



UNIVERSIDAD  
DE MÁLAGA



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA  
INGENIERÍA DE LA SALUD

**ANÁLISIS DE GENOMAS COMPLETOS DE  
ESPECTROS DE MUTANTES DE SARS-CoV-2 A  
PARTIR DE DATOS DE SECUENCIACIÓN MASIVA**

**WHOLE GENOME ANALYSIS OF SARS-COV-2  
MUTANT SPECTRA FROM MASSIVE SEQUENCING  
DATA**

Realizado por  
**María Sánchez Pando**

Tutorizado por  
**Ana Grande Pérez**

Departamento  
**Biología Celular, Genética y Fisiología**

UNIVERSIDAD DE MÁLAGA  
MÁLAGA, SEPTIEMBRE 2023

Fecha entrega: 6 de septiembre de 2023



UNIVERSIDAD  
DE MÁLAGA



# Resumen

El coronavirus 2 del síndrome respiratorio agudo severo (SARS-CoV-2) es un virus RNA que existe en sus hospedadores como variantes víricas denominados cuasiespecies, cuya media se denomina secuencia consenso. La identificación y el estudio de mutaciones de las secuencias consenso ha permitido el seguimiento epidemiológico del virus, y ha asistido al desarrollo de vacunas. Sin embargo, las mutaciones menos frecuentes, no representadas en las secuencias consenso pero sí en las cuasiespecies, pueden ser determinantes en la evolución, emergencia y patología del virus, así como en el desarrollo de tratamientos antivirales. Los estudios de cuasiespecies de SARS-CoV-2 a partir de datos de secuenciación masiva de genoma completo han sido muy escasos. Por ello, en mi Trabajo de Fin de Grado he realizado un estudio de las cuasiespecies de genoma completo a partir de muestras procedentes de un proyecto centrado en el desarrollo de una terapia antiviral de combinación con péptidos inhibidores de proteínas virales y mutagénesis letal contra SARS-CoV-2. Las muestras se amplificaron por RT-PCR mediante amplicones solapantes y se secuenciaron por secuenciación masiva (NGS), por Illumina. Mediante el flujo de trabajo QuasiFlow, se analizó la complejidad y heterogeneidad de las cuasiespecies en cada muestra para entender el efecto de los antivirales en la estructura genética de las cuasiespecies. Los análisis bioinformáticos revelaron que el análogo y los péptidos, por separado y en combinación, aumentaron la variabilidad y heterogeneidad genética en las cuasiespecies. Sin embargo, las secuencias consenso por lo general permanecieron invariables, salvo en ciertas muestras donde se observaron mutaciones en el extremo 5' y en la ORF1a. Estos resultados resaltaron la relevancia de estudiar los cambios a nivel de cuasiespecies, identificando mutaciones indetectables analizando las secuencias consenso, pero que juegan un papel relevante en la patogénesis y evolución del SARS-CoV-2 bajo presión de selección mediada por antivirales.

**Palabras clave:**

SARS-CoV-2, cuasiespecie, secuenciación masiva, Quasiflow, ensayo antiviral.

# Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an RNA virus that exists in its hosts as viric variants called quasispecies, whose mean is called consensus sequence. The identification and study of mutations of the consensus sequences has allowed the epidemiological monitoring of the virus, and has assisted the development of vaccines. However, less frequent mutations, not represented in the consensus sequences but in the quasispecies, may be decisive in the evolution, emergence and pathology of the virus, as well as in the development of antiviral treatments. Studies of SARS-CoV-2 quasispecies from massive whole-genome sequencing data have been very scarce. Therefore, in my Final Degree Project I have carried out a study of genome-wide quasispecies from samples from a project focused on the development of an antiviral therapy in combination with peptides inhibitors of viral proteins and lethal mutagenesis against SARS-CoV-2. The samples were amplified by RT-PCR by overlapping amplicons and sequenced by massive sequencing (NGS), by Illumina. Using the QuasiFlow workflow, the complexity and heterogeneity of the quasispecies in each sample was analyzed to understand the effect of antivirals on the genetic structure of the quasispecies. Bioinformatics analyses revealed that the analogue and peptides, separately and in combination, increased genetic variability and heterogeneity in quasispecies. However, the consensus sequences generally remained unchanged, except in certain samples where mutations were observed in the external 5' and ORF1a. These results highlighted the relevance of studying changes at the quasispecies level, identifying undetectable mutations by analyzing consensus sequences, but which play a relevant role in the pathogenesis and evolution of SARS-CoV-2 under antiviral-mediated selection pressure.

## **Keywords:**

SARS\_CoV\_2, quasispecies, Massive sequencing, Quasiflow, antiviral assay.



# Índice

<b>Introducción</b> .....	<b>1</b>
1.1. Motivación .....	1
1.2. Objetivos.....	2
1.3. Estructura de la memoria .....	3
1.4. Tecnologías empleadas. ....	3
1.5. SARS-CoV-2 .....	4
1.5.1. Estructura genética .....	5
1.5.2. Organización en cuasiespecies .....	6
1.6. Mutagénesis letal como estrategia antiviral.....	8
<b>Material y métodos</b> .....	<b>9</b>
2.1. Muestras .....	9
2.1. Material de laboratorio.....	12
2.2. Amplificación del genoma.....	13
2.2.1. Electroforesis en gel de agarosa .....	17
2.2.2. Visualización de DNA .....	19
2.3. Control del experimento .....	19
2.4. Preparación de pools.....	21
2.5. Análisis bioinformático.....	22
2.5.1. Datos de secuenciación masiva .....	22
2.5.2. Secuencia de Wuhan.....	24
2.5.3. Secuencias consenso .....	25
2.5.3.1. Alineamientos de secuencias .....	26
2.5.3.2. Análisis filogenético .....	28
2.5.4. Cuasiespecies.....	29
2.5.4.1. Análisis de composición .....	30
2.5.4.2. Análisis de variabilidad y heterogeneidad .....	31
2.5.4.2.1. Frecuencia de mutación .....	31
2.5.4.2.2. Entropía de Shannon .....	32
2.5.4.2.3. Diversidad nucleotídica .....	32
<b>Resultados</b> .....	<b>33</b>
3.1. Amplificación del genoma.....	33
3.2. Análisis bioinformático.....	35
3.2.1. Secuencias consenso .....	35
3.2.1.1. Alineamiento de secuencias .....	36

3.2.1.2. Análisis filogenético .....	37
3.2.2. Análisis de cuasiespecie .....	39
3.2.2.1. Análisis de composición.....	39
3.2.2.2. Análisis de variabilidad y heterogeneidad.....	41
<b>Discusión y líneas futuras.....</b>	<b>47</b>
4.1. Análisis filogenético .....	48
4.2. Análisis de composición .....	49
4.3. Análisis de variabilidad y heterogeneidad .....	49
<b>Conclusión.....</b>	<b>53</b>
<b>Referencias.....</b>	<b>55</b>
<b>Apéndice A .....</b>	<b>61</b>
A1. Tablas de sesgos de nucleótidos .....	61
A2. Tablas del análisis de variabilidad y heterogeneidad .....	64
<b>Apéndice B .....</b>	<b>67</b>
B.1. Imágenes geles .....	67

# Índice de figuras

<b>Figura 1.</b> Estructura y genoma del SARS-CoV-2.....	<b>6</b>
<b>Figura 2.</b> Gráfico que representa el protocolo RT-PCR.....	<b>15</b>
<b>Figura 3.</b> Material necesario para electroforesis en gel de agarosa .....	<b>18</b>
<b>Figura 4.</b> Marcador de masa molecular. ....	<b>19</b>
<b>Figura 5.</b> Flujo de trabajo del Quasiflow.....	<b>24</b>
<b>Figura 6.</b> Fichero .vcf en el que aparecen las mutaciones SNP presentes en la muestra A1.....	<b>31</b>
<b>Figura 7.</b> Visualización de gel .....	<b>33</b>
<b>Figura 8.</b> Visualización de gel de los pools 1 .....	<b>34</b>
<b>Figura 9.</b> Visualización de geles de los pools 2 .....	<b>35</b>
<b>Figura 10.</b> Imagen de NextClade .....	<b>35</b>
<b>Figura 11.</b> Ejemplo de la interfaz de salida del software JalView .....	<b>36</b>
<b>Figura 12.</b> Ejemplo de la interfaz de salida del software MEGA11 .....	<b>39</b>
<b>Figura 13.</b> Árbol filogenético de máxima verosimilitud con Mega11 .....	<b>38</b>
<b>Figura 14.</b> Gráficas representativas del porcentaje de mutaciones en cada ensayo antiviral.....	<b>40</b>
<b>Figura 15.</b> Análisis de variabilidad del SARS-CoV-2 en los ensayos antivirales ..	<b>43</b>
<b>Figura 16.</b> Gráficas de frecuencias de mutación en muestras sin péptido .....	<b>44</b>
<b>Figura 17.</b> Gráficas de frecuencias de mutación en muestras con y sin péptido tratadas con diferentes concentraciones de análogo. ....	<b>45</b>
<b>Figura B1.</b> ....	<b>67</b>
<b>Figura B1. 2.</b> ....	<b>67</b>
<b>Figura B1. 3.</b> ....	<b>68</b>
<b>Figura B1. 4.</b> ....	<b>68</b>
<b>Figura B1. 5.</b> ....	<b>68</b>
<b>Figura B1. 6.</b> ....	<b>69</b>
<b>Figura B1. 7</b> .....	<b>69</b>
<b>Figura B1. 8.</b> .....	<b>69</b>

**Figura B1. 9..... 70**  
**Figura B1. 10..... 70**  
**Figura B1. 11..... 70**  
**Figura B1. 12..... 71**  
**Figura B1. 13..... 71**  
**Figura B1. 14..... 71**  
**Figura B1. 15..... 72**  
**Figura B1. 16..... 72**  
**Figura B1. 17..... 72**  
**Figura B1. 18..... 73**  
**Figura B1. 19..... 73**  
**Figura B1. 20..... 73**  
**Figura B1. 21..... 74**  
**Figura B1. 22..... 74**  
**Figura B1. 23..... 74**  
**Figura B1. 24..... 75**  
**Figura B1. 25..... 75**

# Índice de tablas

<b>Tabla 1.</b> Tecnologías usadas para el desarrollo del proyecto .....	<b>4</b>
<b>Tabla 2.</b> Lista de variantes linaje pango y la fecha de designación .....	<b>5</b>
<b>Tabla 3.</b> Listado de muestras usadas en las amplificaciones. Todas ellas son de RNA extracelular .....	<b>11</b>
<b>Tabla 4.</b> Material necesario de laboratorio para desarrollar el proyecto. ....	<b>12</b>
<b>Tabla 5.</b> Reactivos y cantidades RT-PCR un solo paso.....	<b>15</b>
<b>Tabla 6.</b> Programa RT-PCR de un solo paso.....	<b>16</b>
<b>Tabla 7.</b> Listado de cebadores que compone el genoma del SARS-Cov-2. ....	<b>16</b>
<b>Tabla 8.</b> Principales programas que conforman el flujo de trabajo del QuasiFlow junto con una breve explicación de la función que desempeñan .....	<b>23</b>
<b>Tabla 9:</b> Mutaciones presentes en la secuencia consenso, con respecto a la secuencia de referencia de Ómicron del ensayo.....	<b>37</b>
<b>Tabla A1.1</b> Tablas de sesgos de nucleótidos de las muestras tratadas sin péptido .....	<b>61</b>
<b>Tabla A1.2</b> Tablas de sesgos de nucleótidos de las muestras tratadas con péptido #14. ....	<b>62</b>
<b>Tabla A1.3</b> Tablas de sesgos de nucleótidos de las muestras tratadas con péptido #16. ....	<b>63</b>
<b>Tabla A2.1</b> Tabla de análisis de varibilidad y heterogeneidad.....	<b>64</b>
<b>Tabla A2.2</b> Tabla de frecuencia de mutación.....	<b>65</b>



# Agradecimientos

En primero lugar, quería agradecer especialmente a mi tutora la Dra. Ana Grande Pérez por haber confiado en mí, haberme enseñado tanto durante estos meses y haberme guiado en este proyecto y a Sergio Ortega Del Campo por su apoyo, su ayuda, su dedicación, su infinita paciencia, por todo el tiempo dedicado a que este trabajo saliera adelante y por todo lo que me has enseñado. GRACIAS

Gracias a Pepi y a Luis Díaz Martínez por ayudarnos tanto en la secuenciación como en la ejecución del flujo de trabajo Quasiflow.

Gracias a mi compañera María por el apoyo durante estos meses, las horas dedicadas juntas y su amistad y generosidad.

Gracias a mis compañeros de batallas Mario, Cristina y Rafa. Desde primero hasta ahora de la mano. Por haberme hecho crecer como personas y como estudiante durante todos estos años.

Decir gracias a mi familia por su apoyo, comprensión, confianza plena y amor.

Gracias a amigos por todo el apoyo durante todos estos años, por las risas, por lo llantos, por lo buenos y malos momentos, siempre habéis estado ahí, no hubiera llegado a donde estoy ahora mismo sin ninguno de vosotros.



# 1

## Introducción

### 1.1. Motivación

El genoma del SARS-COV-2 consta de unos 30 mil nucleótidos (30 Kb), siendo unas tres veces más largo que la media de los genomas de los virus RNA. Este virus, al igual que otros virus de RNA, se organiza en cuasiespecies. Las cuasiespecies virales son distribuciones dinámicas de genomas virales mutantes y recombinantes no idénticos pero estrechamente relacionados, sometidos a un proceso continuo de variación genética, competencia y selección, y que actúan como una unidad de selección (Domingo et al., 2006). Actualmente, los estudios de cuasiespecies de SARS-CoV-2 con datos de secuenciación masiva (NGS) se centran en regiones concretas (o amplicones) y no en el genoma completo. Esta limitación viene dada por la falta de herramientas bioinformáticas para ello. Sin embargo, el grupo de la Dra. Ana Grande Pérez en colaboración con el del Dr. Gonzalo Claros (ambos de la Facultad de Ciencias de la Universidad de Málaga) ha desarrollado la herramienta bioinformática QuasiFlow que permitiría el análisis de cuasiespecies de genomas completos a partir de datos de NGS (Seoane et al., 2022). Así, el laboratorio de la Dra. Grande Pérez ha conseguido analizar mediante esta herramienta espectros de mutantes de genoma completo de un virus de DNA de cadena sencilla de 3 kb

(Seoane et al., 2022), y de un virus RNA con un genoma bisegmentado de unos 6 Kb (Ortega del Campo et al., 2023) y lo está poniendo a punto para el SARS-CoV-2.

