



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

“Ensamblaje y anotación de los endosimbiontes bacterianos *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* del saltahoja de césped gris (*Exitianus exitiosus*)”

Armas Jiménez, Arly Camila

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Trabajo de titulación, previo a la obtención del título de Ingeniera en Biotecnología

Flores Flor, Francisco Javier PhD.

1 de marzo del 2021

Resultado del análisis de Urkund



Document Information

Analyzed document	Armas_Arly_TRABAJO DE TITULACIÓN_Correcciones.docx (D96781542)
Submitted	2/28/2021 6:39:00 PM
Submitted by	
Submitter email	fjflores2@espe.edu.ec
Similarity	0%
Analysis address	fjflores2.espe@analysis.arkund.com

Sources included in the report



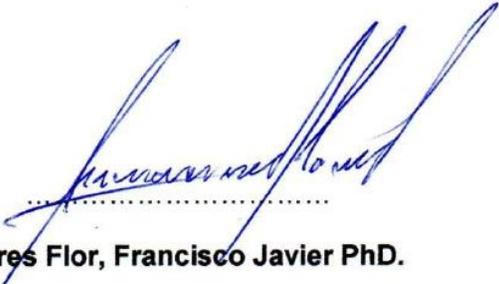
Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Certificación

Certifico que el trabajo de titulación, “**Ensamblaje y anotación de los endosimbiontes bacterianos *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* del saltahojas de césped gris (*Exitianus exitiosus*)**” fue realizado por la señorita **Armas Jiménez, Arly Camila** el cual ha sido revisado y analizado en su totalidad por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 1 de marzo de 2021



Flores Flor, Francisco Javier PhD.

C. C. 1713443479



Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Responsabilidad de autoría

Yo, **Armas Jiménez, Arly Camila**, con cédula de ciudadanía n°1724212087, declaro que el contenido, ideas y criterios del trabajo de titulación: **Ensamblaje y anotación de los endosimbiontes bacterianos *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* del saltahoja de césped gris (*Exítianus exitiosus*)** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 1 de marzo de 2021

Camila - Armas

Armas Jiménez, Arly Camila

C.C.: 1724212087



Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Autorización de publicación

Yo **Armas Jiménez, Arly Camila**, con cédula de ciudadanía n°1724212087, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **Ensamblaje y anotación de los endosimbiontes bacterianos *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* del saltahojas de césped gris (*Exitianus exitiosus*) en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi/nuestra responsabilidad.**

Sangolquí, 1 de marzo de 2021

Arly Camila Jiménez

Armas Jiménez, Arly Camila

C.C.: 1724212087

Dedicatoria

A mis padres Omar Armas, Gabriela Jiménez y a mi abuela Alegría Albán por ser mi apoyo incondicional, por estar a mi lado y velar por mi formación espiritual y académica. Todos mis logros siempre serán dedicados a ustedes por todo su esfuerzo y amor.

Arly Camila Armas Jiménez

Agradecimientos

En primer lugar, quiero agradecer a Astri Wayadande PhD. y al MSc. Christian Ayala por creer en mí y permitirme participar en uno de sus proyectos de investigación en colaboración con la Universidad Estatal de Oklahoma (OSU).

Al Dr. Francisco Flores y M.Sc. Alma Koch del Departamento de Ciencias de la Vida de la Universidad de las Fuerzas Armadas – ESPE por ser unos excelentes docentes y guías para elaborar mi trabajo de titulación. Además, por ser un pilar muy importante en mi formación profesional, les estoy muy agradecida por impartirme sus conocimientos y ser un ejemplo a seguir para todos sus estudiantes.

A mis padres Omar y Gabriela; por darme la vida y desde pequeña siempre velar por mi educación, por enseñarme valores, por sus consejos y amor incondicional. Mis logros son sólo el fruto de su arduo esfuerzo y estaré siempre agradecida con Dios por tenerlos en mi vida.

A mi abuela Alegría por ser mi segunda madre, por siempre darme un espacio en su mesa y ser mi compañía desde pequeña. A toda mi familia por su apoyo y amor.

A mi mejor amiga Lorena, porque desde pequeñas hemos estado juntas y sé que siempre podré contar con ella como una hermana. También a Anita, por su amistad que ha perdurado desde el colegio y por siempre estar pendiente de mí.

A Luis por su amor, por ser mi compañero de aventuras, viajes y comidas, le agradezco por el apoyo incondicional que me ha mostrado desde que entró a mi vida.

A Michelle por su amistad y apoyo durante la elaboración de mi trabajo de titulación, le agradezco su paciencia y por compartir sus conocimientos en bioinformática conmigo.

A mis amigos Brandon, Doménica, Gaby, Alfredo, Tannya, María Fernanda, Xavier, Juan Pablo y Daniela por brindarme su más sincera amistad, por compartir conmigo la universidad y porque a pesar de que muchas veces no compartimos aulas durante toda la carrera, siempre estuvieron a mi lado apoyándome, dándome un consejo, un abrazo o simplemente una sonrisa que bastó para formar la linda amistad que tengo con ustedes, espero seguir siendo su amiga y les agradezco por todos los momentos que hemos pasado y por los que viviremos juntos.

Índice de contenido

Resultado del análisis de Urkund.....	2
Certificación	3
Responsabilidad de autoría.....	4
Autorización de publicación.....	5
Dedicatoria.....	6
Agradecimientos	7
Listado de tablas.....	13
Listado de figuras.....	13
Listado de abreviaturas	15
Resumen.....	17
Abstract.....	18
Capítulo I: Introducción.....	19
Formulación del problema	19
Justificación del problema.....	20
Objetivos.....	22
Objetivo General.....	22
Objetivos Específicos	22
Capítulo II: Marco teórico.....	23
Saltahojas del césped gris	23
Simbiontes bacterianos.....	24

	10
<i>Candidatus Sulcia muelleri</i>	26
<i>Candidatus Nasuia deltocephalinicola</i>	26
Secuenciación de alto rendimiento.....	28
Ensamblaje de genomas.....	29
Anotación de genomas	30
Análisis filogenéticos.....	30
Métodos de inferencia filogenética basados en distancias	32
Métodos de inferencia filogenética basados en caracteres	33
Hipótesis.....	34
Capítulo III: Materiales y métodos	35
Ensamblaje de genomas de <i>Candidatus Sulcia muelleri</i> y <i>Candidatus Nasuia deltocephalinicola</i>	35
<i>Descarga de las lecturas de secuenciación de Exitianus exitiosus</i>	35
<i>Mapeo de las lecturas de secuenciación contra los genomas de Candidatus Sulcia muelleri y Candidatus Nasuia deltocephalinicola</i>	35
<i>Ensamblaje de las lecturas mapeadas</i>	36
Anotación del genoma de <i>Candidatus Sulcia muelleri</i>	36
<i>Comparación de diferentes programas de anotación</i>	36
<i>Visualización del genoma anotado de Candidatus Sulcia muelleri</i>	37
<i>Determinación de rutas metabólicas</i>	37
Análisis filogenético de <i>Candidatus Sulcia muelleri</i>	38
<i>Secuencias incluidas en el análisis</i>	38

	11
<i>Determinación de genes ortólogos</i>	40
<i>Alineamiento múltiple</i>	40
<i>Determinación del mejor modelo de sustitución</i>	41
<i>Construcción del árbol filogenético por máxima verosimilitud</i>	41
<i>Construcción del árbol filogenético por inferencia bayesiana</i>	41
<i>Visualización de los árboles filogenéticos</i>	42
Capítulo IV: Resultados	43
Ensamblaje de genomas de <i>Candidatus Sulcia muelleri</i> y <i>Candidatus Nasuia deltocephalinicola</i>	43
<i>Calidad de las lecturas de secuenciación de Exitianus exitiosus</i>	43
<i>Mapeo de las lecturas de secuenciación contra los genomas de Candidatus Sulcia muelleri y Candidatus Nasuia deltocephalinicola</i>	45
<i>Ensamblaje de las lecturas mapeadas</i>	46
Anotación del genoma de <i>Candidatus Sulcia muelleri</i>	55
<i>Comparación de diferentes programas de anotación</i>	55
<i>Visualización del genoma anotado de Candidatus Sulcia muelleri</i>	56
<i>Determinación de rutas metabólicas</i>	57
Análisis filogenético de <i>Candidatus Sulcia muelleri</i>	59
<i>Determinación de genes ortólogos</i>	59
<i>Alineamiento múltiple</i>	60
<i>Determinación del mejor modelo de sustitución</i>	62

<i>Construcción del árbol filogenético por máxima verosimilitud e inferencia bayesiana</i>	63
Capítulo V: Discusión	66
Capítulo VI: Conclusiones	73
Capítulo VII: Recomendaciones	74
Bibliografía	75

Listado de tablas

Tabla 1 Genomas bacterianos considerados en la filogenia de <i>Candidatus Sulcia muelleri</i>	38
Tabla 2 Número de contigs y scaffolds obtenidos con un tamaño de k-mer de 77 para <i>Candidatus Sulcia muelleri</i> y <i>Candidatus Nasuia deltocephalinicola</i>	48
Tabla 3 Análisis de calidad del ensamblaje de <i>Candidatus Sulcia muelleri</i>	49
Tabla 4 Análisis de calidad del ensamblaje de <i>Candidatus Nasuia deltocephalinicola</i> ..	50
Tabla 5 Identificación de homólogos a los contigs obtenidos del ensamblaje de <i>Candidatus Sulcia muelleri</i>	51
Tabla 6 Identificación de homólogos a los contigs obtenidos del ensamblaje de <i>Candidatus Nasuia deltocephalinicola</i>	52
Tabla 7 Parámetros calculados por ModelTest-NG para determinar el mejor modelo de sustitución	63

Listado de figuras

Figura 1 Morfología del saltahoja de césped gris (<i>Exitianus exitiosus</i>)	23
Figura 2 Transmisión transovárica de endosimbiontes bacterianos.....	25
Figura 3 Distribución de los endosimbiontes en el bacterioma de los insectos	27
Figura 4 Anatomía de un árbol filogenético.....	31
Figura 5 Calidad de las secuencias por base de las lecturas de secuenciación de alto rendimiento del saltahoja del césped gris (<i>Exitianus exitiosus</i>).....	43
Figura 6 Contenido de CG en las lecturas de secuenciación de alto rendimiento del saltahoja del césped gris (<i>Exitianus exitiosus</i>).....	44

Figura 7 Contenido de adaptadores en las lecturas de secuenciación de alto rendimiento del saltahoja del césped gris (<i>Exitianus exitiosus</i>).....	45
Figura 8 Número de lecturas alineadas con los genomas de <i>Candidatus Sulcia muelleri</i> y <i>Candidatus Nasuia deltocephalinicola</i>	46
Figura 9 Longitud del contigs más largo obtenido con diferentes tamaños de k-mers para el ensamblaje de <i>Candidatus Sulcia muelleri</i> y <i>Candidatus Nasuia deltocephalinicola</i>	47
Figura 10 Representación del cromosoma de <i>Candidatus Sulcia muelleri</i> , endosimbionte del saltahoja del césped gris (<i>Exitianus exitiosus</i>).....	53
Figura 11 Interfaz de Geneious v2021.0.3 de la región de 1264 bp no cubierta por los scaffolds de <i>Candidatus Sulcia muelleri</i>	54
Figura 12 Interfaz de Geneious v2021.0.3 de la región de 94 bp no cubierta por los scaffolds de <i>Candidatus Sulcia muelleri</i>	55
Figura 13 Número de genes identificados en el genoma de <i>Candidatus Sulcia muelleri</i> por diferentes programas de anotación.....	56
Figura 14 Representación esquemática del genoma anotado de <i>Candidatus Sulcia muelleri</i>	57
Figura 15 Distribución porcentual de las rutas metabólicas de <i>Candidatus Sulcia muelleri</i>	58
Figura 16 Número de genes codificantes de enzimas involucrados en la síntesis de aminoácidos esenciales en el genoma de <i>Candidatus Sulcia muelleri</i>	59
Figura 17 Resultado de M-Coffee para el alineamiento múltiple del gen infC	61
Figura 18 Resultado de Gblocks 0.91b para el alineamiento múltiple del gen infC	62
Figura 19 Topología del árbol filogenético de <i>Candidatus Sulcia muelleri</i> construido por máxima verosimilitud en RAxML-HPC BlackBox v8.2.12	65

Listado de abreviaturas

ADN: Ácido Desoxirribonucleico

HTS: Secuenciación de alto rendimiento

OLC: Overlap/Layout/Consensus

ORF: Open reading frame o marco de lectura abierto

OTU: Unidades Taxonómicas Operacionales

MCMC: Monte Carlo basado en Cadenas de Markov

SRA: Sequence Read Archive

NCBI: National Center for Biotechnology Information

NA: Número de Acceso

CG: Citosinas y Guaninas

SAM: Sequence Alignment Map

BAM: Binary Alignment Map

BLAST: Basic Local Alignment Search Tool

CDS: Región Codificante

ARNt: Ácido Ribonucleico de Transferencia

ARNr: Ácido Ribonucleico Ribosomal

ARNtm: Ácido Ribonucleico Mensajero de Transferencia

bp: Pares de Bases

BEAST: Bayesian Evolutionary Analysis Sampling Trees

Lys: Lisina

Leu: Leucina

Ile: Isoleucina

Val: Valina

Thr: Treonina

Arg: Arginina

Phe: Fenilalanina

Trp: Triptófano

BIC: Criterio de Información Bayesiano

AIC: Criterio de Información Akaike

AICc: Criterio de Información Akaike Corregido

AT: Adeninas y Timinas

Resumen

El saltahoja del césped gris (*Exitianus exitiosus*) es un representante de la subfamilia Deltocephalinae; cuyos miembros se caracterizan por albergar los endosimbiontes *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* para suministrarle los nutrientes que no pueden adquirir en su dieta. Para caracterizar estas bacterias, la presente investigación tuvo como objetivo ensamblar y anotar los endosimbiontes del saltahoja *Exitianus exitiosus*. En primer lugar, se verificó la calidad de las lecturas de secuenciación de alto rendimiento del insecto, se seleccionaron sólo las lecturas de las bacterias y se ensamblaron con un k-mer de 77 en contigs y scaffolds. Para *Ca. Sulcia muelleri* en total se cubrieron 190,004 bp y sus scaffolds se cerraron en un cromosoma. Para *Ca. Nasuia deltocephalinicola* se obtuvieron sólo 3,396 bp de un total de 112,091 bp que contiene su genoma. Por lo tanto, sus secuencias fueron descartadas de la investigación. Posteriormente, en *Ca. Sulcia muelleri* se identificaron 190 regiones codificantes (CDS), 30 ácidos ribonucleicos de transferencia (ARNt), 1 de transferencia mensajero (ARNtm) y 3 ribosomales (ARNr). Asimismo, se determinó que la bacteria le proporciona ocho aminoácidos esenciales a su hospedador que no puede obtener en su dieta. Se construyó un árbol filogenético, en donde *Ca. Sulcia muelleri* de *Exitianus exitiosus* se agrupó en un clado robusto junto con los demás miembros de la subfamilia Deltocephalinae; comprobando que la filogenia de los simbioses es congruente con la de sus anfitriones, ya que estos se transmiten verticalmente dentro de los insectos.

Palabras clave:

- **CANDIDATUS SULCIA MUELLERI**
- **CANDIDATUS NASUIA DELTOCEPHALINICOLA**
- **AMINOÁCIDOS**
- **FILOGENIA**

Abstract

The gray grass leafhopper (*Exitianus exitiosus*) is a representative of the subfamily Deltocephalinae; whose members are characterized by harboring the endosymbionts *Candidatus Sulcia muelleri* and *Candidatus Nasuia deltocephalinicola* to supply them with nutrients that they cannot acquire in their diet. To characterize these bacteria, the present investigation aimed to assemble and annotate the endosymbionts of the leafhopper *Exitianus exitiosus*. First, the quality of the insect high-throughput sequencing reads was checked, only bacterial reads were selected and assembled with a k-mer of 77 into contigs and scaffolds. For *Ca. Sulcia muelleri* in total 190,004 bp were covered and their scaffolds were closed on one chromosome. For *Ca. Nasuia deltocephalinicola* only 3,396 bp were obtained from a total of 112,091 bp contained in its genome. Therefore, its sequences were discarded from the investigation. Subsequently, 190 coding regions (CDS), 30 transfer ribonucleic acids (tRNA), 1 messenger transfer (mRNA) and 3 ribosomal (rRNA) were identified in *Ca. Sulcia muelleri*. It was also determined that the bacterium provides eight essential amino acids to its host that it cannot obtain in its diet. A phylogenetic tree was constructed, where *Ca. Sulcia muelleri* of *Exitianus exitiosus* was grouped in a robust clade together with the other members of the subfamily Deltocephalinae; proving that the phylogeny of the symbionts is congruent with that of their hosts since these are transmitted vertically within insects.

Key words:

- **CANDIDATUS SULCIA MUELLERI**
- **CANDIDATUS NASUIA DELTOCEPHALINICOLA**
- **AMINO ACIDS**
- **PHYLOGENY**

Capítulo I: Introducción

Formulación del problema

Los saltahojas (Hemíptera: Cicadellidae) son una de las familias de insectos más extensa en el mundo, con más de 22.600 especies descritas hasta la actualidad distribuidas en 19 subfamilias (Bertaccini et al., 2019). No obstante, los miembros de ocho subfamilias son considerados como plagas agrícolas y hortícolas por ser vectores de virus y fitoplasmas patógenos que afectan a cultivos de importancia económica a nivel mundial (Bertaccini et al., 2019; Ishii et al., 2013).

