



**UNIVERSIDAD  
DE ANTIOQUIA**

**CONSTRUCCIÓN Y ANÁLISIS DEL  
PANGENOMA DE *Xanthomonas campestris*: UNA  
CEPA INDUSTRIAL IMPORTANTE EN LA  
PRODUCCIÓN DE BIOPOLIMEROS**

Autor(es)

Alejandra Betancur Sánchez

Universidad de Antioquia

Facultad de Ingeniería, Departamento de Ingeniería  
Química

Medellín, Colombia

2020



Construcción y análisis del pangenoma de *Xanthomonas campestris*: una cepa industrial importante en la producción de biopolímeros”

Alejandra Betancur Sánchez

Trabajo de grado  
como requisito para optar al título de:  
Ingeniería Química

Director: Rigoberto Ríos Estepa  
Co-Director: Carlos Andrés Caicedo Montoya, MSc.

Universidad de Antioquia  
Facultad de Ingeniería, Departamento de Ingeniería Química  
Medellín, Colombia  
2020.

## Resumen

En la actualidad la genómica comparativa desempeña un papel fundamental para el estudio de los genomas microbianos y el entendimiento en la producción de metabolitos secundarios, de interés en la industria química y biotecnológica. Esta área de estudio permite comparar diferencias y similitudes genéticas a gran escala entre diferentes especies, y entre miembros de la misma especie, que difieren en función o fenotipo. Al hacerlo, es posible estudiar la variación del contenido genético, la función genética, sus relaciones y la manera como inciden en la síntesis de metabolitos de interés industrial. En este informe se presenta el análisis genómico de la especie *Xanthomonas campestris* el cual incluye el tamaño y características del pangenoma y el genoma central de quince cepas, así como los genes implicados en la producción del metabolito secundario de gran interés industrial, goma xantana. El análisis incluyó el uso del software BPGA y Micropan. Adicionalmente, haciendo uso de la plataforma antiSMASH, se encontraron clúster o grupos de genes como potenciales candidatos para la producción de metabolitos de interés industrial, por ejemplo, xantoferrina, xantomonadina, bacteriocina y péptidos de lazo. Se observó además que los genes involucrados en la síntesis de la goma xantana que se conservan para las cepas analizadas son *gumD*, *gumH*, *gumI* y *gumK*. Estos resultados permitirán proponer alternativas de mejoramiento de los bioprocesos a partir de la generación de organismos super-productores, haciendo uso de técnicas moleculares, o mediante el desarrollo de diseños experimentales que consideren la manipulación de las condiciones medioambientales, por ejemplo, pH, temperatura y composición del medio de cultivo, con el fin de eventualmente inducir la expresión de los genes de interés en la cepa objetivo, y con ello obtener mejores rendimientos a escala productiva.

## Contenido

1. Introducción .....	4
2. Objetivos.....	6
2.1 General	6
2.2 Específicos	6
3. Marco Teórico.....	7
3.1 <i>Xanthomonas campestris</i> .....	7
3.2 Goma xantano .....	8
3.3 Genómica comparativa - Pangenoma.....	16
4. Metodología.....	18
4.1 Genomas utilizados en este estudio .....	18
4.2 Determinación del pangenoma .....	20
5. Resultados y análisis .....	21
5.1 Resultados obtenidos con BPGA .....	21
5.2 Resultados de MICROPAN.....	34
5.3 Resultados obtenidos en ANTISMASH.....	37
5.4 Resultados MAUVE .....	39
5.5 <i>Propuesta para el mejoramiento en la producción de goma xantana</i>	39
<b>6. Conclusiones</b> .....	<b>40</b>
<b>7. Referencias Bibliográficas</b> .....	<b>41</b>
ANEXOS.....	45
Anexo 1. Descripción de los programas usados en este estudio .....	45
Anexo 2. Resultados de antiSMASH para cada cepa.....	49
Anexo 3. Resultados MAUVE.....	53

## 1. Introducción

*Xanthomonas campestris* es una bacteria que causa diferentes fitopatologías, pero que a nivel industrial es usada en la producción de goma xantana, un polisacárido extracelular con propiedades importantes en el sector alimenticio, por ser viscosificador de soluciones acuosas (Ochoa et al., 2000).

El costo, las propiedades reológicas y el rendimiento de esta goma dependen del medio y las condiciones de cultivo utilizadas en el proceso de fermentación. Por ejemplo, la relación Carbono / Nitrógeno debe considerarse en el proceso de producción del biopolímero; la biomasa está limitada por la concentración de nitrógeno disponible y la tasa de producción de xantana está influenciada por la concentración de carbono. Por lo tanto, los nutrientes requeridos por el microorganismo son diferentes durante el bioproceso, convirtiéndose en necesaria su división en dos etapas: crecimiento celular y producción de la goma (Carignatto et al., 2011).

Varios autores verificaron que el rendimiento y la tasa de producción del biopolímero se ven favorecidos con el aumento de la proporción de glucosa / extracto de levadura en el medio de cultivo. Los almidones parcialmente hidrolizados que se han utilizado en la producción de goma xantana son de arroz hidrolizado, cebada y harina de maíz, suero ácido, melaza de caña de azúcar, mezcla de manosa y glucosa, pulpa de remolacha y pulpa de durazno. Así, se ha reportado que los rendimientos y las cualidades de la goma de xantano son competitivos, sin embargo, la glucosa proporciona mejores rendimientos y calidad del producto (Rosalam & England, 2006).

Algunas de las posibles razones que podrían causar el bajo rendimiento y calidad de la goma son (Rosalam & England, 2006):

- I. La deficiencia de ciertos grupos funcionales en la fuente de carbono lo que resulta en diferentes reacciones a las de la vía metabólica deseada.
- II. La composición de nutrientes varía debido a las diferentes fuentes de carbono utilizadas, por lo tanto, la calidad de la goma de xantano producida también varía.
- III. Formación de otros subproductos.
- IV. Un bajo rendimiento causado por una alta concentración de compuestos que no reaccionan dificultando el proceso posterior a la purificación.

La mayor parte de la investigación previa sobre la producción microbiana de xantano se ha centrado en el tipo de fuente de carbono, los requisitos nutricionales de los microorganismos y la optimización de las condiciones de crecimiento (Niknezhad et al., 2016).

En la revisión bibliográfica se evidencia que la mayoría de trabajos han sido orientados a la optimización proceso fermentativo y de manera particular en la

selección del medio cultivo. Sin embargo, son pocos los trabajos que se enfocan en el estudio de los procesos metabólicos del sistema celular y sus mecanismos de regulación. Las herramientas en genómica permiten comprender mejor las rutas metabólicas asociadas con la producción de metabolitos secundarios y biopolímeros, lo que permitiría un estudio a nivel molecular de la relación entre el medio ambiente de la bacteria, es decir, medio y condiciones de cultivo, y el desempeño del sistema celular. Como fin último, sería posible contribuir a proponer soluciones a las problemáticas planteadas previamente, que limitan la producción y a su vez, estudiar el potencial de la bacteria *Xanthomonas campestris*, para la biosíntesis de metabolitos secundarios de interés en la industria química.

## **2. Objetivos**

### **2.1 General**

Obtener y caracterizar el pangenoma de la bacteria *Xanthomonas campestris*, como herramienta para contribuir al entendimiento del metabolismo celular y los genes implicados en su regulación, con el propósito de plantear estrategias de mejoramiento de las condiciones de cultivo en la síntesis de biopolímeros (goma xantana) y estimar su potencial de producción de nuevos metabolitos.

### **2.2 Específicos**

A través de diferentes metodologías bioinformáticas:

- Identificar el genoma accesorio, los genes únicos y el "Core" genoma de esta especie.
- Estimar la conservación de genes relacionados con la producción de la goma xantana.
- Identificar clúster asociados con la síntesis de metabolitos secundarios y proponer una estrategia de mejoramiento de su producción.

### 3. Marco Teórico

Los biopolímeros se definen como macromoléculas que son sintetizadas a través de una amplia variedad de sistemas biológicos y que son susceptibles a la descomposición primaria a través de la acción enzimática de ciertos microorganismos. En las últimas décadas han sido considerados de gran interés en la industria un grupo de biopolímeros denominados exopolisacáridos de origen microbiano especialmente por parte de las industrias farmacéutica, de alimentos y del petróleo, por las aplicaciones que se les ha encontrado en estos campos. Su uso se fundamenta en dos de sus características más significativas: la capacidad para alterar las propiedades de flujo del agua y la posibilidad de formar geles.

Los microorganismos producen polisacáridos estructurales, como el péptido glucano, que compone la pared celular, polisacáridos intracelulares de reserva energética y polisacáridos extracelulares como el xantano. Los polisacáridos se clasifican en homopolisacáridos y heteropolisacáridos, en función del tipo de unidades de glucosa que contienen (Kariduraganavar et al, 2014).

Los exopolisacáridos permiten a las células adherirse a superficies mediante interacciones electrostáticas. También en ocasiones funcionan como factores de virulencia en los organismos patógenos (Criollo, 2019).

#### 3.1 *Xanthomonas campestris*

*Xanthomonas campestris* (*X. Campestris*) es un fitopatógeno que pertenece a la familia *Pseudomonadaceae*. Es un bacilo Gram negativo con unas dimensiones de 0.4 a 0.7 micrómetros de ancho y 0.7 a 1.8 micrómetros de longitud. Presenta un flagelo polar que le confiere movilidad y es quimiorganotrofo y aerobio obligado. *X. campestris* es capaz de utilizar una gran variedad de carbohidratos y sales de ácidos orgánicos. En medio de cultivo sólido, *X. Campestris* forma colonias amarillentas, lisas, redondas y viscosas. La coloración es debida a la producción de xantomonadinas, que son pigmentos derivados bromados de aril-octanos (Palaniraj & Jayaraman, 2011).

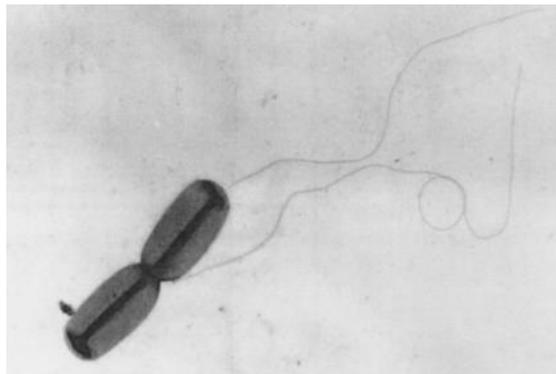


Figura 1. Micrografía electrónica de transmisión de *X. campestris*. Tomado de García & Santos (2000)

Distintas especies de *Xanthomonas* atacan e infectan a una larga selección de plantas incluidas algunas de interés industrial como la col, la alfalfa, así como cultivos de frijol (García & Santos, 2000).

La estructura de la envoltura celular es similar a la de las otras células Gram negativas. Los pigmentos amarillos están presentes en todas las especies de *Xanthomonas*, pero pueden estar ausentes especialmente cuando ocurre la degradación de la cepa. Todos los medios empleados para el crecimiento de *X. campestris* son medios complejos. El medio de cultivo más utilizado es el agar YM, el cual se emplea para aislar microorganismos acidúricos a partir de diversas muestras y que se compone de digerido enzimático de gelatina, extracto de levadura, extracto de malta y dextrosa. *X. campestris* se ha cultivado a diferentes temperaturas que van de 25 a 30 ° C, siendo su temperatura optima de 28°C (Tobón et al, 2008).

### 3.2 Goma xantano

La goma xantana o xantano es un exopolisacárido producido por bacterias del género *Xanthomonas* y de clase Gamma proteobacteria, consiste en un heteropolisacárido cuya estructura es una cadena de celulosa sustituida cada dos unidades de D-glucosa, por una cadena de trisacáridos de D-manosa y ácido D-glucurónico en una relación molar de 2: 2: 1 y proporciones variables de residuos de O-acetilo y piruvilo. Su peso molecular es de  $1.25 \cdot 10^6$  y  $10 \cdot 10^6$  g/mol (Palaniraj & Jayaraman, 2011).

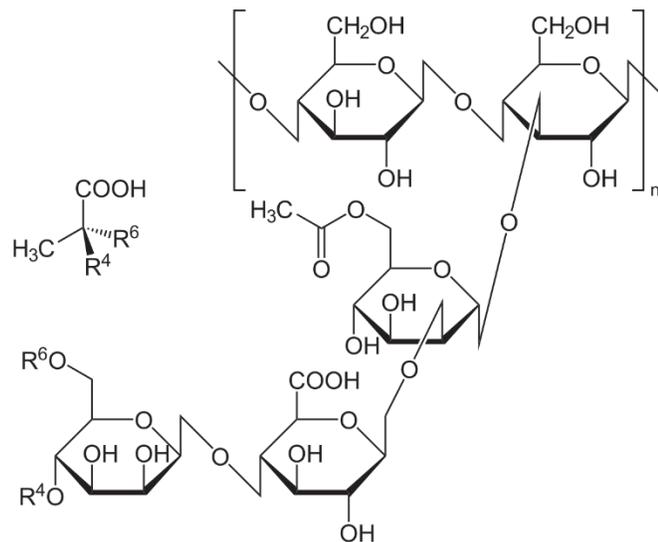


Figura 2. Estructura de la goma Xantana

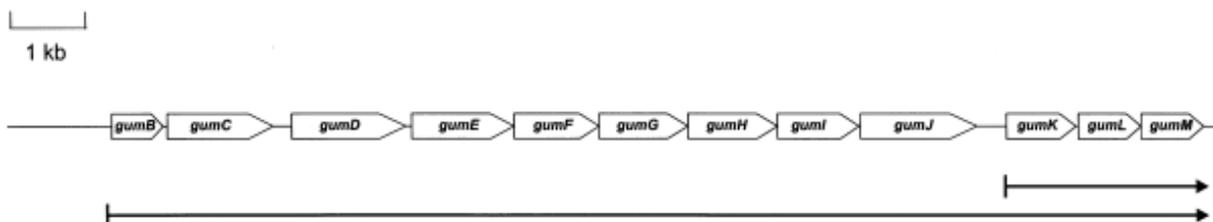
*X. campestris* produce este polisacárido cuando se encuentra sobre una especie vegetal, lo que origina un fluido viscoso que marchita tallo y frutos, debido a que impide la circulación de agua. La goma de xantano provee un reservorio a la

célula para colonizar, adherirse, además de brindarle protección frente a virus, anticuerpos y agentes antimicrobianos.

La síntesis de la unidad de repetición comienza con la transferencia de glucosil 1-fosfato de UDP-glucosa a poliisoprenol fosfato, seguido de la transferencia secuencial de los otros residuos de azúcar para formar la unidad de repetición completa. Los residuos de acetilo y piruvilo se agregan a nivel de pentasacárido unido a lípidos, donados por acetil-CoA y fosfoenolpiruvato respectivamente. El proceso de polimerización indica que las cadenas de xantano crecen en el extremo reductor (Becker et al., 1998).

Becker et al., (1998) propusieron una ruta de síntesis de la goma xantana, basada en la familia de genes *gum* que consiste en las siguientes etapas: la proteína GumD cataliza la adición de glucosa 1- fosfato al poliisoprenol fosfato. Esta reacción reversible es el primer paso en la biosíntesis de intermediarios ligados a lípidos involucrados en la síntesis de xantano. GumM cataliza la adición de una  $\beta$ -1,4-glucosa, seguida de la adición interna de la  $\alpha$ -1,3-manosa por GumH, un ácido  $\beta$ -1,2-glucurónico por GumK, y la terminal  $\beta$ -1,4-manosa por GumI. La proteína GumL incorpora residuos de piruvilo a la  $\beta$ -manosa externa, mientras que los residuos de acetilo se incorporan a la  $\alpha$ -manosa interna por GumF, y a la  $\beta$ -manosa externa por GumG. Además del pentasacárido de poliisoprenil difosfato, el trisacárido unido a lípidos puede actuar como sustrato para GumF. Sin embargo, el acetil trisacárido unido a lípidos no puede actuar como un aceptor de un residuo de ácido glucurónico, lo que sugiere que los residuos de acetilo se incorporan al polímero a través de la unidad de repetición unida a lípidos. A continuación, se presenta la vía metabólica de producción de la goma xantana tomada de Becker et al. (1998).

A)





*glucoquinasa GK, fosfoglucoisomerasa PGI, pirofosforilasa UDPG-PP UDP-glucose, glucoisomerasa UDPG-deH UDP-glucose. (Becker et al., 1998).*

El gen *xanA* codifica una enzima con actividades de fosfoglucomutasa y fosfomannomutasa, *xanB* codifica para una enzima bifuncional con actividades de fosfomanosa isomerasa y PIB-manosa pirofosforilasa (Fig. 2). Por lo tanto, una parte de esta región codifica dos enzimas, una de las cuales es necesaria para la síntesis de los tres precursores de nucleótidos de azúcar de xantano, y la otra para la síntesis de GDP-manosa. A diferencia de otros sistemas biosintéticos de exopolisacáridos, la región de genes que están relacionados con la síntesis de xantano está desvinculada de las requeridas para la síntesis de precursores de nucleótidos de azúcar. El análisis de la secuencia de nucleótidos predijo la presencia de 12 marcos de lectura abiertos (*gumB* a *gumM*) (Becker et al., 1998).

El Xantano se usa en la industria de alimentos como aderezo para ensaladas, en jarabes, condimentos y salsas, productos lácteos y productos horneados. También en la industria farmacéutica en cremas y suspensiones y en productos cosméticos como champús y lociones, otorgando estabilidad (García y Santos, 2000). La goma xantana se utiliza como estabilizador, espesante y emulsionante que se usa ampliamente en las industrias farmacéutica y cosmética. Las propiedades pseudoplásticas de esta goma permiten que las pomadas mantengan su forma y se extiendan fácilmente (Palaniraj & Jayaraman, 2011).

