



Benemérita Universidad Autónoma de Puebla

Instituto de Ciencias BUAP

Centro de Investigaciones en Ciencias Microbiológicas

Laboratorio de Ecología Molecular Microbiana

Título de la tesis

**“Secuenciación del genoma de la cepa *Gluconacetobacter* sp.
UAPS01-405”**

Que para obtener el título de

Maestra en Ciencias (Microbiología)

Con opción en Bioquímica y Genética Microbiana

Presenta

LBM. Rigel Quintero Luis

Director de Tesis

D. en C. Luis Ernesto Fuentes Ramírez

Co-Directora de Tesis

D. en C. Liliana López Pliego

Asesores

D. en C. Miguel Castañeda Lucio

D. en C. Yagul Pedraza Perez

D. en C. José Antonio Munive Hernández

D. en C. María del Rocío Bustillos Cristales

Puebla, Agosto 2025



BUAP

Puebla, Pue. a 10 de abril 2025.

**A LA ACADEMIA DEL POSGRADO
EN MICROBIOLOGÍA
CICM-ICUAP
PRESENTE.**


Por este conducto los abajo firmantes, integrantes del Comité revisor de Tesis de Maestría de la alumna **Rigel Quintero Luis**, les informamos que hemos revisado el escrito titulado:


“Secuenciación del genoma de la cepa *Gluconacetobacter* sp. UAPS01-405”

A nuestro juicio, la alumna **Rigel Quintero Luis** puede proceder a la digitalización de la tesis y a la presentación del examen de grado correspondiente.

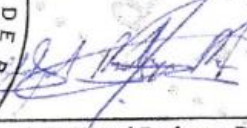
Sin más que agregar, nos despedimos afectuosamente de ustedes.

Atentamente.
“Pensar Bien Para Vivir Mejor”


D.C. María del Rocío Bustillos Cristales


D.C. Miguel Castañeda Lucio


D.C. José Antonio Munive Hernández


D.C. Yagul Pedraza Pérez

Agradecimientos

Agradezco a todas aquellas personas que con su aliento, conocimientos y cariño contribuyeron a la culminación de este esfuerzo académico.

Al consejo nacional de ciencia y tecnología (CONACYT), por la beca otorgada con número 822042 que me permitió cumplir mi sueño de realizar un posgrado. Al Posgrado de Ciencias Microbiológicas por el apoyo para asistir al Congreso Nacional de Bioquímica XXXIV.

A mi mamá, Patricia Luis Torres por su apoyo constante e incondicional, sin ti esto no sería posible, gracias por creer en mí y nunca dejarme sola.

A mis madrinas Carmelita Luis Torres y Ma. de la Luz Luis Torres, por alentarme y animarme cada vez que lo necesite, su cariño me ayudó a continuar esforzándome siempre.

A mis asesores y padres académicos: Dra. Rocío Bustillos Cristales y Dr. Luis Ernesto Fuentes Ramírez, por nutrirme con sus conocimientos y alentarme en mi vocación académica.

A los miembros de mi comité tutorial, sus valiosos aportes y retroalimentación a este proyecto, lo han mejorado y han reforzado mi formación profesional.

A Carlos Codiz Gaona, con tu amable labor y dedicación se nutrió la elaboración de esta investigación, gracias.

A mis compañeras de generación Marina, Fernanda y Estephanie por acompañarme en todo este proceso.

A mi mejor amiga y colega Aleyda Gonzalez Romero, tu amistad ha sido un gran apoyo, gracias por acompañarme en este proceso.

Contenido

1. Resumen.....	6
2. Introducción.....	7
Bacterias Ácido Acéticas.....	7
<i>Conopholis alpina</i>	8
3. Antecedentes	10
<i>Gluconacetobacter</i>	10
Cepa UAPS01-405.....	12
Estrategias bioinformáticas para el análisis de genomas bacterianos	13
4. Justificación.....	15
5. Objetivos	16
Objetivo General	16
Objetivos Específicos	16
6. Metodología	17
Anotación	19
Análisis completo del genoma.....	20
Búsqueda global en BLAST-XYplot viewer	20
Análisis BLASTn con genes Housekeeping	21
Análisis FastANI.....	21
Elaboración de Árbol Filogenético de Pangenoma.....	21
Análisis MLSA con genes relevantes para el género <i>Gluconacetobacter</i>	23
Alineamiento de genomas del BV-BRC.....	23
7. Resultados	24
Ensamble.....	24
• Gráfico de ensamblaje:.....	24
• Reporte de Ensamble de Genoma:	25

•	Reporte QUAST:	25
	Anotación	28
	Análisis Completo de Genoma	30
•	Reporte de análisis de genoma completo	30
•	Anotación del Genoma	31
•	Visualización circular del genoma ensamblado	33
•	Análisis de Subsistemas	34
	BLAST-XYplot viewer	35
	BLASTn del gen16S rRNA	37
	BLASTn con genes Housekeeping	38
	FastANI	39
	Análisis MLSA con genes relevantes para el género <i>Gluconacetobacter</i>	40
	Árbol Filogenético de Pangenoma	41
	Alineamiento de Genomas	42
8.	Discusión	44
9.	Conclusiones	48
10.	Perspectivas	48
11.	Bibliografía	49
12.	Anexos	59
•	Genes de Especialidad	59
•	Genes de Resistencia Antimicrobiana	59
•	BLAST	61
•	Alineamiento de Genomas	65

1. Resumen

Las Bacterias del Ácido Acético (BAA) son miembros de la familia *Acetobacteraceae*. Gracias a su habilidad para metabolizar etanol, glucosa y acetato, tienen un papel clave en la producción de vinagre, kombucha y otros productos de relevancia comercial. Algunas especies han sido aisladas de plantas y participan en interacciones planta-microorganismo, como la promoción del crecimiento vegetal.

Conopholis alpina es una planta parásita, que coloniza las raíces de los árboles en los bosques de encino, cubierta por la sombra de los árboles, no realiza fotosíntesis, sino que se alimenta de la planta anfitriona. Hasta la actualidad, no se han reportado trabajos de investigación relacionados a las interacciones microbiológicas de esta planta. El presente trabajo de investigación tiene como objetivo estudiar un aislado bacteriano ácido acético obtenido de las flores de esta planta.

El aislado UAPS01-405, ha mostrado crecimiento en medio LGI y rasgos fenotípicos característicos del género *Gluconacetobacter*, cuyo ensamblaje, anotación, análisis FastANIs, MLSA, BLAST de genes housekeeping y alineamiento de genomas han mostrado que el microorganismo es cercano a algunos miembros del género como *Ga. diazotrophicus*, *Ga. liquefaciens*, *Ga. dulcium* y *Ga. sacchari*, sin embargo, su identificación a nivel de especie queda por determinarse.

2. Introducción

Bacterias Ácido Acéticas

Las bacterias que, por su capacidad de producir ácidos orgánicos, realizan la oxidación del etanol hacia ácido acético, son comúnmente llamadas “Bacterias Ácido Acéticas” (BAA), de la clase *Alphaproteobacteria*, orden *Rhodospirillales*, familia *Acetobacteraceae*. Una característica que las distingue de otras bacterias es que su metabolismo primario es realizado a pH ácido con un óptimo para su crecimiento entre 5 y 6.5, pero toleran valores menores a 4 (Fuentes-Ramírez *et al.*, 2001; Reis y Texeira, 2015; Raspor y Goranovic, 2008 y Mamlock y Gullo, 2013).

Son bacterias aerobias estrictas, de forma bacilar, gram negativas y mesófilas, con una temperatura óptima de crecimiento varía entre los 28°C y 30°C. Están ampliamente distribuidas en la naturaleza, han sido aisladas de plantas, flores, hierbas, frutas y raíces, alimentos y bebidas fermentadas. Las BAA participan en la producción de alimentos de interés comercial, como vinagre, kombucha, cocoa y otros productos fermentados (Reis y Texeira, 2015). Sin embargo, su actividad metabólica también puede causar deterioro en bebidas alcohólicas como el vino, la cerveza y la sidra, generando sabores avinagrados, turbidez y viscosidad. Además, algunas de estas bacterias producen celulosa bacteriana, un polímero con diversas aplicaciones industriales incluyendo la fabricación de papel, textiles, diafragmas de bocinas, mascarillas cosméticas, gasas médicas, preparación de tabletas y comprimidos (Mamlock y Gullo, 2013; Raspor y Goranovic, 2008; Gomes *et al.*, 2018 y Törün *et al.*, 2019).

Hasta la década del 2000 este grupo se consideraba pequeño desde el punto de vista taxonómico, pues se conformaba de pocos ejemplares. Sin embargo, la adición de nuevos miembros ha sido constante y se espera la incorporación de nuevas especies (Yamada y Yukphan, 2008).

Conopholis alpina

Conocida comúnmente como “*Elote de coyote*” o “*Elotillo*”, *Conopholis alpina*, es una planta holoparásita que habita las raíces de los árboles. Carece de clorofila y no realiza fotosíntesis, por lo que depende del desarrollo de haustorios, que le permitan conectarse al sistema vascular del árbol hospedero para recibir nutrientes orgánicos e inorgánicos directamente de este. No se ha reportado hasta el momento que cause daño en la planta anfitriona (Rodrigues *et al.*, 2013).

Esta planta es residente de los bosques de encino, ubicándose principalmente en las faldas de los robles blancos. Su distribución global se encuentra en Norteamérica, desde los Estados Unidos de América hasta Panamá. A nivel nacional se ha encontrado en los estados de Sonora, Sinaloa, Chihuahua, Michoacán, Jalisco, Tlaxcala, Veracruz y Puebla (El cinturón volcánico trans mexicano) como podemos ver en la figura 2 (Rodrigues *et al.*, 2011).



Figura 1. *Conopholis alpina* Abair, *et al.*, 2011.

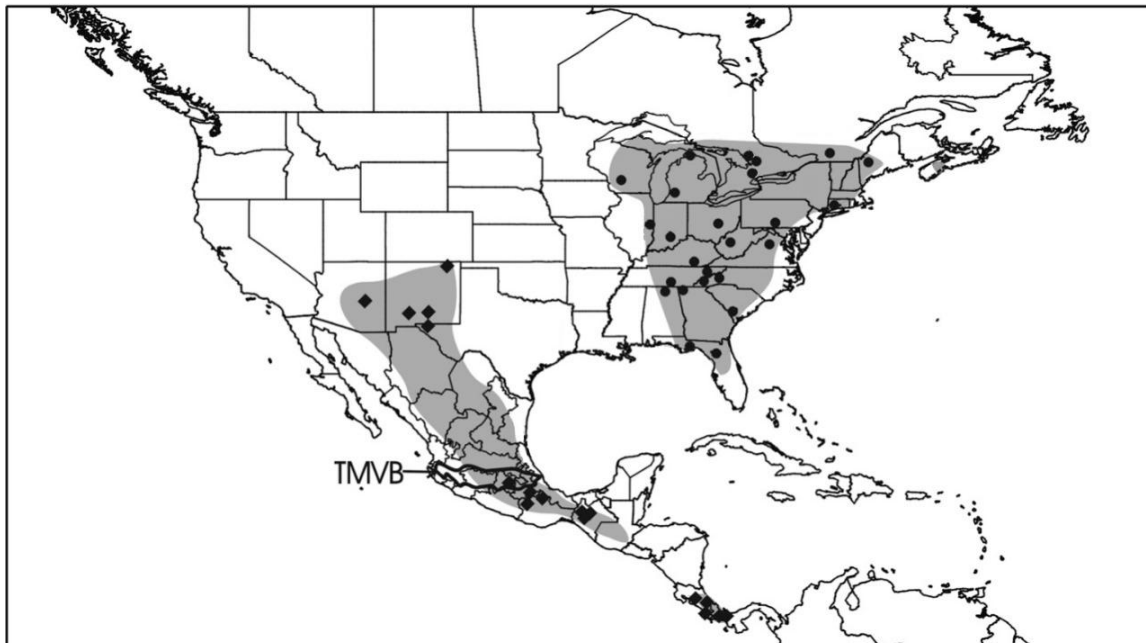


Figura 2. Mapa de la distribución de *Conopholis* (áreas sombreadas). Los círculos representan sitios de muestreo para las poblaciones de *C. americana*, los rombos representan las de *C. alpina* TMVB = Cinturón volcánico Trans-mexicano. Adaptado de Rodrigues, *et al.*, 2011.

La época de floración de *Conopholis alpina* es entre los meses de marzo a agosto, sin embargo, en la zona del altiplano se restringe de marzo a abril. Mientras que la temporada de fructificación es entre junio y noviembre (Calderón de Rzedowski *et al.*, 1998).

Algunos pueblos originarios han utilizado esta planta con fines medicinales desde tiempos ancestrales. Los Acoma de Nuevo México en el tratamiento de tuberculosis, los Rarámuri como relajante durante el parto y en un contexto ritualista los Jémez también de Nuevo México, previo a una carrera frotan el suelo con las flores secas para tener “pies ligeros” y ser más veloces (Swank, 1932; Slattery, 2020; Cook, 1930).

3. Antecedentes

Gluconacetobacter

Dentro de la familia *Acetobacteraceae*, las bacterias pertenecientes al género *Gluconacetobacter* son bacilos, gram negativos, móviles con flagelos peritricos, capaces de crecer a pH 3.5 y productoras de un pigmento café claro soluble en agua. Han sido halladas principalmente en plantas: frutas, flores, caña de azúcar y plantas de café. Algunas especies son conocidas por su capacidad de fijar nitrógeno en condiciones de microaerobiosis, desempeñando un papel importante en la agricultura. Además, algunas otras pueden producir celulosa (Reis y Texeira, 2015, Komagata *et al.*, 2014 y Yamada y Yukphan, 2008).

El metabolismo de *Gluconacetobacter* depende de la fermentación oxidativa, que se lleva a cabo, a través de reacciones sucesivas de la enzima alcohol deshidrogenasa (ADH); que se encarga de la conversión de etanol hacia aldehído; la enzima aldehído deshidrogenasa (ALDH) convierte el aldehído en ácido acético como lo hacen la mayoría de la familia *Acetobacteraceae*. También oxidan lactato y acetato gracias a deshidrogenasas de azúcares. Éstas enzimas se encuentran unidas a la membrana celular orientadas al periplasma, utilizan pirrolquinolina quinona (PQQ) como cofactor. Su función es aprovechar los electrones liberados y dirigirlos a las oxidasas terminales de la cadena respiratoria para contribuir a la formación de un gradiente protón motriz y así generar energía (Matsushita *et al.*, 2016 y Mamlock y Gullo, 2013).

Así mismo, existe un segundo juego de enzimas ADH y ALDH en el citoplasma, dependientes de NAD⁺ y NADP⁺, cuyos productos de oxidación parcial son incorporados al ciclo de los ácidos tricarbónicos por medio de la acción sucesiva de la citrato-sintasa. Esto contribuye a la tolerancia al ácido acético, un rasgo que les da ventaja competitiva ya que puede llegar hasta un 10% en contraste con otras bacterias que sólo toleran un 0.5% (Raspor y Goranovic, 2008 y Román-Camacho *et al.*, 2023).

A nivel genómico, las bacterias del género *Gluconacetobacter* presentan un contenido G+C entre 58 y 65%, con una longitud de genoma promedio entre 3 y 4 GB (millones de bases). Hasta ahora se han reportado genomas de 42 organismos, de los cuales dos corresponden a secuencias de genomas completos, siete a ensamblados en *scaffoldings* y 33 a ensamblados a nivel de *contigs* (Tabla1).

Genomas <i>Gluconacetobacter</i> NCBI	
Genomas completos	2
Ensamblados en <i>scaffoldings</i>	7
Ensamblados en <i>contigs</i>	33
Genomas en total	42

Tabla 1. Genomas encontrados en el NCBI (consulta febrero,2025).

Entre los estudios filogenéticos sobre *Gluconacetobacter* se encuentra el de Matsusani y colaboradores en 2011, dedicado a evaluar las relaciones filogenéticas entre las cepas de *Acetobacter*, *Gluconobacter* y *Gluconacetobacter*. Sin embargo, se inclinan más hacia los análisis proteómicos y no exclusivamente a las secuencias genómicas. Otros se han enfocado a la separación de sus miembros hacia otros géneros, como *Komagataeibacter* (Yamada *et al.*, 2012) y *Novacetimonas* (Brandão *et al.*, 2022).

Los estudios genómicos de *Gluconacetobacter* existentes son limitados. Bertalan y colaboradores (2009) realizaron un análisis de *core-genome*, utilizando para comparar a los genomas de las cepas más cercanas disponibles en ese momento: *Acidiphilium cryptum* JF-5 (ACC), *Gluconobacter oxydans* 621H (GOX) y *Granulibacter bethesdensis* CGDNIH (GRB). Por su parte Mukhtar y colaboradores (2021) compararon a *Gluconacetobacter azotocaptans* DS1 mediante FastANI sólo con la cepa tipo *Gluconacetobacter azotocaptans* LMG21311. Sombolestani y colaboradores (2021) compararon sus aislados de *Gluconacetobacter dulcium* LMG 1728T, LMG 1731 y LMG 22058 con todas las especies del género, pero únicamente a nivel de secuencias del gen 16S ribosomal y rasgos fenotípicos. Finalmente, Brischetto, Rossi y Fedele (2024) realizaron un análisis al microbioma inclinado hacia la diversidad microbológica presente en las uvas vinícolas italianas, limitando su comparación a las secuencias del gen 16S ribosomal.

En la familia *Acetobacteraceae* se tiene registrada una baja estabilidad del genoma, pues bacterias del género *Acetobacter* han mostrado frecuencias altas en mutaciones espontáneas (Azuma *et al.* 2009 y Beppu, 1993). En el género *Gluconacetobacter* se ha reportado que mediante control riguroso y constante de las condiciones de cultivo estas bacterias no presentan inserciones gigantes ni deleciones (Jelencko *et al.*, 2023).

