



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE QUÍMICA

Ensamble del genoma de *Alicyclophilus* sp. BQ1
mediante bioinformática y técnicas de PCR

TESIS

QUE PARA OBTENER EL TÍTULO DE:
QUÍMICO FARMACÉUTICO BIOLÓGO

PRESENTA

EARVIN JEY HERNÁNDEZ GUZMÁN

DIRECTORA DE TESIS

Dra. HERMINIA DE JESÚS LOZA TAVERA



Ciudad Universitaria, CD. MX., 2018



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

PRESIDENTE: Dr. Rogelio Rodríguez Sotres

VOCAL: M. en C. Luz del Carmen Castellanos Román

SECRETARIO: Dra. Herminia de Jesús Loza Tavera

1er. SUPLENTE: M. en C.B. Sara Margarita Garza Aguilar

2do. SUPLENTE: Dr. Alberto Ortega Vázquez

SITIO DONDE SE DESARROLLÓ EL TEMA:

LABORATORIO 105, CONJUNTO E. DEPARTAMENTO DE BIOQUÍMICA, FACULTAD DE QUÍMICA, UNAM.

Asesor del tema:

Dra. Herminia de Jesús Loza Tavera

Supervisora técnica:

Dra. Lilianha Domínguez Malfavón

Sustentante:

Earvin Jey Hernández Guzmán

Agradecimientos técnicos

Este trabajo fue financiado por los proyectos PAPIIT-DGAPA-UNAM IN217114 e IN223317, además de PAIP Facultad de Química, UNAM 5000-9117.

Esta tesis de Licenciatura se realizó bajo la dirección de la Dra. Herminia Loza Tavera en el laboratorio 105 del Departamento de Bioquímica, en el conjunto E de la Facultad de Química de la Universidad Nacional Autónoma de México. Reconocimiento especial para el Dr. Miguel Ángel Cevallos quien brindó amplia asesoría y apoyo para la secuenciación del genoma y para el ensamblado y al Dr. Luis Lozano por su apoyo en el ensamble del genoma de BQ1, ambos investigadores del Centro de Ciencias Genómicas, UNAM, además por darme asistencia y apoyo en el trabajo bioinformático realizado en esta tesis. De igual manera, al Dr. Jaime García Mena quien nos apoyó con una secuenciación con el sistema Ion Torrent y al M. en C. Otoniel Maya Lucas quien realizó también un ensamble, ambos del CINVESTAV, IPN, CDMX.

Gracias a la Dra. Lilianha Domínguez Malfavón y al Dr. Ayixón Sánchez Reyes, posdoctorantes en el grupo de la Dra. Loza Tavera por todas las enseñanzas en técnicas de biología molecular que me brindaron y que me permitieron llevar a cabo mi trabajo.

Se agradece a los miembros del jurado M. en C. Luz del Carmen Castellanos Román, y Dr. Rogelio Rodríguez Sotres, por las sugerencias brindadas para la mejora de este trabajo.

ABREVIATURAS

| | |
|----------------|--|
| ATP | Adenosina trifosfato |
| BLAST | Basic Local Alignment Search Tool/Herramienta básica de búsqueda de alineamiento local |
| pb | Pares de bases |
| c.b.p. | Cuanto baste para |
| cDNA | Ácido desoxirribonucleico complementario |
| CTAB | Bromuro de hexadeciltrimetilamonio |
| ddNTP | Didesoxinucleótido |
| <i>de novo</i> | Desde el principio |
| DMSO | Dimetilsulfóxido |
| DNA | Ácido desoxirribonucleico |
| dNTP | Desoxirribonucleótidos trifosfato |
| D.O. | Densidad óptica |
| E | Expectancia |
| EDTA | Ácido etilendiaminotetraacético |
| E13 | Ensamble de 13 contigs |
| E15 | Ensamble de 15 contigs |
| E98 | Ensamble de 98 contigs |
| <i>g</i> | Gravedades |

| | |
|----------------|--|
| GC | Citosina y Guanina |
| GYT | Glicerol, extracto de levadura, triptona |
| IBT | Instituto de Biotecnología |
| IPTG | Isopropil- β -D-1-tiogalactopiranosido |
| LB | Medio Luria Bertani |
| MM-PUh | Medio mínimo poliéster poliuretano Hydroform |
| N | Nucleótido |
| NCBI | National Center for Biotechnology Information |
| NGS | Next Generation Sequencing/ tecnologías de secuenciación de nueva generación |
| NMP | N-metilpirrolidona |
| ON | Overnight/toda la noche |
| PCR | Reacción en cadena de la polimerasa |
| PU | Poliuretano |
| RNA | Ácido ribonucleico |
| rpm | Revoluciones por minuto |
| rRNA | Ácido ribonucleico ribosomal |
| rxn | Reacción |
| SDS | Dodecilsulfato de sodio |
| SOC | Caldo superóptimo para represión por catabolito |
| T _m | Temperatura de fusión |

| | |
|-----------|---|
| Tris | Tris(hidroximetil)aminometano |
| t-RNA-ala | Ácido ribonucleico de transferencia de alanina |
| t-RNA-ile | Ácido ribonucleico de transferencia de isoleucina |
| UV | Ultravioleta |
| X-Gal | 5-bromo-4-cloro-3-indolil- β -D-galactopiranosido |
| %Id | Porcentaje de identidad |
| %p/v | Porcentaje peso/volumen |
| % v/v | Porcentaje volumen/volumen |

CONTENIDO

| | |
|---|----|
| Abreviaturas | VI |
| Índice de figuras y tablas | XI |
| 1. Resumen | 16 |
| 2. Introducción | 18 |
| 2.1 El poliuretano y la microbiología como alternativa para tratar sus residuo..... | 18 |
| 2.2 Secuenciación de nueva generación..... | 19 |
| 2.2.1 Ion PGM de Ion Torrent..... | 20 |
| 2.2.2 Sistema 454 de Roche..... | 20 |
| 2.2.3 Genome Analyzer de Illumina..... | 21 |
| 2.3 Ensamblés <i>de novo</i> | 21 |
| 2.4 GenBank..... | 23 |
| 3. Antecedentes | 24 |
| 3.1 Aislamiento y características de <i>Alicyclophilus</i> sp. BQ1..... | 24 |
| 3.2 Otras cepas de <i>Alicyclophilus</i> | 24 |
| 3.3 Secuenciación del genoma de <i>Alicyclophilus</i> sp. BQ1..... | 26 |
| 4. Objetivos | 29 |
| 4.1 General..... | 29 |
| 4.2 Particulares..... | 29 |
| 5. Estrategia experimental | 30 |
| 6. Materiales y métodos | 31 |
| 6.1 Cepas bacterianas..... | 31 |
| 6.2 Vector pGEM®-T Easy..... | 31 |
| 6.3 Reactivación de <i>Alicyclophilus</i> sp. BQ1 y extracción de DNA genómico..... | 33 |
| 6.4 Comparación y selección del ensamblé más probable del genoma de <i>Alicyclophilus</i> sp. BQ1..... | 34 |
| 6.5 Reporte de la secuencia del genoma de <i>Alicyclophilus</i> sp. BQ1 a la base de datos del NCBI..... | 35 |

| | | |
|------------|--|-----------|
| 6.6 | Posible orden de los contigs del E15 comparado con otras cepas del género <i>Alicyclophilus</i> | 36 |
| 6.7 | Curación del E15 de <i>Alicyclophilus</i> sp. BQ1 utilizando secuencias del E13..... | 37 |
| 6.8 | Diseño de iniciadores..... | 38 |
| 6.9 | Amplificación por PCR de regiones intercontigs..... | 39 |
| 6.10 | Purificación de los amplicones obtenidos por PCR..... | 40 |
| 6.11 | Reacción de adenilación al amplicón del GAP 4-2 amplificado por Phusion High-Fidelity DNA Polymerase..... | 41 |
| 6.12 | Ligación de los productos de PCR en el vector pGEM®-T Easy y transformación de células electrocompetentes..... | 41 |
| 6.13 | Secuenciación y análisis de secuencias amplificadas del genoma de <i>Alicyclophilus</i> sp. BQ1..... | 44 |
| 7. | Resultados y discusión | 45 |
| 7.1 | Selección del ensamble del genoma de <i>Alicyclophilus</i> sp. BQ1..... | 45 |
| 7.2 | Publicación del E15 de <i>Alicyclophilus</i> sp. BQ1 en el NCB..... | 50 |
| 7.3 | Posible orden de los contigs del E15 tomando como referencias a <i>Alicyclophilus denitrificans</i> BC y K601..... | 51 |
| 7.4 | Curación del E15 de <i>Alicyclophilus</i> sp. BQ1..... | 56 |
| 7.5 | Análisis de los amplicones intercontigs..... | 60 |
| 7.6 | Análisis de los amplicones secuenciados de BQ1..... | 64 |
| 8. | Conclusiones | 68 |
| 9. | Perspectivas | 69 |
| | Apéndice 1 | 70 |
| | Apéndice 2 | 75 |
| | Apéndice 3 | 81 |
| 10. | Referencias | 94 |

Índice de figuras y tablas

| | |
|--|----|
| Figura 1. Propuesta del orden de los contigs del genoma de BQ1 predicho por el programa Mauve..... | 28 |
| Figura 2. Metodología llevada a cabo en esta tesis..... | 30 |
| Figura 3. Mapa de los sitios de restricción del vector pGEM®-T Easy..... | 32 |
| Figura 4. Secuencia del sitio de clonación múltiple del vector pGEM®-T Easy..... | 32 |
| Figura 5. Alineamiento del E15 y el E13 usando mummerplot..... | 47 |
| Figura 6. Alineamiento del E13 y el E98 usando mummerplot..... | 48 |
| Figura 7. Alineamiento del E15 y el E98 usando mummerplot..... | 49 |
| Figura 8. Arreglo genómico de los contigs del E15 de <i>Alicyclophilus</i> sp. BQ1 basado en el genoma de <i>A. denitrificans</i> BC..... | 52 |
| Figura 9. Arreglo genómico de los contigs del E15 de <i>Alicyclophilus</i> sp. BQ1 basado en el genoma de <i>A. denitrificans</i> K601..... | 53 |
| Figura 10. Posible orden de los contigs predicho por Mauve..... | 54 |
| Figura 11. Alineamientos de los extremos del contig 4 y 2 del E15 con el operón rRNA de <i>Alicyclophilus denitrificans</i> BC..... | 55 |
| Figura 12. Secuencia ensamblada en el contig 3..... | 56 |

| | |
|---|----|
| Figura 13. Secuencia ensamblada en el contig 6..... | 57 |
| Figura 14. Secuencia ensamblada en el contig 11..... | 57 |
| Figura 15. Secuencia traslapada entre el contig 5 y 12..... | 58 |
| Figura 16. Secuencia ensamblada intercontigs 11-10..... | 58 |
| Figura 17. Secuencia ensamblada intercontigs 1-8..... | 59 |
| Figura 18. PCRs para amplificar las regiones intercontigs..... | 61 |
| Figura 19. Ajuste experimental de las PCRs para amplificar las regiones intercontigs..... | 63 |
| Figura 20. PCR del GAP4-2 utilizando la enzima Phusion High-Fidelity DNA Polymerase..... | 64 |
| Figura 21. Orden del nuevo ensamble de <i>Alicyclophilus</i> sp. BQ1 propuesto a partir del análisis realizado en este trabajo..... | 67 |
| Figura A1. Electroferograma del GAP9-3 secuenciado con el iniciador m13/pUC -40 FORWARD..... | 81 |
| Figura A2. Electroferograma del GAP3-8 secuenciado con el iniciador m13/pUC -40 FORWARD..... | 82 |
| Figura A3. Electroferograma del GAP3-8 secuenciado con el iniciador m13/pUC REVERSE..... | 83 |
| Figura A4. Electroferograma del GAP5-6 secuenciado con el iniciador m13/pUC -40 FORWARD..... | 84 |

| | |
|--|----|
| Figura A5. Electroferograma del GAP6-10S secuenciado con el iniciador m13/pUC REVERSE..... | 85 |
| Figura A6. Electroferograma del GAP6-10I secuenciado con el iniciador m13/pUC -40 FORWARD..... | 86 |
| Figura A7. Electroferograma del GAP6-10I secuenciado con el iniciador m13/pUC REVERSE..... | 87 |
| Figura A8. Electroferograma del GAP4-2, clona 1 secuenciado con el iniciador m13/pUC -40 FORWARD..... | 88 |
| Figura A9. Electroferograma del GAP4-2, clona 1 secuenciado con el iniciador m13/pUC REVERSE..... | 89 |
| Figura A10. Electroferograma del GAP4-2, clona 2 secuenciado con el iniciador m13/pUC -40 FORWARD..... | 90 |
| Figura A11. Electroferograma del GAP4-2, clona 2 secuenciado con el iniciador m13/pUC REVERSE..... | 91 |
| Figura A12. Electroferograma del GAP4-2, clona 3 secuenciado con el iniciador m13/pUC -40 FORWARD..... | 92 |
| Figura A13. Electroferograma del GAP4-2, clona 3 secuenciado con el iniciador m13/pUC REVERSE..... | 93 |

| | |
|--|----|
| Tabla 1. Cepas del género <i>Alicyclophilus</i> reportadas en el NCBI (Agosto, 2017)..... | 26 |
| Tabla 2. Tamaño de los contigs del E15..... | 27 |
| Tabla 3. Cepas utilizadas en este trabajo..... | 31 |
| Tabla 4. Iniciadores empleados en este trabajo..... | 38 |
| Tabla 5. Cantidades de reactivo para cada reacción de PCR con <i>Taq</i> DNA polymerase y Phusion High-Fidelity DNA Polymerase..... | 39 |
| Tabla 6. Condiciones de reacción de PCR para la <i>Taq</i> DNA polymerase..... | 40 |
| Tabla 7. Condiciones de reacción para la PCR con la Phusion High-Fidelity DNA Polymerase..... | 40 |
| Tabla 8. Condiciones de reacción para la ligación de los amplicones al vector pGEM®-T Easy..... | 42 |
| Tabla 9. Condiciones de reacción de la digestión enzimática con <i>EcoRI</i> del plásmido..... | 43 |
| Tabla 10. Cepas del género <i>Alicyclophilus</i> reportadas en el NCBI (Agosto, 2018)..... | 51 |
| Tabla 11. Análisis de las secuencias de los amplicones..... | 66 |

Tabla A1. Resultados del BLAST del E15 con respecto a

Alicyclophilus denitrificans BC.....75

Tabla A2. Resultados del BLAST del E15 con respecto a

Alicyclophilus denitrificans K601.....78

1. RESUMEN

El poliuretano (PU) es un polímero sintético con elevada versatilidad. Su síntesis se basa en la gran reactividad del isocianato, el cual se une fácilmente a los ésteres o éter-polióles, generando los grupos uretano, mediante una reacción de condensación. Las propiedades fisicoquímicas del PU le confieren resistencia a la degradación bajo condiciones ambientales, por lo que su descomposición es difícil y muy lenta, además de que los métodos para el tratamiento de residuos de este material no son muy eficientes. Dada esta problemática, se busca desarrollar tecnologías para su biodegradación, por microorganismos capaces de emplear este polímero como fuente de carbono.

En nuestro laboratorio se aisló, de una muestra de espuma de PU deteriorada colectada en un basurero, una bacteria que se identificó como *Alicyclophilus* sp., nombrándola cepa BQ1. Esta bacteria es capaz de crecer en un medio mineral con un barniz de PU como única fuente de carbono. Esta cepa es capaz de atacar los grupos carbonilo, uretano y éter del PU, además de degradar la N-metilpirrolidona (NMP), presente como aditivo en algunos barnices. Otras cepas del mismo género, *Alicyclophilus denitrificans* BC y K601, son capaces de degradar compuestos xenobióticos, pero no pueden degradar ni al PU, ni al NMP.

Con el propósito de identificar los genes y las rutas metabólicas que permitan la actividad degradativa de BQ1 sobre el PU, así como de identificar las diferencias genéticas entre BQ1 y las cepas BC y K601 de *A. denitrificans*, se emprendió un proyecto de secuenciación de su genoma. El DNA genómico de BQ1 se secuenció mediante diferentes tecnologías de secuenciación de nueva generación (Next Generation Sequencing, NGS), obteniendo bibliotecas de 200, 300 y 2000 pb secuenciadas por Illumina, más dos secuenciaciones con Ion Torrent y una con Pirosecuenciación 454. Con estas secuencias y utilizando el programa SPAdes 3.9.0, se generó un ensamble de 15 contigs (E15), el cual fue empleado para determinar un posible orden de los contigs empleando el programa Mauve. Posteriormente, se generó

otro ensamble del genoma, utilizando diferentes programas, dando como resultado un ensamble de 13 contigs (E13).

El objetivo de este trabajo era llegar a obtener un solo ensamble del genoma de *Alicyclophilus* sp. BQ1 y de ser posible, cerrar el genoma, utilizando tanto herramientas bioinformáticas como técnicas de biología molecular (PCR) para lograrlo, así como, reportar este genoma al GenBank. Para ello, primero se definió que el mejor ensamble era el E15, con base en la comparación de cada uno de los ensamblados previos (E15 y E13), con un tercer ensamble realizado con otro programa bioinformático. Se realizó el reporte del genoma de *Alicyclophilus* sp. BQ1 en la base de datos del GenBank, obteniendo el número de acceso NZ_NKDB00000000.1. El Ensamble consta de 15 contigs, 69% de GC y 4.6 Mb totales. Posteriormente, se analizó si el orden de los contigs del E15 obtenido por Mauve podría ser soportado haciendo alineamientos de secuencias con los genomas de *A. denitrificans* BC y K601 y si fuera posible identificar regiones ausentes en E15 y presentes en E13, que permitieran ensamblar los contigs. Esta estrategia bioinformática permitió identificar algunos contigs que si estaban contiguos y alargar en pequeñas secuencias otros, con lo que se redujo el número de contigs a 13.

Posteriormente, utilizando las secuencias de los extremos de cada contig, se diseñaron iniciadores para llevar a cabo reacciones de PCR con los primers de los contigs propuestos como contiguos y obtener amplificadores intercontigs. Así se obtuvieron cinco amplicones: 5,700 pb para GAP4-2, 500 pb para GAP9-3, dos bandas cercanas a 1,400 pb para GAP6-10, 300 pb para GAP5-6 y 800 pb para GAP3-8. Se corrieron por electroforesis los productos amplificados por PCR en geles de agarosa; se seleccionaron y recuperaron las bandas de los tamaños antes mencionados, para su clonación, secuenciación e identificación. Ningún amplicón correspondió a las secuencias que se pretendía amplificar lo que sugiere que los primers tuvieron alineamiento inespecífico con el DNA molde. Al final se logró obtener el mejor ensamble del genoma de *Alicyclophilus* sp. BQ1 y se establecieron metodologías bioinformáticas que son de ayuda para posteriores trabajos relacionados con análisis de secuencias, así como de curación de ensamblados de genomas.

2. INTRODUCCIÓN

2.1 El poliuretano y la microbiología como alternativa para tratar sus residuos

El poliuretano (PU) fue desarrollado por Otto Bayer en 1938. Debido a su gama de propiedades, el polímero se usa ampliamente, en revestimientos líquidos, pinturas, adhesivos, selladores, espumas flexibles y rígidas, además de elastómeros (Center for the Polyurethanes Industry, 2007). El incremento en su manufactura y uso ha acarreado la inevitable acumulación de desechos, en México, en el año 2013, se enviaron mas de 4 millones de toneladas de residuos de plástico a los basureros (Instituto Nacional de Estadística y Geografía, 2016). Los PU se degradan lentamente, estimándose que lleva cientos de años para que las moléculas de PU vuelvan al medio ambiente (Gautam *et al.*, 2007), además, no existen sistemas limpios o efectivos para su eliminación.

Entre las técnicas existentes de disposición de residuos de PU se encuentran: 1) la acumulación en vertederos, 2) el reciclaje mecánico, 3) el procesamiento químico, caracterizado por producir polioles mediante hidrólisis, lamentablemente este procedimiento no es viable económicamente debido al elevado consumo de energía y 4) el procesamiento térmico, el cual provoca gran impacto ambiental pues se generan residuos tóxicos derivados de compuestos nitrogenados (Ignatyev *et al.*, 2014).

El descubrimiento de microorganismos capaces de utilizar PU como fuente de carbono y nitrógeno (Nakajima *et al.*, 1999) han abierto la posibilidad de una industria química mas ecológica. Existen algunas cepas bacterianas que crecen en medios de PU suplementados con extracto de levadura o glucosa. Se han detectado actividades de poliuretanasas proteasa en *Pseudomonas fluorescens* y *P. chlororaphis* (Howard y Blake, 1998), de poliuretanasas lipasa en cepas de *Bacillus subtilis* (Rowe y Howard, 2002) y se han identificado actividades de poliuretanasas esterasa en *Corynebacterium* sp., *Comamonas acidovorans* TB-35 y *P. chlororaphis* (Nakajima *et al.*, 1997). *Comamonas acidovorans* TB-35 tiene la capacidad de atacar el PU sólido y usarlo como fuente de carbono (Nakajima *et al.*, 1995). Los PU de tipo poliéster tienen mayor susceptibilidad al ataque microbiano por lo cual, la biodegradación es una alternativa

factible en proceso de desarrollo para el tratamiento de estos desechos (Morton *et al.*, 1991).

Debido a lo anteriormente descrito, es claro que necesitamos hacer uso de nuevas metodologías para el tratamiento de residuos, y una de las opciones es desarrollar procesos biotecnológicos que permitan la biodegradación de polímeros como el PU, entre otros. La clave para esto es estudiar microorganismos con diferentes actividades enzimáticas que ayuden a su biodegradación y el primer paso es conocer su genoma.

2.2 Secuenciación de nueva generación

Gracias a las tecnologías de secuenciación de nueva generación (Next Generation Sequencing, NGS), la información de secuencias genómicas está al alcance de los investigadores. Las tecnologías de secuenciación de DNA podrían ayudar a los biólogos y proveedores de servicios de salud en una amplia gama de aplicaciones como la clonación molecular, la reproducción, la búsqueda de genes patógenos, así como en estudios comparativos y de evolución. Idealmente, las tecnologías de secuenciación de DNA deben ser rápidas, precisas, fáciles de operar y económicas. Actúan como el motor de la era del genoma, que se caracteriza por una gran cantidad de genomas secuenciados y una amplia gama de áreas de investigación y aplicaciones múltiples (Liu *et al.*, 2012).

Las tecnologías NGS son diferentes del método Sanger, el cual utilizaba una tecnología de secuenciación basada en el método de terminación de cadena, en diversos aspectos como el análisis masivo, alto rendimiento y bajo costo. Sin embargo, una de las desventajas de estos métodos es que se tienen dificultades en la secuenciación precisa de regiones ricas en GC, lo cual afecta la secuenciación de algunos géneros bacterianos con esta característica. Aunque las técnicas de NGS hacen que las secuencias de muchos genomas estén disponibles, el análisis de datos y la interpretación del papel biológico de las proteínas codificadas, son todavía el cuello de botella en su comprensión. Los sistemas NGS están típicamente representados por SOLID/Ion Torrent PGM de Life Sciences, Genome Analyzer/HiSeq 2000/MiSeq de Illumina, y GS FLX Titanium/GS Junior de Roche (Liu *et al.*, 2012). Describiré

específicamente los métodos que fueron empleados para secuenciar el genoma de *Alicyclophilus* sp. BQ1, objeto de este trabajo.

2.2.1 Ion PGM de Ion Torrent

Ion PGM fue lanzado por Ion Torrent a finales de 2010. PGM utiliza la tecnología de secuenciación de semiconductores, la cual se basa en la detección de cambio de pH al agregarse un nucleótido incorporado por la DNA polimerasa. Se coloca la muestra a secuenciar en un chip al cual se le agrega un nucleótido a la vez, si no es el nucleótido correcto, no se generara cambios en el pH; si hay 2 nucleótidos añadidos, se detecta el doble de la señal emitida al agregar ese mismo nucleótido. PGM es la primera máquina de secuenciación comercial que no requiere fluorescencia y escaneo de la cámara, resultando en mayor velocidad, menor costo y menor tamaño del instrumento. Actualmente, permite lecturas de 200 pb en 2 horas. (Liu *et al.*, 2012). Esta tecnología ha entrado en desuso debido a los avances logrados con otras tecnologías.

2.2.2 Sistema 454 de Roche

Este secuenciador usa tecnología de pirosecuenciación. En lugar de usar didesoxinucleótidos para terminar la amplificación de la cadena, la pirosecuenciación se basa en la detección de pirofosfato liberado durante la incorporación de nucleótidos. El DNA con adaptadores específicos del sistema 454 se desnaturaliza en cadenas sencillas y se capturan mediante perlas de amplificación seguidas por PCR en emulsión. Luego, en una placa de picotiter, un dNTP se incorpora a la cadena molde con la ayuda de la DNA polimerasa, ATP sulfurilasa, luciferasa, luciferina y la adenosina 5'-fosfosulfato, liberando pirofosfato. El pirofosfato liberado es igual a la cantidad de nucleótidos incorporados. El pirofosfato producido a partir de ATP lleva a la luciferina a la oxiluciferina y genera luz visible, al mismo tiempo, las bases incomparables son degradadas por la apirasa. Actualmente la longitud de lectura podría llegar a 700 pb con precisión 99.9% después de filtrar y salida de 0.7 Gb de datos por ejecución en 24 horas. La ventaja mas destacada de 454 de Roche es su velocidad: solo tarda 10 horas desde el inicio de la secuenciación hasta la finalización. Una de las deficiencias es que

tiene una tasa de error relativamente alta en términos de interpretación de cadenas de poli bases mayor a 6 pb (Liu *et al.*, 2012).

2.2.3 Genome Analyzer de Illumina

En 2006, Solexa lanzó el Genome Analyzer (GA), y en 2007, la empresa fue comprada por Illumina; el secuenciador adopta la tecnología de secuenciación por síntesis. Esta tecnología utiliza cuatro nucleótidos marcados con fluorescencia, durante cada ciclo de secuenciación, se agrega un único desoxinucleósido trifosfato (dNTP) marcado a la cadena de ácido nucleico. La etiqueta de nucleótidos sirve como un terminador para la polimerización, por lo que, después de cada incorporación de dNTP, se toma una imagen del tinte fluorescente para identificar la base, luego se escinde enzimáticamente para permitir la incorporación del siguiente nucleótido. Dado que los cuatro dNTP reversibles unidos al terminador (A, C, T, G) están presentes como moléculas individuales separadas, la competencia natural minimiza el sesgo de incorporación. La identificación de los nucleótidos se lleva a cabo por la intensidad de señal durante cada ciclo (Illumina, 2016). En 2009, la producción de secuenciación era de 85 Gb/corrida. A principios de 2010, Illumina lanzó HiSeq 2000, su salida fue de 200 Gb/corrida inicialmente y mejoró a 600 Gb/corrida (Liu *et al.*, 2012).

2.3 Ensamblaje *de novo*

Los ensamblajes *de novo* intentan reconstruir la secuencia de DNA de un genoma completo a partir de los datos generados por secuenciación, sin ningún tipo de conocimiento previo a cerca del genoma a ensamblar. Busca lecturas cuyo final coincida con el principio de otra lectura de forma que se puedan unir para formar fragmentos de mayor tamaño hasta completar el genoma. A medida que el costo de secuenciación disminuye, aumenta el número de proyectos que intentan el ensamblaje *de novo*, sin embargo, persisten los problemas ya que no se sabe qué combinaciones de datos de secuenciación y algoritmos de ensamblaje conducen a genomas precisos, ni cómo evaluar la precisión de los nuevos (Baker, 2012).

