *peg.1686-FP* (*forward primer*) y *peg.1686-RP* (*reverse primer*), por un lado, y *peg.1246-FP* y *peg.1246-SP*, por el otro (Tabla 2). A cada una de las cuatro mezclas de reacción se añadieron también una disolución tampón, para generar las condiciones adecuadas para la DNA polimerasa (enzima encargada de la síntesis de las cadenas), los dNTPs (los monómeros del DNA), la DNA polimerasa (la enzima encargada de la reacción de polimerización; *Phusion DNA polymerase*; Thermo Scientific) y agua hasta completar un volumen total de 50 µL (Tabla 2).

Tabla 1. Oligonucleótidos empleados para la amplificación de genes *dsr* en Ast1.

| Cepa BAL | RAST-ID (gen) | Secuencia | Peso molecular (g/mol) | Longitud (pb) | T. fusión (°C) | GC % | Estructura secundaria | Primer Dímeros |
|----------|---------------|-------------------------------------|------------------------|---------------|----------------|------|-----------------------|----------------|
| Ast1 | Peg.1686 | FP: 5'-GGTACGTGATTCAATGGGTGTTAGC-3' | 7768,1 | 25 | 68,0 | 48,0 | Ninguna | No |
| | | RP: 5'-GGTAAGAGCTTATGCTGAAGAGGGG-3' | 7851,2 | 25 | 67,7 | 52,0 | Débil | No |
| | Peg.1246 | FP: 5'-GGGTTCTGATCATTGCAAGG-3' | 6492,3 | 21 | 65,5 | 47,6 | Ninguna | No |
| | | RP: 5'-CCAAGTACATCCATGACCC-3' | 7571,0 | 25 | 71,0 | 52,0 | Débil | No |

Tabla 2. Volúmenes de los diferentes componentes que se añaden a las mezclas de reacción de PCR, y las condiciones de temperatura y tiempo fijadas en el termociclador a la hora de realizar la reacción de PCR para la cepa BAL Ast1. Las letras a y b indican la muestra de DNA molde utilizada, los cuales, en ambos casos estaban a una concentración de alrededor de 20 ng/μL. *peg.1686* y *peg.1246* corresponden a los dos genes *dsr*, según nomenclatura RAST. Las disoluciones de ambos cebadores estaban ajustadas a una concentración de 10 μM. Las disoluciones *stock* de dNTPs y del tampón (*buffer*) fueron las comerciales.

| Volúmenes (μL) | 1 a (peg.1686) | 3 a (peg.1246) | 2 b (peg.1686) | 4 b (peg.1246) |
|------------------|--------------------------|----------------|----------------|----------------|
| 1,5 | DNA molde | | | |
| 1,5 | Oligo FP | | | |
| 1,5 | Oligo RP | | | |
| 4 | dNTP | | | |
| 10 | Buffer PS | | | |
| 0,5 | DNA Polymerase (Phusion) | | | |
| 31 | d H ₂ O | | | |
| Total: 50 | | | | |

| T (°C) | Tiempo | |
|-------------------|--------|------|
| Desnaturalización | | |
| 98 | 2' | |
| Elongación | | |
| 98 | 10'' | x 25 |
| 58 | 10'' | |
| 72 | 3'30'' | |
| Terminación | | |
| 72 | 7' | |
| 12 | ∞ | |

4. Resultados

4.1. Tamaños de los *contigs* obtenidos y calidad del ensamblaje

Las lecturas de la secuenciación, proporcionadas por MicrobesNG (ver Materiales y Métodos), se ensamblaron haciendo uso de la herramienta SPAdes (Figura 4).

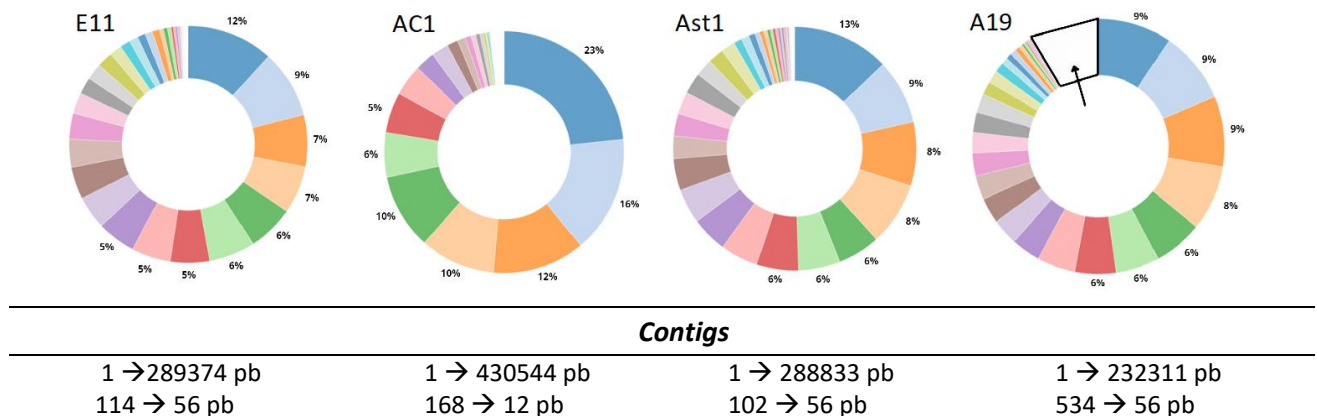


Figura 4. Representación gráfica de los *contigs* según sus longitudes en pb (pares de bases), obtenidos a partir del ensamblaje de las secuencias con SPAdes, para las cepas BAL E11, AC1, Ast1 y A19. En los gráficos circulares se indican los porcentajes de los *contigs* con mayor tamaño con respecto al tamaño total (suma de todos los *contigs*). La zona que está en blanco representa los *contigs* que son demasiado cortos (con longitudes por debajo de 1000 pb), y, por tanto, dificultan o directamente no permiten la búsqueda de genes, y mucho menos *clusters eps*. Esta fracción es más significativa en la cepa A19 (marcada con una flecha). En la tabla inferior se muestran las longitudes del *contig* más grande y pequeño para cada una de las cepas BAL.

Los resultados muestran las relaciones de tamaños de los *contigs* ensamblados. En la cepa AC1 se obtuvieron los *contigs* más largos, por lo que, de todos los ensamblajes, puede haber sido el de mayor calidad. En el caso de la cepa A19 (Figura 4), hay una zona en blanco que representa *contigs* muy cortos que van a dificultar o no van a permitir realizar la búsqueda de genes, ni detectar sus posibles funciones con la anotación. A partir del ensamblaje en SPAdes se extrajo del archivo "*contig_stats*", la lista de *contigs* y sus longitudes (pb) (ver en Anexo III). Teniendo en cuenta que la zona blanca (Figura 4) comienza en el *contig* 47, se calculó el porcentaje de *contigs* muy cortos (a partir y menos de 978 pb-s; unos 210 086 pb-s en total) con respecto a la longitud total del ensamblaje (2446851 pb-s), representando un 8,6% de todo el genoma. Esto sugiere que hay una fracción significativa del genoma de A19 (casi de la misma longitud que el *contig* de mayor longitud para esta cepa; 9%) que no se va a poder analizar con garantías a la hora de identificar posibles *clusters* y genes *eps*.

4.2. Diferencias entre las anotaciones realizadas con Prokka y RAST, y su comparación con la proporcionada por la empresa Microbes NG

Los archivos correspondientes al ensamblaje del genoma de cada cepa se depositaron en las plataformas Prokka y RAST, para realizar sus anotaciones. En este paso quedan identificados los CDS probables dentro de cada uno de los *contigs* (Tablas 3 y 4).

Tabla 3. Datos obtenidos en la anotación de los genomas de las cepas BAL con Prokka.

| Cepas | E11 | AC1 | Ast1 | A19 |
|----------------------------------|------------|------------|-------------|------------|
| Nº de <i>Contigs</i> | 66 | 96 | 61 | 476 |
| Longitud (pb) | 2 461 259 | 1 841 188 | 2 241 092 | 2 440 782 |
| Nº de secuencias de codificación | 2345 | 1818 | 2092 | 2205 |

Tabla 4. Datos obtenidos en la anotación de los genomas de las cepas BAL con RAST.

| Cepas | E11 | AC1 | Ast1 | A19 |
|----------------------------------|------------|------------|-------------|------------|
| Nº de <i>Contigs</i> | 114 | 168 | 102 | 534 |
| Longitud (pb) | 2 466 053 | 1 847 229 | 2 245 463 | 2 446 851 |
| Nº de secuencias de codificación | 2451 | 1950 | 2220 | 2767 |

La anotación realizada con Prokka filtra los resultados, y aquellos *contigs* con tamaños menores de 200 pb, los omite, ya que, de ellos no se puede extraer mucha información debido a que son muy cortos (Tabla 3). En cambio, RAST no establece este filtro, de ahí las diferencias entre ambas plataformas de anotación y el número de CDS predichas (Tabla 4).

Teniendo en cuenta que las longitudes de los genomas son relativamente parecidas para todas las cepas, e independientemente de con qué plataforma se haya realizado la anotación, se puede observar, según los resultados, que en las cepas E11, AC1 y Ast1 el número de *contigs* es mucho menor en comparación con la cepa A19. Además, el porcentaje de la suma de los *contigs* de mayor tamaño con respecto al tamaño total del genoma ensamblado es menor en A19 (y Ast1). Todo ello, sugiere que los ensamblajes de E11, AC1 y Ast1 son de mayor calidad que el de A19, y ello deberá tenerse en cuenta a la hora de describir los *clusters* y genes *eps*.

Ambas anotaciones, con Prokka y RAST, se compararon con la de la empresa MicrobesNG, a quien se le subcontrataron las secuenciaciones de los genomas (Tabla 5). Se alcanzaron resultados similares en las anotaciones realizadas en este TFG con respecto a los obtenidos por la empresa, por lo que se puede concluir que es posible y fiable el realizar la

anotación con estas plataformas de libre acceso disponibles en Internet. Con la anotación de RAST se obtuvo la mayor cantidad de CDS, por lo que para el posterior análisis se podría escoger esta plataforma, además de ofrecer información complementaria que será descrita en la siguiente sección.

Tabla 5. Comparación de los ensamblajes (longitud total) y las anotaciones (número de CDS) realizadas con Prokka y RAST con respecto a la de la empresa MicrobesNG.

| Cepas | | Empresa | Prokka | RAST |
|--------------|---------------------|----------------|---------------|-------------|
| E11 | Longitud total (pb) | 2 131 934 | 2 461 259 | 2 466 053 |
| | Nº de CDS | 2408 | 2345 | 2451 |
| AC1 | Longitud total (pb) | 1 609 334 | 1 841 188 | 1 847 229 |
| | Nº de CDS | 1873 | 1818 | 1950 |
| Ast1 | Longitud total (pb) | 1 881 668 | 2 241 092 | 2 245 463 |
| | Nº de CDS | 2154 | 2092 | 2220 |
| A19 | Longitud total (pb) | 1 984 508 | 2 440 782 | 2 446 851 |
| | Nº de CDS | 2271 | 2205 | 2767 |

4.3. Análisis de genes predictivamente relacionados con la síntesis EPS en el genoma de las cepas BAL de interés

Mediante la anotación con RAST se obtuvo también información acerca de las funciones y procesos celulares más representados que pudieran desempeñar las proteínas codificadas por los genes anotados (Figura 5). En verde oscuro están representados aquellas funciones y procesos relacionados con el ensamblaje de la pared celular y capsula bacteriana, dentro de los que están incluidos los genes relacionados con la síntesis de EPS (Figura 5).

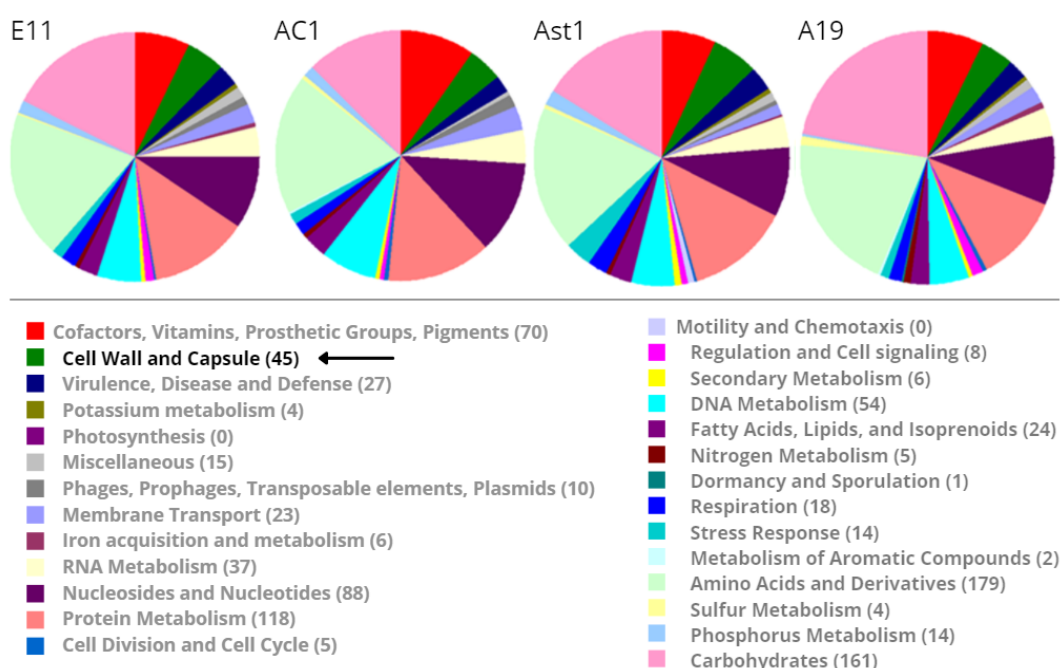


Figura 5. Representación gráfica de las funciones y procesos celulares que desempeñarían las proteínas codificadas por los genes anotados en la plataforma RAST. En verde oscuro están representados los genes que estarían relacionados con la generación de la pared celular y cápsula bacteriana, dentro de los cuales están incluidos los genes relacionados con la síntesis de EPS. En la parte inferior están nombrados los diferentes procesos celulares, con la cantidad de genes asociados a ellos.

Con esta información proporcionada por RAST y haciendo una búsqueda en la lista de genes anotados como involucrados en la generación de la pared celular y la cápsula bacteriana, se comenzó la búsqueda de posibles *clusters* y genes *eps* para cada una de las cepas BAL analizadas, obteniéndose un total de ocho posibles *clusters eps* (Figura 6). De forma general, los genes *eps* encontrados en los genomas codifican proteínas necesarias para la síntesis de HePS, tales como tirosina kinasas (**Wzd**, **Wze**), tirosina fosfatasa (**Wzb**), GT iniciadoras de la síntesis (**WeIE**), GT (**Wel**), flipasas (**wzx**) y reguladores transcripcionales (**Wzr**) (Tablas 6-9).