Cepa UAPS01-405

Dentro del laboratorio de investigación de Ecología Molecular Microbiana anteriormente se ha realizado el aislamiento y la identificación de bacterias del género *Gluconacetobacter* (Fuentes-Ramírez, 2001). En 2019, se realizaron aislamientos bacterianos en la zona boscosa de Amozoc, Puebla, durante la época de floración de *C. alpina*, planta seleccionada por su apariencia exterior. Se recolectaron muestras de los fragmentos de la planta, que luego se transportaron al laboratorio, donde fueron procesadas. A partir de ellas se obtuvieron aislados de bacterias ácido acéticas, mediante la selección con medio LGI semigelificado. Entre estos aislamientos, la cepa UAPS01-405 llamó la atención por su fenotipo en placa (LGI). Ese aislamiento presenta bacilos redondeados y colonias de coloración amarillenta y opaca, similar a otras bacterias del género *Gluconacetobacter* (Sombolestani *et al.*, 2021).



Figura 3. Cultivo de la cepa UAPS01-405 en medio LGI gelificado.

Por su similitud fenotípica a otras especies del género *Gluconacetobacter* y debido a que no hay estudios similares en el abordaje genómico del género, es que el aislado UAPS-01-405 fue seleccionado para su secuenciación y posterior análisis bioinformático.

Estrategias bioinformáticas para el análisis de genomas bacterianos

Las técnicas de secuenciación actuales tienen la capacidad de procesar una gran cantidad de fragmentos de ADN simultáneamente en un período corto de tiempo, mediante una metodología que puede resumirse en cinco pasos:

1. Segmentación del ADN de la muestra en varios fragmentos.
2. Marcaje del ADN con adaptadores que permiten la unión del contenido genético a la plataforma de secuenciación.
3. Amplificación de los fragmentos marcados con adaptadores por métodos basados en PCR.
4. Secuenciación o lectura de los fragmentos de ADN. Después de este punto los fragmentos pasan a ser llamados “lecturas” (*reads*) para efectos prácticos en los flujos de trabajo.
5. Reconstrucción de la secuencia completa mediante el uso de programas especializados y exportación de la información para su almacenamiento (Rubio, 2020).

Previo a su reconstrucción, las lecturas que conforman el genoma deben pasar por un proceso de limpieza, conocido como “*Trimming*”, en el que se analiza la calidad de éstas, descartando las secuencias adaptadoras y posibles errores en el proceso de secuenciación. En seguida puede realizarse el ensamblaje, que consiste en tomar una colección de lecturas, productos de la secuenciación para “rearmar” el genoma mediante programas entrenados que se encargan de buscar la forma más sencilla y lógica en la que las piezas encajan. Un buen ensamble se caracteriza por contar con “pocas” piezas, también conocidas como *contigs*. Finalmente, dentro del ensamblaje se agrega un paso adicional, el pulido, que se encarga de realizar una revisión sobre el acomodo de los datos para corroborar que este ha sido realizado adecuadamente (Krueger, 2012, Kolmogorov *et al.*, 2019 y Vaser *et al.*, 2017).

El siguiente paso es la anotación, la cual permite relacionar los datos de un ensamble con datos compilados previamente de otros genomas,

proporcionándonos la identificación de los genes que conforman al genoma de interés, así como sus posibles funciones y estructuras proteicas, en caso de que estos se encarguen de codificar para alguna proteína. Esta parte es de las más importantes dentro de un análisis genómico, pues se encarga de un análisis profundo de las características del genoma estudiado (Overbeek *et al.*, 2005).

En tiempos recientes se han desarrollado tecnologías y programas computacionales que se dedican a mejorar la forma en la que se realizan los trabajos bioinformáticos, ejemplo de ello es el “Análisis completo de genomas”, en el cual, a partir de *lecturas* cruda, realiza un “meta servicio” que permite encontrar las características claves de genoma estudiado: como un árbol filogenético, la presencia o ausencia de ciertos genes de metabolismo en una descripción general de subsistemas, su posible susceptibilidad a los antimicrobianos, así como su contenido GC, y gráficos que dan una versión visual de la información obtenida. Este tipo de análisis integrales nos dan una idea de la posible variedad de características presentes en una muestra genómica y abre las puertas al conocimiento de otros aspectos que, si bien pueden estar alejados de los objetivos de algún estudio especializado, en el caso de los trabajos dedicados a la ciencia básica enriquecen el panorama de investigación (Davis *et al.*, 2020).

Los análisis evolutivos como la elaboración de árboles filogenéticos y el alineamiento de genomas nos permiten conocer las interacciones evolutivas entre las especies de un mismo género. Por un lado, el análisis filogenético nos brinda nociones del distanciamiento entre las especies. Mientras que el alineamiento de genomas, una herramienta ampliamente usada en biología de poblaciones, busca identificar la ubicación y la homología entre los nucleótidos de dos o más genomas proporcionados a un programa (Darling *et al.*, 2010, Zhu *et al.*, 2019, Wang *et al.*, 2024, Vanhove *et al.*, 2020, Karcher *et al.*, 2020 y Kuismin *et al.*, 2017).

4. Justificación

Conopholis alpina es una planta cuyo estudio es muy limitado, hasta donde se conoce no se han realizado estudios ni muestreos microbiológicos publicados previos. Por lo que esta investigación pretende explorar un aislado bacteriano hallado en esta planta, denominado como cepa UAPS01-405, uno de los microorganismos perteneciente al género *Gluconacetobacter*, conocido por su adaptación al metabolismo vegetal, así como emitir sugerencias acerca de las interacciones ecológicas y microbiológicas de este organismo.

5. Objetivos

Objetivo General

Secuenciar el genoma e identificar al aislado *Gluconacetobacter* UAPS01-405.

Objetivos Específicos

1. Realizar el ensamblaje y anotación de la secuenciación del genoma bacteriano de la cepa *Gluconacetobacter* UAPS01-405.
2. Obtener una aproximación taxonómica a la identificación del aislado UAPS01-405, mediante metodologías bioinformáticas de identificación microbiana y análisis filogenéticos.

6. Metodología

El archivo de secuenciación, obtenido en otoño de 2021 por el grupo de investigación y generado en el Laboratorio de Secuenciación Masiva y Bioinformática del IB-UNAM, mediante la técnica de Nanopore, permitió obtener lecturas largas y pocos *contigs*. Se desarrolló el siguiente flujo de trabajo para cumplir con los objetivos de esta investigación:

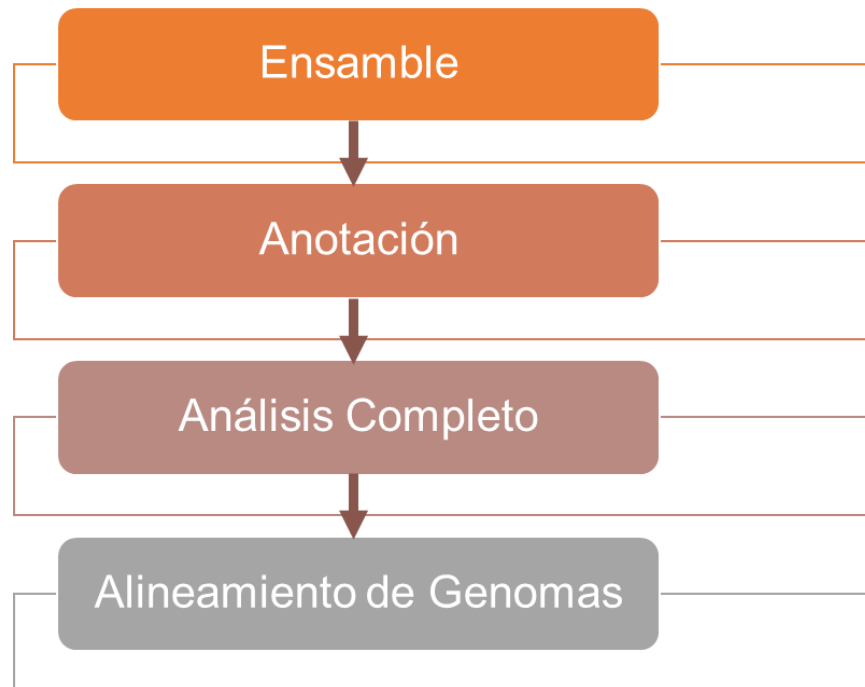


Figura 4. Flujo de trabajo general para el proyecto.

Para su ensamblaje, anotación, análisis completo y alineamiento con otros genomas, se recurrió a las herramientas que provee el Centro de Recursos Bioinformáticos Bacterianos y Virales (*Bacterial and Viral Bioinformatics Resource Center*) (BV-BRC), perteneciente a la Red de Centros de Recursos Bioinformáticos Nacionales de Estados Unidos de Norteamérica del Instituto Nacional de Alergia y Enfermedades Infecciosas (*National Institute of Allergy and Infectious Diseases Bioinformatics Resource Centers*), (Figura 3) (Olson *et al.*, 2023) <https://www.bv-brc.org/>.

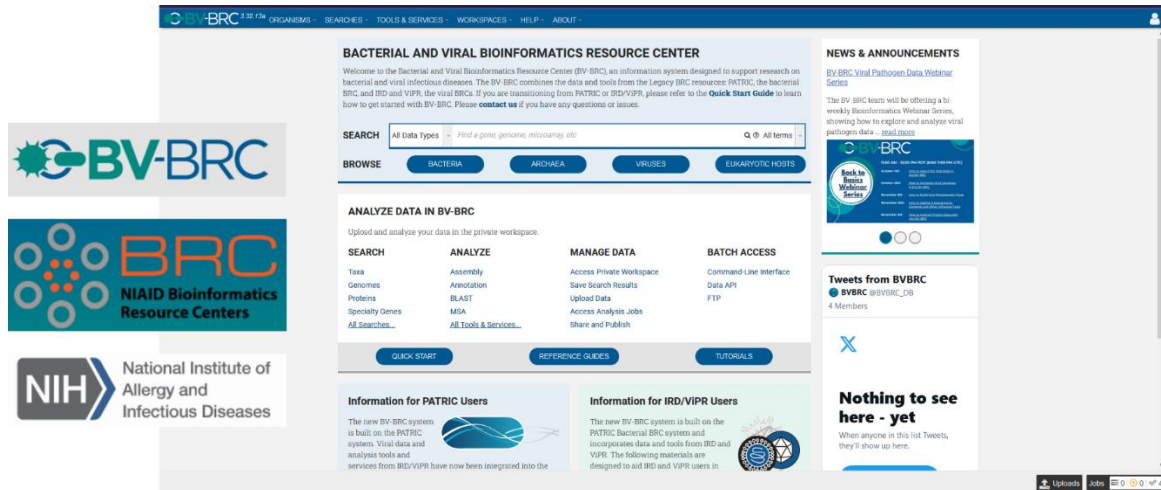


Figura 5. Interfaz del Centro de Recursos Bioinformáticos Bacterianos y Virales (*Bacterial and Viral Bioinformatics Resource Center*), (al centro), y los logos (a la izquierda) del Centro de Recursos Bioinformáticos Bacterianos y Virales (*Bacterial and Viral Bioinformatics Resource Center*), la Red de Centros de Recursos Bioinformáticos Nacionales de Estados Unidos de Norteamérica del Instituto Nacional de Salud (*National Institute of Allergy and Infectious Diseases Bioinformatics Resource Centers*) y el Instituto Nacional de Alergia y Enfermedades Infecciosas de Estados Unidos de Norteamérica (*National Institute of Allergy and Infectious Diseases*) (Olson *et al.*, 2022) <https://www.bv-brc.org/>.

Ensamblaje

La herramienta de ensamblaje que ofrece el BV-BRC permite realizar previo a este, el “*trimming*” o recorte de los fragmentos de las lecturas de baja calidad, a través de la incorporación de la herramienta Trim Galore, que se encarga de tomar las lecturas crudas proporcionadas por la herramienta de secuenciación y automatiza del recorte de calidad y adaptadores (Krueger, 2012).

Para el ensamblaje, esta herramienta cuenta con varias estrategias dependiendo del tipo de lecturas sometidas. En el caso de las lecturas sometidas en esta investigación, el ensamblador Flye fue seleccionado de manera predeterminada, pues se trata de un ensamblador específico de lecturas largas que genera *contigs* precisos (Kolmogorov *et al.*, 2019).

Posterior al ensamblaje la herramienta se apoya de programas que se encargaron de pulir el ensamblaje como Racon, un pulidor especializado para la revisión de ensamblajes de lecturas largas e híbridos (Vaser *et al.*, 2017).

Anotación

La herramienta de anotación disponible en BV-BRC, recibe como datos de ingreso archivos en formato fasta obtenido del ensamblaje realizado dentro de la herramienta de ensamblaje del mismo BV-BRC. Utiliza una versión modular y actualizada del anotador RAST tool kit, una tecnología de anotación rápida (*Rapid Annotation using Subsystem Technology*) que se apoya de distintos algoritmos incorporados a un *pipeline* general. Utiliza los programas Prodigal y Glimmer para determinar las secuencias codificantes (CDS) del genoma analizado (Brettin *et al.*, 2015, Delcher *et al.*, 2007 y Hyatt *et al.*, 2010).

Se utilizan las CDS identificadas en el genoma de interés y se comparan contra bases de datos especializadas; como la Virulence Factor Database, VIOLIN y una selección especial del BV-BRC, las que mediante BLAST de proteínas, permiten identificar los posibles factores de virulencia; mientras que por medio de BLAT (BLAST-like alignment tool), las bases de datos de resistencia a los antibióticos: CARD, NDARO, ARDB y una colección especial de proteínas de relevancia del BV-BRC son usadas para encontrar genes relacionados a la resistencia a los antibióticos; por su parte DrugBank y la Therapeutic Target Database (TTD) son usadas para encontrar dianas farmacológicas (Liu *et al.*, 2022, He *et al.*, 2014, Mao *et al.*, 2015, Altschul *et al.*, 1990, Kent, 2002, Alcock *et al.*, 2020, Davis *et al.*, 2016a, Saier *et al.* 2021, Knox *et al.*, 2024 y Zhou *et al.*, 2024).

Las proteínas familiares se asignan y luego se identifican las proteínas hipotéticas. Posterior a esto, se busca que de ser posible todas sean ubicadas a un subsistema. PubMLST se encarga de nombrar los tipos de secuencia y finalmente PhiSpy busca profagos (Davis *et al.*, 2016b, Overbeek *et al.*, 2005, Overbeek *et al.*, 2014, Jolley *et al.*, 2018 y Akhter *et al.*, 2012).

Análisis completo del genoma

La herramienta de análisis completo de genoma dentro del BV-BRC, se encarga de realizar varios análisis bioinformáticos partiendo de lecturas crudas. Realiza ensamblaje con Flye, Anotación con RAST tool kit y BLAST (*Basic Local Alignment Search Tool*) de todos los genes del genoma contra bases de datos de genes especializados, como los de resistencia a antimicrobianos, factores de virulencia, transportadores y dianas farmacológicas conocidas. Además de un análisis de subsistemas de genes de procesos biológicos específicos, en el que se hipotetiza su actividad y se identifica su parentesco, pues el programa Mash/MinHash selecciona la referencia más cercana y los genomas representativos comparando con las familias de proteínas globales BV-BRC PGFams asignadas por el programa PATtyFams, que se comparten en el grupo de los genomas. Estos genomas son seleccionados por las calidades de sus ensamblajes y anotaciones, para construir un árbol filogenético basado en los alineamientos de esas proteínas, realizados con MUSCLE y RaxML (Overbeek *et al.*, 2014, Ondov *et al.*, 2016, Edgar, 2004 y Stamatakis, 2014).

Búsqueda global en BLAST-XYplot viewer

Con la secuencia obtenida a partir del ensamblaje del genoma del aislado UAPS01-405 se realizó un BLAST, a través del visualizador BLAST-XYplot, una herramienta web que se encarga de una búsqueda masiva y sistemática de BLAST en genomas bacterianos secuenciados completos, y representa los resultados en un gráfico “x,y” para manejar simultáneamente una amplia cantidad de resultados de las comparaciones a nivel de genoma completo (Pedraza-Pérez *et al.*, 2018).

Análisis BLASTn con genes Housekeeping

En la página del NCBI se realizó un BLAST de nucleótidos a las secuencias de los genes housekeeping *dnaA* (708pb), *fabZ* (423pb), *ftsZ* (1671pb), *gyrB* (1254pb), *recA* (1053pb), *rpoD* (1938pb), *dnaK* (1554pb), *groEL* (1641pb), *rpoB* (4173pb) y la secuencia de 1498pb correspondiente al gen 16S rRNA, todas encontradas en el genoma del aislado UAPS01-405, contra las secuencias disponibles en la base de datos del NCBI, utilizando los parámetros por *default*, sin aplicar filtros adicionales (Altschul *et al.*, 1990).

Análisis FastANI

Se realizaron Análisis Rápidos de Identidad de Nucleótidos (*Fast Average Nucleotide Identity*) a través de la página Proksee, que realiza una anotación de los datos del ensamble sometidos en formato fasta. Partiendo de esta anotación, se procedió a realizar un análisis comparativo FastANI contrastando con las secuencias en formato fasta, de los genomas de las cepas pertenecientes a *Gluconacetobacter diazotrophicus*, *G. liquefaciens*, *G. dulcium* y *G. sacchari*, alojadas en el NCBI. Comparando los genomas uno a uno con el genoma del aislado UAPS01-405 (Jain *et al.*, 2018).

Elaboración de Árbol Filogenético de Pangenoma

Se realizó un árbol filogenético de las especies pertenecientes al género *Gluconacetobacter*, a través del flujo de trabajo de la figura 6, incluyendo el uso de algunos programas y herramientas dentro la plataforma bioinformática Galaxy.

Galaxy proporciona un entorno digital para tener un acercamiento hacia el análisis de datos y la programación amigable con el usuario, a través del acceso a múltiples plataformas y herramientas informáticas, *ToolShed* (Blankenberg *et al.*, 2014).