El ensamblaje del genoma para un nuevo organismo debe convertir los datos de secuenciación en bruto en un genoma completo. Los datos de secuenciación en bruto son una biblioteca de lecturas de DNA superpuestas, de un solo lado o de dos extremos, de cualquiera de las dos cadenas y en cualquier sentido. Con los secuenciadores de segunda generación disponibles, las lecturas únicas son de longitud uniforme y precisa, además cubren el genoma subyacente varias veces, con traslape entre las lecturas adyacentes.

Para ensamblar un genoma, las lecturas de DNA en bruto deben ordenarse linealmente de modo que cada par de lecturas adyacentes se traslape con otras de la misma región. Las secuencias continuamente traslapadas de lecturas cortas forman un contig de DNA, con los contigs ordenados se forma un scaffold de DNA y, posteriormente, el ensamble del genoma resultante. Cuanto mas grande sea el genoma y mas profunda sea su cobertura, será mayor la cantidad de datos a analizar, dificultando en algunas ocasiones la computación de estos para su ensamble (Räihä y Ukkonen, 1981).

Los genomas resultantes de un ensamble *de novo* se evalúan principalmente por el número de contigs en el ensamble, la fracción de lecturas en bruto que se pueden ensamblar y la longitud de los contigs. Otro factor para evaluar su calidad es la relación entre el tamaño estimado del genoma, con relación a la fracción de lecturas en bruto que se pueden ensamblar y la longitud de los contigs (Baker, 2012).

El resultado del ensamblaje final de un genoma varía mucho dependiendo del programa utilizado para llevar a cabo el ensamble y de las tecnologías utilizadas para llevar a cabo la secuenciación del genoma; un ensamble obtenido por diferentes métodos de secuenciación es mas probable que sea mas preciso. Cuando existe un genoma de referencia, este es la mejor base para escribir métricas de precisión (por ejemplo, el grado en que la referencia está cubierta por el conjunto y el porcentaje de contigs que se puede alinear en la referencia). Al comparar los ensambles obtenidos con un genoma de referencia, se encuentran algunos problemas como: muchos contigs (de muy corta longitud), contigs duplicados, compresiones de secuencias repetidas y contigs unidos erróneamente por el ensamblador (Bucur, 2017).

2.4 GenBank

GenBank es una base de datos de secuencias genéticas, una colección anotada de casi todas las secuencias de DNA disponibles públicamente. GenBank es parte de la Colaboración Internacional de Base de Datos de Secuencia de Nucleótidos, que comprende el Banco de Datos de DNA de Japón (DDBJ), el Archivo de Nucleótidos Europeo (ENA) y GenBank en el Centro Nacional de Información Biotecnológica (NCBI). Estas tres organizaciones intercambian datos diariamente.

La base de datos de GenBank está diseñada para proporcionar y fomentar el acceso dentro de la comunidad científica a la información de secuencias de DNA mas actualizada y exhaustiva. NCBI no impone restricciones sobre el uso o la distribución de los datos de GenBank. Sin embargo, algunos remitentes pueden reclamar patentes, derechos de autor u otros derechos de propiedad intelectual sobre la totalidad o una parte de los datos que han enviado (NCBI, 2018).

Hay varias formas de buscar y recuperar datos de GenBank.

- Búsqueda en GenBank por identificadores de secuencia y anotaciones con Entrez Nucleotide, que se divide en tres divisiones: CoreNucleotide (la colección principal), dbEST (Etiquetas de secuencia expresadas) y dbGSS (Secuencias de estudio del genoma).
- Búsqueda y alineamiento de secuencias a una secuencia de consulta usando BLAST (herramienta de búsqueda de alineación local básica).
- Buscar, vincular y descargar secuencias utilizando NCBI e-utilities.

3. ANTECEDENTES

3.1. Aislamiento y características de *Alicyclophilus* sp. BQ1

En el año 2002, en nuestro laboratorio se inició un proyecto de investigación con el objetivo de aislar y caracterizar microorganismos capaces de degradar PU. Para ello, se colectaron muestras de hule espuma deteriorado en el basurero Bordo de Xochiaca (Netzahualcóyotl, estado de México) y se inocularon en un medio de cultivo que contenía un barniz comercial de poliéster PU base agua (Hydroform[®], marca Polyform[®]) (MM-PUh), como única fuente de carbono. Después de algún tiempo en incubación, se aislaron ocho cepas bacterianas (Oceguera-Cervantes, 2005), de las cuales se seleccionaron las tres cepas bacterianas (BQ1, BQ5 y BQ8) que mostraron el mejor crecimiento en MM-PUh, para continuar con su estudio. Estas cepas bacterianas son bacilos cortos Gram negativos, que se desarrollan en un intervalo de temperatura de 30-37°C (Oceguera Cervantes, 2005). El análisis de la secuencia parcial del gen 16S rRNA confirmó que las tres cepas pertenecían al género *Alicyclophilus* sp., de la familia *Comamonadaceae* de las β -proteobacterias (Oceguera Cervantes, 2005; Solís González, 2011). *Alicyclophilus* sp. utiliza como fuente de carbono al NMP y también ataca al poliéster PU. Se logró identificar que una de las actividades enzimáticas relacionadas con la degradación de PU es de esterasa (Oceguera *et al.*, 2007). Debido a que la cepa BQ1 mostró la mayor actividad degradativa sobre PU, se decidió que sería la principal cepa a estudiar en cuanto a la degradación de PU.

Posteriormente, en nuestro laboratorio se ha observado que *Alicyclophilus* sp. BQ1 es capaz de crecer en otro barniz de PS-PU, el Impranil (Fuentes-Jaime, 2015) y en un barniz de PU aromático acrílico, el PolyLack (Savín-Gámez, 2016) como única fuente de carbono. Además, se ha encontrado que BQ1 es capaz de degradar NMP empleándolo como única fuente de carbono (Solís-González *et al.*, 2018).

3.2 Otras cepas de *Alicyclophilus*

El género *Alicyclophilus*, cuenta con una sola especie, *Alicyclophilus denitrificans*. La primera cepa reportada fue la K601 (Mechichi *et al.*, 2003) capaz de degradar

ciclohexanol. La siguiente cepa reportada fue *A. denitrificans* BC, la cual degrada benceno acoplada a la reducción de cloratos (Weelink et al., 2008). Para ambas cepas su genoma ya ha sido secuenciado y reportado. En nuestro laboratorio se determinó que *A. denitrificans* BC y K601 no pueden crecer en medios minerales con PU o NMP como única fuente de carbono, lo que indica que la cepa BQ1 debe tener genes específicos responsables de esta capacidad.

Otra cepa cuyo genoma ya ha sido reportado es *Alicyclophilus* sp. B1, la cual se aisló de una muestra de lodo activado recogido de una fábrica de componentes electrónicos en Japón (Okutsu et al., 2015). Se sabe que la cepa B1 es una bacteria productora de N-acil homoserina lactona, que es una molécula señal de quórum que regulan la expresión de exoenzimas implicadas en los sistemas de tratamiento de aguas residuales. La relación entre la detección de quórum mediada por N-acil homoserina lactona en *Alicyclophilus* sp. B1 y la actividad de tratamiento de aguas residuales aún no se ha dilucidado (Morohoshi et al., 2016).

Entre las cepas en las que aún no se ha reportado la secuencia de su genoma se encuentran algunas con capacidades para utilizar compuestos recalcitrantes como fuente de carbono tales como *Alicyclophilus* sp. KN Bun08, aislada de un cultivo enriquecido en butanona y nitrato. Esta bacteria tiene la capacidad para acoplar la degradación de acetona al proceso de reducción de nitratos (Dullius et al., 2011).

Alicyclophilus sp. se ha encontrado formando parte de diferentes comunidades microbianas con capacidades de degradar compuestos como acetato (Aday et al., 2010), triclosán (Lee et al. 2014) y Nonilfenol, un disruptor endocrino (Bai et al. 2017), o tetraciclina (Taşkan et al., 2016). Además este género ya ha sido utilizado en biorreactores para degradar indol (Hong et al., 2010) y quinolina (Zhang et al., 2017); por lo cual, las cepas de este género se encuentran como potenciales candidatas para su uso en estrategias de biorremediación.

A continuación se muestran en la Tabla 1 los genomas secuenciados y reportados hasta antes de que el genoma de *Alicyclophilus* sp. BQ1 fuera registrado en el GeneBank.

Tabla 1. Cepas del género *Alicyclophilus* reportadas en el NCBI (Agosto, 2017).

| Cepas | Tamaño total del genoma (pb) | Nivel de ensamble | Año de reporte en el NCBI |
|--|------------------------------|---------------------------------|---------------------------|
| * <i>Alicyclophilus denitrificans</i> K601 | 5'070,751 | Completo | 2011 |
| * <i>Alicyclophilus denitrificans</i> BC | 4'835,713 | Completo | 2011 |
| <i>Alicyclophilus</i> sp. B1 | 7'465,959 | Contigs (59) | 2015 |
| <i>Alicyclophilus</i> sp. 69-12 | 4'633,122 | Scaffolds (96) Contigs (397) | 2016 |

* Estas cepas cuentan con plásmidos; K601 tiene el plásmido pALIDE201 (75488 pb); BC cuenta con dos plásmidos, pALIDE01 (119718 pb) y pALIDE02 (78982 pb).

Como se ha visto, contar con la secuencia del genoma de diferentes cepas del género *Alicyclophilus* ha ayudado a estudiar a fondo su actividad metabólica, por lo cual es de gran importancia tener un ensamble del genoma de BQ1 para descifrar en que genes podría residir su capacidad degradativa de PU.

3.3 Secuenciación del genoma de *Alicyclophilus* sp. BQ1

Con el propósito de conocer el genoma de *Alicyclophilus* sp. BQ1, para descubrir las vías metabólicas mediante las cuales puede aprovechar el PU como fuente de carbono, se decidió secuenciar su genoma. Para ello, se realizaron varias secuenciaciones en diferentes laboratorios y por diferentes técnicas, tales como pirosecuenciación 454, Ion Torrent e Illumina en dos ocasiones. Analizando los datos obtenidos se generaron dos ensambles a nivel de contigs, pero en ningún caso se logró ensamblar el cromosoma completo. El primer ensamble fue realizado por el Dr. Luis Lozano y el Dr. Miguel Ángel Cevallos, en el Centro de Ciencias Genómicas, UNAM, usando los resultados de secuenciación del genoma de BQ1 obtenidas por la plataforma Illumina, utilizando el programa SPAdes 3.9.0, se generó un ensamble de 15 contigs (E15). El segundo ensamble fue elaborado por el M. en C. Otoniel Maya Lucas y el Dr. Jaime García Mena, en el CINVESTAV, IPN. Para este segundo ensamble se emplearon los resultados de una nueva secuenciación por Ion Torrent y las otras secuenciaciones

anteriormente realizadas mediante Illumina y el sistema 454 de pirosecuenciación, el resultado fue un ensamble de 13 contigs (E13). Posteriormente, durante mi trabajo experimental, el Dr. Miguel Ángel Cevallos realizó un nuevo ensamble utilizando todos los datos de secuenciación hasta la fecha, ayudado con el ensamblador CLC, generó un tercer ensamble de 98 contigs (E98). Todos los ensambles tienen un tamaño total de 4.6 Mb. Se seleccionó al E15 como ensamble para continuar con el proyecto, en la tabla 2 se presentan los tamaños de cada contig numerados de mayor a menor tamaño.

Tabla 2. Tamaño de los contigs del E15.

| Contig | Tamaño (pb) | Contig | Tamaño (pb) |
|--------|-------------|--------|-------------|
| 1 | 1'847,482 | 9 | 23,292 |
| 2 | 1'431,080 | 10 | 6,132 |
| 3 | 417,414 | 11 | 2,100 |
| 4 | 353,217 | 12 | 767 |
| 5 | 323,056 | 13 | 629 |
| 6 | 154,131 | 14 | 622 |
| 7 | 81,368 | 15 | 511 |
| 8 | 34,454 | | |

Se realizaron algunos análisis con el objetivo de determinar el orden de los contigs y demostrar su veracidad. Como resultado y mediante el uso del programa Mauve, se predijo un orden para los contigs de mayor tamaño, del 1 al 10, llevado a cabo por el M. en C. Daniel Ochoa Gutiérrez (Figura 1). Mauve es software que alinea regiones ortólogas entre dos o más secuencias de nucleótidos, determinando la estructura del reordenamiento global con respecto a un organismo de referencia. Debido a que la recombinación puede causar reordenamientos del genoma, Mauve identifica segmentos conservados, los cuales están libres de reordenamientos del genoma (The Darling lab, 2018). En el caso de BQ1, se realizó tomando como genoma de referencia *A. denitrificans* BC y, posteriormente, a K601 con los mismos resultados.

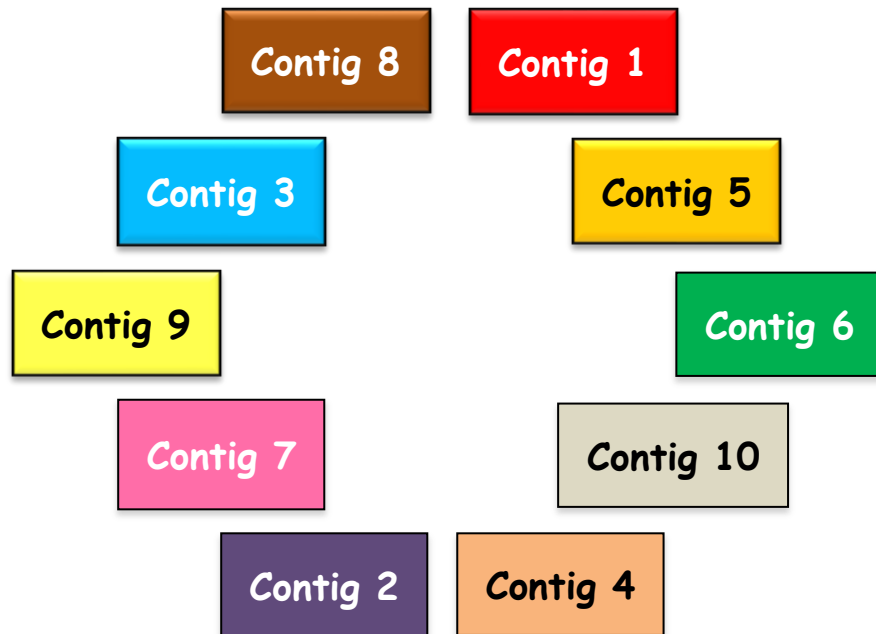


Figura 1. Propuesta del orden de los contigs del genoma de BQ1 predicho por el programa Mauve. Solo se consideraron los contigs del 1 al 10 ya que los contigs del 11 al 15 son de tamaño menor a 5,000 pb y el programa no les puede asignar un orden.

Asimismo, se lograron clonar y secuenciar dos fragmentos entre los contigs 8 y 1 mediante amplificación por PCR y secuenciación por Sanger, llevado a cabo por Ana Bautista Alcántara, en su residencia profesional, lo cual apoya la propuesta de la continuidad de estos dos contigs.

A partir de esta información, este trabajo tuvo como objetivo tratar de determinar el orden de los contigs y cerrar el genoma de BQ1, realizando reacciones de PCR de las regiones intercontigs y reportar el genoma de *Alicyclophilus* sp. BQ1 en el GenBank para que estuviera disponible y anotado para su consulta pública.

4. OBJETIVOS

4.1 General

Determinar la secuencia mas probable del genoma de *Alicyclophilus* sp. BQ1 a partir de las dos propuestas de ensamblajes existentes, empleando herramientas bioinformáticas y PCR, y reportar el genoma al GenBank.

4.2 Particulares

- Proponer un orden de los contigs de *Alicyclophilus* sp. BQ1 mediante alineamientos contra genomas de referencia de bacterias del género *Alicyclophilus*, así como el uso de programas bioinformáticos
- Estimar *in silico* la separación entre los contigs de acuerdo al orden anteriormente propuesto
- Realizar la curación del ensamblaje E15 del genoma de *Alicyclophilus* sp. BQ1
- Diseñar iniciadores para llevar a cabo las reacciones de PCR entre los contigs
- Obtener las condiciones idóneas para cada reacción de PCR, obtener los amplicones y clonarlos para, posteriormente, secuenciar por Sanger
- Analizar bioinformáticamente los resultados de la secuenciación para determinar su ubicación dentro del ensamblaje E15
- Reportar el genoma al GenBank

5. ESTRATEGIA EXPERIMENTAL

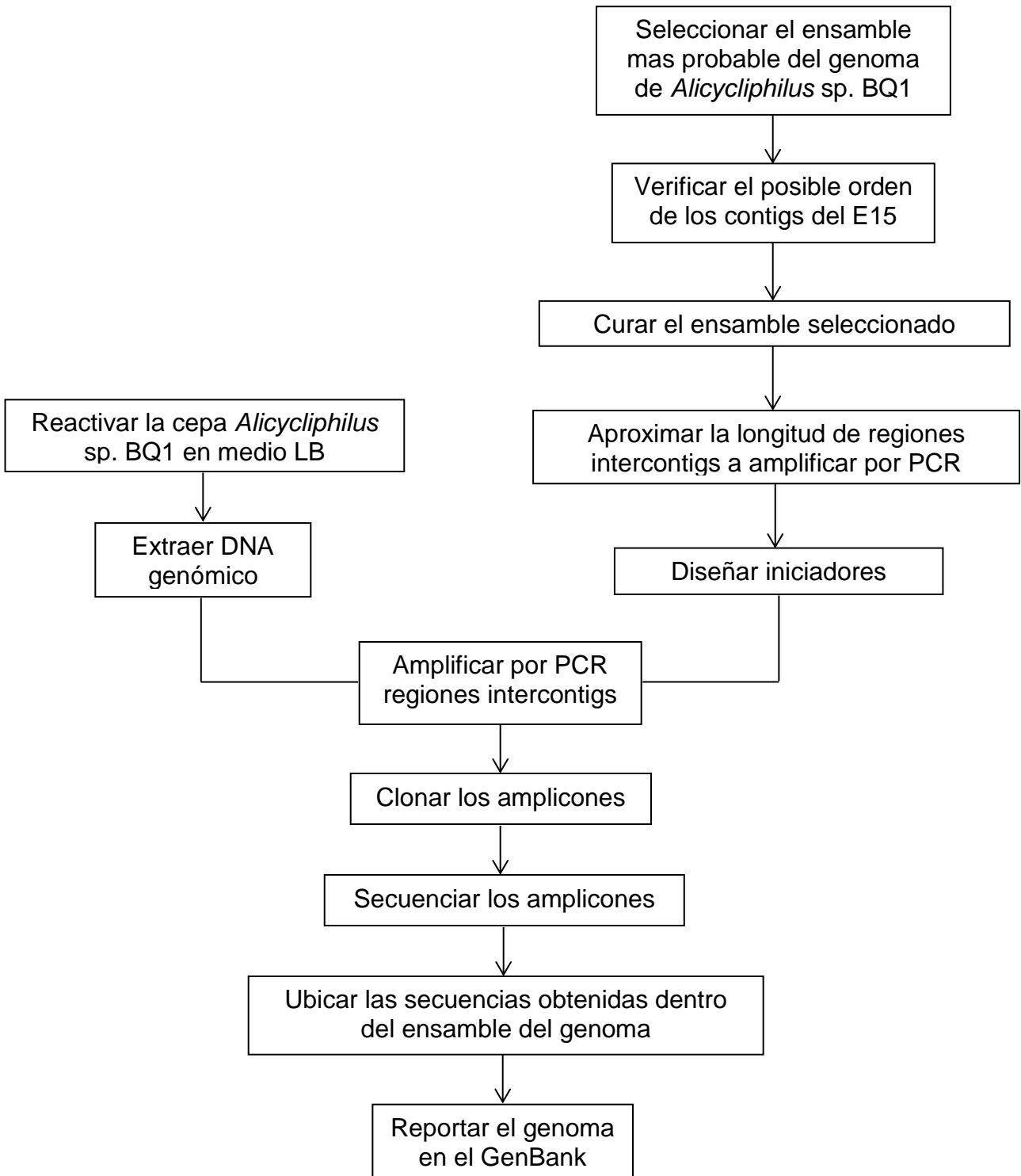


Figura 2. Metodología llevada a cabo en esta tesis.

6. MATERIALES Y MÉTODOS

6.1 Cepas bacterianas

Tabla 3. Cepas utilizadas en este trabajo.

| Cepa | Genotipo | Fuente |
|----------------------------------|--|--|
| <i>Alicyclophilus</i> sp. BQ1 | ----- | Espumas en estado de deterioro, Bordo de Xochiaca, Cd. Nezahualcóyotl, Edo de México, México |
| <i>E. coli</i> cepa DH5 α | F ⁻ <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20</i> ϕ 80d/ <i>lacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>)U169, <i>hsdR17</i> (<i>r_K⁻m_K⁺</i>), λ ⁻ | Cepario del laboratorio 105 del Departamento de Bioquímica, Facultad de Química, UNAM |

6.2 Vector pGEM®-T Easy

El vector pGEM®-T Easy es un vector de alto número de copias que cuenta con un gen de resistencia a ampicilina como marcador de selección (Figura 3). El vector cuenta con una T añadida en ambos extremos 3', lo cual ayuda a que se lleve a cabo la clonación de amplificadores producidos por la *Taq* DNA polimerasa que agrega una adenina al final del amplicón producido (pGEM®-T Easy Vector Systems, 2015). Cuenta también con el sistema de complementación alfa para la identificación de las clonas que adquieren el inserto, en donde tienen el sitio múltiple de clonación con 16 sitios únicos para enzimas de restricción que flanquean la zona donde se inserta el fragmento a clonar (Figura 4).

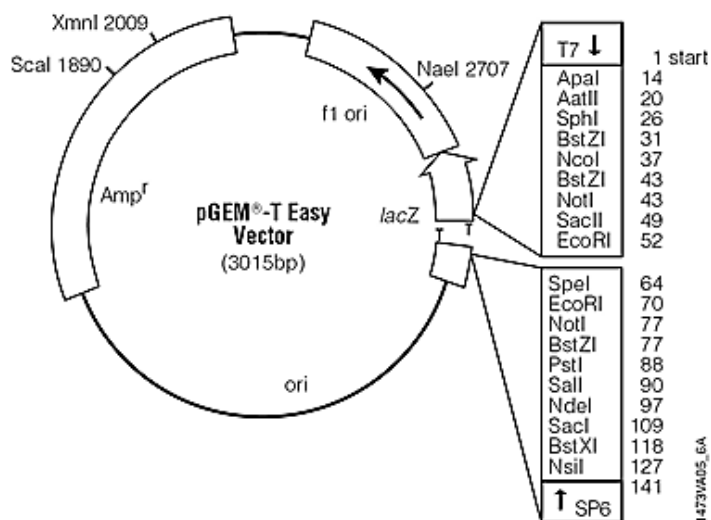


Figura 3. Mapa de los sitios de restricción del vector pGEM®-T Easy.

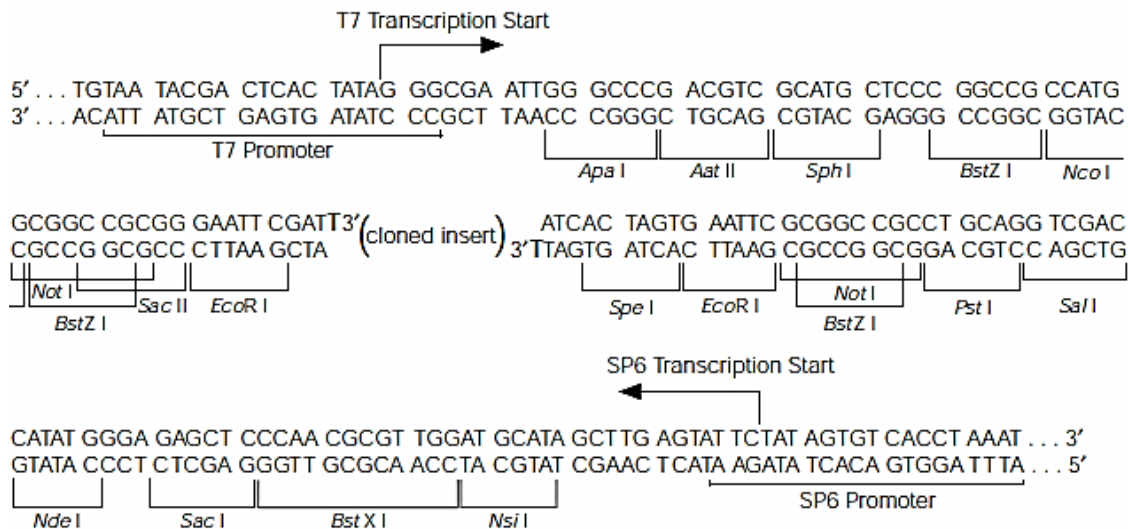


Figura 4. Secuencia del sitio de clonación múltiple del vector pGEM®-T Easy. En el esquema se muestra la ubicación de la T que ayuda a que se lleve a cabo la clonación TA.

6.3 Reactivación de *Alicyclophilus* sp. BQ1 y extracción de DNA genómico

Para tratar de identificar las regiones intercontigs se empleó la técnica de PCR e iniciadores diseñados para cada secuencia de los extremos de cada contig. Para ello primero se extrajo el DNA de la bacteria *Alicyclophilus* sp. BQ1. La extracción de DNA cromosómico se realiza utilizando el método de CTAB (Ausubel *et al.*, 1994) con modificaciones hechas para *Alicyclophilus* sp. BQ1 (Domínguez Cárdenas, 2013). El método consta de los siguientes pasos:

1. Descongelar un microtubo almacenado a $-70\text{ }^{\circ}\text{C}$ con células de *Alicyclophilus* sp. BQ1 preservadas en glicerol al 30% y sembrar por estría en una caja Petri con medio LB sólido. Colocar en una incubadora (THELCO) a $37\text{ }^{\circ}\text{C}$ durante 48 h.
2. A partir de una colonia aislada de BQ1, inocular un tubo con 5 mL de medio LB líquido en un tubo de ensaye. Cultivar overnight (ON) en una incubadora (Thermoscientific™) a $37\text{ }^{\circ}\text{C}$ y 200 rpm.
3. Al día siguiente, inocular estos 5 mL en 20 mL de medio LB en un matraz de 250 mL. Incubar durante 4 horas a $37\text{ }^{\circ}\text{C}$ y 200 rpm, hasta llegar a una absorbencia a 660 nm de 3.
4. Centrifugar el cultivo en cuatro microtubos de 2 mL en una microcentrífuga (Eppendorf) a $10,000\text{ } \times g$, a $4\text{ }^{\circ}\text{C}$, durante 5 min. Decantar el sobrenadante y resuspender completamente el pellet utilizando un vórtex (Scientific Industries) antes de agregar, a cada tubo, 1 mL de buffer TE. Volver a centrifugar, descartar el sobrenadante, y resuspender completamente la pastilla celular por vórtex, hasta que se resuspendan las células.
5. Adicionar 504 μL de buffer TE a cada tubo y resuspender el precipitado con vórtex. Agregar 100 μL de lisozima (2 mg/mL) y 2 μL de RNAsa (25 mg/mL). Incubar a $37\text{ }^{\circ}\text{C}$ y 200 rpm por 60 min.
6. Agregar 30 μL de SDS al 10% e incubar a $37\text{ }^{\circ}\text{C}$ y 200 rpm por 30 min.
7. Agregar 6 μL de proteinasa K (20 mg/mL) e incubar a $37\text{ }^{\circ}\text{C}$ y 200 rpm por 30 min.
8. Adicionar 100 μL de NaCl 5 M, agitar exhaustivamente por 30 s e incubar en un baño de agua (Precision Scientific) a $65\text{ }^{\circ}\text{C}$, por 10 min.

