



Benemérita Universidad Autónoma de Puebla

Instituto de Ciencias

Centro de Investigaciones en Ciencias Microbiológicas

Posgrado en Microbiología

“Estudio genómico de la síntesis de melaninas en *Azotobacter* spp.”

Tesis

Que para obtener el grado de:

Maestra en Ciencias (Microbiología) con opción en Bioquímica y
genética microbiana

Presenta

QFB Nohely Aimeé Pizano Sosa

Director de tesis

D.C. Miguel Castañeda Lucio

Co directora de tesis

D.C. Liliana López Pliego

Puebla, Pue.

Marzo 2023

Agradecimientos

A mi familia muchas gracias por su apoyo incondicional para cumplir mis metas.

A mis amigos por todo su cariño, buenos deseos y apoyo en este trayecto, a Jorge por siempre alentarme a seguir aprendiendo, a esforzarme y no dejarme vencer en días malos, por estar siempre a mi lado para impulsarme.

Al Dr. Miguel Castañeda por siempre tener la mejor disposición para ayudarme cada que era necesario, por sus clases de las cuales aprendí mucho y me quedo con mucho también, por su calidez y amabilidad.

A la Dra. Lili por su tan valiosa ayuda en todos aspectos, por todas las sesiones donde me explicó y me transmitió el gusto por la bioinformática, su asesoría para trabajar en el laboratorio del cual le aprendí mucho, su valiosa ayuda para el desarrollo de mis experimentos y el uso de las herramientas bioinformáticas y su gran apoyo como persona desde que llegué a la ciudad de Puebla me hizo sentir muy bien recibida.

Al comité evaluador que con cada tutorial me facilitaban el comprender de manera más clara el camino a seguir en este proyecto.



BUAP

Puebla, Pue. a 21 de marzo 2023.

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 Nohely Aimee Pizano Sosa, les informamos que hemos revisado el escrito titulado:

"Estudio genómico de la síntesis de melaninas en *Azotobacter* spp."

A nuestro juicio, la alumna Nohely Aimee Pizano Sosa 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. Beatriz Eugenia Baca


D.C. Verónica Quintero Hernández


D.C. Santiago Castillo Ramírez


D.C. Luis Ernesto Fuentes Ramírez



Resumen

El género *Azotobacter* es considerado un modelo de interés biotecnológico por ser de metabolismo versátil, fácil crecimiento, están muy bien estudiadas las condiciones de crecimiento tanto en cultivo continuo como en lote. De las siete especies que lo conforman existe registro de que tres de ellas, *A. salinestris*, *A. chroococcum* y *A. nigricans* tienen la capacidad de sintetizar el biopolímero melanina. En el año 2019 en el Laboratorio de Genética Molecular Microbiana del Instituto de Ciencias en Microbiología de la Benemérita Universidad Autónoma de Puebla se aislaron cepas de *Azotobacter* que presentaron producción de pigmento negro característico en la producción de melaninas. De acuerdo con la secuencia del gen 16S rDNA y la utilización de marcadores específicos se identificaron como *A. chroococcum* y *A. vinelandii*; es remarcable que de esta última no se tiene registro previo de su capacidad para sintetizar melaninas. Dicho biopolímero es de interés por presentar aplicaciones biotecnológicas en diversas áreas como en la farmacéutica, agricultura y nanotecnología. Hasta el momento se conocen cuatro vías de síntesis de melanina, una de ellas tiene como precursor la molécula de malonil-CoA y las otras tres vías utilizan como precursor la tirosina para sintetizar este biopolímero.

El objetivo de esta tesis es estudiar a nivel genómico las rutas biosintéticas de la síntesis de melaninas en cepas del género *Azotobacter*. Se tomó como punto de referencia a los genes involucrados en la síntesis de melaninas de diversas especies bacterianas en las que se había reportado la síntesis de estos biopolímeros, la cual se tomó como *query*. Con esa *query*, se permitió identificar algunos genes que posiblemente se encuentren involucrados en la síntesis de melaninas; a través de comparaciones tipo Blastp teniendo como punto de partida a los genomas completos de la base de datos de NCBI se obtuvieron nueve genes posiblemente implicados en la producción de melaninas en el genoma de *Azotobacter*: 4-hidroxifenilpiruvato dioxigenasa, Fosfoenolpiruvato sintasa, Policétido sintasa tipo III, Policétido sintasa tipo I, Homogentisato-1,2-dioxigenasa, Aspartato aromático tirosina aminotransferasa, Policétido sintasa deshidratasa, Fenilalanina-4-monooxigenasa y D-tirosil-

tRNA (Tir) deasilasa que podrían participar en la vía de síntesis que toma como precursor al malonil-CoA y a una de las vías que parten de la tirosina la cuál es conocida como la vía del homogentisato.

Se llevó a cabo la secuenciación y el análisis genómico de dos cepas aisladas en un trabajo previo de la reserva de Tehuacán Cuicatlán; se realizó el análisis de las secuencias, el ensamblaje y anotación de estos genomas. A partir de estos resultados, encontrando los nueve genes propuestos en la vía de síntesis de melaninas. Posteriormente se realizó un estudio del pangenoma para saber si los genes se encuentran en el núcleo o en el genoma accesorio del organismo y con ello se pudo observar que nueve genes únicamente Policétido sintasa tipo III y Homogentisato-1,2-dioxigenasa se encuentran en el genoma accesorio lo que permite enfocar los futuros estudios en estos dos genes.

Índice

Agradecimientos	2
Resumen	4
Índice	6
Introducción	9
1. Género <i>Azotobacter</i>	9
1.1. <i>Azotobacter vinelandii</i>	9
1.2. <i>Azotobacter chroococcum</i>	11
2. Melanina	13
2.1. Vías de síntesis.....	14
2.2. Enzimas involucradas en la síntesis de melanina.....	15
2.3. Aplicaciones biotecnológicas.....	17
3. Secuenciación masiva y bioinformática	18
3.1. Secuenciación “Illumina”	18
3.2. Secuenciación “Oxford nanopore”.....	20
3.3. Ensamblado genómico	22
3.4. Anotaciones genómicas.....	23
3.5. Análisis de alineamiento y comparación de genomas	23
Antecedentes.....	24
Material biológico.....	27
Justificación	27
Hipótesis	29
Objetivos.....	29

General.....	29
Específicos.....	29
Resultados y discusión.....	30
1. Genomas secuenciados de diferentes especies de <i>Azotobacter</i>	30
2. Alineamiento de secuencias de proteínas (BLAST protein).....	31
2.1 Análisis tipo BLAST en especies de la familia <i>Pseudomonadaceae</i>	31
2.2 Análisis de genomas de especies no relacionadas filogenéticamente con la familia <i>Pseudomonadaceae</i>	32
2.3 Proteoma del género <i>Azotobacter</i>	32
3. Contexto genómico y sintenia.....	33
4. Ensayos fenotípicos	33
5. Extracción de ADN.....	34
6. Secuenciación del genoma.....	35
7. Análisis de calidad de la secuenciación	36
7.1 S4T (Lecturas cortas)	36
7.2 . S4T (Lecturas largas)	43
6.3. SM3.....	44
8. Ensamblado del genoma	44
8.1 Average Nucleotide Identity (ANI) Calculator	45
9. Anotación del genoma	46
10. Blast de proteínas.....	47
11. Pangenoma.....	47
Conclusiones.....	54
Perspectivas	55

Anexos	56
Anexo 1. Ruta metabolismo de tirosina en genoma de <i>Azotobacter vinelandii</i> DJ.	56
Anexo 2. Protocolo Kit Wizard Genetic DNA Purification de Promega para extracción de ADN en bacterias Gram negativas.	57
Anexo 3. Contexto genómico y sintenia.....	59
Bibliografía.....	73

Introducción

1. Género *Azotobacter*

El género *Azotobacter* pertenece a la familia Pseudomonadacea que filogenéticamente está relacionado con el género de las *Pseudomonas* y ambas se incluyen en el dominio de las γ -proteobacterias. Este género se encuentra distribuido en diversos entornos ambientales como suelo, agua, rizosfera de las plantas y en superficies de hojarasca (Mazinani & Asgharzadeh, 2014).

Se conoce que el género *Azotobacter* está conformado por siete especies llamadas, *A. chroococcum*, *A. vinelandii*, *A. beijerinckii*, *A. paspali*, *A. armeniacus*, *A. nigricans* y *A. salinestris*, algunas de las cuales han sido ampliamente documentadas por su capacidad de producción de exopolisacáridos (EPS) de interés biotecnológico, fijación de nitrógeno atmosférico y la promoción general del crecimiento de plantas (PGPR) (Kennedy et al., 2005). Son organismos de fácil crecimiento y muestran un metabolismo versátil, se conocen las condiciones de crecimiento en cultivo y en lote. Este es un género ampliamente utilizado para el estudio de procesos bioquímicos fundamentales, por ejemplo, transcripción, transporte de electrones y almacenamiento de hierro. Los estudios en *A. vinelandii* y *A. chroococcum* han sido fundamentales para poder entender la fisiología, bioquímica y genética de la fijación de nitrógeno y el metabolismo de H₂ (Dev Jnawali et al., 2015).

1.1. *Azotobacter vinelandii*

Azotobacter vinelandii es una bacteria gram negativa, γ -proteobacteria, aerobia. A nivel microscópico presenta un estado pleomórfico que varía desde formas bacilares hasta cocoides y se pueden observar ya sea individuales, en pares o formando agregados irregulares con un diámetro aproximado de 1.6-2.5 x 3-5 μm . En cultivo, *A. vinelandii* crece formando colonias generalmente opacas, lisas, poco convexas y mucoides (Kennedy et al., 2005).

A. vinelandii fue aislada por primera vez de suelo de Vineland, Nueva Jersey (Noar & Bruno Bárcena, 2018). Como parte de la colección bacteriana de la Universidad de

Wisconsin-Madison, se encuentra la cepa *A. vinelandii* O que posee un fenotipo viscoso color blanco. Brush y colaboradores en 1959 asignaron el nombre de *A. vinelandii* OP a una cepa aislada de *Azotobacter* spp. que presentaba fenotipo fluorescente no mucoide, esta cepa también es llamada UW y CA. Existe además la cepa *A. vinelandii* DJ la cual es una variante transformante de la cepa UW generada en 1984 a través de múltiples rondas de transformaciones utilizando ADN cromosómico de cepas con resistencia antibiótica. Su genoma fue secuenciado por Setubal y colaboradores en 2009 y se encuentra disponible en American Type Culture Collection (ATCC) con el número de referencia ATCC BAA-1303.

Existe una gran cantidad de ADN contenido en las células de *A. vinelandii*, el cual en comparación con la cantidad de ADN presente en *E. coli* es cuarenta veces mayor. Su genoma es de 5.36 Mb y es catalogada como una bacteria poliploide por la presencia de múltiples cromosomas, posiblemente alrededor de ochenta copias (Bergey et al., 2001).

A lo largo del tiempo ha sido estudiada como un modelo diazotrófico por su capacidad de fijación de nitrógeno bajo condiciones de aerobiosis al tiempo que protege su complejo de nitrogenasa del daño oxidativo presentando tres versiones de nitrogenasa con variación en el átomo que contienen en el sitio del heterometal, una de ellas contiene un átomo de molibdeno (Mo), otra vanadio (V) y una tercera hierro (Fe), todas ellas reducen el nitrógeno (N₂) a amoníaco (NH₃). Estas enzimas tienen la facilidad de sustitución para permitir el crecimiento cuando uno y otro metal esencial se ausenta o bajo diferentes temperaturas.

Otra capacidad que presenta *A. vinelandii* es la adaptación de su metabolismo a diversas fuentes de nutrientes, incluidos los alcoholes, ácidos orgánicos y azúcares. En ausencia de fuentes de carbono ésta se somete a un proceso de diferenciación para formar quistes (figura 1) que son resistentes a la desecación y a diferentes adversidades ambientales, permitiéndole sobrevivir durante más de 10 años en suelo seco (Setubal et al., 2009). El proceso de enquistamiento se produce naturalmente en la fase estacionaria tardía o por la inducción de células vegetativas con butanol. Dichos quistes presentan en su estructura un cuerpo central conformado por poli b-hidroxibutirato (PHB) rodeado de capas de exinas e intinas en las que el alginato es el componente principal (Rojas Tapias et al., 2013).

La capacidad de producir estos compuestos la convierte en un modelo de estudio con gran interés biotecnológico puesto que el alginato es utilizado como agente viscosante y gelificante en industrias como la farmacéutica y la alimenticia. La capacidad viscosificante y gelificante depende en gran parte de la estructura química, el contenido de unidades de ácido manurónico que pueden ser acetilados y su epímero el ácido gulurónico. Por otro lado, el PHB, un biopolíester intracelular compuesto por unidades de 3-hidroxi-butirato, se emplea como un sustituto de plásticos convencionales derivados de la industria petroquímica (Castillo et al., 2020).

A pesar de que no ha sido descrita la producción de síntesis de melanina en *A. vinelandii* en la plataforma “Kegg” (Kyoto Encyclopedia of Genes and Genomes: <https://www.genome.jp/kegg/>) se presenta la vía metabólica de la tirosina en *A. vinelandii* DJ

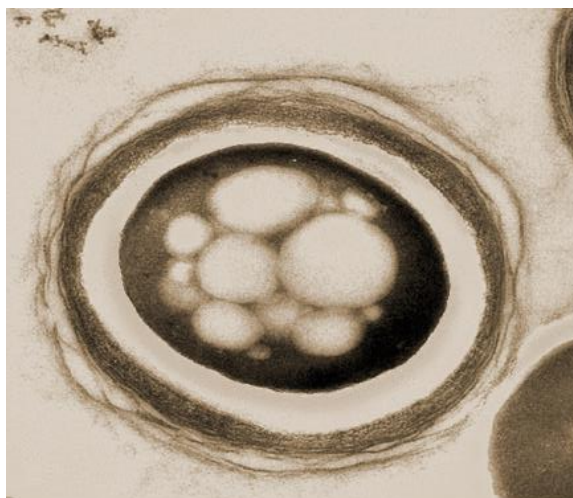


Figura 1. Micrografía de quiste de *Azotobacter vinelandii*. Gránulos de PHB en el interior rodeados por capaz que contienen alginato (tomada de Yañez 2020).

y hace referencia a la posible síntesis de melanina de tipo eumelanina con la participación de enzimas tirosinasas (Anexo 1).

1.2. *Azotobacter chroococcum*

En 1901, el microbiólogo holandés M.W. Beijerinck, descubrió un organismo aeróbico capaz de fijar nitrógeno molecular al que se le dio el nombre de *Azotobacter chroococcum*. *Azotobacter chroococcum* NCIMB 8003 (Ac-8003) aislada en California EE.

UU. en 1934. Es una de las 19 cepas aisladas de *A. chroococcum* caracterizadas a detalle por Thomson y Skerman. Hoy en día, es probablemente la mejor caracterizada que se mantiene en colecciones de cultivo en todo el mundo. Se ha utilizado para el estudio de nitrogenasas con molibdeno y vanadio, los mecanismos fisiológicos que protegen a la nitrogenasa del daño por O₂, nitrato y nitrito reductasas, glutamina sintasa, entre otras enzimas (Robson et al., 2015).

Los miembros de la especie *A. chroococcum* son comunes en suelos neutros o alcalinos, mientras que *A. vinelandii* es menos común en suelos. Son bacterias gram negativas, γ -proteobacteria, aerobias. Son pleomórficas variando su morfología desde bacilos hasta cocos de 1.5-2 μ m, pudiéndose encontrar como células individuales, como pares o formando agregados irregulares. Macroscópicamente presenta colonias mucosas opacas, inicialmente el color del pigmento es claro y brillante, pero a medida que la colonia se desarrolla se torna de color café oscuro. El tamaño de su genoma es de 4.89 Mb (Saribay, 2003).

A partir de números estudios se ha dado a conocer su potencial para estimular el crecimiento de plantas, ya sea a través de producción de sustancias de crecimiento, mineralización o aporte de nitrógeno fijo. *A. chroococcum* puede utilizar diferentes fuentes de nitrógeno inorgánico como amonio, nitrato o dinitrato para asimilar el nitrógeno. Además, esta especie tiene la capacidad de biodegradar compuestos tóxicos y contaminantes, presenta efecto antagónico con patógenos como hongos y nematodos. También metaboliza compuestos fenólicos con alta carga contaminante para el ambiente y puede degradar plaguicidas clorados contaminantes por medio de enzimas deshalogenasas, dioxigenasas e hidroxilasas (Jnawali et al., 2015).

Una característica relevante en *A. chroococcum* es su capacidad para sintetizar melanina. Cultivada en condiciones aerobias produce una gran cantidad de melaninas sin requerir que el medio de cultivo contenga alguna molécula precursora como tirosina o dihidroxifenilalanina, aunque, la presencia de cobre en el medio representa un factor estimulante para estas enzimas. Es posible que *A. chroococcum* emplee la melanogénesis

para mejorar la utilización del nitrógeno y mantener las condiciones reductoras necesarias para unir nitrógeno atmosférico (Plonka & Grabacka, 2006).

2. Melanina

La melanina es un biopolímero derivado de compuestos fenólicos que produce una gama de colores en presencia de oxígeno. Son compuestos hidrofóbicos con carga negativa (Lagunas Muñoz et al., 2006). Presenta diferentes funciones biológicas como la protección contra la radiación UV de la luz solar y la transducción de energía. En los microorganismos como bacterias y hongos, la melanina protege contra metales pesados y agentes oxidativos, a nivel bioquímico funciona como defensa del huésped contra microorganismos invasores ante estrés oxidativo a través de la generación de especies reactivas de oxígeno (ROS, por sus siglas en inglés). La caracterización de las melaninas ha revelado propiedades con aplicaciones industriales potenciales ya que pueden actuar como absorbentes de luz ultravioleta, semiconductores amorfos, intercambiadores de cationes, absorbentes de rayos X y rayos γ (Chavez Béjar et al., 2013). Por lo tanto, existe un interés en el desarrollo de procesos biotecnológicos para la producción de este polímero. La producción de melanina se ha logrado en cultivos líquidos de melanogénicos naturales y con microorganismos recombinantes.

Se ha clasificado a este biopolímero en tres tipos de acuerdo con su naturaleza química, vía de síntesis y las enzimas específicas que participan. La eumelanina que genera pigmentos negros-marrones y la feomelanina generadora de pigmentos amarillos-rojos son los más

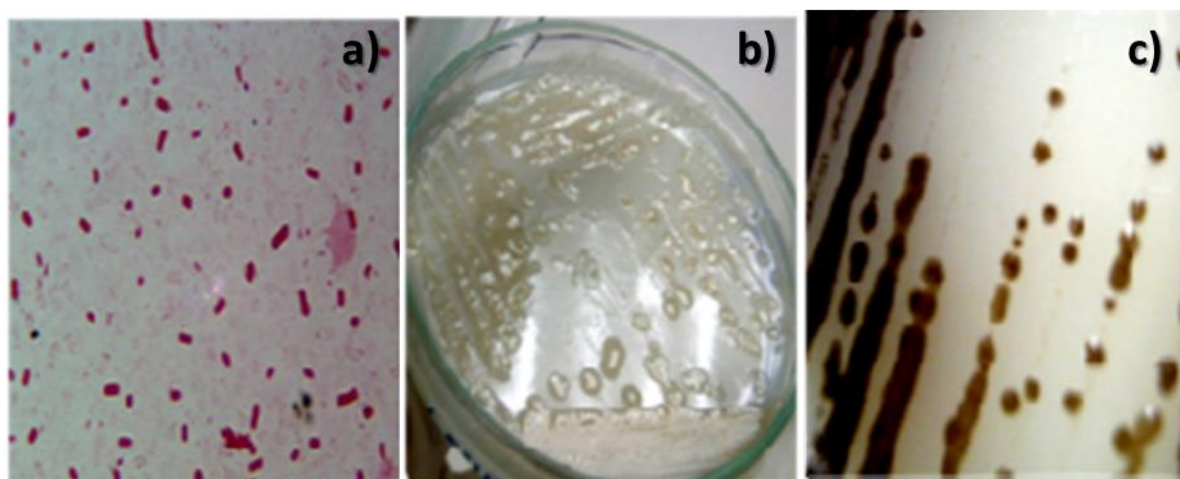


Figura 2. *Azotobacter chroococcum*. a) Bacilos Gram negativos. b) Colonias aisladas con tonalidad blanquecina y mucoides. c) Colonias aisladas con tonalidad marrón oscuro producto del paso del tiempo de la siembra asociado a la producción de melanina (Jimenez, 2011).

abundantes en piel y cabello de humanos y animales. La diferencia entre estos tipos de melanina radica en el contenido de azufre que presenta la feomelanina. El tercer tipo, la alomelanina, comprende un amplio grupo heterogéneo de pigmentos que generalmente son negros y se generan a partir de diferentes precursores en plantas, bacterias y otros microorganismos (Sanchez Amat et al., 2010).