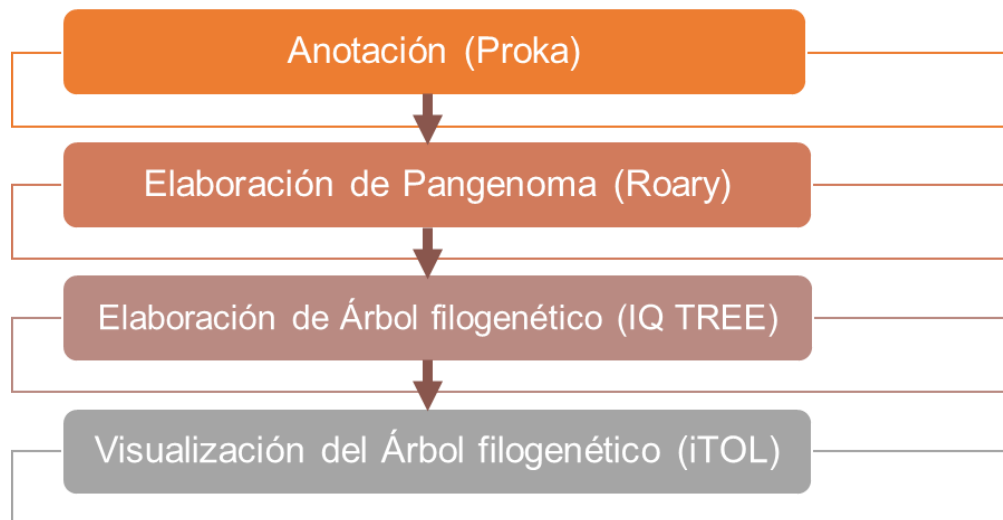


Figura 6. Flujo de trabajo para la obtención de un árbol filogenético de pangenoma.

Se anotó nuevamente del genoma de la cepa UAPS01-405 con el programa anotador Prokka ingresando la secuencia ensamblada en formato fasta, para obtener una anotación en formato gff (*General Feature Format*) (Seemann, 2014).

Dentro de Galaxy se utilizó la herramienta Roary, encargada de la elaboración de pangenomas. Un pangenoma consiste en la descripción de los genes compartidos por un grupo de genomas de interés, en él se analizan las secuencias de genomas completos donde se identifican los genes centrales y accesorios. Como datos de entrada se ingresaron las secuencias de los genomas del género *Gluconacetobacter*, registradas en el NCBI y el genoma del aislado UAPS01-405, todos en formato gff3 y gff respectivamente (Page *et al.*, 2015).

Utilizando como datos de ingreso los datos del alineamiento central o *Core Alignment* resultantes de Roary se utilizó la herramienta IQ TREE, que elabora árboles filogenéticos en Galaxy (Nguyen *et al.*, 2015).

Fuera de Galaxy, se colocó el archivo del árbol final en formato Newick, producto de la herramienta IQ TREE, como dato de ingreso en la página correspondiente a la herramienta iTOL (*Interactive Tree Of Life*), que se encarga de la visualización interactiva de árboles filogenéticos y realiza una representación gráfica del archivo Newark (Letunic y Bork, 2024).

Análisis MLSA con genes relevantes para el género *Gluconacetobacter*

Se seleccionaron las secuencias de los genes *dnaK*, *groEL* y *rpoB*, previamente descritos por Cleenwerck y colaboradores (2010), como los ideales para realizar estudios de filogenia dentro del género *Gluconacetobacter*, para generar un archivo de la concatenación de los tres genes en cinco cepas de la familia *Acetobacteraceae*: *Acetobacter aceti* LMG 1535, *Gluconacetobacter diazotrophicus* LMG 8067, *G. liquefaciens* LMG 1381T, *G. entanii* LTH 4560T y *G. sacchari* LMG 19748. En el programa Geneious Prime se sometió a las secuencias obtenidas de *Gluconacetobacter* UAPS01-405 a un alineamiento multilocus de secuencias, contra las anteriores para determinar sus posibles relaciones filogenéticas (Kearse *et al.*, 2012).

Alineamiento de genomas del BV-BRC

Cuando se trabaja con secuencias de genomas nuevos, una de las recomendaciones del BV-BRC, es que se efectúe la búsqueda de genomas similares entre los registrados dentro de la base de datos de genomas públicos del BV-BRC, mediante el uso de la herramienta *Similar Genome Finder*. Ésta utiliza Mash/MinHash, un programa que busca los genomas relacionados, tomando en cuenta criterios como cambios debidos a mutaciones y valores P de significancia, para arrojar un listado de genomas con similitudes (Ondov *et al.*, 2016).

Partiendo de la tabla de genomas similares se seleccionaron aquellos pertenecientes al género *Gluconacetobacter* y uno de *Acetobacter*, para conformar un grupo de genomas dentro de la plataforma del BV-BRC. A través de la herramienta de alineamiento de genomas que utiliza al programa Mauve, un alineador de genomas ampliamente utilizado en el campo de la biología de poblaciones, se realizó una comparación entre las secuencias de varios genomas tomando como referencia al primer genoma agregado a la herramienta. En este caso, el genoma de interés del aislado UAPS01-405 fue utilizado para efectuar la comparación contra los genomas dentro del grupo de genomas similares (Zhu *et al.*, 2019, Wang *et al.*, 2024, Vanhove *et al.*, 2020, Karcher *et al.*, 2020 y Kuismin *et al.*, 2017).

7. Resultados

Ensamble

Como producto de la herramienta de ensamblaje el BV-BRC se obtienen varios archivos de salida: archivo de los *contigs*, reporte del ensamblaje, gráfico del ensamblaje, archivo Flye stdout, reporte del QUAST, archivo de detalles.json y archivo de *contigs* de cobertura por debajo del umbral.

A continuación, abordaremos los aspectos importantes de este ensamblaje:

- **Gráfico de ensamblaje:**

La aplicación bioinformática Bandage navega en ensamblajes *de novo*, a través de gráficos, lo que mejora la visualización de los datos en comparación a cuando sólo se cuenta con el archivo de *contigs*.

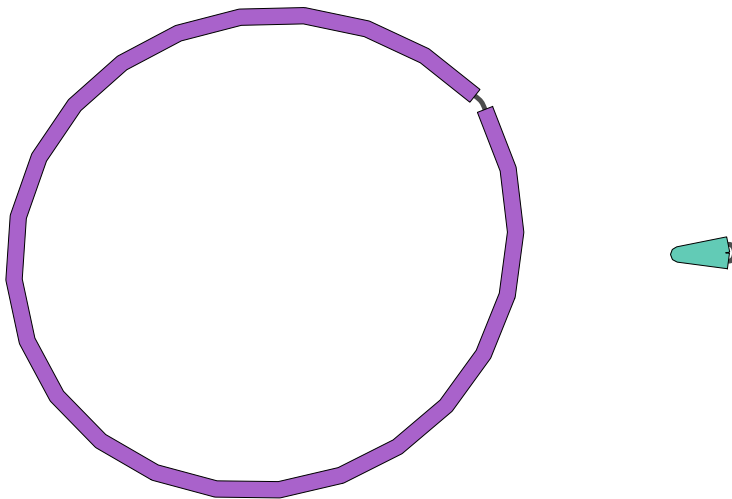


Figura 7. Gráfico de ensamblaje, en el podemos distinguir la presencia de tres contigs dentro del ensamblaje, en morado el más grande con 4,253,699 pb, en azul y verde están marcados los de 284,460 pb y 3,783 pb respectivamente.

- **Reporte de Ensamble de Genoma:**

Adicional a la imagen gráfica del ensamble, obtenemos una serie de datos relacionados a los programas utilizados y las características del ensamble:

Proceso de Ensamble	
Ensamblador	flye
Tamaño de contigs.fasta	4618675
Tiempo transcurrido	0.10 horas
Versión	2.9.1-b1780
Filtrado de <i>Contigs</i> en longitud y cobertura	
Profundidad promedio (lecturas largas)	74.51
Cobertura promedio de lecturas largas	74.511
Umbral de cobertura mínima de <i>contigs</i>	5
Umbral de longitud mínima de <i>contigs</i>	300
Número de <i>contigs</i> por encima de los umbrales	3
Número de <i>contigs</i> por debajo de los umbrales	0
Longitud total de <i>contigs</i> buenos	4541942

Tabla 2. Detalles del ensamble con flye.

Estos datos muestran una calidad de ensamblaje aceptable.

- **Reporte QUASt:**

QUASt (*Quality Assessment Tool for Genome Assemblies*) es una herramienta de evaluación de calidad y comparación de ensamblajes genómicos, muestra las estadísticas de los *contigs* generados.

Número de <i>contigs</i>	3
Número de <i>contigs</i> (≥ 0 bp)	3
Número de <i>contigs</i> (≥ 1000 bp)	3
Número de <i>contigs</i> (≥ 5000 bp)	2
Número de <i>contigs</i> (≥ 10000 bp)	2
Número de <i>contigs</i> (≥ 25000 bp)	2
Número de <i>contigs</i> (≥ 50000 bp)	2
Longitud del <i>contig</i> más largo	4253699
Logitud total	4541942
Logitud total (≥ 0 bp)	4541942
Logitud total (≥ 1000 bp)	4541942
Logitud total (≥ 5000 bp)	4538159

Logitud total (≥ 10000 bp)	4538159
Logitud total (≥ 25000 bp)	4538159
Logitud total (≥ 50000 bp)	4538159
N50	4253699
N90	4253699
auN	4001567
L50	1
L90	1
GC (%)	64.58

Tabla 3. Reporte QUAST de datos acerca de la calidad del ensamble.

De entre estas estadísticas se destacan los valores del N50 que es la longitud del o los *contigs*, tal que usando *contigs* más largos o de igual longitud se obtienen la mitad de las bases del ensamblado. Entre más elevado sea este valor se indica la presencia de *contigs* más largos, y por lo tanto una representación de calidad del genoma. El L50 es el número mínimo de *contigs* que al juntarse producen la mitad (50%) de las bases del ensamble, es decir, el número de *contigs* que en conjunto conforman la longitud de al menos el N50. Entre menor sea este valor, menos *contigs* se requieren para cubrir la mitad de la longitud del genoma, por lo tanto más eficiente se hace el análisis genómico (Gurevich *et al.*, 2013).

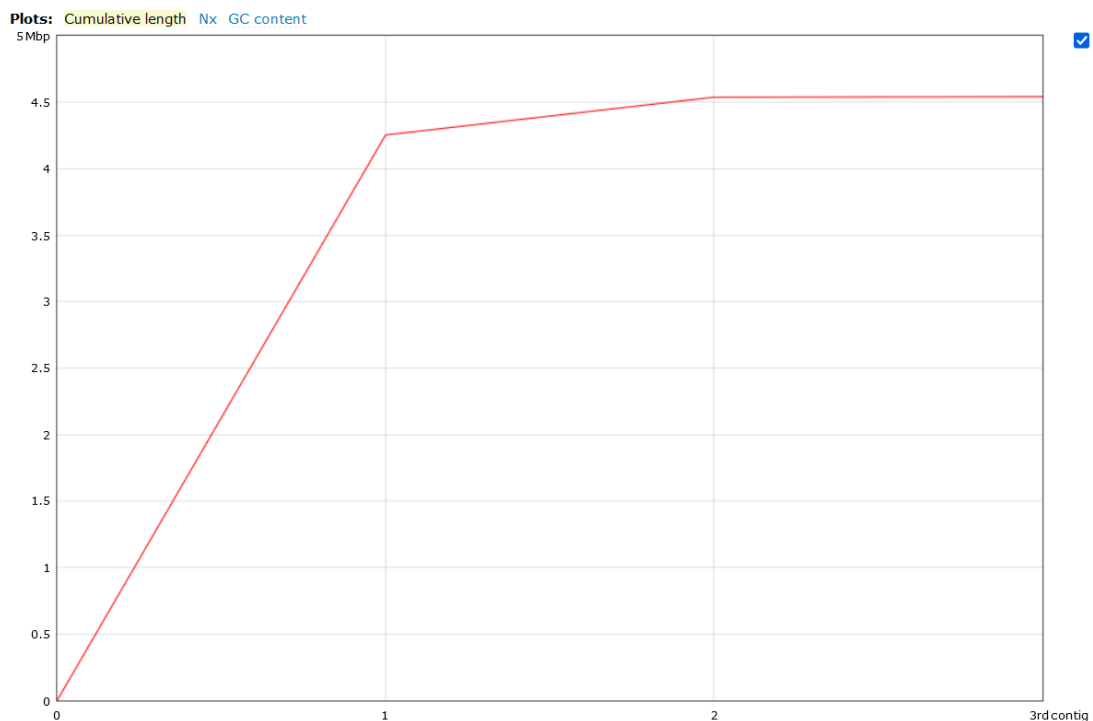


Figura 8. Longitud acumulativa por *contig*, Muestra el número de bases en los primeros *contigs* (eje x, va de 0 al número total de *contigs*) La mayoría de las bases se encuentran en el 1er *contig*.

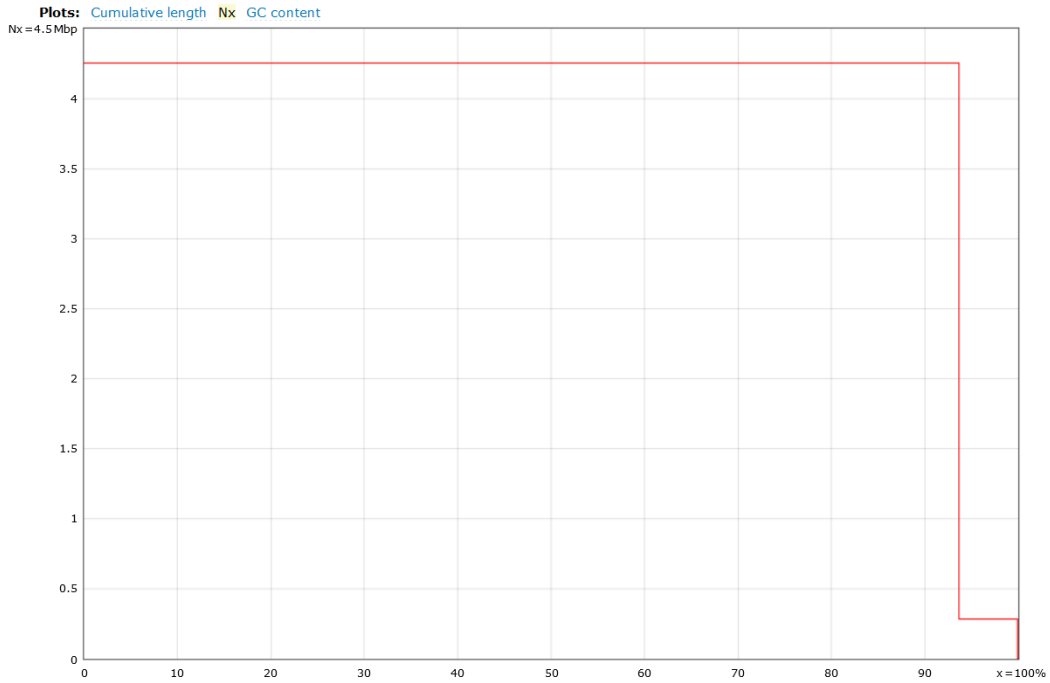


Figura 9. Porcentaje de las bases en cada uno de los *contigs*. El eje x indica el el porcentaje acumulativo del tamaño total del ensamble. El eje y representa la longitud del *contig*.

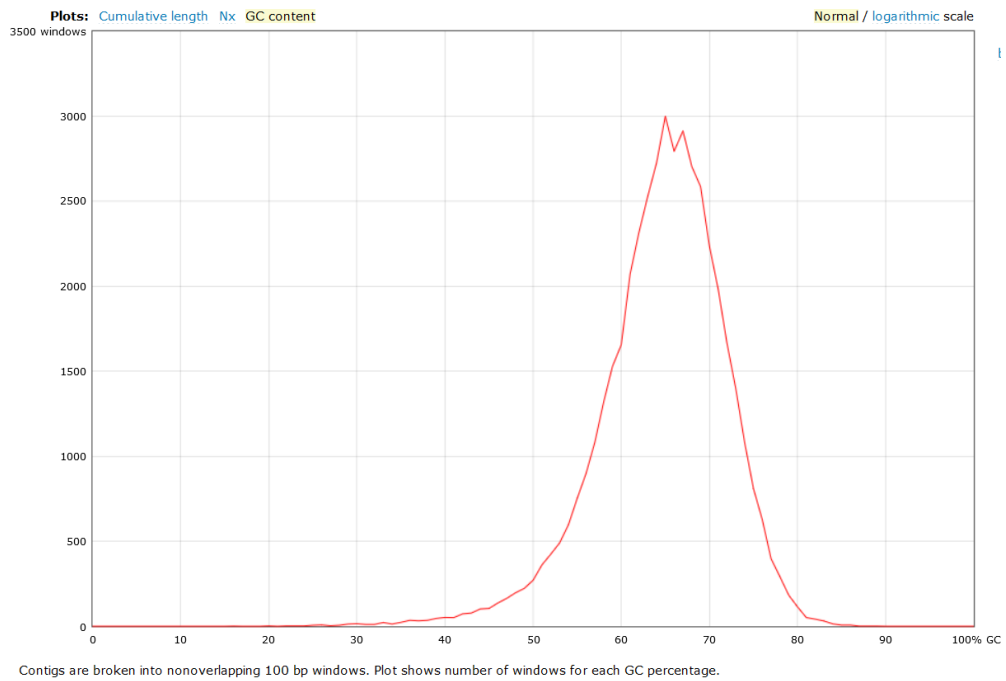


Figura 10. Número total de nucleótidos G y C, dividido entre la longitud total del ensamble. En el eje x se indica el porcentaje, en y se muestra el número de ventana de 100pb no superpuestas.