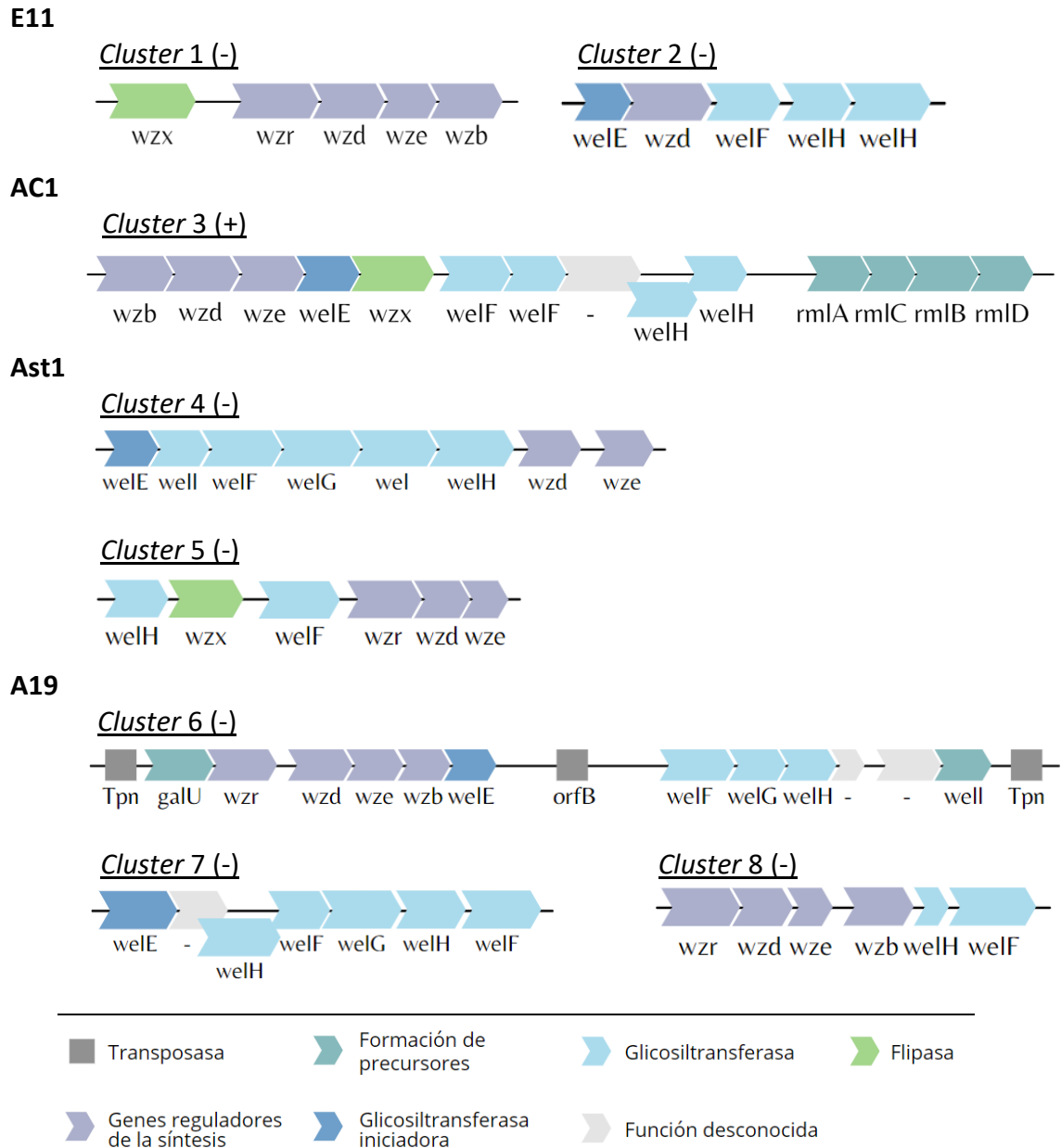


Figura 6. Representación esquemática de los posibles *clusters eps* localizados en los genomas de las cepas BAL E11, AC1, Ast1 y A19. Se tomó el trabajo de tesis doctoral Puertas et al., 2019 como referencia²⁹. Las flechas representan los genes, y su color indica la función posible de las proteínas codificadas por los mismos, en base a la nomenclatura descrita en la parte inferior de la figura.

La cepa A19 es la única que ha sido descrita como productora de HePS (Puertas et al., en redacción). Según los resultados de este análisis, no presentaría ningún *cluster eps* que incluya todos los componentes descritos en la Figura 2 (Introducción), aunque el *cluster* A19-6 tendría casi la totalidad de los genes para las actividades implicadas en la síntesis de HePS, con excepción de aquellos que codifican una polimerasa (*wzy*) y flipasa (*wzx*), que se podrían encontrar, en otros *contigs* debido a trasposiciones, o en ese 8,6% del genoma que

no se ha ensamblado correctamente. Por otro lado, las cepas E11, AC1 y Ast1 han sido descritas como productoras de dextrano (Lamas-Arriba et al., en redacción). Aun así, en este trabajo se han identificado agrupaciones genes que codificarían algunas de las actividades necesarias para la formación de HePS (ver Discusión).

Clusters E11 -1 y -2

Estos posibles *clusters* o fragmentos de *clusters* estarían codificados por la hebra complementaria y tendrían una longitud de 4701 y 4914 pb respectivamente. El *cluster* -1 estaría formado por cuatro genes codificantes de proteínas reguladoras (*wzb*: tirosina fosfatasa, *wze*, *wzd*: tirosina kinasas y *wzr*: regulador transcripcional), seguidos de un gen para un transportador de membrana flipasa (*wzx*) (Figura 6; Tabla 6). Por otro lado, el *cluster* 2 estaría formado por tres genes que codifican GT (dos *welH*, y *welF*), un gen regulador (*wzd*) y un gen para una GT iniciadora (*welE*) relacionado con la iniciación de la síntesis del polisacárido, que se encarga de añadir el pimer residuo desde un azúcar activado a un lípido de membrana (undecaprenil-P).

Cluster AC1 -3

El *cluster* AC1-3 estaría localizado en la hebra molde y tendría una longitud de alrededor de 13 Kb. En él estarían presentes cuatro genes que codifican proteínas relacionadas con la biosíntesis de dTDP-L-ramnosa (*rmlD*, *rmlB*, *rmlC* y *rmlA*), seguidos de cuatro genes que codifican GT-s (dos *welH* y dos *welF*), un gen para una flipasa (*wzx*), un gen para una GT iniciadora (*welE*) y tres genes reguladores (*wze*, *wzd* y *wzb*) (Figura 6; Tabla 7).

Clusters Ast1 -4 y -5

El *cluster* Ast1-4 estaría localizado en la hebra complementaria con una longitud de 7278 pb. Estaría formado en su mayoría por genes que codifican GT, un gen para una GT iniciadora (*welE*) y dos genes para proteínas reguladoras (*wzd* y *wze*). De igual manera, El *cluster* Ast1-5 estaría codificado en la hebra complementaria y tendría una longitud de 5973 pb. Presentaría tres genes que codifican proteínas reguladoras (*wze*, *wzd* y *wzr*), dos genes para GT (*welF* y *welH*) y un gen para una flipasa (*wzx*) (Figura 6; Tabla 8).

Clusters A19 -6, -7 y -8

El *cluster* A19-6 se localizaría en la hebra complementaria con una longitud de 11047 pb. En los extremos estaría flanqueado por genes para Transposasas (Tpn). En este posible *cluster* se encontrarían, un gen relacionado con la síntesis de precursores de la síntesis de EPS, UDP-azúcares, (*galU*), cuatro genes que codificarían reguladores (*wzr*, *wzd*, *wze* y *wzb*), un gen para GT iniciadora (*welE*) y cuatro genes para GT (*welF*, *welG*, *welH* y *welI*) (Figura 6; Tabla 9).

Los *clusters* A19-7 y -8 estarían codificados por la hebra complementaria y tendrían una longitud de 6228 y 4536 pb, respectivamente. El *cluster* -7 estaría formado por cinco genes codificantes de GT (*welF*, *welG* y dos *welH*) y un gen para una GT iniciadora (*welE*). En el *cluster* -8 se encontrarían cuatro genes para proteínas reguladoras (*wzr*, *wzd*, *wze* y *wzb*) y dos genes que codifican GT (*welH* Y *welF*) (Figura 6; Tabla 9).

En las tablas generadas (Tablas 6-9), además de representarse los genes encontrados, se señaló su número (peg.) teniendo en cuenta la anotación con RAST, y se indican las longitudes y funciones predichas de las proteínas correspondientes. Se incluyen también los códigos de los dominios funcionales de dichas proteínas (proporcionados por InterPro). Además, haciendo uso de plataformas como NCBI y su herramienta CD-Search, se determinaron los dominios conservados y superfamilia para cada secuencia analizada, aportando los valores-e, que indican cómo de comparables son dos secuencias (cuanto más cercano a cero es el valor e, mayor conservación de secuencias). Por último, utilizando el algoritmo Blastp, se determinó con qué proteínas bacterianas tienen las secuencias analizadas una mayor identidad, es decir, el porcentaje de aminoácidos que son iguales y conservan la posición en las dos proteínas comparadas. En su mayoría, y como era de esperar, las proteínas analizadas presentan una mayor similitud a otras de especies del género *Lactobacillus* (que con la reciente revisión de la nomenclatura son denominados como *Liquorilactibacillus*, *Lactiplantibacillus*, *Lentilactobacillus*, *Ligilactobacillus* y *Secundilactobacillus* <http://lactobacillus.uantwerpen.be>), a excepción del *cluster* 3 en AC1, en el que la totalidad de las proteínas presentan identidades casi del 100% con proteínas de la especie *Leuconostoc carnosum* (Tabla 9). Este resultado es totalmente lógico teniendo en cuenta que AC1 es una cepa de *Leuconostoc carnosum* (Llamas-Arriba et al., 2019)³⁰.

Tabla 6. Organización génica predicha de los *clusters eps* en la cepa E11 (*Contigs 30* y *16* respectivamente).

| RAST ID | Cluster y genes | Longitud (aa) | Código dominio InterPro | ^a Función predicha | ^b Dom. Cons. (valor-e) / S-Fam | ^c Bacteria / Proteína con la mayor identidad (valor-e) | ^c I (%) |
|------------------|-----------------|---------------|---------------------------|---|--|---|--------------------|
| Cluster 1 | | | | | | | |
| peg.1397 | wzx | 469 | cd13127 | Oligosaccharide flippase | MATE_tuaB_like cd13127 (1,93e-99) MATE_like cl09326 | <i>Enterococcus casseliflavus</i> WP_192415946.1 (5e-174) | 56,13 |
| peg.1396 | wzr | 318 | IPR004474 | Cell envelope-associated transcriptional attenuator LytR-CpsA-Psr | LytR PRK09379 (1,67e-92) LytR_cpsA_psr cl00581 | <i>Liquorilactobacillus mali</i> WP_056990414.1 (0,0) | 88,64 |
| peg.1395 | wzd | 258 | IPR003856 | Tyrosine-protein kinase transmembrane modulator EpsC | YveK COG3944 (1,05e-47) YveK cl34698 | <i>Liquorilactobacillus mali</i> WP_010077715.1 (9e-172) | 92,22 |
| peg.1394 | wze | 256 | IPR005702 | Tyrosine-protein kinase EpsD | BY-kinase cd05387 (1,51e-76) P-loop_NTPase cl38936 | <i>Liquorilactobacillus mali</i> WP_010077716.1 (0,0) | 97,65 |
| peg.1393 | wzb | 266 | IPR016667 | Manganese-dependent protein-tyrosine phosphatase | YwqE COG4464 (4,89e-80) PHP cl23724 | <i>Liquorilactobacillus mali</i> KRN33115.1 (0,0) | 98,11 |
| Cluster 2 | | | | | | | |
| peg.520 | welE | 226 | IPR003362 | Undecaprenyl-phosphate galactosephosphotransferase | Bac_transf pfam02397 (5,57e-96) Bac_transf cl29433 | <i>Liquorilactobacillus mali</i> KRN33101.1 (1e-124) | 91,05 |
| peg.519 | wzd | 405 | IPR017476 | UDP-glucose 6-dehydrogenase | UDP-glucose 6-dehydroge PRK15057 (0,0) cl33071 | <i>Liquorilactobacillus mali</i> WP_010077703.1 (0,0) | 96,53 |
| peg.518 | welF | 389 | IPR015393 | Alpha-D-GlcNAc alpha-1,2-L-rhamnosyltransferase | Glycos_transf pfam00534 (0,0) Glycosyltransferase_GTB cl10013 | <i>Liquorilactobacillus mali</i> WP_010077704.1 (0,0) | 80,99 |
| peg.517 | welH | 293 | IPR001173 | Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase | Glycos_transf cd00761 (3,07e-50) Glyco_tranf_GTA_type cl11394 | <i>Enterococcus faecalis</i> EGO7569190.1 (6e-71) | 47,20 |
| peg.516 | welH | 325 | IPR001173 | Glycosyltransferase | Glycos_transf cd00761 (1,07e-42) Glyco_tranf_GTA_type cl11394 | <i>Enterococcus faecalis</i> EGO7569190.1 (2e-42) | 41,47 |

(a) Función predicha derivada de la anotación de RAST y contrastada con las obtenidas a partir de las comparaciones de los genes *eps* en la base de datos de InterPro.

(b) Resultados obtenidos utilizando la base de datos NCBI y su función CD-Search.

(c) Resultados obtenidos utilizando la función Blastp.