Se alimentan exclusivamente de la savia de las plantas por medio de laceramiento y descarga, una estrategia que consiste en introducir sus estiletes mientras secretan saliva para romper las células y luego consumir su contenido (Radcliffe & Lagnaoui, 2007). Sin embargo, las dietas basadas en xilema o floema son incompletas; porque contienen sólo algunos carbohidratos, ácidos orgánicos y aminoácidos no esenciales. En consecuencia, los insectos sólo pueden sintetizar lípidos a partir de los carbohidratos que consumen, pero no proteínas, debido a la falta de precursores nitrogenados esenciales (Ishii et al., 2013). Por lo tanto, su alimentación no logra satisfacer todos los requerimientos nutricionales para el crecimiento y supervivencia de los mismos (Szklarzewicz et al., 2016).

Para contrarrestar su dieta desequilibrada, los insectos han establecidos relaciones simbióticas obligatorias con bacterias y/o levaduras que les proporcionan aminoácidos esenciales y vitaminas (Szklarzewicz et al., 2016). Además algunos poseen simbioses facultativas encargados de defender a sus hospedadores de parasitoides (Oliver et al., 2003) y proporcionarles resistencia a insecticidas (Kikuchi et al., 2012). A cambio, los simbioses obligatorios se mantienen en un entorno

intracelular estable y se transmiten verticalmente entre generaciones de insectos (Mao et al., 2018).

Los análisis moleculares han demostrado que la mayoría de los saltahojas alojan obligatoriamente la bacteria *Candidatus Sulcia muelleri*; no obstante, coexiste con otros simbiosomas como *Candidatus Zinderia insecticola*, *Candidatus Nasuia deltocephalinicola* o *Candidatus Hodgkinia cicadicola*. Asimismo los insectos también pueden albergar varios endosimbiontes facultativos de los géneros bacterianos *Wolbachia* sp., *Rickettsia* sp. y *Cardinium* sp. (Wang et al., 2019).

Justificación del problema

Dentro de los saltahojas considerados como vectores de fitopatógenos, se conoce que el 80% se encuentran dentro de la subfamilia Deltocephalinae (Bertaccini et al., 2019). Uno de sus representantes es el saltahojas del césped gris (*Exitianus exitiosus*), un insecto que transmite virus y micoplasmas cuando consume la savia de plantas (Jones & Medina, 2020). Específicamente, se alimenta de los cultivos de avena (*Avena sativa* L.), cebada (*Hordem vulgare* L.), uva (*Vitis* sp.) (Saguez et al., 2014), trigo (*Triticum aestivum* L.) (Andreason, 2016), maíz (*Zea mays* L.) (Jones & Medina, 2020) y pasto bermuda (*Cynodon dactylon* L.) (Gahm, 2017).

Por otra parte, el éxito evolutivo de los insectos fitófagos como los saltahojas, se debe a la simbiosis establecida con bacterias y/o levaduras para proveerles nutrientes. Específicamente, para la subfamilia Deltocephalinae con las bacterias *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* (Brentassi et al., 2017).

Antiguamente, estos simbiosomas sólo se podían estudiar por técnicas de microscopía, ya que se encuentran alojados en órganos especializados de los insectos denominados bacteriomas. Pero no podían ser identificados, porque son

microorganismos que dependen por completo de sus hospederos y no pueden ser cultivados fuera de ellos (Kikuchi, 2009).

Sin embargo, con el apareamiento de las nuevas tecnologías de secuenciación y el desarrollo de programas bioinformáticos, algunos simbioses se han identificado y secuenciado por completo. Con esta información, se ha podido determinar las relaciones evolutivas de estos microorganismos y las rutas metabólicas que poseen (Wernegreen, 2012).

Actualmente, con las secuencias obtenidas del ácido desoxirribonucleico (ADN) total del saltahojas o de sus bacteriomas, se puede comprobar bioinformáticamente la presencia de estos microorganismos, ensamblar sus genomas, anotarlos y realizar análisis filogenéticos; para así comprender la genética de las bacterias y cómo son capaces de proveerles los nutrientes necesarios a sus hospederos (Commins et al., 2009; Gil & Latorre, 2019).

Objetivos

Objetivo General

- Ensamblar y anotar los endosimbiontes bacterianos *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* del saltahojas de césped gris (*Exitianus exitiosus*).

Objetivos Específicos

- Ensamblar los genomas de *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* a partir de scaffolds obtenidos por secuenciación de alto rendimiento de ADN total del saltahojas de césped gris (*Exitianus exitiosus*).
- Anotar los genomas de los endosimbiontes *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* del saltahojas de césped gris (*Exitianus exitiosus*).
- Realizar un análisis filogenético de las bacterias *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola*.

Capítulo II: Marco teórico

Saltahojas del césped gris

El saltahojas de césped gris (*Exitianus exitiosus*) es un representante de la subfamilia Deltocephalinae; morfológicamente se lo puede identificar por su color marrón, longitud de 3.5 – 4.7 mm y dos puntos negros en la parte superior de su cabeza, pronoto y escutelo. Tiene alas blanquecinas con venas marrones que generalmente se extienden un poco más allá del abdomen y patas traseras especializadas para saltar, con un recubrimiento de vellosidades que contiene una secreción repelente de agua y compuesta por feromonas (Stiller, 2009) (Figura 1). Mantiene una reproducción sexual y luego de la gestación, las hembras introducen sus huevos en la vena principal o en pecíolo de las hojas (Radcliffe & Lagnaoui, 2007).

Figura 1

Morfología del saltahojas de césped gris (Exitianus exitiosus)



Nota. Tomado de *Diversity and abundance of leafhoppers in Canadian vineyardsm* (p.15), por J. Saguez, C. Oliver, A. Hamilton, T. Lowery, L. Stobbs, J. Lasnier, B. Galka, X. Chen, Y. Mauffette & C. Vincent, 2014, Journal of Insect Science.

Se distribuye en México y Estados Unidos, pero se pueden encontrar también en Canadá y las Antillas (Leung, 2005); donde se alimentan exclusivamente del floema de las plantas; principalmente de gramíneas. Es considerado como una plaga porque

alberga fitopatógenos en sus glándulas salivares como el virus del enanismo clorótico del maíz (MCDV) y *Spiroplasma kunkelii* (Jones & Medina, 2020).

En las plantas infectadas, el virus MCDV produce retraso del crecimiento, acortamiento de los entrenudos superiores, amarillamiento, torsión y desgarro de las hojas y clorosis (Stewart et al., 2017). Mientras que *Spiroplasma kunkelii* produce amarillamiento, atrofia, enrojecimiento de los márgenes de las hojas y disminución en el rendimiento de los cultivos (Andreason, 2016).

Simbiontes bacterianos

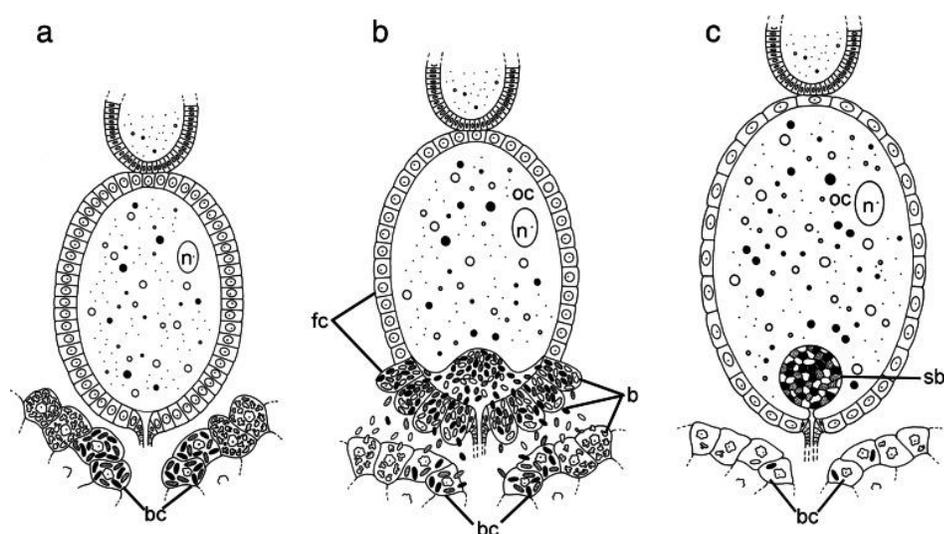
Los insectos que se alimentan de la savia de las plantas del orden Hemiptera han establecido relaciones simbióticas obligatorias con microorganismos capaces de proporcionarles los nutrientes esenciales que no pueden adquirir en el xilema / floema como aminoácidos, esteroides, vitaminas y cofactores (Bennett & Moran, 2013). Por ello, los simbiontes destinan gran parte de su material genético a la síntesis de nutrientes que son más beneficiosos para el huésped (Gil et al., 2004). Estos simbiontes se pueden diferenciar en intra y extracelulares, en función de si se alojan o no dentro de las células de los insectos. Asimismo, se clasifican en primarios y secundarios dependiendo de si el hospedador los necesita para su supervivencia o para proporcionarles beneficios no esenciales (Paniagua Voirol et al., 2018).

Por otra parte, los insectos tienen una estructura similar a un órgano (bacterioma) compuesto por células especializadas (bacteriocitos), cuya función es alojar a los simbiontes primarios y se les denomina endosimbiontes. Debido a que la relación bacteria-insecto es obligatoria, no pueden ser cultivados fuera de su anfitrión y se transfieren verticalmente por transmisión transovárica (Figura 2) (Gil et al., 2004). Por esta razón, las bacterias viven en diferentes especies de insectos y evolucionan independientemente unas de otras; es decir, no realizan intercambio de materia

genético entre microorganismos de otros hospedadores. Por lo tanto, sufren de una reducción en el tamaño de sus genomas durante la evolución (Szklarzewicz et al., 2016).

Figura 2

Transmisión transovárica de endosimbiontes bacterianos



Nota. a) Los bacteriocitos (bc) se aproximan a los ovocitos. b) Los endosimbiontes invaden las células foliculares (fc) e ingresan al ovocito. c) Las bacterias se agrupan formando una bola simbiote (sb). Tomado de *Sulcia symbiont of the leafhopper Macrosteles laevis* (Ribaut, 1927) (*Insecta, Hemiptera, Cicadellidae: Deltocephalinae*) harbors *Arsenophonus* bacteria (p.910), M. Kobiałka, A. Michalik, M. Walczak, Ł. Junkiert & T. Szklarzewicz, 2016, *Protoplasma*.

Actualmente, la disponibilidad de los genomas en bases de datos y análisis filogenéticos han sido muy útiles para comprender el proceso de endosimbiosis en insectos y ver cómo diferentes géneros bacterianos son más predominantes en ciertas familias de insectos (Gil et al., 2004). Hasta la actualidad, se han descritos géneros bacterianos como *Ca. Sulcia* sp., *Ca. Hodgkinia* sp., *Ca. Zinderia* sp., *Ca. Baumannia*

sp., *Ca. Nasuia* sp., *Ca. Vindania* sp., entre otros. Específicamente para la subfamilia Deltoccephalinae, se conoce que albergan obligatoriamente las bacterias *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* (Gil & Latorre, 2019).

Candidatus Sulcia muelleri

Es una bacteria del filo Bacteroidetes, es un simbionte intracelular primario de los insectos que se alimentan de la savia vegetal, se estableció por una infección hace más de 260 millones de años (McCutcheon & Moran, 2007). Es responsable de la biosíntesis de ocho aminoácidos esenciales para su anfitrión como la lisina, arginina, leucina, valina, isoleucina, treonina, triptófano y fenilalanina (Wu et al., 2006), mientras que los demás nutrientes pueden ser proporcionados al insecto por otros simbiontes como *Candidatus Nasuia deltocephalinicola* (Chang et al., 2015).

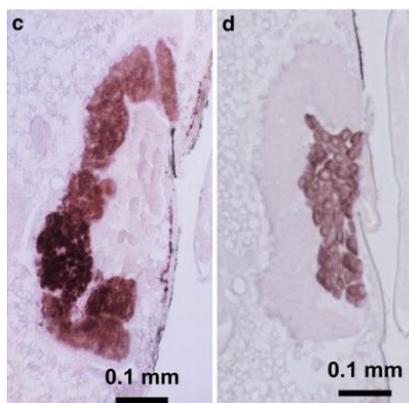
En cuanto a su genoma, tiene un cromosoma circular con un tamaño que varía entre 190 a 277 kb, rango muy bajo para un genoma bacteriano (Chang et al., 2015). No obstante, esto es normal para todos los simbiontes primarios, ya que presentan el síndrome de reducción del genoma; porque durante el tiempo han ido perdiendo los genes necesarios para el mantenimiento de una célula de vida libre (Gil & Latorre, 2019).

Candidatus Nasuia deltocephalinicola

Es una bacteria del filo Proteobacteria, clase Betaproteobacteria; también es un simbionte intracelular al igual que *Candidatus Sulcia muelleri* y coexiste con ella en el mismo bacterioma en diferentes regiones. *Candidatus Nasuia deltocephalinicola* se encuentra en la parte interna del órgano, mientras que *Candidatus Sulcia muelleri* en la parte externa (Figura 3) y también se pueden ubicar en otros órganos como los ovarios de los insectos (Ishii et al., 2013; Noda et al., 2012).

Figura 3

Distribución de los endosimbiontes en el bacterioma de los insectos



Nota. A la izquierda se muestra la distribución de *Candidatus Sulcia muelleri* y a la derecha la de *Candidatus Nasuia deltocephalinicola*. Tomado de *Bacteriome-associated endosymbionts of the green rice leafhopper Nephrotettix cincticeps (Hemiptera: Cicadellidae)* (p.218), H. Noda, K. Watanabe, S. Kawai, F. Yukuhiro, T. Miyoshi, M. Tomizawa, Y. Koizumi, N. Nikoh & T. Fukatsu, 2012, Applied Entomology and Zoology.

En cuanto a su genoma, tiene una longitud aproximada 112 kb con 265 genes, de los cuales 137 codifican para proteínas y tiene un contenido de GC 16,6%. Conserva las vías metabólicas para sintetizar los aminoácidos esenciales histidina y metionina, que *Candidatus Sulcia muelleri* es incapaz de producir (Bennett et al., 2016). Asimismo, a diferencia de *Sulcia*, la bacteria posee genes involucrados en los procesos de replicación, transcripción y traducción. Por ejemplo, para la replicación del ADN tiene los genes *dnaG* y *dnaX* y para la traducción los genes *RlmB*, *rpmCE*, *rsmA*, *rpsP* y *rplAIS* (Mao et al., 2017).

Secuenciación de alto rendimiento

El estudio a nivel genómico de todos los organismos vivos; incluidas las bacterias, se ha ampliado en las últimas décadas debido al rápido desarrollo de las tecnologías de secuenciación (Cao et al., 2017). Específicamente, la secuenciación de alto rendimiento (HTS) comprende un grupo de técnicas diseñadas para secuenciar ácidos nucleicos de forma masiva y en paralelo, a un menor tiempo y costo por base (Rubio et al., 2020).

De manera general, la mayoría de técnicas HTS se basan en: i) segmentar el ADN, ii) marcar con adaptadores los fragmentos para señalar el inicio de la replicación, iii) amplificación de los segmentos marcados, iv) secuenciación del materia genético y v) reporte de las bases identificadas (Rubio et al., 2020).

Entre las plataformas más usadas para HTS se encuentra Illumina; una tecnología que lleva a cabo una “amplificación en puente”, en donde las moléculas de ADN unidas a adaptadores se fijan a un soporte sólido que contiene secuencias complementarias a los adaptadores. Luego el ADN se somete a rondas sucesivas de amplificación para secuenciarse y de esta manera los nucleótidos se van incorporando con un marcador fluorescente diferente para cada base (Slatko et al., 2018).

Por otra parte, HTS se ha usado ampliamente en el estudio de bacterias; inicialmente se enfocaba en la identificación taxonómica mediante la secuenciación de regiones conservadas como el gen de ADNr 16S. Sin embargo, ahora las investigaciones también se están centrando en aplicar estas técnicas de alto rendimiento para secuenciar todo el contenido genético de las muestras. Como resultado de este proceso, se generan millones de lecturas que luego deben ser procesadas computacionalmente para obtener los genomas completos de los organismos de estudio (Cao et al., 2017).

Ensamblaje de genomas

El ensamblaje de genomas es un proceso basado en ordenar correctamente las secuencias de un organismo, para obtener una representación computacional de su genoma. Hasta la actualidad, este proceso es necesario, porque las lecturas generadas por secuenciación son más cortas que todos los genomas y algunos genes reportados (Wajid & Serpedin, 2012).

El proceso se puede llevar a cabo realizando un ensamblaje *de novo* o de mapeo, en función de si ya existe un genoma de referencia para el organismo de estudio. El ensamblaje *de novo* depende completamente de la superposición de las lecturas de secuenciación, mientras que el mapeo se basa en determinar la posición de las lecturas en relación al genoma de referencia para luego ensamblar (F. Li et al., 2019).

En las últimas décadas se han desarrollado varios programas de ensamblaje que utilizan tres categorías de algoritmos computacionales como los gráficos de Overlap/Layout/Consensus (OLC), *De Bruijn* y codiciosos. Específicamente, los gráficos de *De Bruijn* se emplean para lecturas cortas generadas por HTS en plataformas como Illumina (F. Li et al., 2019). En donde el método se basa en probar diferentes tamaños de k-mers para ensamblar las lecturas en segmentos contiguos (contigs) y si hay solapamiento seguirse ensamblando en scaffolds, cromosomas y hasta genomas completos (Rice & Green, 2019).

Generalmente, el flujo de trabajo que se sigue para el ensamblaje de genomas se basa en primero realizar un control de calidad de las lecturas mediante una puntuación de "Phred". Luego se hace un preprocesamiento de los datos originales para eliminar adaptadores y regiones de baja calidad; después se ensamblan las lecturas y si es posible se completan las regiones no cubiertas (Dominguez Del Angel et al., 2018).

Anotación de genomas

Una vez ensamblado un genoma se prosigue a su anotación; un proceso que consiste en identificar la ubicación y el rol biológico de las características genéticas que se encuentra dentro de una secuencia de ADN (Carriço et al., 2018). Este proceso se realiza mediante un análisis estructural y composicional de las secuencias; pero también se pueden usar las secuencias de organismos relacionadas como referencia, debido a su conservación funcional durante la evolución (Dominguez Del Angel et al., 2018; Steward et al., 2017).