Anotación

La herramienta de anotación del BV-BRC produce varios archivos de salida: contigs.fasta,.embl, feature_dna.fasta, feature_protein.fasta, features.txt, gb, genome, gff, merged.gb, tar.gz, txt, xls, genome_quality_details.txt y el reporte de genoma

El reporte del genoma nos arroja un análisis general acerca de la anotación del genoma estudiado (Tabla 4).

Genoma ID	89583.26
Nombre del Genoma	<i>Gluconacetobacter</i> UAPS01405
Genomas de referencia	272568.11
Consistencia Gruesa (%)	97.8
Consistencia Fina (%)	86.3
Integridad (%)	100
Contaminación (%)	7.7
Grupo de evaluación	root
Cuenta de Contigs	3
Tamaño de ADN (bp)	4541942
N50 (bp)	4253699
Contigs de L50	1
Roles Sobrerepresentados	162
Roles Subrepresentados	24
Roles Predecidos	1357
Roles de completitud	24
Total de Roles Distintos	2242
Genes que codifican para proteínas con asignación funcional	3096
Genes que codifican para proteínas sin asignación funcional	2509
% de Cobertura de Características que Codifican Proteínas	123.41
% de Características Hipotéticas	44.76
% de Características que están en Familias de Proteínas Locales	49.99

Tabla 4. Estadísticas de la calidad de la anotación del genoma UAPS01-405.

Los valores como el N50 y L50 se mantienen iguales a los obtenidos a través del ensamble de las lecturas. Mientras que los porcentajes de completitud con un 100% y contaminación con un 7% indican que el ensamble ha permitido generar una anotación aceptable.

En color amarillo en la tabla, el programa señala aquellos valores de puntuación bajos que posiblemente indiquen discrepancias en la anotación y/o el análisis genómico. En este caso se resalta el porcentaje (86.3%) de Consistencia Fina, que cae por debajo del umbral aceptable de 87%, que –a diferencia de la Consistencia gruesa aceptable (97.8%)- también mide la congruencia de la anotación del ensamble respecto de los genomas de referencia utilizados para la anotación de los genes. Esto sugiere la posibilidad de que existan diferencias relevantes entre este aislado y otros pertenecientes al mismo género *Gluconacetobacter*.

Enseguida, se encuentra una lista de los posibles genes problemáticos, pues se hace la comparación del número de ocurrencia distinto al esperado de los genes que integran los genomas ya registrados dentro de la plataforma de anotación y aquellos que el programa ha logrado anotar en el genoma de interés:

Potentially Problematic Roles

Potentially problematic roles in 89583.26

Role	Predicted count	Annotated count	Feature link	Comment
16S rRNA (guanine(527)-N(7))-methyltransferase (EC 2.1.1.170)	1	2	2 features	fig 89583.26.pep.2903 is in contig Ensamble_1_contig_1 and is closest to fig 272568.11.pep.1931 , which performs this role in the reference genome (54 matching kmers) fig 89583.26.pep.2904 is only 180 bases long and is in contig Ensamble_1_contig_1
23S rRNA (adenine(2030)-N(6))-methyltransferase (EC 2.1.1.266)	1	0	0 features	fig 272568.11.pep.299 performs this role in the reference genome. fig 89583.26.pep.701 is the closest protein to the reference feature fig 272568.11.pep.299 with 30 kmers in common.
2-keto-3-deoxy-D-arabino-heptulosonate-7-phosphate synthase II (EC 2.5.1.54)	1	3	3 features	fig 89583.26.pep.433 is in contig Ensamble_1_contig_1 fig 89583.26.pep.434 is in contig Ensamble_1_contig_1 fig 89583.26.pep.435 is in contig Ensamble_1_contig_1 and is closest to fig 272568.11.pep.439 , which performs this role in the reference genome (143 matching kmers)
2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2)	1	2	2 features	fig 89583.26.pep.3220 is in contig Ensamble_1_contig_1 fig 89583.26.pep.3221 is in contig Ensamble_1_contig_1 and is closest to fig 272568.11.pep.2116 , which performs this role in the reference genome (276 matching kmers)
2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase	0	1	1 feature	fig 89583.26.pep.5012 is in contig Ensamble_1_contig_1
2-polyprenyl-6-hydroxyphenyl methylase (EC 2.1.1.222)	1	2	2 features	fig 89583.26.pep.171 is in contig Ensamble_1_contig_1 fig 89583.26.pep.172 is in contig Ensamble_1_contig_1 and is closest to fig 272568.11.pep.135 , which performs this role in the reference genome (60 matching kmers)
3-deoxy-D-manno-octulosonic acid transferase (EC 2.4.99.12)(EC 2.4.99.13)	1	2	2 features	fig 89583.26.pep.4253 is in contig Ensamble_1_contig_1 and is closest to fig 272568.11.pep.3101 , which performs this role in the reference genome (11 matching kmers) fig 89583.26.pep.4255 is in contig Ensamble_1_contig_1
3-hydroxyacyl-CoA dehydrogenase [fadN-fadA-fadE operon] (EC 1.1.1.35)	1	2	2 features	fig 89583.26.pep.1704 is in contig Ensamble_1_contig_1 fig 89583.26.pep.2239 is in contig Ensamble_1_contig_1
3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)	1	2	2 features	fig 89583.26.pep.4850 is only 141 bases long and is in contig Ensamble_1_contig_1 fig 89583.26.pep.4851 is in contig Ensamble_1_contig_1 and is closest to fig 272568.11.pep.3264 , which performs this role in the reference genome (215 matching kmers)

Figura 11. Rasgos problemáticos del genoma de interés (aislado UAPS01-405).

En esta tabla dentro del BV-BRC (Figura 11) se encuentran los enlaces que llevan a las secuencias del genoma analizado y en la columna de comentario se proporciona más información.

La sección final cuenta con una lista de cuatro columnas, siendo en la cuarta donde se encuentran los enlaces a una tabla con los enlaces a los genes problemáticos hallados por *contig*.

Nombre del <i>Contig</i>	Longitud (bp)	<i>Features</i> buenos	Enlaces
<u>Ensamble 1 contig 1</u>	4253699	660	349 <i>features</i>
<u>Ensamble 1 contig 2</u>	284460	3	<u>1 <i>feature</i></u>

Tabla 5. Ejemplo de los enlaces a los genes problemáticos encontrados por *contig*.

Análisis Completo de Genoma

Debido a que se realizan múltiples evaluaciones al genoma de interés, entre los archivos de salida más relevantes en un análisis completo de genoma se obtienen:

- **Reporte de análisis de genoma completo**

El reporte del análisis es un documento donde se mencionan detalles sobre el ensamblaje, anotación y otros análisis realizados por el BV-BRC. En las primeras líneas se nos indica que, basándose en las estadísticas de anotación y la comparación con otros genomas dentro de la base del Centro de Recursos Bioinformáticos Bacterianos y Virales, el genoma sometido es de buena calidad, lo que se muestra a continuación.

Se presenta la primera tabla en la que se arrojan estadísticas relevantes al ensamble del genoma, como que este cuenta con tres *contigs*, una longitud estimada de 4,543,682 bp, contenido G+C de 64.57%, longitud N50 de 4,252,424 bp y un L50 de un *contig*.

Detalles del Ensamble	
Contigs	3
Contenido GC	64.57
Plásmidos	0
Contig L50	1
Longitud del genoma	4543682bp
Contig N50	4252424
Cromosomas	0
ID del trabajo	assembly_1884098
Inicio del trabajo	1 de abril 2024 2:27:11pm
Trabajo completado	1 de abril 2024 3:04:48pm
Tiempo total	37 minutos y 37 segundos
Receta seleccionada	Flye

Tabla 6. Detalles del ensamble de UAPS01-405 realizados por la herramienta de Análisis Completo de Genoma del BV-BRC.

- **Anotación del Genoma**

El genoma de interés ha sido catalogado en el super reino Bacteria

Cuenta con 5997 secuencias codificantes de proteínas (CDS), 53 genes de ARN de transferencia (ARNt) y 12 genes de ARN ribosomal (ARNr). Las características de la anotación se resumen en la siguiente tabla:

Características Anotadas del Genoma	
CDS	5997
tRNA	53
Regiones repetidas	45
rRNA	12
CDS parciales	0
RNA misceláneo	0
ID del trabajo	annotation_1884098
Inicio del trabajo	1 de abril 2024 3:04:48pm
Trabajo completado	1 de abril 2024 3:09:20pm
Tiempo total	4 minutos y 32 segundos

Tabla 7. Características de la anotación del genoma UAPS01-405, realizada por la herramienta de Análisis Completo de Genoma del BV-BRC.

La anotación incluyó 2271 proteínas hipotéticas y 3726 proteínas con asignaciones funcionales, éstas incluyeron 1360 proteínas con números de la Enzyme Commission (EC), 1181 con asignaciones de Ontología Génica (OG) y 1029 de las proteínas se asignaron a las vías KEGG. La anotación realizada incluye dos tipos de familias de proteínas, de las cuales este genoma tiene 2995 proteínas pertenecientes a las proteínas específicas del género (PLFams) y 3827 proteínas de las familias de proteínas de género cruzado (PGFams).

Características de la anotación de proteínas	
Proteínas hipotéticas	2271
Proteínas con asignaciones funcionales	3726
Proteínas con asignaciones de números EC	1360
Proteínas con asignaciones GO	1181
Proteínas con asignaciones de ruta	1029
Proteínas con asignaciones de familia específica de género PATRIC (PLfam)	2995
Proteínas con asignaciones de familia de género cruzado PATRIC (PGfam)	3827

Tabla 8. Características de las proteínas anotadas de UAPS01-405, obtenida de la herramienta de Análisis Completo de Genoma del BV-BRC.

- **Visualización circular del genoma ensamblado**

Mediante el apoyo del software Circos, la herramienta de análisis completo del BV-BRC, ofrece un panorama visual del genoma de UAPS01-405 anotado. Se observan detalles, resultado de comparaciones realizadas contra varias bases de datos. Adicionalmente, en la Figura 12, se tienen los genes de proteínas anotados coloreados dependiendo del subsistema (Krzywinski *et al.*, 2009).

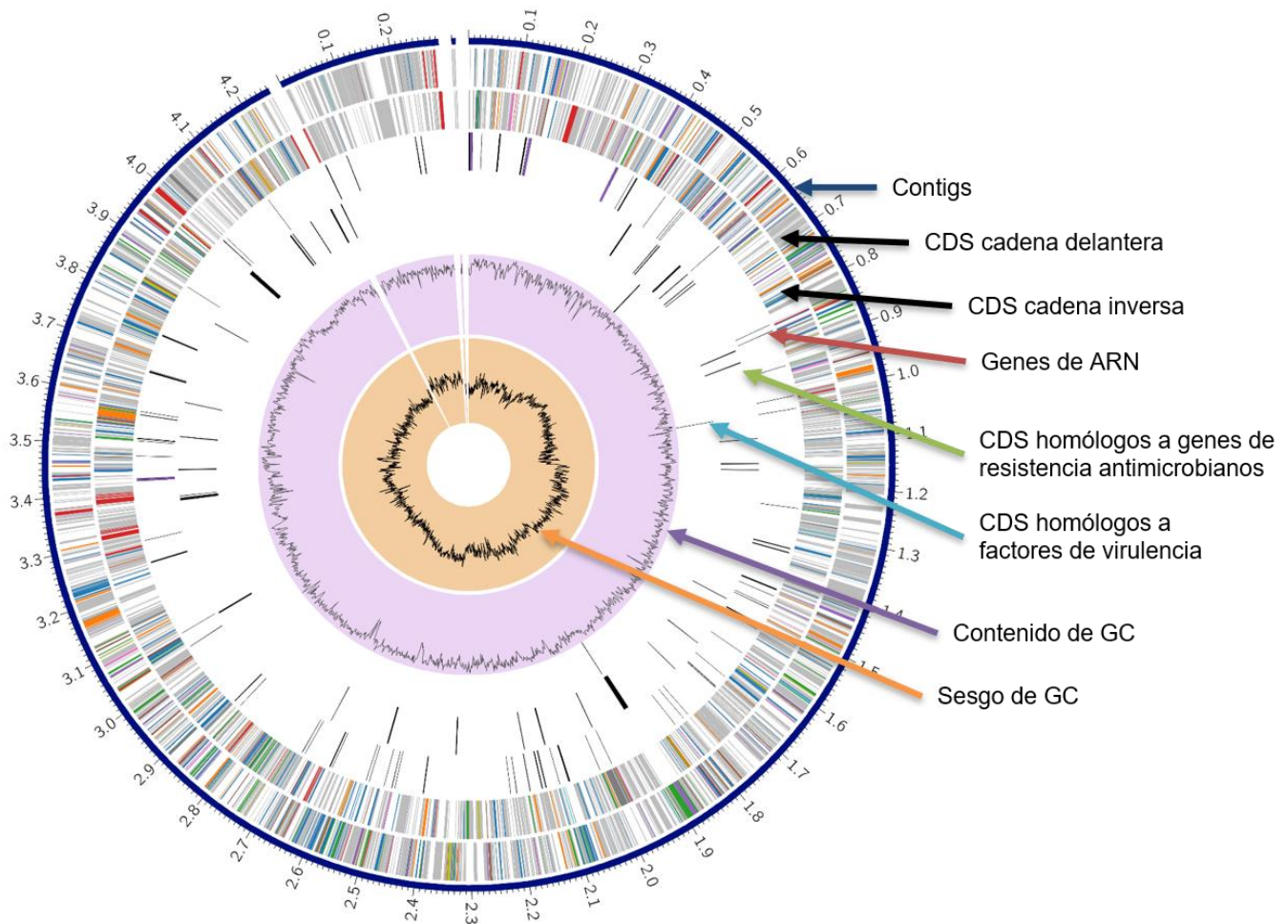


Figura 12. Visualización del genoma UAPS01-405, obtenida con la herramienta de Análisis Completo de Genoma del BV-BRC. De afuera hacia adentro se tienen los contigs, secuencias codificantes tanto en la cadena directa y la inversa, los genes ARN, aquellas secuencias codificantes con homología a genes de resistencia antimicrobiana conocidos y factores de virulencia, así como su contenido GC y la desviación de éste. Los colores observables en la gráfica indican el subsistema al que pertenecen los genes.

- **Análisis de Subsistemas**

El análisis de subsistemas se encarga de realizar una categorización funcional de las proteínas anotadas, basándose en los procesos biológicos específicos en los que están supuestamente activas. El informe resultante de este análisis incluye un gráfico circular (Figura10), que muestra las superclases de subsistemas, así como una indicación de las cantidades de subsistemas dentro de la superclase (primer número) y la cantidad de genes anotados que son parte de la superclase (segundo número).

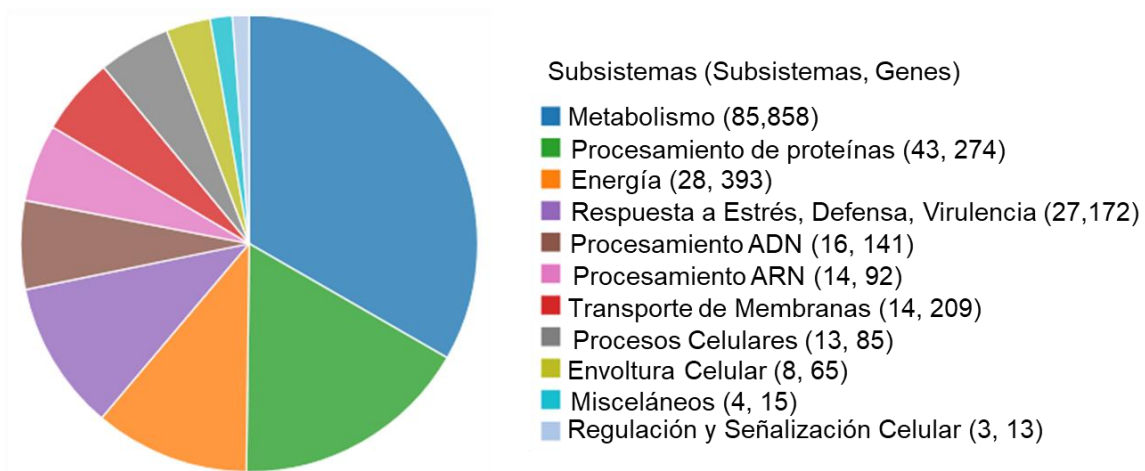


Figura 13. Contenido genético de UAPS01-405, desglosado en subsistemas funcionales, cada color indica un grupo de funciones (subsistemas) al que pertenece las proteínas anotadas.

BLAST-XYplot viewer

Para complementar la anotación como *Gluconacetobacter* y los detalles observados en el análisis filogenético del análisis completo de genoma se procedió a realizar un BLAST con el apoyo de la herramienta BLAST-XYplot viewer, utilizando un E-value de $1e-25$ como punto de corte y sin la aplicación de filtros adicionales. Como resultados se obtienen una tabla y una gráfica, esta última en su versión web otorga una visualización de los datos de la tabla en un formato dinámico, en el que apoyándose del arrastre del cursor permite analizar los datos como una misma distribución entera de miles de resultados de BLAST, o bien con el uso del zoom, en uno en particular (Pedraza-Perez *et al.*, 2018).