Tabla 7. Organización génica predicha de los *clusters eps* en la cepa AC1 (Contig 3).

| RAST ID | Cluster y genes | Longitud (aa) | Código dominio InterPro | ^a Función predicha | ^b Dom. cons. (valor-e) / S-Fam | ^c Bacteria / Proteína con la mayor identidad (valor-e) | ^d I (%) |
|------------------|-----------------|---------------|---------------------------|--|---|---|--------------------|
| Cluster 3 | | | | | | | |
| peg.1130 | <i>rmID</i> | 288 | IPR005913 | dTDP-4-dehydrorhamnose reductase | RmID_sub_bind pfam04321 (2,26e-133) RmID_sub_bind cl37705 | <i>Leuconostoc carnosum</i> WP_135197270.1 (0,0) | 100 |
| peg.1129 | <i>rmIB</i> | 346 | IPR005888 | dTDP-glucose 4,6-dehydratase | RfbB COG1088 (0,0) RfbB cl34098 | <i>Leuconostoc carnosum</i> WP_135197269.1 (0,0) | 100 |
| peg.1128 | <i>rmIC</i> | 199 | IPR000888 | dTDP-4-dehydrorhamnose 3,5-epimerase | RfbC COG1898 (6,16e-54) cupin_RmIC-like cl40423 | <i>Leuconostoc carnosum</i> WP_135198293.1 (5e-144) | 100 |
| peg.1127 | <i>rmIA</i> | 291 | IPR005907 | Glucose-1-phosphate thymidyltransferase | rmIA TIGR01207 (0,0) Glyco_tranf_GTA cl11394 | <i>Leuconostoc carnosum</i> WP_135198292.1 (0,0) | 100 |
| peg.1126 | <i>welH</i> | 328 | IPR001173 | Glycosyltransferase | GT_2_like_d cd04196 (9,73e-56) Glyco_tranf_GTA_type cl11394 | <i>Leuconostoc carnosum</i> WP_135198291.1 (0,0) | 100 |
| peg.1125 | <i>welH</i> | 373 | IPR028098 | Glycosyltransferase | GT4_CapH-like cd03812 (4,52e-89) Glycosyltransf_GTB-type cl10013 | <i>Leuconostoc carnosum</i> WP_135198290.1 (0,0) | 99,73 |
| peg.1124 | - | 412 | PS51257 | Hypothetical protein | - | <i>Leuconostoc carnosum</i> WP_184284944.1 (0,0) | 100 |
| peg.1123 | <i>welF</i> | 308 | IPR001173 | Glycosyltransferase | GT_2_like_b cd04185 (2,84e-52) Glyco_tranf_GTA_type cl11394 | <i>Leuconostoc carnosum</i> WP_150289079.1 (0,0) | 100 |
| peg.1122 | <i>welF</i> | 347 | IPR028098 | Poly(glycerol-phosphate) alpha-glycosyltransferase | GT4_CapM-like cd03808 (7,38e-61) Glycosyltransf_GTB-type cl10013 | <i>Leuconostoc carnosum</i> WP_184284943.1 (0,0) | 100 |
| peg.1121 | <i>wzx</i> | 480 | IPR002797 | Flippase | MATE_Wzx_like cd13128 (7,89e-42) MATE_like cl09326 | <i>Leuconostoc carnosum</i> WP_184284942.1 (0,0) | 100 |
| peg.1120 | <i>welE</i> | 225 | IPR003362 | Undecaprenyl-phosphate galactosephosphotransferase | Bac_transf pfam02397 (9,40e-97) Bac_transf cl29433 | <i>Leuconostoc carnosum</i> WP_184284941.1 (7e-174) | 100 |
| peg.1119 | <i>wze</i> | 244 | IPR005702 | Tyrosine-protein kinase EpsD | BY-kinase cd05387 (8,15e-77) P-loop_NTPase cl38936 | <i>Leuconostoc carnosum</i> WP_184284940.1 (7e-174) | 99,59 |
| peg.1118 | <i>wzd</i> | 259 | IPR003856 | Tyrosine-protein kinase transmembrane modulator EpsC | YveK COG3944 (1,96e-44) YveK cl34698 | <i>Leuconostoc carnosum</i> WP_184284939.1 (1e-169) | 99,58 |
| peg.1117 | <i>wzb</i> | 256 | IPR016667 | Manganese-dependent protein-tyrosine phosphatase | YwqE COG4464 (2,12e-78) PHP cl23724 | <i>Leuconostoc carnosum</i> WP_150289073.1 (0,0) | 98,04 |

Tabla 8. Organización génica predicha de los *clusters eps* en la cepa Ast1 (*Contigs 1 y 24* respectivamente).

| RAST ID | Cluster y genes | Longitud (aa) | Código dominio InterPro | ^a Función predicha | ^b Dom. Cons. (valor-e) / S-Fam | ^c Bacteria / Proteína con la mayor identidad (valor-e) | ^e I (%) |
|------------------|-----------------|---------------|---------------------------|--|--|--|--------------------|
| Cluster 4 | | | | | | | |
| peg.608 | <i>welE</i> | 219 | IPR003362 | Undecaprenyl-phosphate galactosephosphotransferase | EPS_sugtrans TIGR03025 (7.75e-87) EPS_sugtrans cl37214 | <i>Liquorilactobacillus oeni</i> WP_057896533.1 (8e-115) | 72,94 |
| peg.607 | <i>welI</i> | 261 | IPR001173 | Glycosyltransferase | Glycos_transf_2 pfam00535 (6.49e-35) Glycosyl transferase cl11394 | <i>Lactiplantibacillus plantarum</i> WP_024002487.1 (4e-74) | 47,60 |
| peg.606 | <i>welF</i> | 370 | IPR028098 | Glycosyltransferase | Glycos_transf pfam00534 (4.13e-85) Glycosyltransf_GTB cl10013 | <i>Liquorilactobacillus satsumensis</i> WP_056960653.1 (2e-110) | 45,86 |
| peg.605 | <i>welG</i> | 361 | PF14897 | Glycosyltransferase | EpsG pfam14897 (3.21e-21) EpsG cl22544 | <i>Lactiplantibacillus plantarum</i> WP_074029021.1 (4e-58) | 34,59 |
| peg.604 | <i>wel</i> | 338 | IPR001173 | Glycosyltransferase | Putative glycosyl transferase PRK10073 (1.65e-39) cl32454 | <i>Lactobacillus sp.</i> RRG08370.1 (5e-115) | 51,32 |
| peg.603 | <i>welH</i> | 372 | IPR001296 | Glycosyltransferase | GT4_CapH-like cd03812 (8.59e-87) Glycosyltransf_GTB-type cl10013 | <i>Liquorilactobacillus oeni</i> WP_057896532.1 (0,0) | 69,27 |
| peg.602 | <i>wzd</i> | 252 | IPR003856 | Tyrosine-protein kinase transmembrane modulator EpsC | polys_exp_MPA1 TIGR01006 (1.68e-45) polys_exp_MPA1 cl31049 | <i>Liquorilactobacillus oeni</i> WP_057896529.1 (2e-102) | 64,43 |
| peg.601 | <i>wze</i> | 253 | IPR005702 | Tyrosine-protein kinase EpsD | BY-kinase cd05387 (3.00e-79) P-loop_NTPase cl38936 | <i>Liquorilactobacillus satsumensis</i> WP_056960650.1 (9e-137) | 88,53 |
| Cluster 5 | | | | | | | |
| peg.968 | <i>welH</i> | 304 | IPR001173 | Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase | GT_2_like_d cd04196 (6.18e-71) Glyco_tranf_GTA_type cl11394 | <i>Lactiplantibacillus plantarum</i> WP_155430178.1 (4e-102) | 53,11 |
| peg.967 | <i>wzx</i> | 472 | IPR002797 | Membrane protein involved in the export of O-antigen, teichoic acid (Flippase) | MATE_tuaB_like cd13127 (4.60e-37) MATE_like cl09326 | <i>Liquorilactobacillus vini</i> WP_010581022.1 (0,0) | 61,56 |
| peg.966 | <i>welF</i> | 391 | IPR001296 | Glycosyltransferase | GT4_PimA-like cd03801 (4,11e-35) GTB-type cl10013 | <i>Lactobacillaceae</i> WP_070720707.1 (2e-76) | 35,86 |
| peg.965 | <i>wzr</i> | 316 | IPR004474 | Cell envelope-associated transcriptional attenuator LytR-CpsA-Psr | LytR PRK09379 (2.57e-100) LytR_cpsA_psr cl00581 | <i>Lactobacillus uvarum</i> KRL36667.1 (0,0) | 84,52 |
| peg.964 | <i>wzd</i> | 255 | IPR003856 | Tyrosine-protein kinase transmembrane modulator EpsC | YveK COG3944 (3.70e-50) YveK cl34698 | <i>Liquorilactobacillus satsumensis</i> WP_056960649.1 (5e-144) | 83,86 |
| peg.963 | <i>wze</i> | 253 | IPR005702 | Tyrosine-protein kinase EpsD | BY-kinase cd05387 (1.02e-80) P-loop_NTPase cl38936 | <i>Liquorilactobacillus satsumensis</i> WP_056960650.1 (4e-149) | 92,53 |

Tabla 9. Organización génica predicha de los *clusters eps* en A19 (Contigs 18, 27 y 10 respectivamente).

| RAST ID | Cluster y genes | Longitud (aa) | Código dominio InterPro | *Función predicha | ^b Dom. cons. (valor-e) / S-Fam | ^c Bacteria / Proteína con la mayor identidad (valor-e) | ^c l (%) |
|------------------|-----------------|---------------|---------------------------|---|--|---|--------------------|
| <i>Cluster 6</i> | | | | | | | |
| peg.680 | <i>Tpn</i> | 252 | IPR001207 | Transposase | - | - | - |
| peg.679 | <i>well</i> | 336 | IPR001173 | Glycosyltransferase | Glycos_transf_2 pfam00535 (5,74e-33) Glyco_transf_GTA cl11394 | <i>Lactiplantibacillus plantarum</i> WP_138069598.1 (0,0) | 99,39 |
| peg.678 | - | 495 | - | Hypothetical protein | - | - | - |
| peg.677 | - | 94 | - | Hypothetical protein | - | - | - |
| peg.676 | <i>welH</i> | 362 | IPR028098 | Glycosyltransferase | GT4_CapH-like cd03812 (1,84e-91) Glycos_GTB cl10013 | <i>Lactobacillaceae</i> WP_102197015.1 (0,0) | 99,72 |
| peg.675 | <i>welG</i> | 372 | IPR001296 | Glycosyltransferase | GT4_CapH-like cd03812 (3,49e-96) Glycos_GTB cl10013 | <i>Lactiplantibacillus plantarum</i> WP_138069596.1 (0,0) | 99,46 |
| peg.674 | <i>welF</i> | 359 | IPR001296 | Poly(glycerol-phosphate) alpha-glucosyltransferase | GT1_Gtf-like cd03784 (3,89e-66) Glycos_GTB cl10013 | <i>Lactobacillaceae</i> WP_105316413.1 (0,0) | 99,72 |
| peg.673 | <i>orfB</i> | 234 | IPR001584 | Mobile element protein | Putative transposase PHA02517 (1.24e-123) cl33679 | <i>Lactobacillaceae</i> WP_057732102.1 (9e-167) | 99,56 |
| peg.672 | <i>welE</i> | 233 | IPR003362 | Undecaprenyl-phosphate galactosephosphotransferase | Bac_transf pfam02397 (1,74e-83) Bac_transf cl29433 | <i>Lentilactobacillus sunkii</i> WP_070368066.1 (2e-167) | 99,14 |
| peg.671 | <i>wzb</i> | 255 | IPR016667 | Manganese-dependent protein-tyrosine phosphatase | YwqE COG4464 (1,03e-70) PHP cl23724 | <i>Lentilactobacillus buchneri</i> WP_153152602.1 (3e-175) | 92,52 |
| peg.670 | <i>wze</i> | 248 | IPR025669 | Tyrosine-protein kinase EpsD | BY-kinase cd05387 (9,31e-77) P-loop_NTPase cl38936 | <i>Lactobacillus buchneri</i> EEI18213.1 (7e-174) | 97,57 |
| peg.669 | <i>wzd</i> | 287 | IPR003856 | Tyrosine-protein kinase transmembrane modulator EpsC | YveK COG3944 (3,29e-40) YveK cl34698 | <i>Secundilactobacillus collinoides</i> WP_063285228.1 (0,0) | 97,20 |
| peg.668 | <i>wzr</i> | 326 | IPR004474 | Cell envelope-associated transcriptional attenuator LCP | LytR PRK09379 (3,44e-76) LytR_cpsA_psr cl00581 | <i>Secundilactobacillus collinoides</i> WP_063285229.1 (0,0) | 97,85 |
| peg.667 | <i>galU</i> | 311 | IPR005835 | UTP--glucose-1-phosphate uridylyltransferase | GalU COG1210 (4,59e-169) Glyco_transf_GT-A cl11394 | <i>Secundilactobacillus collinoides</i> WP_063285230.1 (0,0) | 99,03 |
| peg.666 | <i>Tpn</i> | 264 | IPR001626 | Zinc ABC transporter, permease protein ZnuB | - | - | - |

| Cluster 7 | | | | | | | |
|-----------|------|-----|---------------------------|---|---|--|-------|
| Peg.1279 | welE | 225 | IPR017475 | Undecaprenyl-phosphate galactosephosphotransferase | Bac_transf pfam02397 (4,45e-100) Bac_transf cl29433 | <i>Lactobacillus vini</i> KRM85207.1 (3e-135) | 82,59 |
| Peg.1278 | - | 255 | IPR025536 | Hypothetical protein | - | <i>Lactobacillus sp.</i> MBD5068713.1 (3e-95) | 54,98 |
| Peg.1277 | welH | 327 | IPR008166 | Glycosyltransferase family 2 | Glyco_tranf_2_4 pfam13704 (1,3e-15) Glyco_tranf_GTA_type cl11394 | <i>Liquorilactobacillus satsumensis</i> WP_056960644.1 (6e-90) | 47,91 |
| Peg.1276 | welF | 307 | IPR001173 | dTDP-rhamnosyl transferase RfbF | GT2_RfbF_like cd02526 (4,64e-48) Glyco_tranf_GTA_type cl11394 | <i>Ligilactobacillus murinus</i> WP_162626441.1 (5e-121) | 55,59 |
| Peg.1275 | welG | 354 | PF14897 | Glycosyltransferase family 2 | EpsG pfam14897 (1.21e-05) EpsG cl22544 | <i>Lactobacillus sp.</i> MBD5430748.1 (8e-49) | 47,91 |
| Peg.1274 | welH | 311 | IPR001173 | Alpha-L-Rha alpha-1,3 - L rhamnosyltransferase | GT_2_like_d cd04196 (3.01e-85) Glyco_tranf_GTA_type cl11394 | <i>Liquorilactobacillus vini</i> WP_010579809.1 (4e-101) | 47,91 |
| Peg.1273 | welF | 297 | IPR001173 | Glycosyl transferase | GT2_RfbF_like cd02526 (8.22e-45) Glyco_tranf_GTA_type cl11394 | <i>Lactiplantibacillus argentoratensis</i> WP_210697326.1 (4e-55) | 93,81 |
| Cluster 8 | | | | | | | |
| peg.86 | wzr | 323 | IPR004474 | Cell envelope-associated transcriptional attenuator LCP | LytR PRK09379 (4,23e-97) LytR_cpsA_psr cl00581 | <i>Liquorilactobacillus vini</i> WP_010579467.1 (0,0) | 92,19 |
| peg.85 | wzd | 255 | IPR003856 | Tyrosine-protein kinase transmembrane modulator EpsC | YveK COG3944 (1,25e-53) YveK cl34698 | <i>Liquorilactobacillus vini</i> WP_034994003.1 (2e-170) | 92,89 |
| peg.84 | wze | 183 | IPR005702 | Tyrosine-protein kinase | BY-kinase cd05387 (8,57e-60) P-loop_NTPase cl38936 | <i>Liquorilactobacillus vini</i> WP_034994000.1 (9e-126) | 97,25 |
| peg.83 | wzb | 273 | IPR016667 | Manganese-dependent protein-tyrosine phosphatase | YwqE COG4464 (1,45e-77) PHP cl23724 | <i>Liquorilactobacillus vini</i> WP_026049403.1 (0,0) | 92,80 |
| peg.82 | welH | 138 | IPR007267 | GtrA family protein | GtrA pfam04138 (1,19e-18) Bac_transf cl19823 | <i>Liquorilactobacillus vini</i> WP_010579463.1 (2e-82) | 86,86 |
| peg.81 | welF | 340 | IPR001173 | Glycosyl transferase | GT_2_like_a cd02522 (2,00e-27) Glyco_transf_GT-A cl11394 | <i>Liquorilactobacillus vini</i> WP_010579462.1 (0,0) | 93,81 |

En la síntesis de HePS por la vía Wzy-dependiente, las flipasas y polimerasas juegan un papel fundamental en el transporte de la unidad repetitiva del polisacárido desde la parte interior de la membrana a la exterior y la síntesis del polisacárido, respectivamente. Por ello, en el caso de la cepa A19 se realizó una búsqueda de otros genes que pudieran codificarlas y que estuvieran alejados físicamente de los *clusters eps* identificados, debido a la posible trasposición que pudieran haber sufrido (Tabla 10).