La anotación de genomas se enfoca principalmente en identificar de forma correcta las secuencias codificantes de proteínas (CDS), ubicados dentro de un marco de lectura abierto (ORF). Un ORF es un tramo continuo de ADN que comienza con un codón de inicio y finaliza con un codón de parada y contiene toda la información para codificar una proteína funcional (Amatore et al., 2020).

Generalmente los programas emplean un conjunto de secuencias de referencia para entrenar un modelo y después predecir genes. Luego cuando se identifican los CDS, estos se pueden alinear con anotaciones de genomas de referencia para encontrar homólogos (Richardson & Watson, 2013). Por otra parte, la anotación también tiene como objetivo predecir más unidades funcionales del genoma como genes de ARN de transferencia (ARNt) y genes de ARN ribosómico (ARNr) (Dominguez Del Angel et al., 2018). No obstante, los programas ahora también pueden identificar características de orden superior como operones, elementos CRISPR e islas genómicas (Carriço et al., 2018).

Análisis filogenéticos

La filogenia es una ciencia derivada de la biología que se basa en reconstruir la historia evolutiva de los organismos. Tradicionalmente, los análisis filogenéticos se

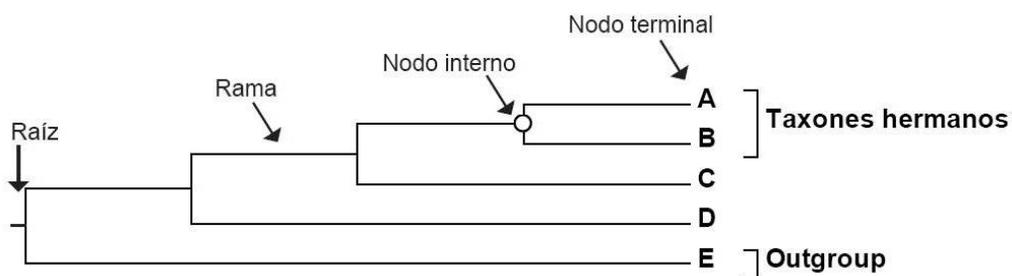
construían en función de características morfológicas; no obstante, la introducción de HTS permitió incluir también la información molecular de los organismos como ADN, ARN o proteínas para así obtener estimaciones más robustas (Young & Gillung, 2020).

Los análisis filogenéticos tienen como objetivo inferir las relaciones filogenéticas entre taxones, obtener información sobre los mecanismos de la evolución molecular y emplear las comparaciones de varios taxones para deducir las funciones putativas de las secuencias de nucleótidos o aminoácidos (Young & Gillung, 2020).

Como resultado de un análisis filogenético (Figura 4) se obtienen representaciones esquemáticas (árboles) que muestran las relaciones entre organismos. En los árboles se pueden visualizar los nodos terminales o unidades taxonómicas operacionales (OTU); que generalmente representan a individuos de una especie o grupos taxonómicos superiores. Estos se enlazan entre sí por medio de ramas hacia su ancestro en común que se identifica como un nodo interno (Gregory, 2008; Mendoza, 2012). Finalmente, también se puede encontrar un “*outgroup*”; una especie lejana pero relacionada con los taxones de estudio “*ingroup*” y sirve para obtener la raíz del árbol (Mendoza, 2012).

Figura 4

Anatomía de un árbol filogenético



Nota. Tomado de *Aportes de la filogenética a la investigación médica* (p.123), por J. Mendoza, 2012, Revista Médica Herediana.

Para construir un árbol filogenético, el flujo de trabajo consiste en primero incluir en el estudio secuencias homólogas. Asimismo, deben ser ortólogas, un tipo de homología en donde los genes de diferentes especies han divergido entre sí debido a un proceso de especiación y de esta manera se recapitula las relaciones entre los taxones de estudio. Después las secuencias se alinean introduciendo huecos (gaps) en las diferentes zonas para permitir que se alineen solo las regiones conservadas. Sin embargo, se debe tener en cuenta que la calidad del alineamiento disminuye con el aumento de la divergencia de las secuencias. Por lo tanto, las regiones ambiguas luego deben ser eliminadas (Kapli et al., 2020). Para los árboles de especies, a continuación, se escogen los enfoques de construcción que pueden ser el enfoque superárbol o supermatriz. El primero es un método basado estimar árboles independientes para varios genes y luego se genera un árbol consenso que representa la filogenia de la especie. Mientras que el segundo enfoque, concatena todos los genes de estudio en un solo alineamiento para estimar un árbol (Janies et al., 2013).

Finalmente, se elige el método de construcción del árbol; estos son del tipo estadísticos y se dividen en dos categorías: i) métodos basados en distancias y ii) métodos basados en caracteres (Kapli et al., 2020).

Métodos de inferencia filogenética basados en distancias

Los métodos basados en distancias calculan la distancia genética entre las especies de estudio según los alineamientos y además emplean una matriz de distancias para construir el árbol. El método más empleado es “Neighbour Joining” (NJ); un algoritmo eficiente con varias implementaciones que permite generar la filogenia de

varias muestras contando las diferencias entre todas las secuencias y uniendo las de menor distancia genética (Mendoza, 2012).

Métodos de inferencia filogenética basados en caracteres

Los métodos basados en caracteres compararan todas las secuencias simultáneamente considerando un carácter/sitio a la vez. Se caracterizan porque son más precisos que los métodos basados en distancias, porque infieren árboles verdaderos o topológicamente similares a los árboles verdaderos. Entre los métodos más empleados se encuentran los de Máxima Verosimilitud (ML) e Inferencia Bayesiana (BI) (Munjal et al., 2018).

La máxima verosimilitud es un método estadístico que se emplea para estimar los parámetros desconocidos de un modelo de probabilidad. En filogenia sirve para inferir árboles, teniendo en cuenta que la topología del árbol se considera como un modelo y los parámetros son la longitud de las ramas y el modelo de sustitución (Yang & Rannala, 2012). El método empieza usando un modelo de evolución y cambiando los parámetros para comparar los posibles árboles que se generan y su capacidad para predecir los datos observados; es decir, busca el árbol con la mayor probabilidad (verosimilitud) de producir las secuencias observadas (Makarenkov et al., 2006).

La inferencia bayesiana es un método que también cambia los parámetros para generar árboles y calcula la verosimilitud de cada uno. Sin embargo, se diferencia de ML porque emplea el algoritmo de Monte Carlo basado en Cadenas de Markov (MCMC) para calcular una probabilidad posterior multiplicando la verosimilitud por una probabilidad a priori. La probabilidad a priori es una probabilidad que se le asignan a los parámetros sin tener en cuenta las secuencias. De esta manera, se obtiene el mejor árbol cuando se maximiza la probabilidad posterior, la cual también le da soporte a los clados que se forman (Kapli et al., 2020).

Hipótesis

Los scaffolds obtenidos de la secuenciación de ADN total del saltahojas de césped gris (*Exitianus exitiosus*) contienen los genomas completos de los endosimbiontes *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola*.

Capítulo III: Materiales y métodos

Ensamblaje de genomas de *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola*

Descarga de las lecturas de secuenciación de *Exitianus exitiosus*

Lecturas pareadas (*paired-end*) de secuenciación de alto rendimiento del saltahojas *Exitianus exitiosus* (Número de acceso: SRR10334086) obtenidas por la plataforma Illumina HiSeq 4000 fueron descargadas del Sequence Read Archive (SRA) del National Center for Biotechnology Information (NCBI) por medio del programa SRA Toolkit v2.10.8. La calidad de las lecturas, presencia de adaptadores y porcentaje de citosinas y guaninas (CG) se evaluaron en FastQC v0.11.9 (Andrews, 2010).

Mapeo de las lecturas de secuenciación contra los genomas de *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola*

Del GenBank se descargaron los genomas de referencia de los endosimbiontes *Candidatus Sulcia muelleri* ALF (NA: CP006060.1) y *Candidatus Nasuia deltocephalinicola* ALF (NA: CP006059.1), con ambos se crearon una base de datos para cada genoma y luego se mapearon todas las lecturas de *Exitianus exitiosus* contra las secuencias de las dos bacterias en Bowtie2 v2.2.6 (Langmead & Salzberg, 2012). Los archivos *Sequence Alignment Map* (SAM) generados se transformaron en *Binary Alignment Map* (BAM) y se filtraron los datos para obtener sólo las secuencias alineadas usando SAMtools v1.3.1 (H. Li et al., 2009) con la opción *samtools view* y después se ordenaron con *samtools sort*. A continuación, los alineamientos se separaron en archivos de lecturas pareadas FastQ en BEDTools v2.29.1 (Quinlan & Hall, 2010).

Ensamblaje de las lecturas mapeadas

Las lecturas filtradas de ambas bacterias se ensamblaron con diferentes tamaños de k-mers de 21, 33, 55 y 77 en contigs y luego en scaffolds en SPAdes v3.14.1 (Bankevich et al., 2012). La calidad del ensamblaje se evaluó en QUAST (Gurevich et al., 2013) dentro del servidor web Galaxy (<https://usegalaxy.org>) (Afgan et al., 2018), subiendo los archivos de scaffolds e indicando al programa que los genomas ALF de cada bacteria sean considerados como la referencia.

A continuación, los dos archivos de contigs obtenidos para *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* se subieron al Basic Local Alignment Search Tool (BLAST) (Camacho et al., 2009), donde se escogió la colección de nucleótidos (nr/nt) como base de datos y la opción para secuencias muy similares (megablast) como selección de programa. De esta manera, se identificaron los homólogos de las secuencias obtenidas para ambas bacterias en función de los valores reportados por el programa como query coverage, e-value y porcentaje de identidad.

Finalmente, como solo el ensamblaje de *Candidatus Sulcia muelleri* fue de buena calidad; sus scaffolds se mapearon contra el genoma ALF (NA: CP006060.1) en Geneious v2021.0.3 (Kearse et al., 2012) y se cerraron en un cromosoma bacteriano, al que se le nombró aislado EXT.

Anotación del genoma de *Candidatus Sulcia muelleri*

Comparación de diferentes programas de anotación

Para la anotación del genoma completo de *Candidatus Sulcia muelleri* de *Exitianus exitiosus* se subió su secuencia en formato .fasta a Prokka (Seemann, 2014) en el servidor web Galaxy (<https://usegalaxy.org>), RAST (Aziz et al., 2008) (<https://rast.nmpdr.org/>) y DFAST (Tanizawa et al., 2018) (<https://dfast.nig.ac.jp/>). Se dejó los parámetros por defecto de cada programa y se especificó el dominio de

búsquedas como Bacteria, el género *Candidatus Sulcia* sp. y la especie *muelleri*. Asimismo, las secuencias identificadas como proteínas hipotéticas se anotaron manualmente mediante búsquedas de homólogos en BLAST (Camacho et al., 2009). Posteriormente, se comparó el número de CDS (región codificante), ARNt (ácido ribonucleico de transferencia), ARNr (ácido ribonucleico ribosomal) y ARNtm (ácido ribonucleico mensajero de transferencia) determinados por cada programa.

Visualización del genoma anotado de Candidatus Sulcia muelleri

A continuación, el archivo de anotaciones del genoma bacteriano se subió en el ejecutable DNA Plotter (Carver et al., 2009) del paquete de Artemis, para visualizar el cromosoma de *Candidatus Sulcia muelleri* junto con los genes anotados en la cadena positiva y negativa, ARNt, ARNr, ARNtm y distribución del contenido de CG en todo el genoma.

Determinación de rutas metabólicas

Se repitió la anotación en RAST (Aziz et al., 2008) (<https://rast.nmpdr.org/>), siguiendo los mismos pasos descritos en la sección de “*Comparación de diferentes programas de anotación*”. Sin embargo, en los parámetros se seleccionó la opción “*Construcción de un modelo metabólico*”; en donde el programa determinó el número y porcentaje de genes involucrados en rutas metabólicas como síntesis y/o metabolismo de aminoácidos, carbohidratos, proteínas, compuestos aromáticos, ARN y otras funciones.

Finalmente, la capacidad de síntesis aminoácidos esenciales se verificó manualmente comprobando los genes codificantes de enzimas involucradas en las rutas biosintéticas empleando las bases de datos *Kyoto Encyclopedia of Genes and Genomes* (KEGG) (<https://www.genome.jp/kegg/>) (Kanehisa & Goto, 2000) y MetaCyc (<https://metacyc.org/>) (Caspi et al., 2020).

Análisis filogenético de *Candidatus Sulcia muelleri*

Secuencias incluidas en el análisis

En el NCBI FTP (<ftp://ftp.ncbi.nlm.nih.gov/>) se accedió a la carpeta genomas, genbank, bacteria y *Candidatus Sulcia muelleri* para visualizar todas las versiones ensambladas hasta la actualidad. De allí se descargaron 32 genomas que fueron subidos como un solo cromosoma bacteriano al GenBank; al igual que el genoma de *Candidatus Uzinura diaspidicola*, que fue escogido como organismo externo (*outgroup*). El detalle de cada secuencia incluida en el análisis, en total 34 se puede observar en la Tabla 1.

Tabla 1

Genomas bacterianos considerados en la filogenia de Candidatus Sulcia muelleri

Endosimbionte			Hospedador	
Género	Especie	Número de accesión	Familia	Especie
<i>Sulcia</i>	<i>muelleri</i>	CP016223.1	Cicadellidae	<i>Nephotettix cincticeps</i>
<i>Sulcia</i>	<i>muelleri</i>	No publicado	Cicadellidae	<i>Exitianus exitiosus</i>
<i>Sulcia</i>	<i>muelleri</i>	CP001981.1	Cicadellidae	<i>Draeculacephala minerva</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029016.1	Cicadidae	<i>Mogannia minuta</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029014.1	Cicadidae	<i>Hyalessa maculaticollis</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029013.1	Cicadidae	<i>Graptopsaltria nigrofuscata</i>
<i>Sulcia</i>	<i>muelleri</i>	CP008986.1	Cicadellidae	<i>Graphocephala atropunctata</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029026.1	Cicadidae	<i>Meimuna kuroiwae</i>

Endosimbionte			Hospedador	
Género	Especie	Número de accesión	Familia	Especie
<i>Sulcia</i>	<i>muelleri</i>	CP001605.1	Cicadidae	<i>Diceroprocta semicincta</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029027.1	Cicadidae	<i>Meimuna opalifera</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029020.1	Cicadidae	<i>Auritibicen bihamatus</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029025.1	Cicadidae	<i>Meimuna iwasakii</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029024.1	Cicadidae	<i>Yezoterpnosia vacua</i>
<i>Sulcia</i>	<i>muelleri</i>	CP007234.1	Cicadidae	<i>Tettigades undata</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029017.1	Cicadidae	<i>Muda kuroiuae</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029012.1	Cicadidae	<i>Graptopsaltria bimaculata</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029019.1	Cicadidae	<i>Platypleura kaempferi</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029015.1	Cicadidae	<i>Kosemia yezoensis</i>
<i>Sulcia</i>	<i>muelleri</i>	CP002163.1	Clastopteridae	<i>Clastoptera arizonana</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029010.1	Cicadidae	<i>Cryptotympana facialis</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029023.1	Cicadidae	<i>Yezoterpnosia nigricosta</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029009.1	Cicadidae	<i>Cryptotympana atrata</i>
<i>Sulcia</i>	<i>muelleri</i>	CP010828.1	Cicadidae	<i>Magicicada tredecim</i>
<i>Sulcia</i>	<i>muelleri</i>	CP000770.2	Cicadellidae	<i>Homalodisca vitripennis</i>
<i>Sulcia</i>	<i>muelleri</i>	CP021172.1	Membracidae	<i>Entylia carinata</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029018.1	Cicadidae	<i>Tanna japonensis</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029022.1	Cicadidae	<i>Vagitanus terminalis</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029011.1	Cicadidae	<i>Euterpnosia chibensis</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029028.1	Cicadidae	<i>Meimuna oshimensis</i>
<i>Sulcia</i>	<i>muelleri</i>	CP029021.1	Cicadidae	<i>Auritibicen japonicus</i>

Endosimbionte			Hospedador	
Género	Especie	Número de accesión	Familia	Especie
<i>Sulcia</i>	<i>muelleri</i>	AP013293.1	Aphrophoridae	<i>Philaenus spumarius</i>
<i>Sulcia</i>	<i>muelleri</i>	CP006060.1	Cicadellidae	<i>Macrosteles quadrilineatus</i>
<i>Sulcia</i>	<i>muelleri</i>	CP010105.1	Cicadellidae	<i>Dalbulus maidis</i>
<i>Uzinura</i>	<i>diaspidicola</i>	CP003263.1	Diaspididae	<i>Aspidiotus nerii</i>

Nota: El genoma CP003263.1 corresponde al *outgroup*

Determinación de genes ortólogos

Todas las secuencias descargadas se unieron junto al genoma EXT en un solo archivo multifasta; este se subió al servidor web AmphotaNet (<https://pitgroup.org/amphoranet/>) (Kerepesi et al., 2014) para identificar genes ortólogos de copia única. En donde se especificó que el archivo de entrada corresponde a secuencias de nucleótidos y que las búsquedas se hagan para marcadores bacterianos. El archivo resultante del análisis .txt se visualizó en AmphoraVizu (<https://pitgroup.org/amphoravizu/>) (Kerepesi et al., 2015) y los archivos individuales con las secuencias de cada gen expresados en residuos de aminoácidos se descargaron del servidor.

Alineamiento múltiple

Las secuencias de cada gen se alinearon en T-Coffee (<http://tcoffee.crg.cat/>) (Di Tommaso et al., 2011), se escogió la opción para alineamientos de proteínas M-Coffee y después se eliminaron los fragmentos ambiguos de cada alineamiento en el servidor web Gblocks (http://phylogeny.lirmm.fr/phylo.cgi/one_task.cgi?task_type=gblocks) (Talavera & Castresana, 2007), marcando todas las opciones para una selección menos estricta. Finalmente, los alineamientos curados se concatenaron en Geneious v2021.0.3

(Kearse et al., 2012), seleccionando la opción *Tools > Concatenate Sequences or Alignments > Concatenating by name*.