Dado que en el laboratorio de la Dra. Grande-Pérez se están realizando estudios sobre una terapia antiviral contra SARS-CoV-2 basada en la mutagénesis letal, surge la necesidad de realizar amplificación y secuenciación de genomas completos de las variantes que componen la cuasiespecie. Ello permitiría hacer un análisis profundo de la variabilidad y evolución del SARS-COV-2 a nivel de variantes intrahospedador de genoma completo lo que permitirá evaluar dicha terapia antiviral.

Los estudios que se llevarán a cabo a lo largo de desarrollo del TFG están enmarcados aquellos que está elaborando en su tesis doctoral Sergio Ortega del Campo bajo la dirección de la Dra. Grande Pérez, en el que estudia la variabilidad genética de las cuasiespecies del SARS- CoV-2.

## 1.2. Objetivos

Este proyecto tiene como objetivo principal la elaboración de un análisis de la variabilidad y evolución del SARS-CoV-2 analizando genoma completo tanto de las secuencias consenso como de las variantes presentes en las cuasiespecies. Dicho objetivo general se ha abordado a través de los siguientes objetivos parciales:

- 1) **Amplificación** del genoma, secuenciación por NGS y ensamblaje del genoma completo.
- 2) **Obtención** de secuencias consenso.
- 3) Determinar la **organización en cuasiespecies** de las poblaciones de SARS-CoV-2 mediante el análisis de la complejidad y heterogeneidad genética tras la secuenciación del genoma y el análisis bioinformático del mismo, a partir de las muestras víricas.
- 4) **Análisis de la variabilidad y diversidad** genética a nivel de genoma completo y comparación con los resultados de otros estudios que se han centrado en regiones concretas del genoma.

### **1.3. Estructura de la memoria**

Al tratarse de un trabajo experimental, y, por lo tanto, no solo tener desarrollo en bioinformática sino también desarrollo en laboratorio, se ha optado por dividir la memoria del proyecto en cuatro apartados claramente diferenciados.

El primer apartado tiene carácter introductorio, con el fin de contextualizar el estudio que se va a desarrollar posteriormente. Además, se detallarán los objetivos y la motivación del desarrollo del estudio que promueve dicho trabajo.

En el segundo apartado, denominado “Material y métodos, propio de artículos científicos con el objetivo de que sirva como manual a otros investigadores para que puedan reproducirlo (Argimón Pallàs & Jiménez Villa, 1998). En él se hace una descripción detallada de cuáles han sido los procesos, la metodología, el tratamiento de los datos y el material que se ha empleado en esta investigación para la obtención de los resultados.

El tercer apartado expone los resultados obtenidos al haber aplicado los procesos mencionados en el apartado anterior. Se ha estructurado de la misma manera que “Material y métodos”, pero presentando y comentando los resultados obtenidos en cada uno de ellos (Jiménez Villa & Argimón Pallàs, 1998).

Como cuarto y último apartado, se presenta una discusión de los resultados en la que se comenta si dichos resultados son coherentes con la teoría, y se relacionan las conclusiones del trabajo con relación a los objetivos detallados en el primer apartado del documento. Además, se exponen posibles futuras líneas que podrían derivar de este trabajo.

### **1.4. Tecnologías empleadas.**

A continuación, se expone una tabla en la que se indican cada uno de los software usados para el desarrollo de este Proyecto Fin de Grado junto con el acceso de su web. Cada uno de ellos se encuentra explicado en el apartado material y métodos en los subapartados donde hayan sido usados.

**Tabla 1.** Tecnologías usadas para el desarrollo del proyecto

<b>Software</b>	<b>Acceso</b>
<i>Mega11</i>	<a href="https://www.megasoftware.net">https://www.megasoftware.net</a>
<i>JalView</i> v2.11.2.6	<a href="https://www.jalview.org">https://www.jalview.org</a>
<i>IGV_2.16.2</i>	<a href="https://igv.org">https://igv.org</a>
<i>NextClade</i>	<a href="https://clades.nextstrain.org">https://clades.nextstrain.org</a>
<i>QuasiFlow</i>	
<i>VirtualBox</i>	<a href="https://www.virtualbox.org/wiki/Documentation">https://www.virtualbox.org/wiki/Documentation</a>
<i>Blast</i>	<a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a>

Además de los software mencionados, cabe mencionar el uso del superordenador Picasso, para poder ejecutar programas como el QuasiFlow y el uso de bases de datos como GenBank (*GenBank Overview*, s. f.) y el NCBI entre otros para la búsqueda y comparación de secuencias.

### 1.5. SARS-CoV-2

En los últimos años, el SARS-CoV-2 ha sido tema de actualidad tanto para la sociedad como en la ciencia pasando en pocos meses de ser un completo desconocido a ser objeto de estudio de muchos investigadores en todo el mundo.

El SARS-CoV-2 fue aislado por primera vez en diciembre de 2019 en Wuhan, China. Gracias a los estudios, podemos afirmar que se trata de un betacoronavirus envuelto, perteneciente al género *Betacoronavirus*, subfamilia *Orthocoronavirinae*, incluida dentro de la familia *Coronaviridae* (*Genus: Betacoronavirus* | ICTV, s. f.; *NCBI Virus*, s. f.).

El virus del SARS-CoV-2 en personas, provoca una enfermedad denominada, COVID-19. En 2020 el COVID-19, causó una pandemia a nivel mundial. Desde que se tiene registros, el virus de SARS-CoV-2 ha provocado cerca de 700 millones casos de COVID-19 de los cuales 6.9 millones de personas han fallecido (*COVID - Coronavirus Statistics - Worldometer*, s. f.).

El genoma del SARS-CoV-2 es uno de los genomas de RNA más grandes conocidos hasta la fecha, con aproximadamente 30.000 nucleótidos. La secuenciación del SARS-CoV-2 sugiere que su tasa de mutación es de  $10^{-6}$ - $10^{-4}$  por ciclo replicativo, algo más lenta en comparación con otros virus de ARN (Expósito Lara et al., 2021). Esta posible ralentización podría atribuirse a la presencia de una enzima exorribonucleasa con actividad proof-reading que se encarga de corregir los errores

que puedan surgir durante la replicación (*El SARS-CoV-2 y sus mutaciones | Vacunas / Asociación Española de Vacunología, s. f.*).

El genoma del virus ha ido evolucionando y adquiriendo mutaciones en su secuencia nucleotídica de una forma significativamente rápida. Para poder diferenciar sus versiones genéticamente distintas, la OMS ha clasificado a los grupos que presentaban las mismas mutaciones relevantes, por **variantes** (*Enfermedad del coronavirus 2019 (COVID-19), 2020*) (*Tabla 2*). Las variantes del SARS-CoV-2 se clasifican en preocupantes, que son las que se detallan en la tabla que podemos ver a continuación (*Tabla 2*) y en variantes de interés. Otra forma de clasificar las variaciones en la secuencia de aminoácidos presentes en el genoma del virus que estamos estudiando, es el sistema de clasificación de linajes, Pango (*Cov-Lineages, s. f.*) (*Tabla 2*). Esta nomenclatura se usa para nombrar todos los linajes y sublinajes descubiertos. A diferencia de la OMS que solo nombra con letras griegas a las variantes de interés sanitario.

**Tabla 2.** Lista de variantes junto a su linaje pango y la fecha de designación

<b>Nomenclatura de la OMS</b>	<b>Linaje Pango</b>	<b>Fecha de designación</b>
<i>Alfa <math>\alpha</math></i>	B.1.1.7	29/12/2020
<i>Beta <math>\beta</math></i>	B.1.35	29/12/2020
<i>Gamma <math>\gamma</math></i>	P.1	29/12/2020
<i>Delta <math>\delta</math></i>	B.1.617.2	15/06/2021
<i>Ómicron <math>\omicron</math></i>	B.1.1.529	26/11/2021

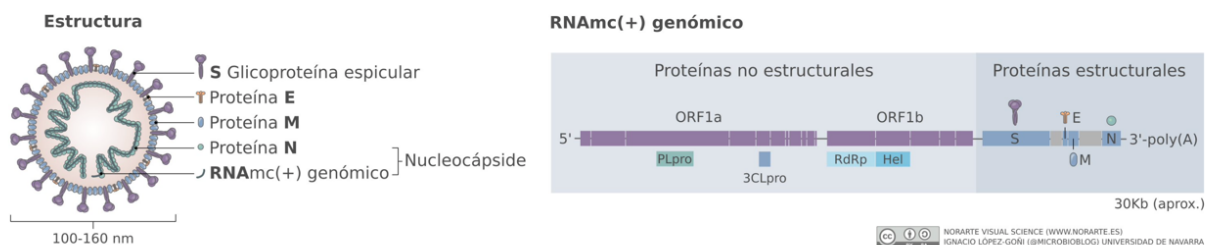
### 1.5.1 Estructura genética

El genoma del SARS-CoV-2 (Wuhan, 2019), está formado por una única cadena de RNA monocatenario y de polaridad positiva compuesta por exactamente 29.903 nucleótidos, que codifican 9860 aminoácidos. Su estructura es muy parecida a la de otros coronavirus. Podemos dividir el genoma del virus en dos partes: las regiones con genes que codifican 16 proteínas no estructurales, que interfieren en la respuesta inmune del innata del hospedero, y las regiones que codifican proteínas estructurales (*Figura 1*).

La cadena de RNA del SARS-CoV-2 presenta inicialmente el gen ORF1ab que codifica proteínas no estructurales (nsp). Tras este, se encuentran los genes que codifican las principales proteínas estructurales: S (proteínas de la espícula), E

(proteínas de la envoltura), M (proteínas de membrana) y N (proteínas de nucleocápside), además de otros ORFs (marcos abiertos de lectura) que codifican proteínas accesorias tales como la proteína hemaglutinina esterasa (HE), la proteína 3, la proteína 7a, entre otras (Pastrian-Soto, 2020). La proteína estructural S, es una de las proteínas estructurales más importante en el proceso de infección. La glicoproteína espicular, S, se encuentra distribuida por la superficie del virus (*Figura 1*). Esta es responsable de la unión y fusión del virus con las membranas celulares. La proteína de membrana, M, es la responsable del transporte transmembrana de nutrientes, liberación de la partícula viral y eventual formación de su envoltura (Oliva Marín, 2020).

### Coronavirus SARS-CoV-2



**Figura 1.** Estructura y genoma del SARS-CoV-2. En primer lugar, se puede observar un esquema de la estructura del SARS-CoV-2, en la que se identifican las principales proteínas estructurales. En segundo lugar, se representa de forma lineal y esquemática los principales ORF del SARS-CoV-2. Figura tomada de (Asensio, 2020).

### 1.5.2 Organización en cuasiespecies

Debido a la alta tasa de mutación que presenta el SARS-CoV-2, el virus existe formando espectros mutantes, denominados cuasiespecies. Las cuasiespecies virales son distribuciones dinámicas de genomas virales mutantes y recombinantes no idénticos, pero estrechamente relacionados y sujetos a un proceso continuo de variación genética, competencia y selección (Domingo et al., 2006). En lugar de pensar en un virus individual, la idea de cuasiespecies considera a la población como un todo, en el que una gran cantidad de variantes genéticas coexisten en equilibrio dinámico. Esta noción, introducida por el biólogo molecular Manfred Eigen en una teoría de evolución molecular para entender el origen de la vida, es particularmente relevante para los virus de ARN debido a su alta tasa de mutación y replicación. (Domingo & Schuster, 2016)

Durante las infecciones virales, las cuasiespecies, desempeñan un papel clave en la adaptabilidad de los virus a los entornos cambiantes y al destino de la población en su conjunto (Domingo et al., 2006).

Los espectros mutantes, que se generan de forma continua durante la replicación de RNA (Domingo et al., 2006), están formados por haplotipos, es decir, cada una de las diferentes secuencias existentes dentro de una cuasiespecie, dentro de una misma población (Díaz Martínez, 2016; Ortega Del Campo, 2020).

Toda cuasiespecie viral es caracterizada por su secuencia consenso, que es la secuencia nucleotídica que presenta en cada posición el nucleótido más frecuente en el conjunto de genomas de la población. Dicha secuencia no tiene por qué estar representada en el espectro de mutantes (Díaz Martínez, 2016).

Para poder estudiar las cuasiespecies de un virus, es necesario secuenciar el genoma del mismo con el fin de poder estudiarlo. La forma mas adecuada de realizar la secuenciación es la secuenciación masiva, NGS.

La secuenciación masiva (Next Generation Sequencing, NGS), se trata de un término genérico que engloba una serie de métodos y herramientas de secuenciación de material genético. Esta técnica presenta una gran precisión y eficacia en la identificación de variantes en el material genético y por ello hoy es una de las tecnologías más usadas en el ámbito de la genómica. La secuenciación masiva permite generar en paralelo y de forma masiva, millones de fragmentos de ADN en un único proceso de secuenciación, con un coste menor y con mayores ventajas que el resto de los sistemas convencionales, como puede ser la secuenciación Sanger. (M. S. González & Rosales, s. f.). En la actualidad existen diversas plataformas de secuenciación masiva, cada una de ellas basada en principios químicos distintos que generan diferencias en los resultados. Las plataformas más usadas son *Illumina* (*Illumina | Sequencing and array-based solutions for genetic research, s. f.*), *Ion Torrent* (*Ion Torrent - ES, s. f.*) y *Helicos BioSciences* (*Helicos BioSciences Corporation, s. f.*). A lo largo de este documento la plataforma que se optado por usar es *Illumina*.

Los datos obtenidos por las técnicas de secuenciación NGS, son principalmente tres tipos de ficheros (M. S. González & Rosales, s. f.):

- Ficheros FASTQ, contienen las puntuaciones de calidad de las llamadas a las bases. Son los ficheros necesarios para realizar la alineación.
- Ficheros BAM, son representaciones de alineamientos de secuencias con un genoma de referencia.
- Ficheros VCF, contienen las diferencias detectadas en cada posición en cualquiera de los dos grupos, considerando tanto las secuencias de lectura en su orientación original como en su orientación inversa.

Para el análisis de los datos obtenidos por las técnicas mencionadas, es necesario el uso de numerosas herramientas computacionales. Para analizar de forma rápida y eficaz los datos obtenidos usando de manera consecutiva y automática las herramientas informáticas, se han desarrollado flujos de trabajo. Estos flujos de trabajo normalmente suelen ser desarrollados por las plataformas de secuenciación o por los propios laboratorios que requieren de los datos de las secuenciaciones (Díaz Martínez, 2016).

