UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO

FACULTAD DE CIENCIAS

ESCUELA PROFESIONAL DE BIOLOGÍA

CARACTERIZACIÓN *IN SILICO* Y CLONACIÓN DEL GEN ASPARTIL AMINOPEPTIDASA A PARTIR DE DATOS NGS (NEXT GENERATION SEQUENCING) DEL HALÓFILO MODERADO *Chromohalobacter salexigens* MP25462.

Tesis presentada por:

Bach. Kevin Brandon Pacheco Capcha.

Para optar al Título Profesional de Biólogo.

Asesora:

Dra. María Antonieta Quispe Ricalde.

Co-asesor:

Blgo. José Luis Sierra Herrera.

Tesis financiada mediante resolución R- N° 1349-2018-UNSAAC

CUSCO – PERÚ

2021

La realización de la presente Tesis ha sido posible gracias a la concesión de la subvención al Proyecto de Investigación "Inhibición de proteasas de *Leishmania* a través de péptidos obtenidos de proteínas de *Lupinus mutabilis* por acción enzimática de proteasas de bacterias halófilas: Nuevas perspectivas de búsqueda de moléculas antileishmanicidas" por la Universidad Nacional de San Antonio Abad del Cusco-UNSAAC bajo el contrato Nº 023-2018-UNSAAC.

Lo experimentado en el desarrollo de esta tesis no lo podría catalogar como algo fácil, pero con el amor de mis semejantes más cercanos hizo que disfrutara de cada proceso en este camino. Es por ello que agradezco primeramente a Dios por darme la oportunidad de haber conocido a aquellas personas que me apoyaron en cada decisión y proyecto.

A mis amados padres Luciano y Victoria, mis principales forjadores de sueños y metas, a ustedes les digo que su soporte y esfuerzo impulsaron cada paso en mí, su comprensión y dedicación lograron la realización de esta meta. A mis hermanos Mazhiun, Yuber y Katerin por darme la motivación madura, que sus palabras de aliento llegaron en los momentos más indicados.

A mis sobrinos Randal, Zucel y Zoe que fueron determinantes para seguir adelante. A mis queridos abuelos que están a mi lado y mis pachos que ya no lo están, agradecerles por ese cariño incondicional que me entregaron.

AGRADECIMIENTOS

Ante todo expresarle mi sincero agradecimiento a mi mentora y asesora Dra María Antonieta Quispe Ricalde, por su apoyo y confianza depositada en mí para la realización de este trabajo de investigación.

Al Blgo. José Luis Sierra Herrera por sus observaciones puntuales y soporte a lo largo de este trabajo.

A los docentes de la Escuela Profesional de Biología de la Universidad Nacional de San Antonio Abad del Cusco que contribuyeron en mi formación académica y personal.

A la Dra. Pilar Foronda Rodríguez y al Dr. Enrique Martínez Carretero por poner a disposición los laboratorios del Instituto Universitario de Enfermedades Tropicales y Salud Publica de Canarias (IUETSPC) de la Universidad de la Laguna (ULL). De igual modo agradecer al Dr. José Antonio Pérez Pérez, Dr. José Manuel Gonzáles Hernández, Mgt. Hugo Gilgardo Castelán Sánchez; que me apoyaron desinteresadamente hasta el término de esta investigación

Al proyecto de investigación "Inhibición de proteasas de Leishmania a través de péptidos obtenidos de proteínas de Lupinus mutabilis por acción enzimática de proteasas de bacterias halófilas: Nuevas perspectivas de búsqueda de moléculas antileishmanicidas"

A mis amigos Marco, Isabel, Lucero, David, Shirley y Yesenia y en especial a mi compañera de trabajo Ilucion.

A todos ustedes muchas gracias.

ÍNDICE

Pág.

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

ÍNDICE DE ANEXOS

ABREVIATURAS

ABREVIATURA	SIGNIFICADO
ADN	**Ácido desoxirribonucleico**
ARN	**Ácido ribonucleico**
LB	**Luria Bertani**
PBS	**Tampón fosfato salino**
TAE	**Tris, acetato y EDTA**
EDTA	**Ácido etilendiaminotetraacético**
SW	**Medio Sea Water**
g.p.m.	**Gravedad por minuto**
PCR	**Reacción en cadena de la polimerasa**
UV	**Ultravioleta**
NGS	**Secuenciación de segunda generación**
DAP	**Aspartil aminopeptidasa**
Kda	**Kilidaltons**

GLOSARIO

PALABRA	SIGNIFICADO
Deduplicación	Compresión de datos para eliminar copias duplicadas de datos repetidos.
Asignación N	Sustitución de un nucleótido por carecer de confianza de asignación

RESUMEN.

Las enzimas proteolíticas son biotecnológicamente catalizadores muy importantes, es por ello que su búsqueda y obtención deben realizarse de manera rápida y eficaz. Es así que las nuevas tecnologías de secuenciación masiva tienen la capacidad de otorgar toda la información genómica de un organismo para la obtención de un gran número de genes de interés. El presente estudio se enfocó en caracterización *in silico* y clonación del gen aspartil aminopeptidasa a partir de datos de secuenciación de próxima generación (NGS) del halófilo moderado *Chromohalobacter salexigens* MP25462.

El aislamiento de *C. salexigens* MP25462 y secuenciación NGS de su material genético se realizó dentro del proyecto "Secuenciación del metagenoma de ambientes salinos del departamento de Cusco". Los datos genómicos obtenidos de secuenciación masiva atravesaron por un control de calidad mediante el programa FastQC. El ordenamiento del genoma se realizó con los ensambladores genómicos de *novo* Velvet, A5 y Spades; y la anotación del genoma ensamblado con el servidor RAST. El modelado de la proteína DAP a partir del gen aspartil aminopeptidasa se efectuó con el servidor en línea I-TASSER y se visualizó con el programa Chimera. El diseño de cebadores se ejecutó con el programa *Gene Runner* para la amplificación, mediante PCR (*Polimerase Chain Reaction*) del gen completo aspartil aminopeptidasa. La clonación del gen se llevó a cabo en el vector pGEM-T utilizando células JM109 competentes para finalmente criopreservarlas.

La anotación dio como resultado 3402 genes codificantes de los cuales 263 implicados en el metabolismo de proteínas. La secuencia nucleotídica del gen aspartil aminopeptidasa presentó una cobertura de 20.7 X y una longitud de 1299 pares de bases traducidas a una secuencia de 432 aminoácidos, dando por resultado su estructura tridimensional similar a la aspartil aminopeptidasa de *Homo sapiens* (DNPEP) y su funcionalidad idéntica a la proteína de la arquea *Thaumarchaeota archaeon* (TET) ubicando su sitio activo en el aminoácido 263 dependiendo de dos iones de cobalto para su actividad. La estandarización de PCR fue a 48, 51 y 55° C amplificando una sola banda de 1283 pb. en las tres temperaturas. La clonación del gen aspartil aminopeptidasa de MP25462 se consiguió en dos colonias JM109: DAP-T2 y DAP-T7, logrando validar el proceso bioinformático y la existencia del gen aspartil aminopeptidasa en MP25462.

Palabras clave: Aspartil aminopeptidasa, bioinformática, ensamble, secuenciación de próxima generación, clonación.

ABSTRACT.

Proteolytic enzymes are biotechnologically very important catalysts, which is why their search and obtaining must be carried out quickly and efficiently. Thus, new massive sequencing technologies have the ability to provide all the genomic information of an organism to obtain a large number of genes of interest. The present study focused on in silico characterization and cloning of the aspartyl aminopeptidase gene from next generation sequencing (NGS) data of the moderate halophile Chromohalobacter salexigens MP25462.

The isolation of C. salexigens MP25462 and NGS sequencing of its genetic material was carried out within the project "Sequencing of the metagenome of saline environments of the department of Cusco". The genomic data obtained from massive sequencing underwent a quality control using the FastQC program. Genome ordering was performed with the de novo Velvet, A5 and Spades genomic assemblers; and annotation of the assembled genome with the RAST server. The modeling of the DAP protein from the aspartyl aminopeptidase gene was carried out with the I-TASSER online server and visualized with the Chimera program. The primer design was executed with the Gene Runner program for the amplification, by PCR (Polimerase Chain Reaction) of the complete aspartyl aminopeptidase gene. The cloning of the gene was carried out in the pGEM-T vector using competent JM109 cells to finally cryopreserve them.

The annotation resulted in 3,402 coding genes of which 263 involved in protein metabolism. The nucleotide sequence of the aspartyl aminopeptidase gene presented a coverage of 20.7 X and a length of 1299 base pairs translated into a sequence of 432 amino acids, resulting in its three-dimensional structure similar to Homo sapiens aspartyl aminopeptidase (DNPEP) and its identical functionality. to the protein of the archaea Thaumarchaeota archaeon (TET) locating its active site at amino acid 263 depending on two cobalt ions for its activity. The PCR standardization was at 48, 51 and 55 ° C amplifying a single band of 1283 bp. in all three temperatures. The cloning of the MP25462 aspartyl aminopeptidase gene was achieved in two JM109 colonies: DAP-T2 and DAP-T7, achieving validation of the bioinformatic process and the existence of the aspartyl aminopeptidase gene in MP25462.

Keywords: Aspartyl aminopeptidase, bioinformatics, assembly, next generation sequencing, cloning.

INTRODUCCIÓN.

A nivel mundial, Perú es uno de los países con mayor diversidad de ecosistemas (Zamora, 1996). Entre ellos tenemos a los ecosistemas extremófilos hipersalinos que variando en factores de salinidad, pH, composición iónica, radiación UV, temperatura y nutrientes; presentan una gran diversidad y dinámica microbiana. Es así, que las adaptaciones de estos microrganismos extremófilos halófilos representan un gran potencial biotecnológico que pueden ser explotados por la industria, siendo en este contexto las proteasas uno de los productos más codiciados de estos microorganismos.

Actualmente un modo eficaz para adquirir toda la información y cualidades de estos microorganismos es mediante la técnica de secuenciación, concretamente la secuenciación de segunda generación o *Next Generation Sequencing* (NGS), que engloba a todas las tecnologías presentes en plataformas como 454 de Roche, Solid, Ion Torrent e Illumina que están destinadas a llevar a cabo la secuenciación masiva a gran escala de cualquier ácido nucleico de un genoma o metagenoma (Jauk, 2019); por tanto, el enfoque más adecuado para interpretar el estudio de un genoma completo es a través de la bioinformática que es una herramienta para el estudio en biología que se vale del uso de programas para procesar miles de mega bases de secuencias de ADN (Garrigues, 2017).

Por lo que se refiere a la tecnología del ADN recombinante, esta permite identificar y aislar un solo gen o segmento de ADN de interés de un genoma. Mediante la clonación, se pueden producir grandes cantidades de copias idénticas de una molécula de ADN específica. Estas copias idénticas, o clones, pueden manipularse para numerosos propósitos, entre ellos, realizar investigaciones sobre la estructura y organización del ADN, estudiar la expresión de genes, estudiar productos de proteínas para comprender su estructura y función, y producir productos comerciales importantes a partir de la proteína codificada por un gen (Klug *et al*., 2016).

Es así, que el trabajo a continuación esta orientado hacia la caracterización *in silico* y clonación del gen aspartil aminopeptidasa a partir de los datos de secuenciación de próxima generación (NGS) del halófilo moderado *Chromohalobacter salexigens* MP25462. Tales datos atravesaran por múltiples programas bioinformáticos para la verificación de calidad, ensamblaje y anotación, descripción y predicción de funcionalidad del gen aspartil aminopeptidasa. Se amplificará la secuencia de la aspartil aminopeptidasa mediante PCR (*Polymerase Chain Reaction*) con cebadores diseñados a partir de los datos bioinformáticos obtenidos. Ya aislado el gen aspartil aminopeptidasa se procederá a la clonación del mismo.

PLANTEAMIENTO DEL PROBLEMA.

Con el avance de la tecnología NGS en la presente década, se tuvo un incremento en las investigaciones que reportan la conformación genómica de una gran diversidad de organismos y microorganismos a nivel mundial. Tal información que ofrece el uso de esta técnica es enorme comparado con el tiempo que toma su aplicación; siendo el trabajo analítico de estos datos cada vez más complejos y extenuantes, por ende esta tecnología se apoya de las capacidades flexibles y sistemáticas de la bioinformática. De ahí que muchos de los procedimientos en el análisis de genomas son críticos, en particular el ensamblaje genómico ya que de él depende el ordenamiento correcto de las secuencias; de modo que se tiene el apremio de utilizar más de un ensamblador de *novo* evaluando sus indicadores de calidad y en consecuencia obtener una anotación verídica del gen o genes de interés para realizar una correcta descripción de estos.

El gen que expresa la aspartil aminopeptidasa se caracterizó en varios organismos tales como mamíferos, plantas, levaduras, parásitos y bacterias como *Pseudomonas aureginosa* de la cual no se tiene datos exactos de su finalidad metabólica (Natarajan & Mathews, 2012) teniendo actividad en presencia de iones cobalto y zinc siendo clonado en células competentes *Escherichia coli* (Wilk *et al*., 2002).

Por otra parte, el gen de la proteasa aspartil aminopeptidasa se ha implicado en la función específica de convertir la angiotensina II en la angiotensina vasoactiva III dentro del cerebro en los seres humanos. Hoy en día, se ha incrementado el empleo de las bacterias extremófilas productoras de proteasas en la síntesis de péptidos bioactivos para su aplicación en muchas áreas. Sin embargo, el gen de la proteasa aspartil aminopeptidasa no se ha descrito hasta el momento en microorganismos halófilos, por lo cual se evidencia la necesidad de datos al respecto (Story *et al*., 2005).

El aislamiento y secuenciación genómica del halófilo moderado *Chromohalobacter salexigens* MP25462 del salar de Maras, Urubamba ha sido realizado; por lo que se plantea efectuar el correcto análisis bioinformático de su genoma, descripción del gen que codifica la proteasa aspartil aminopeptidasa y corroborar esta etapa de caracterización teórica mediante procedimientos experimentales de biología molecular al clonar el gen aspartil aminopeptidasa MP25462 en el vector pGEM-T. Teniendo estos antecedentes es imperativo realizar la caracterización y clonación del gen aspartil aminopeptidasa. Por tanto surge la interrogante: ¿Mediante el uso de programas bioinformáticos será posible caracterizar *in silico* el gen aspartil aminopeptidasa a partir de datos NGS (*Next Generation Sequencing*) del halófilo moderado *C. salexigens* MP25462 y con esta información clonar el gen aspartil aminopeptidasa?

JUSTIFICACIÓN.

Debido a la gran demanda de proteasas en diversas industrias se requiere procedimientos versátiles, rápidos y económicos tal como es la secuenciación que permite ubicar, identificar y aislar genes codificadores de proteasas, que mediante técnicas bioquímicas habituales o por sustratos específicos pasarían desapercibidas. Así también, por su interés biotecnológico, las proteasas se han convertido en enzimas industrialmente importantes. Por lo tanto, la obtención de enzimas puras que sean estables y activas bajo múltiples condiciones extremas (pH alcalino, altas concentraciones de sal y amplia temperatura) es científica e industrialmente significativa, tal potencial de las proteasas de los halófilos aislados de ambientes extremos ha sido revisado previamente (Gomes & Steiner, 2004). Las bacterias que producen proteasas son diversas así tenemos a *Bacillus iranensis* una bacteria moderadamente halófila que produce una proteasa haloalcalina extracelular (Ghafoori *et al*., 2016); *B. circulans* que produce la serín-proteasa alcalina estable a solventes orgánicos (Sari *et al*., 2015); así también *Aeribacillus pallidus* que produce una proteasa alcalina termoestable, que tiene un potencial con un posible aditivo para detergentes (Yildirim *et al*., 2017). Otra bacteria es *Geobacillus toebii* que metaboliza una proteasa termoestable, haloalcalina (Thebti *et al*., 2016). Además que el uso de tecnologías de clonación como medio para obtener proteínas recombinantes otorga una ventaja sobre otros métodos tradicionales. Por esta razón se justifica que hay la necesidad de obtener las proteasas de bacterias extremófilas, en particular las de tipo halófilo, ya que se han convertido en el centro de interés para realizar investigaciones por la importancia biológica y utilidades que pueden ofrecer en el ámbito industrial y biotecnológico.

OBJETIVOS.

OBJETIVO GENERAL.

Caracterizar *in silico* y clonar el gen aspartil aminopeptidasa a partir de datos NGS (*Next Generation Sequencing*) del halófilo moderado *Chromohalobacter salexigens* MP25462.

OBJETIVOS ESPECÍFICOS.

- ✓ Evaluar los datos de ensambles de *novo* del genoma de *Chromohalobacter salexigens* MP25462.
- ✓ Describir *in silico* el gen aspartil aminopeptidasa.
- ✓ Comparar el gen aspartil aminopeptidasa de *C. salexigens* MP25462 con otros genes similares presentes en diferentes organismos.
- ✓ Diseñar y sintetizar cebadores para la amplificación del gen completo aspartil aminopeptidasa de *C. salexigens* MP25462.
- ✓ Amplificar y clonar el gen aspartil aminopeptidasa en un vector de clonación.

HIPÓTESIS.

Es posible caracterizar *in silico* y clonar el gen aspartil aminopeptidasa a partir de datos NGS (*Next Generation Secuencing*) del halófilo moderado *C. salexigens* MP25462.

CAPITULO I
MARCO TEÓRICO Y CONCEPTUAL

1.1. Antecedentes del estudio.

1.1.1. Estudios a nivel internacional.

Wilk *et al.* (1998), en EE.UU. identificaron por primera vez la presencia metaloproteasa de la familia M18 en mamíferos. Realizaron sus trabajos en muestras de tejido cerebral de conejo y en clones con material de ratón y humano. Purificaron del citosol de cerebro de conejo una aminopeptidasa con una preferencia por los residuos de aspartilo N-terminal y glutamilo, pero distinta de la glutamil aminopeptidasa (EC 3.4.11.7). La enzima nativa tiene una masa de 440 kDa y migra en una sola banda de 55 kDa después de la electroforesis en gel de SDS-poliacrilamida. Identificaron y secuenciaron los clones humanos y de ratón descritos como "similares a una aminopeptidasa vacuolar de levadura" y que contenían ADNc de longitud completa. El ADNc humano se expresó en *Escherichia coli*.

Wilk *et al.* (2002), en EE.UU. identificaron 8 secuencias eucariotas altamente homólogas probables de aspartil aminopeptidasas (DAP) que es un miembro de la familia M18 del clan MH de metalopeptidasas cocatalíticas. Las enzimas humanas y de ratón fueron clonadas. Ocho residuos de histidina de la DAP humana se mutaron secuencialmente a fenilalanina. La mutación de His94, His170 y His440 abolió la actividad enzimática. His94 y His440 se postulan como aminoácidos que participan en la unión de los átomos de zinc cocatalítico por homología con otros miembros del clan MH. La mutación de His352 redujo drásticamente la actividad enzimática. Estos estudios revelaron el papel importante de los residuos de histidina tanto en la catálisis como en la integridad estructural de la DAP.

Story *et al.* (2005), en EE.UU. descubrieron que los extractos celulares de la proteína hipertermofílica de *Pyrococcus furiosus* contienen una alta actividad específica (11 U / mg) de lisina aminopeptidasa (KAP). KAP es un homotetramer (38.2 kDa por subunidad) y en cuanto a su actividad esta fue estimulada cuatro veces por la adición de Co^{2+} (0.2 mM). La actividad KAP óptima con Lys-pNA como sustrato ocurrió a pH 8.0 y a una temperatura de 100 ° C. Además identificaron utilizando base de datos del genoma de *P. furiosus*, al gen (PF1861) que codifica un producto correspondiente a 346 aminoácidos basados en este dato, KAP es parte de la familia de peptidasas M18.

Teuscher *et al.* (2007), en Australia identificaron un miembro de la familia M18 de aspartil aminopeptidasas que se expresa en todos los estadios intra-eritrocíticos del parásito de

la malaria humana *Plasmodium falciparum* (PfM18AAP), con los niveles más altos de expresión en la fase de anillos. La enzima recombinante funcionalmente activa, rPfM18AAP y la enzima nativa en extractos citosólicos de parásitos de la malaria son octómeros de 560 kDa que exhiben una actividad óptima a pH neutro y requieren la presencia de iones metálicos para mantener la actividad y estabilidad enzimática. Al igual que la aspartil aminopeptidasa humana, la actividad exopeptidasa de PfM18AAP es exclusiva de los aminoácidos ácidos glutamato y aspartato, lo que hace que esta enzima sea de particular interés y sugiere que puede funcionar junto con la aminopeptidasa neutra citosólica de la malaria en la liberación de aminoácidos.

Kusumoto *et al*. (2008), en Japón sobreexpresaron la DAP (aspartil aminopeptidasa) de *Aspergillus oryzae* bajo un promotor del gen de taka-amilasa, con un conector de cola His. La enzima se extrajo del micelio y se purificó con cromatografía de absorción de ión níquel inmovilizado usando un tampón con ión cobalto e imidazol. La fracción activa se purificó adicionalmente con cromatografía de filtración en gel. La enzima electroforéticamente pura resultante mostró una masa molecular de 520 kDa. La DAP recombinante producida usando un ión cobalto en medios de cultivo de *A. oryzae* y el proceso de purificación permitieron un alto rendimiento de la actividad enzimática. Esta enzima mostró una alta reactividad hacia el sustrato peptídico en lugar de sustratos sintéticos.

Yuga *et al.* (2011), en EE.UU. describieron que la aspartil aminopeptidasa en *Saccharomyces cerevisiae* (Yhr113w / Ape4) que es el tercera elemento del proceso de focalización de citoplasma a vacuola (Cvt), que es similar en estructura primaria y organización de subunidades a la aminopeptidasa I vacuolar (Ape1). Ape4 no tiene propéptido y no se autoensambla en agregados. Sin embargo, se une a Atg19 en un sitio distinto de los sitios de unión de Ape1 y α-manosidasa (Ams1), lo que le permite transportar Ape1en el sistema vacuolar. En condiciones de crecimiento, una pequeña porción de Ape4 se localiza en la vacuola, pero su transporte vacuolar se acelera por la falta de nutrientes, y reside de manera estable en la luz de la vacuola. Es así que propusieron que el Ape4 citosólico se redistribuye a la vacuola cuando las células de levadura necesitan una degradación vacuolar más activa.

Natarajan & Mathews (2012), en Korea estudiaron la aspartil aminopeptidasa de *Pseudomonas aeruginosa* (PaAAP), que está codificada por el gen apeB, que expresaron en *Escherichia coli*, la purificaron y cristalizaron utilizando el método de microbatch. Realizaron un estudio estructural preliminar utilizando el método cristalográfico de rayos X. El cristal de PaAAP difractó a una resolución de 2,0 Å y pertenecía al grupo espacial romboédrico H 3.

Chaikuad *et al*. (2012), en Inglaterra realizaron ensayos de la estructura cristalina de la aspartil aminopeptidasa humana en zinc y un análogo de sustrato aspartato-β-hidroxamato. De acuerdo con los datos de microscopía electrónica revelaron una maquinaria dodecamérica construida por dímeros de intercambio de dominios. Una comparación estructural con homólogos bacterianos identifica las características catalíticas unificadoras entre las enzimas M18 poco conocidas. Los ligandos unidos en el sitio activo también revelan el modo de coordinación del centro de zinc binuclear y una especificidad de sustrato para aminoácidos ácidos.

Nguyen *et al*. (2014), en Korea realizaron estudios cinéticos de la aspartil aminopeptidasa en *Pseudomonas aeruginosa* (PaAP) que pertenece a la familia de peptidasas M18 en el clan MH revelando que su actividad de peptidasas depende de Co^{2+} en lugar de Zn^{2+}: los valores de kcat (s^{-1}) de PaAP fueron 0.006, 5.10 y 0.43 en condiciones sin metal, Co^{2+} y Zn^{2+}, respectivamente. Consistentemente, la adición de bajas concentraciones de Co^{2+} a PaAP previamente saturadas con Zn^{2+} mejoró en gran medida la actividad enzimática, lo que sugiere que el Co^{2+} puede ser el ión metálico cocatalítico de PaAP fisiológicamente relevante. Las estructuras cristalinas de los complejos de PaAP con Co^{2+} o Zn^{2+} mostraron comúnmente dos iones metálicos en el sitio activo coordinados con tres residuos conservados y un ion bicarbonato en una geometría tetragonal. Estos resultados propusieron que las peptidasas MH del clan podrían ser peptidasas cobalto cocatalítica en lugar de una peptidasa dependiente de zinc

Park *et al.* (2017), en EE.UU. realizaron un estudio evolutivo de los dos tipos de aspartil aminopeptidasas M18 (DAP1 y DAP2) señalando que *Arabidopsis thaliana* tiene dos linajes de genes DAP (AtDAP1 y AtDAP2). Localizaron AtDAP1 en el citosol y la vacuola; mientras que AtDAP2 es un cloroplasto localizado albergando un péptido de tránsito N-terminal. Expresaron His6-DAP1 e His6-DAP2 en *Escherichia coli* los cuales fueron enzimáticamente activos y dodecaméricos con masas que excedían los 600 kDa. His6-DAP1 e His6-DAP2 hidrolizaron Asp-p-nitroanilida y Glu-p-nitroanilida. Observaron que AtDAP son metalopeptidasas altamente conservadas activadas por $MnCl_2$ e inhibidas por $ZnCl_2$ y quelantes de iones divalentes. El inhibidor de la proteasa PMSF inhibió y la ditiotreitol (DTT) estimuló tanto las actividades de His6-DAP1 como de His6-DAP2, lo que sugiere un papel para los tioles en el mecanismo catalítico del AtDAP. Las enzimas tuvieron distintos valores óptimos de pH y temperatura, así como distintos parámetros cinéticos. Usando ensayos de chaperona molecular establecidos, AtDAP1 y AtDAP2 previnieron la desnaturalización térmica.

Rout & Mahapatra (2019), en la India aplicaron múltiples enfoques computacionales como la predicción de modelos de proteínas, el estudio 3D QSAR basado en ligandos, el farmacóforo, el cribado virtual basado en estructuras y la simulación de acoplamiento molecular para la identificación de moléculas de plomo potentes contra la enzima aminopeptidasa aspártica M18 (M18AAP) presente en *Plasmodium vivax*. El modelo 3D QSAR se desarrolló utilizando compuestos bioactivos conocidos contra la proteína PvM18AAP que significan estadísticamente el modelo k-NN con q ^ 2 = 0.7654. El estudio informa sobre una molécula líder a partir de un enfoque centrado en el ligando con una buena afinidad de unión y la puntuación de acoplamiento más baja con lo que favorecerá el desarrollo de potentes moléculas antipalúdicas contra las cepas resistentes a los medicamentos del parásito de la malaria.

1.1.2. Estudios en la región Cusco.

Gárate (2016) aisló cepas bacterianas de cinco puntos de muestreo de las pozas de sedimentación salina de Maras, Cusco – Perú. Las cepas halotolerantes N9 Y N10 fueron seleccionadas por su actividad proteolítica en medios de gelatina y caseína. Las proteasas excretadas fueron caracterizadas, teniendo una actividad enzimática de 49.09 U/mL/min y 140.45 U/mL/min respectivamente. Los tamaños moleculares de las proteasas de la cepa N9 fueron de 60.5 kDa, 47.2 kDa y de la cepa N10 de 38.8 kDa aproximadamente. Al realizar ensayos de inhibición de la actividad en sustratos copolimerizados encontró que la cepa N9 presenta metaloproteasas, serín y aspártico proteasas; así también los extractos de la cepa N10 presentan metaloproteasas y serín - cisteín - aspártico proteasas.

Romoacca (2018) aisló 35 cepas bacterianas de tres ambientes salinos de Cusco, resultando 31 cepas Gram negativas y 4 Gram positivas. Así también determino la identidad de las 35 cepas mediante el análisis de las secuencias del gen ribosomal 16S donde llegó a identificar cinco géneros entre ellos *Chromohalobacter*. Evaluó cualitativamente la actividad lipolítica, amilolítica y proteolítica de las cepas en sustratos con tween 80, almidón y gelatina correspondientemente.

Huaihua (2019) purificó, caracterizó e inmovilizó proteasas excretadas de *Staphylococcus sp.* N10 de las salineras de Maras obtenidos por fermentación sumergida, pasaron por sistemas de cromatografía de afinidad, intercambio aniónico y filtración en gel. Además adapto la electroelución como método de purificación y para la inmovilización uso soportes de gelatina y poliacrilamida. La proteasa obtuvo una actividad enzimática de 29.25

U/mL/min y 28.2 U/mL/min en medios de cultivo MH y TSB. Con un peso molecular de 60 kDa y un grado de pureza de 13.56 veces más frente al extracto inicial y un rendimiento de 7.1 %, dicha proteasa se identificó como una metaloproteasa, con un pH óptimo de 7.0 y temperatura de 40° C.

Mamani *et al*. (2019) aislaron al halófilo moderado *Halomonas elongata* MH25661 de un arroyo salino de Huanoquite, en los Andes del Perú para luego realizar la secuenciación genómica de dicha bacteria en el sistema Miseq (Illumina Inc.; GeneLab del Perú SAC) resultando con 375 660 secuencias de lectura y un tamaño entre 2 a 231 pb. Así también realizaron el análisis bioinformático teniendo como ensamblador al programa A5 Miseq Linux resultando el genoma de un tamaño de 3552403 pb agrupados en 117 contigs. Realizaron la anotación funcional del genoma ensamblado revelaron genes que responden al estrés osmótico y al estrés oxidativo, genes que participan en la esporulación y genes que participan en los sistemas de transporte y resistencia a los antibióticos y compuestos tóxicos. Además entre los genes anotados se identificó una aspartil aminopeptidasa de copia única.

1.2. Marco conceptual.

1.2.1. Secuenciación de próxima generación NGS

Es una tecnología innovadora que está redefiniendo el panorama de las pruebas genéticas moleculares humanas. Permite una paralelización sin precedentes de las reacciones de secuenciación, facilitando paradigmas de prueba altamente multiplexados con un tiempo de respuesta relativamente rápido y costos reducidos. Un número creciente de los laboratorios de diagnóstico están adoptando NGS y utilizan para conducir nuevas ofertas de prueba basados en el ADN. Además, las aplicaciones de NGS para la secuenciación de ARN, el perfil epigenético mediante metilación y la secuenciación de inmunoprecipitación de cromatina, y la secuenciación microbiana ofrecen nuevas vías para las pruebas clínicas e investigación. La adopción de esta tecnología conlleva al uso de la bioinformática para el proceso e interpretación de la gran cantidad de datos generados por los instrumentos de secuenciación (Oliver *et al*., 2015).

1.2.2. Secuenciación illumina

Illumina es la tecnología de secuenciación más utilizada hoy en día debido a que ofrece plataformas efectivas como MiSeq, MiniSeq o HiSeq XTen, entre otras; capaces de ofrecer resultados competentes. Esta metodología se caracteriza por el uso de nucleótidos marcados con fluoróforos que bloquean de forma reversible la elongación de la cadena. De este modo, tras la detección de la incorporación del fluoróforo, y la eliminación del mismo, es posible continuar con un nuevo ciclo de adición de un nuevo nucleótido (Garrigues, 2017). Para la amplificación de ADN utiliza la amplificación en fase sólida. *Primers* directos y reversos se unen covalentemente a una secuencia nucleotídica. Los fragmentos de la biblioteca de ADN hibridan mediante la adición de adaptadores con los *primers* sobre el soporte y se realiza una reacción de PCR formándose nuevas cadenas desde ambos extremos en forma de puente. Esto se repite con los *primers* adyacentes generándose grupos aislados que contienen múltiples copias idénticas de un fragmento único. La secuenciación consiste en la técnica de terminación reversible cíclica (TRC). Se agregan los 4 nucleótidos que han sido modificados para unir una molécula fluorescente con un color diferente cada uno, sin permitir la incorporación del siguiente nucleótido a la cadena hasta que no se retire la modificación. Se produce la unión del nucleótido complementario a cada una de las cadenas en formación. Luego se lava el resto de los nucleótidos no unidos y se registra la señal fluorescente en una imagen con cámaras de alta resolución posterior a la excitación mediante láseres. Este proceso se repite cíclicamente y finalmente se analizan las imágenes para obtener la secuencia completa. La señal fluorescente observada será el consenso de las emitidas por todos los nucleótidos añadidos a cada grupo de cadenas idénticas de ADN en cada uno de los ciclos (Bentley *et al*., 2008). Este proceso de secuenciación se observa en la Figura N°1. Los detalles de los pasos y procesos según la plataforma illumina se describen a continuación:

Cuantificación de la plantilla. La cuantificación precisa de la plantilla se ha convertido en una parte fundamental de Illumina preparación de la biblioteca. En lugar de depender de métodos de cuantificación espectrométrica, que son fácilmente influenciados por el pH, nucleótidos libres, sales, proteínas y otros contaminantes que absorber la luz en una longitud de onda de 260 nm, se sugiere el uso de métodos de cuantificación por fluorometría para estimar con precisión las concentraciones de ADN. Estos métodos no son infalibles, ya que otras especies, como grandes cantidades de ARN contaminante, también pueden unirse tintes intercalantes fluorescentes (Bronner & Quail, 2019).

Fragmentación de la plantilla. Las bibliotecas de secuenciación estándar de Illumina tienden a tener un tamaño de fragmento de 200 hasta 500 bp, excluidos los adaptadores. Los fragmentos más grandes de hasta 1000 bp se agruparán, pero con cada vez menor eficiencia y menor rendimiento. En las células de flujo con patrón, estos fragmentos más grandes corren el riesgo de llegar a nanopozos adyacentes, por lo que es valioso mejorar la hermeticidad de los tamaños de los insertos de la biblioteca. El método más práctico en cuanto a laboratorios sin equipamiento es la enzimática (Bronner & Quail, 2019).

Reparación de las terminaciones. La mayoría de los métodos de fragmentación aleatoria producen ADN bicatenario con un mezcla de extremos romos, extremos romos 3′ y extremos romos 5′, con y sin resto fosfato en 5′. Estos deben uniformarse antes de poder ligar los adaptadores, por lo que se utiliza una mezcla de enzimas para generar fragmentos de extremos romos con terminaciones 5′ fosforilados (Bronner & Quail, 2019).

Colas A. La adición de un solo nucleótido A sin plantilla a los extremos 3′ de los fragmentos antes de la ligadura del adaptador impide la concatemerización de plantillas, y ya que los adaptadores tienen cola en T, aumenta la eficiencia de la ligadura del adaptador y disminuye la formación de dímeros adaptadores (Bronner & Quail, 2019).

Ligación de adaptadores. Las hebras de plantilla deben recibir una secuencia de adaptador diferente en cada extremo para participar con éxito en las reacciones de secuenciación y amplificación de grupos. La ligadura del adaptador a las plantillas deben ser lo más eficientes posible, pero al mismo tiempo, la ligadura de adaptadores entre sí deben suprimirse: los dímeros de adaptador generarán clústeres y puede interferir con la secuenciación, lo que reducirá la proporción total de la secuencia deseada (Bronner & Quail, 2019).

Limpieza. La limpieza se realiza originalmente mediante columnas; sin embargo, los flujos de trabajo son generalmente bastante lentos y no escalables para un alto rendimiento. La mayoría de las limpiezas ahora se realizan utilizando la inmovilización reversibles de fase sólida (IRFS) mediante perlas (DeAngelis *et al*., 1995). Estas perlas usan una partícula magnética recubierta de carboxilo que puede unirse reversiblemente al ADN en presencia de polietilenglicol (PEG) y sal (NaCl). Un beneficio adicional de usar perlas IRFS es que cuando se utilizan después de la ligadura, permiten eliminar la mayoría de los dímeros del adaptador de la biblioteca.

Selección de tamaños. La unión de los fragmentos a perlas IRFS (DeAngelis *et al*., 1995) se puede controlar alterando el PEG y concentraciones de NaCl (Borgström et al., 2011; Lundin et al., 2010). Los tratamientos de kits comerciales se pueden utilizar para eliminar fragmentos de ADN por debajo de un cierto tamaño (normalmente 150 hasta 200 pb). También se puede realizar una selección de tamaño doble (superior e inferior) utilizando dos pasos consecutivos de limpieza mediante IRFS (Borgström et al., 2011)

Reacción en cadena de la polimerasa. Después de la selección y limpieza del tamaño, las bibliotecas se pueden amplificar mediante la cadena de polimerasa reacción (PCR) para enriquecer las hebras de molde correctamente ligadas, o aquellas que tienen un adaptador en ambos extremos; aumentar la cantidad de biblioteca disponible para la secuenciación; generar suficiente ADN para una cuantificación precisa; y para agregar secuencias de oligonucleótidos a las hebras plantilla para permitir la hibridación a la superficie de la celda de flujo, como estas secuencias no están contenidas en el adaptador ligado (Bronner & Quail, 2019).

Cuantificación. El número de grupos por carril de una celda de flujo se rige por la concentración de la biblioteca. La cuantificación precisa es vital porque una densidad de grupos o *clúster* demasiado baja reduce el rendimiento de datos y, por lo tanto, aumenta el costo por base de secuenciación, mientras que una densidad de conglomerados da como resultado un rendimiento reducido de datos debido a la superposición de grupos. Para bibliotecas de rutina amplificadas por PCR con tamaños de inserto muy estrictos, la cuantificación fluorométrica o cuantificación mediante un método sensible basado en gel sería suficiente. Para bibliotecas sin PCR o de tamaño menos estricto, sugerimos el tamaño las bibliotecas y cuantificándolas mediante PCR cuantitativa (qPCR). qPCR debe ser capaz de detectar y cuantificar todas las moléculas amplificables, es decir, aquellos con adaptadores en cada extremo (Meyer *et al*., 2008)

Normalización de la biblioteca y agrupación. Las bibliotecas terminadas deberán cargarse en el secuenciador a una concentración adecuada con el fin de lograr una densidad de *clusters* óptima. Para reducir los costos de secuenciación individual bibliotecas o para adaptar la cobertura de la secuencia, se pueden agregar varias bibliotecas indexadas al mismo pool de bibliotecas (multiplexación). En general, las bibliotecas similares del mismo organismo son agrupados en cantidades equimolares y, dependiendo de la cobertura solicitada, el tamaño del genoma, y el rendimiento del secuenciador, estos generalmente se pueden agrupar en 96 plex, 384-plex o superior (Bronner & Quail, 2019).

Denaturación. Las bibliotecas se convierten en monocatenarias mediante incubación con hidróxido de sodio, lo que permite hibridación eficiente de las hebras molde con cebadores unidos a la superficie de la celda de flujo. Después de la desnaturalización, las muestras se cargan en las celdas o clulas de flujo o directamente en el secuenciador dependiendo del fabricante (Bronner & Quail, 2019).

Hibridación y extensión. Como resultado de la preparación de la biblioteca, las cadenas molde poseen una secuencia en un extremo que coincide exactamente con uno de los cebadores de la celda de flujo, y en su extremo opuesto, poseen una secuencia que está en la orientación complementaria inversa al otro cebador de la celda de flujo. Por lo tanto, solo el extremo complementario de cada hebra de plantilla se hibrida con un cebador de la celda de flujo. Los cebadores de celda de flujo se extienden mediante una ADN polimerasa, lo que genera una copia de complemento inverso de la hebra de plantilla original que está atada a la celda de flujo superficie. La hebra de la plantilla original no está atada y se quita enjuagando con formamida(Bronner & Quail, 2019).

Amplificación de grupos. El extremo libre de cada hebra atrapada recién generada es complementaria a la otra imprimación en la superficie de la celda de flujo y puede hibridar con ella. Existen algunas diferencias menores entre la agrupación en una celda de flujo estándar y la agrupación en una celda de flujo con patrón, que Illumina denomina química de exclusión-amplificación (ExAmp). Illumina ha proporcionado plataformas para la amplificación de grupos y celdas de flujo con patrón. En ambos casos, la amplificación en puente es utilizado para amplificar la señal de hebras de biblioteca individuales. Aquí, la imprimación en la celda de flujo en la superficie hibrida la biblioteca. Luego, la imprimación se extiende, generando una doble hebra producto. Las hebras de este producto se desnaturalizan con formamida, produciendo dos hebras el extremo libre de cada lata se hibridan con el otro imprimador en la superficie de la celda de flujo. Los ciclos repetidos de desnaturalización, hibridación y extensión dan como resultado un grupo de 1000 hebras para celdas de flujo normales o continúe hasta que el nanopozo esté lleno para células de flujo patrón. Cada uno de estos grupos / pocillos se deriva de la misma hebra de plantilla original y, por tanto, es clonal. Lamentablemente, la química de Ex-Amp tiene un inconveniente en comparación a la antigua metodología de amplificación de conglomerados: es más susceptible al índice de conmutación (Sinha *et al*., 2017)

Linealización, bloqueo e hibridación. Como consecuencia de la estructura de los adaptadores utilizados para el paso de generación de la biblioteca, las uniones adaptador / inserto tienen 12 nucleótidos idénticos en los extremos opuestos de los fragmentos que se utilizan como plantillas de enlace de secuenciación. Para asegurarse de que cada clúster sea secuenciado solo en una dirección durante cada lectura y para permitir una hibridación más eficiente de secuenciación de cebadores a hebras dentro de un grupo, uno de los cebadores de celda de flujo se escinde, elimina una hebra de forma selectiva y da como resultado grupos de una sola hebra para reducir ruido de fondo en la reacción de secuenciación, sin embargo, los extremos 3´de los cebadores de celda de flujo extendidos y no extendidos están bloqueados por la incorporación de didesoxirribonucleótidos, y los cebadores de secuenciación se hibridan (Bronner & Quail, 2019).

Secuenciación por síntesis. Las células de flujo híbridas se someten a ciclos repetidos de incorporación de nucleótidos, formación de imágenes y escisión (desbloqueo simultáneo del extremo 3 y eliminación del fluoróforo). Originalmente, Illumina usó cuatro colores para etiquetar sus didesoxirribonucleótidos y tomó imágenes de estos individualmente (SBS de cuatro canales). Para algunos secuenciadores, esto se ha reducido a dos canales SBS, que usa dos colores diferentes para T y C, con ambos colores usados en A, mientras que G no es fluorescente. Un sistema de detección aún más avanzado que solo usa iSeq que utiliza SBS de un canal. Aquí, las bases contienen una sonda fluorescente que puede se escinde en la mitad del ciclo (A), es estable (T), no tiene fluorescencia (G) o está biotinilada y emite fluorescencia cuando se ata a una base unida a estreptavidina. La obtención de imágenes a través de iSeq también es muy diferente a otros sistemas, donde la luz se detecta utilizando metal complementario fotodiodos semiconductores de óxido (CMOS) dentro de la celda de flujo (Bronner & Quail, 2019).

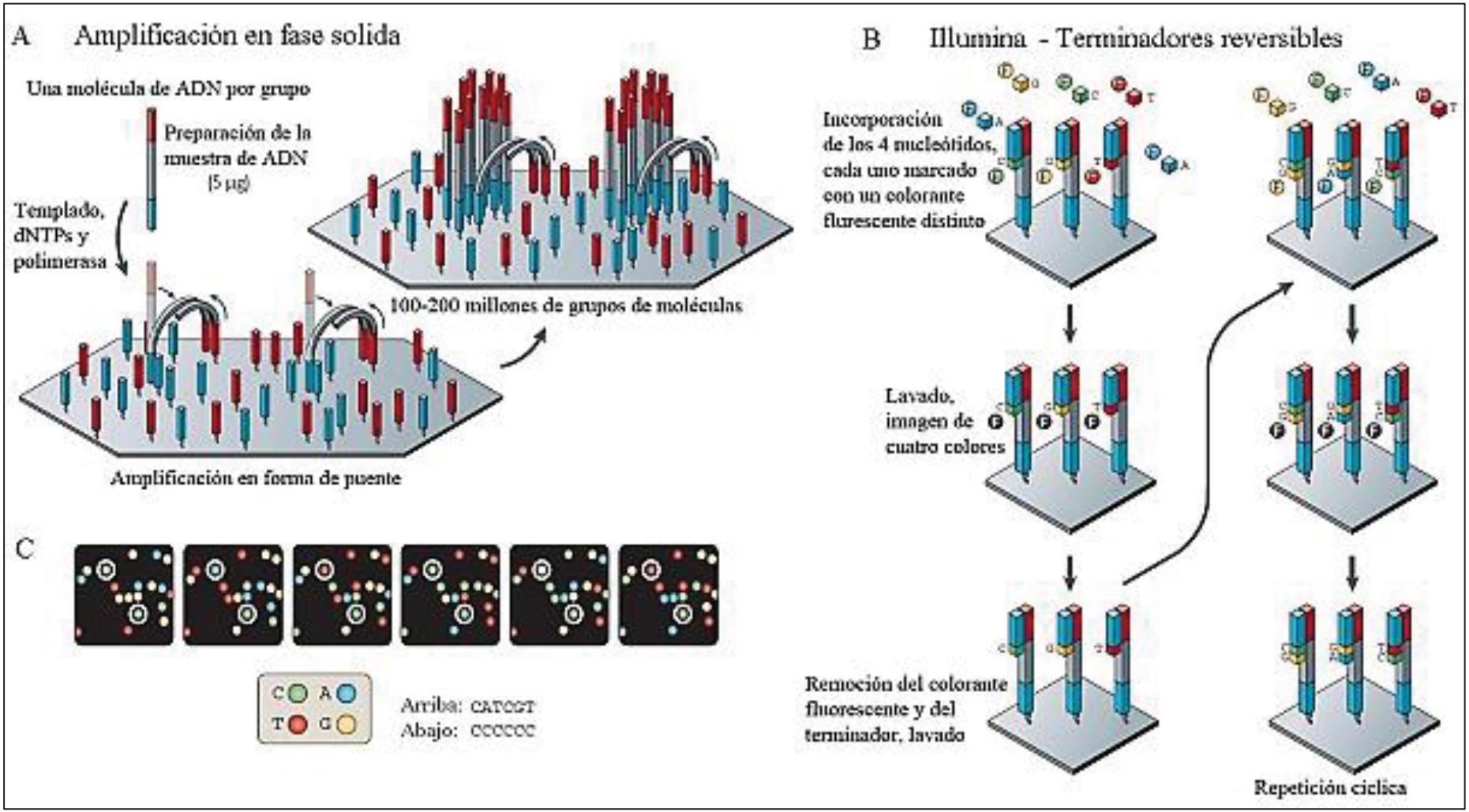


Figura N° 1: Plataforma Illumina. A - Amplificación en fase sólida; B- secuenciación TRC con cuatro colores; C- imágenes de cuatro colores. Se observa datos de secuenciación de dos fragmentos.
Adaptado de Metzker, 2010.

1.2.3. Bioinformática y análisis de datos NGS.

La bioinformática es una disciplina recientemente definida que desarrolla y aplica herramientas computacionales avanzadas para analizar e interpretar datos biológicos de alta dimensión. El papel de los bioinformáticos y los flujos de trabajo de bioinformática son nuevos para los laboratorios de secuenciación y requiere inversiones sustanciales en educación, personal y *hardware*, así como plasticidad en los procesos y partes involucradas en las pruebas. El análisis bioinformático basado en NGS está diseñado para convertir señales en datos, datos en información interpretable e información en conocimiento procesable. Este proceso se puede conceptualizar como análisis primario, secundario y terciario (Figura N° 2). En resumen, el análisis primario consiste en procesar señales de instrumentos de secuenciación sin procesar en base de nucleótidos y datos de lectura corta. El análisis secundario implica la alineación con una secuencia de referencia o el ensamblaje de novo de las lecturas de nucleótidos de NGS y la posterior detección de variantes, y los análisis de bioinformática terciaria proporcionan contexto a la información generada durante un experimento de NGS al asociar el perfil genómico específico de la muestra con anotaciones descriptivas dispares (Oliver *et al*., 2015).

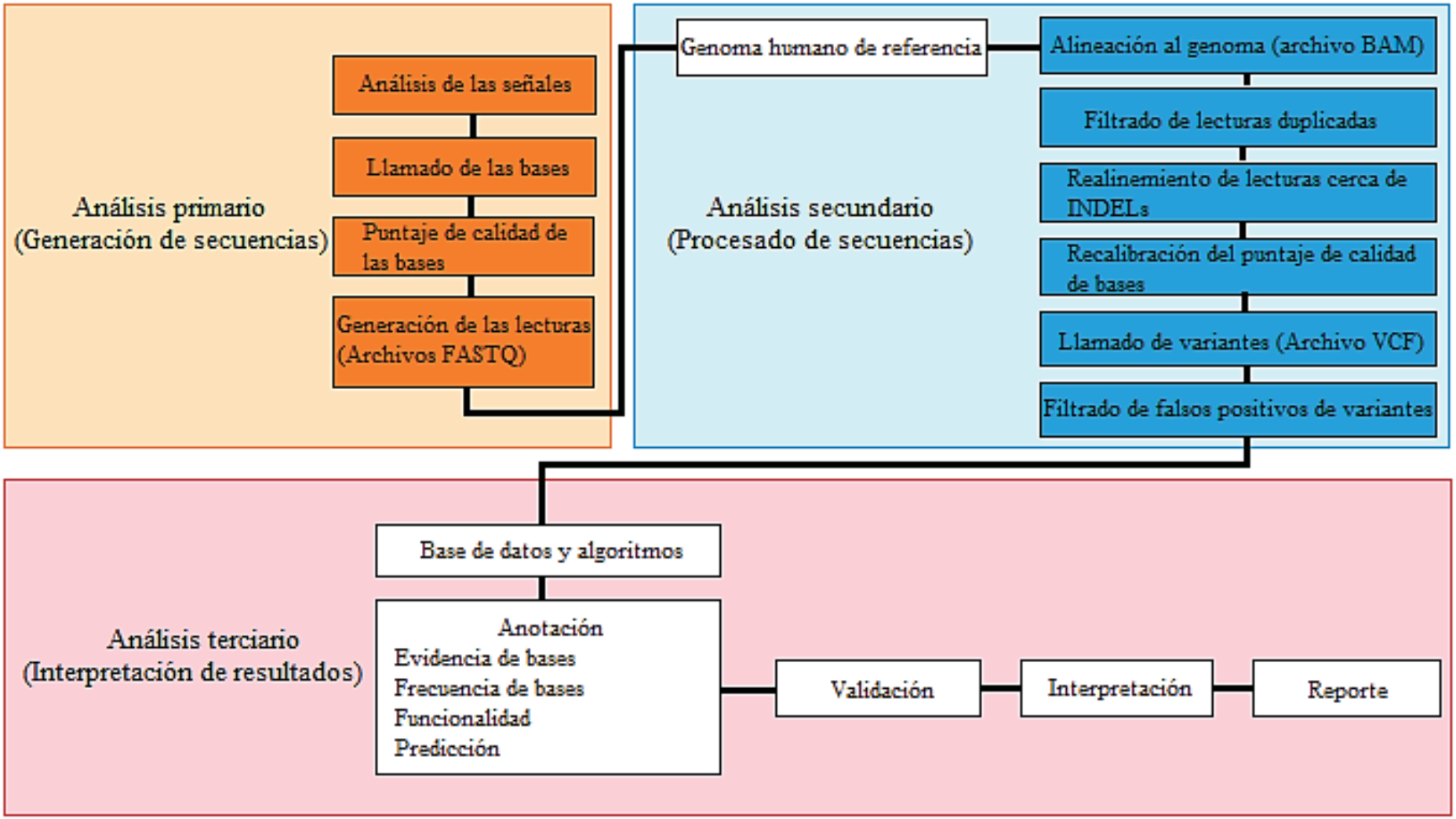


Figura N° 2: Análisis primario, secundario y terciario de datos NGS en librerías genómicas humanas. INDELs (Small insertions and deletions), VCF es una extensión de archivos.
Fuente Oliver et al., 2015, *traducción por Pacheco – Capcha 2019.*

1.2.3.1. Generación de secuencias (Análisis primario).

El análisis primario es un proceso que se ha integrado altamente con los instrumentos de secuenciación y el software asociado. Estas herramientas convierten las señales sin procesar generadas por los instrumentos de secuenciación en bases de nucleótidos con puntuaciones de calidad asociadas y, en última instancia, secuencias de nucleótidos cortas o "lecturas" (Oliver *et al*., 2015).

1.2.3.2. Alineación y detección de variantes (Análisis secundario).

El análisis secundario consiste en una colección variable de métodos que operan juntos para detectar aberraciones genómicas a partir de datos de secuencia con calificación de calidad. Dependiendo del protocolo, este perfil puede ocurrir a nivel del genoma, exoma o paneles de genes enfocados. La clase de variación genómica perfilada puede variar e incluye variantes de un solo nucleótido, pequeñas inserciones y deleciones, o alteraciones mayores como reordenamientos estructurales y cambios en el número de copias. Además, las variaciones genómicas pueden ser constitucionales (de *novo* o heredadas) o somáticas (adquiridas). Aunque cada una de estas consideraciones introduce diferencias sutiles en el protocolo de análisis, los procesos fundamentales son muy similares.

El paso de análisis secundario inicial generalmente implica la alineación colectiva de las lecturas con un genoma de referencia. En esta etapa también es posible realizar el ensamblaje de *novo* de un genoma (Oliver *et al*., 2015).

1.2.3.3. Anotación y visualización (Análisis terciario).

Tras la detección, las variantes deben anotarse para determinar su importancia biológica y permitir la priorización funcional y la interpretación posterior. Esta caracterización se logra generalmente utilizando una combinación de fuentes de anotación biológica que incluyen datos basados en frecuencia, estructurales, predicción o evidencia. Cada clase de anotación tiene beneficios y limitaciones asociados, y cuando se aplica en una interpretación posterior puede introducir más desafíos analíticos (Oliver *et al*., 2015).

Las anotaciones basadas en la predicción utilizan cambios de nucleótidos y / o aminoácidos integrados con datos contextuales adicionales, incluidas las puntuaciones de conservación evolutiva, matrices de sustitución de aminoácidos y el impacto en las estructuras de proteínas 3D para inferir el impacto de la variante en el producto de secuencia resultante. Estos sistemas de software a menudo utilizan modelos de aprendizaje computacional (por ejemplo, redes neuronales, árboles de decisión, modelos ocultos de Markov) (Oliver *et al*., 2015).

1.2.4. Programas Bioinformáticos.

A continuación se describe la función de los programas bioinformáticos empleados en el presente trabajo explicando los diagramas, tablas, valores e índices que expresan tras su análisis.

1.2.4.1. FastQC

FastQC realiza comprobaciones de control de calidad en datos de secuencia sin procesar que provienen de tuberías de secuenciación de alto rendimiento. Proporciona un conjunto modular de análisis que puede usar para dar una impresión rápida de que los datos manejados tienen algún problema que se debe tener en cuenta antes de realizar cualquier análisis posterior. Las funciones principales de FastQC son: importación de datos de archivos BAM, SAM o FastQ (cualquier variante), proporcionar una visión general rápida para tener conocimiento en qué áreas puede haber problemas, evalúa rápidamente los datos mediante sus gráficos y tablas de resumen, exporta los resultados a un informe permanente basado en HTML y además de operar sin conexión para permitir la generación automatizada de informes sin ejecutar la aplicación interactiva (Andrews, 2009).

Un valor importante en FastQC es Phred. Las puntuaciones de calidad de Phred se definen como una propiedad que está relacionada logarítmicamente con las probabilidades de error de llamadas de base. Por ejemplo, si Phred asigna un puntaje de calidad de 30 a una base (Tabla 1), las posibilidades de que esta base se llame incorrectamente son 1 en 1000 (Ewing & Green, 1998).

Tabla N° 1: Índice phred: Probabilidad de que la asignación de una base sea incorrecta.

Nivel de calidad de Phred	Probabilidad de llamada base incorrecta	Precisión de llamada base
10	1 en 10	90%
20	1 en 100	99%
30	1 en 1000	99,9%
40	1 en 10,000	99,99%
50	1 en 100,000	99,999%
60	1 en 1,000,000	99,9999%

FastQC ofrece 12 módulos para el análisis de calidad de los datos de tuberías de secuenciación avanzada. El primer módulo **Estadísticas básicas** genera algunas estadísticas de composición simples para el archivo analizado.

En el módulo **Calidad de las secuencias por base** (Figura N°3) muestra una descripción general del rango de valores de calidad phred de todas las bases. Para cada posición se dibuja un diagrama de cajas o de tipo *BoxWhisker*. Los elementos de dicho diagrama son los siguientes: La línea roja central es el valor de la mediana, el cuadro amarillo representa el rango intercuartil (25-75%), los bigotes superior e inferior representan el 10% y el 90% de los puntos. La línea azul representa la calidad media. El eje en el gráfico muestra los puntajes de calidad. Cuanto mayor sea el puntaje, mejor será la llamada de la base (Figura N° 3).

El fondo del gráfico divide en llamadas de muy buena calidad (verde), llamadas de calidad razonable (naranja) y llamadas de baja calidad (rojo). La calidad de las llamadas en la mayoría de las plataformas se degrada a medida que avanza la ejecución, por lo que es común ver que las llamadas de base caen en el área naranja hacia el final de una lectura (Figura N° 3).

El gráfico del módulo **Calidad de secuencia por mosaico** (Figura N° 4) solo aparece al utilizar los resultados de análisis de una biblioteca Illumina ya que conserva sus identificadores de secuencia originales. Codificado en estos está el mosaico de la celda de flujo de dónde proviene cada lectura. El gráfico permite ver los puntajes de calidad de cada mosaico en todas sus bases para ver si hubo una pérdida de calidad asociada con solo una parte de la celda de flujo.

El gráfico muestra la desviación de la calidad promedio de cada mosaico. Los colores están en una escala de frío a caliente, con colores fríos (azul) en posiciones donde la calidad era igual o superior al promedio de esa base en la ejecución, y los colores más cálidos (rojo) indican que un mosaico tenía peores cualidades que otros mosaicos para esa base. En la figura N° 4 puede observarse que ciertos mosaicos muestran una calidad consistentemente deficiente. Una buena trama debe ser azul por todas partes.

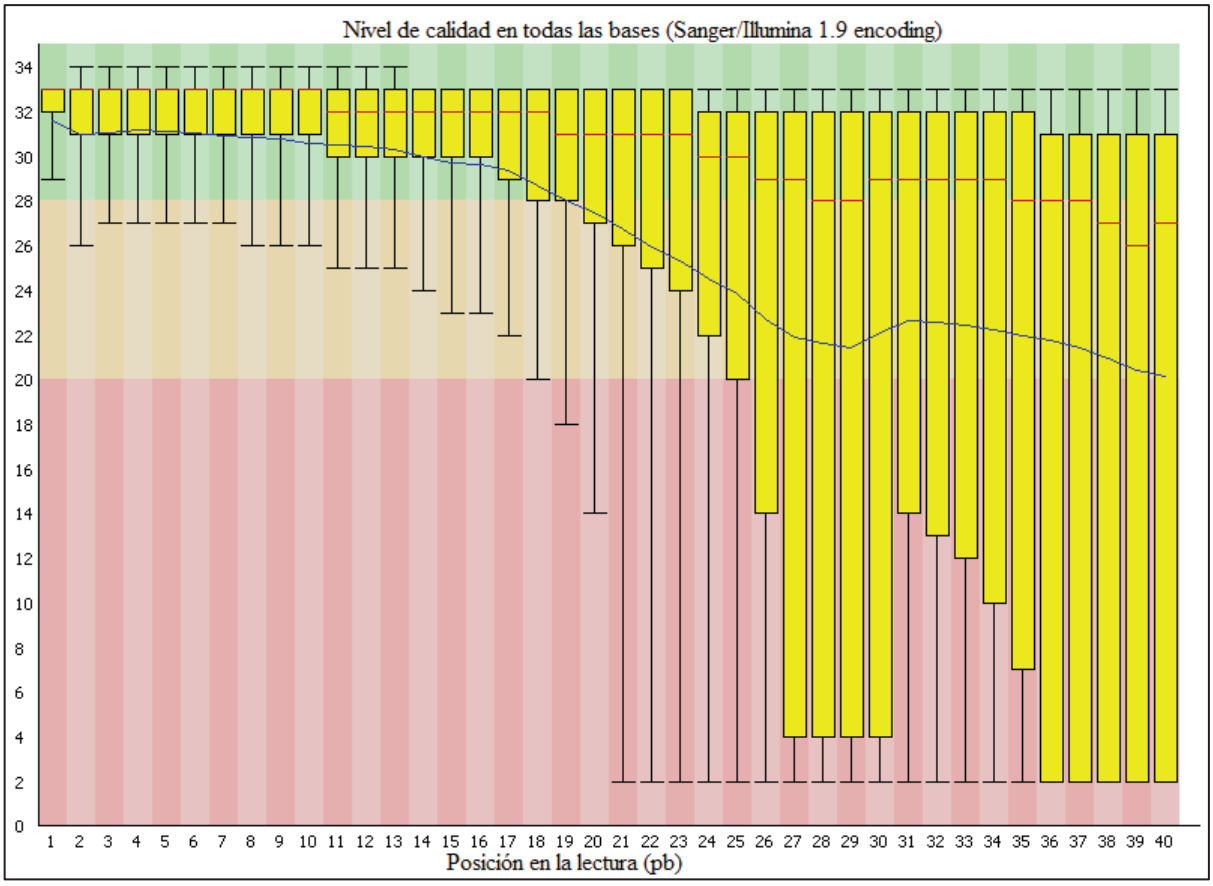


Figura N° 3. Diagrama de cajas o BoxWhisker.
Fuente FastQC, modificado por Pacheco-Capcha 2019.