Determinación del mejor modelo de sustitución

El alineamiento múltiple concatenado se subió al servidor web de CIPRES (<https://www.phylo.org/portal2/home.action>) (Miller et al., 2010), para elegir el mejor modelo de sustitución de los datos en ModelTest-NG v0.1.15 (Darriba et al., 2020), especificando al programa que el archivo de entrada contiene secuencias de aminoácidos.

Construcción del árbol filogenético por máxima verosimilitud

El alineamiento múltiple concatenado se subió al servidor web de CIPRES, para generar el árbol filogenético por máxima verosimilitud en RAxML-HPC BlackBox v8.2.12 (Stamatakis, 2014) empleado el enfoque *SuperMatriz*; el programa se corrió con los siguientes parámetros:

- Número de bootstraps: 1000
- *Outgroup*: CP003263.1
- Número de caracteres: 4715 aminoácidos
- Modelo de sustitución: Reportado por ModelTest-NG v0.1.15

Construcción del árbol filogenético por inferencia bayesiana

El alineamiento múltiple concatenado se cargó en BEAUti para generar un archivo .xml; el programa se corrió con los siguientes parámetros:

- Modelo de sustitución: Reportado por ModelTest-NG v0.1.15
- Reloj molecular: Reloj estricto
- Prior del árbol: Coalescente de tamaño constante
- Número de generaciones: 5000000

- Frecuencia de muestreo: 1000

El archivo .xml se subió a CIPRES para correr el algoritmo MCMC en el programa BEAST (*Bayesian Evolutionary Analysis Sampling Trees*) v1.10.4 (Suchard et al., 2018); del cual se obtuvieron dos archivos con extensión .log y .trees. El .log se cargó en Tracer v1.7.1 para evaluar la calidad del análisis; esto se verificó comprobando que los parámetros calculados en BEAST tengan un valor de ESS > 200. Finalmente, para generar el árbol consenso, se subió el archivo .trees a TreeAnnotator v1.10.4 con un *Burn-in* del 10%.

Visualización de los árboles filogenéticos

Los árboles obtenidos por máxima verosimilitud e inferencia bayesiana se subieron al programa FigTree v1.4.4 (Rambaut, 2014) para visualizar su topología y el soporte estadístico de los clados obtenidos por bootstraps y probabilidades posteriores. Asimismo, los árboles se cargaron en MEGAX (Kumar et al., 2018) seleccionando la opción *User Tree > Display Newick Trees* para eliminar los valores de soporte < 70 (bootstraps) y < 0.7 (probabilidades posteriores). Finalmente, el árbol obtenido por máxima verosimilitud se editó en Inkscape para mostrar la topología y el soporte estadístico de los clados como *bootstraps* debajo de las ramas y las probabilidades posteriores en la parte superior.

Capítulo IV: Resultados

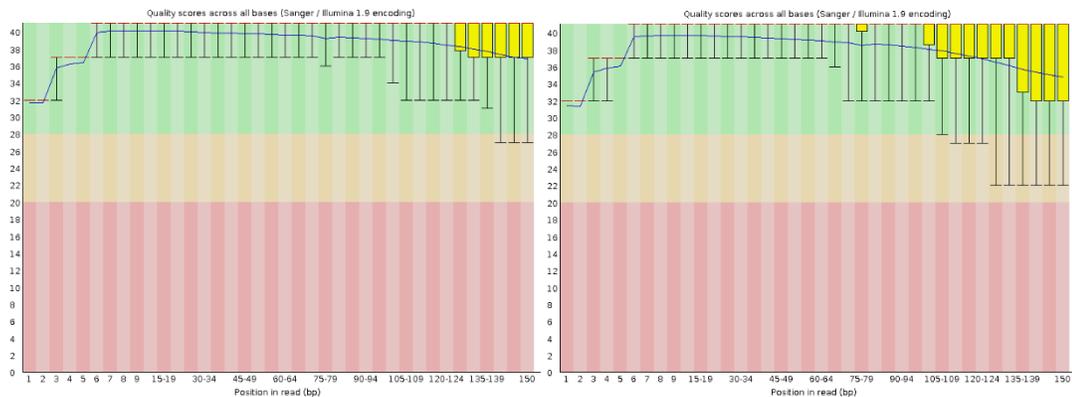
Ensamblaje de genomas de *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola*

Calidad de las lecturas de secuenciación de *Exitianus exitiosus*

Del SRA se descargaron 166,861,468 lecturas *paired-end* con una longitud de 150 bp cada una, obtenidas de la secuenciación de alto rendimiento de *Exitianus exitiosus* por la plataforma Illumina. En FastQC se obtuvo el reporte de la calidad de los datos; en las gráficas de calidad de las secuencias por base (Figura 5), el eje de las abscisas representa las posiciones en las lecturas (bp) y el eje de las ordenadas la puntuación de calidad. Puntuaciones entre 0-20 son consideradas de baja calidad y se representan en color rojo; 20-28 calidad media razonable de color naranja y > 28 buena calidad de color verde. Por otra parte, las cajas amarillas simbolizan el rango intercuartil (25-75%), mientras que los bigotes superiores e inferiores el 10 y 90% y la línea azul representa la media de los datos. Como se puede observar en la Figura 5, la mayoría de las bases se encuentran en la zona verde; es decir, son lecturas de alta calidad. No obstante, las bases de los extremos finales presentaron una calidad media razonable, por encontrarse en la zona naranja del gráfico; tanto para las lecturas de la cadena sentido como para la antisentido.

Figura 5

*Calidad de las secuencias por base de las lecturas de secuenciación de alto rendimiento del saltahoja del césped gris (*Exitianus exitiosus*)*

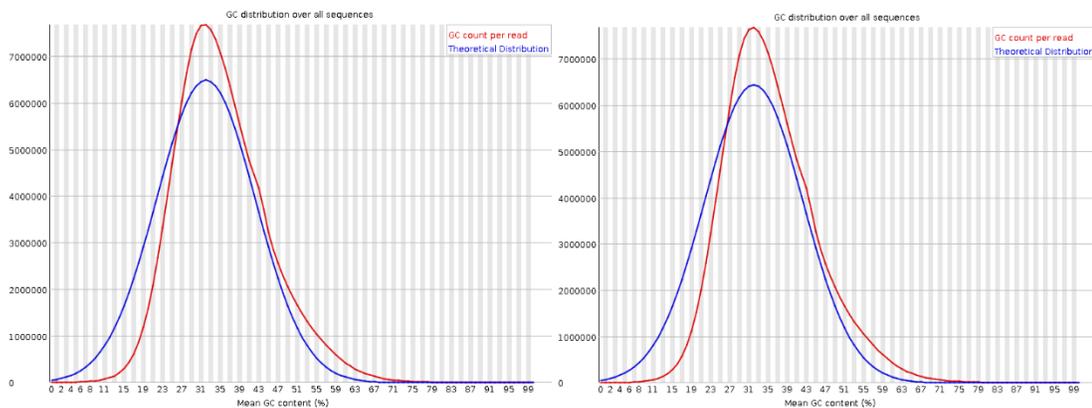


Nota. A la izquierda se muestran los resultados para la cadena sentido y a la derecha para la cadena antisentido

El contenido de GC se evaluó comparando la distribución de CG por lectura (curva roja) con la distribución teórica (curva azul), que se muestra en la Figura 6. En donde el eje de las abscisas representa el promedio de CG % y el eje de las ordenadas el número de lecturas. Para ambas cadenas sentido y antisentido, en promedio sus lecturas presentaron un 31-35% de CG.

Figura 6

Contenido de CG en las lecturas de secuenciación de alto rendimiento del saltahojas del césped gris (Exitianus exitiosus)

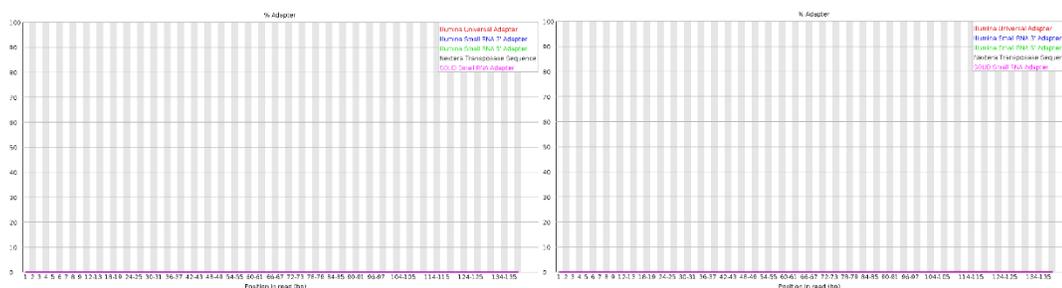


Nota. A la izquierda se muestran los resultados para la cadena sentido y a la derecha para la cadena antisentido

Finalmente, en la gráfica de contenido de adaptadores (Figura 7), el eje de las abscisas muestra la posición de las bases en las lecturas (bp) y el eje de las ordenadas, el porcentaje de adaptadores. Como se puede observar, tanto en las lecturas de la cadena sentido como antisentido no hubo presencia de adaptadores, ya que ninguna curva se levantó. Por lo tanto, al unir estos resultados con los obtenidos anteriormente, se puede inferir que las secuencias subidas al SRA fueron de buena calidad y no necesitaron un preprocesamiento de limpieza.

Figura 7

Contenido de adaptadores en las lecturas de secuenciación de alto rendimiento del saltahojas del césped gris (Exitianus exitiosus)



Nota. A la izquierda se muestran los resultados para la cadena sentido y a la derecha para la cadena antisentido

Mapeo de las lecturas de secuenciación contra los genomas de Candidatus Sulcia muelleri y Candidatus Nasuia deltocephalinicola

Las lecturas de *Exitianus exitiosus* se mapearon a los genomas ALF de *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* en Bowtie2 v2.2.6. Como se puede ver en la Figura 8, de las 166,861,468 lecturas en total (100%), sólo

1318644 (0,79%) se alinearon exactamente una vez de manera concordante al genoma de *Candidatus Sulcia muelleri* y 7345 (0,004%) a *Candidatus Nasuia deltocephalinicola*. Por lo tanto, solo las lecturas mapeadas se filtraron y se separaron en archivos FastQ para proceder a su ensamblaje.

Figura 8

Número de lecturas alineadas con los genomas de Candidatus Sulcia muelleri y Candidatus Nasuia deltocephalinicola



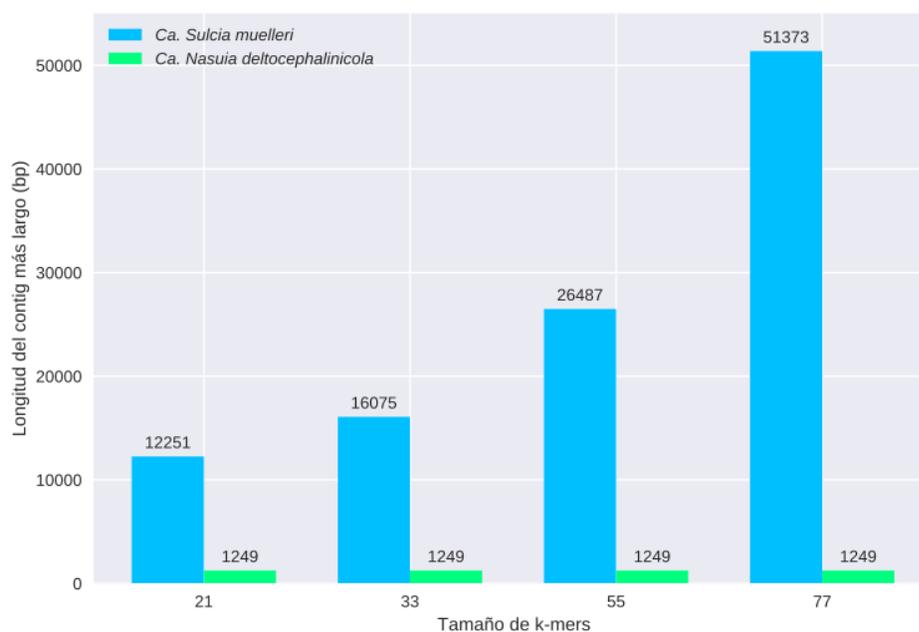
Ensamblaje de las lecturas mapeadas

Las lecturas mapeadas contra los genomas de *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* se procesaron en SPAdes v3.14.1 para ensamblarlas usando diferentes tamaños de k-mer de 21, 33, 55 y 77, determinados previamente por el programa. Como se observa en la Figura 9, el ensamblaje de *Candidatus Sulcia muelleri* generó contigs de longitud creciente de 12251, 16075, 26487 y 51373 bp para k-mers de 21, 33, 55 y 77, respectivamente. Sin embargo, para

Candidatus Nasuia deltocephalinicola, los contigs más largos fueron de la misma longitud de 1249 bp para cada tamaño de k-mer.

Figura 9

Longitud del contigs más largo obtenido con diferentes tamaños de k-mers para el ensamblaje de Candidatus Sulcia muelleri y Candidatus Nasuia deltocephalinicola



Para ambas bacterias, sólo los contigs generados con un tamaño de k-mer de 77 se ensamblaron en scaffolds. Como se puede ver en la Tabla 2, para el caso de *Candidatus Sulcia muelleri*, los 9 contigs se unieron en 5 scaffolds y para *Candidatus Nasuia deltocephalinicola*, los 5 contigs en 4 scaffolds en SPAdes v3.14.1.

Tabla 2

Número de contigs y scaffolds obtenidos con un tamaño de k-mer de 77 para Candidatus Sulcia muelleri y Candidatus Nasuia deltocephalinicola

Endosimbionte	Número de contigs	Número de scaffolds
<i>Ca. Sulcia muelleri</i>	9	5
<i>Ca. Nasuia deltocephalinicola</i>	5	4

Como se puede ver en la Tabla 3, el análisis de calidad del ensamblaje de *Candidatus Sulcia muelleri* presentó una alta calidad, ya que generó contigs y scaffolds de longitudes de hasta 51373 y 110410 bp, respectivamente. Asimismo, en total se obtuvieron 190004 bp, que al compararlas con la longitud del genoma ALF de referencia, se aproxima a los 190,733 bp de la bacteria. De igual manera, sus N50 fueron elevados y no se reportaron secuencias mal ensambladas o alineadas.

Por otra parte, en la Tabla 4 se muestra el reporte del ensamblaje de *Candidatus Nasuia deltocephalinicola*, en donde sus contigs y scaffolds más largos fueron de 1249 bp y además en total sólo se obtuvieron 3386 y 3396 bp, respectivamente. Al comparar estos resultados con la longitud total del genoma ALF de referencia, se puede verificar que el ensamblaje no cubrió el genoma completo de la bacteria, además sus N50 fueron bajos e incluso 4 contigs y 3 scaffolds no se alinearon al genoma de *Candidatus Nasuia deltocephalinicola* ALF; demostrando que el ensamblaje para esta bacteria fue de mala calidad y posiblemente no se encuentre dentro del insecto.

Tabla 3*Análisis de calidad del ensamblaje de Candidatus Sulcia muelleri*

Parámetros de calidad	Contigs	Scaffolds
Número de secuencias	9	5
Tamaño de la secuencia más larga (bp)	51373	110410
Longitud total (bp)	189964	190004
Longitud del genoma de referencia (bp)	190733	190733
CG (%)	23.8	23.8
N50	45597	110410
L50	2	1
Número de secuencias mal ensambladas	0	0
Longitud de secuencias mal ensambladas	0	0
Número de secuencias no alineadas a la referencia	0	0
Longitud de secuencias no alineadas a la referencia	0	0

Nota. N50 se define como la longitud del contig más corto que cubre el 50% del genoma predicho. L50 representa el número de contigs más cortos cuya suma comprende la mitad del genoma predicho.

Tabla 4

Análisis de calidad del ensamblaje de Candidatus Nasuia deltocephalinicola

Parámetros de calidad	Contigs	Scaffolds
Número de secuencias	5	4
Tamaño de la secuencia más larga (bp)	1249	1249
Longitud total (bp)	3386	3396
Longitud del genoma de referencia (bp)	112091	112091
CG (%)	39.10	39.10
N50	770	1104
L50	2	2
Número de secuencias mal ensambladas	0	0
Longitud de secuencias mal ensambladas	0	0
Número de secuencias no alineadas a la referencia	4	3
Longitud de secuencias no alineadas a la referencia	2616	2292

Nota. N50 se define como la longitud del contig más corto que cubre el 50% del genoma predicho. L50 representa el número de contigs más cortos cuya suma comprende la mitad del genoma predicho.

Para verificar que las lecturas ensambladas correspondan a los endosimbiontes *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola*, todos sus contigs se analizaron en BLAST. Para *Candidatus Sulcia muelleri* (Tabla 5) todos sus contigs se alinearon a los genomas de la misma bacteria a los aislados ALF, NC y PUNC con un query coverage > 99%; e-value de 0 y porcentaje de identidad > 97.36%. Por lo tanto, se confirma que los contigs obtenidos en la presente investigación corresponde a la bacteria *Candidatus Sulcia muelleri*.

Por otra parte, para *Candidatus Nasuia deltocephalinicola* (Tabla 6) sus contigs se alinearon a las secuencias publicadas de la misma bacteria, pero también a *Candidatus Rickettsiella viridis*. Además, BLAST reportó un query coverage > 96%, e-values de 0, 2e-119 y 7e-144 y porcentaje de identidad > 86.77%. Debido a que los valores para cada parámetro fueron bajos, sólo se puede confirmar que el primer contig obtenido en la presente investigación corresponde a la región de 16S del ARNr de *Candidatus Nasuia deltocephalinicola*. Por lo tanto, al unir estos resultados con el reporte de calidad del ensamblaje, las secuencias de la bacteria se descartaron de la investigación.