### **1.6. Mutagénesis letal como estrategia antiviral**

Una de las estrategias antivirales actuales contra el SARS-CoV-2 es la administración de análogos de bases mutagénicas. Esta estrategia busca inducir mutaciones en el genoma viral de manera que sean incompatibles con la replicación y la función del virus, con el objetivo de aumentar las tasas de mutación viral y así reducir y eliminar la carga viral.

En la mutagénesis letal, el equilibrio mutación-selección del virus se encuentra alterado. Las mutaciones más abundantes no son las que aparecen en una cuasiespecie en condiciones normales (Díaz-Martínez et al., 2018; Lobinska et al., 2023; Swanstrom & Schinazi, 2022).

# 2

## Material y métodos

Para llevar a cabo este proyecto, se utilizó un Macbook Pro con un procesador 2,3 GHz Intel Core i5 de 4 núcleos, operando bajo un sistema operativo macOS Ventura versión 13.4.1.

Los programas y bases de datos usado para el desarrollo quedan detallados en la *Tabla 1*.

### **2.1. Muestras**

Para llevar a cabo este trabajo, el equipo de la Dr. Ana Grande, el 15 de febrero del 2022 en el hospital Virgen de la Victoria (Málaga), llevó a cabo la toma de muestras de virus de las variantes Alfa, Beta y Ómicron en pacientes mediante hisopos nasofaríngeo. Estas muestras fueron empleadas para llevar a acabo unos ensayos de una terapia antiviral en el marco de un proyectode investigación de la Dra. Ana Grande Pérez. Se realizaron en las instalaciones del Centro de Biología Molecular “Severo Ochoa” (CBMSO-CSIC, Madrid), en el laboratorio de cultivos con nivel de seguridad biológica 3 (P3).

El líquido de los hisopos nasofaríngeos se usó para infectar cultivos celulares formados por células Vero E6, de riñón del mono verde africano, proporcionadas por el Dr. Luis Enjuanes (CNB-CSIC, Madrid). A los 4 días de la infección empezaron a observarse efectos citopáticos. Al analizar la infectividad del sobrenadante de dichos cultivos, se observó que solo uno de los virus presentaba título viral. Tras la extracción del ARN y su secuenciación se comprobó que el virus pertenecía al SARS-CoV-2 de la variante Ómicron, con pango BA.1.17. Por lo tanto, todas las muestras que se analizan en este proyecto derivan de los ensayos realizados con dicho virus.

Para la obtención del virus stock el virus fue sometido a tres pases seriados en Vero E6 . Ese virus stock, denominado SARS-P3 se empleó para ensayar distintos tratamientos antivirales en cultivos. Se estudió el efecto de cada uno de ellos sobre la infectividad del virus. Posteriormente, se tendría que comprobar en el laboratorio de la Dra. Ana Grande Pérez en la UMA mediante secuenciación la aparición de mutaciones en su secuencia nucleotídica. En este segundo objetivo es donde se sitúa este Proyecto Fin de Grado.

El diseño de los distintos tratamientos antivirales, así como el diseño de los péptidos y el número de pases que se llevaron a cabo, constituyen componentes esenciales de un proyecto de investigación previo al inicio de este proyecto y son confidenciales ya que el grupo de investigación está solicitando una patente. Los resultados de este proyecto hasta el momento no han sido publicados, y, en consecuencia, no se puede proporcionar una explicación detallada y exhaustiva en relación con estos aspectos.

A grandes rasgos, se han llevado a cabo dos estrategias antivirales. En primer lugar, el tratamiento de las muestras con un análogo a diferentes concentraciones. Este análogo induciría mutaciones en el genoma del virus, provocando así la eliminación de la infección vírica. Mediante el estudio de genomas completos de SARS-CoV-2 sometidos a los tratamientos con distinta concentración de análogo se pretende comprobar el efecto del análogo sobre la complejidad y la heterogeneidad de las cuasiespecies del virus (Díaz-Martínez et al., 2018; F. J. V. González, s. f.; Martínez, s. f.; Swanstrom & Schinazi, 2022) .

En segundo lugar, el tratamiento de las muestras con péptidos. En el caso del péptido #14, se une a la región nsp14 del genoma del SARS\_CoV\_2 y bloquea la unión con el cofactor que activa la proteína nsp14, por lo tanto, no hay actividad exonucleasa correctora y favorece a la acumulación los errores durante la replicación,

aumentando así el número de mutaciones. En el caso del péptido #16 está relacionado con la actividad metiltransferasa de la nsp16. Esta desempeña un papel fundamental en evitar el reconocimiento del genoma viral y la acción de las enzimas por parte del sistema inmunitario innato que normalmente degradarían el RNA viral (Ferron et al., 2018; F. J. V. González, s. f.)

**Tabla 3.** Listado de muestras usadas en las amplificaciones. Todas ellas son de RNA extracelular.

Péptido		5-FU (µM)	Réplica	Id
Péptido	(µM)			
-	-	-	A	<b>A1</b>
-	-	-	B	<b>A3</b>
-	-	-	C	<b>A5</b>
-	-	300	A	<b>A7</b>
-	-	300	B	<b>A9</b>
-	-	300	C	<b>B2</b>
-	-	400	A	<b>B4</b>
-	-	400	B	<b>B6</b>
-	-	400	C	<b>B8</b>
-	-	500	A	<b>C1</b>
-	-	500	B	<b>C3</b>
-	-	500	C	<b>C5</b>
#14	70	-	A	<b>C7</b>
#14	70	-	B	<b>C9</b>
#14	70	-	C	<b>D2</b>
#14	70	300	A	<b>D4</b>
#14	70	300	B	<b>D6</b>
#14	70	300	C	<b>D8</b>
#14	70	400	A	<b>E1</b>
#14	70	400	B	<b>E3</b>
#14	70	400	C	<b>E5</b>
#14	70	500	A	<b>E7</b>
#14	70	500	B	<b>E9</b>
#14	70	500	C	<b>F2</b>
#16	70	-	A	<b>L4</b>
#16	70	-	B	<b>L6</b>
#16	70	-	C	<b>L8</b>
#16	70	300	A	<b>M1</b>
#16	70	300	B	<b>M3</b>
#16	70	300	C	<b>M5</b>
#16	70	400	A	<b>M7</b>
#16	70	400	B	<b>M9</b>
#16	70	400	C	<b>N2</b>
#16	70	500	A	<b>N4</b>
#16	70	500	B	<b>N6</b>
#16	70	500	C	<b>N8</b>

En la *Tabla 3*, se expone un listado de muestras de los RNA extraídos de cada uno de los ensayos antivirales especificando el tratamiento al que se ha sometido cada uno de ellos. En cada pase se llevaron a cabo tres extracciones de RNA, por ello de cada tratamiento, se pueden identificar tres réplicas A, B y C. Los tipos de tratamiento que podemos encontrar son:

- Muestras con el virus sin tratar.

- Muestras tratadas con el análogo a diferentes concentraciones (300  $\mu\text{M}$ , 400  $\mu\text{M}$  y 500  $\mu\text{M}$ ).
- Muestras tratadas con péptidos con actividad inhibitoria sobre el SARS-CoV-2. Para este proyecto designamos estos péptidos con los nombres #16 y #14 debido a que no se puede detallar la procedencia de estos.
- Muestras tratadas con el péptido ya sea el péptido #16 o #14 más el análogo a 300  $\mu\text{M}$ , 400  $\mu\text{M}$  y 500  $\mu\text{M}$ .

Cada una de las muestras se encuentran identificadas mediante un ID para poder trabajar con ella de una forma más sencilla y rápida.

## 2.1. Material de laboratorio

Todo el material necesario para el desarrollo de este Trabajo Fin de Grado ha sido proporcionado por el grupo de investigación de la Dra. Ana Grande Pérez, en el laboratorio del Área de Genética de la Facultad de Ciencias en la Universidad de Málaga. A continuación, se expone un listado de todo lo necesario para el desarrollo de este.

**Tabla 4.** Material necesario de laboratorio para desarrollar el proyecto.

Material	Uso
Micropipetas	RT-PCR, electroforesis
Puntas con filtro	RT-PCR
Puntas sin filtro	Electroforesis
Tubos de PCR	RT-PCR
Tubos Eppendorf	RT-PCR, Preparación de pools
Matraz, probetas y vasos de precipitado	Electroforesis
Termociclador	Electroforesis
Centrifugadora	RT-PCR
Ultracongelador a -80°C	Conservar las muestras
Muestras de RNA	RT-PCR
<i>Transcriptor One-Step RT-PCR Kit</i>	RT-PCR
Visualizador de geles GelDoc Go de Bio-Rad	Electroforesis
Cubeta y equipo de electroforesis	Electroforesis
TBE 1X y Bloding buffer	Electroforesis
Microondas	Preparación de geles de agarosa
Termobloque	RT-PCR
Nanodrop	Cuantificación de DNA
Máquina para hacer spin	RT-PCR y electroforesis
Parafilm	Electroforesis y preparación de pools
Guantes	Todo el proceso

## 2.2. Amplificación del genoma

Para llevar a cabo la amplificación se ha dividido el genoma del SARS-CoV-2 en secciones las cuales se han amplificado mediante RT-PCR, dando como resultado 12 amplicones solapantes de 3 Kb cada uno que en conjunto abarcan el genoma completo del virus. Los productos de estas 12 reacciones se han combinado en un solo tubo, se han secuenciado por Illumina y analizado mediante el flujo de trabajo QuasiFlow.

El conjunto de cebadores que se han empleado se muestran en la *Tabla 7*. Estos cebadores fueron diseñados por el grupo de la Dra. Ana Grande Pérez empleando el *PrimerBlast* (Primer designing tool, s. f.), programa del NCBI para el diseño de cebadores, y posteriormente se mandaron sintetizar a una empresa externa.

Esta fase del proyecto se ha llevado a cabo en el laboratorio de Genética de la Facultad de Ciencias de la Universidad de Málaga en conjunto con mi compañera de trabajo María Jeréz. Ambas hemos trabajado las mismas muestras de virus, pero cada una se ha centrado en amplificar una parte diferente del genoma. María Jeréz, trabajo con los primeros 6 cebadores desde el nsp1 hasta nsp10-12. Mientras que yo me he centrado en amplificar las muestras con los últimos seis cebadores, es decir, desde el nsp12-13 al N, *Tabla 7*.

Esta sección del estudio ha demandado una gran cantidad de tiempo y esfuerzo, principalmente debido a las duraciones de los procesos implicados y a la naturaleza experimental de la metodología empleada. Resulta pertinente señalar que los resultados obtenidos no siempre han coincidido con las expectativas debido al error inherente a la experimentación.

### Protocolo RT-PCR

El protocolo de transcripción inversa seguida de reacción en cadena de la polimerasa (RT-PCR) es una técnica molecular ampliamente utilizada en biología molecular y genética para amplificar RNA. Entre sus aplicaciones están las de detectar y cuantificar la expresión de genes específicos. Esta técnica combina la amplificación de DNA mediante la reacción en cadena de la polimerasa (PCR) con la conversión previa de RNA a DNA complementario (cDNA) mediante la transcriptasa inversa o

reverso transcriptasa (RT), permitiendo así analizar la expresión génica a nivel de RNA o el amplificado de genomas en el caso de virus de RNA.

Las muestras utilizadas, *Tabla 3*, se conservaron a  $-80^{\circ}\text{C}$  en un ultracongelador localizado dentro del laboratorio. Una vez fuera del ultracongelador, en todo momento debían mantenerse en hielo para conservar el RNA. Para la descongelación de las muestras se llevó a cabo el siguiente método de descongelación. La primera vez descongelación de una muestra se empleó un termobloque. Este dispositivo mantiene y controla la temperatura de las muestras en unos valores establecidos. Su uso fue fundamental para asegurar que la cadena de RNA se desnaturalizase y se mantuviera simple y estable. Esto es debido a que cuando el RNA permanece un tiempo prolongado a temperatura ambiente empieza a hibridar consigo mismo y era necesario mantenerlo como cadena simple para favorecer la hibridación con los cebadores.

El tratamiento de las muestras en el termobloque fue el siguiente. En primer lugar, se debe precalentar la gradilla hasta llegar a  $70^{\circ}\text{C}$ . Una vez la gradilla esté a la temperatura adecuada depositar las muestras en la gradilla y dejarlas 5 minutos. En la mitad del tiempo mezclar manualmente para que las muestras sean homogéneas. Y por último, retirar las muestras del termobloque, introducirlas en una cubeta con hielo y esperar 5 minutos.

En el caso de no ser la primera descongelación las muestras se dejaron descongelar en hielo o en la mano hasta que estuvieran casi descongeladas. Después de agitar las muestras se centrifugaron brevemente en frío y se dejaron en hielo.

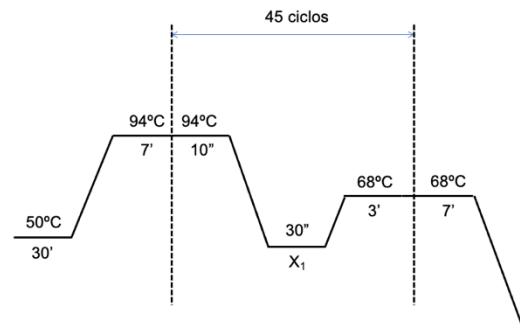
El protocolo RT-PCR se llevó a cabo en reacciones de un solo paso. Para ello, se optó por usar el "*Transcriptor One-Step RT-PCR Kit*" de Roche (Basilea, Suiza)(*Roche España*, s. f.), y se siguieron las condiciones de reacción que se indican en la *Tabla 6*.

Los reactivos necesarios junto a las cantidades fueron calculados de acuerdo con las indicaciones de fabricante. Esta información viene detallada en la *Tabla 5*. Los volúmenes de RNA fueron ajustados dependiendo de las muestras. Esta cantidad varió en un rango de  $2\ \mu\text{L}$  a  $5\ \mu\text{L}$ . La reacción final debía tener un volumen de  $20\ \mu\text{L}$ , por lo tanto, en el caso de que la cantidad de RNA fuera más de  $2\ \mu\text{L}$ , el exceso de

RNA se contrarrestaba quitando *agua* con el fin de que el volumen del mix de reacción fuera el necesario.