2.1. Vías de síntesis

Cada uno de los tipos de melanina son sintetizados a partir de diferentes vías, que tienen en común la formación de quinonas que serán polimerizadas para la formación de melanina (figura2). La síntesis de alomelanina tiene como precursor el malonil-CoA que por medio de la poliketido sintasas se da una oxidación/polimerización de 1,3,6,8-tetrahidroinaftaleno (THN) y posteriormente 1,4,6,7,9,12-hexahidroxiperilen-3,10-quinona (HPQ) que será polimerizado como etapa final. Además, la alomelanina puede ser sintetizada a partir de tirosina, que por desaminación puede producir los precursores 4-hidroxifenilpiruvato (4-HPP) o 4-hidroxifenilacetato (4-HPA) que por acción de enzimas como aspartato-tirosina aminotransferasa o tirosinasa se oxidan a los compuestos aromáticos dihidroxilados homogentisato (HMG) y 3,4-Dihidroxifenilacetato (HPC) que serán oxidados a benzoquinona para su posterior polimerización. La tirosina puede seguir otra vía en la cual por acción de tirosinasas o lacasas se oxida a L-3,4-dihidroxifenilalanina (L-DOPA) y

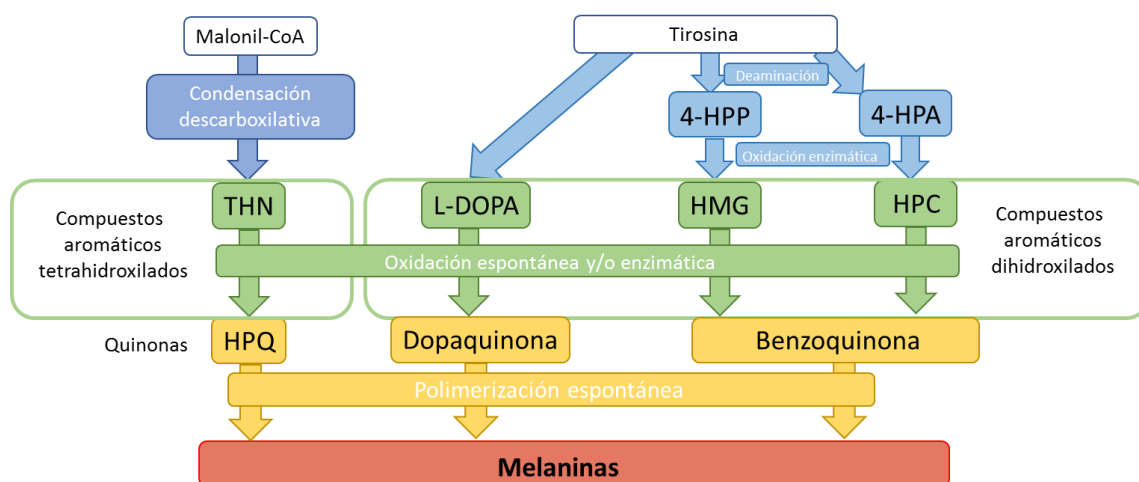


Figura 3. Vías de síntesis de melanina (Pavan, López, & Pettinari, 2019).

dopaquinona que presentará posteriormente una ciclación a 5,6-dihidroxiindol (DHI) polimerizándose y produciendo eumelaninas o fenilmelaninas (Pavan et al., 2019).

2.2. Enzimas involucradas en la síntesis de melanina

A pesar de la complejidad y las diferentes vías que conducen a la formación de la melanina, la mayoría de los pasos son oxidaciones espontáneas y la regulación enzimática solo ocurre en las etapas iniciales de las vías. Dado que los precursores iniciales de este polímero son los fenoles, las enzimas clave se denominaron genéricamente polifenol oxidasas (PPO, por sus siglas en inglés). Casi todas estas PPO relacionadas con la melanogénesis tienen un sitio activo caracterizado por un par de iones de cobre (CuA y CuB) coordinados por seis residuos de histidina (Banerjee et al., 2014).

2.2.1. Tirosinasas y catecol oxidasas

Las tirosinasas (EC 1.14.18.1) juegan un papel fundamental en la síntesis de melanina debido a que regulan directamente la cantidad de melanina producida, mientras que otras enzimas solo modifican el tipo de melanina que es sintetizada en la ruta bioquímica del pigmento (Casañola Martín et al., 2013).

La tirosinasa cataliza la *o*-hidroxilación de tirosina en DOPA, así como la oxidación de DOPA a dopaquinona que por una serie de reacciones enzimáticas y no enzimáticas se puede transformar en pigmentos de melanina. La síntesis de *o*-difenoles es considerada una actividad catalítica muy valiosa generando un gran interés en la industria ambiental, por ejemplo, es utilizada para la detoxificación de compuestos fenólicos contenidos en aguas residuales y suelos contaminados. En la industria farmacéutica se utiliza para la producción de levodopa, medicamento considerado como el más eficaz en el tratamiento de la enfermedad de Parkinson (Zaidi et al., 2014).

Se conoce como periodo de latencia a la etapa de transición de un microorganismo cuando es colocado en una nueva condición de crecimiento produciendo enzimas necesarias para favorecer su crecimiento, algunas tirosinasas muestran periodos de latencia cortos (varios segundos), mientras que otras muestran periodos largos (horas a días), en dicho caso

las tirosinasas pueden considerarse enzimas con actividad de o-difenol oxidasa denominándose catecol oxidasas (EC 1.10.3.1). Estas son habituales en las plantas, por lo que, son responsables del oscurecimiento de las frutas después de daño o maduración.

Tanto la catecol oxidasa como las enzimas tirosinasas son capaces de formar melanina, pero el precursor inicial es, respectivamente, un o-difenol o monofenol.

La principal aplicación de las tirosinasas es como biosensor para la determinación precisa y sensible de fenoles en diferentes industrias que van desde vinos hasta residuos que contienen fenoles contaminantes (Mazinani & Asgharzadeh, 2014).

2.2.2. Lacasas

Denominadas colectivamente dentro de las enzimas fenol oxidasas (PO), y pertenecientes al grupo de la familia de las proteínas multi cobre, las lacasas (EC 1.10.3.2) son capaces de oxidar fenoles, principalmente p-difenoles y monofenoles sustituidos con un grupo metoxilo. (Herter 2011). Debido a que, durante la oxidación de las lacasas, el oxígeno es reducido mediante la transferencia de cuatro electrones al H₂O, se pueden distinguir de otros miembros de las polifenol oxidasas (PPO) como lo son las catecol oxidasas (E.C.1.10.3.1) y monofenol monooxigenasas o tirosinasas (E.C.1.14.18.1) (Singh et al., 2011).

Las lacasas también contienen cuatro dominios de unión a cobre ricos en histidina, que coordinan los tipos 1 a 3 de átomos de cobre que difieren en propiedades espectroscópicas (Sharma et al., 2006).

Se sabe que las lacasas tienen la capacidad de degradar la lignina por lo que la convierte en una enzima de interés para la industria del papel, así como, la degradación de material lignocelulósico para la generación de biocombustibles. Otra aplicación es como biocatalizadores en la modificación y generación de nuevos compuestos con actividades biológicas, por ejemplo, modificaciones de los antibióticos beta-lactámicos (Sanchez Amat et al., 2010).

Con base en las funciones fisiológicas de las PO procariotas, se ha sugerido a lo largo del tiempo una relación directa entre la formación de pigmentos y alteraciones morfológicas de la célula. En este sentido, se sabe que algunas especies del género *Azotobacter* sufren alteraciones morfológicas en el proceso de enquistamiento, que es inducible por condiciones ambientales adversas. En miembros melanogénicos del género *Azotobacter* se ha sugerido la presencia de una PO (Sanchez Amat et al., 2010).

2.3. Aplicaciones biotecnológicas

Las melaninas presentan una amplia gama de aplicaciones biotecnológicas, por ejemplo, en la industria farmacéutica y cosmética son de interés ya que presentan propiedades fotoprotectoras, antioxidantes y antimicrobianas. Las eumelaninas son las que presentan mejores propiedades fotoprotectoras por lo que son las que principalmente se utilizan en la formulación de lociones de protección solar, las fuentes habituales de estas melaninas son sintéticas, dopa-melanina, pelo de mamífero y tinta de sepia. Recientemente, se está implementando el uso de eumelanina de origen bacteriano en filtros fotoprotectores (Lagunas Muñoz et al., 2006).

Gracias a su estructura orgánica polimérica presentan cierta capacidad de conversión de energía, así como propiedades semiconductoras que les proporciona un posible papel en fotoelectroquímica. En este sentido se han hecho algunos intentos para usarlos como componentes de convertidores de energía de paneles solares (Sanchez Amat et al., 2010).

Desde 1990, se reportó la incorporación de dos genes capaces de codificar una tirosinasa de la bacteria *Streptomyces antibioticus* en *Escherichia coli*, que adquirió la capacidad de transformar tirosina en el medio de cultivo a eumelanina. Posteriormente se mejoró la producción de melanina introduciendo en *E. coli* el gen *mela*, de la bacteria melanogénica *Rhizobium etli* que codifica para una enzima tirosinasa lo que la convirtió en una bacteria productora de melanina con alto rendimiento (producción de 6 g/L de eumelanina) que se evalúa como componente de circuitos electrónicos y de bloqueador de luz UV (Martínez et al., 2021).

También son de interés en la industria agrícola pues se pueden utilizar como agentes biocontroladores de plagas por su poder antimicrobiano y como detoxificadores de suelos y aguas contaminados con metales pesados. Probablemente su aplicación más reciente sea la nanotecnología para la síntesis de películas cristalinas, flexibles y termoestables de melanina que pueden ser utilizados en biomedicina (Lagunas Muñoz et al., 2006).

3. Secuenciación masiva y bioinformática

El desarrollo de secuenciadores masivos que permiten leer en paralelo de millones a miles de millones de secuencias o fragmentos de ADN ha revolucionado las áreas biológicas, pasando de un ámbito exclusivamente de laboratorio a uno computacional con la aplicación de la bioinformática posibilitando el llevar a cabo estudios de genómica comparativa y el descubrimiento de genes o variantes de interés. La bioinformática aplica las matemáticas, la estadística y la computación al estudio biológico (Edwards & Holt, 2013).

3.1. Secuenciación “Illumina”

Después de casi 30 años de la aparición de la secuenciación Sanger, surgieron nuevos desarrollos tecnológicos para resolver las limitantes de esta técnica, entre ellas lo poco costeable para secuenciar genomas completos (Tosco Herrera, 2019). El método Sanger implicaba la interrupción del proceso de síntesis de la polimerasa con una mezcla de nucleótidos y nucleósidos modificados de tal manera que la reacción no pudiera continuar la polimerización una vez que dicho nucleótido modificado haya sido incorporado, por lo que se le llamó secuenciación por terminación; la detección de la base se llevaba a cabo en una columna de Sephadex (gel dextrano reticulado) con una marca radioactiva y posteriormente fluorescente.

A partir del año 2000, con la comercialización de tecnologías de secuenciación masiva se presentó un auge de la “Era Genómica” aumentando de manera exponencial los genes y genomas en las bases de datos debido al volumen masivo de información que producen en un tiempo reducido (Swarup & Jong, 2021).

La plataforma Illumina implica la fragmentación aleatoria de ADN y la ligación de los fragmentos de ADN mediante adaptadores presentes en una lámina. Se basa en una secuenciación por síntesis de la cadena complementaria, amplificación en puente y detección a partir de fluorescencia de nucleótidos. Consiste en una máquina con un microscopio y cámara de gran resolución que registra por medio de imágenes, como se incorporan los nucleótidos modificados en cada agrupación de secuencias idénticas inmovilizadas sobre la lámina, que representa a un fragmento de la biblioteca de ADN. La máquina divide el área de la laminilla en cuadrantes de los cuales registra las posiciones de cada grupo por lo que registra una foto por cada cuadrante y esto por cada ciclo. Se generan miles de fotografías digitales de alta resolución que la máquina utiliza para elucidar la secuencia de cada fragmento. Además, tiene una gama de secuenciadores que van desde la plataforma MiSeq, el cual es un secuenciador de “bajo rendimiento” (20 millones de secuencias) hasta conjuntos de secuenciadores como lo es el HiSeq TenX que consta de mínimo 10 máquinas capaces de secuenciar 180 000 genomas humanos en un año.

El proceso de construcción de la biblioteca genómica a partir de ADN total es similar al de otras tecnologías, con la diferencia de la selección del tamaño del fragmento (máximo 1,000 bases) y que la cantidad de material genético necesario es de unos cuantos nanogramos.

Las etapas que constituyen esta metodología son:

- Adhesión de adaptadores a cada uno de los extremos de cada fragmento de cadena sencilla. Los fragmentos se adhieren a una placa o laminilla de vidrio la cual tiene en su superficie una serie de adaptadores complementarios a los de los extremos de las moléculas a secuenciar. Estos adaptadores están esparcidos por la laminilla con una separación suficiente para poder formar las agrupaciones.
- Una vez adheridos los fragmentos a la laminilla, se comienza el proceso de amplificación que requiere que la molécula adherida forme un “puente” para realizar la síntesis de la cadena complementaria. Después de varios ciclos, se obtiene una agrupación de moléculas que provienen de un mismo templado (alrededor de 1,000 por agrupación). Ya formados los agrupamientos se remueven las cadenas

complementarias de tal forma que todas las moléculas vuelvan a ser de cadena sencilla.

- Posteriormente comienza la reacción de secuenciación en la cual se incorporan nucleótidos modificados con una molécula fluorescente que emite un color determinado para diferenciar cada uno de los cuatro nucleótidos que conforman el ADN. El fluoróforo es excitado por un rayo láser y se toma una foto a la laminilla para registrar los grupos donde fue incorporado cada nucleótido. Debido a las modificaciones de los nucleótidos, la reacción de polimerización solo incorpora un nucleótido y solo hasta después de registrar el color de la molécula fluorescente, se remueve el bloqueo del nucleótido para que se pueda incorporar uno nuevo en el siguiente ciclo de polimerización (Hernández et al., 2020).

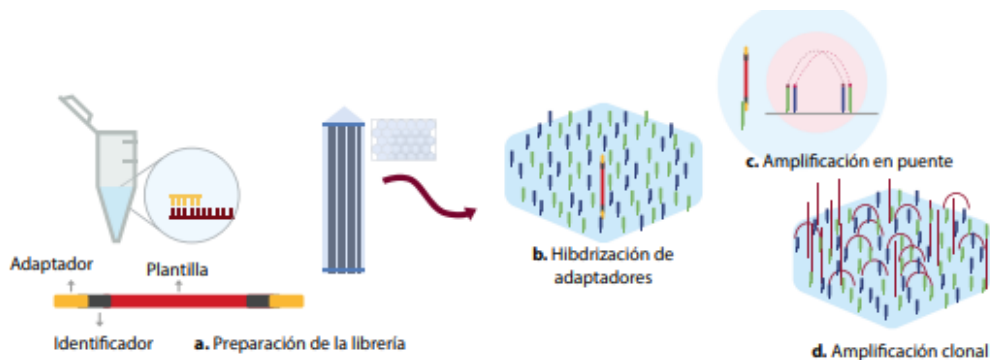


Figura 4. Secuenciación por síntesis Illumina (Swarup & Jong, 2021).

Illumina ofrece un gran número de ventajas, pero sobre todo un gran número de aplicaciones en el área de la Genómica. Se puede utilizar para estudiar genes o regiones particulares de un genoma, genomas completos de cualquier organismo, ADN complementario obtenido de ARN como producto de la expresión genética, análisis de muestras metagenómicas, modificaciones epigenéticas, sitios de unión al ADN, entre otras (Swarup & Jong, 2021).

3.2. Secuenciación “Oxford nanopore”

La tercera generación de tecnologías de secuenciación surge de la necesidad constante de mejorar o evitar ciertas desventajas que presentan las tecnologías previas, permitiendo además tener una mayor cantidad de opciones de métodos para llevar a cabo la secuenciación y de esta manera poder optar por la que mejor se adapte a las necesidades. La tecnología de Illumina realiza secuencias de fragmentos de tamaño corto (hasta 300 pb) y la amplificación clonal a la que se someten estos fragmentos previamente a ser secuenciados introduce sesgos considerables. Para evitar estos problemas, la tercera generación se ha desarrollado con nuevas técnicas mucho más sensibles de detección que evitan la amplificación clonal necesaria en la segunda generación, sumando además la posibilidad de secuenciar fragmentos considerablemente más largos (Camargo Mancipe et al., 2020).

Dentro de sus características principales se encuentra la secuenciación de ADN, ARN y proteínas, el funcionamiento en tiempo real, secuenciación en moléculas individuales, longitudes de lecturas en el orden de decenas de miles de bases, se ha podido solucionar uno de los principales problemas de las lecturas largas el cuál era la calidad debido a su alta tasa de error (~10%).

La tecnología de secuenciación Oxford nanopore comenzó en 1995 donde se utilizó el canal proteico de la alfa-hemolisina de *Staphylococcus aureus*, embebido en una membrana lipídica, donde se midieron los cambios en el voltaje provocados por el paso de las moléculas de ADN y ARN por el poro que formaba el canal. Estos cambios de voltaje son dependientes del cambio conformacional provocado por cada nucleótido y que producen una señal diferente para cada uno (Tosco Herrera, 2019).

Básicamente su fundamento es que a partir de ADN de doble cadena y de alto peso molecular, o bien ARN con buena integridad, se hacen pasar (prácticamente sin fragmentación) una de las cadenas por el interior del canal formado por el poro proteico de dimensiones nanométricas (~10nm). Sin embargo, debido a la velocidad a la cual pasa la molécula, en realidad no se registran los cambios de voltaje generados por un solo nucleótido sino por regiones de 4 a 5 nucleótidos. Debido a los problemas que esto ocasiona, es necesario registrar los cambios de voltaje generados por la cadena complementaria. De esta manera, se

pueden cotejar los registros y elucidar de una manera más exacta, la secuencia de cada fragmento.



MinION es el secuenciador portátil perteneciente a “Oxford Nanopore Technologies” el cuál se comercializa desde 2013 (Guillen, 2019).

Algunas de las diferencias con respecto a la tecnología Illumina son: lecturas de grandes longitudes (3,000 hasta 50,000 bases), permiten secuenciar otras moléculas que no son ácidos nucleicos (proteínas) y teóricamente de cualquier molécula o metabolito, permite realizar estudios de epigenética y regulación en proteínas gracias al registro de moléculas modificadas como bases metiladas y aminoácidos fosforilados o glicosilados. Sin embargo, también cuenta con algunas desventajas, principalmente la tasa de error (~10%) que es significativamente más alta que las encontradas en las otras tecnologías (~1-0.1%) (Tosco Herrera, 2019).

En la figura 5 se muestra un ejemplo del archivo con extensión .fastq con sus características el cuál es producto de la secuenciación masiva y que permite continuar la realización del ensamblado genómico.

3.3. Ensamblaje genómico

Consiste en tomar una colección de secuencias y reconstruir con estas las secuencias del genoma del cual fueron originadas. Una señal de calidad en las secuencias y en el

Figura 5. Características del formato con extensión .fastq

ensamble final es poder localizar de vuelta la mayor cantidad de las lecturas de entrada sobre el resultado del ensamblaje, además de ser localizados en la orientación y distancia esperada. Otra señal de calidad del ensamble es la cantidad de fragmentos en los que queda, en la mayoría de los ensambladores se obtienen finalmente fragmentos que se denominan “*contigs*” (regiones contiguas de lecturas), en otros se llega al nivel de “*scaffolds*” o “*supercontigs*” (conjunto de contigs que se colocan en orden y sentido entre ellos) (Hernández et al., 2020).

En el proceso de reconstrucción de un genoma se puede llegar a obtener varios armados, para poder seleccionar uno existen algunas variables a considerar: la medida más utilizada es el medidor “N50” que representa el menor tamaño de “*contigs*” en el que se encuentra contenido el 50% del ensamble. Otra medida es ver cómo se comportaría el mapeado de las lecturas utilizadas sobre los ensamblados obtenidos. Finalmente es posible medir cuán completo está el ensamblado a través de la identificación de genes altamente conservados (Edwards & Holt, 2013).

3.4. Anotaciones genómicas

Es el proceso de búsqueda de genes que también puede incluir la identificación de ARN ribosomal y de transferencia codificados en el genoma. Existen herramientas automatizadas disponibles en la web como lo es RAST, que facilita las anotaciones. Este tipo de herramientas toman como entrada los contigs ordenados identificando marcos de lectura abiertos que probablemente sean genes y utiliza una serie de técnicas de subsistemas para compararlos con una base de datos sofisticada de genes y secuencias de ARN, produciendo una anotación de alta calidad del ensamblaje (Guillen, 2019).

3.5. Análisis de alineamiento y comparación de genomas

Este se basa en la relación de homología que se presente entre los genomas a comparar. Se dice que existe homología de secuencias cuando al comparar las secuencias de dos o más proteínas o ácidos nucleicos, estas son similares entre si debido a que representan un mismo origen evolutivo. El alineamiento de genomas ayuda a identificar cambios

evolutivos en el ADN mediante la comparación de regiones homólogas de secuencia y a generar hipótesis relacionadas con el origen evolutivo de los genes, entre otras (Swarup & Jong, 2021).

Antecedentes

Diversos microorganismos son productores de melanina, principalmente bacterias marinas y del suelo como especies de *Azospirillum*, algunas cepas de *Rhizobium*, *Streptomyces*, *Klebsiellas*, *Vibrios*, entre otros. Poco se ha estudiado del género *Azotobacter* respecto a la biosíntesis de melanina. Respecto a las especies que conforman el género *Azotobacter* únicamente se ha estudiado la capacidad de síntesis de este polímero en *A. nigricans*, *A. salinestris* y *A. chroococcum*.

Shivprasad y Page estudiaron en 1989 la formación de catecol para llevar a cabo la melanización en *A. chroococcum* dependiente de sodio utilizando la cepa *A. chroococcum* 184, la cual, produjo pigmentos negros cuando se cultivó en condiciones aerobias y de fijación de nitrógeno. El catecol producido por esta cepa desapareció tras la melanización sugiriendo que este es el sustrato para la síntesis de melanina (Shivprasad & Page, 1989).

Teniendo conocimiento de la producción de melanina a partir de catecol en *A. chroococcum*, Page y Shivprasad trabajaron con la cepa *A. salinestris* 184 para estudiar la absorción de hierro a fin de analizar funciones biológicas. Utilizando una variante blanca y

una variante negra, se observó que la variante negra, que debe su color a la síntesis de melanina, unía una mayor cantidad de hierro deduciendo que la melanina sirve como una trampa de hierro para probablemente proteger a las células del daño oxidativo (Page & Shivprasad, 1995).

