



# UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA



## FACULTAD DE CIENCIAS DE LA SALUD CIRUJANO DENTISTA

Tesis

### Aislamiento de aptámeros contra la proteína S mutante del SARS-CoV-2

Que para obtener el título de  
**Cirujano Dentista**

Presenta

**Sandra Alexia Álvarez Meza**

Director(a) de tesis

**Dr. Giovanni Palomino Vizcaino**

Codirector(a) de tesis

**Dra. Lucía Margarita Valenzuela Salas**



# **Aislamiento de aptámeros contra la proteína S mutante del SARS-CoV-2**

El presente trabajo de investigación fue realizado en la **Facultad de Ciencias de la Salud, unidad Valle de las Palmas,** Campus Tijuana de la Universidad Autónoma de Baja California.

Bajo la dirección del  
**Dr. Giovanni Palomino Vizcaino**

Bajo la codirección de  
**Dra. Lucía Margarita Valenzuela Salas**



*Para mi familia*



*“Admira a quien lo intenta, aunque fracase”*

*-Séneca*



# Agradecimientos por los fondos

*“Solo aquel que ha superado sus miedos, es verdaderamente libre”*

*-Aristóteles*

La presente Tesis de Licenciatura se desarrolló en la Facultad de Ciencias de la Salud de la unidad Valle de las Palmas, campus Tijuana de la Universidad Autónoma de Baja California.

Durante toda mi estancia en el Laboratorio del Dr Giovanni Palomino Vizcaino y Sandra Alexia Álvarez Meza pude realizar el presente trabajo gracias a los fondos recibidos de la convocatoria de Apoyo a la incorporación de NPTC PRODEP 2020 (Proyecto UABC-PTC-852 con clave 350/745/E) y Interna de Proyectos de Investigación 2024-2025 (Proyecto con clave 350/2/C/7/24).

Adicionalmente recibí apoyo de la Responsable del Programa Educativo Cirujano Dentista MPA Karina Rodríguez Fuentes, de la Coordinadora de Investigación y Posgrado MC Ofelia Candolfi Arballo y de manera muy especial agradezco a DCS Ana Gabriela Magallanes Rodríguez Directora de la Facultad de Ciencias de la Salud por la oportunidad de realizar mi Servicio Social Profesional en una Plaza Universitaria de Investigación.

Como parte de mi formación científica, asistí al 5to Encuentro de Jóvenes Investigadores que se llevó de manera virtual por la Coordinación General de Investigación y Posgrado en el área de: Medicina y salud, que me permitió poder exponer parte de mi proyecto a otros jóvenes y académicos ayudándome expresar a través de un póster los puntos más importantes e interesantes dentro de mi investigación para así inspirar a otros estudiantes querer pertenecer a este tipo de trabajos dentro de la institución.



# Agradecimientos

*“Lo que con mucho trabajo se adquiere, más se ama”*

*-Aristóteles*

A mi padre por ser la base de mis valores, el motor de mi educación, por hacerme creer que puedo alcanzar con esfuerzo mis metas, por darme todo y mucho más en vida.

A mi madre por dar todo lo que poseía para verme triunfar, por su fe en mí, por permitirme elegir lo que desde pequeña soñé pues se que es un privilegio.

A mi familia por estar orgullosos de mi, son mi motivación número uno, y la razón por quienes deseo triunfar y crecer.

A mi director de tesis pues fue un apoyo fundamental en la elaboración de esta tesis, siempre estuvo dispuesto a ayudarme, a explicarme, a guiarme con mucha paciencia, una actitud positiva, amable y una gran sonrisa así como también motivarnos a llegar más lejos de nuestras expectativas y que el miedo no sea un obstáculo.

A mi novio por ser mi hombro, por su apoyo en todos los sentidos, por darme siempre mucho más de lo que posee, y ayudarme con un escalón más .



# Tabla de Contenido

<b>Introducción</b>	<b>14</b>
1. Coronavirus	14
1.1 SARS-CoV-2	15
1.2 Proteína S	16
1.3 COVID-19	17
2. Aptámeros	18
<b>Justificación</b>	<b>21</b>
<b>Hipótesis</b>	<b>23</b>
<b>Objetivos</b>	<b>25</b>
1. General	25
2. Específicos	25
<b>Materiales y métodos</b>	<b>27</b>
9.1. Oligonucleótidos sintéticos.	27
1. Reactivos biológicos	28
1.1 Enzimas y otros reactivos biológicos	28
1.2. Análisis in silico de la biblioteca “libUABC”	28
1.2.1. Oligonucleótidos	28
1.4 Métodos bioinformáticos	29
1.4.1 Análisis de calidad de datos NGS por FastQC	29
1.4.1.1 Procesamiento bioinformático	29
1.4.1.2 Presentación de resultados	29
1.4.1.2.1 Resumen de todos los datos.	29
1.4.1.2.2 Calidad de secuencia por base.	29
1.4.1.2.3 Calidad de secuencia por mosaico.	30
1.4.1.2.4 Puntuación de calidad por secuencia.	30
1.4.1.2.5 Porcentaje de base por posición en la molécula.	30
1.4.1.2.6 Contenido promedio de GC en todas las moléculas.	30
1.4.1.2.7 Distribución de longitud por secuencia.	31
1.4.1.2.8 Niveles de duplicación de secuencia.	31
1.4.1.2.9 Contenido de adaptador.	31
1.4.2 Eliminación de los adaptadores	31
1.4.2 Eliminación de los iniciadores	32
1.4.2 Rasurado de las secuencias	33
<b>Resultados</b>	<b>35</b>
	12

6.1 Diseño de la biblioteca libUABC para el aislamiento de aptámeros.	35
6.1.1. Diseño de la biblioteca libUABC.	35
7.1. Secuenciación de Ronda 0 (libUABC)	39
7.1.1 Análisis 1 de la Ronda 0	39
7.1.2 Análisis 2 de la Ronda 0	47
7.2. Secuenciación de Ronda 1	53
7.2.1 Análisis 1 de la Ronda 1	53
7.2.2 Análisis 2 de la Ronda 1	59
7.3. Secuenciación de Ronda 2	66
7.3.1 Análisis 1 de la Ronda 2	66
7.3.2 Análisis 2 de la Ronda 2	73
<b>Discusión</b>	<b>82</b>
<b>Conclusiones</b>	<b>85</b>
<b>Perspectivas</b>	<b>87</b>
<b>Referencias Bibliográficas</b>	<b>88</b>

## Capítulo 1

# Introducción

*“Si buscas resultados distintos, no hagas lo mismo siempre”*

*-Albert Einstein*

## 1. Coronavirus

Los coronavirus son virus que pertenecen a la familia *coronaviridae* se caracterizan por tener una alta tasa de recombinación y mutación genética, teniendo por consecuencia que puedan infectar desde aves hasta mamíferos(1). Estos virus pertenecen a la subfamilia Orthocoronavirinae y de acuerdo a las variaciones de la secuencia de su genoma y las pruebas serológicas se dividen en cuatro géneros: *Alfacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* y *Deltacoronavirus*, de ellos los virus clasificados como *Deltacoronavirus* son los responsables por los síndromes respiratorios humanos(2).

El síndrome respiratorio agudo severo (SARS) fue la primera nueva enfermedad infecciosa identificada en el siglo veinte, la cual se originó en Guangdong provincia de China en noviembre de 2002, esta enfermedad a menudo presenta un cuadro clínico agudo y severo(3). Acciones globales coordinadas por la Organización Mundial de la Salud (OMS), conllevaron a la identificación del agente causal del brote de abril 2023, un nuevo coronavirus denominado SARS-coronavirus (SARS-CoV)(4).

Los coronavirus son virus grandes, envueltos, con un genoma de RNA de cadena sencilla y positiva, de importancia médica y veterinaria. Los coronavirus son capaces de infectar mamíferos y aves, causando principalmente enfermedades neurológicas o hepatitis. Se transmiten principalmente por vía respiratoria y/o rutas fecal-oral(5).

Una de las características más distintivas de esta familia es que poseen un genoma no segmentado, de ARN de cadena sencilla positivo y es el genoma de RNA más grande de todos los virus de RNA incluso de aquellos con un genoma fragmentado(6).

Los coronavirus son esféricos, poseen una dimensión de entre 80-120 nm, en la envoltura se presentan tres proteínas de las cuales la glicoproteína S (llamada E2) es la más prominente, es la encargada de mediar el anclaje al receptor celular y la fusión de las membranas. La proteína S es transmembranal, con un extremo amino terminal extramembranal y un extremo carboxilo terminal en el interior de la célula (citoplasma) las espículas de la superficie, poseen una base delgada que se ensancha hasta un ancho de 10 nm en el extremo distal(7).

## 1.1 SARS-CoV-2

El Virus del síndrome respiratorio agudo severo tipo 2 por sus siglas en inglés SARS-CoV-2, es un virus que inicialmente comenzó como un brote de etiología desconocida en la región de Wuhan China a finales del 2019, dicho virus produce una enfermedad con el nombre de COVID-19 (*Novel Coronavirus Infectious Disease 2019*) responsable de una pandemia mundial declarada el 30 de Marzo del 2020 por la Organización mundial de la salud(8).

El mecanismo de transmisión del virus es a través de secreciones provenientes de personas infectadas, las cuales expulsan gotas de saliva con un tamaño mayor a 5 micras, que contaminan superficies u objetos inanimados con estas secreciones y que sean llevadas directamente a mucosas, conjuntiva ocular y/o nasal por un sujeto no infectado(9).

El virus posee un genoma conformado por RNA de cadena sencilla, polaridad positiva y una medida aproximada de 30 kb. El virión se asemeja a una corona que en realidad es una envoltura lipídica con las siguientes cuatro proteínas: E (envoltura), M (membrana), N (nucleocápside) y S (espícula), esta última es la que permite unirse al receptor la proteína convertidora de angiotensina humana 2 (hACE2) de la membrana celular del huésped, dentro del virus también se encuentran proteínas no estructurales conocidas como nsp1 a nsp16 necesarias para la transcripción y replicación(1,10).

## 1.2 Proteína S

Es la proteína encargada de conferir la apariencia de corona solar al virus, con un peso molecular de 180 kDa, se divide a través de una enzima de nombre proteasa en dos subunidades denominadas subunidad S1 y subunidad S2, las cuales tienen el papel de llevar a cabo la unión de la membrana y las células huésped(4).

La subunidad S1 posee un dominio de unión al receptor (RBD) y un dominio (N-terminal) amino-terminal (NTD). El motivo de unión del receptor (RBM) toma una forma parecida a una cuna, la cual ayuda la unión al receptor. El RBD es capaz de adquirir estructuras diferentes como lo son el estado cerrado 'abajo' y el estado abierto 'arriba'. El estado abajo protege las regiones de unión al receptor y el estado arriba adopta una forma de bisagra que expone las regiones determinantes para permitir el reconocimiento de ACE2 en la membrana de la célula del huésped ayudándose de 14 residuos de aminoácidos(7).

La subunidad S2 engloba 4 regiones; 1 péptido de fusión, 2 repeticiones de heptada (HR1 y HR2) y una región transmembrana, las cuales conceden la fusión entre la membrana celular y viral(1,11).

El primer caso de COVID 19 confirmado en Estados Unidos fue el 31 de enero del 2020, el primer caso confirmado en Canadá fue el 25 de enero del 2020, en México se reportaron los primeros casos confirmados el 28 de febrero del 2020, un caso provenía de la Ciudad de México y otro procedía del estado de Sinaloa(12).

La distribución de la enfermedad COVID 19 a través de México durante los primeros 5 meses después del primer caso confirmado era de 32,522 casos confirmados y 3160 defunciones de tal modo los siguientes estados eran los que presentaban una mayor prevalencia a comparación de los demás estados del país, en Ciudad de México se reportaron 8,705 confirmados y 729 defunciones. En el Estado de México se confirmaron 5,418 confirmados y 300 defunciones. En Baja California se confirmaron 2276 confirmados y 365 defunciones. En Tabasco se confirmaron 1,531 confirmados y 201 defunciones. En Sinaloa se confirmaron 1,372 confirmados y 204 defunciones. En Veracruz se confirmaron 1,049 y 112 defunciones(12).

### 1.3 COVID-19

La enfermedad del coronavirus (*Novel Coronavirus Infectious Disease 2019*; COVID-19) provocada por el SARS-CoV-2 presenta manifestaciones clínicas variadas dependiendo del grado de severidad, ya sea, leve, moderado o severo, o completamente asintomático. Los síntomas más frecuentes incluyen fatiga, tos seca, fiebre, mialgias, estornudos, y entre los menos comunes que afectan sistémicamente están la cefalea, rinorrea, dolor abdominal, diarrea, náuseas, insuficiencia renal, trastornos cardiocirculatorios(13).

El tratamiento de COVID-19 se centra en mitigar los síntomas del paciente, ya que no existe un fármaco específico, aunque se llegaron a emplear algunos, como por ejemplo medicamentos antimaláricos como la cloroquina e hidroxiclороquina, que poseen como mecanismo de acción el bloqueo del receptor de la enzima convertidora de angiotensina 2 (ECA2) y reducen la infección viral a través del aumento de pH en la endosoma. También se emplearon otros fármacos como lopinavir/ritonavir, remdesivir, favipiravir, azitromicina, interferón, ivermectina, nitazoxanida, doxiciclina, tocilizumab, sarilumab, siltuximab, bevacizumab, ozono, células madre, óxido nítrico, etc(8,14).

Uno de los profesionales de la salud más expuestos fueron los odontólogos, puesto que mantienen un contacto directo con las mucosas de la cavidad oral y tracto respiratorio dando como resultado una alta exposición ante la enfermedad dentro del

consultorio dental pues los procedimientos dentales conlleva una gran contaminación debido a la intrínseca generación de aerosoles a partir de la saliva del paciente(15).

## 2. Aptámeros

Los aptámeros son oligonucleótidos, pueden ser ADN o ARN de cadena sencilla con la capacidad de plegarse a una estructura tridimensional única que le confiere alta especificidad y afinidad por un blanco(16). Los aptámeros también conocidos como anticuerpos químicos, poseen constantes de afinidad que rivalizan con los de anticuerpos y en algunos casos son mayores(17). Los aptámeros poseen la capacidad de unirse a su blanco con alta estabilidad, la interacción aptámero-blanco es capaz de resistir altas temperaturas. Los aptámeros además tienen bajo peso molecular, baja toxicidad, pueden ser producidos por síntesis química (sin la necesidad de animales) y a un menor costo que los anticuerpos, lo que los convierte en una mejor alternativa que los anticuerpos tanto para ser incorporado en dispositivos de diagnóstico y como agentes terapéuticos(18).

Los aptámeros pueden ser empleados como agentes diagnósticos o terapéuticos, pueden ser incorporados en biosensores para la detección de sus blancos o como agentes de direccionamiento a ciertas partes del organismo(19).

Los aptámeros se obtienen a través de un proceso de selección in vitro denominado evolución sistemática de ligandos por enriquecimiento exponencial (Por sus siglas en inglés, SELEX) y para realizar este procedimiento es necesario una biblioteca de ADN o ARN de cierta longitud, estructurada por una región aleatorizada flanqueada por dos regiones fijas en los extremos. En el centro de la molécula la secuencia se desconoce con precisión, pero permite identificar que en esa región de la biblioteca todas las moléculas son diversas (diferentes), uno de los conceptos necesarios para que exista evolución. Una vez que el procedimiento ha finalizado los oligonucleótidos restantes que no se adhirieron son eliminados, mientras que los que sí se unieron con la molécula objetivo pueden considerarse aptámeros. Para que los aptámeros puedan tener una especificidad y afinidad confiable es necesario que estos

pasen por diferentes ciclos de rondas SELEX, esto con el objetivo de que a través de este proceso se descarten aptámeros con afinidades bajas, es decir de débil unión molecular, para que de esta manera se conserven los mejores candidatos(20).

El desarrollo de la secuenciación de próxima generación (Next Generation Sequencing; NGS) han potenciado el aislamiento de nuevos aptámeros, ya que ha disminuido el tiempo de identificación de aptámeros funcionales(21).

Los aptámeros presentan ventajas sobre los anticuerpos una de las que destaca es que para su producción no se requieren modelos biológicos animales como para los anticuerpos, su tiempo de producción varía de 1 a 3 meses, se pueden modificar químicamente de manera fácil, presentan un amplio campo de adhesión entre blancos inmunogénicos y no inmunogénicos, y su producción es de bajo costo. Algunas desventajas de estas moléculas son que al ser de reciente descubrimiento el grado de madurez de la tecnología y la información disponible de estudios, análisis y experimentos es reducida, el tiempo de vida varía de corta a media, tienden a degradarse por la acción de nucleasas en suero cuando no se han modificado químicamente(22).

Un claro ejemplo dentro de sus aplicaciones se encontrará al anti-VEGF conocido como pegaptanib - Macugen® el cual se destaca por ser el primer aptámero aprobado por la FDA (Food and Drug Administration), cuyo uso se aplica en el tratamiento de la degeneración macular asociada a la edad (DMAE) la cual es una enfermedad de carácter oftalmológica que afecta en forma de deterioro a la mácula, tejido que se encuentra en el centro de la retina. Se caracteriza como la primera causa de ceguera irreversible que manifiesta sus afecciones por encima de la sexta década de vida, entre las cuales se encontró a las drusas, que son depositos blanco amarillos, que interfieren con la interfaz coronoide-retina dañando la irrigación de los fotorreceptores(23).

Entre otras de sus aplicaciones podemos mencionar al aptámero Anti-IgE desarrollado por Yao y colaboradores capaz de identificar a la IgE y útil para las personas con cuadros alérgicos elevados, pues se destaca que el problema radica en un elevado contenido IgE, dicho aptámero es una mejor alternativa pues los estudios requeridos para determinar la presencia de IgE (ELISA o radioinmunoanálisis) suelen

ser muy costosos, dejando así como al aptámero IgE como una mejor, eficiente y más rápida opción pues el sensor basado en aptámeros determina el contenido de IgE en 15 minutos(24).



## Capítulo 2

# Justificación

*“En la vida no existe nada que temer, solo cosas por comprender”*

*–Marie Curie*

El COVID19 es un problema de salud a nivel mundial, esto debido a su alta tasa de contagio y a su difícil control en la población. Según datos de la Dirección General de Epidemiología, en México se han registrado 7,633,355 casos confirmados y 334,336 defunciones desde febrero del 2020 a la fecha. Además del daño en la salud pública, esta crisis sanitaria provocó grandes pérdidas económicas a nivel global, sobre todo en los países en vías de desarrollo. Una de las estrategias más importantes para controlar la pandemia sigue siendo el diagnóstico temprano, el cual consiste en la detección del material genético mediante reacción en cadena de la polimerasa (PCR y RT-PCR), sin embargo, es necesario disponer de otras herramientas que igualmente tengan una alta sensibilidad y precisión, pero que sean más fácil de manejar, más versátiles, más rápidas y sobretodo, más económicas. Al igual que el diagnóstico, las estrategias terapéuticas oportunas han logrado disminuir la mortalidad de esta enfermedad, por lo que el contar con nuevas estrategias de inhibición viral, igualmente resulta de vital importancia. En ambos casos, el diseño de biosensores como son los aptámeros, puede aportar a estas áreas ya que son una alternativa innovadora y vanguardista que puede utilizarse tanto para desarrollo de nuevos terapéuticos, como para el diseño de herramientas diagnósticas.



## Capítulo 3

# Hipótesis

*Nuestra mente se enriquece con lo que recibimos, nuestro corazón con lo que damos*

*-Victor Hugo*

Es posible aislar al menos un aptámero capaz de reconocer selectivamente a la proteína S mutante del SARS-CoV-2 *in vitro*.



## Capítulo 4

# Objetivos

*“El aprendizaje es lo único que la mente nunca agota, nunca teme y nunca lamenta*

*-Leonardo Da Vinci*

### 1. General

Analizar los resultados de secuenciación masiva de un proceso SELEX contra la proteína S del SARS-CoV-2 *in vitro*.

### 2. Específicos

1. Diseño y síntesis de una biblioteca de DNA con 40 posiciones aleatorizadas para el aislamiento de aptámeros por SELEX.
2. Analizar la frecuencia de las las secuencias enriquecidas en el proceso SELEX contra la proteína viral S.