The job: UAPS01405
Blast Type: Genomes, Score Type: Homology, Database: Bacteria

id	Y-Position	X-Position	Query	E-value	BitScore	NewScore	GI Number	Genomic Position	Organism	Product
0	3474	300.928	Ensamble_1_contig_1	0.0	16840	1	-	3249601..3261562	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
1	3469	5796.476	Ensamble_1_contig_1	0.0	16840	1	-	3249601..3261562	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
2	3478	319.588	Ensamble_1_contig_1	0.0	16750	1	-	3502509..3490549	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
3	3471	6247.601	Ensamble_1_contig_1	0.0	16750	1	-	3502509..3490549	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
4	3474	174.349	Ensamble_1_contig_1	0.0	16240	1	-	1882721..1895235	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
5	3469	3358.304	Ensamble_1_contig_1	0.0	16240	1	-	1882721..1895235	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
6	3478	93.329	Ensamble_1_contig_1	0.0	14010	1	-	1011557..1021802	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
7	3471	1804.365	Ensamble_1_contig_1	0.0	14010	1	-	1011557..1021802	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
8	3474	98.519	Ensamble_1_contig_1	0.0	9386	1	-	1063868..1072510	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
9	3469	1897.675	Ensamble_1_contig_1	0.0	9386	1	-	1063868..1072510	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
10	3481	140.349	Ensamble_1_contig_1	0.0	8663	1	-	1222040..1211562	<i>Gluconacetobacter yilius</i> _HBR_C_3288	-
11	3478	14.468	Ensamble_1_contig_1	0.0	8536	1	-	158517..150672	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
12	3471	282.755	Ensamble_1_contig_1	0.0	8536	1	-	158517..150672	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
13	3478	93.887	Ensamble_1_contig_1	0.0	8304	1	-	1028632..1035144	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
14	3471	1834.822	Ensamble_1_contig_1	0.0	8304	1	-	1028632..1035144	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
15	3478	285.056	Ensamble_1_contig_1	0.0	7190	1	-	3123080..3130536	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
16	3471	5570.794	Ensamble_1_contig_1	0.0	7190	1	-	3123080..3130536	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
17	3474	104.064	Ensamble_1_contig_1	0.0	7176	1	-	1123748..1128748	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
18	3469	2004.486	Ensamble_1_contig_1	0.0	7176	1	-	1123748..1128748	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
19	3474	337.952	Ensamble_1_contig_1	0.0	7158	1	-	3649409..3641953	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
20	3469	6509.633	Ensamble_1_contig_1	0.0	7158	1	-	3649409..3641953	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
21	3474	105.053	Ensamble_1_contig_1	0.0	6914	1	-	1134428..1140945	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
22	3469	2023.536	Ensamble_1_contig_1	0.0	6914	1	-	1134428..1140945	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-
23	7529	103.751	Ensamble_1_contig_2	0.0	6597	1	-	5438..10868	<i>Xanthomonas_citri_citri_Aw12879</i> _Plasm:pxcaw19	-
24	3494	335.333	Ensamble_1_contig_1	0.0	6278	1	-	3355585..3361739	<i>Gluconobacter oxydans</i> _H24	-
25	3474	106.853	Ensamble_1_contig_1	0.0	6262	1	-	1153857..1159159	<i>Gluconacetobacter diazotrophicus</i> _PAL_5	-
26	3469	2058.192	Ensamble_1_contig_1	0.0	6262	1	-	1153857..1159159	<i>Gloeocapsa_PCC_7428</i> _Plasm:pgLO7428.02	-

Figura 15. Captura de pantalla de los primeros resultados del genoma de UAPS01-405 con BLASTXY-plot viewer <http://www.blast-xyplot-viewer.icuap.buap.mx/result/table/8dd73508336326b8bb5c6ba32880ae00>.

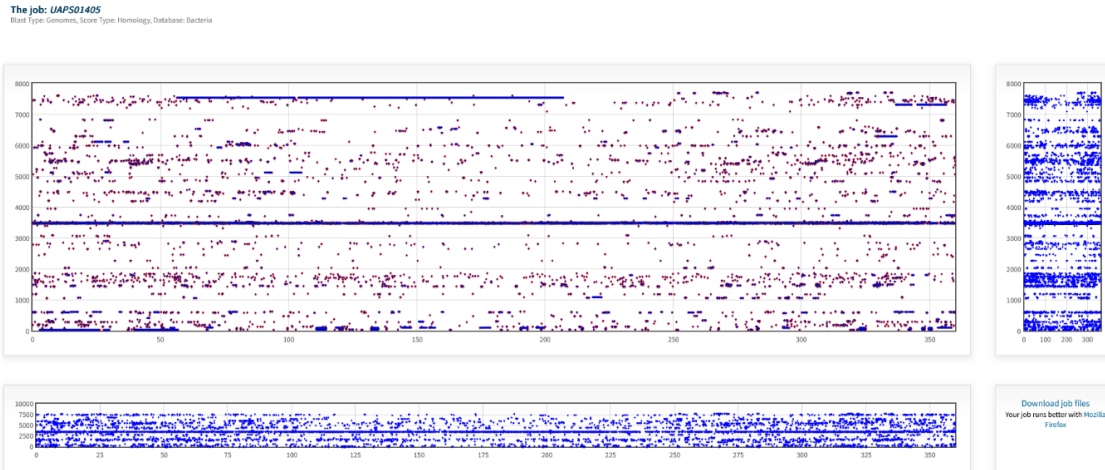


Figura 16. Gráfico “x,y” de los resultados del BLAST con el genoma de UAPS01-405. Se anexa el enlace para poder observar los resultados dentro del gráfico: <http://www.blast-xyplot-viewer.icuap.buap.mx/result/graph/8dd73508336326b8bb5c6ba32880ae00>.

Entre los resultados principales (línea azul en la figura 16) se encuentran los genomas completos correspondientes a *Gluconacetobacter diazotrophicus* PAI 5 y *Gloeocapsa* sp. PCC 7428.

BLASTn del gen16S rRNA

Para revisar los detalles observados en la anotación se procedió a también realizar un BLASTn de la secuencia del gen 16S rRNA. Se resumen los primeros resultados filtrados por porcentaje de identidad, en la Tabla 11.

BLASTn 16S rDNA							
Descripción	Score máximo	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
<i>Gluconacetobacter sacchari</i> isolate IF2-6	2625	2625	95%	0	99.72%	1433	AF127412.1
<i>Gluconacetobacter sacchari</i> isolate SRI1853	2571	2571	93%	0	99.72%	1404	AF127408.1
<i>Gluconacetobacter sacchari</i> isolate IF9701	2612	2612	95%	0	99.65%	1428	AF127411.1
<i>Gluconacetobacter takamatsuzukensis</i> strain T61213-20-1a	2588	2588	94%	0	99.65%	1416	NR_114384.1
<i>Gluconacetobacter sacchari</i> isolate IF9645	2584	2584	94%	0	99.65%	1413	AF127413.1
<i>Gluconacetobacter dulcium</i> strain LMG 22058	2693	2693	98%	0	99.53%	1479	MT379650.1
<i>Gluconacetobacter liquefaciens</i> strain KACC 22064	2639	2639	96%	0	99.52%	1450	ON746603.1
Uncultured bacterium clone MS-248	2627	2627	96%	0	99.51%	1443	GQ477847.1
<i>Gluconacetobacter liquefaciens</i> cepa JCM 17689	2601	2601	95%	0	99.51%	1429	AB626659.1
<i>Gluconacetobacter liquefaciens</i> strain: VTH-Ai57	2584	2584	94%	0	99.51%	1422	LC103264.1

Tabla 11. Resultados del BLASTn 16S rRNA, obtenido del aislado UAPS01-405, se colocan en las columnas los valores de Score máximo, Score Total, Cobertura de Query, E-value, Porcentaje de Identidad, Longitud del fragmento encontrado y Número de acceso.

Se aprecia un porcentaje de similitud de un 99.72% con dos aislados de *Gluconacetobacter sacchari*. A diferencia de lo visto en las figuras 14 y 15 que muestra al aislado UAPS01-405, más cercano a *Ga. diazotrophicus* PAI 5, pues en este análisis se consideran sólo las secuencias del gen 16S rRNA, mientras que en los análisis anteriores se compara únicamente con genomas completos y de calidad.

BLASTn con genes Housekeeping

A continuación, se presenta una tabla en la que se condensan los resultados de los BLAST de nucleótidos, realizados con distintos genes housekeeping, filtrados por el porcentaje de identidad. Tenemos al extremo izquierdo el gen utilizado para realizar el BLAST y en el campo de organismo se colocan los primeros resultados correspondientes a cada uno de los análisis realizados.

Tabla de Resumen BLASTn con genes housekeeping								
Gen	Organismo	Score máx	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
<i>16S rDNA</i>	<i>Gluconacetobacter sacchari</i> aislado IF2-6	2625	2625	95%	0	99.72%	1433	AF127412.1
<i>dnaA</i>	<i>Gluconacetobacter diazotrophicus</i> PAI 5	998%	998	100%	0	92.09 %	3944163	AM889285.1
<i>fabZ</i>	<i>Komagataeibacter xilinus</i> CGMCC 17276	359	359	98%	2.00E-94	82.21%	3527401	CP041348.1
<i>ftsZ</i>	<i>Gluconacetobacter diazotrophicus</i> PAI 5	1482	2189	95%	0	85.37%	3887492	CP001189.1
<i>gyrB</i>	<i>Gluconacetobacter diazotrophicus</i> PAI 5	1094	1094	83%	0	85.67%	3944163	AM889285.1
<i>recA</i>	<i>Gluconacetobacter diazotrophicus</i> PAI 5	1500	1500	99%	0	92.62%	3944163	AM889285.1
<i>rpoD</i>	<i>Gluconacetobacter diazotrophicus</i> PAI 5	2562	2562	100%	0	90.49%	3944163	AM889285.1
<i>dnaK</i>	<i>Gluconacetobacter diazotrophicus</i> PAI 5	2283	2283	99%	0	93.23%	3944163	AM889285.1
<i>groEL</i>	<i>Gluconacetobacter diazotrophicus</i> PAI 5	2481	3621	99%	0	93.97%	3944163	AM889285.1
<i>rpoB</i>	<i>Gluconacetobacter diazotrophicus</i> PAI 5	5860	5860	99%	0	92.07%	3944163	AM889285.1

Tabla 12. BLASTn realizados con distintos de genes housekeeping obtenidos del aislado UAPS01-405, se colocan en las columnas los valores de Score máximo, Score Total, Cobertura de Query, E-value, Porcentaje de Identidad, Longitud del fragmento encontrado, Número de acceso.

Como se puede observar principalmente se encontraron similitudes con los genes de *Gluconacetobacter diazotrophicus* PAI 5.

FastANI

Posteriormente a los resultados observados con los BLASTn se obtuvo Identidad Promedio de Nucleótidos (ANIs) con los genomas reportados de las cepas correspondientes a las especies *G. diazotrophicus*, *G. liquefaciens*, *G. dulcium* y *G. sacchari*. De estos análisis se obtuvieron varios gráficos, los que para propósitos de la distribución de la información serán resumidos en la tabla 13.

Identidad promedio de nucleótidos (ANI) de todo el genoma.			
Cepa	ANI	Query sequence fragments	Orthologous Matches
<i>Gluconacetobacter liquefaciens</i> DSM 5603	91.5939	1377	1146
<i>Gluconacetobacter liquefaciens</i> NRIC 0522	91.2933	617	495
<i>Gluconacetobacter liquefaciens</i> NRBC 12388	91.6286	1350	1120
<i>Gluconacetobacter liquefaciens</i> LGM 1382	91.6685	1379	1126
<i>Gluconacetobacter liquefaciens</i> ERR 4757829	91.0416	1503	936
<i>Gluconacetobacter diazotrophicus</i> PR4	83.4036	1124	764
<i>Gluconacetobacter diazotrophicus</i> PAI5 ATCC	83.3326	1304	774
<i>Gluconacetobacter diazotrophicus</i> PAI5	83.2063	1331	812
<i>Gluconacetobacter diazotrophicus</i> LGM 7603	83.2914	1313	772
<i>Gluconacetobacter diazotrophicus</i> C10	Valor menor al 80 no reportado		
<i>Gluconacetobacter dulcium</i> LGM 1728	91.4688	1438	1079
<i>Gluconacetobacter dulcium</i> LGM 1731	91.5256	1437	1079
<i>Gluconacetobacter dulcium</i> LGM 22058	91.5355	1530	1120
<i>Gluconacetobacter sacchari</i> LMG 1947	87.1003	1560	997
<i>Gluconacetobacter sacchri</i> DSM 12717	87.0256	1421	906

Tabla 13. Identidad Promedio de Nucleótidos (ANIs) realizados contra distintos de los genomas registrados en la base de datos del NCBI, pertenecientes al género *Gluconacetobacter*, y que durante la presente investigación han mostrado similitudes con el genoma de UAPS01-405.

En el caso de *Gluconacetobacter takamatsuzukensis* al sólo contarse con secuencias parciales de su genoma reportadas en *scaffoldings*, el programa no permitió realizar el FastANI, por conflictos con el tipo de archivos de entrada para el programa.

Análisis MLSA con genes relevantes para el género *Gluconacetobacter*

Con el archivo concatenado de los genes housekeeping de cinco cepas pertenecientes a la familia *Acetobacteraceae* que fueron seleccionadas por su similitud observada en los análisis previos realizados en el presente trabajo de investigación. A través del uso del programa Geneious, se obtuvo un árbol filogenético (Figura 17). El aislado UAPS01-405 se ubica entre los clados de *G. liquefaciens* y *G. sacchari*.

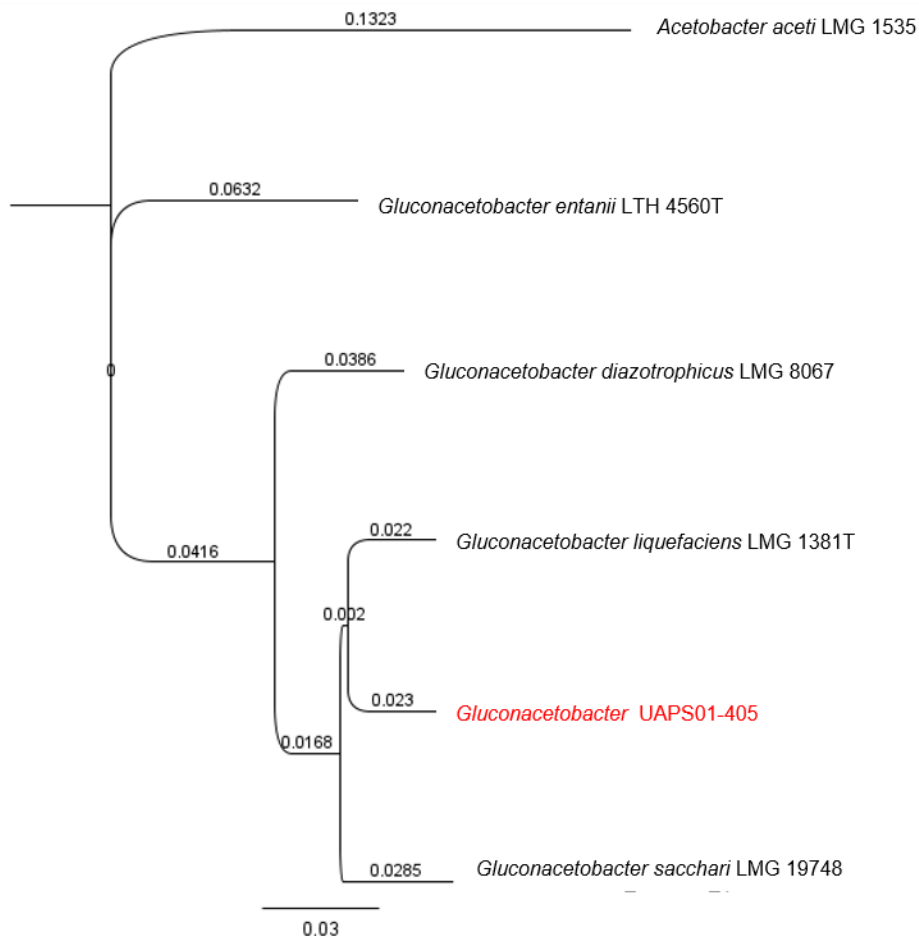


Figura 17. Árbol filogenético obtenido a partir de la realización del MLSA de los genes *dnaK*, *groEL* y *rpoB* con el programa Geneious Prime (Kearse *et al.*, 2012). En la penúltima ubicación se encuentra el aislado UAPS01-405.

Árbol Filogenético de Pangenoma

A partir del Pangenoma de las bacterias registradas del género *Gluconacetobacter*, en el que además de los genomas completos, se incluyeron aquellos genomas reportados parcialmente (en *scaffoldings*), se elaboró el árbol filogenético de la Figura 18. En él podemos apreciar que el genoma del aislado UAPS01-405, se encuentra señalado (amarillo) en un clado separado y cercano entre los de las especies *Gluconacetobacter liquefaciens* y *Gluconacetobacter dulcium*.