Más adelante, se investigaron las condiciones para la producción de melanina por *Azotobacter chroococcum* y algunas de sus propiedades. Gospodaryo y Lushchak llegaron a los resultados que indican que la excreción de melanina por la cepa de *A. chroococcum* solo ocurre cuando esta cultivada en medio Ashby con ácido benzoico, pero no cuando se utiliza manitol como fuente de carbono en el medio. El metabolismo de ácido benzoico en bacterias conduce a la formación de catecol (Gospodaryov & Lushchak, 2011).

Banerjee y colaboradores llevaron a cabo la caracterización espectroscópica de la melanina sintetizada en *A. chroococcum* fijadora de nitrógeno, encontrando claros indicios de la presencia de hidroxindoles en su composición que refieren que la melanina sintetizada en *A. chroococcum* es del tipo eumelanina de acuerdo con las estructuras propuestas y que de acuerdo a las vías de síntesis antes analizadas el precursor es la tirosina siguiendo la vía se conoce como “vía del homogentisato” (Banerjee et al., 2014).

A partir de la secuenciación del genoma de *A. chroococcum* 8003, se sabe que contiene cuatro genes cromosomales codificantes para una enzima polifenol oxidasa multicobre, dentro de los cuales dos de estos genes *cumA* (Achr_10550, 30620) son ortólogos con un alto grado de identidad en dos genes de *A. vinelandii* (Avin_39690, 11950). Como se mencionó anteriormente estas enzimas polifenol oxidasas están directamente involucradas en las vías de síntesis de melaninas dando un parteaguas de que en *A. vinelandii* estas se encuentran presentes también (Robson et al., 2015).

Recientemente en el año 2019, en el Laboratorio de Genética Molecular Microbiana del Centro de Investigaciones Microbiológicas de la Benemérita Universidad Autónoma de Puebla se llevó a cabo un trabajo experimental en el que se aislaron cepas silvestres de *Azotobacter* spp. obtenidas a partir de muestras de suelo, agua y rizosfera de algunas plantas de diferentes zonas ambientales de la Reserva de la Biósfera de Tehuacán-Cuicatlán. Dentro

de estas muestras fue posible aislar tres cepas que presentaron la producción de pigmentos, y se identificaron como S4T, SM3 y S6R.

Con base en la secuenciación del gen 16S rDNA y marcadores genéticos (tabla 1) para identificar género y especie, se determinó que las cepas SM3 y S6R corresponden a la especie *A. chroococcum* en la cual está estudiada la síntesis de melanina y presencia de enzimas polifenol oxidasas. La cepa nombrada como S4T se determinó que corresponde a la especie *A. vinelandii* de la cual no existe investigación ni evidencia de su capacidad de síntesis de melanina (García Gonzalez, 2019).

Tabla 1. Genes utilizados como marcadores moleculares para la caracterización de las cepas aisladas.

Genes	Descripción
<i>arsB</i>	Policétido sintasa tipo III
<i>alyA1</i>	Alginato liasa involucrada en la síntesis de alginato en <i>A. vinelandii</i>
<i>rpoS</i>	Factor sigma y modulador maestro en la regulación de la fase estacionaria
<i>ompA</i>	Gen que codifica para proteína principal de la membrana externa

Material biológico

Tabla 2. Cepas utilizadas en este trabajo

Cepa	Características	Referencia
<i>Azotobacter vinelandii</i> AEIV	Cepa silvestre utilizada como referencia experimental. También llamada cepa E.	Larsen & Haug, 1971
S4T	Muestra del suelo de la Reserva de la Biosfera de Tehuacán-Cuicatlán.	García, 2019
SM3	Muestra de rizosfera de la Reserva de la Biosfera de Tehuacán-Cuicatlán.	García, 2019
S6R	Muestra de suelo del río de la Reserva de la Biosfera de Tehuacán-Cuicatlán.	García, 2019

Justificación

A lo largo del tiempo se ha identificado que diversos géneros de la familia *Pseudomonadaceae* tienen la capacidad de sintetizar melanina, pigmento que a nivel biológico protege a los microorganismos que lo producen del daño oxidativo y que presenta diversas aplicaciones biotecnológicas. Sin embargo, del género *Azotobacter* el estudio de biosíntesis de melanina se limita en mayor medida a *A. salinestrus* y *A. chroococcum* seguidas por *A. nigricans*. Recientemente se aislaron cepas productoras de melanina que de acuerdo con los análisis del gen 16S y la utilización de marcadores específicos se identificaron como *A. chroococcum* y *A. vinelandii* de la cual se desconoce su capacidad de biosíntesis de melanina por lo que el estudio genómico de éstas nos permitirá identificar los genes posiblemente involucrados en la síntesis de melaninas y con base en las posibles rutas de síntesis de éstas y la presencia de enzimas específicas proponer el tipo de compuesto sintetizado. Además, *A. vinelandii* es una especie ampliamente estudiada a nivel biotecnológico de la que ya se conocen sus condiciones de crecimiento en cultivo y en lote, por lo que el poder identificar cepas de dicho género con la capacidad de producir melanina facilitaría la producción a nivel de fermentador del polímero.

Hipótesis

Las cepas de *Azotobacter* spp. aisladas productoras de pigmento poseen enzimas y rutas metabólicas capaces de producir melaninas.

Objetivos

General

Estudiar a nivel genómico las vías de síntesis de melaninas en *Azotobacter* spp.

Específicos

1. Identificar genes posiblemente implicados en la síntesis de melaninas en genomas secuenciados y publicados de cepas del género *Azotobacter*.
2. Secuenciar los genomas de las cepas de *Azotobacter* spp. recientemente aisladas productoras del pigmento.
3. Identificar los genes posiblemente involucrados en la síntesis de melaninas en el genoma de las cepas secuenciadas y proponer las probables vías de síntesis y el tipo de compuesto sintetizado.

Resultados y discusión

1. Genomas secuenciados de diferentes especies de *Azotobacter*.

Tabla 3. Cepas de diferentes especies de *Azotobacter* secuenciadas disponibles en GeneBank-NCBI.

<i>Azotobacter vinelandii</i>	DJ	Circularizado
	CA	Circularizado
	CA6	Circularizado
	PRJEN17910	Scaffold
<i>Azotobacter chroococcum</i>	8003	Circularizado
	PRJNA523901	Scaffold
	353	Circularizado
	PRJNA520333	Scaffold
	PRJNA523901	Scaffold
	PRJNA523901.2	Scaffold
	PRJNA539873	Contig
	PRJNA590128	Circularizado
	PRJNA610299	Contig
	PRJNA523901	Scaffold
<i>Azotobacter salinestris</i>	PRJEB16790	Contig
	PRJEB16809	Scaffold
<i>Azotobacter beijerinckii</i>	PRJEB16980	Scaffold
	PRJEB17088	Scaffold
	PRJEB17433	Scaffold

Hasta la fecha en la que fue descargado el archivo se encuentran disponibles 22 genomas secuenciados del género *Azotobacter* spp. de los cuales 7 ya están circularizados,

10 se presentan como scaffolds y 5 como contigs. Las especies que conforman este grupo son *Azotobacter vinelandii*, *Azotobacter chroococcum*, *Azotobacter beijerinckii* y *Azotobacter salinestris*, en la tabla 3 se muestran las diferentes cepas secuenciadas que en total son 19 debido a que 3 de estos genomas no están anotados.

2. Alineamiento de secuencias de proteínas (BLAST protein)

2.1 Análisis tipo BLAST en especies de la familia *Pseudomonadaceae*

Teniendo conocimiento de las enzimas que se ha reportado participan en la biosíntesis de melanina de acuerdo con la previa investigación bibliográfica, se realizó una búsqueda de estas en el sitio web del *National Center for Biotechnology Information* (<https://www.ncbi.nlm.nih.gov>) en los genomas que se encuentran disponibles de los géneros *Azotobacter* y *Pseudomonas*. Como resultado se obtuvieron 121 secuencias de aminoácidos con las que se llevó a cabo el BLAST de proteínas por línea de comando que comparó estas secuencias contra la base de datos de secuencias del proteoma de *Azotobacter vinelandii* DJ y CA. El filtrado de identidad y cobertura fue de 75 y 97% respectivamente.

Como resultado se obtuvieron nueve secuencias que corresponden a las proteínas que se enlistan a continuación:

- Policétido sintasa deshidratasa
- Policétido sintasa tipo III
- Policétido sintasa tipo I
- Fenilalanina-4-monooxigenasa
- Aspartato aromático tirosina aminotransferasa
- 4-hidroxifenilpiruvato dioxigenasa
- Homogentisato 1,2 dioxigenasa
- Fosfoenolpiruvato sintasa
- D-tirosil-tRNA (Tir) deasilasa

A partir de estos resultados se analizó la función de estas proteínas y su relación en la biosíntesis de melaninas encontrando que las primeras tres, Policétido sintasa deshidratasa, Policétido sintasa tipo III y Policétido sintasa tipo I participan en la condensación

descarboxilativa de malonil-CoA en 1,3,6,8-tetrahidroxinaftaleno (THN) compuesto aromático precursor de quinonas que por polimerización sintetizan melanina.

Las siguientes cuatro proteínas se pueden agrupar como parte de la vía de síntesis de melaninas a partir de tirosina. La fenilalanina-4-monooxigenasa sintetiza la tirosina a partir de fenilalanina la cual es el precursor inicial de la biosíntesis de melanina. La Aspartato aromático tirosina aminotransferasa promueve la conversión de tirosina en 4-hidroxifenilpiruvato el cual es un compuesto intermediario en la biosíntesis de melanina. La 4-hidroxifenilpiruvato dioxigenasa lleva a cabo la conversión de 4-hidroxifenilpiruvato en homogentisato quien se oxida para dar lugar a la benzoquinona y el homogentisato 1,2 dioxigenasa lleva a cabo la degradación de homogentisato, su acumulación, autooxidación espontánea y polimerización de homogentisato conduce a la síntesis de melanina.

Fosfoenolpiruvato sintasa y D-tirosil-tRNA (Tir) deasilasa son proteínas que no tienen relación con la biosíntesis de melanina.

2.2 Análisis de genomas de especies no relacionadas filogenéticamente con la familia Pseudomonadaceae

Otro aspecto importante que considerar en este análisis es la transferencia horizontal de genes por lo que se realizó una búsqueda de las enzimas mostradas en el apartado 2.1 en la plataforma NCBI en los proteomas de *Acinetobacter baumannii*, *Aeromonas salmonicidas*, *Vibrio natriegens*, *Vibrio cholerae*, *Streptomyces lannensis*, *Klebsiella pneumoniae*, *Shewanella oneidensis*, *Burkholderia cepacia*, *Serratia marcescens*, *Mycobacterium leprae*, *Proteus mirabilis*, *Bacillus thuringiensis* y *Bacillus subtilis* que son bacterias no relacionadas filogenéticamente y que de acuerdo con la bibliografía son productoras de melanina. De dicha búsqueda se obtuvieron 19 secuencias de aminoácidos que fueron comparados contra la base de datos que se mencionó en la sección 2.1. El filtrado de identidad y cobertura fue de 50 y 75% respectivamente.

No se obtuvo ninguna proteína como resultado de este análisis.

2.3 Proteoma del género *Azotobacter*

Se llevó a cabo un análisis BLAST de proteínas con los proteomas de *Azotobacter* que se presentan en la tabla 3. Los parámetros de identidad y cobertura fueron de 45 y 50% respectivamente.

Como resultado se obtuvieron nueve secuencias que corresponden a las proteínas que se enlistan en la sección 2.1. Se realizó un análisis para conocer cuantas, y cuáles de las enzimas de interés se encuentran en los proteomas de cada uno de los 19 analizados, estando presentes las nueve proteínas antes mencionadas en la sección 2.1 en los 19 proteomas.

3. Contexto genómico y sintenia

Para el análisis del contexto genómico se utilizó como herramienta el visualizador BLAST-XYplot viewer (<http://www.blast-xyplot-viewer.icuap.buap.mx/>) que realizó un BLAST del grupo deseado de proteínas contra los genomas de *Azotobacter vinelandii* DJ, CA y CA6 obteniendo un gráfico que representa de manera lineal el genoma con una longitud fija de 360 pb (anexo 2), obteniendo como resultado que las nueve proteínas analizadas se conservan en la misma dirección solo *A. vinelandii* CA6 con un ligero desfase ya que esta cepa contiene una delección en su genoma de 42096 bp.

Este visualizador, además, proporciona una tabla con la posición genómica, la cual se ingresó en la página IMG JGI (<https://img.jgi.doe.gov/>) para estudiar la vecindad genómica para *A. vinelandii* DJ y *A. vinelandii* CA encontrándose que están muy bien conservados (anexo 3).

Finalmente utilizando esta misma herramienta pudimos comparar el contexto genómico con *A. chroococcum* donde se pudo comprobar la sintenia de estos genes en los genomas comparados (anexo 3).

4. Ensayos fenotípicos

Con la finalidad de corroborar el fenotipo reportado previamente (García Gonzalez, 2019) se sembraron las cepas S4T, SM3 y S6R, así como la cepa control AEIV en los medios sólidos Ashby manitol, Ashby benzoato y Ashby manitol adicionado con cloruro de cobre

2.5 μM , basándonos en los antecedentes de las enzimas polifenol oxidasa relacionadas con las vías de síntesis de melanina con la finalidad de inducir la expresión de polifenol oxidasas dependiente de cobre (Page & Shivprasad, 1995). Como resultado de este ensayo fenotípico se pudo observar la producción del pigmento oscuro en las colonias de las tres cepas posteriormente a los cinco días de cultivo (figura6). Se logró identificar una diferencia fenotípica en la cepa S4T la cual en el medio Ashby benzoato presento la producción del pigmento no siendo el caso en el medio Ashby manitol adicionado con cloruro de cobre al

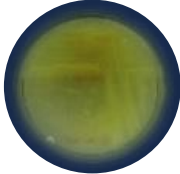

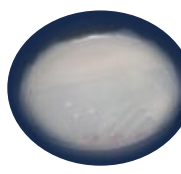
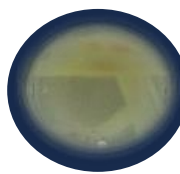
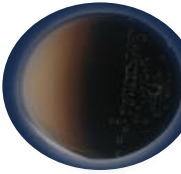

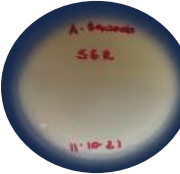
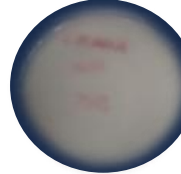


	AEIV	S4T	SM3	S6R
Ashby manitol				
Ashby benzoato				
Ashby manitol-CuCl ₂ (2.5 μM)				

Figura 6. Ensayo fenotípico de las cepas de interés cultivadas en diferentes medios de cultivo.

2.5 μM . Por su parte, las cepas SM3 y S6R sí presentaron la formación de pigmento en dichos medios de cultivo.

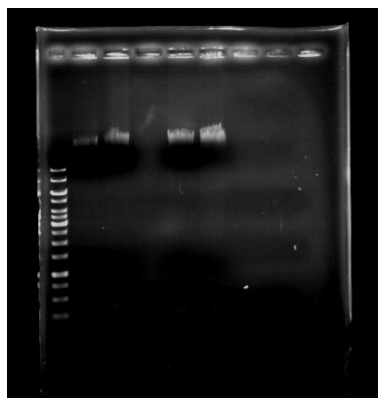
5. Extracción de ADN

Basado en los resultados obtenidos del ensayo fenotípico se decidió continuar trabajando únicamente con las muestras S4T y SM3 por la clara diferencia fenotípica que se presentó.

Las muestras S4T y SM3 que se encontraban conservadas en glicerol fueron sembradas por estriado en los medios sólidos Ashby adicionado con manitol y Burk sacarosa (BS), y se incubó a 30°C. Cuatro días posteriores a la siembra, las cepas presentaron crecimiento con producción de alginato. Una vez seleccionadas las colonias aisladas se

sembraron en medio BS líquido y se incubaron a 37°C en agitación a 200 rpm. Pasados dos días se procesaron las muestras a fin de obtener la materia celular purificada para llevar a cabo la extracción de ADN genómico con el kit “*Wizard Genomic DNA Purification*” de Promega Corporation. Ma, USA. (anexo 1).

para
en la
ARN.
realizó



Una vez realizada la extracción se realizó un corrimiento electroforético en gel de agarosa al 0.5% determinar la integridad del ADN. Como se muestra figura 7 se obtuvo un ADN integro sin residuos de La determinación de calidad y concentración se en el espectrofotómetro NanoDrop 2000 (Thermo scientific). Las muestras utilizadas fueron las que

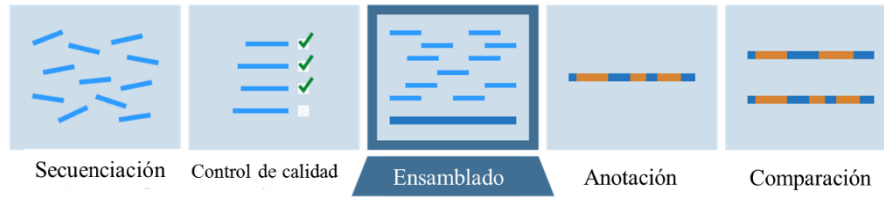
mostraron una relación entre las absorbancias 260/280 nm cercanas a 2.0 ya que indicaba una baja contaminación por proteínas (Tabla 4).

Figura 7. Electroforesis en gel de agarosa al 0.5%. En el primer carril de izquierda a derecha se observa el marcador de peso molecular Thermo scientific 1 kb ladder, segundo y tercer carril corresponde al ADN de SM3 y los carriles quinto y sexto corresponden al ADN de S4T.

Tabla 4. Mediciones de concentración y calidad de ADN realizadas en el espectrofotómetro NanoDrop 2000.

Cepas	260 (ng/μl)	260/280 nm	260/230 nm
S4T-1	22.081	2.05	1.85
S4T-2	24.865	2.05	2.17
SM3-1	8.421	2.08	1.87
SM3-2	8.549	2.01	1.51

6. Secuenciación del genoma



Las muestras de ADN genómico de S4T y SM3 se procesaron en la Unidad de Secuenciación Masiva del Instituto de Biotecnología de la UNAM. S4T se secuenció por la tecnología de MiSeq Illumina y MinION debido a que corresponde como se mencionó con anterioridad a *A. vinelandii* y a la comprobación de la diferencia fenotípica. SM3 se secuenció por MinION esperando en un futuro próximo se cuente con los recursos para mandarse a secuenciar igualmente por MiSeq Illumina. Una vez obtenidas las secuencias en formato fastq, se procedió a realizar el análisis del control de calidad de las secuencias, el ensamble, la anotación y comparación con el genoma de referencia *A. vinelandii* DJ (figura 8).

Figura 8. Esquema de trabajo con las secuencias. Cuando se cuenta con una secuencia el siguiente paso es medir la calidad de ésta para su posterior ensamblado, análisis y comparación con el genoma de referencia.

7. Análisis de calidad de la secuenciación

El análisis de calidad de las secuencias se realizó con el programa FastQC para las secuencias cortas de la muestra S4T, el cual genera diversas métricas que permiten evaluar la calidad. Para las lecturas largas de S4T y SM3 se entregó junto con la secuenciación el análisis de calidad elaborado por la Unidad de Secuenciación Masiva del Instituto de Biotecnología de la UNAM (tabla 5 y 6) reflejándose que son secuencias de calidad.

7.1 S4T (Lecturas cortas)

✓ Basic Statistics

Measure	Value
Filename	S4T_S33_R1_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	3339761
Sequences flagged as poor quality	0
Sequence length	74
%GC	65

Figura 9. Estadística básica de la muestra S4T.

En el gráfico de calidad de secuencia por base (figura 10) se observa que todas las bases se encuentran dentro de la zona color verde (zona de alta calidad >28). Además, se distingue un decaimiento de la señal a medida que avanza la secuenciación posiblemente debida a la disminución de la señal fluorescente.

✓ Per base sequence quality

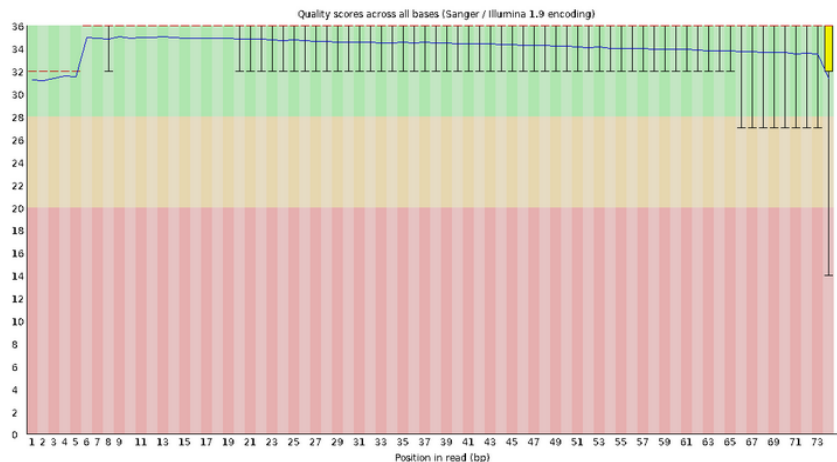


Figura 10. Gráfico de calidad de secuencia por base de S4T. Proporciona la distribución de puntuaciones de calidad en cada posición de la lectura en todas las lecturas. El eje Y proporciona las puntuaciones de calidad, mientras que el eje X representa la posición en la lectura. El código de colores de la gráfica denota lo que se considera puntajes de calidad altos,

El gráfico de la figura 11 nos permite corroborar la calidad asociada a las celdas de flujo.