## Capítulo 5

# Materiales y métodos

*“No hay grandeza, donde faltan la sencillez, la bondad y la verdad,”*

*-León Tolstói*

### 9.1. Oligonucleótidos sintéticos.

Todos los oligonucleótidos usados fueron sintetizados por T4OLIGO (Irapuato, México)

#### 9.1.1 Biblioteca.

Una biblioteca combinatoria de DNA de cadena sencilla bajo el nombre de libUABC fue sintetizada con la siguiente secuencia GACCTGACAGACACCTTGG NNGCCATCACACATCA TCCTG. El diseño consiste en un oligonucleótido de ADN de 80 bases que consta de tres regiones: dos regiones fijas de 19 bases de nucleótidos y una región central de 42 posiciones aleatorizadas, en donde la letra “N” puede representar a cualquiera de 4 bases nitrogenadas del ADN en una proporción equimolar.

### 9.1.2 Iniciadores

Para amplificar y anclar la biblioteca en cada ciclo se utilizaron los siguientes oligonucleótidos: Forward con la secuencia GACCTGACAGACACCTTGG y reverse con la secuencia CAGGATGATGTGTGATGGC, ambos de 19 nucleótidos.

## 1. Reactivos biológicos

### 1.1 Enzimas y otros reactivos biológicos

Para la reacción de cadena de la polimerasa (PCR), se utilizó la enzima Taq DNA polimerasa Amplificasa® (Biotecnologías Moleculares S.A. de C.V.), la mezcla de desoxirribonucleótidos trifosfatados (dNTPs) y los marcadores de peso molecular GeneRuler™ 50bp DNA Ladder, PageRuler™ Plus Prestained Protein Ladder se adquirieron de Thermo Fisher Scientific Inc.

Las proteínas virales, SARS-CoV-2 (2019-nCoV) Spike S1(HV69-70 deletion, Y144 deletion, N501Y, A570D, D614G, P681H)-His Recombinant Protein y SARS-CoV-2 (2019-nCoV) Spike S1(K417N, E484K, N501Y, D614G)-His Recombinant Protein se adquirieron de la empresa SinoBiological (Houston, TX, USA)

### 1.2. Análisis *in silico* de la biblioteca “libUABC”

Con el propósito de analizar la formación potencial de estructuras intramoleculares e intermoleculares que pudiesen dificultar la amplificación por PCR de la biblioteca, se empleó la herramienta bioinformática <https://www.idtdna.com/pages/tools/oligoanalyzer> disponible en línea para calcular la estabilidad de homodímeros y heterodímeros(25).

#### 1.2.1. Oligonucleótidos

Con el propósito de evaluar los dímeros y heterodímeros se usó la herramienta bioinformática oligoanalyzer(25).

## 1.4 Métodos bioinformáticos

### 1.4.1 Análisis de calidad de datos NGS por FastQC

El paquete bioinformático FastQC permite realizar análisis detallados sobre la calidad de los de datos de secuenciación de nueva generación, este análisis tiene como objetivo representar gráficamente la calidad global de los datos. FastQC también calcula un índice de calidad global de los datos y emite una recomendación sobre la calidad de los mismos, aprobándolos cuando poseen buena calidad.

#### 1.4.1.1 Procesamiento bioinformático

Los archivos obtenidos del equipo de secuenciación poseen la terminación `nombredelarchivo.fastq.gz`. Para realizar el procesamiento bioinformático de manera ordenada, es necesario crear una nueva carpeta y colocar el archivo de secuenciación dentro de él. Posteriormente, se debe abrir una terminal que corresponda al directorio de la carpeta que contiene el archivo. En la terminal, se debe emplear el comando `fastqc` seguido del nombre del archivo y `enter`, la terminal procesa el comando y coloca los archivos resultantes del análisis de la calidad en la carpeta contenedora.

#### 1.4.1.2 Presentación de resultados

##### 1.4.1.2.1 Resumen de todos los datos.

Como primer análisis, FastQC mostrará una tabla que resume todos los datos del análisis en ella podemos encontrar el nombre del archivo, número de secuencias totales, número de secuencias marcadas como de mala calidad, rango de longitud de secuencia y el porcentaje del contenido de porcentaje GC.

##### 1.4.1.2.2 Calidad de secuencia por base.

Este análisis muestra por cuartiles la calidad de los datos, graficando la mediana de la calidad en la posición de la molécula con una línea azul, y con una la línea roja, el valor mediano. Para cada posición se asigna un criterio de calidad, dependiendo de los

datos crudos provenientes del equipo, se distinguen 3 zonas dentro del gráfico, una zona verde que refiere buena calidad (28-38), una zona naranja que refiere calidad intermedia (20-27) y por último, una zona roja que refiere mala calidad (>20).

#### 1.4.1.2.3 Calidad de secuencia por mosaico.

Este análisis, permite evaluar si durante la corrida alguna sección de la celda de flujo sufrió algún percance, por ejemplo, la presencia de aire en la celda debido a un mal cargado. Las celdas de flujo de Illumina están conformadas por 8 carriles, con 2 columnas y 50 mosaicos por columna. El gráfico por mosaico permite evaluar si durante la corrida algún mosaico produjo información de secuencia de mala calidad, ya que estos fenómenos se marcan en colores cálidos (rojo-naranja), mientras que los mosaicos de los cuales se obtuvieron lecturas de calidad se representan en colores fríos (tonalidades de azul).

#### 1.4.1.2.4 Puntuación de calidad por secuencia.

La finalidad de esta gráfica es indicar el número de secuencias que presentan cierta puntuación de calidad media del universo analizado.

#### 1.4.1.2.5 Porcentaje de base por posición en la molécula.

Se grafica el porcentaje de cada una de las bases adenina (A), timina (T), guanina (G) y citosina (C) en cada posición de las moléculas. Cuando en una posición de la molécula la proporción de una base es 100%, indica que todas las moléculas secuenciadas poseen esa base en la posición analizada. Por el contrario, si en una posición de la molécula la proporción de las bases está compartida, indica que las secuencias analizadas son distintas entre ellas.

#### 1.4.1.2.6 Contenido promedio de GC en todas las moléculas.

La gráfica muestra la media del porcentaje del contenido de guanina (G) y citosina (C), (contenido GC) de cada molécula analizada. La línea azul representa una distribución poblacional teórica estándar. Es posible encontrar una distribución unimodal cuando

existe dispersión en los datos de secuenciación. Cuanto menos diversa sea la muestra secuenciada, será posible observar una distribución multimodal.

#### 1.4.1.2.7 Distribución de longitud por secuencia.

Este gráfico, permite visualizar la distribución de las moléculas de acuerdo al número de bases (longitud). En el eje de las “x” se grafica la longitud de la molécula indicada en pares de bases y en el de las “y”, se indica su frecuencia, el número de moléculas que presentan esa longitud de todas las moléculas analizadas.

#### 1.4.1.2.8 Niveles de duplicación de secuencia.

Este gráfico representa el nivel de duplicación de la secuencias y se acompaña de un porcentaje de *deduplicación* que muestra la proporción de las secuencias únicas (en este parámetro se consideran a las secuencias repetida una sola vez, independientemente del número de veces que se encuentre repetida).

La línea azul representa el número de moléculas que están repetidas y a línea roja muestra la proporción de las secuencias únicas (en este parámetro se consideran a las secuencias repetida una sola vez “únicas”, independientemente del número de veces que se encuentre repetida) del total de moléculas analizadas. En el eje de las “x” se muestra el número de veces que la moléculas se repite y en el eje de las “y” la proporción de la molécula del total de moléculas únicas (independientemente del número de veces que se encuentre repetida).

#### 1.4.1.2.9 Contenido de adaptador.

En esta gráfica se mostrará la presencia de los adaptadores Illumina (Universal, Small 3' RNA, Small 5' RNA, Nextera) en la molécula y se indica en qué posición y proporción (pares de bases) se encuentra la información de los adaptadores.

### 1.4.2 Eliminación de los adaptadores

Para iniciar con este procesamiento de manera organizada, primeramente se creó un directorio raíz (carpeta) donde se colocaron los archivos \*.fastq.gz.

Adicionalmente, se colocó un archivo \*.fasta que contenía la información de secuencia de los iniciadores empleados, en este caso el archivo contiene la información de los aparadores de Illumina (TrueSeqPCRfreeAdapters.fasta) que deseamos eliminar de nuestras secuencias. Como parte de un proceso traceable, decidimos adicionar una “T” al inicio del archivo resultante para identificar que el archivo está procesado. Abriremos una ventana de la terminal en este directorio, empleando el comando `ls` para verificar que dentro del directorio se encuentre el archivo \*.fastq.gz y el archivos de los adaptadores \*.fasta. Adicionalmente, se empleó el comando `-ltrh` para verificar el tamaño de los archivos, ya que los datos del tamaño pueden ser empleados como un criterio más de verificación de la información. para eliminar la información en los adaptadores en nuestras secuencias se emplea el comando `trimmomatic SE -threads 6 *.fastq.gz T*.fastq.gz ILLUMINACLIP:TrueSeqPCRfreeAdapters.fasta:2:30:10 SLIDINGWINDOW:4:30 LEADING:30 TRAILING:30 MINLEN:80`. Al finalizar el proceso se creará un archivo T\*.fastq.gz el cual contendrá la información de las secuencias pero sin los adaptadores.

#### 1.4.2 Eliminación de los iniciadores

Para eliminar la secuencia de iniciadores empleados para la amplificación de libUABC utilizamos el comando `FAstx_clipper`. Este comando requiere como archivo de entrada un \*.fastq, en caso de tener un archivo \*.fastq.gz es necesario descomprimirlo empleando el comando `gunzip -k *.fastq.gz`. Para emplear este comando es necesario introducir la secuencia de iniciadores que se desea eliminar del archivo, en este caso “CAGGATGATGTGTGATGGC”. Para realizar este procedimiento, debemos abrir la terminal en el directorio contenedor del archivo e introducir el comando `fastx_clipper -a CAGGATGATGTGTGATGGC -d 80 -v -i *.fastq -o PTN*.fastq`.

## 1.4.2 Rasurado de las secuencias

Como parte final del procesamiento realizaremos un rasurado de las secuencias a una longitud exacta de 80 bases, para esto empleamos el comando trimmomatic SE -threads 6 \*.fastq C\*.fastq CROP:80



## Capítulo 6

# Resultados

*“La libertad consiste, en ser dueños de la propia vida”*

*-Platón*

### 6.1 Diseño de la biblioteca libUABC para el aislamiento de aptámeros.

#### 6.1.1. Diseño de la biblioteca libUABC.

Se analizó el iniciador forward con la secuencia GACCTGACAGACACCTTGG de acuerdo a las instrucciones del fabricante, empleando la herramienta en línea OligoAnalyzer con los parámetros estándar una concentración de oligo 0.25 $\mu$ M, dNTPs de 0mM, de Na 50 mM y Mg 0 mM. La Tabla 1 muestra los resultados obtenidos del análisis bioinformático para las estructuras de horquilla y los parámetros fisicoquímicos, la Figura 1 muestra las posibles estructuras de horquilla calculadas.

Se analizó el iniciador reverse con la secuencia CAGGATGATGTGTGATGGC empleando la herramienta en línea OligoAnalyzer con los parámetros estándar una concentración de oligo 0.25 $\mu$ M, dNTPs de 0mM, de Na 50 mM y Mg 0 mM. Siendo una secuencia lineal a 25°C se obtuvieron las siguientes estructuras hairpin. La Tabla 1 muestra los resultados obtenidos del análisis bioinformático para las estructuras de horquilla y los parámetros fisicoquímicos, la Figura 1 muestra las posibles estructuras de horquilla calculadas.

**Tabla 1.** Resultados horquillas de secuencia forward por oligoanalyzer

Estructura	$\Delta G$ (kcal.mol <sup>-1</sup> )	T <sub>m</sub> (·C)	$\Delta H$ (kcal.mol <sup>-1</sup> )	$\Delta S$ (cal.K <sup>-1</sup> mol <sup>-1</sup> )
1	0.51	16.1	-16.8	-58.07
2	0.57	14.9	-16.2	-56.24
3	0.6	16.2	-19.6	-67.75
4	0.67	9.8	-12.5	-44.18
5	0.69	13	-16.5	-57.66
6	0.7	12.7	-16.4	-57.37
7	0.73	15.2	-21.6	-74.91
8	0.99	2.5	-12.1	-43.89
9	1.22	5.7	-17.7	-63.47
10	1.25	-25.6	-6.1	-24.64

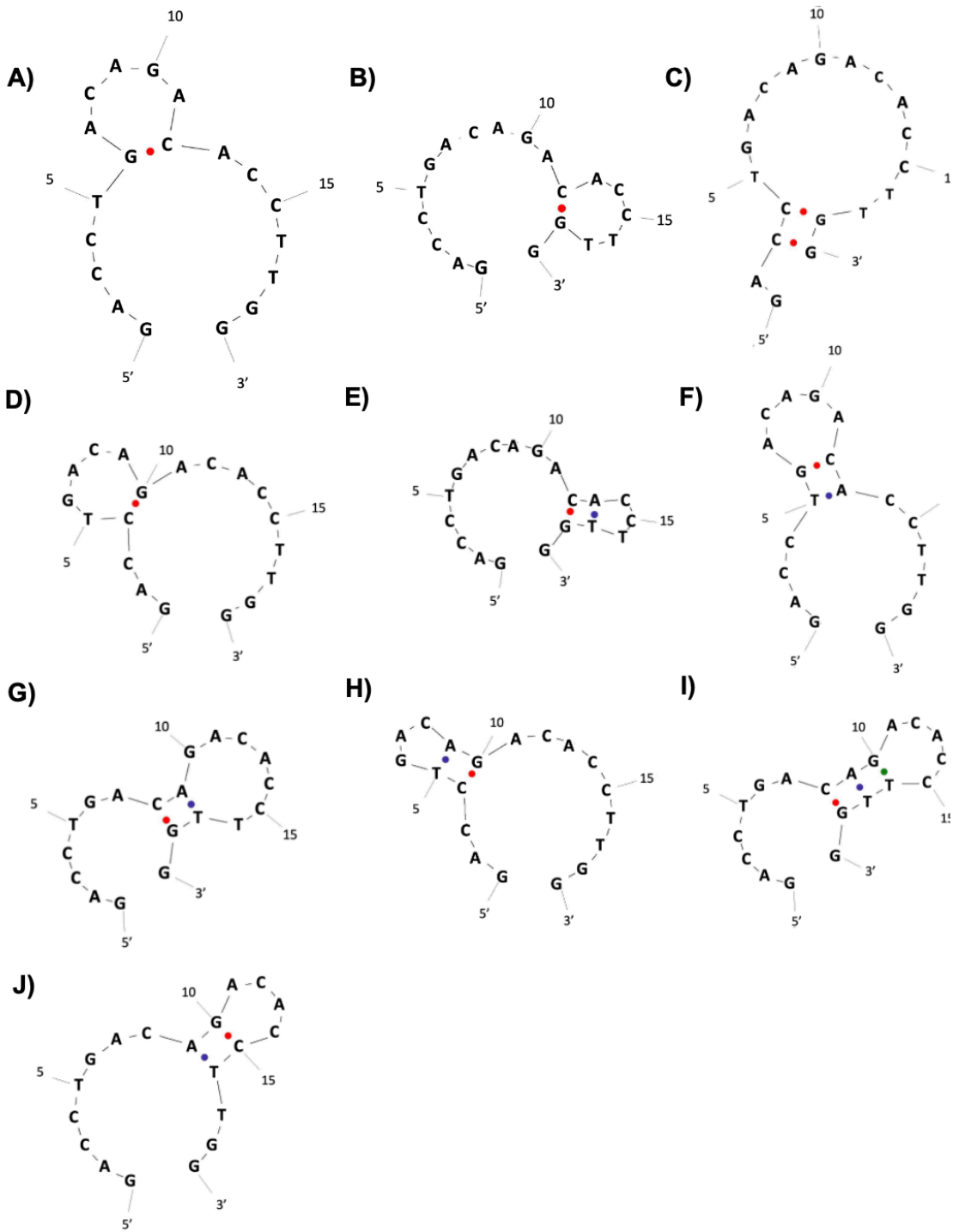


Figura 1. Análisis *in silico* de estructuras secundarias reverse libUABC. A) Estructura 1, B) Estructura 2, C) Estructura 3, D) Estructura 4, E) Estructura 5, F) Estructura 6, G) Estructura 7, H) Estructura 8, I) Estructura 9, J) Estructura 10.

**Tabla 2.** Resultados hairpin de secuencia reverse por oligoanalyzer

Estructura	$\Delta G$ (kcal.mol <sup>-1</sup> )	Tm(·C)	$\Delta H$ (kcal.mol <sup>-1</sup> )	$\Delta S$ (cal.K <sup>-1</sup> mol <sup>-1</sup> )
------------	--------------------------------------	--------	--------------------------------------	---

A	0.3	15.7	-9.4	-32.54
---	-----	------	------	--------

B	1.21	-9.8	-9.2	-34.93
---	------	------	------	--------

Análisis *in silico* de estructuras secundarias reverse libUABC.

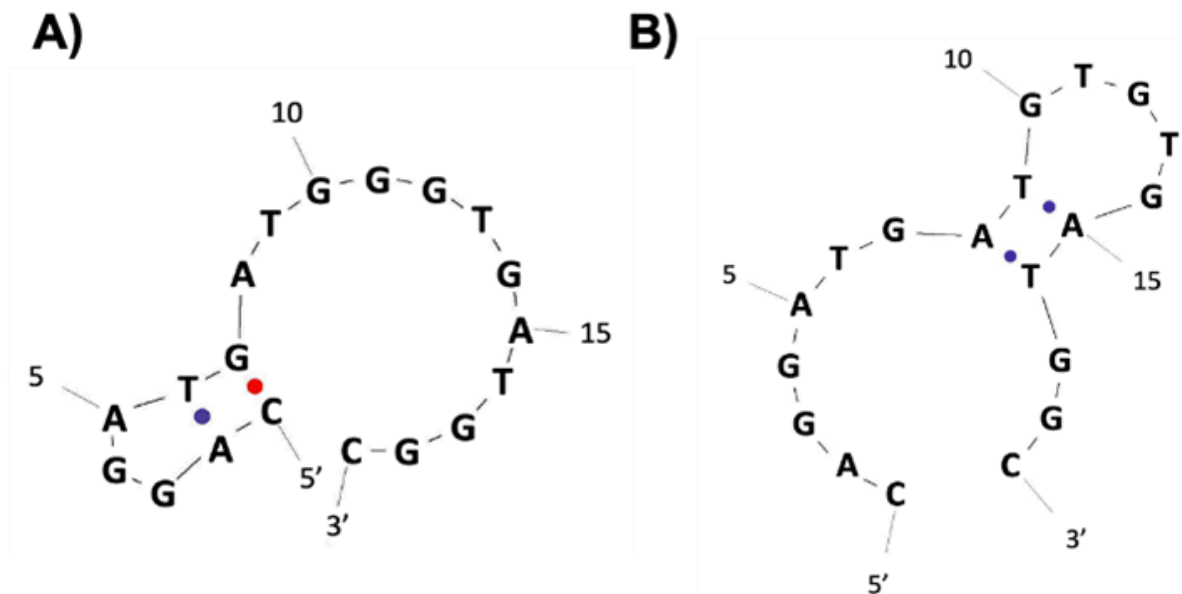


Figura 2. Análisis *in silico* de estructuras secundarias reverse libUABC. A) Estructura 1. B) Estructura 2.

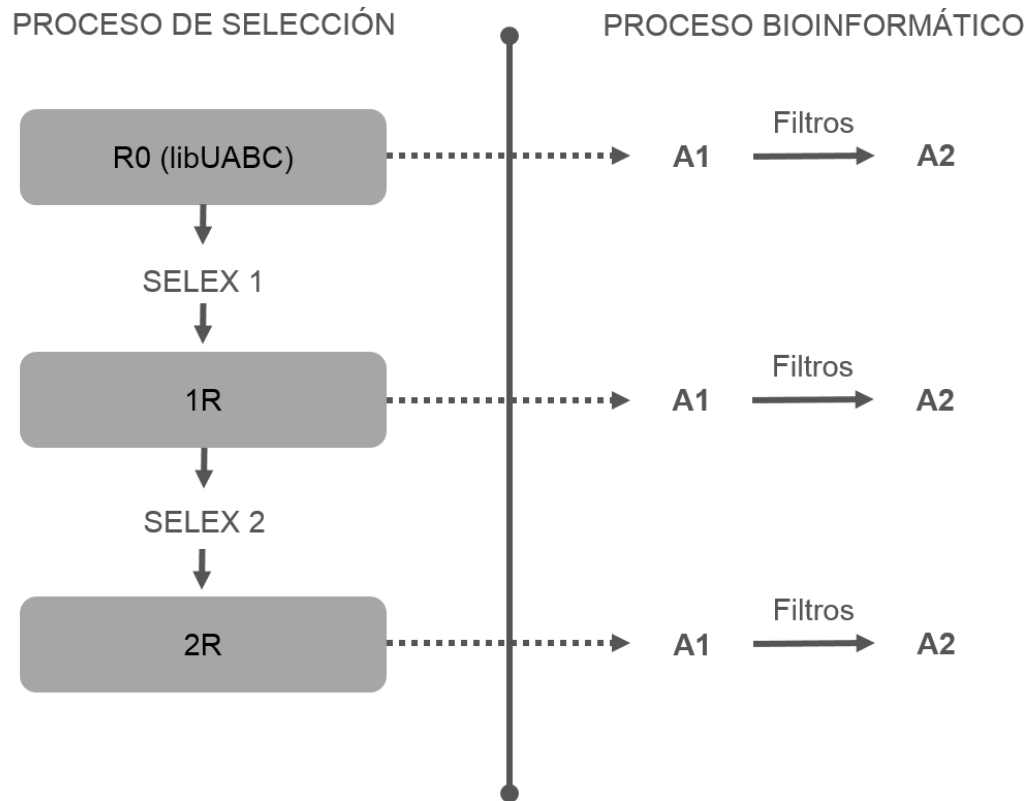


Figura 3. Proceso de selección y bioinformático.