Las razones para ver advertencias o errores en esta gráfica podrían ser problemas transitorios, como burbujas que atraviesan la celda de flujo, o podrían ser problemas más permanentes, como manchas en la celda de flujo o desechos dentro del carril de la celda de flujo. Las advertencias también pueden aparecer cuando una celda de flujo generalmente está sobrecargada

El informe **Niveles de calidad por secuencia** permite ver si un subconjunto de secuencias tiene valores de calidad universalmente bajos. A menudo se da el caso de que un subconjunto de secuencias tendrá una calidad universalmente baja que deberían representar solo un pequeño porcentaje del total de secuencias (Figura N° 5).

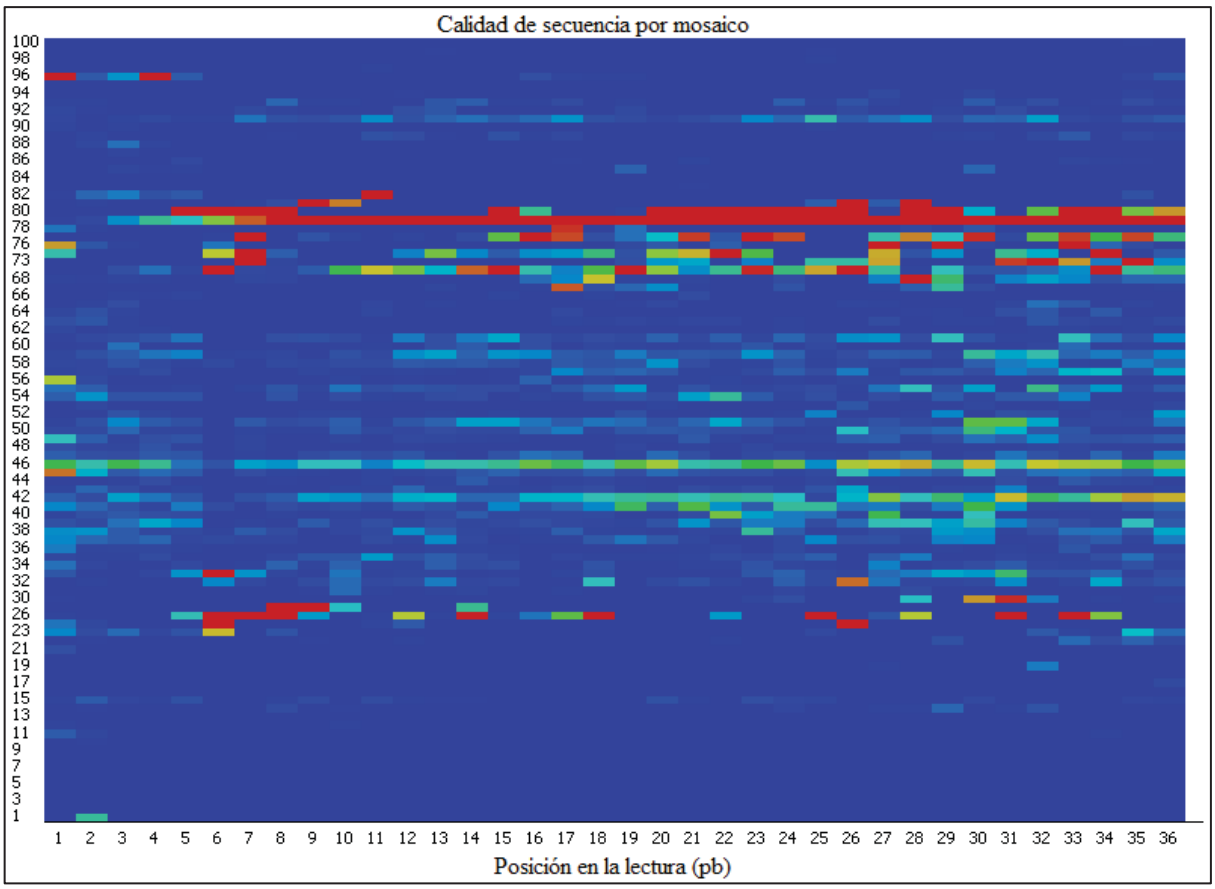


Figura N° 4. Calidad de secuencia por mosaico
Fuente FastQC, modificado por Pacheco-Capcha 2019.

Si una proporción significativa de las secuencias en una ejecución tiene una calidad general baja, esto podría indicar algún tipo de problema sistemático, posiblemente con solo una parte de la ejecución (por ejemplo, un extremo de una celda de flujo).

Para lecturas largas, esto puede aliviarse mediante un recorte de calidad

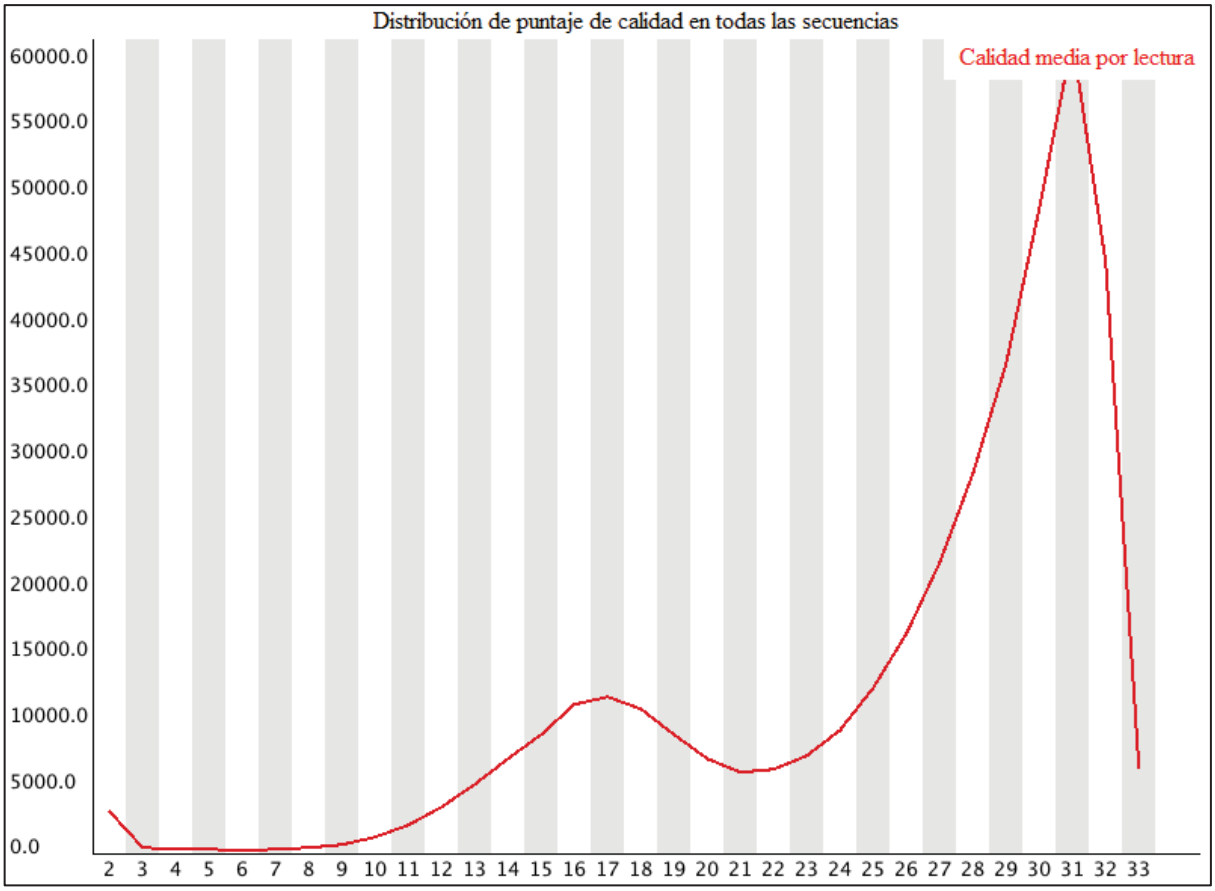


Figura N° 5. Niveles de calidad por secuencia
Fuente FastQC, modificado por Pacheco-Capcha 2019.

El **contenido de las secuencia por base** traza la proporción de cada posición de base en un archivo para el que se ha llamado a cada una de las cuatro bases de ADN normales.

En una biblioteca aleatoria, se espera que haya poca o ninguna diferencia entre las diferentes bases de una ejecución de secuencia, por lo que las líneas en este gráfico deben correr paralelas entre sí. La cantidad relativa de cada base debe reflejar la cantidad total de estas bases en su genoma, pero en cualquier caso no deben estar muy desequilibradas entre sí.

Si hay alguna evidencia de secuencias sobrerrepresentadas, como dímeros adaptadores o ARNr en una muestra, estas secuencias pueden sesgar la composición general y así alterar esta gráfica (Figura N° 6).

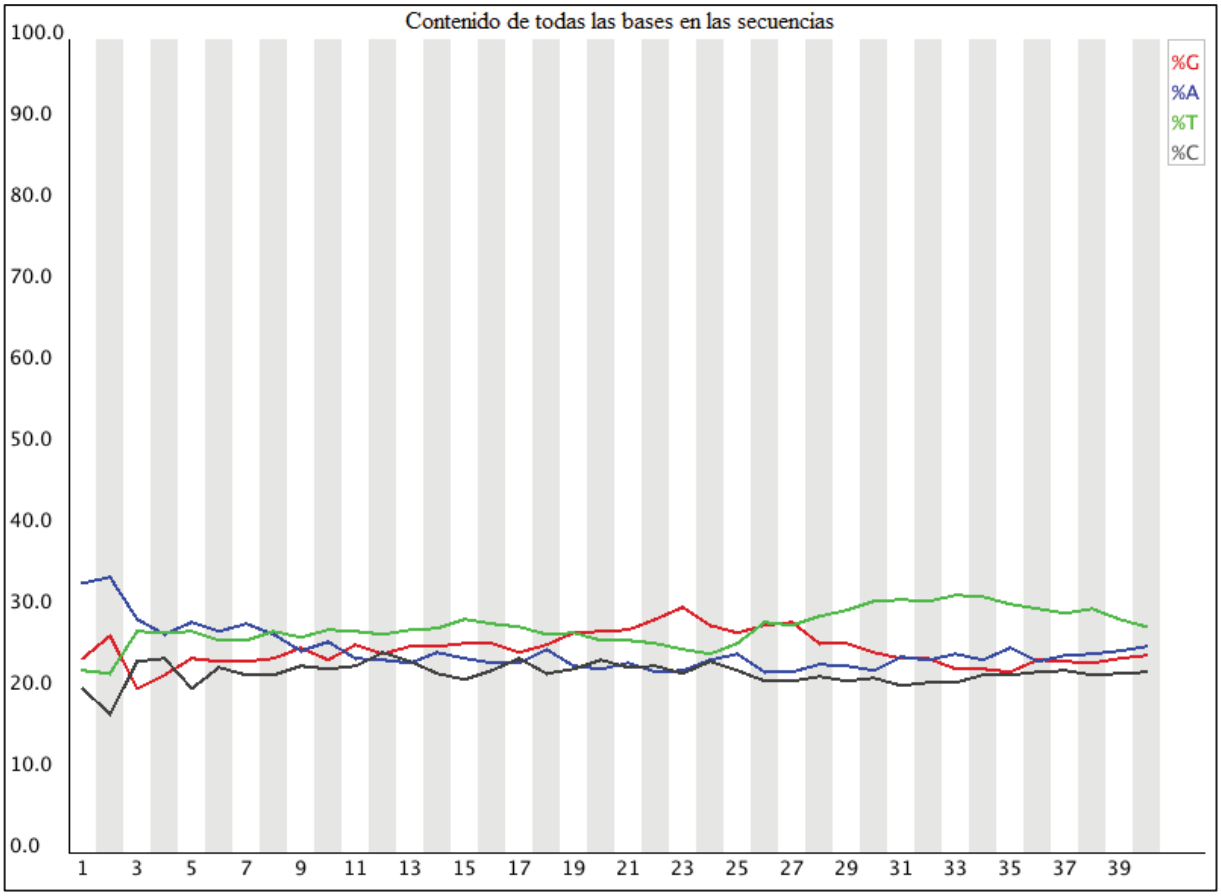


Figura N° 6. Contenido de las secuencia por base.
Fuente FastQC, modificado por Pacheco-Capcha 2019.

En el módulo **Contenido de GC por secuencia** mide el contenido de GC en toda la longitud de cada secuencia en un archivo y lo compara con una distribución normal modelada de contenido de GC por el programa (Figura N° 7).

Las advertencias en este módulo generalmente indican un problema con la biblioteca. Los picos agudos en una distribución uniforme son normalmente el resultado de un contaminante específico (dímeros, adaptadores, por ejemplo), que bien puede ser captado por el módulo de secuencias sobrerrepresentadas. Los picos más amplios pueden representar contaminación con una especie diferente.

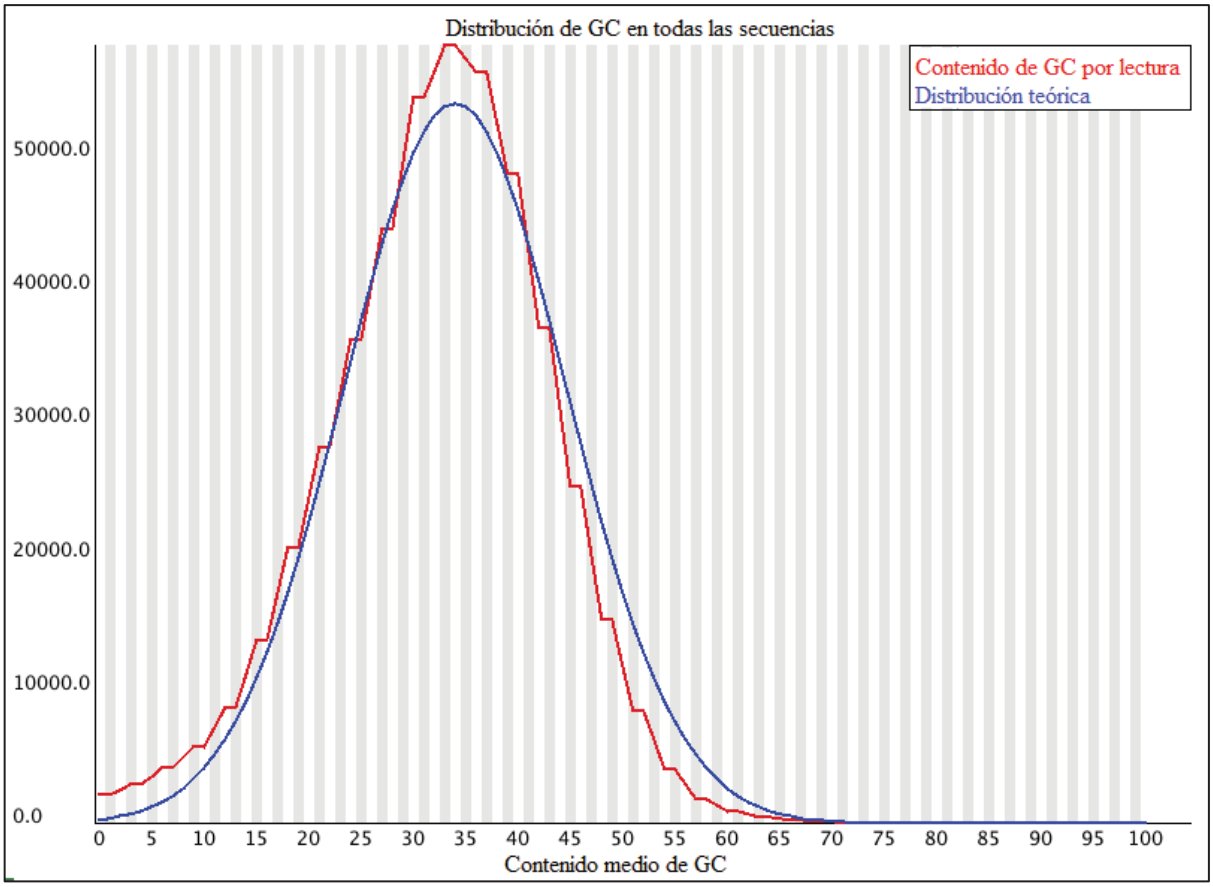


Figura N° 7. Contenido de GC por secuencia.
Fuente FastQC, modificado por Pacheco-Capcha 2019.

Si un secuenciador no puede hacer una llamada de base con suficiente confianza, normalmente sustituirá una llamada N en lugar de una llamada base convencional. Es así que el módulo de **contenido N por base**, traza el porcentaje de llamadas N.

La razón más común para la inclusión de proporciones significativas de Ns es una pérdida general de calidad, por lo que los resultados de este módulo deben evaluarse en concierto con los de los diversos módulos de calidad (Figura N° 8).

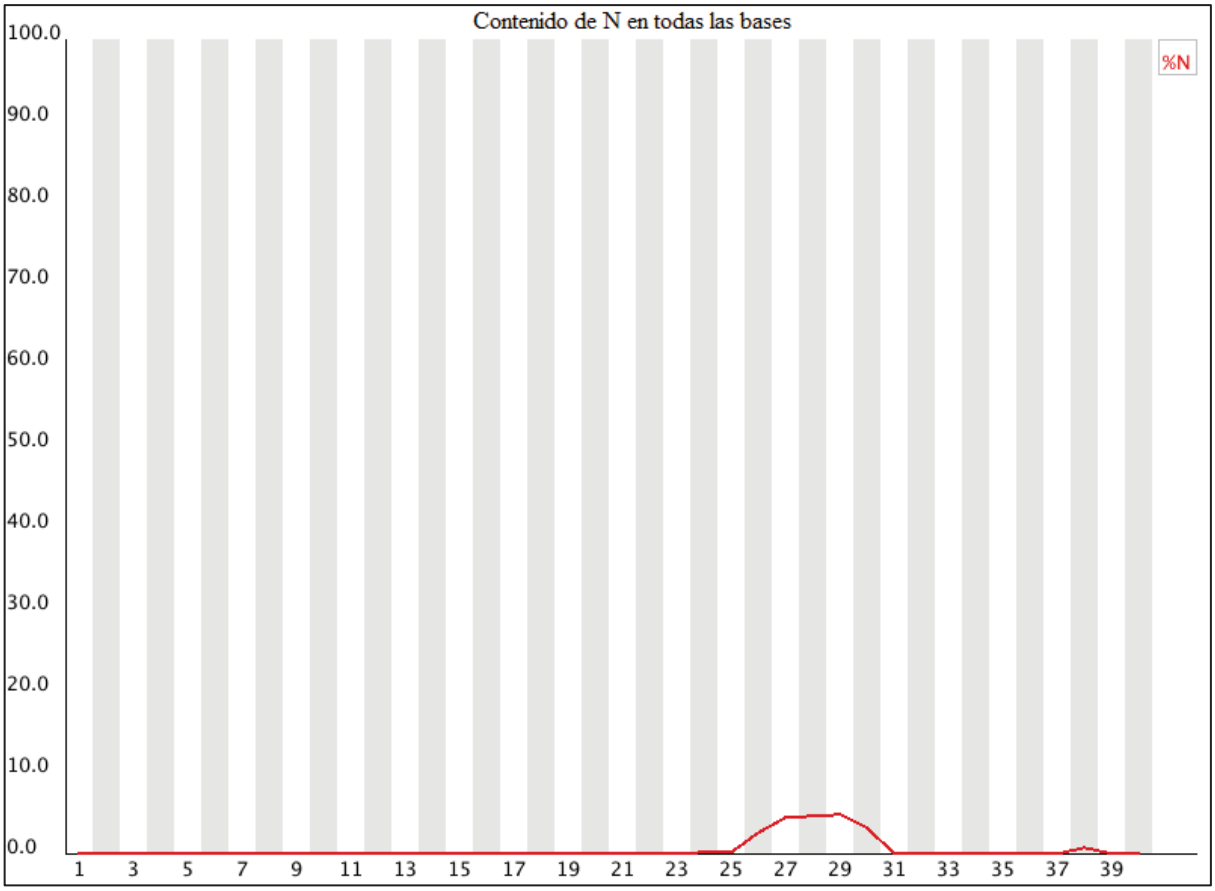


Figura N° 8. Contenido N por base.
Fuente FastQC, modificado por Pacheco-Capcha 2019.

El módulo **distribución de longitud de secuencia** genera un gráfico que muestra la distribución de tamaños de lecturas en el archivo que se analizó (Figura N° 9). Algunos secuenciadores de alto rendimiento generan fragmentos de secuencia de longitud uniforme, pero otros pueden contener lecturas de longitudes muy variables. Incluso dentro de bibliotecas de longitud uniforme, algunas canalizaciones recortarán secuencias para eliminar las llamadas de base de baja calidad del final.

Para algunas plataformas de secuenciación es completamente normal tener diferentes longitudes de lectura, por lo que las advertencias aquí pueden ignorarse.

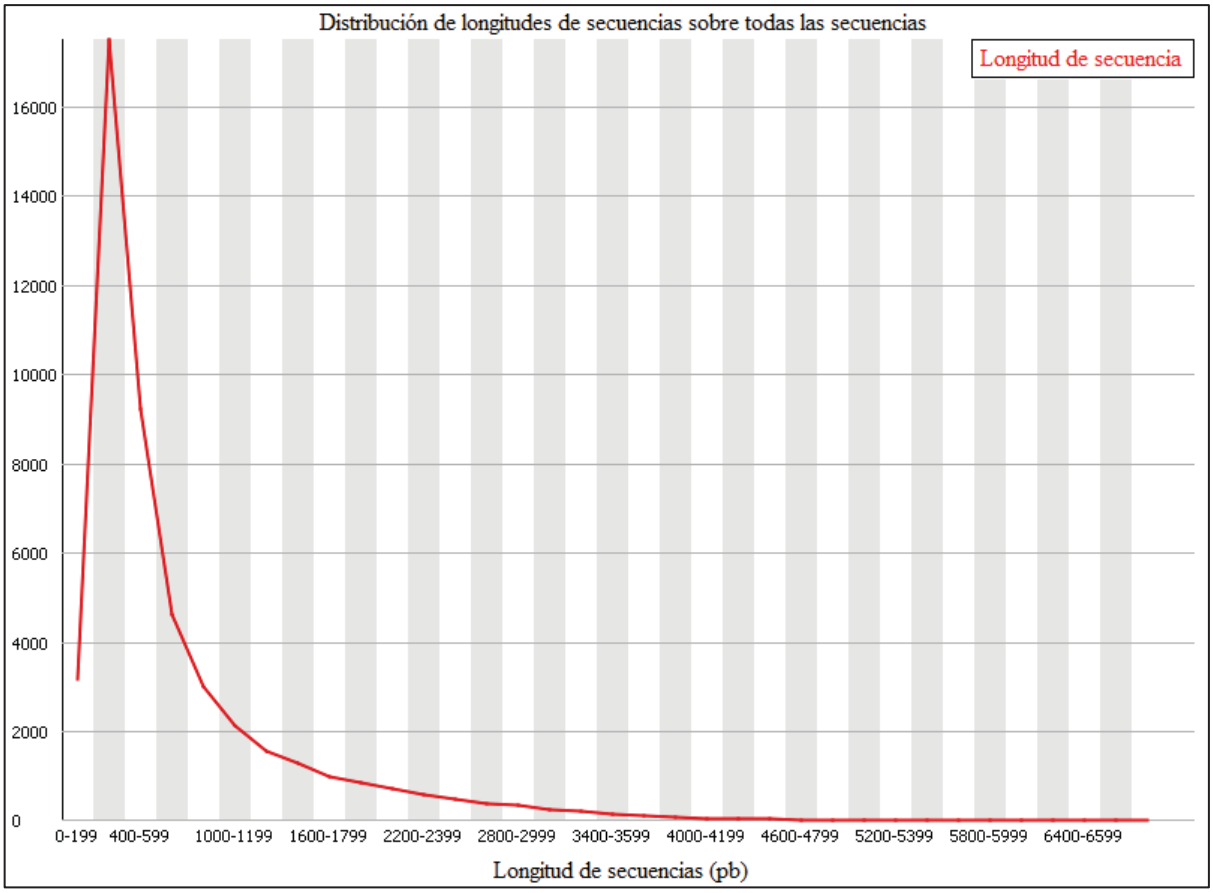


Figura N° 9. Distribución de longitud de secuencia
Fuente FastQC, modificado por Pacheco-Capcha 2019.

En el módulo **niveles de duplicación** (Figura N° 10) mide el grado de duplicación para cada secuencia en una biblioteca y crea un gráfico que muestra el número relativo de secuencias con diferentes grados de duplicación.

El gráfico muestra dos líneas en la trama. La línea azul toma el conjunto de secuencia completa y muestra cómo se distribuyen sus niveles de duplicación. En la gráfica roja, las secuencias se deduplican y las proporciones que se muestran son las proporciones del conjunto deduplicado que provienen de diferentes niveles de duplicación en los datos originales (Figura N° 10).

El módulo también calcula una pérdida general esperada de secuencia si se deduplica la biblioteca. El título de la figura se muestra en la parte superior de la trama y da una impresión razonable del nivel general potencial de pérdida.

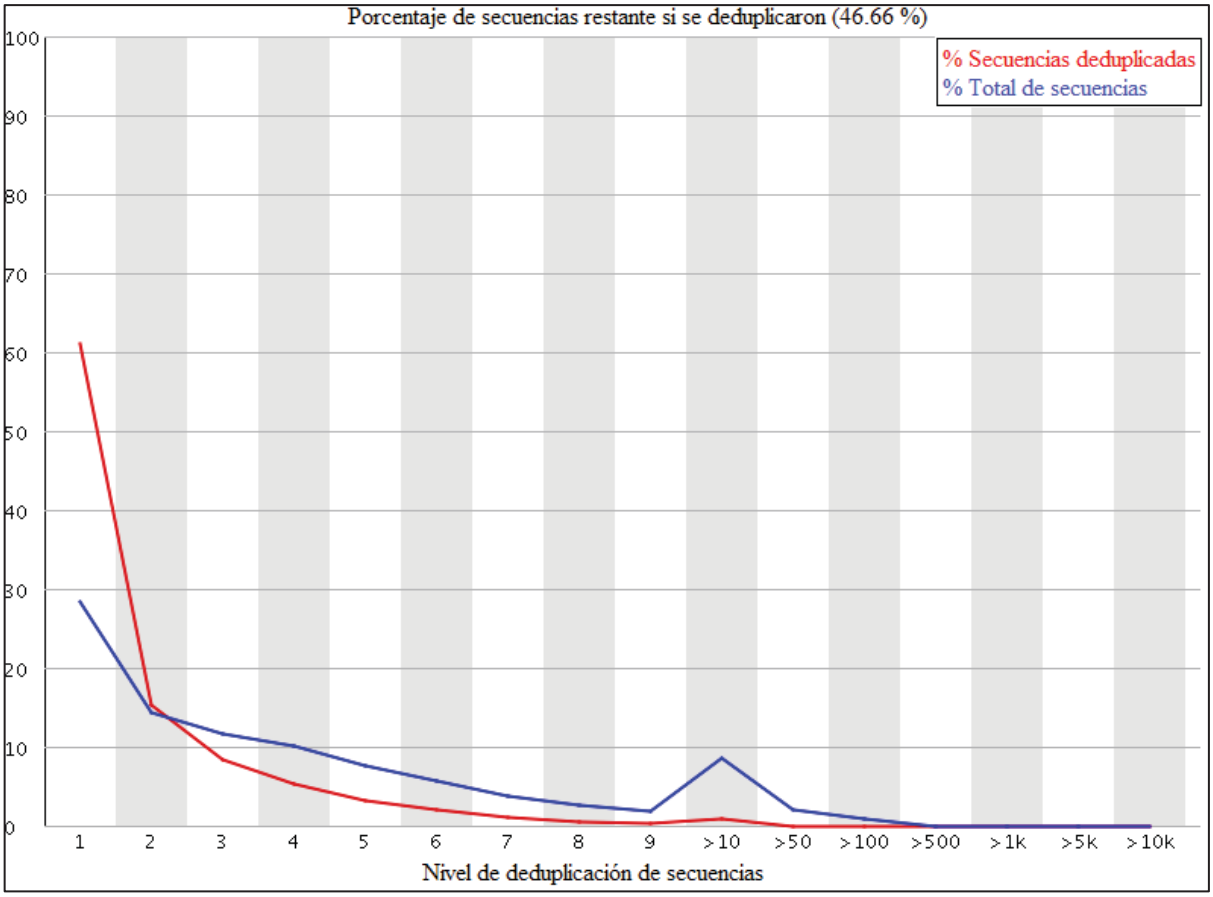


Figura N° 10. Niveles de duplicación.
Fuente FastQC, modificado por Pacheco-Capcha 2019.

En el módulo de **secuencias sobrerrepresentadas** una biblioteca normal de alto rendimiento contendrá un conjunto diverso de secuencias, sin una secuencia individual que constituya una pequeña fracción del conjunto. Descubrir que una sola secuencia está muy sobrerrepresentada en el conjunto indica que es biológicamente significativa o afirma que la biblioteca está contaminada o no es tan diversa como espera.

Un Kmer es una secuencia nucleotídica corta repetitiva o también definido como tamaño de palabra. El **Contenido de adaptadores** hará un análisis genérico de todos los Kmers en su biblioteca para encontrar aquellos que ni siquiera tienen cobertura a lo largo de sus lecturas. Esto puede encontrar varias fuentes diferentes de sesgo en la biblioteca que pueden incluir la presencia de secuencias adaptadoras de lectura que se acumulan al final de sus secuencias.

El módulo **Contenido de Kmer** (Figura N° 11) parte de la suposición de que cualquier fragmento pequeño de secuencia no debe tener un sesgo posicional en su apariencia dentro de una biblioteca diversa. Puede haber razones biológicas por las que ciertos Kmers se enriquecen o agotan en general, pero estos sesgos deberían afectar a todas las posiciones dentro de una secuencia por igual. Por lo tanto, este módulo mide el número de cada 7-mer en cada posición

en su biblioteca y luego utiliza una prueba binomial para buscar desviaciones significativas de una cobertura uniforme en todas las posiciones. Se informa cualquier Kmers con enriquecimiento posicionalmente posicionado. Los 6 mejores Kmer más sesgados se trazan adicionalmente para mostrar su distribución (Figura N° 11).

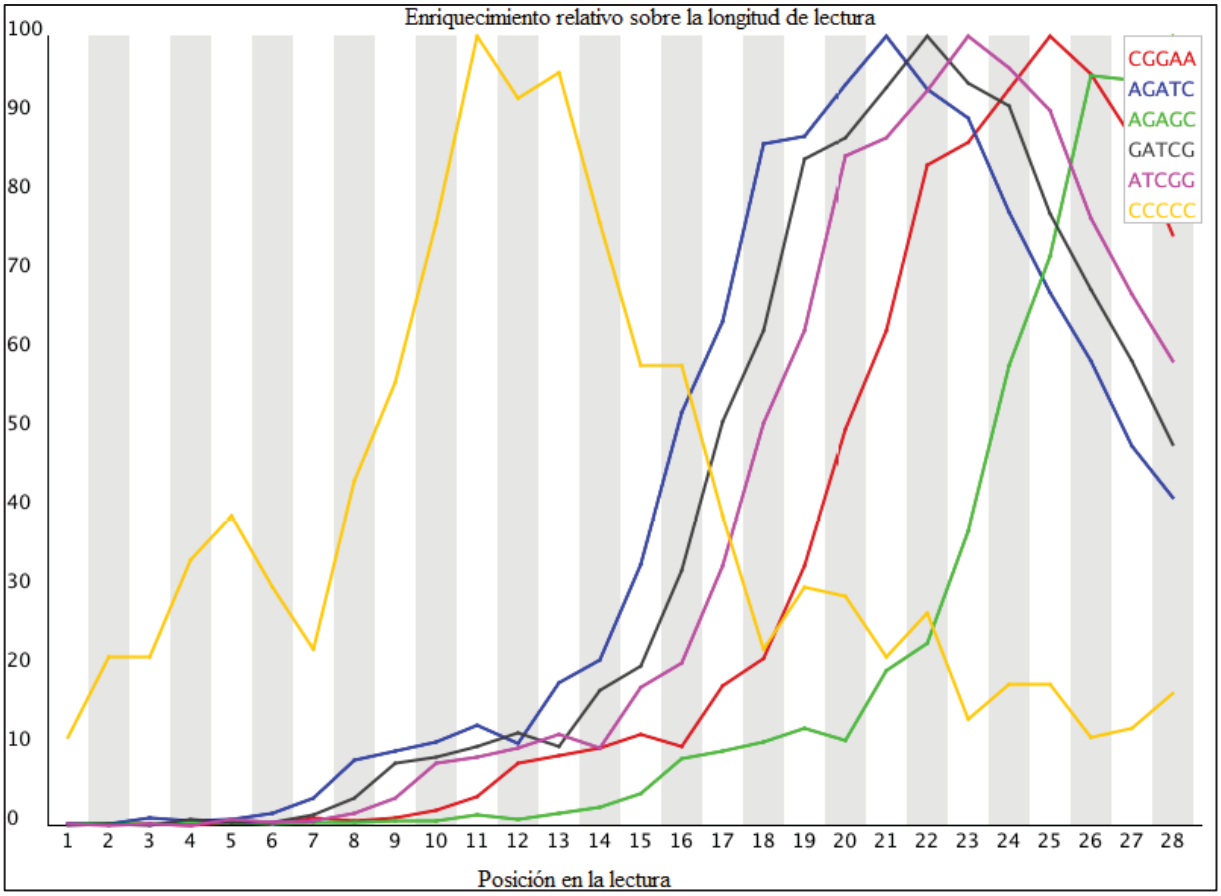


Figura N° 11: Contenido de Kmer
Fuente FastQC, modificado por Pacheco-Capcha 2019.

1.2.4.2. KmerGenie

Estima la mejor longitud de k-mer para el ensamblaje del genoma de novo. Dado un conjunto de lecturas, KmerGenie primero calcula el histograma de abundancia k-mer para muchos valores de k. Luego, para cada valor de k, predice el número de k-mers genómicos distintos en el conjunto de datos, y devuelve la longitud de k-mer que maximiza este número. Los experimentos muestran que las elecciones de KmerGenie conducen a ensamblajes cercanos a los mejores posibles en todas las longitudes de k-mer. Las predicciones de KmerGenie se pueden aplicar a ensambladores de genoma de una sola k (por ejemplo, Velvet, SOAPdenovo 2, ABySS, Minia). Sin embargo, los ensambladores de genoma multi-k (por ejemplo, SPAdes, IDBA) generalmente funcionan mejor con parámetros predeterminados (usando múltiples valores de k), en lugar del mejor k único predicho por KmerGenie.

En los ensambladores basados en De Bruijn, el parámetro más significativo es k, que determina el tamaño de los k -mers en los que se cortan las lecturas. Las repeticiones más largas que k nucleótidos pueden enredar el gráfico y romper contigs; por lo tanto, se desea un gran valor de k. Por otro lado, cuanto más larga sea la k, mayores serán las posibilidades de que un k -mer tenga un error; por lo tanto, hacer k demasiado grande disminuye el número de k -mers correctos presentes en los datos. Otro efecto es que cuando dos lecturas se superponen en menos de k caracteres, no comparten un vértice en el gráfico y, por lo tanto, crean una brecha de cobertura que rompe un *contig*. Por lo tanto, la elección de k representa una compensación entre varios efectos (Chikhi & Medvedev, 2014).

1.2.4.3. Velvet

Velvet es un ensamblador genómico de *novo* especialmente diseñado para tecnologías de secuenciación de lectura corta, Velvet actualmente toma secuencias de lectura cortas, elimina errores y luego produce contigs únicos de alta calidad. Luego, utiliza la información de lectura final y lectura larga, cuando está disponible, para recuperar las áreas repetidas entre contigs (Zerbino & Birney, 2008).

1.2.4.4. A5

Esta herramienta ejecuta una tubería A5-miseq diseñada para ensamblar datos de secuencia de ADN generados en la plataforma de secuenciación Illumina. Hay muchas situaciones en las que A5-miseq no es la herramienta adecuada para el trabajo. Para producir resultados precisos, A5-miseq requiere datos de Illumina con ciertas características. A5-miseq probablemente no funcionará bien con lecturas de Illumina más cortas que alrededor de 80 nt, o lecturas donde las cualidades básicas son bajas en todas o la mayoría de las lecturas antes de 60 nt. Esta herramienta necesita un par de archivos de lectura de extremo emparejado con formato fastq (Coil *et al*., 2015).

1.2.4.5. Spades

Spades es un ensamblador basado en gráficos de Bruijn, que es notable por su enfoque en la aplicación de múltiples gráficos de Bruijn (cada uno construido con diferentes tamaños de k-mer) para manejar mejor las grandes variaciones en la cobertura, a través del genoma que son características de una sola celda de secuenciación, así como un método novedoso para manejar la información final emparejada. Comienza su proceso de ensamblaje mediante el uso de gráficos de Bruijn multisized para construir el gráfico de ensamblaje mientras detecta y elimina las lecturas quiméricas. A continuación, las distancias entre los k-mers se estiman para

mapear los bordes del gráfico de ensamblaje. Posteriormente, se construye un gráfico de ensamblaje emparejado y Spades genera un conjunto de secuencias de ADN contiguas denominadas *contigs* (Bankevich *et al*., 2012).

1.2.4.6. QUAST

QUAST es una herramienta de evaluación de calidad de ensambles de genomas. Esta herramienta puede evaluar ensamblajes tanto con un genoma de referencia como sin referencia. QUAST produce muchos informes, tablas de resumen y diagramas (Gurevich *et al*., 2013).

1.2.4.7. I-TASSER

El servidor I-TASSER es una plataforma en línea que implementa una diversidad de algoritmos para el modelado estructural de proteínas y predecir de funciones. Permite a los usuarios de generar automáticamente predicciones de modelos de alta calidad de la estructura 3D y la función biológica de las moléculas de proteínas a partir de secuencias de aminoácidos (Yang *et al*., 2015).

El servidor primero intenta recuperar las proteínas de plantilla de pliegues similares (o estructuras súper secundarias) de la biblioteca PDB por LOMETS (*Local Meta-Threading-Server*) que es un método de meta-servidor para la predicción de la estructura de la proteína.

En el segundo paso, los fragmentos continuos obtenidos de las plantillas PDB se vuelven a ensamblar en modelos completos mediante simulaciones Monte Carlo de intercambio de réplicas con las regiones no alineadas (principalmente bucles) de subprocesamiento construidas por el modelado ab initio. En los casos en que LOMETS no identifica una plantilla apropiada, I-TASSER construirá las estructuras completas mediante el modelado ab initio. SPICKER, que es un algoritmo de agrupamiento para identificar los modelos casi nativos de un conjunto de señuelos de estructura de proteínas, identifica los estados de baja energía libre agrupando los señuelos de simulación (Zhang & Skolnick, 2004).

En el tercer paso, la simulación del ensamblaje de fragmentos se realiza nuevamente a partir de los centroides del grupo SPICKER, donde las restricciones espaciales recogidas tanto de las plantillas LOMETS como de las estructuras PDB por TM-align se utilizan para guiar las simulaciones. El propósito de la segunda iteración es eliminar el choque estérico, así como refinar la topología global de los centroides del grupo. Los señuelos generados en las segundas simulaciones se agrupan y se seleccionan las estructuras de energía más bajas. REMO, otro algoritmo del servidor, obtiene los modelos atómicos completos finales, que construye los detalles atómicos de los señuelos I-TASSER seleccionados a través de la optimización de la red de enlace de hidrógeno (Roy *et al*., 2010).

Para predecir la función biológica de la proteína, el servidor I-TASSER compara los modelos 3D predichos con las proteínas en 3 bibliotecas independientes que consisten en proteínas de número de clasificación enzimática conocida (CE), ontología génica (GO) vocabulario y sitios de unión a ligandos. Los resultados finales de las predicciones de funciones se deducen del consenso de los principales emparejamientos estructurales con los puntajes de función calculados con base en el puntaje de confianza de los modelos estructurales I-TASSER, la similitud estructural entre el modelo y las plantillas según lo evaluado por el puntaje TM , y la secuencia identidad en las regiones estructuralmente alineadas (Zhang, 2008). Todo este proceso se puede observar en la figura N° 12.

El puntaje C es un puntaje de confianza para estimar la calidad de los modelos predichos por I-TASSER. Se calcula en función de la importancia de las alineaciones de plantillas de subprocesos y los parámetros de convergencia de las simulaciones de ensamblaje de estructuras. El puntaje C está típicamente en el rango de [-5,2], donde un puntaje C de mayor valor significa un modelo con una alta confianza y viceversa.

TM-score es una escala recientemente propuesta para medir la similitud estructural entre dos estructuras. El propósito de proponer TM-score es resolver el problema de RMSD que es sensible al error local. Debido a que RMSD es una distancia promedio de todos los pares de residuos en dos estructuras, un error local (por ejemplo, una desorientación de la cola) generará un gran valor de RMSD aunque la topología global es correcta. En TM-score, sin embargo, la pequeña distancia se pondera más fuerte que la gran distancia, lo que hace que la puntuación sea insensible al error de modelado local. Una puntuación TM> 0.5 indica un modelo de topología correcta y una puntuación TM <0.17 significa una similitud aleatoria. Este límite no depende de la longitud de la proteína (Yang *et al*., 2015).

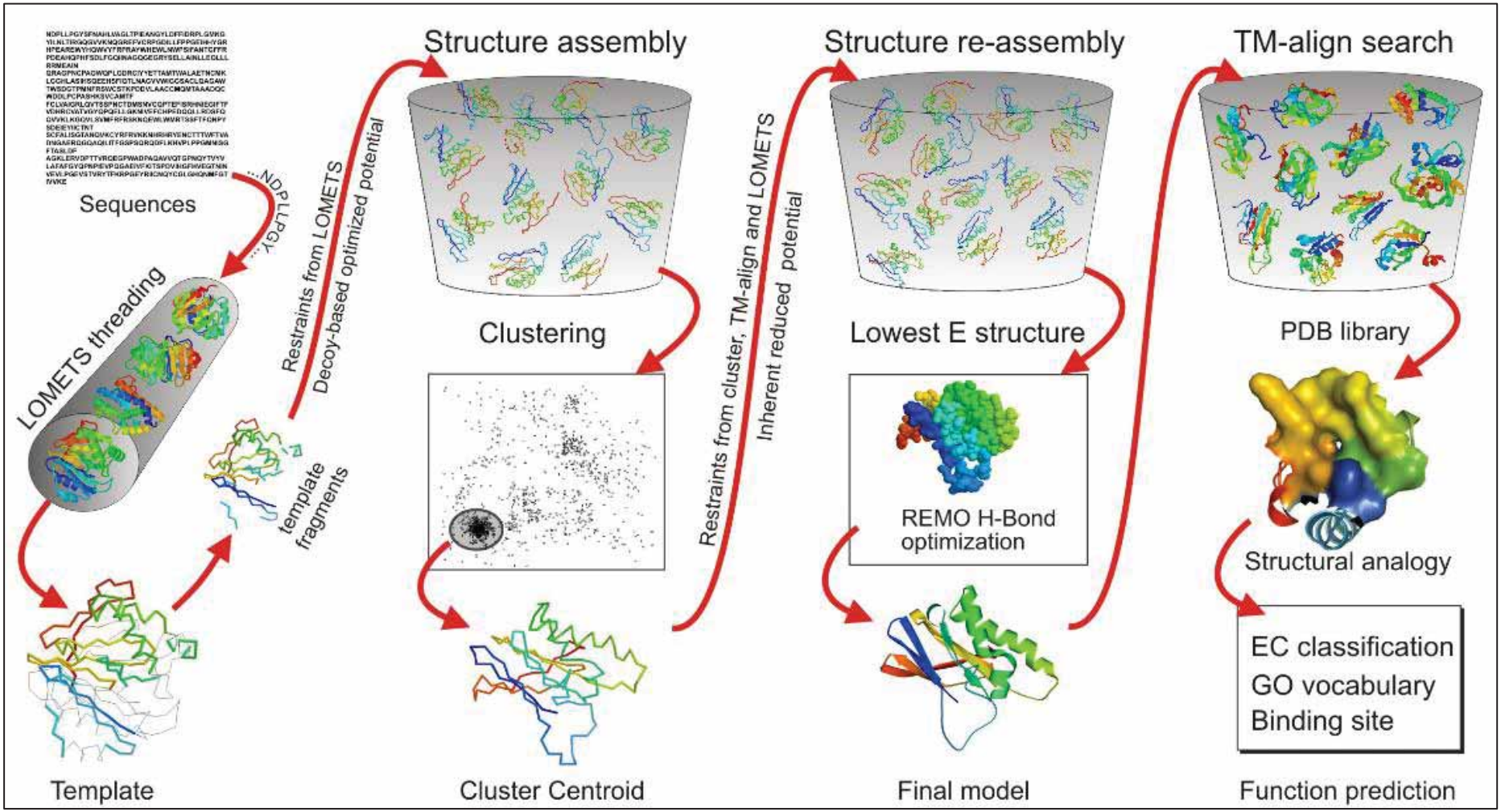


Figura N° 12: Protocolo I-TASSER para la estructura de proteínas y la predicción de funciones.
Fuente Yang et al., 2015.

1.2.4.8. UCSF Chimera.

UCSF Chimera es un programa altamente extensible para la visualización interactiva y el análisis de estructuras moleculares y datos relacionados, incluidos mapas de densidad, ensamblajes supramoleculares, alineaciones de secuencias, resultados de acoplamiento, trayectorias y conjuntos conformacionales. Se pueden generar imágenes y animaciones de alta calidad (Pettersen *et al*., 2004).

1.2.5. Bacteria Chromohalobacter salexigens

Chromobacterium marismortui que fue aislado del mar muerto y presentado en la tesis Doctoral de Elazari-Volcani en 1940. El género *Chromobacterium* paso a formar parte del género *Chromohalobacter* por la descripción realizada por Ventosa *et al.* en 1989. *Chromohalobacter marismortui* fue descrito en el manual de Berge en las 7th y 8th ed. Ventosa caracterizo las cepas de *C. marismortui* como varillas Gram negativas, de 0.6 a 1.0 por 1.5 a 4.0 μm. Las células fueron móviles por la presencia de flagelos peritrícos. No observaron la producción de esporas. Las características de las colonias fueron: circular, liso, entero y pigmentado característicamente con centros marrones oscuros seguidos de marrón azulado, grisáceo anillos marrones y amarillos. Produjeron un pigmento amarillo y un pigmento azul violeta que no era violaceína. Tuvo crecimiento en medio líquido con 10% de NaCl, todas las cepas produjeron turbidez y películas y generalmente marrones. Todas las cepas crecieron en medios sólidos que contenían del 2 al 30% de NaCl y creció óptimamente en medios que contenían aproximadamente 10% de NaCl y a una temperaturas entre 5 y 45 ° C con un rango de pH de 5 a 10. Las cepas de *C. marismortui* produjeron ácido a partir de glucosa en el medio de oxidación-fermentación de Hugh y Leifson que tenía NaCl al 8%, en medio marino de oxidación-fermentación al 10% de NaCl y en medio estándar al 10% de NaCl. Todas las cepas fueron positivas en tubos aeróbicos y negativas en condiciones anaerobias. Su contenido de G + C obtenido fue de 62.1 a 64.9 % (Ventosa *et al*., 1989).

La especie *Chromohalobacter salexigens* es una bacteria Gram negativa de 0.7-1.0x2.0-3.0 μm no formadora de esporas. Las células aparecen en pares y solas presentando motilidad. Las colonias son cremas, opacas, circulares y de menos de 2 mm de diámetro. En medio líquido produce una turbidez homogénea. Su crecimiento es posible cuando el medio presenta de 0.9 a 25 % de NaCl, a un pH de 5 a 10 entre 15 a 45 °C siendo su temperatura óptima de 37 °C. Es un aerobio estricto. Catalasa y oxidasa positiva. Citrato positivo. Reduce nitrato a nitrito. No hidroliza gelatina, almidón, esculina, ADN y Tween 80. Hidroliza caseína. Así también utiliza un gran número de compuestos carbonados como fuente de energía. Los siguientes

componentes son utilizados como única fuente de carbono, nitrógeno y energía: L-arginina, asparagina, betaina, glicina, L-glutamina, L-lisina, L-ornitina, L-prolina y L-serina. Su contenido de G + C en el ADN es de 64.2 % (Arahal et al., 2001).

1.2.6. Enzimas.

Las enzimas son moléculas orgánicas que actúan como catalizadores de reacciones químicas, es decir, aceleran la velocidad de reacción. Comúnmente son de naturaleza proteica, pero también ribonuleica (Walter & Engelke, 2002).

1.2.6.1. Clasificación enzimática

Existe un sistema internacional para la nomenclatura y clasificación de las enzimas creado por la *Enzyme Commission* (EC) de la *International Union of Biochemistry and Molecular Biology* (IUBMB) que evita imprecisiones y ambigüedades. Según este sistema, las enzimas pueden clasificarse en varios grupos, según el tipo de reacciones que catalicen (Feduchi et al., 2010) (Tabla N° 2).

Tabla N° 2: Clasificación de las enzimas según la IUBMB.

Clasificación	Tipos de reacción catalizada
EC. 1. Oxirreductasas	Reacciones de óxido-reducción
EC. 2. Transferasas	Transferencia de grupos funcionales
EC. 3. Hidrolasas	Reacciones de hidrolisis
EC. 4. Liasas	Eliminación de grupo para formar enlaces dobles
EC. 5. Isomerasas	Isomerización
EC. 6. Ligasas	Formación de enlace acoplada con la hidrolisis de ATP.

Feduchi et al., 2010

1.2.6.1.1. Hidrolasas

Esta clasificación alberga a un gran número de enzimas tales como peptidasas o también llamadas proteasas, nucleasas, lipasas, fosfatasas, glucosidasas entre otras. Esta denominación depende del tipo de enlace en que actúa (McDonald, 1985).

1.2.6.2. Proteasas

Las proteasas son la única clase de enzimas que ocupan una posición central con respecto a sus aplicaciones en campos tanto fisiológicos como comerciales. Las enzimas proteolíticas catalizan la escisión de enlaces peptídicos en otras proteínas. Las proteasas son enzimas degradantes que catalizan la hidrólisis total de las proteínas. Los avances en las técnicas analíticas han demostrado que las proteasas realizan modificaciones altamente

específicas y selectivas de proteínas tales como la activación de formas zimogénicas de enzimas por proteólisis limitada, coagulación de la sangre y lisis de coágulos de fibrina, y procesamiento y transporte de proteínas secretoras a través de las membranas (Godfrey & West, 1996).

1.2.6.3. Proteasas bacterianas

La incapacidad de las proteasas de plantas y animales para satisfacer las demandas mundiales actuales ha llevado a un mayor interés en las proteasas microbianas. Los microorganismos representan una excelente fuente de enzimas debido a su amplia diversidad bioquímica y su susceptibilidad a la manipulación genética. Las proteasas microbianas representan aproximadamente el 40% de las ventas mundiales totales de enzimas(Godfrey & West, 1996). Las proteasas de fuentes microbianas son preferidas a las enzimas de fuentes vegetales y animales ya que poseen casi todas las características deseadas para sus aplicaciones biotecnológicas.

1.2.6.4. Clasificación de las proteasas

Actualmente las proteasas son clasificadas en base a tres criterios principales: el tipo de reacción catalizada, la naturaleza química del sitio catalítico y la relación evolutiva.

Las proteasas se dividen en dos grupos principales: exopeptidasas y endopeptidasas, dependiendo del sitio de acción. Las exopeptidasas cortan el enlace peptídico en los extremos amino o carboxi-terminal, liberando un solo aminoácido, mientras que las endopeptidasas cortan las proteínas en enlaces peptídicos internos de la molécula. Basándose en el grupo funcional del sitio activo se clasifican en cuatro grupos: serín-proteasas, aspártico-proteasas, cisteín-proteasas y las metaloproteasas. Dentro de estos cuatro grupos se clasifican en familias dependiendo de su secuencia de aminoácidos. Hay un grupo que no entraría en la clasificación estándar, son las proteasas que necesitan ATP para su actividad (Barrett, 1994).

Exopeptidasas. Las aminopeptidasas suelen ser enzimas intracelulares y generalmente cortan el aminoácido N-terminal de proteínas inmaduras. Están ampliamente extendidas en bacterias y hongos. Las carboxipeptidasas se clasifican en serincarboxipeptidasas, cisteincarboxipeptidasas y metalocarboxipeptidasas, basándose en la naturaleza de los aminoácidos del sitio activo. Se han aislado metalocarboxipeptidasas en *Pseudomonas spp*., que requieren Zn^{2+} o Co^{2+} para su actividad.

Endopeptidasas. Se clasifican en serín-proteasas, cisteín-proteasas, aspártico-proteasas y metaloproteasas presentes en un número considerable de bacterias (Rao *et al*., 1998) (Tabla N° 3).

❖ Serín proteasas: Las serín proteasas se caracterizan por la presencia de un grupo serina en su sitio activo. Son numerosos y extendido entre virus, bacterias y eucariotas, sugiriendo que son vitales para los organismos. Serina proteasas se encuentran en la exopeptidasa, endopeptidasa, oligopeptidasa, y grupos omega peptidasa. Basado en su similitudes estructurales, las serina proteasas se han agrupado en 20 familias, que se han subdividido en alrededor de seis clanes con ancestros comunes (Barett, 1994).

✓ Serín-proteasas alcalinas: Con un pH óptimo alrededor de 10 y un tamaño entre 15 y 30 kDa. *Arthrobacter, Streptomyces y Flovobacterium* spp. las sintetizan.

✓ Subtilisinas. Sintetizadas en su mayoría por el género bacteriano *Bacillus*. Con pH óptimo de 10 y temperatura óptima de actividad de 60 °C. Hay dos subtipos: Calsberg (utilizada en detergentes) y Nagase o BPN, sintetizadas por *B. liquenifornis* y *B. amyloliquefaciens*, respectivamente.

❖ Aspártico-proteasas o proteasas ácidas: pH óptimo entre 3-4 y masa molecular entre 30 y 45 KDa. El sitio activo está formado por Asp-Ser/Thr-Gly (Aspartato-Serina/Treonina-Glicina). Son sintetizadas por los mohos como *Aspergillus, Pinicillum, Rhizopus* y *Nuerospora.*

❖ Cisteín-proteasas. Se encuentran tanto en procariotas como en eucariotas. Se han caracterizado, por ejemplo, las de *Clostridium histolyticum y Streptococcus spp.*

❖ Metaloproteasas. Son las proteasas más diversas. Se caracterizan porque requieren un ion metálico divalente para su actividad. Se clasifican en neutras, alcalinas, de *Myxobacter* I y de *Myxobacter* II. Las neutras tienen especificidad por los aminoácidos hidrófobos mientras que las alcalinas muestran gran variedad de dianas. Todas son inhibidas por agentes quelantes como el EDTA.

✓ La termolisina es una proteína neutra, sintetizada por *B. stearothermophilus*, con un átomo de Zn^{2+} y cuatro de Ca^{2+} que le confieren la termorresistencia. Es muy estable a altas temperaturas (resiste una hora a 80 °C).

✓ La colagenasa de *Clostridium histolyticum* o la elastasa de *Pseudomonas aeruginosa* y *Serratia spp.* son activas entre pH 7 y 9, y ocupan un rango de 48-60 kDa de peso molecular (Rao *et al*., 1998).

Como se ha citado, las proteasas más comúnmente encontradas en hongos son cisteín-proteasas y proteasas acidas, mientras que en bacterias están más entendidas las proteasas alcalinas, las neutras y las serín-proteasas (Kalisz, 1988).

Tabla N° 3: Principales endopeptidasas bacterianas.

Tipo Catalítico	Subgrupo	Subclase	Grupos bacterianos que incluyen productores de proteasas (con secuencias de genes en NCBI)	Especies bacterianas
Serín proteasas	Serín-proteasas alcalinas; proteasas tipo subtilisina		Grupo CFB (*Cythophaga, Flavobacterium, Bacteroidetes*); alfa y gamma-Proteobacteria; *Bacillus*/ grupo *Clostridium; Streptomyces spp.*	-
	Subtilisinas	Carlsberg, BPN	*Bacillus spp*.	*B. licheniformis, B. amyloliquefaciens, B subtilis.*
Cisteín-proteasas			*Porphyromonas gingivalis; Streptococcus pyogenes.*	-
Metaloproteasas	Neutral	Tipo termolisin	*Bacillus; Streptococcus spp.; Bacteriodes; Erwinia chrysanthemi; Serratia marcescens; Vibrio spp.; B brevis B subtilis*	*B megaterium, Paenibacillus polymyxa, Lactobacillus sp., B stearothermophilus, B caldolyticus, B cereus, B thruringiensis, Alicyclobacillus acidocaldarius, Staphilococcus epidermis, B thermoproteolyticus, B amyloliquefaciens, B brevis, Clostridium perfringens, Listeria monocytogenes*
		Colagenasa	Grupo CFB; *Clostridium spp*.; *Vibrio spp.; B cereus*.	-
		Elastasa	*P aeruginosa; Aeromonas spp., Vibrio spp.*	-
	Alcalino		*Gamma-Proteobacteria*	*Pseudomonas fluorescens, P tolaasii, P aeruginosa, Erwinia chrysanthemi, Serratia marcescens, Serratia sp.*

Fuente: Bach et al., 2001; adaptado de Rao et al., 1998.

1.2.7. Peptidasa M18 aspartil aminopeptidasa.

La familia de la peptidasa M18, la subfamilia de aspartil aminopeptidasa (DAP; EC 3.4.11.21), está ampliamente distribuida en bacterias y eucariotas. La DAP corta solamente los residuos de aminoácidos ácidos N-terminales no bloqueados. Es una enzima citosólica y está altamente conservada; por ejemplo, la enzima humana tiene una identidad del 51% con una proteína similar a la aspartil aminopeptidasa en *Arabidopsis thaliana* (Park *et al*., 2017). La DAP de mamífero es altamente selectiva para la hidrólisis de los residuos de aspartato o glutamato N-terminales de los péptidos. A diferencia de la glutamil aminopeptidasa (M42), la

DAP no corta sustratos de aminoaril-arilamida simples. Aunque no se comprende la función de esta enzima, se piensa que actúa en concierto con otras aminopeptidasas para facilitar el recambio de proteínas debido a sus especificidades restringidas para el ácido aspártico y glutámico N-terminal, que no puede ser escindido por ninguna otra aminopeptidasa. La aspartil aminopeptidasa de mamífero posiblemente contribuye al catabolismo de los péptidos, incluidos los producidos por el proteasoma. También puede recortar el extremo N de los péptidos destinados al sistema de clase I del complejo mayor de histocompatibilidad (MHC). En los seres humanos, la DAP se ha implicado en la función específica de convertir la angiotensina II en la angiotensina vasoactiva III dentro del cerebro. La aminopeptidasa I de *Saccharomyces cerevisiae* (Ape1) participa en la degradación de las proteínas en las vacuolas (los lisosomas de levadura) donde se transporta por la vía única de focalización del citoplasma a la vacuola (Cvt) en condiciones de crecimiento vegetativo y por vía autofágica durante la inanición. Su región propeptídica N-terminal, que media en la formación de complejos de orden superior, sirve como una carga de andamiaje crítica para el ensamblaje de la vesícula Cvt para el suministro de vacuolar. La aminopeptidasa de *Pseudomonas aeruginosa* (PaAP) muestra que su actividad depende de Co^{2+} en lugar de Zn^{2+}a, y por lo tanto es una peptidasa cocatalítica de cobalto en lugar de una peptidasa dependiente de zinc (Natarajan & Mathews, 2012).