**Tabla 5.** Reactivos y cantidades RT-PCR un solo paso.

Reactivos	Volumen
Agua para PCR	12 $\mu$ L
5x buffer	4 $\mu$ L
Fwd primer	0,8 $\mu$ L
Rev primer	0,8 $\mu$ L
Mix enzimas	0,4 $\mu$ L
Muestra RNA	2 $\mu$ L
	<b>20 <math>\mu</math>L</b>



**Figura 1.** Gráfico que representa el protocolo RT-PCR.

Antes de usar los reactivos se agitaron suaaavemente con el fin de homogeneizar y se centrifugaron durante 1 minuto a 4°C. La reacción en la que se unen todos los reactivos recibe el nombre de *mix de reacción*. En primer lugar, se realizaba un mix con todos los reactivos detallados en la *Tabla 5*, a excepción del RNA, para poder realizar RT-PCR a más de una muestra, *Tabla 3*, con un mismo cebador y agilizar así el proceso. Por ello, el *mix de reacción* sin el RNA se elaboraba en tubos Eppendorf y posteriormente una vez centrifugado, se depositaban 18  $\mu$ L de dicha reacción en cada tubo de PCR en el que ya se encontraban los 2  $\mu$ L de RNA.

En el momento en el que cada tubo de PCR contuvieran 20  $\mu$ L, se sometían a un spin, con el fin de homogeneizar la solución y se introducían en el termociclador con el programa adecuado a cada primer (*Tabla 6*).

Como podemos observar, en el “*Transcriptor One-Step RT-PCR Kit*”, *Tabla 5*, encontramos un mix de enzimas el cual contiene todas las enzimas necesarias para la reacción. Para lograr la amplificación del material genético, es esencial emplear la DNA polimerasa. No obstante, esta enzima tiene capacidad de actuar únicamente sobre el DNA, mientras que la muestra con la que trabajamos es de RNA. Por ello, es necesario convertir previamente el RNA viral en cDNA, mediante el uso de la transcriptasa inversa. Una vez formada la cadena de DNA, es necesario la desnaturalización mediante calor con el fin de separar la hebra de RNA y la hebra de cDNA donde posteriormente el cebador se unirá a su zona complementaria y con la actuación de DNA polimerasa crearán doble cadena de DNA la cual se amplificará en los ciclos determinados por el programa indicada en el termociclador, *Tabla 6*. Los

ciclos de amplificación constan de tres fases. Una primera fase de desnaturalización donde la doble hebra de DNA se separa debido a un aumento de temperatura. Seguida de una fase de hibridación, en la que los cebadores se unen a sus secuencias complementarias. Y finalmente una fase de elongación, en la que la DNA polimerasa sintetiza una nueva cadena de DNA complementaria.

**Tabla 6.** Programa RT-PCR de un solo paso.

<b>Pasos</b>	<b>Temperatura</b>	<b>Tiempos</b>	<b>Nº de ciclos</b>
<i>Transcripción inversa</i>	50 °C	30 min	1
<i>Desnaturalización inicial</i>	94 °C	7 min	1
<i>PCR estándar</i>	<i>Desnaturalización</i>	94 °C	45
	<i>Hibridación</i>	X <sub>1</sub> °C *	
	<i>Elongación</i>	X <sub>2</sub> *	
	<i>Elongación final</i>	68 °C	7 min

X<sub>1</sub>\*: Temperatura de hibridación propias de cada oligonucleótido. X<sub>2</sub>\*: Tiempo de elongación, 60 s por kb, teniendo en cuenta el número de kb que presentan las muestras, el tiempo en todos los casos es 3 minutos.

En la etapa tanto de hibridación como de elongación la temperatura varía dependiendo del par de cebadores con el que estemos trabajando. Las temperaturas y el número de pares de bases necesarias para cada par de cebadores se encuentran recogidos en la *Tabla 7*.

**Tabla 7.** Listado de cebadores que compone el genoma del SARS-Cov-2. En cada amplicón encontramos 2 cebadores, FP\_ y RP\_ correspondientes a la hebra directa (5' → 3') e indirecta (3' → 5').

<b>Nº de amplicón</b>	<b>Cebador_id</b>	<b>Posición</b>		<b>T hibridación (°C)</b>	<b>Amplicón (pb)</b>
<i>Control S Fw</i>	Fw	22853	22871	47,00	-
<i>Control S Rev</i>	Rv	23288	23269		
<i>nsp1</i>	FP_nsp1	19	40	55,00	2935,00
	RP_nsp1	2953	2932		
<i>nsp3-1</i>	FP_nsp3-1	2394	2415	55,00	3055,00
	RP_nsp3-1	5448	5425		
<i>nsp3-2</i>	FP_nsp3-2	4880	4902	53,00	3058,00
	RP_nsp3-2	7937	7917		
<i>nsp4</i>	FP_nsp4	7373	7392	53,00	2976,00
	RP_nsp4	10348	10328		
<i>nsp5-9</i>	FP_nsp5-9	9836	9855	52,00	3026,00
	RP_nsp5-9	12861	12841		
<i>nsp10-12</i>	FP_nsp10-12	12406	12426	53,00	3046,00
	RP_nsp10-12	15451	15431		
<i>nsp12-13</i>	FP_nsp12-13	14901	14921	53,00	2988,00
	RP_nsp12-13	17888	17868		
<i>nsp14-15</i>	FP_nsp14-15	17313	17332	54,00	3160,00

	RP_nsp14-15	20472	20452		
<i>nsp16-S</i>	FP_nsp16-S	19787	19808	55,00	3056,00
	RP_nsp16-S	22842	22822		
<i>S</i>	FsP_S	22172	22191	54,00	3043,00
	RP_S	25214	25193		
<i>ORF6</i>	FP_ORF6	24690	24709	55,00	2896,00
	RP_ORF6	27585	27566		
<i>N</i>	FP_N	26962	26981	53,00	2845,00
	RP_N	29806	29785		

### 2.2.1. Electroforesis en gel de agarosa

La electroforesis es una técnica esencial en laboratorio de biología molecular utilizada para separar y analizar moléculas de ácidos nucleicos. Esta técnica se basa en la migración de moléculas cargadas en un gel de agarosa que actúa como una matriz tridimensional y resiste el movimiento de las moléculas según su tamaño y carga. Por lo tanto, nos permite separar y caracterizar diferentes fragmentos de ácidos nucleicos presentes en una muestra. Las bandas observadas en el gel representan los diferentes tamaños de las moléculas. Para estimar el tamaño de las moléculas usamos un marcador de peso molecular (*Figura 4*).

#### Protocolo electroforesis en gel de agarosa

Para poder visualizar los tamaños moleculares de los resultados de la RT-PCR, realizamos electroforesis en gel de agarosa a todas las muestras con cada uno de los cebadores que componen el genoma del SARS-CoV-2. El método que se debe seguir es el siguiente:

##### 1. Preparación del gel de agarosa

Para la preparación del gel de agarosa es necesario un matraz de Erlenmeyer, en el que se depositan 100 mL de TBE 1X, Tris-Borato-EDTA, un tampón necesario para que se produzca la electroforesis. Y 1 g de agarosa en polvo. Esta solución, debe introducirse en el microondas hasta observarse que toda la agarosa en polvo esté fundida. En el laboratorio pudimos observar que el tiempo exacto era 1,5 minutos. Una vez homogeneizada la solución y comprobando previamente que toda la agarosa está fundida, se incorporan 5µL de greensafe DNA gel Stain (Canvax Biotech, Valladolid, España) (*GreenSafe DNA Gel Stain - Canvax Reagents SL, s. f.*).

El matraz a de moverse para homogeneizar la solución y se debe dejar enfriar, hasta una temperatura de unos 50°C. Para agilizar el proceso, se introduce el matraz en agua fría con cuidado de que el agua no entre en la solución de agarosa. Una vez alcanzada la temperatura especificada, se vierte el contenido del matraz en un molde nivelado específico para la cubeta de electroforesis, colocando el peine en la posición deseada y dejando que solidifique durante una hora aproximadamente (*Figura 3*).

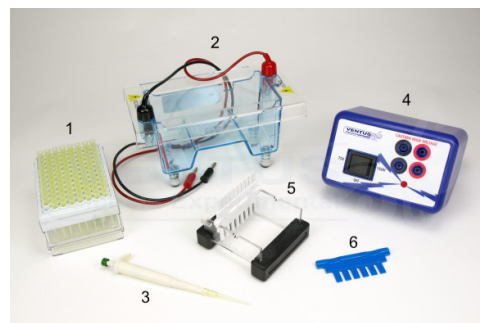
## 2. Preparación de las muestras

Una vez acabado el protocolo RT-PCR, las muestras son retiradas del termociclador y deben ser sometidas a un pequeño spin con el fin de garantizar que la muestra sea homogénea en todo su contenido. En cada pocillo del gel van a ser cargados 7µL, 5µL de muestra y 2µL de blocking buffer del laboratorio Thermo Fisher (Massachusetts, EE. UU.). En esta parte del protocolo, se mezclan 5µL de muestra y 2µL de tampón de carga de electroforesis (loading buffer) con el fin de que la muestra de DNA entre al fondo del pocillo del gel.

## 3. Carga de las muestras en el gel y electroforesis

Cuando el gel ya se encuentra solidificado, se introduce en la cubeta de electroforesis (*Figura 3*), y se retira cuidadosamente el peine del gel. La cubeta debe estar llena de TBE 1x hasta la marca del máximo volumen para que el gel pueda correr. Posteriormente se cargan los 7µL, indicados en el apartado anterior, en los pocillos. Hay que tener en cuenta, que el primer pocillo siempre será un marcador de peso molecular que nos ayudará determinar el tamaño de las bandas de la muestra a estudiar.

Cuando ya este todo listo, se conectan los cables de la fuente de alimentación y se aplica un voltaje de 100 voltios durante 45 minutos.

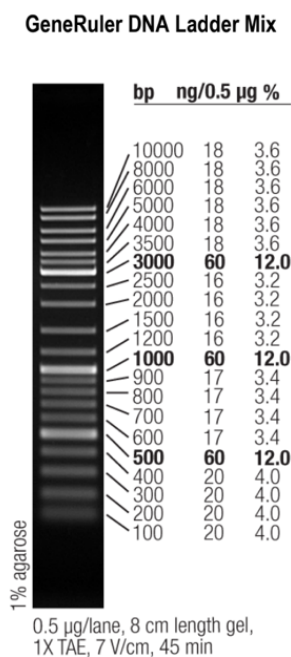


**Figura 3:** Material de laboratorio necesario para electroforesis en gel de agarosa. 1.Puntas sin filtro 2.Cubeta de electroforesis 3.Micropipetas 4.Fuente de alimentación 5.Molde 6.Peine. Figura tomada de <https://www.ventusciencia.com/equipo-base-para-practicas-de-electroforesis.html>

### 2.2.2. Visualización de DNA

Una vez pasados los 45 minutos de la electroforesis, podemos retirar el gel de la cubeta de electroforesis e introducir el gel en el visualizador de geles. Gracias al marcador de peso molecular podemos saber el tamaño molecular de cada una de las bandas en los carriles comparándolos con la figura que nos proporciona la casa comercial, *Figura 4*.

Las imágenes obtenidas eran exportadas a un pendrive y almacenadas en una carpeta para posteriormente usarlas en el estudio.



**Figura 4.** Marcador de masa molecular.

### 2.3. Control del experimento

Como en todo experimento científico, debe haber una serie de controles negativos y controles positivos que garanticen y validen la confiabilidad de los resultados obtenidos.

Los controles negativos desempeñan un papel esencial en la evaluación de posibles efectos no deseados. Al utilizar una muestra en la que no se espera que ocurra el resultado deseado, es posible determinar si hay factores externos influyan en los resultados. Estos ayudan también a garantizar que los efectos observados sean genuinos y no sean resultados de fluctuaciones aleatorias o interferencias. En el caso de este trabajo, todos los protocolos RT-PCR que se han realizado contaban con un

control negativo, a excepción de las RT-PCR de los pools. Este se trataba del mix de reacción sin el reactivo de RNA. El resultado esperado sería que al realizar la electroforesis en el gel de agarosa y llevar a cabo la visualización correspondiente, no se observase ninguna banda, ya que no hay fragmentos de RNA que amplificar.

Por otro lado, los controles positivos brindan una base para comprobar los resultados obtenidos en el experimento. En el caso de que el control positivo de un resultado esperado confirma que el sistema está operando correctamente y que las metodologías y condiciones empleadas son las adecuadas. En este proyecto, como control se ha utilizado una muestra control de RNA de SARS-CoV-2 cedida amablemente por el grupo de Celia Perales y Esteban Domingo, del Centro Nacional de Biotecnología (Madrid, España).

Además de los controles positivos y negativos en las reacciones RT-PCR, se concluyó que también sería conveniente realizar un control de calibración en el proceso de secuenciación para garantizar que las variaciones en el genoma eran producidas por la cuasiespecie y no por errores en el proceso de manipulación de las muestras en el laboratorio (reacciones RT-PCR, preparación de pools, purificación de las muestras...). Por ello se realizó una última tanda de RT-PCR con las siguientes muestras.

1. Control negativo
  2. RNA control A
  3. RNA control B
  4. RNA control C
- } Réplica exacta de Wuhan
5. Virus sin tratar 1 Replica A
  6. Virus sin tratar 2 Replica A
- } Secuencia muestra A1, *tabla 3*.

Para realizar este control solo se usaron el primer y último par de cebadores, *Tabla 7*.

Las muestras 2, 3 y 4 corresponden que con los controles A, B y C, provienen de un RNA con la secuencia exacta de Wuhan, Twist Synthetic SARS-CoV-2 RNA control 2 (MN908947.3). Este RNA fue comprado a la casa comercial Twist Bioscience Corporation (San Francisco, EE. UU.).

En los resultados de estas amplificaciones y posteriores secuenciaciones en dicho RNA control no deberían aparecer ninguna mutación ya que al RNA no se le ha aplicado ningún tratamiento, salvo las inherentes a los propios procesos de transcripción de dicho RNA control o de la secuenciación masiva del mismo que deberrían estar muy por debajo de la tasa de error del virus (Klempt et al., 2020).

#### **2.4. Preparación de pools**

Tal y como he mencionado en el apartado 2.3, para llevar a cabo la amplificación del genoma, es necesaria la división por regiones de este, es decir, se ha amplificado el genoma por partes, obteniéndose amplicones de 3Kb. Pero para poder ejecutar la secuenciación, es necesario que, para cada muestra (*Tabla 3*) estén todas las partes del genoma unidas en una misma solución y por ello es necesario elaborar pools de cada muestra de RNA que se extrajo.

Cada pool esta compuesto por las amplificaciones de las diferntes partes del genoma que hemos realizado entre mi compañera María Jeréz y yo.