9. Agregar 80 μL de la disolución CTAB/NaCl (CTAB 10% en NaCl 0.7 M), precalentada a 65 °C e incubar en baño de agua, por 20 min.
10. Adicionar 1 volumen de cloroformo: alcohol isoamílico (24:1) y agitar vigorosamente en vórtex (2-5 min). Centrifugar por 10 min en una microcentrífuga (Beckman Coulter) a 10,000 x g a temperatura ambiente.
11. Tomar el sobrenadante y pasarlo a tubos limpios. Adicionar 1 volumen igual de fenol: cloroformo: alcohol isoamílico (25:24:1). Agitar en vórtex por 5 min, formando una emulsión constante. Centrifugar 5 min a 14,000 x g a 4 °C.
12. Tomar el sobrenadante y pasarlo a un tubo limpio. Adicionar 1 volumen de cloroformo: alcohol isoamílico (24:1). Agitar por 3 min en vórtex, formando una emulsión constante. Centrifugar 5 min a 14,000 x g a temperatura ambiente.
13. Tomar la fase acuosa y pasarla a un tubo nuevo. Adicionar 2 volúmenes de etanol frío. Agitar el tubo por inversión e incubar ON a 4 °C. Centrifugar a 10,000 x g , durante 10 min, a temperatura ambiente.
14. Retirar el sobrenadante sin perder el precipitado. Lavar el pellet 2 veces con 500 μL de etanol al 70% frío. Centrifugar 5 min a 14,000 x g a temperatura ambiente y decantar completamente. Resuspender en 100 μL de agua inyectable.
15. Correr una alícuota en un gel de agarosa al 1% para confirmar la presencia de DNA genómico y la ausencia de RNA. Leer la absorbencia de una alícuota a 230, 260, 280 nm en un espectrofotómetro BioDrop.

6.4 Comparación y selección del ensamble mas probable del genoma de *Alicyclophilus* sp. BQ1

Con el objetivo de reportar el genoma de *Alicyclophilus* sp. BQ1 en una base de datos oficial (GenBank) para que estuviera disponible para ser consultado por la comunidad científica y como contábamos con dos ensamblajes del genoma era necesario elegir el mejor de ellos para reportarlo. Para esto, se implementó una metodología nueva basada en la comparación con otro alineamiento *de novo* con el cual se alinearon cada uno de los ensamblajes, así como comparaciones entre ellos para observar similitudes y diferencias entre ellos.

1. Realizar alineamientos de nucleótidos con el programa mummerplot, en los cuales se tomen en cuenta todos los contigs de ambos ensamblajes del genoma de *Alicyclophilus* sp. BQ1 (E15 y E13).
2. Decidir en qué partes pueden complementarse ambos ensamblajes por métodos bioinformáticos (coincidencia de secuencia de nucleótidos por alineamientos).
3. Realizar un nuevo alineamiento de nucleótidos usando como referencia E15 y E13 contra otro ensamblaje del genoma de *Alicyclophilus* sp. BQ1 que consta de 98 contigs.
4. Seleccionar el ensamblaje que mejor se acople al ensamblaje de 98 contigs.

6.5 Reporte de la secuencia del genoma de *Alicyclophilus* sp. BQ1 a la base de datos del NCBI

1. Entrar a Submissions | Genome | Submission Portal, que se encuentra en la siguiente dirección electrónica: <https://submit.ncbi.nlm.nih.gov/subs/genome/>.
2. Iniciar sesión en caso de contar con una cuenta, de no ser así, crearla.
3. En la esquina superior derecha se encuentran varias opciones, de las cuales se debe elegir la llamada *MY SUBMISSIONS*.
4. En la esquina superior derecha se encuentra un apartado donde dice Genome y enseguida un botón llamado *New submission*, pulsarlo.
5. Lo primero que hay que elegir es el modo en el cual se desea enviar los datos, se debe elegir la opción *Single genome*, lo cual nos permite llenar un formulario manualmente sobre los datos que vamos a enviar.
6. El primer apartado se refiere a los datos de la persona que envía los datos. Los datos requeridos en este apartado son: Nombre, dos correos electrónicos, teléfono, fax y la dirección de la institución a la que pertenece la persona que sube los datos (en nuestro caso el Departamento de Bioquímica; Facultad de Química, UNAM).
7. En información general, se pregunta si ya se cuenta con un BioProject y un BioSample relacionados al genoma en cuestión, además de cuándo debe de liberarse públicamente la información del genoma.
8. Mas adelante se encuentra la información sobre el genoma, en la cual se pregunta por: fecha en la que se realizó el ensamblaje, el programa con el cual se realizó y que versión es, nombre del ensamblaje, cobertura del genoma, método de secuenciación

del genoma, si en la muestra se incluye el genoma completo, si se trata de la versión final, si es un *ensamble de novo* y si es una actualización de un genoma ya existente.

9. Al no contar con un BioProject se llena la información para que este se genere. Los datos requeridos son: descripción de los objetivos de estudio, así como su relevancia.
10. A continuación se llenan los datos del BioSample si no se cuenta con uno. Primeramente, se elige una opción que describa la muestra, posteriormente, los datos que se necesitan son: nombre de la muestra, organismo, cepa, hospedero, identificación del aislamiento, lugar de aislamiento, fecha de muestreo, localización geográfica del lugar de la toma de muestra y tipo de muestra.
11. Fuente es el siguiente apartado. En esta pregunta si se desea que el NCBI anote el genoma antes de ser liberado. La opción seleccionada fue que sí.
12. La siguiente pregunta acerca de cuál de tres opciones describe mejor el envío del genoma (si hay uno o más cromosomas y si se agregan secuencias adicionales a ellos). A continuación, se selecciona el archivo que contiene la secuencia del genoma en formato FASTA o archivo de texto.
13. En asignación se pregunta si alguna de esas secuencias pertenece a un cromosoma completo o si pertenece a un plásmido, seleccionando cromosoma completo ya que se realizó la secuenciación del genoma completo, sabiendo que BQ1 no contiene plásmidos.
14. En referencias se coloca el nombre del autor(es) de la secuenciación, estado de una publicación que haga referencia a este genoma, el nombre de la publicación y sus autores.
15. Por último, se muestran algunos datos previamente llenados durante el formulario, si todos son correctos, se pulsa el botón *Submit* para finalizar.

6.6 Posible orden de los contigs del E15 comparando con otras cepas del género *Alicyclophilus*

El siguiente paso después de elegir un ensamble es tratar de darle un orden lógico a los contigs ensamblados. La metodología utilizada se basa en una comparación por BLAST

(*Basic Local Alignment Search Tool*/Herramienta básica de búsqueda de alineamiento local) de la secuencia del genoma de BQ1 con los genomas de otras cepas del género *Alicyclophilus* cuyos genomas estén ensamblados en un solo cromosoma.

1. Seleccionar fragmentos de diferentes longitudes de cada uno de los contigs dependiendo del tamaño de cada uno de estos. Para los contigs 1-6 se utilizaron 50,000 pb; para el contig 7, 200,000 pb; para el contig 8, 7,000 pb; para los contigs 9 y 10, 6,000 pb; y para los contigs 11-15 se alinearon por completo (2,100-511 pb).
2. En la página base del NCBI, realizar BLAST de los fragmentos de cada uno de los contigs, usando como genomas de referencia, dos diferentes cepas de *Alicyclophilus denitrificans* (BC y K601).
3. Realizar una lista de la posición que ocupa cada fragmento de cada contig con respecto a cada genoma de referencia dependiendo de las coordenadas de los alineamientos. Se tomará como criterio de selección la opción con el menor valor de expectancia (E), el mayor porcentaje de cobertura y con un valor de similitud mayor al 95%.
4. Asignar un orden a cada segmento alineado de acuerdo a las coordenadas obtenidas en cada genoma de referencia.
5. De acuerdo al orden anterior, calcular la posible longitud de los segmentos de nucleótidos que se podrían encontrar entre los extremos de los diferentes contigs que se estima son contiguos.

6.7 Curación del E15 de *Alicyclophilus* sp. BQ1 utilizando secuencias del E13

Debido a que se encontró que existen regiones que complementan ambos ensambles entre sí (sección 7.1), se realizaron alineamientos de nucleótidos entre secuencias concordantes de ambos ensambles. A partir de ellos, se realiza una curación del ensamble E15, lo cual quiere decir en este caso que se mejorará al lograr disminuir el número de contigs pertenecientes a este.

1. Ubicar las secuencias terminales de cada contig del E15 dentro del E13.
2. Buscar coincidencias en las cuales dos extremos terminales de diferentes contigs del E15 se encuentren dentro de un solo contig del E13.

3. Realizar alineamientos de nucleótidos con el programa Vector NTI de las secuencias que coinciden en ambos ensamblados. En caso de tener un alineamiento con un porcentaje de similitud del 95% o mayor, las secuencias de los dos diferentes contigs se unirán y formarán un nuevo contig.
4. Realizar un BLAST con cada secuencia nueva agregada al ensamblado.

6.8 Diseño de iniciadores

Se diseñaron iniciadores para los extremos terminales de los contigs 1-10 para realizar la PCR para amplificar posibles regiones desconocidas intercontigs del E15. Utilizando el programa Vector NTI se consideraron diversos criterios para su diseño, entre los que se encuentran: una longitud entre 17-22 pb, Tm entre 52-55 ° C, el par de iniciadores para cada PCR no debe tener una diferencia mayor a 2 ° C, %GC entre 45-70%, no formar más de 3 dímeros y no formar más de 1 loop (Tabla 4).

Tabla 4. Iniciadores empleados en este trabajo.

| Primer | Secuencia | Tm | %GC | Dímeros | Loops | pb |
|---------|-------------------------|------|------|---------|-------|----|
| J5Rout | TCGATATTCTTGTTTCAGCAGCC | 53.7 | 45.5 | 2 | 0 | 22 |
| J5Fout | GTGGTCGAGGGCATCCAG | 53.3 | 66.7 | 3 | 1 | 18 |
| J6Rout | AGAACTGGTGCGGGTAGGC | 54.0 | 63.2 | 0 | 0 | 19 |
| J6Fout | CTGGCGGTGAAGTCCAGC | 53.7 | 66.7 | 2 | 1 | 18 |
| J10Rout | CTTGCCGTCAGCACCTC | 53.7 | 66.7 | 1 | 0 | 18 |
| J10Fout | GCTTCTTGCTGCCTTGCC | 53.3 | 61.1 | 0 | 0 | 18 |
| J4Rout | TCTTGCGGTCCATCTCCG | 54.1 | 61.1 | 2 | 1 | 18 |
| J4Fout | TTGGAAGTGCATGGGCG | 53.3 | 58.8 | 2 | 0 | 17 |
| J2Rout | AACTTGATGGTTGCGGCG | 54.3 | 55.6 | 2 | 1 | 18 |
| J2Fout | CCAGACGGCATTGCTCG | 53.0 | 64.7 | 2 | 0 | 17 |
| J7Rout | CCCTCCAGCAGAAAGAAGCTC | 54.7 | 57.1 | 3 | 1 | 21 |
| J7Fout | CGCTGATGGTGTAGGTGGC | 53.7 | 63.2 | 0 | 0 | 19 |
| J9Rout | GAGCTTGATGGTTGCGGC | 53.9 | 61.1 | 1 | 0 | 18 |
| J9Fout | TTGGGCTTTAGGGCTTGC | 52.8 | 55.6 | 0 | 0 | 18 |
| J3Rout | CCTGGGCTTGCTGGTAAGC | 53.8 | 60.0 | 2 | 1 | 20 |
| J3Fout | GCCTGAGAGCCTGTTCAAGATC | 54.6 | 54.5 | 3 | 1 | 22 |
| J8Rout | CACTGGCTGCTGCTGGC | 53.2 | 70.6 | 0 | 0 | 17 |
| J1Fout | ACAGATCCGCCTCCAGTCC | 53.6 | 63.2 | 1 | 0 | 19 |

6.9 Amplificación por PCR de regiones intercontigs

Con base en los resultados obtenidos del análisis bioinformático sobre el posible orden de los contigs, se emplearon los iniciadores anteriormente descritos para llevar a cabo reacciones de PCR para amplificar regiones intercontigs en el genoma de BQ1.

1. Utilizar DNA genómico extraído de *Alicyclophilus* sp. BQ1 mediante el método descrito en la sección 6.3. Para este trabajo se emplearon dos diferentes polimerasas: Taq DNA polymerase (Thermoscientific™) y Phusion High-Fidelity DNA Polymerase (Thermoscientific™).
2. Realizar las reacciones de PCR en el termociclador (Applied Biosystems™, Veriti, Modelo 9902) empleando un volumen total de 10 µL con cantidades de reactivo y condiciones preestablecidas que han demostrado amplificar genes de *Alicyclophilus* sp. BQ1, según se describe en las Tablas 5, 6 y 7.

Tabla 5. Cantidades de reactivo para cada reacción de PCR con Taq DNA polymerase y Phusion High-Fidelity DNA Polymerase.

| Reactivo | Taq DNA polymerase | | Phusion High-Fidelity DNA Polymerase | |
|----------------------|--------------------|--------------|--------------------------------------|--------------|
| | [Stock] | [rxn] | [Stock] | [rxn] |
| Taq Buffer | 10X | 1X | --- | --- |
| 5X Phusion GC Buffer | --- | --- | 5X | 1X |
| dNTPs | 10 mM | 0.2 mM | 10 mM | 0.2 mM |
| *Iniciador directo | 20 µM | 1 µM | 20 µM | 1 µM |
| *Iniciador reverso | 20 µM | 1 µM | 20 µM | 1 µM |
| MgCl ₂ | 25 mM | 2.5 mM | 25 mM | 1 mM |
| DNA molde | 362 ng/ µL | 90 ng | 362 ng/ µL | 90 ng |
| DNA polimerasa | 5 U/ µL | 0.05 U/ µL | 2 U/ µL | 0.02 U/ µL |
| DMSO | 100% | 10% | 100% | 10% |
| H ₂ O | --- | c.b.p. 10 µL | --- | c.b.p. 10 µL |

*Las secuencias son alguna de las parejas de iniciadores de la tabla 4.

Tabla 6. Condiciones de reacción de PCR para la *Taq* DNA polymerase.

| Reacción | Temperatura °C | Tiempo | Ciclos |
|---------------------------|----------------|--------|--------|
| Desnaturalización inicial | 95 | 2 min | 1 |
| Desnaturalización | 95 | 30 s | 40 |
| Alineamiento | 53-59 | 30 s | |
| Extensión | 72 | 4 min | |
| Extensión final | 72 | 7 min | 1 |

Tabla 7. Condiciones de reacción para la PCR con la Phusion High-Fidelity DNA Polymerase.

| Reacción | Temperatura °C | Tiempo | Ciclos |
|---------------------------|----------------|---------|--------|
| Desnaturalización inicial | 98 | 2 min | 1 |
| Desnaturalización | 98 | 1 min | 30 |
| Alineamiento | 60-65 | 30 s | |
| Extensión | 72 | 2.4 min | |
| Extensión final | 72 | 7 min | 1 |

3. En caso de que los resultados obtenidos no sean los esperados, se variará la temperatura de alineamiento y posteriormente la cantidad de cloruro de magnesio, hasta obtener un amplificado único.
4. Purificar y ligar de los amplicones obtenidos para cada par de iniciadores en el vector de clonación.
5. Separar mediante electroforesis los amplicones obtenidos por PCR en un gel de agarosa al 1.5% utilizando 60 V durante 1 h (condiciones controladas por una fuente de poder, BIO-RAD™). Teñir el gel con SYBR Safe y visualizar los resultados en un transiluminador de luz UV (Invitrogen).

6.10 Purificación de los amplicones obtenidos por PCR

Para obtener los amplicones puros y poder llevar a cabo las reacciones de ligación al vector de clonación, estos se purifican por columnas utilizando el kit GeneJET™ Gel Extraction.

1. Después de visualizar los amplicones en el transiluminador, cortar las bandas deseadas con una navaja nueva en cada ocasión y colocarlos en un microtubo estéril de 1.5 mL.
2. Utilizar el kit GeneJET Gel Extraction (Thermoscientific™) para purificar los amplicones. Los pasos seguidos en el procedimiento se realizan de acuerdo a las instrucciones del fabricante.
3. Una vez purificados, realizar la cuantificación de los productos de PCR (espectrofotómetro BioDrop a una longitud de onda de 260 nm) y observar la relación 260/280, la cual debe ser mayor a 1.8.

6.11 Reacción de adenilación al amplicón del GAP 4-2 amplificado por Phusion High-Fidelity DNA Polymerase

Debido a que la enzima Phusion High-Fidelity DNA Polymerase genera amplicones con extremos romos, se lleva a cabo una reacción de poliadenilación del amplicón para así poderlo clonar en el vector pGEM®-T Easy.

1. Una vez purificado y cuantificado el amplicón del GAP 4-2, el cual fue producido por una reacción de PCR con la enzima Phusion DNA pol; se toman 600 ng del amplicón y se adiciona Buffer Taq 1X, 2.5 mM MgCl₂, 0.2 mM dATP y 1 U de Taq polimerasa, esta última fue la utilizada para agregar las adeninas.
2. La reacción se lleva a cabo de acuerdo al procedimiento indicado por el fabricante de la Phusion High-Fidelity DNA Polymerase (Thermoscientific™) en el apartado de recomendaciones para clonación, utilizando la Taq DNA Polimerasa.
3. Al haber completado la reacción, se purifica el producto de la reacción utilizando el kit GeneJET Gel Extraction (Thermoscientific™), de acuerdo a las especificaciones del fabricante.

6.12 Ligación de los productos de PCR en el vector pGEM®-T Easy y transformación de células electrocompetentes

Los siguientes pasos consisten en clonar los amplicones obtenidos en el vector pGEM®-T Easy y transformar *E. coli* DH5α para la transformación por electroporación, después de esto las células son cultivadas para seleccionarlas transformantes por

complementación alfa. De las colonias positivas, se purifica el plásmido mediante el kit GeneJET Plasmid Miniprep y posteriormente, se verifica que contenga el inserto deseado.

1. Calcular la relación de vector: inserto para la ligación la cual fue de 1:3 molar; para 50 ng de plásmido.
2. Agregar a la reacción las cantidades de reactivo descritas en la Tabla 8.

Tabla 8. Condiciones de reacción para la ligación de los amplicones al vector pGEM®-T Easy.

| Reactivo | [Stock] | [rxn] |
|--------------------------|----------|--------------|
| 2X Rapid Ligation Buffer | 2X | 1X |
| Vector | 50 ng/μL | 50 ng |
| Producto de PCR | --- | * |
| T4 DNA Ligasa | 3 U/μL | 3 U |
| H ₂ O | --- | c.b.p. 10 μL |

*La cantidad de producto de PCR depende de la relación molar vector: inserto 1:3.

3. Colocar en la incubadora (Thermoscientific™) durante 1 h a 22 °C las reacciones de ligación cuyo tamaño de inserto va de 0.3-1.4 Kb; mientras que para el inserto de 5.7 Kb, la incubación se lleva a cabo durante 16 h a 4 °C.
4. Emplear la mitad del volumen de ligación (5 μL) para transformar 200 μL de células electrocompetentes de *E. coli* DH5α mediante electroporación (Apéndice 1).
5. Sembrar 100 y 150 μL de las células transformadas en cajas con medio LB con ampicilina (concentración final de 100 μg/mL), 40 μL de X-Gal al 2% y 4 μL de IPTG al 20%, para la selección de las clonas de interés. Las cajas se colocan en una incubadora (THELCO™) durante 14 h, a 37 °C.
6. Seleccionar 3 colonias blancas de las cajas de transformantes y realizar la PCR de colonia para confirmar que contienen el inserto del tamaño de interés para cada caso.
7. Inocular cada colonia confirmada por PCR en 5 mL de medio LB con ampicilina (100 μg/mL). Posteriormente, colocar en una incubadora a 37 °C y 200 rpm, por 14 h.

8. Extraer el plásmido utilizando el kit GeneJET Plasmid Miniprep siguiendo las instrucciones del fabricante. Posteriormente, cuantificar el plásmido (espectrofotómetro BioDrop) a una longitud de onda de 260 nm y observar la relación 260/280, la cual debe ser mayor a 1.8.
9. Realizar la PCR utilizando como molde el plásmido extraído (20 ng), siguiendo las condiciones de PCR anteriormente descritas. Correr una alícuota de 3 µL del plásmido y 4 µL de cada PCR realizada, en un gel de agarosa al 1% teñido con bromuro de etidio y posteriormente visualizar su presencia en el CHEMIDOC (BIO-RAD), para verificar la presencia y calidad del plásmido extraído, además de verificar de nuevo la presencia del inserto deseado.
10. Para el plásmido que contiene el inserto de 5.7 Kb, se realiza una digestión enzimática con la enzima *EcoRI*, en la Tabla 9 se muestran las cantidades de reactivo utilizadas.

Tabla 9. Condiciones de reacción de la digestión enzimática con *EcoRI* del plásmido.

| Reactivo | [Stock] | [rxn] |
|------------------|---------|--------------|
| Buffer Tango | 10X | 2X |
| Plásmido | --- | 1 µg |
| <i>EcoRI</i> | 5 U/ µL | 0.2 U |
| H ₂ O | --- | c.b.p. 20 µL |

11. Colocar las reacciones en una incubadora (THELCO) durante 4 h a 37 °C.
12. Correr en un gel de agarosa al 1% teñido con bromuro de etidio la reacción de digestión enzimática y visualizar en el CHEMIDOC para verificar la presencia de inserto.

6.13 Secuenciación y análisis de secuencias amplificadas del genoma de *Alicyclophilus* sp. BQ1

Se comprueba que las secuencias obtenidas sean confiables de acuerdo a los electroferogramas, para esto se espera que los picos formados por las señales de cada nucleótido no se encimen y la señal sea única. Posteriormente, mediante un

alineamiento de nucleótidos dentro de cada contig del E15 se ubican las secuencias amplificadas por PCR.

1. Mandar a secuenciar los plásmidos a la Unidad de Síntesis y Secuenciación de ADN del Instituto de Biotecnología (IBT), UNAM, Cuernavaca, Morelos. Utilizar los iniciadores m13/pUC -40 Forward y m13/pUC Reverse de la lista de iniciadores que proporciona el IBT para la secuenciación.
2. Verificar la calidad de las secuencias observando los electroferogramas, utilizando el programa SnapGene. Localizar la zona comprendida entre los iniciadores usados para cada reacción de PCR.

Analizar la ubicación de las secuencias de *Alicyclophilus* sp. BQ1 dentro del E15 mediante alineamientos de nucleótidos, utilizando el programa Vector NTI.

7. RESULTADOS Y DISCUSIÓN

7.1 Selección del ensamble del genoma de *Alicyclophilus* sp. BQ1

La cepa BQ1, debido a su actividad enzimática de tipo esterasa capaz de degradar el PU podría tener diversas aplicaciones biotecnológicas (Oceguera *et al.*, 2007). Es pues imperativo contar con un ensamble del genoma que sea de fácil acceso, en bases de datos disponibles como el NCBI. Como se explicó antes (sección 3.4), en nuestro grupo de trabajo se cuenta con dos ensambles *de novo* del genoma de *Alicyclophilus* sp. BQ1, por lo que, es necesario hacer una depuración para tener un solo bloque de datos sobre el cual planear los posteriores experimentos.

En este trabajo, se implementó una nueva aproximación para el análisis del ensamble del genoma que será reportado. Utilizando ambos ensambles, se realizaron alineamientos de nucleótidos utilizando el programa mummerplot, con el E13 como referencia (Figura 5). En este alineamiento se aprecian regiones en las cuales las cadenas de ambos ensambles alinean en la misma dirección (sentido 5'-3'; líneas verticales y círculos rojos), así como regiones en las cuales alinean pero en sentido contrario (líneas verticales y círculos azules). Esto indica que a pesar de que ambos ensambles poseen ciertas regiones idénticas en el mismo sentido, como en los contigs 9, 3, 8 y 1; no son exactamente iguales, por lo cual es necesario contar con otro ensamble de referencia para comparar.

Con apoyo del Dr. Miguel Ángel Cevallos (Centro de Ciencias Genómicas, UNAM), se realizó un tercer ensamble *de novo* del genoma utilizando la información de todas las secuenciaciones realizadas a *Alicyclophilus* sp. BQ1 y el ensamblador CLC, dicho ensamble consta de 98 contigs (E98). Se empleó la misma estrategia de alineamientos de secuencias con el programa mummerplot, utilizando el E98 como genoma de referencia comparándolo con E13 y E15 (Figuras 6 y 7). La comparación E13 con E98 muestra que ambos ensambles son semejantes en algunas regiones ya que coinciden los alineamientos de nucleótidos, apreciándose una línea roja que cubre casi por completo el E13; sin embargo, presenta algunos alineamientos en antisentido y no hay continuidad en los alineamientos. A diferencia de este, en el alineamiento con E15 y

E98 (Figura 7), se observan una mayor cantidad de alineamientos en el mismo sentido, los cuales son menos fragmentados en comparación con el anterior, además de tener una menor cantidad de alineamientos en sentido contrario (líneas azules).

Por lo tanto, debido a que la similitud entre el E15 con E98 es mayor que la que existe entre E13 con E98 se decidió elegir al E15 como el ensamble con el que se trabajará durante este proyecto y que, posteriormente, será reportado en la base de datos del GenBank (NCBI).

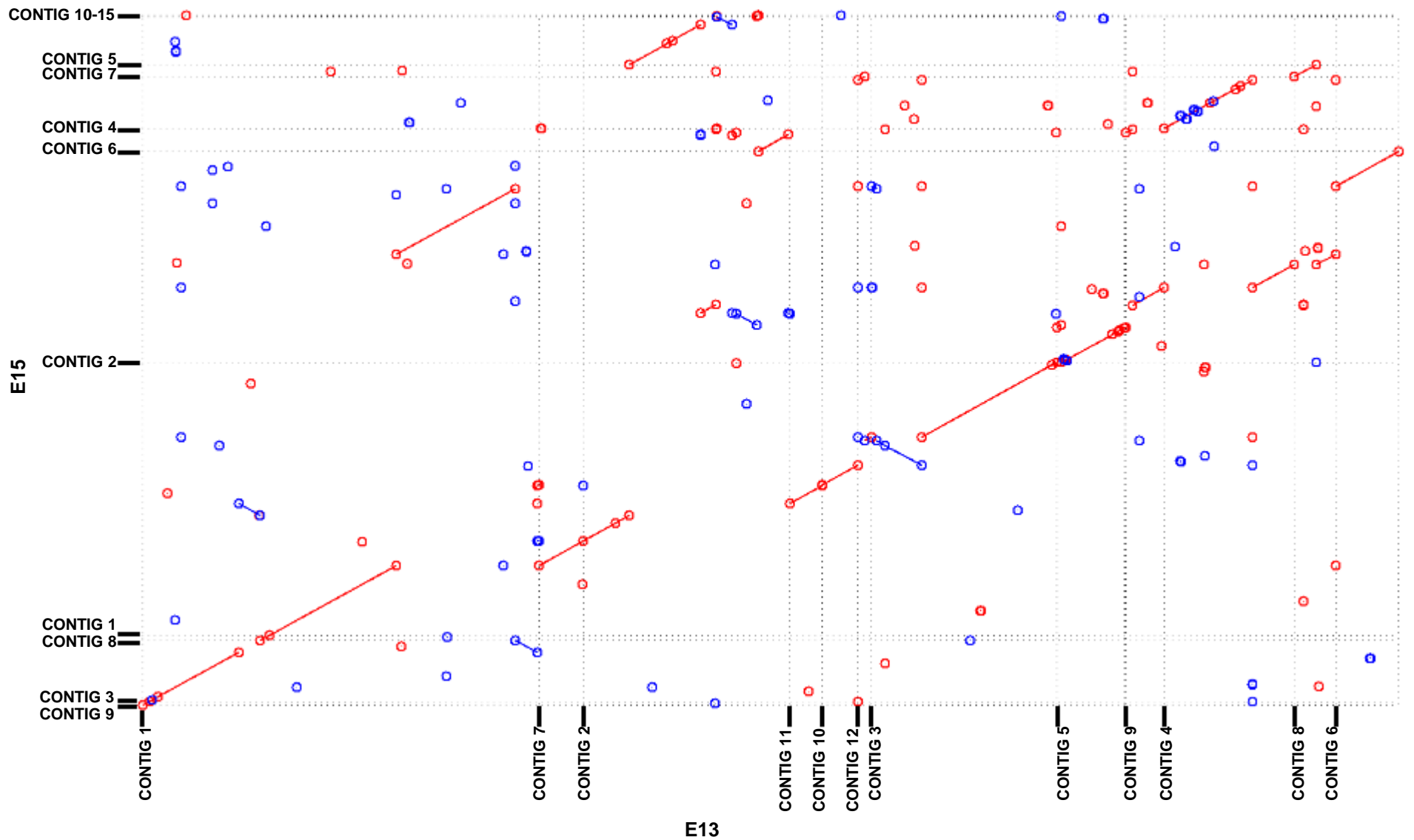


Figura 5. Alineamiento del E15 y el E13 usando mummerplot. En la lateral izquierda se encuentran los contigs del E15, mientras que en la parte inferior se muestran los contigs del E13. Las líneas y círculos rojos muestran alineamiento de nucleótidos en el mismo sentido, mientras que los azules representan alineamiento en antisentido.