1.2.8. Extracción de ADN.

La aplicación de técnicas moleculares inicia con la extracción de ADN y la obtención exitosa de datos confiables y reproducibles depende, en gran medida, de la extracción de ADN íntegro y puro. La extracción consiste en el aislamiento y purificación de moléculas de ADN y se basa en las características fisicoquímicas de la molécula. Los protocolos tradicionales consisten de cinco etapas principales: homogeneización del tejido, lisis celular, separación de proteínas y lípidos, precipitación y redisolución del ADN (Alejos *et al*., s. f.).

1.2.8.1. Cuantificación de ADN.

Una vez obtenido el material genético, es importante determinar el rendimiento mediante espectrofotometría. La ley de Beer-Lambert indica que la concentración de una molécula en solución depende de la cantidad de luz absorbida de las moléculas disueltas. Una característica del ADN es que absorbe la luz ultravioleta (UV) a 260 nm y permite estimar su concentración mediante espectrofotometría. Cuando la longitud de la celda en que se disuelve el ADN, es de 1 cm, la absorbancia es igual a la densidad óptica (DO). En el caso de ADN genómico o de doble cadena, una densidad óptica equivale a 50 ug/ml (Leninger, 1975). Se debe considerar el factor de dilución para obtener la concentración en nanogramos/microlitro

(ng/ul). En el caso de algunos equipos como el espectrofotómetro Nanodrop, no es necesario diluir la muestra y el equipo nos proporciona directamente la concentración en ng/ul. Considerando que para una reacción de PCR se requieren de 10 a 200 ng debemos obtener al menos, 5 ng/ul de ADN en cada muestra, si se recuperaron 50 ul, el rendimiento neto de la extracción sería de 0.25 ug. Concentraciones menores dificultan la estandarización de la PCR u otras técnicas. La absorbancia a 260 nm se utiliza para calcular la concentración de ácidos nucleicos. Para estimar la pureza del ADN se considera la proporción de la absorbancia a 260 nm y 280 nm. Una proporción de 1.8 es aceptada como ADN puro, proporciones menores a este valor indican la presencia de proteínas. Una segunda valoración de la pureza de ácidos nucleicos es la proporción 260/230, los valores aceptados se encuentran en el rango de 2.0 a 2.2, si la relación es menor indican la presencia de contaminantes como carbohidratos o fenol (Alejos *et al*., s. f.).

1.2.9. Amplificación del ADN mediante PCR convencional.

La reacción en cadena de la polimerasa, conocida como PCR por sus siglas en inglés (*polymerase chain reaction*), es una técnica de biología molecular desarrollada en 1986 por Kary Mullis. Su objetivo fue obtener un gran número de copias de un fragmento de ADN particular, partiendo de un mínimo; partiendo de una única copia de ese fragmento original, o molde. Esta técnica sirve para amplificar un fragmento de ADN; su utilidad es que tras la amplificación resulta mucho más fácil identificar el fragmento de ADN con una muy alta probabilidad (Mullis *et al*., 1986). Esta técnica se basa en tres etapas (Figura N° 13):

1. Desnaturalización: El ADN bicatenario se desnaturaliza en cadenas simples. El ADN puede provenir de muchas fuentes, incluido el ADN genómico. Calentar a 92-95 ° C durante aproximadamente 1 minuto desnaturaliza el ADN bicatenario en cadenas simples.

2. Hibridación: La temperatura de la reacción se reduce a una temperatura entre 45 ° C y 65 ° C, lo que provoca la unión del cebador, también llamado hibridación o recocido, al ADN desnaturalizado y monocatenario. Los cebadores sirven como puntos de partida para que la ADN polimerasa sintetice nuevas cadenas de ADN complementarias al ADN objetivo. Factores como la longitud del cebador, la composición de la base de los cebadores, y si todas las bases en un cebador son complementarias o no de las bases en la secuencia objetivo, están entre las consideraciones principales al seleccionar una temperatura de hibridación para un experimento.

3. Extensión: la temperatura de reacción se ajusta entre 65° C y 75° C, y la ADN polimerasa usa los cebadores como punto de partida para sintetizar nuevas cadenas de ADN al agregar nucleótidos a los extremos de los cebadores en dirección 5’ a 3’ (Klug *et al*., 2016).

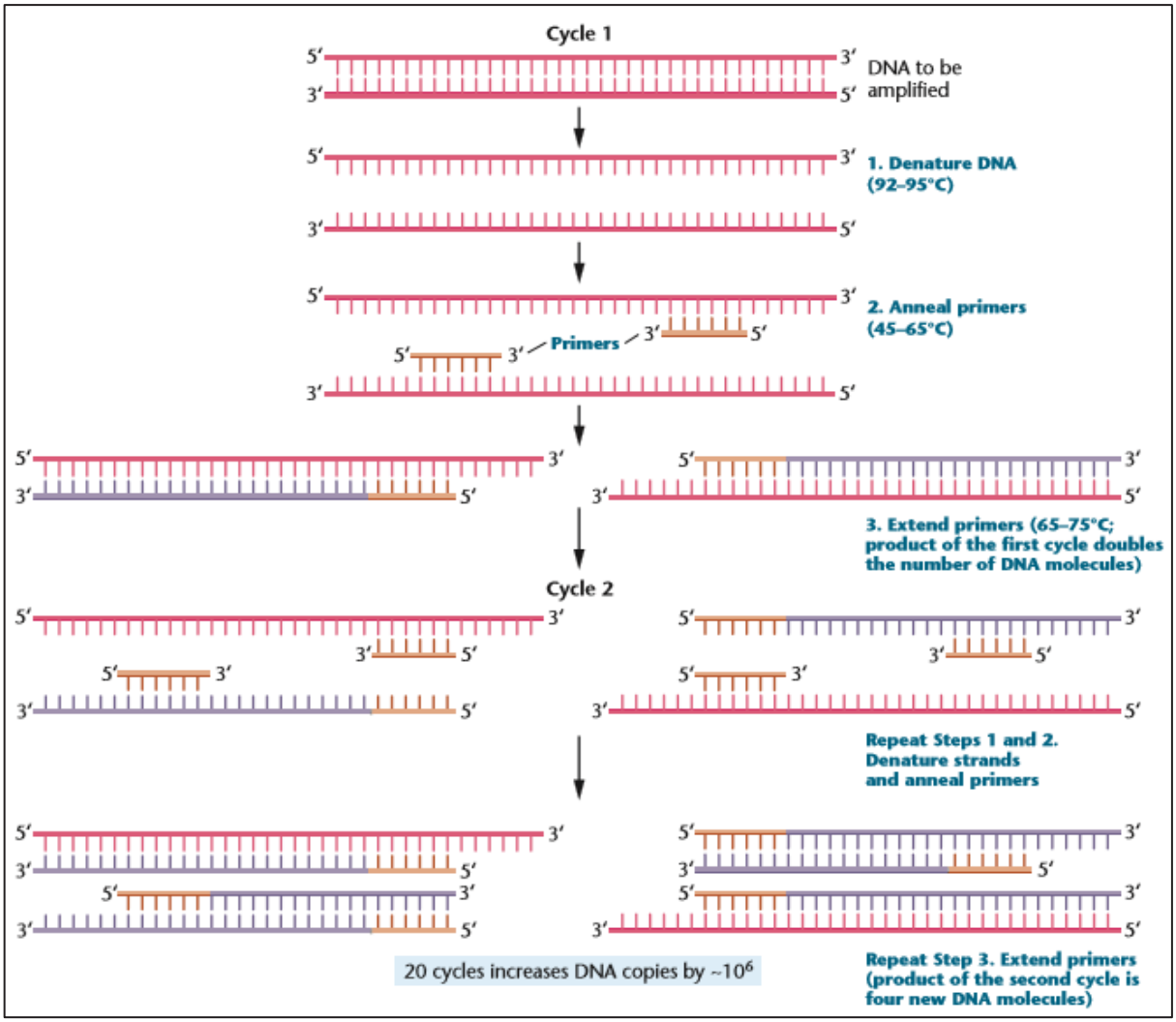


Figura N° 13: PCR con sus etapas detalladas.

1.2.10. Electroforesis

La visualización del resultado de la Reacción en Cadena de la Polimerasa se realiza normalmente mediante electroforesis. Dependiendo del tamaño de la amplificación y de la resolución que se desee, se utilizan diferentes geles (agarosa, poliacrilamida) a distintas concentraciones. La posterior observación se puede realizar con bromuro de etidio (lámpara de luz UV), tinción de plata, fluorescencia, radioactividad (Mas *et al*., 2016)

1.2.11. Tecnología del ADN recombinante

La era del ADN recombinante comenzó en la década de los ochenta cuando las investigaciones por entonces, describieron el aislamiento de una enzima bacteriana y el uso de

esta enzima para cortar el ADN vírico por secuencias de nucleótidos específicos (Danna & Nathans, 1971) a lo que más adelante se llamarían enzimas de restricción.

Utilizando enzimas de restricción y una serie de otros recursos, los investigadores de mediados y finales de la década de 1970 desarrollaron varias técnicas para crear, replicar y analizar moléculas de ADN recombinante, ADN creado al unir piezas de ADN de diferentes fuentes. Los métodos utilizados para copiar o clonar ADN, llamados tecnología de ADN recombinante y a menudo conocido como "*gene splicing*" en sus comienzos, marcaron un gran avance en la investigación en biología molecular y genética, permitiendo a los científicos aislar y estudiar secuencias específicas de ADN (Klug *et al*., 2016).

1.2.12. Clonaje

La clonación se puede definir como el proceso para conseguir, la producción masiva de copias idénticas de una molécula, célula u organismo. Los plásmidos bacterianos genéticamente modificados fueron los primeros vectores desarrollados, y todavía se usan ampliamente para la clonación. Los vectores de clonación plásmidicos son moléculas de ADN bicatenarias extracromosómicas que se replican independientemente de los cromosomas dentro de las células bacterianas. Los plásmidos han sido ampliamente modificados por ingeniería genética para servir como vectores de clonación. Muchos plásmidos preparados comercialmente están fácilmente disponibles con una gama de características útiles. Los plásmidos se introducen en las bacterias mediante el proceso de transformación. Dos técnicas principales son ampliamente utilizadas para la transformación bacteriana. Solo uno o unos pocos plásmidos generalmente ingresan a una célula huésped bacteriana por transformación. Debido a que los plásmidos tienen un origen de replicación (ori) que permite la replicación de plásmidos, muchos plásmidos pueden aumentar su número de copias para producir varios cientos de copias en una sola célula huésped. Estos plásmidos mejoran en gran medida la cantidad de clones de ADN que se pueden producir. Los vectores plasmídicos también se han modificado genéticamente para contener una serie de sitios de restricción para enzimas de restricción de uso común en una región llamada sitio de clonación múltiple. La clonación del ADN con un plásmido generalmente comienza cortando tanto el ADN del plásmido como el ADN que se va a clonar con la misma enzima de restricción. Típicamente, el plásmido se corta una vez dentro del sitio de clonación múltiple para producir un vector lineal. Los fragmentos de restricción de ADN del ADN que se va a clonar se añaden al vector linealizado en presencia de ADN ligasa. Los extremos adhesivos de los fragmentos de ADN se unen, uniéndose al ADN que se va a clonar y al plásmido. Luego, la ADN ligasa se usa para crear enlaces fosfodiéster

para sellar mellas en el esqueleto del ADN, produciendo así ADN recombinante, que luego se introduce en las células huésped bacterianas por transformación. Una vez dentro de la célula, los plásmidos se replican rápidamente para producir múltiples copias (Klug *et al*., 2016).

1.2.13. Transformación de células

La transformación se define como la transferencia de información genética a una bacteria receptora utilizando ADN desnudo, sin ningún requisito de contacto con una bacteria donante. La capacidad de transformar o aceptar ADN exógeno generalmente se conoce como competencia, aunque el término se ha utilizado tan ampliamente en diferentes sistemas que es difícil generar una definición de competencia que lo incluya todo. La competencia natural ocurre en un subconjunto definido de especies bacterianas que tienen la capacidad de absorber ADN lineal, y a veces circular, generalmente dependiente de un sistema de absorción específico. Como la competencia natural se limita a un subconjunto de bacterias, los métodos para la inducción química de un estado competente en bacterias no transformables son una herramienta importante en genética bacteriana. Para estas especies, la competencia se refiere a la capacidad de captar y propagar ADN plasmídico, generalmente sin especificidad de secuencia para la captación (Swords, 2003).

1.2.14. PCR de colonias

La PCR de colonias es la amplificación por PCR directa de secuencias diana de células sin aislamiento previo o purificación de ADN. La PCR de colonias es posible si suficientes células se lisan como consecuencia de la alta temperatura en el paso inicial de desnaturalización del molde solo o en combinación con procedimientos adicionales para hacer que el ADN sea más accesible. El material que contiene las células o las propias células tampoco deben presentar inhibición de PCR en un grado que impida la amplificación de PCR. La ventaja de la PCR de colonias sobre el uso de ADN purificado es principalmente la velocidad y el costo, ya que se omite el paso de extracción de ADN que lleva mucho tiempo; pero omitir la purificación de ADN y, por lo tanto, minimizar el manejo de la muestra también puede aumentar la sensibilidad si el material de partida es limitante, como podría ser, por ejemplo, en aplicaciones forenses. La necesidad de detectar la presencia o ausencia de secuencias específicas dentro de las células se necesita habitualmente en una amplia gama de disciplinas, como la microbiología clínica, la ingeniería genética y las ciencias forenses. La aplicación más común de PCR de colonias en ingeniería genética es probablemente la amplificación de secuencias de productos de ligadura dentro de los transformantes de *Escherichia coli* después de cortar y pegar la clonación (Azevedo *et al*., 2017).

CAPITULO II
MATERIALES Y MÉTODOS

2.1. Lugar de ejecución

La etapa Bioinformática de este trabajo se realizó en el laboratorio de transferencia tecnológica C-234 de la escuela profesional de Biología de la Universidad de San Antonio Abab del Cusco UNSAAC desde el febrero del 2018 hasta junio del 2019 y la etapa de clonación se ejecutó en los laboratorios número 9 y 10 del Instituto Universitario de enfermedades tropicales y salud pública de Canarias de la Universidad de La Laguna desde el mes de octubre hasta diciembre del 2018.

2.2. Materiales

2.2.1. Datos de secuenciación genómica

Son los archivos *paired-end* en formato FASTQ producto de la secuenciación NGS de *C. salexigens* los cuales presentan 594952 lecturas, cada una de estas con longitud entre 31 a 301 pares de bases.

- ✓ Archivo FASTQ *forward*. Nombre pe_1_S15_L001_R1_001.fastq
- ✓ Archivo FASTQ *reverse*. Nombre pe_1_S15_L001_R2_001.fastq

2.2.2. Material biológico

- ✓ Cepa de *Chromohalobacter salexigens* MP25462.
- ✓ Plásmidos pGEM-T (Promega).
- ✓ Cepa de *E. coli* JM109 (Promega).

2.2.3. Material de laboratorio

- ✓ Mechero de alcohol.
- ✓ Encendedor.
- ✓ Placas Petri de 100 mm x 20 mm.
- ✓ Probetas graduadas de 50 ml, 100 ml, y 1000 ml.
- ✓ Vasos de precipitados de 50 ml, 100 ml y 500 ml.
- ✓ Matraces de 50 ml y 500 ml.
- ✓ Puntas para micropipetas de 1-20 µl, 1-200 µl, 100-1000 µl.
- ✓ Puntas para micropipetas con filtro de 1-20 µl, 1-200 µl, 100-1000 µl.
- ✓ Asas de siembra descartables.
- ✓ Papel toalla.
- ✓ Jeringas de 10 ml.

- ✓ Filtros de jeringa de 0.22 µm.
- ✓ Tubos de PCR de 0.2 ml
- ✓ Tubos de microcentrífuga de 1.5 ml.
- ✓ Tubos Falcon de 15 ml 50 ml.
- ✓ Tubos de criopreservación de 1.8 ml.
- ✓ Gradillas para tubos de 1.5 ml.
- ✓ Materiales de protección personal (barbijo, mandil, guantes).
- ✓ Mondadientes.

2.2.4. Equipos

- ✓ Multi-Rotator PTR-35 (Grant-bio)
- ✓ Microcentrífuga Micro Star 17 (VWR).
- ✓ Microcentrífuga 5418 (Eppendorf)
- ✓ Baño maria (Univeba)
- ✓ Lector de microplacas 680 (Bio-Rad).
- ✓ Lector de Imágenes ChemiDoc XRS+ (Bio-Rad).
- ✓ Vórtex (VWR).
- ✓ Centrifuga Megafuge 16R (Thermo Scientific).
- ✓ Cámara de flujo Laminar Bio-II-A (TelStar).
- ✓ Congeladora (Corbero).
- ✓ Microondas (Beko).
- ✓ Thermo Mixer C (Eppendorf).
- ✓ Cámara de electroforesis (Bio-Rad)
- ✓ Termociclador T Professional Basic (Biometra).
- ✓ Termociclador T Professional Gradient (Biometra).
- ✓ Termociclador My Cycler (Bio-Rad).
- ✓ Fuente de energía de Electroforesis (Hoefer).
- ✓ Cubeta de electroforesis Sub-Cell GT Basic (Bio-Rad).
- ✓ Fuente de energía de Electroforesis (Bio-Rad).
- ✓ Cubeta de electroforesis Wide Mini-Sub Cell GT Basic (Bio-Rad).
- ✓ Micropipeta de 2-20 µl Corning (Lambda).
- ✓ Balanza electrónica JT-1200 (Cobos).
- ✓ Fotodocumentador de imágenes (Hoefer).
- ✓ Agitador magnético (Bunsen).

- ✓ Espectrofotómetro DS-11 + (DeNovix).
- ✓ Congelador de temperaturas ultra bajas (Innova).
- ✓ Autoclave S100 (Matachana).
- ✓ Autoclave (Raypa).
- ✓ Destilador de agua MiliQ Paranity TU (VWR).
- ✓ Micropipetas de 0.5-10 µl, 10-100 µl y 100-1000 µl (Eppendorf).

2.2.5. Reactivos

- ✓ PureYield™ Plasmid Miniprep System (Promega).
- ✓ QIAEX II Gel Extraction Kit (Qiagen).
- ✓ E.Z.N.A. Gel Extraction Kit (VWR).
- ✓ E.Z.N.A Plasmid DNA Mini Kit I (VWR)
- ✓ MicroElute Cycle-Pure Kit (VWR).
- ✓ Ilustra™ Tissue & Cells genomic Prep Mini Spin Kit (GE Healthcare).
- ✓ Agarosa de grado molecular (BIOLINE).
- ✓ Marcador molecular Lambda Hind III (BIOLINE).
- ✓ Marcador molecular Hyper Ladder II (BIOLINE).
- ✓ Marcador molecular 100bp DNA Step Ladder (Promega).
- ✓ T4 DNA ligasa (Thermo Scientific).
- ✓ Tampón de reacción 2X (Thermo Scientific).

2.2.6. Softwares

- ✓ Gene Runner (6.5.51 Beta).
- ✓ MEGA VII y X.
- ✓ Sistema operativo GNU/Linux (Versión 16.04LTS)
- ✓ FastQC (Versión 0.11.4)
- ✓ Trimmomatic (Versión 0.36)
- ✓ Kmergenie (Versión 1.7044)
- ✓ Velvet (Versión 1.2.10)
- ✓ Spades (Versión 3.13.1)
- ✓ A5 (Versión 1.2.10)
- ✓ Quast (Versión 4.6.1)

2.2.7. Base de datos

- ✓ NCBI: Centro Nacional de Información Biotecnológica. https://www.ncbi.nlm.nih.gov/
- ✓ RAST: *Rapid Annotation using Subsystem Technology* http://rast.nmpdr.org/
- ✓ I-TASSER. Servidor en línea.

2.3. Métodos.

2.3.1. Tipo de investigación.

Descriptiva - experimental

2.3.2. Variables.

Tabla N° 4: Variables dependientes e independientes

	Caracterización *in silico*	**Clonaje**
7.1. Variables dependientes.	• Secuencias de comparación. • Ensambladores.	• Temperatura de amplificación. • Secuencia del gen aspartil aminopeptidasa.
7.2. Variables independientes.	• Datos NGS. • Programas bioinformáticos. • Genoma.	

2.3.3. Flujograma de la metodología de investigación.

2.3.3.1. Etapa 1: Caracterización *in silico* del gen DAP.

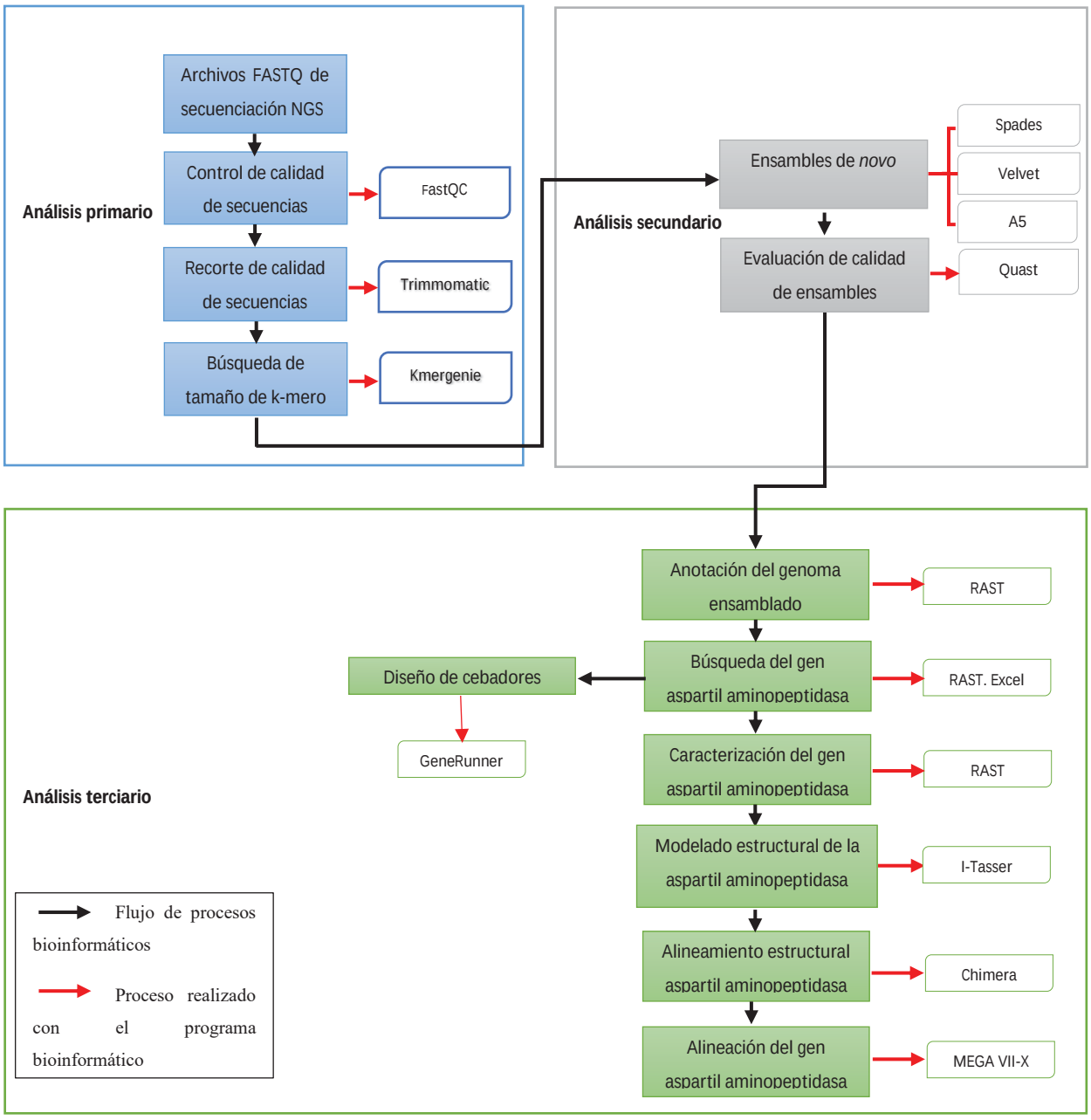


Figura N° 14: Flujograma de la primera etapa de caracterización in silico de la aspartil aminopeptidasa de Chromohalobacter salexigens MP25462

2.3.3.2. Etapa 2: Clonación del gen aspartil aminopeptidasa.

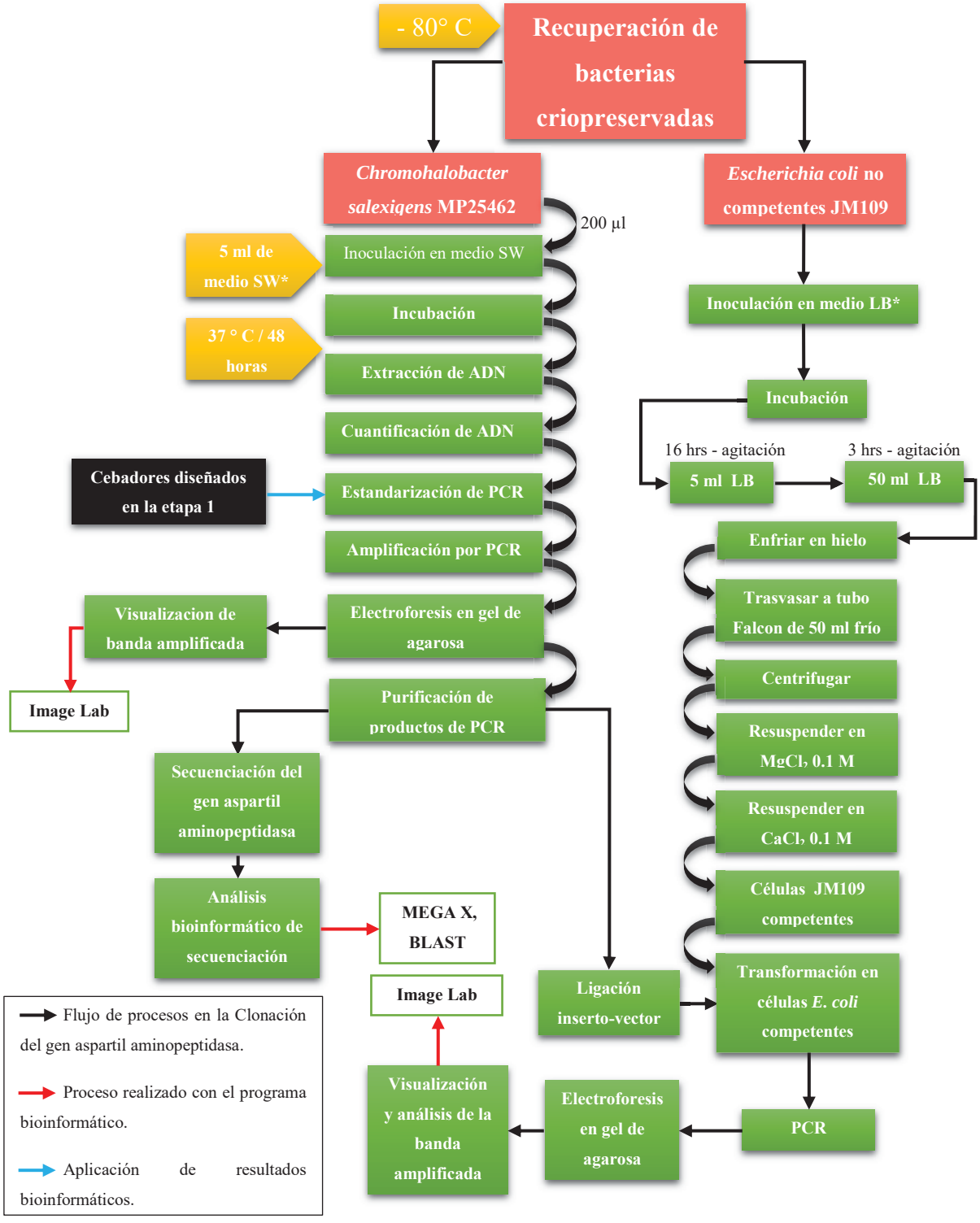


Figura N° 15: Flujograma de la segunda etapa: Clonación de la aspartil aminopeptidasa

2.3.4. Caracterización *in silico* del gen aspartil aminopeptidasa

Para conseguir la caracterización *in silico* del gen aspartil aminopeptidasa se realizó un análisis bioinformático primario, secundario y terciario de los datos de secuenciación masiva así como se observó en el apartado 2.3.3.1.

2.3.4.1. Análisis primario: Generación de secuencias

2.3.4.1.1. Control y recorte por calidades genómicas

El aislamiento de *C. salexigens* MP25462 y secuenciación NGS de su material genético se realizó dentro del proyecto "Secuenciación del metagenoma de ambientes salinos del departamento de Cusco". Es desde ahí que se inicia el análisis de todos los datos proporcionados de la secuenciación masiva NGS del genoma de *Chromohalobacter salexigens* MP25462 (Anexo 9). Este genoma atravesó por un control de la calidad, para lo cual se utilizó el programa FASTQC que fue ejecutado en la terminal del sistema operativo LINUX. Se seleccionó el archivo FASTQ *forward* pe_1_S15_L001_R1_001.fastq, terminado el análisis de calidad FASTQC otorgó 12 módulos con datos representativos de las secuencias NGS.

Luego de la observación del estado en que se encontró las lecturas de secuenciación se procedió a la limpieza o recorte de las mismas a través del programa TRIMMOMATIC con la comanda de ejecución en la terminal LINUX (Anexo 1.A). Para la visualización de las secuencias genómicas después del recorte, se volvió a utilizar el programa FASTQC, los informes del antes y después del recorte con tal programa fueron comparados.

2.3.4.1.2. Obtención del tamaño de K-mero

Las secuencias genómicas recortadas fueron introducidas en el programa KMERGENIE para determinar el tamaño de k-mero requerido para el ensamblaje del genoma de *Chromohalobacter salexigens* MP25462. El comando para la ejecución de este software se visualiza en el Anexo 1.B.

2.3.4.2. Análisis secundario: Procesado de secuenciación

2.3.4.2.1. Ensamble de *novo* y evaluación de parámetros

Para el acoplamiento o ensamblaje genómico se utilizó tres ensambladores de *novo*: **VELVET**, **A5** y **SPADES**. La ejecución del ensamblador **VELVET** se realizó con las secuencias recortadas y la información del tamaño de k-mer. Tal proceso se efectuó en dos fases: indexación y ensamble (Anexo 1.C).

El ensamblador de *novo* **A5** se ejecutó desde la terminal LINUX con los dos archivos de la secuenciación NGS (Anexo 1.D), de igual forma se realizó el proceso del ensamblaje genómico de *Chromohalobacter salexigens* MP25462 con el programa **SPADES** (Anexo 1.E).

Cada producto de los tres ensambladores mencionados fueron evaluados y comparados en cuanto a su calidad mediante el programa **QUAST** el cual se efectuó mediante las comandas que se muestran en el Anexo 1.F.

2.3.4.3. Análisis terciario: Interpretación de resultados.

2.3.4.3.1. Anotación del genoma ensamblado.

Las secuencias nucleotídicas ensambladas del genoma completo de *Chromohalobacter salexigens* MP25462 se ingresó al servidor RAST (http://rast.nmpdr.org/rast.cgi) para su anotación para lo cual se creó una cuenta de usuario en este anotador.

Con la cuenta creada se ingresó en el menú principal del servidor RAST y se eligió la opción *"Upload a new job"* para seleccionar la opción *"File formats"* donde se eligió el archivo del genoma ensamblado de *Chromohalobacter salexigens* MP25462 en formato FASTA. Para subir dicho archivo en la base de datos de RAST fue necesario agregar información sobre el genoma del microorganismo a anotar: identidad taxonómica, dominio, género, especie, cepa, etc.

La anotación dio inicio al escoger el tipo de esquema de anotación RAST, así como otras configuraciones ya establecidas por el servidor. RAST otorga un código de acceso al trabajo de anotación y normalmente otorga los resultados de anotación entre 24-48 horas posteriores a la presentación, pero en última instancia, la calidad de dicho servicio se evalúa en términos de precisión, consistencia e integridad de las anotaciones producidas (Aziz *et al*., 2008).

Terminado el proceso de anotación se comenzó la búsqueda de la secuencia del gen de la aspartil aminopeptidasa ubicando el código de anotación de *C. salexigenes* 6666666.448892 para luego seleccionar la opción *View details* en el apartado *Anotation Progress.* En la parte superior de estas características ya señaladas se ubican 5 accesos directos: *Browse annotated genome in SEED Viewer* en donde se encuentra los resultados de las características del genoma de *C. salexigens* tales como longitud del genoma, contenido de G + C, N50, L50, número de contigs, número de subsistemas, número de secuencias codificantes y número de ARNs. También produjo una gráfica circular de los subsistemas detallando información sobre esta así como su cobertura, su distribución categórica y funciones como motilidad, metabolismo de nitrógeno, nucleótidos, nucleótidos, respuesta al estrés, etc. También existen las opciones de

Browse y *Compare* en donde se realizó la búsqueda y la comparación de la vía metabólica de la aspartil aminopeptidasa respectivamente*; Available downloads for this job* donde se realizó la descarga de la anotación genómica de *C. salexigens* en formato Excel*; Share this genome with selected users* con el cual se puede compartir la anotación genómica y ensamblado tanto al servidor como usuarios; *View Close Strains for this job* donde se observó a las cepas cercanas a la estudiada y *Back to the Jobs Overview* que regresa a la ventana de progreso de trabajos. Como ya se detalló en el acceso directo *Browse annotated genome in SEED Viewer* existe la opción *Browse* que contiene la información completa del genoma anotado al cual se tuvo acceso. Dicha aplicación contiene una tabla con diferente opciones de búsqueda de genes, entre estas: el *contig* de ubicación del gen, ID de función, comienzo, longitud, subsistema y función. Se utilizó la opción de búsqueda *Función* donde se escribió *aspartyl aminopeptidase* dando como resultado su ubicación, código de identificación, descripción y visualización en el genoma. Para la obtención de la secuencia de nucleótidos y aminoácidos fue necesario ingresar al código de identificación del gen aspartil aminopeptidasa. Otra forma de búsqueda del gen aspartil aminopeptidasa fue mediante del archivo de anotación en formato Excel en el cual fue necesario realizar la búsqueda con la opción del panel de herramientas *Buscar y seleccionar*. La descarga de secuencias nucleotídicas y aminoacídicas de la aspartil aminopeptidasa se realizó en formato FASTA para su manejo en los siguientes métodos.

2.3.4.3.2. Descripción del gen aspartil aminopeptidasa

Las opciones del anotador RAST *Browse* y *Compare* fueron utilizadas para la descripción del gen aspartil aminopeptidasa haciendo uso de datos, visualizaciones y buscador genómico como se explicó en el apartado 2.3.4.3.1.

2.3.4.3.3. Descripción de la proteína aspartil aminopeptidasa.

2.3.4.3.3.1. Modelado de DAP MP25462.

Luego de haber realizado la búsqueda de la secuencia nucleotídica del gen que expresa la aspartil aminopeptidasa se comenzó con la descripción de la misma con el servidor I-TASSER (Iterative Threading ASSEmbly Refinement) http://zhang.bioinformatics.ku.edu/I-TASSER. Dicho servidor generó predicciones estructurales automatizadas de la proteína aspartil aminopeptidasa de longitud completa en 3D así como la ubicación del sitio activo.

Se subió la secuencias nucleotídicas expresadas en aminoácidos en un archivo formato FASTA en la opción "*Or upload the sequence from your local computer*", señalando la identificación o nombre de la proteína. Finalmente se seleccionó el botón "*Run I-TASSER*" para terminar de subir la secuencia de interés. El proceso de identificación mediante las

plantillas estructurales del PDB (Protein Data Base) demoró 72 horas debido a la carga de trabajo en el sistema del servidor.

2.3.4.3.3.2. Alineamiento estructural de DAP.

Con los datos obtenidos del modelado 3D de la proteína aspartil aminopeptidasa se procedió a realizar las simulaciones visuales del alineamiento estructural de la misma con el programa bioinformático Chimera 1.13.

El archivo en formato PDB de la proteína aspartil aminopeptidasa obtenido del modelado en I-TASSER fue ejecutado en la terminal Chimera donde automáticamente se visualizó la estructura tridimensional de DAP.

Se obtuvo datos del modelado de la proteína en estudio la cual tuvo alta similitud con una aspartil aminopeptidasa del PDB (*Protein Data Base*). Con tal dato se realizó la búsqueda de la proteína los datos del PDB y se procedió a descargar dicha información. Este archivo se ejecutó como se describió para la aspartil aminopeptidasa de *C. salexigens* MP25462, tomando en cuenta que debe ser realizado en la misma terminal.

Con las dos proteínas proyectadas en 3D se procedió a seleccionar la proteína bacteriana para diferenciarla de la descargada mediante el botón de la barra de herramientas *Select* opción *Chain* opción *A Model1DAP* para luego resaltarlo con un color que lo diferencie. Esto mediante el botón *Actions* opción *Color* y escogió el color deseado. Con las proteínas diferenciadas se realizó el alineamiento estructural de ambas mediante la selección del botón *Tools* opción *Structure Comparison* y luego *Match Maker*. Realizado este proceso apareció la ventana en donde se podrá visualizar en la parte superior izquierda los nombres de los dos archivos empleados. En la opción *Structure Reference* se seleccionó el archivo descargado de PDB mientras que en la opción *Structure to Match* la proteína bacteriana DAP de MP25462 y se ejecutó el alineamiento estructural.

Chimera otorga la posibilidad de visualizar la alineación estructural tal y como se presentan en las publicaciones científicas, para esto se seleccionó el botón *Presets* opción *Publication3*.

A partir de los datos del artículo de la aspartil aminopeptidasa cristalizada se comenzó a realizar la descripción de la proteína DAP de *C. salexigens* realizado manualmente.

2.3.4.3.3.3. Identificación del sitio activo de DAP

De la información obtenida en el apartado 2.3.4.3.3.1 del modelado con respecto al sitio activo, el servidor I-TASSER mediante su base de datos otorgo la predicción de funcionalidad de la proteína DAP de *C.salexigens* MP25462 realizando una comparación estructural con las

proteínas cristalizadas contenidas en PDB tomando en cuanto el C-score, tamaño de Cluster, y comparaciones de ligando, dichos índices fueron explicados en el apartado 1.2.4.7. Terminado el proceso del servidor otorgó el archivo PDB de la enzima cristalizada con mayor probabilidad de ubicación de sitio activo así como ligandos implicados, con lo cual se realizó la descarga del archivo en formato PDB.

El archivo de las proteína DAP y la predicha por I-TASSE fueron ejecutadas en Chimera tal como se explicó en el apartado 2.3.4.3.3.2 de alineamiento estructural de DAP.

Con el alineamiento realizado se procedió a ubicar los sitios de unión de ligando y sitio activo teniendo como referencia la información del artículo científico de la enzima cristalizada.

Al encontrar la similaridad predicha se procedió a resaltar los aminoácidos comprometidos en el sitio activo de la proteína para luego realizar la simulación de su superficie proteica.

2.3.4.3.4. Diseño de cebadores

El diseño de los cebadores para la amplificación del gen aspartil aminopeptidasa se realizó con el programa *Gene Runner* versión 6.5.51 Beta el cual es de descarga libre.

La ejecución de una ventana de trabajo en *Gene Runner* se realizó mediante el botón *File* luego la opción *New* y *Nucleic acid sequence*. En la ventana de trabajo se llevó la plantilla del gen aspartil aminopeptidasa descargado en formato FASTA, abriendo la opción *open* del botón *File* seleccionando el archivo descargado. Con la plantilla de nucleótidos del gen en la ventana de *Gene Runner* se comenzó la búsqueda y ensayos para la obtención del mejor par de cebadores. Tal proceso se llevó seleccionando secuencias de nucleótidos en ambos extremos 3´ de las cadenas sentido y antisentido de la secuencia nucleotídica del gen diana. Se ensayaron la longitud del par de cebadores variando entre 15 a 30 nucleótidos. También se probaron diferentes ubicaciones para los cebadores. Se realizó la ejecución la opción *Oligo Analysis* que se encuentra en las herramientas del apartado *Analysis,* que se basa en el método de temperatura de fusión (*Tm* del inglés *Temperature melting*) *Nearest Neighbor*. Este algoritmo supervisa la clasificación de elementos, en este caso la pareja de cebadores, para que la función de estos se aproximen a la localización donde teóricamente deben hibridarse correctamente tomando como plantilla la secuencia del gen aspartil aminopeptidasa.

La ejecución de *Oligo Analysis* aperturó una ventana donde mostró los parámetros utilizados para el diseño de la pareja de cebadores entre ellos: hebra, longitud, comienzo y término en la plantilla, temperatura de fusión (*Tm*), porcentaje de G + C y energía libre de Gibbs (dG). En la parte inferior de la ventana de *Oligo Analysis* se tiene 5 accesos a las

estructuras secundarias que se forman o formarían al ensayar una pareja de cebadores los cuales se evitó su aparición. *Hairpin loops* o tallo-bucle que es una estructura secundaria muy común en los ARN, *Dimers* o dímeros que son las hibridaciones de cebadores entre sí por complementariedad de sus bases, *Bulge loops* o bucle protuberante que se forman cuando uno o más nucleótidos están presentes en una cadena en una región dúplex (Strom *et al*., 2015), además pueden encontrarse pequeños bucles de protuberancia en la mayoría de los ARN grandes y pequeños (Woese & Gutell, 1989; Cevec *et al*., 2010), *Internal loops* o bucles internos que se forman donde el ARN bicatenario se separa debido a que no hay emparejamiento entre los nucleótidos y *Match sites* o sitios de coincidencia que por lo general resultan del ensayo de la pareja de cebadores.

Se revisó la no conformación de las estructuras secundarias en ambos cebadores individualmente seleccionando la opción *Single* y su simulación funcional en pareja con la opción *Other*.

Teniendo en cuenta estos parámetros se procuró que la *Tm* de cada cebador este entre el rango de 50 a 65° C y que tengan entre ellos una diferencia de *Tm* no mayor a 5° C. También se vigiló que no formen estructuras secundarias demasiado estables observando que su *Tm* de formación de estos no esten en el rango de las temperaturas empleadas en la reacción en cadena de la polimerasa ó PCR (*Polimerase Chain Reaction*) y que su dG sea mayor a 1.

En el diseño de los cebadores también se procuró que el % de G + C no sea menor a 55 ya que este confiere una cierta especificidad al momento de la hibridación al templado. Se evitó el alto contenido de GC en el extremo 3', porque si la unión de este extremo es demasiado estable probablemente el cebador dará problemas de amplificación.

2.3.5. Clonación del gen aspartil aminopeptidasa

2.3.5.1. Recuperación de la bacteria criopreservada *C. salexigens* MP25462

La bacteria halófila *C. salexigens* MP25462 se recuperó a partir de tubos de criopreservación que se encontraban a una temperatura de -80° C. Se inoculó 200 µl de la cepa bacteriana en 5 ml de medio líquido SW al 15 % de NaCl (Anexo 2.A) y se incubó a 37° C durante 48 horas con agitación. La recuperación de la cepa bacteriana también se realizó en medio sólido, esto mediante la siembra por agotamiento de asa (Elorrieta, 2018).

2.3.5.2. Preparación de células competentes

Como se explicó en el Capítulo I apartado 1.2.12, una bacteria competente es aquella que está apta para recibir ADN exógeno, tal estado fisiológico se llama competencia. Es así que las células *E.coli* no competentes JM109 fueron inducidas para convertirse en competentes

mediante el siguiente protocolo: Se incubaron las células *E.coli* no competentes JM109 en 5 ml de medio LB líquido (Anexo 2.B) toda la noche a 37° C con agitación. Se tomó 1 ml de cultivo que se inoculó en un matraz con 50 ml de LB líquido. Luego se incubó con agitación a 37° C hasta que alcance una longitud de onda de 0.6 nm. Esto duró entre 3 a 4 horas. Se enfrió en hielo durante 10 minutos y se pasó a tubos de centrifuga de 50 ml previamente enfriados. Se centrifugó a 4° C a 4000 rpm por 17 minutos y retiró el sobrenadante. Se resuspendió con el mismo volumen (50 ml) de $MgCl_2$ 0.1 M frío (primero se resuspendió con 1 ml para disolver el precipitado y luego se añadió el resto de $MgCl_2$ 0.1 M). Se mantuvo en hielo por 15 minutos. Luego se centrifugó a 4000 rpm durante 7 minutos a 4 ° C y se desechó el sobrenadante. Se añadió 25 ml de $CaCl_2$ 0.1 M. Seguidamente se resuspendió y dejó en hielo 1 hora. Se centrifugó a 4000 rpm por 5 minutos a 4° C. Se retiró el sobrenadante y resuspendió en 5 ml de $CaCl_2$ 0.1 M más 20% de glicerol. Finalmente se hizo alícuotas de 200 µl y almacenó a –80° C.

2.3.5.3. Extracción del ADN genómico.

El método que se describe a continuación se realizó para la extracción del ADN genómico de *Chromohalobacter salexigens* MP25462 siguiendo el protocolo del *Kit Ilustra Tisue & cells genomicPrep Mini Spin (GE Healthcare)*. Primero se agregó 1.5 ml de la cepa MP25462 en un tubo de microcentrífuga. Se centrifugó a 16000 g por 3 minutos descartando el sobrenadante para así obtener el precipitado celular. Esto se realizó en el mismo tubo de microcentrífuga hasta utilizar los 5 ml incubados de MP25462, luego se agregó 50 µl de buffer fosfato salino (conocido también por sus siglas en inglés, PBS, de *Phosphate Buffered Saline*) para resuspender, homogenizar las células en la solución y eliminar los componentes del medio SW, después se centrifugó a 2000 g por 10 segundos descartando el sobrenadante, luego se adicionó 100 µl del tampón de lisis tipo 1. Así también se adicionó 10 µl de la Proteinasa K (20 mg/ml). Se mezcló durante 15 segundos, luego se incubó por 15 minutos a 56° C en la termoplaca y se centrifugó a 2000 g por 10 segundos.

Para la purificación se agregó 500 µl del tampón de lisis tipo 4 y se mezcló por 15 segundos para luego incubar la mezcla por 10 minutos a temperatura ambiente.

Se centrifugó a 11000 g por 10 segundos para llevar el contenido celular al fondo del tubo de microcentrífuga. Se ensambló la mini columna al tubo de recolección y se le añadió 700 µl de la muestra para centrifugarla por 1 minuto a 11000 g descartando el contenido del tubo de recolección.

Para el lavado y secado del ADN genómico se adicionó 500 µl del tampón de lisis tipo 4 y se centrifugó a 11000 g por 1 minuto descartando el exceso. Se agregó 500 µl de tampón de lavado tipo 6, se centrifugó a 11000 g por 3 minutos y descartó el tubo colector. Para la elución se transfirió la mini columna a un tubo de 1.5 ml libre de DNasas. Finalmente se agregó 200 µl del tampón de elución tipo 5 directamente al centro de la membrana de la columna. Se incubó la columna durante un minuto a temperatura ambiente y se centrifugó a 11000 g por 1 minuto para colectar el ADN genómico. Se almacenó el ADN genómico de *C. salexigens* a 4° C para su cuantificación y posterior uso. Todo este procedimiento se visualiza en la figura N° 16.

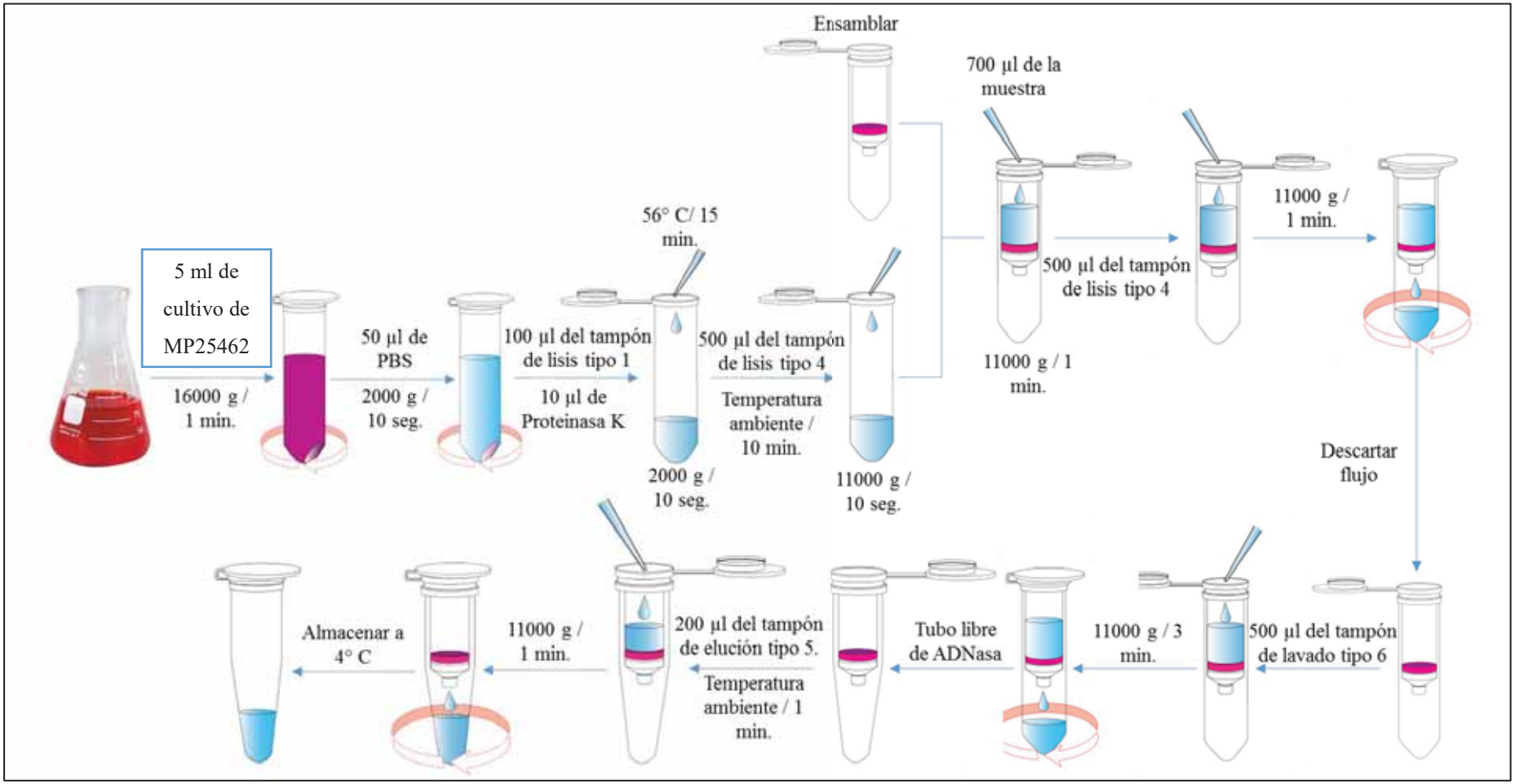


Figura N° 16: Proceso de extracción del ADN genómico de C. salexigens MP25462
Fuente propia

2.3.5.4. Cuantificación del ADN genómico.

La cuantificación del ADN genómico de *C. salexigens* MP25462 se realizó al nanoespectrofotómetro y en gel de agarosa. Para el primer caso se utilizó 1µl del ADN extraído y se colocó en el nanoespectrofotómetro marca DeNovix modelo DS-11+ para medir la concentración del ADN así como la pureza del mismo. La cuantificación en el gel de agarosa se realizó en una matriz al 0.8% para un volumen de 25 ml (Anexo 3.A). El ADN genómico y el marcador molecular Lambda HindIII se prepararon para la corrida electroforética. El volumen total para la carga en el gel fue de 12 µl: 8 µl de agua ultra pura, 2 µl de tampón de carga 6X y 2 ul del marcador, del mismo modo para la muestra de ADN genómico. se realizó la electroforesis a 90 V constantes durante 30 min. La electroforeisis se visualizó en el lector de imágenes ChemiDoc XRS+ (Bio-Rad). Los análisis y cálculos se realizaron con el programa Imagen Lab TM (Bio-Rad Laboratories, Inc.).

2.3.5.5. Estandarización de PCR convencional para la amplificación del gen aspartil aminopeptidasa.

La amplificación del gen aspartil aminopeptidasa se realizó mediante la técnica de PCR descrita en el capítulo 1. Se realizó la estandarización de la temperatura de hibridación de los cebadores diseñados. Cada reacción de PCR (20 µl) contenía de 2µl de ADN genómico extraído de *C. salexigens* MP25462 (5ng/µl), Taq polimerasa (Promega) 0.25 U, tampón de PCR a una concentración final de 1X, $MgCl_2$ a 1.5 mM, dNTPs a 0.2 mM y cada cebador a 10 ρmol (Tabla N° 5).

La reacción de PCR se llevó a cabo con los ciclos de amplificación que se explican a continuación: 5 minutos de denaturación inicial a 94 °C, 31 ciclos de 30 segundos a 94 °C, 30 segundos de hibridación a 48, 51 y 55° C por 30 segundos, extensión a 72°C por 30 segundos y finalmente una extensión final de 5 minutos a 72 °C

La temperatura de fusión escogida para la amplificación del gen aspartil aminopeptidasa fue de 51° C. Se realizó los cálculos para 10 reacciones de PCR. En este caso efectuó 5 reacciones en el termociclador Biometra y las otras 5 reacciones en uno de marca BioRad. La finalidad de esta PCR es obtener la suficiente cantidad de muestra para la secuenciación y clonación de la misma.

.

Tabla N° 5: Sistema de reacción de PCR para la estandarización de amplificación del gen aspartil aminopeptidasa.

Reactivo	Concentración Inicial	Concentración Final	Volumen por reacción
H_2O	-	-	9.35 µl
Tampón	10X	1X	4 µl
$MgCl_2$	50 mM	1.5 mM	0.6 µl
dNTPs	2 mM	0.2 mM	2 µl
Primer Sentido	10 ρmol/µl	10 ρmol	1 µl
Primer Antisentido	10 ρmol/µl	10 ρmol	1 µl
Taq Polimerasa	5 U/µl	0.25 U	0.05 µl
ADN	5 ng/µl	10 ng	2 µl
		Volumen Total	20 µl

X = Es las veces que está concentrado el tampón respecto al de su uso corriente.

2.3.5.6. Electroforesis en gel de agarosa.

Se preparó el gel de agarosa al 0.8 % para la visualización del ADN genómico y al 1% (Anexo 3.B) para las reacciones de amplificación por PCR. Para preparar un gel de agarosa al 1% se pesó 1g de agarosa y se disolvió en 100 ml de buffer TAE 1X. (Anexo 4) La mezcla contenida en un matraz se calentó, en un horno de microondas dando pulsos de 30 segundos hasta lograr una mezcla homogénea evitando la ebullición violenta de la mezcla. Se colocó la bandeja en el soporte del gel o "*gel caster*" y se giró la palanca de ajuste para cerciorar el cierre hermético. Antes de colocar la bandeja se niveló el soporte. Se añadió 1 ul de *Real safe* a la solución. Se vació con cuidado la mezcla caliente de agarosa/buffer evitando hacer burbujas y se colocó un peine en un extremo. Se dejó enfriar hasta que polimerizó el gel. Se desajusto el soporte para retirar el molde y sumergirlo en la cámara electroforética que contenía el tampón TAE al 1X y finalmente se retiró cuidadosamente el peine del gel. Los estándares moleculares utilizados fueron Lambda DNA Hind III e Hyper Ladder II los cuales fueron preparados con 2 µl de tampón de muestra, 8 µl de agua ultra pura y 2 µl del marcador molecular. Estos fueron colocados independientemente en el primer o último pozo del gel de agarosa.

Para la preparación de las muestras amplificadas se realizó combinando 2 µl del ADN producto de la amplificación por PCR agregándole 2 µl del tampón de carga y 8 µl de agua ultra pura. Las muestras se colocaron en los pozos del gel. En el último pozo se colocó el blanco que es el control de contaminación en el proceso de la reacción de PCR, porque contiene agua

en vez de la muestra de ADN. La electroforesis se realizó a 90 voltios constantes durante 1 hora. El resultado de la electroforesis se visualizó el lector de imágenes ChemiDoc XRS+ (Bio-Rad). Las bandas amplificadas fueron analizadas y calculadas por el programa Imagen Lab TM (Bio-Rad Laboratories, Inc.).

2.3.5.7. Purificación y secuenciación del gen aspartil aminopeptidasa.

Para comprobar la identidad del fragmento amplificado se procedió a la purificación de la banda amplificada. Se realizó los cortes de las bandas de interés mediante un bisturí lo más rápido posible para que el ADN del gen Asp no se formen dímeros de timina ante la exposición a la luz UV. A continuación se realizó el protocolo del *Kit Ilustra^{TM} GFX PCR DNA and Gel Band Purification* que consistió en pesar un tubo de microcentrífuga para luego colocar dentro de la misma las rebanadas de geles que contienen la banda amplificada del gen aspartil aminopeptidasa, luego se volvió a pesar el tubo conteniendo los geles, calculando por diferencia el peso de la muestra a procesar.

Según el protocolo del ya mencionado kit se tuvo que agregar 300 µl del tampón de captura tipo III a la muestra. Después se mezcló por inversión e incubó a 60° C durante 15 minutos en la termoplaca. Luego se transfirió 800 µl de la mezcla tampón tipo III y muestra del amplicón aspartil aminopeptidasa a una columna GFX ensamblado previamente a un tubo recolector, luego se incubo a temperatura ambiente por 1 minuto y centrifugó a 16000 g por 30 segundos descartando el exceso del tubo colector. A continuación se añadió 500 µl del tampón de lavado tipo I a la columna GFX centrifugándola a 16000 g por 30 segundos. Luego se transfirió la columna GFX a un tubo de microcentrífuga libre de DNasas añadiendo de 10 µl del tampón de elución tipo IV al centro de la membrana de la columna dejando incubar la columna a temperatura ambiente por 1 minuto se centrifugó a 16000 g por 1 minuto para recuperar el ADN purificado. Este proceso se observa en la figura N° 17.

La muestra se envió para su secuenciación con la técnica de Sanger al servicio de secuenciación de la ULL. Teniendo en cuenta lo recomendado por el servicio.