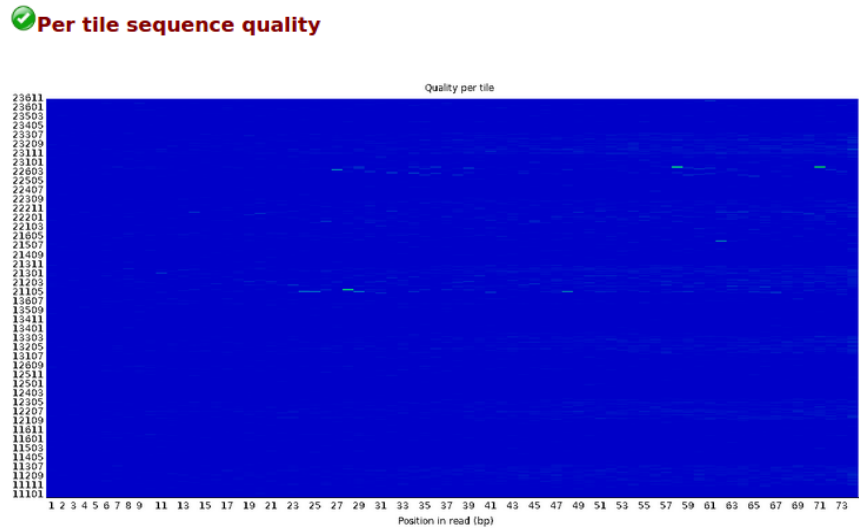


Figura 11. Gráfico de calidad de la secuencia por mosaico de S4T. Muestra las celdas de flujo, canales donde se coloca la secuencia. Solo aparece si se usa una biblioteca de Illumina.

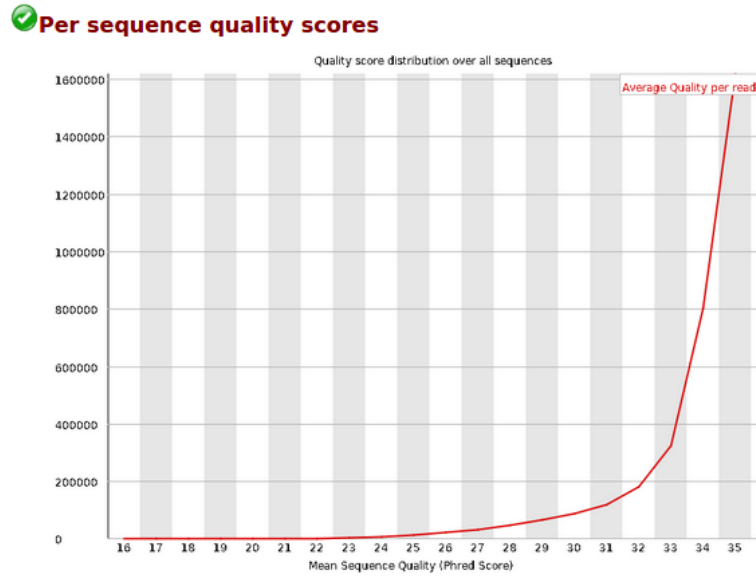


Figura 12. Gráfico de puntuaciones de calidad por secuencia de S4T. El eje de las X representa la media de la calidad de las secuencias, el eje de las Y representa el número de secuencias o lecturas al que corresponde esa media.

El gráfico puntuaciones de calidad por secuencia (figura 12) indica que existen más de 1,600,000 lecturas que representan una calidad en torno a 35 existiendo menos con baja calidad.

Per base sequence content

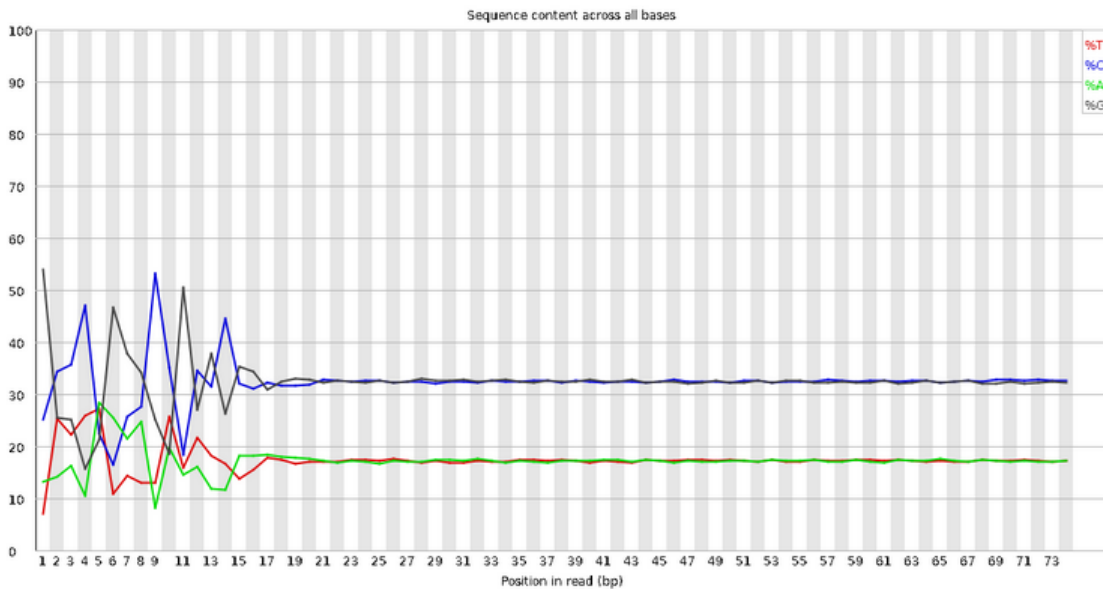


Figura 13. Gráfico de contenido de secuencia por base de S4T. Proporción de cada una de las bases nucleotídicas de la secuencia.

El gráfico de contenido de secuencia por base (figura 13) siempre da un fallo debido a que las primeras 10-12 bases resultan del cebado hexámico “aleatorio” que se produce durante la preparación de la biblioteca. En la figura 10 correspondiente a este gráfico se observa que las líneas correspondientes a las bases A, T, G y C corren en paralelo entre sí. La cantidad de A es igual a la cantidad de T y la cantidad de G igual que la de C.

En la figura 14 correspondiente al gráfico de contenido de Guanina y Citocina (GC) por secuencia se observa una distribución normal. El contenido de GC del pico central corresponde al 65% de GC esperado para *A. vinelandii* de acuerdo con lo reportado en NCBI.

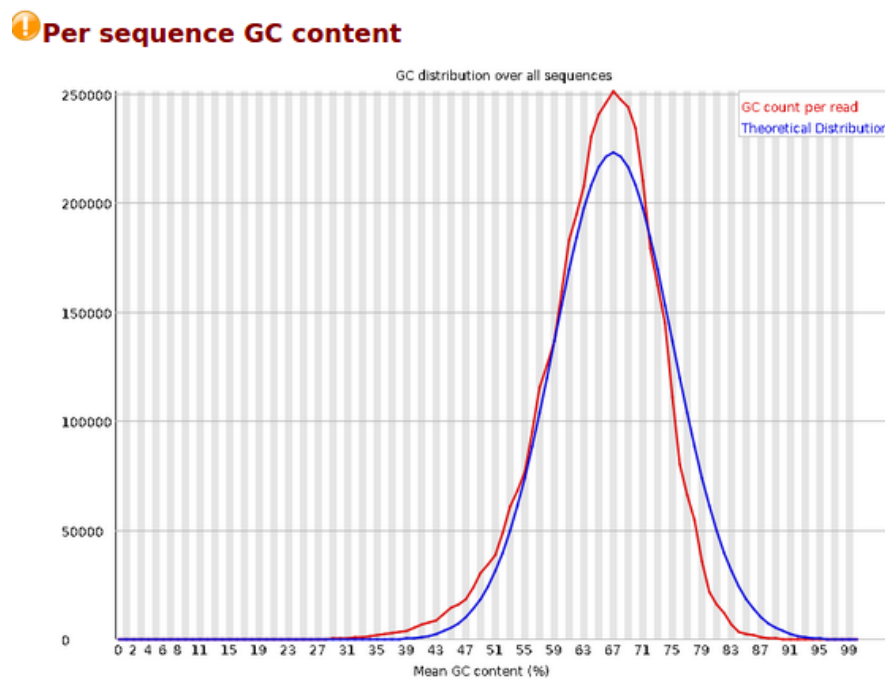


Figura 14. Gráfico contenido de GC por secuencia de S4T. Contenido de GC de la secuencia (línea roja) comparado con una distribución normal teórica del contenido de GC (línea azul). El eje X muestra el porcentaje medio del contenido de G y C, el eje Y muestra el número de lecturas.

El contenido de N es nulo como se puede apreciar en el gráfico de contenido de N por base (figura 15).

Per base N content

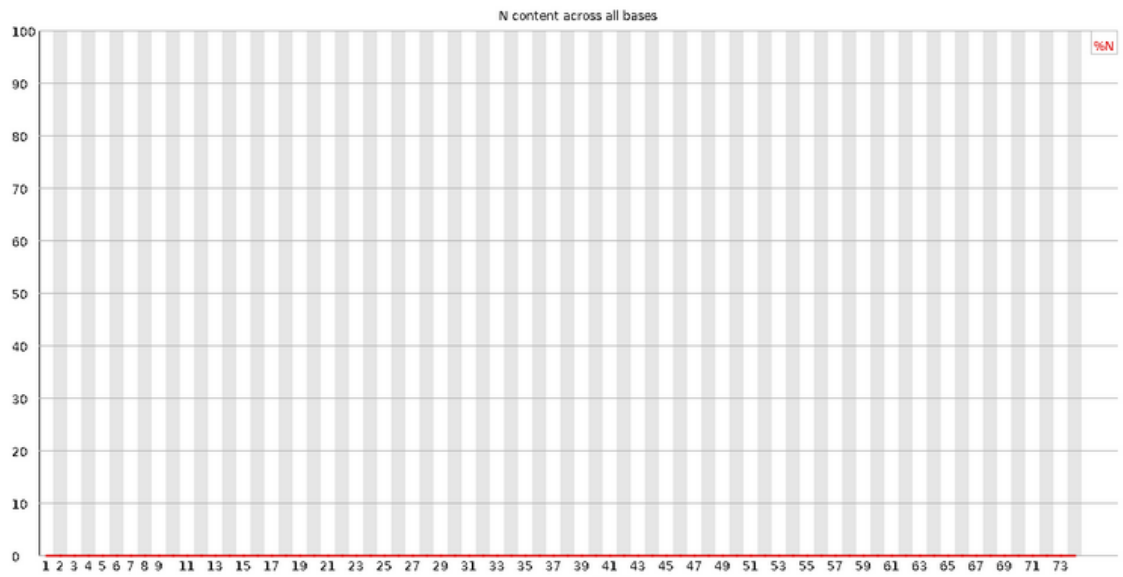


Figura 15. Gráfico de contenido de N por base de SAT.

La distribución de tamaños de los fragmentos de la secuenciación se muestra en la figura 16 en el gráfico de distribución de la longitud de las secuencias donde se observa una distribución homogénea.

Sequence Length Distribution

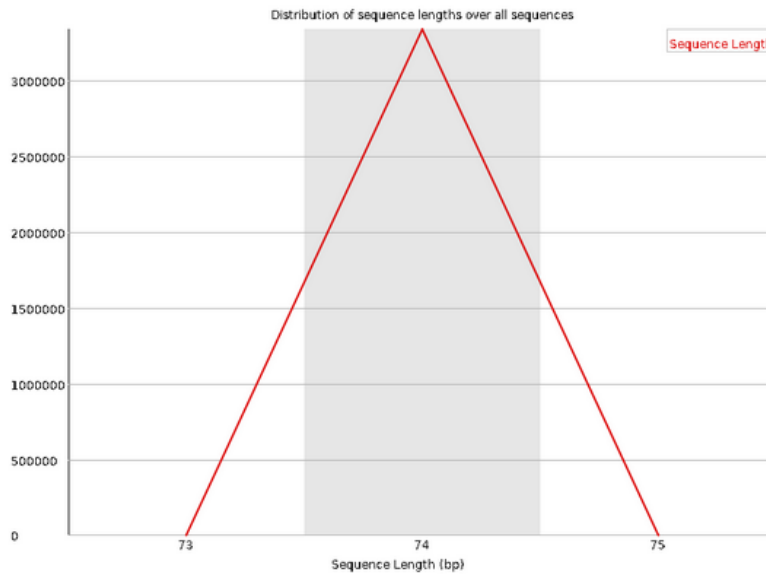


Figura 16. Gráfico de distribución de tamaño de fragmentos de secuenciación de SAT.

Sequence Duplication Levels

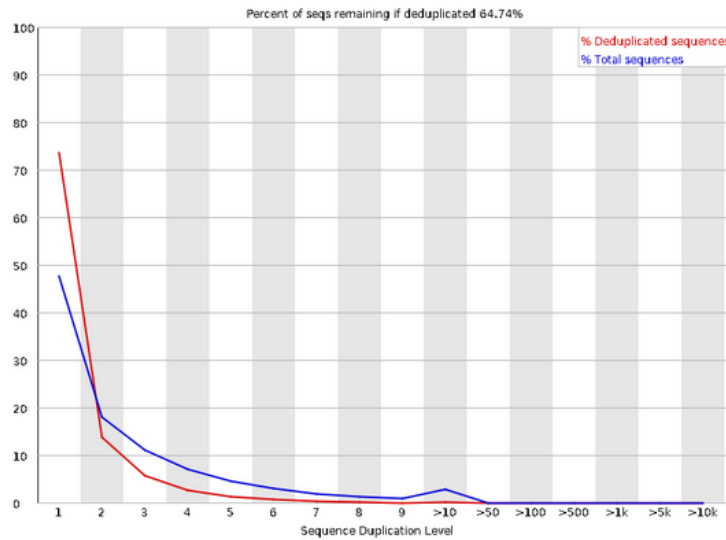


Figura 17. Gráfico de niveles de secuencias duplicadas de SAT. La línea azul muestra el total de las secuencias y la línea roja las secuencias duplicadas.

El grado de duplicación para cada secuencia en una biblioteca se representa en el gráfico de la figura 17 que muestra el número relativo de secuencias con diferentes grados de duplicación. Como se puede ver, la mayoría de las secuencias caen en el extremo izquierdo

del gráfico para ambas líneas, esto indica que se trata de una biblioteca apropiadamente diversa. La presencia de un pico en >10 de la línea azul se refiere a que hay más de un 10% de secuencias que tienen el mismo fragmento desde el principio hasta el final 10 veces.



Figura 18. Secuencias sobrerrepresentadas de S4T.

No se encontraron secuencias sobrerrepresentadas (figura 18) lo que significa que se tiene un conjunto diverso de secuencias sin que ninguna secuencia individual constituya una pequeña fracción del total.

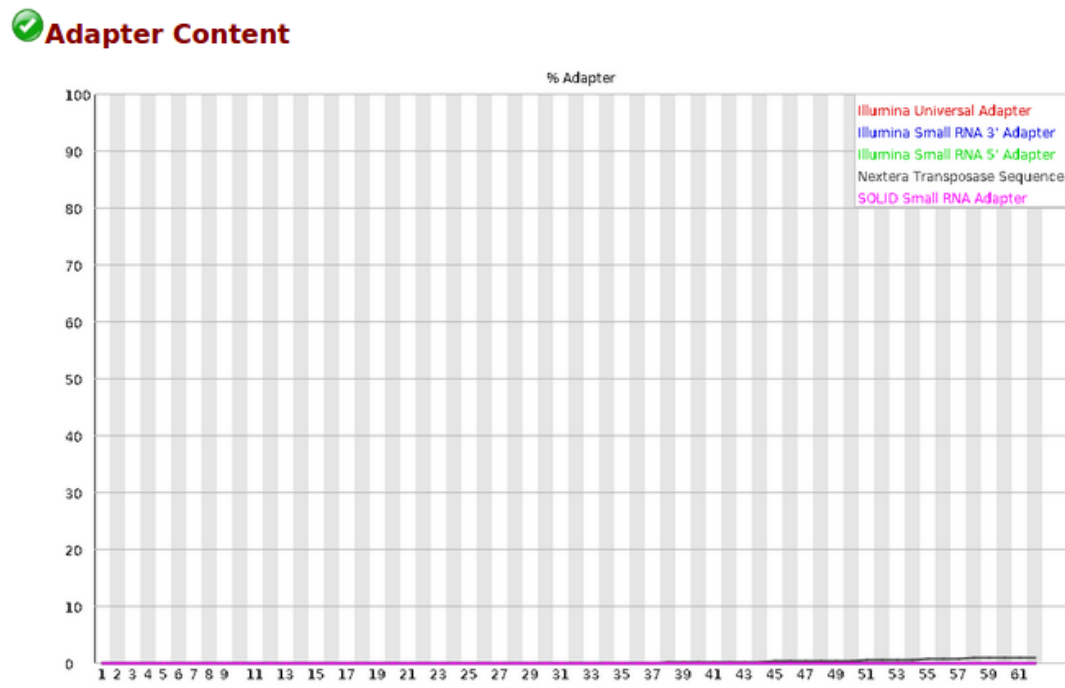


Figura 19. Gráfico de contenido de adaptadores de S4T

No se encontraron adaptadores de acuerdo con lo observado en el gráfico de la figura 19.

7.2. S4T (Lecturas largas)

Tabla 5. Métricas de evaluación de calidad para secuencias largas de S4T.

Longitud media de las lecturas	8,730.3
Calidad media de las lecturas	11.8
Mediana de la longitud de las lecturas	5,354.0
Mediana de la calidad de las lecturas	12.5
Número de las lecturas	118,446.0
N50	14,619.0
Total de bases	1,034,068,722.0

7.3. SM3

Tabla 6. Métricas de evaluación de calidad para secuencias largas de SM3.

Longitud media de las lecturas	3,576.9
Calidad media de las lecturas	11.4
Mediana de la longitud de las lecturas	1,831.0
Mediana de la calidad de las lecturas	12.0
Número de las lecturas	301,620.0
N50	7,381.0
Total de bases	1,078,854,371.0

8. Ensamblaje del genoma

Como se mencionó previamente, las muestras con las que se trabajó a lo largo de este proyecto se secuenciaron por diferentes tecnologías, S4T por MiSeq Illumina y MinION y SM3 por MinION. Se llevó a cabo utilizando la herramienta Unicycler que produce ensamblajes más precisos y para el caso de S4T permite realizar ensamblado híbrido ya que primero ensambla lecturas cortas (MiSeq Illumina), siguiendo el método de resolución de gráficos de De Bruijn y luego utiliza las lecturas largas (MinION) para encontrar las mejores rutas a través del gráfico.

En la tabla 7 se muestran los datos obtenidos del ensamblaje para ambas muestras.

Tabla 7. Resultados del ensamblado por Unicycler.

	S4T	SM3
Contigs	3	9
N50	4173288	7,381
Tamaño del ensamblaje	5.4 Mb	5.6 Mb

8.1 Calculador de la Identidad de nucleótidos promedio

Para comparar las dos secuencias genómicas, es decir las recién ensambladas contra la de *Azotobacter vinelandii* y *Azotobacter chroococcum* se utilizó el calculador “ANI” por sus siglas en inglés (Average Nucleotide Identity) de EZBioCloud (<https://www.ezbiocloud.net/tools/ani>) que nos permitió observar el alto grado de similitud entre las secuencias con la comparación nucleótido a nucleótido que realiza, obteniendo como resultado un 97.75% de identidad entre *A. vinelandii* y S4T y un 96.16% de identidad entre *A. chroococcum* y SM3 (figura 20 y 21).

OrthoANIu Results

Metric	Value
OrthoANIu value (%)	97.75
Genome A length (bp)	5,365,200
Genome B length (bp)	5,481,480
Average aligned length (bp)	3,439,149
Genome A coverage (%)	64.10
Genome B coverage (%)	62.74

Figura 20. Valores obtenidos del calculador de ANI para el genoma de *Azotobacter vinelandii* (Genome A) comparado con S4T (Genome B).

OrthoANli Results	
Metric	Value
OrthoANli value (%)	96.16
Genome A length (bp)	5,021,460
Genome B length (bp)	5,640,600
Average aligned length (bp)	2,979,949
Genome A coverage (%)	59.34
Genome B coverage (%)	52.83

Figura 21. Valores obtenidos del calculador de ANI para el genoma de *Azotobacter chroococcum* (Genome A) comparado con *SM3* (Genome B).

9. Anotación del genoma

Una vez que se tuvieron los ensamblados de las muestras se llevaron a cabo las anotaciones por medio de la herramienta de software Prokka, que nos permite encontrar los marcos de lectura abiertos tomando como genomas de referencia a los del género *Azotobacter*. Genera como resultado, una serie de diferentes archivos (tabla 8) los que utilizaremos en el análisis posterior para encontrar los genes que se conservan relacionados con las vías de síntesis de melanina. Se propuso generar un análisis de tipo pangenoma tomando en cuenta las secuencias de los genomas completos en la base de datos de NCBI para poder discernir si los genes pertenecientes a la síntesis de melaninas encontrados previamente pertenecían al genoma núcleo (genes que son compartidos entre las especies) o genes accesorios (genes únicos de la especie particular).

Tabla 8. Extensiones producto de Prokka

Extensión	Descripción
.gff	Anotación maestra en formato GFF3, contiene las secuencias y anotaciones.
.gbk	Es el archivo estándar en Genbank, deriva de .gff.
.fna	Archivo FASTA de nucleótidos de las secuencias contigs de entrada.
.faa	Archivo FASTA de proteínas de las secuencias CDS traducidas.
.ffn	Archivo FASTA de nucleótidos de todas las transcripciones predichas.
.sqn	Archivo ASN1 con formato "Sequin" para enviar a Genbank. Necesita ser editado para establecer la taxonomía, autores, publicaciones relacionadas, etc. correctas.
.fsa	Archivo FASTA de nucleótido de las secuencias contigs, usadas por "tbl2asn" para crear el archivo .sqn. Muy similar al archivo .fna pero con las etiquetas Sequin en las líneas de descripción de la secuencia.

.tbl	Archivo de tabla con características, usado por “tbl2asn” para crear el archivo. Sqn.
.err	Anotaciones que no se aceptan.
.log	Contiene toda la información de salida que produce Prokka mientras trabaja.
.txt	Estadísticas relativas a las características de las anotaciones encontradas.
.tsv	Archivo separado por “tab” de las características: locus_tag.ftype, len_bp.gene,EC_number,COG,producto.