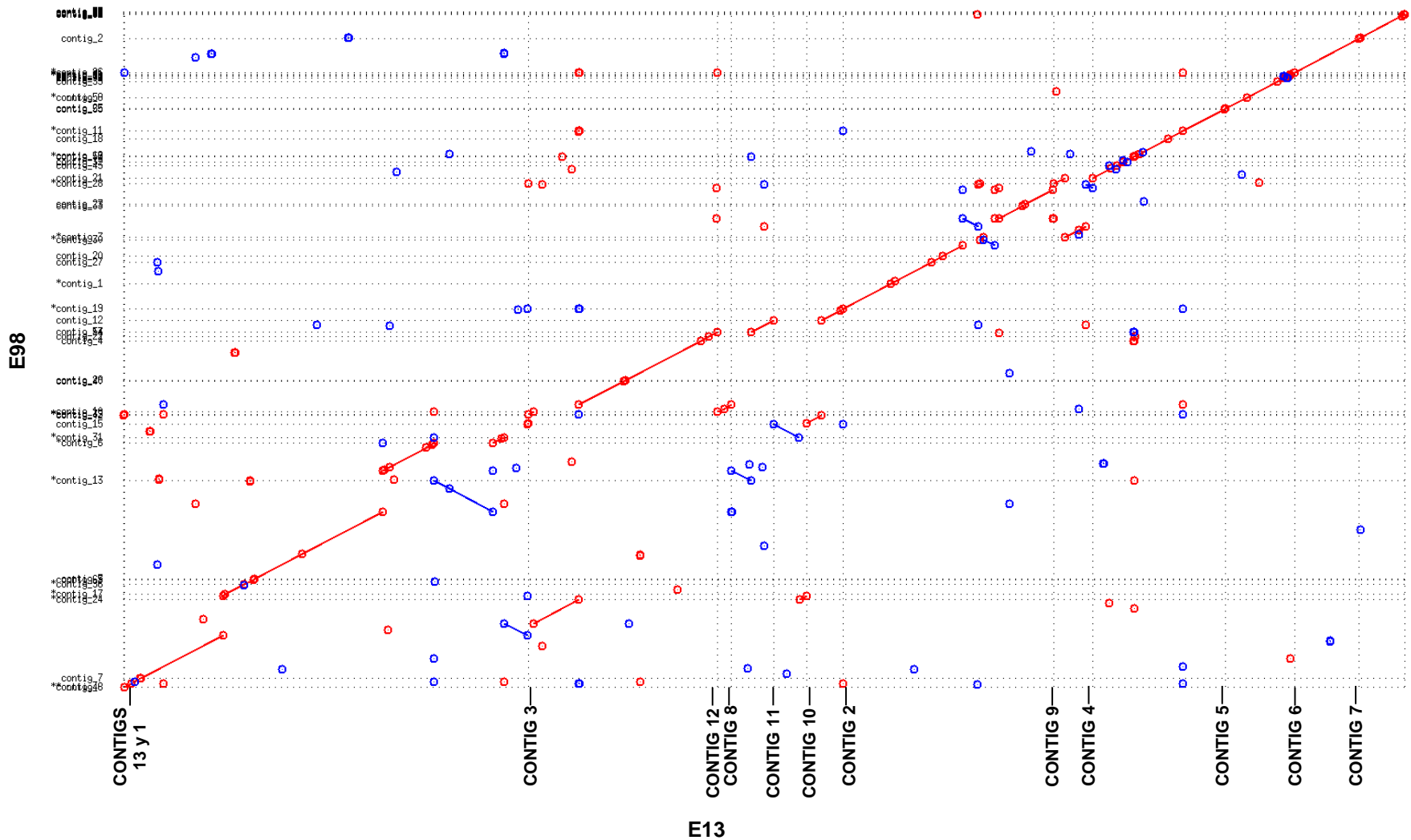


Figura 6. Alineamiento del E13 y el E98 usando mummerplot. En la lateral izquierda se encuentran los contigs del E98, mientras que en la parte inferior se muestran los contigs del E13. Las líneas y círculos rojos muestran alineamiento de nucleótidos en el mismo sentido, mientras que los azules representan alineamiento en antisentido.

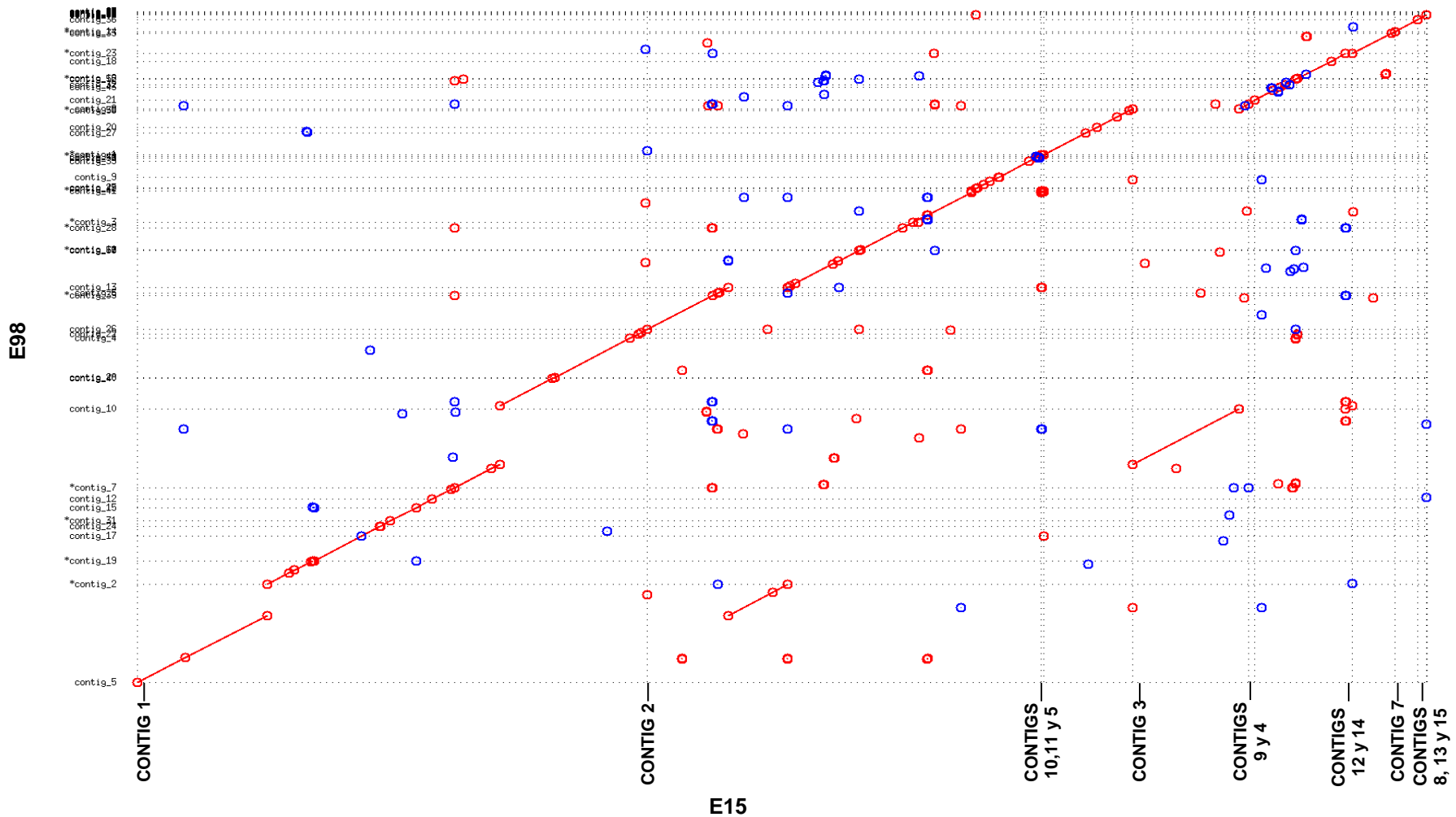


Figura 7. Alineamiento del E15 y el E98 usando mummerplot. En la lateral izquierda se encuentran los contigs del E98, mientras que en la parte inferior se muestran los contigs del E15. Las líneas y círculos rojos muestran alineamiento de nucleótidos en el mismo sentido, mientras que los azules representan alineamiento en antisentido.

7.2 Publicación del E15 de *Alicyclophilus* sp. BQ1 en el NCBI

Tras determinar que el E15 es el ensamblaje más adecuado para el genoma de *Alicyclophilus* sp. BQ1, se realizaron los procedimientos para poder subir al NCBI los datos de su secuencia (sección 6.5). El número de acceso asignado fue NZ_NKDB000000000.1. El Ensamblaje consta de 15 contigs, 69% de GC y 4'676,255 pb totales. La anotación del genoma estuvo a cargo del NCBI mediante NCBI *Prokaryotic Genome Annotation Pipeline* (programa anotador). Se encontraron 4,372 genes totales, de los cuales 4,251 codifican proteínas, 55 codifican RNAs y se identificó solo un operón para rRNA. Es importante señalar que, antes de que se reportara el ensamblaje de BQ1, en la base de datos del NCBI solo se contaba con cuatro genomas reportados (sección 3.3), viendo un gran aumento en la cantidad de genomas reportados de este género después de BQ1, apareciendo ocho nuevas cepas (Tabla 10).

La anotación del genoma de BQ1 así como su reporte en el GenBank es un gran avance en el campo de estudio de esta bacteria. A partir de estos datos se podrán realizar mayores investigaciones tanto bioinformáticas como bioquímicas acerca de las actividades enzimáticas de esta cepa capaz de actuar sobre el PU. Asimismo, se podrán realizar estudios de genómica comparativa entre las diferentes especies del género *Alicyclophilus* y otros géneros bacterianos implicados en la degradación de xenobióticos. La información que se obtenga de esos estudios puede ser trasladada al campo de la biotecnología para poder plantear alternativas al tratamiento de residuos, destacando en este caso el del PU.

Tabla 10. Cepas del género *Alicyclophilus* reportadas en el NCBI (Agosto, 2018).

| Cepas | Tamaño total del genoma (pb) | Nivel de ensamble | Año de reporte en el NCBI |
|--|-------------------------------------|----------------------------------|----------------------------------|
| * <i>Alicyclophilus denitrificans</i> K601 | 5'070,751 | Completo | 2011 |
| * <i>Alicyclophilus denitrificans</i> BC | 4'835,713 | Completo | 2011 |
| <i>Alicyclophilus</i> sp. B1 | 7'465,959 | Contigs (59) | 2015 |
| <i>Alicyclophilus</i> sp. 69-12 | 4'633,122 | Scaffolds (96) Contigs (397) | 2016 |
| <i>Alicyclophilus</i> sp. BQ1 | 4'676,255 | Contigs (15) | 2017 |
| <i>Alicyclophilus</i> sp. UBA880 | 3'034,705 | Scaffolds (141) Contigs (359) | 2017 |
| <i>Alicyclophilus</i> sp. UBA2076 | 2'271,502 | Scaffolds (267) Contigs (420) | 2017 |
| <i>Alicyclophilus</i> sp. UBA3448 | 2'383,256 | Scaffolds (108) Contigs (125) | 2017 |
| <i>Alicyclophilus</i> sp. UBA4115 | 2'909,939 | Scaffolds (265) Contigs (314) | 2017 |
| <i>Alicyclophilus</i> sp. UBA6560 | 3'475,643 | Scaffolds (13) Contigs (15) | 2017 |
| <i>Alicyclophilus</i> sp. UBA6761 | 2'179,381 | Scaffolds (244) Contigs (343) | 2017 |
| <i>Alicyclophilus</i> sp. UBA7619 | 4'738,868 | Scaffolds (20) Contigs (28) | 2017 |
| <i>Alicyclophilus denitrificans</i> CD02 | 4'671,828 | Contigs (52) | 2018 |

* Estas cepas cuentan con plásmidos; K601 tiene el plásmido pALIDE201 (75488 pb); BC cuenta con dos plásmidos, pALIDE01 (119718 pb) y pALIDE02 (78982 pb).

7.3 Posible orden de los contigs del E15 tomando como referencias a *Alicyclophilus denitrificans* BC y K601

Al contar con un ensamble a nivel de contigs, no se sabe el orden que estos presentan en el genoma, por lo cual es importante emplear algún método para encontrar el orden mas probable y así poder realizar las PCR que permitan amplificar regiones entre ellos para tratar de ensamblarlos en un solo cromosoma. Para ello, utilizando como

referencia los genomas completos de dos cepas de la especie *A. denitrificans*, cepas BC y K601 reportadas en la base de datos del NCBI, se llevaron a cabo alineamientos de nucleótidos para observar las similitudes en el orden de dichas secuencias de nucleótidos con respecto a *Alicyclophilus* sp. BQ1, y poder darle un orden a los 15 contigs. Al realizar alineamientos de regiones de las secuencias de cada contig de BQ1 utilizando el programa BLAST, con los genomas de las cepas de referencia BC y K601 de *A. denitrificans* (sección 6.6), se obtuvieron dos diferentes arreglos de los contigs de E15 (Figuras 8 y 9). Los resultados de los alineamientos realizados por BLAST en los que se muestra la posición de cada secuencia dentro de los genomas de ambas cepas de *A. denitrificans* se encuentran en el Apéndice 2, Tablas A1 y A2.

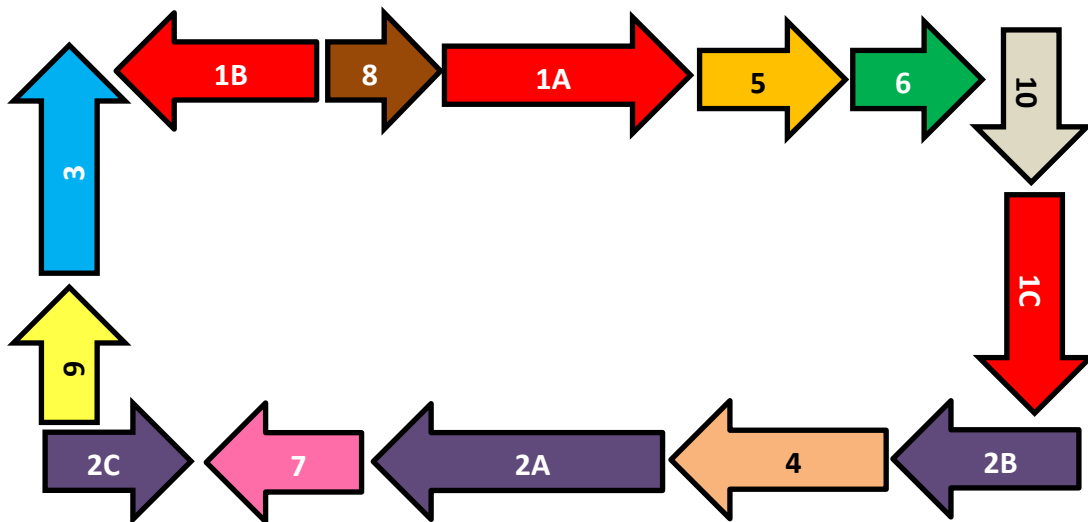


Figura 8. Arreglo genómico de los contigs del E15 de *Alicyclophilus* sp. BQ1 basado en el genoma de *A. denitrificans* BC. Cada flecha representa un contig, excepto los contigs 1 y 2 que se encuentran divididos en tres partes (A, B y C) de acuerdo a que los genes contenidos en ellos presentan otro orden en el genoma de *A. denitrificans* BC.

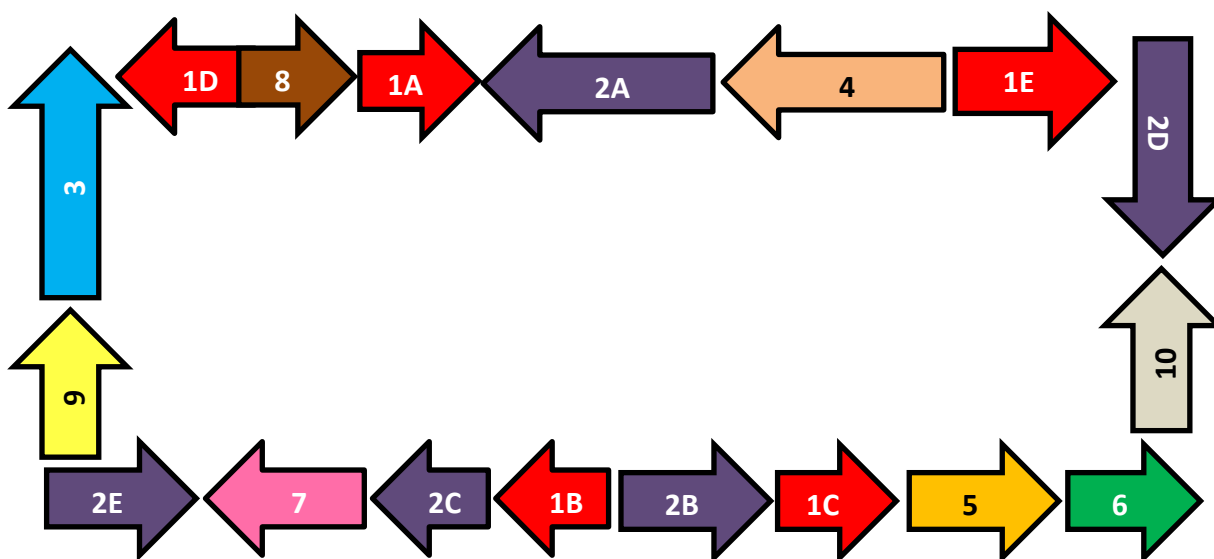


Figura 9. Arreglo genómico de los contigs del E15 de *Alicyclophilus* sp. BQ1 basado en el genoma de *A. denitrificans* K601. Cada flecha representa un contig, excepto los contigs 1 y 2 que se encuentran divididos en cinco partes (A, B, C, D y E) de acuerdo a que los genes contenidos en ellos presentan otro orden en el genoma de *A. denitrificans* K601.

Los contigs 11 al 15 no fueron localizados mediante el BLAST, debido quizá a que al tener tamaños de 2,100-511 pb no pudieron ser alineados con los genomas de referencia. En esta comparación también se observó que los contigs 1 y 2 están fragmentados, lo que sugiere un reordenamiento de los genes en BQ1 con respecto a ambas cepas de *A. denitrificans*. Las secuencias que no pudieron ser ubicadas en BC y K601 quizá codifican genes únicos de BQ1.

Cuando se realizó la comparación del orden que presentaron los contigs de E15 en los genomas de referencia (Figuras 8 y 9), con el orden encontrado utilizando Mauve, se observó una coincidencia en la continuidad de los contigs 9-3, 8-1, 5-6-10 y 4-2; por lo que, se esperaba que estas regiones estén conservadas y presenten el mismo orden en *Alicyclophilus* sp. BQ1 (Figura 10). En el caso de la continuidad entre el contig 3 y 8 se observó una región del contig 1 entre ellos, lo cual podría deberse a un rearrreglo de las secuencias de nucleótidos entre las cepas comparadas. Tampoco fue posible sustentar la continuidad de los contigs 1-5, 10-4, 2-7 y 7-9 detectada con el programa

Mauve al hacer las comparaciones contra los genomas de K601 y BC. Sin embargo, en este trabajo se tratará de probar si el orden propuesto por Mauve está presente en el genoma de BQ1.

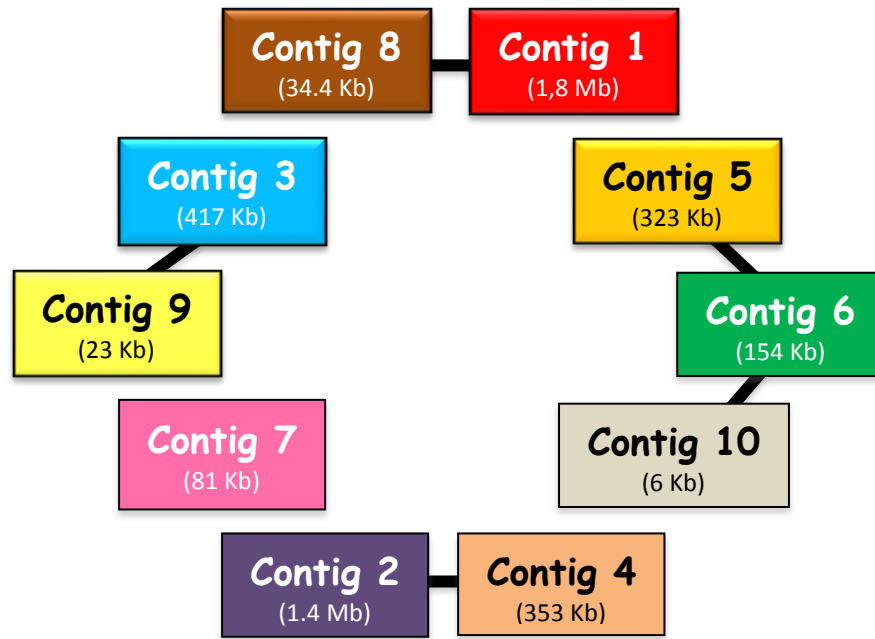


Figura 10. Posible orden de los contigs predicho por Mauve. Los contigs unidos con líneas negras indican que las evidencias obtenidas por alineamientos realizados por BLAST con los genomas de *A. denitrificans* BC y K601 apoyan esta organización. Entre paréntesis se colocaron los tamaños de cada uno.

Con base en la estrategia propuesta, se realizarán las PCR para amplificar las regiones intercontigs, secuenciar los amplicones resultantes, confirmar el orden de estos y agregar las secuencias faltantes al ensamble del genoma, verificando los resultados bioinformáticos. Para ello era necesario tener una aproximación de la longitud de las regiones a amplificar, a partir de la posición de los extremos de cada contig en los genomas conocidos, para lo cual se tomaron los 1,000 nt de principio y fin de cada contig que se propone está contiguo y se compararon con los genomas de BC y K601. Solo fue posible encontrar alineamientos entre los contigs 4 y 2. En todos los demás casos esta longitud no se pudo estimar. Los últimos 1,000 nt de la secuencia final del contig 4 y los 1,000 nt de la inicial del contig 2 se encontraron en los genomas de *A.*

denitrificans K601 y BC con porcentajes de similitud de 96.3% para la secuencia del contig 4 y 95% en la del contig 2 para ambos genomas (Figura 11).

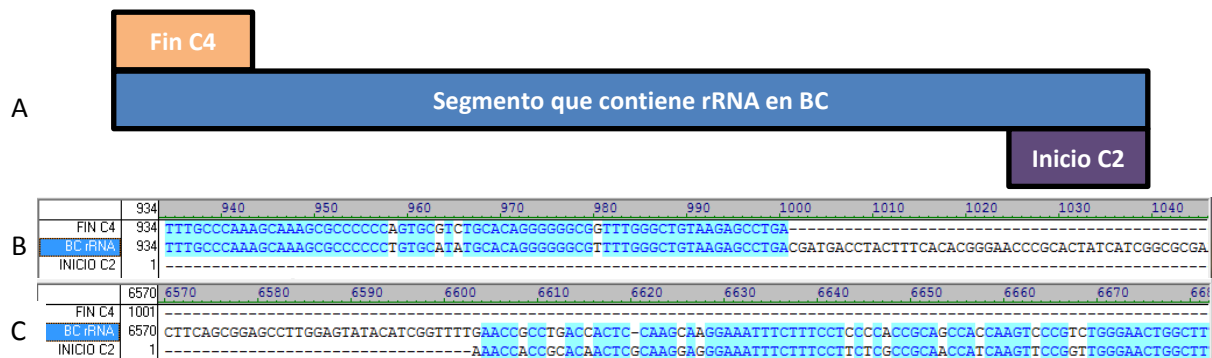


Figura 11. Alineamientos de los extremos del contig 4 y 2 del E15 con el operón rRNA de *A. denitrificans* BC. A) representación gráfica de las regiones en las cuales se llevó a cabo el alineamiento de nucleótidos. B) Alineamiento entre el extremo final del contig 4 y un extremo flanqueante del operón de rRNA de *A. denitrificans* BC. C) Alineamiento entre el extremo inicial del contig 2 y un extremo flanqueante del operón de rRNA de *A. denitrificans* BC. Las zonas donde no se presentan alineamientos corresponden a la secuencia del operón de rRNA. Se obtuvieron resultados similares con *A. denitrificans* K601.

Entre estas dos secuencias encontradas en ambas cepas de *A. denitrificans* BC y K601 se localizan 5,698 pb que codifican el operón de rRNA que contiene los genes: 5S rRNA, 23S rRNA, t-RNA-ala, t-RNA-ile y 16S RNA. Por lo que se supondría que en la región intercontigs 4-2 se encuentre este operón. Se ha reportado que las cepas BC y K601 cuentan con tres copias de este operón. Sin embargo, cada operón está flanqueado por distintos genes. El operón de rRNA con los genes flanqueantes detectados en el contig 4 y el contig 2 de BQ1 fueron identificados entre los nucleótidos 3'138,630 y 3'144,328 de *A. denitrificans* BC, mientras que para K601 la secuencia va del 1'500,125 hasta el 1'505,823. En el E15 de BQ1, una copia del operón de rRNA se localiza en el contig 2, a mas de 1.1 Mb de los 1,000 nt finales que se utilizaron para este análisis. Esta copia del operón rRNA está flanqueado por genes diferentes que los que tiene la copia que se predice se encuentra entre los contigs 4 y 2. Esta copia está localizada entre los nucleótidos 4'277,568 y 4'283,266 en la cepa BC, mientras que en K601 se ubica del nucleótido 4'657,978 al 4'663,676. Esto sugiere que, si efectivamente

la región intercontigs 4-2 tuviera estos genes, serían una copia de los ya identificados en el contig 2, lo cual sería algo semejante de lo observado en *A. denitrificans* BC. Por otro lado, también se buscó la secuencia del operón de rRNA en los contigs 11-15 y no se encontró.