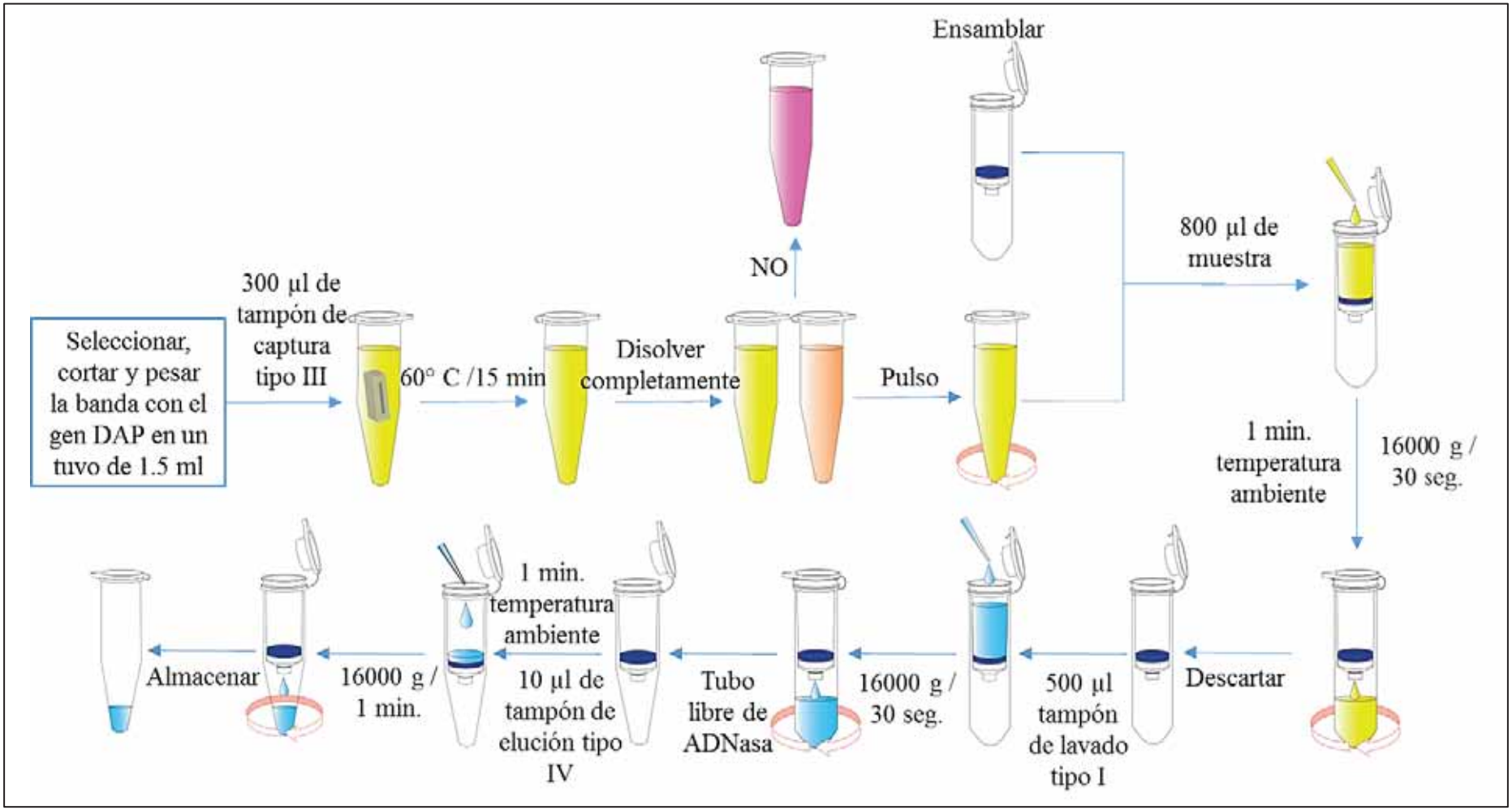


Figura N° 17: Proceso de purificación del gen aspartil aminopeptidasa
Fuente propia.

2.3.5.8. Análisis bioinformático de las secuencias del gen aspartil aminopeptidasa.

Los datos obtenidos de los archivos de secuenciación capilar automática del gen aspartil aminopeptidasa fueron revisados manualmente, tanto su secuencia como su cromatograma para determinar nucleótidos codificados erróneamente, usando el programa MEGA versión 10.0 (Figuras N° 18 y 19). La secuencia del gen Asp obtenida en el presente estudio fue comparada contra la base de datos del GenBank, usando el algoritmo BLAST (*Basic Local Alignment Search Tool*) http://www.ncbi.nlm.nih.gov/BLAST/ para obtener la identidad de la región amplificada y su comparación de secuencias (Figura N° 20).

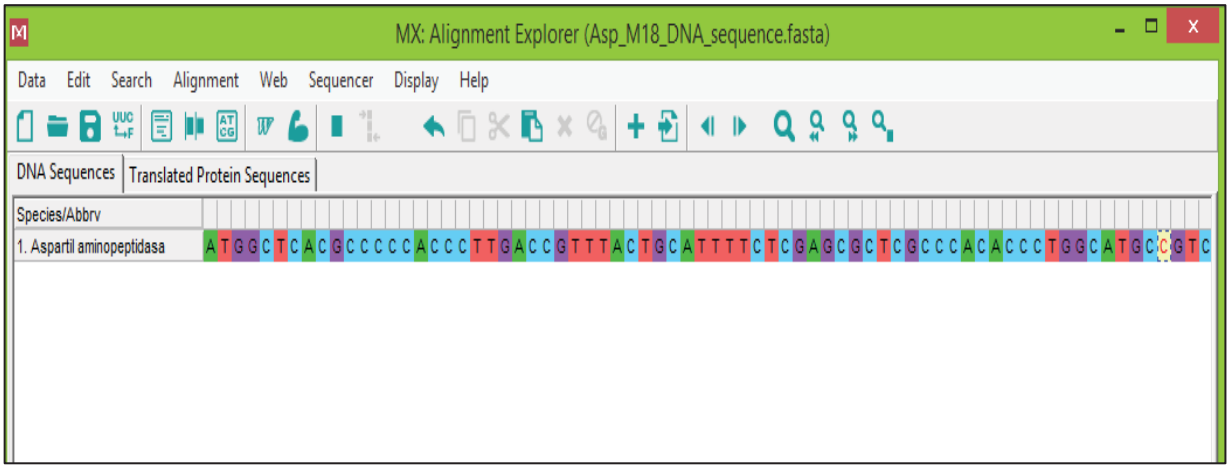


Figura N° 18. Visualización de una secuencia nucleotídica en MEGA 10.
Fuente MEGA 10

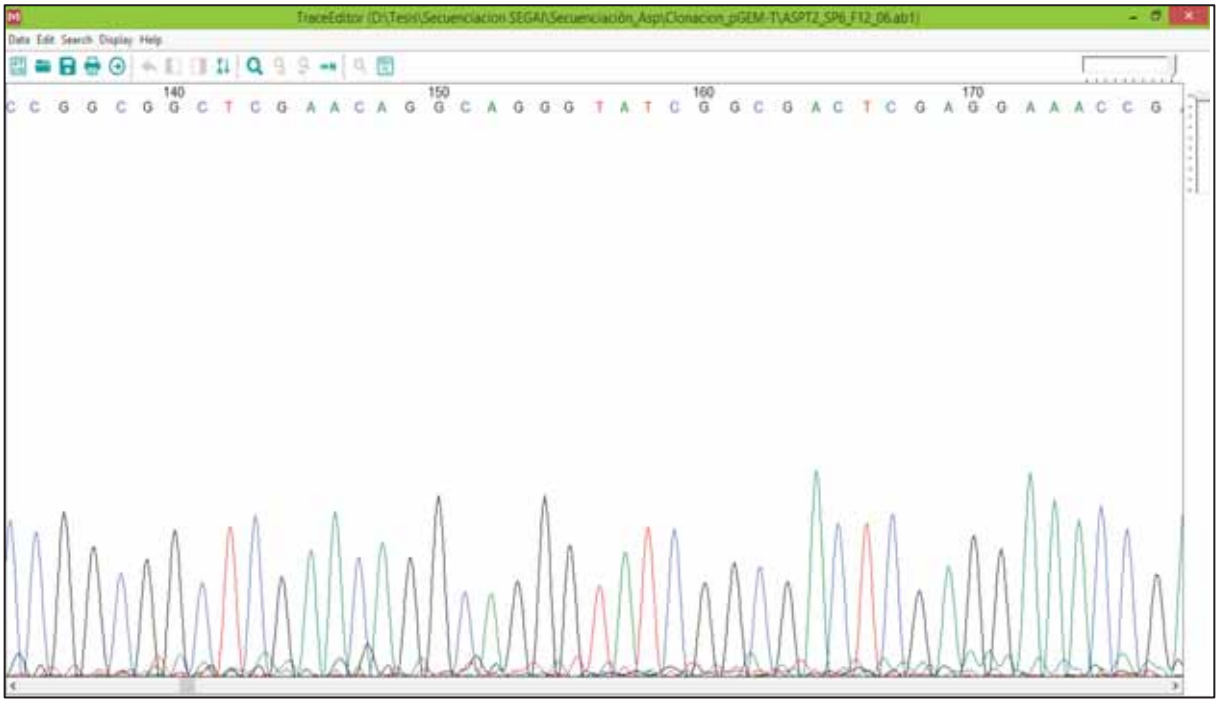


Figura N° 19. Cromatograma nucleotídica en MEGA 10.
Fuente MEGA 10

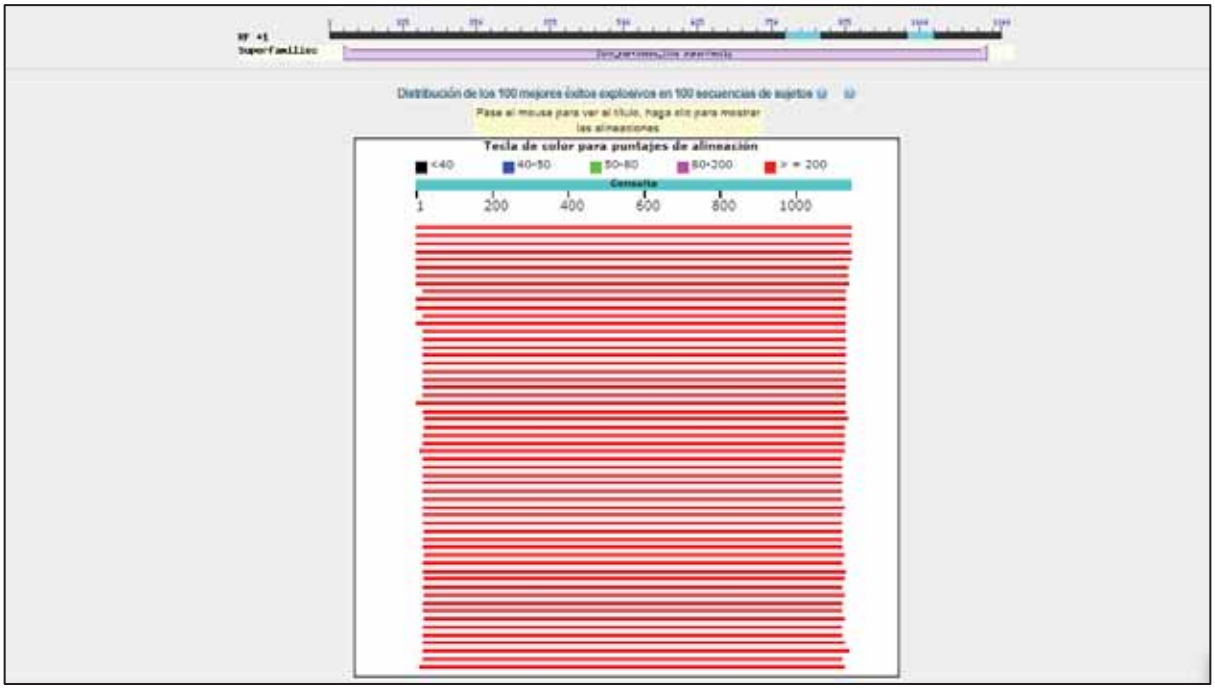


Figura N° 20. Resultados de alineamiento BLAST.
Fuente [*https://blast.ncbi.nlm.nih.gov/Blast.cgi*](https://blast.ncbi.nlm.nih.gov/Blast.cgi)

2.3.5.9. Clonación del gen aspartil aminopeptidasa en el vector plasmídico pGEM-T.

El procedimiento de clonación del gen aspartil aminopeptidasa se realizó en *pGEM-T Vector System I*, para lo cual se procedió con el protocolo de este sistema diferenciado en dos pasos: ligación inserto-vector y transformación.

Primeramente se realizó la ligación de los productos purificados del gen aspartil aminopeptidasa con el vector pGEM-T. El sistema de vectores pGEM-T son convenientes para clonar productos de PCR generados por ciertas polimerasas termoestables. En este caso la Taq polimerasa añadió una única desoxiadenosina, de forma independiente del molde de aspartil aminopeptidasa, a los extremos 3' de los fragmentos amplificados. El vector pGEM-T está listo para ser usado en reacciones de ligación, el cual ya viene preparado a partir del vector pGEM®-5Zf (+) cortado con la enzima de restricción EcoRV y con una timina agregada en el terminal 3´ en ambos extremos. Estos salientes 3´-T únicos en el sitio de inserción mejoran en gran medida la eficiencia de la ligación de un producto de PCR en el plásmido al evitar la recircularización del vector y proporcionar un saliente compatible para la ligación de los productos de PCR del gen aspartil aminopeptidasa con salientes A

2.3.5.9.1. Ligación inserto-vector

La ligación de los productos purificados del gen aspartil aminopeptidasa (inserto) y el vector pGEM-T se realizó centrifugando vigorosamente el vector pGEM-T y el inserto del

ADN control. Luego se calculó la cantidad requerida del ADN del inserto para una relación de 1:3 del vector e inserto respectivamente. Tal cálculo se realizó la siguiente formula:

$$Productos\ de\ PCR\ requeridos = \frac{50\ ng\ x\ Longitud\ de\ inserto}{Longitud\ de\ vector} x \frac{3}{1}$$

Se mezcló vigorosamente el tampón de ligación rápida 2X antes de su uso. Para preparar la reacción de ligación se usó tubos de microcentrífuga de 1.5 ml de capacidad. Los componentes así como el volumen de los mismos se observan en la Tabla N° 6.

Cabe resaltar que el vector pGEM-T no se requiere un pretratamiento con una enzima de restricción para linealizarse porque ya viene en esta condición y además con una timina en los extremos 3´ los cuales se ligan con las desoxiadenosinas dejadas por la Taq polimerasa en los extremos 3´del amplicón aspartil aminopeptidasa.

Tabla N° 6: Componentes de reacción de ligación

Componentes de reacción	Reacción estándar	Control positivo	Control de fondo
Tampón de ligación rápida ADN T4 ligasa	5 µl	5 µl	5 µl
Vector pGEM-T (50 ng)	1 µl	1 µl	1 µl
Productos de PCR	X µl*	-	-
Inserto de ADN control	-	2 µl	-
ADN T4 ligasa	1 µl	1 µl	1 µl
Agua libre de nucleasas para un vol. final	10 µl	10 µl	10 µl

* La cantidad de productos de PCR no debe sobrepasar los 3 µl ya que el volumen final de la reacción de ligación es de 10 µl.

Mezclar las reacciones y dejar en incubación durante toda la noche a 4° C.

2.3.5.9.2. Transformación.

La transformación de la reacción de ligación inserto-vector se llevó a cabo con las células competentes *E. coli* JM109. La reacción de ligación se centrifugó brevemente al mismo tiempo que se descongelaba las células competentes JM109 en una cubeta con hielo. Se agregó 2 µl de la reacción de ligación a 50 µl de células JM109, se mezcló suavemente para combinar los dos componentes en un tubo de 1.5 ml para luego mantenerlo en hielo durante 20 minutos.

Después se realizó el shock térmico que consistió en llevar las células JM109 exactamente a 42° C durante 1 minuto y luego 5 minutos en hielo. Para luego agregar 100 µl de medio SOC (Anexo 2.C) e incubar a 37° C durante 90 minutos.

Finalmente se sembró por diseminación 100 µl de las células transformadas en una placa con medio LB con ampicilina a una concentración de 100 µg/ml previamente preparada y se incubó a 37° C por 24 horas.

2.3.5.10. PCR de las colonias clonadas

Se seleccionaron colonias al azar que crecieron en la placa con medio LB con ampicilina. Las colonias transformadas se reconocen por el haber adquirido la resistencia a la ampicilina ya que el vector pGEM-T presenta el gen de resistencia a este antibiótico. Luego con el proceso de PCR se determinan que colonias contienen el vector con el inserto aspartil aminopeptidasa o en las que se recirculó.

Las bacterias transformadas se cultivaron en 1000 µl de medio líquido LB/ampicilina contenido en tubos de microcentrífuga. Se llevó a incubación durante 16 horas a 37 °C con agitación. Terminado el tiempo de incubación se tomó 100 µl de cada tubo para llevarlo a otro tubo libre de ADNasas para centrifugarlas a 6000 rpm por 10 minutos descartando el sobrenadante. Se agregó 50 µl de agua ultra pura a cada tubo y se llevaron a baño maría durante 5 minutos a 100 °C, luego se centrifugó a 13000 rpm durante 5 minutos para recuperar el sobrenadante. Todo este proceso se realizó para todas las colonias seleccionadas.

Se utilizó 2 µl del sobrenadante como el ADN molde para cada reacción de PCR utilizando los cebadores diseñados para la amplificación del gen DAP y los cebadores T7 y SP6 del vector pGEM-T. La PCR con los cebadores T7 y SP6 (vector) fue de 5 segundos a 94 °C, 31 ciclos de 30 segundos a 94 °C, 30 segundos a 50 °C, 30 segundos a 72 °C y finalizando con 5 minuto a 72 ° C. De igual manera se llevó a cabo con los cebadores diseñados (inserto) solamente variando en la temperatura de hibridación: 5 minutos de denaturación inicial a 94 °C, 31 ciclos de 30 segundos a 94 °C, 30 segundos de hibridación a 51° C, extensión a 72°C por 30 segundos y finalmente una extensión final de 5 minutos a 72 °C. Terminado el proceso de PCR se realizó la electroforesis de las muestras en un gel de agarosa al 1% a 90 V constantes durante 1 hora. La visualización de los resultados se realizó con el lector de imágenes ChemiDoc XRS+ (Bio-Rad). Con el software Imagen Lab ™ (Bio-Rad Laboratories, Inc.) se calculó el tamaño de las bandas amplificadas.

2.3.5.11. Extracción del ADN plasmídico por lisis alcalina.

Este método se realizó con el *Kit E.Z.NA.® Plasmid DNA Mini Kit I D6942-02* (Omega) siguiendo las instrucciones del fabricante para extraer el ADN plasmídico pGEM-T clonado en las células JM109, dicho plásmido contendrá el gen aspartil aminopeptidasa de *C. salexigens* MP25462.

Se aisló una única colonia de la placa con el antibiótico selectivo ampicilina 100 mg/µl e inoculó a un cultivo de 5 ml de medio LB/ampicilina. Se incubó durante 16 horas a 37 ° C con agitación vigorosa de 300 rpm. Se trasvasó 2 ml del medio incubado a un tubo de 10 ml para luego centrifugarlo a 10.000 x g durante 1 minuto a temperatura ambiente. Se decantó al precipitado, se agregó 250 µl de solución I / ARNasa A, se mezcló bien y resuspendió completamente el sedimento celular para obtener buenos rendimientos. Esta suspensión se transfirió a un tubo nuevo de microcentrífuga de 1.5 ml y se agregó 250 µL de Solución II e invirtió y giró suavemente el tubo varias veces para obtener un lisado transparente y se incubó por 2 minutos. En este paso se evitó la mezcla vigorosa porque el ADN cromosómico habría sufrido cortes que resultarían en pequeños fragmentos que afectaría a la pureza del plásmido. En este paso se procuró que la reacción de lisis no continúe durante más de 5 minutos. Se agregó 350 µl de solución III e inmediatamente se invirtió varias veces hasta que se formó un precipitado blanco floculante. Es importante que la solución se mezcle bien para evitar la precipitación localizada, después se centrifugó a la velocidad máxima (≥13,000 x g) durante 10 minutos.

Se colocó una mini columna de ADN HiBind® en un tubo de recolección de 2 ml y se le agregó 100 µL de NaOH 3M para luego centrifugarla a 16000 g durante 60 segundos y desechar el filtrado y reutilizar el tubo de recogida. Después se transfirió el sobrenadante aspirándolo cuidadosamente a la mini columna de ADN HiBind®. En este paso se tuvo mucho cuidado de no perturbar el sedimento para no transferir residuos celulares a la mini columna de ADN HiBind® y se centrifugó a 16000 g durante 1 minuto desechando el filtrado y reutilizando el tubo de recolección para luego añadir 500 µL de tampón HB y centrifugar a la velocidad máxima durante 1 minuto. Se desechó el filtrado y reutilizó el tubo de recogida añadiendo 700 µL del tampón de lavado de ADN para después centrifugarlo a velocidad máxima durante 1 minuto desechando el filtrado y reutilizando el tubo de recolección.

Se realizó un paso de centrifugación de la mini columna de ADN de HiBind® a velocidad de 16000 g durante 2 minutos para secar la matriz de la columna. Después se transferió la mini columna que contiene el ADN a un tubo de microcentrífuga nuevo de 1.5 ml y se agregó 30-100 µl de tampón de elución directamente al centro de la membrana de la columna que se dejó reposar a temperatura ambiente durante 1 minuto.

Finalmente se centrifugó a 16000 g durante 1 minuto y almacenó el ADN plasmídico a - 20 ° C.

2.3.5.12. Secuenciación del gen aspartil aminopeptidasa clonado

El producto de la extracción del ADN plasmídico clonado fue secuenciado en el Servicio General de Apoyo a la Investigación de la Universidad de La Laguna (SEGAI) para asegurar la clonación del gen aspartil aminopeptidasa y conocer como ingreso el gen en el plásmido pGEM-T, recto o invertido. Se envió un volumen de 10 µl a una concentración de 20 ng/µl. Los resultados de este paso fueron analizados en el programa MEGA 10 como se describió en el apartado 2.3.5.9 de análisis bioinformático.

2.3.5.13. Criopreservación de células clonadas.

La criopreservación es el proceso en el cual las células o tejidos son congelados a muy bajas temperaturas, generalmente entre -80 °C y -196 °C (el punto de ebullición del nitrógeno líquido) para disminuir las funciones vitales de una célula o un organismo y poderlo mantener en condiciones de *vida suspendida* por mucho tiempo. En cuanto a la congelación sus rangos van de temperaturas iguales o menores a 0 °C (Sambu, 2015). Es así que las cepas clonadas con el inserto aspartil aminopeptidasa fueron criopreservadas en tubos especiales de criopreservación de 2 ml de capacidad. Primero se preparó las cepas bacterianas clonadas al realizar su cultivo en 5 ml de medio LB/ampicilina durante 16 horas a 37° C con agitación. Se colocó en cada tubo 160 µl de medio LB con ampicilina con crecimiento bacteriano, después se colocó glicerol estéril a una concentración de 20 %, es decir, 40 µl. Luego cuidadosamente se invirtió los tubos con la finalidad de homogenizar esta mezcla. Finalmente estos tubos fueron almacenados en refrigeradores de -80 °C.

CAPITULO III
RESULTADOS Y DISCUSIÓN

3.1. Etapa 1: Caracterización *in silico* del gen aspartil aminopeptidasa

3.1.1. Análisis primario

La calidad de la secuenciación masiva del genoma de MP25462 mediante el programa FastQC mostró módulos o controles de calidad. Todos estos resultados se encuentran en un archivo en formato Html. Cada uno de estos módulos fueron analizados.

El primer módulo del programa FastQC es el de **estadísticas básicas**. Tal parámetro muestra los datos de los archivos producto de la secuenciación genómica de *C. salexigens.* Los valores más resaltantes fueron el % de GC que fue de 62, el número de secuencias producidas fue de 594952 cada una de ellas con una longitud entre 35 a 301 pb. En la tabla N° 7 se puede ver dichos datos para las secuencia del archivo *forward*. Los datos para el archivo *reverse* son iguales al de *forward* en este control.

Tabla N° 7: Estadísticas básicas del archivo forward

Características	Valores
Nombre del archivo	pe_1_S15_L001_R1_001.fastq
Tipo de archivo	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total de secuencias	594952 lecturas
Secuencias marcadas como de baja calidad	0
Longitud de secuencias	35-301 pb
%GC	62

En el módulo de **calidad de las bases secuenciadas** se esperó que las mismas tengan un índice *Phred* mayor a 20 (probabilidad de error de 1 en 100). Los resultados de este control para cada posición de las secuencias del archivo *forward* se observaron en un diagrama de cajas (Figura N° 21) en donde las secuencias o lecturas con una longitud de entre 35 hasta 277 pb obtuvieron valores phred mayores de 20 hasta 38. A partir de secuencias de tamaños entre 278 hasta 301 pb empieza a descender la calidad, el índice phred se presenta desde un valor 12 hasta 19. Los valores de la mediana (línea roja) van desde 17 a 38 y en cuanto a la calidad media (línea azul) presenta datos desde 18 a 37 a lo largo de las lecturas (Figura N° 21).

Por otra parte el control de calidad de las bases secuenciadas del archivo *reverse* las secuencias de longitud de 35 a 200 pb presentaron un índice phred de 20 hasta 38. Desde las

secuencias de longitud de 201 a 301 pb se observó un índice phred comenzó a decaer presentando valores de 7 hasta 19. Los valores de la mediana van desde 8 a 38 y en cuanto a la calidad media presenta datos desde 11 a 37 a lo largo de las lecturas (Figura N° 22). Vale aclarar que las calidades de las bases secuenciadas menores a Phred 20 no se presentan en todas las lecturas del tamaño correspondiente.

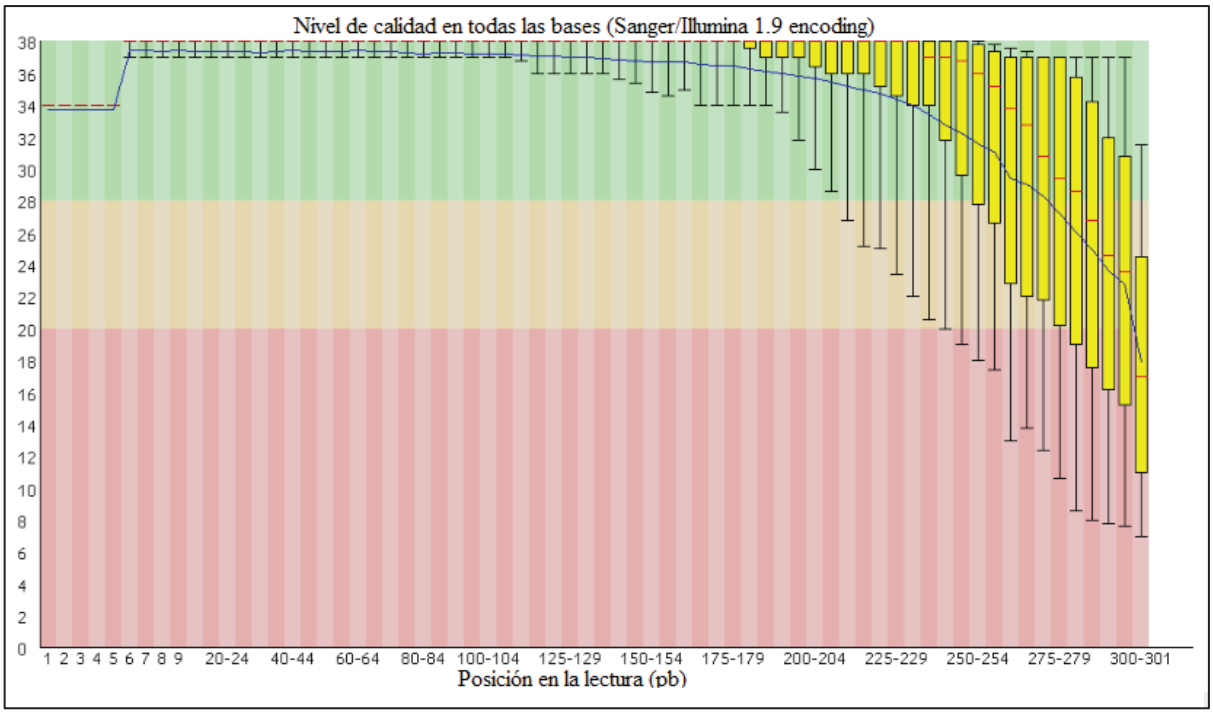


Figura N° 21. Calidad de las bases secuenciadas forward. Eje X: longitud de lecturas en pares de base (pb). Eje Y: índice Phred

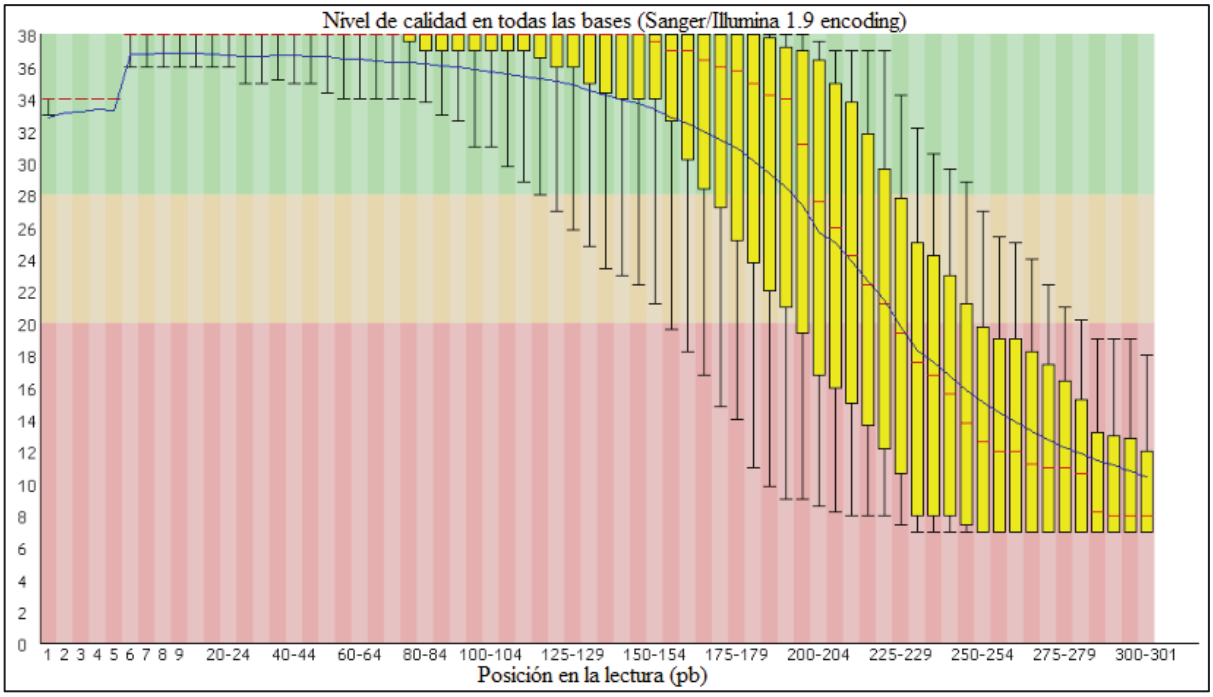


Figura N° 22. Calidad de las bases secuenciadas reverse. Eje X: longitud de lecturas en pares de base (pb). Eje Y: índice Phred

El módulo de **calidad de las secuencias por azulejo** que presenta el programa FastQC solamente aparece cuando se utiliza una biblioteca de Illumina, siendo este el caso, muestra la célula de flujo (*flow cell*), canales donde se pone la secuencia. En este gráfico se observó la aparición de manchas en el fondo azul tanto en las lecturas del archivo *forward* y *reverse* (Figura N° 23 y 24). Esto indica que se obtuvo una mala calidad en ciertas regiones de célula de flujo a causa de dos motivos: primero puede que no se haya filtrado los reactivos respectivos o segundo que no se haya hecho el vacío. De forma que las burbujas quedan en la célula de flujo y de esta forma los espectros se interfieren, advirtiendo de fallas en el proceso experimental de la secuenciación masiva.

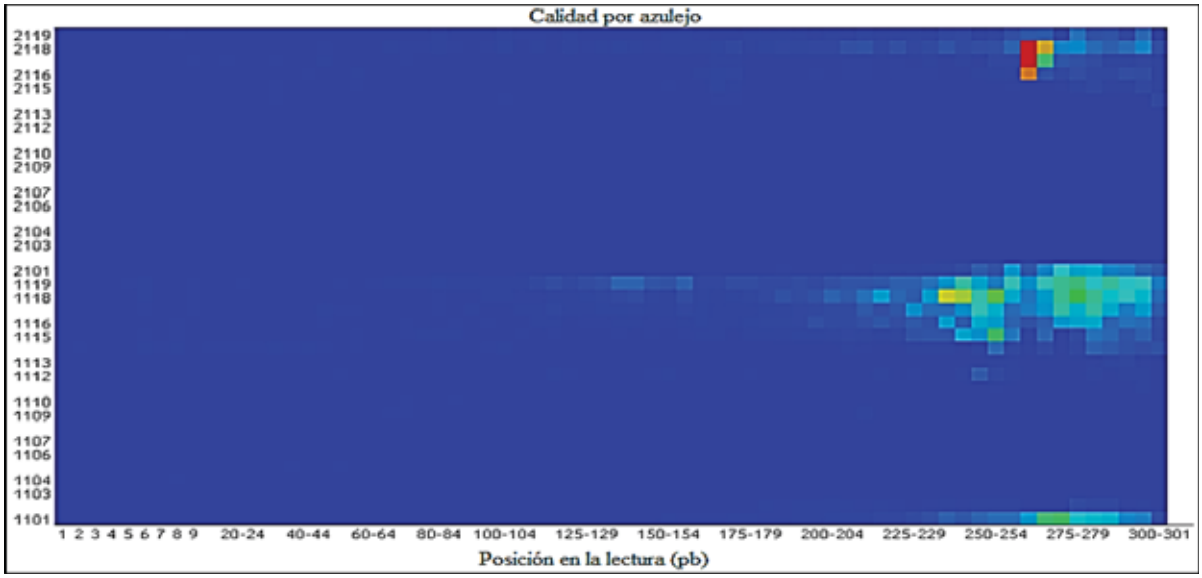


Figura N° 23. Calidad por azulejo forward. Eje X: longitud de lecturas en pares de base (pb). Eje Y: Tamaño de la celda de flujo (mm)

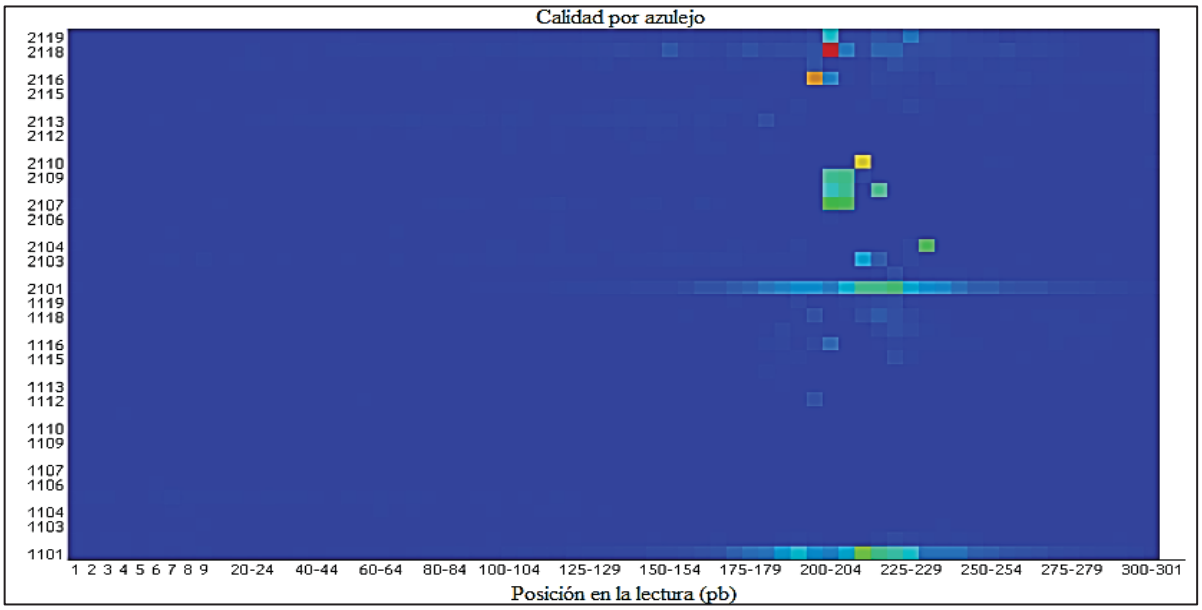


Figura N° 24. Calidad por azulejo reverse. Eje X: longitud de lecturas en pares de base (pb). Eje Y: Tamaño de la celda de flujo (mm)

La calidad de **secuencias por base** muestra de antemano de cuantas lecturas se van a eliminar ya que permite ver si un subconjunto de sus secuencias tiene valores con baja calidad.

En el gráfico resultante en la calidad de secuencias por base (Figura N° 25 y 26) muestran en el eje X la media de la calidad de las secuencias con el índice phred. Mientras que en el eje Y, se representa el número se secuencias o lecturas al que corresponde esa media.

En el caso de las secuencias *forward* se tiene más de 300000 lecturas con una calidad media de entre 36 y 37 (Figura N° 25). Mientras que en el gráfico de las secuencias *reverse* se observa que el número de secuencias a partir de una calidad 30 desciende hasta menos de 25000 lecturas para luego incrementar a más de 200000 lecturas con calidades medias de entre 36 y

37, siendo una discontinuidad en la cantidad de lecturas según su calidad (Figura N° 26). Así estos resultados dieron a para ambos casos, la existencia de menos lecturas con calidad media baja.

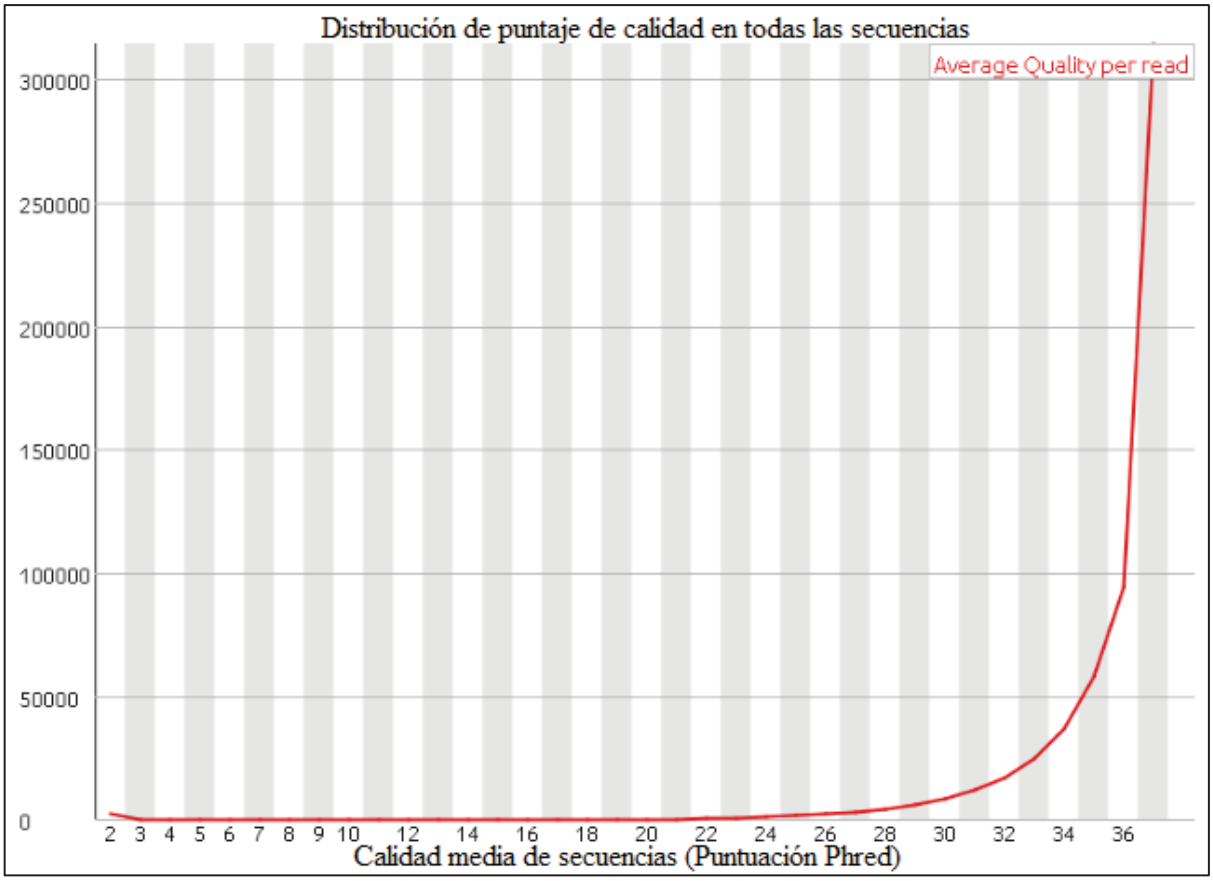


Figura N° 25. Calidad de secuencias por base del archivo forward.

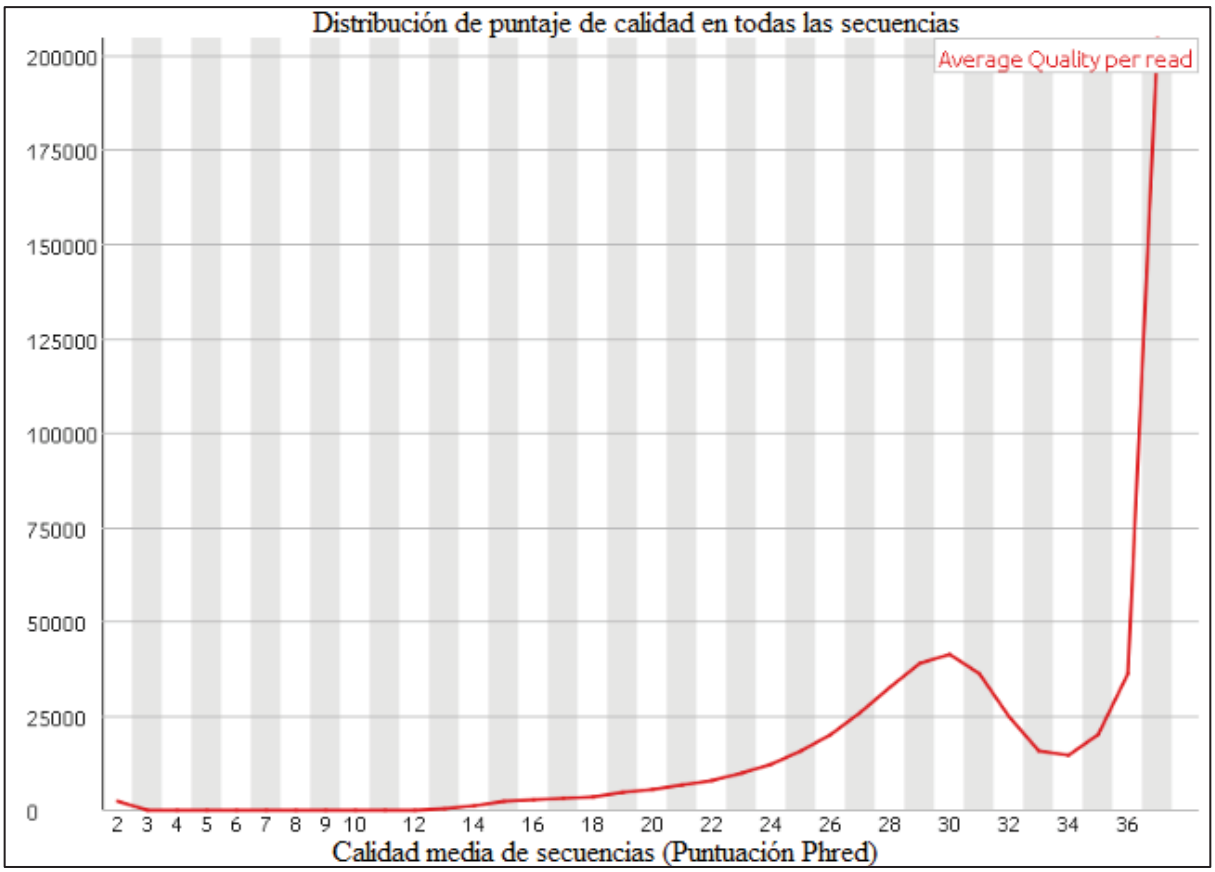


Figura N° 26. Calidad de secuencias por base del archivo reverse.

En el módulo del **contenido de la secuencia por base** mostró la proporción de cada una de las bases que existen en la secuencias según su tamaño. En las secuencias *forward*, se observó que las líneas negra y azul, guanina (G) y citosina (C) respectivamente, corrieron en paralelo a lo largo de las lecturas (Figura N° 27). De igual manera sucedió con las líneas verde y azul correspondientes a adenina (A) y timina (T). El grafico para las secuencias *reverse* muestran que las bases nitrogenadas de GC tienen una proporción lineal hasta lecturas de longitud menos de 200 pb mientras que la proporción de A T se muestran en paralelo hasta casi el total de lecturas (Figura N° 28).

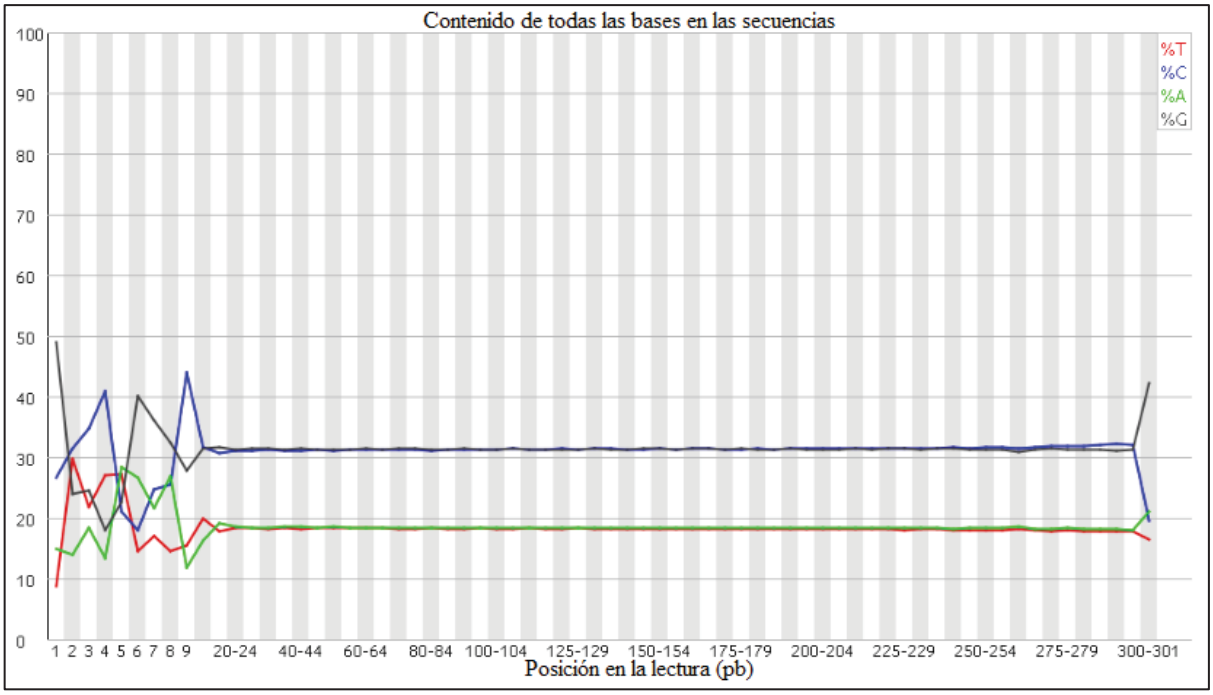


Figura N° 27. Contenido de la secuencia por base forward

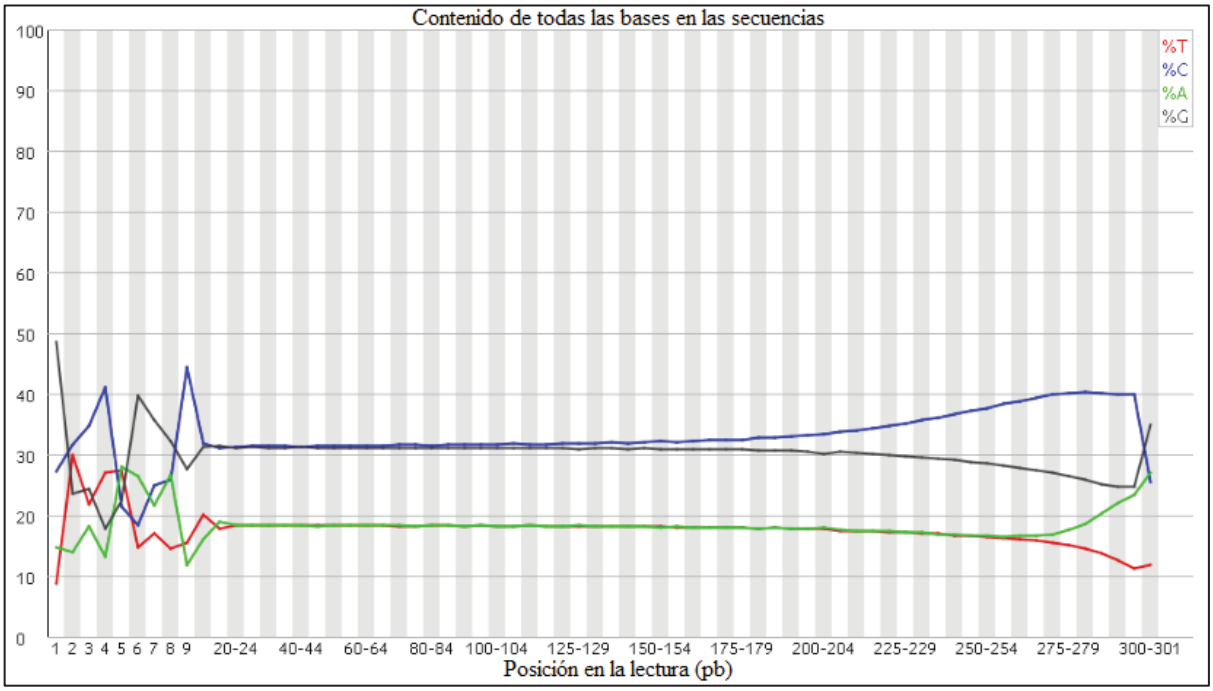


Figura N° 28. Contenido de la secuencia por base reverse

En el gráfico N° 29 y 30 se muestra el **contenido de guanina y citosina (GC) por secuencia**, y se observó que el valor teórico (curva azul) es compatible con el valor de las lecturas trabajadas (curva roja). Tanto las secuencias *forward* y *reverse* el porcentaje de GC es de 63.5% (Figura N° 29 y 30).

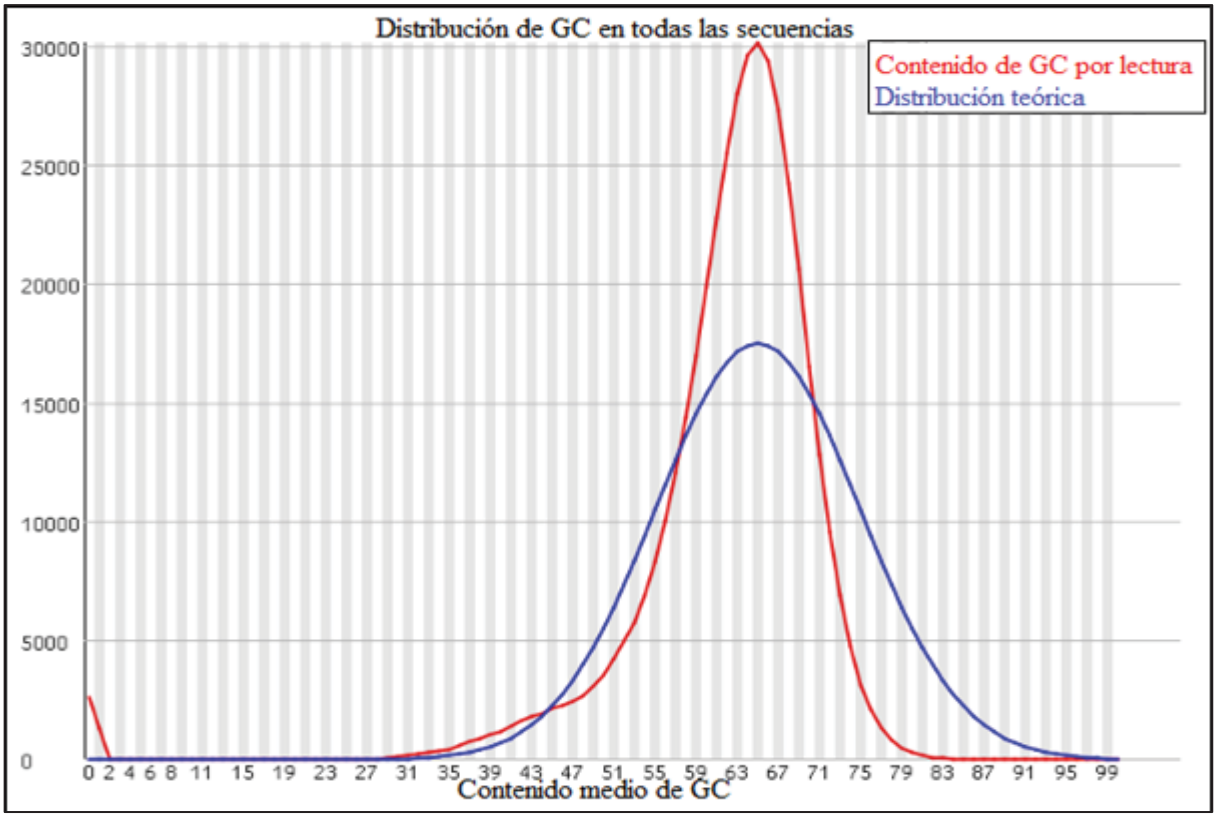


Figura N° 29. Contenido de guanina y citosina del archivo forward. Eje X: % de GC. Eje Y: número de lecturas.

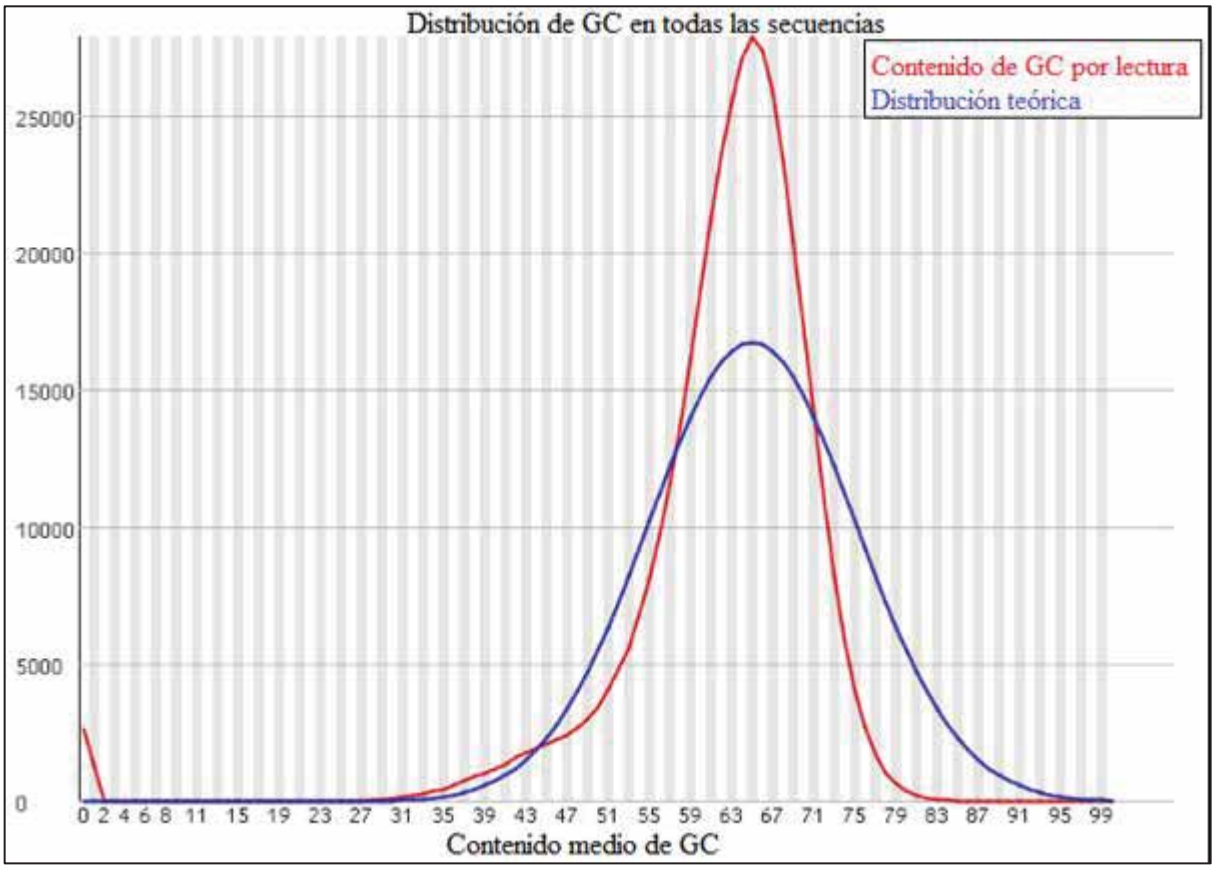


Figura N° 30. Contenido de guanina y citosina del archivo reverse. Eje X: % de GC. Eje Y: número de lecturas.

En el contenido N (N=asignación de una base nucleotídica por carecer de calidad) por base para los diagramas de secuencias *forward* y *reverse* mostró que el contenido de N es prácticamente nulo, es decir, que no se han asignado N en las bases (Figura N° 31 y 32). Esto indica que las secuencias no tienen una mala la calidad como para poner una N.

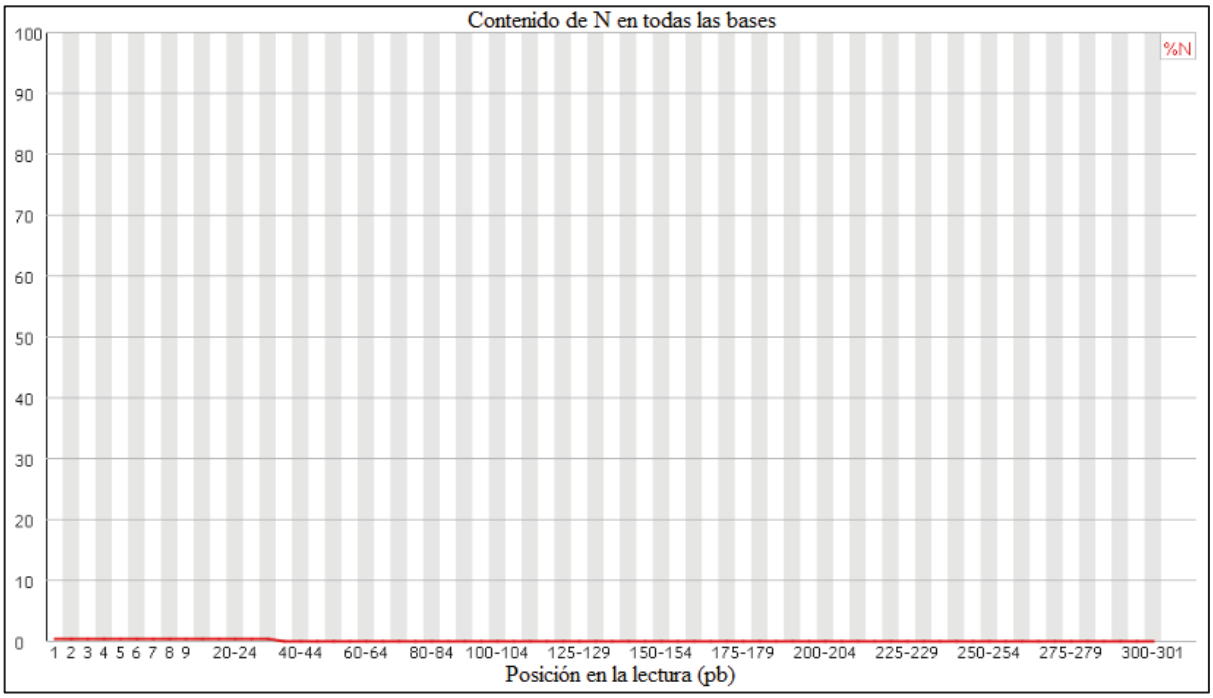


Figura N° 31. Contenido N por base del archivo forward

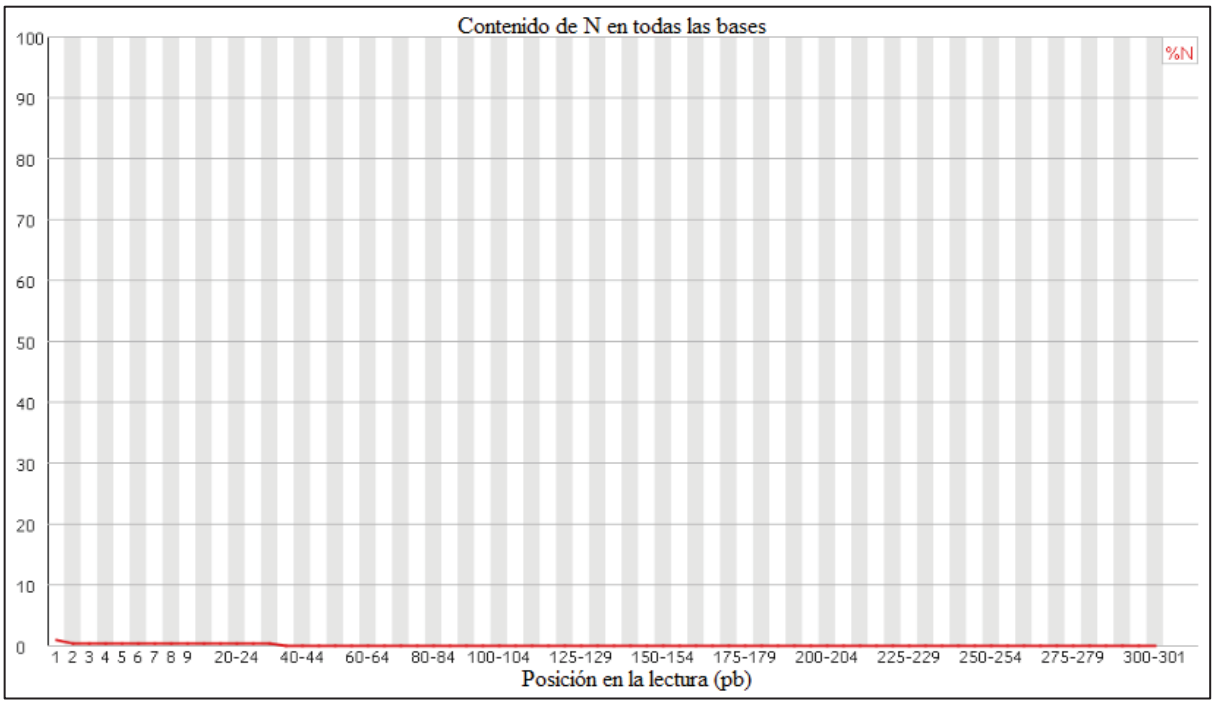


Figura N° 32. Contenido N por base del archivo reverse

En la distribución de la longitud de las secuencias para la biblioteca de MP25462 presenta secuencias heterogéneas en las lecturas *forward* y *reverse*. Hasta 250000 secuencias para ambos casos, presentan en un tamaño 299 pb (Figura N° 33 y 34).