Tabla 10. Detección de genes codificantes de posibles flipasas y polimerasas en la cepa A19, y que estuvieran alejados físicamente de los posibles *clusters eps* descritos previamente

| <i>RAST</i> | Función predicha (Código Dominio InterPro) | Dom. cons. (valor-e) / S-Fam | Bacteria / Proteína con la mayor identidad (valor-e) (%) |
|--------------------------|--|--|--|
| FLIPASAS | | | |
| <i>Contig 2</i> | | | |
| peg.1421 | Membrane protein involved in the export of O-antigen, teichoic acid lipoteichoic acids | MATE_Wzx_like cd13128 (1.18e-35) MATE_like cl09326 | <i>Ligilactobacillus murinus</i> WP_112194038.1 (0,0) (58,19) |
| <i>Contig 134</i> | | | |
| peg.283 | Membrane protein involved in the export of O-antigen, teichoic acid lipoteichoic acids | MATE_Wzx_like cd13128 (2.87e-11) MATE_like cl09326 | <i>Lactiplantibacillus plantarum</i> TBX47097.1 (1e-65) (55,38) |
| POLIMERASAS | | | |
| <i>Contig 1</i> | | | |
| peg.960 | DNA polymerase III delta prime subunit (IPR004622) | DNA polymerase III delta PRK08058 (2.87e-102) DNA polymerase cl32265 | <i>Liquorilactobacillus vini</i> WP_010581438.1 (0,0) (87,65) |
| <i>Contig 2</i> | | | |
| peg.1491 | DNA polymerase III beta subunit (IPR001001) | DNA polymerase III beta PRK05643 (3.25e-135) DNA polymerase cl35355 | <i>Liquorilactobacillus vini</i> WP_010581278.1 (0,0) (90,01) |
| <i>Contig 4</i> | | | |
| peg.2010 | DNA polymerase III polC-type (IPR006308) | DNA polymerase III PolC PRK00448 (0.0) DNA polymerase cl35100 | <i>Liquorilactobacillus vini</i> WP_056970404.1 (0,0) (95,02) |
| <i>Contig 4</i> | | | |
| peg.2061 | DNA polymerase III alpha subunit (IPR004805) | DNA polymerase III, alpha COG0587 (0.0) DNA polymerase cl33964 | <i>Liquorilactobacillus vini</i> WP_141567395.1 (0,0) (88,81) |

La detección de estos genes que codificarían flipasas y polimerasas, podría justificar la posible síntesis de HePS para la cepa A19, en el caso de que por trasposición se hayan alejado de los *clusters eps* (ver discusión).

4.4. Identificación de genes que codificarían proteínas con actividad dextransacarasa

Para detectar la presencia de genes que pudieran codificar actividades *dsr*, se han comparado secuencias *dsr* contenidas en bases de datos (*query*) del NCBI, haciendo uso de la función BLAST en la propia anotación RAST de cada una de las cepas BAL aquí analizadas (Tabla 11). Se obtuvieron un total de cinco genes (*hit-s*) que pudieran codificar posibles actividades *dsr* (uno en E11, dos en AC1, dos en Ast1, y ninguno en A19). La distribución de estos genes en cada uno de los genomas estaría de acuerdo con la producción de dextrano descrita para E11, AC1 y Ast1, y para la producción de HePS (y no dextrano) por parte de A19. El análisis posterior de las secuencias de aminoácidos de las proteínas correspondientes en la plataforma InterPro permitió identificar y localizar la posición de los probables dominios *dsr* en cada uno de los *hit-s* (Figura 7). A partir de las secuencias de genes *dsr* encontradas, se diseñaron oligonucleótidos para realizar reacciones de PCR dirigidas a confirmar experimentalmente la presencia de los genes identificados en la cepa Ast1 (Tabla 1).



Figura 7. Identificación de los posibles dominios dextransacarasa (en verde) en las secuencias de aminoácidos codificadas por los posibles genes *dsr*. Los gráficos fueron exportados desde la plataforma InterPro.

Tabla 11. Detección de genes con posible actividad dextranasacarasa.

| Cepas | Secuencias RAST | Tamaño (aa) | Query (NCBI) | Score | Valor E | Identidad de los aminoácidos (%) | Positivos de los aminoácidos (%) | Proteína con la mayor identidad (I%) (valor-e) (score) |
|-------|-----------------|-------------|--------------------------|-------|---------|----------------------------------|----------------------------------|--|
| E11 | peg.1356 | 1054 | Dextranucrase CCK33644.1 | 1015 | 0,0 | 55 | 72 | YG repeat-containing glycosyl hydrolase family 70 WP_010078043.1 (99,34) (0,0) (2154) |
| | | | Dextranucrase CCK33643.1 | 1034 | 0,0 | 51 | 67 | |
| | | | Dextranucrase QDK48460.1 | 976 | 0,0 | 55 | 69 | |
| | | | Dextranucrase KEK16955.2 | 957 | 0,0 | 54 | 70 | |
| AC1 | peg.1102 | 1342 | Dextranucrase ATN28243.1 | 1169 | 0,0 | 54 | 70 | Dextranucrase MBB6432305.1 (97,67) (0,0) (2563) |
| | | | Dextranucrase BAX70498.1 | 1934 | 0,0 | 69 | 80 | |
| | peg.278 | 1603 | Dextranucrase ATN28243.1 | 1230 | 0,0 | 54 | 69 | Glycosyl hydrolase BAX71365.1 (64,85) (0,0) (1993) |
| | | | Dextranucrase BAX70498.1 | 1511 | 0,0 | 51 | 66 | |
| Ast1 | peg.1686 | 1081 | Dextranucrase CCK33644.1 | 1044 | 0,0 | 56 | 73 | YG repeat-containing glycosyl hydrolase family 70 WP_010078043.1 (76,04) (0,0) (1644) |
| | | | Dextranucrase QDK48460.1 | 1020 | 0,0 | 57 | 71 | |
| | peg.1246 | 1022 | Dextranucrase ADB43097.3 | 1008 | 0,0 | 51 | 68 | Membrane associated glycosyl hydrolase with glucan-binding domain ABJ61965.1 (72,38) (0,0) (1555) |
| | | | Dextranucrase CCK33643.1 | 1006 | 0,0 | 55 | 69 | |
| | | | Dextranucrase BAX70498.1 | 1536 | 0,0 | 72 | 85 | |
| | | | Dextranucrase BAX70498.1 | 1536 | 0,0 | 72 | 85 | |

4.5. PCR para la amplificación de los genes *dsr* en la cepa Ast1

Tras el diseño de los cebadores (Tabla 1) para amplificar los dos posibles genes *dsr* encontrados en la cepa Ast1, peg.1686 (1) y peg.1246 (2) (Figura 8), se realizaron las reacciones de PCR utilizando como molde dos muestras de DNA de dicha cepa (a y b), obtenidas previamente por miembros del grupo GlycoBAL. La preparación de las muestras y las condiciones de la reacción se muestran en la Tabla 2. Ambos amplicones presentaban los tamaños predichos (3.0 Kb en el caso de peg.1686, y 2.8 Kb en el de peg.1246; ver Anexo II), por lo que se concluyó que la amplificación fue correcta. Todo esto, confirmaría la presencia de dichos genes en los genomas analizados, por lo que la cepa Ast1 podría ser productora de dextrano.

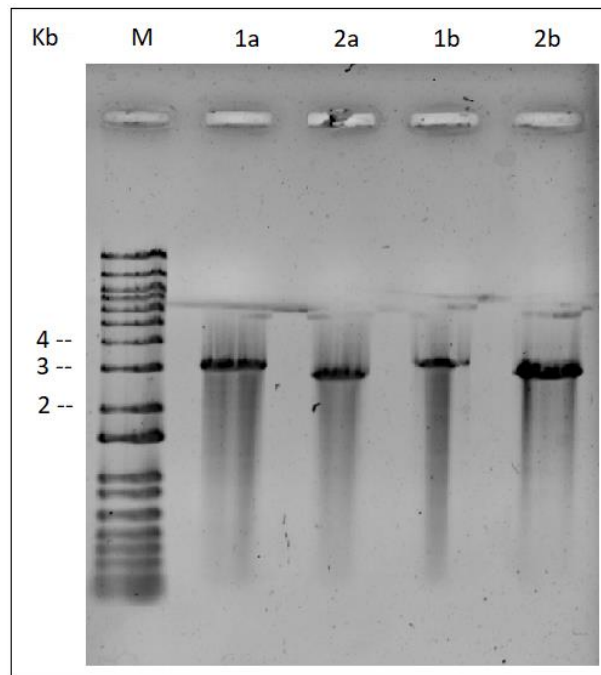


Figura 8. Detección por PCR de los genes *dsr* en Ast1. Análisis en gel de agarosa de las reacciones de PCR para los genes *dsr*: peg.1686 (1) y peg.1246 (2) realizadas con las preparaciones de DNA (a y b) de la cepa Ast1. M es el marcador de peso molecular, el cual, incluye varias bandas de tamaño conocido que se toman como referencia.

5. Discusión

El proceso de ensamblaje de las lecturas correspondientes a la secuenciación Illumina es el paso inicial del estudio, por lo que conviene que dichas lecturas sean lo más numerosas y lo más largas posible, de modo que haya una mayor probabilidad de que algunas de ellas solapen y se puedan ensamblar *contigs* lo más largos posible. En la cepa A19 no se ha conseguido ensamblar una parte significativa (8,6%) del genoma, (Figura 4), quedando una zona con lecturas muy cortas que no codificarán, por lo que de esa parte se pierde la información y no se puede anotar ni analizar posteriormente, limitando los resultados del estudio. El resto de las cepas presentan igualmente una parte del genoma que tampoco ha sido ensamblada, pero la fracción con respecto a la longitud total del genoma ensamblado es mucho menor.

Los EPS que producen las BAL difieren en su composición al presentar diferentes tipos de enlaces y unidades de monosacáridos (HePS o HoPS), lo cual estará condicionado por su mecanismo de producción. La capacidad de producción de HePS está mediada por *clusters* que presentan una serie de genes específicos que actúan de manera concertada, codificando enzimas y proteínas que serán las encargadas de la síntesis³¹. De este modo, se describieron un total de ocho posibles *clusters eps* en los genomas estudiados, de los que seguramente algunos no les proporcionen a sus respectivas bacterias la capacidad para producir HePS. Esto se debe a que no presentan todos los genes fundamentales para la síntesis, es decir, los relacionados con funciones de regulación, de formación de precursores, de transporte a través de la membrana y, además, de genes relacionados con la formación de dNDP-azúcares³².

En el caso de las cepas AC1 y A19, los *clusters eps* -3 y -6, respectivamente, parecen más completos con respecto a la estructura consenso. Sólo poseen todos los genes reguladores (*wzr*, *wzd*, *wze* y *wzb*) los *clusters* E11-1, y A19-6 y -8. Los genes *welE* están presentes para todas las cepas, en al menos uno de los *clusters* encontrados para ellas. Son los encargados de codificar una GT iniciadora implicada en el primer paso de la síntesis de unidades repetidas de HePS, siendo fundamentales, por lo que se suelen hacer PCR dirigidas a amplificar ese gen³³. Los *clusters* E11-1, AC1-3 y Ast-5 son los únicos que tienen los genes para flipasas (Figura 6). Los genes encargados de la síntesis de flipasa se han visto en

regiones muy variables y a veces hasta alejadas de los *clusters eps*¹⁰, lo cual podría explicar la falta de estos en algunos de los *clusters* encontrados.

Se podría justificar la posible síntesis de HePS, si entre los *clusters* de una misma cepa, existiese una función cooperativa³⁴. Sin embargo, no se han identificado genes que codifiquen enzimas polimerasas, por lo que el hecho de que exista una cooperación entre los *clusters* no sugiere una respuesta a su falta. En el caso de la cepa A19, como está descrita como productora de HePS (Puertas et al., en redacción), se realizó un análisis más profundo en zonas alejadas de los *clusters* encontrados. En primer lugar, se utilizaron los códigos de las enzimas polimerasas con mayor identidad utilizadas para los *clusters eps* de CUPV237 (Puertas et al., 2019), pero no se encontraron similitudes significantes. Y en segundo lugar, se realizó una búsqueda en el genoma anotado de A19 y se encontró, según las funciones predictivas que proporciona RAST, cuatro posibles genes para polimerasa (Tabla 10). Estos genes que codifican polimerasa, junto con los de flipasa, también encontrados en zonas alejadas de los *clusters*; pueden haberse separado de los *clusters* debido a una fragmentación por transposasas que habitualmente se encuentran formando parte, y/o próximas a los *clusters eps*³⁵, tal y como pasa con el *cluster* 6 (Tabla 9). Este tipo de secuencias puede hacer que los *clusters* estén fragmentados, debido a la transposición, lo que dificulta el análisis. Todas estas hipótesis, y el hecho de presentar un *cluster* prácticamente entero, podrían justificar que la cepa A19, tal y como está descrita, sintetice HePS. Además, que la cepa A19 tenga una parte bastante significativa del genoma (8,6%) no ensamblada correctamente (Figura 4), nos sugiere que los genes *wzy* y *wzx* podrían encontrarse en esa zona del genoma no ensamblada. En ese caso, una PCR con oligonucleótidos degenerados podría ser una opción para detectar polimerasas de este tipo.

Los genes *eps* encontrados en E11 codifican en su mayoría proteínas con una identidad entre el 80 y 98 % con proteínas de la especie *Liquorilactobacillus mali*. Los de AC1 codifican proteínas con identidades muy cercanas al 100% con respecto a proteínas de *Leuconostoc carnosum*. Por último, los genes de los *clusters* en Ast1 y A19 mostraron una similitud con proteínas de especies pertenecientes al género *Lactobacillus*, con identidades no muy altas para Ast1, pero sí para la cepa A19, sobre todo en el caso de los genes de los *clusters* 6 y 8.

Las cepas E11, AC1 y Ast1 están descritas como productoras de dextrano (Lamas-Arriba et al., en redacción), lo cual es consecuente debido a los posibles genes *dsr* encontrados (Tabla 11). Para finalizar, mediante PCR se amplificaron en la cepa Ast1, los posibles genes *dsr*, determinándose así, según los resultados obtenidos (Figura 8), la presencia en el genoma de ambos genes *dsr*, que podrían estar implicados en la síntesis del dextrano.

6. Conclusiones & Conclusions

De los resultados obtenidos en este trabajo, se pueden extraer las siguientes conclusiones:

1. La secuenciación de nueva generación y sus tecnologías de análisis masivo, permiten, a partir de la secuenciación de un genoma, su ensamblaje y anotación, para así extraer información sobre las funciones moleculares que pueda desarrollar el organismo estudiado.

2. Haciendo uso de plataformas de análisis masivo, se puede determinar la presencia de *clusters* o genes *eps* para conocer si las cepas u organismos analizados podrían ser o no productores HePS y/o de HoPS.