*Tabla 1. Ejemplos de aplicaciones industriales, alimentarias y farmacéuticas de xantano. Tomado de Becker et al., 1998.*

<b>Aplicaciones industriales</b>	<b>Aplicaciones en la industria de alimentos y la industria farmacéutica</b>
<ul style="list-style-type: none"> <li>• Abrasivos (control de viscosidad)</li> <li>• Esmaltes cerámicos, esmaltes, pinturas tixotróficas (estabilización, pseudoplasticidad)</li> <li>• Explosivos (gelificantes)</li> <li>• Fire-uids (estabilizador de espuma)</li> <li>• Fracturas hidráulicas (viscosidad / reticulación)</li> <li>• Productos químicos de lavandería</li> <li>• Productos químicos agrícolas incluidos herbicidas, pesticidas, fertilizantes y fungicidas (suspensión)</li> <li>• Lodos de perforación de petróleo (control de viscosidad, cizallamiento)</li> <li>• Teñido de textiles (pseudoplasticidad)</li> <li>• Clarificación del agua (Floculante)</li> </ul>	<ul style="list-style-type: none"> <li>• Cerveza (estabilizador de espuma)</li> <li>• Queso (inhibidor de sinéresis)</li> <li>• Confitería (recubrimiento)</li> <li>• Helado (estabilizador, control de cristalización)</li> <li>• Glaseados y esmaltes (adhesivo)</li> <li>• Mermeladas, salsas (agente espesante)</li> <li>• Bebidas (suspensión)</li> <li>• Alimentos para mascotas (agente aglutinante)</li> <li>• Productos farmacéuticos (liberación retardada de fármacos)</li> <li>• Saborizantes en polvo (encapsulación)</li> <li>• Aderezo para ensaladas, rellenos de panadería (agente emulsionante)</li> <li>• Envolturas de embutidos (formación de película)</li> <li>• Jarabes (pseudoplasticidad)</li> </ul>

Los exopolisacáridos bacterianos como el xantano proporcionan una alternativa a los polisacáridos de origen vegetal. Sin embargo, los productos de polisacáridos bacterianos son relativamente caros en comparación con los polisacáridos vegetales a granel, como el alginato procedente de plantas y almidón. Una ventaja de xantano es que la calidad de este producto bacteriano puede garantizarse mediante el uso de cepas de producción específicas y las condiciones de fermentación (Becker et al., 1998).

El éxito comercial de xantano no solo se basa en sus propiedades reológicas, sino también en factores económicos. *X. campestris* puede utilizar una amplia gama de sustratos diferentes. La tasa de conversión del sustrato es alta, sin embargo, la producción de esta goma necesita un cuidadoso monitoreo y control de las condiciones debido a su baja concentración y alta viscosidad, así al comportamiento reológico (Kuppuswami, 2014).

### **Producción de la goma xantana a escala industrial**

La producción a escala industrial de xantano se lleva a cabo utilizando sustratos y nutrientes de bajo costo. Fuentes de carbohidratos como sacarosa, melaza de caña de azúcar y suero. Las fuentes de nitrógeno inorgánico como las sales de amonio o nitrato son adecuadas, y una amplia variedad de fuentes complejas de nitrógeno como el extracto de levadura, la peptona de harina de soja y el suero de soja también son útiles para la producción de xantano. La producción beneficiosa se logra utilizando granos de cereales. Generalmente se favorece el cultivo discontinuo con medios complejos. Sin embargo, se pueden usar medios sintéticos simples. El cultivo de crecimiento por lotes requiere 2 días. El mantenimiento adecuado del cultivo stock de *X. campestris* es importante por su consistencia en la producción de xantano. Durante la fase de crecimiento inicial, la acumulación de polisacáridos comienza y continúa después del crecimiento. El pH disminuye durante la fermentación debido a la formación de ácidos orgánicos. Si el pH cae por debajo de 5.0, la formación de xantano se reduce drásticamente. Por lo tanto, es necesario controlar el medio de fermentación al pH óptimo de 7.0. otra variable importante de controlar en la producción de la goma es la agitación del medio, ya que es útil para mejorar la velocidad de transporte de nutrientes a través de la membrana celular, lo que a su vez apoya la tasa de crecimiento del microorganismo (Palaniraj & Jayaraman, 2011).

En la producción de exopolisacáridos, además de las condiciones de operación, el pH del cultivo y la fuente y los niveles del sustrato de carbono juegan un papel importante en las productividades de fermentación generales (Papagianni et al.,

2001). La concentración de la fuente de carbono afecta la eficiencia con la que se convierte en polisacárido.

Las principales operaciones involucradas en la producción microbiana de xantano son las siguientes:

- Organismo y preparación del inóculo
- Preparación de medio de cultivo
- Fermentación
- Procesamiento aguas abajo

Como se ha dicho a lo largo de este documento, la goma xantana es producida por la bacteria *X. campestris*, mantenido en un medio de sacarosa-triptona y extracto de levadura (STYA). Se necesita un inóculo de 5 a 10% para una producción óptima de polisacáridos. Los inóculos se preparan transfiriendo células desde el medio STYA (a 28 ° C) a un tubo que contiene 7 ml de medio STYA (pH 7,0), e incubando a 28 ° C y 160 rpm durante 24 h. El inóculo se transfiere luego a matraces de 250 ml y, 24 h después, a los fermentadores de laboratorio (Kuppuswami, 2014).

Los componentes de los medios utilizados en la industria son principalmente económicos y complejos, siendo materias primas naturales. El agua corriente generalmente se usa para la dilución. La fuente de carbono puede ser glucosa, sacarosa o almidón, en el rango de concentración de 1 a 5%. Concentraciones superiores al 5% tienden a inhibir tanto el crecimiento como la producción de goma de xantano. En la fermentación la reología dinámica del caldo tiene una profunda influencia en las tasas de biorreacción, el consumo de energía, la transferencia de calor y masa, y la mezcla. La cinética está influenciada significativamente por variaciones espaciales en las concentraciones de sustrato, biomasa y el propio polímero, que son atribuibles a la reología del polímero y su producción (Kuppuswami, 2014).

En conclusión, el rendimiento y las propiedades del producto están influenciados por la cepa microbiana utilizada, el medio de crecimiento y otros factores ambientales. La recuperación del producto se complica por la alta viscosidad del caldo. Las propiedades de las soluciones de xantano se ven afectadas por la temperatura de disolución, la temperatura de medición y la presencia de otros polímeros no xantanos (García et al., 2000).

A continuación, se presenta una breve revisión de otros metabolitos secundarios producidos por *X. campestris*.

### *Xantoferrina*

Los sideróforos son pequeñas moléculas orgánicas producidas por microorganismos en condiciones limitantes de hierro que mejoran la absorción de hierro en los microorganismos. En el medio ambiente, la forma férrica de hierro es insoluble e inaccesible a pH fisiológico. Bajo esta condición, los microorganismos sintetizan sideróforos que tienen alta afinidad por el hierro férrico. Estos complejos férrico-sideróforos de hierro se transportan luego al citosol. En el citosol, el hierro férrico se reduce a hierro ferroso y se vuelve accesible para los microorganismos. En los últimos tiempos, los sideróforos han llamado mucho la atención debido a sus roles potenciales en diferentes campos. Los sideróforos tienen aplicación en la ecología microbiana para mejorar el crecimiento de varios microorganismos no cultivables y pueden alterar las comunidades microbianas. En el campo de la agricultura, los diferentes tipos de sideróforos promueven el crecimiento de varias especies de plantas y aumentan su rendimiento al mejorar la absorción de Fe por las plantas. Los sideróforos actúan como un posible agente de biocontrol contra fitopatógenos nocivos y tienen la capacidad de sustituir pesticidas peligrosos. Las muestras contaminadas con metales pesados pueden desintoxicarse aplicando sideróforos, lo que explica su papel en la biorremediación. Los sideróforos pueden detectar el contenido de hierro en diferentes entornos, exhibiendo su papel como biosensor. En el campo de la medicina, el sideróforo utiliza la "estrategia del caballo de Troya" para formar complejos con antibióticos y ayuda en el suministro selectivo de antibióticos a las bacterias resistentes a los antibióticos. Ciertas enfermedades de sobrecarga de hierro, por ejemplo, la anemia falciforme, pueden tratarse con la ayuda de sideróforos. Otras aplicaciones médicas de sideróforos incluyen actividad antipalúdica, eliminación de elementos transuránicos del cuerpo y actividad anticancerígena (Saha et al., 2016).

También, los sideróforos y sus derivados tienen una gran aplicación en la agricultura para aumentar la fertilidad del suelo y como biocontrol para patógenos fúngicos en plantas (Sah & Singh, 2015).

### *Xantomonadina*

Las xantomonadinas son pigmentos amarillos unidos a la membrana que típicamente son producidos por la bacteria fitopatógena *Xanthomonas* spp. Una investigación considerable ha revelado que son un grupo único de pigmentos halogenados, aril-poliene, insolubles en agua. Se ha demostrado que las xantomonadinas desempeñan papeles importantes en la supervivencia epifítica y las interacciones huésped-patógeno en el fitopatógeno *Xanthomonas campestris* pv. *campestris* (He et al., 2020). Las xantomonadinas protegen a las *Xanthomonas* contra el daño fotooxidativo y la oxidación lipídica y por lo tanto

juegan un papel en el establecimiento o mantenimiento de una relación estrecha entre las bacterias y sus plantas hospedadoras, He et al. (2020) han demostrado que el mecanismo biosintético de aril polieno se conserva entre estos diversos taxones bacterianos. Por otro lado, estos investigadores concluyen que los mecanismos biosintéticos detallados de las xantomonadininas y los pigmentos similares a las xantomonadininas todavía no se comprenden completamente. En consecuencia, se necesitan más análisis genéticos y bioquímicos de grupos de genes biosintéticos de pigmentos y sus productos.

### *Bacteriocina*

Las bacteriocinas son péptidos antimicrobianos, que pueden considerarse seguros ya que pueden ser fácilmente degradados por enzimas proteolíticas del tracto gastrointestinal de los mamíferos, siendo atractivos en la industria láctea para extender la vida útil y evitar el deterioro de los productos lácteos. Como no plantean problemas de riesgo para la salud, las bacteriocinas, ya sea purificadas o excretadas por cepas productoras de bacteriocina, son una gran alternativa al uso de conservantes químicos en los productos lácteos (Silva et al., 2018).

Una estrategia común para la conservación de los alimentos que se comen crudos o sin cocción adicional es la aplicación de películas o recubrimientos comestibles que contienen sustancias antimicrobianas. La incorporación de compuestos antimicrobianos como las bacteriocinas en recubrimientos y películas comestibles se presenta como una alternativa interesante para garantizar el control de microorganismos patógenos en productos alimenticios. Los recubrimientos y películas comestibles están compuestos de capas delgadas de biopolímeros que modifican la atmósfera circundante de los alimentos, formando una barrera entre los alimentos y el medio ambiente, mejoran la seguridad, la calidad y la funcionalidad de los productos alimenticios sin cambiar las propiedades organolépticas y nutricionales (Valdés et al., 2017).

### *Lassopéptido*

Los lassopéptidos o péptidos lazo son una clase de productos naturales sintetizados en el ribosoma y modificados después de su translación. Los péptidos lazos se caracterizan tanto por la selectividad como por la potencia de las proteínas biológicas más grandes, pero con poca o baja inmunogenicidad, estabilidad y biodisponibilidad de moléculas pequeñas. Además, la diversa funcionalidad de los péptidos de lazo y su extraordinaria estabilidad frente a la degradación química, térmica y proteolítica los convierten en candidatos atractivos para el descubrimiento de fármacos (Martin & Tulla, 2018).

### 3.3 Genómica comparativa - Pangenoma

El mejoramiento continuo de cepas productoras de bio-productos de importancia a nivel industrial requiere una aproximación multinivel para el estudio, tanto a nivel de las vías metabólicas que producen dichos bio-productos, como el entendimiento de las redes de regulación y de señalización que modulan la síntesis dependiendo de las condiciones medioambientales y el estadio de crecimiento del organismo bajo consideración. Esta aproximación multinivel se ve beneficiada por las diferentes ciencias "ómicas" (genómica, transcriptómica, proteómica, metabolómica, etc.) que brindan enormes cantidades de datos sobre las especies de interés y que permiten detallar de forma precisa las interacciones entre las condiciones de cultivo y fermentación, la expresión génica, la producción de proteínas y enzimas y por último la adecuada síntesis del producto de interés. Por lo tanto, una primera aproximación en la caracterización de las capacidades metabólicas de una especie determinada y su adaptación a diferentes nichos, es el uso de técnicas de genómica comparativa que permitan caracterizar todo el acervo genético presente en una especie; adicionalmente, debido a la gran disponibilidad de datos de genomas en las bases de datos públicas de bioinformática que existen en el mundo, es posible extender estos estudios de genómica comparativa, hacia análisis más globales, como lo es el caso de la caracterización del pangenoma de una especie, es decir el conjunto global de todos los genes que conforman esta especie (aunque la definición de puede extender a otros niveles de la clasificación taxonómica como géneros y familias), considerando las cepas conocidas de la misma y con el propósito de identificar las diferencias fenotípicas, esto es por ejemplo, su adaptación a diferentes ambientes o su capacidad de producir diversos metabolitos, a partir del genotipo de cada cepa.

Por lo tanto, desde la propuesta para el estudio del pangenoma, se han utilizado activamente un gran número de herramientas computacionales; a fines de 2014, Panseq y PGAP (*Pan-Genomes Analysis Pipeline*) se clasificaron como los dos paquetes más populares según las citas acumuladas de publicaciones científicas revisadas por pares. Las funciones de los paquetes de software, aunque variables entre ellas, incluyen categorizar genes ortólogos, calcular perfiles pangenómicos, integrar anotaciones genéticas y construir filogenias (Xiao et al., 2015). Actualmente, existen muchas otras herramientas desarrolladas que permiten analizar un pangenoma desde diferentes aproximaciones como el contenido proteico, el contenido de dominios o el contenido de regiones reguladoras de los genomas.

Según la definición previa, el perfil genético (contenido) de un pangenoma se divide en tres grupos: núcleo (compartido por todos los genomas), genes prescindibles y específicos de cepa (o aislado). Los avances recientes en las tecnologías de secuenciación de nueva generación y la metagenómica han llevado a un cambio de paradigma en la genómica microbiana de pocas comparaciones genómicas a estudios pangenómicos a gran escala en diferentes escalas de resolución filogenética (Tettelin et al., 2005).

Un pangenoma se puede definir como abierto o cerrado (infinito o finito), de acuerdo con la capacidad de la especie para adquirir ADN exógeno, es decir genes adicionales que implicarían tener la maquinaria para usarlos y poseer una gran cantidad de ARN ribosomal para interpretar dicha información, lo que en el caso del pangenoma cerrado no ocurriría. La naturaleza abierta o cerrada de un pangenoma está ligada al estilo de vida de las especies bacterianas estudiadas, por lo cual, las especies alopatricas que viven aisladas en un nicho estrecho generalmente tienen un genoma pequeño y un pangenoma cerrado, porque están especializadas, mientras que especies simpátricas, que viven en una comunidad, tienden a tener grandes genomas y un pangenoma abierto, además de una alta transferencia horizontal de genes y varios operones ribosómicos (Rouli et al., 2015). Además, según lo citado por Dagan y Martin (2006), un árbol basado en un solo gen o en genes que codifican proteínas ribosomales completas es demasiado simplista y no representa la realidad. Por el contrario, el estudio de pangenomas con diferentes herramientas puede ayudar a definir las especies (Rouli et al., 2015).

El concepto de pangenoma tiene consecuencias importantes para la forma en que entendemos la evolución bacteriana, la adaptación y la estructura de la población, así como para cuestiones más aplicadas, como el diseño de vacunas o la identificación de genes de virulencia (Mira et al., 2010). En síntesis, los estudios de pangenoma se han vuelto casi esenciales para las para explorar el genoma bacteriano y entender el funcionamiento metabólico del mismo para aprovecharlo en el ámbito industrial.

En el Anexo 1, se hace una descripción de los programas utilizados en este estudio.

De acuerdo con lo expuesto anteriormente, hace mucho tiempo se reconoció que la diversidad entre las cepas procariotas es mucho mayor que entre, por ejemplo, individuos en una subpoblación humana. Todos los humanos comparten, más o menos, los mismos genes, y las diferencias entre los individuos a menudo se encuentran en las partes reguladoras del genoma. En contraste, dos cepas diferentes de *E. coli* pueden ser muchas veces más diversas que los humanos y los

chimpancés, con tal vez el 20% de diferencia (Snipen & Liland, 2015). La genómica comparativa microbiana se ha centrado en la mayoría de los casos en los genes centrales, es decir, los genes conservados en todas las cepas. En la pangenómica se consideran todos los genes, también los denominados genes accesorios. En todos los casos, un "gen" es en realidad una familia de genes o un grupo de secuencias similares, y solo son considerados los genes codificadores de proteínas. Una familia de genes suele ser sinónimo de una colección de ortólogos, es decir, genes que desempeñan papeles idénticos en diferentes genomas (Snipen & Liland, 2015).

#### **4. Metodología**

Se describe cada uno de los pasos realizados y técnicas empleadas para el desarrollo del proyecto. Se redacta en pasado.

##### *4.1 Genomas utilizados en este estudio*

Se seleccionaron 15 genomas de la especie *Xanthomonas campestris* en los formatos que se requieren para el análisis en los softwares a usar (FASTA, Genbank, GFF) disponibles en la base de datos *Assembly* del Centro Nacional para la Información Biotecnológica (NCBI) del Instituto Nacional de Salud de los Estados Unidos.

El criterio de selección fue escoger solo aquellos genomas que estuvieran reportados como completos con el propósito de garantizar una buena calidad en el análisis. Para evitar sesgos por las diferentes metodologías de anotación utilizadas en los trabajos de investigación que generaron estos genomas, se decidió hacer las anotaciones propias de cada uno utilizando Prokka (*Prokaryotic genome annotation*) disponible en Galaxy (<https://usegalaxy.eu/>). (Cuccuru et al, 2014; Seemann, 2014).

Las indicaciones en Prokka para los archivos de salida fueron: base de datos *Xanthomonas campestris* anotaciones de tRNA, rRNA, ncRNAs, archivo FASTA de proteínas de las secuencias CDS traducidas (.faa), anotación en formato GFF3, que contiene tanto secuencias como anotaciones (.gff), archivo GenBank estándar, con un registro para cada secuencia (.gbk), archivo FASTA con anotación en la secuencia de nucleótidos, anotaciones en formato tabular, incluidos los COG y estadísticas relacionadas con las características anotadas encontradas.