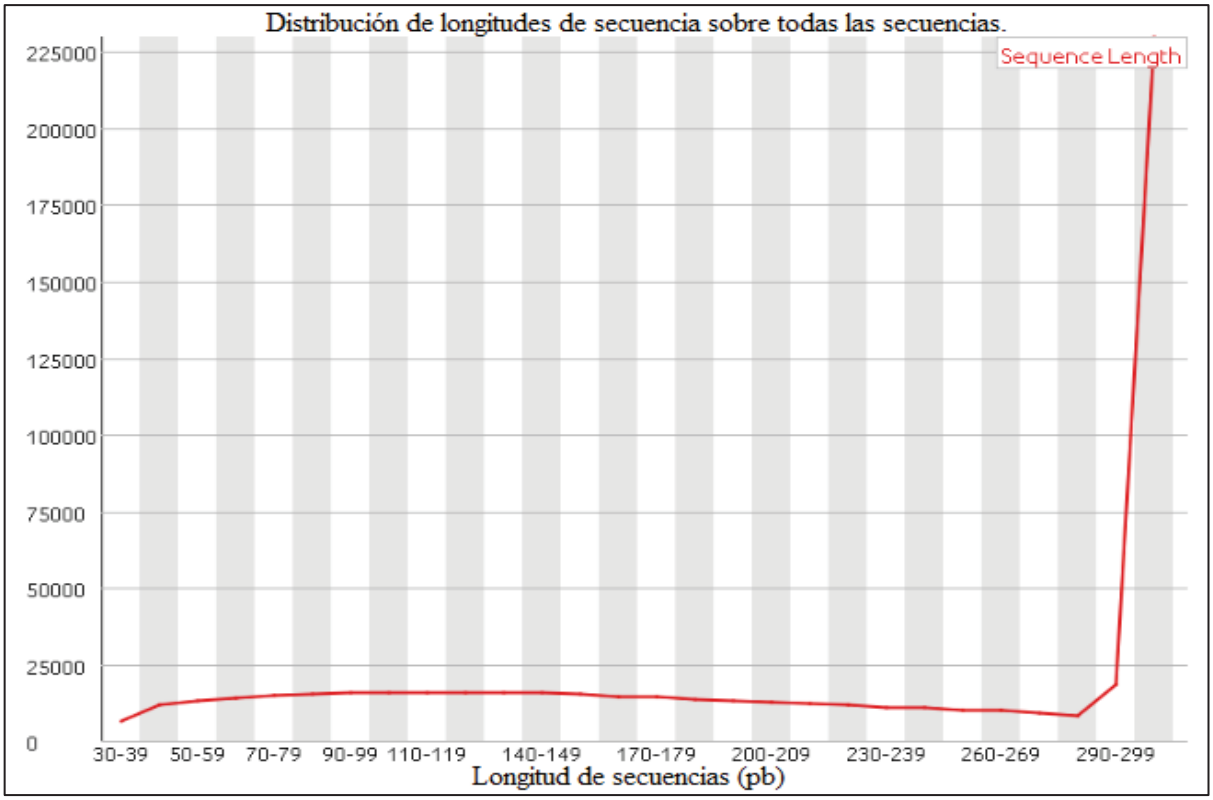


Figura N° 33. Distribución de la longitud de las secuencias del archivo forward

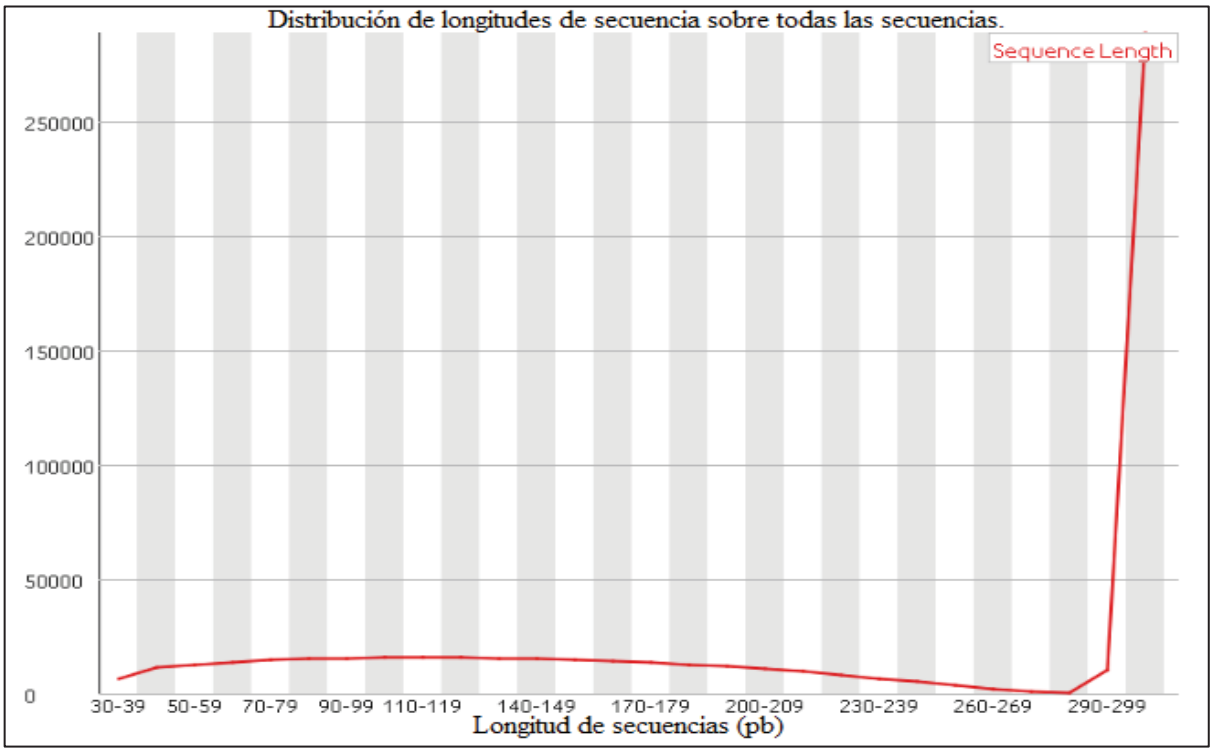


Figura N° 34. Distribución de la longitud de las secuencias del archivo reverse.

En cuanto a los niveles de secuencias duplicadas este módulo cuenta el grado de duplicación para cada secuencia en la biblioteca y crea un gráfico que muestra el número relativo de secuencias con diferentes grados de duplicación. Hay dos líneas: La línea azul muestra el porcentaje total de las secuencias y la línea roja muestra las secuencias deduplicadas. No observa picos de duplicación, es decir, se produjo aleatoriedad en la formación de las lecturas *forward* y *reverse* a lo largo del genoma de *Chromohalobacter salexigens* MP25462 (Figura N° 35 y 36).

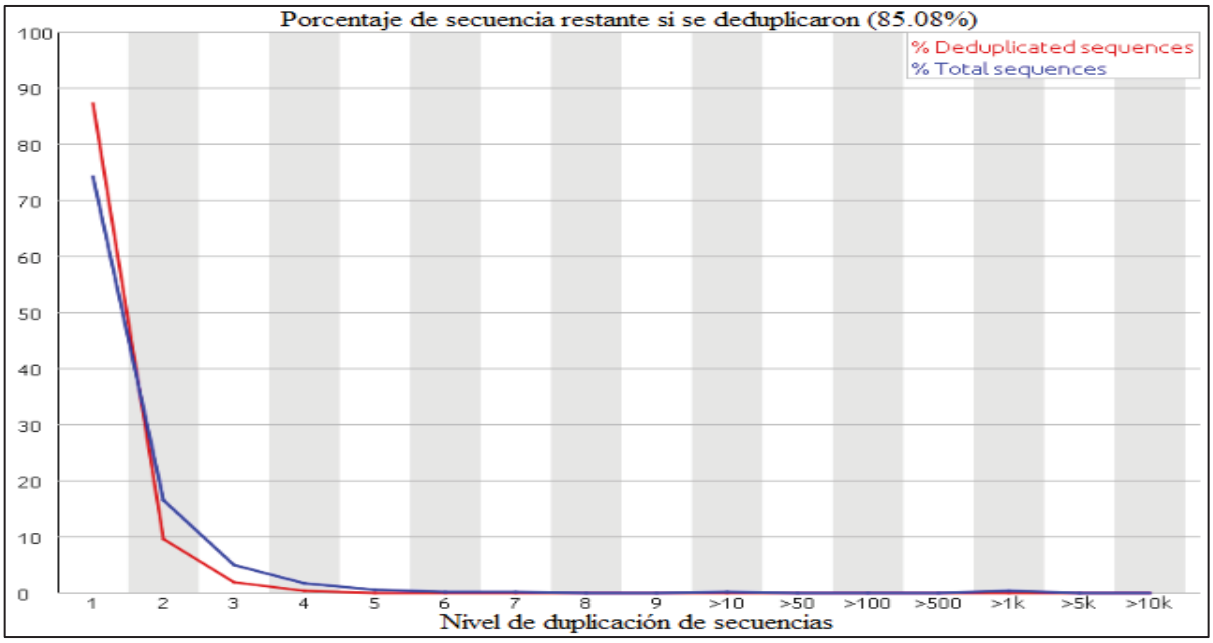


Figura N° 35. Porcentaje de niveles de secuencias duplicadas del archivo forward

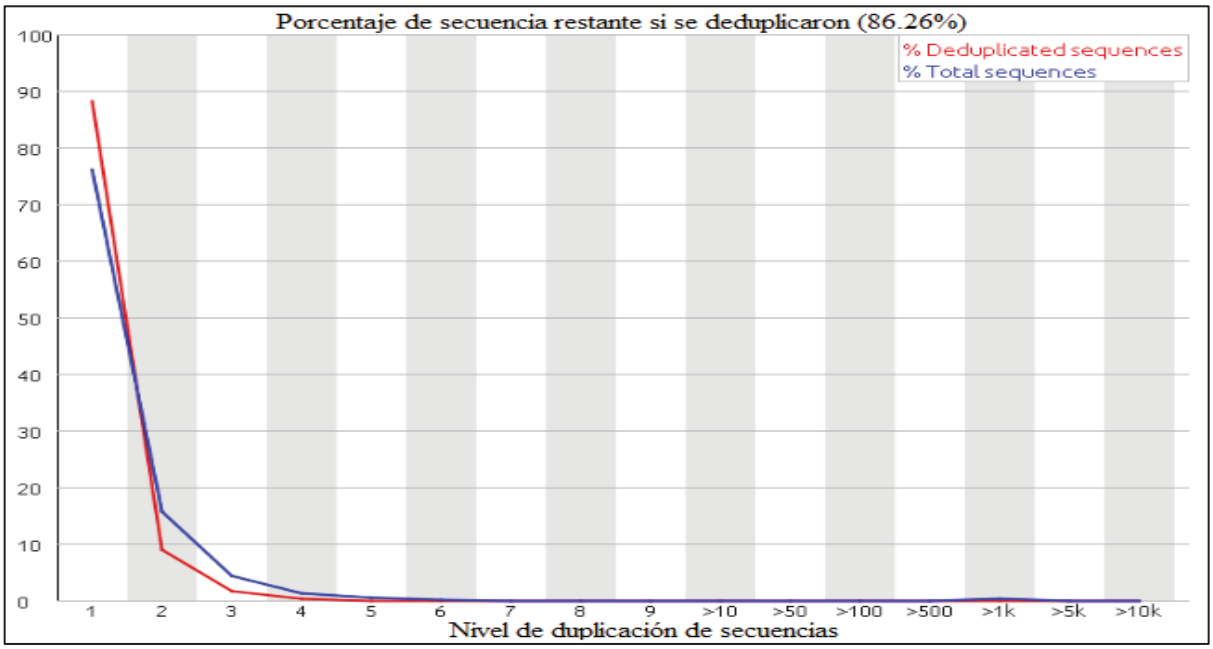


Figura N° 36. Porcentaje de niveles de secuencias duplicadas del archivo reverse

En cuanto las tablas N° 8 y 9 el módulo de **secuencias sobrerrepresentadas** mostró la valoración del número de secuencias que pueden dar problemas al realizar el ensamblado del genoma. En la biblioteca de MP25462 se encontró una secuencia con estas características con una posible fuente no identificada ya que este dato se obtiene al mapear el genoma con otro de referencia al momento de realizar la secuenciación masiva.

Tabla N° 8: Secuencias sobrerrepresentadas del archivo forward

Secuencia	Conteo	Porcentaje	Posible fuente
NNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNN	2639	0.44	No encontrada

Tabla N° 9: Secuencias sobrerrepresentadas del archivo reverse

Secuencia	Conteo	Porcentaje	Posible fuente
NNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNN	2644	0.44	No encontrada

En el contenido de los adaptadores se realizó la búsqueda de secuencias adaptadoras. En este caso no se encontró la presencia de estos en la biblioteca genómica *forward* y *reverse* (Figura N° 37 y 38).

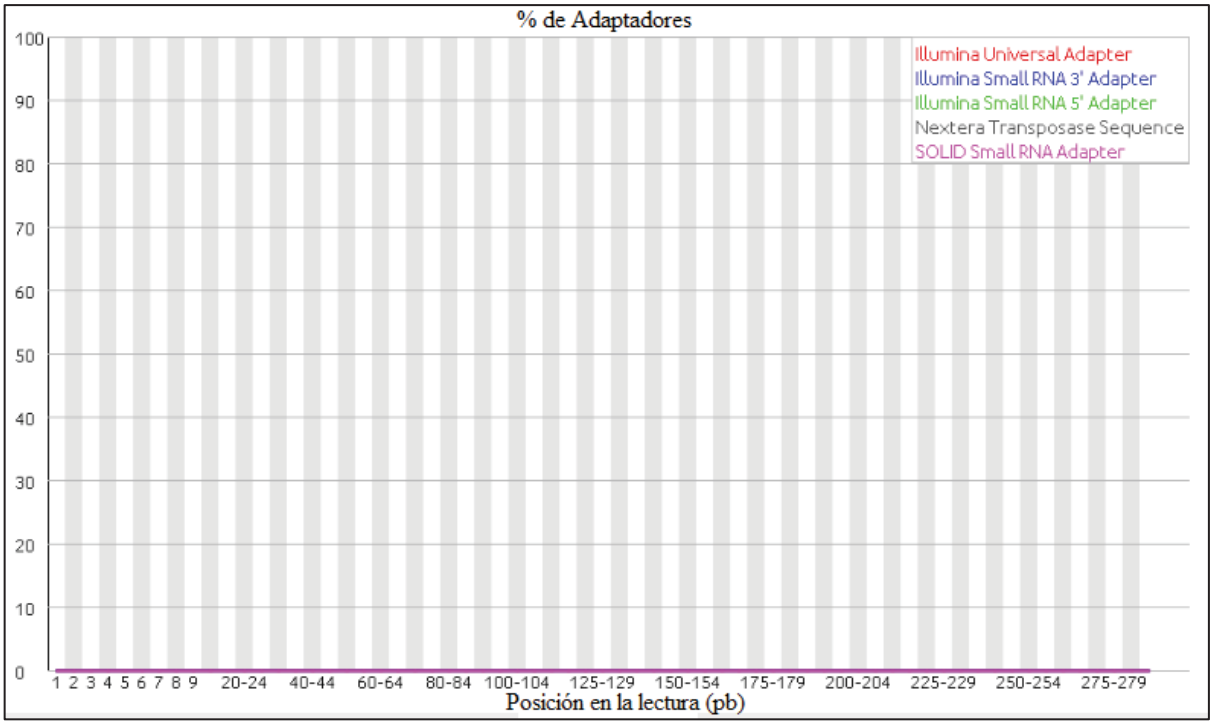


Figura N° 37. Contenido porcentual de los adaptadores de secuencias forward

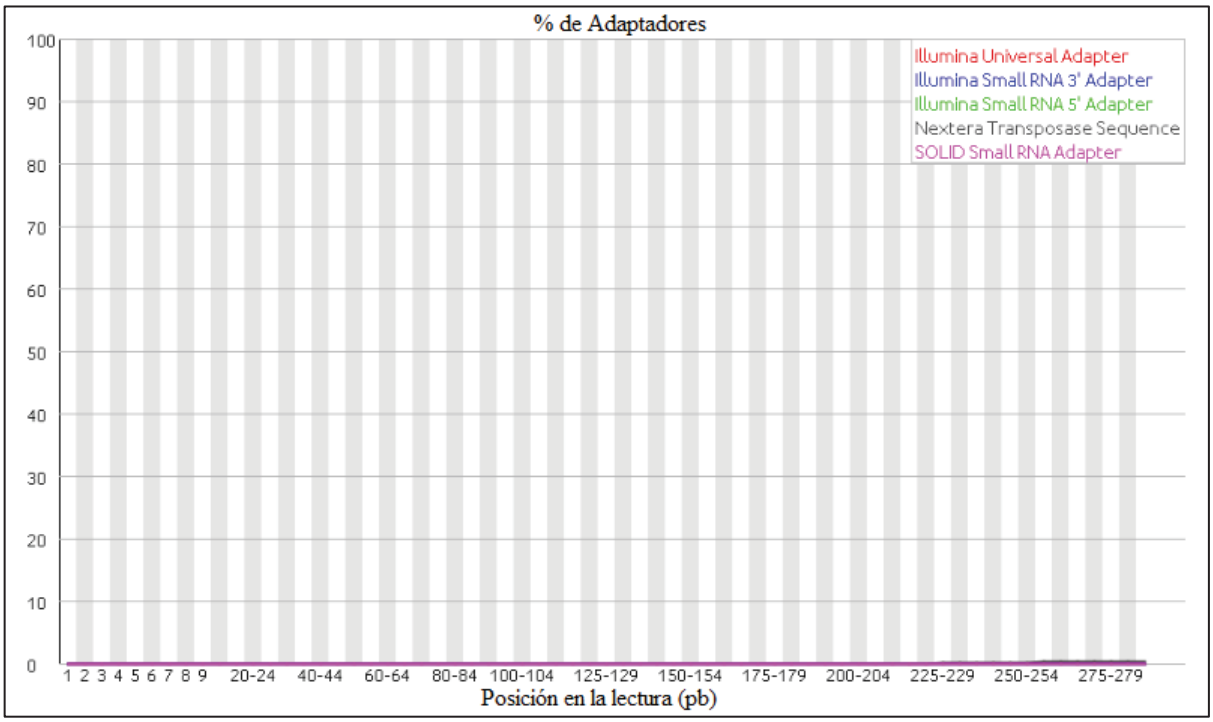


Figura N° 38. Contenido porcentual de los adaptadores de secuencias reverse.

En el apartado del contendido K-mer, indicó la existencia de subfragmentos de ADN de un tamaño menores a 7 pares de base para las lecturas *forward* y *reverse* como se aprecia en la parte superior derecha de las figuras N° 39 y 40.

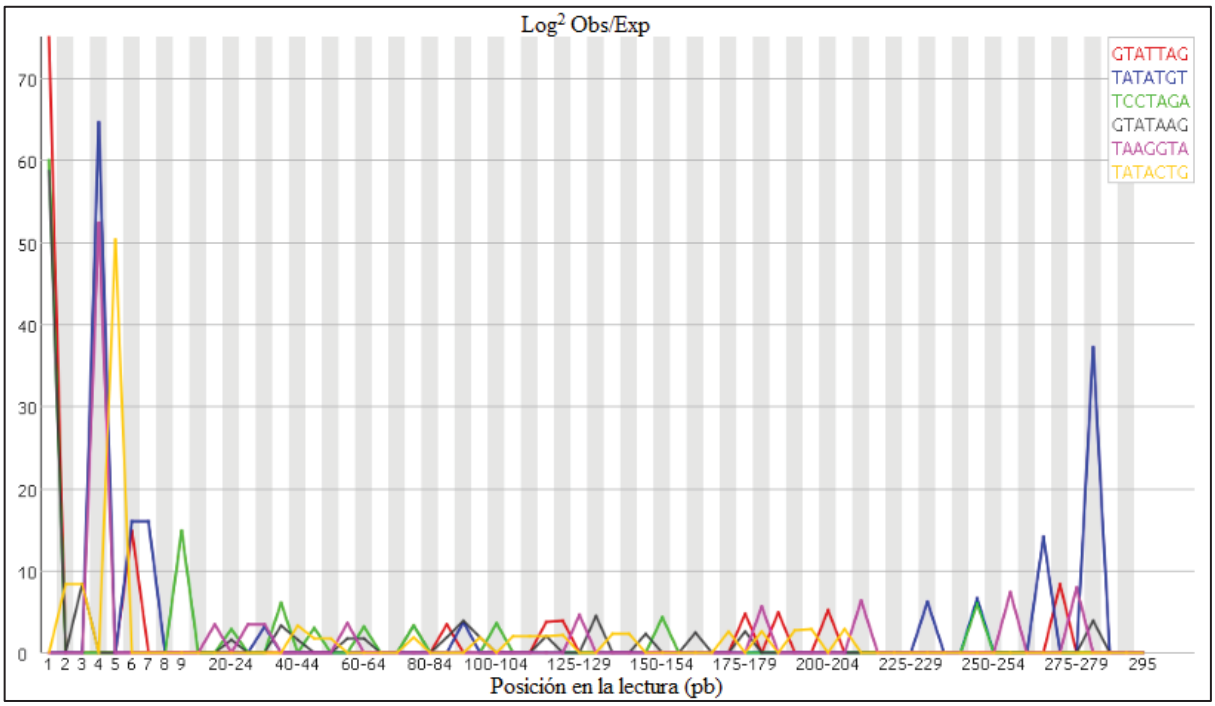


Figura N° 39. Contendido K-mer de secuencias forward.

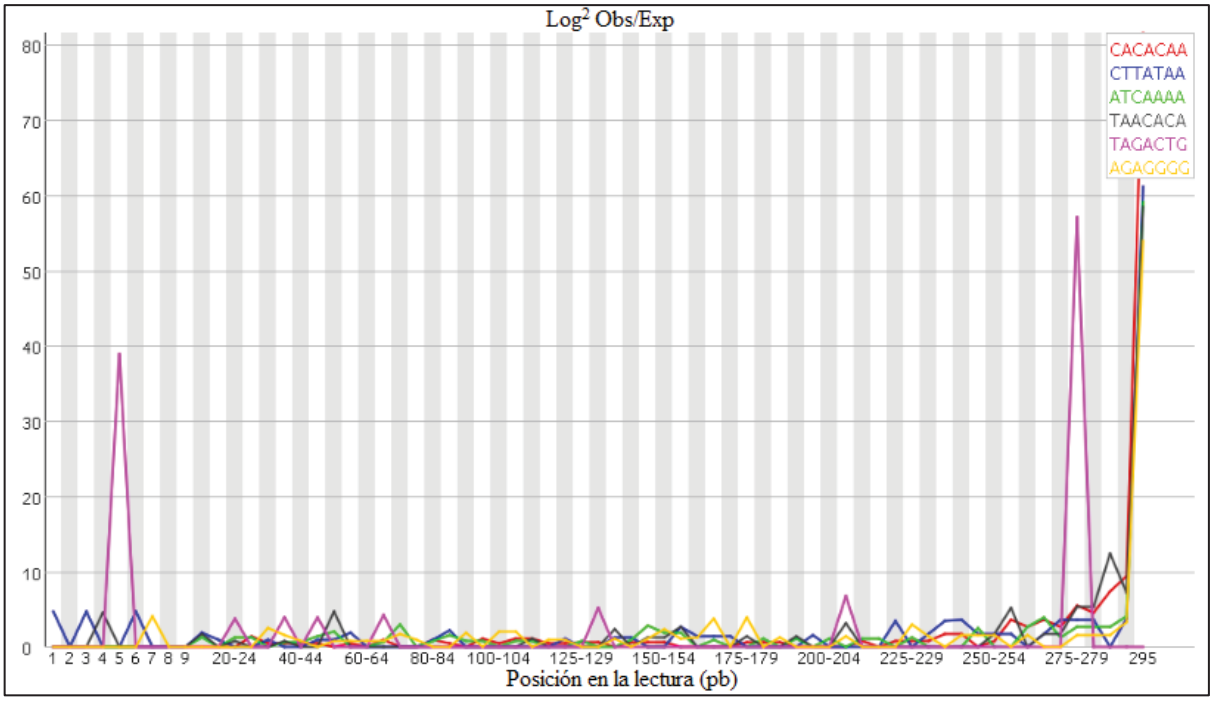


Figura N° 40. Contendido K-mer de secuencias reverse.

La valoración final de la biblioteca genómica de *Chromohalobacter salexigens* MP25462 tras el análisis con FastQC el producto de secuenciación masiva es alta, ya que siendo una secuenciación completa de un organismo esta presenta el 97.47 % de lecturas

presentan un índice phred mayor a 20. Es el archivo *reverse* donde existen una cantidad de menos de 30000 lecturas con índice Phred menor a 30 y que descienden hasta un índice Phred 14.

Análisis en Trimmomatic.

Un total de 1189904 lecturas de las cuales 594952 son secuencias *forward* y 594952 son *reverse* fueron introducidas y analizadas en el programa Trimmomatic. Se hizo el ensayo de recorte de las secuencias genómicas tomando en cuenta no perder el mayor número de secuencias es así que se obtuvieron los valores de más alta calidad con los siguientes parámetros utilizados: PHRED: 33, LEADING: 3, TRAILING: 28, HEADCROP: 20, MINLEN: 100 y SLIDINGWINDOW: 4:15. Donde:

- SLIDINGWINDOW: Realizar una ventana deslizante de recorte, corte una vez que la calidad media dentro de la ventana cae por debajo de un umbral.
- LEADING: Corta las bases el inicio de una lectura, si está por debajo del umbralde calidad.
- TRAILING: Corta las bases al extremo de la lectura, si está por debajo del umbral de calidad.
- HEADCROP: Corta el número especificado de bases desde el comienzo de la lectura
- MINLEN: Excluye la lectura si está por debajo de una longitud especificada
- TOPHRED33: Convierte las puntuaciones de calidad de Phred-33

El proceso de Trimmomatic se completó exitosamente llegando a prevalecer 406511 lecturas (68.33%) del total de las lecturas. Para su visualización de estos resultados del recorte de calidad de las secuencias se hizo uso de FastQC.

Los módulos que se tomaron en cuenta para el análisis después del recorte de calidad de las secuencias hecho por Trimmomatic fueron: el control de calidad de las bases secuenciadas, contenido de la secuencia por base y contendido K-mer porque que estos módulos en el análisis inicial con FastQC presentaron índices bajos.

En el diagrama de cajas del control de la calidad de las bases secuenciadas para la biblioteca *forward* se obtuvo que el tamaño de las secuencias presentan una longitud de hasta 281 pb. El índice phred va desde 25 hasta 38, la mediana de 30 a 38 y la calidad media desde 29 a 38; presentándose de esta forma casi el total de lecturas en la parte superior verde que representa una alta calidad (Figura N° 41). En el archivo *reverse* después del recorte de calidad presento lecturas de una longitud de hasta 277 pb teniendo una calidad phred desde 19 hasta valores de 38, mediana de 25 a 38 y calidad media de 25 hasta 37 a lo largo de todo el

secuenciado reverse. La calidad de las lecturas se ubica en un estado intermedio (naranja) y alto (verde) (Figura N°42).

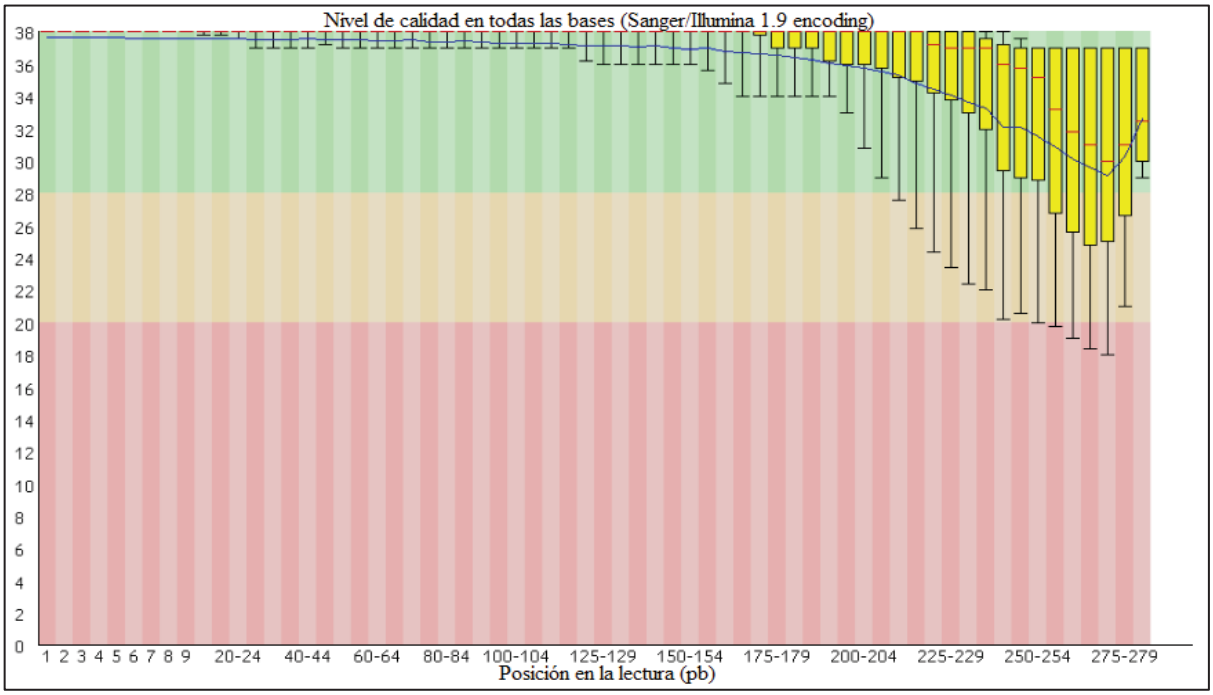


Figura N° 41. Calidad de las bases secuenciadas forward después del recorte de calidad. *En la figura no se llega a visualizar el tamaño 281 por criterios de diseño de diagrama del programa FastQC.

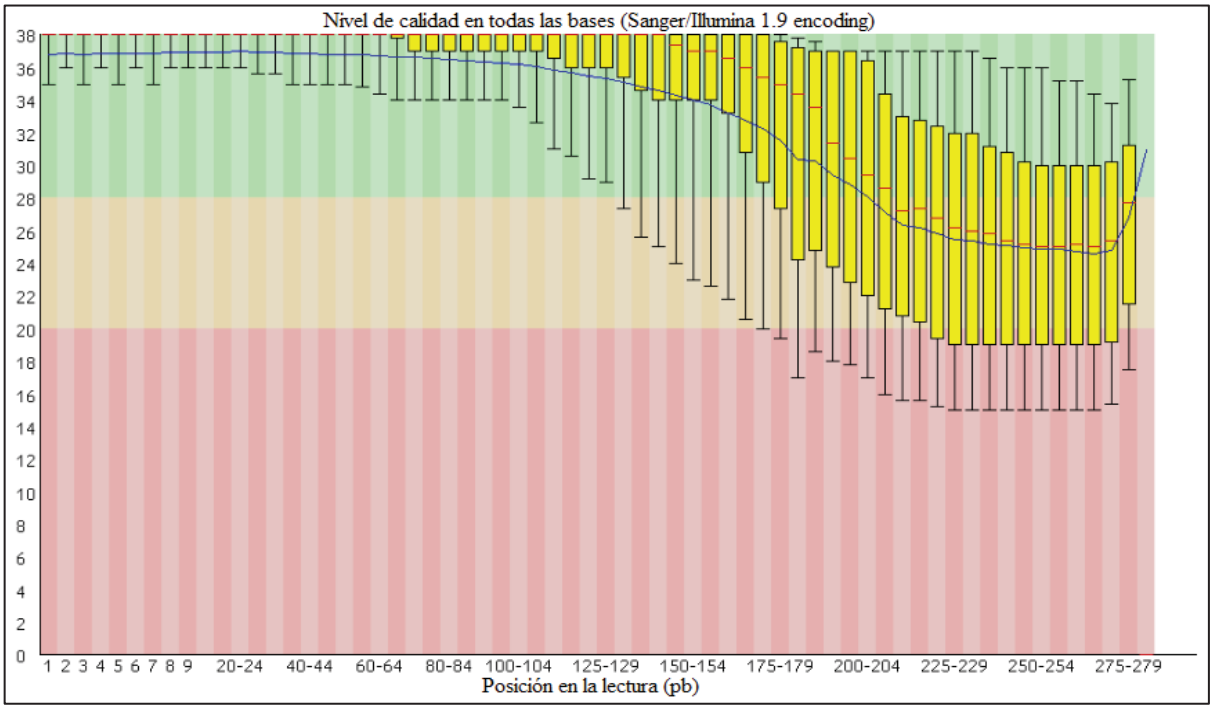


Figura N° 42. Calidad de las bases secuenciadas forward después del recorte de calidad

La proporcionalidad de las bases nitrogenadas en la biblioteca genómica se mantiene casi en su totalidad en *forward* y *reverse*, variando tan solamente en las lecturas de mayor longitud. Las pertenecientes a *forward* no son significativas ya que se observa que se trazan

líneas paralelas, teniendo una proporcionalidad entre las bases hasta lectura de longitud de 277 pb. pero decayendo la hacia el final de la secuenciación. Esto es común ver en este tipo de tecnologías que la calidad de secuenciación descienda hacia el final (Figura N° 43). Por otro lado, la proporcionalidad de bases en *reverse* describe dos líneas paralelas desde las lecturas cortas, al incremento del tamaño de estas comienza a observarse anormalidades en su representación de la proporcionalidad. Llegando a tener AT y GC en la misma proporcionalidad hasta secuencias de longitud 254 pb (Figura N° 44).

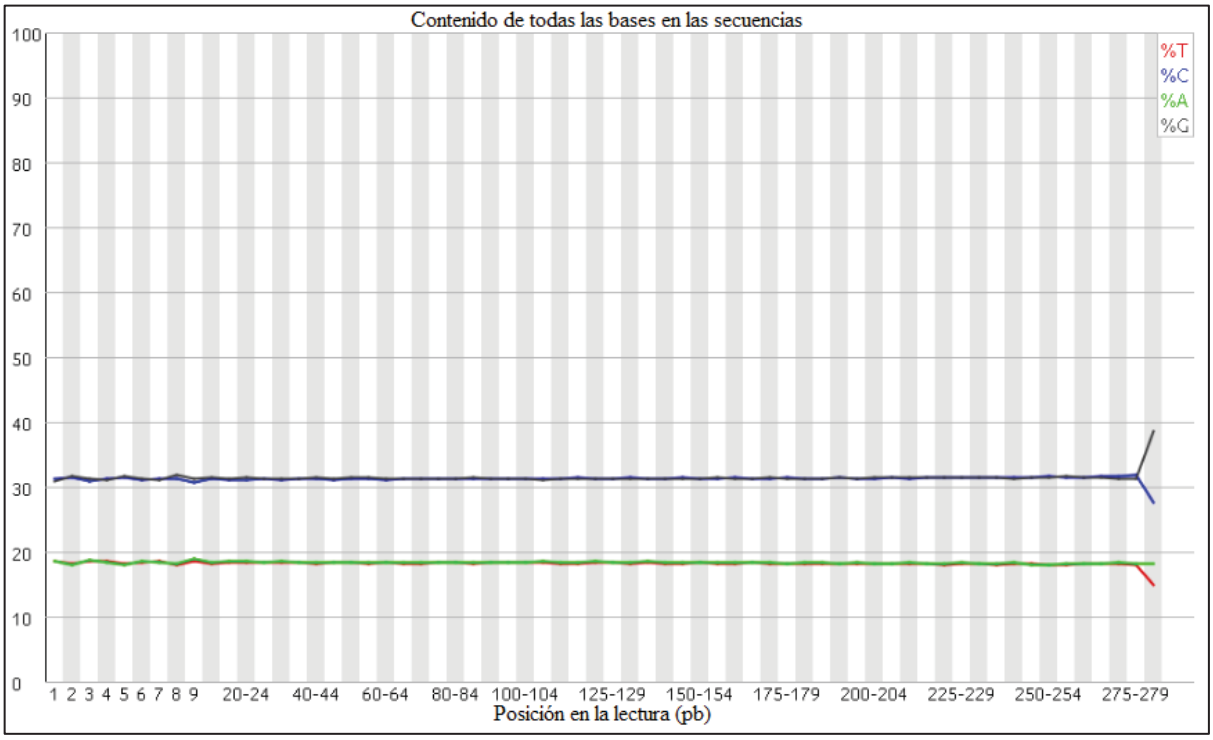


Figura N° 43. Contenido de la secuencia por base forward después del recorte de calidad

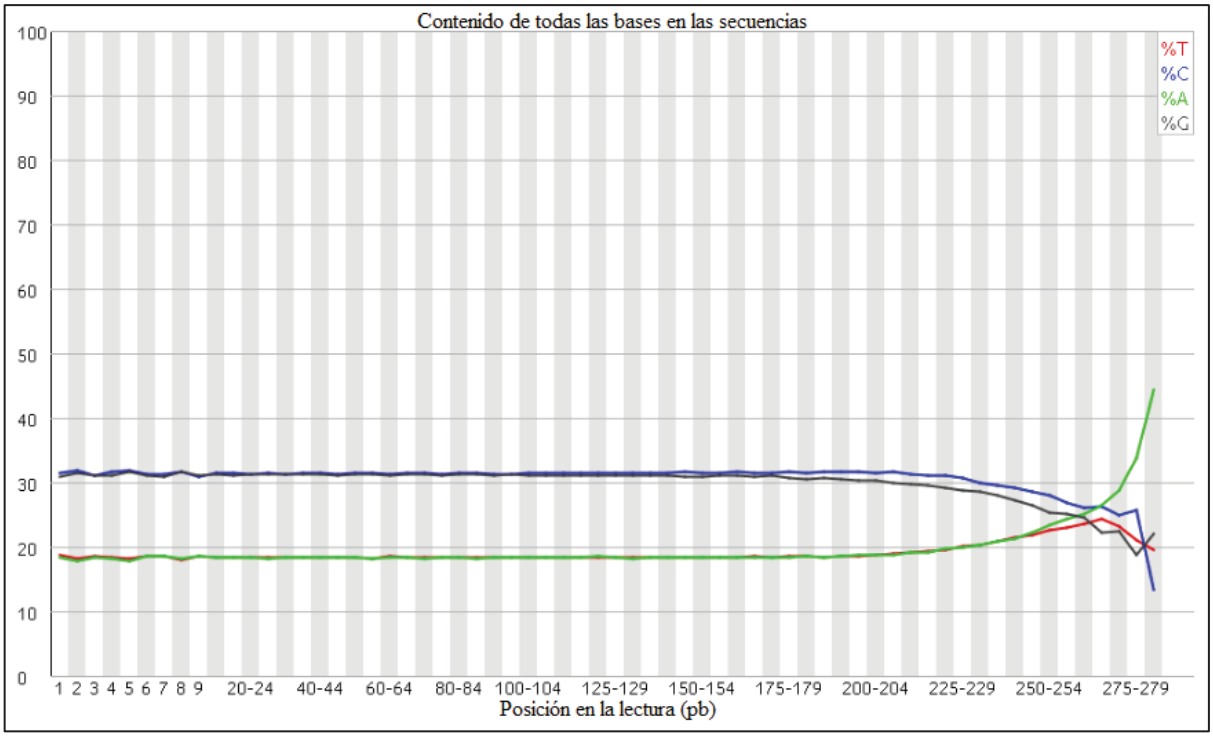


Figura N° 44. Contenido de la secuencia por base reverse después del recorte de calidad

El contenido de K-meros desciende en ambos componentes de la biblioteca genómica *paired-end.* En *forward* se observa la presencia de k-meros en lecturas de longitud de 275pb (Figura N° 45). En cuanto al archivo *reverse* no se observa la presencia de k-meros hasta la conformación de secuencias mayores de 254 pb (Figura N° 46).

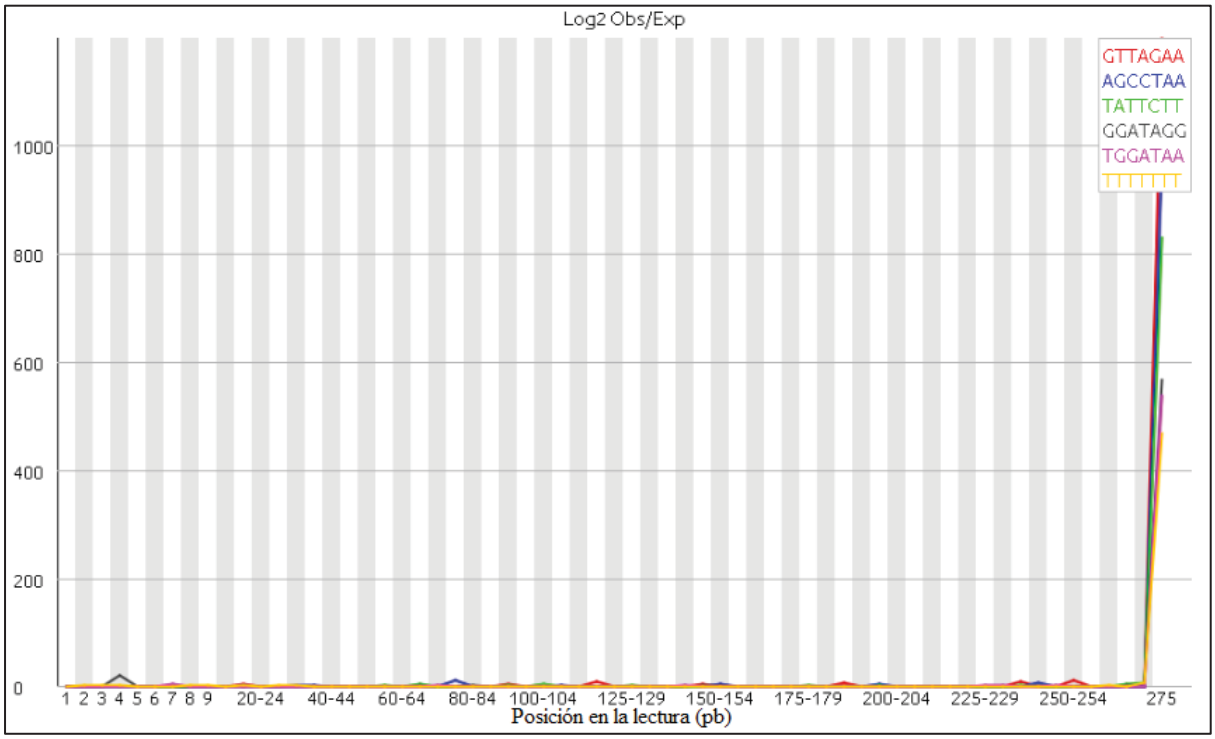


Figura N° 45. Contenido de K-mer de secuencias forward después del recorte de calidad

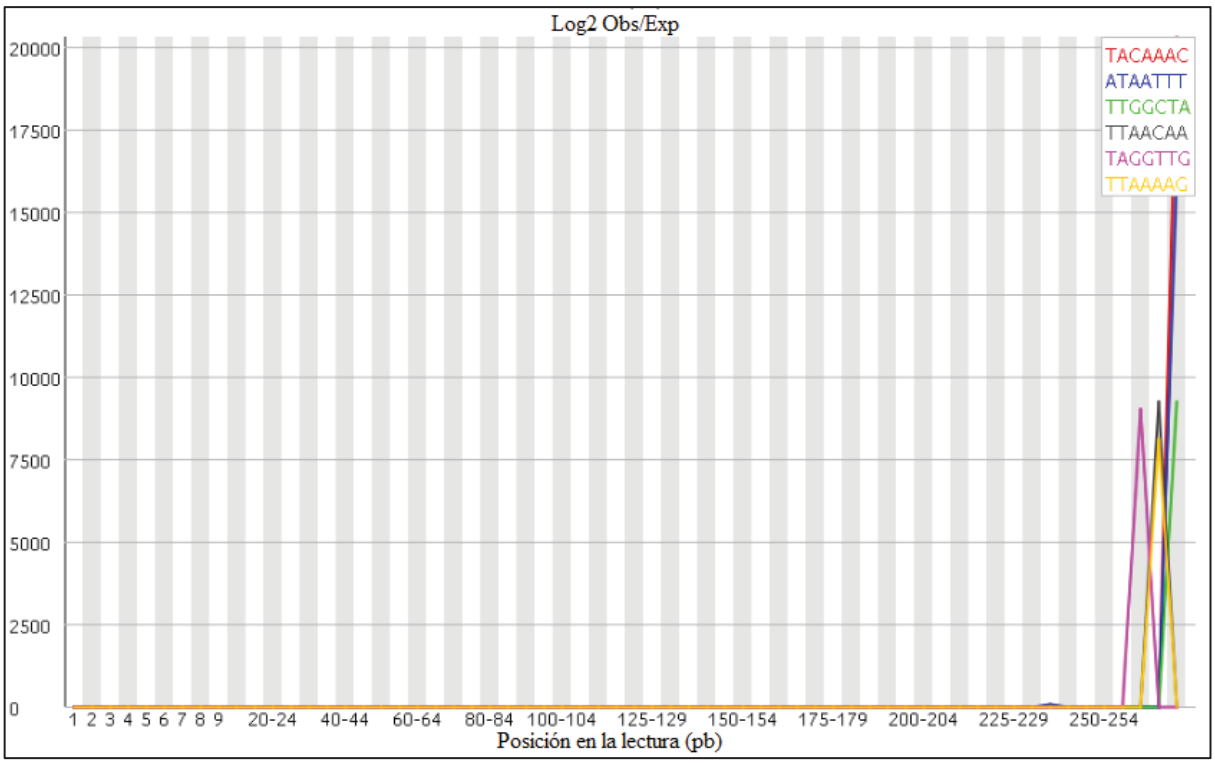


Figura N° 46. Contenido de K-mer de secuencias reverse después del recorte de calidad.

Se obtuvo 4 archivos al realizar el recorte por calidad de la biblioteca genómica. El resultado del análisis del programa Kmergenie otorgó el valor del tamaño de K-mer (k) el cual fue de 31. Tal programa presentó la comparación total de los valores k empleados con las secuencias genómicas de MP25462. En la figura N° 52 se aprecia que el valor k 31 logra producir un ensamble genómico de longitud mayor a 3680000 pb. Así también se observó la abundancia que logra cubrir el valor k 31 frente a otros valores de K-mer (Figura N° 47)

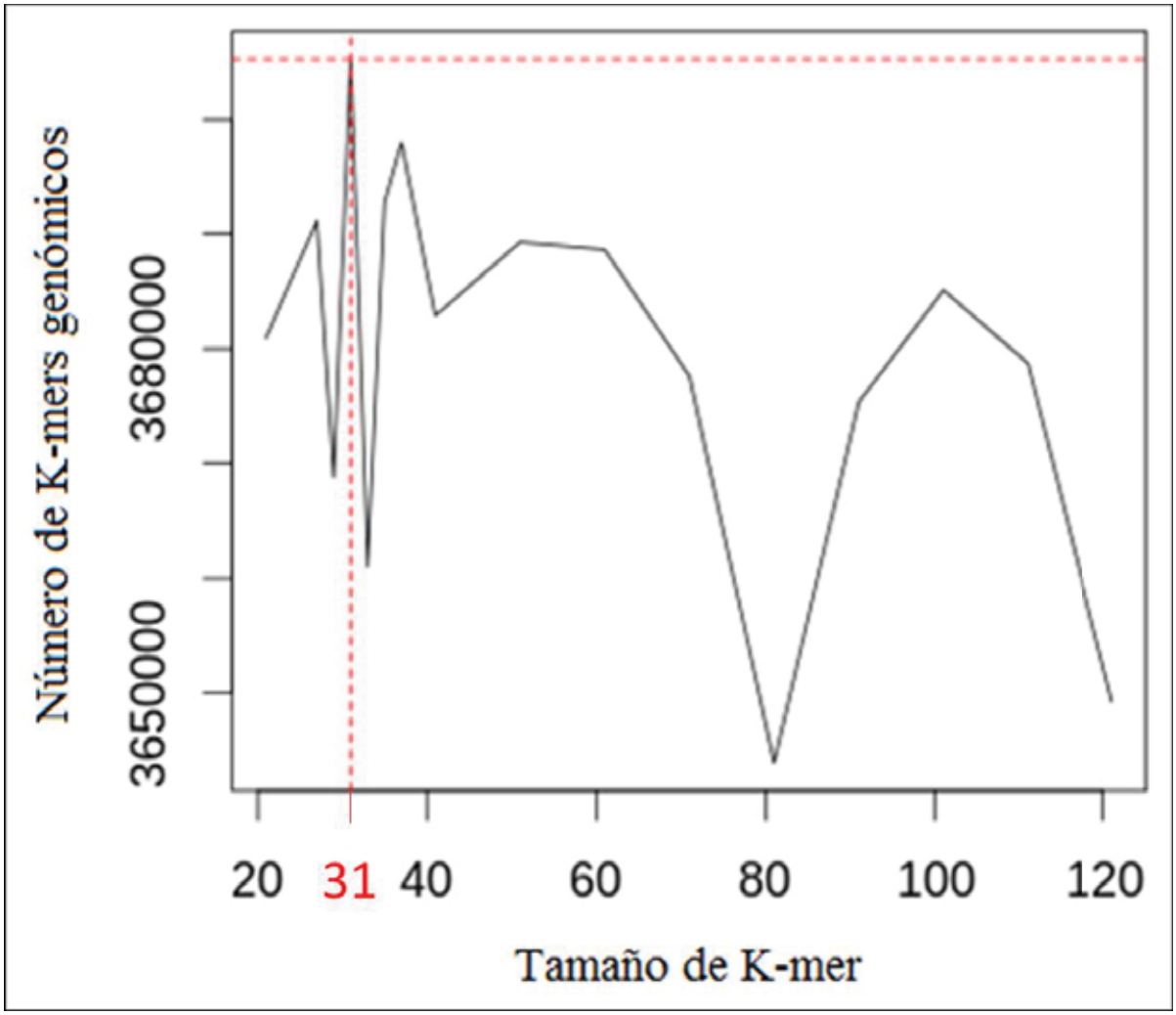


Figura N° 47. Trama principal producida por KmerGenie (# genomic k-mers vs. k). El eje x es el tamañode k- mers. El eje y es el tamaño estimado del genoma (en pb).

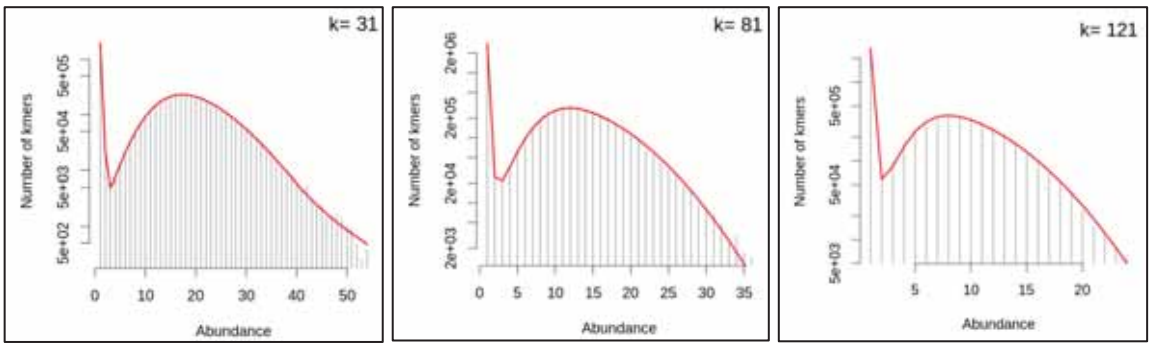


Figura N° 48. Comparación del mejor tamaño de K-mer (k=31) con otros valores escogidos indistintamente (k=81, k=121).

3.1.2. Análisis secundario

Los resultados de los ensamblados genómicos de *novo* en los programas Velvet, A5 y Spades se presentan en tablas, curvas y diagramas. En la Tabla N° 10 se muestra la longitud total del genoma ensamblado. Para el programa Velvet fue de 3´284,795; para A5 3´706,402 y Spades 3´712,216 pb. El número de *contigs* formados en el proceso del ensamblaje de *novo* fue

de 2147 con el ensamblador Velvet, 18 en A5 y 42 con Spades; siendo el ensamblador automático A5 quien presenta el mejor valor. En el parámetro del *contig* más grande para el ensamblador Velvet fue de una longitud de 8789 pb, para A5 fue de 620693 pb y en Spades de 491015 pb. El estadístico N50 en Velvet fue de 1886, en A5 482144 y Spades 447018.

En los tres ensambladores se presentó similares valores en el porcentaje de guanina y citosina que conforman el genoma: Velvet 64.05 %, A5 64.01 % y Spades 63,99 %. En la Tabla N° 10 menciona también los desajustes que presentan los ensambles realizados por los ya citados programas. Velvet y Spades no presenta bases N adicionadas en el genoma ensamblado de *novo,* mientras que en el ensamblador A5 existe la presencia de bases N en un numero de 18, en cuanto a la presencia de Ns en 100 kpb (cien mil pares de bases).

Anteriormente en el análisis primario con FastQC el contenido de GC fue de 63% mientras que en el análisis secundario con los ensambladores resulto: Velvet 64.05 %, A5 64.01 % y Spades 63,99 %. Siendo la media de los tres 64.02%, es decir, el contenido de GC aumentó en 1.02%. Esta variación se debe a que en la fase primaria se realizó el recorte de calidad con el programa Trimmomatic en donde se tomó en cuenta el número de adaptadores presentes a lo largo de la biblioteca genómica. Estos adaptadores son los responsables de esta variación en el contenido de GC probablemente porque la mayoría de estos *index* o adaptadores contenían en sus secuencias A y T. Después de su recorte y eliminación de estas cortas lecturas repetitivas, la proporcionalidad entre bases cambio. Siendo suficiente para que el porcentaje de GC aumente en 1.02%.

Tabla N° 10: Datos de los tres ensambles con los ensambladores de novo velvet, A5 y Spades.

Parámetros de calidad	Ensambladores de *novo*		
	Velvet	A5	Spades
Largo total	**3284795**	**3706402**	**3712216**
Longitud total (> = 0 pb)	3999969	3706402	3732734
Longitud total (> = 1000 pb)	2655898	3704878	3699448
Longitud total (> = 5000 pb)	177136	3704878	3691396
Longitud total (> = 10000 pb)	0	3692976	3691396
Longitud total (> = 25000 pb)	0	3681031	3654840
Longitud total (> = 50000 pb)	0	3681031	3654840
# contigs	**2147**	**18**	**42**
# contigs (> = 0 pb)	7295	18	89
# contigs (> = 1000 pb)	1288	16	21
# contigs (> = 5000 pb)	29	16	17
# contigs (> = 10000 pb)	0	14	17
# contigs (> = 25000 pb)	0	13	15
# contigs (> = 50000 pb)	0	13	15
Contig más grande	8789	620693	491015
N50	**1886**	**482144**	**447018**
N75	1136	191821	169102
L50	551	4	4
L75	1109	6	8
GC (%)	64.05	64.01	63,99
Desajustes			
# N's	**0**	**18**	**0**
# N por 100 kpb	**0**	**0.49**	**0**

Contig es la agrupación de secuencias o lecturas, N50 es la longitud del contigs que se encuentra en la mitad del total genómico, N75 es la longitud del *contig* que se encuentra en el 75% del tamaño total del genoma, L50 es el número de contigs cuya longitud sumada es N50, L75 es el número de contigs cuya longitud sumada es N75, #N´s es la sustitución de un nucleótido por una llamada N por carecer de confianza de asignación y #N por 100 kpb es el número de N´s por 100 kpb (100,000 pb).

Calidad de ensambles con el programa QUAST.

El visualizador de calidad de ensambles QUAST también proporciono diagramas para cada ensamble realizado donde se visualizó la curva que evidencia el número de contigs que forman parte del genoma ensamblado. En este caso se escogió el contig número 8 para los tres

ensambladores de *novo* y poder visualizar su ubicación en el genoma. En el ensamblador Velvet localizó el contig 8 en los 57421 pb (Figura N° 49), para A5 en 3150444 pb (Figura N° 50) y en Spades 2832236 pb (Figura N° 51) esto debido a que los ensambladores difieren en cuanto a su composición de contigs como se mencionó en la anteriormente.