Al juntar las diferentes partes del genoma de cada muestra en una única solución, estas deben ser cantidades equimoleculares, con el objetivo de que la cobertura de secuenciación sea lineal. Por lo tanto, se ha de buscar una forma para que la concentración de DNA sea igual de todas las amplificaciones de las muestras. Para ello, comparamos todas las visualizaciones de las RT-PCR que hemos ido haciendo durante el proceso, con el propósito de que cada una de las bandas correspondientes a cada amplicón tengan todas con la misma cantidad de ADN.

Una vez seleccionadas las 12 reacciones de cada amplicón para cada muestra, el material amplificado de cada uno de los tubos de PCR se transfirió a un tubo Eppendorf. Todos los pools contenían aproximadamente 160  $\mu$ L. En previsión de posibles contratiempos en el laboratorio donde se llevó a cabo la secuenciación, se decidió dividir el pool en dos pools de 80  $\mu$ L, uno se mandaría a secuenciar y el otro se conservaría en el laboratorio siempre en las condiciones adecuadas. La finalidad es asegurar que, en caso de dificultades, contáramos con la posibilidad de repetir el procedimiento y no poner en riesgo el proyecto.

Como se expondrá a continuación, la secuenciación del genoma se ha llevado a cabo en un laboratorio externo. Para que se pudiera garantizar un buen resultado, desde el laboratorio nos exigían una serie de requisitos para cada uno de los pools:

Se necesitaba que cada pool contuviera, al menos 20  $\mu$ l a una concentración como mínimo de 10 ng/ $\mu$ l y como control de calidad del DNA un índice OD260/280 de entre 1.8 y 2.

Tanto para medir la concentración como para determinar el índice OD260/280, es necesario el uso del espectrofotómetro, en este caso usamos el Nanodrop.

En el momento en el que se comprobó que los pools cumplían los requisitos impuestos, estos se purificaron por precipitación de ADN con etanol y sales de acetato de potasio. Una vez todos los pools estuvieron listos, los tubos se sellaron con Parafilm, se colocaron en cajas con gradillas y se enviaron al Servicio de Secuenciación del SCBI de la Universidad de Málaga, dentro una caja térmica, donde se secuenciaron por Illumina.

## **2.5. Análisis bioinformático**

El análisis bioinformático de este proyecto está hecho en colaboración con el Centro de Bioinnovación y Supercomputación (SCBI), de la Universidad de Málaga. Ellos han sido los encargados llevar a cabo la secuenciación y la ejecución del flujo de trabajo QuasiFlow mediante el cual estudio de la variabilidad de cuaespecie virales a partir de datos de secuenciación masiva, NGS.

### **2.5.1. Datos de secuenciación masiva**

Los datos de secuenciación masiva que se han usado para el desarrollo del análisis informático fueron proporcionados por el Centro de Bioinnovación y Supercomputación (SCBI), de la Universidad de Málaga, encargados de llevar a cabo la secuenciación de las muestras que fueron amplificadas en el laboratorio. La tecnología usada en la secuenciación NGS fue por Illumina net set 550.

Para poder ver la lecturas mediante los archivos BAM, proporcionados por Illumina se uso el software IGV\_2.16.2.

Los análisis de los datos obtenidos de la secuenciación fueron tratados con el flujo de trabajo QuasiFlow, ejecutado desde una terminal y se encuentra instalado en la supercomputadora PICASSO (*Picasso | RES - Red Española de Supercomputación*, s. f.), ubicado en el SCBI de la Universidad de Málaga.

La herramienta bioinformática de código libre QuasiFlow, consiste en un flujo de trabajo, *Figura 5*, el cual usa un conjunto de programas especialmente diseñados para el análisis de secuencia nucleotídicas procedentes de secuenciación masiva, *Tabla 8*. Todos los programas que usa dicho entorno están creados para Linux y funcionan bajo el entorno UNIX. Esta herramienta solo puede ser utilizada a través del supercomputador *PICASSO* (Seoane et al., 2022).

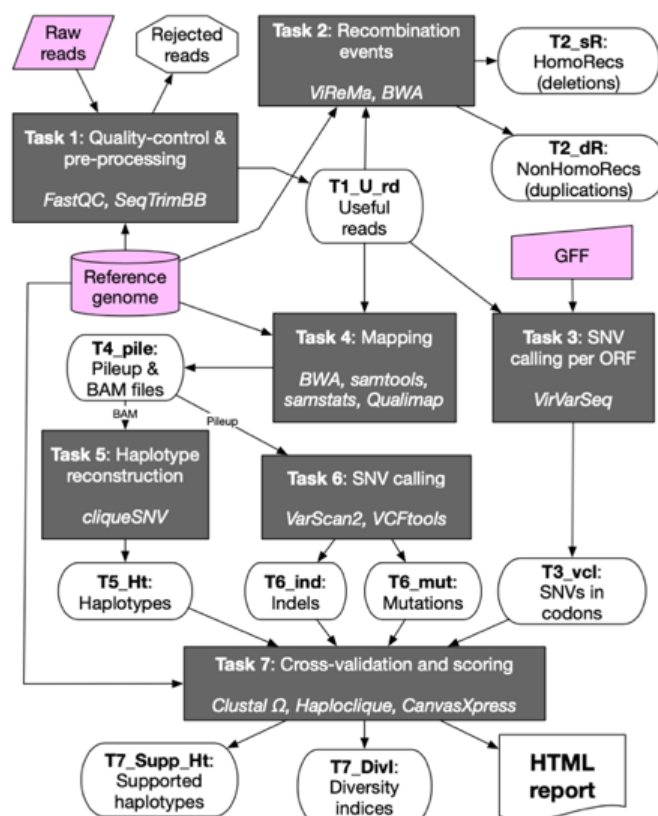
Gracias a la estructura a modo de flujo de trabajo reduce el tiempo de análisis de los datos, aumenta el número de muestras a estudiar en paralelo y automatiza al máximo este tipo de análisis.

El QuasiFlow presenta una estructura dividida en dos bloques, un primer bloque en el que se estudian las muestras individualmente y un segundo bloque en el que se realiza un estudio estadístico comparativo de todas las muestras de forma conjunta.

**Tabla 8.** Principales programas que conforman el flujo de trabajo del QuasiFlow junto con una breve explicación de la función que desempeñan.

<b>Programas</b>	<b>Uso</b>
<i>Autoflow</i>	Gestor de flujos de trabajo, base del QuasiFlow.
<i>Seqtrim Next</i>	Filtrado de secuencias de baja calidad, búsqueda de secuencias similares a una referencia dada y eliminación de indeterminaciones en las secuencias y otros contaminantes
<i>BWA</i>	Indexado de las secuencias de referencia y posterior mapeo de las secuencias obtenidas por secuenciación masiva.
<i>SamTools</i>	Alineamiento de secuencias en formato SAM, selección, fusión e indexación de secuencias previamente alineadas.
<i>Bowtie</i>	Indexado de secuencias de referencia.
<i>Samstats</i>	Obtener estadísticas a partir de archivos SAM.
<i>Vcftools</i>	Con un archivo VCF ofrece métodos accesibles para trabajar con variaciones genéticas procedentes de secuenciones complejas.
<i>Varscan</i>	Análisis de variantes dentro de datos de secuenciación masiva. Detecta deleciones de SNPs e InDels.
<i>Haploclique</i>	Reconstrucción de la estructura de cuasiespecie virales a partir de datos de secuenciación masiva. Reconstrucción de haplotipos y estimación de su abundancia.
<i>R-Bioconductor</i>	Análisis y comprensión de datos obtenidos por otros programas de análisis genómicos de NGS.
<i>Clustal omega</i>	Estudio filogenético de secuencias procedentes de la reconstrucción de los haplotipos de la cuasiespecie viral.

<i>VirVarSeq</i>	Estudio de las poblaciones minoritarias de codones.
<i>ViReMa</i>	Detecta y enumera cualquier evento de recombinación o fusión en genomas virales secuenciados por NGS.
<i>Circos</i>	Visualización de datos en un plano circular.



**Figura 5.** Flujo de trabajo del Quasiflow. Tomada de (Seoane et al., 2022).

Gracias a esta herramienta podremos obtener datos tales como la frecuencia de mutación, la diversidad genética, el número de haplotipos, índice de Shannon y número de mutaciones a partir del cual podremos hacer un análisis de sesgos de nucleótidos.

### 2.5.2. Secuencia de Wuhan

Tal y como se ha introducido en la primera parte del proyecto. El SARS-CoV-2 fue aislado por primera vez en la ciudad de Wuhan (China), el 13 de enero de 2020. A partir de la secuencia que se obtuvo en este momento esta se determinó como la de referencia para posteriores estudios.

En este estudio, la secuencia de Wuhan es de gran importancia debido a que prácticamente todos los estudios que se han realizado son con respecto a esta. Para

obtener la secuencia completa es necesario acceder a la base de datos del *NCBI Virus* que usa como referencia la base de datos *gen bank* (*GenBank Overview*, s. f.). El número de acceso de la secuencia es NC\_045512, presenta 29903 pares de bases y el hospedador del virus donde fue aislado es el *Homo sapiens*.

### 2.5.3. Secuencias consenso

Las secuencias consenso, son secuencias de DNA que representan patrones altamente conservados en una colección de secuencias relacionadas, derivadas de un análisis de muchas secuencias similares. Son cruciales para comprender las características estructurales y funcionales compartidas por un conjunto de genes. Las secuencias consenso recogen en cada posición la base más frecuente, las bases que estén presentes en más del 50% de las secuencias víricas de la población (Ortega Del Campo, 2020).

Esta secuencia al tratarse de una secuencia aproximada puede ser real o no, es decir, puede estar presente o no en la cuasiespecie.

En este proyecto, las secuencias consenso son extraídas gracias al flujo de trabajo QuasiFlow, que como se ha detallado en apartados anteriores esta fase del trabajo se ha desarrollado en colaboración con el SCBI, centro de Bioinnovación de la Universidad de Málaga. Uno de los programas más destacables del flujo de trabajo es SAMtools para la obtención de las secuencias consenso. Este se trata de un software mediante el cual se manipulan archivos y se realizan alineación en formato SAM y BAM.

El SAMtools se trata de software muy útil y ampliamente utilizado para manipular archivos y alineaciones en el formato SAM y BAM, utilizado en una amplia gama de análisis genéticos (Ramirez-Gonzalez et al., 2012).

Para el análisis de estas, se usarán las herramientas bioinformáticas *Mega11* , *JalView* (Tamura et al., 2021; Waterhouse et al., 2009), *NextClade* (Aksamentov et al., 2021) y *Blast* (*BLAST: Basic Local Alignment Search Tool*, s. f.) además de otras que se mencionarán en los demás apartados.

*NextClade* se trata de una plataforma desarrollada para analizar y visualizar las secuencias genéticas de virus, en particular del SARS-CoV-2. Esta herramienta permite realizar análisis filogenéticos, detectar mutaciones siempre en relación con la

secuencia de referencia (Secuencia de Wuhan, Apartado 2.6.2), asignar clados y linajes, identificar motivos genéticos, visualizar mutaciones en cada ORF...

La herramienta *BLAST* detecta áreas con similitudes en secuencias biológicas. Este programa compara secuencias de nucleótidos o proteínas frente a bases de datos de secuencias, y realiza cálculos de relevancia estadística.

### **2.5.3.1. Alineamientos de secuencias**

El alineamiento de secuencias es una técnica fundamental que permite identificar regiones conservadas y patrones de similitud entre secuencias de ADN.

La alineación múltiple de secuencias biológicas se ha llevado a cabo con el programa informativo Mega11 (Stecher et al., 2020; Tamura et al., 2021) en un terminal macOS. Este ofrece diferentes estrategias para realizar el proceso de alineación. En este caso, se ha optado por seguir un alineamiento múltiple tipo MUSCLE (Multiple Sequence Comparison by Log-Expectation), considerándose el dominio completo de cada familia de proteínas puede alinearse globalmente.

El alineamiento de secuencias genéticas mediante el método MUSCLE sigue un proceso iterativo y progresivo que busca identificar relaciones evolutivas y homologías entre secuencias (Edgar, 2004). Los pasos principales del método son:

- Selección de secuencias: Se recopilan secuencias de ADN relacionadas evolutivamente.
- Construcción de perfiles: Este método crea perfiles individuales para cada secuencia utilizando matrices de puntuación que evalúan la similitud entre los nucleótidos de las secuencias.
- Alineación progresiva: MUSCLE realiza un alineamiento progresivo comenzando con las secuencias más similares entre sí. Luego, agrega gradualmente las secuencias restantes utilizando la información de los perfiles y las matrices de puntuación para establecer alineamientos coherentes.
- Refinamiento iterativo: El proceso se somete a varias iteraciones en las que se ajustan las puntuaciones y penalizaciones para mejorar la precisión del alineamiento.

- Generación de alineamiento final: Una vez completadas las alineaciones, MUSCLE produce un alineamiento múltiple final en el que las secuencias están ordenadas de manera que las regiones homólogas se alinean, revelando patrones de similitud.

Tras la obtención de las secuencias consenso en archivos con formatos FASTA (.fasta), es necesario el tratamiento de estos ficheros para poder determinar las mutaciones que han sufrido las diferentes muestras con respecto a las otras variantes.

El laboratorio donde se llevó a cabo la secuencia elaboró ficheros por separado de cada una de las muestras en las que cada fichero contenía la secuencia consenso. El programa informático MEGA11 solo admite un fichero como entrada para poder realizar el alineamiento, por lo que se ha generado un fichero en el que podemos encontrar las secuencias consenso de todas las muestras concatenadas. En el fichero se encuentra cada secuencia introducida por un ">" seguido del nombre de la muestra. Para ello se ha utilizado VirtualBox (*Documentation – Oracle VM VirtualBox*, s. f.) con la distribución Ubuntu 20.04 (sistema operativo Linux). En la terminal de Linux se ha utilizado el siguiente código:

Para comprobar que el directorio de trabajo es el correcto se usa el comando 'pwd'. En el caso de que el directorio no sea el de la carpeta donde se encuentran todas las secuencias .fasta, este ha de cambiarse usándose el comando:

```
cd ruta/al/directorio
```

El comando 'cd' se usa seguido de la ruta del directorio específico al que se desea cambiar y permite navegar entre diferentes carpetas y ubicaciones en el sistema de archivos, con el fin de cambiar el directorio de trabajo de la terminal. Una vez se ha comprobado que el directorio de la terminal es el correcto, ejecutamos el siguiente código:

```
cat * >> SARS_Consensus_ALL.fasta
```

Dicho comando (cat) permite concatenar todo el contenido (\*) del directorio en el que se está trabajando, redirigiendo (>>) y creando un nuevo archivo (SARS\_Consensus\_ALL.fasta) con todas las secuencias agrupadas en un solo archivo .fasta.