## 7.1. Secuenciación de Ronda 0 (libUABC)

### 7.1.1 Análisis 1 de la Ronda 0

Se analizaron los datos de secuenciación de nueva generación en la plataforma Illumina para la biblioteca no seleccionada (R0) empleando el paquete bioinformático FastQC, obteniendo los siguientes resultados. En la Tabla 3 se muestran los datos globales del análisis, identificando 2,159,420 secuencias con buena calidad, 0 secuencias de mala calidad, las secuencias varían en longitud entre 35-151 bases y el porcentaje de GC es de 49%.

Para visualizar la calidad de las secuencias por base de acuerdo a su longitud, en la figura R0V1-1A se grafica en el eje “x” la posición de la base en la secuencia y en el eje “y” la calidad de cada base en la secuencia en un rango de 0 a 38, la línea azul identifica la mediana y la línea roja marca la media de la calidad, para cada posición se asigna un criterio de dispersión de la calidad, dependiendo de los datos crudos provenientes del equipo. Se distinguen 3 zonas dentro del gráfico, una zona verde que abarca de 28 a 38 lo cual refiere buena calidad, zona naranja de 20 a 28 de calidad intermedia y por último la zona roja de 0 a 20 que corresponde a una mala calidad.

La Figura R0V1-1A corresponde a R0 (libUABC) y muestra las secuencias desde la posición 1 a la 151. Los datos de la posición 1 a la 80 poseen buena calidad ya que sus puntuaciones son superiores a 34, estas posiciones además presentan una mediana y promedio de buena calidad. Las secuencias desde las posición 81 a la 151 poseen una distribución muy amplia como para identificarlas con buena calidad, ya que la distribución de la puntuación de calidad por posición de lectura, varía desde 11 hasta 38 con medianas y medias dentro de las zonas de buena calidad, con la excepción de la última lectura donde la mediana y media se encuentran por debajo de 28.

Los datos de secuenciación obtenidos empleando la plataforma illumina para la muestra R0 (libUABC) poseen una buena calidad de la base 1 a la 80, pero son de mala calidad a partir de la base 81. Estos datos son congruentes con a la muestra secuenciada, ya que los fragmentos analizados son de 80 bases, por lo cual la información de la posición 81-151 corresponde a extensiones de las secuencias analizadas, por lo que estas no son de interés y deberán ser removidas para poder realizar un análisis de los resultados del proceso de evolución *in vitro*.

Para evaluar la calidad de los datos obtenidos sobre la celda de flujo se emplea la Figura R0V1-1B. Este gráfico de mosaico de calidad de la secuencia, muestra en el eje “x” la posición de la lectura de la secuencia y en el eje “y” la distribución espacial en la celda de flujo. Este gráfico marca en tonos fríos las secuencias de buena calidad y en tonos cálidos las áreas de el chip que poseen baja calidad.

La Figura R0V1-1B se muestra que los datos con longitudes de 1 a 80 muestran mosaicos en colores fríos de la posición 81 en adelante, algunos mosaicos muestran

colores más cálidos que se identifican como datos de no tan buena calidad, estos son abundantes al final de la gráfica.

En general la Figura R0V1-1B muestra que el cargado de la celda de flujo fue muy bueno ya que no es posible identificar mosaicos dañados. Las secuencias con longitudes superiores a 80 mostraron algunos mosaicos con colores menos fríos, pero no lo suficientemente cálidos (naranja o rojo) como para identificarlos de mala calidad.

En la Figura R0V1-2A se muestra el gráfico de nivel de calidad por secuencia, el cual en el eje “x” muestra la posición de la lectura y en el eje “y” el número de secuencias que corresponden a esa media, los datos muestran que existen más de 1.4 millones de lecturas que presentan una calidad de 38. Los datos muestran que la mayoría de las secuencias son de buena calidad, de manera consistente con los análisis anteriores, de manera global la calidad de las secuencias es buena.

La Figura R0V1-2B muestra el contenido de GC de las secuencia, en esta gráfica muestra el eje “x” muestra el porcentaje de GC en cada secuencia y en el eje “y” el número de secuencias con ese porcentaje de GC. Los datos muestran una distribución donde la mayoría de las secuencias tiene un porcentaje de GC entre 35 y 63 por ciento, con una media de distribución de 49%. Este gráfico unimodal, con un pico simétrico y de base estrecha y reflejando consistencia en los datos.

En la Figura R0V1-3A muestra el contenido de bases por cada posición en la molécula de la muestra analizada, en el eje de las “x” se grafica la posición en la molécula en pares de bases y en el eje “y” el contenido de cada una de las bases en porcentaje. Los resultados muestran tres secciones bien definidas en la gráfica, de la posición 1 a 19 con identidades bien definidas, 20-61 en esta región la identidad de las secuencias es de 25% para cada base aproximadamente y finalmente 43 en adelante donde la identidad de la base en la posición está más definida. Los resultados son consistente con el diseño de la libUABC, el cual presenta una región fija de 19 bases, una región aleatorizada de 20 a la 61 donde de manera proporcional se combinan las cuatro bases y finalmente una región fija de la 61 a la 80.

La Figura R0V1-3B muestra la distribución de tamaño de las moléculas analizadas, se grafica en el eje de las “x” la longitud de las moléculas y en el eje de las “y” el número de moléculas con esa longitud en pares de bases. Los resultados muestran un único pico a 80-84 pares de bases con una frecuencia de 1.8 millones. Estos resultados son consistentes con lo esperado, ya que las moléculas analizadas presentan un tamaño de 80 bases.

La Figura R0V1-4A muestra el nivel de duplicación de secuencias de libUABC, se grafica en el eje de las “x” el nivel de duplicación de las secuencias y en el eje de las “y” el porcentaje de secuencias restantes si se eliminan los duplicados. Los resultados muestran que de las secuencias totales existe un 2% que muestra duplicación y manteniendo un 98% de secuencias únicas.

La Figura R0V1-4B muestra el porcentaje de los adaptadores en las secuencias, se grafica en el eje de las “x” la posición de la lectura en la secuencia que corresponde a la identidad de los adaptadores y en el eje de las “y” el porcentaje de los adaptadores. Los resultados muestran presencia de los adaptadores después de 80 bases, lo que corresponde a lecturas fuera de la secuencia de interés.

Tabla 3. Tabla resumen de los datos	
Total de secuencias	2159420
Secuencias de baja calidad	0
Longitud de las secuencias	35-151
%GC	49

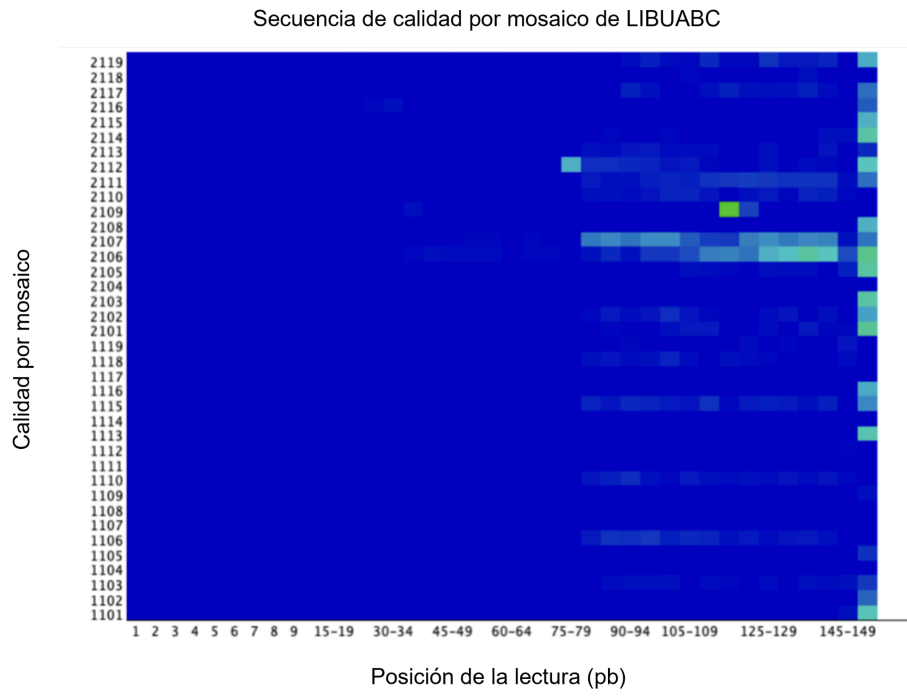
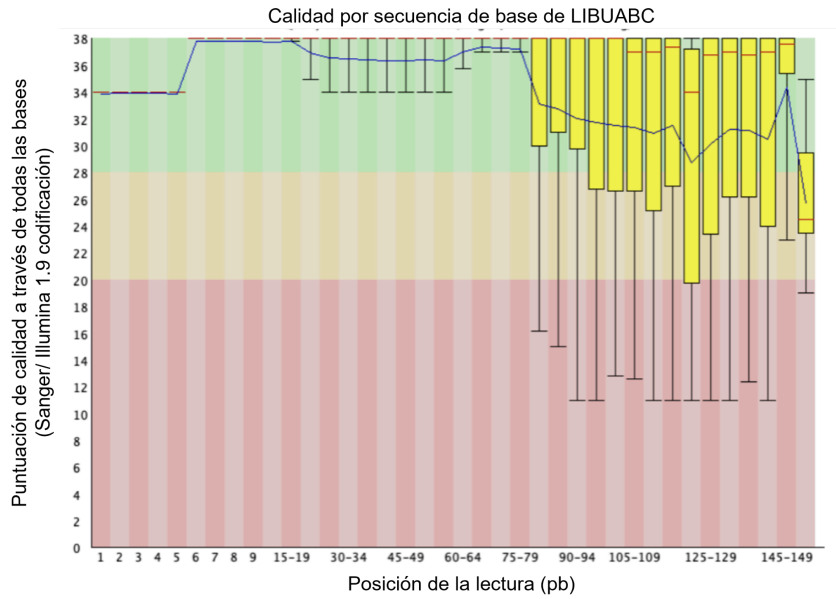


Figura R0V1-1. Análisis de la calidad de las secuencias por base y mosaico. A) Por base, las bases de 1-80 poseen muy buena calidad en todas las moléculas. B) Por mosaico, no se identifican secciones de la celda de flujo dañadas o contaminadas.

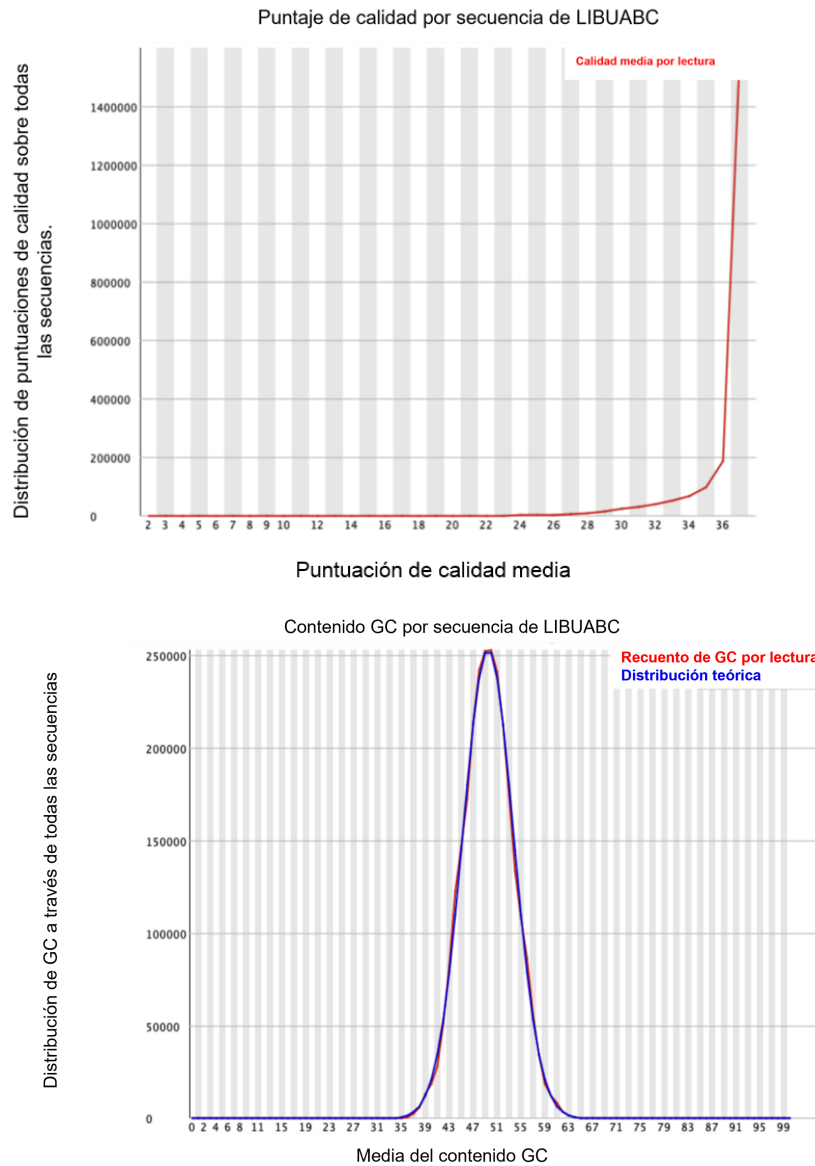


Figura R0V1-2. Análisis de la calidad por secuencia y contenido GC. A) Por secuencia, más de 1.4 millones de secuencias presentan la máxima calidad. B) Contenido GC, distribución unimodal con una media de 49%.

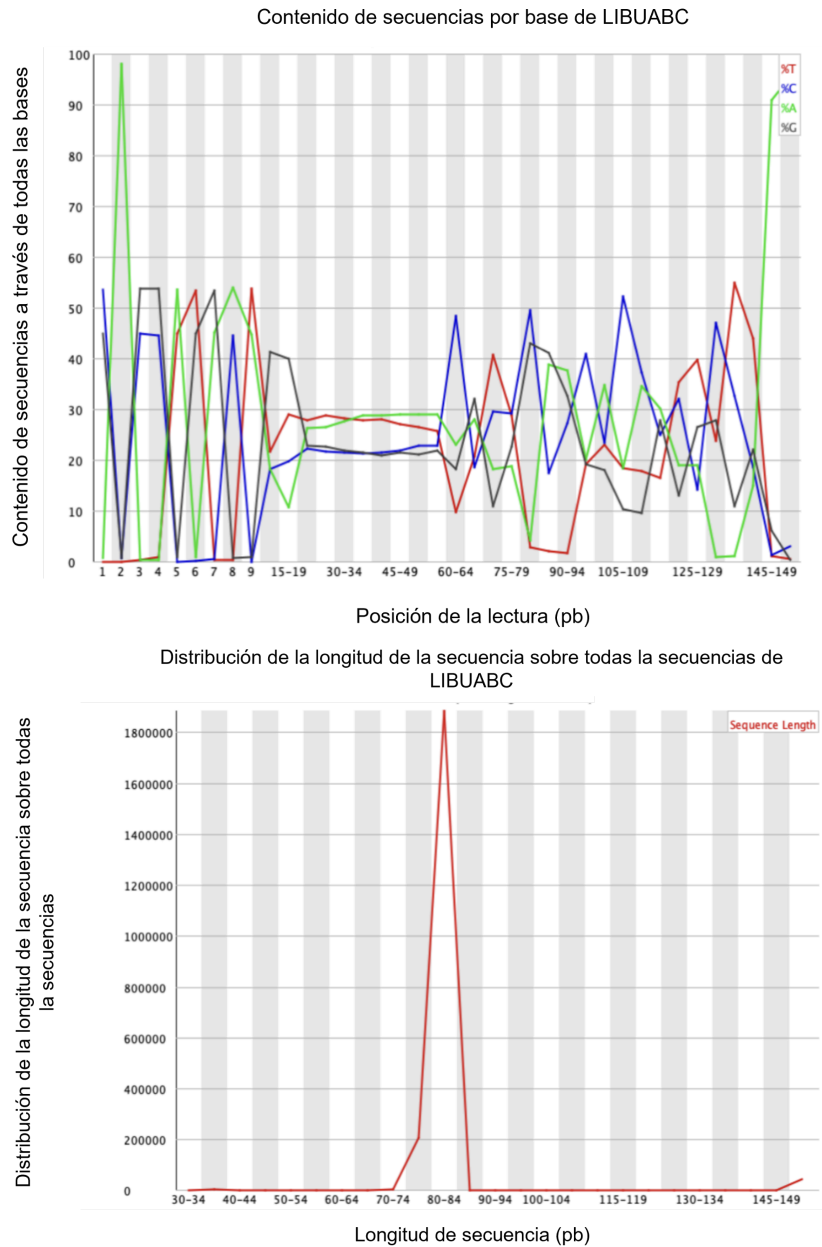


Figura R0V1-3. Análisis de las bases por secuencia y su longitud. A) Porcentaje de base por posición en la lectura. B) distribución por longitud de las moléculas.

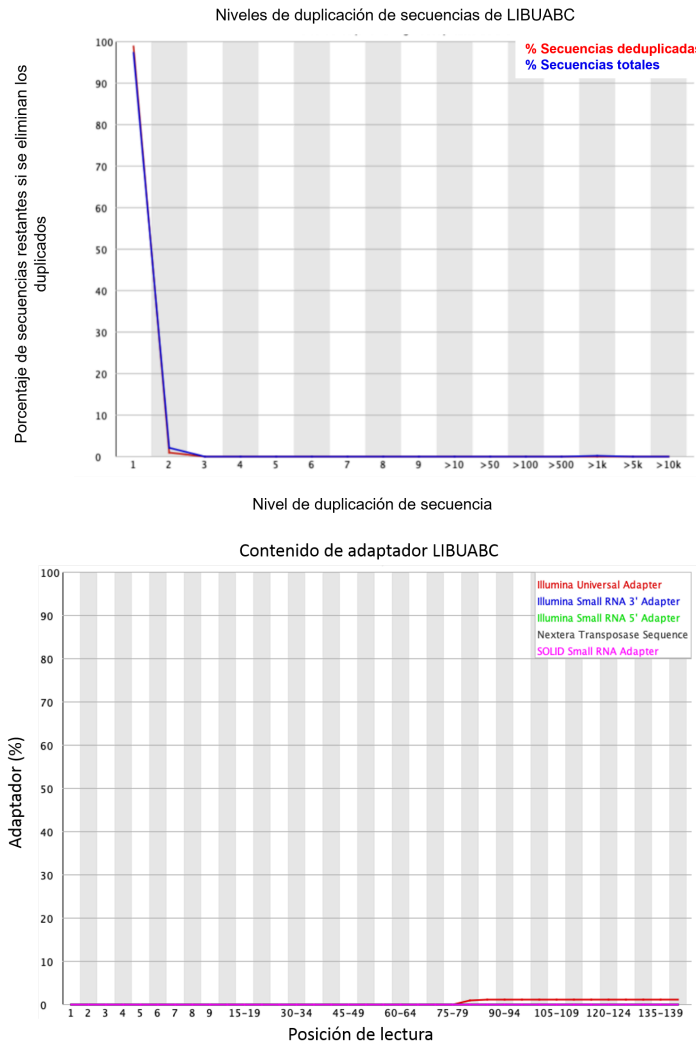


Figura R0V1-4. Análisis de nivel de duplicación y adaptadores. A) Duplicación de las secuencias, nivel de duplicación 2% . B) porcentaje de adaptadores en las moléculas.