7.4 Curación del E15 de *Alicyclophilus* sp. BQ1

Los resultados arrojados al realizar el alineamiento con mummerplot utilizando el E15 y el E13 (Figura 5), mostraron ciertas regiones de similitud entre ambos ensamblados. Esta situación permitió identificar algunas regiones no existentes en E15 pero presentes en E13. Utilizando el programa Vector NTI se realizaron alineamientos que ayudaron a añadir en algunos casos nuevas secuencias de nucleótidos a algunos contigs y en otros se logró ensamblar dos contigs del E15 junto con nucleótidos del E13. Una de las regiones identificadas se localiza justo antes del contig 3, ya que parte de sus secuencias se encuentran ensambladas en el contig 1 del E13. La secuencia consiste en 128 nt localizados antes del inicio del contig 3 (Figura 12).

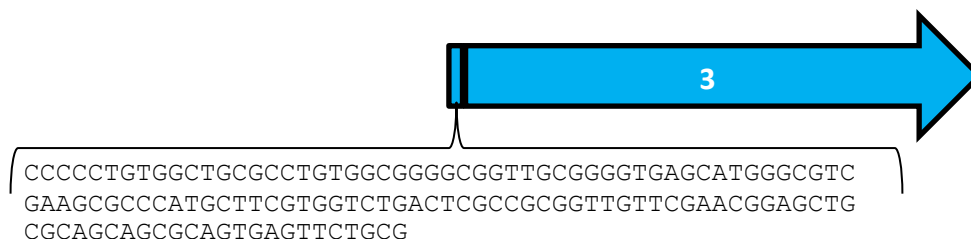


Figura 12. Secuencia ensamblada en el contig 3. Secuencia de 128 nt perteneciente al contig 1 del E13 ensamblada en el inicio del contig 3 del E15.

También, se encontraron secuencias del contig 9 del E15 dentro del contig 1 del E13, de la misma forma, se pudieron agregar secuencias a los contigs 6 y 11 del E15. Las secuencias agregadas fueron encontradas en el contig 2 del E13. En el caso del contig 6 se agregaron 392 nt inmediatamente después del final de este (Figura 13). Mientras, para el contig 11 se agregaron 516 pb antes del inicio de este (Figura 14). Aunque estos análisis permitieron añadir algunas secuencias a ciertos contigs, no se ha conseguido eliminar regiones intercontigs dentro del E15. Sin embargo, agregar estas

secuencias servirá para identificarlas en los amplicones que se logren obtener y secuenciar en este trabajo, ayudando a la identificación de secuencias.

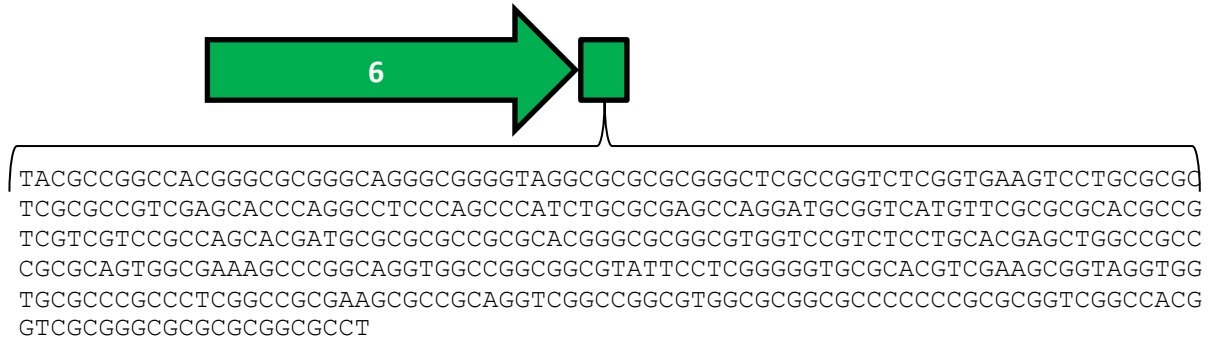


Figura 13. Secuencia ensamblada en el contig 6. Secuencia de 392 nt perteneciente al contig 2 del E13 ensamblada en al final del contig 6 del E15.

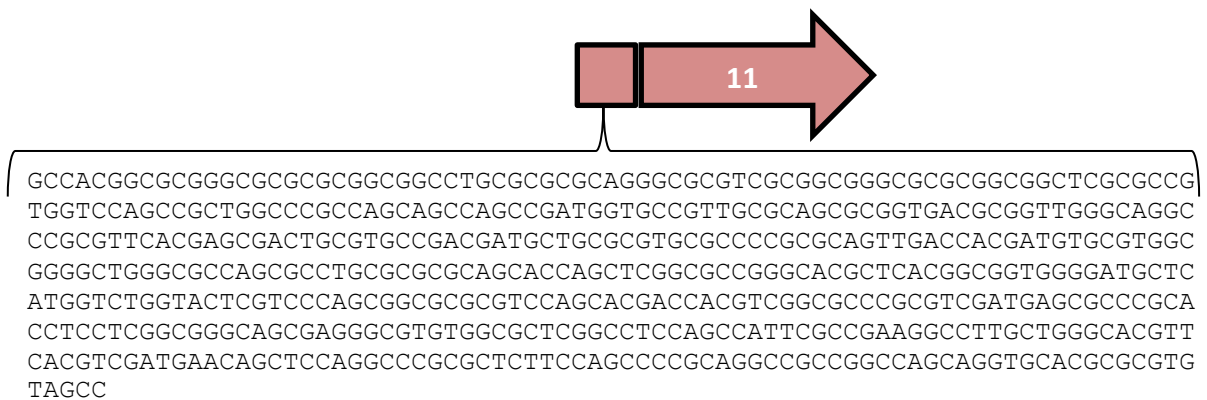


Figura 14. Secuencia ensamblada en el contig 11. Secuencia de 516 nt perteneciente al contig 2 del E13 ensamblada al inicio del contig 11 del E15.

Un caso especial fue la unión del contig 5 con el 12 del E15. Los alineamientos de secuencias entre ensambles, usando el contig 2 del E13 y los contigs 5 y 12 del E15, demostraron que realmente no faltaba ninguna secuencia entre ambos contigs, sino que había una diferencia de 7 nt intercalados entre los primeros 41 nt del inicio del contig 12 con respecto a las secuencias correspondientes del contig 5 del E15 y la secuencia del contig 2 del E13. La región que abarcaban estos nucleótidos se encontraba distribuida en 29 nt consecutivos, logrando ensamblar una sola secuencia entre ambos contigs (Figura 15).



Figura 15. Secuencia traslapada entre el contig 5 y 12. Ensamble de los contigs 5 y 12 del E15 con base en un traslape de secuencias entre sus extremos.

Adicionalmente, se logró la unión de dos contigs del E15, el contig 11 con el 10. Esto fue posible al identificar una secuencia de 191 nt en el contig 2 del E13, que correspondía al final del contig 11 y al inicio del contig 10 (Figura 16). Un logro adicional al disminuir el número de contigs del ensamblaje es que se pudo asignar ubicación al contig 11, ya que este no pudo ser colocado por Mauve en orden dentro del ensamblaje.

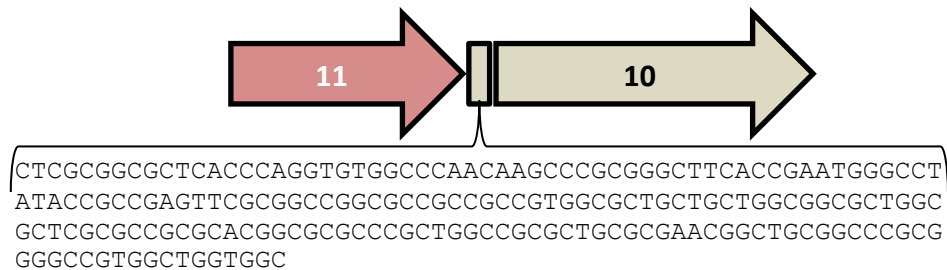


Figura 16. Secuencia ensamblada intercontigs 11-10. Secuencia de 191 nt perteneciente al contig 2 del E13 ensamblada en el final del contig 11 y al inicio del contig 3 del E15, formando un nuevo contig.

Para finalizar, se lograron determinar las secuencias faltantes entre los contigs 1 y 8. Previamente, en nuestro laboratorio se había logrado amplificar un fragmento de 1,800 pb empleando primers diseñados para el final contig 8 y el inicio del contig 1. Para secuenciar este fragmento se clonó en pBS en el sitio *Pst*I para lo cual se digirió este fragmento con esa enzima de restricción. Desafortunadamente, el fragmento fue digerido en una zona intermedia generando dos fragmentos los cuales fueron denominados fragmento A, de 849 pb y fragmento B de 655 pb. Esos fragmentos fueron secuenciados y comparados con los genomas de *A. denitrificans* BC y K601, encontrando que para ambas comparaciones, faltaban dos pequeñas regiones de 121 y 276 pb entre el contig 8 y el fragmento B y entre el fragmento B y el fragmento A,

respectivamente (Bautista Alcántara, 2015). Las secuencias de los fragmentos A y B fueron localizadas en el contig 1 del E13, lo que permitió identificar las secuencias faltantes con tamaños de 120 y 96 pb las cuales se habían predicho ser de 121 y 276 pb, respectivamente (Figura 17). La falta de coincidencia entre el tamaño encontrado del segundo fragmento (96 pb), y el tamaño predicho (276 pb) puede ser debida a que los 180 pb faltantes no se encuentran en BQ1, pero si en BC y K601, o bien a que esa región es un faltante en el E13.

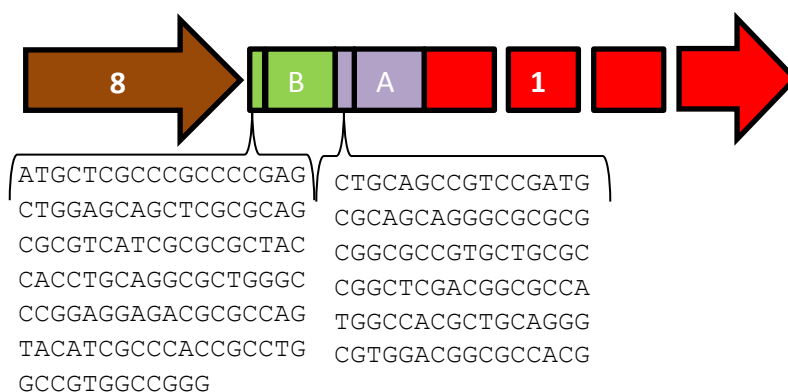


Figura 17. Secuencia ensamblada intercontigs 1-8. En el contig 1 del E13 se logró identificar la región faltante entre el final del contig 8 del E15 y el inicio del fragmento B. Entre el fragmento B y el fragmento A se identificó una secuencia de 96 nt localizada también en el contig 1 del E13. En este contig de E13 también fue posible identificar final del fragmento A y el inicio del contig 1 del E15 contiguos. Los contig 8 y 1 se ensamblaron en uno solo.

Para comprobar si estas secuencias se encuentran también en los genomas de *A. denitrificans* BC y K601 se llevó a cabo un BLAST, ya que en trabajos anteriores se comprobó que la región intercontigs entre 8 y 1 se conserva en estas (Bautista Alcántara, 2015). La secuencia de 120 nt comprendida entre el contig 8 y el fragmento B fue ubicada entre los nucleótidos 528,759 y 528,640 en K601, la cual, en BC, comprende al segmento del nucleótido 579,570 al 579,451. La secuencia de 96 nt entre el fragmento B y el fragmento A fue ubicada entre los nucleótidos 529,490 y 529,396 en K601, la cual, en BC, comprende al segmento del nucleótido 580,301 al 580,207.

Se logró respaldar con los datos del análisis bioinformático, la unión de los contigs 8 con el 1, el 5 con el 12 y el 11 con el 10. Además, se agregaron secuencias a los

contigs 3, 6 y 11, con lo que se logró reducir el número de contigs a 12. Solo los contigs 13, 14 y 15 siguen sin tener una ubicación asignada dentro del genoma de BQ1.

7.5 Análisis de los amplicones intercontigs

Debido a que empleando métodos de secuenciación masiva de diferentes tipos no fue posible ensamblar el genoma de BQ1 en un solo cromosoma, se decidió emplear la técnica de PCR para tratar de amplificar las regiones desconocidas entre los contigs para posteriormente clonaras y secuenciarlas por el método de Sanger. Para ello, se diseñaron iniciadores para cada una de las regiones intercontigs (sección 6.8) de acuerdo al orden identificado bioinformáticamente por Mauve (Figura 10), que sugirió el orden de contigs 5-6-10-4-2-7-9-3-8-1-5. Es importante recordar que al comparar por BLAST este orden propuesto contra los genomas de BC y K601 se observaron evidencias a favor de la continuidad de los contigs 9 con 3, 8 con 1, 5 con 6, 6 con 10 y 4 con 2.

En las reacciones de PCR se probaron distintas condiciones de amplificación considerando que debido a que el contenido de GC del genoma de *Alicyclophilus* sp. BQ1 es del 69%, era necesario agregar 10% de DMSO, con lo que se logra que la temperatura de fusión del complejo iniciador-molde disminuya significativamente. A las reacciones y amplicones que se obtengan se les nombrará de acuerdo a la región que se está amplificado, por ejemplo a la banda producida por iniciadores que unen el contig 9 con el 3 se le llamará GAP9-3 y así para cada región. Las secuencias y nombres de cada par de iniciadores se encuentran en la tabla 4, en la sección 6.8. Se probaron temperaturas de alineamiento de 53 a 59 °C en las reacciones con los pares de iniciadores J10Fout-J4Rout (GAP10-4), J1Fout-J5Rout (GAP1-5), J2Fout-J7Rout (GAP2-7) y no se obtuvieron bandas de amplificación (Figura 18-1, 17-2, 17-3). Para el par J7Fout-J9Rout (GAP7-9) se observaron bandas tenues a la temperatura de alineamiento mas baja, probablemente inespecíficas, pues desaparecieron al aumentar la temperatura de alineamiento (Figura 18-4). Por lo tanto, no se continuó trabajando con el análisis de estas cuatro regiones intercontigs. La ausencia de amplicones en los cuatro casos descritos puede deberse a que estos contigs no se encuentran contiguos

o a que la región entre los distintos contigs es muy grande y la PCR no puede amplificarla.

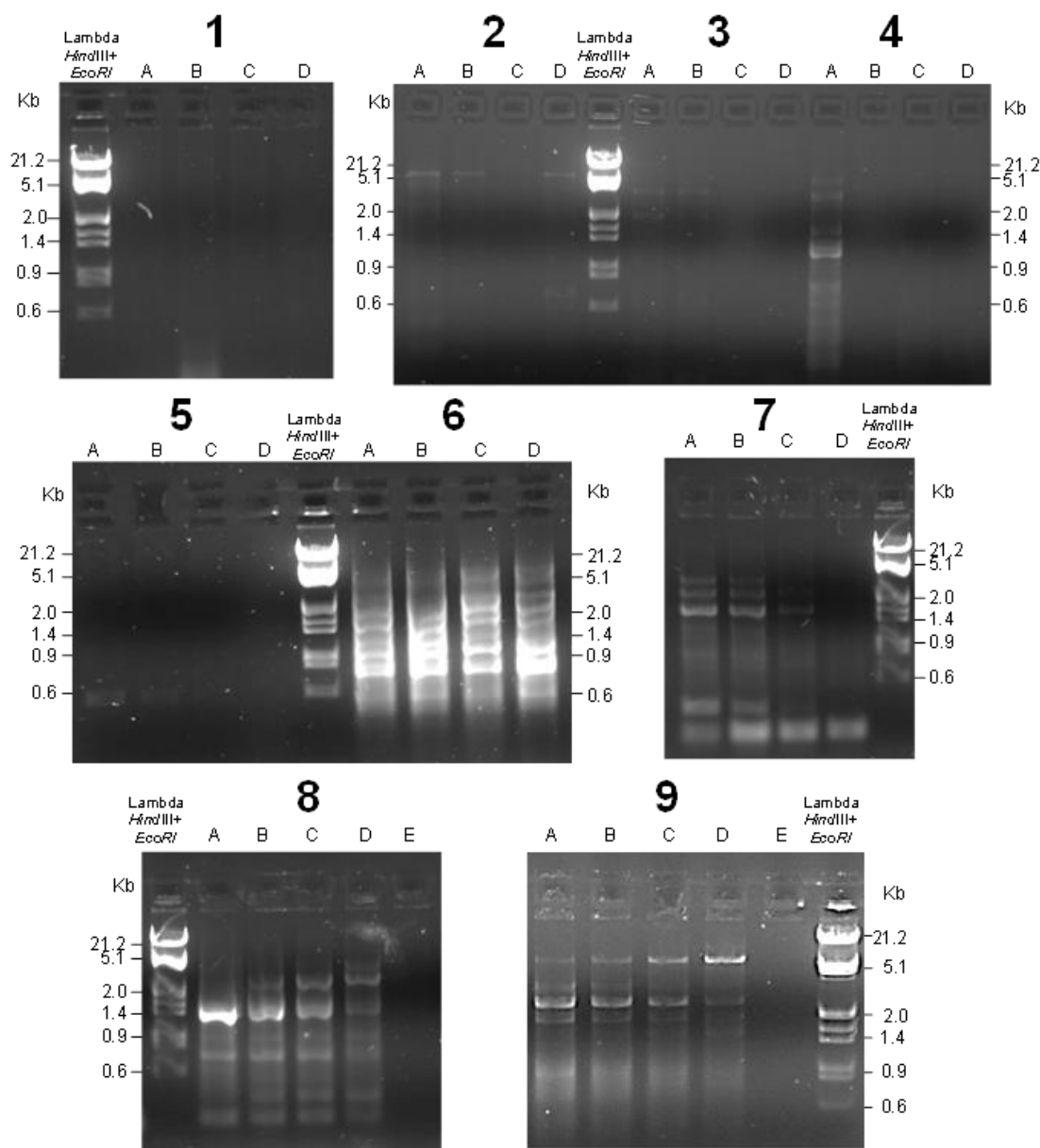


Figura 18. PCRs para amplificar las regiones intercontigs. 1) GAP10-4, 2) GAP1-5, 3) GAP2-7, 4) GAP7-9, 5) GAP9-3, 6) GAP3-8, 7) GAP5-6, 8) GAP6-10 y 9) GAP4-2. Se probaron diferentes temperaturas de alineamiento: A) 53 °C, B) 55 °C, C) 57 °C y D) 59 °C. E) En el control de PCR se sustituyó el DNA molde por agua. Gel de agarosa al 1.5%. En cada carril se cargaron 4 μ L de la reacción de PCR.

En el caso del GAP9-3 se obtuvo una banda muy tenue pero única (Figura 18-5), en los casos de GAP5-6 (Figura 18-7) y el GAP4-2 (Figura 18-9) se observó la desaparición de bandas inespecíficas y la persistencia de una. Para los casos de los GAP3-8 (Figura 18-6) y GAP6-10 (Figura 18-8) se observaron varias bandas inespecíficas. Para todos estos casos, se probaron diferentes condiciones para las reacciones de PCR, con el objetivo de mejorar su desempeño y lograr amplificar una sola banda.

Debido a que se espera un buen rendimiento de la reacción así como obtener bandas únicas al usar cada par de iniciadores, todas las reacciones de PCR en las cuales se obtuvieron amplificadas fueron ajustadas experimentalmente probando diferentes concentraciones de $MgCl_2$ en cada reacción. Esto se analizó debido a que el Mg^{2+} se une a dNTPs, iniciadores y DNA molde, disminuyendo su concentración lo cual puede afectar la polimerización ya que la DNA polimerasa requiere Mg^{2+} para su correcta actividad. Las cantidades seleccionadas fueron concentraciones de 1.5 a 4 mM, ya que son las óptimas para que trabaje la *Taq* DNA Polimerasa; resultando idóneas 1.5 mM para los GAPs 5-6 y 9-3; 2.5 mM para los GAPs 6-10 y 3-8; y de 4 mM para el GAP4-2.

Los resultados finales de estos ajustes fueron amplicones de diversos tamaños. Para el GAP5-6 de aproximadamente 300 pb (Figura 19, 2A); para el GAP9-3 de 500 pb aproximadamente (Figura 19, 4A), para el GAP3-8 de 800 pb (Figura 19, 5B) y para el GAP6-10 se observaron dos bandas cercanas a 1400 pb muy intensas que se analizaron por igual (amplicón GAP6-10S y amplicón GAP6-10I) (Figura 19, 1C). Para el GAP4-2, el tamaño fue mayor a 5.1 Kb (Figura 19, 3D), coincidiendo con el tamaño esperado bioinformáticamente (5,698 pb), para los genes ribosomales predichos en ese segmento de acuerdo a lo descrito previamente.

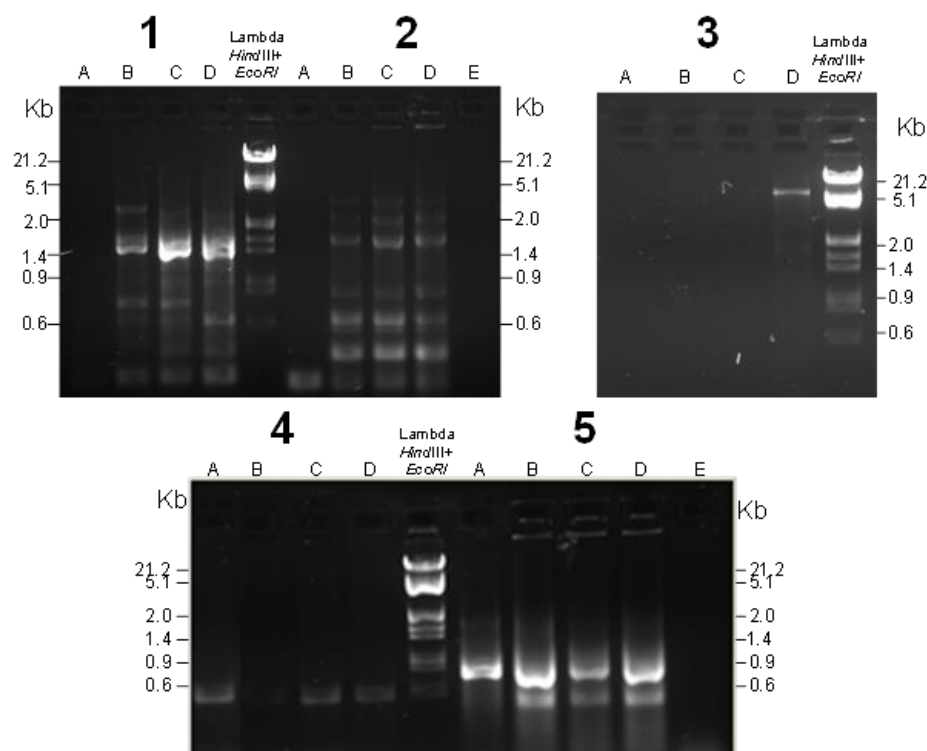


Figura 19. Ajuste experimental de las PCRs para amplificar las regiones intercontigs. 1) GAP6-10, 2) GAP5-6, 3) GAP4-2, 4) GAP9-3 y 5) GAP3-8. Se probaron diferentes concentraciones de $MgCl_2$: A) 1.5 mM, B) 2.5 mM, C) 3.5 mM y D) 4 mM. E) En el control de PCR se substituyó el DNA molde por agua. Gel de agarosa al 1.5%. En cada carril se cargaron 4 μ L de la reacción de PCR.

A pesar de obtener un solo amplificado con *Taq* DNA polimerasa, se optó por usar la enzima Phusion High-Fidelity DNA Polymerase, debido a que puede amplificar fácilmente fragmentos de DNA de hasta 7.5 Kb, además de tener una mayor fidelidad con una tasa de error 50 veces menor que la de la *Taq* DNA polimerasa (Phusion High-Fidelity DNA Polymerase, 2016).

En la Figura 20, se muestran los resultados de PCR con la enzima Phusion, al utilizar diferentes temperaturas de alineamiento. Al final no se realizaron mas cambios en las condiciones de PCR y se utilizó la temperatura de alineamiento de 65 °C, ya que se producía una mayor cantidad de amplicón y esa misma banda se generaba en reacciones con ambas polimerasas. Posteriormente, este amplicón se purificó y se llevó

a cabo una reacción de adición de adeninas sobresalientes (sección 6.9) para poder llevar a cabo la clonación en el vector pGEM®-T Easy como con los demás amplicones.

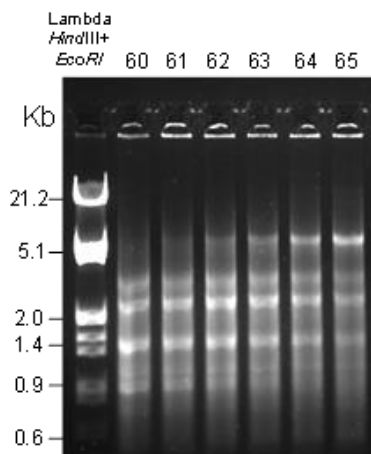


Figura 20. PCR del GAP4-2 utilizando la enzima Phusion High-Fidelity DNA Polymerase. Se probaron diferentes temperaturas de alineamiento, expresadas en °C, mostradas en la parte superior de cada carril. Gel de agarosa al 1%. En cada carril se cargaron 4 µL de la reacción de PCR.

Se lograron obtener bandas únicas para las regiones intercontigs 5-6, 4-2, 9-3 y 3-8. En el caso de la región entre los contigs 6-10 se obtuvieron dos bandas que no desaparecen al variar las condiciones de la PCR por lo cual ambas fueron analizadas. Todos los amplicones fueron clonados en el vector pGEM®-T Easy y secuenciados para determinar si corresponden a las regiones intercontigs, con lo que se lograría así disminuir el número de contigs y acercarnos un poco mas a cerrar el genoma de BQ1.

7.6 Análisis de los amplicones secuenciados de BQ1

Los amplicones intercontigs obtenidos por PCR, usando los diferentes pares de iniciadores, fueron purificados usando el kit GeneJET™ Gel Extraction para ligarlos al vector de clonación pGEM®-T Easy. Este vector fue seleccionado debido a que facilita la ligación de amplicones obtenidos al utilizar la *Taq* DNA polimerasa. Estos amplicones poseen al final una adenina agregada por la *Taq* DNA Polymerase, lo que permite ligar el amplicón al vector pGEM®-T Easy linealizado, gracias a que cuenta con una timidina 3'-terminal en ambos extremos (pGEM®-T Easy Vector Systems, 2015). Los plásmidos

con cada inserto clonado fueron secuenciados con iniciadores cuyas secuencias con las que alinean se encuentran fuera del amplicón ligado, para obtener una lectura completa de las señales pertenecientes al producto ligado. Los iniciadores empleados fueron m13/pUC -40 Forward (iniciador directo) y m13/pUC Reverse (iniciador reverso). Por su tamaño, los amplicones GAP5-6 y GAP9-3 fueron secuenciados solo con el iniciador m13/pUC -40 Forward, mientras que para los amplicones con un tamaño mayor a 700 pb se utilizaron ambos iniciadores para secuenciarlos.

El análisis de secuencias comenzó con la depuración de cada uno de los electroferogramas, buscando la región correspondiente a los iniciadores utilizados en cada reacción de PCR y revisando que las señales (picos en el electroferograma) de cada nucleótido fueran definidas, utilizando el programa SnapGene. Posteriormente, se realizaron alineamientos de secuencias con el programa Vector NTI con el objeto de ubicar cada amplicón secuenciado en el E15. En los casos en los que la longitud del amplicón (GAP3-8 y GAP6-10I) requirió emplear los dos iniciadores, directo y reverso, los resultados de las secuenciaciones fue posible ensamblarlos en una sola secuencia. Todos los resultados de los electroferogramas pueden ser consultados en el Apéndice 3.