10. Blast de proteínas (BlastP)

Se llevó a cabo el análisis del alineamiento local de la proteína homogentisato 1,2 dioxigenasa de *A. vinelandii* DJ y la cepa S4T como se muestra en la figura 22 en donde podemos apreciar que, aunque la diferencia es poca representa una diferencia significativa al tratarse del mismo género, en este caso *Azotobacter*.

Score	Expect	Method	Identities	Positives	Gaps
880 bits(2274)	0.0	Compositional matrix adjust.	427/433(99%)	431/433(99%)	0/433(0%)
Query 1		MHTAEPSAPGYQTGFGNTFSSEALPGALPIGQNSPQRLPYGLYAEQLSGTAFTVPRNEAR			60
Sbjct 1		MHTAEPSAPGYQTGFGNTFSSEALPGALPIGQNSPQRLPYGLYAEQLSGTAFTVPRSEAR			60
Query 61		HAWLYRIRPAANHPEFKRLGSGISGIEQGPITPNRLRWAPFEVPAEHTDFLDGLIRLAAT			120
Sbjct 61		HAWLYRIRPSANHPDFKRLGSGISGIEQGPITPNRLRWAPFEVPAEHTDFLDGLIRLAAT			120
Query 121		AAAEQADGVSLYVYRANASMESVFFDADGELLLVPESGRLGIDTELGRLEIGPLQIAVVP			180
Sbjct 121		AAAEQADGVSLYVYRANASMESVFFDADGELLLVPESGRLGIDTELGRLEIGPLQIAVVP			180
Query 181		RGMRFRVELLDATARGYLCENHGSPRLPELGPISNGLANPRDFLVPVARYEDRDGPVQ			240
Sbjct 181		RGVRFREVELLDATARGYLCENHGSPRLPELGPISNGLANPRDFLVPVARYEDRDGPVQ			240
Query 241		LVQKFLGELWACELNHSPLDVVAWHGNHVPYSYDLRRFNTLGTVSFDHPDPSIFTVLTSP			300
Sbjct 241		LVQKFLGELWACELNHSPLDVVAWHGNHVPYSYDLRRFNT GTVSFDHPDPSIFTVLTSP			300
Query 301		GSPVPGQANVDFVIFPPRWMVAEHTFRPPWFHRNLMNEFMGLIQGVYDAKAGGF L PGGASL			360
Sbjct 301		GSPVPGQANVDFVIFPPRWMVAEHTFRPPWFHRNLMNEFMGLIQGVYDAKAGGF L PGGASL			360
Query 361		HNRM SAHGPDAAATTRQAI AADLQPKLENTMAFMFETGQVLRPSRLAIDCPQRQTDYDGC			420
Sbjct 361		HNRM SAHGPDAAATTRQAI AADLQPKLENTMAFMFETGQVLRPSRLAIDCPQRQTDYDGC			420

Figura 22. Blast de proteínas del género *Azotobacter* para la proteína Homogentisato 1,2 dioxigenasa. Las secuencias identificadas como “Sbjct” corresponden a *A. vinelandii* DJ y las identificadas como “Query” corresponden a la cepa S4T.

11. Pangenoma

Se realizó un análisis con el software Roary que permite estudiar el pangenoma, el cual se refiere al conjunto de todos los genes de todas las cepas que forman un clado filogenético adquiriendo importancia en un contexto evolutivo. Esto con el fin de conocer si los genes de interés se encuentran en el genoma núcleo o en el genoma accesorio de *Azotobacter* ya que de encontrarse en el genoma núcleo la probabilidad de que estos se encuentren también en las muestras S4T y SM3 son altas.

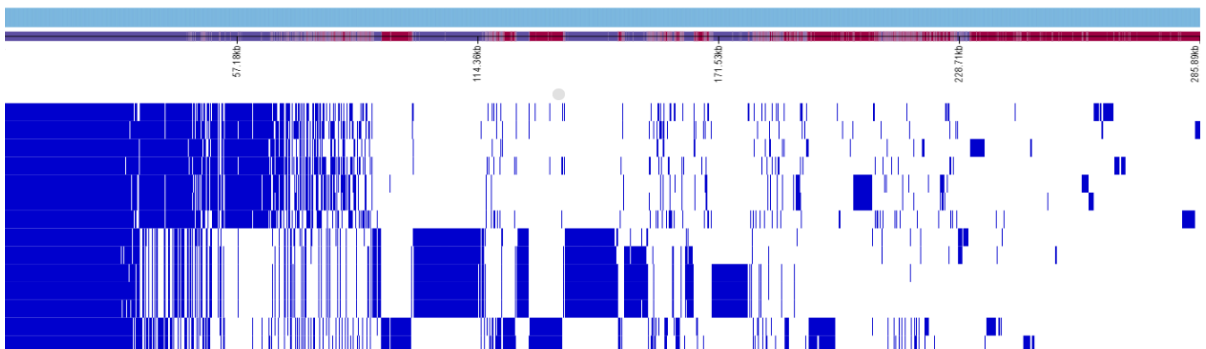
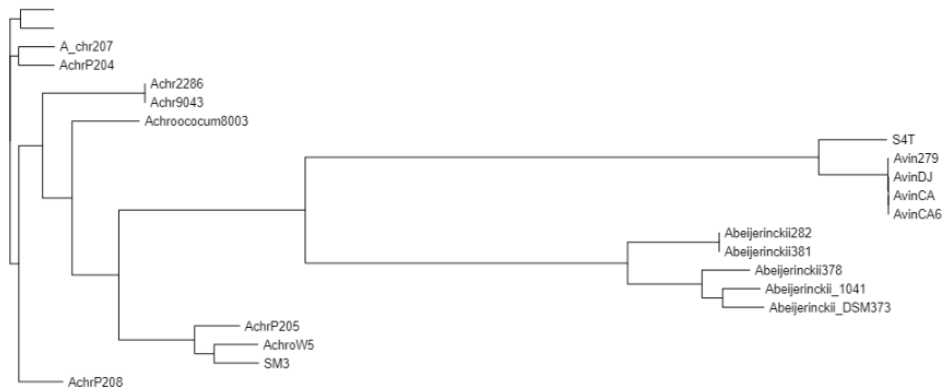
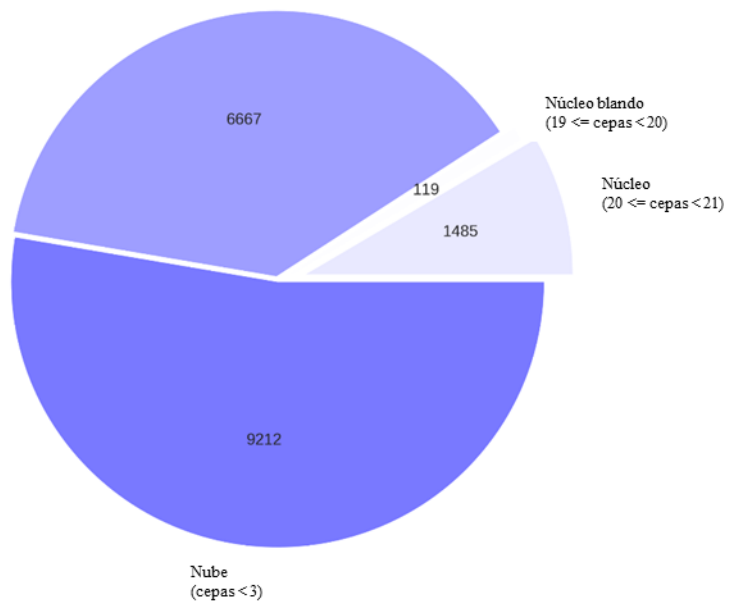
Roary proporciona las métricas del genoma núcleo, así como del genoma accesorio el cual está dividido en “nube” que hace referencia a los genes raros presentes solo en unos pocos genomas y “caparazón” que son los genes restantes moderadamente conservados presentes en varios genomas. Aparece también el “núcleo blando” que son los genes contenidos en el 95% de los genomas considerados.

Los resultados obtenidos al llevar a cabo el análisis con los genomas anotados disponibles en NCBI (sección 3.6) junto con los recién anotados en este proyecto con 90% de identidad se presentan en la tabla 9 y gráficamente en la figura 23.

Tabla 9. Métricas obtenidas con Roary con los parámetros de 90% de identidad y 75% de cobertura.

Genoma núcleo	1485
Núcleo blando	119
Caparazón	6667
Nube	9212
Genes totales	17483

Caparazón
(3 <= cepas < 19)



La figura 24 muestra el dendrograma en comparación con una matriz con la presencia y ausencia de genes básicos y accesorios. Derivado de esto se generó una tabla (figura 24) que muestra el desglose de cada genoma analizado y la presencia o ausencia de los genes de interés para efecto de este proyecto.

De los genes analizados únicamente el fosfoenolpiruvato sintasa tiene presencia en los 21 genomas analizados. Para el genoma de S4T hay ausencia de Policétido sintasa tipo III (FCG41_04075) y Homogentisato-1,2-dioxigenasa (hmgA, E0E53_01385, SAMN04244573_01687). SM3 tiene ausencia de los genes Policétido sintasa tipo III (arsC) y Ácido graso sintasa tipo I (arsA)

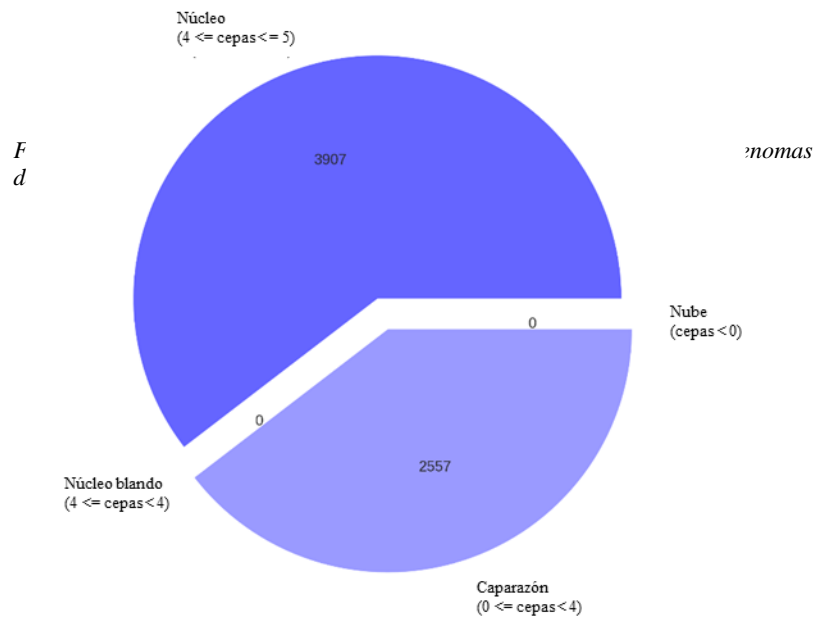
A partir de los resultados obtenidos del análisis del pangenoma se observó el agrupamiento de S4T con los genomas de *Azotobacter vinelandii* que se analizaron por lo que se realizó un análisis de pangenoma con Roary únicamente con los genomas de *Azotobacter vinelandii* DJ y S4T. Los resultados se muestran en la tabla 10 y gráficamente en la figura 25.

Tabla 10. Métricas obtenidas con Roary con los parámetros de 90% de identidad y 75% de cobertura para los genomas de *Azotobacter vinelandii* y ST4.

Genoma núcleo	3907
---------------	------

Figura 31. Dendrograma de los grupos de genomas analizados en Roary junto con la matriz de presencia y ausencia de genes.

Núcleo blando	0
Caparazón	2557
Nube	0
Genes totales	6464



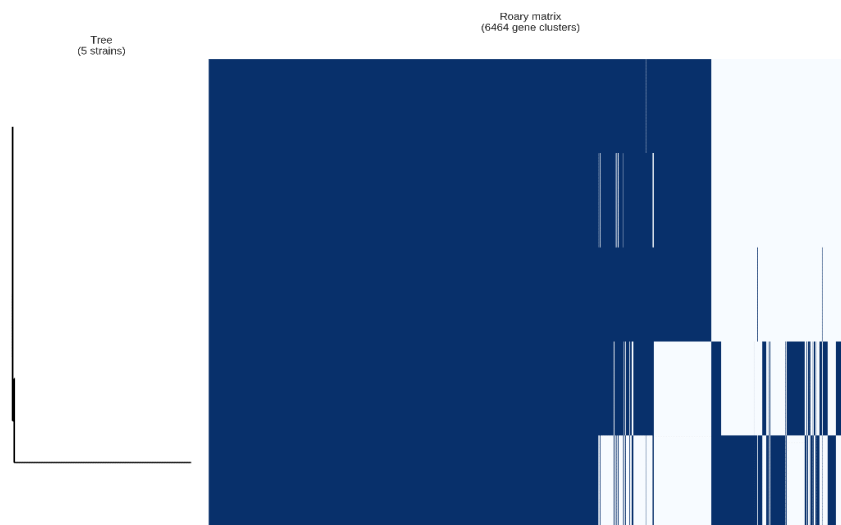


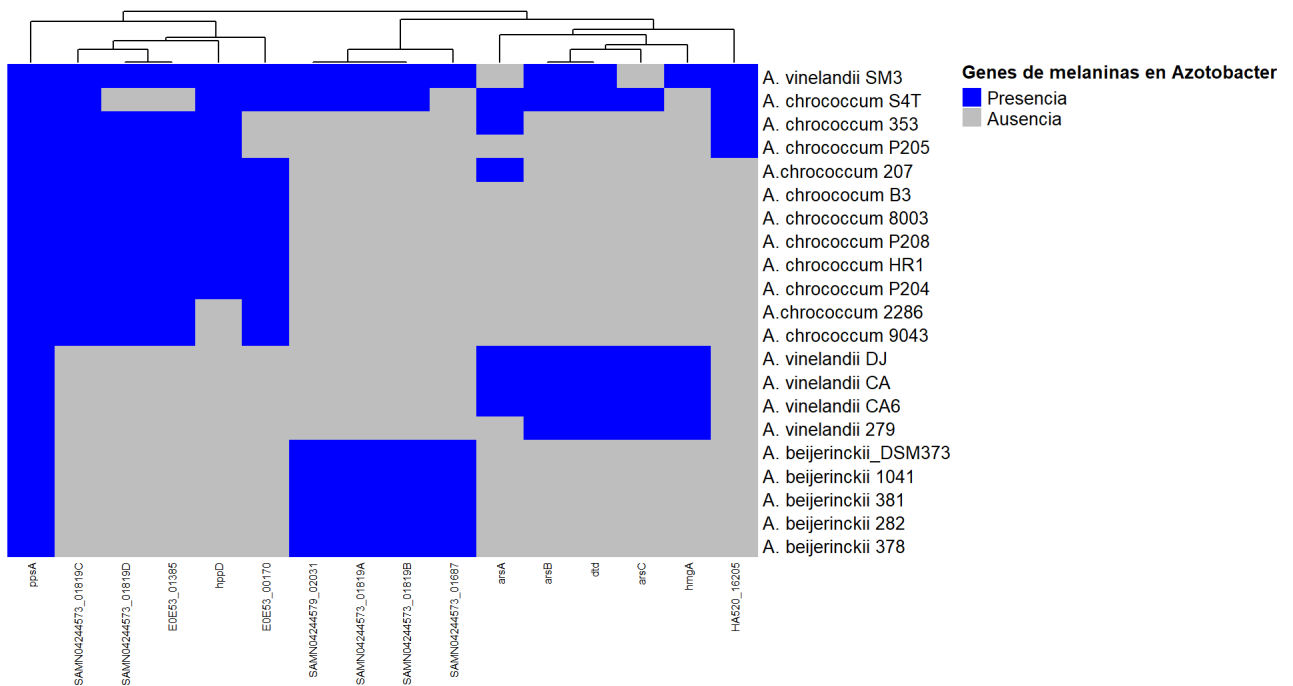
Figura 26. Dendrograma de los grupos de genomas analizados en Roary junto con la matriz de presencia y ausencia de genes.

En la figura 26 y 27 se presenta el dendrograma junto con la matriz de presencia y ausencia de los genes estudiados, la nomenclatura de los genes se presenta en la tabla 11. Basado en estos resultados, se identificó que los genes marcados con ausencia en el genoma núcleo de S4T, Policétido sintasa tipo III y Homogentisato-1,2-dioxigenasa se encuentran en el genoma accesorio marcando de esta manera una diferencia en la conservación de los genomas posiblemente involucrada en la capacidad de síntesis de melanina.

Tabla 11. Nomenclatura de genes analizados en los genomas de *Azotobacter*.

Genes	Nombre
ppsA	Fosfoenolpiruvato sintasa
SAMN04244573_01819C	Policétido sintasa tipo III
SAMN04244573_01819D	Policétido sintasa tipo III
E0E53_01385	Homogentisato 1,2-dioxigenasa
hppD	4- hidroxifenilpiruvato dioxigenasa
E0E53_00170	D- tirosil- Trna (Tir) deacilasa

SAMN04244579_02031	D- tirosil- Trna (Tir) deacilasa
SAMN04244573_01819A	Fenilalanina-4-monooxigenasa
SAMN04244573_01819B	Aspartato tirosina aminotransferasa
SAMN04244573_01687	Homogentisato 1,2-dioxigenasa
arsA	Ácidos grasos sintasa tipo I
arsB	Policétido sintasa tipo III
dtd	D- tirosil- Trna (Tir) deacilasa
arsC	Policétido sintasa tipo III
hmgA	Homogentisato 1,2-dioxigenasa
HA520_16205	D- tirosil- Trna (Tir) deacilasa



Conclusiones

De los BLAST de proteínas se encontró que son nueve los genes que posiblemente están involucrados en la síntesis de melanina, 4-hidroxifenilpiruvato dioxigenasa, Fosfoenolpiruvato sintasa, Policétido sintasa tipo III, Policétido sintasa tipo I, Homogentisato-1,2-dioxigenasa, Aspartato aromático tirosina aminotransferasa, Policétido

sintasa dehidratasa, Fenilalanina-4-monooxigenasa y D-tirosil-tRNA (Tir) deasilasa. Dichos genes se pueden identificar dentro de la vía que tiene como precursor al malonil-CoA y la vía conocida como “vía del homogentisato” que parte de una molécula de tirosina y produce el compuesto homogentisato.

A partir de los ensayos fenotipos se observó que la muestra SM3 y S6R produce el pigmento melanina cuando el medio de cultivo Ashby manitol está adicionado con Cloruro de cobre al 2.5 μM lo que sugiere una vía de síntesis donde intervienen enzimas polifenol oxidasas.

De acuerdo con el análisis del Pangenoma, de los nueve genes objeto de estudio únicamente Policétido sintasa tipo III que forma parte de la vía de síntesis de melanina a partir del malonyl-coA y Homogentisato-1,2-diooxigenasa que forma parte de la vía que utiliza como precursor a la tirosina para la síntesis del biopolímero, se encuentran en el genoma accesorio, los demás se encuentran en el genoma núcleo, esto indica que de alguna manera la secuencia de nucleótidos presente en esos genes no se encuentra conservada en un 90 % que fue el parámetro que se tomó como punto de corte para realizar el pangenoma.

Perspectivas

Comprobar la actividad de los genes identificados mediante mutagénesis dirigida en los genes Policétido sintasa tipo III y Homogentisato-1,2-diooxigenasa con el fin de demostrar si su mutagénesis hace incapaz al microorganismo de producir melanina.

Realizar mutagénesis azarosa en las cepas S4T y SM3 en búsqueda de las mutantes que conserven la capacidad de sintetizar melanina, amplificar y aislar su ADN. De esta

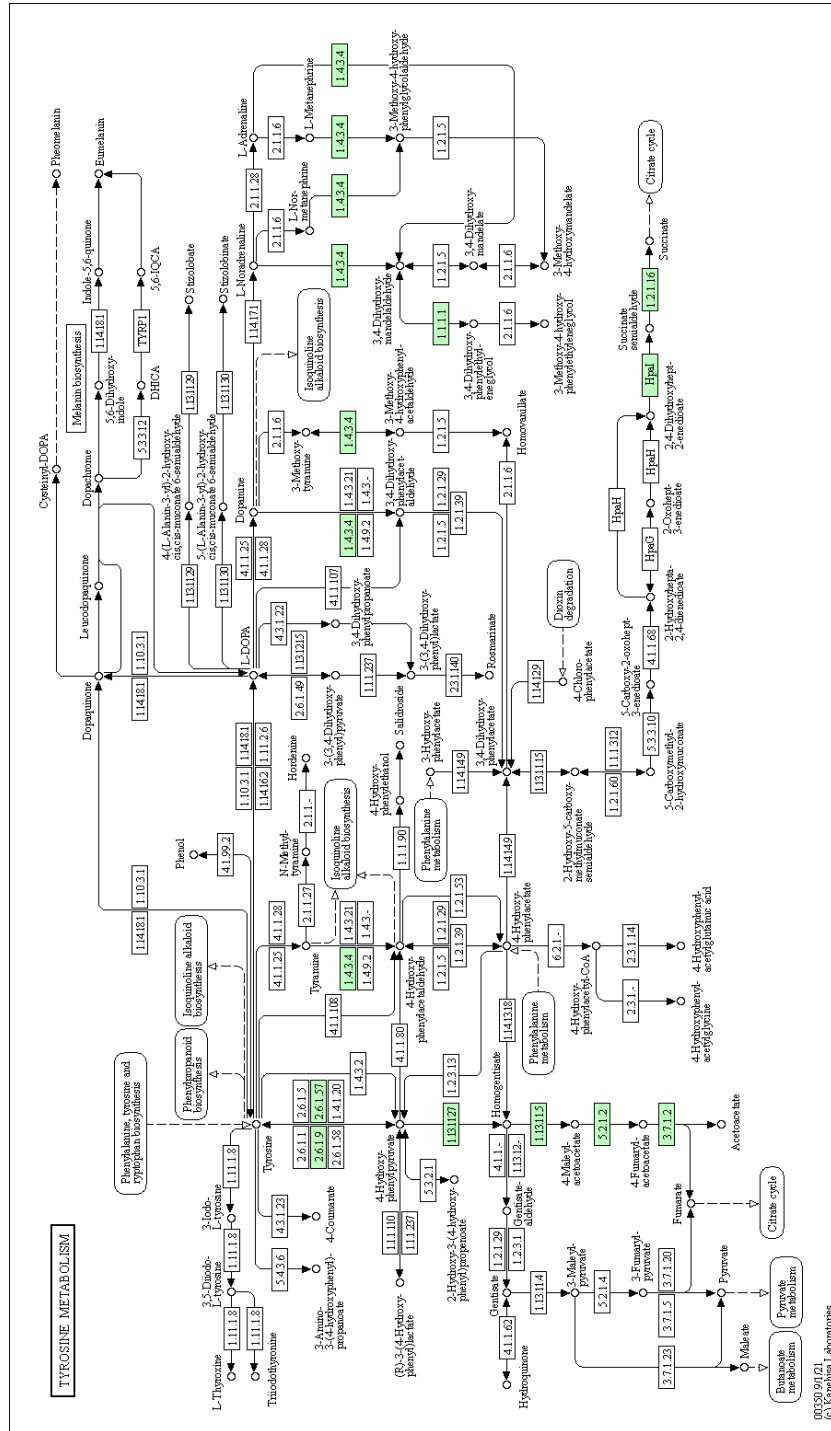
manera poder secuenciar su ADN para la identificación de genes involucrados en la síntesis del biopolímero.