### 7.1.2 Análisis 2 de la Ronda 0

Se analizaron los datos de secuenciación de nueva generación en la plataforma Illumina para la biblioteca no seleccionada (R0) después del filtrado de secuencias de baja calidad (R0V2) empleando el paquete bioinformático FastQC, obteniendo los siguientes resultados. En la tabla 4 se muestran los datos globales del análisis, identificando 685,038 secuencias con buena calidad, 0 secuencias de mala calidad, secuencias con longitud de 80 bases y un porcentaje de GC de 49%.

La Figura R0A2-1A muestra el análisis de calidad de la R0V2 (libUABC) donde las secuencias de la posición 1 a la 80 poseen buena calidad ya que sus puntuaciones son superiores a 34, estas posiciones además presentan una mediana y promedio de buena calidad, indicando que el procesamiento de datos permitió conservar sólo secuencias de buena calidad con una longitud de 80 bases.

La Figura R0A2-1B muestra que la mayoría de los datos poseen calidad, ya que se grafican con mosaicos de colores fríos, con la excepción de dos mosaicos con longitudes entre 74 a 75 ya que se presentan con mosaicos de colores más cálidos, de la posición 75 a la 77 se grafican con colores fríos que se identifican como datos de buena calidad, el análisis global del gráfico por mosaico indica que la celda de flujo se empleó de manera exitosa, y podemos constatar que se tratan de datos de buena calidad.

En la Figura R0A2-2A los datos muestran que existen más de 50,000 secuencias con una calidad de 36 y más de 600,000 con una calidad de 37.

En la Figura R0A2-2B los datos muestran una distribución donde la mayoría de las secuencias tiene un porcentaje de GC de 35 a 63 %, con una media de distribución de 49 %. Este gráfico unimodal, con un pico simétrico y de base estrecha, indica consistencia en los datos de secuenciación obtenidos.

En la Figura R0A2-3A muestra el contenido de bases por cada posición en la molécula de la muestra analizada, los resultados muestran tres secciones bien definidas en la gráfica, la primera de la posición 1 a 19 con identidades bien definidas que alcanzan el 100% de congruencia, la segunda de 20-61 en esta región la identidad

de las secuencias es de 25% para cada base, mostrando que en esta sección de las moléculas coexisten de manera equilibrada y finalmente la tercera sección de 62 a 80 donde la identidad de la base en la posición está definida. Los resultados son consistentes con el diseño de la libUABC, el cual presenta regiones fijas de 19 bases en los extremos de las moléculas y una región central aleatorizada de la posición 20 a la 61 (42 bases) donde se combinan las cuatro bases de manera proporcional.

La Figura R0A2-3B muestra la distribución de tamaño de las moléculas analizadas donde los resultados indican un pico de 80 pares de bases con una frecuencia mayor a 600,000 consistente con el número total de secuencias analizadas.

En la Figura R0A2-4A los resultados muestran que de las secuencias totales existe un 2% que muestra duplicación y por lo tanto el 98% de las secuencias son únicas.

En la Figura R0A2-4B no muestra presencia de adaptadores.

Tabla 4. Tabla resumen de los datos	
Total de secuencias	685038
Secuencias de baja calidad	0
Longitud de las secuencias	80
%GC	49

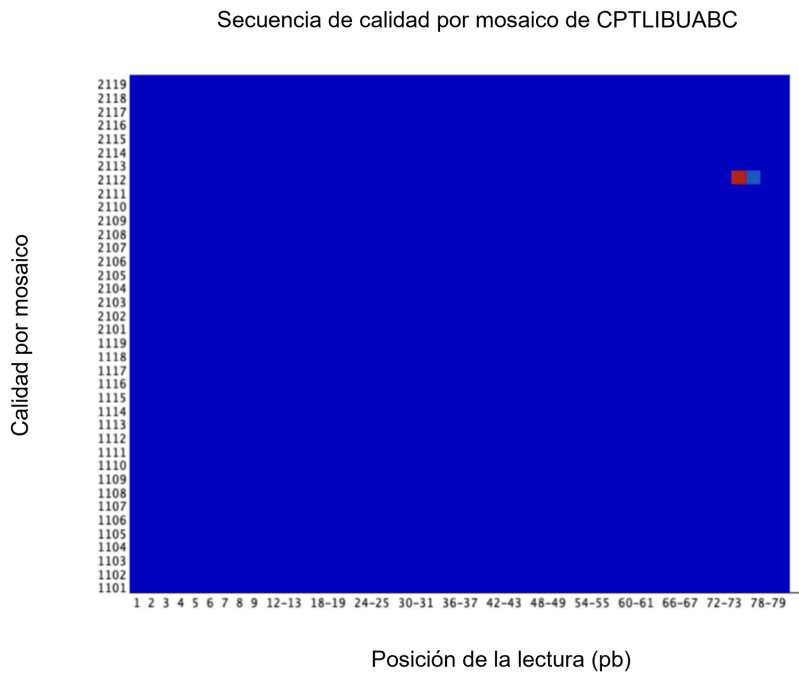
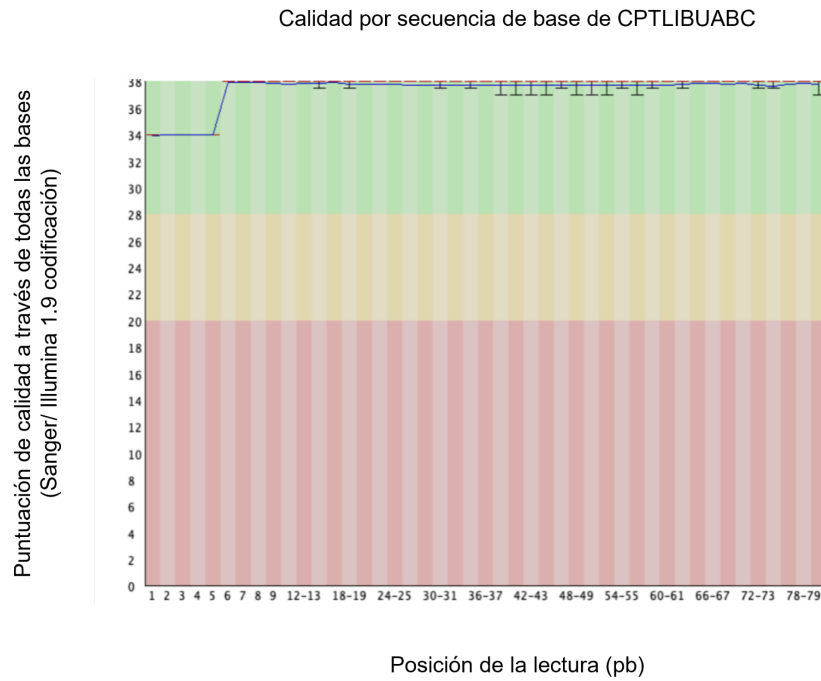


Figura R0A2-1. Análisis de la calidad de las secuencias por base y mosaico. A) calidad de secuencias por base, la información de calidad de secuencia es mayor a 34, identificado como de buena calidad. B) calidad de secuencias por mosaico, mayoritariamente los mosaicos son azules indicando buena calidad en la información

de secuencia.

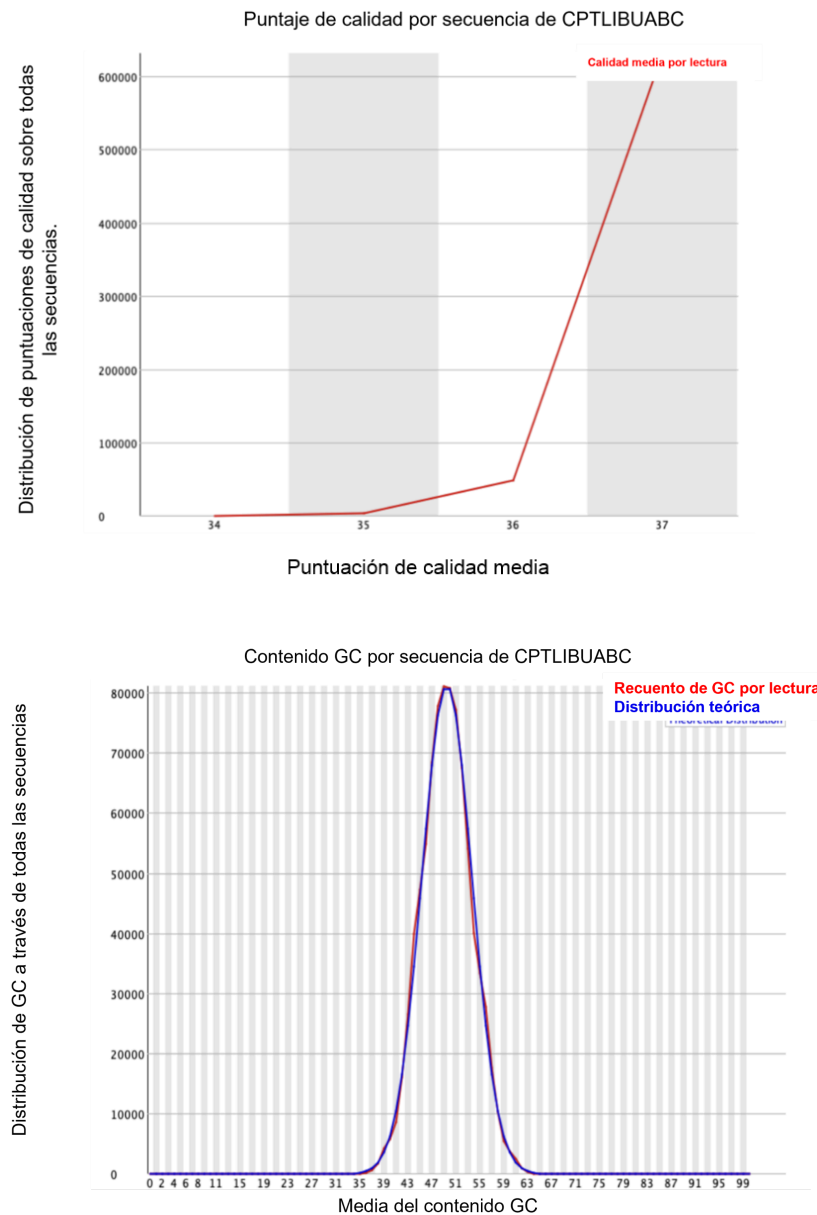


Figura R0A2-2. Análisis de la calidad por secuencia y contenido GC. A) Más de 600,000 poseen una calidad de 37, por lo que describirnos que todas poseen una

excelente calidad. B) La media de contenido GC es 49%, obteniendo un pico unimodal.

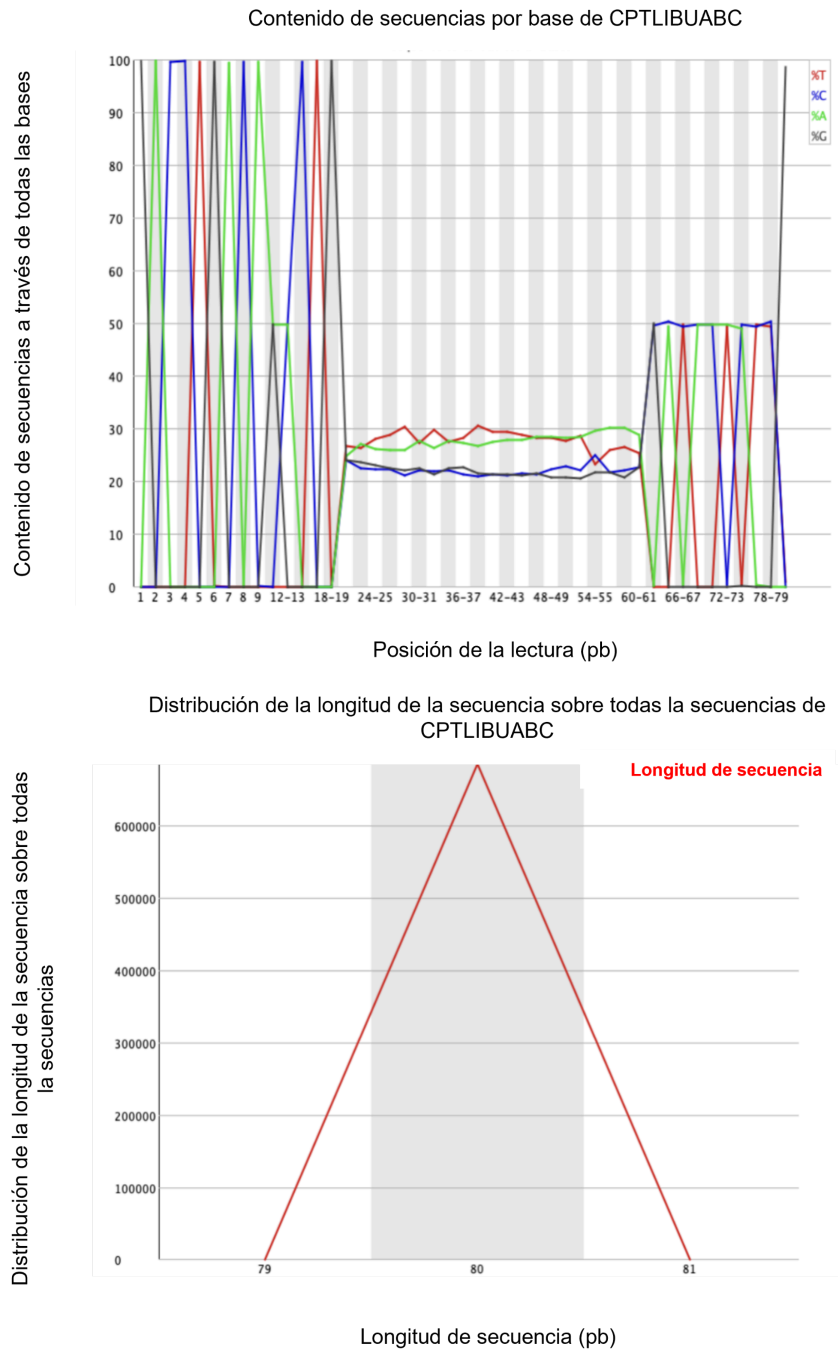


Figura R0A2-3. Análisis de las bases por secuencia y su longitud. A) tres secciones

bien definidas, 2 secciones definidas y 1 aleatoria. B) Más de 600,000 poseen una longitud de 80.

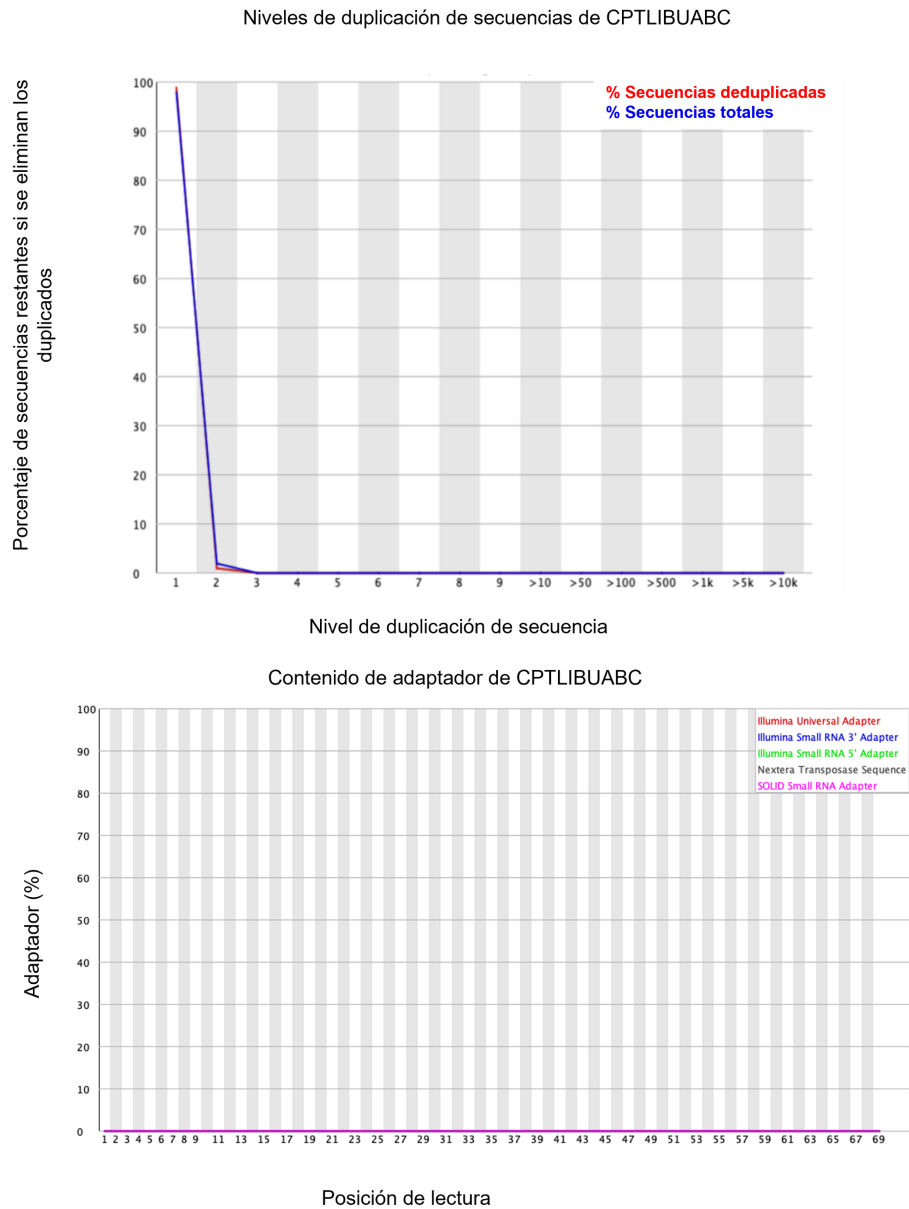


Figura R0A1-4. Análisis de nivel de duplicación y adaptadores. A) Duplicación de las secuencias, en la gráfica se indica un porcentaje de 2% de duplicación. B) Ausencia de adaptadores.

La secuenciación de las moléculas de la libUABC antes de ser sometida a un proceso de selección, reveló la identidad nucleotídica de 2,159,420 secuencias. Después del filtrado de calidad, recuperamos un 31.72% (se eliminaron 68.27%) de secuencias identificadas con buena calidad, este porcentaje corresponde a 685,038 secuencias con buena calidad.

## 7.2. Secuenciación de Ronda 1

### 7.2.1 Análisis 1 de la Ronda 1

En la Tabla 5 se muestran los datos globales del análisis por secuenciación del ADN recuperado después del SELEX contra las proteínas virales, mismas que se han identificado como Ronda 1 (R1). El análisis identificó 2,381,547 secuencias con buena calidad, 0 secuencias de mala calidad, las secuencias varían en longitud entre 35-151 bases y el porcentaje de GC es de 49%.

En la Figura R1A1-1A, los datos de secuenciación obtenidos empleando la plataforma Illumina para la muestra R1 (libUABC) poseen una buena calidad de la base 1 a la 80, pero son de mala calidad a partir de la base 81. Estos datos son congruentes con la muestra secuenciada, ya que los fragmentos analizados son de 80 bases, por lo cual la información de la posición 81-151 corresponde a extensiones de las secuencias analizadas, por lo que estas no son de interés y deberán ser removidas para poder realizar un análisis posterior empleando los resultados del proceso de evolución *in vitro*.

La Figura R1A1-1B gráfica la calidad de secuencia por mosaico, los resultados mostraron que la calidad de las las secuencias de la posición 1 a la 80, poseen buena calidad ya que los mosaicos son azules. La información de secuencia de las moléculas a partir de la base 80 mostraron algunos mosaicos con colores menos fríos (cian, verde, etc.), pero no cálidos (naranja o rojo) como para identificarlos de mala calidad.

En la Figura R1A1-2A correspondiente al gráfico de nivel de calidad por secuencia, los datos muestran que existen más de 200,000 moléculas que presentan una calidad de 36, mientras que más de 1.6 millones poseen el valor de calidad máxima (38). Los datos demuestran de manera consistente que la mayoría de las secuencias presentan una puntuación de calidad muy alta.

La Figura R1A1-2B gráfica el contenido de GC de las moléculas, los datos analizados señalan una distribución donde la mayoría de las secuencias tiene un porcentaje de GC entre 35 y 63 por ciento, con una media de distribución de 48%. El análisis de los datos es consistente, esto sustentado por el gráfico unimodal, con un pico simétrico y de base estrecha.