De esta forma, se consigue concatenar todos los archivos .fasta donde se encontrabas las secuencias consenso de nuestras muestras por separado, en un único archivo "SARS\_Consensus\_ALL.fasta" donde se todas las secuencias.

Para hacer el alineamiento se han usado dos secuencias de la variante alfa, dos de la beta, dos de la delta, dos de la gamma, tres de la ómicron y las treinta y seis muestras indicadas de la *Tabla 3*, con el fin poder hacer un estudio filogenético completo y con más variedad de muestras.

Las secuencias externas al experimento que se usan en la alineación son obtenidas de la base de datos del *NCBI Virus* (*NCBI Virus*, s. f.).

Haciendo uso del alineamiento y apoyándome en las bases de datos del NextClade (Aksamentov et al., 2021) y la herramienta bioinformática JalView (Waterhouse et al., 2009). Se podrá realizar una identificación y clasificación de las mutaciones presentes en las secuencias consenso.

Es importante hacer estudio de las mutaciones presentes en las secuencias consenso con respecto a la secuencia de referencia de Omicron para este ensayo. Esta fase del trabajo será realizada mediante el uso de JalView, una herramienta que permite la alineación de secuencias de manera similar a Mega11, pero con una interfaz gráfica que incluye un gráfico el cual muestra el porcentaje de conservación de la consenso de todas las secuencias alineadas. Esto facilita la detección de posibles variaciones en las secuencias de manera más sencilla.ello es am más fácil el reconocimeinto de variaciones en las secuencias.

### **2.5.3.2. Análisis filogenético**

El análisis filogenético consiste en un estudio de las relaciones filogenéticas entre diferentes especies, poblaciones o secuencias genéticas.

Para llevar a cabo el análisis, se ha realizado la construcción de árboles filogenéticos mediante el programa bioinformático Mega11 (Stecher et al., 2020; Tamura et al., 2021). En este programa se ha optado por uso del método de máxima verosimilitud. Este método de reconstrucción filogenética está basado en modelos probabilísticos. Busca encontrar el árbol que maximice la probabilidad de observar las secuencias dadas una serie de hipótesis del árbol y los modelos evolutivos. Se

asignan probabilidades a las secuencias observadas en función de un modelo de evolución que describe cómo cambian las secuencias a lo largo del tiempo.

Además, para que el árbol filogenético de una información más sólida, se han incluido los valores de Bootstrap para cada uno de los nodos. Estos son estimaciones estadísticas que proporcionan una medida de la confiabilidad o robustez de las relaciones filogenéticas representadas en el árbol filogenético. Los nodos que cuentan con valores cercanos a 100 indican una mayor probabilidad de ser correctos, mientras que aquellos que se alejan de ese valor muestran una mayor probabilidad de haber surgido al azar. Se podrá considerar que los nodos son robustos siempre que el valor de Bootstrap sea superior a 60.

Para realizar el análisis filogenético se han utilizado las secuencias alineadas en el apartado anterior y además una secuencia externa. Esta secuencia estaría más alejada filogenéticamente que cualquiera de las secuencias de SARS-CoV-2 recogidas en el alineamiento de secuencias y hará que este análisis sea más completo. Para la obtención de esta, se ha usado la base de datos del ICTV, Comité Internacional de Taxonomía de Virus (*Current ICTV Taxonomy Release* | ICTV, s. f.). En esta base de datos se ha seleccionado una especie, y mediante un enlace directo a la base de datos del NCBI (*NCBI Virus*, s. f.), se ha obtenido la secuencia exacta de su genoma completo. En concreto se ha escogido la especie *Lucheng Rn rat coronavirus, Lucheng-19*, de la familia Coronaviridae con género Alphacoronavirus y de subgénero Luchacovirus.

#### **2.5.4. Cuasiespecies**

El estudio de la secuencia consenso proporciona una información muy fragmentaria de la composición genética y del potencial evolutivo del virus. Por lo tanto, es importante centrarse en el análisis de secuencias genómicas individuales de espectros mutantes (Domingo et al., 2006). Dado el extenso tamaño del genoma del SARS-CoV-2 y la notable cantidad de mutaciones presentes, este proyecto se enfoca aún más en la evaluación de los datos fundamentales de cuasiespecies.

### 2.5.4.1. Análisis de composición

Una parte muy importante en la caracterización genética de cuasiespecies consiste en la identificación de posibles sesgos en los tipos de mutaciones nucleotídicas a lo largo del genoma del conjunto de cuasiespecies.

Dentro del conjunto de datos proporcionados por el QuasiFlow, se encuentran los ficheros .vcf que contienen todas las mutaciones presentes en la cuasiespecie con respecto a la secuencia de referencia de Ómicron, sin incluir las mutaciones fijadas en el consenso. Para cada muestra se obtienen dos archivos .vcf diferentes. En uno se encuentran las mutaciones correspondientes a deleciones e inserciones y en otro las mutaciones SNP, estas son las variaciones en la secuencia de DNA que afectan solo a una base del genoma. En la *Figura 6* se puede observar un ejemplo de uno de los ficheros .vcf.

El estudio de sesgos de nucleótidos está centrado únicamente en aquellas mutaciones con una frecuencia de mutación mayor o igual al 1%. En el caso de que una mutación contenga una frecuencia mayor del 50%, quiere decir que se encuentra fijada en la población y sería una mutación correspondiente al consenso, no a la cuasiespecie, debido a que la secuencia consenso es la secuencia en la que cada posición del genoma representa la principal base de toda la cuasiespecie. Los archivos .vcf del QuasiFlow, permitirán la generación de tablas en las que, para cada muestra, se podrán ver representadas por separado cada tipo de mutación junto con los porcentajes de representación en todo el genoma. A partir de los datos recogidos en las tablas, se obtendrán conclusiones significativas con relación a los tratamientos antivirales a los que están sometidos cada muestra.

```

SARS_1_A_SNP.vcf
##fileformat=VCFv4.1
##source=VarScan2
##INFO=
##INFO=
##INFO=
##INFO=
##INFO=
##FILTER=
##FILTER=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
##FORMAT=
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT Sample
SARS_reference 15 - T C - PASS ADP=2154;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:17652:17652:17801:591:3:354:9:BE-1:37:36:9536:7465:391:200
SARS_reference 222 - A C - PASS ADP=49689;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:49689:49689:49686:6081:224:9:BE-1:40:38:39546:9514:491:117
SARS_reference 2495 - T C - PASS ADP=56251;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:56251:56251:55392:8401:494:9:BE-1:39:36:23845:31547:296:544
SARS_reference 2884 - T C - PASS ADP=48344;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:48344:48344:39806:459:1:149:9:BE-1:38:35:24317:15483:304:155
SARS_reference 3186 - C A - PASS ADP=34568;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:34568:34568:33848:1490:4:314:9:BE-1:37:37:28538:12518:826:664
SARS_reference 3464 - G T - PASS ADP=34778;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:34778:34778:33723:1823:2:294:9:BE-1:39:37:22807:11656:786:317
SARS_reference 3783 - T C - PASS ADP=37695;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:37695:37695:36819:452:1:214:9:BE-1:38:37:26918:9981:329:123
SARS_reference 4155 - G A - PASS ADP=35695;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:35695:35695:35152:525:1:474:9:BE-1:40:39:22923:12229:359:166
SARS_reference 4486 - C T - PASS ADP=33365;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:33365:33365:32864:4731:3:424:9:BE-1:37:37:12280:20784:164:389
SARS_reference 8108 - T C - PASS ADP=20969;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:20969:20969:20641:285:1:364:9:BE-1:38:36:91531:11481:149:136
SARS_reference 8554 - T C - PASS ADP=18193;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:18193:18193:17852:283:1:324:9:BE-1:38:36:8884:8966:183:100
SARS_reference 9524 - A G - PASS ADP=9886;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:9886:9886:9811:931:1:824:9:BE-1:39:39:4869:4842:44:49
SARS_reference 9575 - T C - PASS ADP=8653;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:8653:8653:8445:101:1:104:9:BE-1:42:40:5163:3285:65:36
SARS_reference 9593 - G A - PASS ADP=9239;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:9239:9239:9031:286:2:234:9:BE-1:40:38:5512:3519:146:600
SARS_reference 14655 - A G - PASS ADP=38674;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:38674:38674:38803:685:1:574:9:BE-1:37:36:21333:10985:285:300
SARS_reference 18881 - T C - PASS ADP=35452;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:35452:35452:31436:3989:11:264:9:BE-1:38:36:17735:13781:2582:1487
SARS_reference 18984 - T C - PASS ADP=36445;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:36445:36445:36823:366:1:834:9:BE-1:38:36:25835:10188:245:121
SARS_reference 19822 - T C - PASS ADP=33936;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:33936:33936:29267:530:1:784:9:BE-1:38:36:20846:9221:347:103
SARS_reference 19892 - C T - PASS ADP=31119;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:31119:31119:30122:994:3:194:9:BE-1:40:41:19054:11868:664:330
SARS_reference 19922 - T C - PASS ADP=31478;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:31478:31478:29561:1892:6:684:9:BE-1:39:37:31807:11374:969:923
SARS_reference 19922 - T C - PASS ADP=28373;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:28373:28373:27268:785:2:774:9:BE-1:39:38:17809:10178:335:458
SARS_reference 19922 - T C - PASS ADP=32566;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:32566:32566:29962:2512:7:774:9:BE-1:39:38:17655:12297:1246:1266
SARS_reference 19922 - T C - PASS ADP=32491;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:32491:32491:31916:571:1:764:9:BE-1:38:36:17361:14555:272:299
SARS_reference 19922 - T C - PASS ADP=8714;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:8714:8714:8818:295:3:309:9:BE-1:35:35:2459:5999:92:283
SARS_reference 19922 - T C - PASS ADP=6589;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:6589:6589:6514:75:1:144:9:BE-1:41:39:4660:2454:40:35
SARS_reference 19922 - T C - PASS ADP=36488;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:36488:36488:35878:589:1:434:9:BE-1:42:40:25965:9113:389:128
SARS_reference 21990 - T A - PASS ADP=29236;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:29236:29236:29208:1813:3:464:9:BE-1:41:39:18739:9481:594:419
SARS_reference 22610 - T C - PASS ADP=23176;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:23176:23176:22535:583:2:524:9:BE-1:38:36:10935:11680:249:334
SARS_reference 22610 - T C - PASS ADP=33978;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:33978:33978:33142:778:2:274:9:BE-1:40:39:9237:23985:212:558
SARS_reference 23526 - C T - PASS ADP=38379;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:38379:38379:38121:3438:18:844:9:BE-1:38:36:14673:13574:1668:1762
SARS_reference 24882 - A C - PASS ADP=29611;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:29611:29611:28666:846:2:864:9:BE-1:38:36:14478:14188:549:297
SARS_reference 25196 - C T - PASS ADP=24948;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:24948:24948:28561:4384:17:574:9:BE-1:35:35:16348:4213:3185:1279
SARS_reference 26029 - A G - PASS ADP=37128;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:37128:37128:36624:465:1:349:9:BE-1:37:37:24487:12137:338:152
SARS_reference 26220 - T C - PASS ADP=31444;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:31444:31444:30876:725:2:314:9:BE-1:38:37:18969:19687:258:467
SARS_reference 28235 - T C - PASS ADP=38229;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:38229:38229:29732:389:1:294:9:BE-1:38:37:19585:18227:251:138
SARS_reference 28285 - T C - PASS ADP=38793;WT=0;HET=1;HOM=0;NC=0 GT:GQ:SDP:DP:RD:AD:FREQ:PVAL:RBQ:ABQ:PDF:RDR:ADF:ADR 0/1:0:38793:38793:38319:468:1:524:9:BE-1:38:35:16755:13564:288:180

```

**Figura 6.** Fichero .vcf en el que aparecen las mutaciones SNP presentes en la muestra A1. En la columna REF aparece la secuencia original y en la columna ALT aparece la mutación, también se indica la posición de la mutación incluyendo la frecuencia de cada una de ellas.

## 2.5.4.2. Análisis de variabilidad y heterogeneidad

Otro de los archivos proporcionados por QuasiFlow es un report, que contiene datos, útiles para el análisis de la variabilidad genética. Como son el número de lecturas realizadas, así como la cobertura, número de haplotipos, la frecuencia de mutación, la entropía de Shannon, el índice de diversidad genética y el ratio TsTv además de gráficas que ayudan a entender de una más visual el contenido. Estos datos sirven de soporte para poder llevar a cabo un estudio de la complejidad y heterogeneidad de las cuasiespecies del ensayo.

### 2.5.4.2.1. Frecuencia de mutación

Esta se estudia con el de analizar la evolución de las poblaciones y la complejidad de las cuasiespecies La frecuencia mutación describe la proporción de un mutante que presenta en una población genómica, durante un periodo de tiempo determinado. Esta dependerá de la velocidad a la que se genera y de su capacidad de replicación en relación con otros genomas de la población.

La frecuencia de mutación puede variar entre diferentes regiones del genoma, por ello en el QuasiFlow se pueden diferenciar las frecuencias mutación de cada ORF y

la del genoma completo. Este trabajo solo se centra en la frecuencia de mutación del genoma completo (Domingo et al., 2006, 2012; Ortega Del Campo, 2020).

#### **2.5.4.2.2. Entropía de Shannon**

La entropía de Shannon proporciona una medida de diversidad basada en la frecuencia de haplotipos (Domingo et al., 2006), midiendo la porción de genomas diferentes presentes en una población de mutantes. En un conjunto de secuencias de ADN, se calcula teniendo en cuenta la distribución de los diferentes nucleótidos en las posiciones específicas.

En el caso del QuasiFlow el índice de Shannon se encuentra normalizado. Esto es, se obtendrá un valor máximo de 1 cuando todos los clones son diferentes y un mínimo de 0 cuando todos los clones sean idénticos entre sí (Gregori et al., 2016).

#### **2.5.4.2.3. Diversidad nucleotídica**

La diversidad nucleotídica se define como la media del número de nucleótidos distintos entre dos genomas cualquiera de las cuasiespecie, las cuales se toman aleatoriamente del mismo espectro de mutantes y para una misma región genómica. Esta medida cuantifica la cantidad de diferencias en los nucleótidos entre las secuencias. Cuanto mayor sea el número promedio de nucleótidos diferentes, mayor será la diversidad nucleotídica.