Las características generales de las cepas y genomas analizados se informan en la Tabla 2.

Tabla 2. Información sobre las cepas y genomas de *Xanthomonas campestris* investigadas en el presente estudio

CEPA	ORIGEN	ORF*	% GC	TAMAÑO (BP)	NÚMERO DE ACCESO	REFERENCIA
ATCC 33913	<i>Brassica y Arabidopsis</i>	4067	65.7	5076188	GCF_000007145.1	da Silva et al., 2002
8004	<i>Brassica</i>	4355	65.1	5148708	GCF_000012105.1	Quian et al., 2005
CN18	<i>Brassica juncea var. Foliosa</i>	4315	65.1	5261244	GCF_002776835.1	Denancé et al., 2018
CN17	<i>Brassica rapa subsp. Chinensis</i>	4499	65	5122268	GCF_002776715.1	Denancé et al., 2018
CN12	<i>Brassica napus subsp. Oleifera</i>	4351	65.1	5115137	GCF_002776775.1	Denancé et al., 2018
B100	<i>Brassica</i>	4368	65.1	5079002	GCF_000070605.1	Vorhölter et al., 2008
3811	<i>Brassica oleracea</i>	4341	65.2	5072566	GCF_002879955.1	Kong et al., 2019
MAFF302021	<i>Brassica oleracea</i>	4148	65.4	5048651	GCF_009177345.1	Takeuchi & Mitsuahara (2020).
CN03	<i>Brassica rapa subsp. Pekinensis</i>	4138	65.4	5111737	GCF_002776735.1	Denancé et al., 2018
CN14	<i>Brassica juncea var. Foliosa</i>	4393	65.1	5082555	GCF_000401735.2	Denancé et al., 2018
CN15	<i>Brassica rapa subsp. Chinensis</i>	4368	65.2	5114892	GCF_000403575.2	Denancé et al., 2018
pv. raphani 756C	<i>Brassica spp., Raphanus sativus, Capsicum annuum, Lycopersicon esculentum, Nicotiana tabacum</i>	4136	65.4	4941214	GCF_000221965.1	Bogdanove et al., 2011
ICMP 21080	<i>Brassica</i>	4310	65.1	4911121	GCF_001186415.1	Desai et al., 2015
ICMP 4013	<i>Brassica oleracea var. Capitata</i>	4344	65.1	4908383	GCF_001186465.1	Desai et al., 2015
pv. badrii	<i>Xanthium strumarium, Pisum sativum</i>	4287	65.2	4955824	GCF_012848175.1	Patel et al., 1950

\*ORF: Open Reading Frame - marco abierto de lectura.

## 4.2 Determinación del pangenoma

### **Determinación del pangenoma a través de BPGA y MICROPAN**

BPGA usa USEARCH como la herramienta de agrupación de ortólogos predeterminada. En este estudio se eligieron porcentajes de similitud de 70% y 90%, ya que en la literatura se sugieren dichos porcentajes para hacer comparación genómica a nivel de especie (Barh et al., 2020). Con la información de los clúster se determinó la matriz de pangenoma basado en la presencia (1) o ausencia (0) de genes. A partir de esta información se clasificaron los genes que pertenecen al genoma central (presentes en las 15 cepas seleccionadas), genoma accesorio (presentes entre 2 y 14 cepas) o al genoma único (genes presentes en una sola cepa). Adicionalmente tomando 20 genes aleatorios del Core o genoma central se construyó la filogenia la cual es más robusta que la rRNA 16S porque tiene más marcadores. La secuencia de rRNA 16S se ha utilizado durante mucho tiempo sin sentido crítico como un reloj molecular para inferir las relaciones filogenéticas entre los procariotas sin dilucidar completamente los cambios evolutivos que experimenta esta molécula, por lo que la funcionalidad de 16S rRNA que utilizan un entorno experimental diferente, necesita ser probada, puesto que los genes 16S rRNA tienen una capacidad de evolución neutral universal. Es indudablemente cierto que los organismos evolucionan siguiendo el modelo evolutivo en forma de árbol, pero las moléculas o ensamblajes (es decir, los genomas) de los mismos no necesariamente siguen este modelo; de hecho, a menudo lo violan (Tsukuda et al., 2017).

BPGA utiliza el algoritmo de Neighbor-Joining para generar los árboles filogenéticos (Chaudhari et al., 2016). También con el propósito de caracterizar la funcionalidad del pangenoma se hicieron anotaciones para los cluster determinados utilizando la base de datos Kegg y se hicieron anotaciones de COG.

Adicionalmente se caracterizó el pangenoma basado en la presencia de dominios en las proteínas usando modelos ocultos de Markov a través del software HMMER para contrastar con la metodología de BPGA, lo anterior se llevó a cabo utilizando el paquete R, micropan versión 1.2.

Con lo anterior se determinó si el pangenoma de la especie es abierto o cerrado, su tamaño y su fluidez y variabilidad a través de las distintas cepas analizadas.

### **Caracterización de metabolitos secundarios con ANTISMASH**

Con esta herramienta se buscaron y caracterizaron los grupos de genes para la producción eventual de metabolitos secundarios, con lo que se podría revelar el

potencial de *Xanthomonas campestris*, para producir nuevos compuestos de interés a nivel industrial.

Para llevar a cabo este análisis se ingresó al siguiente link <https://antismash.secondarymetabolites.org/#!/start>, una vez cargada la secuencia en formato GenBank, se seleccionó detección estricta y activación de las características extra.

### **Alineación múltiple de secuencia genómica conservada con reordenamientos con MAUVE**

Una vez instalado el software MAUVE disponible en <http://darlinglab.org/mauve/download.html>, se procedió a cargar las 15 secuencias de las cepas de estudio en formato gbk para hacer el alineamiento, los parámetros del software se dejaron por defecto.

## **5. Resultados y análisis**

### **5.1 Resultados obtenidos con BPGA**

#### **Características genómicas generales**

Los genomas de las cepas de *X. campestris* analizadas en este estudio muestran heterogeneidad en el tamaño de sus genomas. El genoma más grande lo tiene la cepa 8004 con un tamaño de 5148708 bp, mientras que los genomas más pequeños pertenecen a las cepas pv. raphani 756C, con un tamaño de 4941214 bp, e ICMP 21080, con un tamaño de 4911121 bp. El tamaño promedio de las 15 cepas trabajadas es de 5063299 bp. Son muchos los mecanismos mutacionales que pueden producir cambios en el tamaño del genoma. Algunos de ellos ocurren a gran escala (duplicación de todo el genoma), mientras que otros ocurren a una escala muy pequeña (pérdida o ganancia de unos pocos nucleótidos). Para el caso de estudio, esto cobra relevancia en el hecho de a futuro, poder establecer las relaciones en cuanto al estilo de vida de este microorganismo, el hábitat y sus mecanismos de patogenicidad

El contenido G+C varía de 65.1% (en la mayoría de las cepas) a 65.7 % (para la cepa ATCC 33913). El contenido promedio de G+C es de 65.2%.

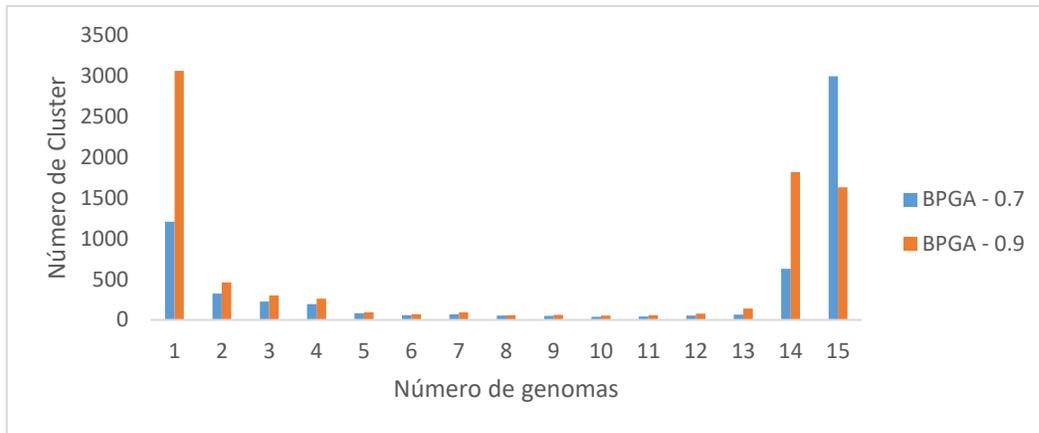
#### **Análisis del pangenoma en BPGA**

El pangenoma de la especie *X. campestris* es abierto de acuerdo al análisis realizado en BPGA con un porcentaje de similitud de 70% y 90% entre las cepas bajo estudio. La distribución de los genes que se comparten entre las cepas, los

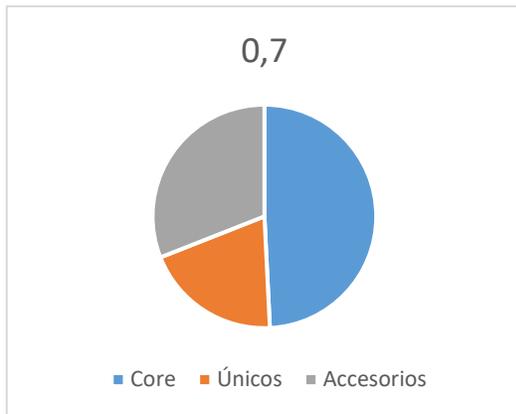
genes únicos y los genes accesorios se muestran en la Figura 4, de acuerdo con el porcentaje de similitud.

Con 70% de similitud se tienen 1206 genes únicos y 2997 genes que se comparten entre las 15 cepas. A su vez, con 90% de similitud se tiene 3064 genes únicos y 1632 genes que se encuentran en todas las cepas analizadas.

A)



B)



C)

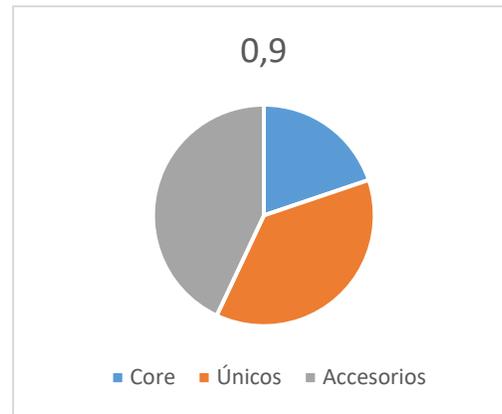
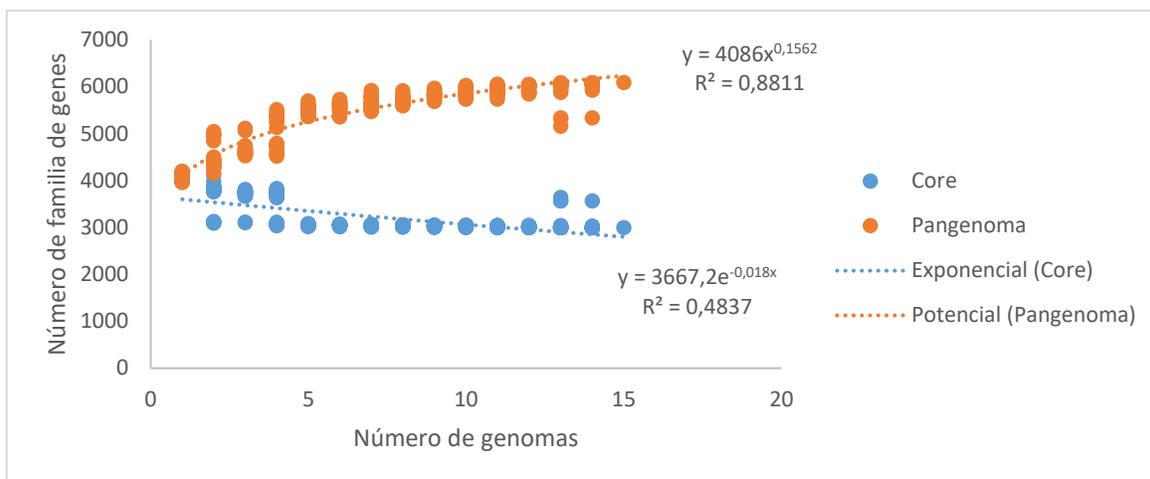


Figura 4. Análisis del pangenoma. A) Número de genomas versus número de clúster. Porcentaje de genes del core, únicos y accesorios con B) 70% y C) 90% de similitud

A)



B)

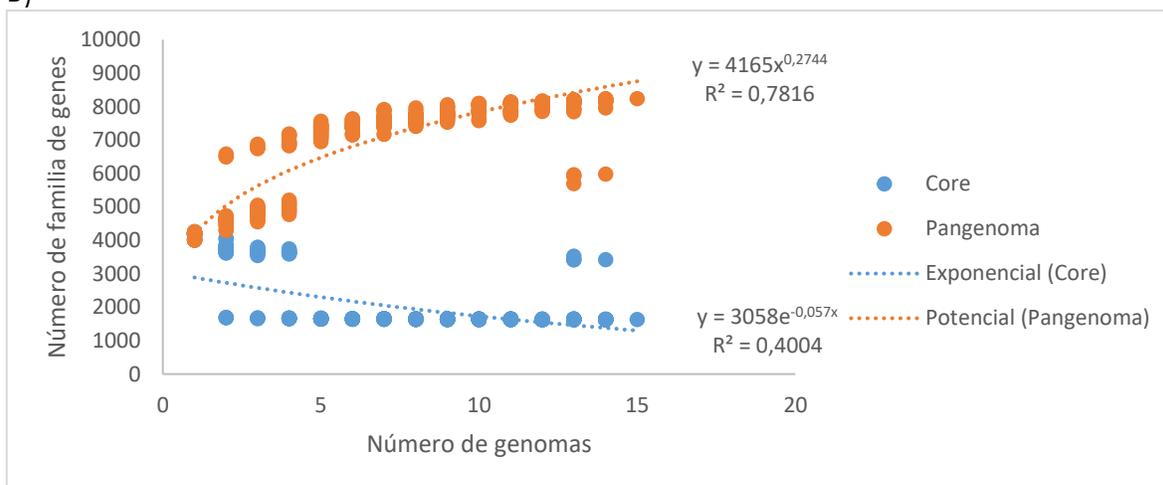


Figura 5. Resultados del core y el pan genoma en A) 70% y B) 90% de similitud

De acuerdo con la Ley de Heaps, cuando  $a > 1$  ( $\gamma < 0$ ), el pangenoma se considera cerrado, y la adición de nuevos genomas no aumentará significativamente el número de nuevos genes. Por otro lado, cuando  $a < 1$  ( $0 < \gamma < 1$ ), el pangenoma está abierto, y para cada genoma recién agregado, el número de genes aumentará significativamente (Guimaraes et al., 2015).

A partir de los resultados obtenidos haciendo uso de BPGA se observa:

70 % similitud		
	PAN GENOME	CORE GENOME
Fit law	POWER	EXPONENTIAL
Equation	$f(x)=a.x^b$	$f1(x)=c.e^{(d.x)}$
Parameters	$a= 4120.19$	$c= 3736.56$
	$b= 0.152211$	$d= -0.0202102$
Expected Size	6089	0

Estimated Size	6222.02	2759.4
----------------	---------	--------

90 % similitud		
	PAN GENOME	CORE GENOME
Fit law	POWER	EXPONENTIAL
Equation	$f(x)=a.x^b$	$f1(x)=c.e^{(d.x)}$
Parameters	a= 4385.92	c= 3833.48
	b= 0.250218	d= -0.0862478
Expected Size	8238	0
Estimated Size	8636.54	1051.33

En ambos análisis se obtuvo un pangenoma abierto con un  $\gamma$  de 0.15 para 70% de similitud y de 0.25 para un 90% de similitud. Esto no significa que la cantidad de genes de esta cepa sea infinita; lo que indican los resultados es que se requieren más genomas para una caracterización completa del conjunto de genes que definen la especie *X. campestris*.

Un pangenoma abierto es típico de grupos taxonómicos, que pueden colonizar múltiples ambientes teniendo la oportunidad de intercambiar material genético con una variedad de diferentes fuentes, enriqueciendo efectivamente su reserva genética con genes nuevos que pudieran conferirle características evolutivas en términos de resistencia a factores ambientales, patogenicidad o producción de metabolitos secundarios, lo cual es el tema central de este estudio (Tian et al., 2016).

Además, esto podría tener importantes implicaciones evolutivas, ya que podría sugerir que se ha obtenido un número relativamente alto de genes a través de eventos como la transferencia horizontal de genes, que podría haber jugado un papel clave en la configuración de la diversidad genómica entre las cepas analizadas.

### **Caracterización funcional del pangenoma de *Xanthomonas campestris***

El pangenoma es un concepto poderoso que se puede utilizar para representar eficazmente la taxonomía de una bacteria y que proporciona información sobre las categorías funcionales. La figura 6 muestra para cada sección de genes (core, accesorio y genoma único) una categoría funcional.