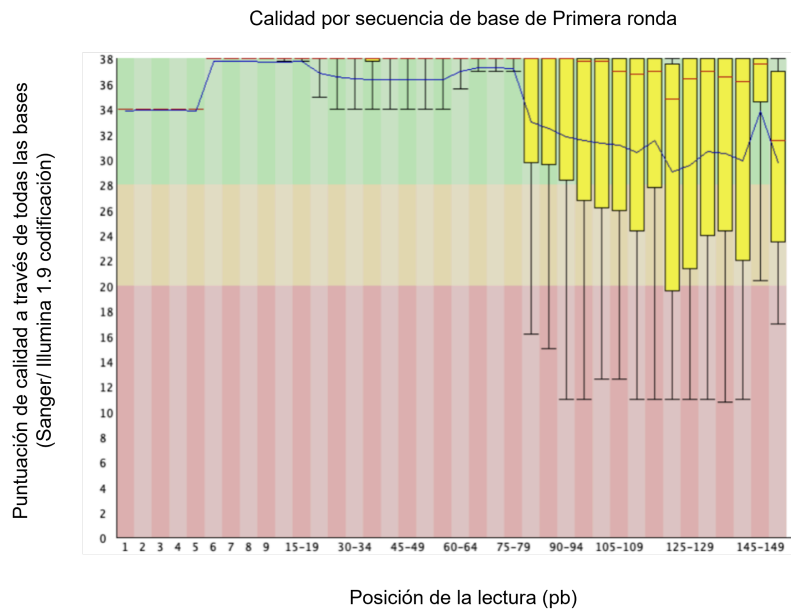
En la Figura R1A1-3A la gráfica muestra el contenido de bases por cada posición en la molécula de la muestra analizada. En la gráfica se identifican tres secciones bien definidas, de la posición 1 a 19 donde la identidad de la base es de 100%, 20-61 donde la identidad de las secuencias es de 25% para cada base aproximadamente y finalmente 62 en adelante donde la identidad de la base en la posición está definida. Los resultados son consistente con el diseño de la libUABC, el cual presenta una región fija de 19 bases, una región aleatorizada de 20 a la 61 donde de manera proporcional se combinan las cuatro bases y finalmente una región fija de la 61 a la 80.

La Figura R1A1-3B gráfica la distribución de tamaño de las moléculas analizadas. Los resultados indican un único pico a 76-87 pares de bases con una frecuencia de 2 millones. Estos resultados son consistentes con lo esperado, ya que las moléculas analizadas poseen un tamaño de 80 bases.

La Figura R1A1-4A indica el nivel de duplicación de secuencias de libUABC, Los resultados muestran que de las moléculas totales existe un 11% de moléculas únicas, 11% de moléculas de más de 10 copias, 45% de moléculas con más de 50 copias y un 28% de moléculas con más de 100 copias. Demostrando un número reducido de moléculas únicas y un marcado incremento en las secuencias con más de 10, 50 y 100 duplicados.

La Figura R1A1-4B analiza el porcentaje de adaptadores en las moléculas, los datos de la gráfica indican la presencia de adaptadores en las moléculas analizadas en posiciones de lectura superiores a 80 bases, posiciones que se encuentran fuera del rango de interés.

Tabla 5. Resumen de todos los datos	
Total de secuencias	2381547
Secuencias de baja calidad	0
Longitud de las secuencias	35-151
%GC	49



Secuencia de calidad por mosaico de Primera ronda

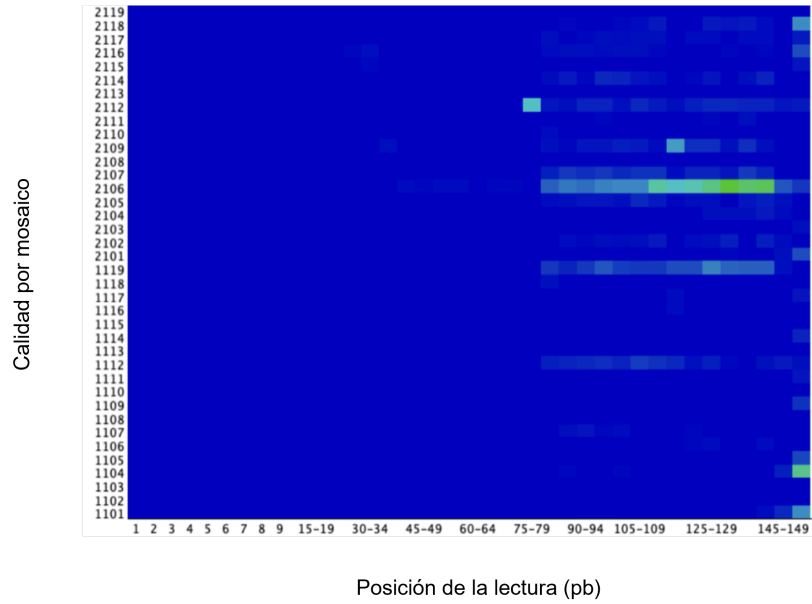
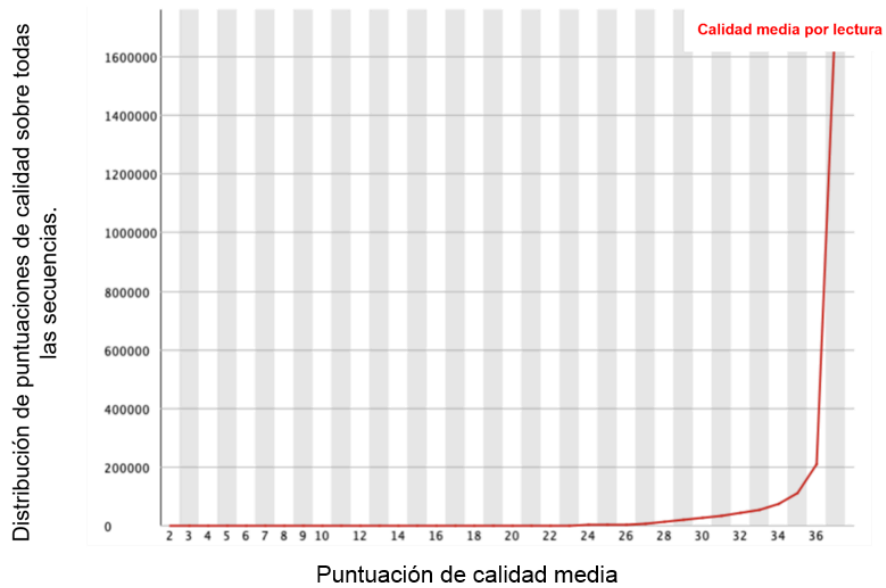


Figura R1A1-1. Calidad de secuencia por base y mosaico. A) De las secuencias 1 a la 80 hay presencia de buena calidad. B) Ausencia de celdas contaminadas en mosaicos.

Puntaje de calidad por secuencia de Primera ronda



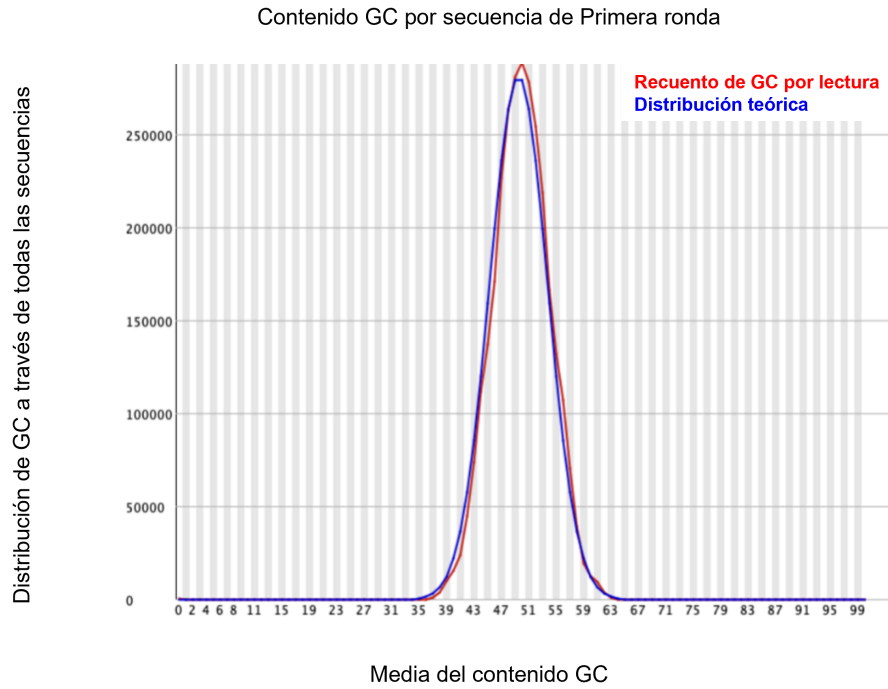


Figura R1A1-2. Puntaje de calidad por secuencia y GC. A) Más de 1,600,000 poseen una calidad superior a 36. B) La distribución es unimodal.

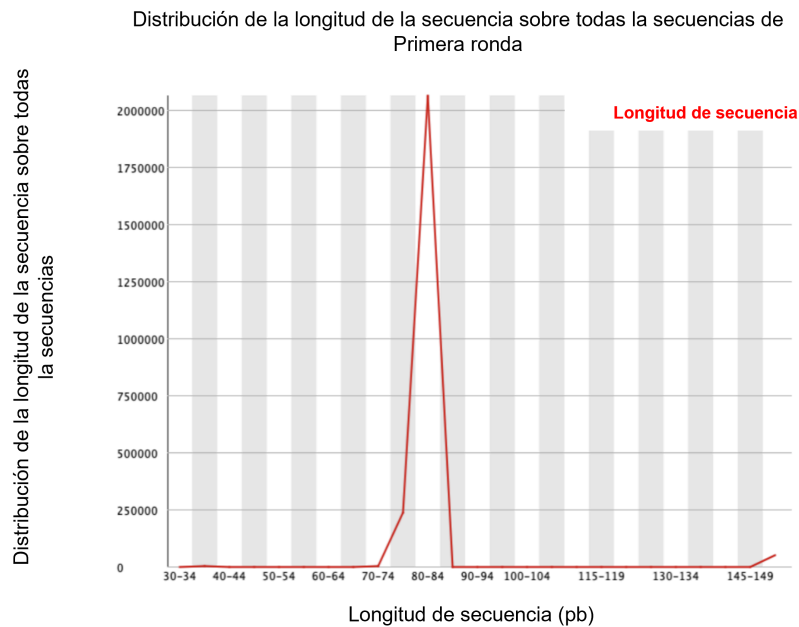
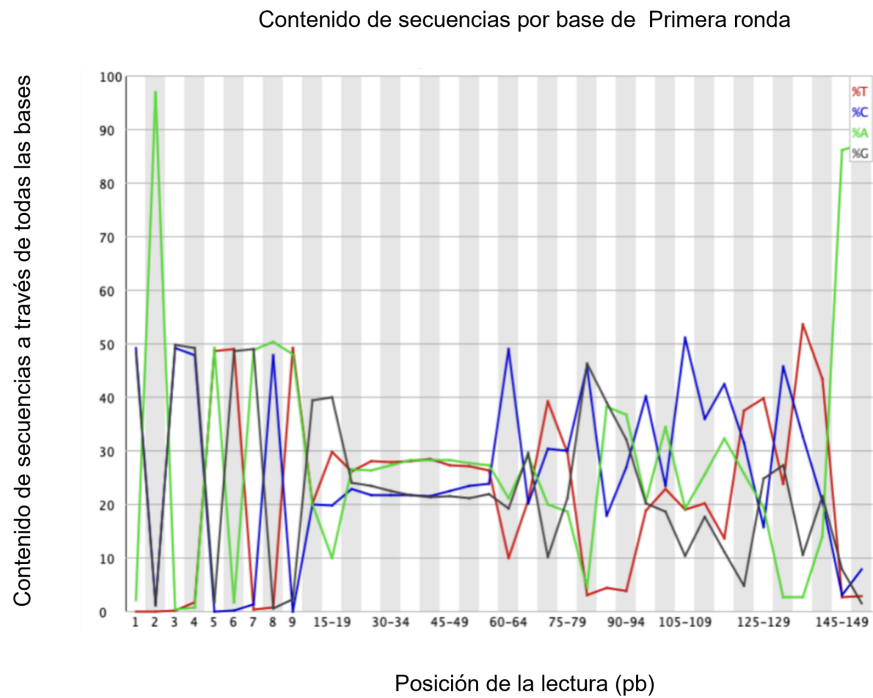


Figura R1A1-3. Contenido de secuencia por base y distribución de longitud.. A) Se identifican tres secciones bien definidas. B) Más de 2,000,000 de secuencias presentan una longitud de 80.

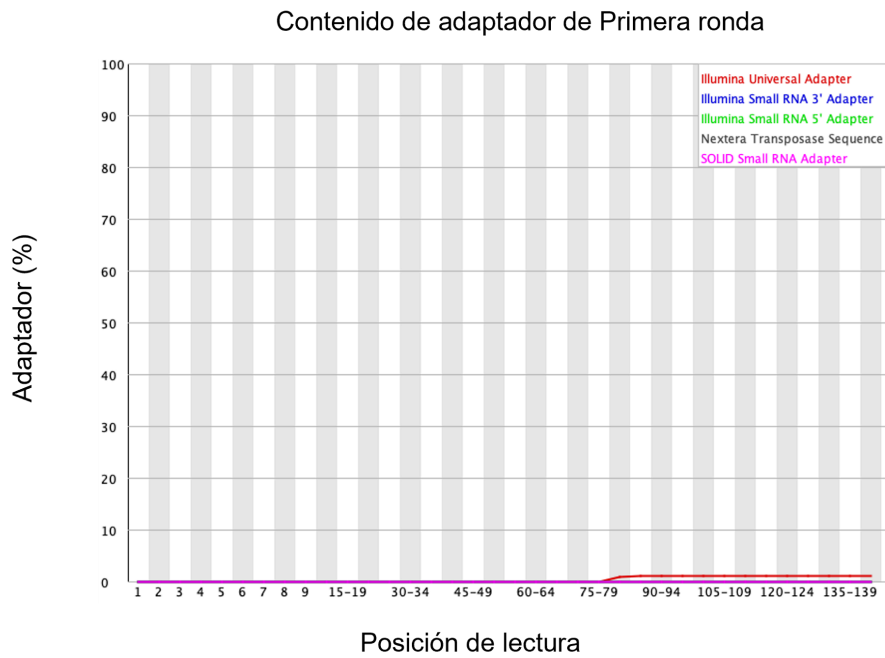
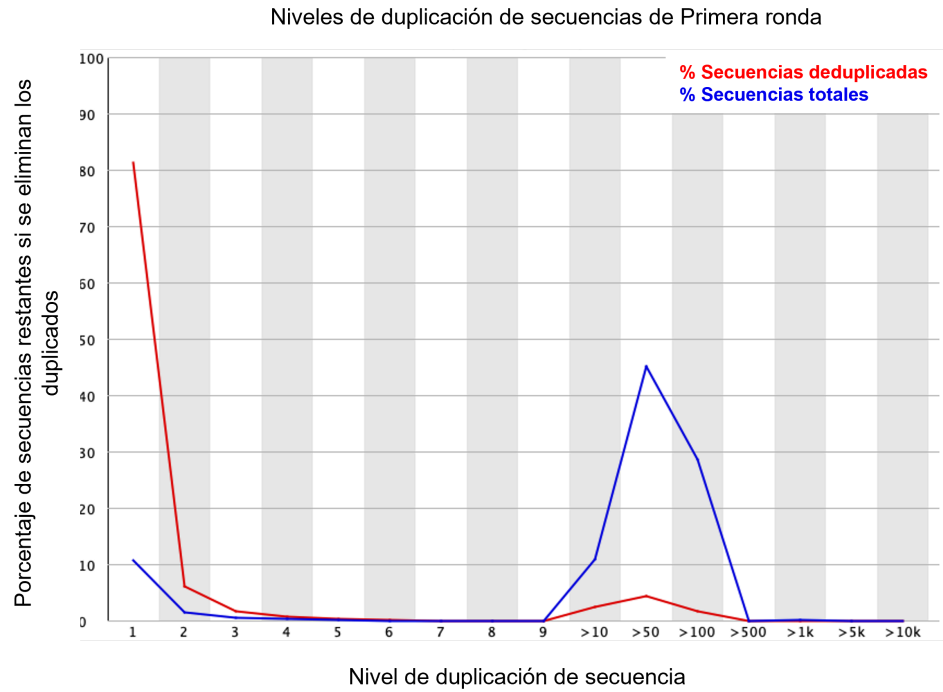


Figura R1A1-4. Análisis de nivel de duplicación y adaptadores. A) Duplicación de las secuencias. B) Presencia de adaptadores.

## 7.2.2 Análisis 2 de la Ronda 1

En la Tabla 6 se muestran los datos globales del análisis de secuenciación de las moléculas recuperadas después del primer SELEX empleando como blanco una combinación de las proteínas SARS-CoV-2 Spike Glycoprotein South African Mutation B.1.351, HEK293. SARS-CoV-2 (2019-nCoV) Spike 100µg S1(K417N, E484K, N501Y, D614G)-His Recombinant Protein Y SARS-CoV-2 Spike Glycoprotein UK Mutation B.1.1.7, HEK293. SARS-CoV-2 (2019-nCoV) Spike S1(HV69-70 deletion, Y144 deletion, N501Y, A570D, D614G, P681H)-His Recombinant Protein, identificando 820,143 secuencias con buena calidad, 0 secuencias de mala calidad, todas las secuencias poseen una longitud de 80 bases y un porcentaje de GC de 49%.

La Figura R1A2-1A muestra el análisis de calidad de la R1A2 donde las secuencias de la posición 1 a la 80 poseen buena calidad ya que sus puntuaciones son superiores a 34, estas posiciones además presentan una mediana y promedio de buena calidad, indicando que el procesamiento de datos permitió conservar sólo secuencias de buena calidad con una longitud de 80 bases.

La Figura R1A2-1B representa la secuencia de calidad por mosaico y muestra que la mayoría de los datos poseen calidad, ya que se grafican con mosaicos de colores fríos, con la excepción de dos mosaicos con longitudes entre 74 a 75 con mosaicos de colores más cálidos, de la posición 75 a la 77 se grafican con colores fríos que se identifican como datos de buena calidad, el análisis global del gráfico por mosaico indica que la celda de flujo se empleó de manera exitosa y indicando que los datos obtenidos del uso de esta, poseen buena calidad.

En la Figura R1A2-2A se indica el puntaje de calidad por secuencia cuyos datos muestran que existen 50,000 secuencias con una calidad de 36 y más de 700,000 con una calidad de 37.

En la Figura R1A2-2B los datos muestran una distribución donde la mayoría de las secuencias tiene un porcentaje de GC de 35 a 63 %, con una media de distribución de 49 %. Este gráfico unimodal de pico simétrico y con un base estrecha, indica consistencia en los datos de secuenciación obtenidos.

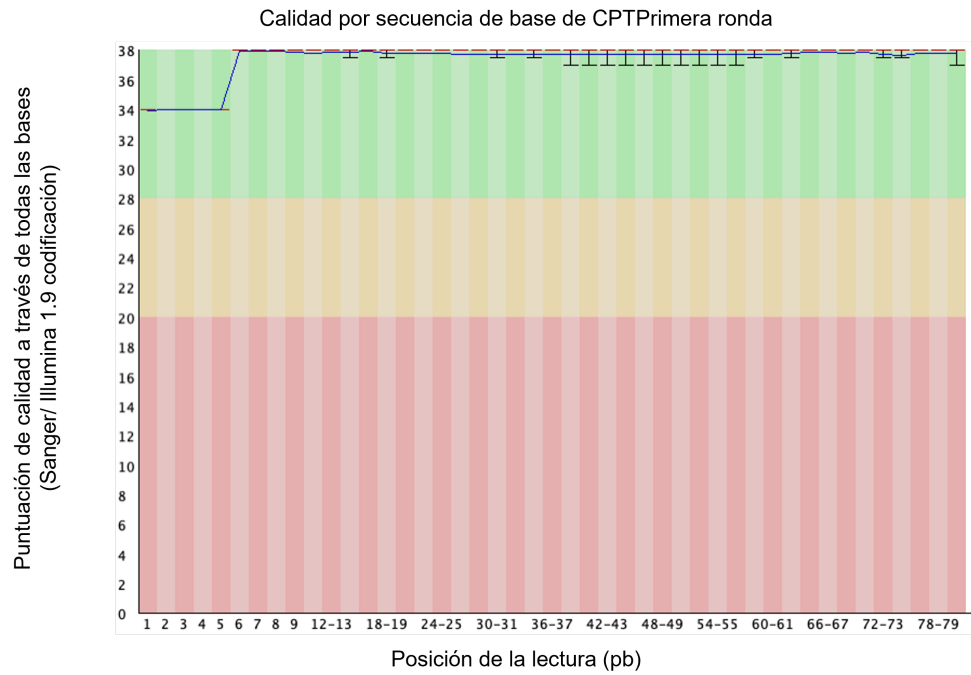
En la Figura R1A2-3A muestra el contenido de bases por cada posición en la molécula, los resultados muestran tres secciones bien definidas en la gráfica, la primera de la posición 1 a 19 con identidades bien definidas que alcanzan el 100 % de congruencia, la segunda de 20-61 en esta región la identidad de las secuencias es de 30% para cada base, mostrando que en esta sección de las moléculas coexisten de manera equilibrada y finalmente la tercera sección de 62 a 80 donde la identidad de la base en la posición está definida. Los resultados son consistentes con el diseño de la libUABC, el cual presenta regiones fijas de 19 bases en los extremos de las moléculas y una región central aleatorizada de la posición 20 a la 61 (42 bases) donde se combinan las cuatro bases de manera proporcional.