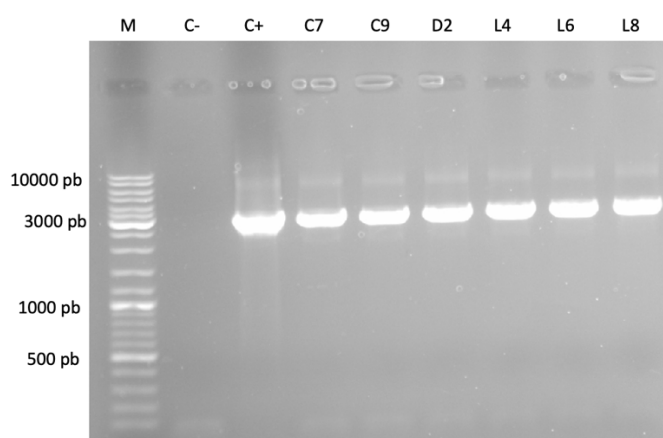
Esta indica cuán variable es una población en términos de secuencias de ADN. Una alta diversidad nucleotídica puede sugerir adaptación a diferentes ambientes, resistencia a patógenos o una historia evolutiva rica en cambios genéticos. Por otro lado, una baja diversidad nucleotídica puede indicar una población más homogénea o sometida a una selección fuerte (Domingo, s. f.; Gregori et al., 2016).

# 3

## Resultados

### 3.1 Amplificación del genoma

Como parte de los experimentos en el laboratorio, una vez aplicado el protocolo RT-PCR a las muestras, *Tabla 3*, con cada uno de los cebadores, *Tabla 7*, se han obtenido unas bandas (amplicones) de aproximadamente 3 Kb, lo cual fue corroborado en las electroforesis posteriores. A continuación, se expone una imagen de una de las visualizaciones de los gels generados a partir de las electroforesis.



**Figura 7.** Visualización de gel de las muestras C7, C9, D2, L4, L6 y L8 amplificadas con el primer nsp3-2.

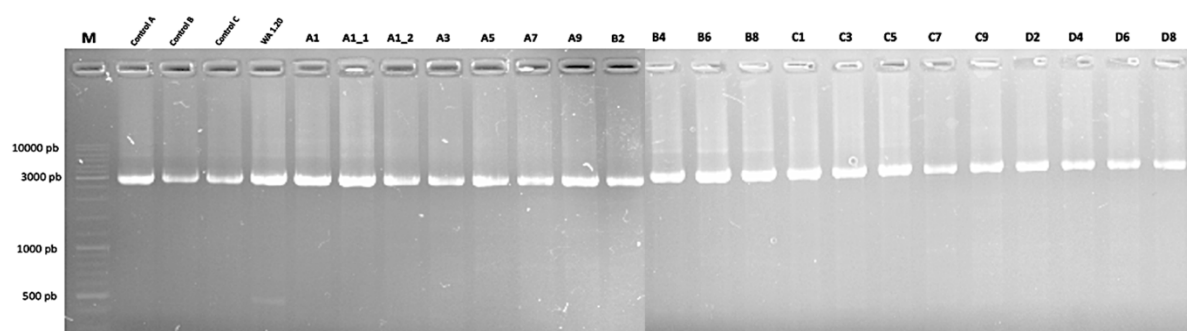
Como se puede observar en la *Figura 7*, en primer lugar, en el gel contamos con un control negativo, en el que no debe haber amplificación ya que no hay RNA y en este caso tampoco contaminación externa, con un control positivo y las muestras C7, C9, D2, L4, L6 y L8 que amplifican una clara banda de 3 Kb. Los dos primeros carriles

los podremos encontrar en todas las amplificaciones que se han realizado en todo el proceso. El resto de los geles con todas las amplificaciones se encuentran en el *Apéndice B1*.

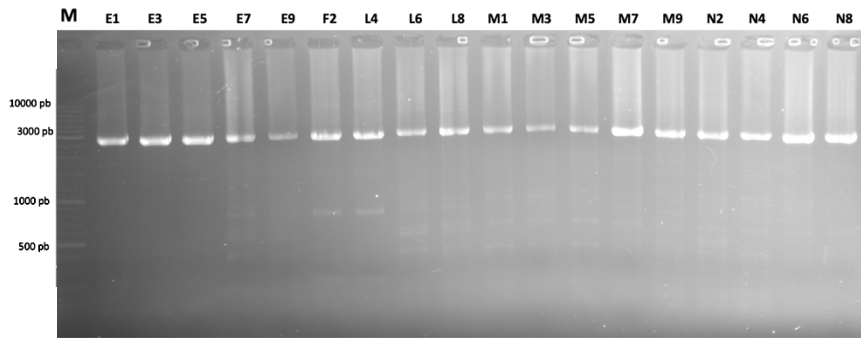
Tal y como se especifica en el *Apartado 2.5*, para poder realizar la secuenciación es necesario la preparación pools de cada muestra además de los controles indicados en el apartado 2.4, un total de 42 pools.

Una vez elaborados los pools, estos se amplificaron con el fin de comprobar la calidad de las bandas. Como podemos observar en la *Figura 8*, *Figura 9* y *Figura 10* en general no se observan bandas intermedias y todas las bandas que aparecen bien marcadas son las de 3Kb que son las correspondientes a los amplicones obtenidos. En la amplificación de los pools no encontramos ni controles positivos ni negativos, debido a que todos los amplicones que forman los pools ya se habían sometido a electroforesis por separados cada una con sus controles negativos y positivos correspondientes, *Apéndice B1*.

Cada uno de los pools contaba con 80  $\mu\text{L}$  y con una concentración de aproximadamente 350  $\text{ng}/\mu\text{L}$ . Valores que cumplen los requisitos impuestos por el SCBI. En el caso del índice OD260/280, los resultados no fueron los esperados, fue aproximadamente de 1.70 y el valor mínimo impuesto era de 1.80. Pero aun así al entregarlos en el laboratorio consideraron que las muestras no se encontraban lo suficientemente contaminadas, y aceptaron este valor (Koetsier et al., s. f.).



**Figura 8.** Visualización de gel de los pools 1. En la parte superior de cada carril se puede ver a que muestra pertenece. Aquí se encuentran tanto los amplicones obtenidos por María Jeréz como los míos ya que cada pools abarca el genoma completo.

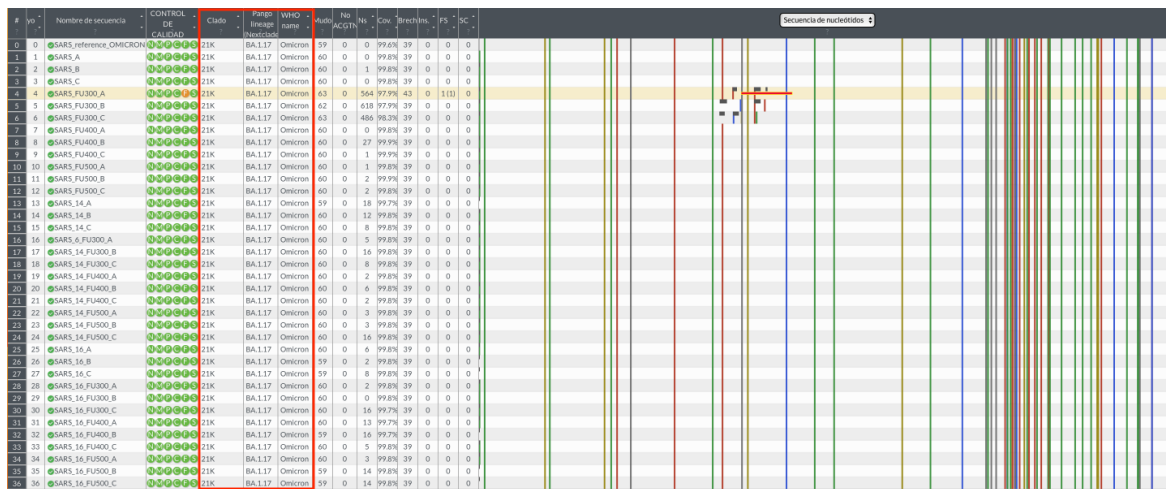


**Figura 9.** Visualización de geles de los pools 2. En la parte superior de cada carril se puede ver a que muestra pertenece. Aquí se encuentran tanto los amplicones obtenidos por María Jeréz como los míos ya que cada pools abarca el genoma completo.

## 3.2 Análisis bioinformático

### 3.2.1 Secuencias consenso

Una vez conseguidas las secuencias consenso, proporcionadas por el Centro de Bioinnovación SCBI, se han introducido en la herramienta bioinformática *NextClade* (Aksamentov et al., 2021) con el fin de verificar que tanto la variante como el pango y el clado de las secuencias son los esperados.



**Figura 10.** Imagen de NextClade en la que podemos confirmar que todas las muestras que tenemos pertenecen a la variante Ómicron del SARS-CoV-2 y que su pango es BA.1.17.

Como podemos comprobar en la *Figura 10*, tanto la variante como el pango coinciden con el mencionado en el apartado 2.1. Es decir, las muestras corresponden con la variante Omicrón con pango BA.1.17.

### 3.2.1.1 Alineamiento de secuencias

Las secuencias se realizaron un alineamiento múltiple mediante MUSCLE haciendo uso tanto del software JalView (Waterhouse et al., 2009) , *Figura 11*, como de MEGA11(Stecher et al., 2020; Tamura et al., 2021), *Figura 13*. Haciendo uso del alineamiento realizado con herramienta bioinformática de JalView, se ha llevado a cabo la construcción de la *Tabla 9* en la que aparecen las mutaciones que hay presentes en las distintas secuencias de nucleótidos con respecto a la secuencia de referencia de Ómicron.

Las mutaciones de las que hablaremos en este apartado, son del genoma ya amplificado y secuenciado y por lo tanto, el RNA viral pasa a ser DNA. Por ello no aparece uracilo (U) y aparece timina(T).

Como se puede observar en la *Tabla 9*, hay muy pocas mutaciones, un total de 19, en las secuencias consenso con respecto a la secuencia de referencia de Ómicron. La única región del genoma que presenta variaciones en la ORF1a desde la posición 265 a la base 13.215 del genoma del virus. Las muestras tratados sin péptido con análogo a 300µM son las que más mutaciones presentan en comparación con las otras muestras.

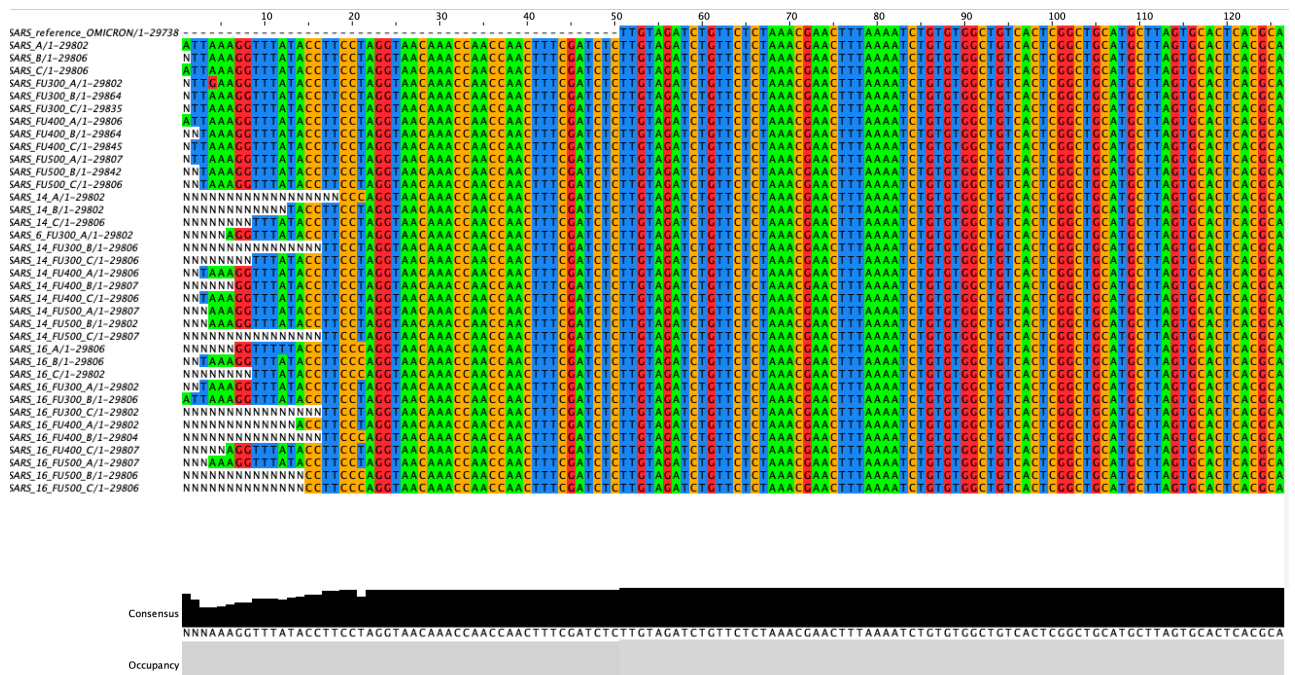


Figura 11. Ejemplo de la interfaz de salida del software JalView

**Tabla 9:** Mutaciones presentes en la secuencia consenso, con respecto a la secuencia de referencia de Ómicron del ensayo.