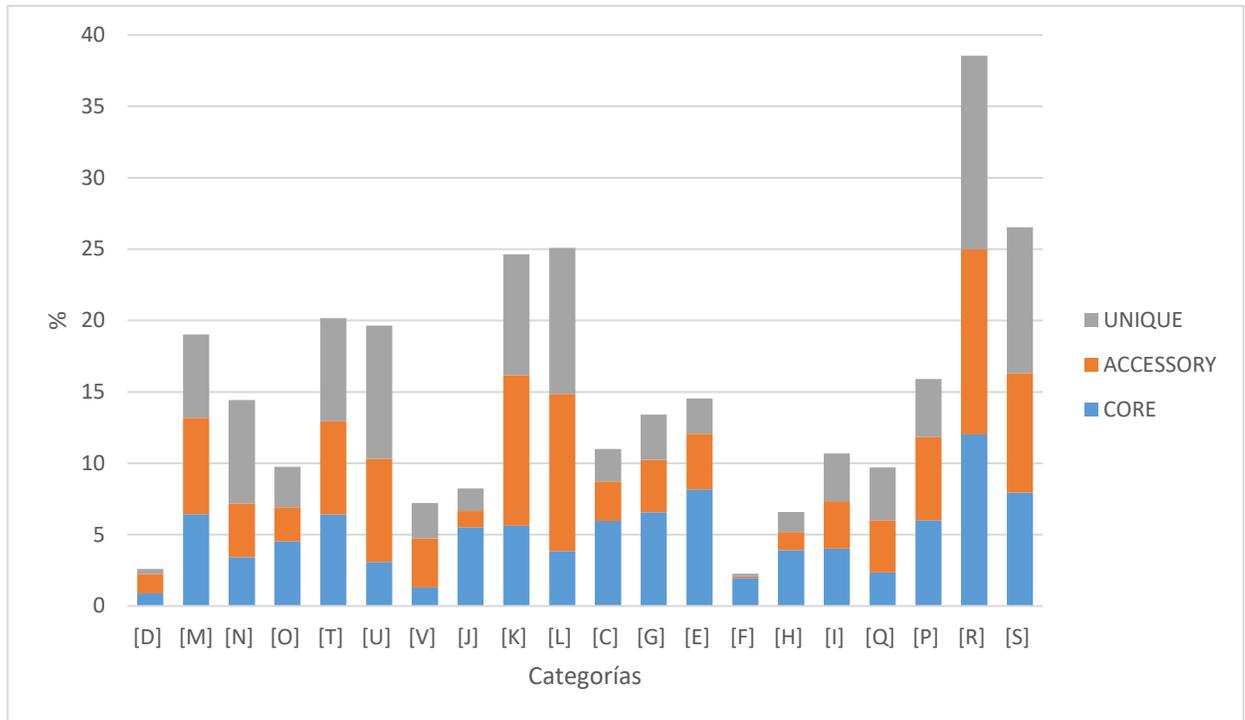


Figura 6. Grupos de genes ortólogos (COG). Categorías funcionales del núcleo (CORE), genoma accesorio (ACCESSORY) y genes únicos (UNIQUE) para 70% de similitud

- [D] Control del ciclo celular, división celular, partición cromosómica
- [M] Biogénesis de pared celular / membrana / envoltura
- [N] Motilidad celular
- [O] Modificación postraduccional, recambio proteico y chaperonas
- [T] Mecanismos de transducción de señales
- [U] Tráfico intracelular, secreción y transporte vesicular
- [V] Mecanismos de defensa
- [J] Traducción, estructura ribosómica y biogénesis
- [K] Transcripción
- [L] Replicación, recombinación y reparación.
- [C] Producción y conversión de energía.
- [G] Transporte de carbohidratos y metabolismo
- [E] Transporte de aminoácidos y metabolismo.
- [F] Transporte de nucleótidos y metabolismo
- [H] Transporte de coenzima y metabolismo
- [I] Transporte de lípidos y metabolismo
- [Q] Metabolitos secundarios, biosíntesis, transporte y catabolismo
- [P] Transporte de iones inorgánicos y metabolismo
- [R] Solo predicción de función general

[S] Función desconocida

La mayor parte de los genes centrales están involucrados principalmente en procesos metabólicos y estructurales (p. ej., COG categorías R, E, M, S, G), además de procesos como la transcripción, traducción y replicación, como es de esperarse.

En las categorías de metabolismo y transporte de carbohidratos y nucleótidos (G y F) se presenta una mayor proporción en el genoma central y de manera equitativa para el genoma accesorio y único.

Se tiene una alta proporción en las familias de genes accesorios y únicos, con respecto al genoma central para la categoría de metabolitos secundarios (Q), lo cual induce a que las cepas propuestas en este estudio pudieran ser productoras de sustancias de interés en diferentes industrias, no solamente en la producción de la goma xantana como producto de interés industrial.

A continuación, se presenta el análisis KEGG de rutas metabólicas que se obtuvo con BPGA con 70% de similitud. Cabe destacar que se suprimieron algunas de las categorías que por defecto analiza el software, ya que se tomó la decisión omitir aquellas que tuvieran menor del 1% de proporción en las tres familias de genes (core, único y accesorio). En la categoría de biosíntesis de metabolitos secundarios se observa muy baja proporción en las 3 familias de genes. Sin embargo, en la categoría de metabolismo de carbohidratos las proporciones para los genes centrales y accesorios es de 13 y 11%, respectivamente, lo que permite inferir que esta categoría podría presentar variabilidad entre las cepas analizadas lo cual a su

vez se relaciona con la producción de la goma xantana, constituida de unidades de carbohidratos.

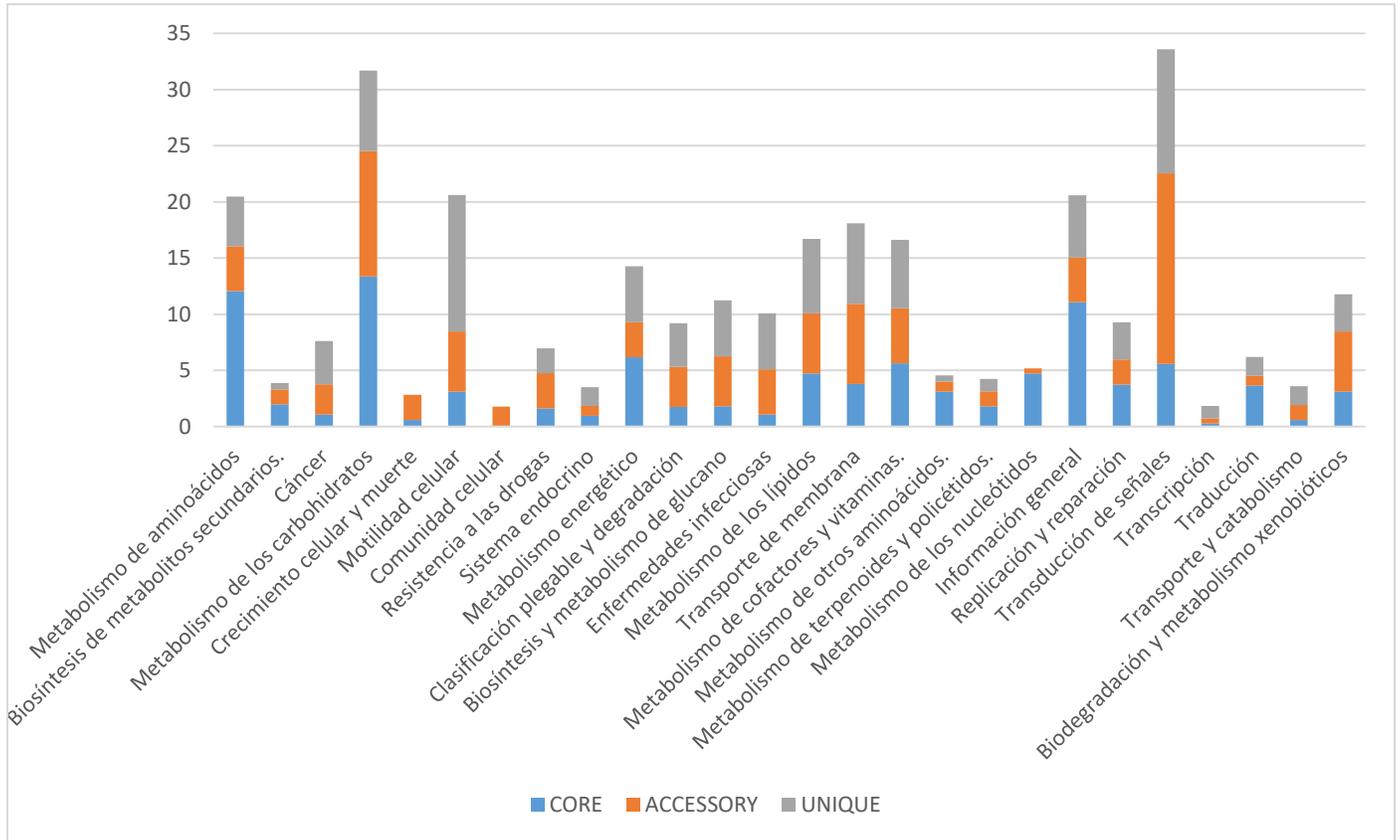


Figura 7. Análisis KEGG. Para rutas del genoma núcleo (CORE), genoma accesorio (ACCESSORY) y genoma único (UNIQUE) para 70% de similitud

Una vez asociadas a una superficie, las bacterias pueden adherirse a ella mediante estructuras de unión específicas como fibrillas de celulosa, proteínas de membrana, pili o polisacáridos extracelulares (EPS). Este aspecto se relaciona bastante con el hábitat del microorganismo; teniendo en cuenta que *X. campestris* se puede aislar de suelos y de plantas infectadas, es importante considerar su mecanismo de acción en plantas, cómo en el caso del movimiento de la bacteria en el suelo. Se ha planteado la hipótesis de que las bacterias emplean adhesinas específicas en entornos donde hay fuertes fuerzas físicas. Debido a la improbabilidad de las fuerzas de cizallamiento fuertes y continuas en el suelo, las bacterias probablemente no usan adherencia extensivamente, excepto cuando ocupan sitios que ofrecen cierta protección, o son ricos en nutrientes, o en las raíces, porque el movimiento de las raíces podría proporcionar fuerzas físicas suficientemente fuertes para causar su eliminación (Hillel & Hatfield, 2005).

La categoría de biosíntesis y metabolismo de glucanos alcanzó la mayor proporción para genes únicos con un valor de 4.9%.

El metabolismo de lípidos presenta la mayor proporción para los genes únicos, con un valor de 6.6%.

Lo anterior se relaciona con la producción de la goma xantana de acuerdo con la cepa de estudio. Es importante determinar y relacionar la expresión de dichos genes bajo diferentes condiciones ambientales para evaluar qué podría propiciar la sobreexpresión de estas regiones de acuerdo con el tipo de sustrato, pH y temperatura en la que se cultive el microorganismo en cuestión.

Por último, para la categoría de transducción de señal se tienen los siguientes valores 5.57%, 16.96% y 11.04% para el genoma central, accesorio y único, respectivamente.

A continuación, se presentan los resultados obtenidos del análisis en 90% de similitud para COG y KEGG.

El análisis de categorías COG con el 90% de identidad muestra altas proporciones en las mismas categorías que para el análisis realizado con 70% de similitud, pero mostrando valores más altos para la familia de genes accesorios y únicos. Este aspecto es evidente para la categoría de biogénesis y desarrollo de pared y membrana celular, al igual que para el metabolismo y transporte de carbohidratos.

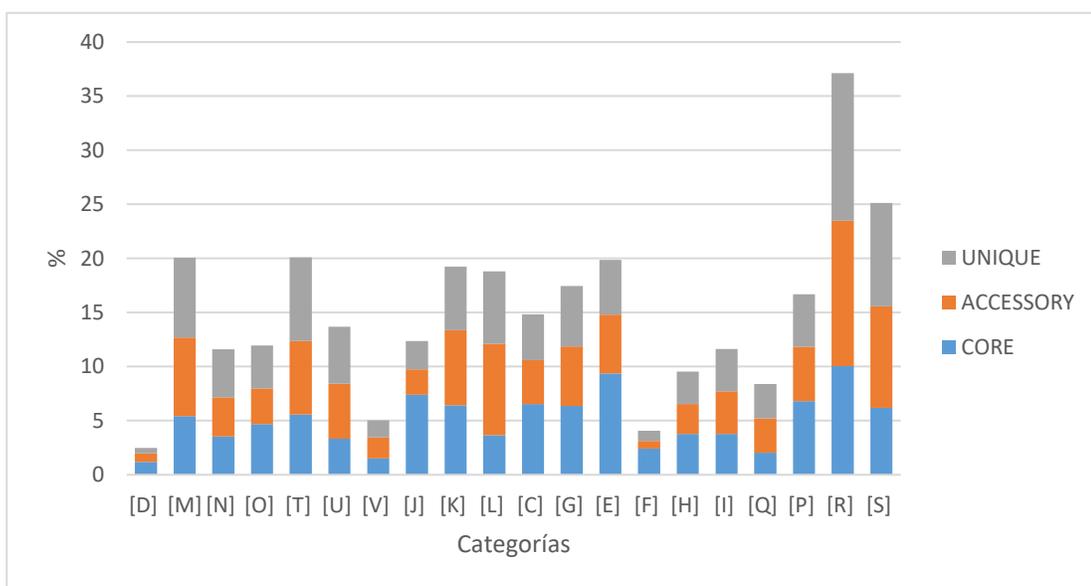


Figura 8. Grupos de genes ortólogos (COG). Categorías funcionales del núcleo (CORE), genoma accesorio (ACCESSORY) y genes únicos (UNIQUE) para 90% de similitud

De acuerdo con el análisis KEGG, se observa una alta proporción para la categoría de metabolismo de carbohidratos, en lo que respecta a genes únicos. Así mismo, se observa, una alta proporción para el metabolismo de aminoácidos. Este resultado se puede interpretar si se considera el hecho de que, si hay genes únicos en esta categoría, las cepas podrían usar diferentes sustratos y de esta manera colonizar diferentes ambientes. Adicionalmente, en la producción a escala de laboratorio y a escala industrial es importante contar con varias fuentes de carbono para evaluar la viabilidad técnica y económica del proceso.

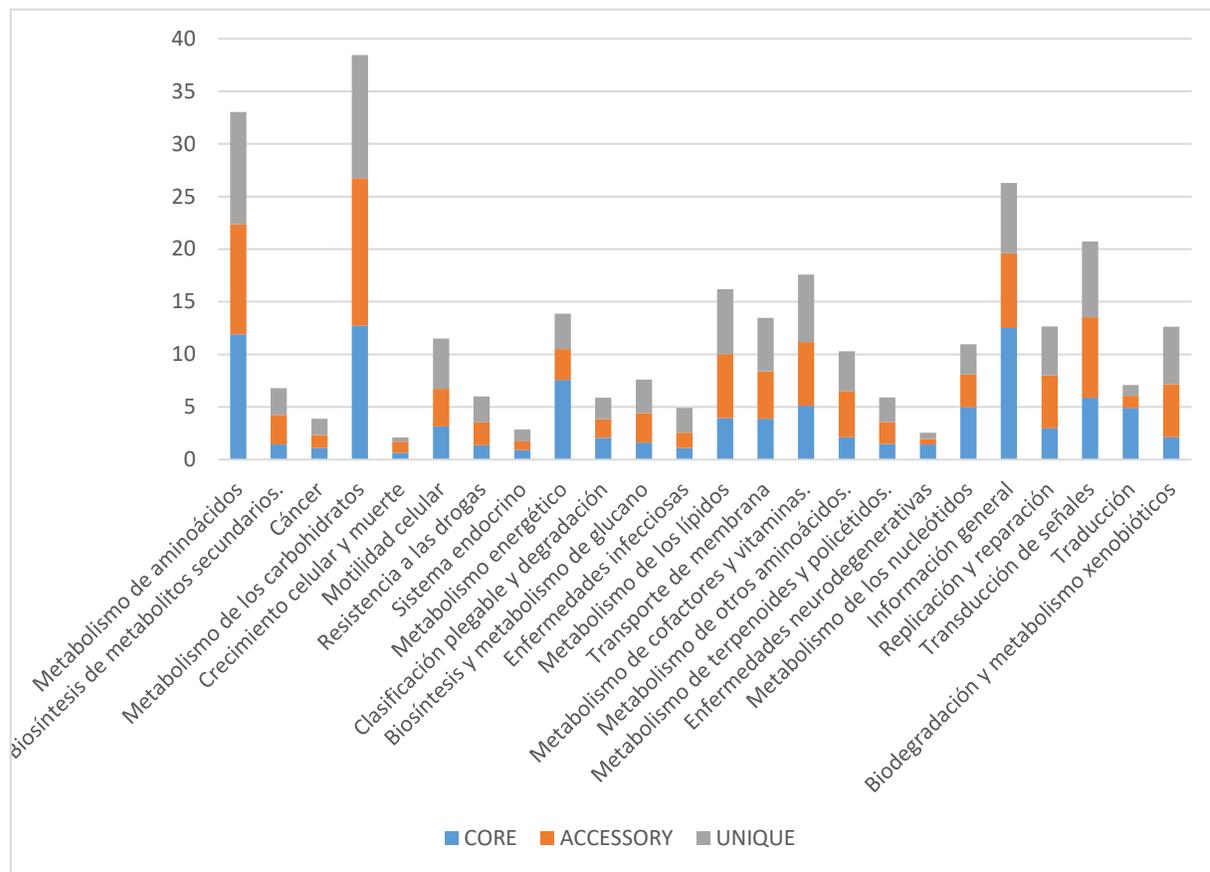


Figura 9. Análisis KEGG. Para rutas del genoma núcleo (CORE), genoma accesorio (ACCESSORY) y genoma único (UNIQUE) para 90% de similitud

## Análisis de genes responsables de la producción de goma xantana

### Análisis para un 70% de similitud

Se observó la presencia de los genes *gumD*, *gumH*, *gumI*, *gumK* en las quince cepas analizadas. Sin embargo, sólo algunas de ellas han sido reportadas como

productoras de goma xantana. Lo cual sugiere la necesidad de disponer de regiones de regulación en el genoma para su síntesis.

Cluster	Gi_Name	Gene	Annotation
1096	Prokka:CNOKAHHB_01764	<i>gumD</i>	UDP-glucose:undecaprenyl-phosphate glucose-1-phosphate transferase
1599	Prokka:GLKPIJAN_01744	<i>gumK</i>	UDP-glucuronate:glycolipid 2-beta-glucuronosyltransferase
1827	Prokka:KCDCCFJN_02104	<i>gumH</i>	GDP-mannose:cellobiosyl-diphosphopolyprenol alpha-mannosyltransferase
2102	Prokka:NFGADIDN_01711	<i>gumI</i>	GDP-mannose:glycolipid 4-beta-D-mannosyltransferase

La ruta de síntesis de la goma xantana involucra UDP-glucosa, GDP-manosa y UDP-glucurónico.