Tabla 5

Identificación de homólogos a los contigs obtenidos del ensamblaje de Candidatus Sulcia muelleri

Contig	Homólogo identificado	Query coverage	E value	% de Identidad	Número de accesoión
1	<i>Ca. Sulcia muelleri</i> ALF	99%	0.0	97.69%	CP006060.1
2	<i>Ca. Sulcia muelleri</i> NC	99%	0.0	98.92%	CP016223.1
3	<i>Ca. Sulcia muelleri</i> NC	99%	0.0	98.67%	CP016223.1

Contig	Homólogo identificado	Query coverage	E value	% de Identidad	Número de accesión
4	<i>Ca. Sulcia muelleri</i> PUNC	99%	0.0	98.59%	CP013212.1
5	<i>Ca. Sulcia muelleri</i> NC	100%	0.0	99%	CP016223.1
6	<i>Ca. Sulcia muelleri</i> NC	99%	0.0	99.28%	CP016223.1
7	<i>Ca. Sulcia muelleri</i> NC	99%	0.0	97.44%	CP016223.1
8	<i>Ca. Sulcia muelleri</i> ALF	99%	0.0	96.48%	CP006060.1
9	<i>Ca. Sulcia muelleri</i> NC	99%	0.0	97.36%	CP016223.1

Tabla 6

Identificación de homólogos a los contigs obtenidos del ensamblaje de Candidatus Nasuia deltocephalinicola

Contig	Homólogo identificado	Query Coverage	E value	% de Identidad	Número de accesión
<i>Ca. Nasuia</i>					
1	<i>deltocephalinicola</i> 16S rRNA	99%	0.0	94.75%	AB702994.1
2	<i>Ca. Nasuia deltocephalinicola</i> ALF	96%	0.0	91.81%	CP006059.1
3	<i>Ca. Nasuia deltocephalinicola</i> ALF	100%	0.0	86.77%	CP006059.1
4	<i>Ca. Nasuia deltocephalinicola</i> NAS-MSEV	97%	2e-119	91.82%	CP060019.1

Contig	Homólogo identificado	Query Coverage	E value	% de Identidad	Número de accesión
5	<i>Ca. Rickettsiella viridis</i> Ap-RA04	100%	7e-144	98.97%	AP018005.1

Como sólo el ensamblaje de *Candidatus Sulcia muelleri* fue de buena calidad, sus scaffolds se mapearon contra el genoma ALF de la bacteria en Geneious v2021.0.3 y los scaffolds se cerraron en un cromosoma de 191,362 bp, como se puede observar en la Figura 10. Sin embargo, dos regiones de 1,264 y 94 bp no fueron cubiertas por los scaffolds, por lo que dentro de la secuencia pueden ser identificadas con signos de interrogación (Figura 11 y 12). Además, como el programa permite ver las anotaciones dentro una secuencia, se pudo identificar que la región faltante de 1,264 bp corresponde al gen *lpdA*, mientras que la segunda región no presentó anotaciones en la referencia.

Figura 10

Representación del cromosoma de Candidatus Sulcia muelleri, endosimbionte del saltahoja del césped gris (Exitianus exitiosus)

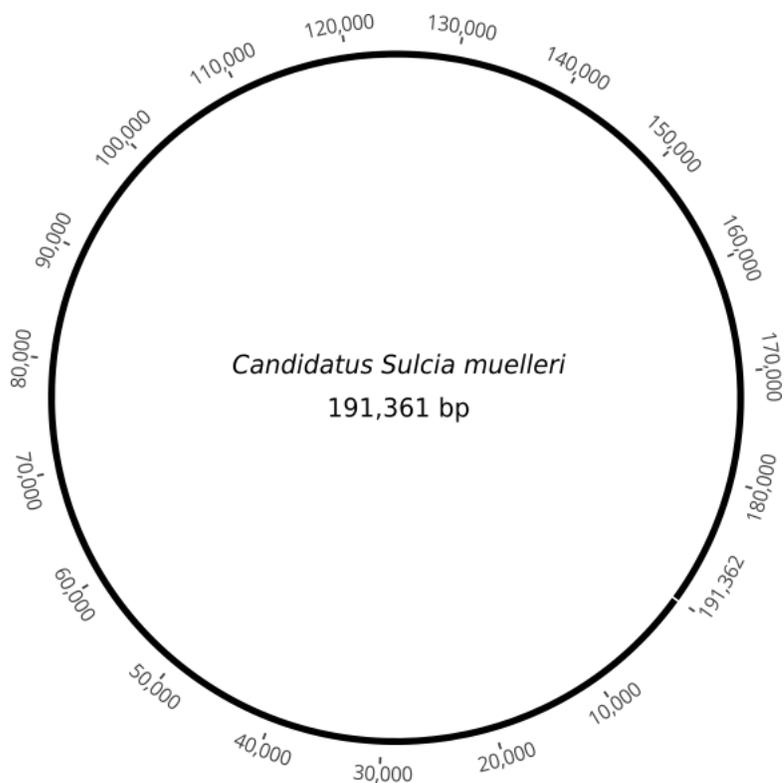


Figura 11

Interfaz de Geneious v2021.0.3 de la región de 1264 bp no cubierta por los scaffolds de Candidatus Sulcia muelleri

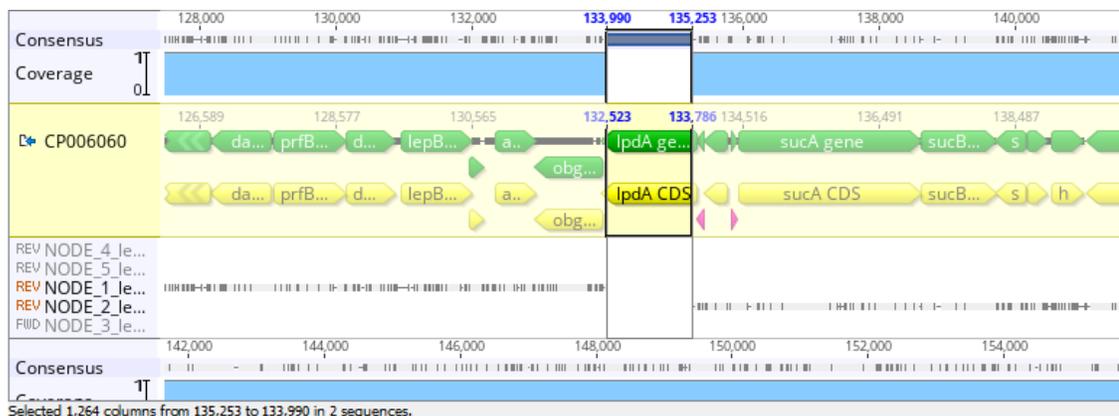
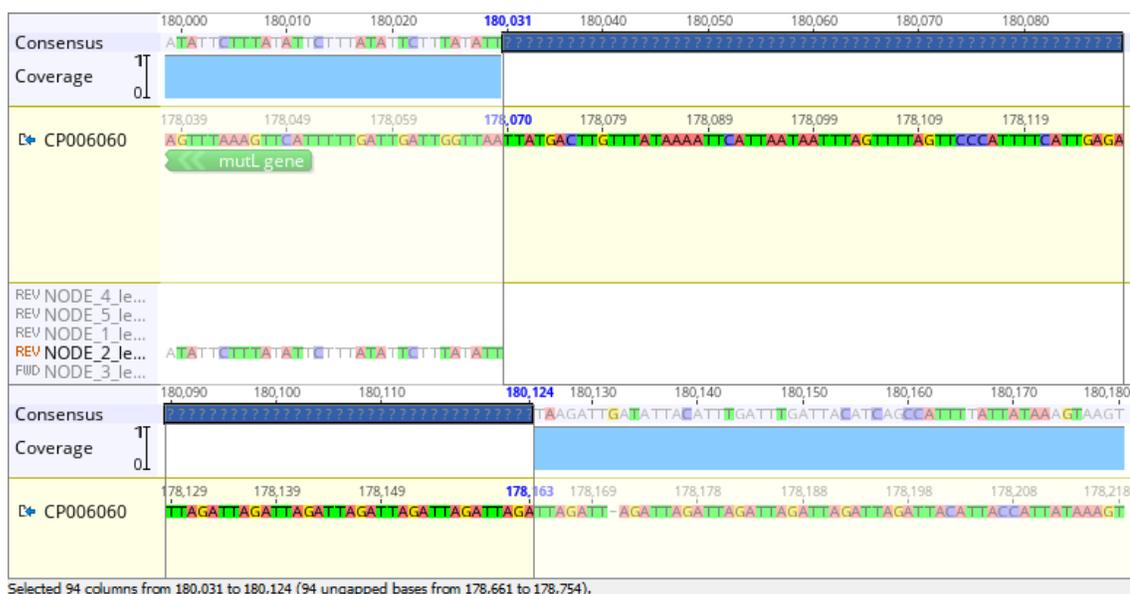


Figura 12

Interfaz de Geneious v2021.0.3 de la región de 94 bp no cubierta por los scaffolds de Candidatus Sulcia muelleri



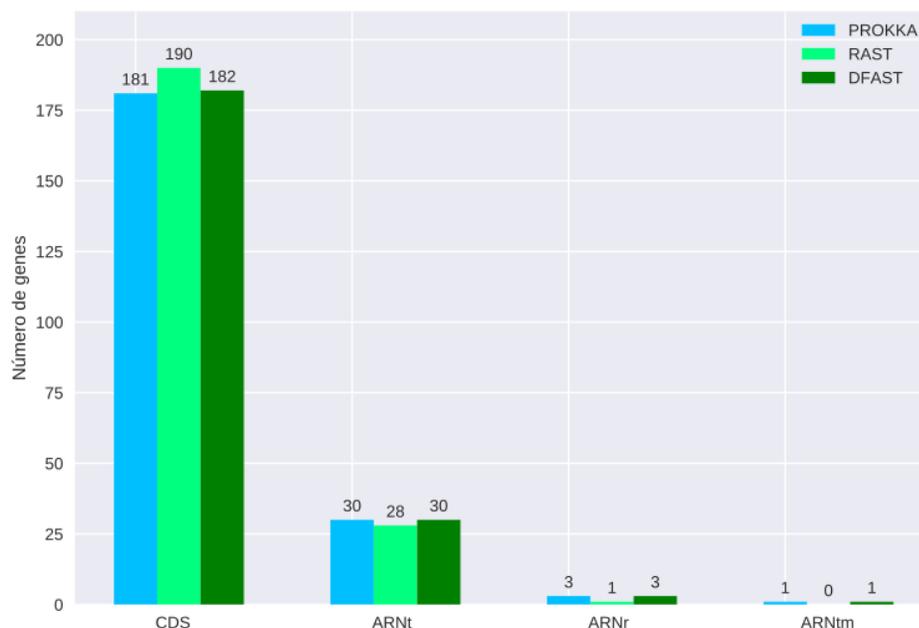
Anotación del genoma de *Candidatus Sulcia muelleri*

Comparación de diferentes programas de anotación

Para la anotación del genoma de *Candidatus Sulcia muelleri* se empleó los programas Prokka, RAST y DFAST; en donde para los CDS se identificaron 181, 190 y 182 genes; para los ARNt 30, 28 y 30 genes; para los ARNr 3, 1 y 3 genes y para los ARNtm 1, 0 y 1 gen por cada programa. Como se puede ver en la Figura 13, RAST identificó más regiones codificantes; no obstante, para los ARNt, ARNr y ARNtm los programas Prokka y DFAST anotaron más genes que RAST en el genoma de la bacteria.

Figura 13

Número de genes identificados en el genoma de *Candidatus Sulcia muelleri* por diferentes programas de anotación

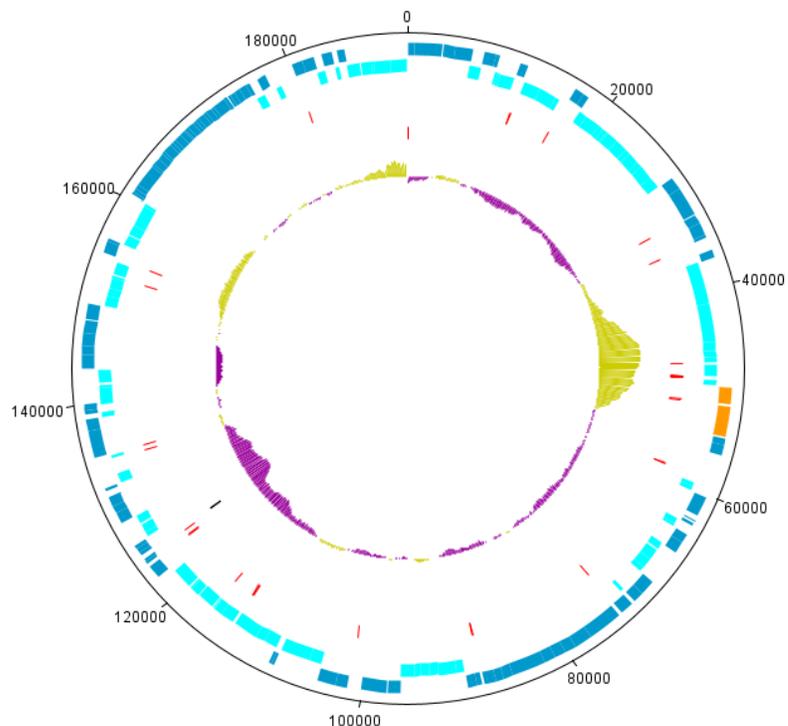


Visualización del genoma anotado de *Candidatus Sulcia muelleri*

El archivo con las anotaciones de *Candidatus Sulcia muelleri* se subió al DNA Plotter para obtener una representación esquemática del cromosoma del endosimbionte (Figura 14); en donde los CDS que se encuentran en la cadena positiva se pueden identificar en color azul y en celeste los que están en la cadena negativa. Por otra parte, los ARNr (operón 16S-23S-5S), ARNt y ARNtm se pueden distinguir con los colores naranja, rojo y negro, respectivamente. Finalmente, en todo el genoma las regiones que se encuentran bajo el promedio del contenido de CG (23.8%) se pueden identificar con color morado y las que se encuentran sobre el promedio en color amarillo.

Figura 14

Representación esquemática del genoma anotado de Candidatus Sulcia muelleri

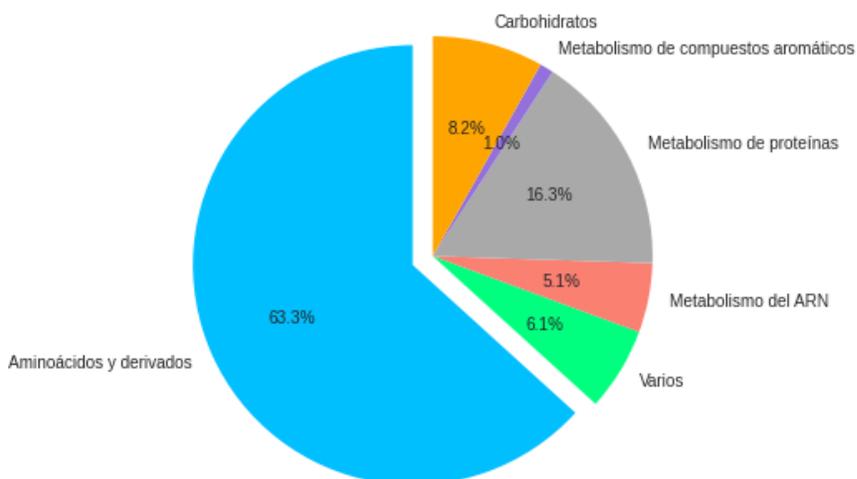


Determinación de rutas metabólicas

De todos los genes anotados, RAST identificó 98 genes involucrados en rutas metabólicas (Figura 15); de los cuales 63.3% CDS se encargan de la síntesis de aminoácidos y sus derivados, 8.2% del metabolismo central de los carbohidratos, 1.0% del metabolismo de compuestos aromáticos, 16.3% del metabolismo de proteínas, 5.1% del metabolismo del ARN y 6.1% de otras funciones.

Figura 15

Distribución porcentual de las rutas metabólicas de Candidatus Sulcia muelleri

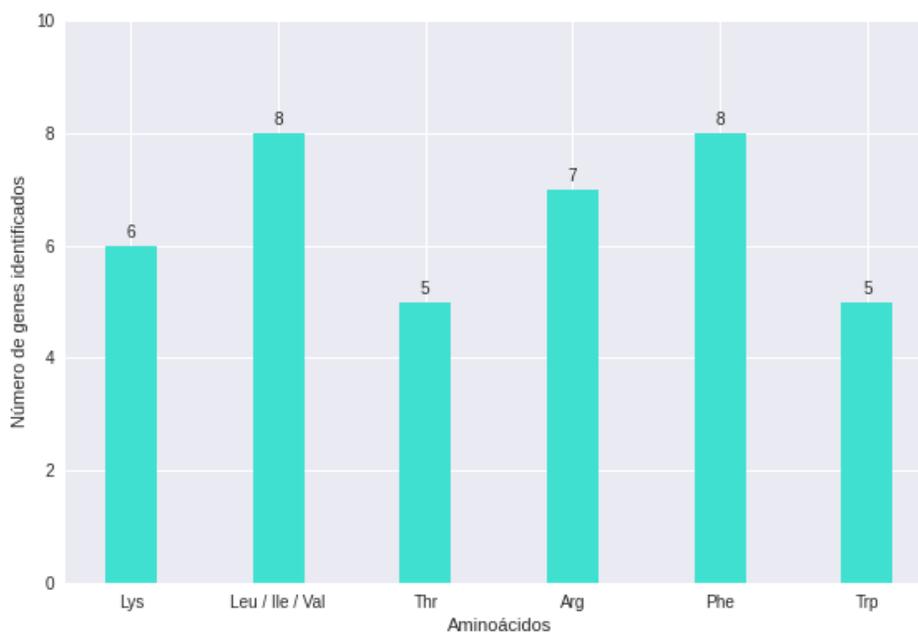


Mediante la verificación manual de genes codificantes de enzimas involucradas en la síntesis de aminoácidos, se determinó que el genoma de *Candidatus Sulcia muelleri* posee 38 genes que se encargan únicamente de la síntesis de ocho aminoácidos esenciales. Como se puede ver en la Figura 16, seis genes corresponden a la lisina (Lys), ocho a aminoácidos de cadena ramificada [leucina (Leu), isoleucina (Ile) y valina (Val)], cinco a treonina (Thr), siete a arginina (Arg), ocho a fenilalanina (Phe) y cinco a triptófano (Trp).