Anexos

Anexo 1. Ruta metabolismo de tirosina en genoma de *Azotobacter vinelandii*

DJ.

Anexo 2.
Kit
DNA



Protocolo
Wizard
Genetic

Purification de Promega para extracción de ADN en bacterias Gram negativas.

1. Adicionar 1 ml del cultivo en microtubo de 1.5 ml para centrifuga.
2. Centrifugar a 13,000–16,000 rpm por 2 minutos, se formará una pastilla en el fondo del microtubo. Remover el sobrenadante.
3. Adicionar 600 µl de la solución “Lisis de núcleos”. Pipetear gentilmente para resuspender las células.
4. Incubar a 80°C por 5 minutos para la lisis de las células; enfriar a temperatura ambiente.
5. Adicionar 3 µl de solución de “RNAsa” para lisis celular. Invertir el tubo de 2 a 5 veces para agitar.
6. Incubar a 37°C de 15 a 60 minutos. Atemperar la muestra.
7. Adicionar 200 µl de la solución “Precipitante de proteínas”. Vortexear vigorosamente 20 segundos.
8. Incubar en hielo 5 minutos.
9. Centrifugar a 13,000-16,000 rpm por 3 minutos.
10. Trasferir el sobrenadante a un microtubo de 1.5 ml para centrifuga limpio.
11. Adicionar 600 µl de isopropanol a temperatura ambiente.
12. Mezclar por inversión del microtubo para que los hilos de ADN se unan en una masa visible.
13. Centrifugar a 13,000-16,000 rpm por 2 minutos.
14. Verter el sobrenadante escurriendo el microtubo en un papel absorbente. Adicionar 600 µl de etanol al 70% a temperatura ambiente y mezclar por inversión del microtubo.
15. Centrifugar a 13,000-16,000 rpm por 2 minutos.
16. Drenar el microtubo en un papel absorbente limpio y dejar que la pastilla formada seque por 10 a 15 minutos.
17. Adicionar 100 µl de solución “Rehidratación de ADN” en el microtubo y rehidratar el ADN por incubación a 65°C por 1 hora. Periódicamente mezclar

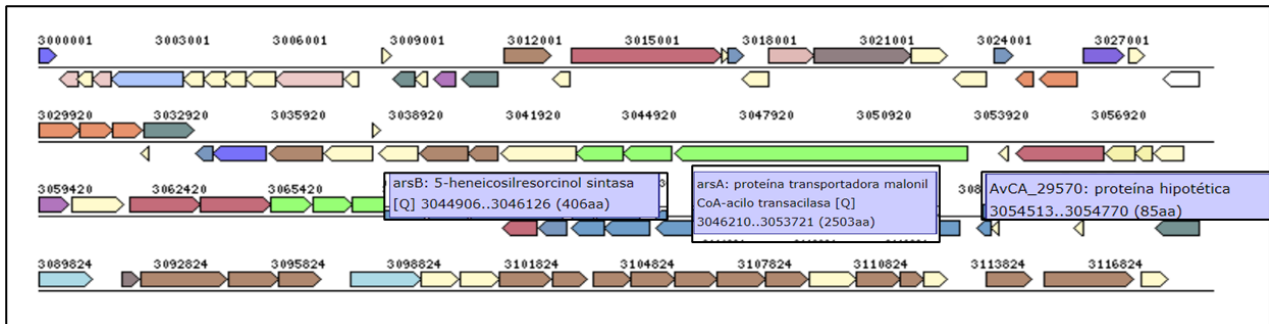
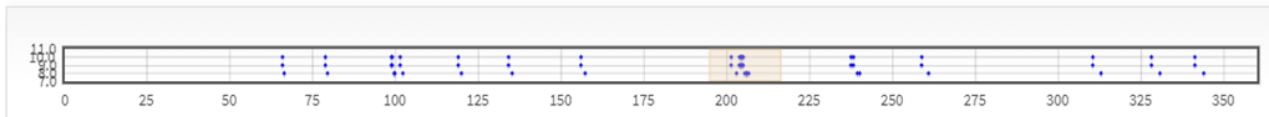
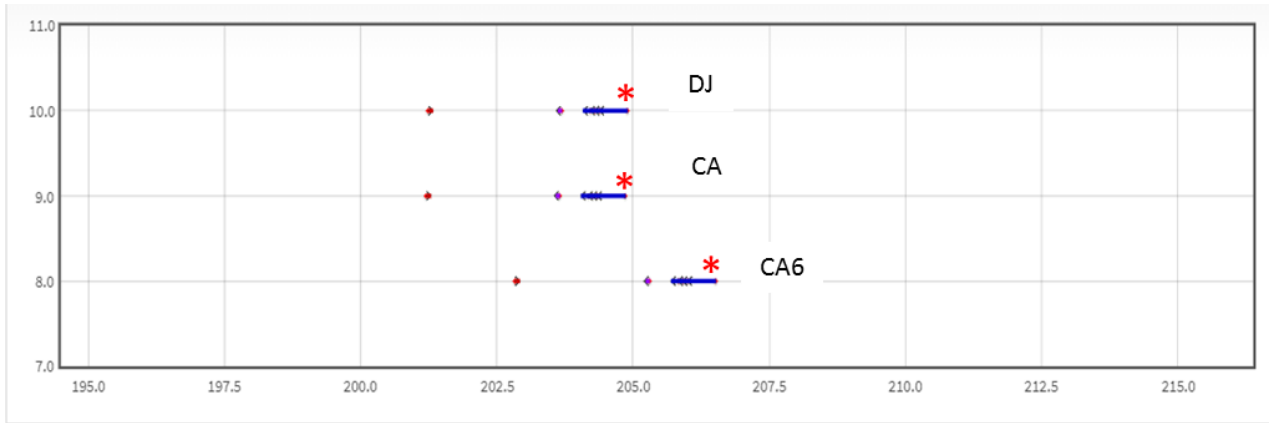
por inversión del microtubo. Alternativamente, rehidratar el ADN por incubación de la solución a 40°C.

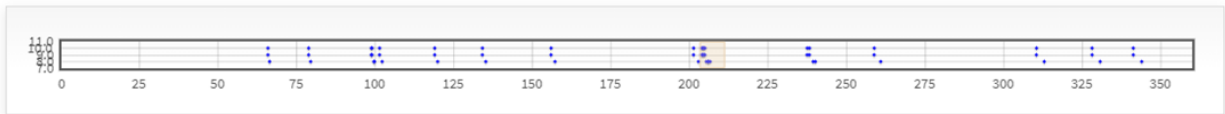
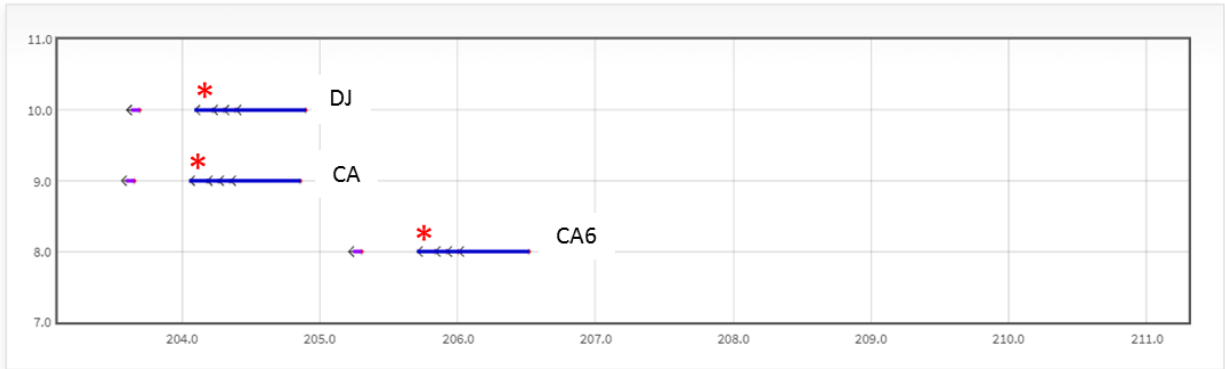
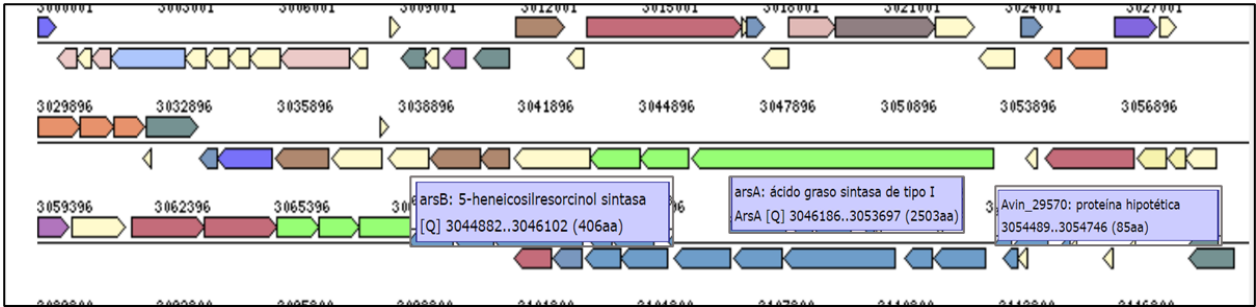
18. Almacenar el ADN a 2-8°C.

Anexo 3. Contexto genómico y sintenia

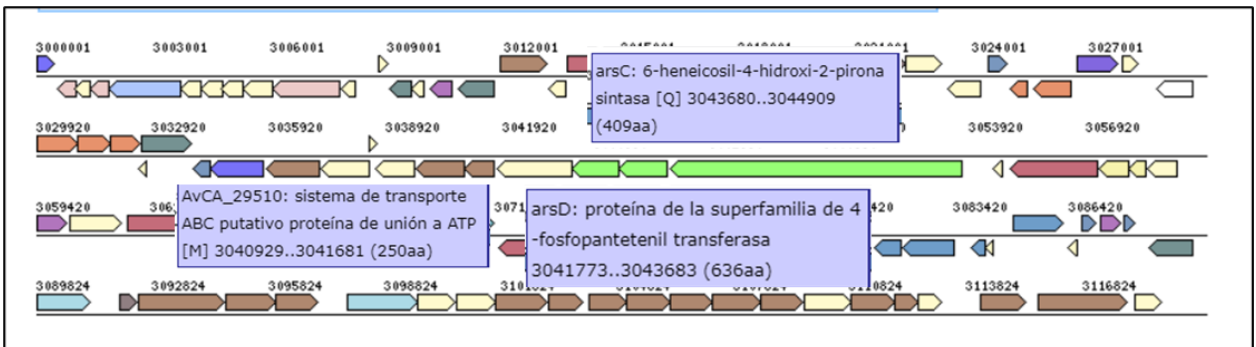
Contexto genómico de los genes de interés en los genomas de *Azotobacter vinelandii* DJ, CA y CA6 obtenidos del visualizador BLAST-XYplot viewer, así como vecindad genómica analizada en los genomas de *A. vinelandii* DJ y CA en la página del IMG JGI.

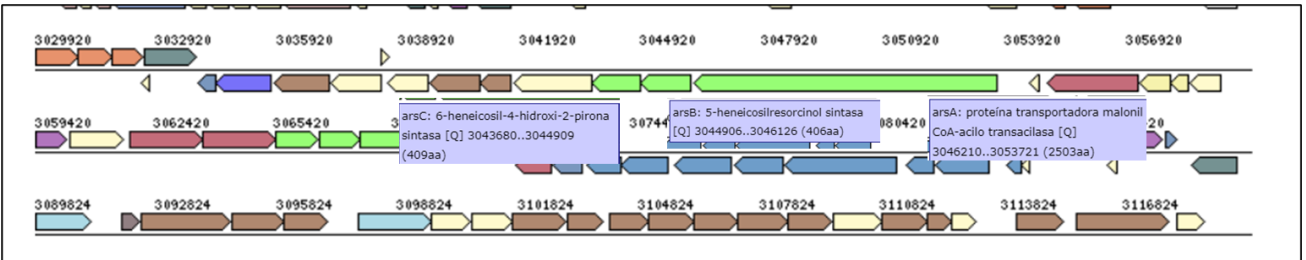
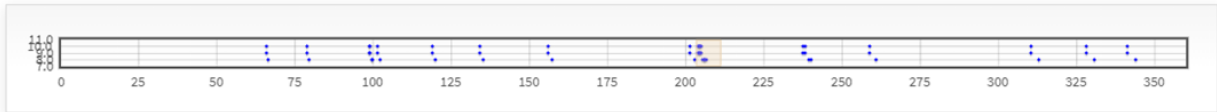
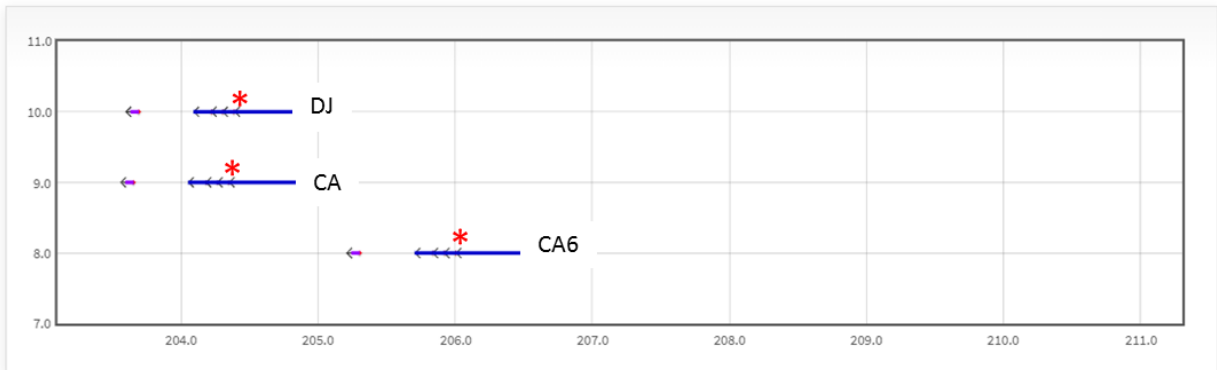
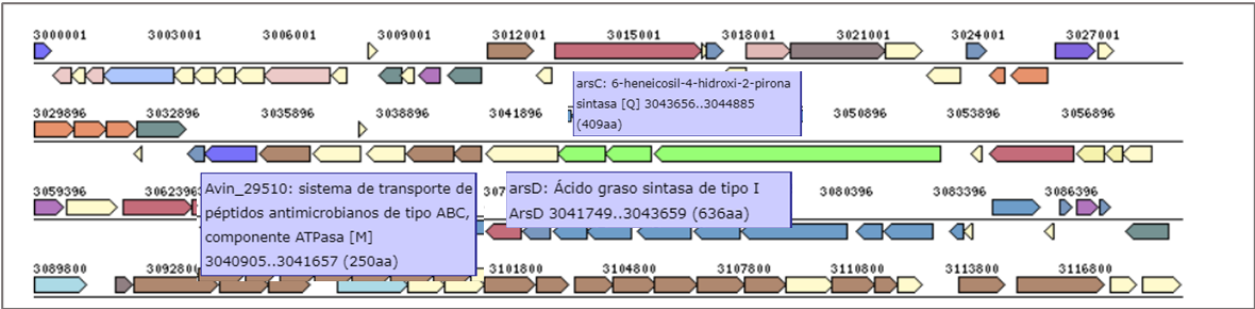
1. Policétido sintasa tipo I



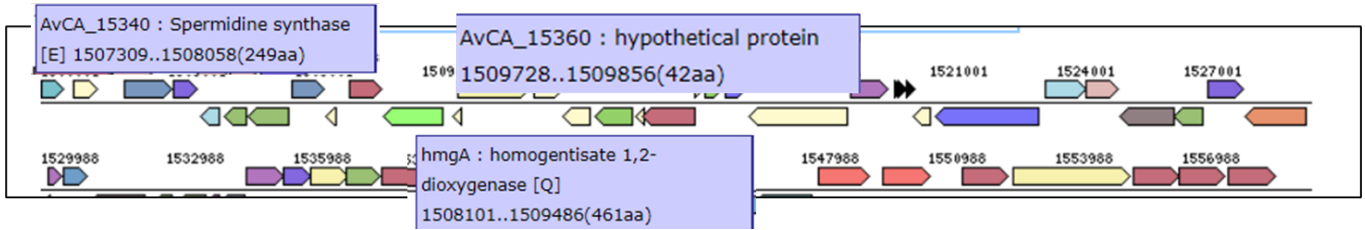
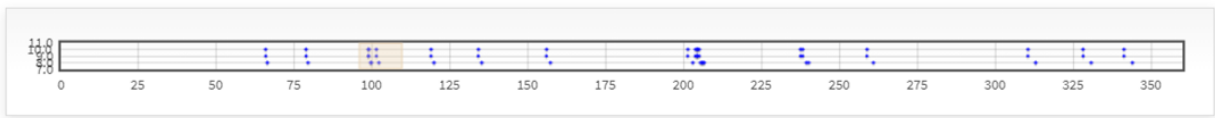
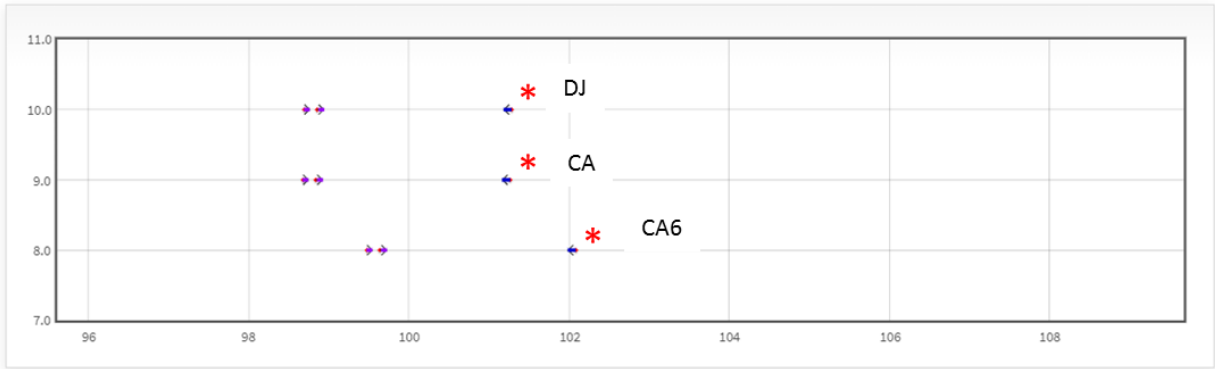


2. Policétido sintasa deshidratasa

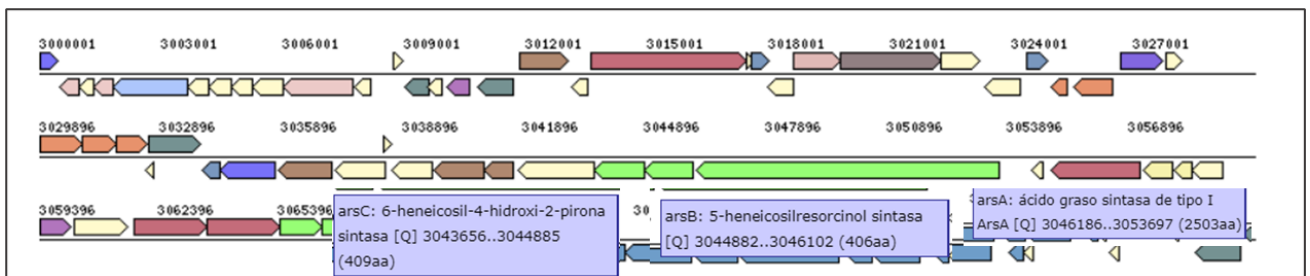
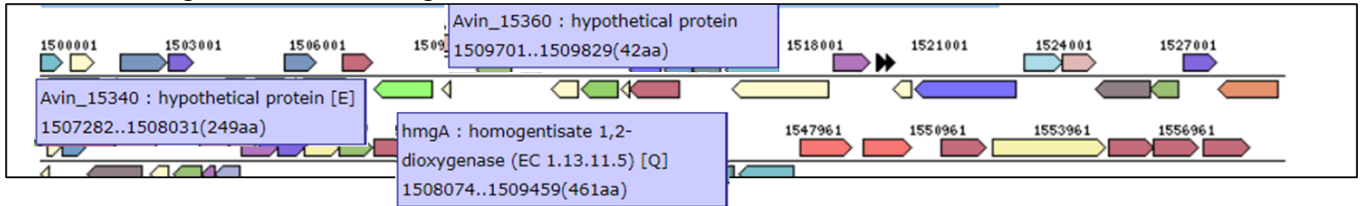