Todos los amplicones clonados, secuenciados y analizados resultaron inespecíficos para la región que se pretendía amplificar (Tabla 11). Las causas de estas amplificaciones inespecíficas pueden ser variadas, como una baja temperatura de alineamiento, una cantidad inadecuada de $MgCl_2$ que pudieron permitir la unión de los iniciadores a secuencias no específicas, generando amplicones inesperados.

Tabla 11. Análisis de las secuencias de los amplicones.

| Amplicón intercontig | Tamaño (pb) | Ubicación en el E15 | % Id | Posición en el contig |
|------------------------------------|------------------|---------------------|-------|-----------------------|
| GAP9-3 | 543 | Contig 1 | 98.8% | 389,948-390,490 |
| GAP3-8 | 813 | Contig 1 | 99.2% | 1'079,528-1'080,340 |
| GAP5-6 | 362 | Contig 6 | 99.0% | 46,861-47,222 |
| GAP6-10S | 593 ^a | Contig 1 | 99.0% | 1'339,350-1'339,942 |
| GAP6-10I | 1,415 | Contig 1 | 99.6% | 482,183-483,597 |
| ^b GAP4-2F (Clona 1 y 2) | 883 | Contig 1 | 99.2% | 1'129,082-1'129,964 |
| GAP4-2R (Clona 1 y 2) | 835 | Contig 1 | 99.4% | 1'133,054-1'133,888 |
| GAP4-2F (Clona 3) | 217 | Contig 5 | 86.6% | 42,781-42,997 |
| GAP4-2R (Clona 3) | 804 | Contig 5 | 99.4% | 38,329-39,132 |

^a El tamaño del amplicón corresponde a la secuencia con un solo iniciador pues con el otro no se logró obtener secuencia.

^b Para analizar el GAP4-2 se tomaron 3 colonias transformantes. Las clonas 1 y 2 correspondieron al mismo amplicón, no así la clona 3.

Todos los amplicones secuenciados fueron ubicados en el genoma de BQ1, es decir en ningún caso correspondieron a secuencias no incluidas en el ensamble. Debido a que todos los amplicones resultaron ser inespecíficos, un nuevo diseño de pares de iniciadores de mayor longitud, con temperaturas de alineamiento mayores a las utilizadas en este trabajo, así como probar diferentes DNA polimerasas podrían ayudar a amplificar las regiones intercontig faltantes.

Las contribuciones de este proyecto fueron la selección de un ensamble del genoma de *Alicyclophilus* sp. BQ1, el cual fue reportado en el GenBank, logrando tener la información de su secuencia, así como la anotación de todos sus genes para posteriores investigaciones para tratar de identificar las actividades enzimáticas responsables de la degradación del PU. Se logró mejorar el ensamble, reduciéndolo de 15 contigs a 12 contigs, logrando ubicar y ensamblar a los contigs 11 y 12 junto a los contigs 10 y 5 respectivamente. Asimismo, se logró unir los contigs 1 y 8 en uno solo.

Las 5,698 pb de la secuencia entre los contigs 2 y 4 no pudieron ser confirmadas bioinformáticamente, por lo que no se pueden añadir al tamaño del genoma. Sin embargo, esta propuesta es susceptible de ser probada en un futuro experimento diseñando iniciadores específicos para amplificar por PCR esa región, confirmando así su presencia en esa región del genoma de BQ1. Por otro lado, las secuencias que fueron adicionadas a los contigs del E15 analizando el E13 (2,712 nt), son secuencias que no fueron identificadas en ninguno de los contigs del E15, por lo que serán agregadas al tamaño del genoma, que inicialmente era de 4'676,255 pb, quedando en 4'678,967 pb. Por último, se muestra una representación final de los nuevos contigs obtenidos en este trabajo, a los que se les agregaron secuencias y los que permanecen iguales (Figura 21).

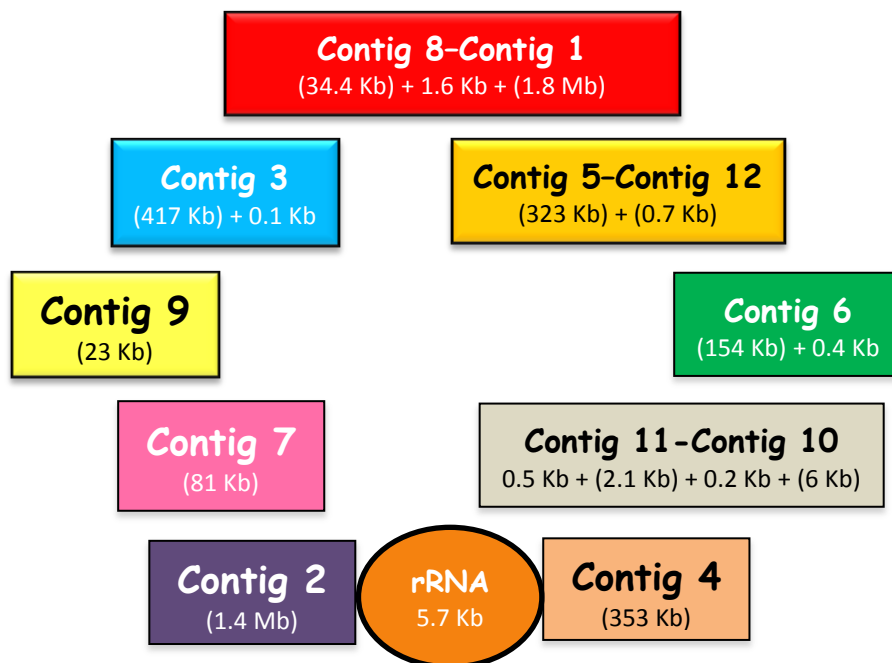


Figura 21. Orden del nuevo ensamblaje de *Alicyclophilus sp. BQ1* propuesto a partir del análisis realizado en este trabajo. Cada rectángulo representa un contig, los números entre paréntesis indican los tamaños iniciales de cada contig, los números sin paréntesis representan las secuencias que se pudieron adicionar a partir de este trabajo. Se incluye la región que se propone existe entre los contigs 4 y 2 correspondiente a un operón de rRNA. No se incluyen los contigs 13, 14 y 15 ya que aún no se conoce la posición que ocupan en el ensamblaje.

8. CONCLUSIONES

El ensamble de 15 contigs de *Alicyclophilus* sp. BQ1 fue reportado en la base de datos del GenBank.

Se logró obtener un ensamble del genoma de *Alicyclophilus* sp. BQ1 que consta de 12 contigs, 69% de GC y 4.6 Mb totales.

Mediante diversas herramientas bioinformáticas como el programa Mauve y los alineamientos de secuencias usando como referencia cepas de *Alicyclophilus denitrificans*, se logró proponer un orden posible para los contigs: 8, 1, 5, 6, 10, 4, 2, 7, 9 y 3. Se lograron unir bioinformáticamente el contig 8 con el 1, el 5 con el 12 y el 11 con el 10, además se agregaron secuencias a los contigs 3, 6 y 11, con lo que se logró reducir el número de contigs a 12.

Asimismo, debido a que otras cepas del género *Alicyclophilus* tienen más de una copia de los genes ribosomales y en BQ1 solo se encontró una, los resultados sugieren que la región faltante entre el contig 4 y 2 podría corresponder a una copia de estos genes debido a la semejanza de secuencia que hay entre los extremos final e inicial respectivamente de estos contigs y las regiones que rodean estos genes en *A. denitrificans*.

Todos los amplificadores obtenidos por PCR resultaron ser inespecíficos con respecto a la secuencia para la cual fueron diseñados. Lo anterior puede deberse al porcentaje de GC total del genoma del 69%, lo cual complicaría la amplificación de ciertas regiones.

9. PERSPECTIVAS

Diseñar nuevos iniciadores para cada extremo de los nuevos contigs, con una mayor longitud y temperaturas de alineamiento mayores a las utilizadas en este proyecto, para así favorecer la especificidad de estos.

Utilizar nuevamente todas las secuenciaciones realizadas e intentar usar nuevos ensambladores de genomas para seguir curando el ensamble del genoma utilizando el mismo método realizado en este trabajo, para con ello reducir aún mas el número de contigs y obtener mas información acerca de los genes y regiones reguladoras del genoma de *Alicyclophilus* sp. BQ1.

Debido a su importancia ambiental con respecto a la degradación del PU, realizar trabajos de genómica comparativa con respecto a otras cepas del mismo género para plantear cuales genes presentes en BQ1 pueden ser los responsables de su actividad degradativa sobre el poliuretano.

Identificando algunos genes que pudieran estar implicados en la degradación del PU, se podrían clonar para producir proteína recombinante, con el fin de montar un proceso alternativo en el tratamiento de residuos del PU.

APÉNDICE 1

Metodologías complementarias, medios de cultivo, soluciones y reactivos

Protocolo de transformación de *E. coli* por electroporación

1. Colocar en hielo 40 μ L de células competentes en microtubos de 1.5 mL y las celdas para electroporar estériles (BIO-RAD).
2. Calentar el medio SOC a 37 °C en un baño de agua (Precision Scientific) (1 tubo con 1 mL por cada tubo de células que se transformarán).
3. Agregar 25 ng del plásmido que se desea introducir en las bacterias competentes a cada tubo de bacterias. Incubar tubos de células en hielo durante 1 min.
4. Colocar la mezcla DNA + células en la celda de electroporación y verificar que no se formen burbujas. Limpiar la humedad de la celda, colocar en el electroporador (BIO-RAD) y generar el pulso a 2.00 KV.
5. Sacar la celda de electroporación y agregar 1.0 mL de medio SOC a 37 °C. Resuspender suavemente con micropipeta.
6. Colocar en un microtubo de 1.5 mL y colocar en una incubadora (Thermo Scientific) durante 2 h a 37 °C a 210 rpm.

Preparación de células competentes de *E. coli* DH5 α

1. Inocular una colonia aislada de DH5 α de un agar fresco en un matraz que contiene 50 mL de medio LB. Cultivar por una noche a 37°C con agitación (250 rpm).
2. Inocular dos alícuotas de 500 mL de medio LB en frascos separados (matraces de 2 L) con 25 mL del cultivo por una noche.
3. Incubar los matraces a 37°C, con agitación, 300 ciclos (250 rpm). Medir la densidad óptica a 600 nm cada 20 min hasta 0.35-0.4 (alrededor de 25 h).
4. Colocar los matraces en hielo por 15-30 min. (Agitando los matraces para que el frío se distribuya). Se van colocando los tubos de centrifuga en hielo para mantenerlos fríos. Transferir cultivos a botellas frías. Cosechar células por centrifugación (2500 rpm por 20 min a 4°C).

5. Decantar sobrenadante y resuspender el pellet con 500 mL de agua fría. Cosechar células por centrifugación a 2500 rpm durante 20 min a 4°C. Decantar el sobrenadante y resuspender el pellet en 250 mL de glicerol al 10 % v/v.
6. Cosechar células por centrifugación a 2500 rpm durante 20 min a 4°C. Decantar sobrenadante y resuspender el pellet en 10 mL de glicerol al 10 % v/v (decantar cuidadosamente, las células se adhieren poco en glicerol).
7. Cosechar células por centrifugación a 2500 rpm durante 20 min a 4°C. Decantar sobrenadante en su totalidad usando micropipeta para remover el buffer. Resuspender en 1 mL de GYT frío.
8. Medir la D.O. a 600 nm de una dilución 1:100. Diluir la suspensión hasta una concentración 2×10^{10} - 3×10^{10} (1.0 D.O. a 600 nm= 2.5×10^8 células/mL) con medio GYT frío.
9. Almacenar a -70°C en alícuotas de 40 µL.

Medio LB. Composición por litro

| | |
|----------------------|--------|
| Peptona | 10.0 g |
| Extracto de levadura | 5.0 g |
| NaCl | 10.0 g |

1. Disolver todos los componente en el orden que se indica en 700 mL de agua desionizada. Cuando la solución esté totalmente cristalina, ajustar a pH7 y aforar hasta un volumen final de un litro.
2. En caso de requerir medio sólido, se agregan 16 g de agar.
3. Esterilizar por autoclave a 120°C/15 minutos.
4. Las placas de LB sólido se preparan vertiendo el medio ya estéril y caliente sobre cajas Petri estériles desechables. Se deja enfriar y solidificar el medio por 20-30 minutos.

Nota: El medio LB convencional fue modificado ya que se adicionó el doble de sal para reducir la probabilidad de contaminación ambiental.

Medio SOC. Composición por litro

| | |
|----------------------|--------|
| Triptona | 20.0 g |
| Extracto de levadura | 5.0 g |
| NaCl | 0.5 g |

1. Disolver en 950 mL de agua desionizada los siguientes componentes.
2. Añadir 10 mL de solución de KCl 250 mM. Ajustar pH a 7.0 con NaOH 5N y esterilizar por autoclave a 120°C por 20 minutos.
3. Antes de utilizarla, añadir 5 mL de una solución estéril de MgCl₂ 2M (esterilizada también por autoclave a 120°C por 20 minutos) y 20 mL de una solución de glucosa 1M, esterilizada por filtración con membrana de poro de 0.22 µM (Millipore™).
4. Finalmente, ajustar el volumen final a 1L con agua desionizada estéril.

Aislamiento de DNA genómico

Solución amortiguadora TE

| | |
|------------------------|-------|
| Tris-HCl 10 mM pH= 8.0 | 1 mL |
| EDTA 10 mM | 2 mL |
| SDS 1% | 10 mL |

1. Colocar las tres soluciones en agua desionizada y aforar a un volumen final de 100 mL.
2. Para preparación el Tris-HCl 10 mM pH= 8.0, se disuelven 121 g de Tris base en 800 mL de agua desionizada, ajustar al pH deseado con HCl concentrado. Agitar y aforar a 1L con agua desionizada. Conservar a temperatura ambiente.

Solución de lisozima (2 mg/mL)

1. Preparar una solución Stock de lisozima a una concentración de 50 mg/mL en agua desionizada.
2. Distribuir en alícuotas y diluir hasta una concentración de 20 mg/mL, almacenar las alícuotas a -20°C.

Solución SDS (10 %)

1. Disolver 100 g de SDS en 900 ml de agua desionizada.
2. Calentar a 68°C hasta disolver. Ajustar a pH 7.2 adicionando gotas de HCl.
3. Ajustar el volumen final con agua hasta un litro.

Solución NaCl 5 M

1. Agregar 292.2 g de NaCl en 900 ml de agua desionizada.
2. Disolver por agitación y ajustar el volumen final con agua hasta un litro.

CTAB/NaCl (10% de CTAB en NaCl 0.7 M)

NaCl 4.1g

CTAB 10 g

1. Disolver el NaCl en 80 mL de agua desionizada.
2. Agregar lentamente 10 g de CTAB con agitación y calentamiento constante de la solución.
3. Ajustar a un volumen final de 100 mL.

Electroforesis de DNA en agarosa

Solución amortiguadora TAE 50X. Composición por litro

Tris base 242.0 g

Ácido acético glacial 57.1 mL

Solución de EDTA 0.5 M 100.0 mL

Agregar las componentes anteriores y llevar a un volumen final de un litro con agua desionizada con un pH final de 8.0. Esterilizar en autoclave a 120°C por 20 minutos. Almacenar a temperatura ambiente.

Solución amortiguadora de carga para geles de agarosa 6X

Azul de bromofenol 0.25% p/v

Xilencianol 0.25% p/v

Glicerol en agua 30.0 % v/v

1. Disolver en la solución de glicerol con agua desionizada estéril lo necesario de cada colorante.
2. Mezclar bien la solución y mantener en refrigeración a 4°C.

Selección de clonas transformantes

Solución de ampicilina

1. Hacer una solución stock de ampicilina a una concentración de 100 mg/mL en agua desionizada.
2. Esterilizar por filtración con membrana de poro de 0.22 µM (Millipore™).
3. Almacenar a -20 °C.

Solución de 5-bromo-4-cloro-3-indolil-β-D-galactopiranosido (X-Gal)

1. Disolver X-gal en dimetilformamida a una concentración de 20 mg/mL (2% p/v) en un tubo de vidrio o polipropileno.
2. Realizar alícuotas y envolver el tubo con papel aluminio y almacenar a -20 °C. No es necesario esterilizar por filtración.

Solución de isopropil tio-β-D-galactopiranosido (IPTG)

1. Disolver IPTG en agua desionizada a una concentración de 200 mg/mL.
2. Esterilizar por filtración con membrana de poro de 0.22 µM (Millipore™).
3. Realizar alícuotas y almacenar a -20 °C.

Apéndice 2

Resultado de BLAST para los contigs de E15 empleando como referencia los genomas de *Alicyclophilus denitrificans* BC y K601.

Tabla A1. Resultados del BLAST del E15 con respecto a *Alicyclophilus denitrificans* BC.

| Coordenadas dentro del genoma de BC | pb del contig correspondiente | Contig | % Identidad |
|-------------------------------------|-------------------------------|----------|-------------|
| 5,627-21,543 | 350,000-417,414 | Contig 3 | 97 |
| 97,986-121,616 | 1'250,000-1'300,000 | Contig 1 | 98 |
| 129,917-144,271 | 1'200,000-1'250,000 | Contig 1 | 97 |
| 167,260-173,514 | 1'150,000-1'200,000 | Contig 1 | 95 |
| 203,051-209,758 | 1'100,000-1'150,000 | Contig 1 | 97 |
| 320,144-339,936 | 1'050,000-1'100,000 | Contig 1 | 96 |
| 339,937-346,740 | 1'000,000-1'050,000 | Contig 1 | 96 |
| 407,305-414,822 | 950,000-1'000,000 | Contig 1 | 95 |
| 426,119-445,383 | 900,000-950,000 | Contig 1 | 97 |
| 478,247-492,627 | 850,000-900,000 | Contig 1 | 97 |
| 533,033-545,470 | 800,000-850,000 | Contig 1 | 98 |
| 549,051-554,464 | 1-7,000 | Contig 8 | 96 |
| 554,465-561,470 | 7,000-14,000 | Contig 8 | 95 |
| 561,474-568,468 | 14,000-21,000 | Contig 8 | 97 |
| 569,072-572,993 | 21,000-28,000 | Contig 8 | 97 |
| 572,994-579,450 | 28,000-34,454 | Contig 8 | 98 |
| 580,962-596,087 | 1-50,000 | Contig 1 | 96 |
| 629,264-645,939 | 50,000-100,000 | Contig 1 | 98 |
| 693,975-715,802 | 100,000-150,000 | Contig 1 | 97 |
| 765,508-786,218 | 150,000-200,000 | Contig 1 | 95 |
| 803,627-820,613 | 200,000-250,000 | Contig 1 | 96 |
| 833,129-850,979 | 250,000-300,000 | Contig 1 | 98 |
| 880,722-915,746 | 300,000-350,000 | Contig 1 | 97 |
| 927,726- 961,042 | 350,000-400,000 | Contig 1 | 97 |
| 961,050-975,237 | 400,000-450,000 | Contig 1 | 96 |
| 1'040,853-1'050,702 | 450,000-500,000 | Contig 1 | 97 |
| 1'069,638-1'076,980 | 500,000-550,000 | Contig 1 | 98 |
| 1'082,450-1'130,547 | 550,000-600,000 | Contig 1 | 97 |
| 1'132,533-1'148,065 | 600,000-650,000 | Contig 1 | 96 |
| 1'185,053-1'204,622 | 650,000-700,000 | Contig 1 | 97 |
| 1'235,255-1'250,079 | 700,000-750,000 | Contig 1 | 97 |
| 1'274,219-1'287,542 | 750,000-800,000 | Contig 1 | 97 |
| 1'354,819-1'377,878 | 1-50,000 | Contig 5 | 97 |
| 1'409,011-1'431,960 | 50,000-100,000 | Contig 5 | 97 |

| | | | |
|---------------------|---------------------|-----------|----|
| 1'469,955-1'481,124 | 100,000-150,000 | Contig 5 | 96 |
| 1'516,026-1'527,363 | 150,000-200,000 | Contig 5 | 97 |
| 1'531,023-1'552,156 | 200,000-250,000 | Contig 5 | 97 |
| 1'605,284-1'629,328 | 250,000-300,000 | Contig 5 | 97 |
| 1'637,579-1'641,962 | 300,000-323,056 | Contig 5 | 94 |
| 1'659,840-1'680,528 | 1-50,000 | Contig 6 | 97 |
| 1'797,743-1'807,756 | 50,000-100,000 | Contig 6 | 97 |
| 1'819,597-1'857,020 | 100,000-154,131 | Contig 6 | 97 |
| 1'876,851-1'879,350 | 1-6,132 | Contig 10 | 94 |
| 1'904,147-1'917,935 | 1'300,000-1'350,000 | Contig 1 | 97 |
| 1'977,012-2'006,916 | 1'350,000-1'400,000 | Contig 1 | 97 |
| 2'006,917-2'023,810 | 1'400,000-1'450,000 | Contig 1 | 94 |
| 2'045,881-2'084,292 | 1'450,000-1'500,000 | Contig 1 | 96 |
| 2'084,293-2'107,321 | 1'500,000-1'550,000 | Contig 1 | 97 |
| 2'198,249-2'216,519 | 1'550,000-1'600,000 | Contig 1 | 94 |
| 2'244,232-2'261,236 | 1'600,000-1'650,000 | Contig 1 | 97 |
| 2'289,426-2'304,800 | 1'650,000-1'700,000 | Contig 1 | 96 |
| 2'337,040-2'350,437 | 1'700,000-1'750,000 | Contig 1 | 96 |
| 2'368,734-2'377,119 | 1'750,000-1'800,000 | Contig 1 | 96 |
| 2'382,365-2'397,707 | 1'800,000-1'847,482 | Contig 1 | 96 |
| 2'425,151-2'442,605 | 750,000-800,000 | Contig 2 | 97 |
| 2'468,324-2'508,738 | 800,000-850,000 | Contig 2 | 98 |
| 2'508,739-2'530,401 | 850,000-900,000 | Contig 2 | 96 |
| 2'573,178-2'587,104 | 900,000-950,000 | Contig 2 | 98 |
| 2'617,901-2'630,927 | 950,000-1'000,000 | Contig 2 | 97 |
| 2'634,154-2'648,173 | 1'000,000-1'050,000 | Contig 2 | 96 |
| 2'703,566-2'717,001 | 1'050,000-1'100,000 | Contig 2 | 97 |
| 2'751,818-2'761,653 | 1'100,000-1'150,000 | Contig 2 | 97 |
| 2'818,592-2'824,405 | 100,000-150,000 | Contig 4 | 84 |
| 2'872,519-2'898,815 | 1-50,000 | Contig 4 | 97 |
| 2'898,816-2'918,731 | 50,000-100,000 | Contig 4 | 97 |
| 2'992,276-3'012,846 | 150,000-200,000 | Contig 4 | 96 |
| 3'031,338-3'057,323 | 200,000-250,000 | Contig 4 | 96 |
| 3'070,954-3'077,573 | 250,000-300,000 | Contig 4 | 95 |
| 3'091,261-3'109,264 | 300,000-353,217 | Contig 4 | 97 |
| 3'144,328-3'173,356 | 1-50,000 | Contig 2 | 97 |
| 3'197,546-3'215,080 | 50,000-100,000 | Contig 2 | 97 |
| 3'267,791-3'285,276 | 100,000-150,000 | Contig 2 | 97 |
| 3'285,353-3'306,642 | 150,000-200,000 | Contig 2 | 97 |
| 3'322,738-3'349,893 | 200,000-250,000 | Contig 2 | 95 |
| 3'398,256-3'410,306 | 250,000-300,000 | Contig 2 | 96 |
| 3'458,482-3'496,031 | 300,000-350,000 | Contig 2 | 96 |
| 3'497,437-3'547,474 | 350,000-400,000 | Contig 2 | 97 |
| 3'565,877-3'576,799 | 400,000-450,000 | Contig 2 | 96 |

| | | | |
|---------------------|---------------------|----------|----|
| 3'601,339-3'618,789 | 450,000-500,000 | Contig 2 | 98 |
| 3'670,525-3'685,181 | 500,000-550,000 | Contig 2 | 96 |
| 3'747,694-3'765,743 | 550,000-600,000 | Contig 2 | 96 |
| 3'776,117-3'796,460 | 600,000-650,000 | Contig 2 | 99 |
| 3'814,289-3'836,361 | 650,000-700,000 | Contig 2 | 97 |
| 3'859,880-3'885,798 | 700,000-750,000 | Contig 2 | 97 |
| 3'914,838-3'922,563 | 1-20,000 | Contig 7 | 97 |
| 3'934,046-3'950,538 | 20,000-40,000 | Contig 7 | 97 |
| 3'960,377-3'968,872 | 40,000-60,000 | Contig 7 | 97 |
| 3'973,393-3'985,324 | 60,000-81,368 | Contig 7 | 97 |
| 3'998,497-4'018,863 | 1'150,000-1'200,000 | Contig 2 | 98 |
| 4'039,116-4'048,508 | 1'400,000-1'431,080 | Contig 2 | 98 |
| 4'075,835-4'089,590 | 1'350,000-1'400,000 | Contig 2 | 97 |
| 4'114,760-4'145,061 | 1'300,000-1'350,000 | Contig 2 | 97 |
| 4'218,236-4'230,328 | 1'250,000-1'300,000 | Contig 2 | 96 |
| 4'235,208-4'255,345 | 1'200,000-1'250,000 | Contig 2 | 97 |
| 4'283,294-4'289,284 | 1-6,000 | Contig 9 | 97 |
| 4'289,285-4'293,326 | 6,000-12,000 | Contig 9 | 97 |
| 4'297,950-4'301,486 | 12,000-18,000 | Contig 9 | 98 |
| 4'301,487-4'305,278 | 18,000-23,292 | Contig 9 | 97 |
| 4'342,185-4'353,036 | 1-50,000 | Contig 3 | 97 |
| 4'362,812-4'371,675 | 50,000-100,000 | Contig 3 | 97 |
| 4'401,108-4'417,109 | 100,000-150,000 | Contig 3 | 98 |
| 4'434,567-4'458,715 | 150,000-200,000 | Contig 3 | 98 |
| 4'500,870-4'522,222 | 200,000-250,000 | Contig 3 | 97 |
| 4'569,856-4'580,409 | 250,000-300,000 | Contig 3 | 95 |
| 4'605,456-4'619,547 | 300,000-350,000 | Contig 3 | 97 |

Tabla A2. Resultados del BLAST del E15 con respecto a *Alicyclophilus denitrificans* K601.