Al filtrar las anotaciones con UDP – glucosa se encontró lo siguiente:

Cluster	Gi_Name	Gene	Annotation
1096	Prokka:CNOKAHHB_01764	<i>gumD</i>	UDP-glucose:undecaprenyl-phosphate glucose-1-phosphate transferase
1305	Prokka:GLKPIJAN_02726	<i>tuaD</i>	UDP-glucose 6-dehydrogenase <i>TuaD</i>
1599	Prokka:GLKPIJAN_01744	<i>gumK</i>	UDP-glucuronate:glycolipid 2-beta-glucuronosyltransferase
1765	Prokka:KCDCCFJN_03000	<i>ugd</i>	UDP-glucose 6-dehydrogenase
1796	Prokka:DGFEFBEB_03894	<i>galE</i>	UDP-glucose 4-epimerase
2387	Prokka:NPELLJKA_00408	-	UDP-glucose 4-epimerase

El gen *tuaD* es central, lo cual implica que se encuentra en las quince cepas. A su vez, el gen *ugd* es único de la cepa ATCC 33913.

Regiones denominadas xpsIII, xpsIV y xpsVI, así como un grupo de genes de 35,3 kb contienen varias regiones codificantes necesarias para la biosíntesis de nucleótidos requeridos para la producción de xantano y lipopolisacáridos (Becker et al., 1998).

Al filtrar el gen *xps* en la matriz de anotaciones se obtiene lo siguiente:

Cluster	Gi_Name	Gene	Annotation
366	Prokka:ICODJPOJ_03650	xpsD	Secretin XpsD
780	Prokka:NPELLJKA_00761	xpsE	Type II secretion system protein E
3027	Prokka:KCDCCFJN_00397	xpsN	General secretion pathway protein N

Los tres genes detectados en el análisis son del genoma central.

Según Becker et al. (1998) la mayoría de los genes *gum* podrían ser alterados dentro de la cepa de tipo salvaje. Sin embargo, los genes *gumB*, *gumC*, *gumE*, *gumM* y *gumJ* solo podrían mutar cuando se expresan en una cepa defectuosa de UDP-glucosa ya que su inactivación en cepas de tipo salvaje parecía ser letal. Sin embargo, en este análisis se observa la ausencia de dichos genes. Adicionalmente, la deficiencia de dichos genes podría recuperarse por la introducción de un plásmido que lleva la codificación de la región para el dominio C-terminal de GumD, que parece ser responsable de su actividad glucosil-1-fosfato transferasa Becker et al. (1998).

Dado que los genes *gumB*, *gumC* y *gumE* parecen acumular subunidades completas de xantano in vitro y no pueden sintetizar el polímero, los productos de estos genes pueden ser necesarios para la polimerización o exportación del polímero.

Aunque la función del producto *gumJ* no se puede asociar con un paso biosintético de goma particular, no se puede descartar un papel de secreción para *gumJ*. Alternativamente, podría ser necesario para evitar la acumulación de un producto nocivo o para reciclar sustratos esenciales (Becker et al., 1998).

La síntesis de estas y otras enzimas extracelulares y xantano se activa por los productos de un grupo genes denominados *rpf*. Al buscar en la matriz de anotaciones dichos genes se obtuvo:

Cluster	Gi_Name	Gene	Annotation
438	Prokka:DKBECLLH_02313	rpfC_1	Sensory/regulatory protein RpfC
1842	Prokka:DKBAFDHH_02149	rpfG_2	Cyclic di-GMP phosphodiesterase response regulator RpfG
2309	Prokka:KCDCCFJN_02004	rpfG_2	Cyclic di-GMP phosphodiesterase response regulator RpfG
2721	Prokka:BLGKLHJB_03180	rpfC_3	Sensory/regulatory protein RpfC
3615	Prokka:KCDCCFJN_02207	trpF	N-(5'-phosphoribosyl)anthranilate isomerase
3911	Prokka:CNOKAHHB_0239 3	rpfC_3	Sensory/regulatory protein RpfC

Todos ellos del genoma central. La secuencia de uno de estos genes, *rpfC*, mostró homología con los miembros de la clase de proteínas reguladoras de dos componentes. Estos factores parecen estar equilibrados por al menos un gen no vinculado que regula negativamente la síntesis de las enzimas extracelulares y el xantano (Tang et al. 1990).

#### **Análisis para un 90% de similitud**

El gen *gumD* y *gumH* son del genoma central, mientras que *gumK* y *gumI* son únicos de la cepa ATCC 33913.

Cluster	Gi_Name	Gen	Annotation
1469	Prokka:CNOKAHHB_01764	gumD	UDP-glucose:undecaprenyl-phosphate glucose-1-phosphate transferase
2125	Prokka:GLKPIJAN_01744	gumK	UDP-glucuronate:glycolipid 2-beta-glucuronosyltransferase
2191	Prokka:KCDCCFJN_02101	gumK	UDP-glucuronate:glycolipid 2-beta-glucuronosyltransferase
2424	Prokka:KCDCCFJN_02104	gumH	GDP-mannose:cellobiosyl-diphosphopolyprenol alpha-mannosyltransferase
2800	Prokka:NFGADIDN_01711	gumI	GDP-mannose:glycolipid 4-beta-D-mannosyltransferase
2849	Prokka:KCDCCFJN_02103	gumI	GDP-mannose:glycolipid 4-beta-D-mannosyltransferase

Al filtrar la matriz de anotación con UDP – glucosa se encuentran los genes *gumD* (genoma central), *gumK* en el genoma accesorio y único en la cepa ATCC 33913. El gen *tuaD* es accesorio, presente en 14 cepas; no está presente en la ATCC 33913. El gen *ugd* es único de ATCC 33913 mientras que el gen *galE* es accesorio de 14 cepas, sin estar presente en la ATCC 33913.

Cluster	Gi_Name	Gen	Annotation
1469	Prokka:CNOKAHNB_01764	<i>gumD</i>	UDP-glucose:undecaprenyl-phosphate glucose-1-phosphate transferase
1736	Prokka:GLKPIJAN_02726	<i>tuaD</i>	UDP-glucose 6-dehydrogenase <i>TuaD</i>
2125	Prokka:GLKPIJAN_01744	<i>gumK</i>	UDP-glucuronate:glycolipid 2-beta-glucuronosyltransferase
2191	Prokka:KCDCCFJN_02101	<i>gumK</i>	UDP-glucuronate:glycolipid 2-beta-glucuronosyltransferase
2345	Prokka:KCDCCFJN_03000	<i>ugd</i>	UDP-glucose 6-dehydrogenase
2387	Prokka:DGFEFBFB_03894	<i>galE</i>	UDP-glucose 4-epimerase
3186	Prokka:NPELLJKA_00408	-	UDP-glucose 4-epimerase
8118	Prokka:KCDCCFJN_01219	<i>tuaD</i>	UDP-glucose 6-dehydrogenase <i>TuaD</i>

Adicionalmente, para la anotación GDP-Manosa se encontraron los siguientes cluster:

Cluster	Gi_Name	Gen	Annotation
2424	Prokka:KCDCCFJN_02104	<i>gumH</i>	GDP-mannose:cellobiosyl-diphosphopolyprenol alpha-mannosyltransferase
2433	Prokka:KCDCCFJN_00719	<i>mgtA</i>	GDP-mannose-dependent alpha-mannosyltransferase
2448	Prokka:NFGADIDN_03243	<i>mgtA</i>	GDP-mannose-dependent alpha-mannosyltransferase
2800	Prokka:NFGADIDN_01711	<i>gumI</i>	GDP-mannose:glycolipid 4-beta-D-mannosyltransferase
2836	Prokka:KCDCCFJN_03013	<i>gmd</i>	GDP-mannose 4,6-dehydratase
2849	Prokka:KCDCCFJN_02103	<i>gumI</i>	GDP-mannose:glycolipid 4-beta-D-mannosyltransferase
2858	Prokka:NIDCJLAI_00746	<i>gmd</i>	GDP-mannose 4,6-dehydratase
3463	Prokka:BLGKLIJB_03775	<i>gmd_1</i>	GDP-mannose 4,6-dehydratase
5117	Prokka:DKBECLLH_01596	<i>nudK</i>	GDP-mannose pyrophosphatase <i>NudK</i>
7846	Prokka:NFGADIDN_03601	<i>gmd</i>	GDP-mannose 4,6-dehydratase

Del análisis anterior, se destaca la presencia del gen *gmd*, el cual es único para la cepa ATCC 33913 y CMP 4013. Además, el gen *gmd\_1* es único para la cepa 8004.

En cuanto a la presencia de genes *xps*, se observó que *xpsE* es del genoma central. Los genes *xpsD* y *xpsN* son únicos de la cepa ATCC 33913.

Cluster	Gi_Name	Gen	Annotation
480	Prokka:ICODJPOJ_03650	<i>xpsD</i>	Secretin XpsD
514	Prokka:KCDCCFJN_00398	<i>xpsD</i>	Secretin XpsD
517	Prokka:OJGHHOKI_03708	<i>xpsD</i>	Secretin XpsD
1035	Prokka:NPELLJKA_00761	<i>xpsE</i>	Type II secretion system protein E
4067	Prokka:KCDCCFJN_00397	<i>xpsN</i>	General secretion pathway protein N
4133	Prokka:DGFEFBEB_00703	<i>xpsN</i>	General secretion pathway protein N

## 5.2 Resultados de MICROPAN

A continuación, se presenta la estimación del tamaño del pangenoma y la distribución de los clúster en las diferentes cepas de estudio.

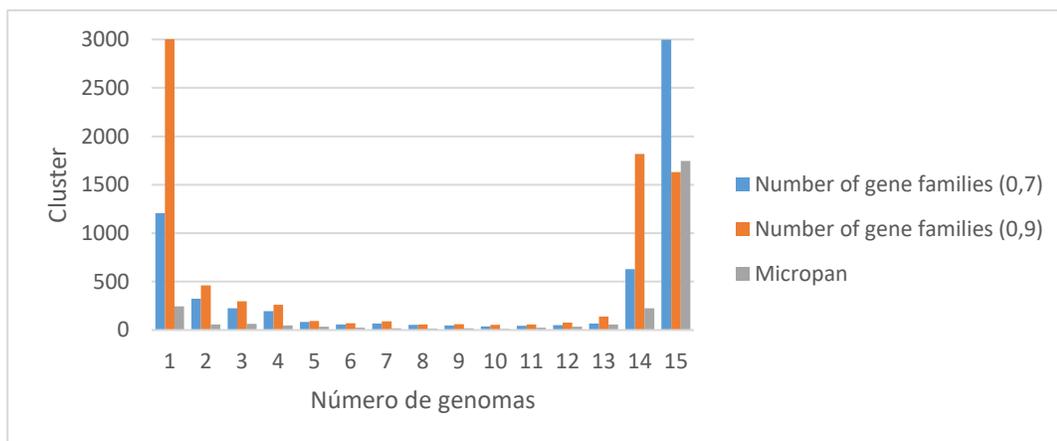


Figura 10. Comparación del análisis en BPGA y Micropan

El tamaño estimado del core fue de 418 clúster y el pangenoma de 4178 clúster. La ley de Heaps en micropan permitió obtener un intercepto de 229.64 y un alfa de 1.0906, lo que indica un pangenoma cerrado.

Adicionalmente se analizó la distribución de los genes en el genoma central, accesorio y único obteniendo el siguiente esquema:

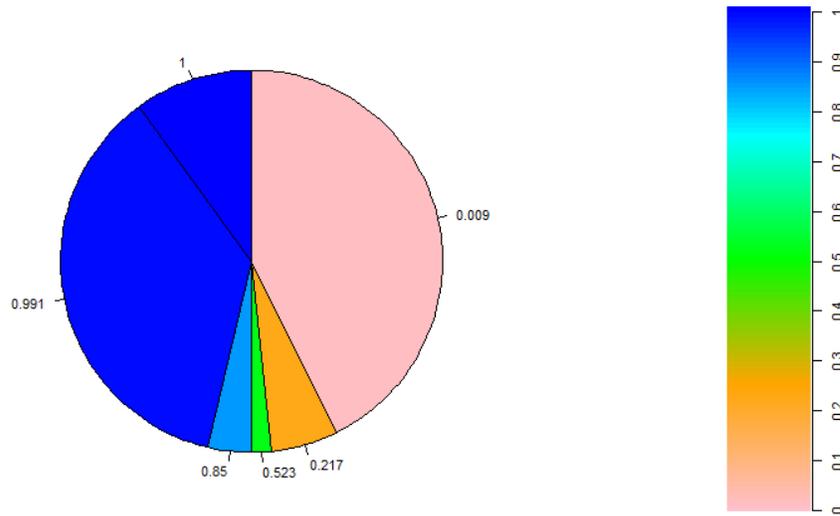


Figura 11. Distribución de los genes en el genoma central, accesorio y único

En la Figura 7 se observa que la mitad de los genes se encuentran en el genoma de las 15 cepas, lo que representa el genoma central (área azul oscura), mientras que los genes que tienen una proporción de 0.009 (área rosada) indica la probabilidad de los genes de pertenecer a un solo genoma, es decir la categoría de genes únicos. Los otros genes (representados por azul claro, verde y naranja) corresponden al genoma accesorio.

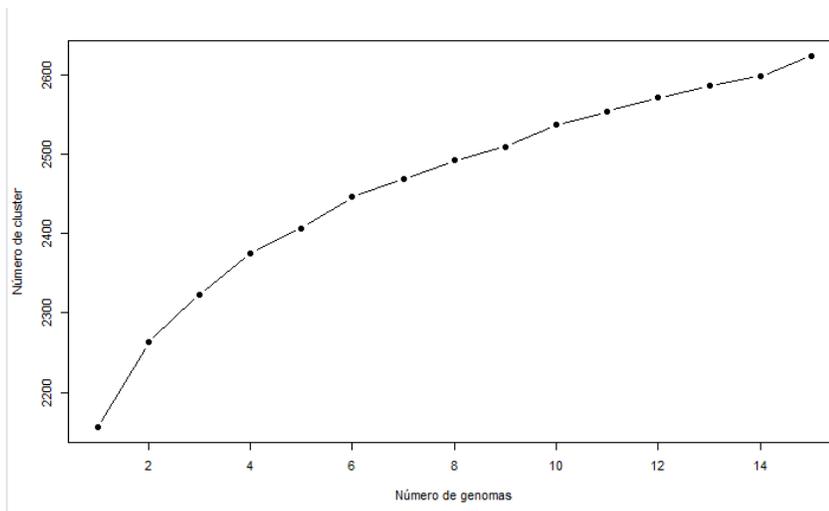


Figura 12. Número de clúster de acuerdo a los genomas de las cepas analizadas

En la figura 12 se observa que hay un aumento de los clúster en la medida que aumentan los genomas.

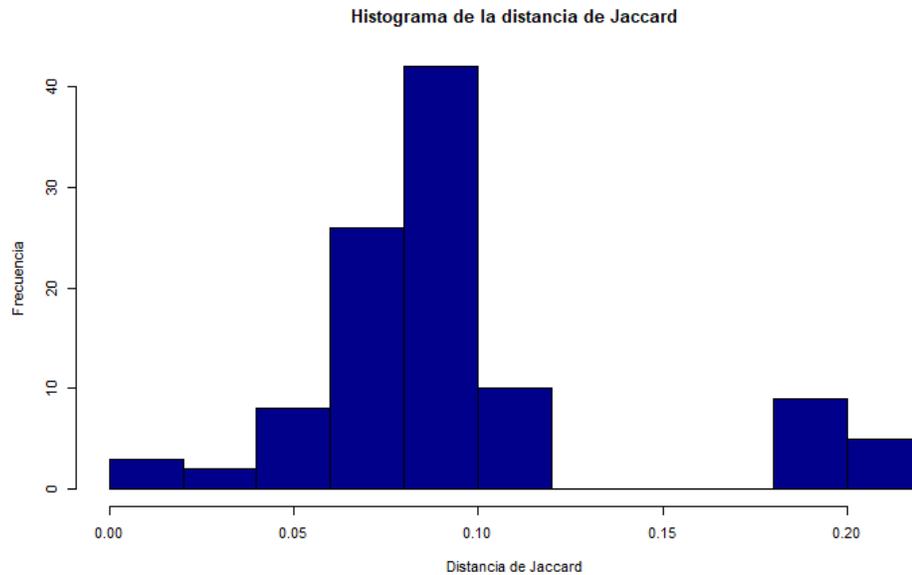


Figura 13. Distancia de Jaccard

Adicionalmente se calculó la distancia de Jaccard, la cual define una distancia de cero para genomas idénticos y una distancia de 1 para genomas completamente diferentes. Con este análisis se buscó determinar en qué porcentaje difiere un genoma de otro. Como resultado que se obtuvo que la mayoría de los genomas difiere en promedio en un 9.4%, teniendo genomas que son casi idénticos (con valores de Jaccard muy cercanos a cero) y genomas que difieren en un 20%, como se muestra en la Figura 13.

La diferencia de genomas puede deberse a que se trata de organismos procariotas que son mucho más diversos y que intercambian rápidamente material genético usando diferentes mecanismos que promueven la diversidad genética, por ejemplo, la transferencia horizontal de genes. La recombinación provoca reordenamientos frecuentes del genoma, la transferencia horizontal introduce nuevas secuencias en los cromosomas bacterianos y las delecciones eliminan segmentos del genoma. En consecuencia, cada genoma es un mosaico de segmentos únicos específicos del linaje, regiones compartidas con un subconjunto

de otros genomas y segmentos conservados entre todos los genomas en consideración (Darling et al., 2004).

Se sabe que los genomas sufren varios tipos de eventos evolutivos a gran escala. La duplicación de genes puede dar lugar a la existencia de genes parálogos, mientras que la pérdida de genes puede eliminar una copia y ocultar el supuesto de la ortología. El reordenamiento de los elementos genéticos ocurre por mecanismos tales como la inversión repetida o la translocación. La transferencia horizontal introduce nuevos elementos genéticos en genomas bacterianos (Darling et al., 2004).

### 5.3 Resultados obtenidos en ANTISMASH

#### **Identificación de grupos de genes biosintéticos**

La diversidad metabólica especializada se hizo con antiSMASH; los resultados se procesaron mediante Excel. Se identificaron clúster de sideróforos (xantoferrina), enzimas de tipo péptido sintasa no ribosomal (NPRS) y arilpolienos (xanthomonadina I) en la mayoría de las cepas. Los resultados se muestran en la Figuras 14 y 15, y anexo 2.