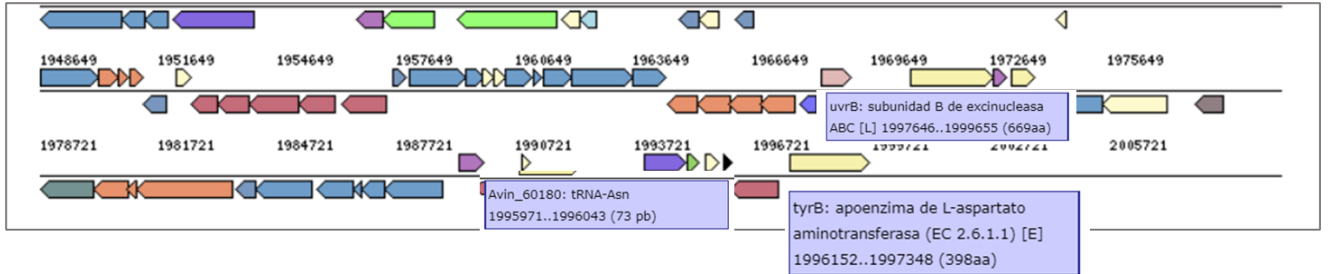
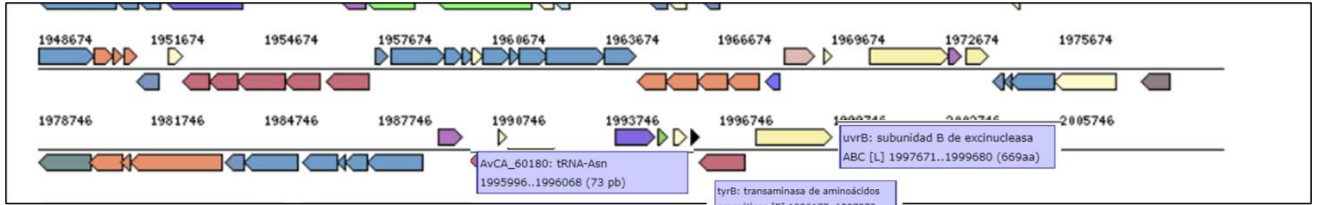


3. Policétido sintasa tipo III



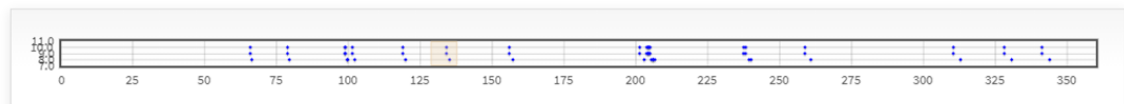
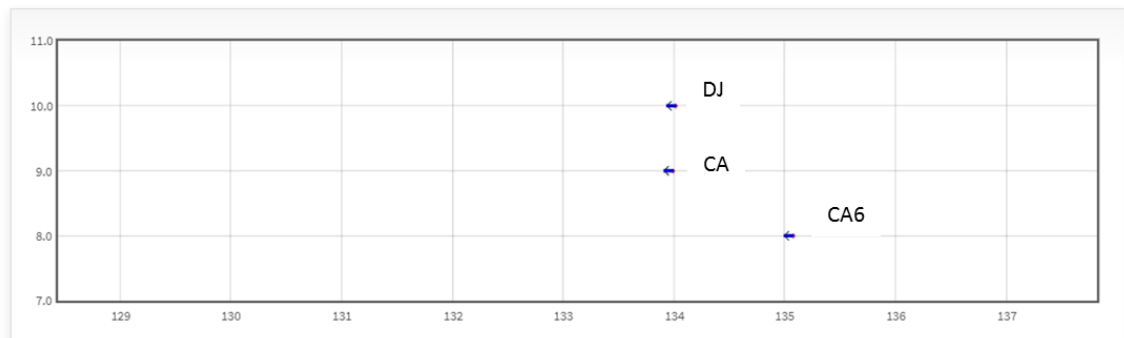
4. Homogentisato 1,2-dioxygenasa

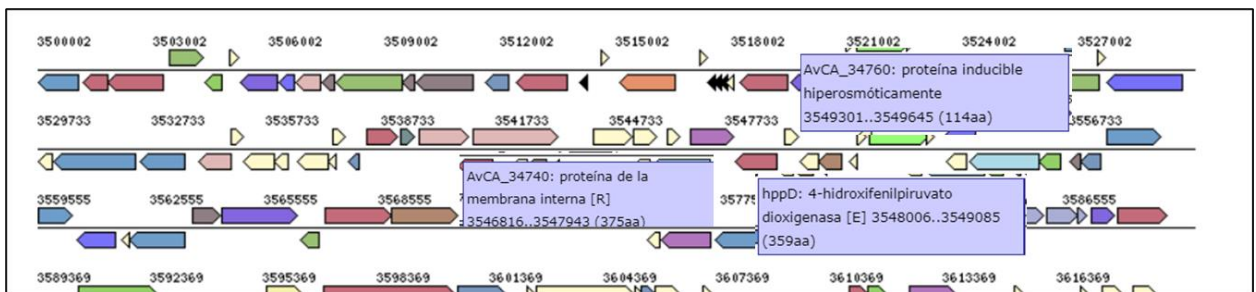
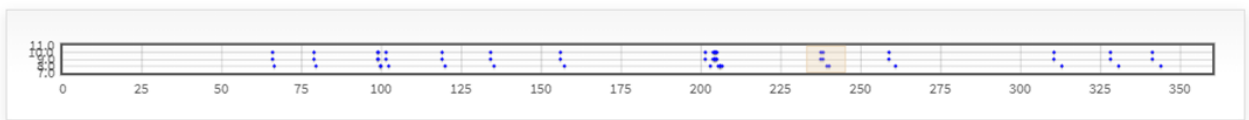
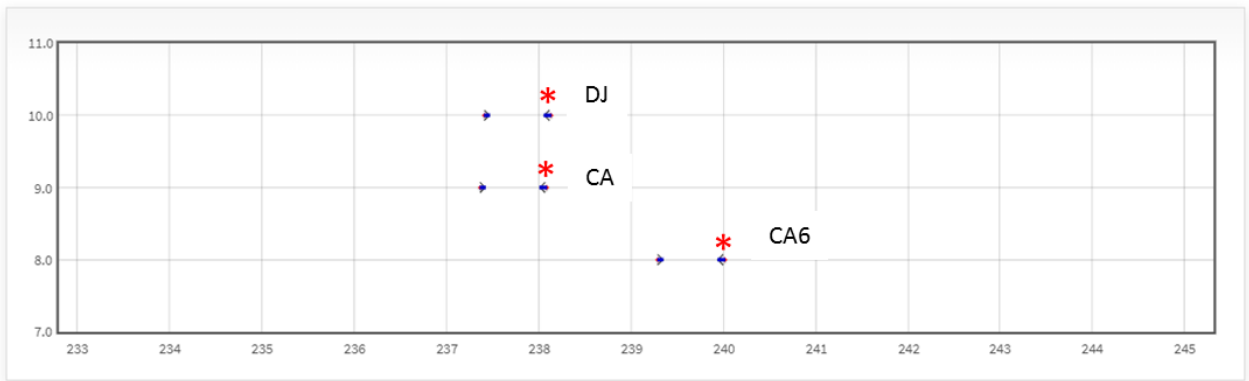


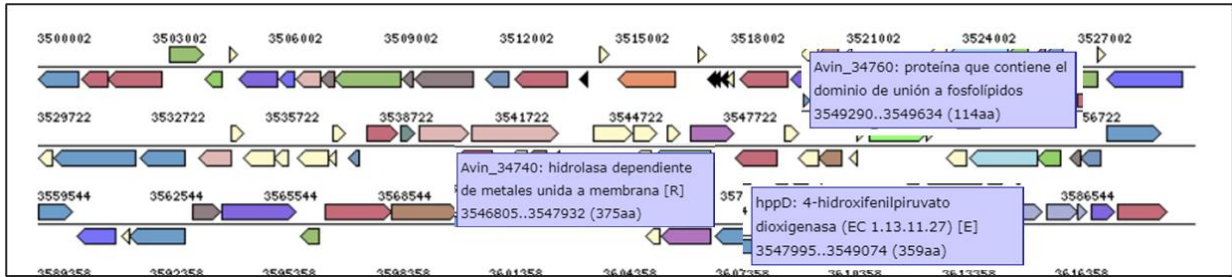


5. Aspartato aromático tirosina aminotransferasa

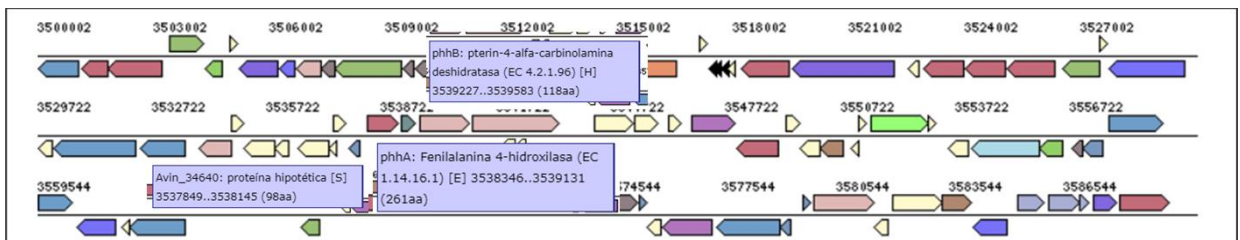
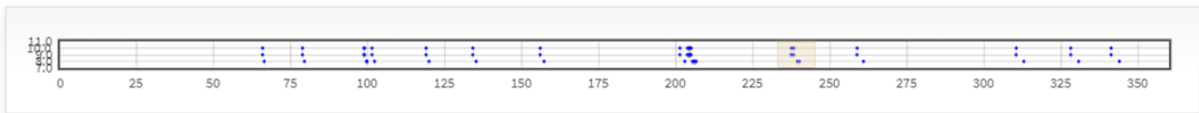
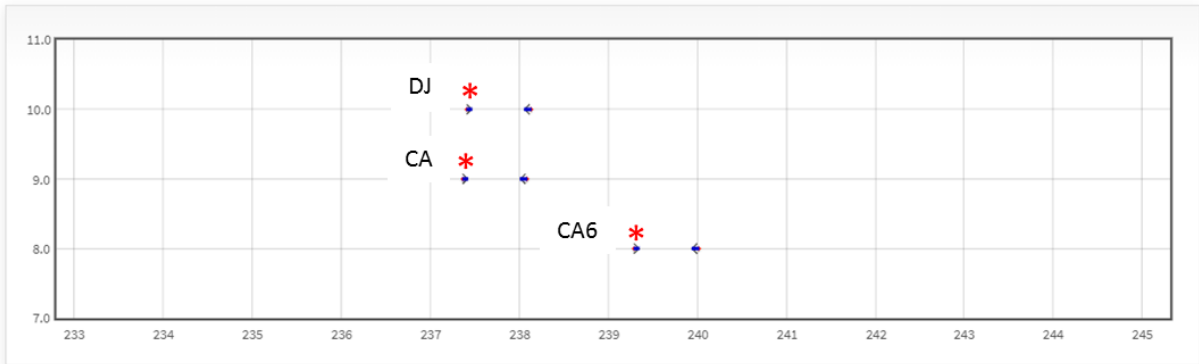
6. 4- hidroxifenilpiruvato dioxigenasa



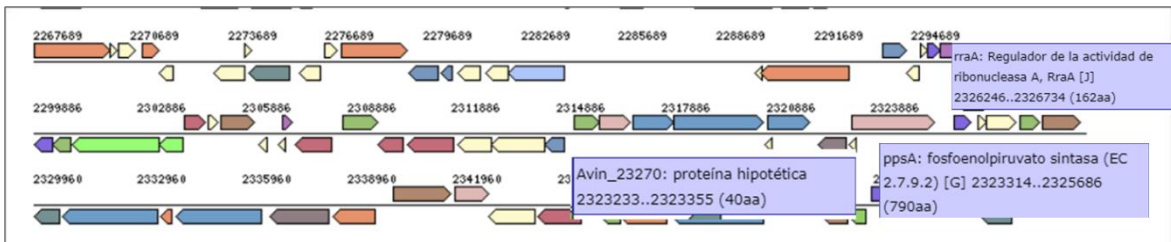
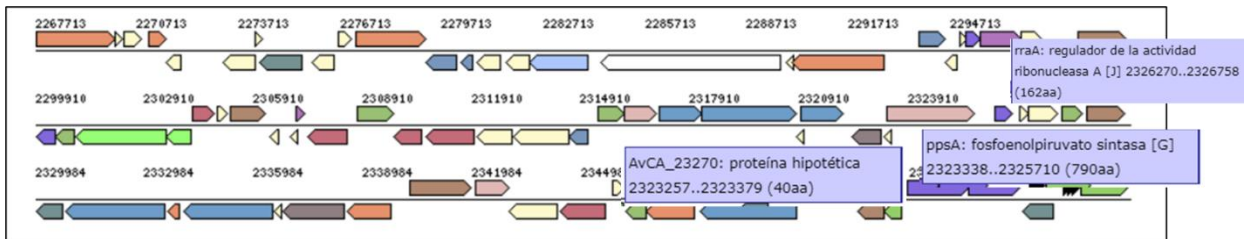
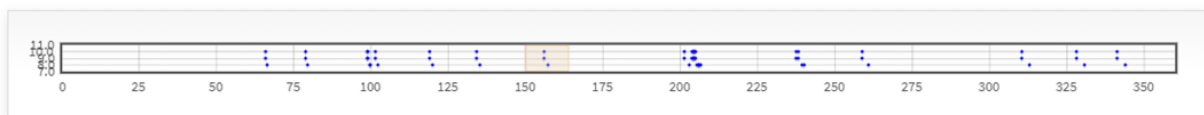
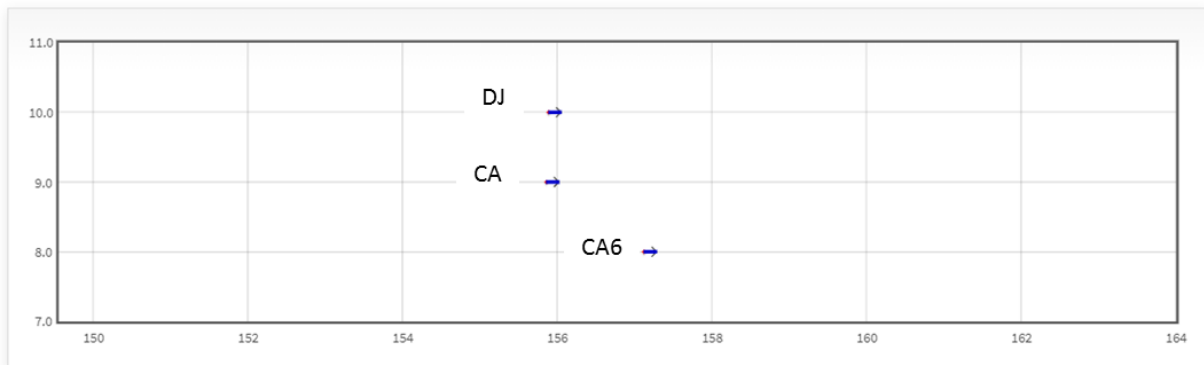