| Coordenadas dentro del genoma de K601 | pb del contig correspondiente | Contig | % Identidad |
|--|--------------------------------------|---------------|--------------------|
| 10,457-26,379 | 350,000-417,414 | Contig 3 | 97 |
| 98,937-113,642 | 1'250,000-1'300,000 | Contig 1 | 97 |
| 121,943-136,285 | 1'200,000-1'250,000 | Contig 1 | 96 |
| 158,320-166,951 | 1'150,000-1'200,000 | Contig 1 | 96 |
| 192,788-199,495 | 1'100,000-1'150,000 | Contig 1 | 96 |
| 256,242-286,314 | 1'050,000-1'100,000 | Contig 1 | 96 |
| 286,315-293,118 | 1'000,000-1'050,000 | Contig 1 | 96 |
| 352,348-359,865 | 950,000-1'000,000 | Contig 1 | 96 |
| 371,162-390,426 | 900,000-950,000 | Contig 1 | 97 |
| 425,788-440,171 | 850,000-900,000 | Contig 1 | 97 |
| 457,873-471,428 | 800,000-850,000 | Contig 1 | 95 |
| 498,327-503,737 | 1-7,000 | Contig 8 | 96 |
| 503,738-510,743 | 7,000-14,000 | Contig 8 | 95 |
| 510,747-515,011 | 14,000-21,000 | Contig 8 | 97 |
| 518,262-522,183 | 21,000-28,000 | Contig 8 | 97 |
| 522,184-528,639 | 28,000-34,454 | Contig 8 | 99 |
| 530,151-541,251 | 1-50,000 | Contig 1 | 96 |
| 584,054-600,731 | 50,000-100,000 | Contig 1 | 97 |
| 646,381-663,703 | 100,000-150,000 | Contig 1 | 97 |
| 704,828-725,548 | 150,000-200,000 | Contig 1 | 94 |
| 753,145-774,497 | 200,000-250,000 | Contig 1 | 97 |
| 774,498-792,348 | 250,000-300,000 | Contig 1 | 98 |
| 822,100-857,123 | 300,000-350,000 | Contig 1 | 97 |
| 869,103-902,420 | 350,000-400,000 | Contig 1 | 97 |
| 902,428-916,617 | 400,000-450,000 | Contig 1 | 96 |
| 986,141-995,990 | 450,000-500,000 | Contig 1 | 97 |
| 1'014,211-1'031,263 | 450,000-500,000 | Contig 2 | 98 |
| 1'055,817-1'066,733 | 400,000-450,000 | Contig 2 | 97 |
| 1'085,142-1'122,572 | 350,000-400,000 | Contig 2 | 97 |
| 1'148,271-1'192,560 | 300,000-350,000 | Contig 2 | 96 |
| 1'234,115-1'246,165 | 250,000-300,000 | Contig 2 | 96 |
| 1'294,533-1'321,709 | 200,000-250,000 | Contig 2 | 95 |
| 1'337,805-1'359,094 | 150,000-200,000 | Contig 2 | 97 |
| 1'359,171-1'376,653 | 100,000-150,000 | Contig 2 | 97 |
| 1'429,368-1'446,907 | 50,000-100,000 | Contig 2 | 97 |
| 1'471,096-1'500,125 | 1-50,000 | Contig 2 | 97 |
| 1'535,189-1'553,193 | 300,000-353,217 | Contig 4 | 97 |
| 1'566,709-1'573,329 | 250,000-300,000 | Contig 4 | 95 |
| 1'586,960-1'612,928 | 200,000-250,000 | Contig 4 | 96 |
| 1'631,320-1'645,810 | 150,000-200,000 | Contig 4 | 96 |

| | | | |
|---------------------|---------------------|-----------|----|
| 1'682,022-1'688,712 | 100,000-150,000 | Contig 4 | 89 |
| 1'733,546-1'753,461 | 50,000-100,000 | Contig 4 | 97 |
| 1'776,587-1'792,425 | 1-50,000 | Contig 4 | 97 |
| 2'072,609-2'086,397 | 1'300,000-1'350,000 | Contig 1 | 97 |
| 2'209,605-2'239,510 | 1'350,000-1'400,000 | Contig 1 | 97 |
| 2'239,511-2'255,533 | 1'400,000-1'450,000 | Contig 1 | 95 |
| 2'344,513-2'367,655 | 1'450,000-1'500,000 | Contig 1 | 97 |
| 2'367,656-2'390,685 | 1'500,000-1'550,000 | Contig 1 | 97 |
| 2'441,185-2'459,470 | 1'550,000-1'600,000 | Contig 1 | 94 |
| 2'487,182-2'504,188 | 1'600,000-1'650,000 | Contig 1 | 97 |
| 2'534,220-2'549,441 | 1'650,000-1'700,000 | Contig 1 | 96 |
| 2'581,681-2'595,078 | 1'700,000-1'750,000 | Contig 1 | 96 |
| 2'613,381-2'621,766 | 1'750,000-1'800,000 | Contig 1 | 96 |
| 2'627,012-2'642,354 | 1'800,000-1'847,482 | Contig 1 | 96 |
| 2'669,805-2'687,252 | 750,000-800,000 | Contig 2 | 97 |
| 2'718,235-2'736,192 | 800,000-850,000 | Contig 2 | 97 |
| 2'736,193-2'757,855 | 850,000-900,000 | Contig 2 | 96 |
| 2'799,271-2'813,197 | 900,000-950,000 | Contig 2 | 98 |
| 2'982,898-2'995,923 | 950,000-1'000,000 | Contig 2 | 97 |
| 3'013,099-3'023,044 | 1'000,000-1'050,000 | Contig 2 | 97 |
| 3'068,487-3'081,922 | 1'050,000-1'100,000 | Contig 2 | 97 |
| 3'154,054-3'163,889 | 1'100,000-1'150,000 | Contig 2 | 97 |
| 3'216,306-3'217,395 | 1-6,132 | Contig 10 | 96 |
| 3'237,209-3'274,631 | 100,000-154,131 | Contig 6 | 97 |
| 3'286,471-3'296,474 | 50,000-100,000 | Contig 6 | 97 |
| 3'345,851-3'366,535 | 1-50,000 | Contig 6 | 97 |
| 3'382,349-3'386,729 | 300,000-323,056 | Contig 5 | 94 |
| 3'391,281-3'420,693 | 250,000-300,000 | Contig 5 | 96 |
| 3'596,801-3'617,938 | 200,000-250,000 | Contig 5 | 97 |
| 3'635,190-3'646,959 | 150,000-200,000 | Contig 5 | 97 |
| 3'653,585-3'664,743 | 100,000-150,000 | Contig 5 | 97 |
| 3'712,506-3'737,200 | 50,000-100,000 | Contig 5 | 97 |
| 3'777,450-3'794,236 | 1-50,000 | Contig 5 | 97 |
| 3'828,637-3'843,103 | 750,000-800,000 | Contig 1 | 98 |
| 3'875,316-3'890,131 | 700,000-750,000 | Contig 1 | 98 |
| 3'906,786-3'926,374 | 650,000-700,000 | Contig 1 | 98 |
| 3'966,345-3'988,445 | 650,000-700,000 | Contig 2 | 97 |
| 4'035,112-4'046,347 | 600,000-650,000 | Contig 2 | 98 |
| 4'056,733-4'074,794 | 550,000-600,000 | Contig 2 | 96 |
| 4'074,768-4'089,482 | 500,000-550,000 | Contig 2 | 96 |
| 4'134,949-4'141,419 | 500,000-550,000 | Contig 1 | 97 |
| 4'146,889-4'194,989 | 550,000-600,000 | Contig 1 | 97 |
| 4'196,963-4'212,491 | 600,000-650,000 | Contig 1 | 96 |
| 4'233,018-4'258,955 | 700,000-750,000 | Contig 2 | 97 |

| | | | |
|---------------------|---------------------|----------|----|
| 4'287,990-4'295,713 | 1-20,000 | Contig 7 | 97 |
| 4'301,713-4'318,761 | 20,000-40,000 | Contig 7 | 97 |
| 4'328,600-4'337,095 | 40,000-60,000 | Contig 7 | 97 |
| 4'341,616-4'353,547 | 60,000-81,368 | Contig 7 | 97 |
| 4'366,768-4'387,131 | 1'150,000-1'200,000 | Contig 2 | 98 |
| 4'407,384-4'416,777 | 1'400,000-1'431,080 | Contig 2 | 98 |
| 4'444,104-4'457,840 | 1'350,000-1'400,000 | Contig 2 | 97 |
| 4'481,684-4'504,214 | 1'300,000-1'350,000 | Contig 2 | 96 |
| 4'590,905-4'602,331 | 1'250,000-1'300,000 | Contig 2 | 98 |
| 4'644,137-4'655,414 | 1'200,000-1'250,000 | Contig 2 | 97 |
| 4'663,705-4'669,697 | 1-6,000 | Contig 9 | 97 |
| 4'669,698-4'673,739 | 6,000-12,000 | Contig 9 | 97 |
| 4'678,327-4'681,863 | 12,000-18,000 | Contig 9 | 98 |
| 4'681,864-4'685,655 | 18,000-23,292 | Contig 9 | 97 |
| 4'710,789-4'721,640 | 1-50,000 | Contig 3 | 97 |
| 4'731,309-4'740,166 | 50,000-100,000 | Contig 3 | 97 |
| 4'769,599-4'785,594 | 100,000-150,000 | Contig 3 | 98 |
| 4'803,053-4'827,203 | 150,000-200,000 | Contig 3 | 98 |
| 4'849,333-4'882,812 | 200,000-250,000 | Contig 3 | 98 |
| 4'930,257-4'940,810 | 250,000-300,000 | Contig 3 | 95 |
| 4'963,603-4'977,694 | 300,000-350,000 | Contig 3 | 97 |

APÉNDICE 3

Electroferogramas de las secuencias para cada amplicón obtenido por PCR

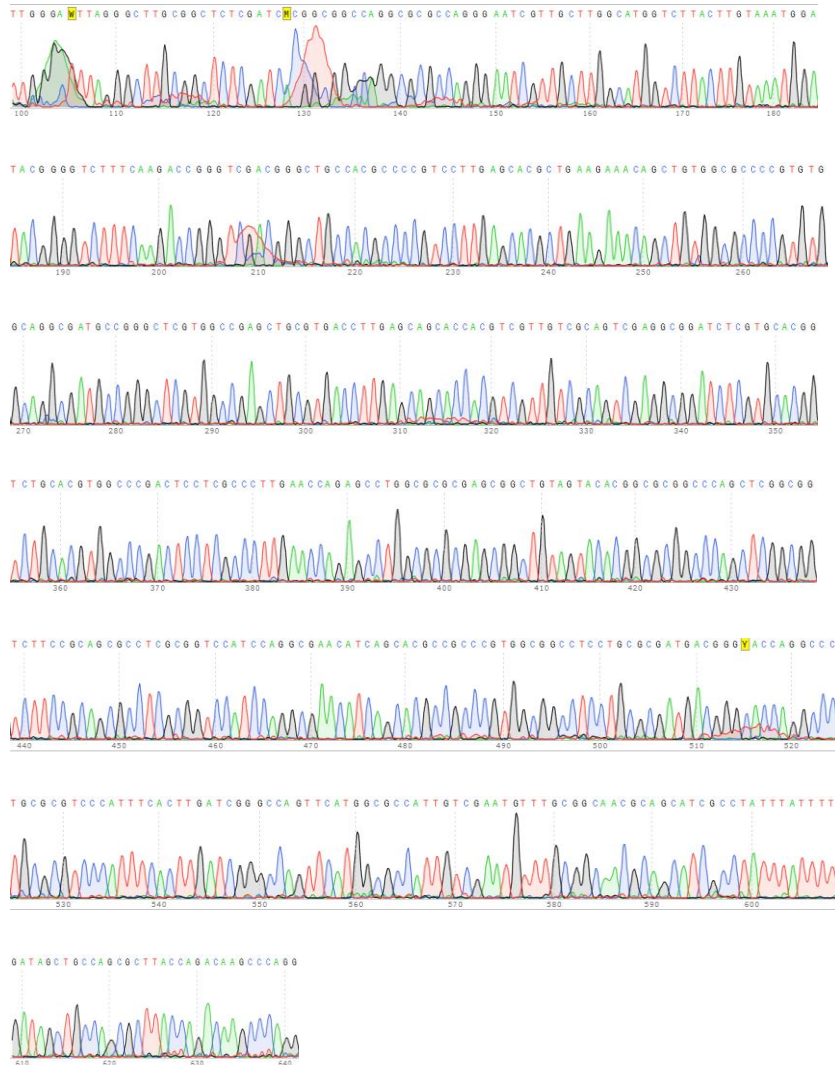


Figura A1. Electroferograma del GAP9-3 secuenciado con el iniciador m13/pUC -40 FORWARD.

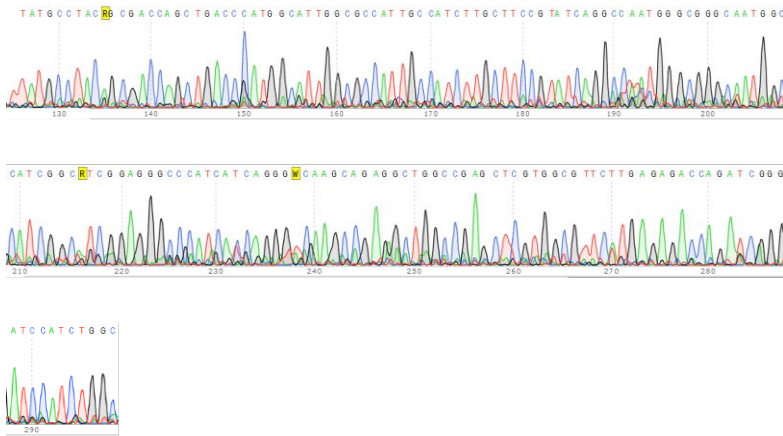


Figura A2. Electroferograma del GAP3-8 secuenciado con el iniciador m13/pUC -40 FORWARD.

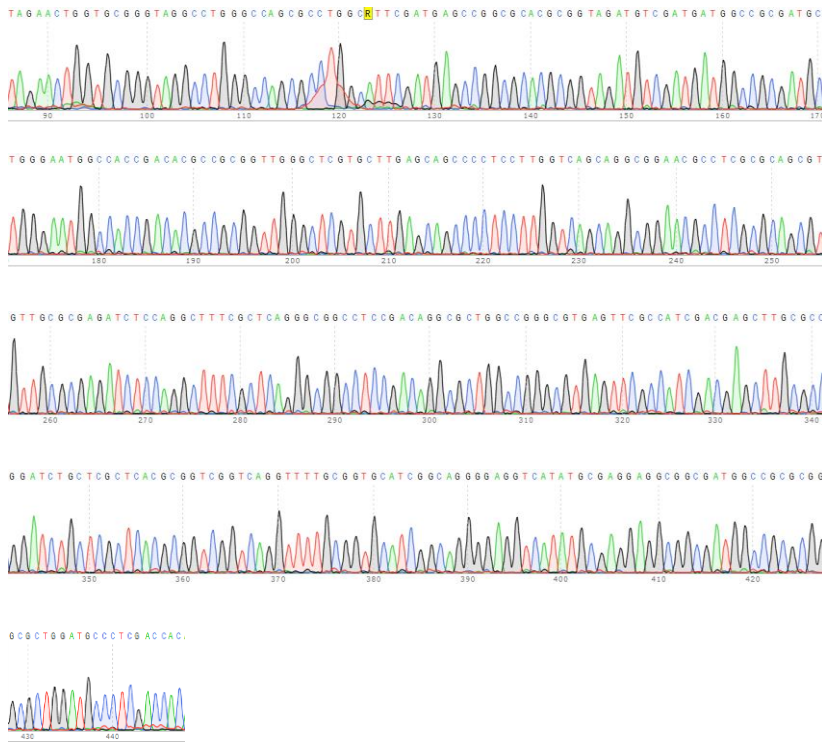


Figura A4. Electroferograma del GAP5-6 secuenciado con el iniciador m13/pUC -40 FORWARD.

No se obtuvo un electroferograma para el GAP6-10S secuenciado con el iniciador m13/pUC -40 FORWARD

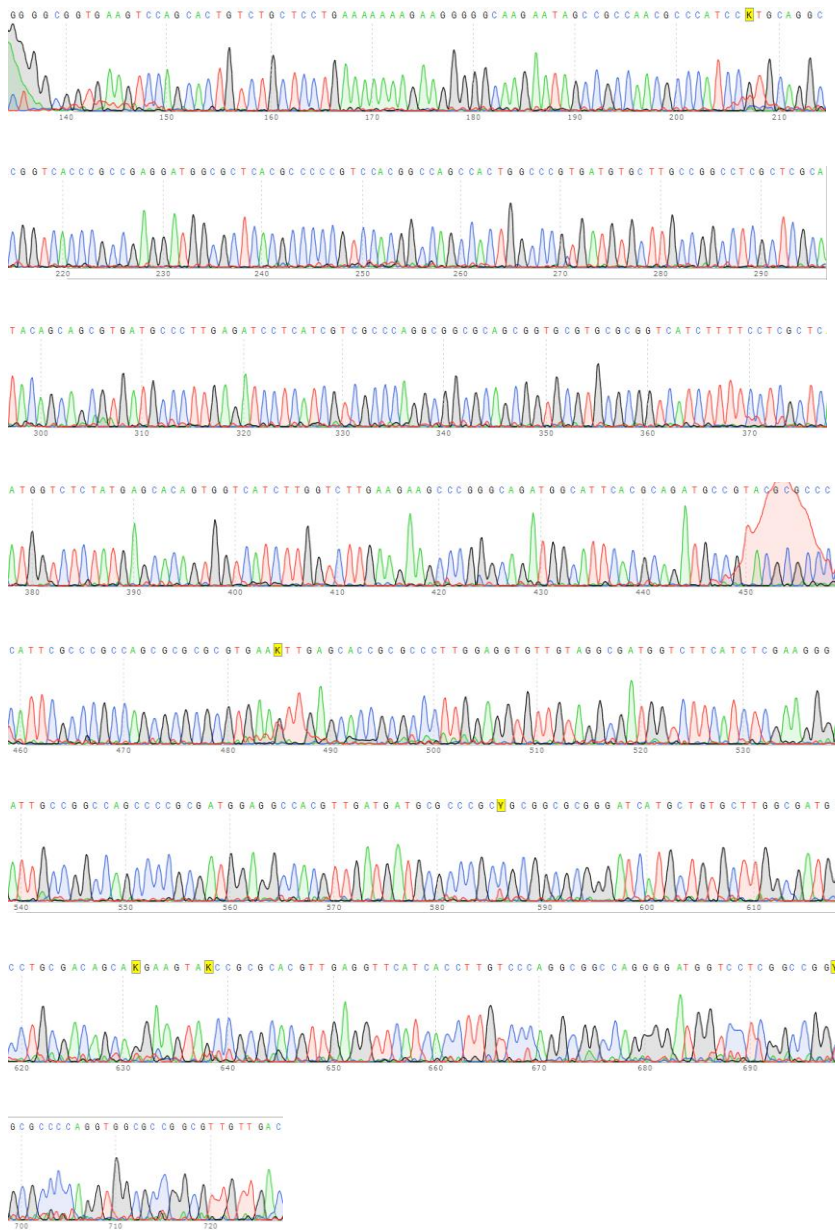


Figura A5. Electroferograma del GAP6-10S secuenciado con el iniciador m13/pUC REVERSE.

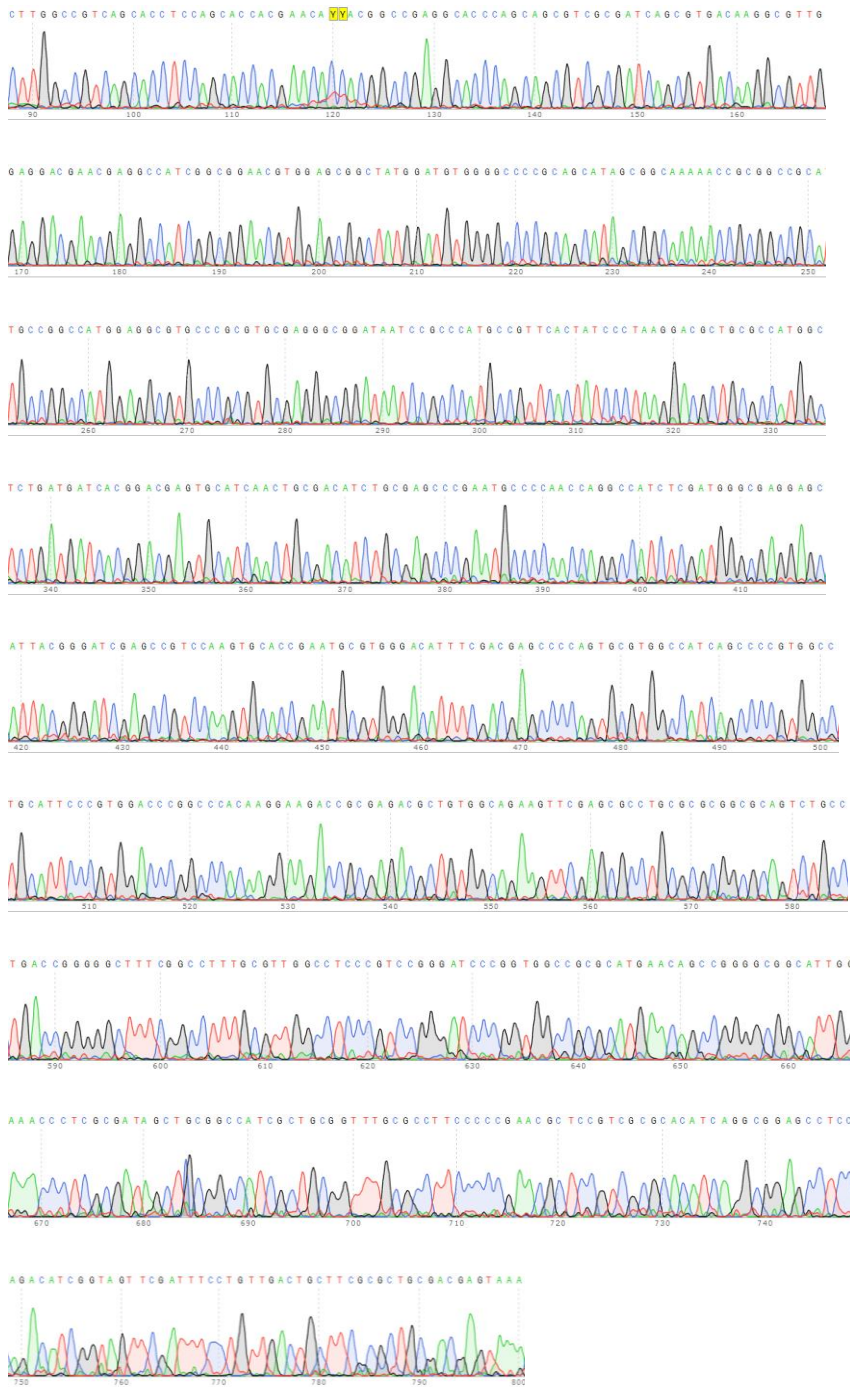


Figura A6. Electroferograma del GAP6-10l secuenciado con el iniciador m13/pUC - 40 FORWARD.

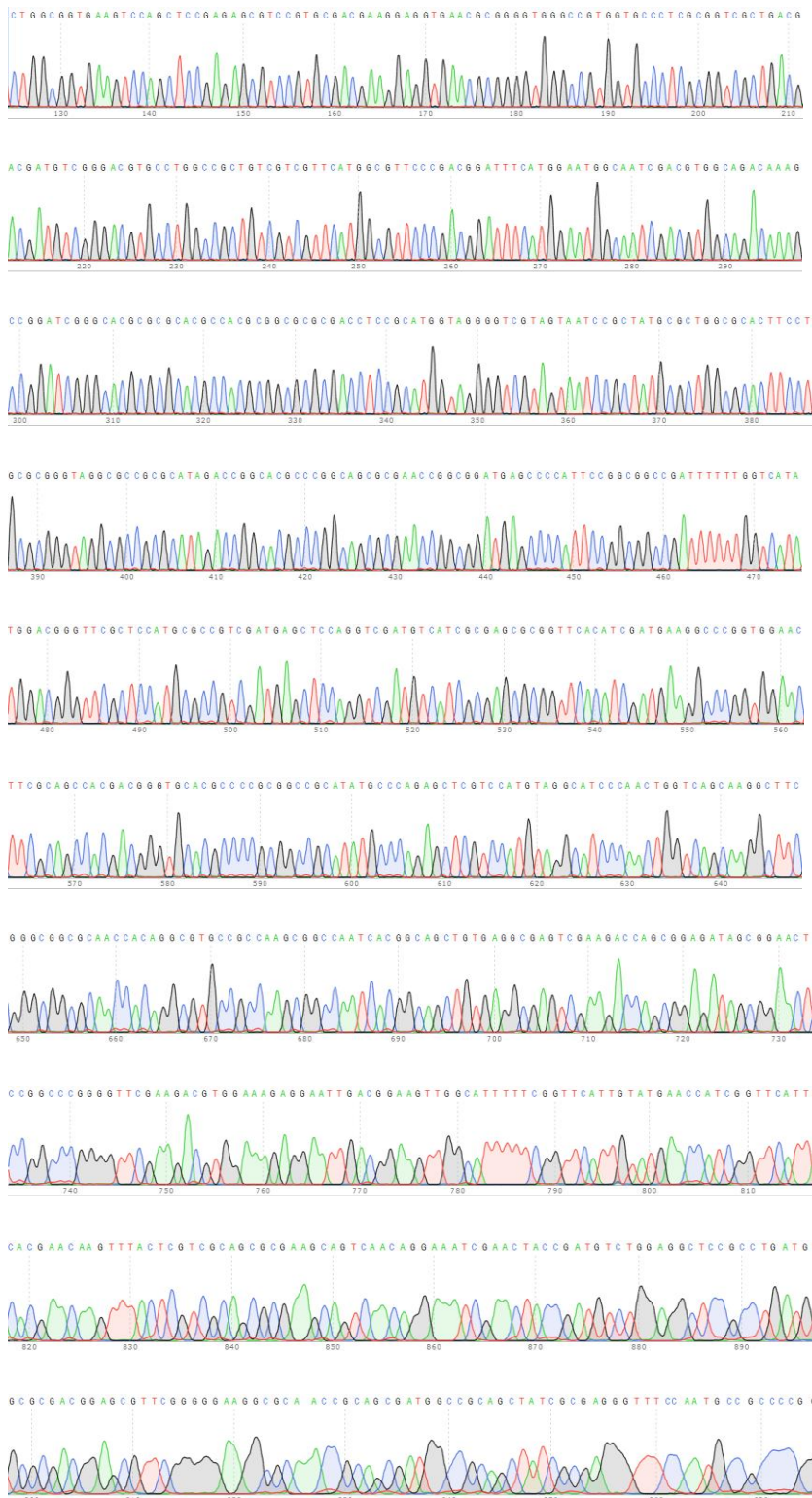


Figura A7. Electroferograma del GAP6-10I secuenciado con el iniciador m13/pUC REVERSE.