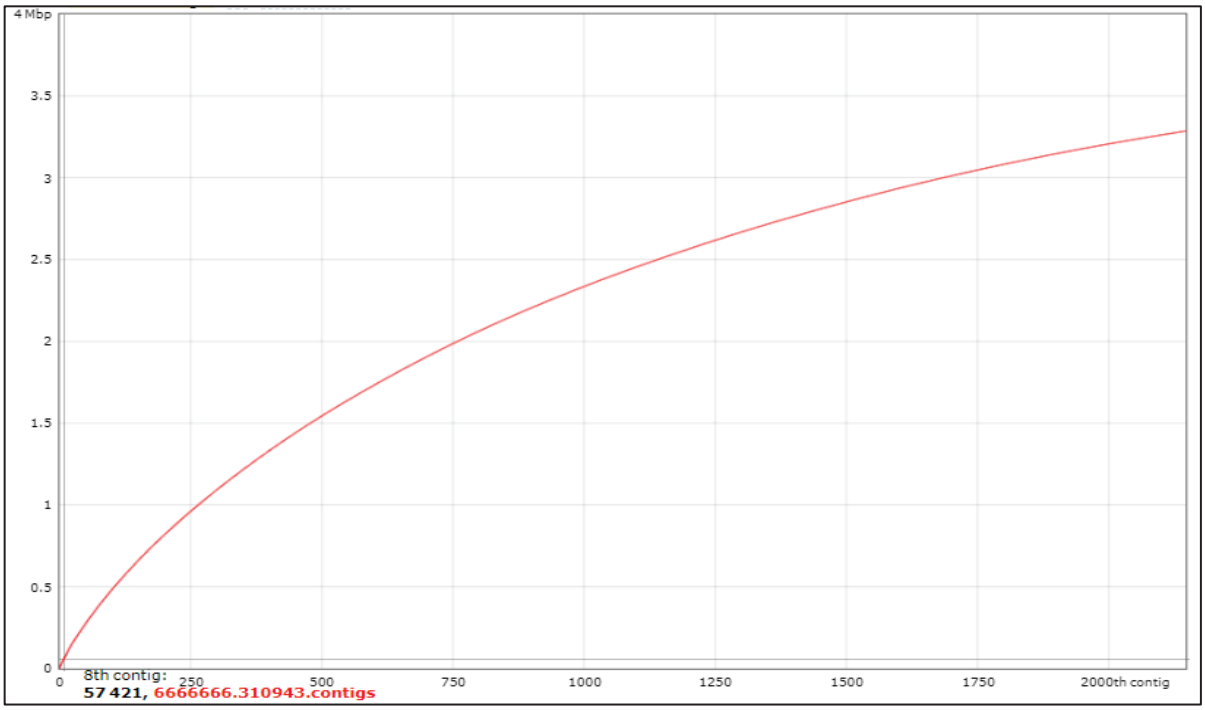


Figura N° 49. Longitud acumulada: Eje y longitud total del genoma vs eje x # de contigs con el programa Velvet.

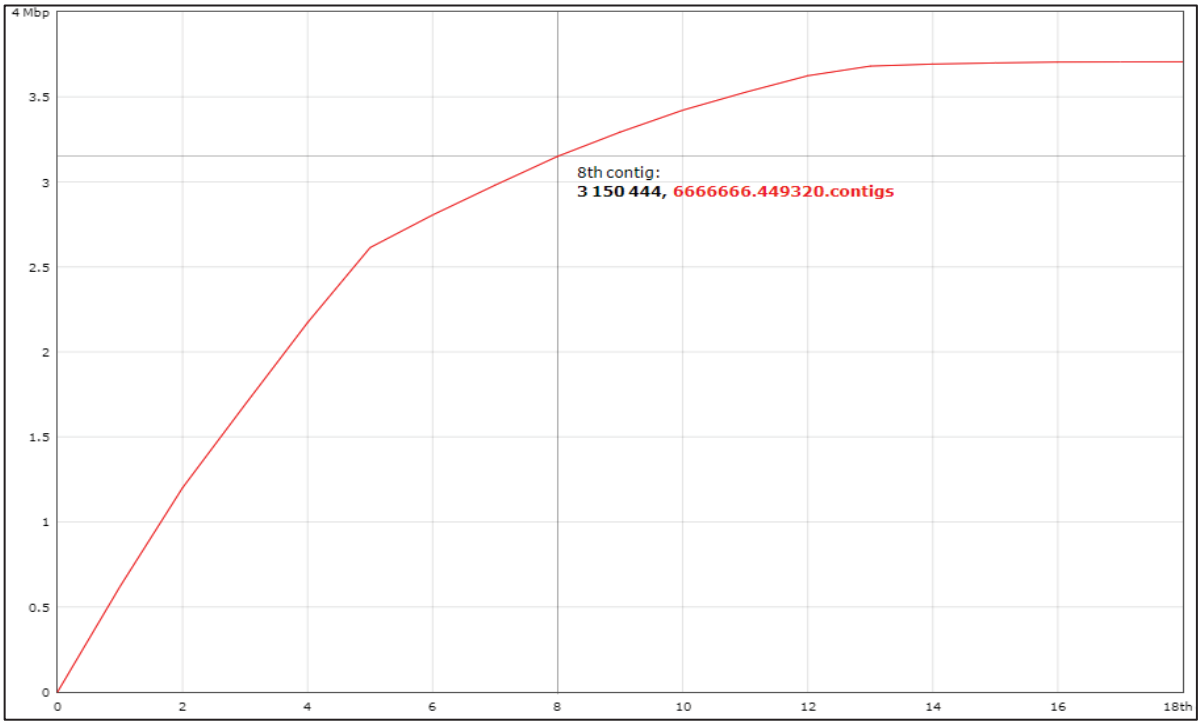


Figura N° 50. Longitud acumulada: Eje y longitud total del genoma vs eje x # de contigs con el programa A5.

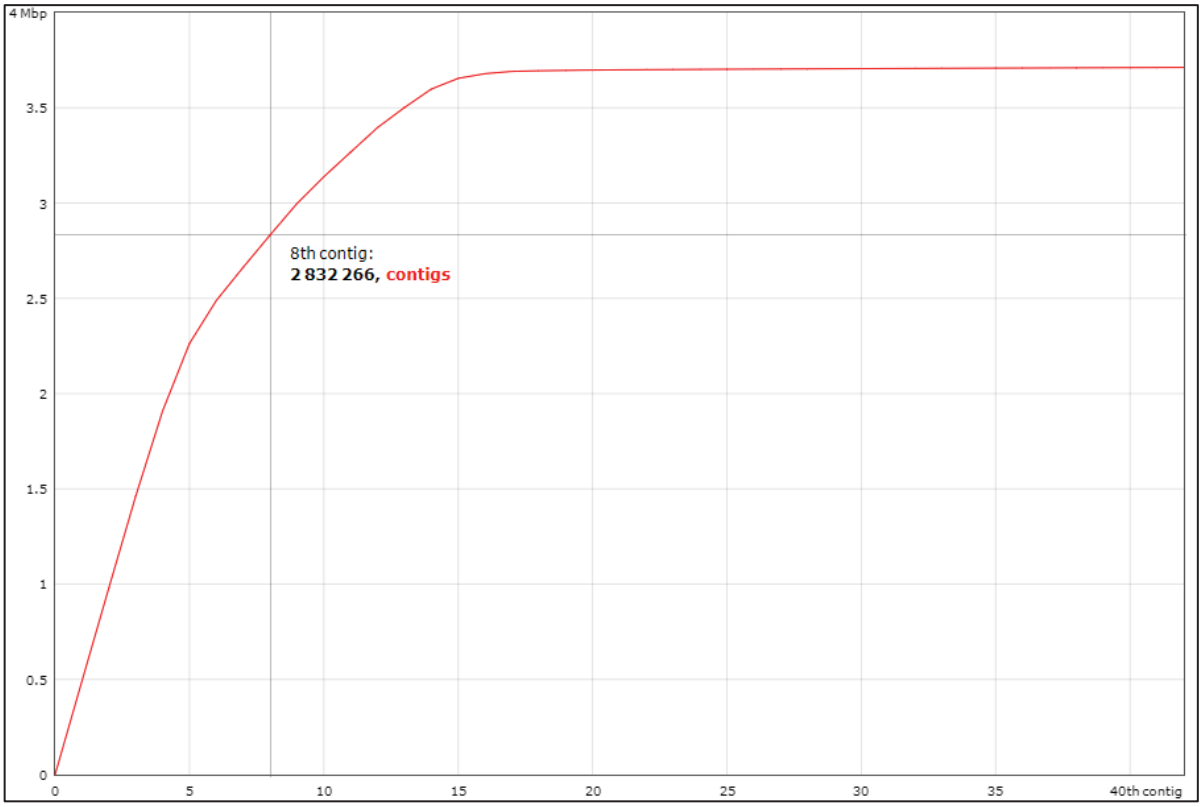


Figura N° 51. Longitud acumulada: Eje y longitud total del genoma vs eje x # de contigs con el programa Spades.

Así mismo la aplicación *Icarus Contigs Browser* mostró de una forma panorámica todos los contigs presentes en el genoma ensamblado con Velvet, A5 y Spades. Al lado superior izquierdo de las figuras N° 52, 53 y 54 se puede observar los valores como el largo total, número de contigs y el estadístico N50 los cuales ya mencionamos.

En este formato se puede navegar en los diferentes *contigs* y saber su tamaño, ubicación y cobertura. Además se puede observar la ubicación de los valores estadísticos N50 y N75.

Para el caso del ensamblaje con Spades se tiene un número de contigs igual a 42 mientras que en la tabla de parámetros nos da un valor de 89. Tal desigualdad se debe a que todas las estadísticas se basan en contigs de tamaño > = 500 pb, a menos que se indique lo contrario (por ejemplo, "# contigs (> = 0 pb)" y "Longitud total (> = 0 pb)" incluyen todos los contigs).

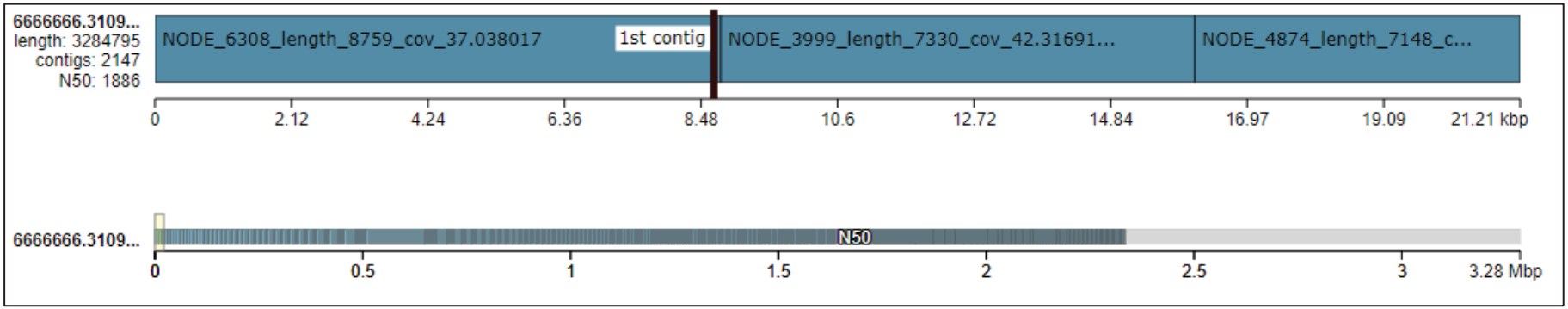


Figura N° 52. Visualización de contigs del genoma ensamblado con Velvet en Icarus Contigs Browser.

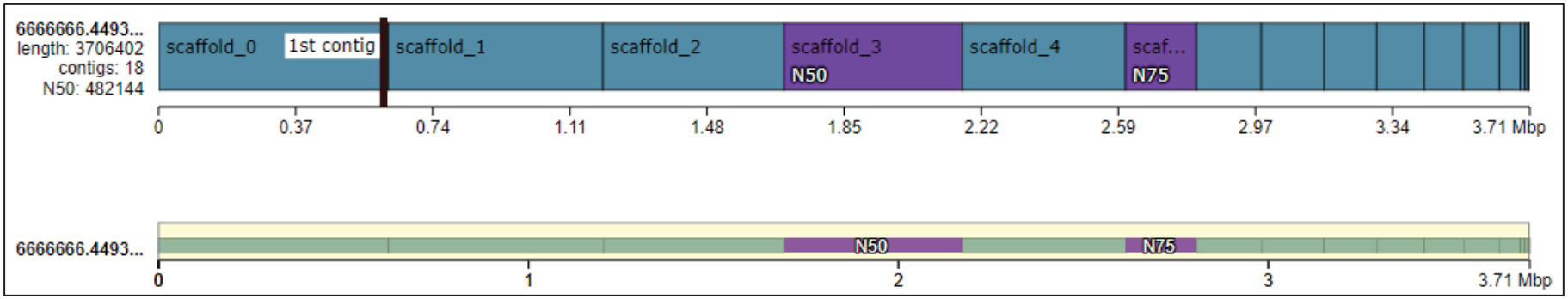


Figura N° 53. Visualización de contigs del genoma ensamblado con A5 en Icarus Contigs Browser.

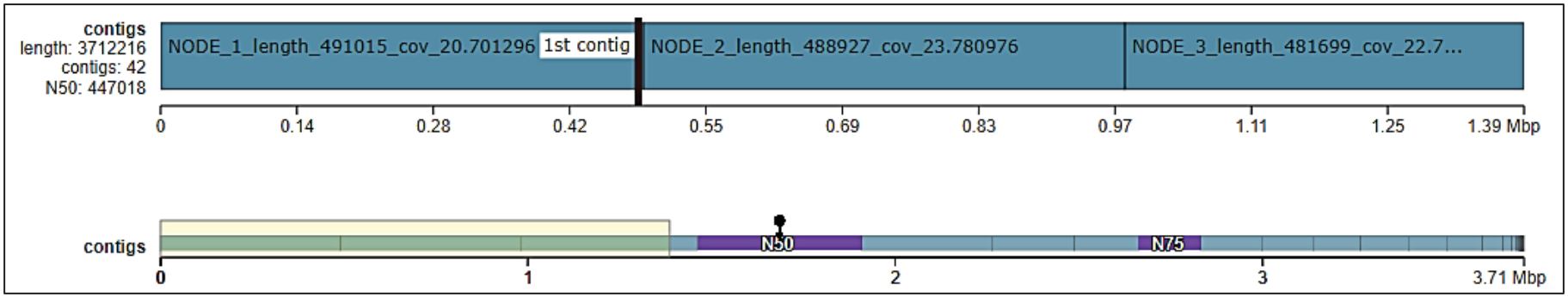


Figura N° 54. Visualización de contigs del genoma ensamblado con Spades en Icarus Contigs Browser.

La evaluación del ensamblado genómico define al ensamblador de novo Spades como el mejor ya que este no presenta una cantidad alta de contigs y además por no mostrar presencia de secuencias N´s a lo largo del genoma. La información del genoma reconstruido con Spades se utilizó para continuar con el análisis terciario.

3.1.3. Análisis terciario.

El servidor de anotación RAST realizó la anotación del archivo FASTA producto del ensamblador de *novo* Spades. El análisis funcional del genoma completo de MP25462 predijo 3402 genes. La cantidad de secuencias que codifican ARN fue de 69. El contenido de G + C proporcionado por RAST fue de 63.99 % este dato se compara con otras especies de *Chromohalobacter morismortui* con un contenido de G + C de 63.5 % (Ventosa et al., 1989), *Chromohalobacter salexigens* 1H11 con un contenido de G + C de 63.91 % (Copeland et al., 2011) .

Los 3402 genes codificantes fueron organizados en 486 de subsistemas. Los subsistemas son agrupaciones de genes resultado del servidor de anotación RAST. La cobertura de los genes que están dentro de los subsistemas fue de 59.52% y 40.48% secuencias codificadoras que no están dentro del mismo de acuerdo a los estándares de RAST. Los genes dentro del subsistema fueron un total de 2025, de los cuales 1909 son genes no hipotéticos y 116 genes hipotéticos. Los genes no presentes en el subsistema fueron de un total de 1377, de los mismos 781 son no hipotéticos y 596 hipotéticos

En la figura N° 55 se observa hacia el lado derecho el listado de 27 funciones de los subsistemas generados por RAST tomando en cuenta sus criterios. En este formato se puede comprobar la composición de la pared celular y determinar si la bacteria es Gram – o +. En la segunda subcategoría de pared celular y capsula existen 133 genes implicados de los cuales ninguno se asocia con bacterias grampositivas. Así también existen 9 genes implicados en la invasión y resistencia intracelular, resistencia a antibióticos y compuestos tóxicos donde existen 64 genes relacionados con la motilidad flagelar en procariotas, 2 genes para la dormancia y esporulación. Se observó 40 genes de respuesta ante el estrés osmótico y 68 genes para el estrés oxidativo. Y en cuanto a la función metabolismo de proteínas se encontró 263 genes implicados de los cuales 25 se asocian a proteínas plegables, 4 a selenoproteínas, 171 implicadas en la biosíntesis de proteínas, 24 en el procesamiento y modificación de proteínas y 37 en la degradación de proteínas. Esto por solo mencionar algunas funciones de los subsistemas.

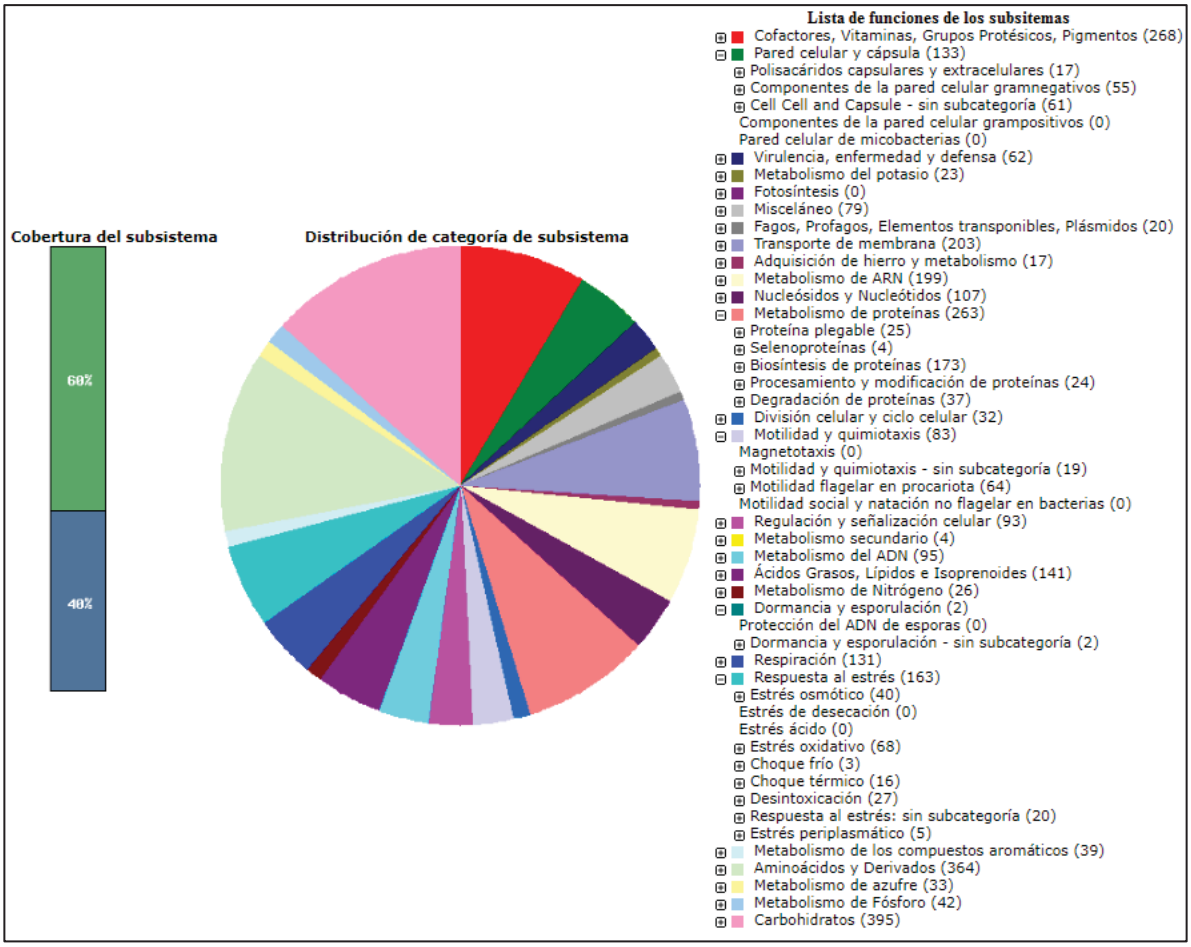


Figura N° 55. Subsistemas funcionales del genoma de Chromohalobacter salexigens MP25462 anotado del ensamble con el programa Spades.

La anotación con los tres ensambladores resulto diferenciándose en cuanto a la distribución de los genes en las 27 funciones de los subsistemas. En la tabla N° 11 se observa la variación de los genes en las distintas funciones encontradas para *Chromohalobacter salexigens* MP25462.

Tabla N° 11: Comparación de los genes agrupados en las 27 funciones de los subsistemas para los ensambladores Velvet, A5 y Spades.

Funciones de los subsistemas	Anotación con los ensambladores		
	Velvet	**A5**	**Spades**
Cofactores, vitaminas, grupos prostéticos, pigmentos	172	150	268
Pared celular y capsula	98	23	133
Virulencia, enfermedad y defensa	58	29	62
Metabolismo del potasio	20	4	23
Fotosíntesis	0	0	0
Misceláneo	41	34	79
Fagos, profagos, elementos transponibles, plásmidos	13	9	20
Transporte de menbrana	174	126	203
Adquisición de hierro y metabolismo	12	27	17
Metabolismo de ARN	142	42	199
Nucleósidos y nucleótidos	72	74	107
Metabolismo de proteínas	151	211	263
División celular y ciclo celular	26	0	32
Motilidad y quimiotaxis	35	15	83
Regulación y señalización celular	69	26	93
Metabolismo secundario	4	4	4
Metabolismo del ADN	75	63	95
Ácidos grasos, lípidos e isoprenoides	90	76	141
metabolismo de nitrógeno	19	19	26
Dormancia y esporulación	2	2	2
Respiración	88	96	131
Respuesta al estrés	131	73	163
Metabolismo de los compuestos aromáticos	24	29	39
Aminoácidos y derivados	306	255	364
Metabolismo de azufre	17	4	33
Metabolismo del fósforo	30	39	42
Carbohidratos	293	184	395

3.1.3.1. Descripción del gen aspartil aminopeptidasa de MP25462.

Para dar comienzo a la descripción del gen aspartil aminopeptidasa se realizó la búsqueda de aspartil aminopeptidasa. Como se mencionó en el apartado 3.1.1.3 la anotación del ensamblaje con el programa *Spades* de *Chromohalobacter salexigens* MP25462 contiene 263 genes implicados en el metabolismo proteico. Los ensambladores *Velvet* y *A5* también presentan genes implicados en tal función 151 y 211 genes respectivamente (Tabla N° 11). Al ingresar en este módulo funcional de proteínas en las tres anotaciones realizadas no se halló al gen que codifica al gen aspartil aminopeptidasa en *C. salexigens*. Producto del análisis terciario

y anotación del genoma con el servidor de anotación RAST, se generó un archivo en formato Excel XLS en las tres anotaciones realizadas (Anexo 8). Se realizó la descarga del archivo Excel XLS de la anotación con los ensambladores *Spades, Velvet* y *A5* para buscar la existencia del gen aspartil aminopeptidasa en los datos del archivo. Los nombres de los archivos anotados para el programa *Spades, Velvet* y *A5* fueron 6666666.448892, 6666666.449320 y 6666666.310943 respectivamente. Los tres archivos Excel XLS fueron ejecutados buscando el nombre del gen en inglés *Aspartyl aminopeptidase* teniendo como resultado la ubicación, el inicio y final del gen, la cobertura de su secuencia en el genoma, su sentido en el genoma y su composición en nucleótidos (Figura N° 56) y aminoácidos (Figura N°57).

La cobertura del gen aspartil aminopeptidasa resulto de 20.7 X. Se ubica en la hebra antisentido del ADN. Dentro del genoma de MP25462 comienza en el nucleótido 71771 y termina en el 70473 teniendo una longitud de 1299 pares de bases traduciéndose a una secuencia de 432 aminoácidos (Figura N° 58).

La dirección de expresión génica se presenta en la figura N° 58. Siendo la cadena sentido 5'→3' de izquierda a derecha y la antisentido 3'→5'de derecha a izquierda. El gen aspartil aminopeptidasa esta flanqueado por el gen de la proteína de unión a GTP y ácido nucleico YchF (azul), su ubicación en el genoma va desde el 69109 al 70200 nucleótido con una longitud de 1092 pb o 264 aminoácidos. Entre el gen de unión a GTP/ácido nucleico YchF y el gen aspartil aminopeptidasa se ubica el ARNt-Met-CAT (magenta) comenzado en el nucleótido 70288 y terminando en 70364 siendo de un tamaño de 77 pb. El espacio intergénico entre el gen aspartil aminopeptidasa y el ARNt es de 109 pb y entre el ARNt y el gen para la proteína de unión a GTP es de 88 pb. Por el lado derecho de la aspartil aminopeptidasa se presenta el gen que codifica para una proteasa periplásmica que comienza en 72066 y termina en 74162 con un tamaño de 2097 pb o 699 aminoácidos. El intervalo de la zona intergénica entre ambos genes es de 295 pb. El gen aspartil aminopeptidasa no forma parte de un operón.

La descripción del gen aspartil aminopeptidasa con los datos del ensamblador *A5* resulto muy similar al realizado con la anotación del ensamblador *Spades.* El tamaño del gen aspartil aminopeptidasa, composición de nucleótidos y aminoácidos, dirección antisentido en el genoma, tamaño de espacios intergénicos, genes aledaños a la aspartil aminopeptidasa son los mismos. La única variación entre estos datos es la ubicación de la aspartil aminopeptidasa y genes contiguos en el genoma de *Chromohalobacter salexigens* MP25462. Existiendo una variación de 5 nucleótidos en su ubicación por ejemplo: Si aspartil aminopeptidasa de *C. salexigens* tiene comienzo en el nucleótido 71771 de la cadena antisentido con los datos de la

anotación con el ensamblador *Spades*, con el programa *A5* el gen aspartil aminopeptidasa comienza en el nucleótido 71776 (Figura N°59).

Los resultados con el programa *Velvet* fueron muy diferentes a los ensambladores *Spades* y *A5*. La aspartil aminopeptidasa se encuentra en la cadena sentido de ADN con una cobertura de 33.11 X. Tiene un tamaño de 1299 nucleótidos traducidos a 432 aminoácidos. En el genoma de *C. salexigens* MP25462 la aspartil aminopeptidasa comienza en el nucleótido 362 y termina en el 1660 y este no se encuentra flanqueado por ningún gen (Figura N° 60).

```
>fig|6666666.448892.peg.708 Aspartyl aminopeptidase [Chromohalobacter salexigens
MP25462]
atggctcacgcccccacccttgaccgtttactgcattttctcgagcgctcgcccacaccctggcatgccgtcgacaacat
ggcccggcggctcgaacaggcagggtatcggcgactcgaggaaaccgaagcgtggcaattggcgcccggcgatcgtttct
acgtcacgcgcaacgcgtcgtcattgatcgccatgcaggtgccgacggatcccttgagcgggctgcgcatgatcggggcg
cataccgacagcccggggttgcgcttgaaaccccaacccgtggtggccaagaaggattggctgcagttgagcgtcgaggt
ctacggcggtgcgctgctggcaccatggttcgaccgcgatctggggctggccgggcgcatccatgtacgacgcgaggatg
gacgcttgcagggcgtattattgcatgtcgatcgtcccgtcgcgatcattcccagcctggccatccacctggatcgcgag
gccaacaacgggcgtgccctgaatgcccagacgcagatgctgccggtcgtgctgcaaggcggtggcgaagccgatctcga
gcgttggctcaagcgctggctgtacgaacagcatgggctggagaacattcagttgttggattacgaactctcgctttacg
acatgcagcggccgtcgcgtgtcgggatcgaggggggaactgatcgccagtgcgcgcctcgacaatctgctgtcgtgcttc
accggtatcgaggccttgctggccggcgacgggcgacaggggcgctctttgtcgccaacgatcacgaagaagtgggcag
tgccagtgcgtgcggcgcccagggccccttcctgggagatgtgctgcgtcgcgtgcatgcgcaactgggtgagggcggcg
aagacggctgggtgcgtctgatccagggctcgcgcatgatttcctgcgacaacgcccatgccgtgcaccccaactttccc
gagaaacacgacgaacaccacggcccggcgatcaatggcgggcccgtgatcaaggtgaacgccaaccagcgctatgccac
caacagcgcgacagcggccatgttccgggatatctgtcgcgaggcaggaacacccgtgcagaccttcgtgacgcgtgccg
acatgggctgcggcagcaccatcgggcccatcaccgccaccgaactcggggtgccgacgctggatgtgggtatcccgcag
tggggcatgcactcgattcgcgaaaccgccggcagccgcgatgccgattacttgatccgcgcgctgacggccttcgtcaa
tcgcaccgagctggactag
```

Figura N° 56. Secuencia de nucleótidos en formato FASTA del gen aspartil aminopeptidasa de Chromohalobacter salexigens MP25462.

```
>fig|6666666.448892.peg.708 Aspartyl aminopeptidase [Chromohalobacter salexigens
MP25462]
MAHAPTLDRLLHFLERSPTPWHAVDNMARRLEQAGYRRLEETEAWQLAPGDRFYVTRNASSLIAMQVPTDPLSGLRMIGA
HTDSPGLRLKPQPVVAKKDWLQLSVEVYGGALLAPWFDRDLGLAGRIHVRREDGRLQGVLLHVDRPVAIIPSLAIHLDRE
ANNGRALNAQTQMLPVVLQGGGEADLERWLKRWLYEQHGLENIQLLDYELSLYDMQRPSRVGIEGELIASARLDNLLSCF
TGIEALLAGDGRQGALFVANDHEEVGSASACGAQGPFLGDVLRRVHAQLGEGGEDGWVRLIQGSRMISCDNAHAVHPNFP
EKHDEHHGPAINGGPVIKVNANQRYATNSATAAMFRDICREAGTPVQTFVTRADMGCGSTIGPITATELGVPTLDVGIPQ
WGMHSIRETAGSRDADYLIRALTAFVNRTELD
```

Figura N° 57. Secuencia de aminoácidos en formato FASTA de la aspartil aminopeptidasa de Chromohalobacter salexigens MP25462.

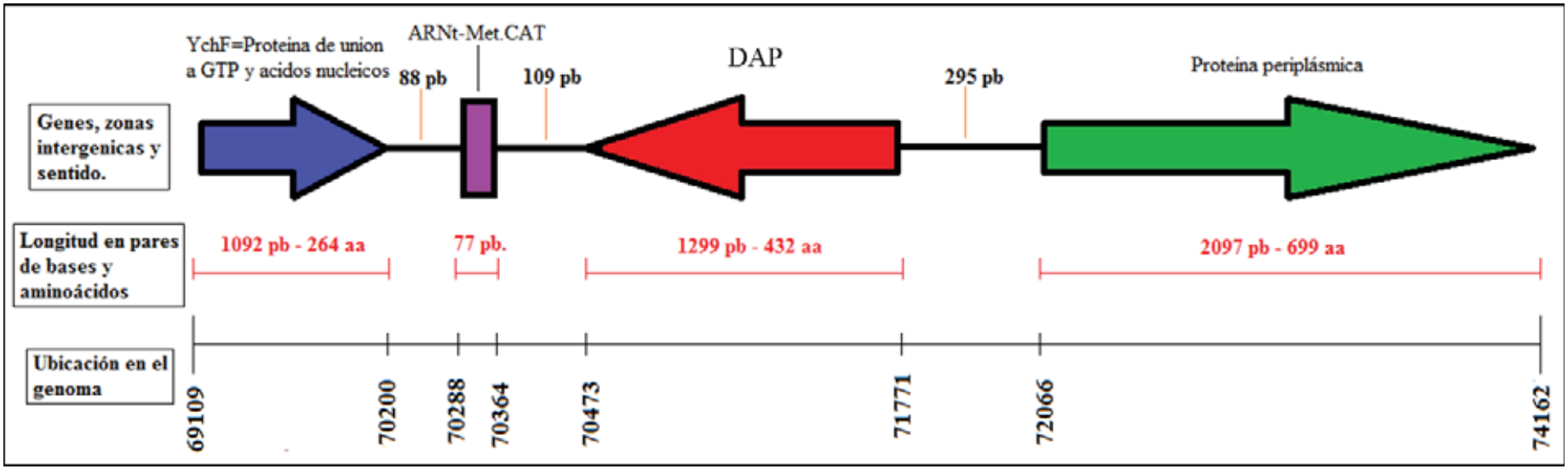


Figura N° 58.Ubicación del gen aspartil aminopeptidasa en el genoma de Chromohalobacter salexigens con el ensamblador Spades.

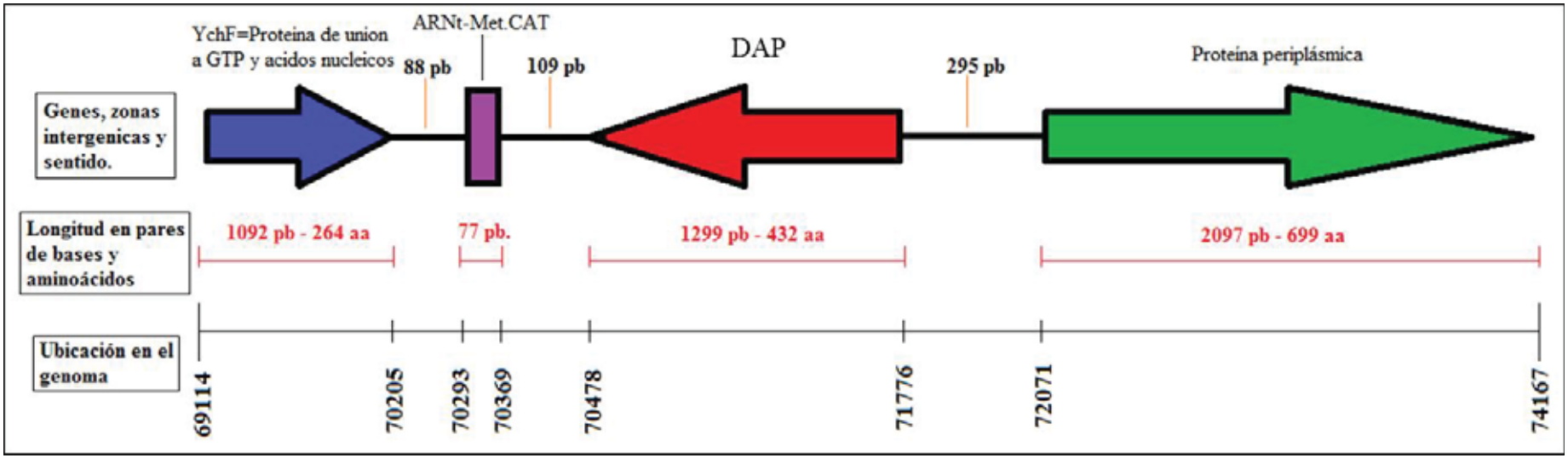


Figura N° 59. Ubicación del gen aspartil aminopeptidasa en el genoma de Chromohalobacter salexigens con el ensamblador A5.

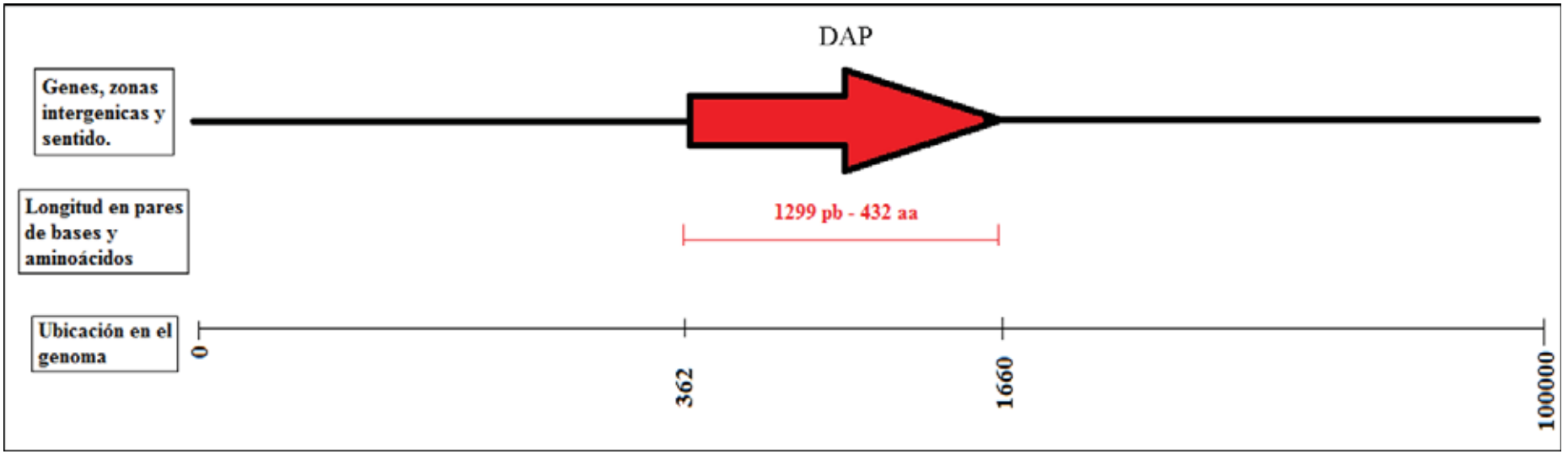


Figura N° 60. Ubicación del gen aspartil aminopeptidasa en el genoma de Chromohalobacter salexigens con el ensamblador Velvet.

Las secuencias de genes de ARNt con un anticodón CAT están presentes en casi todos los seres vivos pero estos varían en el número de copias en el gen. Anteriormente en un estudio bioinformático analizaron 234 genomas bacterianos para clasificar en los tres tipos de CAT conocidos como es el Ile, iniciador y alargador Met. En este trabajo clasificaron los tres tipos conocidos y desarrollaron métodos para diferenciación entre ellos, especialmente para identificar los genes de ARNt Ile CAT encontrando hasta un número máximo de 15 genes ARNt con anticodon CAT en *Photobacterium profundum,* 5 corespondiendo a ARNt Met CAT, es así que crearon un método en la que se puedan diferenciar el iniciador ARNt Met CAT contra el elongador tRNA Met CAT. En los análisis bioinformático se usó información genómica de gammaproteobacterias (Silva *et al*., 2006).

Los datos de descripción del gen aspartil aminopeptidasa con los programas *Spades* y *A5* ofrecen resultados similares y además con una precisión similar concluyendo que ambas herramientas son útiles para la descripción del gen aspartil aminopeptidasa.

3.1.3.2. Modelado estructural de la aspartil aminopeptidasa y su comparación con otros genes similares.

El modelado estructural de la proteína se realizó a partir de la secuencia de aminoácidos obtenido de la anotación. La secuencia aminoacídica de DAP MP25462 fue subida al servidor I-TASSER resultando un archivo PDB con el cual se inició el análisis de la estructura 3D de DAP. El modelado de DAP MP25462 otorgó la similaridad estructural con la proteína con el código de PDB 316SA que corresponde a la hidrolasa aspartil aminopeptidasa humana (DNPEP) expresada en *Escherichia coli* BL21 (DE3) con un índice de Cov de 1.0 y Z-score de 2.99. La simulación entre los diez mejor homologías estructurales se expresaron en modelo final para DAP MP25462 con un C-score de 1.73 que es un puntaje de confianza de las alineaciones y ensamblaje de las estructuras. Su puntaje está en el rango de -5 a 2 y cuando el valor de C es mayor, más alto es el valor de confianza del modelo, TM-score igual a 0.96 ± 0.05 mide la similitud estructural con una puntuación TM > 0.5 indica una topología correcta mientras que un valor TM < 0.17 significa que la proteína se emparejo aleatoriamente y un RMSD estimado de 3.4 ± 2.4 Å que es la distancia promedio de los átomos entre ambas proteínas, las cuales entre más pequeñas indicarán un modelado preciso y seguro. El modelado de DAP MP25462 obtuvo puntajes muy altos en cuanto la confianza, similitud y precisión en el alineamiento estructural, demostrando que la proteína existe.

Las 10 proteínas estructuralmente cercanas al modelado de DAP MP25462 se muestran en la Tabla N° 12. Entre ellos se encuentran proteínas cristalizadas de *Homo sapiens* (Chaikuad

et al., 2012), *Plasmodium falciparum* (Sivaraman et al., 2012), *Chaetomium thermophilum* (Bertipaglia et al., 2016), *Saccharomyces cerevisiae* (Su *et al*., 2015), *Clostridium acetobutylicum* (Min & Shapiro, s. f.-b), *Thermotoga maritima* (Min & Shapiro, s. f.-a), *Pseudomonas aeruginosa* (Min, Burley, et al., s. f.), *Borrelia burgdorferi* (Min, Gorman, et al., s. f.), *Desulfurococcus amylolyticcus* (Petrova et al., 2015) y *Pyrococcus horikoshii* (Russo & Baumann, 2004)

Los resultados del alineamiento o superposición estructural se visualizaron en el programa Chimera como se describe en el capítulo II (Figura N°61).

Las características de la estructura proteica de DAP MP25462 fueron comparables a lo descrito anteriormente por Chaikuad *et al*., 2012. El dominio proteolítico globular en *Chromohalobacter salexigens* MP25462 se ubica del 1-85 y del 219-432 aa (celeste y magenta) que contiene once cadenas de hoja beta intercaladas entre nueve hélices α. Además se diferenció la existencia de un subdominio α (celeste) y un pequeño subdominio β (magenta) de cuatro cadenas que descansa en la parte superior del dominio proteolítico que es muy similar entre las estructuras M18 descritas en publicaciones anteriores. El extremo amino y carboxilo se localizan en este dominio.

Así también se pudo identificar el dominio de dimerización ubicado del 86 al 218 aa. su forma que se compara a alas de mariposa presenta tan solamente una hélice α3 que conforma un bucle más pequeño que lo descrito en el humano donde existen hélices α3 y α4 formando un bucle mucho mayor. El sitio activo fue localizado entre el interfaz de ambos dominios en el 263 aa (Figura N° 62).

Tabla N° 12: Proteínas estructuralmente cercanas al modelado de DAP MP25462

PDB encontrado	TM-score	Organismo	RMSD	IDEN	Cov	Alineación
1	3l6sA	*Homo sapiens*	0.986	0.85	0.407	0.995
2	4emeA	*Plasmodium falciparum*	0.98	0.92	0.298	0.993
3	5jm6A	*Chaetomium thermophilum*	0.925	1.57	0.329	0.958
4	4r8fA	*Saccharomyces cerevisiae*	0.896	1.84	0.289	0.935
5	2gljX	*Clostridium acetobutylicum*	0.882	2.45	0.221	0.954
6	2glfA	*Thermotoga maritima*	0.877	2.53	0.236	0.951
7	2ijzA	*Pseudomonas aeruginosa*	0.83	1.97	0.308	0.873
8	1y7eA	*Borrelia burgdorferi*	0.799	2.32	0.194	0.859
9	4wwvA	*Desulfurococcus amylolyticcus*	0.717	2.62	0.17	0.792
10	1xfoA	*Pyrococcus horikoshii*	0.697	2.64	0.169	0.768

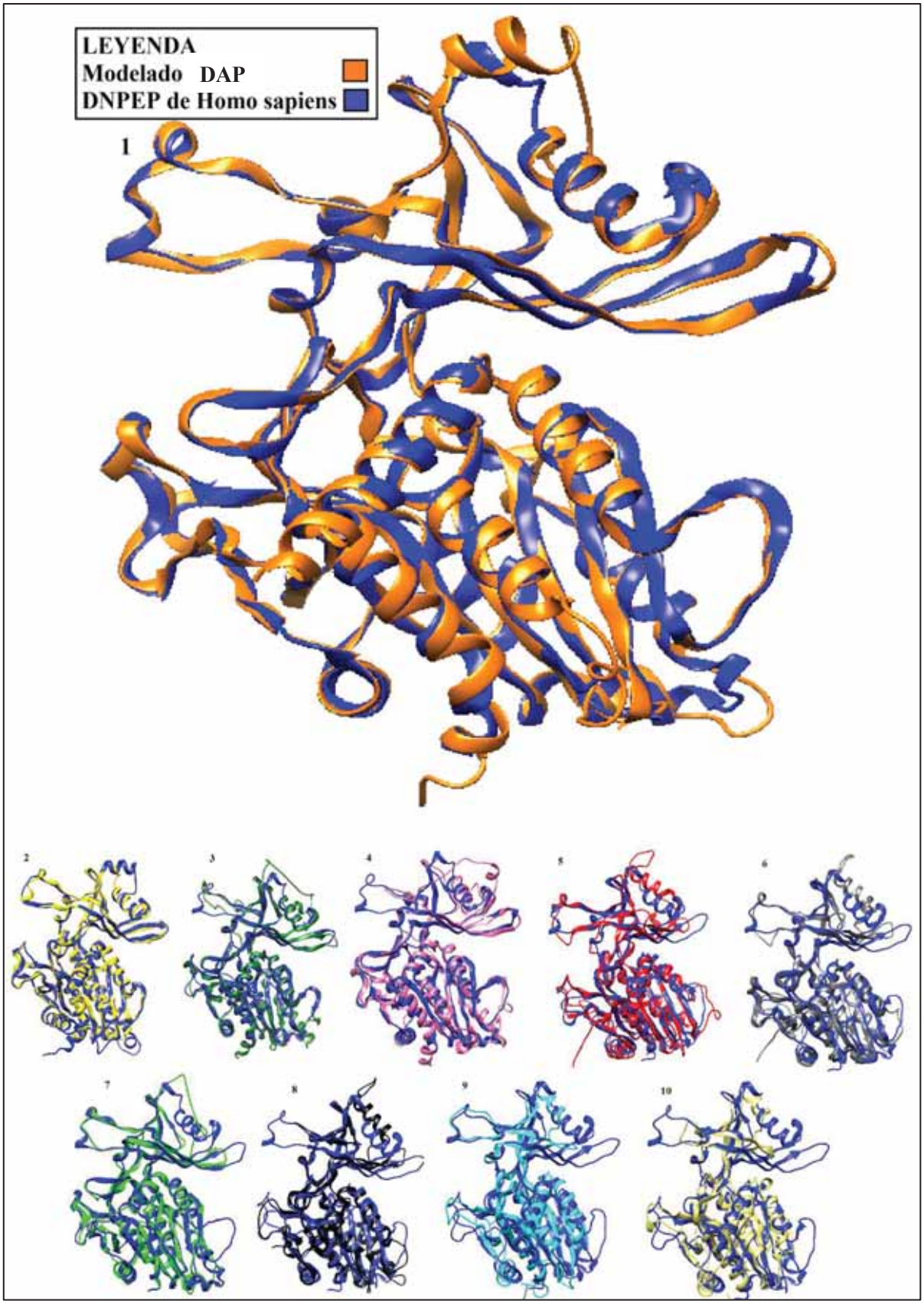


Figura N° 61. Visualización de alineamientos estructurales entre la aspartil aminopeptidasa de Chromohalobacter salexigens MP25462 y Homo sapiens (1), Plasmodium falciparum (2), Chaetomium thermophilum (3), Saccharomyces cerevisiae (4), Clostridium acetobutylicum (5), Thermotoga marítima (6), Pseudomonas aeruginosa (7), Borrelia burgdorferi (8), Desulfurococcus amylolyticcus (9) y Pyrococcus horikoshii (10).

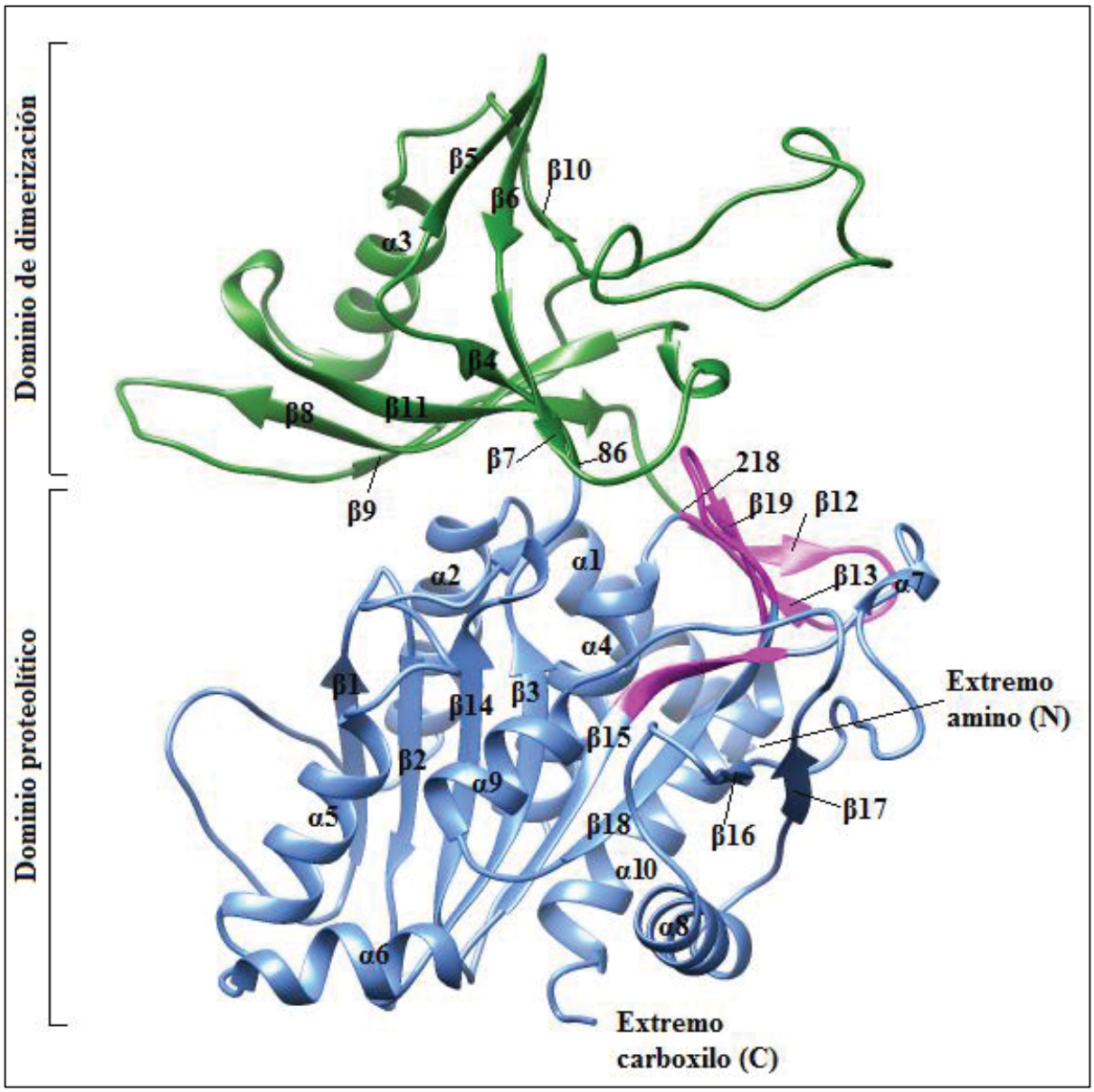


Figura N° 62. Características estructurales de la proteína aspartil aminopeptidasa

I-TASSER mediante su base de datos otorgo la predicción de funcionalidad de la proteína DAP *Chromohalobacter salexigens* MP25462 comparándola estructuralmente con las proteínas cristalizadas contenidas en PDB. La peptidasa TET de *Thaumarchaeota archaeon* (Michalska et al., s. f.) resulto ser muy similar en cuanto a la conformación de los sitios de unión al ligando con un C-score de 0.74 de un tamaño tamaño de Cluster 71 (Tabla N° 13) ubicando a los residuos del sitio de unión al ligando en 81H (histidina), 234D (ácido aspártico o aspartato), 235N (asparagina), 263G (glutamina) como sitio activo (Tabla N°14), 264G (glutamina), 310D (aspartato), 404H (histidina) deduciendo el ligando como dos iones de cobalto (Figura N° 63), definiendo la topología del sitio activo de la proteína DAP *C. salexigens* MP25462 similar a la de *Homo sapiens* y la composición de ligandos del sitio activo idénticos a *Thaumarchaeota archaeon*. También se pudo visualizar la superficie proteica de estos residuos catalíticos (Figura N° 64).

Tabla N° 13: Sitios de unión al ligando de la aspartil aminopeptidasa de MP25462

Rango	C-score	Tamaño de cluster	PDB Hit	Organismo	Nombre de ligando	Residuos del sitio de unión al ligando
1	0.74	71	5ds0A	*Thaumarchaeota archaeon*	Co	81; 234; 235; 263; 264; 310; 404.
2	0.64	63	4njqB	*Pseudomonas aeruginosa*	Co	234; 264; 403; 404.
3	0.35	13	4dyoA	*Homo sapiens*	Zn	81; 234; 263; 264; 310; 311; 313; 338; 345; 378; 379; 403; 404.

Tabla N° 14: Sitio activo producido por I-TASSER.

Rango	C-score^{EC}	PDB	TM-score	RMSD	IDEN	Cov	Numero EC	Residuo del sitio activo
1	0.555	3dljA	0.511	3.93	0.133	0.616	3.4.13.20	263
2	0.538	2zofA	0.512	3.82	0.109	0.609	3.4.13.18	263
3	0.537	1ampA	0.5	3.24	0.146	0.569	3.4.11.10	263
4	0.513	1lfwA	0.482	4.03	0.147	0.574	3.4.13.3	263
5	0.512	3gb0A	0.493	3.01	0.134	0.558	3.4.11	263

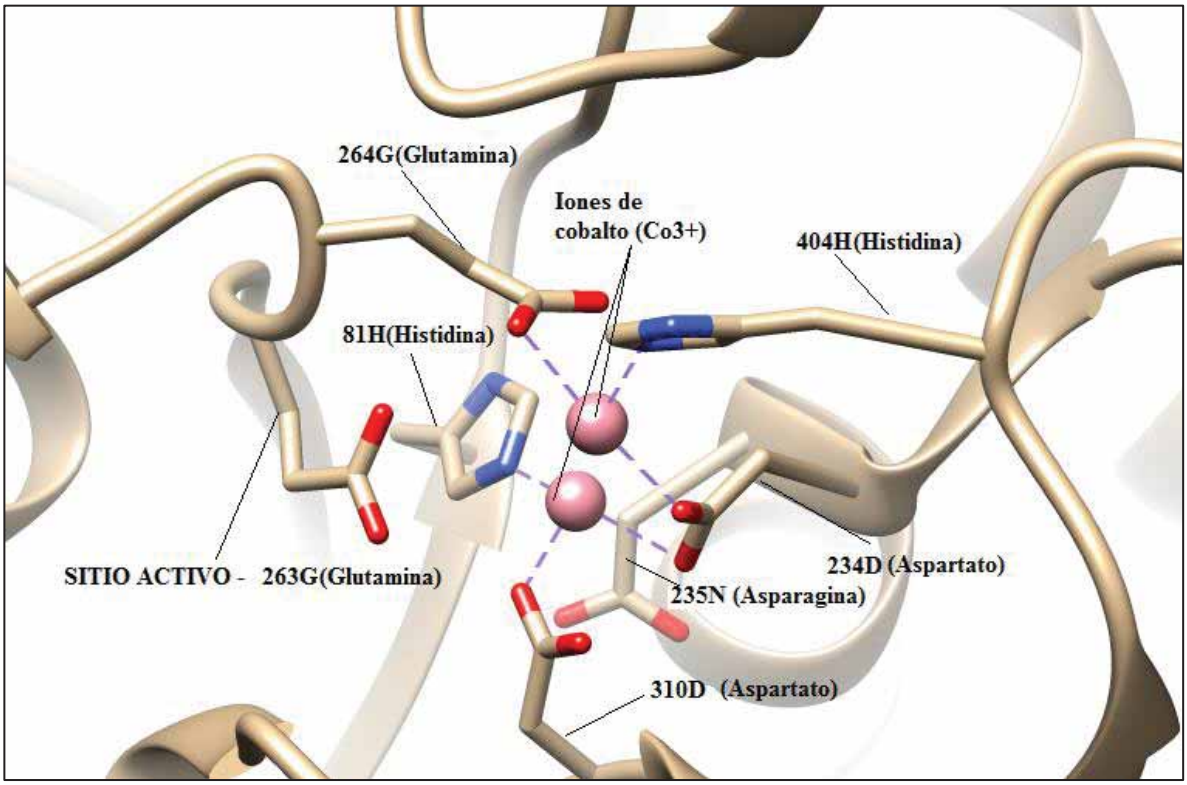


Figura N° 64: Visualización de los sitios de unión a los ligandos de cobalto y sitio activo de las aspartil aminopeptidasa.

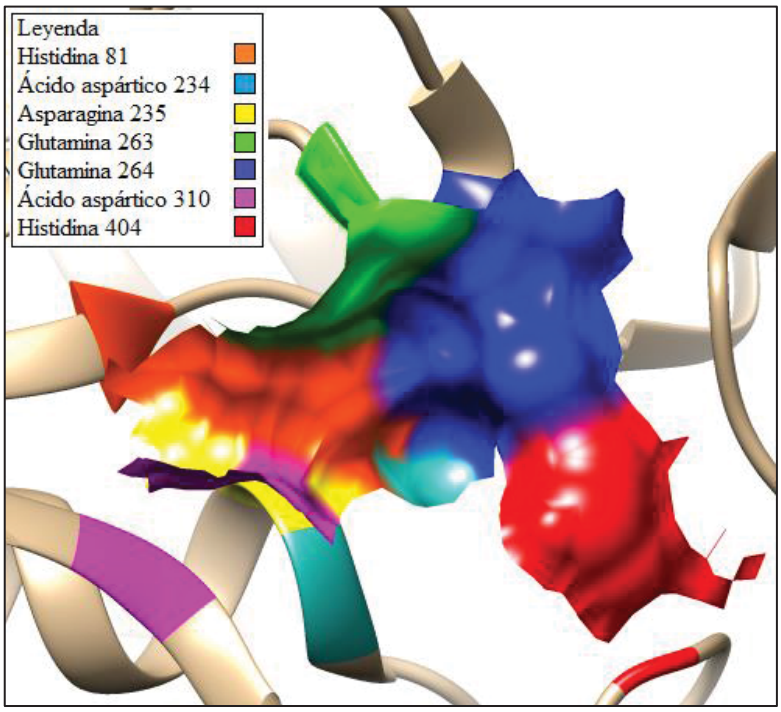


Figura N° 63: Superficie proteica de los residuos catalíticos de la aspartil aminopeptidasa

Del alineamiento de la secuencia de aminoácidos con otras especies que presentan la proteasa DAP MP25462 se observó que los aminoácidos pertenecientes a los residuos catalíticos se ubican en lugares conservados deduciendo el ligando como dos iones de cobalto (Figura N° 63). También se pudo visualizar la superficie proteica de estos residuos catalíticos (Figura N° 64).

El alineamiento de la secuencia de aminoácidos con otras especies que presentan la proteasa DAP MP25462 se observó que los aminoácidos pertenecientes a los residuos catalíticos se ubican en lugares conservados (Figura N° 65).

En las figuras N° 63 y 64 se observa la distribución de los sitios de unión al ligando cobalto y al sitio activo **Glutamina 263**. Los sitios resaltados con recuadro de color rojo en el alineamiento de la figura N° 65 son los 7 sitios implicados en la funcionalidad de la proteína DAP *C. salexigens* que se aprecia también en todas las proteínas aspartil aminopeptidasas alineadas en los diferentes organismos comparados como: *Pseudomonas aeruginosa, Aspergillus oryzae, Saccharomyces cerevisiae, Homo sapiens, Mus musculus, Zea mays, Arabidopsis thaliana* y *Micromonas pusilla* dando así a entender que estos son sitios conservados en el tiempo y que la proteína presenta actividad proteolítica que caracteriza a la familia de las metaloproteasas dependientes de cobalto (M18).