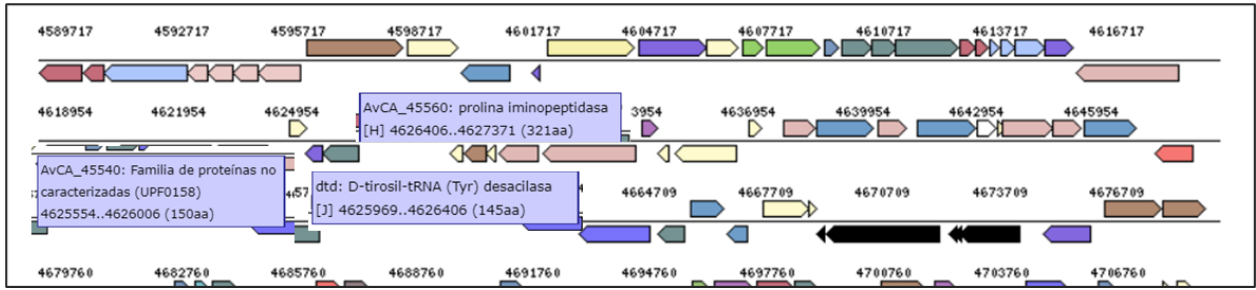


7. Fenilalanina-4-monooxigenasa

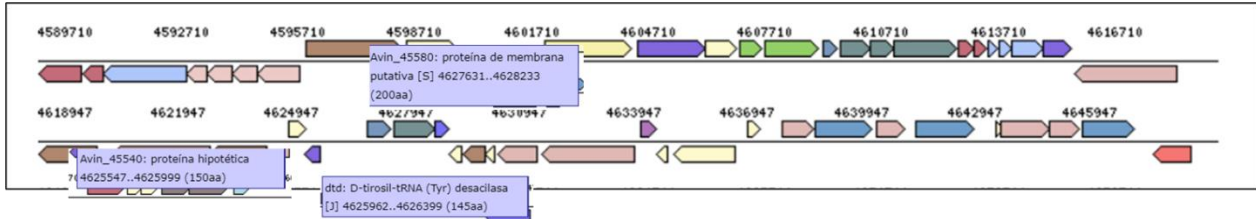


8. Fosfoenolpiruvato sintasa



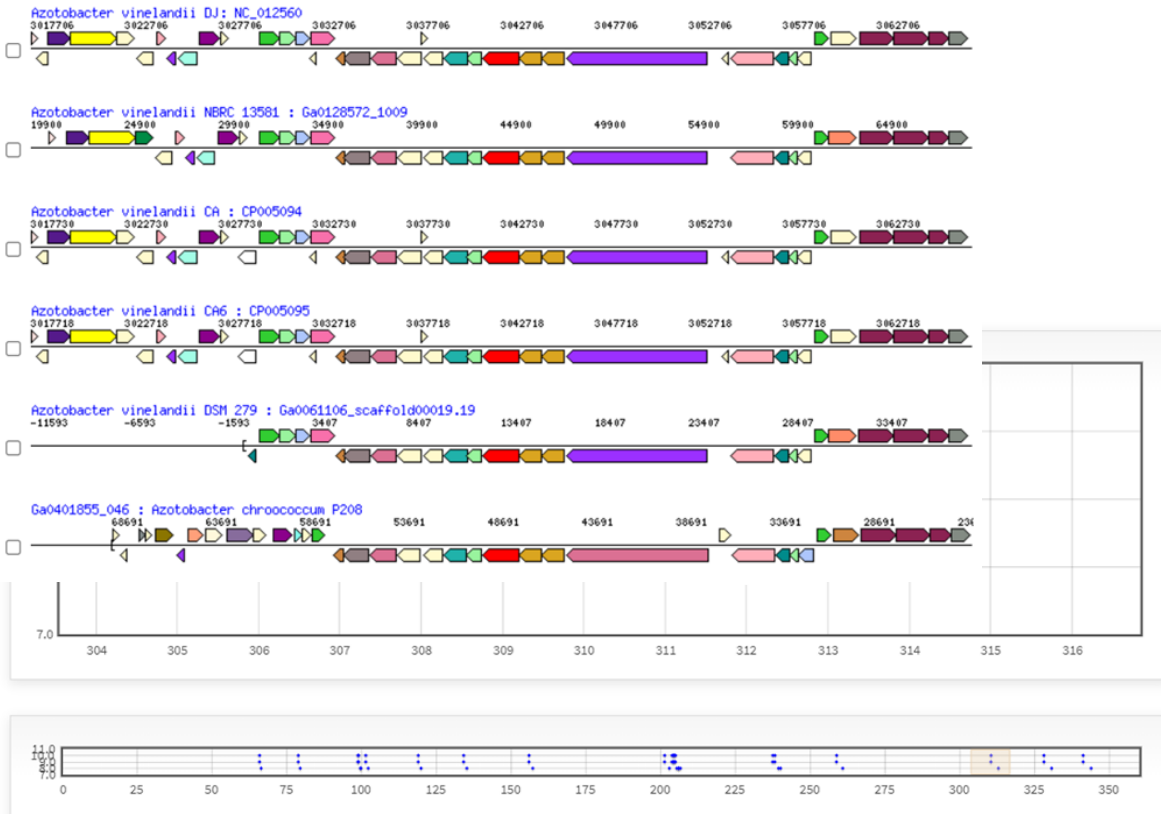


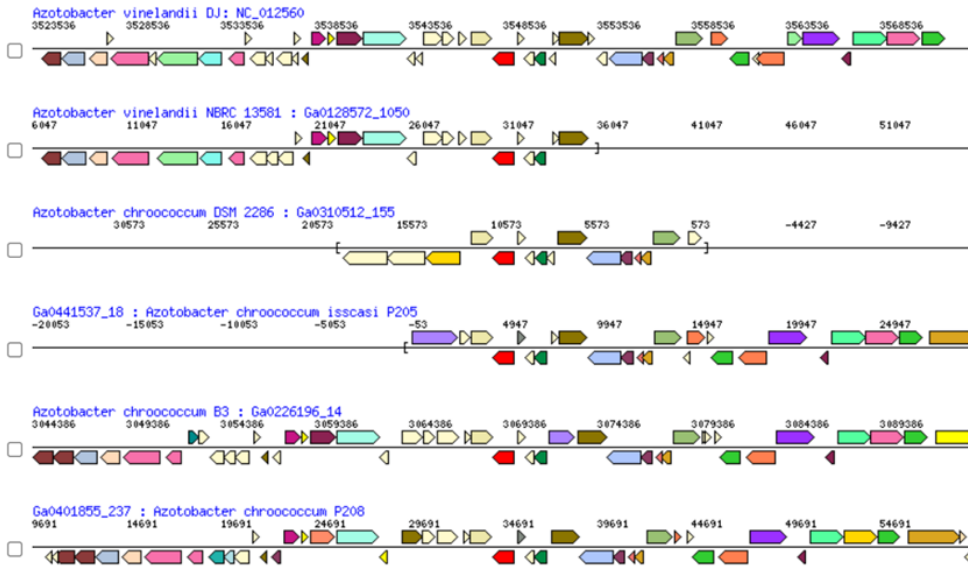
9. D-tirosil-tRNA- (Tir) deacilasa



Sintenia de los nueve genes analizados en la página del IMG-JGI

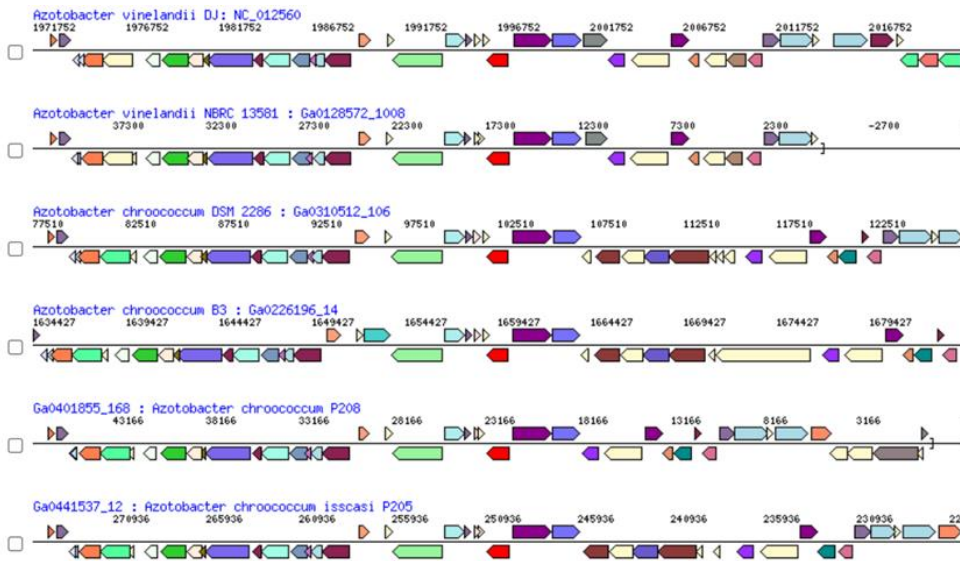
1. Policétido sintasa tipo I



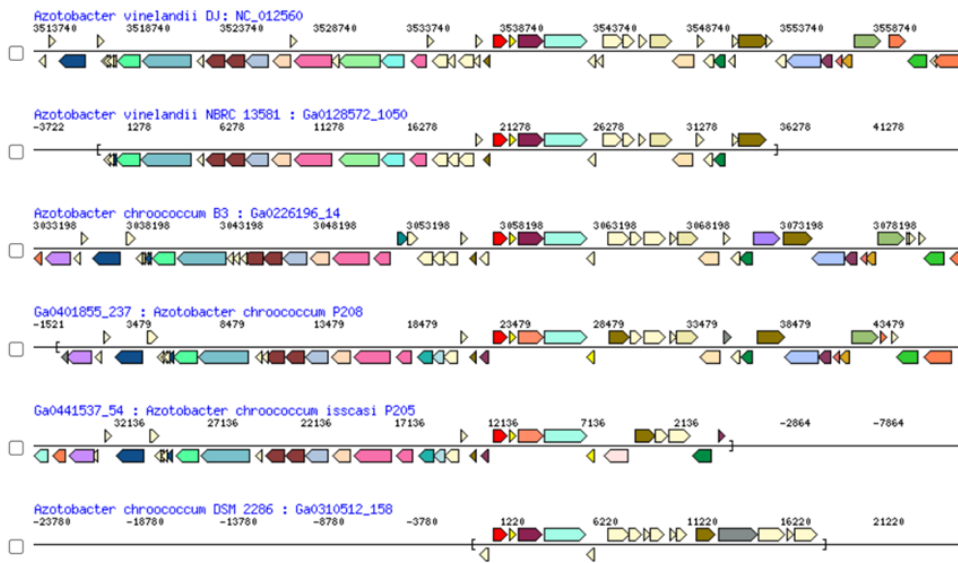


2. Policétido

sintasa deshidratasa

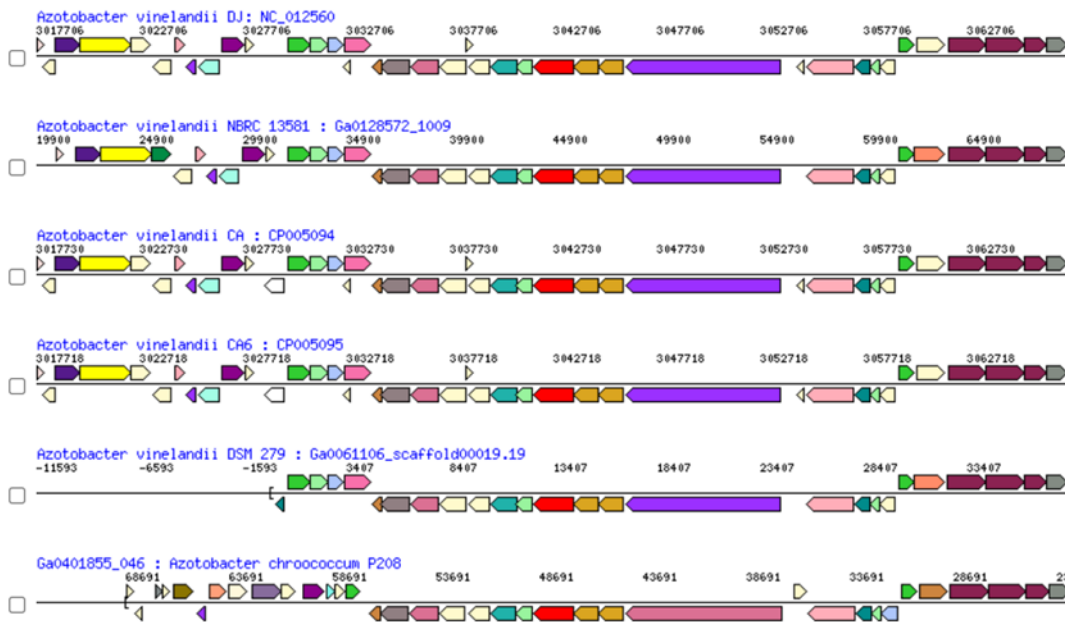


3. Policétido sintasa tipo III



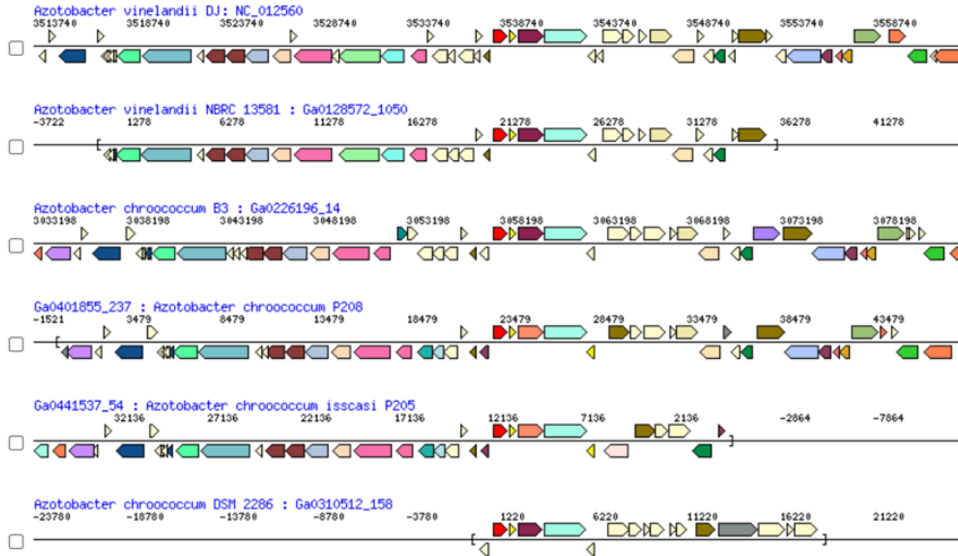
4. Homogentisato 1,2-dioxigenasa

5. Aspartato aromático tirosina aminotransferasa

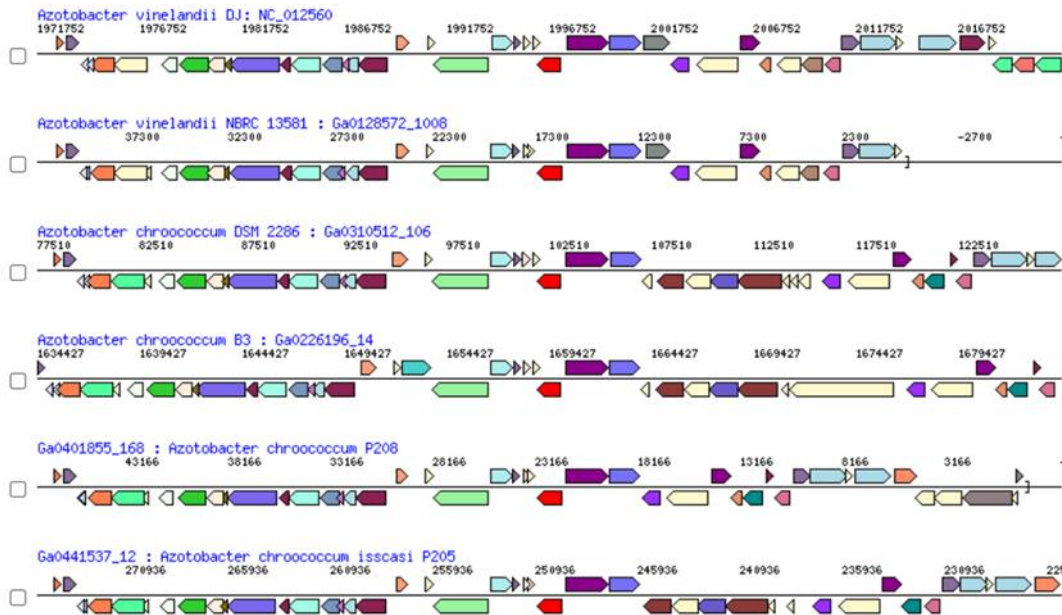


6.4-

hidroxifenilpiruvato dioxigenasa

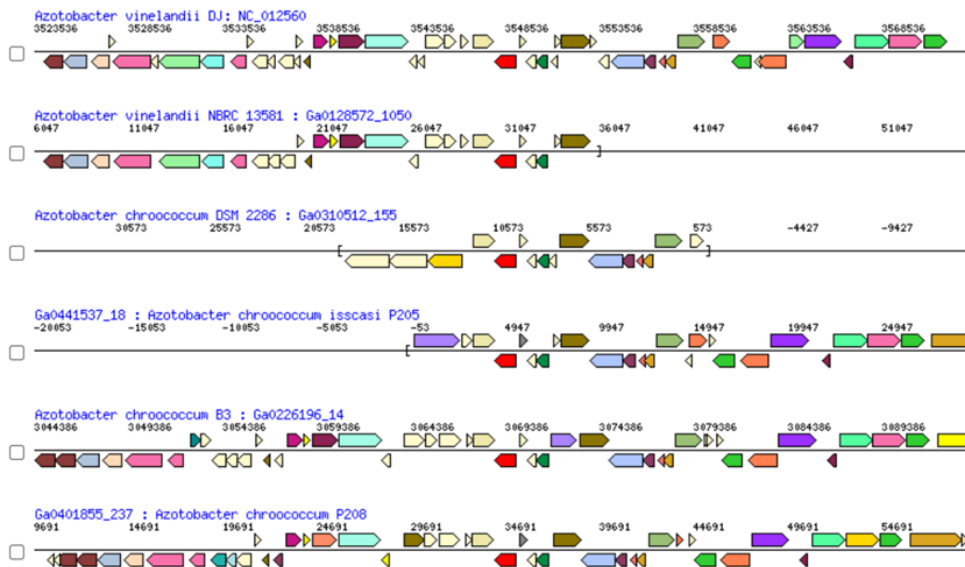


7. Fenilalanina-4-monooxigenasa



8. Fosfoenolpiruvato sintasa

9. D-tirosil-tRNA- (Tir) deacilasa



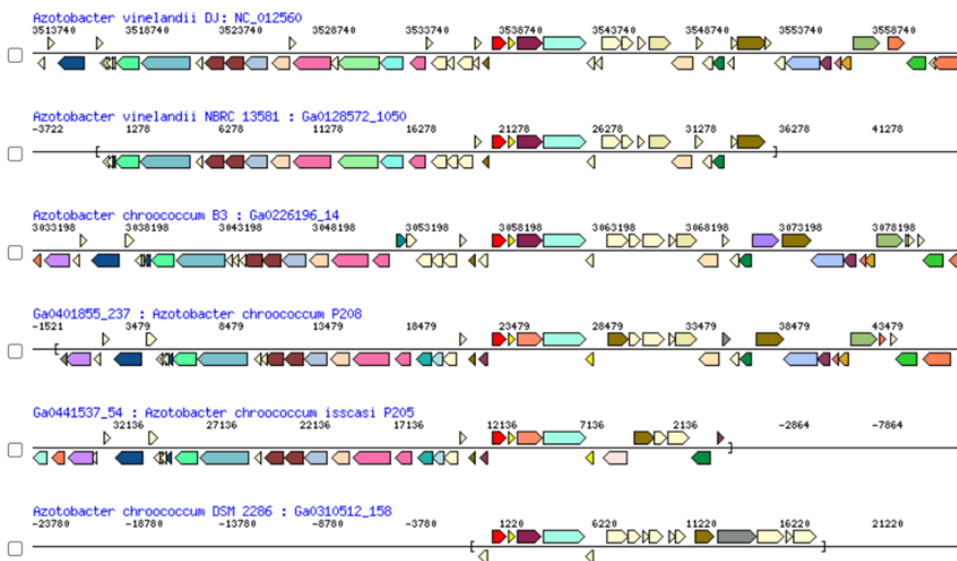
Bibliografía

Banerjee, A., Subhrangshu, S., & Banerjee, R. (2014). Melanin from the Nitrogen-Fixing Bacterium *Azotobacter chroococcum*: A spectroscopic characterization. *PLOS ONE*,

1-7.

Banerjee, A., Supakar, S., & Banerjee, R. (2014).

Melanin from the nitrogen-fixing bacterium *Azotobacter*



chroococcum: A spectroscopic characterization. *PLOSone*.

- Bergey, D. H., Garrity, D. M., Boone, D. R., Castenholz, R. W., & Brenner, D. J. (2001). *Bergey's manual of systematic bacteriology*. Nueva York : Springer.
- Camargo Mancipe, A., Valero González, K., Gomez Rodriguez, A., Camargo Mancipe, D., Suárez Martínez, C., & Cuy Chaparro, L. (2020). Fundamentos y aplicaciones biomédicas de las principales tecnologías de secuenciación . *Revista de Investigación en Salud. Universidad de Boyaca* , 138-172.
- Casañola Martin, G., Marrero Ponce, Y., Le Thi Thu, H., & Torrens, F. (24 de Abril de 2013). *Centro de Información y gestión tecnológica (CITMA)*. Obtenido de 273744-Article%20Text-373579-1-10-20140210.pdf
- Castillo, T., García, A., Padilla, C., Díaz, A., & Peña, C. (2020). Respiration in *Azotobacter vinelandii* and its relationship with the synthesis of biopolymers. *Electronic Journal of Biotechnology*, 36-45.
- Chavez Béjar, M., Balderas Hernández, V., Gutierrez Alexandre, A., Martínez, A., Bolívar, F., & Gosset, G. (2013). Metabolic engineering of *Escherichia coli* to optimize melanin synthesis from glucose. *Microbial cell Factories*, 1-12.
- Dev Jnawali, A., Babu Ojha, R., & Marahatta, S. (2015). Role of *Azotobacter* in soil fertility and sustainability . *MedCrave* , 250-256.
- Edwards, D. J., & Holt, K. E. (2013). Beginner's guide to comparative bacterial genome analysis using next generation sequence data. *Microbial informatics Journal*, 1-9.
- García Gonzalez, D. (2019). *Tesis: Aislamiento y caracterización de cepas de Azotobacter spp. y el efecto de la sobreexpresión de los sRNA: RsmZ1 y RsmZ6 sobre la producción de alginato.*
- Gospodaryov, D., & Lushchak, V. (2011). Some properties of melanin produced by *Azotobacter chroococcum* and its possible application in biotechnology. *Articulos experimentales Biotecnologia*.

- Guillen, R. (2019). Next-generation sequencing technology in relation to our understanding and tackling of Antimicrobial Resistance. *Journal of Sciences and Engineering*, 1-7.
- Hernández, M., Quijada, N., Rodríguez Lázar, D., & Eiros, J. M. (2020). Aplicación de la secuenciación masiva y la bioinformática en el diagnóstico microbiológico clínico . *ScienceDirect*, 150-161.
- Instituto de Biotecnología UNAM. (2012). *Analisis de datos de Secuenciación masiva*. Ciudad de México.
- Jnawali, A. D., Babu Ojha, R., & Marahatta, S. (2015). Role of Azotobacter in soil fertility and sustainability . *Advances in Plants & Agriculture Research*, 250-253.
- Kennedy, C., Rudnick, P., MacDonald, M. L., & Melton, T. (2005). Azotobacter. *Bergey's Manual of Systematics of Archaea and Bacteria*, 3-7.
- Lagunas Muñoz, V. H., Cabrera Valladares, N., Bolívar, F., Gosset, G., & Martínez, A. (2006). Optimum melanin production using recombinant Escherichia coli. *Journal of Applied Microbiology* , 1002-1008.
- Martínez, L., Martínez, A., & Gosset, G. (2021). El enorme potencial de las melaninas como biomateriales en salud y ambiente. *Bioteconología en movimiento*, 3-8.
- Mazinani, Z., & Asgharzadeh, A. (2014). Genetic Diversity of Azotobacter Strains Insolated from Soils by Amplified Ribosomal DNA Restriction Analysis. *Cytology and Genetics* , 293-301.
- Nagpal, P., Jafri, S., Reddy, M. A., & Das, H. K. (1989). Multiple Chromosomes of Azotobacter vinelandii. *Journal of Bacteriology*, 3133-3138.
- Noar, J. D., & Bruno Bárcena, J. M. (2018). Azotobacter vinelandii: the source of 100 years of discoveries and many more to come. *Microbiology society*, 421-436.
- Page, W. J., & Shivprasad, S. (1995). Iron binding to Azotobacter salinestris melanin, iron mobilization and uptake mediated by siderophores. *BioMetals*, 59-64.

- Pavan, M. E., López, N. I., & Pettinari, M. J. (2019). Melanin biosynthesis in bacteria, regulation and production perspectives. *Springer Nature 2019*.
- Plonka, P. M., & Grabacka, M. (2006). Melanin synthesis in microorganisms - biotechnological and medical aspects. *Acta Biochimica Polonica*, 433-450.
- Robson, R. L., Jones, R., Robson, M., Schwartz, A., & Richardson, T. (2015). Azotobacter Genomes: The Genome of Azotobacter chroococcum NCIMB 8003 (ATCC 4412). *PLOS ONE*, 1-35.
- Rojas Tapias, D. F., Ortiz Vera, M. P., Rivera Bolla, D., Kloepper, J., & Bonilla Bultrago, R. R. (2013). Evaluation of three methods for preservation of Azotobacter chroococcum and Azotobacter vinelandii. *Universitas Scientiarum*, 129-139.
- Sanchez Amat, A., Solano, F., & Lucas Elío, P. (2010). Finding New Enzymes from Bacterial Physiology: A Successful Approach Illustrated by the Detection of Novel Oxidases in Marinomonas mediterranea. *Journal of Marine Drugs*, 519-541.
- Saribay, G. F. (2003). *Growth and nitrogen fixation dynamics of azotobacter chroococcum in nitrogen-free and omw containing medium*. The Middle East Technical University.
- Setubal, J., dos Santos, P., Goldman, B. S., Ertesvag, H., Espín, G., Rubio, L. M., . . . Wood, D. (2009). Genome Sequence of Azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *Journal of Bacteriology*, 4534-4545.
- Sharma, P., Goel, R., & Capalash, N. (2006). Bacterial Laccases. *World J Microbiol Biotechnol*, 823-832.
- Shivprasad, S., & Page, W. J. (1989). Catechol formation and melanization by Na-dependent Azotobacter chroococcum: a protective mechanism for aeroadaptation? *Applied and environmental microbiology*, 1811-1817.
- Singh, G., Bhalla, A., Kaur, P., Capalash, N., & Sharma, P. (2011). Laccase from prokaryotes: a new source for an old enzyme. *Sci Biotechnol*, 309-326.

Swarup, R., & Jong, R. (2021). New insight into melanin for food packaging and biotechnology applications. *Taylor & Francis Group*, 2-28.

Tosco Herrera, E. (Julio de 2019). *Análisis de datos en Biocomputación: Tecnología minION y herramientas bioinformáticas*. Obtenido de Universidad de la Laguna: <https://riull.ull.es/xmlui/bitstream/handle/915/15764/Analisis%20de%20datos%20en%20Biocomputacion%20tecnologia%20MinION%20y%20herramientas%20bioinformaticas..pdf;jsessionid=6F10A4DC829B0ED9985BC454C7773647?sequence=1>

Zaidi, K., Ali, A. S., Ali, S. A., & Naaz, I. (2014). Microbial Tyrosinase: Promising Enzymes for Pharmaceutical, Food Bioprocessing and Environmental Industry. *Biochemistry Research International*, 1-17.