3. El análisis de las posibles funciones codificadas en cada uno de los *clusters eps* identificados en las cuatro cepas BAL estudiadas en este trabajo, sugiere que la cepa A19 podría sintetizar HePS. Esta cepa presenta uno de los *clusters eps* más completos, pero carece de algunos genes fundamentales para la producción de HePS, como los que codifican las enzimas flipasa y polimerasa. En el caso de la función flipasa, esta podría codificarse en otras regiones del genoma de A19. En el caso de la actividad polimerasa, esta podría encontrarse codificada en una región del genoma no ensamblada correctamente.

4. Todas las cepas, excepto la A19, tendrían la capacidad de sintetizar HoPS, concretamente dextranos, debido a que sus genomas contienen posibles genes *dsr*. En el caso del genoma Ast1, se ha confirmado mediante PCR la presencia de los dos genes identificados.

Conclusions

From the results obtained in this work, the following conclusions are drawn:

1. New generation sequencing and massive analysis technologies allow accessible assembly, annotation and analyses of genomes, in order to extract information on the molecular functions that an organism could carry out.

2. Using massive analysis platforms, the presence of *eps clusters* and genes can be determined, and predict if the strains analysed could have the ability to produce HePS and/or HoPS.

3. The analysis of the hypothetical functions encoded by each gene within the identified *eps clusters* in the four BAL strains analysed in this work suggests that the A19 strain could synthesize HePS. This strain has one of the most complete *eps clusters*, but it lacks some fundamental genes for the production of HePS, such as those encoding the enzymes flippase and polymerase. In the case of the flippase function, this could be encoded in other regions of the A19 genome. In the case of polymerase activity, it could be encoded in a region of the genome that has not been correctly assembled.

4. All the strains, with the exception of A19, would have the ability to synthesize HoPS, specifically dextrans, due to the presence in E11, AC1 and Ast1 genomes of probable *dsr* genes. In the case of Ast1 genome, the presence of the two identified *dsr* genes has been confirmed by PCR.

BIBLIOGRAFÍA

- ¹ Zhang, H., & Cai, Y. (2014). *Lactic Acid Bacteria*. Springer
- ² Horvath, P., Coûté-Monvoisin, A. C., Romero, D. A., Boyaval, P., Fremaux, C., & Barrangou, R. (2009). Comparative analysis of CRISPR *loci* in lactic acid bacteria genomes. *International Journal of Food Microbiology*, *131*(1), 62–70.
- ³ Schroeter, J., & Klaenhammer, T. (2009). Genomics of lactic acid bacteria. *FEMS Microbiology Letters*, *292*(1), 1–6.
- ⁴ Börner, R. A., Kandasamy, V., Axelsen, A.M., Nielsen, A. T., & Bosma, E. F. (2019). Genome editing of lactic acid bacteria: opportunities for food, feed, pharma and biotech. *Cartas de microbiología de la FEMS*, *366*(1), fny291.
- ⁵ Buron-Moles, G., Chailyan, A., Dolejs, I., Forster, J., & Mikš, M. H. (2019). Uncovering carbohydrate metabolism through a genotype-phenotype association study of 56 lactic acid bacteria genomes. *Applied Microbiology and Biotechnology*, *103*(7), 3135–3152.
- ⁶ Schär-Zammaretti, P., & Ubbink, J. (2003). The cell wall of lactic acid bacteria: surface constituents and macromolecular conformations. *Biophysical Journal*, *85*(6), 4076–4092.
- ⁷ Lebeer, S., Vanderleyden, J., & De Keersmaecker, S. C. (2008). Genes and molecules of *Lactobacilli* supporting probiotic action. *Microbiology and Molecular Biology Reviews: MMBR*, *72*(4), 728–764.
- ⁸ Delcour, J., Ferain, T., Deghorain, M., Palumbo, E., & Hols, P. (1999). The biosynthesis and functionality of the cell-wall of lactic acid bacteria. *Antonie van Leeuwenhoek*, *76*(1-4), 159–184.
- ⁹ Korakli, M., & Vogel, R. F. (2006). Structure/function relationship of homopolysaccharide producing glycosyltransferases and therapeutic potential of their synthesised glycans. *Applied Microbiology and Biotechnology*, *71*(6), 790–803.
- ¹⁰ Zeidan, A. A., Poulsen, V. K., Janzen, T., Buldo, P., Derkx, P., Øregaard, G., & Neves, A. R. (2017). Polysaccharide production by lactic acid bacteria: from genes to industrial applications. *FEMS Microbiology Reviews*, *41*(Supp_1), S168–S200.

- 11 Zarour, K., Llamas, M. G., Prieto, A., Rúas-Madiedo, P., Dueñas, M. T., de Palencia, P. F., Aznar, R., Kihal, M., & López, P. (2017). Rheology and bioactivity of high molecular weight dextrans synthesised by lactic acid bacteria. *Carbohydrate Polymers*, *174*, 646–657.
- 12 Leemhuis, H., Pijning, T., Dobruchowska, J. M., van Leeuwen, S. S., Kralj, S., Dijkstra, B. W., & Dijkhuizen, L. (2013). Glucansucrases: three-dimensional structures, reactions, mechanism, β -glucan analysis and their implications in biotechnology and food applications. *Journal of Biotechnology*, *163*(2), 250–272.
- 13 de Palencia, P. F., Werning, M. L., Sierra-Filardi, E., Dueñas, M. T., Irastorza, A., Corbí, A. L., & López, P. (2009). Probiotic properties of the 2-substituted (1,3)-Beta-D-glucan-producing bacterium *Pediococcus parvulus* 2.6. *Applied and Environmental Microbiology*, *75*(14), 4887–4891.
- 14 De Vuyst, L., & Degeest, B. (1999). Heteropolysaccharides from lactic acid bacteria. *FEMS microbiology reviews*, *23*(2), 153–177.
- 15 Stingle, F., Neeser, J. R., & Mollet, B. (1996). Identification and characterization of the eps (Exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. *Journal of bacteriology*, *178*(6), 1680–1690.
- 16 Aguilar-Toalá, J.E., García-Varela, R., García, H., Mata-Haro, V., González-Córdova, A., Vallejo-Cordoba, B., & Hernández-Mendoza, A. (2018). Postbiotics: An evolving term within the functional foods field. *Trends in Food Science and Technology*, *75*, 105-114.
- 17 Nikolic, M., López, P., Strahinic, I., Suárez, A., Kojic, M., Fernández-García, M., Topisirovic, L., Golic, N., & Ruas-Madiedo, P. (2012). Characterisation of the exopolysaccharide (EPS)-producing *Lactobacillus paraplantarum* BGCG11 and its non-EPS producing derivative strains as potential probiotics. *International Journal of Food Microbiology*, *158*(2), 155–162.
- 18 Wegh, C., Geerlings, S. Y., Knol, J., Roeselers, G., & Belzer, C. (2019). Postbiotics and Their Potential Applications in Early Life Nutrition and Beyond. *International Journal of Molecular Sciences*, *20*(19), 4673.

- ¹⁹ Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: a Journal of Computational Molecular Cell Biology*, *19*(5), 455–477.
- ²⁰ Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544.
- ²¹ Cuccuru, G., Orsini, M., Pinna, A., Sbardellati, A., Soranzo, N., Travaglione, A., Uva, P., Zanetti, G., & Fotia, G. (2014). Orione, a web-based framework for NGS analysis in microbiology. *Bioinformatics* (Oxford, England), *30*(13), 1928–1929.
- ²² Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., ... Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, *9*, 75.
- ²³ Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., & Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, *44*(14), 6614–6624.
- ²⁴ Marchler-Bauer, A., & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic acids research*, *32*(Web Server issue), W327–W331.
- ²⁵ Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*, *47*(W1), W636–W641.

- ²⁶ Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., Mi, H., ... Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic acids research*, *49*(D1), D344–D354.
- ²⁷ Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A., & Eddy, S. R. (2015). HMMER web server: 2015 update. *Nucleic acids research*, *43*(W1), W30–W38.
- ²⁸ Mount D. W. (2009). Using hidden Markov models to align multiple sequences. *Cold Spring Harbor protocols*, *2009*(7), pdb. top41.
- ²⁹ Puertas, A. I. (2019). *Caracterización de los heteropolisacáridos producidos por Lactobacillus aislados de sidra natural y de las bacterias productoras: Lactobacillus collinoides y Lactobacillus sicerae*. [Tesis de Doctorado, Universidad del País Vasco].
- ³⁰ Llamas-Arriba, M. G., Puertas, A., Prieto, A., López, P., Cobos, M., Miranda, J. I., Marieta, C., Ruas-Madiedo, P., & Dueñas, M. T. (2019) Characterization of dextrans produced by *Lactobacillus mali* CUPV271 and *Leuconostoc carnosum* CUPV411. *Food Hydrocolloids*, *89*, 613-622.
- ³¹ Laws, A., Gu, Y., & Marshall, V. (2001). Biosynthesis, characterisation, and design of bacterial exopolysaccharides from lactic acid bacteria. *Biotechnology Advances*, *19*(8), 597–625.
- ³² Prechtel, R. M., Wefers, D., Jakob, F., & Vogel, R. F. (2018). Structural characterization of the surface-associated heteropolysaccharide of *Lactobacillus plantarum* TMW 1.1478 and genetic analysis of its putative biosynthesis cluster. *Carbohydrate Polymers*, *202*, 236–245.
- ³³ Puertas, A. I., Ibarburu, I., Elizaquivel, P., Zuriarrain, A., Berregi, I., López, P., Prieto, A., Aznar, R., Dueñas, M. T. (2018) Disclosing diversity of exopolysaccharide-producing lactobacilli from Spanish natural ciders. *LWT*, *90*, 469-474.

- ³⁴ Noda, M., Sugimoto, S., Hayashi, I., Danshiitsoodol, N., Fukamachi, M., & Sugiyama, M. (2018). A novel structure of exopolysaccharide produced by a plant-derived lactic acid bacterium *Lactobacillus paracasei* IJH-SONE68. *Journal of Biochemistry*, *164*(2), 87–92.
- ³⁵ Berger, B., Pridmore, R. D., Barretto, C., Delmas-Julien, F., Schreiber, K., Arigoni, F., & Brüssow, H. (2007). Similarity and differences in the *Lactobacillus acidophilus* group identified by polyphasic analysis and comparative genomics. *Journal of Bacteriology*, *189*(4), 1311–1321.

ANEXOS

ANEXO

Anexo I:

Tool Parameters

| Input Parameter | Value |
|--|--|
| Single-cell? | False |
| Run only assembly? (without read error correction) | False |
| Careful correction? | True |
| Automatically choose k- mer values | false |
| K-mers to use, separated by commas | 21,33,55 |
| Coverage Cutoff | off |
| Libraries are IonTorrent reads? | False |
| Library type | Paired-end / Single reads |
| Orientation | -> <- (fr) |
| Select file format | separate |
| Forward reads | <ul style="list-style-type: none">1: 32942_A19_1_trimmed.fastq |
| Reverse reads | <ul style="list-style-type: none">2: 32942_A19_2_trimmed.fastq |


Figura I. Parámetros seleccionados en SPAdes para la realización del ensamblaje del genoma de las cuatro cepas BAL, en este caso, la A19.

Anexo I.I:

Tool Parameters

| Input Parameter | Value |
|----------------------------------|---------------------|
| File Format | auto |
| file_count | 1 |
| Specify Files for Dataset (auto) | 1 uploaded datasets |
| force_composite | false |
| Genome | ? |
| File Format | auto |

Job Information

| | |
|----------------------|---|
| Galaxy Tool ID: | upload1  |
| Command Line | empty |
| Tool Standard Output | empty |
| Tool Standard Error | empty |
| Tool Exit Code: | 0 |
| Job API ID: | bbd44e69cb8906b53a6d66d02d8bb03b |

Dataset Storage

This dataset is stored in a Galaxy object store with id **corral-repl**.

Inheritance Chain

| |
|---|
| Galaxy5-[SPAdes_on_data_2_and_data_1_contigs_(fasta)].fasta |
|---|

Figura I.I. Parámetros seleccionados en Prokka para la realización de la anotación del genoma de las cuatro cepas BAL, en este caso, la A19.

Anexo I.II:

File formats: You can either use [FASTA](#) or Genbank format.

- If in doubt about FASTA, [this service](#) allows conversion into FASTA format.
- Due to limits on identifier sizes imposed by some of the third-party bioinformatics tools that RAST uses, we limit the size of contig identifiers to 70 characters or fewer.
- If you use GenBank, you have the option of preserving the original GenBank gene calls in the options block below. By default, genes will be recalled.

File Upload:

Sequences File Galaxy5-[SP... (fasta)] fasta

Genome information:

Taxonomy ID:

Taxonomy string:

Domain: Bacteria Archaea Virus

Genus:

Species:

Strain:

Genetic Code: 11 (Archaea, most Bacteria, most Virii, and some Mitochondria) 4 (Mycoplasmata, Spiroplasmata, Ureoplasmata, and Fungal Mitochondria)

- If you enter a valid NCBI taxonomy-ID and click "Fill in form based on NCBI taxonomy-ID," RAST will attempt to automatically fill in the form below. You may then edit any incorrect field values before going to the next step.
- If you do not know the taxonomy-ID of your genome, please leave the taxonomy-ID field blank, and fill in the fields manually.
- If you leave this field blank, RAST will fill in a dummy taxonomy string of the form "Domain; genus species strain.", based on the form entries below.
- E.g., "Escherichia". If you do not know the genus, leave blank, and it will default to "Unknown".
- E.g., "coli". If you do not know the species, leave blank, and it will default to "sp.".
- E.g., "str. K12 substr. MG1655". This field is optional. (May also be used as a comment.)

RAST Annotation Settings:

Choose RAST annotation scheme: *Choose "RASTtk" for the current modular customizable production RAST pipeline, or "Classic RAST" for the old pipeline.*

Customize RASTtk pipeline: Yes *Customize the RASTtk pipeline*

Automatically fix errors? Yes *The automatic annotation process may run into problems, such as gene candidates overlapping RNAs, or genes embedded inside other genes. To automatically resolve these problems (even if that requires deleting some gene candidates), please check this box. If you wish for the pipeline to fix frameshifts, check this option. Otherwise frameshifts will not be corrected.*

Fix frameshifts? Yes *If you wish RAST to build a metabolic model for this genome, check this option.*

Build metabolic model? Yes *If you wish debug statements to be printed for this job, check this box.*

Turn on debug? Yes *Set this to the verbosity level of choice for error messages.*

Set verbose level: *Even if this job is identical to a previous job, run it from scratch.*

Disable replication: Yes

Figura I. II. Parámetros seleccionados en RAST para la realización de la anotación del genoma de las cuatro cepas BAL, en este caso, la A19.

Anexo II:

Secuencias de los genes que sintetizan proteínas con actividad dextransacarasa y la ubicación en ellas, de las secuencias de los oligonucleótidos seleccionados para realizar la amplificación de los genes mediante PCR.