La Figura R1A2-3B muestra la distribución de tamaño de las moléculas analizadas donde los resultados indican un pico de 80 pares de bases con una frecuencia mayor a 800,000 consistente con el número total de secuencias analizadas.

En la Figura R1A2-4A se indica el total de secuencias duplicadas, donde los resultados muestran que de las secuencias totales existe un 20% que muestra duplicación hasta 10 veces, un 57% hasta 50 veces y un 12% hasta 100 veces.

En la Figura R1A2-4B la cual identifica la presencia de adaptadores, observamos la ausencia de adaptadores.

Tabla 6. Resumen de todos los datos.	
Total de secuencias	820143
Secuencias de baja calidad	0
Longitud de las secuencias	80
%GC	49



Secuencia de calidad por mosaico de CPTPrimera ronda

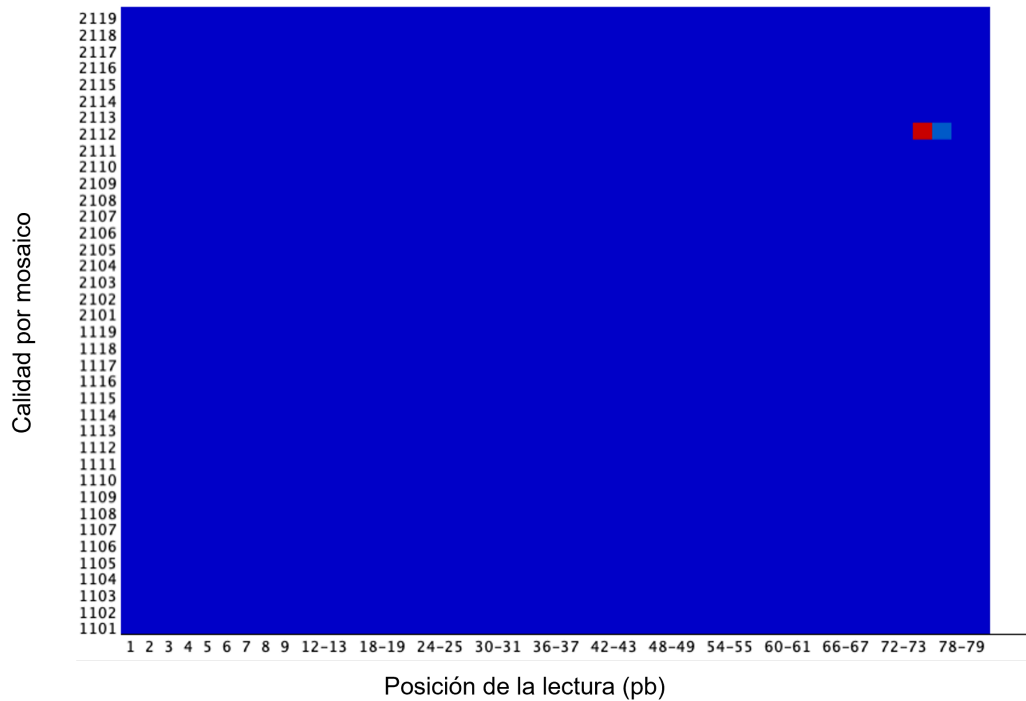


Figura R1A2-1. Calidad de secuencia por base y mosaico. A) Todas las secuencias poseen buena calidad. B) Celdas libres de contaminación.

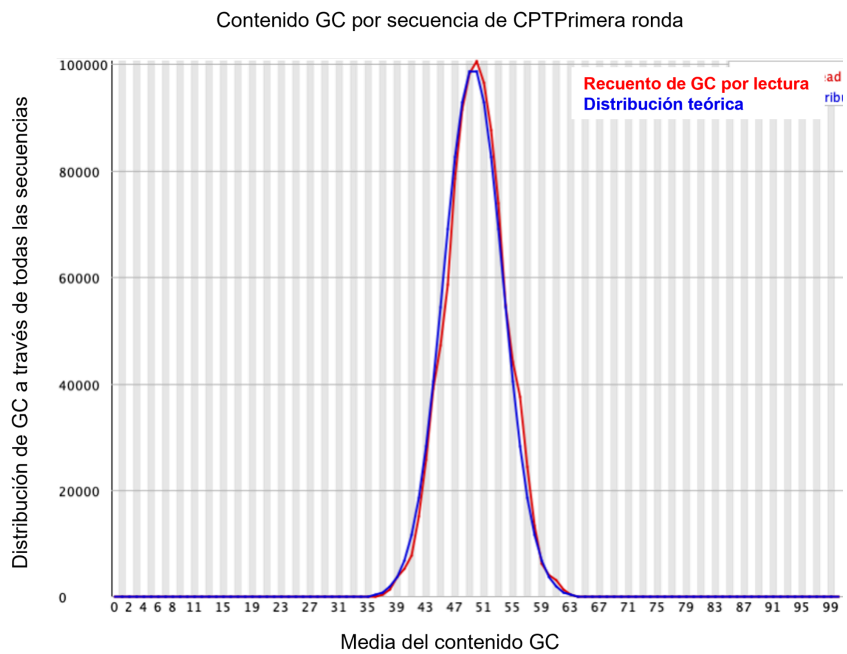
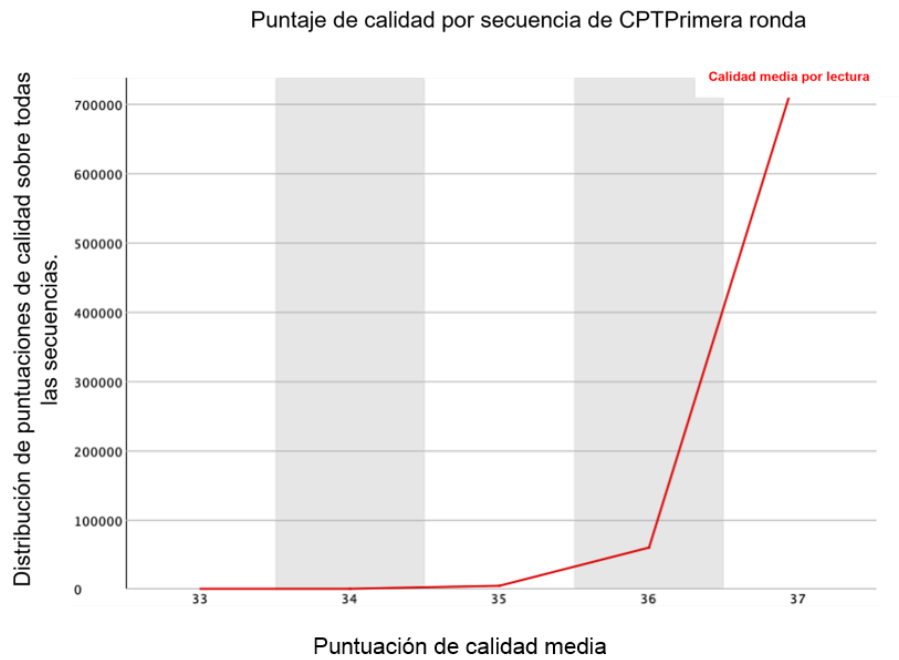


Figura R1A2-2. Puntaje de calidad por secuencia y GC A) Más de 700,000i poseen calidad de 37. B) Media de contenido GC 49.

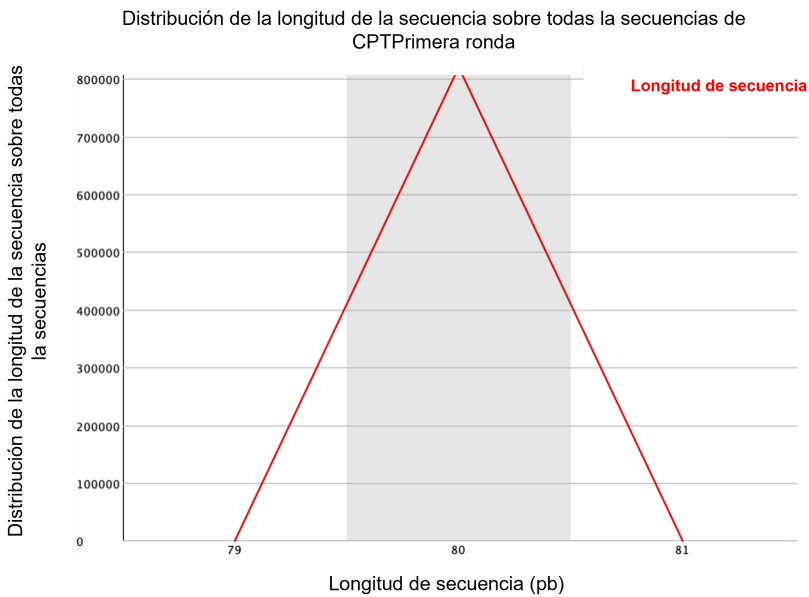
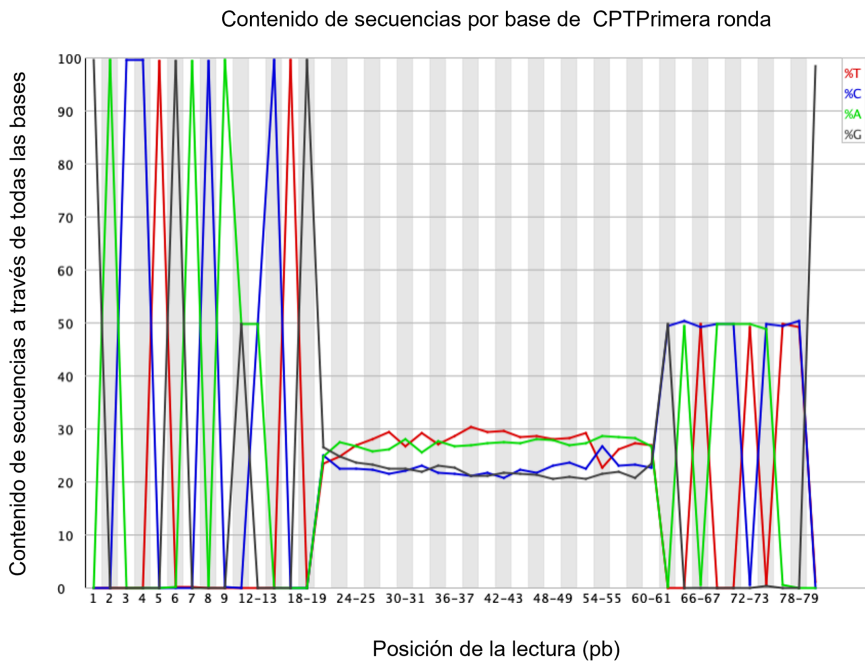


Figura R1A2-3. Contenido de secuencia por base y longitud de secuencia. A) Tres secciones bien definidas. B) Todas las secuencias poseen una longitud de 80 bases.

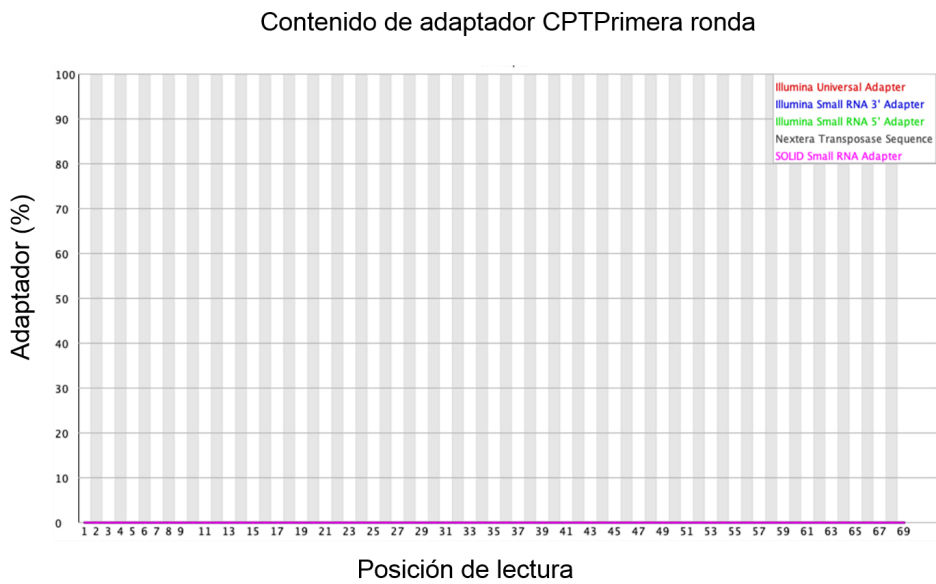
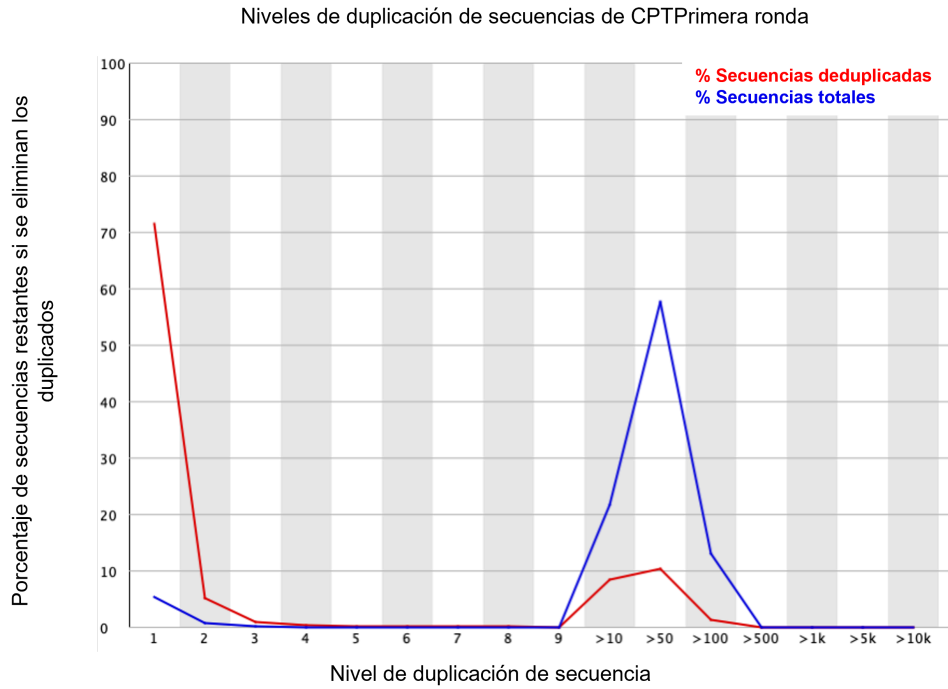


Figura R1A2-4. Niveles de duplicación de secuencias y adaptadores. A) Duplicación de las secuencias. B) Adaptadores ausentes.

La secuenciación de las moléculas de la libUABC después de la primera ronda SELEX, reveló la identidad nucleotídica de 2,381,547 secuencias. Después del filtrado de calidad, recuperamos un 34.43% (se eliminaron 65.56%) de secuencias identificadas con buena calidad, este porcentaje corresponde a 820,143 secuencias con buena calidad.

## 7.3. Secuenciación de Ronda 2

### 7.3.1 Análisis 1 de la Ronda 2

En la tabla 7 se muestran los datos globales del análisis por secuenciación del ADN recuperado de la segunda ronda de selección SELEX contra las proteínas virales, mismas que se han identificado como Ronda 2 (R2). El análisis identificó un total de 2,254,319 secuencias con buena calidad, 0 secuencias de mala calidad, las secuencias varían en longitud entre 35-151 bases y con un porcentaje de GC de 50%

En la Figura R2A1-1A, empleando el paquete bioinformático FastQC para visualizar los datos de secuenciación de nueva generación obtenidos de la plataforma illumina para la muestra R2, al analizar la calidad de las secuencias por posición de lectura, nos grafica que de la posición 1 a la 80 hay buena calidad, y los datos de la identidad de las bases de la posición 81 en adelante no poseen calidad. Estos datos, son consistentes con la muestra secuenciada, ya que los fragmentos analizados de interés tiene una longitud correspondientes a 80 bases, por lo cual la información de la posición 81 a 151 son extensiones de las secuencias analizadas, dichas bases no son de relevantes y serán eliminadas en el procesamiento bioinformático.

La Figura R2A1-1B análisis de la calidad de secuencia por mosaico, los resultados indican que de la posición 1 a la 80 los mosaicos son azules en su totalidad, lo que nos quiere indica que la calidad es buena, a partir de la posición 81 hay presencia de mosaicos más claros (azul claro), pero ausencia de mosaicos rojos o naranjas lo que nos señala que no hay mala calidad.

En la Figura R2A1-2A correspondiente al puntaje de calidad por secuencia, los datos muestran que existen más de 200,000 moléculas con una calidad de 36, mientras que más de 1.6 millones poseen calidad de 38. Los datos nos reflejan niveles de calidad altos en su totalidad.

La Figura R2A1-2B gráfica el contenido de GC, los resultados señalan una distribución donde la mayoría de las secuencias tiene un porcentaje de GC entre 35 y 63 por ciento, los datos reflejan una distribución bimodal, encontramos dos picos no simétricos de diferente altura. El análisis de los datos es consistente con una diversidad baja.

En la Figura R2A1-3A se muestra el contenido de bases por cada posición en la muestra analizada. En la gráfica, se identifican tres secciones bien definidas, de la posición 1 a 19 donde la identidad de la base es de 50%, 20-61 donde la identidad de las secuencias es de 25% para cada base aproximadamente y finalmente 62 en adelante donde la identidad de la base es de 50%. Los resultados corresponden con el diseño de la libUABC, el cual presenta una región fija de la base 1 a la 19, una región aleatorizada de 20 a la 61 donde de manera proporcional se combinan las cuatro bases y finalmente una región fija de la 61 a la 80.

La Figura R2A1-3B gráfica la distribución de tamaño de las moléculas analizadas. Los resultados indican un pico a 72-89 pares de bases con una frecuencia de 800,000 mil y un segundo pico a 145 con una baja frecuencia. Estos resultados son lo esperado, ya que las moléculas analizadas poseen un tamaño de 80 bases.

La Figura R2A1-4A indica el nivel de duplicación de secuencias de 2R, Los resultados grafican que de las moléculas totales existe un 4% de moléculas que son únicas, 5% de moléculas de más de 10 copias, 3% de moléculas con más de 100 copias y un 52% de moléculas con más de 5,000 copias. Demostrando un número reducido de moléculas únicas y un marcado incremento en las secuencias con más de 10, 100 y 5,000 duplicados.

La Figura R2A1-4B muestra el porcentaje de adaptadores en las moléculas, los datos de la gráfica señalan la presencia de adaptadores en las secuencias analizadas

en posiciones de lectura superiores a 80 bases, lo cual esclarece la información de secuencia de las moléculas de la posición 80.