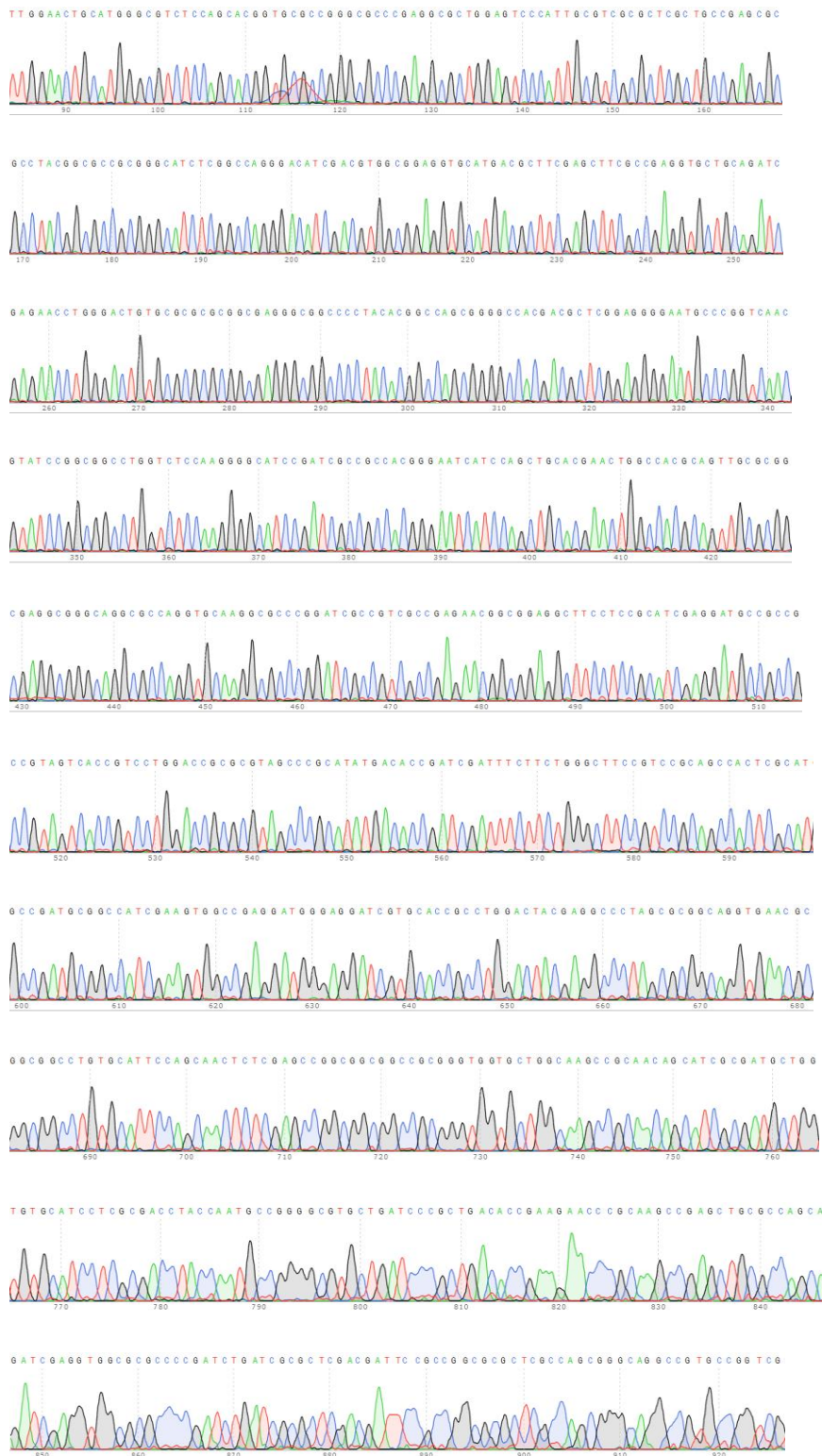


Figura A8. Electroferograma del GAP4-2, clona 1 secuenciado con el iniciador m13/pUC -40 FORWARD.

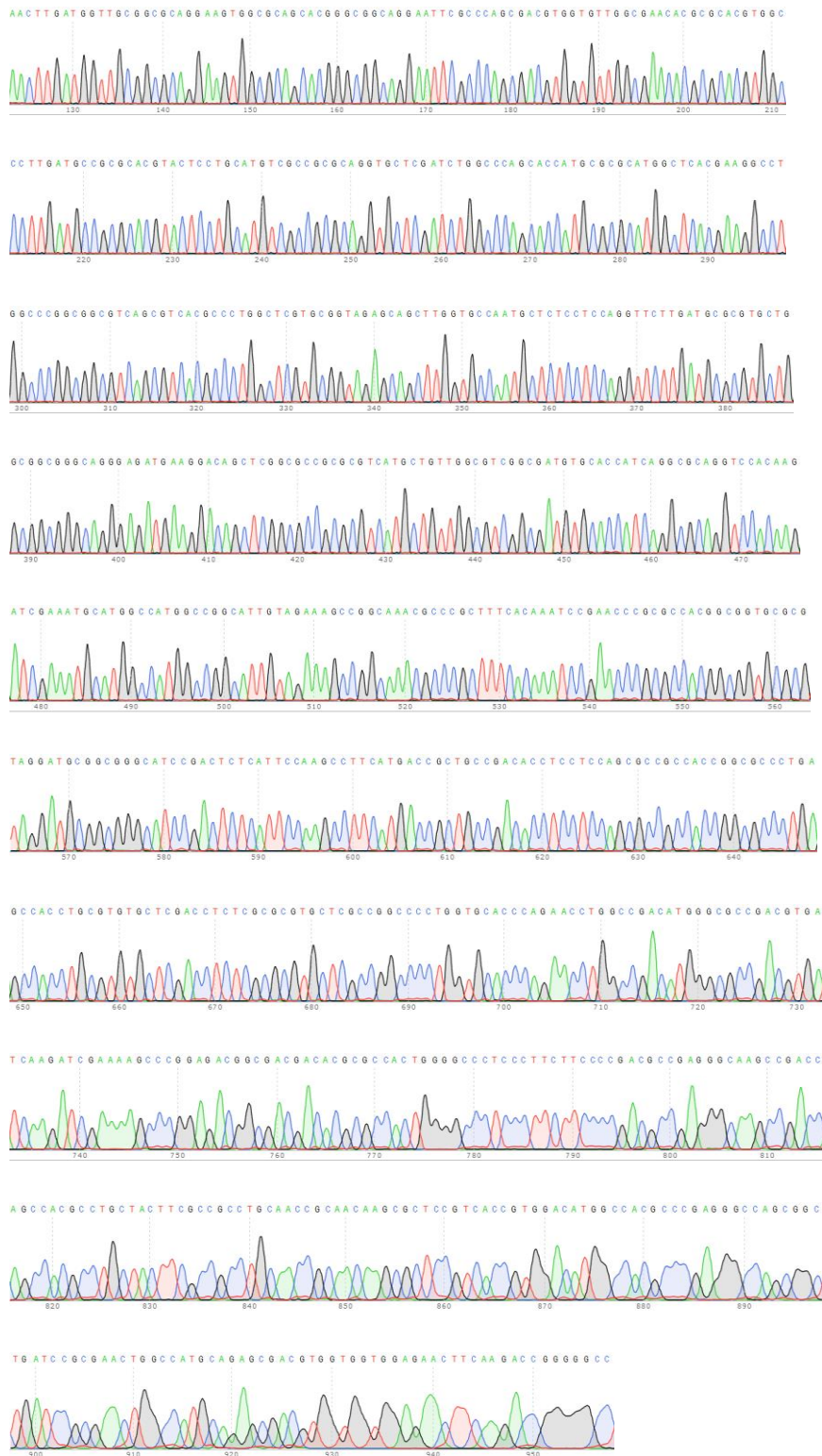


Figura A9. Electroferograma del GAP4-2, clona 1 secuenciado con el iniciador m13/pUC REVERSE.

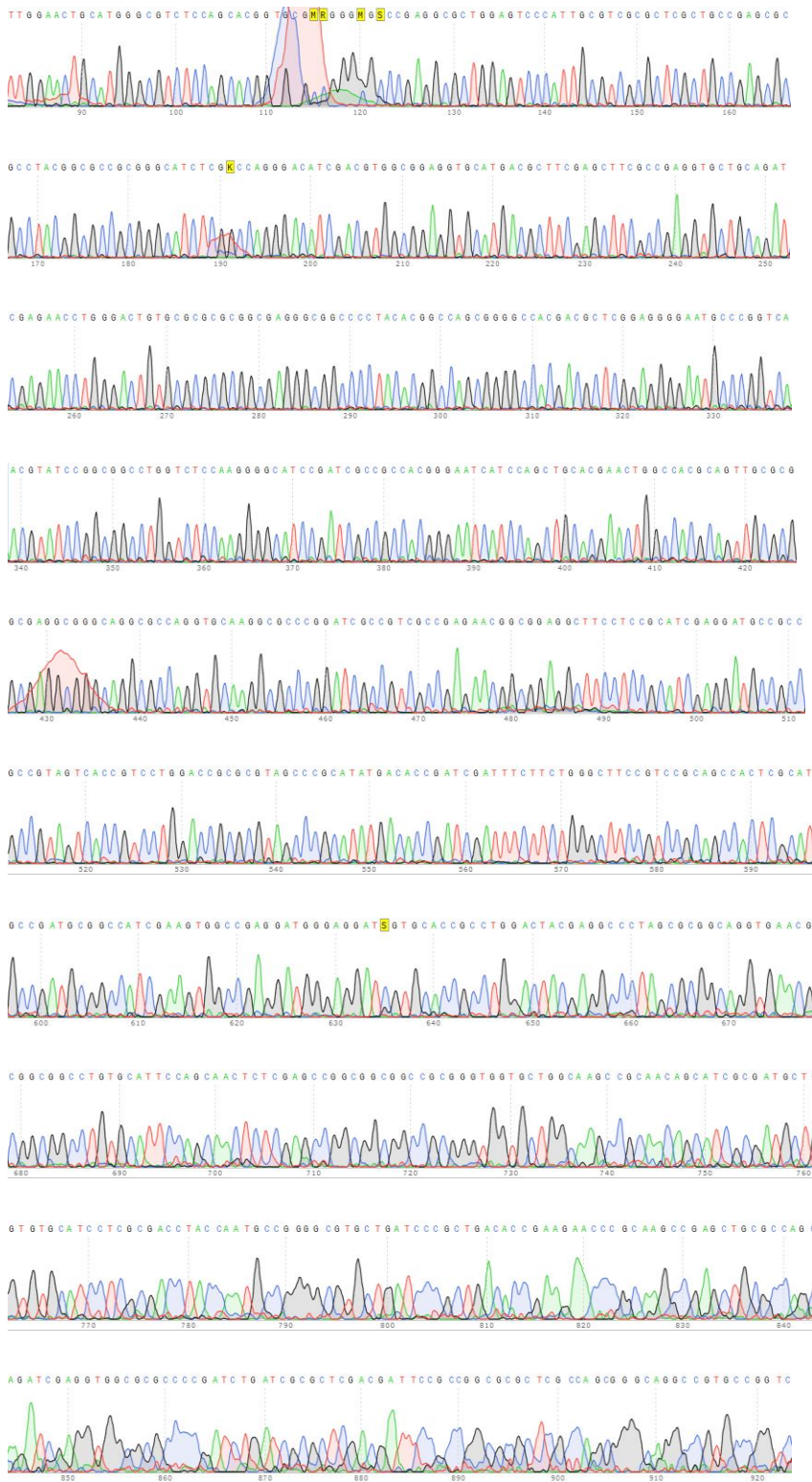


Figura A10. Electroferograma del GAP4-2, clona 2 secuenciado con el iniciador m13/pUC -40 FORWARD.

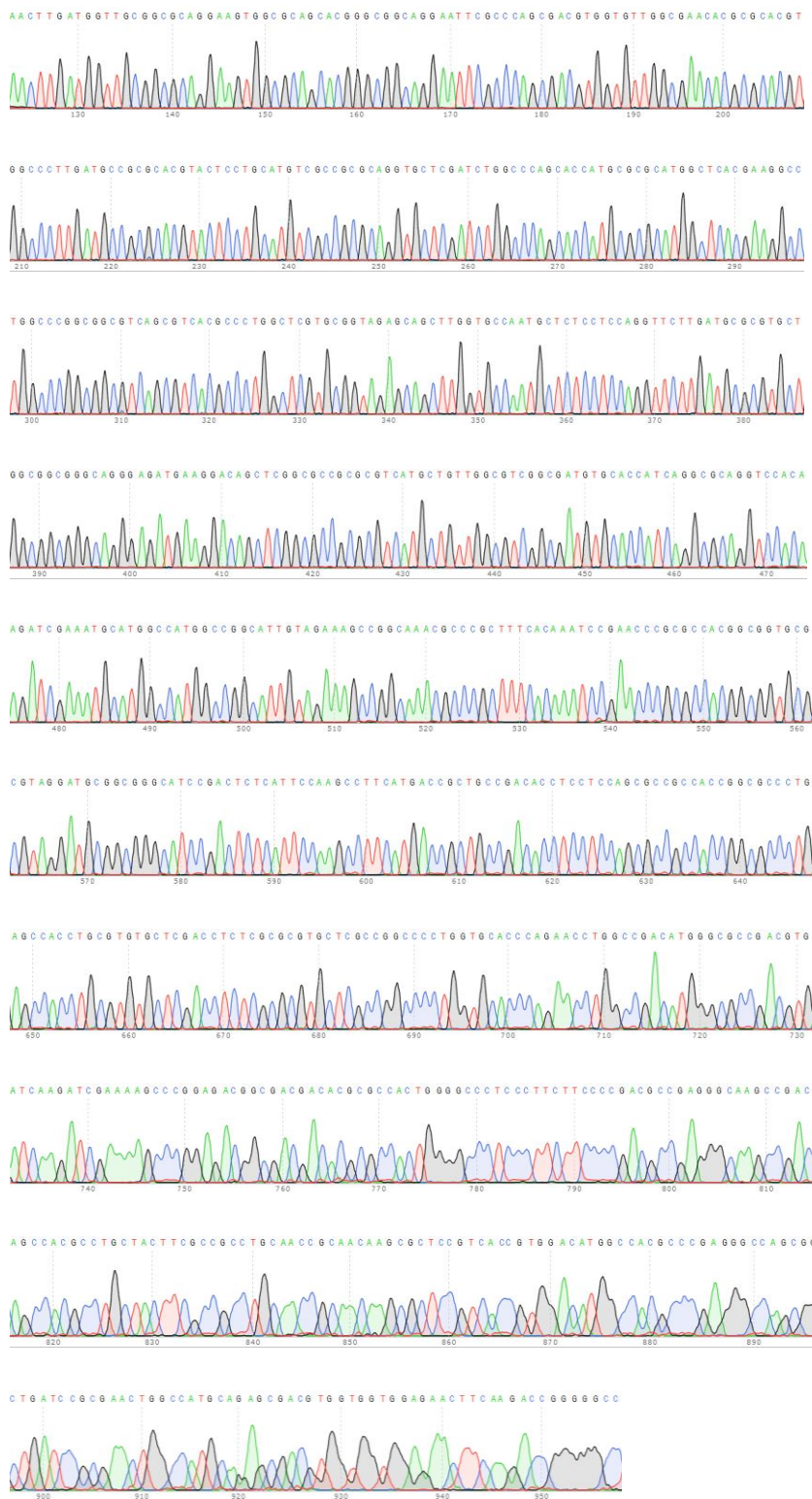


Figura A11. Electroferograma del GAP4-2, clona 2 secuenciado con el iniciador m13/pUC REVERSE.

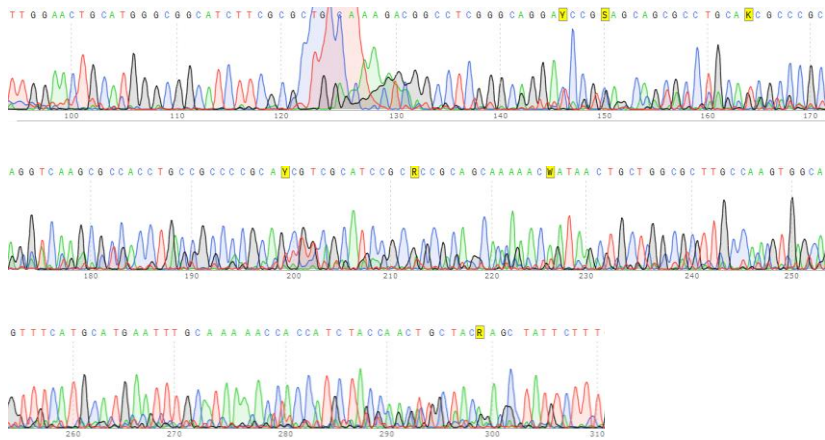


Figura A12. Electroferograma del GAP4-2, clona 3 secuenciado con el iniciador m13/pUC -40 FORWARD.

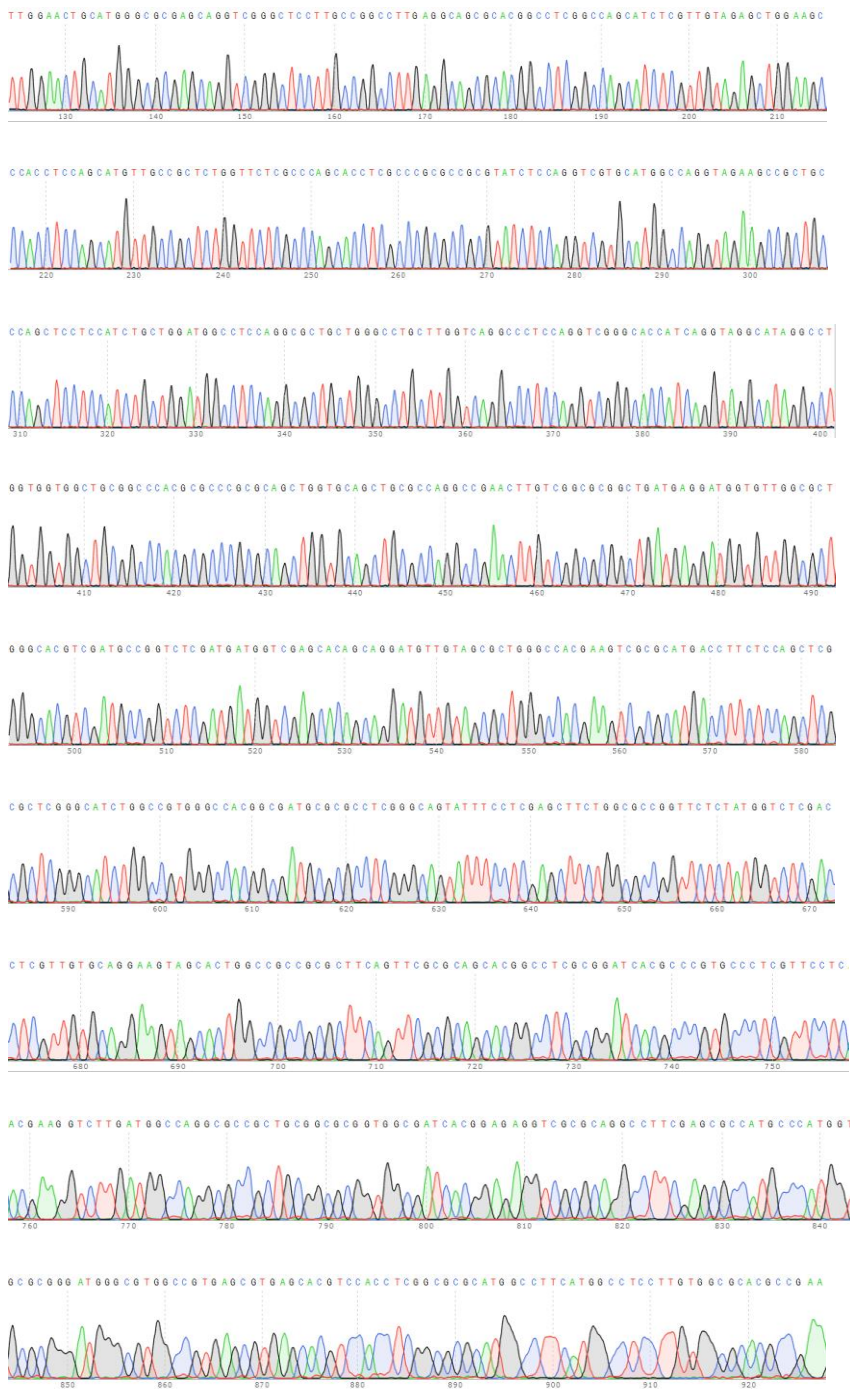


Figura A13. Electroferograma del GAP4-2, clona 3 secuenciado con el iniciador m13/pUC REVERSE.

10. REFERENCIAS

Adav S.S., Lee D.J., Lai J.Y. 2010. Microbial community of acetate utilizing denitrifiers in aerobic granules. *Appl Microbiol Biotechnol.* 85:753-762.

Ausubel F.M., R. Brent, R. Kingston, D. Moore, J.G. Seidman, J.A. Smith y K. Struhl. 1994. *Current protocols in molecular biology.* New York. John Wiley & Sons. 12-19.

Bai N., Abuduaini R., Wang S., Zhang M., Zhu X., Zhao Y. 2017. Nonylphenol biodegradation characterizations and bacterial composition analysis of an effective consortium NP-M2. *Environ Pollut.* 220:95-104.

Baker M. 2012. *De novo* genome assembly: what every biologist should know. *Nat Meth.* 9:333-337.

Bautista-Alcántara, A. 2015. Amplificación por PCR y secuenciación de regiones del genoma de *Alicyclophilus* sp. BQ1 que no pudieron ser identificadas por secuenciación masiva. Residencia profesional. Departamento de Bioquímica, Facultad de Química, UNAM.

Bej A.K. 1994. Gerhardt P. *Polymerase Chain Reaction in Methods for general and molecular bacteriology.* Second Edition. ASM Press. Washington, D.C. Pp 418-435.

Bucur D. 2017. A stochastic *de novo* assembly algorithm for viral-sized genomes obtains correct genomes and builds consensus. *Inf Sci.* 420:184-199.

Center for the Polyurethanes Industry. 2007. Polyurethane: history. Center for the Polyurethanes Industry, Arlington, VA. http://www.polyurethane.org/s_api/sec.asp?CID_853&DID_3487. Consultado el 20 de febrero de 2018.

Domínguez Cárdenas, M. 2013. Clonación del gen de una porina esterasa de membrana de *Alicyclophilus* sp. BQ1 en un vector de expresión, para la producción de proteína recombinante. Tesis licenciatura QA, Facultad de Química, UNAM.

Dullius C. H., Chen C. Y., Schink B. 2011. Nitrate-dependent degradation of acetone by *Alicyclophilus* and *Paracoccus* strains and comparison of acetone carboxylase enzymes. *Appl Environ Microbiol.* 77:6821-6825.

Fuentes-Jaime, J. 2015. Caracterización de una esterasa de *Alicyclophilus* sp. BQ1 inducida por N-metil-2-pirrolidona y análisis *in silico* de su interacción con posibles substratos. Tesis de Maestría. Facultad de Química, UNAM.

Gautam R., Bassi A.S., Yanful E.K. 2007. A review of biodegradation of synthetic plastic and foams. *Appl Microbiol Biotechnol.* 141:85-108.

Hong X., Zhang X., Liu B., Mao Y., Liu Y., Zhao L. 2010. Structural differentiation of bacterial communities in indole-degrading bioreactors under denitrifying and sulfate-reducing conditions. *Res Microbiol.* 161:687-693.

Howard G. T., Blake R.C. 1998. Growth of *Pseudomonas fluorescens* on a polyester-polyurethane and the purification and characterization of a polyurethanase-protease enzyme. *Int Biodeterior Biodegrad.* 42:213–220.

Ignatyev I.A., Thielemans W., Vander Beke B. 2014. Recycling of polymers: a review. *ChemSusChem.* 7:1579-1593.

Illumina, Illumina sequencing technology, Illumina Inc., San Diego, 2016.

Instituto Nacional de Estadística y Geografía. 2016. Anuario estadístico de los Estados Unidos Mexicanos 2016. Instituto Nacional de Estadística y Geografía, México.

Kay M.J., Morton L.H.G., Prince E.L. 1991. Bacterial Degradation of Polyester Polyurethane. *Int Biodeterior.* 27:205-222.

Lee D.G., Cho K.C., Chu K.H. 2014. Identification of triclosan-degrading bacteria in a triclosan enrichment culture using stable isotope probing. *Biodegradation.* 25:55-65.

Liu L., Li Y., Li S., Hu N., He Y., Pong R., Li D., Lu L., Law M. 2012. Comparison of Next-Generation Sequencing Systems. *J Biomed Biotechnol.* 2012:1-11.

Mechichi T., Stackebrandt E., Fuchs G. 2003. *Alicyclophilus denitrificans* gen. nov., sp. nov., a cyclohexanol-degrading, nitrate-reducing β -proteobacterium. Int J Syst Evol Microbiol. 53:147-152.

Morohoshi T., Okutsu N., Xie X., Ikeda, T. 2016. Identification of quorum-sensing signal molecules and a biosynthetic gene in *Alicyclophilus* sp. isolated from activated sludge. Sensors. 16:1218.

Nakajima-Kambe T., Onuma F., Akutsu Y., Nakahara T. 1997. Determination of the polyester polyurethane breakdown products and distribution of the polyurethane degrading enzyme of *Comamonas acidovorans* strain TB-35. J Ferment Bioeng 83:456–460.

Nakajima-Kambe T., Onuma F., Kimpara N., Nakahara T. 1995. Isolation and characterization of a bacterium which utilizes polyester polyurethane as a sole carbon and nitrogen source. FEMS Microbiol Lett. 129:39–42.

Nakajima-Kambe T., Shigeno-Akutsu Y., Nomura N., Onuma F., Nakahara T. 1999. Microbial degradation of polyurethane, polyester polyurethanes and polyether polyurethanes. Appl Microbiol Biotechnol. 51:134–140.

NCBI. 2018. GenBank Overview. <https://www.ncbi.nlm.nih.gov/genbank/>. Consultado el 3 de marzo del 2018.

Oceguera-Cervantes, A. 2005. Aislamiento de bacterias capaces de crecer en un medio con poliuretano (Hydroform) como única fuente de carbono y caracterización de las actividades enzimáticas relacionadas con su degradación. Tesis de licenciatura. Facultad de Química, UNAM.

Oceguera-Cervantes A., Carrillo-García A., López N., Bolaños-Nuñez S., Cruz-Gómez M.J., Wachter C., Loza-Tavera H. 2007. Characterization of the polyurethanolytic activity of two *Alicyclophilus* sp. strains able to degrade polyurethane and N-Methylpyrrolidone. Appl Environ Microbiol. 73:6214-6223.

Okutsu N., Morohoshi T., Ikeda T. 2015. Draft genome sequence of *Alicyclophilus* sp. B1, an N-Acylhomoserine lactone-producing bacterium, isolated from activated sludge. *Genome Announc* 3. e004:24-15.

Promega, pGEM®-T and pGEM®-T Easy Vector Systems, Promega Corporation, USA, 2015.

Raiha K., Ukkonen E. 1981. The shortest common supersequence problem over binary alphabet is NP-complete. *Theor Comput Sci.* 16: 187–198.

Rowe L., Howard G.T. 2002. Growth of *Bacillus subtilis* on polyurethane and the purification and characterization of a polyurethanase-lipase enzyme. *Int Biodeterior Biodegrad.* 50:33–40.

Savín-Gámez, A. 2016. Identificación de miembros bacterianos de consorcios capaces de crecer en un barniz de poliéter poliuretano. Tesis de Maestría. Facultad de Química, UNAM.

Solís-González, J. 2011. Identificación y clonación de algunos de los genes de *Alicyclophilus* sp. relacionados con su capacidad para utilizar poliuretano como única fuente de carbono. Tesis de Maestría. Facultad de Química, UNAM.

Taşkan B., Hanay Ö., Taşkan E., Erdem M., Hasar H. 2016. Hydrogen-based membrane biofilm reactor for tetracycline removal: biodegradation, transformation products, and microbial community. *Environ Sci Pollut Res.* 23:21703-21711.

The Darling lab. 2018. mauve multiple genome alignment. <http://darlinglab.org/mauve/mauve.html>. Consultado el 14 de junio de 2018.

Thermo Scientific, *Taq* DNA Polymerase (recombinant), Thermo Fisher Scientific Inc., 2015.

Thermo Scientific, Thermo Scientific Phusion High-Fidelity DNA Polymerase, Thermo Fisher Scientific Inc., 2016.

Weelink S.A., Tan N.C., ten Broeke H., van den Kieboom C., van Doesburg W., Langenhoff A.A., Gerritse J., Junca H., Stams A.J. 2008. Isolation and characterization of *Alicyclophilus denitrificans* strain BC, which grows on benzene with chlorate as the electron acceptor. *Appl Environ Microbiol.* 74:6672-6681.

Zhang X., Hua X., Yue X. 2017. Comparison of bacterial community characteristics between complete and shortcut denitrification systems for quinoline degradation. *Appl Microbiol Biotechnol.* 101:1697-1707.