E11:

Peg.1356:

```
>fig|6666666.742337.peg.1356 hypothetical protein [Unknown sp.]
Atggatcaaataacagtagaaaggccaaaacatattttatgatggggaaggcccgaaagttaaagaaactgctcgattaggctc
ttttgagagaatttttgagcagtttgggaattccagtagttttaggaggcgaatatgtaaacgttaatgatacgtggttttttg
ttgatgatcagggaaataagattaaaggattaaatttgattgatggtaatttaccagtagttttgatcaaacctcaggcgcacaa
atcaaaggcgagtagtataaaaagcagtgcaacgactttttattttgataaaagttcaggaaacgccattccgtatattttaga
agaaaatggtaaaaaactgcttatcaggcaaatgggtcaaccagcgggctcaggctttttaccgatcaaaaataagatgattt
attactttgataagaatggttcttttgaacaggaatgcaaaaatattgatggaaaagtttattactttgacgaggatggttat
ctgattaaagatcagatagaacagtcgaaaggtaagcatatttctgaaagctggaactggaaaagttaaacggctagtttc
acagcatataccaaagcagggttaattcaagcaagctgttaacagttttgcaactcacaataatgtaaaaaaaggaatatcg
aaagttttgagacgctagatggttatttaactgcccattcttggtagcctcctaaagaaatattacttgggtggtaaaacttgg
agaaaatctaaatctgaggatttccggcactcctgatggcttgggtggccaagtaagagtattcagtcagattatgtaaacta
tatgatgaaagctggaattgtggaaagtaaatcatgtattttcaacaactgatagctgggtgatactcaatgggtgaggcgcaga
ctattcaaatgcttttagaaaaagataataactttgacaaagtctaccaattgggtacgacagttaatgcatagttttatataaa
acagaatcaatttggaaataaatcaacagaagatgtcagttcagatggattacagtcctttgcaaggtggatttttggcacaaca
gaaatgtaaaaggattcctaaagcaaaactctgattatcgcattttaggctacgcttcctcatatggttaaaaaatcatgaaaata
aaggttccgaattcttattggcaaatgatattgataattctaataccaattgttcaagcagaagagttgaattggattcattat
ttgatgaaactttgggttcaattactgcaaatgattcatctgcaaattttgatgggtattcgtgttgatgctgataaatgttga
tgcagatctactaagattgacagctgattacttcaaggatgtttatcatgtcaatgaaaatgaacaaacagcgaatgctcatt
tatcaattttagaagattggggcgggtgatgatccacagtatatatatgatcaaggtgctaatacaactaacgatggatgattat
tttgcgggcaaatgaaactcttctctagaaaaatagccaggttccgaatgatacaatggagcgttttttacaatgggtatttgat
tgaccgttctgaagatagtagtataacagcaattccgaattacacatttggtagagccatgatttcaattcacaagatc
aaatccaaacggctgttcaggctgcaactgggtggaatattgggtattttacatgggatcaacttgatcaaggtttgaagcta
tattatgaagaccaagcatcaacttctaagcaatataatcggtataatagccagtgcttatgcaactgctgtaacgaataa
agatacagttccgcatctactatggggatattgtatcttgaaggtggacaatataatggctgataagacgatttactatgata
caatcgtaacttattgacaagtcgtgttaaatacgttgcaaggtggcagacgatggcgggtgatagtaatgatattttaacc
agtggtcgttttcggaaaagggtgcaatgacgggtgacgacaggggtacggcagaaacgcggacggaaggtgctcggagtcattgt
tagcaacaatactagcctgagtttaaaagcagcgaacaggtagtagttgcatatgggagcagccattgtaatacaaacctacc
gcgagctgttctaactacaacagcaggtggttaattataacttctgatgaggggtgcaccaatgcaataactgataaat
ggggatttgatttttcaaatcatgatttactagtttaattggtaaggatgaggataatacaaggtttcaaggttatgctaacc
tgaagttacaggttatttagcagtttgggtacctgttgggtgcaacagctgagcaagatgtacgcagggcggcaacaacaacgg
cgtcaaccgatggtaagattttcactcaaatgcccactagattcaaatgttatttttgaaggatttcaaacctttattttc
tatcctgattcacatgataataatacaaatgtaattatcgccaaaaatcgtctttattcaaaagccttaggttttacaagttt
tgaattggcaccacaataccgatcaagtggtgacaagacattcttggactcaagaattgataatggctatgctttcactgac
gttatgatctcggtttttgaacaccacaaaaatattggtacagtagacgatttacgaaatgctattaaggcgttgcatgtggcc
ggattcaggtgattgctgatttgggtaccagatcagatttcaaacctgctggtgaagagttagtaacagtttaactcgggga
tgaagaaggaaatcaattagcaaaactcggagatacatgagatgctgtatggttggcaataccattgggtggcggaaaatatcaaa
agaagtaggtgggtgaattcctagacttattggaagttgaaacatccagatttatttaaatagtttcaagtagcagaggtaaa
acaattgattcaagtgcaaaagttgatcaatggatggcaaaatacttcaatgggacaaatatttttagatcgtggaatcggcta
tgttttacgtaatgacgacaatgaatatttcttttgggcaaatctaatccagttccacaaaaattattgcccaggtaaactag
agatttaa
```

GGACTGGATTAGATTCGCC

Ast1:

Peg.1686:

```
>fig|6666666.731431.peg.1686 hypothetical protein [Unknown sp.]
atggctcaattagtagtaggaacaaaagaagtcttacgggtgataagaaggtacgtgattcaatgggtgtagcaatcaatt
acaaaacactcaaaactcaggttccttcagtaggggggtgacattgtcagttccaatcagcagggttcaaatactagcgttattg
atagttctgcaatttctaaaaattttgggggtgaaatattgtaatttaaatgacacctggttttatggtgatcgacaaggtaac
aaaattaagggttaataaagggtcagcggaaatttgcatttttcgaccaaacttcaggagcacaataaaaagggtgaatgata
aaaagtggagaacactttactattttcgataaaaaattcaggaatgctattccgtacgttttaacaaaaaatggcaagataa
ctgcctatcaaaccaatggggcagcagcaggacaaggcttttttactgatcaaaaatacaaaaatttattacttttaattgatgat
ggttcatatgcaactggattaaaaatgattgctggaaaaacttattattttgatgataatggctatttggcacaagataccgt
```

agagactgttaatggtaaagcttattattccgataacacaatcggcaaagttagccatttaattttctcaaaaaatcccaaac
aaactagttcggcgagagccattatggaatttgcaagtcataacattgcaaatttctaatagacacgaaaagttttgagaatatt
gatggatatttaaccctgatcttggtagccgaccaaagtatattctatttagcggacggacatggataaaatccaagataaa
tgattttcgaccactgtaaatggtttggtagccaaataaagagtatacaagcaaattacttaaattacatgattgaacgaggaa
ttgtccaagcagtcataattctctactagtgtatgatttagttcttaataagagcagttcaaaatgtgcagattgcatta
gaaaaagatctaacgatgacgggttccaactaactggctacggcaattaatgcatagttttgttaaaacagaacctatttgaa
taaagcctcagaggatgttagttcagatggtttgcacatcattgcaaggtggcttttaacctatcaaaatagcaaatcaac
cgaagcaaattctgagatcggatcttaggttatacaccttcatacgttaaagattttgaaaataaaggttcagaattttta
ttagctaatagatatcgacaattctaatacagttgtacaagctgaagaactaaattggcttattatttaatagaattttggtc
aattacagccaatgatgctcagctaattttgatggcatcaggggtgatgacagtagataatggtgatgctgatttgcgtaaca
ttgtagtaattatttcaaggatgctcatggtgatcaaaatgataaaattgctaattcacaacttatcaattttgggaagac
tggatcactgatgatccgcaatataatttatgatcaaggtgcaaatcagttacaatggatgattatgtggctagtcaaatgaa
ctattcctggagaatgcaccagacgcaaatgataaaatggctcgtttttgcaatggattttgattaatcgggctaataata
gtactgaaaatgttgcatttccaaattatacatttgcactgacatgattccaactctcaggatcaaaatcaaaaagcaatt
caagcagctactggagaagagtatggtgtgtttacttgggagcaactgcagcaggggttagatttgcactataaggatcaagc
ttcgaccaataagcaatataatcgttacaatattccaagtaacctacgcatgcttttgactaataaggacactgttccgcggy
tttattatggtgatgtatcttgaaggtggtcagtcacatggctaataaaaccatttattatgatatacattgctaatttattg
aaaacaagaatcaagtaactgctgggtggccagaccatgctcagttgacagtaataatgattttaaactagcgttcggtttggtaa
aggatccatgactggcagtgatcaaggtactactgaaacgcgaactgaaaggtattggggttattattagtaataatcagact
tgagtttaaaagatggtgaaaaagtagtgttgcataatgggtgctgctcaccgcaatcaaaacttatcgcgacagtagtttaact
actaagaaggagtggttaactatacagcagacgaaaaatgctccaattgcttacactgatgataatggagacttacttttac
caactgcgatttagcgcataatggacaaggagaagccaatactagcgttcaaggttataataaccctgaggtaacagggtatc
tggctgtttgggttctgttggagcaaaaatcaatcaagatatacgaacagctgcaactacgaatgcttcaactgatggtaaa
acttttcaattcaaatgctgcatttggattctaatgttatttttgaaggtcttcaaaactttatttcttatccagacagcagta
cgataatgctaactgctgtaatttaccaaaatactcactatccaactttggaatttacgagctttgaatttagcaccccaat
atcgttcaagtggtgactcggactttcttggattcaaccattgataatggttatgcttctactgaccggttagcactaggtttt
ggcactccaacaaaataggcagtagatgatttacggaacgcaatataagctttgcatggtgctggaattcaagcaattgc
cgattgggtaccagatcaaatctataatttccaactgaagaactggtagcgttaatcgacagatgaaagaggcaatcaat
tatcaagttctaaaattcatgatattgtgtatgctcgaatacaattggaagcgggtgattaccaaaagaaatcggaggagt
tttctaaacttactggaattgaatatcctggtttatttgcgtgatactcaagcatcaataacaagaagatttaattcaaaagt
taagattactagttggtcgcgaagtattttaacggaacaaatatttttagggcgcggaatcgattacgttttacgtaacaatg
atcacaattatttcttgcgggtgaatcaa**ccctcttcagcataagctcttacc**agagaacttggttgctaagaattttggtt
aatga

GGTAAGAGCTTATGCTGAAGAGGGG

Peg.1246:

>fig|666666.731431.peg.1246 hypothetical protein [Unknown sp.]
atggttagttcaagagaatagaaaaagaattagctctgtgggtaatacagattgggtgaaaaacaaatattatgcaattcgt
tgactcacagtctaattggaatattaacagtgaatcta**gggtctctgatacatttgaagg**tggggcattgctttatgtaata
gtgacttgactcctgatgctaattcttcttataggttactagatagaactcccactaatcaaagcggacaaataacagataca
tcaaaacaaggcggatagaaatgctacttgctaattgatgttgataattctaatccagtagtccaggcagaaacagctaaactg
gctatattatgatgaaccttggcactatcactgcaaatgattcaagcgcgaactttgatggatagatagatagatgctgttg
ataatgtagatgctgatttattgcaaatgctgggagactattttaaaggcgcatacggaaactaacaaggtgatgctaacgct
aatagccacctttctatattggaagattgggaaactagtgattcagattacattaataaaacaaggaaataaccagataacaat
ggactttccagctcatttagcaattaaatattctttaaataatgccaacttcggaaagaagtggtatagaacctgagttacaac
ctagcttagttaacagaaccgaagattcaactgaagatgttgcacaacctaatatttcttctatcagagcacatgatagtgaa
gttcaaaactgtaattgctgcaaaattataaaggataaaataaaatcctaactcggatggttcaacagtgagtgagaatgaga
acaagcatttaagatttataatcaagatgagttgcaggctgataaaaaataactgcttataatcccagcgtcttatgcca
ttctacttagcaataaggatactgttccctcgcgtttattatggggatttattcaccgatgatggacaatataatgctgctcat
tcaccatactatgatgctatcacagctttattgaaagcagagaattaagtagctgctgctggcggacaaagtaggggtgttcaata
tgttcatggcgatgatagtagtcaagtgatagttataaaggcgttttaacatcgggttcggtatggaaaagatgcaatgactg
ccgaatccactgggtgatactgatactgaaactgaaggatggcgggttattttaaagtaacaatccagacttgaactaagta
aatgacgaaatcgttgttaatatgggtgctgctcataaaaaatcaagcctatcgtccgcttttgttaacaacaaagaaggtat
gaaaatttatacttctgatacttttgcacctgttatttataactaatgatcaagggcagttaatttttaaattccaattttattt
ttggagtaagtgatcctcaggtttcaggatattcttgcagtggtgggttccggtagggtgcaagcagcactcaggatgctaggact
gcaagcagtaacttctgtttcgcactgatggaaaaacgtatcattctaattcggcttttagattcacaagtaatttatgaaggtt
ctctaattttcaggcattccaactcaagagagcgaataatactaatgtttaaattgcacaaaatgcacaattattttaaagaat
tagggattactagttttgagttagctccgcagatcgttcaagtagtaaatagtttcttggattctgttattcaaaatggc
tatgctttactgactgttatgatgttggatataataacccaacgaaatgggaactgttgatcaattatttagcgttttgcg
agcacttcatgctcagggattcaagcaattgatgattgggttccgcatcagatataataattaccaggtgaagaaattgtga
ctgctacacgtacaaaatgggttccggtagctatgactatgattcggtaattgataatactctttatgattctcgcacagttggt
ggtggtgaatatcagtcgaatttatggtggagcatttttagataaaacttaaggaactttatccagatttggttgaaataaaaca
gatattctacgggccagcaatggaccctgatgtaaaaaataaagaatggtctgctaaatattttaaaggttcaaatattcaag
gacgtggagcagttgatgttttaaaggattggggaactaataaatattttagcgtttcaagtgataatctttatcttctctaaa
caattagaaggtgaaaaactagtcagaggttttgaggatgtcaataataatgagactgaattttataccactagcgggtatca

agctaaaagtactttttattcaagacggaataattggtactatthttgatggtaatggatataatggtaacaggagctcaaacta
taaatgataaaaactattatthtttacctaattggtatagagctacaagatgcatatthttgctggatagtaacaaagaatattat
tatgataaaaatggtaagcaaatcagggatcagttatcgaagattcaaggaggatggagatactthttcaaagatgggtc
catggctgtaggaatcacaanaattatthttcatcagatgggtgaaagtgtactcagttatthttgattctgtatgggtgcaagcca
aaggaatcagtggttaagatgaaaatggcaatacacggctactthttgatggtaaacgggtaaatatgattacaaattcctthttggg
gaattagcagatggatcgtggatatacttgggataacaacggcgtagctgtcaccgggtgaacagaagattaacgggtcagacatt
atacttcaatgctgatgggaaacaagtaagaatgatattgggttaaaaatagtgcggaagcataaattattataatggattat
cagggtgaacaaattaagaatgactthttggtgaactgccagat**gggtcatggatgtacttgg**ataacaatggagtagctgtcacg
agtagacagaaaattaatggtcaaacattatacttcaatgctgatgggaaacaagtaagaatgatattgggttaaaaat