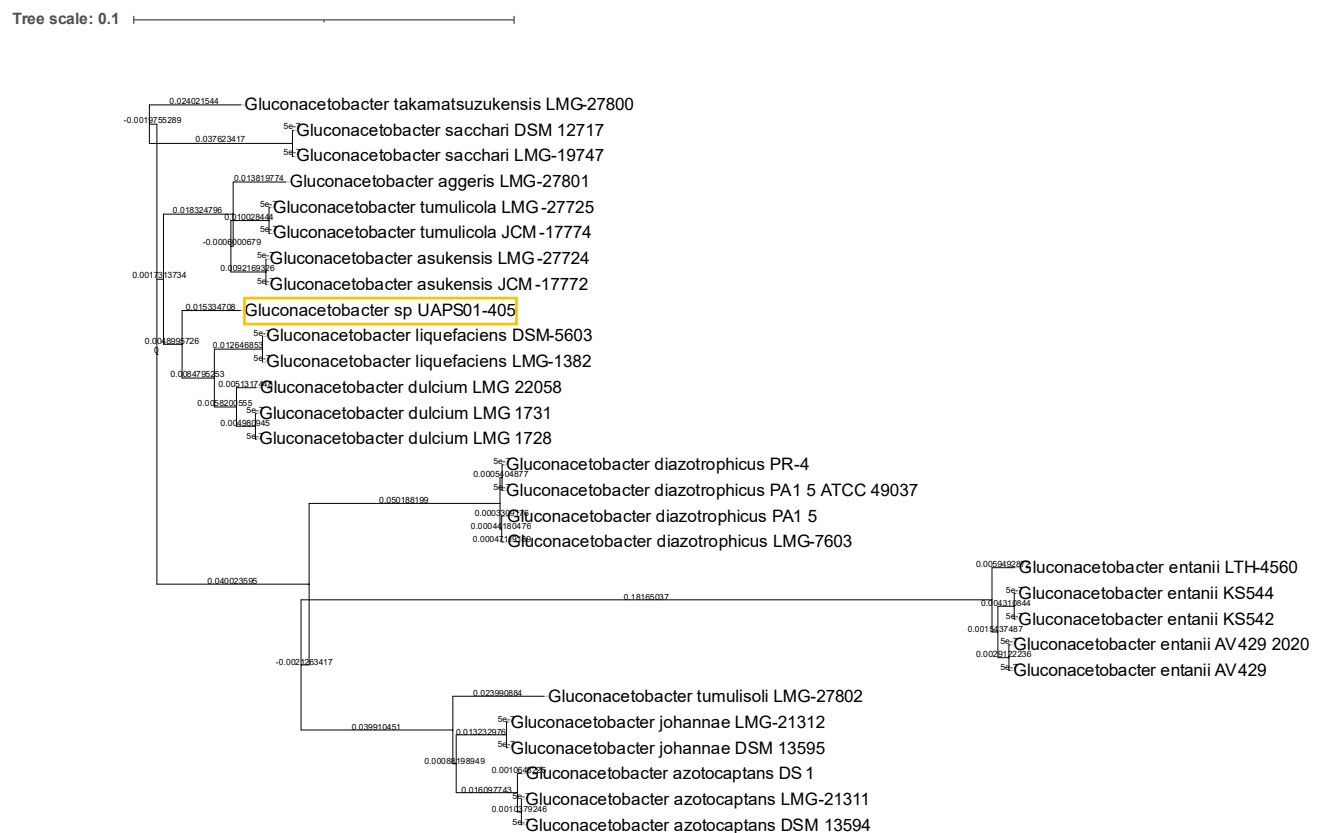


Figura 18. Árbol filogenético realizado en iTOL, a partir del Pangenoma obtenido con Roary en Galaxy. Elaborado con las secuencias de los genomas de *Gluconacetobacter* registradas en el NCBI, en amarillo se encuentra el señalado el genoma de interés UAPS01-405.

Alineamiento de Genomas

Como resultado de la Búsqueda de genomas similares se obtuvieron alrededor de 51 genomas similares, de los cuales al inicio y por mayor similitud se encontraron 12 del género *Gluconacetobacter*; (tabla en anexos). Estos se seleccionaron para establecer un grupo de genomas (tabla 14), se adicionó al grupo el genoma de *Acetobacter aceti* ATCC 23746 como elemento externo dentro de la familia *Acetobacteraceae*.

ID	Nombre del genoma	Estatus del genoma	Contigs	Longitud del genoma	Contenido GC	Distancia	Valor P	Counts
89584.5	<i>Gluconacetobacter liquefaciens</i> strain DSM 5603	WGS	31	4176944	64.38373	0.07371	0	119/1000
2729096.3	<i>Gluconacetobacter</i> sp. 1a LMG 1728	WGS	76	4415112	64.65086	0.08032	0	102/1000
1017177.4	<i>Gluconacetobacter tumulicola</i> strain LMG 27725	WGS	90	4296956	65.13765	0.09749	0	69/1000
1017181.4	<i>Gluconacetobacter asukensis</i> strain LMG 27724	WGS	64	4387527	65.16641	0.0988	5.19725e-317	67/1000
1286186.4	<i>Gluconacetobacter aggeris</i> strain LMG 27801	WGS	61	4324546	65.209366	0.0988	3.16973e-317	67/1000
1286190.4	<i>Gluconacetobacter takamatsuzukensis</i> strain LMG 27800	WGS	35	3783863	67.03943	0.10766	1.48E-257	55/1000
92759.5	<i>Gluconacetobacter sacchari</i> strain LMG 19747	WGS	114	4832047	65.955154	0.12572	2.74E-165	37/1000
272568.11	<i>Gluconacetobacter diazotrophicus</i> PAI 5 (Prj:21071)	Complete	2	3914947	66.3	0.15915	2.28E-76	18/1000
112140.5	<i>Gluconacetobacter johannae</i> strain LMG 21312	WGS	49	3632583	67.31461	0.16182	4.35E-72	17/1000
142834.5	<i>Gluconacetobacter azotocaptans</i> strain LMG 21311	WGS	57	4135095	66.518135	0.17093	5.12E-58	14/1000
1286189.4	<i>Gluconacetobacter tumulisoli</i> strain LMG 27802	WGS	49	3962674	66.63465	0.18676	1.75E-40	10/1000
663932.3	<i>Acetobacter aceti</i> ATCC 23746	WGS	8	3690268	57	0.23011	1.90E-15	4/1000

Tabla 14. Grupo de genomas pertenecientes al género *Gluconacetobacter* y que durante la presente investigación mostraron similitudes con el genoma de UAPS01-405.

Como resultado de la herramienta de alineamiento se obtuvo la visualización en una página web dentro del servicio del BV-BRC. En ella se puede interactuar con la alineación de los genomas y observar en listones las ubicaciones de los genes similares entre genomas, se muestran en colores las secuencias de los genes de cada genoma, dependiendo del subsistema al que pertenecen, en los casos en los que los genes están en distintas hebras se distinguen las ubicaciones en escalones. La selección gen por gen refleja su ubicación en cada uno de los genomas alineados. Otras funcionalidades, como el zoom, se aprecian mejor en la versión web del alineamiento. A continuación, se muestra como resultado en este escrito, la visualización de descarga del alineamiento, en acercamiento. Para fines ilustrativos sólo se observa el listado de dos genomas alineados (Figura 19) y el alineamiento completo en listones (Figura suplementaria 1), ambos en vista general sin zoom.

Tanto el enfoque que brinda el árbol filogenético obtenido con el pangenoma, como el alineamiento de genomas nos brindan un panorama general que para propósitos de esta investigación permite ubicar al aislado UAPS01-405 entre el resto de las especies registradas como *Gluconacetobacter*

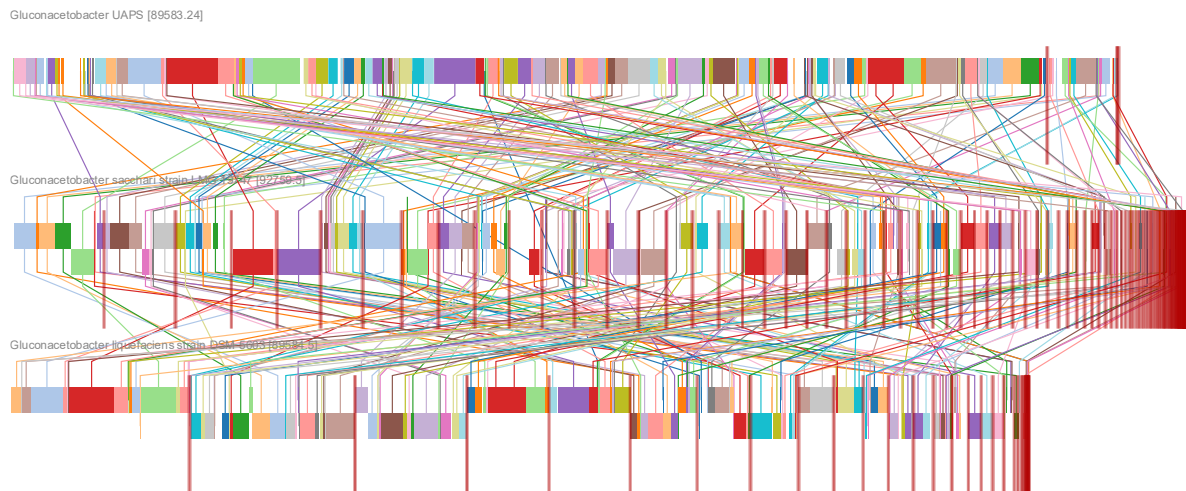


Figura 19. Imagen del alineamiento de genomas. Se presenta el listado de genomas comparados con genes coloreados de acuerdo a los subsistemas. Los listones indican la presencia del mismo gen en cada genoma, las líneas rojas representan la división del contenido genético en contigs.

8. Discusión

Son escasos los estudios relacionados a la biología de la planta *Conopholis alpina*, el presente estudio realiza una aproximación a sus interacciones biológicas a nivel microbiológico enfocándose en uno de los aislados bacterianos obtenido durante un muestreo de diversidad microbiana: *Gluconacetobacter* UAPS01-405 (Rodríguez *et al.*, 2013, Rodríguez *et al.*, 2011 y Calderón de Rzedowski *et al.*, 1998).

El ensamblaje obtenido en el presente trabajo de investigación consta de un tamaño total de 4,541,942 pb y contenido G+C de 64.58%, similares a los registrados en otras especies del género *Gluconacetobacter*. Así mismo en la anotación se cuenta con valor alto de consistencia gruesa de 97.8%, lo que nos indica una alta calidad tanto del trabajo de secuenciación y ensamblaje, como una adecuada anotación dentro del género *Gluconacetobacter*. Sin embargo, el 86.3%, de consistencia fina, por debajo del umbral aceptable de 87%, señala posibles discrepancias al compararse el ensamblaje con otros genomas completos de calidad registrados en la base de datos del BV-BRC, que se encuentra conectada al NCBI (Sombolostani *et al.*, 2021).

A continuación, se presenta una comparación entre los valores estadísticos claves para medir la calidad de un ensamblaje y poder considerarlo apto para usos taxonómicos, establecidos por Riesco y Trujillo (2024), como continuación del trabajo de Chun y colaboradores (2018), en contraste con los obtenidos como resultado del ensamblaje de la secuenciación del aislado UAPS01-405.

Característica	Riesco y Trujillo, 2024	Aislado UAPS01-405
Número de <i>Contigs</i>	Menor a 1000 <i>contigs</i>	3 <i>contigs</i>
N50	Mayor a 5000 bp	4,253,699 pb
L50	Menor a 1000 <i>contigs</i>	1 <i>contig</i>
Profundidad de la secuenciación	Mayor a 50 veces el genoma	74.51 veces el genoma
Estimación de la integridad	Mayor al 50%	100%
Estimación de la contaminación	Menor al 10%	7.7%

Tabla 15. Estadísticas de un ensamble apto para uso en taxonomía, creada con datos establecidos por Riesco y Trujillo (2024), en contraste con los obtenidos en la presente investigación.

Se decidió realizar varios análisis para confirmar la ubicación a nivel de género y posiblemente a nivel de especie del aislado UAPS01-405. El primero de los cuales es un BLASTXY, que al comparar genomas completos y de calidad obtenidos de la base de datos del NCBI, mostró una relación alta entre los genomas de *Gluconacetobacter diazotrophicus* y el genoma de interés. A diferencia del resto de los análisis realizados en el presente trabajo de investigación, en el BLASTXY se compara contra una base de datos exclusivamente de genomas completos, debido a que se cuentan con pocos registros de otras especies de *Gluconacetobacter* y a que en su mayoría, los genomas reportados corresponden a ensamblajes parciales o de baja calidad. Es por ello que tampoco se ven incluidos en el segmento de análisis filogenético dentro del Análisis de genoma completo. El análisis del gen 16S ribosomal mostró mayores similitudes con los genes registrados de *Gluconacetobacter sacchari*. En contraste con los otros BLAST realizados contra varios genes housekeeping (*dnaA*, *fabZ*, *ftsZ*, *gyrB*, *recA*, *rpoD*, *dnaK*, *groEL*, *rpoB*, ver Tabla 11), en los que se obtuvieron mayores similitudes *G. diazotrophicus* y *G. liquefaciens*.

Debido a lo anterior se procedió a realizar un análisis de identidad de nucleótidos (FastANI) entre las cepas registradas dentro del NCBI, que exhibieron similitudes en los BLAST realizados previamente. Sin embargo, como se ha podido observar en la tabla 13, contamos con índices de identidad bajos, menores al 95-

97%, que es el umbral recomendado para poder ubicar un aislado nuevo dentro de alguna de las especies evaluadas (Chun *et al.*, 2018, Riesco y Trujillo, 2024).

Para continuar con la identificación se realizó un análisis multilocus de secuencias (MLSA), estándar recomendado para el análisis de genomas nuevos. Éste fue realizado con los genes *groEL*, *rpoB* y *dnaK*, recomendados por Cleenwerck y colaboradores (2010) para la identificación de nuevas especies dentro del género. En la presente investigación, encontraron similitudes entre los clados de *G. liquefaciens* y *G. sacchari*.

En contraste, en el Pangenoma realizado, que incluyen todos los registros pertenecientes al NCBI de los genomas ensamblados, o ensamblajes parciales, anotados bajo el género *Gluconacetobacter*, se puede apreciar en la figura 18 que el aislado UAPS01-405 se ubica en el mismo grupo que los clados de *Gluconacetobacter liquefaciens* y *Gluconacetobacter dulcium*.

En el caso de *Gluconacetobacter dulcium*, identificado por el grupo de Sombolostani y colaboradores (2021), es necesario señalar que ésta especie carece del gen *groEL*, pues su identificación se realizó por medio de pruebas bioquímicas, MALDI-TOF, Ortho ANI y dDDH (un método de hibridación DNA-DNA in silico) comparándolo contra los otros genomas de *Gluconacetobacter* resecuenciados en esa investigación. Por ello es que no fue incluido dentro del grupo de clados para el MLSA.

Por su parte, el alineamiento de genomas otorga una visualización sobre las similitudes del ensamblaje del aislado UAPS01-405, en relación al resto de los miembros contrastados de *Gluconacetobacter*. Como se aprecia en la figura 19, algunos de los genomas ensamblados se encuentran fragmentados entre muchos contigs, cuyas longitudes varían. Sin embargo, a nivel de similitudes se puede ver cómo gran parte del genoma ensamblado en la presente investigación se encuentra seccionado en colores y muchos de sus genes se encuentran conservados dentro del resto de especies del género.

Finalmente, en el marco del establecimiento de la identificación del genoma del aislado UAPS01-405, se ha podido concluir que, si bien éste se encuentra dentro del género *Gluconacetobacter*, aún queda pendiente su identificación a nivel de especie. Entre la literatura revisada, se encuentran sugerencias sobre análisis adicionales para poder considerarlo dentro de una especie nueva, pues como hemos podido apreciar a través de los distintos análisis realizados en esta investigación. La gran mayoría de los genomas registrados muestran sólo acercamientos parciales, esto puede deberse a que varios de los genomas registrados se encuentran solamente en secuencias parciales o ensamblajes que no han logrado alcanzar los estándares de calidad, por lo que el presente ensamblaje al mostrar información completa queda fuera de los clados establecidos.

Chun y colaboradores (2018) sugieren que para nombrar una nueva especie en la que se logren colocar genomas nuevos, es necesario secuenciar nuevamente el gen 16S ribosomal, mediante la metodología Sanger y con éste repetir el BLAST, tal que puedan descartarse posibles contaminantes dentro de la secuencia, así como hacer un alineamiento contra el gen encontrado en el genoma ensamblado. Sin embargo, en una actualización presentada en 2024, Riesco y Trujillo comentan la necesidad de realizar análisis con secuencias de aminoácidos adicionales a los que emplean sólo nucleótidos, para permitir el establecimiento de una nueva especie, pues consideran que no basta con los resultados obtenidos mediante ANI o dDDH, porque estos pese a su importancia, varían entre estudios, por ello pueden considerarse sólo como guías hacia “un acercamiento más balanceado que mezcle múltiples análisis y métricas”.

9. Conclusiones

- A través de los resultados obtenidos con el apoyo en las herramientas del BV-BRC, se logró construir un ensamblado adecuado, con una anotación aceptable dentro del género *Gluconacetobacter*.
- Se cuentan con pocas secuencias de genomas de *Gluconacetobacter* anotados y publicados dentro del NCBI, cuya secuenciación sea completa y de calidad, lo que complicó la identificación filogenética del aislado UAPS01-405 a nivel de especie.
- Es posible que las inconsistencias encontradas durante los análisis bioinformáticos: la anotación, análisis completo de genoma (en especial en el segmento de árbol filogenético), BLASTs y FastANIs, se deban a que, como puede observarse en el árbol filogenético obtenido a partir de la elaboración del pangenoma, el aislado UAPS01-405 pertenezca a una especie que aún no ha sido descrita.
- El genoma presentado en este trabajo representa un aporte de calidad a las bases de datos de genómicas de *Gluconacetobacter*.

10. Perspectivas

Los resultados de esta investigación, indican la necesidad de continuar con la línea de trabajo para el discernimiento taxonómico del aislado UAPS01-405 y una mejor comprensión de la diversidad del género *Gluconacetobacter*.

Así mismo se incentiva a la profundización de las características propias del aislado UAPS01-405, ya que al encontrarse dentro del género *Gluconacetobacter* podría involucrarse en tareas de interés comercial agrícola e industrial (Mukhtar *et al.*, 2021).

Adicional a lo anterior sugiere realizar un análisis a profundidad acerca de los genes de resistencia a antibióticos que podría contener *Gluconacetobacter* UAPS01-405.

11. Bibliografía

Akhter, S., Aziz, R. K., y Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity-and composition-based strategies. *Nucleic Acids Research*, 40(16), e126-e126.

Alcock, B. P., Raphenya, A. R., Lau, T. T., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen A. V., Cheng, A. A., Liu, S., Min S. Y., Miroshnichenko, A., Tran, H., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A. N., Bordeleau, E., Pawlowski, A. C., Zubyk, F. L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G. L., Beiko, R. G., Brinkham, F. S. L., Hsiao, W. W. L., Doonselaar, G. V. y McArthur, A. G. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 48(D1), D517|-D525.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., y Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410.