Tabla 7. Resumen de todos los datos.	
Total de secuencias	2254319
Secuencias de baja calidad	0
Longitud de las secuencias	35-151
%GC	50

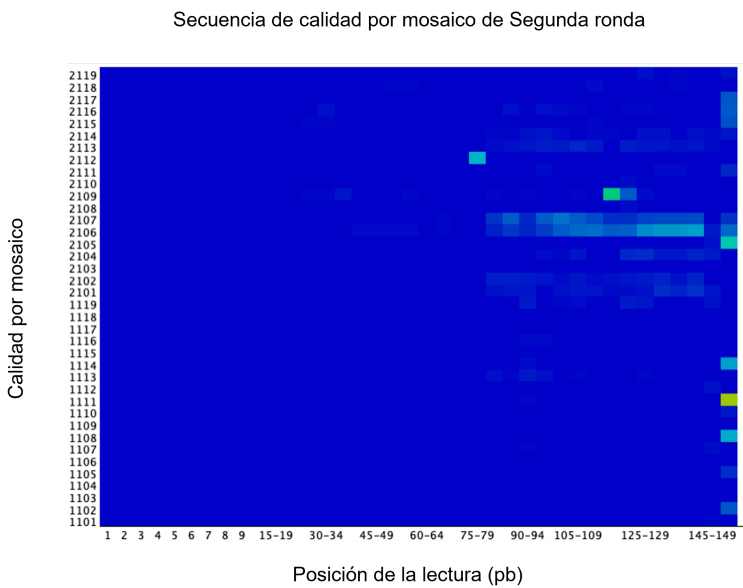
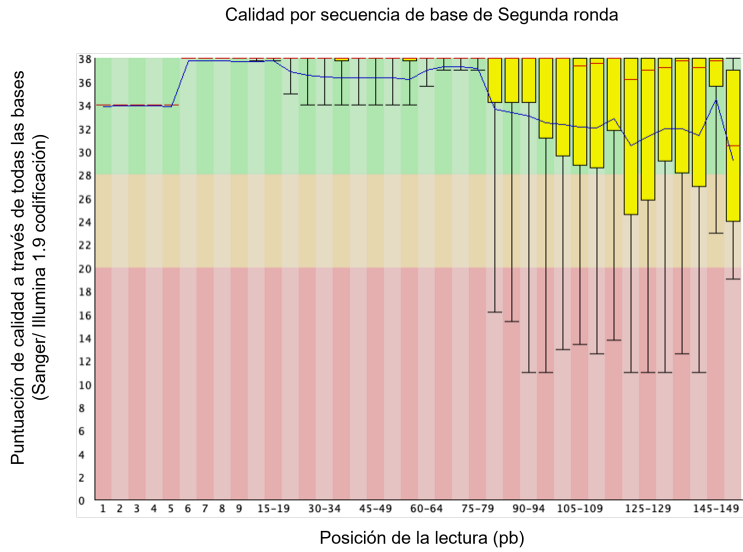


Figura R2A1-1. Calidad de secuencia por base y mosaico. A) De las secuencias 1 a la 80 hay presencia de buena calidad. B) Ausencia de celdas contaminadas en mosaicos, los mosaicos poseen buena calidad.

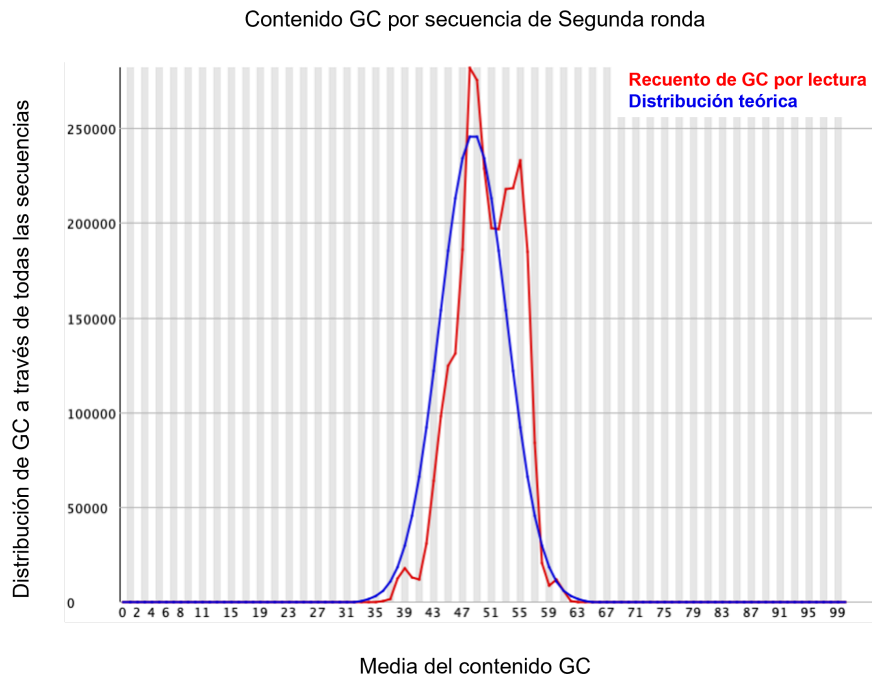
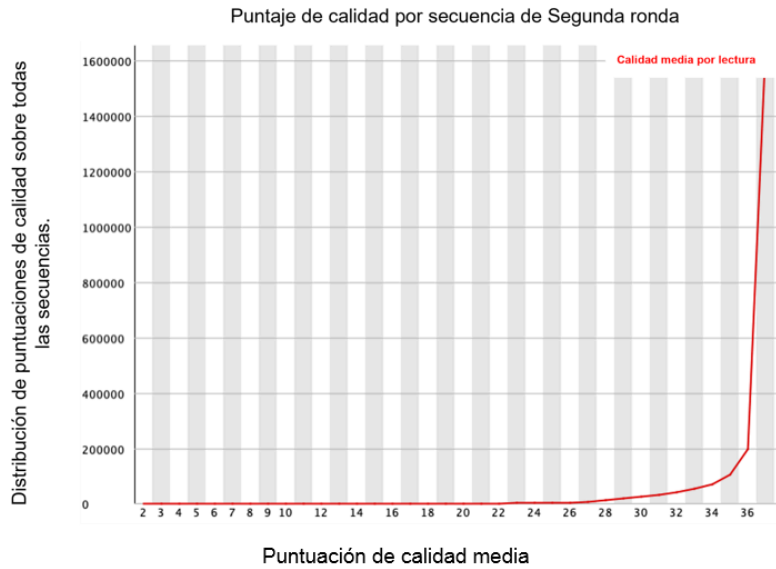


Figura R2A1-2. Puntaje de calidad por secuencia y GC. A) Más de 1,600,000 poseen una calidad superior a 36. B) La distribución es unimodal.

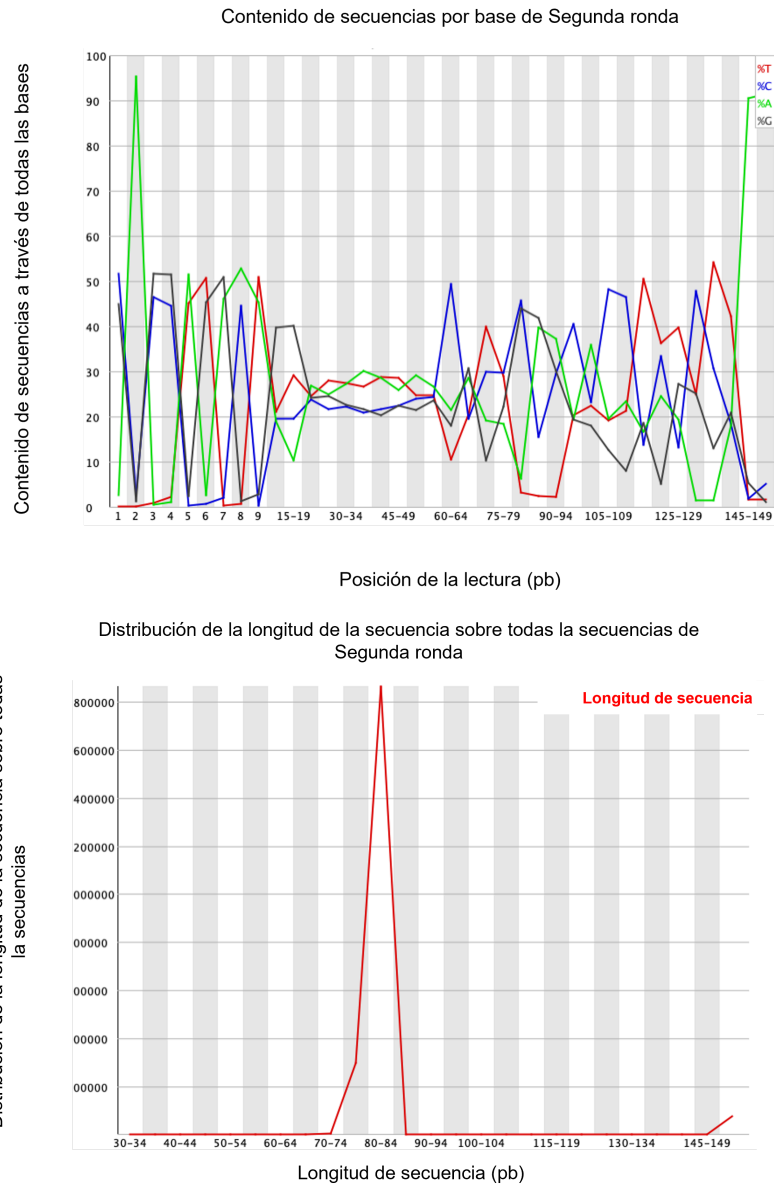


Figura R2A1-3. Contenido de secuencia por base y distribución de longitud. A) Se identifican tres secciones bien definidas. B) Más de 2,000,000 de secuencias presentan una longitud de 80.

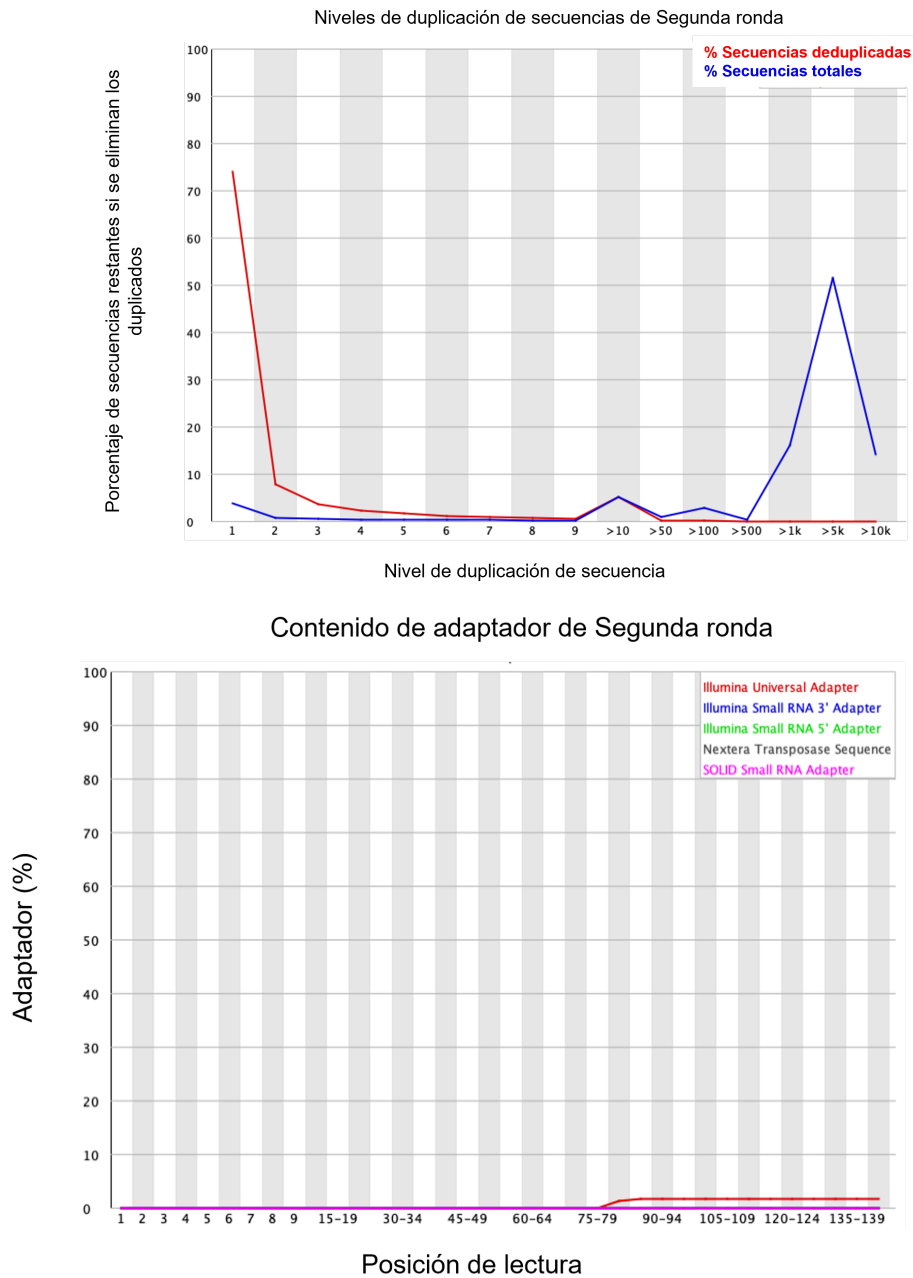


Figura R2A1-4. Análisis de nivel de duplicación y adaptadores. A) Duplicación de las secuencias. B) Presencia de adaptadores de la posición 81 en adelante.

### 7.3.2 Análisis 2 de la Ronda 2

En la tabla 8 se muestran los datos del análisis global por secuenciación de nueva generación en la plataforma Illumina del ADN recuperado de la ronda 2R después del SELEX contra las proteínas virales, y sometidos al procesamiento bioinformático antes descrito. El análisis recuperó 713,950 secuencias con buena calidad, 0 secuencias de mala calidad, las secuencias poseen una longitud de 80 bases y un porcentaje de GC de 50%.

La Figura R2A2-1A muestra el análisis de calidad donde las secuencias de la posición 1 a la 80 poseen puntuaciones arriba de 34, indicando que son de buena calidad, estas posiciones además presentan una mediana y promedio de buena calidad, señalando que el procesamiento de datos resultó exitoso al conservar sólo secuencias de buena calidad con una longitud de 80 bases.

La Figura R2A2-1B representa la secuencia de calidad por mosaico y muestra que la mayoría de los datos poseen buena calidad, ya que la mayoría de los mosaicos son de color azul, en la posición de lectura de las pares de base 74 a 76 se presenta un mosaico de color calido (rojo) y en la longitud 78 a 79 existe un mosaico de color frio (azul claro) dado que en la mayoría de los datos hay presencia de mosaicos color azul, los resultados se considera de buena calidad.

En la Figura R2A2-2A se señala el puntaje de calidad por secuencia, dichos datos muestran que existen 50,000 secuencias con una calidad de 36 y más de 600,000 con una calidad de 37.

En la Figura R2A2-2B los datos muestran una distribución donde la mayoría de las secuencias tiene un porcentaje de GC de 35 a 63%, con una distribución bimodal a 48 y 54. Este gráfico posee dos picos de diferente intensidad y proporción, los datos indican que el proceso de selección operó adecuadamente y después de él tenemos al menos dos grandes poblaciones de secuencias.

En la Figura R2V2-3A se gráfica el contenido de bases por cada posición en las secuencias, los resultados identifican tres secciones bien definidas en la gráfica, la

primera de la posición 1 a 19 con identidades bien definidas que alcanzan el 100 % de congruencia, la segunda de 20-61 en esta región la identidad de las posiciones varía entre 20 y 30% para cada base, mostrando que en esta sección de las moléculas la identidad es diversa y finalmente la tercera sección de 62 a 80 donde la identidad de la base en la posición está definida alcanza el 100%. Los resultados concuerdan con el diseño de la libUABC, el cual presenta 2 regiones fijas de 19 bases ubicadas en los extremos de las moléculas y una aleatorizada ubicada en el centro de la posición 20 a la 61.

La Figura R2A2-3B muestra la distribución de tamaño de las moléculas analizadas donde los resultados indican un pico de 80 pares de bases con una frecuencia mayor a 700,000 consistente con el número total de secuencias analizadas.

En la Figura R2A2-4A se indica el total de secuencias duplicadas, donde los resultados señalan que de las secuencias totales existe un 5% duplicado hasta 10 veces, un 30% hasta 1000 veces y un 55% hasta 5,000 veces, reflejando que el proceso de evolución se llevó acabo.

En la Figura R2A2-4B la cual identifica la presencia de adaptadores, señala la ausencia de adaptadores.

Tabla 8. Resumen de todos los datos.	
Total de secuencias	713950
Secuencias de baja calidad	0
Longitud de las secuencias	80
%GC	50

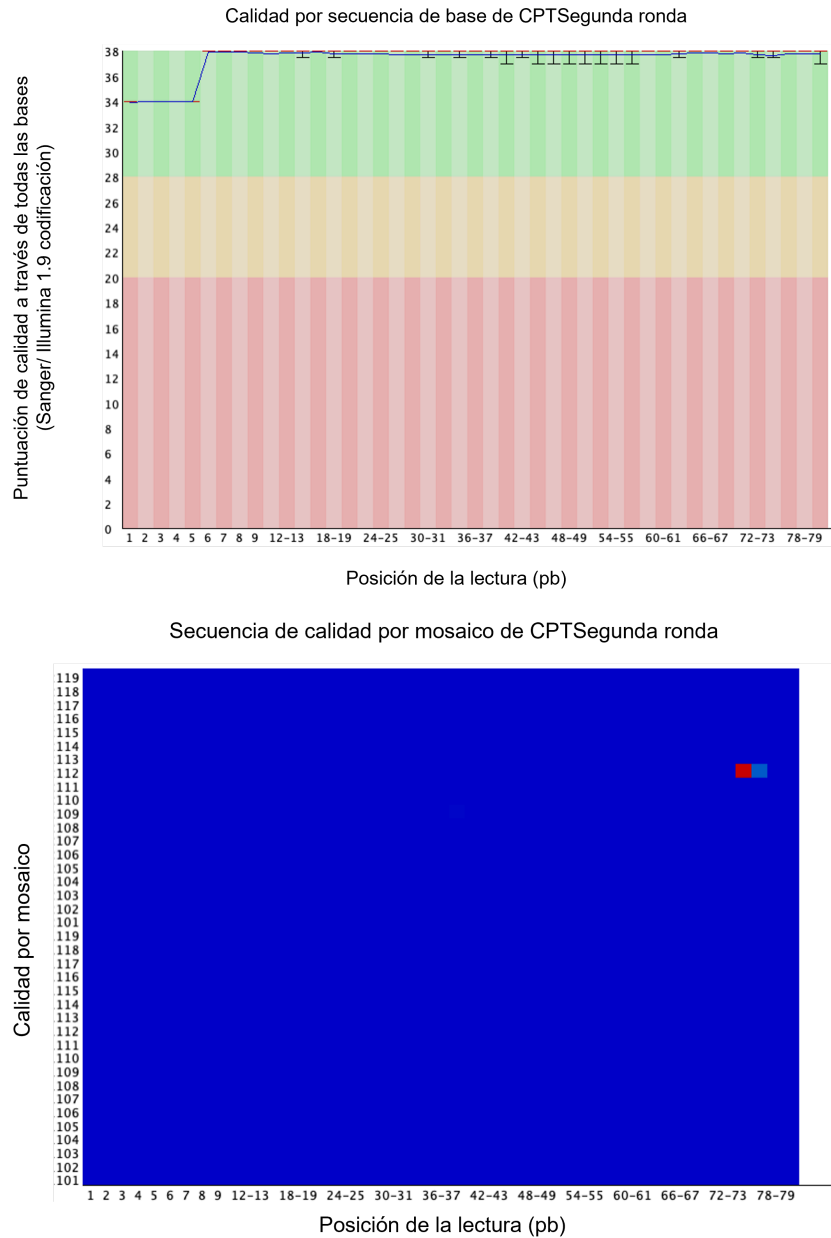


Figura R2A2-1. Calidad de secuencia por base y mosaico. A) Todas las secuencias poseen buena calidad. B) Celdas libres de contaminación resultados de buena calidad..