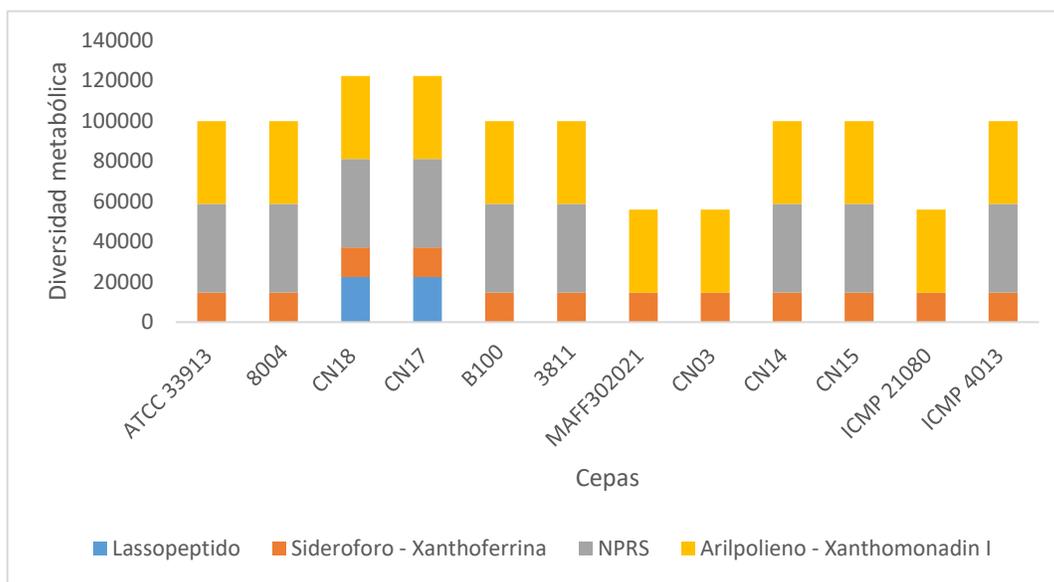


Figura 14. Resultados de antiSMASH (I)

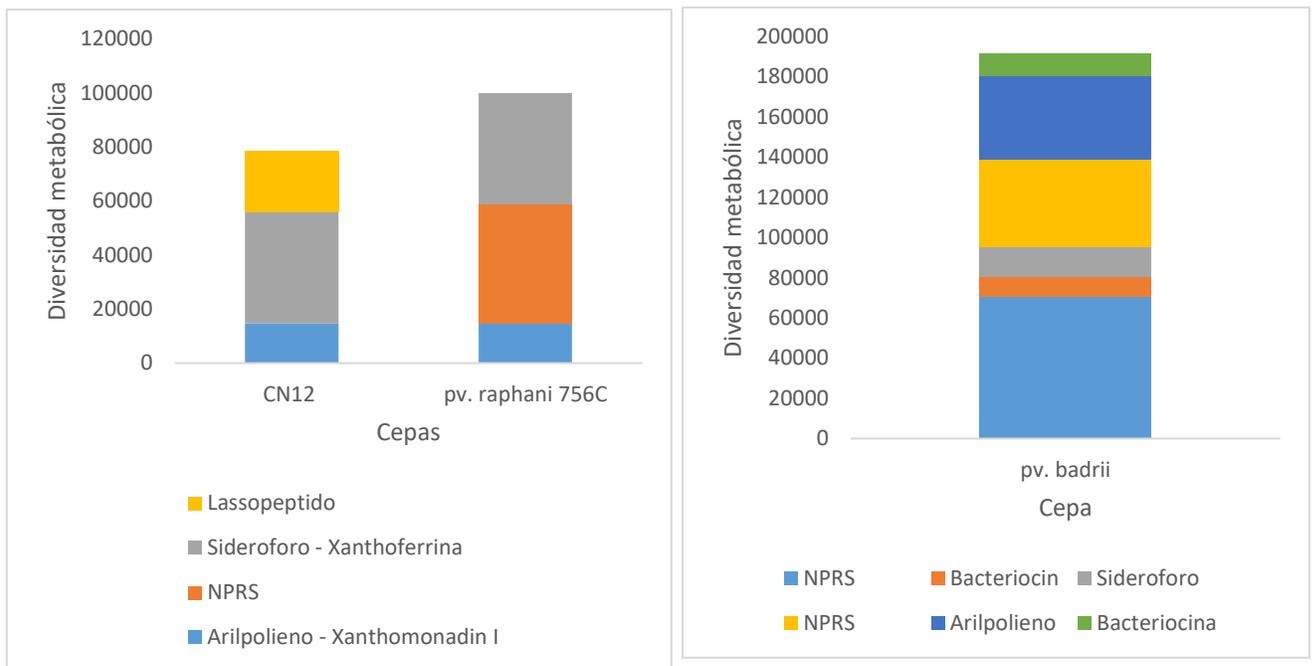


Figura 15. Resultados de antiSMASH (II)

Las cepas CN12 y pv. Raphani 756C presentaron posiciones diferentes de los clúster en el genoma. pv. Badrii fue la cepa que presentó mayor número de metabolitos secundarios (Figura 15b).

En particular, las xantomonas promueven la biosíntesis de sideróforos y la expresión de un clúster de genes de captación xss (síntesis de sideróforos de *Xanthomonas*) implicados en la producción de un sideróforo de tipo vibrioferrina. La Xanthoferrina es un sideróforo de tipo hidroxycarboxilato que se requiere para el crecimiento en condiciones de baja temperatura y para una virulencia óptima. Una mutación en el gen xssA de síntesis de sideróforos causa deficiencia en la producción de sideróforos en condiciones de bajo contenido de hierro (Pandey et al., 2017).

La captación y el metabolismo del hierro son cruciales para el crecimiento y la supervivencia de los patógenos bacterianos dentro de sus huéspedes. En condiciones bajas en hierro, las bacterias producen y segregan sideróforos.

La homeostasis celular del hierro es crítica para la supervivencia y el crecimiento. Las bacterias emplean una variedad de estrategias para secuestrar el hierro del medio ambiente y almacenar el excedente de hierro intracelular que se puede utilizar en condiciones de restricción de hierro, al tiempo que limitan el potencial para la producción de especies reactivas de oxígeno inducidas por hierro (ROS) (Javvadi et al., 2018).

#### 5.4 Resultados MAUVE

Los genes *gum* se encuentran en regiones conservadas de los genomas de las quince cepas (Ver región color rojo en el Anexo 3). Con lo anterior, se concluye que hay sintenia entre los genomas, es decir que se conserva la ubicación espacial de dichos genes en los genomas analizados.

#### 5.5 Propuesta para el mejoramiento en la producción de goma xantana

La fermentación exitosa de la goma xantana requiere una comprensión clara del entorno microbiano y la cinética involucrada. El estudio del sistema celular se ve ampliamente favorecido si se conocen los genes que se conservan en las diferentes cepas y su potencial aplicaciones en la producción de metabolitos secundarios. Este hecho permite alcanzar avances a escala de laboratorio, a partir del mejoramiento de especies celulares, así como el mejoramiento de los bioprocesos y la optimización en la producción de los mismos.

De acuerdo a lo anterior, se recomienda realizar ensayos experimentales sobre las cepas tenidas en cuenta en este estudio para evaluar el efecto de diferentes variables ambientales y operaciones que podrían contribuir con la producción de sustancias de interés. Avanzar en el entendimiento del sistema celular, sus condiciones de cultivo y sus rutas metabólicas, podrá permitir a futuro la eliminación de subproductos no deseados, a partir de modificaciones genéticas de las cepas de producción logrando una mejora en los procesos de separación y con ello una mejor recuperación de xantano del líquido de fermentación.

Se ha trabajado en múltiples intentos para optimizar las variables de cultivo en la fermentación de goma de xantano, ya sea, la composición de nutrientes y la técnica de alimentación, la temperatura, pH, agitación y la adición de antiespumante. La glucosa sigue siendo el mejor sustrato en términos de rendimiento, suministro y calidad del producto. En cuanto a la producción convencional de xantano, la mezcla es el principal problema de cultivo, ya que el caldo producido durante la etapa de producción es muy viscoso (Rosalam & England, 2006).

Se han reportado diferentes cepas mutantes capaces de producir mejores rendimientos de producto. Por ejemplo, la cepa mutante *Xanthomonas campestris* WXLB-006 alcanzó un rendimiento de 17,8 g de goma por litro de cultivo en matraz. Con la adopción del control del pH, la aireación y agitación variadas, y la

estrategia de alimentación de glicerol variada, la producción de xantano alcanzó 33.9 g / L en un fermentador de 7 L y en 60 horas (Wang et al., 2017).

En otro estudio, se eliminó el gen *Xcc4015*, que codifica para una AMP-ligasa, relacionado con la síntesis xantomonadinas; se observó que las cepas de *Xanthomonas* que carecen de xantomonadinas presentan una disminución en la producción de xantano. Los investigadores concluyeron que puede haber una relación entre la síntesis de xantomonadinas y el nivel de polisacáridos extracelulares obtenidos. Los resultados demostraron que el uso de bacterias, manipuladas genéticamente, en la producción industrial de goma de xantano podría reducir en gran medida el uso de etanol para la recuperación de la goma y el costo de producción (Dai et al., 2019). Estudios como estos se ven grandemente favorecidos en la medida en que se conozca el role particular de diferentes genes de clústers, previamente identificados, en los procesos de síntesis del metabolito de interés.

En este sentido, es importante promover la formulación y desarrollo de diseños experimentales que consideren el mejoramiento genético a partir del conocimiento de su genoma y los potenciales blancos metabólicos. Para el caso particular, la manipulación genética que permita sobre expresar los genes *gum* y la producción de las xantomonadinas, contribuye a la producción de xantano y al mejoramiento de las condiciones en las que se produce, para reducir significativamente problemas de purificación, separación y manipulación del producto.

## 6. Conclusiones

- Haciendo uso del análisis en BPGA se observó que el pangenoma de la especie *Xanthomas campestris* fue abierto, mientras que en micropan fue cerrado. Esto hecho se debe a que el primer software hace el alineamiento de secuencias con base en la presencia o ausencia de genes, mientras que el segundo lo hace a través de la agrupación de dominios.
- Las proporciones de genes centrales, accesorios y únicos fueron altas para las categorías involucradas con la ruta de carbohidratos en KEGG como en COG, para 70 y 90% de similitud. Este aspecto contribuye al entendimiento de la biosíntesis de goma xantana, la cual requiere de azúcares para su fabricación como se observó en la ruta metabólica presentada en la Figura 3B.
- Se podría inferir que la goma xantana es fundamental para el estilo de vida fitopatógeno de las especies del género *Xanthomonas*, ya que los genes

involucrados en su producción se encuentran conservado en todos los genomas analizados.

- Una perspectiva evolutiva sobre los componentes del genoma de *X. campestris* permitirá tener nociones más claras respecto a la manipulación genética de este microorganismo, a la comprensión de su sistema celular, así como los mecanismos adaptativos de las características específicas de las cepas, incluida la producción de nuevos metabolitos secundarios, patogénesis y respuestas al estrés ambiental y fisiológico.
- El pangenoma logró agrupar cada uno de los genomas con relación a la presencia y ausencia de los genes implicados en las rutas metabólicas de biosíntesis de la goma xantana, lo cual podrá aprovecharse para realizar manipulación genética a escala de laboratorio y lograr así altos rendimientos en los procesos de producción.

## 7. Referencias Bibliográficas

Armitage, J. P. (1999). Bacterial tactic responses. In *Advances in microbial physiology* (Vol. 41, pp. 229-289). Academic Press.

Barh, D., Soares, S. C., Tiwari, S., & Azevedo, V. (Eds.). (2020). *Pan-genomics: Applications, Challenges, and Future Prospects*. Academic Press.

Becker, A., Katzen, F., Pühler, A., & Ielpi, L. (1998). Xanthan gum biosynthesis and application: a biochemical/genetic perspective. *Applied microbiology and biotechnology*, 50(2), 145-152.

Blin, K., Kazempour, D., Wohlleben, W., & Weber, T. (2014). Improved lanthipeptide detection and prediction for antiSMASH. *PLoS One*, 9(2), e89420.

Bogdanove, A. J., Koebnik, R., Lu, H., Furutani, A., Angiuoli, S. V., Patil, P. B., ... & Aparna, G. (2011). Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. *Journal of Bacteriology*, 193(19), 5450-5464.

Carignatto, C. R. R., Oliveira, K. S. M., de Lima, V. M. G., & de Oliva Neto, P. (2011). New culture medium to xanthan production by *Xanthomonas campestris* pv. *campestris*. *Indian journal of microbiology*, 51(3), 283-288.

Carlos Guimaraes, L., Benevides de Jesus, L., Vinicius Canario Viana, M., Silva, A., Thiago Juca Ramos, R., de Castro Soares, S., & Azevedo, V. (2015). Inside the pan-genome-methods and software overview. *Current genomics*, 16(4), 245-252.

Chaudhari, N. M., Gupta, V. K., & Dutta, C. (2016). BPGA-an ultra-fast pan-genome analysis pipeline. *Scientific reports*, 6(1), 1-10.

- Criollo Alcántara, A. (2019). Evaluación de la producción de xantano en cultivos por lote por *Xanthomonas campestris* 1459 empleando subproductos agroindustriales como fuente de carbono.
- Cuccuru, G., Orsini, M., Pinna, A., Sbardellati, A., Soranzo, N., Travaglione, A., ... & Fotia, G. (2014). Orione, a web-based framework for NGS analysis in microbiology. *Bioinformatics*, 30(13), 1928-1929.
- da Silva, A. R., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., ... & Do Amaral, A. M. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature*, 417(6887), 459-463.
- Dagan, T., & Martin, W. (2006). The tree of one percent. *Genome biology*, 7(10), 118.
- Dai, X., Gao, G., Wu, M., Wei, W., Qu, J., Li, G., & Ma, T. (2019). Construction and application of a *Xanthomonas campestris* CGMCC 15155 strain that produces white xanthan gum. *MicrobiologyOpen*, 8(2), e00631.
- Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research*, 14(7), 1394-1403.
- Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*, 5(6), e11147.
- Denancé, N., Szurek, B., Doyle, E. L., Lauber, E., Fontaine-Bodin, L., Carrère, S., ... & Poussier, S. (2018). Two ancestral genes shaped the *Xanthomonas campestris* TAL effector gene repertoire. *New Phytologist*, 219(1), 391-407.
- Desai, D., Li, J. H., de Jong, E. V. Z., Braun, R., Pitman, A., Visnovsky, S., ... & Christey, M. (2015). Draft genome sequences of two New Zealand *Xanthomonas campestris* pv. *campestris* isolates, ICMP 4013 and ICMP 21080. *Genome announcements*, 3(5).
- Garcia-Ochoa, F., Santos, V. E., Casas, J. A., & Gómez, E. (2000). Xanthan gum: production, recovery, and properties. *Biotechnology advances*, 18(7), 549-579.
- Garcia-Ochoa, F., Santos, V. E., Casas, J. A., & Gómez, E. (2000). Xanthan gum: production, recovery, and properties. *Biotechnology advances*, 18(7), 549-579.
- He, Y. W., Cao, X. Q., & Poplawsky, A. R. (2020). Chemical Structure, Biological Roles, Biosynthesis and Regulation of the Yellow Xanthomonadin Pigments in the Phytopathogenic Genus *Xanthomonas*. *Molecular Plant-Microbe Interactions*, 33(5), 705-714.
- Hillel, D., & Hatfield, J. L. (Eds.). (2005). *Encyclopedia of Soils in the Environment* (Vol. 3). Amsterdam: Elsevier.
- Javvadi, S., Pandey, S. S., Mishra, A., Pradhan, B. B., & Chatterjee, S. (2018). Bacterial cyclic  $\beta$ -(1, 2)-glucans sequester iron to protect against iron-induced toxicity. *EMBO reports*, 19(1), 172-186.

- Kariduraganavar, M. Y., Kittur, A. A., Kamble, R. R., Kumbar, S. G., Laurencin, C. T., & Deng, M. (2014). Polymer Synthesis and Processing. Natural and Synthetic Biomedical Polymers. *Nat Synth Biomed Polym*, 25-26.
- Kong, C., Horta de Passo, V., Fang, Z., Yang, L., Zhuang, M., Zhang, Y., ... & Lv, H. (2019). Complete Genome Sequence of Strain WHRI 3811 Race 1 of *Xanthomonas campestris* pv. *campestris*, the Causal Agent of Black Rot of Cruciferous Vegetables. *Molecular Plant-Microbe Interactions*, 32(12), 1571-1573.
- Kuppuswami, G. M. (2014). Fermentation (Industrial)| production of xanthan gum.
- Lancheros, R., Caicedo, L., & de Navarro, Y. N. (2002). Polisacáridos rhizobianos, una nueva alternativa. *Ingeniería e Investigación*, (49), 4-8.
- Martin-Gómez, H., & Tulla-Puche, J. (2018). Lasso peptides: chemical approaches and structural elucidation. *Organic & biomolecular chemistry*, 16(28), 5065-5080.
- Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., ... & Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic acids research*, 39(suppl\_2), W339-W346.
- Mira, A., Martín-Cuadrado, A. B., D'Auria, G., & Rodríguez-Valera, F. (2010). The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol*, 13(2), 45-57.
- Niknezhad, S. V., Asadollahi, M. A., Zamani, A., & Biria, D. (2016). Production of xanthan gum by free and immobilized cells of *Xanthomonas campestris* and *Xanthomonas pelargonii*. *International Journal of Biological Macromolecules*, 82, 751-756.
- Palaniraj, A., & Jayaraman, V. (2011). Production, recovery and applications of xanthan gum by *Xanthomonas campestris*. *Journal of Food Engineering*, 106(1), 1-12.
- Papagianni, M., Psomas, S. K., Batsilas, L., Paras, S. V., Kyriakidis, D. A., & Liakopoulou-Kyriakides, M. (2001). Xanthan production by *Xanthomonas campestris* in batch cultures. *Process Biochemistry*, 37(1), 73-80.
- Pandey, S. S., Patnana, P. K., Rai, R., & Chatterjee, S. (2017). Xanthoferrin, the  $\alpha$ -hydroxycarboxylate-type siderophore of *Xanthomonas campestris* pv. *campestris*, is required for optimum virulence and growth inside cabbage. *Molecular plant pathology*, 18(7), 949-962.
- Patel, M. K., Kulkarni, Y. S., & Dhande, G. W. (1950). *Xanthomonas badrii* sp. nov., on *Xanthium strumarium* L. in India. *Indian Phytopathology*, 3(1).
- Qian, W., Jia, Y., Ren, S. X., He, Y. Q., Feng, J. X., Lu, L. F., ... & Wu, W. (2005). Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome research*, 15(6), 757-767.
- Rosalam, S., & England, R. (2006). Review of xanthan gum production from unmodified starches by *Xanthomonas campestris* sp. *Enzyme and Microbial Technology*, 39(2), 197-207.