Azuma, Y., Hosoyama, A., Matsutani, M., Furuya, N., Horikawa, H., Harada, T., Hirakawa, H., Kuhara, S., Matsushita K., Fujita, N y Shirai, M. (2009). Whole-genome analyses reveal genetic instability of *Acetobacter pasteurianus*. *Nucleic Acids Research*, 37(17), 5768-5783.

Beppu, T. (1993). Genetic organization of *Acetobacter* for acetic acid fermentation. *Antonie Van Leeuwenhoek*, 64, 121-135.

Bertalan, M., Albano, R., de Pádua, V., Rouws, L., Rojas, C., Hemerly, A., Teixeira, K., Schwab, S., Araujo, J., Oliveira, A., França, L., Magalhães, V., Alquéres, S., Cardoso, A., Almeida, W., Martins M. L., Nogueira, E., Cidade, D., Oliveira, D., Simão, T., Macedo, J., Valadão, A., Drechsel, M., Freitas, F., Vidal, M., Guedes, H., Rodrigues, E., Meneses, C., Brioso, P., Pozzer, L., Figueiredo, D., Montano, H., Junior, J., de Souza, G., Quintana, V. M. F., Ferreira, B., Branco, A.,

Gonzalez, P., Guillobel, H., Lemos, M., Seibel, L., Macedo, J., Alves-Ferreira, M., Sachetto-Martins, G., Coelho, A., Santos, E., Amaral, G., Neves, A., Pacheco, A. B., Carvalho, D., Lery, L., Bisch, P., Rössle, S. C., Ürményi, T., Pereira, A. R., Silva, R., Rondinelli, E., von Krüger, W., Marins, O., Baldani, J. I. y Ferreira, P. C. (2009). Complete genome sequence of the sugarcane nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus* Pal5. *BMC Genomics*, *10*, 1-17.

Blankenberg, D., Von Kuster, G., Bouvier, E., Baker, D., Afgan, E., Stoler, N., the Galaxy Team, Taylor, J. y Nekrutenko, A. (2014). Dissemination of scientific software with Galaxy ToolShed. *Genome Biology*, *15*, 1-3.

Brandão, P. R., Crespo, M. T., y Nascimento, F. X. (2022). Phylogenomic and comparative analyses support the reclassification of several *Komagataeibacter* species as novel members of the *Novacetimonas gen. nov.* and bring new insights into the evolution of cellulose synthase genes. *International Journal of Systematic and Evolutionary Microbiology*, *72*(2), 005252.

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason J. A., Stevens, R., Vonstein, V., Wattam, A. R. y Xia, F. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, *5*(1), 1-6.

Brischetto, C., Rossi, V., y Fedele, G. (2024). The microbiome analysis of ripen grape berries supports the complex etiology of sour rot. *Frontiers in Microbiology*, *15*, 1450443.

Calderón de Rzedowski, G. (1998). *Orobanchaceae*. Flora del Bajío y de Regiones Adyacentes, (69), 1-11. Instituto de Ecología, A.C., Centro Regional del Bajío.

Cepec, E., y Trček, J. (2022). Antimicrobial resistance of *Acetobacter* and *Komagataeibacter* species originating from vinegars. *International journal of Environmental Research and Public Health*, *19*(1), 463.

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., Rooney, A. P., Yi, H., Xu, X., Meyer, S. y Trujillo, M. E. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, 68(1), 461-466.

Cleenwerck, I., De Vos, P., y De Vuyst, L. (2010). Phylogeny and differentiation of species of the genus *Gluconacetobacter* and related taxa based on multilocus sequence analyses of housekeeping genes and reclassification of *Acetobacter xylinus* subsp. *sacrofermentans* as *Gluconacetobacter sacrofermentans* (Toyosaki et al. 1996) sp. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 60(10), 2277-2283.

Cook, S. L. (1930). The Ethnobotany of Jemez Indians.

Darling, A. E., Mau, B., y Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*, 5(6), e11147.

Davis, J. J., Boisvert, S., Brettin, T., Kenyon, R. W., Mao, C., Olson, R., Overbeek, R., Santerre, J., Shukla, M., Wattam, A. R., Will, R., Xia, F. y Stevens, R. (2016a). Antimicrobial resistance prediction in PATRIC and RAST. *Scientific Reports*, 6(1), 27930.

Davis, J. J., Gerdes, S., Olsen, G. J., Olson, R., Pusch, G. D., Shukla, M., Vonstein, V., Wattam, A. R. y Yoo, H. (2016b). PATtyFams: protein families for the microbial genomes in the PATRIC database. *Frontiers in Microbiology*, 7, 118.

Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E. M., Gabbard, J. L., Gerdes, S., Guard, A., Keynon, R. W., Machi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E. K., Olsen, G. J., Olson, R. D., Overbeek, J. C., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomas, C., VanOeffelen, M., Vonstein, V., Warren, A. S., Xia, F., Xie, D., Yoo, H. y Stevens, R. (2020). The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Research*, 48(D1), D606-D612.

Delcher, A. L., Bratke, K. A., Powers, E. C., y Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, 23(6), 673-679.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792-1797.

Fuentes-Ramírez, L. E., Bustillos-Cristales, R., Tapia-Hernández, A., Jiménez-Salgado, T., Wang, E. T., Martínez-Romero, E., y Caballero-Mellado, J. (2001). Novel nitrogen-fixing acetic acid bacteria, *Gluconacetobacter johannae* sp. nov. and *Gluconacetobacter azotocaptans* sp. nov., associated with coffee plants. *International Journal of Systematic and Evolutionary Microbiology*, 51(4), 1305-1314.

Gomes, R. J., de Fatima Borges, M., de Freitas Rosa, M., Castro-Gómez, R. J. H., y Spinoso, W. A. (2018). Acetic acid bacteria in the food industry: systematics, characteristics and applications. *Food Technology and Biotechnology*, 56(2), 139.

Gurevich, A., Saveliev, V., Vyahhi, N., y Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072-1075.

He, Y., Racz, R., Sayers, S., Lin, Y., Todd, T., Hur, J., Li, X., Patel, M., Zhao, B., Chung, M., Ostrow, J., Sylora, A., Dungarani, P., Ulysse, G., Kochhar, K., Vidri, B., Strait, K., Jourdain, G. W. y Xiang, Z. (2014). Updates on the web-based VIOLIN vaccine database and analysis system. *Nucleic Acids Research*, 42(D1), D1124-D1132.

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., y Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11, 1-11.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., y Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, 9(1), 5114.

Jelenko, K., Cepec, E., Nascimento, F. X., y Trček, J. (2023). Comparative genomics and phenotypic characterization of *Gluconacetobacter entanii*, a highly acetic acid-tolerant bacterium from vinegars. *Foods*, *12*(1), 214.

Jolley, K. A., Bray, J. E., y Maiden, M. C. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST. org website and their applications. *Wellcome Open Research*, *3*.

Karcher, N., Pasolli, E., Asnicar, F., Huang, K. D., Tett, A., Manara, S., Armanini, F., Bain, D., Duncan, S. H., Louis, P., Zolfo, M., Manghi, P., Valles-Colomer M., Raffaetà, R., Rota-Stabelli, O., Collado, M. C., Zeller, G., Falush, D., Maixner, F., Walker, A. W., Huttenhower, C. y Segata, N. (2020). Analysis of 1321 *Eubacterium rectale* genomes from metagenomes uncovers complex phylogeographic population structure and subspecies functional adaptations. *Genome Biology*, *21*, 1-27.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. y 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.

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Research*, *12*(4), 656-664.

Kim, S. H., Jang, H. W., Park, J. J., Nam, D. G., Lee, S. J., Yeo, S. H., y Kim, S. Y. (2024). Antibiotic Resistance in Acetic Acid Bacteria Originating from Vinegar. *Antibiotics*, *13*(7), 626.

Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A., ... y Wishart, D. S. (2024). DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic Acids Research*, *52*(D1), D1265-D1275.

Kolmogorov, M., Yuan, J., Lin, Y., y Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540-546.

Komagata, K., Iino, T., y Yamada, Y. (2014). The family acetobacteraceae. *The prokaryotes*, 3-78.

Krueger, F. (2012). Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisulfite-Seq) libraries. URL http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. (Date of access: 28/04/2016), 71.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. y Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome research*, 19(9), 1639-1645.

Kuismin, M. O., Ahlinder, J., y Sillanpää, M. J. (2017). CONE: community oriented network estimation is a versatile framework for inferring population structure in large-scale sequencing data. *G3: Genes, Genomes, Genetics*, 7(10), 3359-3377.

Letunic, I., y Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research*, gkae268.

Liu, B., Zheng, D., Zhou, S., Chen, L., y Yang, J. (2022). VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Research*, 50(D1), D912-D917.

Mamlouk, D., y Gullo, M. (2013). Acetic acid bacteria: physiology and carbon sources oxidation. *Indian Journal of Microbiology*, 53, 377-384.

Mao, C., Abraham, D., Wattam, A. R., Wilson, M. J., Shukla, M., Yoo, H. S., y Sobral, B. W. (2015). Curation, integration and visualization of bacterial virulence factors in PATRIC. *Bioinformatics*, 31(2), 252-258.

Matsushita, K., y Matsutani, M. (2016). Chapter 7: Distribution, evolution and physiology of oxidative fermentation. En K. Matsushita, *Acetic Acid Bacteria* (pp. [159-178]). Springer Japan DOI 10.1007/978-4-431-55933-7_7.

Matsutani, M., Hirakawa, H., Yakushi, T., y Matsushita, K. (2011). Genome-wide phylogenetic analysis of *Gluconobacter*, *Acetobacter*, and *Gluconacetobacter*. *FEMS Microbiology Letters*, 315(2), 122-128.

Mukhtar, S., Farooq, M., Baig, D. N., Amin, I., Lazarovits, G., Malik, K. A., Yuan, Z., y Mehnaz, S. (2021). Whole genome analysis of *Gluconacetobacter azotocaptans* DS1 and its beneficial effects on plant growth. *3 Biotech*, 11(10), 450.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., y Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268-274.

Olson, R. D., Assaf, R., Brettin, T., Conrad, N., Cucinell, C., Davis, J. J., Dempsey, D. M., Dickerman, A., Dietrich, E. M., Kenyon, R. W., Kuscuoglu, M., Lefkowitz, E. J., Lu, J., Machi, D., Macken, C., Mao, C., Niewiadomska, A., Nguyen, M., Olsen, G. J., Overbeek, J. C., Parrello, B., Parrello, V., Porter, J. S., Pusch, G. D., Shukla, M., Singh, I., Stewart, L., Tan, G., Thomas, C., VanOeffelen, M., Vonstein, V., Wallace, Z. S., Warren, A. S., Wattam, A. R., Xia, F., Yoo, H., Zhang, Y., Zmasek, C. M., Scheuermann, R. H. y Stevens, R. L. (2023). Introducing the bacterial and viral bioinformatics resource center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research*, 51(D1), D678-D689.

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., y Phillippy, A. M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology*, 17, 1-14.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O. y Vonstein, V. (2005). The subsystems approach to genome

annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, 33(17), 5691-5702.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. y Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, 42(D1), D206-D214.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., Fookes, M., Falush, D., Keane, J. A. y Parkhill, J. (2015). Roary: rapid large-scale prokaryote pangenome analysis. *Bioinformatics*, 31(22), 3691-3693.

Raspor, P., y Goranovič, D. (2008). Biotechnological applications of acetic acid bacteria. *Critical reviews in biotechnology*, 28(2), 101-124.

Pedraza-Pérez, Y., Cuevas-Vede, R. A., Canto-Gómez, Á. B., López-Pliego, L., Gutiérrez-Ríos, R. M., Hernández-Lucas, I., Rubín-Linares, G., Martínez-Laguna, Y., López-Olguín, J. F. y Fuentes-Ramírez, L. E. (2018). BLAST-XYPlot viewer: a tool for performing BLAST in whole-genome sequenced bacteria/archaea and visualize whole results simultaneously. *G3: Genes, Genomes, Genetics*, 8(7), 2167-2172.

Reis, V. M., y Teixeira, K. R. D. S. (2015). Nitrogen fixing bacteria in the family Acetobacteraceae and their role in agriculture. *Journal of Basic Microbiology*, 55(8), 931-949.

Riesco, R., y Trujillo, M. E. (2024). Update on the proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, 74(3), 006300.

Rodrigues, A. G., Colwell, A. E., y Stefanović, S. (2011). Molecular systematics of the parasitic genus *Conopholis* (Orobanchaceae) inferred from plastid and nuclear sequences. *American Journal of Botany*, 98(5), 896-908.

Rodrigues, A., Shaya, S., Dickinson, T. A., y Stefanović, S. (2013). Morphometric analyses and taxonomic revision of the North American holoparasitic genus *Conopholis* (Orobanchaceae). *Systematic Botany*, 38(3), 795-804.

Román-Camacho, J. J., García-García, I., Santos-Dueñas, I. M., García-Martínez, T., y Mauricio, J. C. (2023). Latest trends in industrial vinegar production and the role of acetic acid bacteria: classification, metabolism, and applications—a comprehensive review. *Foods*, 12(19), 3705.

Rubio, S., Pacheco-Orozco, R. A., Milena Gómez, A., Perdomo, S., y García-Robles, R. (2020). Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. *Universitas Médica*, 61(2), 49-63.

Saier Jr, M. H., Reddy, V. S., Moreno-Hagelsieb, G., Hendargo, K. J., Zhang, Y., Iddamsetty, V., Jing, K. K. L., Tian, N., Russum, S., Wang, J. y Medrano-Soto, A. (2021). The transporter classification database (TCDB): 2021 update. *Nucleic Acids Research*, 49(D1), D461-D467.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068-2069.

Slattery, J. (2020). Flor de tierra, *Southwest Medicinal Plants* (p. 183). Timber Press, Inc.

Sombolestani, A. S., Cleenwerck, I., Cnockaert, M., Borremans, W., Wieme, A. D., Moutia, Y., Spaepen, S., De Vuyst, L. y Vandamme, P. (2021). *Gluconacetobacter dulcium* sp. nov., a novel *Gluconacetobacter* species from sugar-rich environments. *International Journal of Systematic and Evolutionary Microbiology*, 71(1), 004569.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312-1313.

Swank, G. R. (1932). The ethnobotany of the Acoma and Laguna Indians.

Törün, B., Çoban, E. P., Bozdoğan, B., y Bıyık, H. H. (2019). Determination of genetic diversity of *Gluconobacter hansenii* with RAPD and ERIC-PCR. *European Journal of Biotechnology and Bioscience*, 7(6), 49-59.

Vanhove, M., Sicard, A., Ezennia, J., Leviten, N., y Almeida, R. P. (2020). Population structure and adaptation of a bacterial pathogen in California grapevines. *Environmental Microbiology*, 22(7), 2625-2638.

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, 27(5), 737-746.

Wang, X., Wang, D., Zhang, R., Qin, X., Shen, X., y You, C. (2024). Morphological Structure Identification, Comparative Mitochondrial Genomics and Population Genetic Analysis toward Exploring Interspecific Variations and Phylogenetic Implications of *Malus baccata* 'ZA' and Other Species. *Biomolecules*, 14(8), 912.

Yamada, Y., y Yukphan, P. (2008). Genera and species in acetic acid bacteria. *International Journal of Food Microbiology*, 125(1), 15-24.

Yamada, Y., Yukphan, P., Vu, H. T. L., Muramatsu, Y., Ochaikul, D., Tanasupawat, S., y Nakagawa, Y. (2012). Description of *Komagataeibacter* gen. nov., with proposals of new combinations (Acetobacteraceae). *The Journal of General and Applied Microbiology*, 58(5), 397-404.

Zhou, Y., Zhang, Y., Zhao, D., Yu, X., Shen, X., Zhou, Y., ... y Zhu, F. (2024). TTD: Therapeutic Target Database describing target druggability information. *Nucleic Acids Research*, 52(D1), D1465-D1477.

Zhu, D., He, J., Yang, Z., Wang, M., Jia, R., Chen, S., Liu, M., Zhao, X., Yang, Q., Wu, Y., Zhang, S., Liu, Y., Zhang, L., Yu, Y., You, Y., Chen, X. y Cheng, A. (2019). Comparative analysis reveals the Genomic Islands in *Pasteurella multocida* population genetics: on Symbiosis and adaptability. *BMC Genomics*, 20, 1-11.

12. Anexos

- **Genes de Especialidad**

Como parte de los resultados arrojados por las herramientas de anotación y análisis de genoma completo el BV.BRC proporciona un análisis preliminar de los genes de especialidad y genes de resistencia antimicrobiana. A través de un BLAT (BLAST Like Alingment Tool), una herramienta que realiza una comparación de todos los genes del genoma de interés con bases de datos de genes especializados; incluyendo los genes conocidos por proporcionar resistencia a los antibióticos, factores de virulencia, transportadores y dianas farmacológicas conocidas; se obtuvo un listado de la cantidad de genes en el genoma de interés que tienen homología con los genes de esas bases de datos (Tabla 9).

Genes de Especialidad	Fuente	Genes
Resistencia antibióticos	CARD	1
Resistencia antibióticos	PATRIC	67
Diana Farmacéutica	DrugBank	4
Diana Farmacéutica	TTD	1
Transportador	TCDB	3
Factor de virulencia	Victors	3

Tabla suplementaria 1. Cantidad de genes de especialidad encontrados en el genoma de UAPS01-405, se incluye la fuente de la base de datos

- **Genes de Resistencia Antimicrobiana**

El BV-BRC provee un método de detección para genes de resistencia a antimicrobianos basado en k-meros (fragmento de la secuencia de nucleótidos de longitud k utilizado para realizar análisis bioinformáticos), que muestra el número de genes que comparten estos k-meros. A través de una anotación funcional se obtiene una predicción del fenotipo de resistencia a antimicrobianos, refiriéndose a la resistencia o susceptibilidad de un organismo dado, hacia uno o más antibióticos. En la tabla 9 se proporciona el listado de los genes de resistencia encontrados en el genoma de UAPS01-405, junto a su mecanismo de resistencia (izquierda).