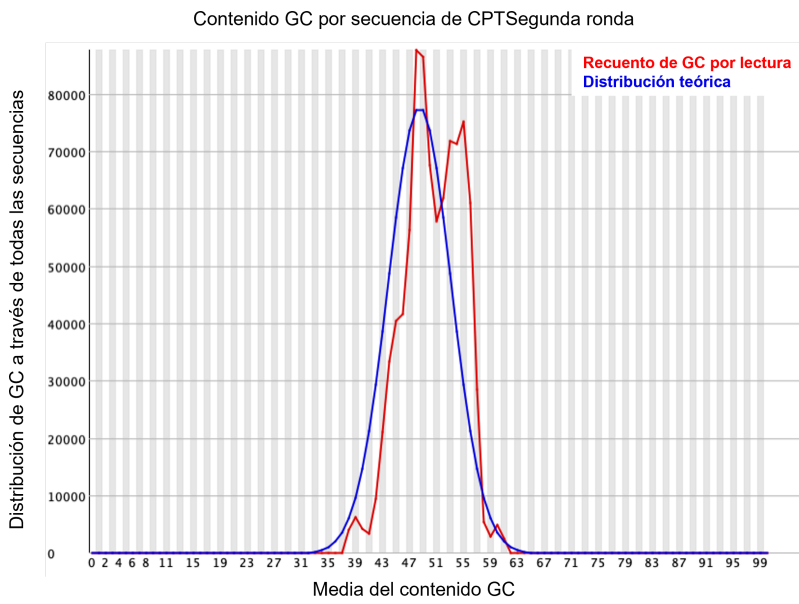
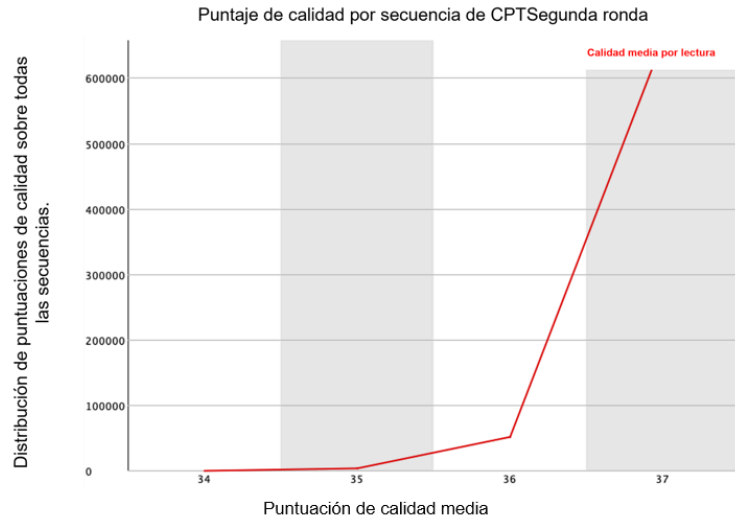
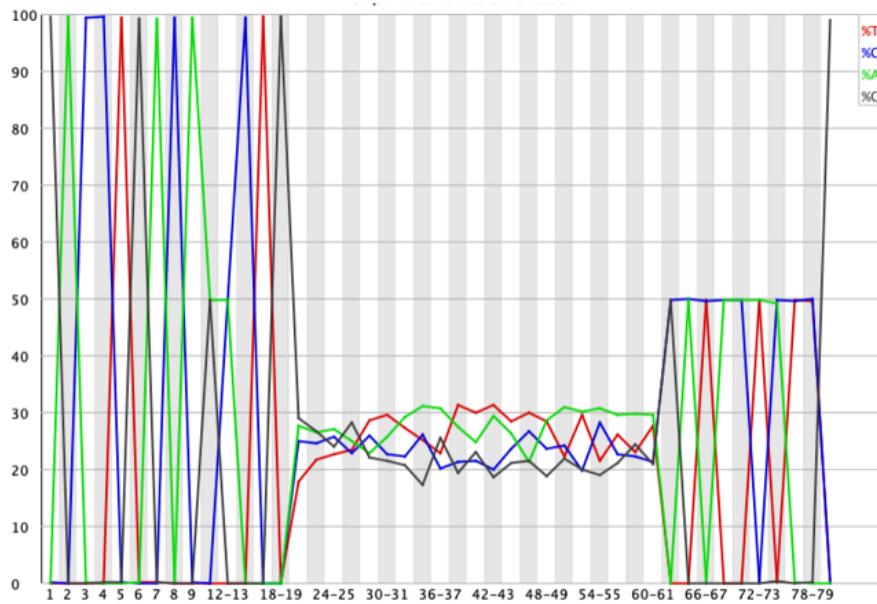


Figura R2A2-2. Puntaje de calidad por secuencia y GC A) Más de 600,000 poseen calidad de 37. B) Media de contenido GC 49.



Distribución de la longitud de la secuencia sobre todas la secuencias de CPTSegunda ronda

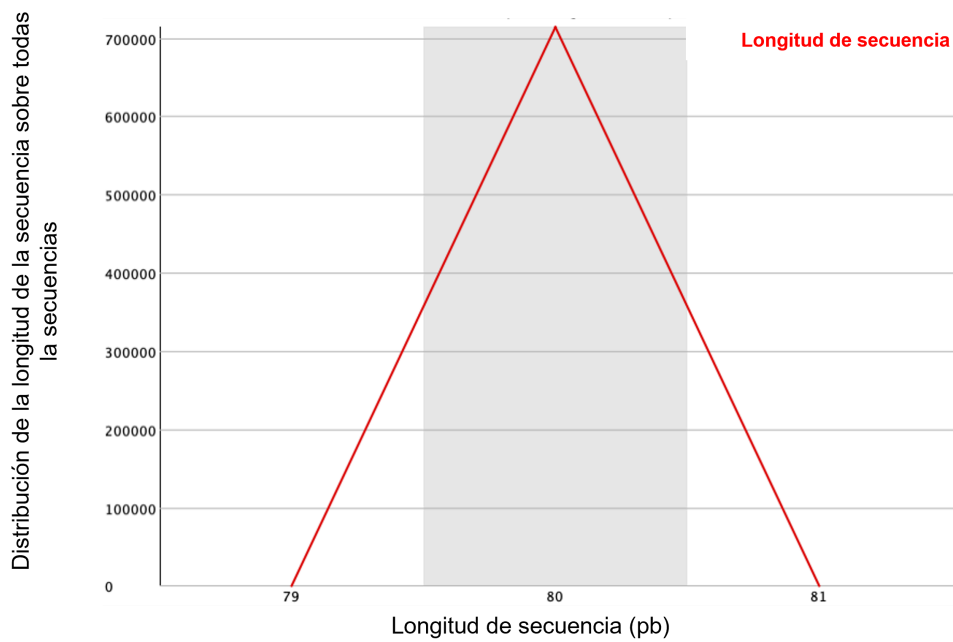


Figura R2A2-3. Contenido de secuencia por base y longitud de secuencia. A) Tres secciones bien definidas, 2 fijas y 1 aleatorizada. B) Todas las secuencias poseen una longitud de 80 bases.

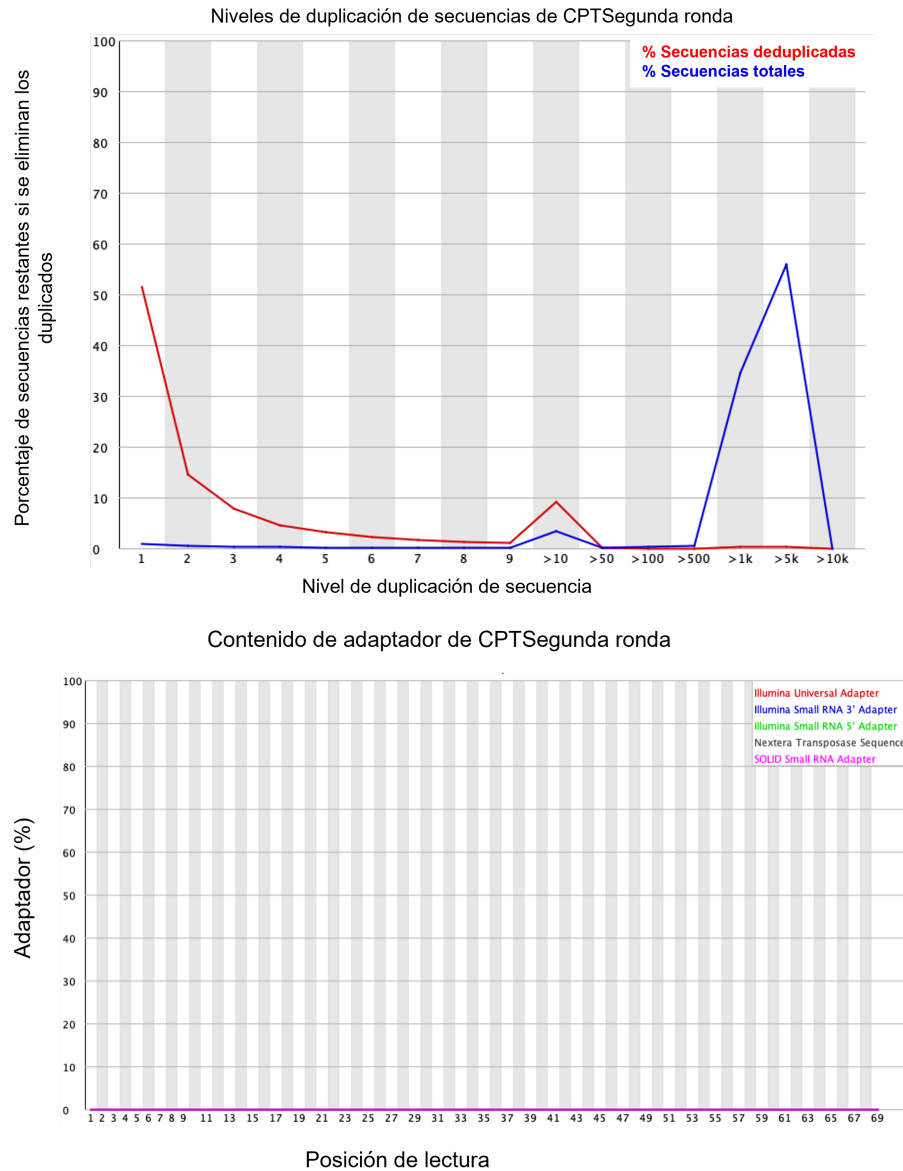


Figura R2A2-4. Niveles de duplicación de secuencias y adaptadores. A) Duplicación de las secuencias. B) Adaptadores ausentes

La secuenciación de las moléculas de la libUABC después de la segunda ronda SELEX, reveló la identidad nucleotídica de 2,254,319 secuencias. Después del filtrado de calidad, recuperamos un 31.67% (se eliminaron 68.32%) de secuencias

identificadas con buena calidad, este porcentaje corresponde a 713,950 secuencias con buena calidad.



## Capítulo 7

# Discusión

*“Sin memoria no existimos; sin responsabilidad quizá no merezcamos existir “*

*- José Saramago*

Se diseñó y evaluó una nueva biblioteca, denominada libUABC, para el aislamiento de aptámeros con 42 posiciones aleatorizadas, flanqueada por dos regiones físicas para su amplificación por PCR. Para el diseño de la biblioteca y de el par de oligonucleótidos para su amplificación, se emplearon herramientas bioinformáticas de análisis de estructura secundaria, con el objetivo de verificar que no se formaran estructuras que intervinieran con la amplificación. Las bibliotecas de ADN con alrededor de 40 posiciones aleatorizadas han sido empleadas previamente en la literatura calibrando el número de ciclos para evitar la formación de subproductos(26,27).

El análisis de las secuencias recuperadas del proceso SELEX, empleando proteínas S mutantes del SARS-CoV-2, demuestra que de cada ronda SELEX (Ronda 1 y Ronda 2) y muestra inicial (Ronda 0) se recuperaron más de 2.1 millones de secuencias. Los datos recuperados de cada ronda se procesaron para eliminar las secuencias de baja calidad, por lo cual se pasaron por tres filtros donde en cada uno de ellos se eliminaron secuencias de baja calidad, resultando en aproximadamente una eliminación del 68% de las secuencias. Patel y colaboradores reportaron una análisis de calidad de secuencias obtenidas en la misma plataforma de secuenciación, donde eliminaron un 45% de las secuencias obtenidas(28). La diferencia entre la cantidad de

secuencias eliminadas por baja calidad puede deberse a la naturaleza de nuestras secuencias. En el caso de la bibliotecas empleadas para el aislamiento de aptámeros son secuencias cortas regularmente de no más de 100 nucleótidos, al tratarse de secuencias cortas que cada paso durante el proceso de preparación de la muestra para secuenciación puede acarrear una serie de eventos que den como resultado una secuencia de baja calidad, como la incorporación doble de adaptadores para el chip. Sin embargo, este tipo de secuenciación ha sido empleado previamente para el aislamiento de aptámero con éxito(27).

Martínes-Roque y colaboradores aislaron un aptámero contra la proteína S de SARS-CoV-2 con una biblioteca de ADN 40 posiciones aleatorizadas, realizando dos rondas de selección y empleando electroforesis capilar como método de partición. Identificaron dos aptámeros capaces de unirse a la proteína S de SARS-CoV-2 con alta especificidad, la cuales posteriormente fueron incorporadas en dos biosensores uno electroquímico y otro fluorescente, obteniendo 0.07 fM y 41.87 nM respectivamente(27). Este reporte confirma que es factible aislar, aptámeros contra la proteína S de SARS-CoV-2 y que las moléculas aisladas son específicas.

El número de moléculas analizadas de cada muestra es consistente: aprox 2.1 millones, de cada uno de estos, después del filtrado de calidad, se recuperó aprox 30% de secuencias de buena calidad.

## Capítulo 8

# Conclusiones

*“Aprendí que no se puede dar marcha atrás, que la esencia de la vida es ir hacia adelante “*

*- Agatha Christie*

Se diseñó y evaluó una nueva biblioteca para el aislamiento de aptámeros con 42 posiciones aleatorizadas, la cual permitió el aislamiento de aptámeros para la proteína S de SARS-CoV-2, demostrando un proceso de enriquecimiento en el proceso SELEX empleado.

A través del procesamiento informático de las secuencias obtenidas por secuenciación de próxima generación fue posible manejar el gran número de datos obtenidos y reducir el número de moléculas a evaluar posteriormente.



## Capítulo 9

# Perspectivas

*“Felicidad es mirarse hacia dentro y sonreír al verse “*

*– Ron Lorent*

Como perspectiva, se plantea el análisis de las secuencias recuperadas para identificar las moléculas más enriquecidas. Este análisis permitirá proponer que aptámeros se deberían evaluar biológicamente. Mientras más se haya enriquecido la molécula, mientras más veces se haya repetido esa secuencia en la secuenciación implica que es más “apta” y por ende poseen una mayor probabilidad de ser un mejor aptámero. Al identificar las secuencias más aptas, se deberá proceder a verificar *in vitro* que el aptámero se una a la proteína S del SARS-CoV-2.

Como perspectiva a largo plazo y habiendo verificado que los aptámeros posean una buena afinidad y especificidad, estas moléculas podrán incorporarse como elementos de bioreconocimiento en biosensores para el diagnóstico de la infección, o por otra parte también se podrán verificar si estas moléculas poseen la capacidad de bloquear la interacción entre el SARS-CoV-2 y su receptor celular, la proteína ACE2, para funcionar como un elemento de protección contra la infección viral.



# Referencias Bibliográficas

1. Yang H, Rao Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nat Rev Microbiol*. 2021;19(11):685–700.
2. Reina J. El SARS-CoV-2, una nueva zoonosis pandémica que amenaza al mundo. *Vacunas*. 2020;21(1):17–22.
3. Zhong N, Zheng B, Li Y, Poon L, Xie Z, Chan K, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet*. 2003;362(9393):1353–8.
4. Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV — a target for vaccine and therapeutic development. *Nat Rev Microbiol*. 2009;7(3):226–36.
5. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol*. 2021;19(3):155–70.
6. Masters PS. The Molecular Biology of Coronaviruses. *Adv Virus Res*. 2006;66:193–292.
7. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020;367(6483):1260–3.
8. Vargas CM, Acosta RG, Bernilla AT. El nuevo Coronavirus y la pandemia del Covid-19. *Revista Medica Hered*. 2020;31(2):125–31.
9. Calvo C, López-Hortelano MG, Vicente JC de C, Martínez JLV, Sanidad G de trabajo de la AE de P para el brote de infección por C colaboradores con el Ministerio de, AEP M del G de E de la, et al. Recomendaciones sobre el manejo clínico de la infección por el «nuevo coronavirus» SARS-CoV2. Grupo de trabajo de la Asociación Española de Pediatría (AEP). *Anales De Pediatría*. 2020;92(4):241.e1-241.e11.
10. Santos-López G, Cortés-Hernández P, Vallejo-Ruiz V, Reyes-Leyva J. SARS-CoV-2: basic concepts, origin and treatment advances. *Gaceta De M Xico*. 2021;157(1):84–9.
11. Cabanillas ERL, Risco AOL, Risco KBL, Hoyos GLL, Zavaleta RML, Tirado E del RL, et al. Bases moleculares de la patogénesis de COVID-19 y estudios in silico de posibles tratamientos farmacológicos. *Revista De La Fac De Medicina Humana*. 2021;21(2):417–32.
12. Ramiro-Mendoza MS. Epidemiología del SARS-CoV-2. *Acta Pediatr Mex*. 2020;1(41):8–14.
13. Diseases CNCRC for R, China NC for CH Beijing, Association G of R Chinese Pediatric Society, Chinese Medical, Pediatrics CMDAC on R, Pediatrics CMEAC on, Pediatrics CRHAC on, et al. Diagnosis, treatment, and prevention of 2019 novel coronavirus infection in children: experts' consensus statement. *World J Pediatr*.

2020;16(3):223–31.

14. Vargas MR, Melendrez-Arango EC, Durán-Aguirre ML, Lucero LQ, Peralta-Peña SL. Tratamiento farmacológico en pacientes con COVID-19: una revisión integradora.

Sanus. 2021;6:e250.

15. Camacho YTC, Benavides VK, Rivadeneira KGC, Cárdenas KRH, Ramírez WGL, Heredia EWB, et al. COVID-19 y su impacto en la odontología. Rev Estomatológica Hered. 2021;31(3):199–207.

16. Leija-Montoya AG, Benítez-Hess ML, Toscano-Garibay JD, Alvarez-Salas LM. Characterization of an RNA aptamer against HPV-16 L1 virus-like particles. Nucleic Acid Therapeutics (Formerly Oligonucleotides). 2014 Oct;24(5):344–55.

17. Toscano-Garibay JD, Benítez-Hess ML, Alvarez-Salas LM. Isolation and Characterization of an RNA Aptamer for the HPV-16 E7 Oncoprotein. Archives of Medical Research. 2011 Feb;42(2):88–96.

18. Nakamura Y. Aptamer: Biology to Applications. In Springer Berlin Heidelberg; 2011. p. 135–52.

19. Ospina JD. Los aptámeros como novedosa herramienta diagnóstica y terapéutica y su potencial uso en parasitología. Biomédica. 2020;40(Supl. 1):148–65.

20. Bruno JG, Kiel JL. In vitro selection of DNA aptamers to anthrax spores with electrochemiluminescence detection. Biosensors and Bioelectronics [Internet]. 1999 May 31;14(5):457–64. Available from:

<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=10451913&retmode=ref&cmd=prlinks>

21. Rubio S, Pacheco-Orozco RA, Gómez AM, Perdomo S, García-Robles R. Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. Univ Médica. 2020;61(2).

22. Ruckman J, Green LS, Beeson J, Waugh S, Gillette WL, Henninger DD, et al. 2'-Fluoropyrimidine RNA-based Aptamers to the 165-Amino Acid Form of Vascular Endothelial Growth Factor (VEGF165): INHIBITION OF RECEPTOR BINDING AND VEGF-INDUCED VASCULAR PERMEABILITY THROUGH INTERACTIONS REQUIRING THE EXON 7-ENCODED DOMAIN. Journal of Biological Chemistry. 1998 Aug 7;273(32):20556–67.

23. Gilead. Macugen. Gilead. 2008 Apr 11;1–2.

24. Gokulrangan G, Unruh JR, Holub DF, Ingram B, Johnson CK, Wilson GS. DNA Aptamer-Based Bioanalysis of IgE by Fluorescence Anisotropy. Anal Chem. 2005;77(7):1963–70.

25. Owczarzy R, Tataurov AV, Wu Y, Manthey JA, McQuisten KA, Almabrazi HG, et al. IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic Acids Res. 2008;36(suppl\_2):W163–9.

26. Musheev MU, Krylov SN. Selection of aptamers by systematic evolution of ligands by exponential enrichment: Addressing the polymerase chain reaction issue. *Analytica Chimica Acta*. 2006 Mar;564(1):91–6.
27. Martínez-Roque MA, Franco-Urquijo PA, García-Velásquez VM, Choukeife M, Mayer G, Molina-Ramírez SR, et al. DNA aptamer selection for SARS-CoV-2 spike glycoprotein detection. *Anal Biochem*. 2022;645:114633.
28. Patel RK, Jain M. NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *PLoS ONE*. 2012;7(2):e30619.