El detalle de los genes y enzimas destinadas a la biosíntesis de cada aminoácido, se encuentran descritos en la matriz de presencia/ausencia en el Anexo 1. En donde se puede verificar que, en la ruta biosintética de la lisina (Lys), no se encontró el gen codificante para una diaminopimelato aminotransferasa; asimismo para la arginina (Arg) faltó una N-acetiltransferasa de aminoácidos.

Figura 16

Número de genes codificantes de enzimas involucrados en la síntesis de aminoácidos esenciales en el genoma de Candidatus Sulcia muelleri



Análisis filogenético de *Candidatus Sulcia muelleri*

Determinación de genes ortólogos

El archivo multifasta con los 34 genomas completos de *Candidatus Sulcia muelleri*, incluido el outgroup *Candidatus Uzinura diaspidicola* se subió al servidor AmphoraNet; este determinó que para la filogenia se pueden trabajar con un total de 29 genes. De los cuales se obtuvieron archivos separados por cada gen con secuencias expresadas en residuos de aminoácidos. Sin embargo, cada archivo se inspeccionó manualmente y se descartaron del análisis los que no incluían al genoma de *Candidatus Sulcia muelleri* de *Exitianus exitiosus* (ensamblado en la presente investigación) ni el *outgroup*. La matriz de evaluación de presencia/ausencia por cada gen se puede observar en el Anexo 2. Por lo tanto, se trabajó solo con 24 genes

ortólogos de copia única (*infC*, *rplA*, *rplB*, *rplD*, *rplE*, *rplF*, *rplK*, *rplL*, *rplM*, *rplN*, *rplP*, *rplS*, *rplT*, *rpmA*, *rpoB*, *rpsB*, *rpsC*, *rpsE*, *rpsI*, *rpsJ*, *rpsK*, *rpsM*, *rpsS* y *smpB*).

Alineamiento múltiple

Se obtuvieron 24 alineamientos múltiples usando M-Coffee dentro del servidor T-Coffee; en donde también se pudo visualizar la calidad del alineamiento por residuo de cada secuencia. Por ejemplo, como se puede ver en la Figura 17 para el gen *infC*; al igual que para los demás, el programa añadió gaps (-) que representan inserciones o deleciones dentro de las secuencias. Asimismo, M-Coffee marcó en color azul y verde las secciones pobremente alineadas; en amarillo los alineamientos promedio y en rosado las regiones correctamente alineadas.

A continuación, Gblocks 0.91b eliminó los fragmentos ambiguos para los 24 genes y conservó únicamente las regiones correctamente alineadas (en azul); como se observa en la Figura 18. Luego los alineamientos múltiples individuales de los 24 genes se concatenaron en un solo archivo de 4715 residuos de aminoácidos usando Geneious v2021.0.3.

Figura 18

Resultado de Gblocks 0.91b para el alineamiento múltiple del gen *infC*

CP029017.1_1356	-----	LMKR	-----	KPAKR	--I--	SYKINKKISATELRV		
CP029012.1_395	-----	K	LMKR	-----	KPAKR	--R--	SYKINKKISATELRV	
CP029019.1_390	----NKFGVIKI	LK	LMKR	-----	KPAK	-----	SYKINKKISATELRV	
CP029015.1_381	-----	VW	VMKR	KPAKRI	-----	-----	SYKINKNISATELRV	
CP002163.1_412	-----NNKIIA	LK	-----	KKYN--N	--SQRL	--KKNELHRV	NKRITATKLRV	
CP029010.1_1769	-----	II	LK	-----	LMKR	-----	KPTKR--R--	SYKINKKISATELRV
CP029023.1_399	-----	K	LMKR	-----	KPAKR	--R--	-----	SYKINKKISATELRV
CP029009.1_1772	-----	II	LK	-----	LMKR	-----	KPTKR--R--	SYKINKKISATELRV
CP016223.1_273	-----KSKNIA	LK	-----	KTKN--N	--SQRY	--KKEELHRV	NKRITTTKLRV	
CP010828.1_368	-----NNIVII	LK	-----	LMKR	-----	KPAKR	--I--	SYKINKKISATELRV
CP000770.2_325	-----KYKNIA	LK	-----	KTKN--N	--SQRY	--KKEELHRV	NKRITATKLRV	
CP021172.1_299	-----KYKNIA	LK	-----	KTKN--N	--SQRY	--KKEELHRV	NKRITATKLRV	
CP029018.1_416	-----	K	LMKR	-----	KPAKR	--R--	-----	SYKINKKISATELRV
CP029022.1_352	-----RSKISLVI	-----	MMKR	KPA	-----	-----	-----	YKINKNISATELRV
CP029011.1_406	-----	K	LMKR	-----	KPAKR	--R--	-----	SYKINKKISATELRV
CP029021.1_1857	-----	II	LK	-----	LMKI	-----	KPTKR--R--	SYKINKKISATELRV
AP013293.1_404	-----NYKIIA	LK	-----	KKYN--N	--SQRF	--KKNELHRV	NKRITATKLRV	
CP006060.1_280	-----KDKNIA	LK	-----	KKTN--N	--SQRY	--KKEELHRV	NKRITATKLRV	
CP010105.1_269	-----KSKNIA	LK	-----	KKQN--N	--SQRY	--KKEELHRV	NKRITATKLRV	
CP029028.1_389	-----	K	LMKR	-----	KPAKR	--R--	-----	SYKINKKISATELRV

Determinación del mejor modelo de sustitución

El programa ModelTest-NG calculó la puntuación para los criterios de información bayesiano (BIC), de información akaike (AIC) e información akaike corregido (AICc); junto con su máxima verosimilitud (lnL) (Tabla 7). De esta manera, considerando los menores valores de puntuación y lnL se determinó que el mejor modelo de sustitución que se ajusta a los datos fue CPREV+F.

Tabla 7

Parámetros calculados por ModelTest-NG para determinar el mejor modelo de sustitución

Criterios	Modelo de sustitución	InL	Puntuación
BIC	CPREV+F	-39754.7645	80220.0434
AIC	CPREV+F	-39754.7645	79677.5291
AICc	CPREV+F	-39754.7645	79680.5291

Construcción del árbol filogenético por máxima verosimilitud e inferencia bayesiana

La filogenia de *Candidatus Sulcia muelleri* se construyó con el alineamiento múltiple concatenado de 24 genes ortólogos de copia única con 4715 residuos de aminoácidos en total y usando el modelo de sustitución CPREV+F. En primer lugar, RAxML-HPC BlackBox v8.2.12 generó un árbol aditivo o también conocido como filograma (Figura 19); en donde el soporte estadístico de los clados está especificado en la parte superior por probabilidades posteriores, calculadas por el paquete de BEAST v1.10.4 para inferencia bayesiana y en la parte inferior por *bootstraps* calculados por RAxML-HPC para máxima verosimilitud. Sin embargo, no se pudo mostrar el soporte de todos los clados, ya que algunos no presentaron cambios significativos entre ellos y en la gráfica sólo se señalaron los valores para los grupos más representativos. Asimismo, para identificar cada OTU se colocó en el árbol consenso el número de accesión, nombre del endosimbionte y nombre del hospedador para cada taxón.

Como se puede ver en la Figura 19, en la filogenia de *Candidatus Sulcia muelleri* se distinguen dos clados principales y la raíz en color gris que corresponde al organismo seleccionado como outgroup *Candidatus Uzinura diaspidicola*, simbiote del insecto *Aspidiotus nerii* de la familia Diaspididae.

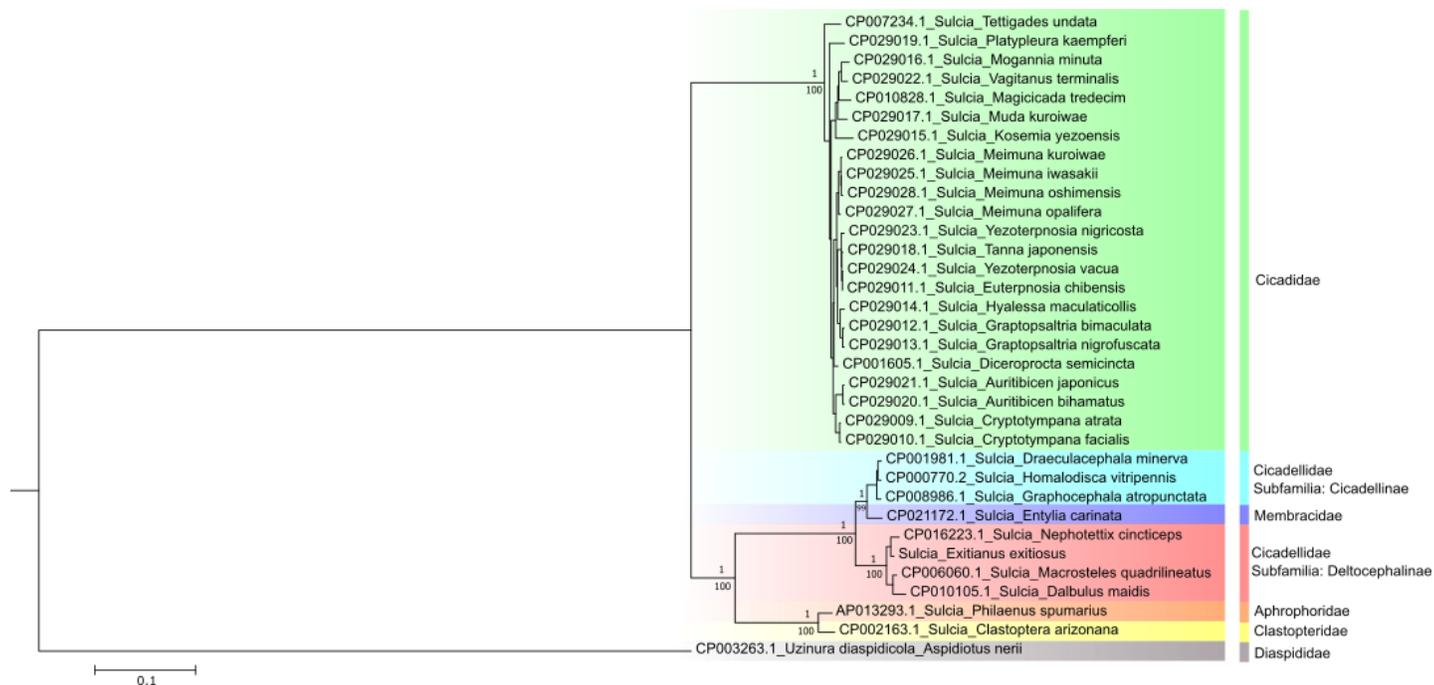
El primer clado en color verde presentó un soporte robusto de 1 y 100% de probabilidad posterior y *bootstrap*, respectivamente. En este se agruparon en total 23 OTUs, que se caracterizan por ser secuencias de endosimbiontes (*Candidatus Sulcia muelleri*) de insectos que pertenecen a la familia Cicadidae.

El segundo clado (probabilidad posterior =1 / bootstrap =100%) a su vez se subdividió en más clados; en el primer grupo de color celeste se agruparon tres simbiotes de la familia Cicadellidae, subfamilia Cicadellinae y bajo este en color azul se identifica al único miembro de la familia Membracidae; que en conjunto forman un clado con soporte de 1 y 99% de probabilidad posterior y *bootstrap*, respectivamente.

En color rojo se observa un clado (probabilidad posterior =1 / bootstrap =100%) con cuatro OTUs de la familia Cicadellidae, subfamilia Deltocephalinae; en donde se distingue que un taxón no tiene número de accesión. Este corresponde a las secuencias del genoma de *Candidatus Sulcia muelleri* de *Exitianus exitiosus* obtenidas en la presente investigación; donde se puede confirmar que la bacteria se agrupó con los demás miembros Deltocephalinae.

Finalmente, el último clado (probabilidad posterior =1 / bootstrap =100%) se formó con dos taxones que pertenecen a familias diferentes como Aphrophoridae y Clastopteridae y se distinguen en color naranja y amarillo, respectivamente.

Figura 19

Topología del árbol filogenético de *Candidatus*

Nota. La barra de 0.1 representa las sustituciones por sitio. Sobre las ramas se encuentra la probabilidad posterior y bajo las misma su valor de bootstrap

Capítulo V: Discusión

Los genomas completos de endosimbiontes bacterianos de insectos se han obtenido en previas investigaciones por medio de la secuenciación del ADN total de los insectos (Chang et al., 2015) o de sus bacteriomas (Shigenobu et al., 2000) mediante plataformas como Illumina, PacBio y MinION (Bennett et al., 2016; Waneka et al., 2020). Las lecturas descargas en la presente investigación del SRA fueron obtenidas por Illumina HiSeq 4000 del ADN total de *Exitianus exitiosus*. A pesar de que Illumina es una de las plataformas más utilizadas para la generación de lecturas cortas; usualmente su calidad tiende a ser muy baja al final de las lecturas (Tan et al., 2019). La calidad puede disminuir al aumentar la longitud de las lecturas, por bases identificadas como N y presencia de adaptadores, que en conjunto pueden originar genomas mal ensamblados y por ello es necesario remover las secuencias de baja puntuación de calidad y adaptadores para evitar obtener conclusiones biológicas erróneas (Pfeifer, 2017).

Del análisis de calidad se determinó que en conjunto las lecturas de *Exitianus exitiosus* fueron de buena calidad; debido a que no presentaron bases de baja calidad, ni adaptadores y su distribución en el contenido de CG se aproximó a la distribución normal teórica. En el estudio de Chen et al. (2017) las lecturas originales de secuenciación fueron sometidas a un filtrado de calidad y de adaptadores usando la herramienta Trim Galore para luego continuar con el ensamblaje. No obstante, como en este estudio las lecturas fueron descargas del SRA, la alta calidad de las secuencias sugiere que fueron subidas al repositorio procesadas previamente.

Con respecto a *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola*; sus genomas se encuentran disponibles en las bases de datos (Bennett & Moran, 2013). De manera que se puede emplear el enfoque basado en

mapeos; en donde las lecturas se alinean contra un genoma de referencia y luego se filtran (Pfeifer, 2017). Del mapeo de *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* se encontró que, del total de lecturas sólo el 0.79% y 0.004% corresponden a cada bacteria, respectivamente. Guyomar et al. (2019) explican que estos bajos porcentajes se deben a que las lecturas de secuenciación de los insectos como *Exitianus exitiosus* son datos metagenómicos desequilibrados, porque en su mayoría pertenecen al hospedador, pero una parte también a sus simbioses. Los cuales se caracterizan por poseer genomas diminutos que pueden tener una gran profundidad de lectura en la secuenciación, lo que permite extraer información relevante sobre estos microorganismos a partir de las muestras de los insectos.

En cuanto al ensamblaje de *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola* se empleó el programa SPAdes v3.14.1; en donde el tamaño de k-mer de 77 ensambló las lecturas en contigs y luego en scaffolds para ambas bacterias. SPAdes es un programa que emplea un enfoque *multi-k-mer*, es decir en función de la longitud de las lecturas determina el tamaño de los k-mers a probar. Además utiliza la información de los contigs preensamblados en cada paso para generar el ensamblaje final (Vollmers et al., 2017; Voshall & Moriyama, 2018). Por esta razón, se ha utilizado en previas investigaciones para el ensamblaje de ambas bacterias (Bennett et al., 2016; Bennett & Mao, 2018; Mao et al., 2017; Waneka et al., 2020).

Del análisis de calidad, el ensamblaje de *Candidatus Sulcia muelleri* logró cubrir hasta 190,004 bp; valor aproximado al de su genoma de referencia de 190,733 bp (Bennett & Moran, 2013). De hecho, Mao et al. (2017) reportan que los endosimbiontes *Candidatus Sulcia muelleri* de la subfamilia Deltoccephalinae; a la que pertenece *Exitianus exitiosus*, tienen un genoma cuya longitud es < 200,000 bp. En cuanto al %CG, se encontró que la bacteria tiene un 23.8%, valor dentro del rango aproximado de

23,5 – 24% reportado por los mismos investigadores. Asimismo, cabe mencionar que el bajo %GC se debe a que los genomas de los endosimbiontes son abundantes en AT (adeninas y timinas), porque las bacterias carecen de mecanismos de reparación del ADN y tienen altas tasas de transiciones de GC por AT (Wernegreen, 2012). El correcto ensamblaje se confirmó analizando los demás parámetros como N50, L50 y porque no se generó ninguna secuencia mal ensamblada y todas se alinearon a la referencia en QUAST. Por lo tanto, al subir sus contigs a BLAST, todos identificaron como homólogos a *Candidatus Sulcia muelleri*; al igual que Chang et al. (2015) y Bennett et al. (2016) cuando ensamblaron el genoma completo de la bacteria para los aislados ML y PUNC; respectivamente.