Existen pocos reportes sobre resistencia antimicrobiana en las BAA ya que debido a su susceptibilidad a realizar transferencia horizontal de material genético se dificulta la reproducibilidad de los ensayos. Dentro de los genes de resistencia hallados en este y otros análisis (Kim *et al.*, 2024 y Cepec y Trček, 2021) se destacan los relacionados a los sistemas de bombas de flujo, por su relación con la liberación de compuestos tóxicos hacia el exterior de la célula, pues parte de su metabolismo se encarga de liberar ácido acético al medio de cultivo.

Genes relacionados a Resistencia Antimicrobiana	
Mecanismo de Resistencia	Genes
Enzima de inactivación de antibióticos	<i>AAC(6')-Ic,f,g,h,j,k,l,r-z</i>
Diana de antibióticos en especies susceptibles	<i>Alr, Ddl, dxr, EF-G, EF-Tu, folP, gyrA, gyrB, inhA, fabI, Iso-tRNA, kasA, MurA, rho, rpoB, rpoC, S10p, S12p</i>
Bomba de flujo que confiere resistencia antibiótica	<i>EmrAB-ToIC, MdtABC-OMF, MdtABC-ToIC, MexAB-OprM, MexHI-OpmD, MexHI-OpmD system</i>
Gen que confiere resistencia antibiótica por ausencia	<i>gidB</i>
Proteína que altera la carga de la pared celular confiriendo resistencia antibiótica	<i>PgsA</i>
Proteína que modula la permeabilidad a antibióticos	<i>OprB</i> , familia <i>OprB</i>
Genes reguladores de la modulación de la expresión de resistencia a antibióticos	<i>OxyR</i>

Tabla suplementaria 2. Genes de resistencia antimicrobiana encontrados en el genoma de

Como se observa en la tabla 10 y en la tabla de Genes de resistencia en los anexos, la mayoría de los genes de resistencia antimicrobiana se encuentran relacionados a sistemas bombas de expulsión y proteínas implicadas en la permeabilidad de la membrana, esto puede deberse a que, por su metabolismo, las bacterias del género *Gluconacetobacter* requieren de la expulsión de ácidos orgánicos desde el interior celular hacia el medio. Por lo que es posible que el programa considere relacionar a estas bombas con la expulsión de antibióticos.

- **BLAST**

A continuación se muestran las tablas obtenidas de los BLAST realizados con genes Housekeeping en la plataforma del NCBI.

BLASTn <i>dnaA</i>							
Descripción	Score máximo	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
Gluconacetobacter diazotrophicus PA1 5	998	998	100%	0	92.09%	3944163	AM889285.1
Gluconacetobacter diazotrophicus PA1 5	998	998	100%	0	92.09%	3887492	CP001189.1
Nguyenibacter sp. L1	926	926	100%	0	90.25%	4294433	CP127015.1
Nguyenibacter vanlangensis	915	915	100%	0	89.97%	4751868	CP152276.1
Komagataeibacter oboediens	837	837	100%	0	87.99%	3495026	CP117861.1
Komagataeibacter oboediens	837	837	100%	0	87.99%	2642352	CP043481.1
Komagataeibacter europaeus	815	815	100%	0	87.45%	3448767	CP021467.1
Komagataeibacter diospyri	809	809	100%	0	87.29%	3527157	CP171836.1
Komagataeibacter medellinensis NBRC 3288	782	782	100%	0	86.62%	3136818	AP012159.1
uncultured Alphaproteobacteria bacterium	765	765	99%	0	86.38%	39640	KY010302.1

Tabla suplementaria 3. BLAST obtenido con *dnaA* de *Gluconacetobacter* UAPS01-405

BLAST <i>fabZ</i>							
Descripción	Score máximo	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
Komagataeibacter xylinus	359	359	98%	2.00E-94	82.21%	3527401	CP041348.1
Komagataeibacter xylinus	348	348	98%	4.00E-91	81.73%	3353346	CP025269.1
Komagataeibacter rhaeticus	311	311	99%	6.00E-80	80.00%	3161283	LT575493.1
Komagataeibacter rhaeticus	311	311	99%	6.00E-80	80.00%	3645912	CP061369.1
Komagataeibacter rhaeticus	311	311	99%	6.00E-80	80.00%	3597563	CP050139.1
Komagataeibacter rhaeticus	311	311	99%	6.00E-80	80.00%	3563316	CP139093.1
Acetobacter aceti	291	291	98%	8.00E-74	79.38%	3725037	CP014692.1
Acetobacter aceti	268	268	98%	4.00E-67	78.33%	3743357	AP023326.1
Acetobacter aceti NBRC 14818	257	257	98%	8.00E-64	77.86%	3596270	AP023410.1
Gluconobacter albidus	226	226	98%	2.00E-54	76.56%	3107922	CP014689.1

Tabla suplementaria 4. BLAST obtenido con *fabZ* de *Gluconacetobacter* UAPS01-405

BLAST <i>ftsZ</i>							
Descripción	Score máximo	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
Gluconacetobacter diazotrophicus PA1 5	1482	2189	95%	0	85.37%	3887492	CP001189.1
Gluconacetobacter diazotrophicus PA1 5	1476	2316	95%	0	85.30%	3944163	AM889285.1
Nguyenibacter sp. L1	1321	1377	59%	0	90.65%	4294433	CP127015.1
Nguyenibacter vanlangensis	1299	1454	66%	0	90.25%	4751868	CP152276.1
Lichenicola cladoniae	1101	1101	57%	0	87.27%	4799761	CP053708.1
uncultured Lichenicoccus sp.	1083	1083	58%	0	86.72%	704535	OY843755.1
Komagataeibacter rhaeticus	1055	1055	61%	0	85.34%	3161283	LT575493.1
Komagataeibacter rhaeticus	1055	1055	61%	0	85.34%	3645912	CP061369.1
Komagataeibacter rhaeticus	1055	1055	61%	0	85.34%	3597563	CP050139.1
Komagataeibacter rhaeticus	1055	1055	61%	0	85.34%	3563316	CP139093.1

Tabla suplementaria 5. BLAST obtenido con *ftsZ* de *Gluconacetobacter* UAPS01-405

BLASTn <i>gyrB</i>							
Descripción	Score máximo	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
Gluconacetobacter diazotrophicus PA1 5	1094	1094	83%	0	85.67%	3944163	AM889285.1
Gluconacetobacter diazotrophicus PA1 5	1066	1066	83%	0	85.22%	3887492	CP001189.1
Nguyenibacter sp. L1	1007	1007	83%	0	84.13%	4294433	CP127015.1
Nguyenibacter vanlangensis	979	979	83%	0	83.67%	4751868	CP152276.1
Komagataeibacter sucrofermentans	758	758	83%	0	79.94%	2947958	CP137152.1
Komagataeibacter sucrofermentans	758	758	83%	0	79.94%	2948989	CP137147.1
Komagataeibacter sucrofermentans DSM 15973	758	758	83%	0	79.94%	2947970	CP137157.1
Komagataeibacter medellinensis NBRC 3288	701	701	83%	0	78.90%	3136818	AP012159.1
Komagataeibacter rhaeticus	656	656	83%	0	78.15%	3161283	LT575493.1
Komagataeibacter rhaeticus	656	656	83%	0	78.15%	3645912	CP061369.1
Komagataeibacter rhaeticus	656	656	83%	0	78.15%	3597563	CP050139.1

Tabla suplementaria 6. BLAST obtenido con *gyrB* de *Gluconacetobacter* UAPS01-405

BLASTn <i>recA</i>							
Descripción	Score máximo	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
Gluconacetobacter diazotrophicus PA1 5	1500	1500	99%	0	92.62%	3944163	AM889285.1
Gluconacetobacter diazotrophicus PA1 5	1495	1495	99%	0	92.52%	3887492	CP001189.1
Nguyenibacter vanlangensis	1434	1434	99%	0	91.40%	4751868	CP152276.1
Nguyenibacter sp. L1	1434	1434	99%	0	91.40%	4294433	CP127015.1
Komagataeibacter rhaeticus	1134	1134	96%	0	86.81%	3161283	LT575493.1
Komagataeibacter rhaeticus	1134	1134	96%	0	86.81%	3645912	CP061369.1
Komagataeibacter rhaeticus	1134	1134	96%	0	86.81%	3597563	CP050139.1
Komagataeibacter rhaeticus	1134	1134	96%	0	86.71%	3563316	CP139093.1
Gluconacetobacter polyoxogenes	1112	1112	97%	0	86.31%	1200	D13183.1
Komagataeibacter nataicola	1107	1107	97%	0	86.22%	3485191	CP019875.1

Tabla suplementaria 7. BLAST obtenido con *recA* de *Gluconacetobacter* UAPS01-405

BLASTn <i>rpoD</i>							
Descripción	Score máximo	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
Gluconacetobacter diazotrophicus PA1 5	2562	2562	100%	0	90.49%	3944163	AM889285.1
Gluconacetobacter diazotrophicus PA1 5	2562	2562	100%	0	90.49%	3887492	CP001189.1
Nguyenibacter sp. L1	2412	2412	100%	0	89.12%	4294433	CP127015.1
Nguyenibacter vanlangensis	2401	2401	100%	0	89.02%	4751868	CP152276.1
Komagataeibacter nataicola	1818	1818	100%	0	83.64%	3452065	CP118675.1
Komagataeibacter nataicola	1805	1805	100%	0	83.53%	3533141	CP134885.1
Komagataeibacter nataicola	1801	1801	100%	0	83.48%	3485191	CP019875.1
Komagataeibacter rhaeticus	1801	1801	100%	0	83.50%	3161283	LT575493.1
Komagataeibacter rhaeticus	1801	1801	100%	0	83.50%	3645912	CP061369.1
Komagataeibacter rhaeticus	1801	1801	100%	0	83.50%	3597563	CP050139.1

Tabla suplementaria 8. BLAST obtenido del *rpoD* de *Gluconacetobacter* UAPS01-405

BLASTn <i>dnaK</i>							
Descripción	Score máximo	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
Gluconacetobacter diazotrophicus PA1 5	2283	2283	99%	0	93.23%	3944163	AM889285.1
Gluconacetobacter diazotrophicus PA1 5	2283	2283	99%	0	93.23%	3887492	CP001189.1
Nguyenibacter vanlangensis	2097	2097	99%	0	91.10%	4751868	CP152276.1
Nguyenibacter sp. L1	2097	2097	99%	0	91.10%	4294433	CP127015.1
Novacetimonas hansenii	1770	1770	99%	0	87.35%	3424868	CP094848.1
Komagataeibacter xylinus	1762	1762	99%	0	87.38%	3353346	CP025269.1
Komagataeibacter sucrofermentans	1760	1760	99%	0	87.38%	2947958	CP137152.1
Komagataeibacter sucrofermentans	1760	1760	99%	0	87.38%	2948989	CP137147.1
Komagataeibacter sucrofermentans DSM 15973	1760	1760	99%	0	87.38%	2947970	CP137157.1
Novacetimonas hansenii	1753	1753	99%	0	87.16%	3530618	CP062147.1

Tabla suplementaria 9. BLAST obtenido del *dnaK* de *Gluconacetobacter* UAPS01-405

BLASTn <i>groEL</i>							
Descripción	Score máximo	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
Gluconacetobacter diazotrophicus PA1 5	2481	3621	99%	0	93.97%	3944163	AM889285.1
Gluconacetobacter diazotrophicus PA1 5	2481	3621	99%	0	93.97%	3887492	CP001189.1
Nguyenibacter vanlangensis	2418	2418	99%	0	93.29%	4751868	CP152276.1
Nguyenibacter sp. L1	2407	3544	99%	0	93.17%	4294433	CP127015.1
Komagataeibacter europaeus	1930	1930	99%	0	88.03%	3448767	CP021467.1
Komagataeibacter rhaeticus	1908	1908	100%	0	87.67%	3161283	LT575493.1
Komagataeibacter rhaeticus	1908	1908	100%	0	87.67%	3645912	CP061369.1
Komagataeibacter rhaeticus	1908	1908	100%	0	87.67%	3597563	CP050139.1
Komagataeibacter rhaeticus	1908	1908	100%	0	87.67%	3563316	CP139093.1
Komagataeibacter diospyri	1903	1903	99%	0	87.73%	3527157	CP171836.1

Tabla suplementaria 10. BLAST obtenido del *groEL* de *Gluconacetobacter* UAPS01-405

BLASTn <i>rpoB</i>							
Descripción	Score máximo	Score Total	Cobertura de Query	E-value	Porcentaje de Identidad	Longitud	Número de acceso
Gluconacetobacter diazotrophicus PA1 5	5860	5860	99%	0	92.07%	3944163	AM889285.1
Gluconacetobacter diazotrophicus PA1 5	5854	5854	99%	0	92.05%	3887492	CP001189.1
Nguyenibacter vanlangensis	5574	5574	100%	0	90.78%	4751868	CP152276.1
Nguyenibacter sp. L1	5568	5568	100%	0	90.75%	4294433	CP127015.1
Komagataeibacter sucrofermentans	4554	4554	100%	0	86.40%	2947958	CP137152.1
Komagataeibacter sucrofermentans	4554	4554	100%	0	86.40%	2948989	CP137147.1
Komagataeibacter sucrofermentans DSM 15973	4554	4554	100%	0	86.40%	2947970	CP137157.1
Komagataeibacter xylinus	4554	4554	100%	0	86.41%	3353346	CP025269.1
Komagataeibacter nataicola	4554	4554	100%	0	86.40%	3452065	CP118675.1
Komagataeibacter nataicola	4549	4549	100%	0	86.38%	3533141	CP134885.1

Tabla suplementaria 11. BLAST obtenido del *rpoB* de *Gluconacetobacter* UAPS01-405

- **Alineamiento de Genomas**

A continuación se muestra la imagen completa obtenida de la herramienta de alineamiento de genomas del BV-BRC (Mauve).

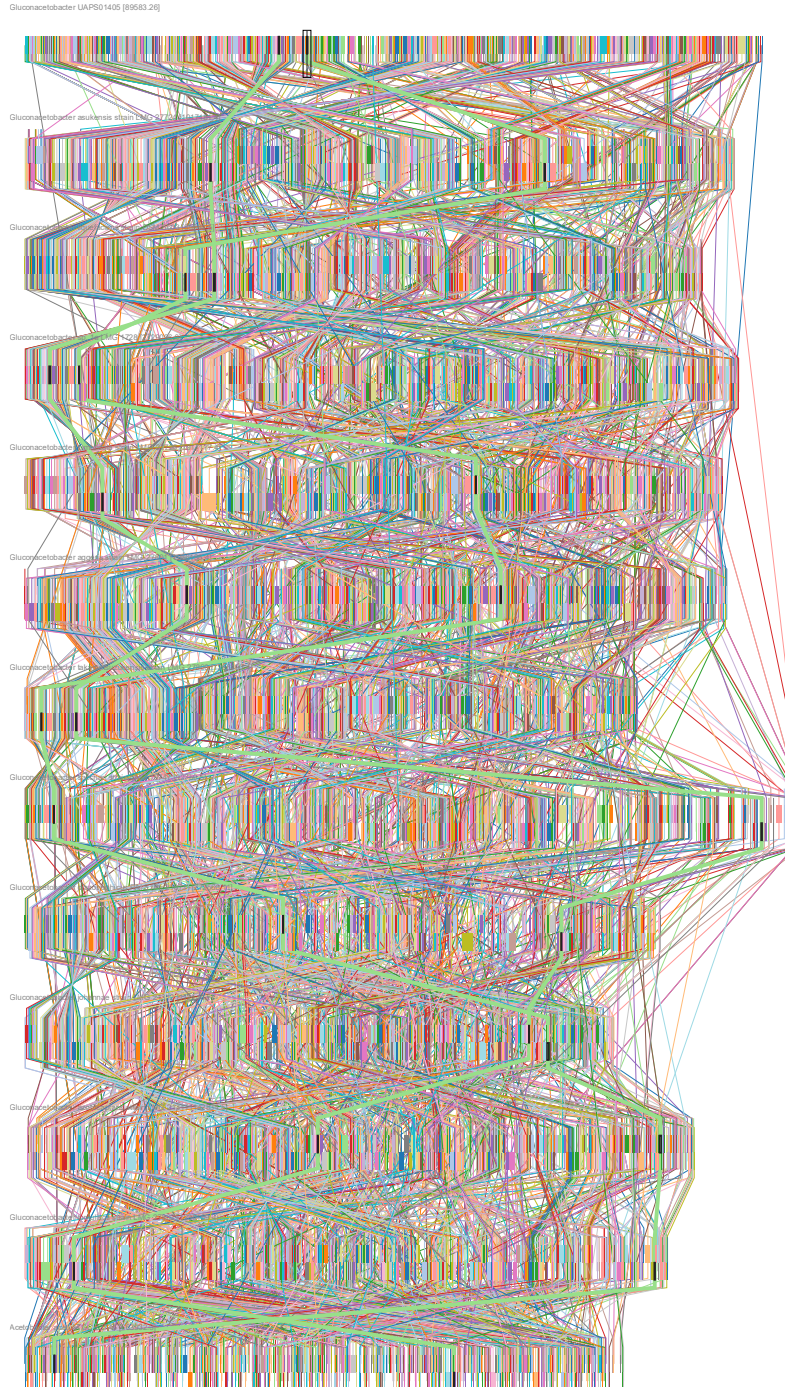


Figura suplementaria 1. Alineamiento de genomas (extendido). Presenta el listado de genomas comparados con genes coloreados de acuerdo a los subsistemas, el alineamiento se aprecia en líneas entre las secuencias de los genomas.