- Rouli, L., Merhej, V., Fournier, P. E., & Raoult, D. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New microbes and new infections*, 7, 72-85.
- Sah, S., & Singh, R. (2015). Siderophore: Structural and functional characterisation—A comprehensive review. *Agriculture (Pol'nohospodárstvo)*, 61(3), 97-114.
- Saha, M., Sarkar, S., Sarkar, B., Sharma, B. K., Bhattacharjee, S., & Tribedi, P. (2016). Microbial siderophores and their potential applications: a review. *Environmental Science and Pollution Research*, 23(5), 3984-3999.
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068-2069.
- Silva, C. C., Silva, S. P., & Ribeiro, S. C. (2018). Application of bacteriocins and protective cultures in dairy food preservation. *Frontiers in microbiology*, 9, 594.
- Snipen, L., & Liland, K. H. (2015). micropan: an R-package for microbial pan-genomics. *BMC bioinformatics*, 16(1), 79.
- Takeuchi, K., & Mitsuhashi, I. (2020). Complete Genome Sequences of Two Strains of *Xanthomonas campestris* pv. *campestris* Isolated in Japan. *Microbiology Resource Announcements*, 9(2).
- Tang, J. L., Gough, C. L., & Daniels, M. J. (1990). Cloning of genes involved in negative regulation of production of extracellular enzymes and polysaccharide of *Xanthomonas campestris* pathovar *campestris*. *Molecular and General Genetics MGG*, 222(1), 157-160.
- Tettelin, H., Maignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., ... & DeBoy, R. T. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences*, 102(39), 13950-13955.
- Tian, X., Zhang, Z., Yang, T., Chen, M., Li, J., Chen, F., ... & Wu, J. (2016). Comparative genomics analysis of *Streptomyces* species reveals their adaptation to the marine environment and their diversity at the genomic level. *Frontiers in Microbiology*, 7, 998.
- Tobón, R. A. S., Arcila, J. D. Z., & Cardona, M. E. R. (2008). Producción de goma de xantano empleando cáscara de piña. *Investigaciones Aplicadas*, 2(2), 30-34.
- Tsukuda, M., Kitahara, K., & Miyazaki, K. (2017). Comparative RNA function analysis reveals high functional similarity between distantly related bacterial 16 S rRNAs. *Scientific reports*, 7(1), 1-8.
- Valdés, A., Ramos, M., Beltrán, A., Jiménez, A., & Garrigós, M. C. (2017). State of the art of antimicrobial edible coatings for food packaging applications. *Coatings*, 7(4), 56.
- Vorhölter, F. J., Schneiker, S., Goesmann, A., Krause, L., Bekel, T., Kaiser, O., ... & Sidhu, V. K. (2008). The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. *Journal of biotechnology*, 134(1-2), 33-45.

Wang, Z., Wu, J., Gao, M. J., Zhu, L., & Zhan, X. B. (2017). High production of xanthan gum by a glycerol-tolerant strain *Xanthomonas campestris* WXLB-006. *Preparative Biochemistry and Biotechnology*, 47(5), 468-472.

Xiao, J., Zhang, Z., Wu, J., & Yu, J. (2015). A brief review of software tools for pangenomics. *Genomics, proteomics & bioinformatics*, 13(1), 73-76.

## ANEXOS

### Anexo 1. Descripción de los programas usados en este estudio

#### **BPGA – Bacterial Pan Genome Analysis pipeline**

BPGA es un paquete de software ultrarrápido que proporciona análisis pangenómico completo de microorganismos. Dentro de las características de BPGA es que permite al usuario seleccionar entre tres herramientas diferentes para la agrupación de ortólogos: USEARCH (usando el 50% de corte de identidad de secuencia por defecto; el usuario puede cambiar este valor de corte), CD-HIT y OrthoMCL, siendo la primera la herramienta de agrupación predeterminada (Chaudhari et al., 2016).

BPGA realiza un paso de preprocesamiento para preparar datos de secuencia para la agrupación. La salida agrupada se procesa para generar una matriz binaria de presencia/ ausencia de genes.

BPGA tiene siete módulos funcionales: (i) Análisis del perfil del pangenoma, (ii) Extracción de la secuencia del pangenoma, (iii) Análisis la familia de genes exclusivos, (iv) Análisis filogenético de las especies, (v) Análisis funcional del pangenoma, (vi) Análisis de contenido de GC atípico y (vii) Análisis de subconjuntos (Chaudhari et al., 2016). Los primeros tres módulos se integran como un solo paso llamado Análisis de pangenoma predeterminado, que incluye la clasificación de grupos ortólogos en genes centrales, accesorios y únicos, extracción de sus secuencias, construcción y extrapolación de las gráficas del genoma central y del pangenoma e identificación de familias de genes que están presentes o ausentes en un genoma específico o subconjunto de genomas.

El módulo sobre análisis filogenético de especies proporciona opciones para la construcción de árboles sobre la base de la matriz del pangenoma, el genoma central y la Tipificación de Secuencia de Múltiples locus *in silico* (MLST – por sus siglas en inglés). El módulo para el Análisis Funcional del Pangenoma realiza las asignaciones COG (*Clusters of Orthologous Groups of proteins*) y KEGG (*Kyoto*

*Encyclopedia of Genes and Genomes* - colección de bases de datos en línea de genomas y rutas enzimáticas) para secuencias representativas de todas las familias de genes ortólogos y realiza un análisis funcional comparativo para genoma central, accesorio y único.

El módulo de Análisis de contenido de GC atípico es una característica única de BPGA, extrae las secuencias de genes que tienen contenidos atípicos de Guanina y Citocina, es decir, genes con contenidos de G + C considerablemente más altos o bajos del contenido promedio de G + C de los genomas respectivos (es decir, casos potenciales de transferencia horizontal de genes) (Chaudhari et al., 2016).

### **R-MICROPAN**

R es una herramienta gratuita e independiente del sistema operativo muy utilizada en bioinformática. El paquete micropan ofrece un conjunto de funciones específicamente diseñadas para la pangenómica microbiana. Una estructura de datos central en un análisis pangenoma es la matriz de pangenoma. Cualquier análisis de pangenoma se divide típicamente en dos fases; primero, los cálculos pesados necesarios para establecer la matriz de pangenoma, y luego lo más interesante analizar dónde se usa a menudo la matriz como entrada.

Cualquier estudio de pangenoma requiere un conjunto de archivos FASTA que contengan las secuencias de proteínas de todos los genes en cada genoma. La agrupación de genes en familias de genes clásicas (grupos de ortólogos) es un enfoque común, y significa que todas las secuencias deben compararse con todas las secuencias de forma directa (alineaciones por pares).

Una alternativa a la agrupación basada en ortólogos es escanear todos los genes en busca de dominios de proteínas utilizando HMMER y luego agruparlos por su secuencia ordenada de dominios no superpuestos. La idea completa de esta comparación indirecta es comparar primero todas las proteínas con algún conjunto de secuencias de referencia. Finalmente, la matriz se construye a partir de los resultados de agrupamiento basados en comparaciones de HMMER.

El tamaño del pangenoma puede estimarse mediante métodos sugeridos previamente. La estimación del límite inferior del tamaño del pangenoma es una estimación conservadora, es decir, tiende a estar en el lado más pequeño del tamaño real. El ajuste de un modelo de mezcla binomial también producirá una estimación conservadora del tamaño del pangenoma, así como una estimación del tamaño del genoma central. Tales modelos también dan una buena vista gráfica de la composición del pangenoma con respecto a los tipos de genes, que van desde genes centrales (siempre o casi siempre presentes), genes accesorios (a menudo presentes) hasta genes únicos (observados en solo unos pocos genomas).

La estimación de que el pangenoma sea cerrado o abierto se puede hacer mediante un tipo de modelo de ley de Heaps.

La ley de Heaps es una ley empírica que describe el número de palabras distintas en un documento (o conjunto de documentos) en función de la longitud del documento, y está representada por la siguiente ecuación:

$$n = kN^{-\alpha}$$

En un contexto genético,  $n$  es el número esperado de genes para un número dado de genomas,  $N$  es el número de genomas y  $k$  y  $\alpha$  ( $\alpha = 1-\gamma$ ) son parámetros libres que se determinan empíricamente (Guimaraes et al., 2015)

Otra característica que describe la diversidad pangenómica es la fluidez genómica también implementada como una función en el paquete de micropan. La fluidez es muy similar a la distancia de Jaccard, que es una medida de superposición entre genomas con respecto a sus grupos de genes. Se pueden calcular tanto las distancias de Jaccard como las distancias de Manhattan entre genomas (Snipen & Liland, 2015), siendo esta última definida como la suma de las diferencias entre cada elemento de 2 genomas (Barh et al., 2020); ambas distancias consideran por defecto solo la presencia y ausencia de cualquier grupo de genes en un genoma.

### **AntiSMASH - Antibiotic and Secondary Metabolite Analysis Shell**

Los metabolitos secundarios producidos por microorganismos son la principal fuente de compuestos bioactivos que se usan como medicamentos antimicrobianos y anticancerígenos, fungicidas, herbicidas y pesticidas. La creciente disponibilidad de genomas microbianos ha establecido los análisis genómicos como métodos muy importantes para la identificación de sus grupos de genes biosintéticos (BGC). La rápida disminución en el costo de la secuenciación del genoma ahora permite el descubrimiento de cientos o incluso miles de grupos de genes que codifican la maquinaria biosintética para estos compuestos, además, la identificación efectiva *in silico* de los objetivos más prometedores dentro de los genomas es esencial para la extracción exitosa de las riquezas genómicas disponibles (Medema et al., 2011).

antiSMASH es un software para la identificación de grupos biosintéticos de metabolitos secundarios a partir de la entrada genómica y la predicción de productos producidos por los grupos identificados (Blin et al., 2014). En antiSMASH, el usuario, que no necesita capacitación especializada en bioinformática, puede cargar secuencias del genoma microbiano que luego se extraen para vías

biosintéticas de metabolitos secundarios de una manera totalmente automatizada.

antiSMASH permite identificar loci biosintéticos que cubren todo el rango de clases conocidas de compuestos de metabolitos secundarios (policétidos, péptidos no ribosomales, terpenos, aminoglucósidos, aminocoumarinas, indolocarbazoles, lantibióticos, bacteriocinas, nucleósidos, betalactámicos, butirolactonas, sideróforos, melaninas y otros). Alinea las regiones identificadas a nivel del grupo de genes con sus parientes más cercanos de una base de datos que contiene todos los otros grupos de genes conocidos, e integra o entrecruza todos los métodos de análisis de genes específicos de metabolitos secundarios previamente disponibles en una vista interactiva (Medema et al., 2011)

Con los rápidos avances en las técnicas de secuenciación de nueva generación y la metagenómica, la genómica bacteriana ha cambiado. Desde estudios de un solo genoma hasta la comparación cruzada de cientos de miles de organismos a diferentes escalas de resolución filogenética. Un pangenoma se define como el conjunto de todas las familias de genes únicos que se encuentran en una o más cepas de una especie procarionta. Debido a la amplia diversidad dentro de las especies en el mundo microbiano, el pangenoma es a menudo muchas veces más grande que un solo genoma. Los estudios de pangenomas se han vuelto populares debido al fácil acceso a los datos de la secuencia del genoma completo para procariontas. Un estudio pangenómico revela diversidad de especies y familias de genes (Snipen & Liland, 2015).

### **MAUVE - Alineación múltiple del genoma con ganancia, pérdida y reordenamiento genético**

La alineación precisa del genoma representa un requisito previo necesario para innumerables análisis genómicos comparativos. Durante el curso de la evolución, los genomas experimentan procesos mutacionales tanto locales como a gran escala. Las mutaciones locales afectan solo un pequeño número de nucleótidos e incluyen la sustitución e inserción o eliminación de nucleótidos. Las mutaciones a gran escala pueden incluir ganancia, pérdida o duplicación de segmentos grandes, generados por recombinación desigual u otros procesos. La tarea de alineación del genoma busca identificar los nucleótidos homólogos en dos o más genomas, es decir, una alineación del genoma identifica los nucleótidos que descienden de un solo sitio en algún organismo ancestral (Darling et al., 2010).

Se espera que la comprensión de las tasas y patrones de cada tipo de cambio produzca información sobre varios procesos biológicos. El advenimiento de la secuenciación del genoma proporciona los datos necesarios para caracterizar las tasas y los patrones de evolución del genoma. Las alineaciones del genoma

pueden identificar cambios evolutivos en el ADN alineando regiones homólogas de secuencia. Mauve es un paquete de software que intenta alinear regiones *ortólogas* (entre secuencias en la que la divergencia acontece tras un evento de especiación), *parálogas* (condición evolutiva en la que la divergencia observada acontece tras un evento de duplicación génica) y *xenólogas* (relación entre secuencias dada por un evento de transferencia horizontal) entre dos o más secuencias del genoma que han sufrido cambios tanto locales como a gran escala. El software está disponible en: <http://darlinglab.org/mauve/user-guide/introduction.html> (Darling et al., 2010). Mauve representa el primer sistema de alineación que integra el análisis de eventos evolutivos a gran escala con la alineación tradicional de múltiples secuencias. Al integrar estos pasos de análisis previamente separados, Mauve proporciona facilidad de uso y sensibilidad adicional sobre otros sistemas, al comparar genomas con reordenamientos significativos. Mauve identifica y alinea regiones de colinealidad local llamadas bloques colineales locales (LCB). Cada bloque localmente colineal es una región homóloga de secuencia compartida por dos o más de los genomas en estudio, y no contiene ningún reordenamiento de secuencia homóloga (Darling et al., 2004).

## Anexo 2. Resultados de antiSMASH para cada cepa

Cepa	Resultado ANTISMASH																								
ATCC 33913	<p>NC_003902.1 (Xanthomonas campestris pv. campestris str. ATCC 33913)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>siderophore</td> <td>3,607,435</td> <td>3,622,159</td> <td>xanthoferrin</td> <td>100%</td> </tr> <tr> <td>Region 2</td> <td>NRPS</td> <td>4,570,582</td> <td>4,614,562</td> <td></td> <td></td> </tr> <tr> <td>Region 3</td> <td>arylpolyene</td> <td>4,721,273</td> <td>4,762,469</td> <td>xanthomonadin I</td> <td>71%</td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	siderophore	3,607,435	3,622,159	xanthoferrin	100%	Region 2	NRPS	4,570,582	4,614,562			Region 3	arylpolyene	4,721,273	4,762,469	xanthomonadin I	71%
	Region	Type	From	To	Most similar known cluster	Similarity																			
Region 1	siderophore	3,607,435	3,622,159	xanthoferrin	100%																				
Region 2	NRPS	4,570,582	4,614,562																						
Region 3	arylpolyene	4,721,273	4,762,469	xanthomonadin I	71%																				
8004	<p>NC_007086.1 (Xanthomonas campestris pv. campestris str. 8004)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>siderophore</td> <td>1,325,930</td> <td>1,340,654</td> <td>xanthoferrin</td> <td>100%</td> </tr> <tr> <td>Region 2</td> <td>NRPS</td> <td>4,645,246</td> <td>4,689,226</td> <td></td> <td></td> </tr> <tr> <td>Region 3</td> <td>arylpolyene</td> <td>4,795,391</td> <td>4,836,587</td> <td>xanthomonadin I</td> <td>71%</td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	siderophore	1,325,930	1,340,654	xanthoferrin	100%	Region 2	NRPS	4,645,246	4,689,226			Region 3	arylpolyene	4,795,391	4,836,587	xanthomonadin I	71%
Region	Type	From	To	Most similar known cluster	Similarity																				
Region 1	siderophore	1,325,930	1,340,654	xanthoferrin	100%																				
Region 2	NRPS	4,645,246	4,689,226																						
Region 3	arylpolyene	4,795,391	4,836,587	xanthomonadin I	71%																				