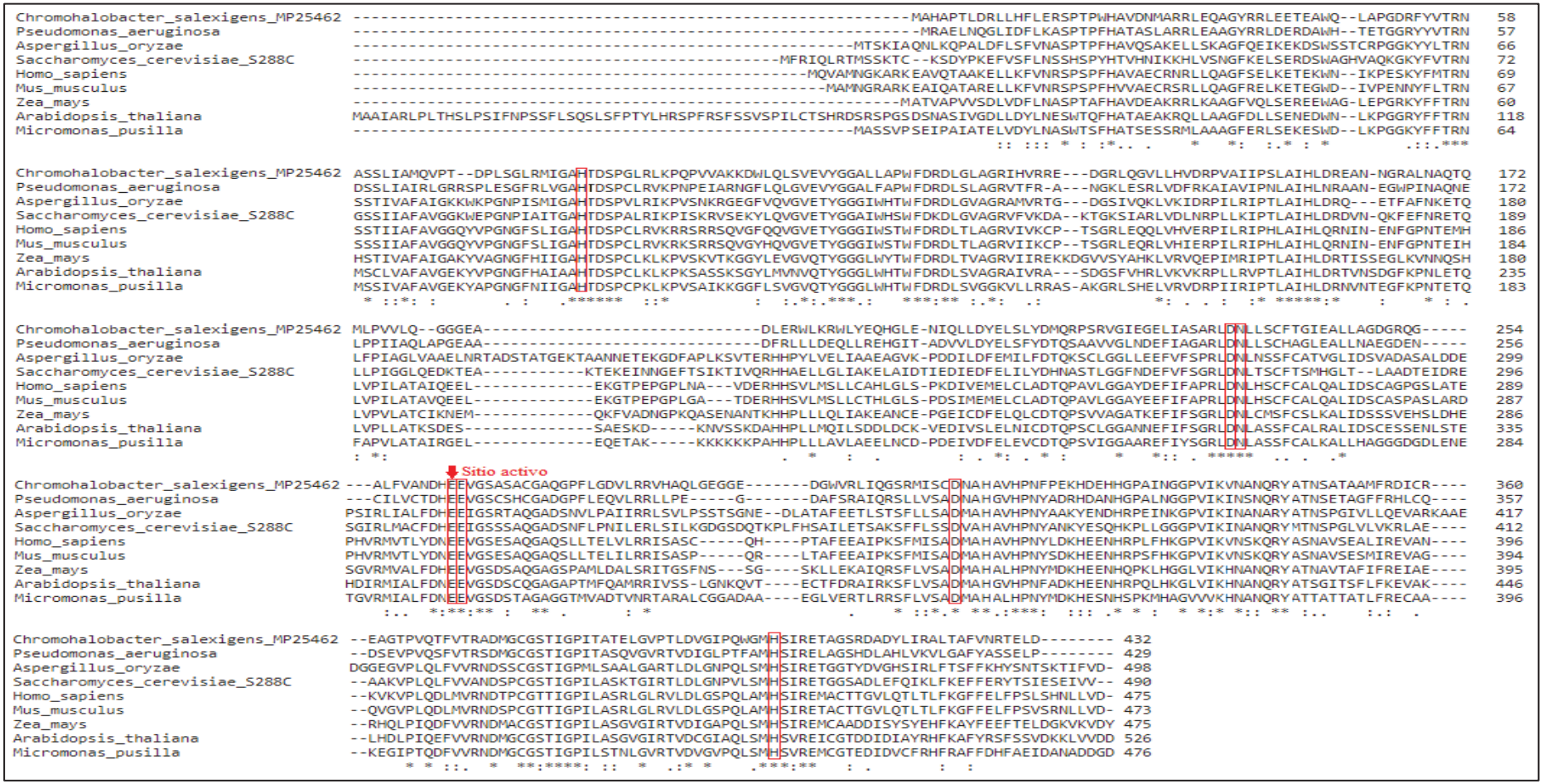


Figura N° 65: Alineamiento de la DAP Chromohalobacter salexigens MP25462 con la de otras organismo; resaltado con recuadros de color rojo los aminoácidos de ligando y el sitio activo señalado con una flecha. (-) Espacios vacíos para lograr el alineamiento, (*) alineamiento de todos los aminoácidos, (.) existencia de al menos 2 aminoácidos diferentes en la alineación y (:) presencia de un aminoácido diferente en la alineación.

Como se observó en el apartado 3.1.1.3.2 el modelado de DAP MP25462 y su visualización contra la proteína humana tienen gran similitud estructural. En el alineamiento podemos observar que la DAP MP25462 tiene una longitud de 432 aminoácidos y la human 475 aminoácidos con una diferencia de 43. Dicha afirmación se sustenta al analizar el modelado de superficie proteica de DAP MP25462 (Figura N° 64) en donde se puede observar la existencia de dos espacios donde se ubican los iones de cobalto los cuales están formados por los 7 sitios implicados en la actividad enzimática. Siendo glutamina 263 (sitio activo) glutamina 264 histidina 404, acido aspártico 234 y la histidina 81 los que forman los soportes principales para unirse al metal en la parte superior; en cambio el lugar ubicación del segundo ion cobalto en la parte inferior no está bien definida la superficie proteica. Estudios realizados en la obtención de la aspartil aminopeptidasa de *P. aeruginosa* determinaron que era una proteasa cocatalítica con una actividad en presencia de iones de zinc y que su actividad aumentaba en presencia de cobalto (Nguyen *et al*., 2014). Esta podría ser una posibilidad para la proteasa aspartil aminopeptidasa de *C salexigens* porque se constató que tiene una gran homología con la proteína aspartil aminopeptidasa humana como se observó en la tabla N° 12 la cual presenta actividad enzimática en presencia de zinc (Tabla N° 13) (Chaikuad *et al*., 2012).

3.1.4. Síntesis de cebadores.

La pareja de cebadores fueron sintetizados para continuación de los procesos de la segunda etapa de clonación del gen aspartil aminopeptidasa La secuencia del cebador sentido diseñado fue 5’ CCCTTGACCGTTTACTGC 3’ este comienza en el nucleótido 17 y termina en el 34 con una Tm de 56° C y el cebador antisentido 5’ CTAGTCCAGCTCGGTGCG 3’ comienza en el nucleótido 1299 hasta el 1282 con Tm de 60° C a los cuales se les nombró DAP.F2 y DAP.R2 respectivamente. Esta pareja de cebadores amplifica *in silico* un fragmento de 1283 pb.

3.2. Etapa 2: Clonación del gen aspartil aminopeptidasa

3.2.1. Recuperación de la bacteria criopreservada *C. salexigens* MP25462

Las colonias de *Chromohalobacter salexigens* MP25462 crecieron en el medio SW 15% de NaCl en placa a los 4 días de incubar a 37° C (Figura N° 66) mientras que en medio liquido SW al 15% de NaCl se tornó turbio a los 2 días. El cultivo en medio liquido se utilizó para la extracción del ADN genómico.

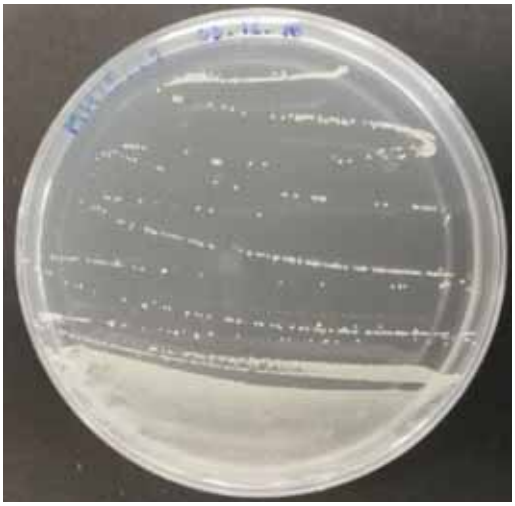


Figura N° 66. Crecimiento de colonias de Chromohalobacter salexigens MP25462

3.2.2. Extracción del ADN genómico de *C. salexigens* MP25462.

Se realizó la extracción del ADN genómico de acuerdo a las indicaciones del fabricante como se detalló en el capítulo II apartado 2.3.3.3. La cuantificación en gel de agarosa resultó **9.45 ng/µl** observando una banda mayor a 20000 pb (figura N° 67).

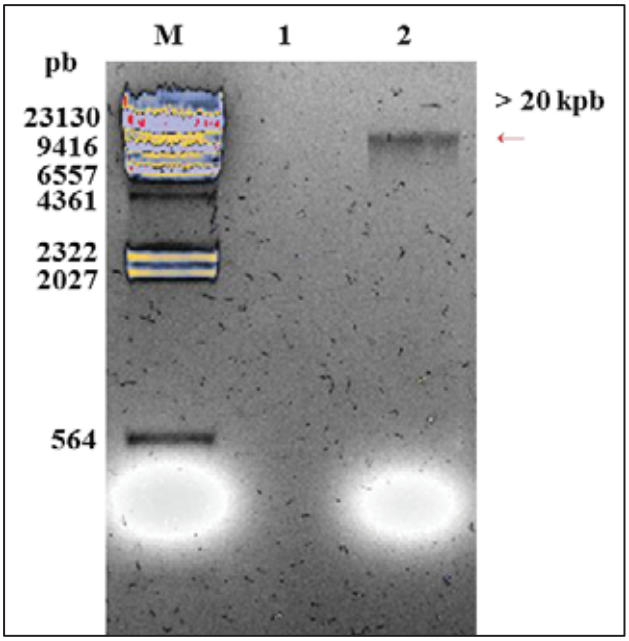


Figura N° 67: Visualización y cuantificación del ADN genómico de Chromohalobacter salexigens en gel de agarosa al 0.8%. M: Marcador de peso molecular Lambda HindIII. 1: Pozo sin muestra. 2: Banda del ADN genómico de Chromohalobacter salexigens MP25462.

En cuanto a la cuantificación al nanoespectrofómetro se utilizó 2 µl de ADN genómico. La absorbancia a 260 nm resulto **0.342**, tal índice es usado por el equipo para calcular la concentración de los ácidos nucleicos genómicos de MP25462 el cual fue de **17.08 ng/µl**. Para la proporción de la absorbancia a 260 nm y 280 nm resultó **2.21**. La segunda valoración de la pureza de ácidos nucleicos de proporción 260/230 resulto en **0.66**. Estos parámetros de la cuantificación del ADN genómico en el espectrofotómetro correspondientes a *Chromohalobacter salexigens* MP25462 se aprecian en la tabla N° 15.

Tabla N° 15. Valores de la cuantificación del ADN genómico de Chromohalobacter salexigens.

	ng/µl	A260	260/230	260/280
MP25462	17.08	0.342	0.66	2.21

3.2.3. Estandarización de PCR para la amplificación del gen aspartil aminopeptidasa.

Los cebadores DAP.F2 y DAP.R2 fueron utilizados para la estandarización de la PCR para la amplificación del gen aspartil aminopeptidasa. Para lo cual se tuvo en cuenta la diferencia entre sus temperaturas de hibridación. Al presentar DAP.R2 la Tm de 60° C y DAP.F2 Tm de 56° C se optó por realizar un gradiente térmico de hibridación con las temperaturas 48, 51 y 55° C porque se procura que la Tm en la hibridación no sea mayor a 60°

C al no tener una unión especifica de los cebadores al templado. Al visualizar las reacciones de PCR se obtuvo amplificación a las tres temperaturas empleadas de (Figura N° 68).

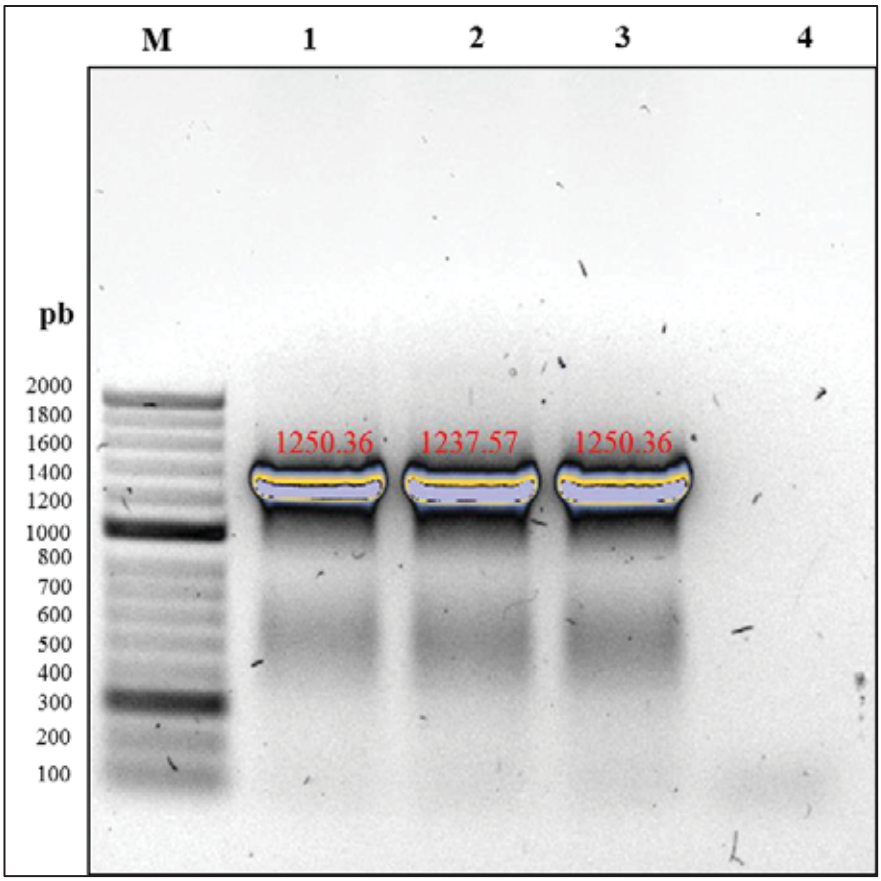


Figura N° 68: Visualización de PCR estandarizada. M: Marcador molecular Hiper Lader II. 1: Reacción de PCR con Tm de 48° C. 2: reacción de PCR con Tm de 51° C. 3: reacción de PCR a Tm de 55° C. 4: blanco.

Se logró amplificar una sola banda de aproximadamente 1237 pb. y el tamaño esperado es de 1283 pb. con lo cual se tuvo una alta probabilidad de que sea el gen de la aspartil aminopeptidasa (Tabla N° 16). Los cebadores diseñados pueden amplificar a temperaturas de fusión desde 48 a 55° C. Teniendo en cuenta este resultado se procederá con las subsiguientes PCRs.

Tabla N° 16: Tamaños de las bandas amplificadas de la PCR estandarizada.

Pozo	Banda N°	Pares de base (pb)
M (marcador Hiper Lader II)	1	2,000.00
	2	1,800.00
	3	1,600.00
	4	1,400.00
	5	1,200.00
	6	1,000.00
	7	800
	8	700
	9	600
	10	500
	11	400
	12	300
	13	200
	14	100
1	1	1,250.36
2	1	1,237.57
3	1	1,250.36

3.2.4. Amplificación por PCR convencional del gen aspartil aminopeptidasa.

A continuación se muestra en la tabla N° 17 el sistema de reacción de la PCR construido.

Tabla N° 17: Sistema de reacción de PCR con los cebadores DAP.F2 y DAP.R2.

REACTIVO	CONCENTRACIÓN INICIAL	CONCENTRACIÓN FINAL	VOLUMEN POR 1 REACCIÓN	VOLUMEN POR 5 REACCIÓN
H_2O	-	-	8.75 µl	43.75 µl
TAMPÓN	10X	1X	4 µl	20 µl
$MgCl_2$	50mM	1.5 mM	1.2 ul	6 µl
dNTPs	2mM	0.2 mM	2 µl	10 µl
PRIMER DAP.F2	10pmol/ µl	10pmol	1 µl	5 µl
PRIMER DAP.R2	10pmol/ µl	10pmol	1 µl	5 µl
TAQ POLIMERASA	5U/ µl	0.25U	0.05 µl	0.25 µl
ADN	5ng/ µl	10ng	2 µl	-
	Volumen Total		20 µl	90 µl

Para visualizar el producto de amplificación, se cargó 20 ul de las reacciones de PCR en un gel de agarosa al 1 %, se llevo a electroforesis durante 1 hora a 90 V constantes (Figura N° 69).

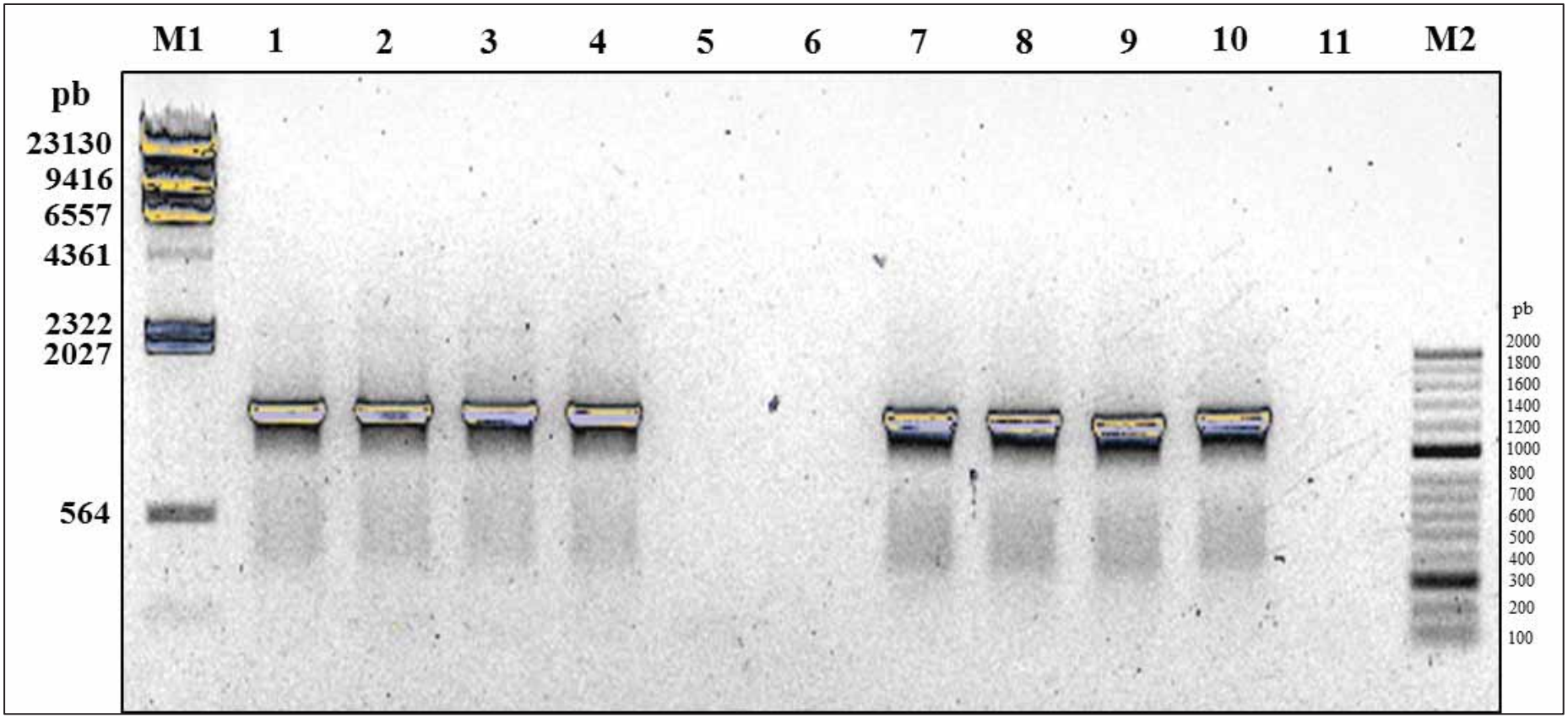


Figura N° 69: PCR del gen aspartil aminopeptidasa para purificación a partir de geles. M1: marcador molecular Lambda Hind III. 1, 2, 3, 4: amplificaciones en termociclador Biometra 5: Blanco 6: Pozo sin muestra, 7, 8, 9, 10: amplificaciones en termociclador BioRad, 11: Blanco M2: Marcador Hyper Lader II

La visualización de las bandas y su análisis se realizó con Image Lab (Tabla N° 18). Para obtener la cantidad suficiente de las secuencia del gen aspartil aminopeptidasa MP25462 para los procesos posteriores, se purificó las 8 bandas de amplificadas mediante el protocolo del *Kit Ilustra^{TM} GFX PCR DNA and Gel Band Purification* como se explicó en el capítulo de materiales y métodos.

Se realizó la lecturas al espectrofotómetro del ADN purificado resultando **48.32 ng/µl.** La cuantificación en gel resulto de **22.3 ng/µl.** Siendo este último valor tomado en cuenta más adelante en los cálculos para la clonación.

Tabla N° 18: Cuantificación en gel de las bandas de amplificadas.

Pozo	Banda N°	Pares de base (pb)	Cuantificación (ng/µl)
M1	1	23,130.00	97
	2	9,416.00	67.08
	3	6,557.00	51.8
	4	4,361.00	N/A
	5	2,322.00	20.8
	6	2,027.00	21
	7	564	6.3
1	1	1,212.23	114.5
2	1	1,212.23	99
3	1	1,197.82	112
4	1	1,183.59	112.1
7	1	1,128.32	109.3
8	1	1,114.91	106.4
9	1	1,088.57	107.4
10	1	1,141.89	107

3.2.5. Purificación y secuenciación del gen aspartil aminopeptidasa.

Para comprobar la identidad del fragmento amplificado corresponda con la secuencia del gen de la proteasa DAP MP25462, se mandó a secuenciar dicha fragmento al Servicio General de Apoyo a la Investigación (SEGAI) de la Universidad de La Laguna anexo 5.

Los resultados de secuenciación fueron analizados mediante BLAST para conocer la identidad del fragmento amplificado. Dicho análisis otorgo 100% de identidad contra el gen

que codifica la proteasa aspartil aminopeptidasa de *Chromohalobacter salexigens* DSM 3043. Con el programa MEGA X, mediante un cromatograma se constató la asignación de bases en el proceso de secuenciación Sanger del gen aspartil aminopeptidasa (Figura N° 70). Tomando en cuenta estos datos se procedieron con seguridad los siguientes experimentos.

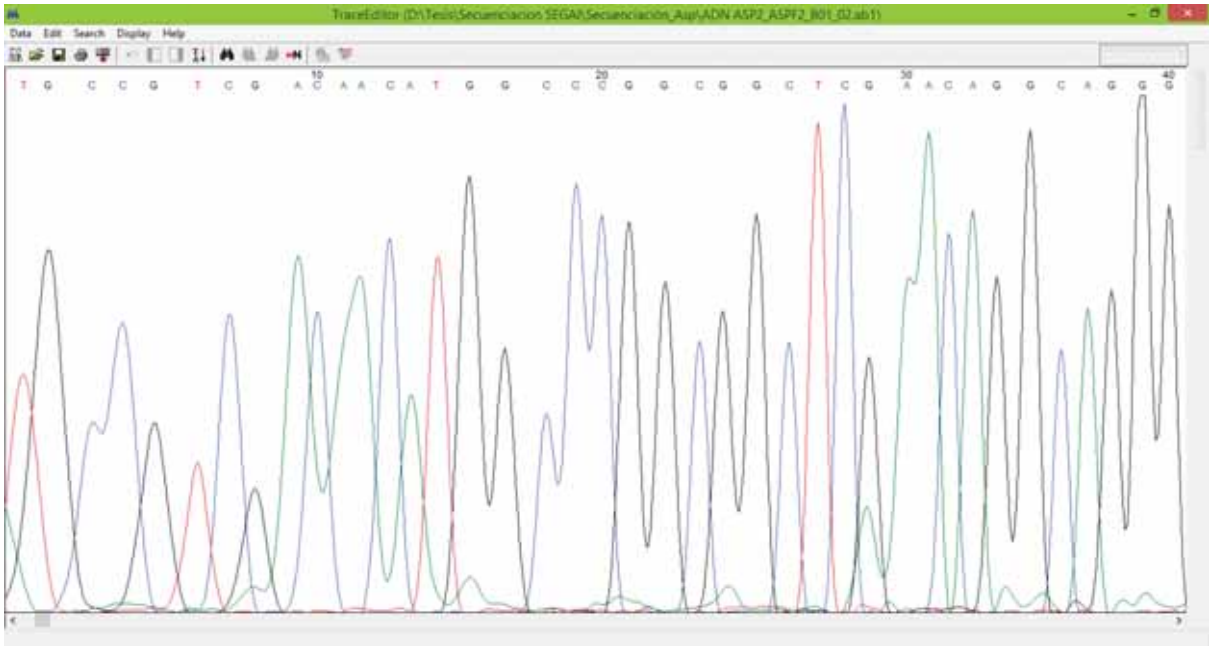


Figura N° 70: Cromatograma de la secuenciación del gen aspartil aminopeptidasa amplificado

3.2.6. Clonación en el vector plásmido pGEM-T.

Para la clonación se tuvo en cuenta el valor de cuantificación en gel de la muestra del gen aspartil aminopeptidasa el cual fue de **22.3 ng/µl**. A partir de este dato se hizo los cálculos con lo indica el protocolo del sistema de vectores pGEM-T. Todas las indicaciones para el uso del sistema de vectores pGEM-T se detallaron en el apartado 2.3.5.9.

Como resultado de la transformación del inserto-vector en las células competentes JM109 se obtuvo el crecimiento de 6 colonias a 22 horas después de su cultivo. Tales colonias presentan el vector pGEM-T ya que el vector presenta el gen para a resistencia a la ampicilina el cual permite a las células JM109 crecer en el medio LB/ampicilina. A partir de las 6 colonias obtenidas de la transformación se realizó PCR, esto para garantizar el ingreso del inserto aspartil aminopeptidasa en el vector ya que la recircularización del vector pGEM-T para ciertas colonias es probable. Este método de transformación se explicó en apartado 2.3.5.10. Se visualizaron las bandas amplificadas de tal proceso (Figura N° 71).

Se observó que 4 de las 6 colonias presentan el vector pGEM-T con el inserto del gen aspartil aminopeptidasa. Estas son las colonias DAP.T-2, DAP.T-3, DAP.T-6 y DAP.T-7.

En la PCR con los cebadores del vector para las 4 colonias positivas ya mencionadas, se observó la amplificación de dos bandas. La mayor de una longitud de aproximadamente 1484 pb y la menor de un tamaño de 866 pb. Esto puede deberse a que gen aspartil aminopeptidasa poseía una secuencia apropiada para la hibridación de uno de los cebadores propios del plásmido ocurriendo la aparición de esta banda de 866 pb.

En los pozos número 3 y 6 se observa que la transformación se llevó a cabo pero con un vector plasmídico pGEM-T recirculado. La banda de 177 y 165 pb respectivamente, es la demostración de tal caso.

El caso de la PCR con los cebadores de la aspartil aminopeptidasa constata que el plásmido contiene a la secuencia que codifica a la aspartil aminopeptidasa en las colonias DAP.T-2, DAP.T-3, DAP.T-6 y DAP.T-7. En el poso diez y catorce amplifico una banda de 566 y 594 lo que da a entender que los cebadores diseñados DAP.F2 y DAP.R2 tienen secuencias diana para su hibridación en el genoma de *E. coli* JM109.

El tamaño de las bandas obtenido por el programa *Image Lab* se puede apreciar en la tabla N° 19.

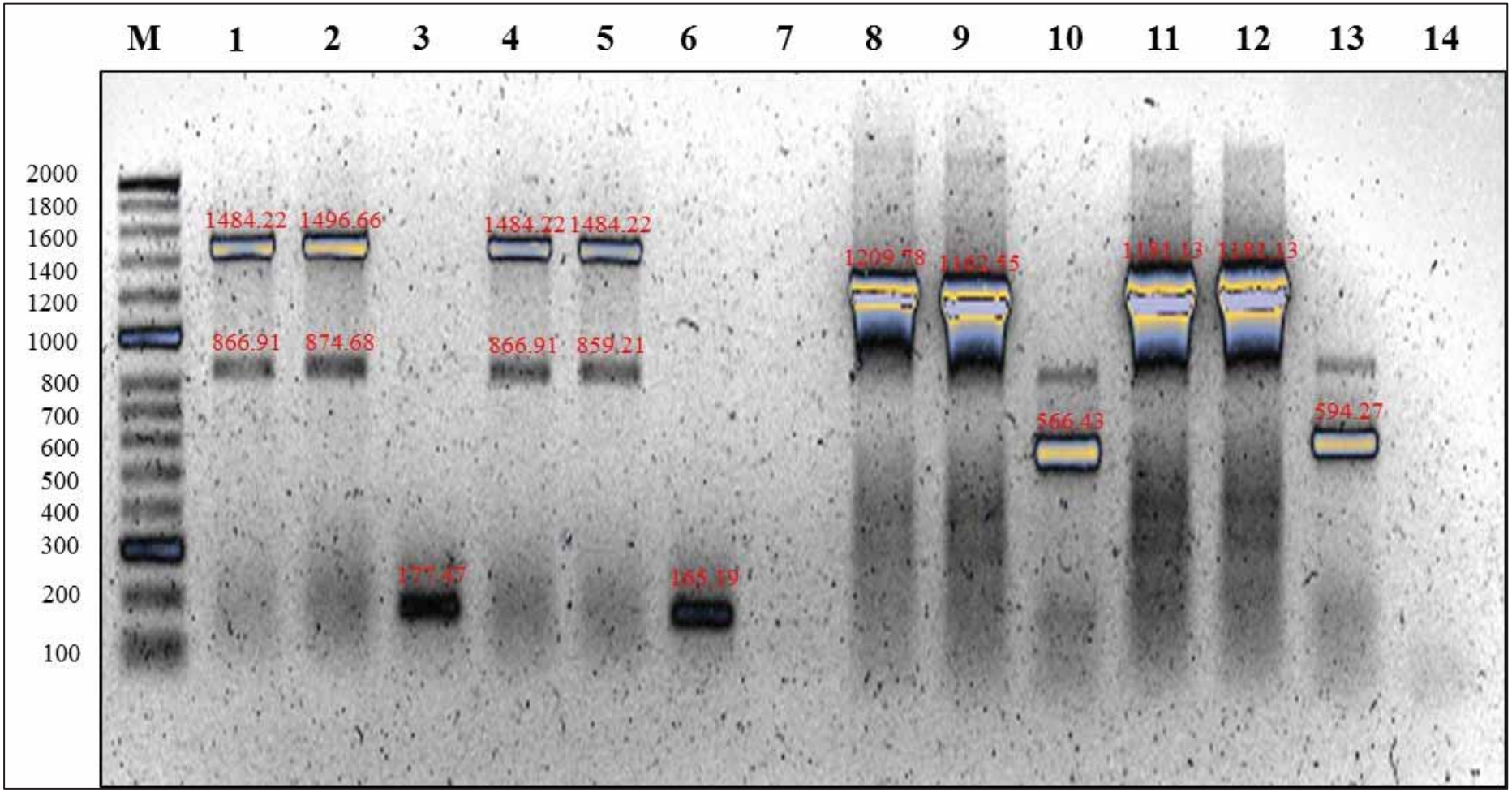


Figura N° 71: Visualización del gen aspartil aminopeptidasa clonado en el vector pGEM-T. M: Marcador Hyper Lader II. ***1, 2, 3, 4, 5, 6, 7****: Amplificación con T7/SP6,* ***7*** *(blanco).* ***8, 9, 10, 11, 12, 13, 14.*** *Amplificación con DAP.F2/DAP.R2,* ***15*** *(blanco).*

Tabla N° 19: Tamaño de las bandas amplificadas por PCR.

Pozo	Banda N°	Pares de base (pb)
M: Marcador Hyper Lader II	1	2,000.00
	2	1,800.00
	3	1,600.00
	4	1,400.00
	5	1,200.00
	6	1,000.00
	7	800
	8	700
	9	600
	10	500
	11	400
	12	300
	13	200
	14	100
1	1	1,484.22
	2	866.91
2	1	1,496.66
	2	874.68
3	1	177.47
4	1	1,484.22
	2	866.91
5	1	1,484.22
	2	859.21
7	1	165.19
8	1	1,209.78
9	1	1,162.55
10	1	566.43
11	1	1,181.13
12	1	1,181.13
13	1	594.27

3.2.7. Secuenciación del gen aspartil aminopeptidasa clonado

Las colonias DAP-T2 y DAP-T7 fueron seleccionadas para la extracción del plásmido clonado (apartado 2.3.5.11). Los datos de secuenciación fueron analizados como se explica en el apartado 2.3.5.12.

Se pegaron los resultados de secuenciación del plásmido de la colonia DAP-T2 en la ventana de alineación de MEGA X, ubicando al resultado de secuenciación por el cebador T7 en la primera fila de la ventana de alineación y al resultado por SP6 en la quinta fila. Luego se llevaron a las cadenas nucleotídicas sentido y antisentido del gen aspartil aminopeptidasa a la ventana de alineación para tratar de alinearlas con uno de los resultados de secuenciación del plásmido clonado. Por alineamiento de nucleótidos de manera manual se pudo identificar que la secuencia antisentido del gen aspartil aminopeptidasa es igual al resultado de secuenciación con T7 y este cebador amplifica en la cadena sentido del plásmido pGEM-T, mientras que la cadena sentido del gen Asp se alinea con el resultado de secuenciación con SP6 que amplifica en la cadena antisentido del vector. Se tuvo el mismo análisis con la muestra de DAP-T7 llegando a la misma conclusión. El resultado de secuenciación del plásmido clonado en ambas muestras DAP-T2 y DAP-T7, el inserto ingreso de forma invertida en el sitio de multiclonaje. Asi se puede apreciar en las figura N° 72 y 73.

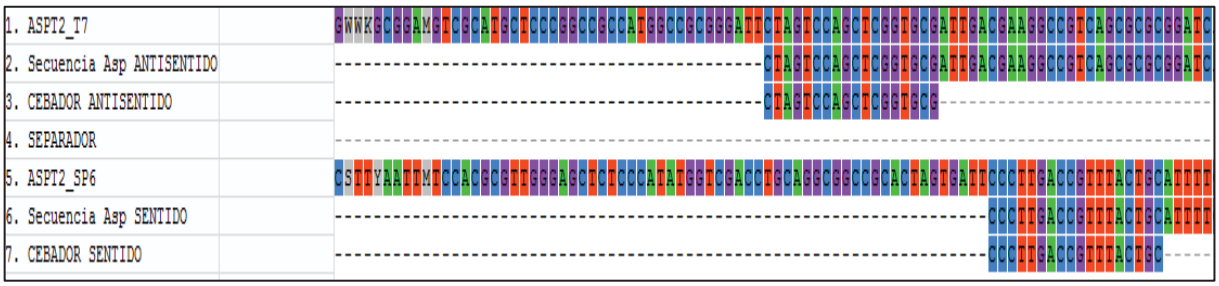


Figura N° 72: Alineamiento entre secuenciación del plásmido clonado, gen aspartil aminopeptidasa y cebadores DAP.F2/DAP.R2: Fila 1: secuenciación de plásmido con T7, Fila 2: secuencia antisentido del gen aspartil aminopeptidasa, Fila 3: cebador antisentido DAP.R2, Fila 5: secuenciación de plásmido con SP6, Fila 6: secuencia sentido, Fila 7: cebador sentido DAP.F2.

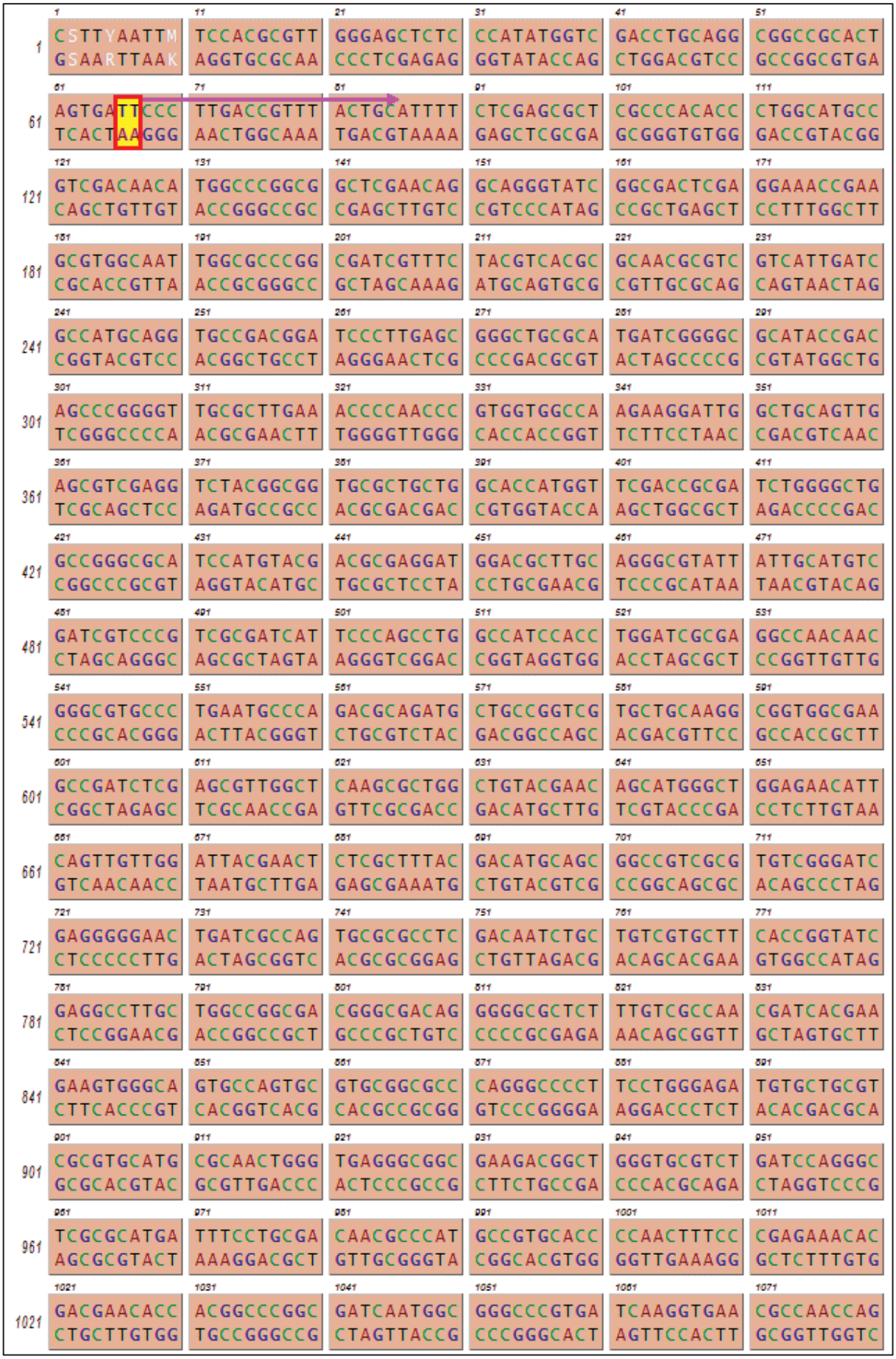

1 CSTTYAATTM GSAARTTAAK TCCACGCGTT AGGTGCGCAA GGGAGCTCTC CCCTCGAGAG CCATATGGTC GGTATACCAG GACCTGCAGG CTGGACGTCC CGGCCGCACT GCCGGCGTGA
61 AGTGATTCCC TCACTAAGGG TTGACCGTTT AACTGGCAAA ACTGCATTTT TGACGTAAAA CTCGAGCGCT GAGCTCGCGA CGCCCACACC GCGGGTGTGG CTGGCATGCC GACCGTACGG
121 GTCGACAACA CAGCTGTTGT TGGCCCGGCG ACCGGGCCGC GCTCGAACAG CGAGCTTGTC GCAGGGTATC CGTCCCATAG GGCGACTCGA CCGCTGAGCT GGAAACCGAA CCTTTGGCTT
181 GCGTGGCAAT CGCACCGTTA TGGCGCCCGG ACCGCGGGCC CGATCGTTTC GCTAGCAAAG TACGTCACGC ATGCAGTGCG GCAACGCGTC CGTTGCGCAG GTCATTGATC CAGTAACTAG
241 GCCATGCAGG CGGTACGTCC TGCCGACGGA ACGGCTGCCT TCCCTTGAGC AGGGAACTCG GGGCTGCGCA CCCGACGCGT TGATCGGGGC ACTAGCCCCG GCATACCGAC CGTATGGCTG
301 AGCCCGGGGT TCGGGCCCCA TGCGCTTGAA ACGCGAACTT ACCCCAACCC TGGGGTTGGG GTGGTGGCCA CACCACCGGT AGAAGGATTG TCTTCCTAAC GCTGCAGTTG CGACGTCAAC
361 AGCGTCGAGG TCGCAGCTCC TCTACGGCGG AGATGCCGCC TGCGCTGCTG ACGCGACGAC GCACCATGGT CGTGGTACCA TCGACCGCGA AGCTGGCGCT TCTGGGGCTG AGACCCCGAC
421 GCCGGGCGCA CGGCCCGCGT TCCATGTACG AGGTACATGC ACGCGAGGAT TGCGCTCCTA GGACGCTTGC CCTGCGAACG AGGGCGTATT TCCCGCATAA ATTGCATGTC TAACGTACAG
481 GATCGTCCCG CTAGCAGGGC TCGCGATCAT AGCGCTAGTA TCCCAGCCTG AGGGTCGGAC GCCATCCACC CGGTAGGTGG TGGATCGCGA ACCTAGCGCT GGCCAACAAC CCGGTTGTTG
541 GGGCGTGCCC CCCGCACGGG TGAATGCCCA ACTTACGGGT GACGCAGATG CTGCGTCTAC CTGCCGGTCG GACGGCCAGC TGCTGCAAGG ACGACGTTCC CGGTGGCGAA GCCACCGCTT
601 GCCGATCTCG CGGCTAGAGC AGCGTTGGCT TCGCAACCGA CAAGCGCTGG GTTCGCGACC CTGTACGAAC GACATGCTTG AGCATGGGCT TCGTACCCGA GGAGAACATT CCTCTTGTAA
661 CAGTTGTTGG GTCAACAACC ATTACGAACT TAATGCTTGA CTCGCTTTAC GAGCGAAATG GACATGCAGC CTGTACGTCG GGCCGTCGCG CCGGCAGCGC TGTCGGGATC ACAGCCCTAG
721 GAGGGGGAAC CTCCCCCTTG TGATCGCCAG ACTAGCGGTC TGCGCGCCTC ACGCGCGGAG GACAATCTGC CTGTTAGACG TGTCGTGCTT ACAGCACGAA CACCGGTATC GTGGCCATAG
781 GAGGCCTTGC CTCCGGAACG TGGCCGGCGA ACCGGCCGCT CGGGCGACAG GCCCGCTGTC GGGGCGCTCT CCCCGCGAGA TTGTCGCCAA AACAGCGGTT CGATCACGAA GCTAGTGCTT
841 GAAGTGGGCA CTTCACCCGT GTGCCAGTGC CACGGTCACG GTGCGGCGCC CACGCCGCGG CAGGGCCCCT GTCCCGGGGA TCCTGGGAGA AGGACCCTCT TGTGCTGCGT ACACGACGCA
901 CGCGTGCATG GCGCACGTAC CGCAACTGGG GCGTTGACCC TGAGGGCGGC ACTCCCGCCG GAAGACGGCT CTTCTGCCGA GGGTGCGTCT CCCACGCAGA GATCCAGGGC CTAGGTCCCG
961 TCGCGCATGA AGCGCGTACT TTTCCTGCGA AAAGGACGCT CAACGCCCAT GTTGCGGGTA GCCGTGCACC CGGCACGTGG CCAACTTTCC GGTTGAAAGG CGAGAAACAC GCTCTTTGTG
1021 GACGAACACC CTGCTTGTGG ACGGCCCGGC TGCCGGGCCG GATCAATGGC CTAGTTACCG GGGCCCGTGA CCCGGGCACT TCAAGGTGAA AGTTCCACTT CGCCAACCAG GCGGTTGGTC

Figura N° 73: Secuenciación del gen aspartil aminopeptidasa visualizado en GeneRunner. Delineado con color magenta el cebador sentido y por detrás de este la región T7 del plásmido. Delineado con color verde el cebador antisentido y por detrás de este la región SP6 del plásmido. Resaltado en amarillo las colas de adenina.

3.2.8. Criopreservación de células clonadas

Se realizó la criopreservación de las dos colonias positivas secuenciadas que contienen el plásmido con el inserto aspartil aminopeptidasa (Figura N° 74) tal como se detalló en 2.3.3.13. En este caso fueron las colonias DAP-T2 y DAP-T7.

Figura N° 74: Crecimiento de colonias clonadas con el gen aspartil aminopeptidasa. Fecha de criopreservación: 9/11/2018

CONCLUSIONES.

1. El ensamblador Spades fue el de mejor calidad en la evaluación de los datos de ensambles de *novo* del genoma de *Chromohalobacter salexigens* MP25462 caracterizado por una longitud de 3712216 pb, número de contigs 42, estadístico N50 con 447018 pb, y contenido de G + C del 63.99%.
2. Se describe el gen aspartil aminopeptidasa de *Chromohalobacter salexigens in silico* con una longitud de 1299 pb. traducidas a 432 secuencias de aminoácidos, presentando una similitud estructural con la proteasa DNPEP de *Homo sapiens* con Cov de 0.97, Z-score de 5.72, C-score de 1.73, TM-score igual a 0.96 ± 0.05 y un RMSD estimado de 3.4 ± 2.4 Å con actividad enzimática en presencia de dos iones $Co^{2+.}$
3. Al comparar el gen aspartil aminopeptidasa de *C. salexigens* MP25462 con otros genes aspartil aminopeptidasa presentes en eucariotas superiores, inferiores, bacterias y archeas, se determinó la presencia de sitios conservados en el sitio activo 263G (glutamina) y ligandos al metal vinculante Co^{2+} siendo 81H (histidina), 234D (ácido aspártico o aspartato), 235N (asparagina), 264G (glutamina), 310D (aspartato) y 404H (histidina).
4. Al diseñar y sintetizar la pareja cebadores DAP.F2 5'CCCTTGACCGTTTACTGC3' y DAP.R2 5'CTAGTCCAGCTCGGTGCG3' con *Tm* de 56 y 60° C respectivamente, se amplificó el gen completo aspartil aminopeptidasa de *C. salexigens* MP25462 en un rango de temperatura de entre 48 a 55° C.
5. Al clonar el gen completo aspartil aminopeptidasa en el vector de clonación pGEM-T se obtuvo 2 colonias DAP-T2 y DAP-T7 las que fueron criopreservadas.

RECOMENDACIONES

Investigar la biblioteca genómica de *Chromohalobacter salexigens* MP25462 así como la importancia de la información de anotación completa

Ampliar la biblioteca genómica de bacterias halófilas presentes en el Cusco.

REFERENCIAS

Alejos, L. P. A., Aragón, M. del C., & Cornejo, A. (s. f.). *Extracción y purificación de ADN*. 26.

Andrews, S. (2009). *FastQC Una herramienta de control de calidad para datos de secuencia de alto rendimiento*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arahal, D. R., García, M. T., Vargas, C., Cánovas, D., Nieto, J. J., & Ventosa, A. (2001). Chromohalobacter salexigens sp. Nov., a moderately halophilic species that includes Halomonas elongata DSM 3043 and ATCC 33174. *International Journal of Systematic and Evolutionary Microbiology*, *51*(Pt 4), 1457-1462. https://doi.org/10.1099/00207713-51-4-1457

Azevedo, F., Pereira, H., & Johansson, B. (2017). Colony PCR. En L. Domingues (Ed.), *PCR* (Vol. 1620, pp. 129-139). Springer New York. https://doi.org/10.1007/978-1-4939-7060-5_8

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., … Zagnitko, O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, *9*, 75. https://doi.org/10.1186/1471-2164-9-75

Bach, H.-J., Hartmann, A., Schloter, M., & Munch, J. C. (2001). PCR primers and functional probes for amplification and detection of bacterial genes for extracellular peptidases in single strains and in soil. *Journal of Microbiological Methods*, *44*(2), 173-182. https://doi.org/10.1016/S0167-7012(00)00239-6

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of*

Computational Biology: A Journal of Computational Molecular Cell Biology, *19*(5), 455-477. https://doi.org/10.1089/cmb.2012.0021

Barett, A. J. (1994). *Proteolytic enzymes: Serine and cysteine peptidases. 244*, 1-700.

Barrett, A. J. (1994). [1] Classification of peptidases. En *Methods in Enzymology* (Vol. 244, pp. 1-15). Elsevier. https://doi.org/10.1016/0076-6879(94)44003-4

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Cheetham, R. K., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., … Smith, A. J. (2008). Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. *Nature*, *456*(7218), 53-59. https://doi.org/10.1038/nature07517

Bertipaglia, C., Schneider, S., Jakobi, A. J., Tarafder, A. K., Bykov, Y. S., Picco, A., Kukulski, W., Kosinski, J., Hagen, W. J., Ravichandran, A. C., Wilmanns, M., Kaksonen, M., Briggs, J. A., & Sachse, C. (2016). Higher-order assemblies of oligomeric cargo receptor complexes form the membrane scaffold of the Cvt vesicle. *EMBO reports*, *17*(7), 1044-1060. https://doi.org/10.15252/embr.201541960

Borgström, E., Lundin, S., & Lundeberg, J. (2011). Large Scale Library Generation for High Throughput Sequencing. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0019119

Bronner, I. F., & Quail, M. A. (2019). Best Practices for Illumina Library Preparation. *Current Protocols in Human Genetics*, *102*(1). https://doi.org/10.1002/cphg.86

Cevec, M., Thibaudeau, C., & Plavec, J. (2010). NMR structure of the let-7 miRNA interacting with the site LCS1 of lin-41 mRNA from Caenorhabditis elegans. *Nucleic Acids Research*, *38*(21), 7814-7821. https://doi.org/10.1093/nar/gkq640

Chaikuad, A., Pilka, E. S., De Riso, A., von Delft, F., Kavanagh, K. L., Vénien-Bryan, C., Oppermann, U., & Yue, W. W. (2012). Structure of human aspartyl aminopeptidase complexed with substrate analogue: Insight into catalytic mechanism, substrate specificity and M18 peptidase family. *BMC Structural Biology*, *12*(1), 14. https://doi.org/10.1186/1472-6807-12-14

Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics (Oxford, England)*, *30*(1), 31-37. https://doi.org/10.1093/bioinformatics/btt310

Coil, D., Jospin, G., & Darling, A. E. (2015). A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics (Oxford, England)*, *31*(4), 587-589. https://doi.org/10.1093/bioinformatics/btu661

Copeland, A., O'Connor, K., Lucas, S., Lapidus, A., Berry, K. W., Detter, J. C., Del Rio, T. G., Hammon, N., Dalin, E., Tice, H., Pitluck, S., Bruce, D., Goodwin, L., Han, C., Tapia, R., Saunders, E., Schmutz, J., Brettin, T., Larimer, F., … Woyke, T. (2011). Complete genome sequence of the halophilic and highly halotolerant Chromohalobacter salexigens type strain (1H11T). *Standards in Genomic Sciences*, *5*(3), 379-388. https://doi.org/10.4056/sigs.2285059

Danna, K., & Nathans, D. (1971). Specific Cleavage of Simian Virus 40 DNA by Restriction Endonuclease of Hemophilus Influenzae. *Proc. Nat. Acad. Sci. USA*, 5.

DeAngelis, M. M., Wang, D. G., & Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Research*, *23*(22), 4742-4743.

Elorrieta Mamani, E. J. P. E. M. (2018). *Caracterización molecular de bacterias halófilas aisladas a partir de tres salares del departamento del Cusco.* [Tesis]. Universidad de San Antonio Abad del Cusco.

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, *8*(3), 186-194.

Feduchi, Blasco, Romero, & Yánez. (2010). *Bioquímica. Conceptos esenciales*. Editorial Médica Panamericana.

Gárate Palomino, A. W. (2016). *Obtención y caracterización de proteasas de excreción a partir de bacterias halófilas aisladas de las salineras de Maras* [Tesis]. Universidad de San Antonio Abad del Cusco.

Garrigues, F. (2017, mayo 23). NGS: Secuenciación de Segunda Generación. *Genética Médica Blog*. https://revistageneticamedica.com/blog/ngs-secuenciacion/

Ghafoori, H., Askari, M., & Sarikhan, S. (2016). Purification and characterization of an extracellular haloalkaline serine protease from the moderately halophilic bacterium, Bacillus iranensis (X5B). *Extremophiles*, *20*(2), 115-123. https://doi.org/10.1007/s00792-015-0804-8

Godfrey, T., & West, S. (1996). Proteases. En *Industrial Enzymology* (Segunda, pp. 1-8). Macmillan.

Gomes, J., & Steiner, W. (2004). The biocatalytic potential of extremophiles and extremozymes. *Food Technol Biotechnol*, *42*, 223-235.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072-1075. https://doi.org/10.1093/bioinformatics/btt086

Huaihua Puma, P. (2019). *Purificación, caracterización e inmovilización de proteasas de excreción de Staphylococcus sp. De las salineras de Maras—Cusco* [Tesis]. Universidad de San Antonio Abad del Cusco.

Jauk, F. (2019). *Next-Generation Sequencing (NGS) basic concepts and applications*. *23*, 21-38.

Kalisz, H. M. (1988). *Microbial Proteinases*. *36*, 1-65.

Klug, W. S., Cummings, M. R., Spencer, C. A., & Palladino, M. A. (2016). *Concepts of Genetics* (XIth edition). Pearson.

Kusumoto, K.-I., Matsushita-Morita, M., Furukawa, I., Suzuki, S., Yamagata, Y., Koide, Y., Ishida, H., Takeuchi, M., & Kashiwagi, Y. (2008). Efficient production and partial characterization of aspartyl aminopeptidase from Aspergillus oryzae. *Journal of Applied Microbiology*, *105*(5), 1711-1719. https://doi.org/10.1111/j.1365-2672.2008.03889.x

Leninger, A. L. (1975). *Biochemestry*.

Lundin, S., Stranneheim, H., Pettersson, E., Klevebring, D., & Lundeberg, J. (2010). Increased Throughput by Parallelization of Library Preparation for Massive Sequencing. *PLoS ONE*, *5*(4). https://doi.org/10.1371/journal.pone.0010029

Mamani, J. I., Pacheco, K. B., Elorrieta, P., Romoacca, P., Castelan, H., Davila, S., Sierra, J. L., & Quispe-Ricalde, M. A. (2019). Draft Genome Sequence of Halomonas elongata MH25661 Isolated from a Saline Creek in the Andes of Peru. *Microbiology Resource Announcements*, *8*(1). https://doi.org/10.1128/MRA.00934-18

Mas, E., Poza, J., Ciriza, J., Zaragoza, P., Osta, R., & Rodellar, C. (2016). Fundamento de la Reacción en Cadena de la Polimerasa (PCR). *Revista AquaTIC*, *0*(15). http://www.revistaaquatic.com/ojs/index.php/aquatic/article/view/139

McDonald, J. K. (1985). An overview of protease specificity and catalytic mechanisms: Aspects related to nomenclature and classification. *The Histochemical Journal*, *17*(7), 773-785. https://doi.org/10.1007/BF01003313

Metzker, M. L. (2010). Sequencing technologies—The next generation. *Nature Reviews. Genetics*, *11*(1), 31-46. https://doi.org/10.1038/nrg2626

Meyer, M., Briggs, A. W., Maricic, T., Höber, B., Höffner, B., Krause, J., Weihmann, A., Pääbo, S., & Hofreiter, M. (2008). From micrograms to picograms: Quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Research*, *36*(1), e5. https://doi.org/10.1093/nar/gkm1095

Michalska, K., Chhor, G., Mootz, J., Endres, M., Jedrzejczak, R., Babnigg, G., Steen, A., Lloyd, K., Joachimiak, A., & Genomics (MCSG), M. C. for S. (s. f.). Crystal structure of TET aminopeptidase from marine sediment archaeon Thaumarchaeota archaeon SCGC AB-539-E09. *TO BE PUBLISHED*. https://doi.org/10.2210/pdb5ds0/pdb

Min, T., Burley, S. K., & Shapiro, L. (s. f.). Crystal structrue of putative aminopeptidase 2 from Pseudomonas Aeruginosa. *TO BE PUBLISHED*. https://doi.org/10.2210/pdb2ijz/pdb

Min, T., Gorman, J., & Shapiro, L. (s. f.). The crystal structure of aminopeptidase I (yscI) from Borrelia burgdorferi. *TO BE PUBLISHED*. https://doi.org/10.2210/pdb1y7e/pdb

Min, T., & Shapiro, L. (s. f.-a). Crystal structure of Aminipeptidase (M18 family) from Thermotoga Maritima. *TO BE PUBLISHED*. https://doi.org/10.2210/pdb2glf/pdb

Min, T., & Shapiro, L. (s. f.-b). Crystal structure of aminopeptidase I from Clostridium acetobutylicum. *TO BE PUBLISHED*. https://doi.org/10.2210/pdb2glj/pdb

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, *51 Pt 1*, 263-273.

Natarajan, S., & Mathews, R. (2012). Cloning, expression, crystallization and preliminary X-ray crystallographic analysis of aspartyl aminopeptidase from the apeB gene of Pseudomonas aeruginosa. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *68*(Pt 2), 207-210. https://doi.org/10.1107/S1744309111054388

Nguyen, D. D., Pandian, R., Kim, D., Ha, S. C., Yoon, H.-J., Kim, K. S., Yun, K. H., Kim, J.-H., & Kim, K. K. (2014). Structural and kinetic bases for the metal preference of the M18 aminopeptidase from Pseudomonas aeruginosa. *Biochemical and Biophysical Research Communications*, *447*(1), 101-107. https://doi.org/10.1016/j.bbrc.2014.03.109

Oliver, G. R., Hart, S. N., & Klee, E. W. (2015). Bioinformatics for Clinical Next Generation Sequencing. *Clinical Chemistry*, *61*(1), 124-135. https://doi.org/10.1373/clinchem.2014.224360

Park, S.-Y., Scranton, M. A., Stajich, J. E., Yee, A., & Walling, L. L. (2017). Chlorophyte aspartyl aminopeptidases: Ancient origins, expanded families, new locations, and secondary functions. *PLoS ONE*, *12*(10). https://doi.org/10.1371/journal.pone.0185492

Petrova, T. E., Slutskaya, E. S., Boyko, K. M., Sokolova, O. S., Rakitina, T. V., Korzhenevskiy, D. A., Gorbacheva, M. A., Bezsudnova, E. Y., & Popov, V. O. (2015). Structure of the dodecamer of the aminopeptidase APDkam598 from the archaeon Desulfurococcus kamchatkensis. *Acta Crystallographica Section F: Structural Biology Communications*, *71*(3), 277-285. https://doi.org/10.1107/S2053230X15000783

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605-1612. https://doi.org/10.1002/jcc.20084

Rao, M. B., Tanksale, A. M., Ghatge, M. S., & Deshpande, V. (1998). *Molecular and biotechnical aspects of microbial proteases*. *62*, 597-635.

Romoacca Hancco, P. (2018). *Identificación molecular mediante el gen ribosomal 16s y evaluacion de las capacidades enzimaticas de bacterias halófilas aisladas en zonas*

salinas del departamento del Cusco [Tesis]. Universidad de San Antonio Abad del Cusco.