Por otra parte, para *Candidatus Nasuia deltocephalinicola* no se obtuvo un buen ensamblaje de sus lecturas ya que en total se logró cubrir sólo 3,386 bp y la longitud de su genoma de referencia es de 112,091 bp (Bennett & Moran, 2013). Asimismo, su %CG fue de 39.10 y en otras investigaciones se han obtenido valores de 16.6 o 17.1% (Bennett et al., 2016; Bennett & Moran, 2013). Confirmándose que el genoma completo del endosimbionte de *Exitianus exitiosus* no pudo ensamblarse. Incluso en QUAST cuatro de sus cinco contigs no se alinearon a la referencia, lo que se corrobora con el análisis en BLAST. En donde, el último contig no identificó como homólogo a la secuencia de *Candidatus Nasuia deltocephalinicola*, sino a *Candidatus Rickettsiella viridis*. Posiblemente, para encontrar más regiones del genoma de la bacteria, se pueda secuenciar los bacteriomas del insecto para eliminar la gran cantidad de lecturas de hospedador y obtener más de los simbiontes; como se ha realizado en otras investigaciones (Husnik et al., 2020; Woyke et al., 2010). Asimismo, la identificación de *Candidatus Rickettsiella viridis* podría deberse a que aparte del sistema dual “*Sulcia-Nasuia*”, el saltahojas *Exitianus exitiosus* podría tener simbiontes secundarios, como se

ha reportado para otros Deltoccephalinae que de igual manera contienen otras bacterias como *Wolbachia* sp., *Rickettsia* sp., *Burkholderia* sp. y *Diplorickettsia* sp. (Kobialka et al., 2016). También existe la posibilidad de una pérdida o reemplazo de *Candidatus Nasuia deltocephalinicola* por alguna bacteria o levadura que le permita al anfitrión superar los desafíos de las dietas desequilibradas basadas en la savia de las plantas. Posiblemente por su alta deriva genética la bacteria con el tiempo perdió los genes esenciales para la simbiosis y fue reemplazada por otras bacterias co-primarias, como ha ocurrido para algunos insectos (Bennett & Moran, 2013).

Debido a las incongruencias encontradas en las secuencias de *Candidatus Nasuia deltocephalinicola* se decidió descartarla para los siguientes análisis y no se pudo comprobar que posee las rutas metabólicas involucradas en la síntesis de metionina e histidina como se ha publicado en los estudios de Bennett & Moran (2013) y Bennett et al. (2016).

En cuanto a la anotación del genoma de *Candidatus Sulcia muelleri*, se encontraron en total hasta 190 CDS, 30 ARNt y 3 ARNr; valores similares a los reportados por Bennett et al. (2016) del mismo endosimbionte, pero del saltahoja *Macrosteles quadripunctulatus* (Cicadellidae: Deltoccephalinae). Además, en el presente estudio también se encontró un ARNtm; este es un ARN bacteriano análogo en estructura y función a los ARNt y ARNm. Específicamente, los ARNtm intervienen en el proceso de *trans-traducción*; reciclando los ribosomas estancados en ARNm que han perdido su codón de parada, también añaden una etiqueta de degradación al polipéptido incompleto y favorecen la degradación del ARNm defectuoso (Doherty & Guo, 2016; Routh & Sankaranarayanan, 2017).

Asimismo, se encontró que los CDS de *Candidatus Sulcia muelleri* están involucrados en el metabolismo de aminoácidos, carbohidratos, proteínas, compuestos

aromáticos y ARN. Sin embargo, cabe mencionar que la bacteria carece de genes implicados en procesos esenciales como la síntesis de ADN, ARN y pared celular, reparación del ADN, transcripción, traducción y aminoacilación del ARNt; presentes en otros microorganismos de vida libre. De hecho, en los estudios de endosimbiontes bacterianos, se puede destacar que la mayoría de sus genomas son < 500,000 bp (Mao et al., 2018). Es decir, tienen una longitud diminuta comparada con otras bacterias como *Escherichia coli* de 4,639,221 bp (Gil et al., 2004). Esto se debe a que, dentro de sus anfitriones las bacterias experimentan una reducción de su genoma hasta el 90%. En donde empiezan a perder genes esenciales debido a su entorno de crecimiento inusualmente estable, transmisión vertical, pequeño tamaño de las poblaciones y baja recombinación hasta asexualidad completa que favorece las deleciones. No obstante, en su mayoría mantienen las rutas metabólicas necesarias para su hospedador (Mao et al., 2018; McCutcheon & Moran, 2010; Wernegreen, 2012). Esto se pudo comprobar ya que, de todos los genes involucrados en procesos metabólicos, el 63.3% correspondía al metabolismo de aminoácidos, compuestos esenciales para compensar la deficiencia nutricional de la dieta de *Exitianus exitiosus* por medio de su simbiote, que destina parte de su genoma a procesos de mayor utilidad para el insecto que para la bacteria (Gil et al., 2004). Bajo el mismo contexto, se confirmó que *Candidatus Sulcia muelleri* mantiene los genes codificantes de enzimas implicadas en la síntesis de ocho aminoácidos esenciales como la lisina, leucina, isoleucina, valina, treonina, arginina, fenilalanina y triptófano; al igual que McCutcheon & Moran (2010), McCutcheon et al. (2009) y Bennett & Moran (2013) encontraron en sus estudios.

No obstante, cabe mencionar que no se pudo hallar un gen involucrado en la síntesis de lisina y uno para la arginina. Posiblemente, como no se pudieron cubrir dos regiones del cromosoma de la bacteria de 1,264 y 94 bp (Figura 13 y 14), ambos genes

podrían encontrarse allí. Sin embargo, si al cubrir estos espacios los genes no son anotados, el anfitrión podría estar abasteciéndole de maquinaria genética a su simbionte para compensar sus funciones ausentes a través del fenómeno conocido como colaboración metabólica hospedador / endosimbionte. Por ejemplo, se ha demostrado que en bacterias que han perdido el gen *ilvE* encargado de la síntesis de aminoácidos de cadena ramificada, sus huéspedes complementan la función mediante la expresión de un gen que cumple el mismo rol (Wilson & Duncan, 2015).

Para el análisis filogenético de *Candidatus Sulcia muelleri* no se empleó el enfoque tradicional de construir árboles basados en las secuencias del 16S de varias bacterias (Brentassi et al., 2017) o de pocos genes (Koga & Moran, 2014; Tesneem & Rebecca, 2017). En su lugar se emplearon 24 genes ortólogos determinados por AmphoraNet; este es un enfoque que usa múltiples loci independientes que se concatenan y convergen en un solo árbol de especies robusto (Aguileta et al., 2008). Como Sloan & Moran (2012) hicieron para el endosimbionte *Candidatus Carsonella ruddii*; en donde AMPHORA2 determinó los 29 genes a usarse en el análisis filogenético.

Después cada gen se alineó, se eliminó sus fragmentos ambiguos, se concatenaron y se construyó un árbol filogenético mediante máxima verosimilitud e inferencia bayesiana, como se ha reportado en previas investigaciones para *Candidatus Sulcia muelleri* (Kobińska et al., 2015; Nishino et al., 2016). El árbol resultante se expresó como filograma, en donde la longitud de sus ramas es proporcional al número de cambios de caracteres en los datos (Elliott et al., 2018). Por lo tanto, se puede destacar que todas las *Candidatus Sulcia muelleri* de la familia Cicadidae no tienen muchos cambios entre sí y por ello la longitud de sus ramas fue corta en comparación con las demás. De igual manera, se puede resaltar que las secuencias de los

endosimbiontes se asociaron en clados con todos los representantes de su misma familia o subfamilia de huésped. En particular, la bacteria de *Exitianus exitiosus* formó un clado bien definido junto a los demás simbioses de los saltahojas de la subfamilia Deltocephalinae.

Esto también se ha observado en los análisis del ARNr 16S; en donde se ha comprobado que la filogenia de las bacterias es congruente con la filogenia de sus insectos hospedadores. Demostrando la antigua adquisición de *Candidatus Sulcia muelleri* por infección a un ancestro de los insectos hace aproximadamente 260 millones de años (Moran et al., 2005). Para establecer esta simbiosis, posiblemente la bacteria tuvo que superar las barreras físicas, celulares y moleculares del insecto para luego internalizarse, sobrevivir y replicarse dentro de la célula eucariota. Asimismo, tuvo que desarrollar mecanismos para garantizar su propagación a la descendencia y es por eso que actualmente los endosimbiontes primarios como *Candidatus Sulcia muelleri* se transfieren verticalmente por transmisión transovárica (Gil et al., 2004). Esto también ocurre en los simbioses obligatorios como *Candidatus Buchnera* sp., *Candidatus Wigglesworthia* sp., *Candidatus Blochmannia* sp. y *Candidatus Baumannia* sp. que viven en pulgones, moscas tsetsé, hormigas y francotirados; respectivamente. En donde su filogenia también refleja la filogenia del anfitrión, demostrando que estas asociaciones son estables durante el tiempo. No obstante, este fenómeno sólo se observa en simbioses obligatorios, mas no en simbioses secundarios; porque no son esenciales para la vida de sus insectos y por ello no se los puede encontrar dentro de bacteriomas, sino en los demás órganos o hemolinfa del animal. Incluso al construir su filogenia, esta no concuerda con la de su huésped, lo que sugiere que provienen de asociaciones recientes y no antiguas como la de *Candidatus Sulcia muelleri* (Kikuchi et al., 2009).

Capítulo VI: Conclusiones

- El mapeo de las lecturas de secuenciación de alto rendimiento de *Exitianus exitiosus* mostró que el 0,79% de los datos corresponden a *Candidatus Sulcia muelleri* y el 0,004% a *Candidatus Nasuia deltocephalinicola*.
- El ensamblaje de *Candidatus Sulcia muelleri* produjo nueve contigs que se agruparon en cinco scaffolds con un total de 190,004 bp y %GC de 23,8. Además, debido a la alta calidad del ensamblaje los scaffolds se cerraron en un cromosoma con la ayuda del genoma ALF; demostrando que la hipótesis se comprobó para la bacteria. Sin embargo, no se lograron cubrir dos regiones de 1,264 y 94 bp.
- El ensamblaje de *Candidatus Nasuia deltocephalinicola* produjo cinco contigs que se agruparon en cuatro scaffolds con un total de 3,396 bp y %GC de 39,10. Por lo tanto, no se logró encontrar el genoma completo de la bacteria que aproximadamente es de 112,091 bp; demostrando que la hipótesis no se comprobó para el endosimbionte y sus secuencias fueron descartadas para la anotación y filogenia.
- En total se anotaron 190 CDS, 30 ARNt, 3 ARNr y 1 ARNtm en el genoma de *Candidatus Sulcia muelleri* de *Exitianus exitiosus*. Además, se identificó que 98 CDS están involucrados en rutas metabólicas y de estos el 63,3% corresponden al metabolismo de aminoácidos.
- Se identificaron genes codificantes de enzimas involucradas en la biosíntesis de ocho aminoácidos esenciales en el genoma de *Candidatus Sulcia muelleri*. Por lo tanto, se pudo comprobar que el endosimbionte proporciona lisina, leucina, isoleucina, valina, treonina, arginina, triptófano y fenilalanina a *Exitianus exitiosus* para compensar su dieta deficiente en nutrientes.

- Se construyó un árbol filogenético con las secuencias de 24 genes ortólogos de 34 bacterias; incluido el outgroup y *Candidatus Sulcia muelleri* de *Exitianus exitiosus*. Se obtuvieron diferentes clados con un alto soporte de bootstrap (99-100%) y probabilidad posterior (1).
- El análisis filogenético demostró que las secuencias de *Candidatus Sulcia muelleri* de *Exitianus exitiosus* se agruparon en un solo clado robusto junto a todos los miembros de la subfamilia Deldocephalinae. De esta manera, se comprobó que la filogenia de los endosimbiontes primarios es congruente con la de sus hospederos.

Capítulo VII: Recomendaciones

- Ensamblar todas las lecturas de *Exitianus exitiosus* y comparar los contigs con los obtenidos mediante el mapeo de lecturas a *Candidatus Sulcia muelleri* y *Candidatus Nasuia deltocephalinicola*.
- Emplear las lecturas de secuenciación de los bacteriomas, para aumentar el número de lecturas correspondientes a los endosimbiontes y eliminar la gran cantidad de datos de los huéspedes como *Exitianus exitiosus*.
- Emplear algún software de identificación taxonómica como Kraken, para determinar la presencia de simbioses secundarios y no sólo de primarios como *Candidatus Sulcia muelleri*.
- Diseñar primers, amplificar las zonas no cubiertas del cromosoma de *Candidatus Sulcia muelleri* y secuenciar los fragmentos para rellenar los espacios faltantes.
- Anotar también las secuencias de *Exitianus exitiosus* para determinar si posee los genes para completar las rutas biosintéticas de lisina y arginina.

Bibliografía

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>
- Aguileta, G., Marthey, S., Chiapello, H., Lebrun, M.-H., Rodolphe, F., Fournier, E., Gendrault-Jacquemard, A., & Giraud, T. (2008). Assessing the Performance of Single-Copy Genes for Recovering Robust Phylogenies. *Systematic Biology*, *57*(4), 613–627. <https://doi.org/10.1080/10635150802306527>
- Amatore, Z., Gunn, S., & Harris, L. K. (2020). An Educational Bioinformatics Project to Improve Genome Annotation. *Frontiers in Microbiology*, *11*. <https://doi.org/10.3389/fmicb.2020.577497>
- Andreason, S. (2016). Discrimination and transcriptional response analysis of hemipteran phytopathogen vectors. *Oklahoma State University*, 1-184.
- Andrews, S. (2010). *FastQC: A Quality Control tool for High Throughput Sequence Data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., ... Zagnitko, O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, *9*(1), 75. <https://doi.org/10.1186/1471-2164-9-75>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Pribelski, A. D., Pyshkin, A. V., Sirotkin, A. V.,

- Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477.
<https://doi.org/10.1089/cmb.2012.0021>
- Bennett, G. M., Abbà, S., Kube, M., & Marzachi, C. (2016). Complete Genome Sequences of the Obligate Symbionts “*Candidatus Sulcia muelleri*” and “*Ca. Nasuia deltocephalinicola*” from the Pestiferous Leafhopper *Macrostelus quadripunctulatus* (Hemiptera: Cicadellidae). *Genome Announcements*, *4*(1).
<https://doi.org/10.1128/genomeA.01604-15>
- Bennett, G. M., & Mao, M. (2018). Comparative genomics of a quadripartite symbiosis in a planthopper host reveals the origins and rearranged nutritional responsibilities of anciently diverged bacterial lineages. *Environmental Microbiology*, *20*(12), 4461–4472. <https://doi.org/10.1111/1462-2920.14367>
- Bennett, G. M., & Moran, N. A. (2013). Small, Smaller, Smallest: The Origins and Evolution of Ancient Dual Symbioses in a Phloem-Feeding Insect. *Genome Biology and Evolution*, *5*(9), 1675–1688. <https://doi.org/10.1093/gbe/evt118>
- Bertaccini, A., Weintraub, P. G., Rao, G. P., & Mori, N. (Eds.). (2019). *Phytoplasmas: Plant Pathogenic Bacteria - II: Transmission and Management of Phytoplasma - Associated Diseases*. Springer Singapore. <https://doi.org/10.1007/978-981-13-2832-9>
- Brentassi, M. E., Franco, E., Balatti, P., Medina, R., Bernabei, F., & Marino de Remes Lenicov, A. M. (2017). Bacteriomes of the corn leafhopper, *Dalbulus maidis* (DeLong & Wolcott, 1923) (Insecta, Hemiptera, Cicadellidae: Deltocephalinae) harbor *Sulcia* symbiont: molecular characterization, ultrastructure, and transovarial transmission. *Protoplasma*, *254*(3), 1421–1429.
<https://doi.org/10.1007/s00709-016-1033-4>

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. <https://doi.org/10.1186/1471-2105-10-421>
- Cao, Y., Fanning, S., Proos, S., Jordan, K., & Srikumar, S. (2017). A Review on the Applications of Next Generation Sequencing Technologies as Applied to Food-Related Microbiome Studies. *Frontiers in Microbiology*, *8*. <https://doi.org/10.3389/fmicb.2017.01829>
- Carrico, J. A., Rossi, M., Moran-Gilad, J., Van Domselaar, G., & Ramirez, M. (2018). A primer on microbial bioinformatics for nonbioinformaticians. *Clinical Microbiology and Infection*, *24*(4), 342–349. <https://doi.org/10.1016/j.cmi.2017.12.015>
- Carver, T., Thomson, N., Bleasby, A., Berriman, M., & Parkhill, J. (2009). DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics*, *25*(1), 119–120. <https://doi.org/10.1093/bioinformatics/btn578>
- Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., Ong, W. K., Paley, S., Subhraveti, P., & Karp, P. D. (2020). The MetaCyc database of metabolic pathways and enzymes—A 2019 update. *Nucleic Acids Research*, *48*(D1), D445–D453. <https://doi.org/10.1093/nar/gkz862>
- Chang, H.-H., Cho, S.-T., Canale, M. C., Mugford, S. T., Lopes, J. R. S., Hogenhout, S. A., & Kuo, C.-H. (2015). Complete Genome Sequence of “*Candidatus Sulcia muelleri*” ML, an Obligate Nutritional Symbiont of Maize Leafhopper (*Dalbulus maidis*). *Genome Announcements*, *3*(1). <https://doi.org/10.1128/genomeA.01483-14>
- Chen, X., Hitchings, M. D., Mendoza, J. E., Balanza, V., Facey, P. D., Dyson, P. J., Bielza, P., & Del Sol, R. (2017). Comparative Genomics of Facultative Bacterial Symbionts Isolated from European Orius Species Reveals an Ancestral

Symbiotic Association. *Frontiers in Microbiology*, 8.