CCAAGTACATCCATGACC

AC1:

Peg.1102:

>fig|6666666.731432.peg.1102 hypothetical protein [Unknown sp.]

gacaaaacagttgaagcaccagtaaaagatactacaacagtagcagcaacagacaaaacagttgaagcaccagtaaaagatac
tacaacagtagcagcaacagataaaaacagttgaagcaccagtagatactacaacagcaactaaa**gcagttgacacgcaaacac**
cagcgacagtttacaaaaaactgatgccacaacaacagcaacagacaaaacagttgaagaaaacttcagaacacagcaacaacg
aataaaagtgatgacgcaaaaactgaatcaagcagaagcacatcaccaaatatcaacaaaattgatggaaaacactatthttgt
aggtaatgatgggtcaaatgaagaagaattcaccgctgtcattgatggcaagtagctthttatthttgataagcaactgggtgaat
tgatatacgaatgatcatcaatattggtcaaggccttacgaatataagataataagcacaacgctgcatattcattatcatctgat
agcttcacaaatggtgatgggttacctgacagctagtagctggatcgcactaaagataattctaaagtcaggtaccacatggac
accgtcaacaggaactgatthttcgaccattggttaatgcatgggtggcctgacaaagatacacaattgcttacctcaaatata
tgcagtcactthttggctgctagaagatgatgggtcatttcaataatacggcaacaagtaatttaactaataatgcatggtca
gtccaaaaaaataattgaacagagaattagttgtctggtaatactgatthttgctaaaaacagatatacagtcattggattaagac
acaacaaaactggaatattacaagtgaaatcaaatcaaatgatcatctacaagggggagcgttggttatatgttaaacagtgaca
aaacaccagatgcaaatcaggtgaccgattatthttaaatcgcacaccgattaatcaaaaagggtcaaatcactgatthttggcaaa
caggggtggttatgagatggtgtagctaacgatgtggataattcaaatccagttggttcaagctgagcaattaaatgggtgta
ctacatgatgaacattggaactatcacacaaaatgatccaagtgctaatthttgatgggtaccaggttgatgctggtgataacg
tggatgctgacttattacaattgctagcagattacttcaagcagcttataaaatagatcaagtgatagtcaggtcaatgctg
catctctcaatthtttagaggattgggaatccaaagatgcccgttataatcaaaagcagtggttaataatcaattgcaaatggactt
tccaattcatctggattgaaatagcgttaaatgtaccagtcagcacaacagtagtggttagaacaacaaattacgaatagcc
tgggtgacgcacagaagatgcaaccgaaaatacagctcaacaaattactcgtttatccgtgacatgatagtggaagtacaa
acagttattgcacaaatgtgcaagataaaaatcaatcctaattcagatgggttaactggttacacctgatgaaatcgaaaaagc
thttgcaatatataatgcagatgaattgaaagcgggataaagaatatactgctthttcaatataccttcttcttatgcattactat
taactaataaggatacaaatcctcgtggtatattatgggtgatttatatactgatgatgggcaatatatgggtcaaaaatcacca
tattatgatgccattactcattattgcaatctcgtgtaaaaatattgctctggtgggtcaagtagggcatgcttattctaca
tgataatcaaggtgtagtacttacttccggtccgtthttggaaccgggttctatgaaagcaacagatgctgggaatgataagacacgta
cagaaggtattggctggttattagtaacaagactgatttgaatcctaataacgatgaaaaggttggtgtaaatgggtgct
gcgcaaaaaatcaagattatcgcgccttgatggttaagtactaaagatggattgaaaatataactgatgatgctggagcacc
aattggtcgcacaaatgatattgggacaattgattthttcaaatcagatattggtctacgggtatcaatgatccgcaagtttctggtt
atcttctgctgcatgggtacctggtgggtgctagcgataatcaagatgtgagaacagcaagtagtacggcaaaaatcaactgatgggt
gcaactatcattcgaatgaagcgttagattctcaagtcattatgaaggattthttcaatthttcaagccatgccaaaccaatc
cgacgagtagataactaacgtcaaaaatagtagcaaaaacggcggcattggttaagagccttggtatcacaactthttgaaatgacccgc
aatactggtcaagtagcagacacaagtttcttagattcaattggttcaaaaatgggttacgctthttagcgtatgattgatattggt
tacaatacacaacgaatattggtacagttgatcaattggtcaataacttggagggccttacatgacagaaggaattcaagccat
taacgattgggtacctgaccaaatataaatttaccaggggaacaaattgtgacagcagatcgtacaaaatagttctgaaaaag
atgacgcggattccgtcatcaaccacactttatgatattcatcaacaatcggaggtggccaatatcaggcacagtagcgtgggt
gcattcttggatcaattgaagcagactatccgactthttggttgaacaaaagcaaatthttctacaggaaagccaatgaatccaaa
tgthtaaaatcacagatgggtcagctaaagtattthttatggatccaatattcaaggtcgtggtgcttgggtatgtagtaaaagatt
gggcaacgaatcagttatthttatggtcagtagtataacgattthttacccaaaaatggttaggtgaaaagacaagcactggc
thttcaactgatgataatggtaaaaatcattthttatthttcaacaagtggttatcaagctaaagaatacctthttatcaagataatac
aaattgggtattatthttcgacaacgatggttacatggtaattgggagcgcacaacaaatgaatggtaaaaagattacttcttaccta
atggcgttgaaattacaagatgcttacttgtctgatgggcaaaaatcagattatttacaacaagactggtaaaagcagttctcaaat
caatactatthttaggacctgataataatggcgttattthttcttggctgacgggcacatggcctgggactaacaacaataactgc
agataatggatcgacgtctgagcaatattthttgatcagaatggcattcaggttaagggaaactgaaattagagatactgccccga
atthttacgatattthttcgatgggtcagacaggttaattctattgactaatacctggcaggaattagcaaaaataatcatggattthttg
aatgatcaaggaattgctgtagctgggtcaacaaaacatcaacgggtcaaatggtgattthttcaattatgatggcatgcaaattaa
aaacgattthttcaaacgthttggataatgggtgcatgggtatacttgatgacaaagcgttgctgtagcaggggaacagacaatthtt
atggcaaaaagtagtactthttatcagaatggcgaacaagtcaaaataacaaagttaaaataagtagtagcacaataaagctat
tatgcaggtattatgggtgaaaaactaacaatgactthttggtgaattaccagatgggttcttgggttatthttggataatcaggg
tcatgccataacaggtgaacaaataatthttatcaagtcctthttacttcaatgataacgggtatacaattgaagggcggtagc
atattgatgaagcaggtcatatgactattatgat**gctgattctggcgagaggg**tttcgaaccaattthttgacaaaattaacggg
caaaaagttacttccgataatgcaggcaatgctgtagattga

CCCTCTGCCAGAATCAGC

Peg.278:

>fig|6666666.731432.peg.278 hypothetical protein [Unknown sp.]
atgagaaataaagaaacgacttgtcgttaaaaaattatataagt**ccggcaaatatgggttactg**cggggattgttatcatcatt
agggcttggtttctgtaccacaagtttatgctgatacaacacagagtggacaaattgctcaagtaaaagtgtacacctgata
atccgaaaagtgttcagaatagcgtatgattatattgtcacaattgacgtaccgaaagataaaaaagtaaccggacaataat
cgtgctggttgggtggtcaagcgttgcacaaatcaagatcaatcaaaattaagtagtgacgataaattctgctaataaaattga
taaaagtagtcaagtagctgatgctaacaaaactgaacaaaagatacagttacaataatagtagtagaccaataaacgatg
atgattctgttaacaaagcagataaaaaatggtcaagttgctgccaacgatatacagaaggcaaccaagataataacccgacg
ccgttacgagataagaatagtgctgatgatagagtagaagagctgatcaggtagatggtggtgtaaaaccagatgagaatca
agaccctgcaactagtcagtaatagattcaaacagtgatcaagacaacaaaaagaacaagttccaataaatctatctcaa
cgccacaagatgagaaacaaacggcatcactcactcagcaccataataattctaaaaatattcaaacgatttgatggtaag
acatattatattgacgataatgggtcaaaaaagaagaattttacaacaatagtcgatggacatgtattgtatttcgacaaaaga
taatggttttcttgtaccaaccaatgattatcaatttaagaaggattgacaagtcacaaaataaattttaccacacataatg
ctgtacatgggtgatactgcagataaatttactgaaattgatggctatttaacagcggatagttgggtatcgacctaggatata
ttagttaatggtaagatggacatcctcgtccaataatgattttcgtccacttttgatgacttgggtggcctgataaaagtac
tcaaatatactatcttaattatgaaaagcattcggttaagtaatagtcagttgaatattctaatcaagatagtcacaaagt
tcttaaatcaagccgcacaggatgtacaaattaaaaatagaacaaaagattagccacgatgggcaaacctgagtggtcaaaaaat
aatgtttctgattttgttaaacgtcaaccctaattggaaatattgcccagtaaatatgcaacaactggaaaatgataaagatcct
acaagtggtgacattattatgtaaatagtgataaaacacctgatgcaaatcagattaccgacttttgatcgtacaccaa
caaatcaaaaagggtattattcttactgacccaactcaaggtggttatgatttcttattagctaatgatggtgataat
tcaaacctgctgttcaagccgaacaataaattggatgtattatctatataaatttggggcaattaccaatcaagatgctga
cgcaaatttgatagatatacgtggtgatgctggtgataatgctgatgctgatttattacaattgcaacagattatatagaaag
cagcatatggtggtgataaaagtgcagtggtgctaatacagcatttgcacattttagaagattggagtgataacgatgcagaa
tatgtcaagataatcatgataaccaattatcaatggataacaaattacgcttgcgtaagatttctttgacctgcccac
ggctgatcaaacggcaatagtggttctgtagcgggttggaaaccactaattactaataagcttagttgaccgtacaatagacacaa
ctaacaatgtagcagtcacaaattattcatttctgctgacatgatagtgaaagtacaaaccgtaattgctgaaattatcaag
caacgaattgatcctaattcagatgggttaacgccaacctggatcagttaacacaggcatttaaaatctacaatgcccgatca
attaaaaacaaataaagagtttacacaatataatattcctagcacaatgctgctattgctaaccaataaagacacagtagccg
gagcatattatggcgatctctttaccgatgatggccaatataatggcaacaaaatcaccgactatgatgcaattaacacattg
ttgcagctcctgtttaaagtatggtgctgggtcaacaaatggcgatgcattataggtggtgatagcagtagcctgcaaa
tgggaccggtggcctactgactctgtacgttatggtaaaagcgcctatgacagcagacgagcaggggagtaattgacacaaga
cacaggggatagctgttatcgaggctaataatcctgacctgaagttaaatcaactcgatcaagtagttgtaaaaaatgggagct
gccataaaaaatcaagcttatcgaccagtagtactcactacaaaaaatggcattggccacataataatctgattcggatggttag
tgatgcattaataaaaatatacgaataataatgggtgaactaacttttgatgcttctgaaataaaaggggtttgcaaatcgcaag
tttctggtcatttagctgtttgggtaccagttggggctgataataaacaacacagcagttaccgaaagttattctgtgatgaa
gttaaaaaatggacaacatttcattcaaatgctgcactagattctcaggttaatttacgaaggattctcaaattttcaatcttt
tccaactaatgaagatgaatagctaatgtgattatgcgcaaaaattcagatttatataagagctgggggattactagttttg
agtttccaccacaatatcgactcaagtagtgaaggagattctggattctattatccaaaatgggttagctttacagatcgc
tatgatctaggatttgatacaccaacaaaatattgggacaattgaccaactacgtgatgcaattaaatcactgcaccaagcagg
aattcaagcagtaggctgatgggtacctgatcaaatatataatttgactggacaacaaattgtaactgcgcaacgtgttaata
attcgggcatttataatgaaaactctgtgattaataaaacgctttatgacgcccacaaactgtagggtggcgagaatatcaaaaa
ttataggtggcgaattcttagacacaattaaaaaattatccatcattatttacaactaagcaaatctactgggggtacc
catggatccaagtgaaaagatgaaagaatggtcagcaagatattttaatggtagtaataacagggacgtggtgcttattatg
tcttaaaagattgggtaccaatgattatttcaaggtgagttcatcaaatcaaaatcaagcatttttacctaaccagttactt
aatcaaacctgctagtagcgggttttggtaatgataataaaggcatgacttattttcaactagtggtaccagcaaaaagatac
gtttatccaagatgaacaatcaaatgggtactactttaaataaaaaatgggtatattgacttatggatttcaaaaaatgaatggtt
ttgggtactattttttgccaatgggtatgaattgcaagatgcttatatggaaaatagtaacggaaatgctctattatttcaat
aagcaaggtgaagcaattgattaatgctctattttatggatgataataagcaatggcactattttgataagaatgggtactatggc
aaacaaaggcatgacaacaatacacattaatgggtgaagaatagtagcagatttttaaccaagatgggtgacaagcaaaaagatc
agtttataactactagtgatggtaagatgattatttctctgatgattctggggatagtagtaatacaatttaaaaaatga
gctgacgggtcatgggttactttggttcagatggtagtgcaattaaagggttctcagcaaataaaggacaagcactttattt
tgatgatgattatcgtcaaataaaggacgcaaaaagttgatacacaaggcttaatacgttattatcatgctgattccggcg
aatgatagttaatcgatttgaaaaactgctgacgggtcatgggctacttccggggtcgatgggtggtgctgcaagggtaac
cgaatcattgaagggcaaaatcttttcttgaagataattatagtcaggtcaagggaaaaacgggtgactgatcagcaaggcag
aattcgttactatgatgcccattcaggtgaaatgatagtaaacggtttcgaaaaactgctaacggttcatgggcttattttg
gtgcagatgggtgtagcgggttaaagggtatcaaacgattaatgggtcagcaactttactttgatgataaattatgctcaataaaa
ggtaacaagattggtgatgtacaaggtcacttacgattctttgatgctgattccgggtgagatgggtgtaatacaatttagaca
actgtcagatggttcatgggcttattttggtgcagatgggtgtagcgggttaaaaggcatcaaacgattaatggtaaaaagatgt
attttgataataatgctcaagtaaaaggtaagaagttattgatagccaaggatcagtagcattattatgatgctgattcc
ggtagatgggtgtaatacagtcaggaattatcgagcggcatttttgatattgatggtaaaagcatga

CCATGTGCCGTCGATAATTGC

Anexo III:

Contigs obtenidos a partir del ensamblaje en la cepa A19 y sus longitudes (pb). A partir del Node (Contig 47) comienzan los contigs a considerarse muy cortos (zona blanca).