Rout, S., & Mahapatra, R. K. (2019). In silico study of M18 aspartyl amino peptidase (M18AAP) of Plasmodium vivax as an antimalarial drug target. *Bioorganic & Medicinal Chemistry*, *27*(12), 2553-2571. https://doi.org/10.1016/j.bmc.2019.03.039

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725-738. https://doi.org/10.1038/nprot.2010.5

Russo, S., & Baumann, U. (2004). Crystal Structure of a Dodecameric Tetrahedral-shaped Aminopeptidase. *Journal of Biological Chemistry*, *279*(49), 51275-51281. https://doi.org/10.1074/jbc.M409455200

Sambu, S. (2015). A Bayesian approach to optimizing cryopreservation protocols. *PeerJ*, *3*. https://doi.org/10.7717/peerj.1039

Sari, E., Loğoğlu, E., & Öktemer, A. (2015). Purification and characterization of organic solvent stable serine alkaline protease from newly isolated *Bacillus circulans* M34: Purification of Bacillus circulans M34. *Biomedical Chromatography*, *29*(9), 1356-1363. https://doi.org/10.1002/bmc.3431

Silva, F. J., Belda, E., & Talens, S. E. (2006). Differential annotation of tRNA genes with anticodon CAT in bacterial genomes. *Nucleic Acids Research*, *34*(20), 6015-6022. https://doi.org/10.1093/nar/gkl739

Sinha, R., Stanley, G., Gulati, G. S., Ezran, C., Travaglini, K. J., Wei, E., Chan, C. K. F., Nabhan, A. N., Su, T., Morganti, R. M., Conley, S. D., Chaib, H., Red-Horse, K., Longaker, M. T., Snyder, M. P., Krasnow, M. A., & Weissman, I. L. (2017). *Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing* [Preprint]. Molecular Biology. https://doi.org/10.1101/125724

Sivaraman, K. K., Oellig, C. A., Huynh, K., Atkinson, S. C., Poreba, M., Perugini, M. A., Trenholme, K. R., Gardiner, D. L., Salvesen, G., Drag, M., Dalton, J. P., Whisstock, J. C., & McGowan, S. (2012). X-ray Crystal Structure and Specificity of the Plasmodium falciparum Malaria Aminopeptidase PfM18AAP. *Journal of Molecular Biology*, *422*(4), 495-507. https://doi.org/10.1016/j.jmb.2012.06.006

Story, S. V., Shah, C., Jenney, F. E., & Adams, M. W. W. (2005). Characterization of a Novel Zinc-Containing, Lysine-Specific Aminopeptidase from the Hyperthermophilic Archaeon Pyrococcus furiosus. *Journal of Bacteriology*, *187*(6), 2077-2083. https://doi.org/10.1128/JB.187.6.2077-2083.2005

Strom, S., Shiskova, E., Hahm, Y., & Grover, N. (2015). Thermodynamic examination of 1- to 5-nt purine bulge loops in RNA and DNA constructs. *RNA*, *21*(7), 1313-1322. https://doi.org/10.1261/rna.046631.114

Su, M.-Y., Peng, W.-H., Ho, M.-R., Su, S.-C., Chang, Y.-C., Chen, G.-C., & Chang, C.-I. (2015). Structure of yeast Ape1 and its role in autophagic vesicle formation. *Autophagy*, *11*(9), 1580-1593. https://doi.org/10.1080/15548627.2015.1067363

Swords, W. E. (2003). Chemical Transformation of E. coli. En N. Casali & A. Preston, *E. coli Plasmid Vectors* (Vol. 235, pp. 49-54). Humana Press. https://doi.org/10.1385/1-59259-409-3:49

Teuscher, F., Lowther, J., Skinner-Adams, T. S., Spielmann, T., Dixon, M. W. A., Stack, C. M., Donnelly, S., Mucha, A., Kafarski, P., Vassiliou, S., Gardiner, D. L., Dalton, J. P., & Trenholme, K. R. (2007). The M18 Aspartyl Aminopeptidase of the Human Malaria Parasite *Plasmodium falciparum*. *Journal of Biological Chemistry*, *282*(42), 30817-30826. https://doi.org/10.1074/jbc.M704938200

Thebti, W., Riahi, Y., & Belhadj, O. (2016). Purification and Characterization of a New Thermostable, Haloalkaline, Solvent Stable, and Detergent Compatible Serine Protease

from Geobacillus toebii Strain LBT 77. *BioMed Research International*, *2016*. https://doi.org/10.1155/2016/9178962

Ventosa, A., Gutierrez, M. C., Garcia, M. T., & Ruiz-Berraquero, F. (1989). *Classification of ‘‘Chromobacteriummarismortui” in a New Genus, Chromohalobacter gen. Nov. , As Chromohalobacter marismortui comb. Nov., norn. Rev.* 5.

Walter, N. G., & Engelke, D. R. (2002). Ribozymes: Catalytic RNAs that cut things, make things, and do odd and useful jobs. *Biologist (London, England)*, *49*(5), 199-203.

Wilk, S., Wilk, E., & Magnusson, R. P. (1998). Purification, Characterization, and Cloning of a Cytosolic Aspartyl Aminopeptidase. *Journal of Biological Chemistry*, *273*(26), 15961-15970. https://doi.org/10.1074/jbc.273.26.15961

Wilk, S., Wilk, E., & Magnusson, R. P. (2002). Identification of histidine residues important in the catalysis and structure of aspartyl aminopeptidase. *Archives of Biochemistry and Biophysics*, *407*(2), 176-183. https://doi.org/10.1016/S0003-9861(02)00494-0

Woese, C. R., & Gutell, R. R. (1989). Evidence for several higher order structural elements in ribosomal RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(9), 3119-3122. https://doi.org/10.1073/pnas.86.9.3119

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7-8. https://doi.org/10.1038/nmeth.3213

Yildirim, V., Baltaci, M. O., Ozgencli, I., Sisecioglu, M., Adiguzel, A., & Adiguzel, G. (2017). Purification and biochemical characterization of a novel thermostable serine alkaline protease from Aeribacillus pallidus C10: A potential additive for detergents. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *32*(1), 468-477. https://doi.org/10.1080/14756366.2016.1261131

Yuga, M., Gomi, K., Klionsky, D. J., & Shintani, T. (2011). Aspartyl Aminopeptidase Is Imported from the Cytoplasm to the Vacuole by Selective Autophagy in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, *286*(15), 13704-13713. https://doi.org/10.1074/jbc.M110.173906

Zamora, C. (1996). *Las regiones ecológicas del Perú*. Rodriguez L.O.

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821-829. https://doi.org/10.1101/gr.074492.107

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*, 40. https://doi.org/10.1186/1471-2105-9-40

Zhang, Y., & Skolnick, J. (2004). SPICKER: A clustering approach to identify near-native protein folds. *Journal of Computational Chemistry*, *25*(6), 865-871. https://doi.org/10.1002/jcc.20011

ANEXOS

Anexo 1: Comandas de programas bioinformáticos.

A. Trimmomatic.

```
kevinpc@kevinpc-Satellite-L755:~$ /Descargas/Trimmomatic-0.36$ java -jar trimmomatic-0.36.jar PE -phred33 /home/kevinpc/Escritorio/secuencias_C_salexigens/pe_1_S15_L001_R1_001.fastq /home/kevinpc/Escritorio/secuencias_C_salexigens/pe_1_S15_L001_R2_001.fastq /home/kevinpc/Escritorio/secuencias_C_salexigens/pe_1_S15_L001_R1_PE.fastq /home/kevinpc/Escritorio/secuencias_C_salexigens/pe_1_S15_L001_R1_UP.fastq /home/kevinpc/Escritorio/secuencias_C_salexigens/pe_1_S15_L001_R2_PE.fastq /home/kevinpc/Escritorio/secuencias_C_salexigens/pe_1_S15_L001_R2_UP.fastq ILLUMINACLIP:/home/kevinpc/Descargas/Trimmomatic-0.36/adapters/NexteraPE-PE.fa:2:30:10 LEADING:3 TRAILING:28 HEADCROP:20 MINLEN:100 SLIDINGWINDOW:4:15
```

Figura N° 75. Comando de ejecución del programa Trimmomatic en la terminal del sistema operativo Linux con sus parámetros de recorte de calidad.
Fuente propia

B. Kmergenie.

```
kevinpc@kevinpc-Satellite-L755:~$ /home/kevinpc/Descargas/kmergenie-1.7051/kmergenie-1.7051/kmergenie /media/kevinpc/KEVIN/DATOSchrmo_replanteo/secuencias_C_salexigens/pe_1_S15_L001_R1_001.fastq
```

Figura N° 76. Comando de ejecución del programa KmerGenie en la terminal del sistema operativo Linux.
Fuente propia

C. Velvet.

```
7                                                                1     2   3   4       5       6 8
kevinpc@kevinpc-Satellite-L755:~/Descargas/Aplicaciones/velvet_1.2.10$ ../velveth ensamble 21 -fastq -shortPaired -separate '/media/kevinpc/KEVIN/DATOSchrmo_replanteo/secuencias_C_salexigens/secuencias_conTrimming/pe_1_S15_L001_R1_PE.fastq' '/media/kevinpc/KEVIN/DATOSchrmo_replanteo/secuencias_C_salexigens/secuencias_conTrimming/pe_1_S15_L001_R2_PE.fastq'
```

Figura N° 77. Comando de ejecución del ensamblador Velvet (fase de indexación) en la terminal del sistema operativo Linux; 1: velveth (indexación de secuencias); 2: ensamble (nombre de la carpeta de salida); 3: 21 (tamaño de k-mer); 4: -fastq (formato de los archivos de entrada); 5: shortPaired (); 6: -separate (); 7 y 8 dirección de los archivos fastq del genoma de Chromohalobacter salexigens MP25462.
Fuente propia.

```
                                                                        1        2
kevinpc@kevinpc-Satellite-L755:~/Descargas/Aplicaciones/velvet_1.2.10$ ../velvetg ensamble
```

Figura N° 78. Comanda de ejecución del ensamblador Velvet (fase de ensamblado) en la terminal del sistema operativo Linux. 1: velvetg (ensamblado de secuencias indexadas); 2: ensamble (nombre de la carpeta de salida).
Fuente propia.

D. A5.

```
                  4                      3          1         2
kevinpc@kevinpc-Satellite-L755:~/Descargas/Aplicaciones/a5$ perl a5_pipeline.pl '/media/kevinpc/KEVIN/DATOSchrmo_replanteo/secuencias_C_salexigens/pe_1_S15_L001_R1_001.fastq' '/media/kevinpc/KEVIN/DATOSchrmo_replanteo/secuencias_C_salexigens/pe_1_S15_L001_R2_001.fastq' ensambleA5
```

Figura N° 79. Comanda de ejecución del ensamblador A5 en la terminal del sistema operativo Linux. 1: a5_pipeline.pl (); 2 y 3 ubicación de las secuencias fastq; 4: ensambleA5 archivo de salida.
Fuente propia.

E. Spades.

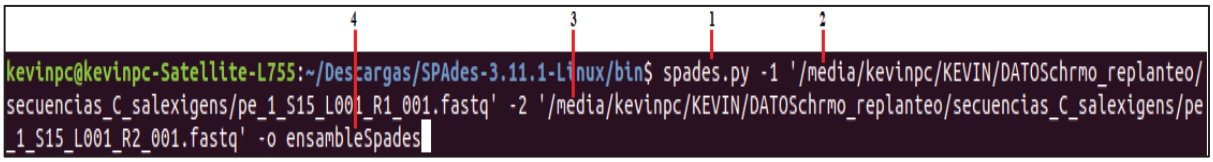


Figura N° 80. Comanda de ejecución del ensamblador Spades en la terminal del sistema operativo Linux. 1: spades.py (); 2 y 3 ubicación de las secuencias fastq; 4: ensambleSpades archivo de salida.
Fuente propia.

F. Quast.

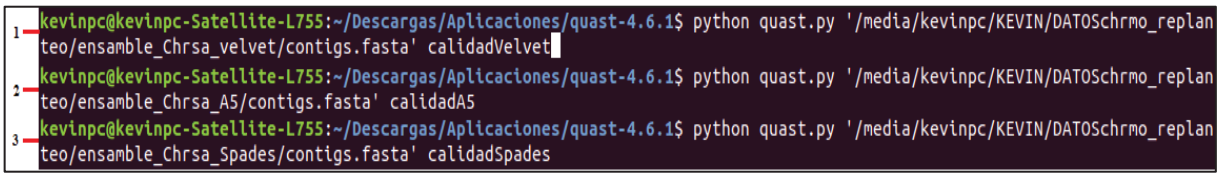


Figura N° 81. Comandas del programa Quast en los productos ensamblados de Velvet (1), A5 (2) y Spades (3).
Fuente propia

Anexo 2: Medios de cultivo.

A. Medio SW.

Componentes del medio SW al 15% de NaCl

Componentes	gr/1000ml
NaCl	150 gr
Extracto de levadura	5 gr
Cloruro de magnesio hexahidratado	5 gr
Sulfato de magnesio heptahidratado	5,83 gr
Cloruro de potasio	1,17 g
Bicarbonato de sodio	0,03 g
Cloruro de calcio dihidratado	0,083g
pH final	8,0
Agar	20 gr

Preparación

- Preparar un stock de solución de sales a 5X para un volumen de 100 ml (cloruro de sodio, cloruro de magnesio hexahidratado, sulfato de magnesio, cloruro de potasio heptahidratado, bicarbonato de sodio, cloruro de calcio dihidratado) y los demás componentes (Agar y extracto de levadura) en recipientes diferentes. Autoclavar por 15 minutos a 121° C
- Agregar la solución de sales autoclavada al frasco de los componentes para una concentración de 1X
- Mezclar y plaquear en cámara de flujo laminar.

B. Medio LB (*Luria-Bertani*; Sambrook *et al*., 1989).

i. Componentes del medio LB

Componentes	gr/200ml
NaCl	2 gr
Extracto de levadura	1 gr
Triptona	2 gr
Agar	4 gr

Autoclavar por 15 minutos a 121° C a una presión de 15 libras.

ii. Medio LB + ampicilina A 100 mg/ml.

- Terminado el proceso de autoclavado del medio LB dejar enfriar hasta que la mezcla tenga una temperatura menor a 50° C
- Agregar 200 µl de ampicilina 100 mg/ml para 200 ml de medio LB.
- Todo esto debe ser llevado a cabo en una cámara de flujo laminar.

C. Medio SOC (Hanahan *et al.*, 1991).

Componentes del medio SOC

Componentes	gr/100ml
Triptona	2 gr
Extracto de levadura	0.5 gr
NaCl 1M	1 gr
KCl 1M	0.25 gr
$MgCl_2$ 2M	1 gr
Glucosa 2M	1 gr

Preparación:

- Agregar triptona, extracto de levadura, NaCl y KCl a 98 ml de agua bidestilada
- Agitar y autoclavar por 15 min a 121° C y dejar enfriar a temperatura ambiente
- Agregar 2M de Mg y Glucosa, cada uno de estos a concentración final de 20 mM
- El $MgCl_2$ y la glucosa se esterilizan por filtración con filtros de 0.2 µm para ambos casos

Anexo 3: Gel de agarosa.

A. Al 0.8%

Componentes del gel de agarosa al 0.8%

Componentes	Cantidad
Agarosa	0.8 gr
Agua bidestilada	100 ml
Real safe	1 µl

Preparación:

- Pesar 0.8 gr de agarosa y disolver en 100ml de tampón TAE 1X,
- Homogenizar, disolver y llevar a ebullición al microondas
- Esperar a que enfrié lo suficiente y agregar 1 µl *Real safe* y homogenizar.
- Verter a la bandeja con los peines y dejar gelificar aproximadamente de 15 a 30 minutos.
- Colocar el gel en la cámara de electroforesis y cubrir con TAE 1X y retirar con cuidado los peines.

B. Al 1 %

Componentes del gel de agarosa al 0.8%

Componentes	Cantidad
Agarosa	1 gr
Agua bidestilada	100 ml
Real safe	1 µl

Preparación:

- Pesar 1 gr de agarosa y disolver en 100ml de tampón TAE 1X,
- Homogenizar, disolver y llevar a ebullición al microondas
- Esperar a que enfrié lo suficiente y agregar 1 µl *Real safe* y homogenizar.
- Verter a la bandeja con los peines y dejar gelificar aproximadamente de 15 a 30 minutos.
- Colocar el gel en la cámara de electroforesis y cubrir con TAE 1X y retirar con cuidado los peines.

Anexo 4: Tampón de electroforesis tris/ácido acético/EDTA (TAE 50x).

Componente del tampón TAE 50X

Componentes	Contidad
Tris base	242g
Ácido acético glacial	57.1ml
EDTA 0.5M (pH 8.0)	100ml

Preparación:

- Medir 57.1 ml de ácido acético glacial en un envase de 500 ml
- Agregar 100 ml de una solución 0.5 M a pH 8.
- Pesar 242.28 g de tris base y agregar a la solución
- La solución de trabajo se obtuvo diluyendo el stock a una concentración 1X

Anexo 5: Correo electrónico de resultados de secuenciación.

ULL | Universidad de La Laguna
Vicerrectorado de Investigación

SEGAI

Página 1 de 1

INFORME DE PRESTACIÓN DE SERVICIO

Referencia	R.69/PO-SG
Ref.:	85
Nº:	47444

Resumen de datos Técnicos:

TÉCNICAS	PROCEDIMIENTOS	CANTIDAD
☐ Purificación de ADN + electroforesis	**Kit de extracción de ADN:** ☐ DNeasy Blood & Tissue Kit (Qiagen) ☐ DNeasy Plant Mini Kit (Qiagen) ☐ Otros:	
☐ PCR estándar + electroforesis	**Kit de PCR utilizado:** ☐ AmpONE taq DNA Polymerase (GenoAll) ☐ Otros:	
☐ Electroforesis ADN	**Marcador de peso molecular:** ☐ 1 Kb DNA Ladder (Promega) ☐ Otros: 1 Kb Ladder (New England Biolabs)	
☐ Purificación de producto de PCR	**Kit de purificación utilizado:** ☐ ExoSAP-IT (Affimetrix) ☐ Zymoclean™ Gel DNA Recovery Kit (ZimoR) ☐ Otros:	
☐ PCR a tiempo real	**Kit de PCR a tiempo real utilizado:** ☐ iQ™ SYBR® Green Supermix (BIORAD) ☐ Otros:	
☐ Síntesis de cDNA	**Kit de síntesis de cDNA utilizado:** ☐ iScript™ cDNA Synthesis Kit (BIORAD) ☐ Otros:	
☒ Secuenciación de ADN	**Kit de reacción de secuenciación:** ☒ Big Dye Terminator v3.1 (Applied Biosystems) ☐ Otros: **Purificación reacción de secuenciación:** ☒ Etanol + acetato amónico + EDTA ☐ Otros: **Polímero en Genetic Analyzer AB 3500:** ☒ POP-7 ☐ Otros:	75
☐ Análisis de fragmentos de ADN	**Estándar de peso molecular:** ☐ GeneScan 600 LIZ Size Standard ☐ Otros: **Polímero en Genetic Analyzer AB 3500:** ☐ POP-7 ☐ Otros:	
☐ Otro tipo de análisis	☐ Otros:	

Observaciones a los resultados:	Personal del Servicio:
Se ha procedido al análisis de 75 secuencias, cuyo pago ha sido realizado a través del sistema U2T (hoja de encargo: PS-A1810069-003).	Fecha 23/11/2018 Fdo.: Mario Andrés González Cancedo.

Universidad de La Laguna, Vicerrectorado de Investigación, SEGAI. Apartado: 455 CP: 38200
San Cristóbal de La Laguna/ Santa Cruz de Tenerife. Tel.: 922318523 - 922318940.
E-mail: segainve@ull.edu.es. CIF: Q3818001D.

EFQM

Anexo 6: Certificados.

ULL | Universidad de La Laguna

225 ANIVERSARIO 1792 - 2017

INSTITUTO UNIVERSITARIO DE ENFERMEDADES TROPICALES Y SALUD PÚBLICA DE CANARIAS
Universidad de La Laguna

ENRIQUE MARTINEZ CARRETERO, PROFESOR TÍTULAR DE PARASITOLOGÍA Y DIRECTOR DEL INSTITUTO UNIVERSITARIO DE ENFERMEDADES TROPICALES Y SALUD PÚBLICA DE CANARIAS DE LA UNIVERSIDAD DE LA LAGUNA.

CERTIFICA

Que KEVIN BRANDON PACHECO CAPCHA, Bachiller en Biología de la Universidad Nacional de San Antonio Abad del Cusco (UNSAAC), e integrante como tesista de pre-grado del proyecto "Inhibición de proteasas de *Leishmania* a través de péptidos obtenidos de proteínas de *Lupinus mutabilis* por acción enzimática de proteasas de bacterias halófilas: Nuevas perspectivas de búsqueda de moléculas antileishmanicidas", ha realizado una pasantía en el Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias (IUETSPC) de la Universidad de La Laguna (ULL), desde el 02 de octubre hasta el 21 de diciembre de 2018. Las tareas realizadas han sido experimentos de amplificación, secuenciación y clonaje de proteasas de bacterias halófilas.

La pasantía se desarrolló en el marco del proyecto "Inhibición de proteasas de *Leishmania* a través de péptidos obtenidos de proteínas de *Lupinus mutabilis* por acción enzimática de proteasas de bacterias halófilas: Nuevas perspectivas de búsqueda de moléculas antileishmanicidas", convenio N° 023-2018-UNSAAC autorizado a través de la Resolución N° R-1348-2018-UNSAAC, donde la Universidad de La Laguna participa como entidad colaboradora y en el marco del Convenio Marco entre la UNSAAC y la ULL.

En La Laguna, 21 de diciembre de 2018

Enrique Martínez Carretero

Avda. Astrofísico Francisco Sánchez, s/n C.P. 38203 La Laguna – Tenerife – España (Teléfono (922) 31-8484

Universidad de La Laguna

Dña. Pilar Foronda Rodríguez, Profesora Titular de Parasitología, e Investigadora del Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias, de la Universidad de La Laguna

CERTIFICA

Que KEVIN BRANDON PACHECO CAPCHA, Bachiller en Biología de la Universidad Nacional de San Antonio Abad del Cusco (UNSAAC), e integrante como tesista de pre-grado del proyecto "Inhibición de proteasas de Leishmania a través de péptidos obtenidos de proteínas de Lupinus mutabilis por acción enzimática de proteasas de bacterias halófilas: Nuevas perspectivas de búsqueda de moléculas antileishmanicidas", ha realizado una pasantía en el Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias (IUETSPC) de la Universidad de La Laguna (ULL), desde el 02 de octubre hasta el 21 de diciembre de 2018. Las tareas realizadas han sido experimentos de amplificación, secuenciación y clonaje de proteasas de bacterias halófilas.

La pasantía se desarrolló en el marco del proyecto "Inhibición de proteasas de Leishmania a través de péptidos obtenidos de proteinas de Lupinus mutabilis por acción enzimática de proteasas de bacterias halófilas: Nuevas perspectivas de búsqueda de moléculas antileishmanicidas", convenio Nº 023-2018-UNSAAC autorizado a través de la Resolución Nº R-1348-2018-UNSAAC, donde la Universidad de La Laguna participa como entidad colaboradora y en el marco del Convenio Marco entre la UNSAAC y la ULL.

.

En La Laguna, 21 de diciembre de 2018

Fdo: Pilar Foronda Rodríguez

Anexo 7: Resolución.

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO

RESOLUCION NRO. R- 1349 -2018-UNSAAC.

CUSCO, 25 SET. 2018

EL RECTOR DE LA UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO.

VISTO, *el Oficio Nro. 741-2018-VRIN-UNSAAC registrado con el Expediente Nro. 846594 presentado por el* **DR. GILBERT ALAGÓN HUALLPA**, *Vicerrector de Investigación de la Institución, solicitando subvención económica para pasantia internacional que se indica, y;*

CONSIDERANDO:

Que, con Resolución Nro. R-0267-2018-UNSAAC de fecha 22 de febrero de 2018, se toma conocimiento de la Resolución de Dirección Ejecutiva Nro. 016-2018-FONDECYT-DE de fecha 02 de febrero de 2018, emitida por el FONDO DE DESARROLLO CIENTÍFICO, TECNOLÓGICO Y DE INNOVACIONES TECNOLÓGICA –FONDECYT, RATIFICA LOS RESULTADOS DEL CONCURSIO DEL ESQUEMA FINANCIERO EO41-2017-UNSAAC-02, denominado "PROYECTOS DE INVESTIGACIÓN";

Que, asimismo con Resolución Nro. R-0392-2018-UNSAAC de fecha 21 de marzo de 2018, se aprueba el presupuesto en la suma de S/500,000.00 soles (Quinientos Mil con 00/100 soles) para ejecución de gastos del Proyecto de Investigación Tipo Aplicada Avanzado titulado "INHIBICIÓN DE PROTEASAS DE LEISHMANIA A TRAVÉS DE PÉPTIDOS OBTENIDOS DE PROTEÍNAS DE LUPINUS MUTABILIS POR ACCIÓN ENZIMÁTICA DE PROTEASAS DE BACTERIAS HALÓFILAS: NUEVAS PERSPECTIVAS DE BÚSQUEDA DE MOLÉCULAS ANTILEISHMANICIDAS" Esquema Financiero E041-2017-UNSAAC-02 denominado "PROYECTOS DE INVESTIGACION", de conformidad con la Resolución de Dirección Ejecutiva Nro. 016-2018/FONDECYT-DE de fecha 02 de febrero de 2018, emitida por el FONDO NACIONAL DE DESARROLLO CIENTÍFICO, TECNOLÓGICO Y DE INNOVACION TECNOLÓGICA que APRUEBA LOS RESULTADOS DEL CONCURSO DEL Esquema Financiero E041-2017UNSAAC-02, denominado "PROYECTOS DE INVESTIGACION" y ratificada por Resolución Nro. R-0267-2018-UNSAAC de fecha 22 de febrero de 2018, bajo responsabilidad de la DRA. MARIA ANTONIETA QUISPE RICALDE, Profesora Auxiliar a Tiempo Parcial de 20 horas en el Departamento Académico de Biología de la Facultad de Ciencias;

Que, mediante expediente del Visto, el Vicerrector de Investigación de la Institución, en atención al Oficio Nro. 018-2018-Proy-Contrato Nro. 23-208-UNSAAC remitido por la Dra. María Antonieta Quispe Ricalde, Coordinadora General del Proyecto de Investigación "INHIBICIÓN DE PROTEASAS DE LEISHMANIA A TRAVÉS DE PÉPTIDOS OBTENIDOS DE PROTEINAS DE LUPINUS MUTABILIS POR ACCIÓN ENZIMÁTICA DE PROTEASAS DE BACTERIAS

Calle Tigre 127 - Telefax: 084 - 224891 - Apdo. 921 CUSCO - PERÚ E-mail: secretariageneral@unsaac.edu.pe

HALÓFILAS: NUEVAS PERSPECTIVAS DE BÚSQUEDA DE MOLÉCULAS ANTILEISHMANICIDAS", solicita subvención económica para el estudiante integrante del equipo de investigación BR. KEVIN BRANDON PACHECO CAPCHA, Tesista de Pregrado quien realizará pasantía internacional en el Instituto de Enfermedades Tropicales y Salud Pública de Canarias en la Universidad de La Laguna (Tenerife-España) del 29 de septiembre al 20 de diciembre de 2018 estancia que tiene como objetivo realizar el clonaje de una proteasa halófila Halomonas elongata;

Que, por tal motivo solicita autorización de viaje al exterior a favor del BR. KEVIN BRANDON PACHECO CAPCHA, Tesista de Pregrado, quien realizará pasantía internacional en el Instituto de Enfermedades Tropicales y Salud Pública de Canarias en la Universidad de La Laguna (Tenerife-España) del 29 de setiembre al 20 de diciembre de 2018 (83 días), previsto en el Plan Operativo del referido Proyecto de Investigación, concediéndosele apoyo económico por la suma de S/12,000.00 (doce mil con 00/100 soles), importe que será utilizado para la compra de pasajes aéreos y gastos de manutención durante el periodo de estancia;

Que, al respecto a través del Informe Nro. 042-2018-FEDR, la Mgt. Fiorella Edilma Díaz Roca, Monitor Externo FONDECYT, considera viable se otorgue la subvención económica para la realización de la PASANTÍA INTERNACIONAL a efectuarse del 29 de septiembre al 20 de diciembre de 2018 (83 días) con un monto aprobado de S/12,000.00 soles (doce mil con 00/100 soles), a favor del estudiante KEVIN BRANDON PACHECO CAPCHA, Tesista de pregrado del Proyecto de Investigación citado precedentemente, conforme aparece en el Plan Operativo;

Que, conforme a lo señalado la Jefa de la Unidad de Presupuesto de la Dirección de Planificación expide la Certificación de Crédito Presupuestario Nro. 3512-2018, indicando la afectación presupuestal para atender lo solicitado, conforme al anexo que forma parte de la presente Resolución;

Que, la Autoridad Universitaria ha tomado conocimiento del referido expediente y ha dispuesto la emisión de la resolución correspondiente;

Estando a lo solicitado, Ley Nro. 30693, en uso de las atribuciones conferidas por la Ley y el Estatuto Universitario;

RESUELVE:

***PRIMERO.- AUTORIZAR** el viaje del **BR. KEVIN BRANDON PACHECO CAPCHA** integrante del equipo de investigación, en calidad de Tesista de Pregrado del Proyecto de Investigación denominado: "INHIBICIÓN DE PROTEASAS DE LEISHMANIA A TRAVÉS DE PÉPTIDOS OBTENIDOS DE PROTEÍNAS DE LUPINUS MUTABILIS POR ACCIÓN ENZIMÁTICA DE PROTEASAS DE BACTERIAS HALÓFILAS: NUEVAS PERSPECTIVAS DE BÚSQUEDA DE MOLÉCULAS ANTILEISHMANICIDAS", a la ciudad de Tenerife-España, con el propósito de*

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO

realizar pasantía internacional en el Instituto de Enfermedades Tropicales y Salud Pública de Canarias en la Universidad de La Laguna (Tenerife-España) a desarrollarse del 29 de setiembre al 20 de diciembre de 2018; estancia que tiene como objetivo realizar el clonaje de una proteasa halófila Halomonas elongata, en mérito a lo señalado en la parte considerativa de la presente Resolución.

***SEGUNDO.- DISPONER** que la Dirección General de Administración, otorgue apoyo económico a favor del **BR. KEVIN BRANDON PACHECO CAPCHA**, Tesista de Pregrado del referido Proyecto de Investigación, en el importe de S/ 12,000.00 soles (DOCE MIL CON 00/100 SOLES), para cubrir gastos de pasajes aéreos y manutención considerado dentro del Plan Operativo de dicho Proyecto.*

***TERCERO.- EL EGRESO** que se origine por la aplicación de la presente Resolución se atenderá con cargo a la Certificación de Crédito Presupuestario Nro. 3512-2018, de acuerdo al anexo que forma parte de la presente Resolución.*

***CUARTO.- DEJAR ESTABLECIDA,** la obligación del estudiante de presentar ante la Dirección General de Administración, rendición de cuenta con los comprobantes de pago que constituyen la sustentación del gasto, en el término de quince (15) días calendario de su retorno a la ciudad del Cusco, perdiendo el derecho luego del citado periodo, asimismo informe ante el Vicerrectorado de Investigación y Despacho Rectoral sobre su participación en la referida estancia.*

El Vicerrectorado de Investigación y la Dirección General de Administración, adoptarán las acciones complementarias para el cumplimiento de la presente Resolución.

REGISTRESE, COMUNIQUESE Y ARCHIVESE

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
RECTORADO
Dr. BALTAZAR NICOLÁS CÁCERES HUAMBO
RECTOR

***TR:** VRAC.-VRIN.-VRAD.-OCI.-DIRECCIÓN DE PLANIFICACIÓN.- UNIDAD DE PRESUPUESTO.- DIGA.- U. FINANZAS.- A. TESORERIA.- A. EJECUCIÓN PRESUPUESTAL.- A. INTEGRACIÓN CONTABLE.- U. LOGISTICA.- A. MANTENIMIENTO Y SERVICIOS.- DIRECCIÓN DE GESTIÓN DE LA INVESTIGACIÓN.- UNIDAD DE TALENTO HUMANO.- A. ESCALAFON (02).- FONCECYT.- **DRA. MARÍA ANTONIETA QUISPE RICALDE**.- **BR. KEVIN BRANDON PACHECO CAPCHA**.- ASESORIA JURÍDICA.- RED DE COMUNICACIONES.- ARCHIVO CENTRAL.- ARCHIVO S.G.- BNCH/LPFP/MCCH/JGPF/JGL.*

Lo que transcribo a Ud. Para su conocimiento y fines consiguientes.

Atentamente,

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
Mgt. LINO PRISCILIANO FLORES PACHECO
Secretario General

Calle Tigre 127 - Telefax: 084 - 224891 - Apdo. 921 CUSCO - PERÚ E-mail: secretariageneral@unsaac.edu.pe

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
DIRECCION DE PLANIFICACION
UNIDAD DE PRESUPUESTO

CERTIFICACION DE CRÉDITO PRESUPUESTARIO N° 3512

DEPENDENCIA SOLICITANTE : Proyecto de Investigación "INHIBICIÓN DE PROTEASAS DE LEISHMANIA A TRAVES DE PEPTIDOS OBTENIDOS DE PROTEINAS DE LUPINUS MUTABILIS POR ACCION ENZIMATICA DE PROTEASAS DE BACTERIAS HALOFILAS: NUEVAS PERSPECTIVAS DE BUSQUEDA DE MOLECULAS ANTILEISHMANICIDAS"

RESPONSABLE : Dra. MARIA ANTONIETA QUISPE RICALDE

REFERENCIA : EXP. 846594 Oficio N° 018-2018-Proy-Contrato N° 23-2018-UNSAAC
Informe N° 042-2018-FEDR
Oficio N° 018-2018-VRIN-UNSAAC

La Unidad de presupuesto de la Dirección de Planificación, CERTIFICA que existe crédito presupuestal, debiendo afectarse de la siguiente manera:

REQUERIMIENTO	Solicita Subvención a estudiante para pasantía internacional a KEVIN BRANDON PACHECO CAPCHA, SEGÚN LA RESPONSABLE DEL Proyecto y la monitora externa del Fondecyt, determinan que se encuentra considerado dentro del Plan operativo el importe de S/ 12,000.00 para la pasantía en Tenerife-España del 29/09/2018 al 20/12/2018.
CATEGORIA PRESUPUESTAL	Asignaciones presupuestarias que no resultan en Productos
PROGRAMA PRESUPUESTAL	9002 Asignaciones presupuestarias que no resultan en Productos
PRODUCTO/PROYECTO	3.999999 Sin Producto
ACTIV/OBRA/ACC INVERSION	5000894 Investigación Científica y Desarrollo Tecnológico
META	0046
ESPECIFICA DE GASTO	25.31.11
MONTO DE CERTIFICACION	S/ 12,000.00
FUENTE DE FINANCIAMIENTO	RECURSOS DETERMINADOS
FECHA DE CERTIFICACION	14 de Setiembre de 2018
CERTIFICACION VALIDA POR 15 DIAS	

La presente certificación es requisito para financiar y ejecutar única y exclusivamente para lo solicitado, por lo que es pertinente el acto administrativo correspondiente.
NOTA.- La determinación de PCA no convalida los actos o acciones que no se ciñan a la normativa vigente; y está destinado a comprometer un gasto, previo cumplimiento de las disposiciones legales vigentes que regulan el objeto materia del compromiso

DIRECCIÓN DE PLANIFICACIÓN
UNIDAD DE PRESUPUESTO
Mgt. Mercedes Pinto Castillo
JEFE

OFICINA DE PLANIFICACION UNIVERSITARIA
17 SET. 2018
RECIBIDO
HORA .....

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
VICERRECTORADO DE INVESTIGACIÓN
17 SET. 2018
RECIBIDO
HORA: .....

Aprobado por Resolución R. N° 1349-2018 -UNSAAC

c.c.
Archivo
MPC/goh

Anexo 8: Anotación de MP25462

Localización	Inicio	Final	Hebra	Función
NODE_10_length_140745_cov_22.094447_1958_474	1958	474	-	2-methylcitrate dehydratase (EC 4.2.1.79)
NODE_10_length_140745_cov_22.094447_3139_2012	3139	2012	-	2-methylcitrate synthase (EC 2.3.3.5)
NODE_10_length_140745_cov_22.094447_4085_3204	4085	3204	-	Methylisocitrate lyase (EC 4.1.3.30)
NODE_10_length_140745_cov_22.094447_4843_4082	4843	4082	-	Propionate catabolism operon transcriptional regulator of GntR family [predicted]
NODE_10_length_140745_cov_22.094447_5850_4987	5850	4987	-	alpha/beta hydrolase
NODE_10_length_140745_cov_22.094447_6066_7412	6066	7412	+	Para-aminobenzoate synthase, aminase component (EC 2.6.1.85)
NODE_10_length_140745_cov_22.094447_7477_8547	7477	8547	+	Iron(III) dicitrate transport system permease protein FecD (TC 3.A.1.14.1)
NODE_10_length_140745_cov_22.094447_8544_9326	8544	9326	+	Iron(III) dicitrate transport ATP-binding protein FecE (TC 3.A.1.14.1)
NODE_10_length_140745_cov_22.094447_9488_10108	9488	10108	+	Phosphoadenylyl-sulfate reductase [thioredoxin] (EC 1.8.4.8)
NODE_10_length_140745_cov_22.094447_11158_10187	11158	10187	-	Cys regulon transcriptional activator CysB
NODE_10_length_140745_cov_22.094447_11417_11740	11417	11740	+	Thioredoxin
NODE_10_length_140745_cov_22.094447_12894_11812	12894	11812	-	protein of unknown function UPF0118
NODE_10_length_140745_cov_22.094447_14853_13042	14853	13042	-	Acyl-CoA dehydrogenase (EC 1.3.8.7)
NODE_10_length_140745_cov_22.094447_15612_15001	15612	15001	-	Transporter, LysE family
NODE_10_length_140745_cov_22.094447_15719_16627	15719	16627	+	Chromosome initiation inhibitor
NODE_10_length_140745_cov_22.094447_16954_16649	16954	16649	-	protein of unknown function DUF77
NODE_10_length_140745_cov_22.094447_17717_17085	17717	17085	-	Outer membrane lipoprotein carrier protein LolA
NODE_10_length_140745_cov_22.094447_21166_17909	21166	17909	-	Cell division protein FtsK
NODE_10_length_140745_cov_22.094447_21513_22283	21513	22283	+	Leucyl/phenylalanyl-tRNA--protein transferase (EC 2.3.2.6)
NODE_10_length_140745_cov_22.094447_22280_23023	22280	23023	+	Arginine-tRNA-protein transferase (EC 2.3.2.8)
NODE_10_length_140745_cov_22.094447_23208_23426	23208	23426	+	Translation initiation factor 1
NODE_10_length_140745_cov_22.094447_25804_23543	25804	23543	-	ATP-dependent Clp protease ATP-binding subunit ClpA
NODE_10_length_140745_cov_22.094447_26279_25917	26279	25917	-	ATP-dependent Clp protease adaptor protein ClpS
NODE_10_length_140745_cov_22.094447_26540_27139	26540	27139	+	Ribosomal large subunit pseudouridine synthase E (EC 4.2.1.70)
NODE_10_length_140745_cov_22.094447_27136_27585	27136	27585	+	Nudix-like NDP and NTP phosphohydrolase YmfB
NODE_10_length_140745_cov_22.094447_27671_28825	27671	28825	+	tRNA-specific 2-thiouridylase MnmA
NODE_10_length_140745_cov_22.094447_28822_29466	28822	29466	+	FIG002903: a protein of unknown function perhaps involved in purine metabolism
NODE_10_length_140745_cov_22.094447_29512_30876	29512	30876	+	Adenylosuccinate lyase (EC 4.3.2.2)
NODE_10_length_140745_cov_22.094447_31034_32257	31034	32257	+	FIG002776: hypothetical protein
NODE_10_length_140745_cov_22.094447_32259_32681	32259	32681	+	GNAT family acetyltransferase YjcF
NODE_10_length_140745_cov_22.094447_32968_34278	32968	34278	+	Isocitrate lyase (EC 4.1.3.1)
NODE_10_length_140745_cov_22.094447_35415_34351	35415	34351	-	Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11)
NODE_10_length_140745_cov_22.094447_36662_35523	36662	35523	-	3-isopropylmalate dehydrogenase (EC 1.1.1.85)
NODE_10_length_140745_cov_22.094447_37352_36705	37352	36705	-	3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
NODE_10_length_140745_cov_22.094447_38818_37355	38818	37355	-	3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)
NODE_10_length_140745_cov_22.094447_39022_39915	39022	39915	+	Chromosome initiation inhibitor

Se muestra 36 de 3402 genes anotados del Genoma de *C. salexigens* MP25462

NOMBRE DE ARCHIVO: 6666666.448892

Anexo 9: Genoma secuenciado de MP25462 (Región del gen aspartil aminopeptidasa).

>fig|6666666.448858.peg.717 Aspartyl aminopeptidase (EC 3.4.11.21) [Chrmohalobacter salexigens SpadesMP25462]

tcataatgatggggtcccctgttcgagtcagggcgtagccaccacttccttcctttcgcggttttcctccccggtgtcatttt
tcgtcaggcgtcgttccgcattcctctggattgcgtccctggatggcactcctcgactgtcttcgtcgtgcatgacgcatgtg
cagaccattgtttcgcaacgggattgttcgaggcttgcggcgtcttttccgggaaggaattcactttcaaccttgacgctggc
gatcgaatccgtatactacgccgcgtgctccggcatcgttccccgatagctcagttggtagagcaaatgactgttaatcattg
ggtcgcaggttcgagtcctgctcggggagccaaaccggatcgcgcactcggcgatccccgatagctcagttggtagagcaaat
gactgttaatcattgggtcgcaggttcgagtcctgctcggggagccaacttcccgccgagtcacatcacgaatgttccccgat
agctcagttggtagagcaaatgactgttaatcattgggtcgcaggttcgagtcctgctcggggagccaagtggtgccattggc
ctgcttgacatcgaccctccgtacgtctccttttctgctcgtcccttgcgtttgccctgtgtcgtttctgaagcgttctcgtgc
tttccagtgacgccagccagcgccgaatgtccgtgtctcgaaaaaagcgcgtcagctgcatggcacgtgtttcgccgatccct
gaatagtcacgccagtccgatggcgcgagagtgccgagtccggccaggccgtatcgctcgagggcatcggtcagtgcctgtcg
aggcacggcgggaacgcccagggcacgtaaccagcgcaccgtagagcgcgtgtcggcgagggcgaaatgccgctgccaggacc
gggcggtcttcttgcctacgcggggcaaggcttcaagggctgcggtatcgagcgtgcgccagtccagaagccccttgaccagg
cctgcatccaccagctgattccatgtcgcgggaccgatgccgggaaagtccagcccctcgtcgccgcccagccactcgagacg
cgcgaggaattgatcgcgacatgcgggcgtgaaccgcagacaggaaaggcgatcgtattggtctgcctcgggagcgtcgacct
gcggtcgcggccgagtacggaccacgacctcctgcagttgcgggatggtgagtccggccaggcgcaggcgtacttgatcgccg
gggcggacatccgtgcgtcgccagtgcgccagggatccgaggctgacgcgggtcacgtgcttgtcccccaggaggaccggttc
cagctcggcgacgaccgcgatgcgcccggtgcgaccgaccgagaattcgagtgcatcgacacgcgcgagcgcctcgcgcgcag
ggtacttccaagccattgcccagtccggcggcgaggcctgccagtcgcttcccggcggccggtctccgcgcttgacgacgacg
ccatcggtcgcgaagggcagggcatggcggtaccagcgttggcgccaggcggcgatatcttccgccgtggcgaccccatgcgt
catctcggcgacgtcgccgaatccccagccagccagcgtctccaggcgttcgggcatggagtcgggtccgttgggccacgccc
agacgaacaggccgaccgtgctgccttcacgatccgtcaaggtgtcgcgtgccatcaggccggcgatccgcgcgcgtgcaccg
tcggtgccgtcgcgcgattgccggtgtgccgtgcgtctggcgtagagctcgccttgcagtatgaccagcggcggggcatctgc
cggcaggtgcgagggaatcgccgcgatatgcagcgccttggtgtaccaatcctggccagagacgccatcgccgcggctgatgg
cggctgccagccggccatgccgatagaccagagtcacggcgacgccgtcgaccttgggttgcacccacagcgggtgatcgtcc
tgacgtgcgacccagtcggcaacggcgtcgcggtcgggcagtttggcgataccggtctgtgcgaccggatgatgcttcgcccc
cgtgggttcgtgcgtgttgacagatggcttcgcgagcgaggggaagcactccgcccagtgcgcgagccggcgtttggcgctgt
cgtaggtgccgtcgtcgaccaggcgctgtccctgtcggtagtaggcgcgatcccagcgtgcgatccgagcgtcgagggcctgg
tagtcgttctgcatggcctcgggcgtcatctgggggcagggtggcgcgctcgcccagccgggcatggcgagggtgctcaggcc
gaccagcagacacgctctcagccaggctgcggttcggagtggcgcgggacgcatgcctgaaccgtcaatccacgaaggtcgtc
gcgacatgtcgagacatccagttattaggtttttcttatttaactgaaatcgaggccccaggccaagcccgggaggggcctgt
tcaaaaaaatgccgccggcgagcgccggcggcaggagagtacgatgagacaggttctgaaataggttctgaaacaggttcggg
cgtccgcgtgctggcggccggaccgaaccaggcgggtcaggccgcgtcctgcaggtgggcataatccagcaggatctccgctg
cctcgacgacctgggcgcggtcgaccggttcgtcgtcttctgctgacgcgtcatccgactccggagcatcttccgccgaagac
gcctcggcgctttcgcgggcatcgatccactcgtccagcggttcgaggcccagcgcccgccggcgctggttttccagcgacag
ctgctcggcttcctgggcttccatctcgcgttggcgctgttcccggttcaggctgatgctggtctgctgttcctgaaggcgct
tggcgagcctggcttcacgctcgagatagacgaagttggggttctgctgaacgcgcgcctcgtgcttgtcacgcagggtggcg
agatagcgctggggcgtgccgtactgactgtattgcacggcgcgcacctgatcccagggcagggcgttgtcgagcgcgctttc
cccgatgtcgtccttgccgatcaggctggggaaagcgatgtcgggttcgacgccgcgcagttgggtgctttcaccggtgatac
ggtagaacttggcgcgggtcagcttgatctgaccatggctgagatcattgagcgtctgcacggtacccttgccgaaggtctgg
ctgccgacgaccaggccgcgaccgtaatcctggatcgcgccggcgaagatttccgaagcggaagccgatagccgattgaccag
gacgctgagcggcccgtcgtaatgcgtgccgctgtccgagtcgccatacaggttgatgcgaccacgggcgtcgcggacctgaa
ccgtggggccgcgatcgatgaacaggccgaccagggaattggcttcctgcaaggcgccgccgccgttgttgcgcagatccagc
acgatgccttcgacgttctccgccttgagcttgtcgatttccttggcgacgtcccgcgtgctgctgcggtagtcgctgtcgcc
ggcctgccaggcatcgaagtcgacgtagaaggtggggacgtcgatgacgccaatcttgtgctcctcgccgtcgcgggtcacgg
tcacgacctcgctgctggcggcctggtcctccagtttgacggtgtcgcgcgtgatgcgcacggtatgcgtacgcgtcacgtcc
acggccttggcgggaatgacgtcaaggcgcacggtggagcccttgggaccgcggatcaggtcgacaacttcgtcgaggcgcat
gcccacgacattgaccatttcgccgtcctcgccttgaccgaccgcgacgatacggtccgcgggctgcagggcgtcgctcttgt
ccgcggggccgcccggcaccaggctggaaaccttgacgtactcgccttcgctttgcagcagggccccgatgccctccagcgag
agcttcatctggatgtcgaacgattcgctctggctcggagacagatactcggtgtgcggatcgatggtgccggtgacggcgct
caggaacagcaccagcacgtcttcggcattggtctgctcgacgcgcgacaactgattttcataccgctcgcgcagtgtctcgg
cgatgtccgtatcgtcttcgtcgtccggggcgctcagggactgggtcagggcggcgttcttcaggcgcttttcccatagccga
tcgagcgcgtcctggtgcttggcccaggcttcgttctcgcggtccaacgccagtcgctcgttgccttgatagtcatagtcgat
gccgtcttccagacggtcgatcaaccattcgaggcgggcttccaggcgctcctgataacgggcgtaaagttcgtagacgcgct
ccagttcaccgttgtcgagcgcttcgtcgaagcgggtctcgaggtcctgaaagtcgtcgacgtcccgcgacaacaaataggcg
cgttgtgggtccagtacatcgagataacgctgaaacgctctcgacgaccactcatcatccaggttgatatcggcgtaatgtcc
atatcgcagcgattcagcgacttccgccgctgcttgccgctgttcgtcggtaggtgtcgggccctcggccaaggccggggtac
acaccaccagcaacagtgacaagatcgccgcgcgccgagcggctgcaaacaggctcatcaatcaagcactcccggtttcatgt
acactctcgttcgttcctagggcctgttgacgtttcatcgcgaagggccgggccagtttttgcacggtgctgcgttattcgtc
gttcatttggaatgaccaaactacgctcctcacgccttgccccgcacaaaacatggctccggcgcggccgtgcatgaaatgtc
aacaggccctagattgccggctctttgagttgagttttctcagaccggactggcattgtagcagtctgtatcatccgtcacag
acacttctccgaggaacgccATGGCTCACGCCCCCACCCTTGACCGTTTACTGCATTTTCTCGAGCGCTCGCCCACACCCTGG
CATGCCGTCGACAACATGGCCCGGCGGCTCGAACAGGCAGGGTATCGGCGACTCGAGGAAACCGAAGCGTGGCAATTGGCGCC
CGGCGATCGTTTCTACGTCACGCGCAACGCGTCGTCATTGATCGCCATGCAGGTGCCGACGGATCCCTTGAGCGGGCTGCGCA

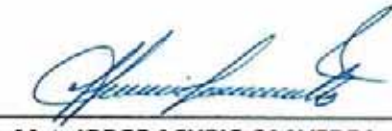
TGATCGGGGCGCATACCGACAGCCCGGGGTTGCGCTTGAAACCCCAACCCGTGGTGGCCAAGAAGGATTGGCTGCAGTTGAGC
GTCGAGGTCTACGGCGGTGCGCTGCTGGCACCATGGTTCGACCGCGATCTGGGGCTGGCCGGGCGCATCCATGTACGACGCGA
GGATGGACGCTTGCAGGGCGTATTATTGCATGTCGATCGTCCCGTCGCGATCATTCCCAGCCTGGCCATCCACCTGGATCGCG
AGGCCAACAACGGGCGTGCCCTGAATGCCCAGACGCAGATGCTGCCGGTCGTGCTGCAAGGCGGTGGCGAAGCCGATCTCGAG
CGTTGGCTCAAGCGCTGGCTGTACGAACAGCATGGGCTGGAGAACATTCAGTTGTTGGATTACGAACTCTCGCTTTACGACAT
GCAGCGGCCGTCGCGTGTCGGGATCGAGGGGGAACTGATCGCCAGTGCGCGCCTCGACAATCTGCTGTCGTGCTTCACCGGTA
TCGAGGCCTTGCTGGCCGGCGACGGGCGACAGGGGGCGCTCTTTGTCGCCAACGATCACGAAGAAGTGGGCAGTGCCAGTGCG
TGCGGCGCCCAGGGCCCCTTCCTGGGAGATGTGCTGCGTCGCGTGCATGCGCAACTGGGTGAGGGCGGCGAAGACGGCTGGGT
GCGTCTGATCCAGGGCTCGCGCATGATTTCCTGCGACAACGCCCATGCCGTGCACCCCAACTTTCCCGAGAAACACGACGAAC
ACCACGGCCCGGCGATCAATGGCGGGCCCGTGATCAAGGTGAACGCCAACCAGCGCTATGCCACCAACAGCGCGACAGCGGCC
ATGTTCCGGGATATCTGTCGCGAGGCAGGAACACCCGTGCAGACCTTCGTGACGCGTGCCGACATGGGCTGCGGCAGCACCAT
CGGGCCCATCACCGCCACCGAACTCGGGGTGCCGACGCTGGATGTGGGTATCCCGCAGTGGGGCATGCACTCGATTCGCGAAA
CCGCCGGCAGCCGCGATGCCGATTACTTGATCCGCGCGCTGACGGCCTTCGTCAATCGCACCGAGCTGGACTAGccgaagaga
gcagcgcccggccgaggacgggcgcccggggcagtcttgccgggccttccagatgaaaaaagcccgtcgacacaggtgtcgac
gggctttttcgtatcttggtggctacgccctgactcgaacaggggaccccatcattatgagtgatgtgctctaaccaactgag
ctacgtagccgttcaatgcgggcgtatactacgcataactccctgatgtgtcaactgtatcgtcgacacgaacgcatgatccg
ccgcgtggcgtcacctagacgttgaaacggaagtgcacgacatcgccgtcggcgaccacgtaatccttgccttccagacgcca
tttgcccgcgtccttggcgccttgctcgccccccagggccacgaagtcttcgtagccgaccacctcggcacgaatgaatccct
tctggaagtcggtgtggatcgcgccggcggcttccggggcggtggcaccttgcttgacggtccaggcgcggacttccttcacg
ccggcggtgaagtaggtctgcaggccgagcaggtcatagccggcgcggatgacgcggtccaggccgggctcttccatgcccat
ctcctcgaggaagacgctgcgttcttcttcgtcgagctcggcgatttcggcctcgatctggttgcacaccgggaccaccacgg
cgccttcctcgcgcgcgatgtcattgaccacgtcgaggtaggggttgttgtcgaacccgtcttcgttgacgttggcgatgtac
atggtgggcttcagggtcaagaagccgaagctcttgagctggcgcgtctcgtccgcgtcgaggccgaagctgcgcagcggctg
gccctcggcgaggtgcggctggatgcggtcgagaatggccttggtggccatggcgtccttgtcgccgcccttgaccacgcgcg
ccagacgctggatggcgcgttccacggtatccaggtcggccagggcaagctccagattgatggtctcgatatcggcgcgcgga
tcgacctgattggcgacgtggatgacgttgtcattgtcgaagcagcgcacgacatgcgcgatcgcctgggtttcgcggatatt
ggccaggaactggttgcccaggccttccccccttggatgccccggcgaccaggccggcgatatccacgaactccatggtcgtcg
ggatcacctttttccggcttgacgatcgccgcgagcttgtccagacgcgggtccggcatgggcacgatgcccacgttgggttcg
atggtgcagaaaggaaagttttcggcgtcgattcccgatttggtcaaggcattgaacagcgtggatttgccgacattgggaag
gccgacgataccgcaattgaaacccatgggatcttcctgaatcgtatgcgagaaggcgtgatgaacgtggcctgcattctacc
cggcagacgggcgttgggccatgggcgagacgaggtcgccgcgaaaggcgggtgcgatcaggcgttgaagctgtgcaagcggt
tcatggcgcctgcccagttaccgtcgagaacgtcggggagggtccgcaggcattcgtcgatggcggcatcgatggcggcgcgt
tccgccttgccggggcgtcccagaacgtaattgaccacttgcttggccgagccgggatggccgatgccgatgcgcaagcgatg
gaaggacttgtcgttgcccagcgcgctgacggtatcgcgcaggccgttgtgaccgccgtggccaccgccgtgcttgtagcgtg
ccgtgccgggagcgatatccagctcgtcgtgggcgatgagcagttcgtcgggcgtgagcttgaagaactgacaaagcgcggcg
acagccttgccgctgcggttcatgtaggtgctgggcaccagcaagtgaacgtcgtgaccggcgtagcgaaccttggcgtactg
accaaggaacttgcgctcggcacgcagcggcgtcaggctgtcccgcgccagggcctcgacgacccaggcgcccgcattgtgtc
gcgtggcatcatattcggtgcctggattgcccaggccgataatcgctttgacgtcactcatgcgggcacctgggaagaccgat
tactcggacttctcgccgtcttcgtcggtggcttcttcgccgccttcctcgtcaccctcggtgccggtgtcggcacgggtcac
gctgaccacgccgaagtcgtgttcctcgccatggctcaaggccacgagggaaacgccttccggcagcggcaggtcggacaggt
gcagcgtctcgcccagacccaggttggtgacgtcgacttcgaggtattcgggcagcttgccgggcaggcagctgatctcgacg
tcgttggccagcacgtgcagcacgccgtcttcatccttgacgcccttgcaggcttcttcaccgttgatgtgaagcgggacgtt
catggtgatttcgtgcgtggcatcgacgcgcatgaagtcggcatgcgtcaccagcgccttgtacgggtggcgctggatgtcgc
gaatgatgacctgttccgactggccttcgaggtcgatgttgatcaccgacgagaagaaggcttcatcctcgatggccttgtag
aaggcggacttttcgagggtgatcggttgcggcgtcttgctgccaccgtagatgatggcggcgactttcttgttctcacgacg
caggcggcggctcgcacctttccccaggtcgtgacgaacctcggcggatagagtgaattgggacattgattgtcacctcatag
gttatgtaaggaaatcggcaccgcccgcgaccaggcgatgccgtttccgtgagggcctcgcggcccgcgtatccgtgcatgtt
cccggtgacgcatcactcttttcagtgatgcatcgcctcagtggaacatggcactgacggattcttcgttgctgacgcggcgg
atggcctccgcgatcagaccggcgaccgagagctggcgaatgcgaccgctgcgggacgccgtctcggaaagcggaatggtatc
ggccacgacgagttcgtcaagtaccgaaccgacgatattgtcgacggcggggcccgagagaatggggtgcgtggcgtaggcca
cgacacgccgggcaccatgacccttgagcgcttcggcggccttgcacagcgtgccggcggtatcgaccatgtcgtcgaccacc
acgcaggtgcggtcctggatgtcgccgatgatgtgcatgacctgggcctggttggcctgcgggcgacgcttgtcgatgatcgc
caggtcggcgttgagctgcttggcaatggcgcgggcacgcaccacgccgccgacgtcgggggagaccaccaccagatcgtcgt
aattctggcgctcgatgtcgtcgagcaggatcggcgatccatagacgttgtcgacgggcacgtcgaaaaagccctgaatctga
tcggcgtggaggtccatggtcatgacgcggtcgacgccggccttgaccatcatgtcggcgaccagcttggccgagatcgggac
gcgcgcggagcgcacgcggcggtcctggcgagcatagccgaaatacggcacgacggcggtgattcgcgtggcggaggcacgac
gcagtgcatccaccatgagaatcagttccatcaagttgtcgttggtcggtgcgcaggtggactgcagtacgaagacgtccttg
ccgcgaacgttttcgttgatctcgaccgcgatttcaccatcgctgaattgaccgacggttgcatggcccagccggttgtcgag
gctttcggccaccttgcgggcgagttcgggattggcatttcccgcgaacaccatcaatttagacacgcgcagccacctttgga
gtcttggaggatctcggagtggaattcgtcgggcgccctgaagaattggctggggtaccaggatttgaacctgggaatgccgg
taccagaaaccggtgccttaaccacttggctataccccagcaaccatcgactcgtttcgttatccacatagaacacgatgccg
atttcaggcgtccccgtattggctcatgggtgaccattcgtcgtgaatgcccagagcgtcatggaggggagaagtgttgcagc
cgcgtgccttgaacgcatgccagtggctgccaacacggcgcaaaatcctatcggcttgccgttcgctcgtcaagggacagaaa
aggcaactgcccgttcccgtgagcatggccggaccgaaggccgataaccagtcgagcgcatgtgcgacgtccggcgagaggcg
tcgcactacggcttcgcagtcattgcgccaggcgcgcccctgttcggctccccctcgcagtgcgcgcgccatactaatcggcg
gggtgtcgcgtgtcaattcggggtgcccgaacacggcgggcgtggcgatttcctcgcccgggtgtatgaccacgaaccagggg
gtgtcgagcgtcaccggcgtaaggcgttcgccgataccttcggcccaggcgctgtgcccgcgcacgaagaccggcacgtcggc
gccgagcgtgagacccagctcggcgagacggggcagccccaggtcgagcgaccacaagcggtcgagcgcgagcagggtggtgg

ccgcgtcggagctgccgccgcccaggccgcctcccatggggagccgtttgtcgaggtgaatgtccacgccctgatgcgtgccg
cttgcgtgctgcagcaggcgggcggcgcgcacgatgaggttggcgtcgtgatcgacgcccgcgatggccggggccaggtgaat
ggcgccgtcagcgcgggggctgaagtgcagcgtgtcgctccggtcgagaaactggaacagtgtctgcagctcgtggtagccgt
cggcacggcggcccacgatatgcagcatgcgattgagcttggcgggcgccggcaggctgaggcgttgcatcatgtatcgctca
gcagctgccactggttggcgaccagagtggcgcgcagctcgccgtaggtcatgaccagccggcgcggcatccacagcgtgtcg
gcgcgtgtcca

Se muestra el gen aspartil aminopeptidasa resaltado en color magenta contenido en el genoma de MP25462 flanqueado por 5000 nucleótidos (El genoma completo de *Chromohalobacter salexigens* MP25462 es de una total de 3712216 pb).

Mgt. JULIA GRISELDA MUÑIZ DURÁN
PRESIDENTA DEL JURADO-
PRIMER REPLICANTE

Blga. OLGA CJUNO HUANCA
SEGUNDO REPLICANTE



Mgt. JORGE ACURIO SAAVEDRA
PRIMER DICTAMINANTE

Mgt. YOLANDA CALLO CHOQEVILCA
SEGUNDO DICTAMINANTE