<https://doi.org/10.3389/fmicb.2017.01969>

Commins, J., Toft, C., & Fares, M. A. (2009). Computational Biology Methods and Their Application to the Comparative Genomics of Endocellular Symbiotic Bacteria of Insects. *Biological Procedures Online*, 11(1), 52. <https://doi.org/10.1007/s12575-009-9004-1>

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution*, 37(1), 291–294. <https://doi.org/10.1093/molbev/msz189>

Di Tommaso, P., Moretti, S., Xenarios, I., Orobittg, M., Montanyola, A., Chang, J.-M., Taly, J.-F., & Notredame, C. (2011). T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, 39(Web Server issue), W13–W17. <https://doi.org/10.1093/nar/gkr245>

Doherty, J., & Guo, M. (2016). Transfer RNA. En R. A. Bradshaw & P. D. Stahl (Eds.), *Encyclopedia of Cell Biology* (pp. 309–340). Academic Press. <https://doi.org/10.1016/B978-0-12-394447-4.10039-2>

Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., Amselem, J., Bouri, L., Bocs, S., Klopp, C., Gibrat, J.-F., Vlasova, A., Leskosek, B. L., Soler, L., Binzer-Panchal, M., & Lantz, H. (2018). Ten steps to get started in Genome Assembly and Annotation. *F1000Research*, 7. <https://doi.org/10.12688/f1000research.13598.1>

Elliott, M. J., Knerr, N. J., & Schmidt-Lebuhn, A. N. (2018). Choice between phylogram and chronogram can have a dramatic impact on the location of phylogenetic

- diversity hotspots. *Journal of Biogeography*, 45(9), 2190–2201.
<https://doi.org/10.1111/jbi.13399>
- Gahm, N. (2017). The ecology of grassland leafhoppers (hemiptera: cicadellidae) with emphasis on the genus *balclutha*. *Oklahoma State University*, 91.
- Gil, R., & Latorre, A. (2019). Unity Makes Strength: A Review on Mutualistic Symbiosis in Representative Insect Clades. *Life*, 9(1), 21.
<https://doi.org/10.3390/life9010021>
- Gil, R., Latorre, A., & Moya, A. (2004). Bacterial endosymbionts of insects: Insights from comparative genomics. *Environmental Microbiology*, 6(11), 1109–1122.
<https://doi.org/10.1111/j.1462-2920.2004.00691.x>
- Gregory, T. R. (2008). Understanding Evolutionary Trees. *Evolution: Education and Outreach*, 1(2), 121–137. <https://doi.org/10.1007/s12052-008-0035-x>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075.
<https://doi.org/10.1093/bioinformatics/btt086>
- Guyomar, C., Delage, W., Legeai, F., Mougel, C., Simon, J.-C., & Lemaitre, C. (2019). MinYS: Mine Your Symbiont by targeted genome assembly in symbiotic communities. *BioRxiv*, 2019.12.13.875021.
<https://doi.org/10.1101/2019.12.13.875021>
- Husnik, F., Hypsa, V., & Darby, A. (2020). Insect—Symbiont Gene Expression in the Midgut Bacteriocytes of a Blood-Sucking Parasite. *Genome Biology and Evolution*, 12(4), 429–442. <https://doi.org/10.1093/gbe/evaa032>
- Ishii, Y., Matsuura, Y., Kakizawa, S., Nikoh, N., & Fukatsu, T. (2013). Diversity of Bacterial Endosymbionts Associated with Macrosteles Leafhoppers Vectoring Phytopathogenic Phytoplasmas. *Applied and Environmental Microbiology*, 79(16), 5013–5022. <https://doi.org/10.1128/AEM.01527-13>

- Janies, D. A., Studer, J., Handelman, S. K., & Linchangco, G. (2013). A comparison of supermatrix and supertree methods for multilocus phylogenetics using organismal datasets. *Cladistics*, *29*(5), 560–566.
<https://doi.org/10.1111/cla.12014>
- Jones, T. L., & Medina, R. F. (2020). Corn Stunt Disease: An Ideal Insect–Microbial–Plant Pathosystem for Comprehensive Studies of Vector-Borne Plant Diseases of Corn. *Plants*, *9*(6), 747. <https://doi.org/10.3390/plants9060747>
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30.
- Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, *21*(7), 428–444. <https://doi.org/10.1038/s41576-020-0233-0>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kerepesi, C., Bánky, D., & Grolmusz, V. (2014). AmphoraNet: The webserver implementation of the AMPHORA2 metagenomic workflow suite. *Gene*, *533*(2), 538–540. <https://doi.org/10.1016/j.gene.2013.10.015>
- Kerepesi, C., Szalkai, B., & Grolmusz, V. (2015). Visual Analysis of the Quantitative Composition of Metagenomic Communities: The AmphoraVizu Webserver. *Microbial Ecology*, *69*(3), 695–697. <https://doi.org/10.1007/s00248-014-0502-6>
- Kikuchi, Y. (2009). Endosymbiotic Bacteria in Insects: Their Diversity and Culturability. *Microbes and Environments*, *24*(3), 195–204.
<https://doi.org/10.1264/jsme2.ME09140S>

- Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., Tago, K., & Fukatsu, T. (2012). Symbiont-mediated insecticide resistance. *Proceedings of the National Academy of Sciences*, *109*(22), 8618–8622. <https://doi.org/10.1073/pnas.1200231109>
- Kikuchi, Y., Hosokawa, T., Nikoh, N., Meng, X.-Y., Kamagata, Y., & Fukatsu, T. (2009). Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biology*, *7*(1), 2. <https://doi.org/10.1186/1741-7007-7-2>
- Kobińska, M., Michalik, A., Walczak, M., Junkiert, Ł., & Szklarzewicz, T. (2015). Symbiotic microorganisms of the leafhopper *Deltocephalus pulicaris* (Fallén, 1806) (Insecta, Hemiptera, Cicadellidae: Deltocephalinae): molecular characterization, ultrastructure and transovarial transmission. *Polish Journal of Entomology*, *84*(4), 289–304. <https://doi.org/10.1515/pjen-2015-0025>
- Kobińska, M., Michalik, A., Walczak, M., Junkiert, Ł., & Szklarzewicz, T. (2016). *Sulcia* symbiont of the leafhopper *Macrostelus laevis* (Ribaut, 1927) (Insecta, Hemiptera, Cicadellidae: Deltocephalinae) harbors *Arsenophonus* bacteria. *Protoplasma*, *253*, 903–912. <https://doi.org/10.1007/s00709-015-0854-x>
- Koga, R., & Moran, N. A. (2014). Swapping symbionts in spittlebugs: Evolutionary replacement of a reduced genome symbiont. *The ISME Journal*, *8*(6), 1237–1246. <https://doi.org/10.1038/ismej.2013.235>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Leung, R. (2005). *Species Exitianus exitiosus—Gray Lawn Leafhopper—BugGuide.Net*. <https://bugguide.net/node/view/10225>

- Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., Li, Z., & Walters, J. R. (2019). Insect genomes: Progress and challenges. *Insect Molecular Biology*, 28(6), 739–758.
<https://doi.org/10.1111/imb.12599>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
<https://doi.org/10.1093/bioinformatics/btp352>
- Makarenkov, V., Kevorkov, D., & Legendre, P. (2006). 3—Phylogenetic Network Construction Approaches. En D. K. Arora, R. M. Berka, & G. B. Singh (Eds.), *Applied Mycology and Biotechnology* (Vol. 6, pp. 61–97). Elsevier.
[https://doi.org/10.1016/S1874-5334\(06\)80006-7](https://doi.org/10.1016/S1874-5334(06)80006-7)
- Mao, M., Yang, X., & Bennett, G. M. (2018). Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. *Proceedings of the National Academy of Sciences*, 115(50), E11691–E11700.
<https://doi.org/10.1073/pnas.1811932115>
- Mao, M., Yang, X., Poff, K., & Bennett, G. (2017). Comparative Genomics of the Dual-Obligate Symbionts from the Treehopper, *Entylia carinata* (Hemiptera: Membracidae), Provide Insight into the Origins and Evolution of an Ancient Symbiosis. *Genome Biology and Evolution*, 9(6), 1803–1815.
<https://doi.org/10.1093/gbe/evx134>
- McCutcheon, J. P., McDonald, B. R., & Moran, N. A. (2009). Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proceedings of the National Academy of Sciences of the United States of America*, 106(36), 15394–15399.
<https://doi.org/10.1073/pnas.0906424106>

- McCutcheon, J. P., & Moran, N. A. (2007). Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proceedings of the National Academy of Sciences*, *104*(49), 19392. <https://doi.org/10.1073/pnas.0708855104>
- McCutcheon, J. P., & Moran, N. A. (2010). Functional Convergence in Reduced Genomes of Bacterial Symbionts Spanning 200 My of Evolution. *Genome Biology and Evolution*, *2*, 708–718. <https://doi.org/10.1093/gbe/evq055>
- Mendoza, J. (2012). Aportes de la filogenética a la investigación médica. *Revista Medica Herediana*, *23*(2), 119–127. <https://doi.org/10.20453/rmh.v23i2.1042>
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, 1–8. <https://doi.org/10.1109/GCE.2010.5676129>
- Moran, N. A., Tran, P., & Gerardo, N. M. (2005). Symbiosis and Insect Diversification: An Ancient Symbiont of Sap-Feeding Insects from the Bacterial Phylum Bacteroidetes. *Applied and Environmental Microbiology*, *71*(12), 8802–8810. <https://doi.org/10.1128/AEM.71.12.8802-8810.2005>
- Munjal, G., Hanmandlu, M., & Srivastava, S. (2018). Phylogenetics Algorithms and Applications. *Ambient Communications and Computer Systems*, *904*, 187–194. https://doi.org/10.1007/978-981-13-5934-7_17
- Nishino, T., Tanahashi, M., Lin, C.-P., Koga, R., & Fukatsu, T. (2016). Fungal and bacterial endosymbionts of eared leafhoppers of the subfamily Ledrinae (Hemiptera: Cicadellidae). *Applied Entomology and Zoology*, *51*(3), 465–477. <https://doi.org/10.1007/s13355-016-0422-7>
- Noda, H., Watanabe, K., Kawai, S., Yukuhiro, F., Miyoshi, T., Tomizawa, M., Koizumi, Y., Nikoh, N., & Fukatsu, T. (2012). Bacteriome-associated endosymbionts of the green rice leafhopper *Nephotettix cincticeps* (Hemiptera: Cicadellidae). *Applied*

- Entomology and Zoology*, 47(3), 217–225. <https://doi.org/10.1007/s13355-012-0110-1>
- Oliver, K. M., Russell, J. A., Moran, N. A., & Hunter, M. S. (2003). Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, 100(4), 1803–1807. <https://doi.org/10.1073/pnas.0335320100>
- Paniagua Voirol, L. R., Frago, E., Kaltenpoth, M., Hilker, M., & Fatouros, N. E. (2018). Bacterial Symbionts in Lepidoptera: Their Diversity, Transmission, and Impact on the Host. *Frontiers in Microbiology*, 9. <https://doi.org/10.3389/fmicb.2018.00556>
- Pfeifer, S. P. (2017). From next-generation resequencing reads to a high-quality variant data set. *Heredity*, 118(2), 111–124. <https://doi.org/10.1038/hdy.2016.102>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Radcliffe, E. B., & Lagnaoui, A. (2007). Chapter 25—Insect Pests in Potato. En D. Vreugdenhil, J. Bradshaw, C. Gebhardt, F. Govers, D. K. L. Mackerron, M. A. Taylor, & H. A. Ross (Eds.), *Potato Biology and Biotechnology* (pp. 543–567). Elsevier Science B.V. <https://doi.org/10.1016/B978-044451018-1/50067-1>
- Rambaut, A. (2014). *FigTree v1.4.2, A Graphical Viewer of Phylogenetic Trees*. <http://tree.bio.ed.ac.uk/software/figtree/>
- Rice, E. S., & Green, R. E. (2019). New Approaches for Genome Assembly and Scaffolding. *Annual Review of Animal Biosciences*, 7(1), 17–40. <https://doi.org/10.1146/annurev-animal-020518-115344>
- Richardson, E. J., & Watson, M. (2013). The automatic annotation of bacterial genomes. *Briefings in Bioinformatics*, 14(1), 1–12. <https://doi.org/10.1093/bib/bbs007>

- Routh, S. B., & Sankaranarayanan, R. (2017). Editing and Proofreading in Translation ☆. En *Reference Module in Life Sciences*. Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.06374-3>
- Rubio, S., Pacheco-Orozco, R. A., Gómez, A. M., Perdomo, S., & García-Robles, R. (2020). Secuenciación de nueva generación (NGS) de ADN: Presente y futuro en la práctica clínica: Presente y futuro en la práctica clínica. *Universitas Medica*, 61(2), Article 2. <https://doi.org/10.11144/Javeriana.umed61-2.sngs>
- Saguez, J., Olivier, C., Hamilton, A., Lowery, T., Stobbs, L., Lasnier, J., Galka, B., Chen, X., Mauffette, Y., & Vincent, C. (2014). Diversity and abundance of leafhoppers in Canadian vineyards. *Journal of Insect Science*, 14. <https://doi.org/10.1093/jis/14.1.73>
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., & Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, 407(6800), 81–86. <https://doi.org/10.1038/35024074>
- Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next Generation Sequencing Technologies. *Current protocols in molecular biology*, 122(1), e59. <https://doi.org/10.1002/cpmb.59>
- Sloan, D. B., & Moran, N. A. (2012). Genome Reduction and Co-evolution between the Primary and Secondary Bacterial Symbionts of Psyllids. *Molecular Biology and Evolution*, 29(12), 3781–3792. <https://doi.org/10.1093/molbev/mss180>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>

- Steward, C. A., Parker, A. P. J., Minassian, B. A., Sisodiya, S. M., Frankish, A., & Harrow, J. (2017). Genome annotation for clinical genomic diagnostics: Strengths and weaknesses. *Genome Medicine*, *9*(1), 49. <https://doi.org/10.1186/s13073-017-0441-1>
- Stewart, L. R., Jarugula, S., Zhao, Y., Qu, F., & Marty, D. (2017). Identification of a maize chlorotic dwarf virus silencing suppressor protein. *Virology*, *504*, 88–95. <https://doi.org/10.1016/j.virol.2016.11.017>
- Stiller, M. (2009). Leafhoppers associated with grasslands of South Africa. Part I. Grassland Biome endemics. *The Grassland Society of Southern Africa*, *9*(4), 3.
- Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(vey016). <https://doi.org/10.1093/ve/vey016>
- Szklarzewicz, T., Grzywacz, B., Szwedo, J., & Michalik, A. (2016). Bacterial symbionts of the leafhopper *Evacanthus interruptus* (Linnaeus, 1758) (Insecta, Hemiptera, Cicadellidae: Evacanthinae). *Protoplasma*, *253*, 379–391. <https://doi.org/10.1007/s00709-015-0817-2>
- Talavera, G., & Castresana, J. (2007). Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology*, *56*(4), 564–577. <https://doi.org/10.1080/10635150701472164>
- Tan, G., Opitz, L., Schlapbach, R., & Rehrauer, H. (2019). Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific Reports*, *9*(1), 2856. <https://doi.org/10.1038/s41598-019-39076-7>
- Tanizawa, Y., Fujisawa, T., & Nakamura, Y. (2018). DFAST: A flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics*, *34*(6), 1037–1039. <https://doi.org/10.1093/bioinformatics/btx713>

- Tesneem, N., & Rebecca, C. (2017). A novel groel gene from the endosymbiont of beet leafhopper, *Candidatus Sulcia muelleri*. *African Journal of Microbiology Research*, 11(44), 1586–1599. <https://doi.org/10.5897/AJMR2017.8725>
- Vollmers, J., Wiegand, S., & Kaster, A.-K. (2017). Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective—Not Only Size Matters! *PLOS ONE*, 12(1), e0169662. <https://doi.org/10.1371/journal.pone.0169662>
- Voshall, A., & Moriyama, E. N. (2018). Next-Generation Transcriptome Assembly: Strategies and Performance Analysis. *Bioinformatics in the Era of Post Genomics and Big Data*. <https://doi.org/10.5772/intechopen.73497>
- Wajid, B., & Serpedin, E. (2012). Review of General Algorithmic Features for Genome Assemblers for Next Generation Sequencers. *Genomics, Proteomics & Bioinformatics*, 10(2), 58–73. <https://doi.org/10.1016/j.gpb.2012.05.006>
- Waneka, G., Vasquez, Y. M., Bennett, G. M., & Sloan, D. B. (2020). Mutational pressure drives differential genome conservation in two bacterial endosymbionts of sap feeding insects. *BioRxiv*, 2020.07.29.225037. <https://doi.org/10.1101/2020.07.29.225037>
- Wang, H., Wu, N., Liu, Y., Kundu, J. K., Liu, W., & Wang, X. (2019). Higher Bacterial Diversity of Gut Microbiota in Different Natural Populations of Leafhopper Vector Does Not Influence WDV Transmission. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.01144>
- Wernegreen, J. J. (2012). Strategies of genomic integration within insect-bacterial mutualisms. *The Biological Bulletin*, 223(1), 112. <https://doi.org/10.1086/BBLv223n1p112>
- Wilson, A. C. C., & Duncan, R. P. (2015). Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proceedings of the National*

Academy of Sciences, 112(33), 10255–10261.

<https://doi.org/10.1073/pnas.1423305112>

Woyke, T., Tighe, D., Mavromatis, K., Clum, A., Copeland, A., Schackwitz, W., Lapidus, A., Wu, D., McCutcheon, J. P., McDonald, B. R., Moran, N. A., Bristow, J., & Cheng, J.-F. (2010). One Bacterial Cell, One Complete Genome. *PLOS ONE*, 5(4), e10314. <https://doi.org/10.1371/journal.pone.0010314>

Wu, D., Daugherty, S. C., Van Aken, S. E., Pai, G. H., Watkins, K. L., Khouri, H., Tallon, L. J., Zaborsky, J. M., Dunbar, H. E., Tran, P. L., Moran, N. A., & Eisen, J. A. (2006). Metabolic Complementarity and Genomics of the Dual Bacterial Symbiosis of Sharpshooters. *PLoS Biology*, 4(6).

<https://doi.org/10.1371/journal.pbio.0040188>

Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: Principles and practice. *Nature Reviews Genetics*, 13(5), 303–314. <https://doi.org/10.1038/nrg3186>

Young, A. D., & Gillung, J. P. (2020). Phylogenomics—Principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, 45(2), 225–247.

<https://doi.org/10.1111/syen.12406>