| #name | length | | | | |
|---------|--------|-----------------------|-------------------|-----------------|------------|
| NODE_1 | 232311 | NODE_42 | 1199 | NODE_83 | 692 |
| NODE_2 | 219150 | NODE_43 | 1130 | NODE_84 | 683 |
| NODE_3 | 202484 | NODE_44 | 1079 | NODE_85 | 670 |
| NODE_4 | 153641 | NODE_45 | 1069 | NODE_86 | 663 |
| NODE_5 | 129121 | NODE_46 | 1026 | NODE_87 | 661 |
| NODE_6 | 118910 | <u>NODE_47</u> | <u>978</u> | NODE_88 | 660 |
| NODE_7 | 116569 | NODE_48 | 958 | NODE_89 | 660 |
| NODE_8 | 106759 | NODE_49 | 938 | NODE_90 | 658 |
| NODE_9 | 91834 | NODE_50 | 920 | NODE_91 | 658 |
| NODE_10 | 81485 | NODE_51 | 914 | NODE_92 | 658 |
| NODE_11 | 74356 | NODE_52 | 883 | NODE_93 | 657 |
| NODE_12 | 72110 | NODE_53 | 879 | NODE_94 | 655 |
| NODE_13 | 69156 | NODE_54 | 877 | NODE_95 | 650 |
| NODE_14 | 68149 | NODE_55 | 876 | NODE_96 | 648 |
| NODE_15 | 63247 | NODE_56 | 864 | NODE_97 | 647 |
| NODE_16 | 62339 | NODE_57 | 838 | NODE_98 | 644 |
| NODE_17 | 58740 | NODE_58 | 835 | NODE_99 | 644 |
| NODE_18 | 58083 | NODE_59 | 831 | NODE_100 | 642 |
| NODE_19 | 39412 | NODE_60 | 811 | NODE_101 | 641 |
| NODE_20 | 38609 | NODE_61 | 797 | NODE_102 | 640 |
| NODE_21 | 31999 | NODE_62 | 795 | NODE_103 | 633 |
| NODE_22 | 20484 | NODE_63 | 792 | NODE_104 | 631 |
| NODE_23 | 19929 | NODE_64 | 788 | NODE_105 | 630 |
| NODE_24 | 18307 | NODE_65 | 785 | NODE_106 | 625 |
| NODE_25 | 16614 | NODE_66 | 779 | NODE_107 | 623 |
| NODE_26 | 15633 | NODE_67 | 778 | NODE_108 | 621 |
| NODE_27 | 10445 | NODE_68 | 773 | NODE_109 | 621 |
| NODE_28 | 7068 | NODE_69 | 771 | NODE_110 | 617 |
| NODE_29 | 4732 | NODE_70 | 766 | NODE_111 | 614 |
| NODE_30 | 4011 | NODE_71 | 763 | NODE_112 | 610 |
| NODE_31 | 3846 | NODE_72 | 762 | NODE_113 | 609 |
| NODE_32 | 3329 | NODE_73 | 760 | NODE_114 | 606 |
| NODE_33 | 2600 | NODE_74 | 751 | NODE_115 | 605 |
| NODE_34 | 2559 | NODE_75 | 744 | NODE_116 | 604 |
| NODE_35 | 2538 | NODE_76 | 741 | NODE_117 | 603 |
| NODE_36 | 2367 | NODE_77 | 734 | NODE_118 | 600 |
| NODE_37 | 2050 | NODE_78 | 724 | NODE_119 | 593 |
| NODE_38 | 1907 | NODE_79 | 714 | NODE_120 | 589 |
| NODE_39 | 1746 | NODE_80 | 708 | NODE_121 | 586 |
| NODE_40 | 1406 | NODE_81 | 702 | NODE_122 | 584 |
| NODE_41 | 1227 | NODE_82 | 698 | NODE_123 | 584 |

| | | | | | |
|----------|-----|----------|-----|----------|-----|
| NODE_124 | 583 | NODE_171 | 492 | NODE_218 | 467 |
| NODE_125 | 582 | NODE_172 | 490 | NODE_219 | 466 |
| NODE_126 | 578 | NODE_173 | 490 | NODE_220 | 466 |
| NODE_127 | 577 | NODE_174 | 490 | NODE_221 | 464 |
| NODE_128 | 575 | NODE_175 | 487 | NODE_222 | 464 |
| NODE_129 | 575 | NODE_176 | 484 | NODE_223 | 464 |
| NODE_130 | 574 | NODE_177 | 481 | NODE_224 | 463 |
| NODE_131 | 571 | NODE_178 | 481 | NODE_225 | 463 |
| NODE_132 | 569 | NODE_179 | 480 | NODE_226 | 463 |
| NODE_133 | 566 | NODE_180 | 480 | NODE_227 | 461 |
| NODE_134 | 564 | NODE_181 | 479 | NODE_228 | 460 |
| NODE_135 | 560 | NODE_182 | 479 | NODE_229 | 459 |
| NODE_136 | 560 | NODE_183 | 478 | NODE_230 | 458 |
| NODE_137 | 559 | NODE_184 | 478 | NODE_231 | 458 |
| NODE_138 | 558 | NODE_185 | 477 | NODE_232 | 458 |
| NODE_139 | 558 | NODE_186 | 476 | NODE_233 | 457 |
| NODE_140 | 557 | NODE_187 | 475 | NODE_234 | 457 |
| NODE_141 | 552 | NODE_188 | 475 | NODE_235 | 456 |
| NODE_142 | 549 | NODE_189 | 474 | NODE_236 | 456 |
| NODE_143 | 545 | NODE_190 | 474 | NODE_237 | 454 |
| NODE_144 | 544 | NODE_191 | 474 | NODE_238 | 454 |
| NODE_145 | 542 | NODE_192 | 473 | NODE_239 | 453 |
| NODE_146 | 541 | NODE_193 | 473 | NODE_240 | 452 |
| NODE_147 | 535 | NODE_194 | 473 | NODE_241 | 452 |
| NODE_148 | 532 | NODE_195 | 472 | NODE_242 | 452 |
| NODE_149 | 530 | NODE_196 | 472 | NODE_243 | 451 |
| NODE_150 | 528 | NODE_197 | 472 | NODE_244 | 451 |
| NODE_151 | 525 | NODE_198 | 472 | NODE_245 | 451 |
| NODE_152 | 524 | NODE_199 | 472 | NODE_246 | 449 |
| NODE_153 | 523 | NODE_200 | 472 | NODE_247 | 449 |
| NODE_154 | 521 | NODE_201 | 471 | NODE_248 | 449 |
| NODE_155 | 521 | NODE_202 | 471 | NODE_249 | 448 |
| NODE_156 | 520 | NODE_203 | 471 | NODE_250 | 448 |
| NODE_157 | 517 | NODE_204 | 471 | NODE_251 | 447 |
| NODE_158 | 516 | NODE_205 | 470 | NODE_252 | 447 |
| NODE_159 | 515 | NODE_206 | 470 | NODE_253 | 447 |
| NODE_160 | 510 | NODE_207 | 470 | NODE_254 | 447 |
| NODE_161 | 509 | NODE_208 | 469 | NODE_255 | 447 |
| NODE_162 | 508 | NODE_209 | 469 | NODE_256 | 446 |
| NODE_163 | 505 | NODE_210 | 469 | NODE_257 | 446 |
| NODE_164 | 505 | NODE_211 | 468 | NODE_258 | 446 |
| NODE_165 | 503 | NODE_212 | 468 | NODE_259 | 445 |
| NODE_166 | 501 | NODE_213 | 467 | NODE_260 | 445 |
| NODE_167 | 498 | NODE_214 | 467 | NODE_261 | 443 |
| NODE_168 | 494 | NODE_215 | 467 | NODE_262 | 443 |
| NODE_169 | 494 | NODE_216 | 467 | NODE_263 | 442 |
| NODE_170 | 493 | NODE_217 | 467 | NODE_264 | 442 |

| | | | | | |
|----------|-----|----------|-----|----------|-----|
| NODE_265 | 441 | NODE_312 | 418 | NODE_359 | 395 |
| NODE_266 | 440 | NODE_313 | 417 | NODE_360 | 394 |
| NODE_267 | 440 | NODE_314 | 417 | NODE_361 | 393 |
| NODE_268 | 440 | NODE_315 | 417 | NODE_362 | 393 |
| NODE_269 | 439 | NODE_316 | 417 | NODE_363 | 393 |
| NODE_270 | 438 | NODE_317 | 416 | NODE_364 | 393 |
| NODE_271 | 438 | NODE_318 | 416 | NODE_365 | 392 |
| NODE_272 | 437 | NODE_319 | 416 | NODE_366 | 391 |
| NODE_273 | 437 | NODE_320 | 416 | NODE_367 | 390 |
| NODE_274 | 436 | NODE_321 | 416 | NODE_368 | 390 |
| NODE_275 | 435 | NODE_322 | 416 | NODE_369 | 389 |
| NODE_276 | 435 | NODE_323 | 415 | NODE_370 | 388 |
| NODE_277 | 433 | NODE_324 | 415 | NODE_371 | 386 |
| NODE_278 | 433 | NODE_325 | 414 | NODE_372 | 386 |
| NODE_279 | 433 | NODE_326 | 414 | NODE_373 | 386 |
| NODE_280 | 432 | NODE_327 | 414 | NODE_374 | 385 |
| NODE_281 | 432 | NODE_328 | 413 | NODE_375 | 384 |
| NODE_282 | 430 | NODE_329 | 413 | NODE_376 | 383 |
| NODE_283 | 430 | NODE_330 | 412 | NODE_377 | 383 |
| NODE_284 | 429 | NODE_331 | 412 | NODE_378 | 382 |
| NODE_285 | 429 | NODE_332 | 412 | NODE_379 | 382 |
| NODE_286 | 428 | NODE_333 | 412 | NODE_380 | 381 |
| NODE_287 | 428 | NODE_334 | 411 | NODE_381 | 380 |
| NODE_288 | 427 | NODE_335 | 411 | NODE_382 | 378 |
| NODE_289 | 426 | NODE_336 | 411 | NODE_383 | 378 |
| NODE_290 | 426 | NODE_337 | 409 | NODE_384 | 377 |
| NODE_291 | 426 | NODE_338 | 407 | NODE_385 | 377 |
| NODE_292 | 425 | NODE_339 | 406 | NODE_386 | 376 |
| NODE_293 | 425 | NODE_340 | 405 | NODE_387 | 376 |
| NODE_294 | 425 | NODE_341 | 405 | NODE_388 | 375 |
| NODE_295 | 425 | NODE_342 | 405 | NODE_389 | 373 |
| NODE_296 | 425 | NODE_343 | 404 | NODE_390 | 372 |
| NODE_297 | 424 | NODE_344 | 404 | NODE_391 | 372 |
| NODE_298 | 424 | NODE_345 | 403 | NODE_392 | 372 |
| NODE_299 | 424 | NODE_346 | 403 | NODE_393 | 371 |
| NODE_300 | 423 | NODE_347 | 402 | NODE_394 | 371 |
| NODE_301 | 423 | NODE_348 | 401 | NODE_395 | 371 |
| NODE_302 | 423 | NODE_349 | 401 | NODE_396 | 370 |
| NODE_303 | 422 | NODE_350 | 400 | NODE_397 | 370 |
| NODE_304 | 422 | NODE_351 | 400 | NODE_398 | 368 |
| NODE_305 | 421 | NODE_352 | 399 | NODE_399 | 367 |
| NODE_306 | 421 | NODE_353 | 398 | NODE_400 | 364 |
| NODE_307 | 421 | NODE_354 | 398 | NODE_401 | 363 |
| NODE_308 | 420 | NODE_355 | 398 | NODE_402 | 362 |
| NODE_309 | 419 | NODE_356 | 396 | NODE_403 | 362 |
| NODE_310 | 418 | NODE_357 | 396 | NODE_404 | 362 |
| NODE_311 | 418 | NODE_358 | 395 | NODE_405 | 361 |

| | | | | | |
|----------|-----|----------|-----|----------|-----|
| NODE_406 | 361 | NODE_453 | 318 | NODE_500 | 111 |
| NODE_407 | 358 | NODE_454 | 316 | NODE_501 | 111 |
| NODE_408 | 357 | NODE_455 | 316 | NODE_502 | 111 |
| NODE_409 | 357 | NODE_456 | 316 | NODE_503 | 109 |
| NODE_410 | 356 | NODE_457 | 314 | NODE_504 | 109 |
| NODE_411 | 355 | NODE_458 | 314 | NODE_505 | 108 |
| NODE_412 | 355 | NODE_459 | 313 | NODE_506 | 108 |
| NODE_413 | 354 | NODE_460 | 313 | NODE_507 | 107 |
| NODE_414 | 353 | NODE_461 | 312 | NODE_508 | 107 |
| NODE_415 | 352 | NODE_462 | 312 | NODE_509 | 103 |
| NODE_416 | 348 | NODE_463 | 311 | NODE_510 | 103 |
| NODE_417 | 348 | NODE_464 | 311 | NODE_511 | 103 |
| NODE_418 | 347 | NODE_465 | 309 | NODE_512 | 103 |
| NODE_419 | 346 | NODE_466 | 307 | NODE_513 | 103 |
| NODE_420 | 346 | NODE_467 | 307 | NODE_514 | 95 |
| NODE_421 | 345 | NODE_468 | 305 | NODE_515 | 94 |
| NODE_422 | 345 | NODE_469 | 298 | NODE_516 | 91 |
| NODE_423 | 344 | NODE_470 | 297 | NODE_517 | 91 |
| NODE_424 | 343 | NODE_471 | 294 | NODE_518 | 89 |
| NODE_425 | 342 | NODE_472 | 292 | NODE_519 | 89 |
| NODE_426 | 342 | NODE_473 | 267 | NODE_520 | 89 |
| NODE_427 | 341 | NODE_474 | 219 | NODE_521 | 77 |
| NODE_428 | 341 | NODE_475 | 219 | NODE_522 | 77 |
| NODE_429 | 340 | NODE_476 | 210 | NODE_523 | 77 |
| NODE_430 | 340 | NODE_477 | 194 | NODE_524 | 75 |
| NODE_431 | 339 | NODE_478 | 172 | NODE_525 | 72 |
| NODE_432 | 337 | NODE_479 | 171 | NODE_526 | 72 |
| NODE_433 | 337 | NODE_480 | 155 | NODE_527 | 72 |
| NODE_434 | 337 | NODE_481 | 153 | NODE_528 | 63 |
| NODE_435 | 336 | NODE_482 | 149 | NODE_529 | 61 |
| NODE_436 | 335 | NODE_483 | 149 | NODE_530 | 60 |
| NODE_437 | 334 | NODE_484 | 141 | NODE_531 | 59 |
| NODE_438 | 331 | NODE_485 | 127 | NODE_532 | 58 |
| NODE_439 | 330 | NODE_486 | 121 | NODE_533 | 56 |
| NODE_440 | 328 | NODE_487 | 120 | NODE_534 | 56 |
| NODE_441 | 328 | NODE_488 | 117 | | |
| NODE_442 | 327 | NODE_489 | 117 | | |
| NODE_443 | 326 | NODE_490 | 114 | | |
| NODE_444 | 325 | NODE_491 | 112 | | |
| NODE_445 | 325 | NODE_492 | 111 | | |
| NODE_446 | 324 | NODE_493 | 111 | | |
| NODE_447 | 323 | NODE_494 | 111 | | |
| NODE_448 | 323 | NODE_495 | 111 | | |
| NODE_449 | 323 | NODE_496 | 111 | | |
| NODE_450 | 321 | NODE_497 | 111 | | |
| NODE_451 | 319 | NODE_498 | 111 | | |
| NODE_452 | 319 | NODE_499 | 111 | | |