<p><b>CN18</b></p>	<p>NZ_CP017319.1 (Xanthomonas campestris pv. campestris str. CN18)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1.1</td> <td>lassopeptide <a href="#">↗</a></td> <td>733,040</td> <td>755,453</td> <td></td> <td></td> </tr> <tr> <td>Region 1.2</td> <td>siderophore <a href="#">↗</a></td> <td>1,329,589</td> <td>1,344,313</td> <td>xanthoferrin <a href="#">↗</a></td> <td>Other 100%</td> </tr> <tr> <td>Region 1.3</td> <td>NRPS <a href="#">↗</a></td> <td>4,547,189</td> <td>4,591,169</td> <td></td> <td></td> </tr> <tr> <td>Region 1.4</td> <td>arylpolylene <a href="#">↗</a></td> <td>4,695,325</td> <td>4,736,521</td> <td>xanthomonadin I <a href="#">↗</a></td> <td>Other 71%</td> </tr> </tbody> </table> <p>No secondary metabolite regions were found in these records:            NZ_CP017321.1 (Xanthomonas campestris pv. campestris str. CN18)            NZ_CP017320.1 (Xanthomonas campestris pv. campestris str. CN18)            NZ_CP017322.1 (Xanthomonas campestris pv. campestris str. CN18)</p>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1.1	lassopeptide <a href="#">↗</a>	733,040	755,453			Region 1.2	siderophore <a href="#">↗</a>	1,329,589	1,344,313	xanthoferrin <a href="#">↗</a>	Other 100%	Region 1.3	NRPS <a href="#">↗</a>	4,547,189	4,591,169			Region 1.4	arylpolylene <a href="#">↗</a>	4,695,325	4,736,521	xanthomonadin I <a href="#">↗</a>	Other 71%
Region	Type	From	To	Most similar known cluster	Similarity																										
Region 1.1	lassopeptide <a href="#">↗</a>	733,040	755,453																												
Region 1.2	siderophore <a href="#">↗</a>	1,329,589	1,344,313	xanthoferrin <a href="#">↗</a>	Other 100%																										
Region 1.3	NRPS <a href="#">↗</a>	4,547,189	4,591,169																												
Region 1.4	arylpolylene <a href="#">↗</a>	4,695,325	4,736,521	xanthomonadin I <a href="#">↗</a>	Other 71%																										
<p><b>CN17</b></p>	<p>NZ_CP017307.1 (Xanthomonas campestris pv. campestris str. CN17)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>lassopeptide <a href="#">↗</a></td> <td>716,279</td> <td>738,692</td> <td></td> <td></td> </tr> <tr> <td>Region 2</td> <td>siderophore <a href="#">↗</a></td> <td>1,328,574</td> <td>1,343,298</td> <td>xanthoferrin <a href="#">↗</a></td> <td>Other 100%</td> </tr> <tr> <td>Region 3</td> <td>NRPS <a href="#">↗</a></td> <td>4,534,541</td> <td>4,578,521</td> <td></td> <td></td> </tr> <tr> <td>Region 4</td> <td>arylpolylene <a href="#">↗</a></td> <td>4,682,669</td> <td>4,723,865</td> <td>xanthomonadin I <a href="#">↗</a></td> <td>Other 71%</td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	lassopeptide <a href="#">↗</a>	716,279	738,692			Region 2	siderophore <a href="#">↗</a>	1,328,574	1,343,298	xanthoferrin <a href="#">↗</a>	Other 100%	Region 3	NRPS <a href="#">↗</a>	4,534,541	4,578,521			Region 4	arylpolylene <a href="#">↗</a>	4,682,669	4,723,865	xanthomonadin I <a href="#">↗</a>	Other 71%
Region	Type	From	To	Most similar known cluster	Similarity																										
Region 1	lassopeptide <a href="#">↗</a>	716,279	738,692																												
Region 2	siderophore <a href="#">↗</a>	1,328,574	1,343,298	xanthoferrin <a href="#">↗</a>	Other 100%																										
Region 3	NRPS <a href="#">↗</a>	4,534,541	4,578,521																												
Region 4	arylpolylene <a href="#">↗</a>	4,682,669	4,723,865	xanthomonadin I <a href="#">↗</a>	Other 71%																										
<p><b>CN12</b></p>	<p>NZ_CP017310.1 (Xanthomonas campestris pv. campestris str. CN12)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>arylpolylene <a href="#">↗</a></td> <td>309,463</td> <td>350,659</td> <td>xanthomonadin I <a href="#">↗</a></td> <td>Other 71%</td> </tr> <tr> <td>Region 2</td> <td>siderophore <a href="#">↗</a></td> <td>3,738,932</td> <td>3,753,656</td> <td>xanthoferrin <a href="#">↗</a></td> <td>Other 100%</td> </tr> <tr> <td>Region 3</td> <td>lassopeptide <a href="#">↗</a></td> <td>4,376,942</td> <td>4,399,355</td> <td></td> <td></td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	arylpolylene <a href="#">↗</a>	309,463	350,659	xanthomonadin I <a href="#">↗</a>	Other 71%	Region 2	siderophore <a href="#">↗</a>	3,738,932	3,753,656	xanthoferrin <a href="#">↗</a>	Other 100%	Region 3	lassopeptide <a href="#">↗</a>	4,376,942	4,399,355								
Region	Type	From	To	Most similar known cluster	Similarity																										
Region 1	arylpolylene <a href="#">↗</a>	309,463	350,659	xanthomonadin I <a href="#">↗</a>	Other 71%																										
Region 2	siderophore <a href="#">↗</a>	3,738,932	3,753,656	xanthoferrin <a href="#">↗</a>	Other 100%																										
Region 3	lassopeptide <a href="#">↗</a>	4,376,942	4,399,355																												
<p><b>B100</b></p>	<p>NC_010688.1 (Xanthomonas campestris pv. campestris)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>siderophore <a href="#">↗</a></td> <td>1,302,241</td> <td>1,316,965</td> <td>xanthoferrin <a href="#">↗</a></td> <td>Other 100%</td> </tr> <tr> <td>Region 2</td> <td>NRPS <a href="#">↗</a></td> <td>4,563,198</td> <td>4,607,178</td> <td></td> <td></td> </tr> <tr> <td>Region 3</td> <td>arylpolylene <a href="#">↗</a></td> <td>4,711,711</td> <td>4,752,907</td> <td>xanthomonadin I <a href="#">↗</a></td> <td>Other 71%</td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	siderophore <a href="#">↗</a>	1,302,241	1,316,965	xanthoferrin <a href="#">↗</a>	Other 100%	Region 2	NRPS <a href="#">↗</a>	4,563,198	4,607,178			Region 3	arylpolylene <a href="#">↗</a>	4,711,711	4,752,907	xanthomonadin I <a href="#">↗</a>	Other 71%						
Region	Type	From	To	Most similar known cluster	Similarity																										
Region 1	siderophore <a href="#">↗</a>	1,302,241	1,316,965	xanthoferrin <a href="#">↗</a>	Other 100%																										
Region 2	NRPS <a href="#">↗</a>	4,563,198	4,607,178																												
Region 3	arylpolylene <a href="#">↗</a>	4,711,711	4,752,907	xanthomonadin I <a href="#">↗</a>	Other 71%																										
<p><b>3811</b></p>	<p>NZ_CP025750.1 (Xanthomonas campestris pv. campestris)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>siderophore <a href="#">↗</a></td> <td>1,295,466</td> <td>1,310,190</td> <td>xanthoferrin <a href="#">↗</a></td> <td>Other 100%</td> </tr> <tr> <td>Region 2</td> <td>NRPS <a href="#">↗</a></td> <td>4,556,627</td> <td>4,600,607</td> <td></td> <td></td> </tr> <tr> <td>Region 3</td> <td>arylpolylene <a href="#">↗</a></td> <td>4,705,126</td> <td>4,746,322</td> <td>xanthomonadin I <a href="#">↗</a></td> <td>Other 71%</td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	siderophore <a href="#">↗</a>	1,295,466	1,310,190	xanthoferrin <a href="#">↗</a>	Other 100%	Region 2	NRPS <a href="#">↗</a>	4,556,627	4,600,607			Region 3	arylpolylene <a href="#">↗</a>	4,705,126	4,746,322	xanthomonadin I <a href="#">↗</a>	Other 71%						
Region	Type	From	To	Most similar known cluster	Similarity																										
Region 1	siderophore <a href="#">↗</a>	1,295,466	1,310,190	xanthoferrin <a href="#">↗</a>	Other 100%																										
Region 2	NRPS <a href="#">↗</a>	4,556,627	4,600,607																												
Region 3	arylpolylene <a href="#">↗</a>	4,705,126	4,746,322	xanthomonadin I <a href="#">↗</a>	Other 71%																										

MAFF302021	<p>NZ_AP019684.1 (<i>Xanthomonas campestris</i> pv. <i>campestris</i>)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>siderophore</td> <td>1,303,829</td> <td>1,318,553</td> <td>xanthoferrin</td> <td>100%</td> </tr> <tr> <td>Region 2</td> <td>NRPS</td> <td>4,538,069</td> <td>4,582,049</td> <td></td> <td></td> </tr> <tr> <td>Region 3</td> <td>arylpolyene</td> <td>4,685,290</td> <td>4,726,486</td> <td>xanthomonadin I</td> <td>71%</td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	siderophore	1,303,829	1,318,553	xanthoferrin	100%	Region 2	NRPS	4,538,069	4,582,049			Region 3	arylpolyene	4,685,290	4,726,486	xanthomonadin I	71%
Region	Type	From	To	Most similar known cluster	Similarity																				
Region 1	siderophore	1,303,829	1,318,553	xanthoferrin	100%																				
Region 2	NRPS	4,538,069	4,582,049																						
Region 3	arylpolyene	4,685,290	4,726,486	xanthomonadin I	71%																				
CN03	<p>NZ_CP017308.1 (<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. CN03)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1.1</td> <td>siderophore</td> <td>1,275,219</td> <td>1,289,943</td> <td>xanthoferrin</td> <td>100%</td> </tr> <tr> <td>Region 1.2</td> <td>arylpolyene</td> <td>4,668,802</td> <td>4,709,998</td> <td>xanthomonadin I</td> <td>71%</td> </tr> </tbody> </table> <p>No secondary metabolite regions were found in these records:  NZ_CP017309.1 (<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. CN03)</p>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1.1	siderophore	1,275,219	1,289,943	xanthoferrin	100%	Region 1.2	arylpolyene	4,668,802	4,709,998	xanthomonadin I	71%						
Region	Type	From	To	Most similar known cluster	Similarity																				
Region 1.1	siderophore	1,275,219	1,289,943	xanthoferrin	100%																				
Region 1.2	arylpolyene	4,668,802	4,709,998	xanthomonadin I	71%																				
CN14	<p>NZ_CP017317.1 (<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. CN14)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1.1</td> <td>siderophore</td> <td>1,306,643</td> <td>1,321,367</td> <td>xanthoferrin</td> <td>100%</td> </tr> <tr> <td>Region 1.2</td> <td>NRPS</td> <td>4,512,739</td> <td>4,556,718</td> <td></td> <td></td> </tr> <tr> <td>Region 1.3</td> <td>arylpolyene</td> <td>4,661,690</td> <td>4,702,886</td> <td>xanthomonadin I</td> <td>71%</td> </tr> </tbody> </table> <p>No secondary metabolite regions were found in these records:  NZ_CP017318.1 (<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. CN14)</p>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1.1	siderophore	1,306,643	1,321,367	xanthoferrin	100%	Region 1.2	NRPS	4,512,739	4,556,718			Region 1.3	arylpolyene	4,661,690	4,702,886	xanthomonadin I	71%
Region	Type	From	To	Most similar known cluster	Similarity																				
Region 1.1	siderophore	1,306,643	1,321,367	xanthoferrin	100%																				
Region 1.2	NRPS	4,512,739	4,556,718																						
Region 1.3	arylpolyene	4,661,690	4,702,886	xanthomonadin I	71%																				
CN15	<p>NZ_CP017323.1 (<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. CN15)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1.1</td> <td>siderophore</td> <td>1,306,639</td> <td>1,321,363</td> <td>xanthoferrin</td> <td>100%</td> </tr> <tr> <td>Region 1.2</td> <td>NRPS</td> <td>4,509,366</td> <td>4,553,345</td> <td></td> <td></td> </tr> <tr> <td>Region 1.3</td> <td>arylpolyene</td> <td>4,658,316</td> <td>4,699,512</td> <td>xanthomonadin I</td> <td>71%</td> </tr> </tbody> </table> <p>No secondary metabolite regions were found in these records:  NZ_CP017325.1 (<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. CN15)  NZ_CP017324.1 (<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. CN15)</p>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1.1	siderophore	1,306,639	1,321,363	xanthoferrin	100%	Region 1.2	NRPS	4,509,366	4,553,345			Region 1.3	arylpolyene	4,658,316	4,699,512	xanthomonadin I	71%
Region	Type	From	To	Most similar known cluster	Similarity																				
Region 1.1	siderophore	1,306,639	1,321,363	xanthoferrin	100%																				
Region 1.2	NRPS	4,509,366	4,553,345																						
Region 1.3	arylpolyene	4,658,316	4,699,512	xanthomonadin I	71%																				
pv. raphani 756C	<p>NC_017271.1 (<i>Xanthomonas campestris</i> pv. <i>raphani</i> 756C)</p>  <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>arylpolyene</td> <td>282,716</td> <td>323,912</td> <td>xanthomonadin I</td> <td>71%</td> </tr> <tr> <td>Region 2</td> <td>NRPS</td> <td>428,313</td> <td>472,296</td> <td></td> <td></td> </tr> <tr> <td>Region 3</td> <td>siderophore</td> <td>3,584,285</td> <td>3,599,009</td> <td>xanthoferrin</td> <td>100%</td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	arylpolyene	282,716	323,912	xanthomonadin I	71%	Region 2	NRPS	428,313	472,296			Region 3	siderophore	3,584,285	3,599,009	xanthoferrin	100%
Region	Type	From	To	Most similar known cluster	Similarity																				
Region 1	arylpolyene	282,716	323,912	xanthomonadin I	71%																				
Region 2	NRPS	428,313	472,296																						
Region 3	siderophore	3,584,285	3,599,009	xanthoferrin	100%																				

ICMP 21080	<p>NZ_CP012145.1 (<i>Xanthomonas campestris</i> pv. <i>campestris</i>)</p> <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>siderophore <a href="#">↗</a></td> <td>1,292,542</td> <td>1,307,266</td> <td>xanthoferrin <a href="#">↗</a> Other</td> <td>100%</td> </tr> <tr> <td>Region 2</td> <td>arylpolyene <a href="#">↗</a></td> <td>4,561,822</td> <td>4,603,018</td> <td>xanthomonadin I <a href="#">↗</a> Other</td> <td>71%</td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	siderophore <a href="#">↗</a>	1,292,542	1,307,266	xanthoferrin <a href="#">↗</a> Other	100%	Region 2	arylpolyene <a href="#">↗</a>	4,561,822	4,603,018	xanthomonadin I <a href="#">↗</a> Other	71%																								
	Region	Type	From	To	Most similar known cluster	Similarity																																					
Region 1	siderophore <a href="#">↗</a>	1,292,542	1,307,266	xanthoferrin <a href="#">↗</a> Other	100%																																						
Region 2	arylpolyene <a href="#">↗</a>	4,561,822	4,603,018	xanthomonadin I <a href="#">↗</a> Other	71%																																						
ICMP 4013	<p>NZ_CP012146.1 (<i>Xanthomonas campestris</i> pv. <i>campestris</i>)</p> <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>siderophore <a href="#">↗</a></td> <td>1,304,600</td> <td>1,319,324</td> <td>xanthoferrin <a href="#">↗</a> Other</td> <td>100%</td> </tr> <tr> <td>Region 2</td> <td>NRPS <a href="#">↗</a></td> <td>4,411,511</td> <td>4,455,491</td> <td></td> <td></td> </tr> <tr> <td>Region 3</td> <td>arylpolyene <a href="#">↗</a></td> <td>4,560,331</td> <td>4,601,527</td> <td>xanthomonadin I <a href="#">↗</a> Other</td> <td>71%</td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	siderophore <a href="#">↗</a>	1,304,600	1,319,324	xanthoferrin <a href="#">↗</a> Other	100%	Region 2	NRPS <a href="#">↗</a>	4,411,511	4,455,491			Region 3	arylpolyene <a href="#">↗</a>	4,560,331	4,601,527	xanthomonadin I <a href="#">↗</a> Other	71%																		
Region	Type	From	To	Most similar known cluster	Similarity																																						
Region 1	siderophore <a href="#">↗</a>	1,304,600	1,319,324	xanthoferrin <a href="#">↗</a> Other	100%																																						
Region 2	NRPS <a href="#">↗</a>	4,411,511	4,455,491																																								
Region 3	arylpolyene <a href="#">↗</a>	4,560,331	4,601,527	xanthomonadin I <a href="#">↗</a> Other	71%																																						
pv. <i>badrii</i>	<p>NZ_CP051651.1 (<i>Xanthomonas campestris</i> pv. <i>badrii</i>)</p> <table border="1"> <thead> <tr> <th>Region</th> <th>Type</th> <th>From</th> <th>To</th> <th>Most similar known cluster</th> <th>Similarity</th> </tr> </thead> <tbody> <tr> <td>Region 1</td> <td>NRPS <a href="#">↗</a></td> <td>1,321,921</td> <td>1,392,844</td> <td>syringomycin <a href="#">↗</a> NRP</td> <td>17%</td> </tr> <tr> <td>Region 2</td> <td>bacteriocin <a href="#">↗</a></td> <td>2,052,036</td> <td>2,061,572</td> <td></td> <td></td> </tr> <tr> <td>Region 3</td> <td>siderophore <a href="#">↗</a></td> <td>3,155,860</td> <td>3,170,590</td> <td>xanthoferrin <a href="#">↗</a> Other</td> <td>100%</td> </tr> <tr> <td>Region 4</td> <td>NRPS <a href="#">↗</a></td> <td>3,922,535</td> <td>3,966,515</td> <td></td> <td></td> </tr> <tr> <td>Region 5</td> <td>arylpolyene <a href="#">↗</a></td> <td>4,080,639</td> <td>4,121,835</td> <td>xanthomonadin I <a href="#">↗</a> Other</td> <td>71%</td> </tr> <tr> <td>Region 6</td> <td>bacteriocin <a href="#">↗</a></td> <td>4,123,883</td> <td>4,134,722</td> <td></td> <td></td> </tr> </tbody> </table>	Region	Type	From	To	Most similar known cluster	Similarity	Region 1	NRPS <a href="#">↗</a>	1,321,921	1,392,844	syringomycin <a href="#">↗</a> NRP	17%	Region 2	bacteriocin <a href="#">↗</a>	2,052,036	2,061,572			Region 3	siderophore <a href="#">↗</a>	3,155,860	3,170,590	xanthoferrin <a href="#">↗</a> Other	100%	Region 4	NRPS <a href="#">↗</a>	3,922,535	3,966,515			Region 5	arylpolyene <a href="#">↗</a>	4,080,639	4,121,835	xanthomonadin I <a href="#">↗</a> Other	71%	Region 6	bacteriocin <a href="#">↗</a>	4,123,883	4,134,722		
Region	Type	From	To	Most similar known cluster	Similarity																																						
Region 1	NRPS <a href="#">↗</a>	1,321,921	1,392,844	syringomycin <a href="#">↗</a> NRP	17%																																						
Region 2	bacteriocin <a href="#">↗</a>	2,052,036	2,061,572																																								
Region 3	siderophore <a href="#">↗</a>	3,155,860	3,170,590	xanthoferrin <a href="#">↗</a> Other	100%																																						
Region 4	NRPS <a href="#">↗</a>	3,922,535	3,966,515																																								
Region 5	arylpolyene <a href="#">↗</a>	4,080,639	4,121,835	xanthomonadin I <a href="#">↗</a> Other	71%																																						
Region 6	bacteriocin <a href="#">↗</a>	4,123,883	4,134,722																																								

### Anexo 3. Resultados MAUVE