Péptido		5-FU (µM)	Réplica	Id	Mutaciones en las secuencias consenso con respecto a la secuencia de referencia de Ómicron
Péptido	(µM)				
-	-	-	A	A1	
-	-	-	B	A3	
-	-	-	C	A5	
-	-	300	A	A7	A--4→G, G--10898→A, G--11815→A
-	-	300	B	A9	T--11224→C, G--11815→A, G--11819→A, G--12253→A
-	-	300	C	B2	T--10922→C, A--11267→G, G--11819→A, C--11901→T
-	-	400	A	B4	
-	-	400	B	B6	
-	-	400	C	B8	
-	-	500	A	C1	
-	-	500	B	C3	
-	-	500	C	C5	
#14	70	-	A	C7	T--21→C
#14	70	-	B	C9	
#14	70	-	C	D2	
#14	70	300	A	D4	
#14	70	300	B	D6	
#14	70	300	C	D8	
#14	70	400	A	E1	
#14	70	400	B	E3	
#14	70	400	C	E5	
#14	70	500	A	E7	
#14	70	500	B	E9	
#14	70	500	C	F2	
#16	70	-	A	L4	A--12→T, T--21→C
#16	70	-	B	L6	T--21→C
#16	70	-	C	L8	T--21→C
#16	70	300	A	M1	
#16	70	300	B	M3	
#16	70	300	C	M5	
#16	70	400	A	M7	
#16	70	400	B	M9	T--21→C
#16	70	400	C	N2	
#16	70	500	A	N4	
#16	70	500	B	N6	T--21→C
#16	70	500	C	N8	T--21→C

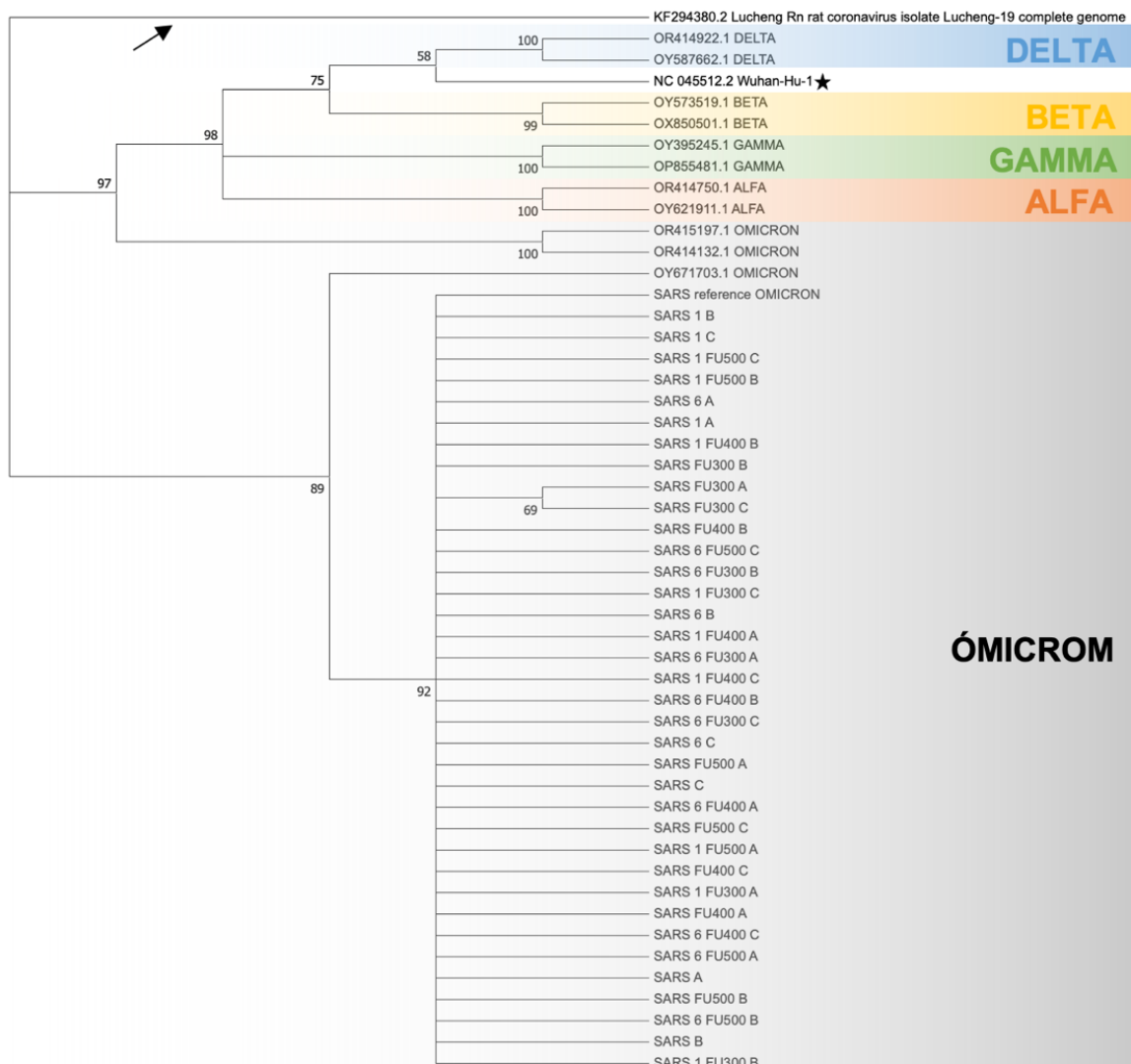
### 3.2.1.2 Análisis filogenético

Una vez alineadas las secuencias consenso se pudieron establecer las relaciones filogenéticas usando el método de máxima verosimilitud, con un Bootstrap de 600 repeticiones. El resultado de dicho estudio queda reflejado en el árbol filogenético de la *Figura 12*.

El árbol obtenido mediante Mega11 (Stecher et al., 2020; Tamura et al., 2021), ha sido enraizado manualmente seleccionando como grupo externo al género Alphacoronavirus. Como se puede observar en la *Figura 12*, el árbol presenta seis grupos claramente diferenciado. La secuencia externa se encuentra en primer lugar, debido a que es la secuencia que se ha enraizado y posteriormente se pueden encontrar las principales cinco variantes del SARS\_CoV\_2. Entre la variante Delta y Beta, se encuentra la secuencia original de Wuhan.

Por lo general, las ramas del árbol presentan una robustez considerablemente buena, lo que quiere decir que hay una alta probabilidad de ser datos correctos y de no haber surgido al azar.

La muestra SARS\_Reference, muestra de Ómicron que se usa como referencia para el estudio de las secuencias consenso y de cuasiespecies, junto con las muestras de los distintos ensayos antivirales, se encuentran englobadas en una misma rama del árbol. Pero como se puede observar, hay secuencias de Ómicron que se han englobado con secuencias del grupo Alfa, Gamma y Beta, ya que se encuentran en la misma rama. Estas secuencias son algunas de las que se han escogido para hacer un análisis filogenético más completo.



**Figura 12.** Árbol filogenético de máxima verosimilitud realizado con Mega11. En él se encuentran todas las secuencias con las que se ha realizado el alineamiento de nucleótidos. Se observan las cinco principales variantes de SARS-CoV-2 además de la secuencia de referencia de Wuhan original, señalada con una estrella y una secuencia externa de Alphacoronavirus señalada con una fecha. En cada rama del árbol aparece el valor de Bootstrap.

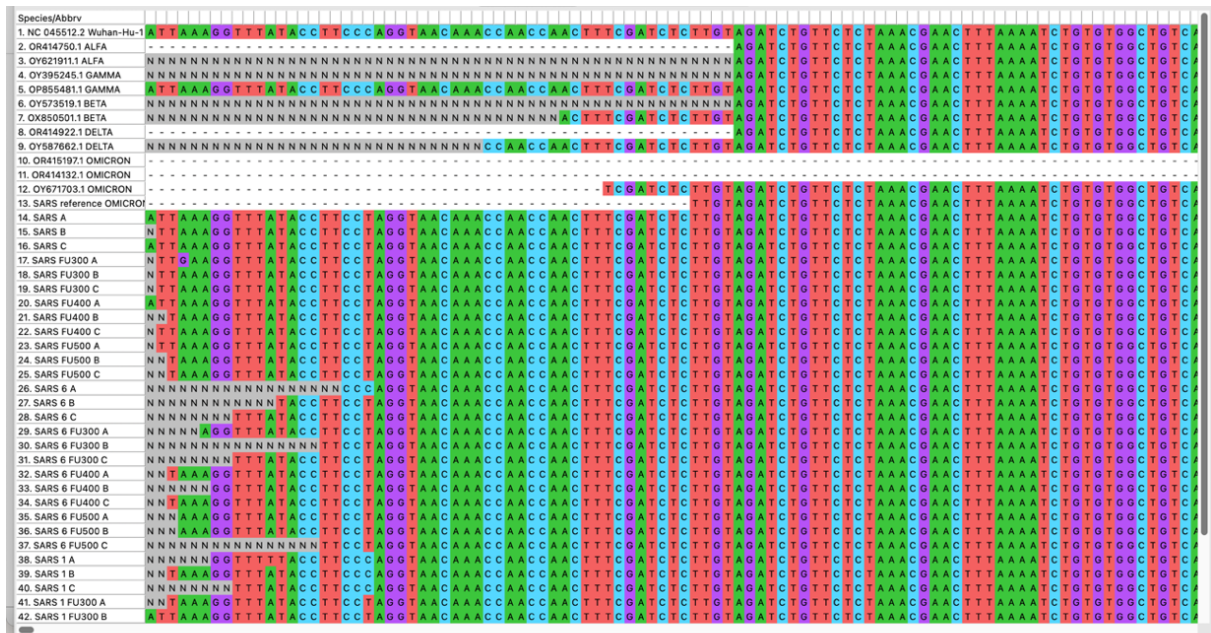


Figura 13. Ejemplo de la interfaz de salida del software MEGA11

## 3.2.2 Análisis de cuasiespecie

### 3.2.2.1 Análisis de composición

Con el fin de poder estudiar la evolución de las cuasiespecies de SARS-CoV-2 se examinaron los tipos de mutaciones para identificar posibles sesgos de sustitución, es decir, que mutaciones predominan sobre otras dependiendo del tratamiento antiviral.

Para el análisis de los sesgos de nucleótidos de todas las muestras, se han llevado a cabo tablas en las que quedan reflejadas todas las mutaciones. Estas se encuentran en el *Apéndice A.1*. Sin embargo, para poder extraer conclusiones de una forma más visual se han diseñado gráficos en los que podemos ver las mutaciones con sus porcentajes incluidos. Los datos utilizados para construir las gráficas representan un promedio obtenido de cada replica que compartía el mismo tratamiento antiviral, *Figura 13*.

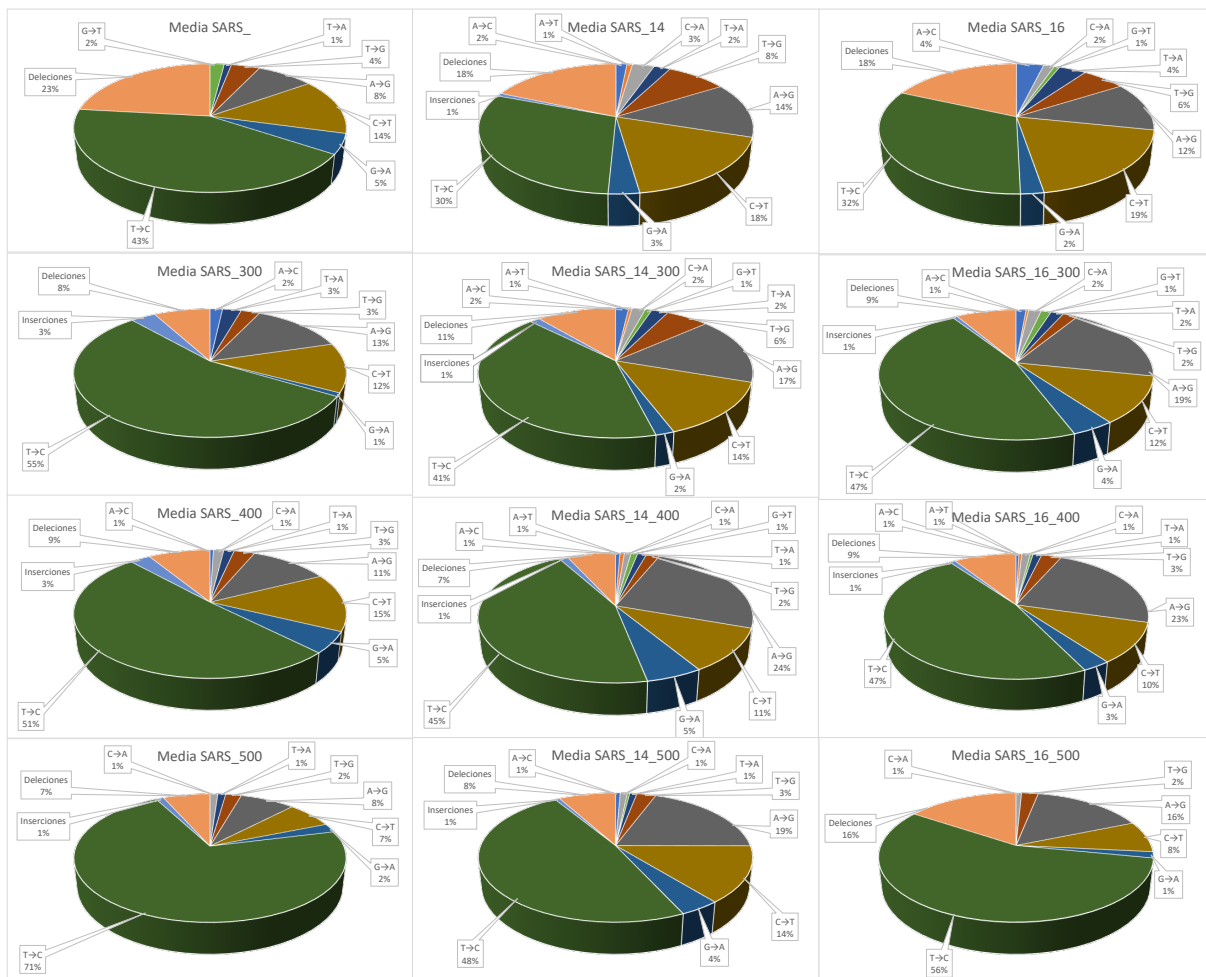
Las mutaciones de las que hablaremos en este apartado, son del genoma ya amplificado y secuenciado y por lo tanto, el RNA viral pasa a ser DNA. Por ello no aparece uracilo (U) y aparece timina (T).

Como podemos observar en la *Figura 14*, a medida que aumenta a la concentración de análogo en las muestras, el porcentaje de transición T→C aumenta considerablemente. Es importante señalar que este porcentaje es más alto en las

muestras que no se encuentran tratadas con péptido. Además, las transiciones C→T y A→G, también están significativamente sobrepresentadas, pero en menor cantidad.

Las transversiones T→G también se puede observar que están presentes en todas las gráficas, pero en este caso, al añadir mayor análogo al cultivo, la presencia de estas sustituciones se ve reducida. En el caso de las transversiones A→G, ocurre el efecto contrario, al tratar las muestras con mayores concentraciones de análogo la frecuencia de estas experimenta un pequeño aumento.

Es importante señalar que las mutaciones InDels, inserciones y deleciones han sido abundantes en el virus. En este caso las deleciones predominan. Al tratar con análogo el porcentaje de deleciones desciende, la concentración de este no implica una variación significativa en este tipo de mutaciones.



**Figura 14.** Gráficas representativas del porcentaje de mutaciones en cada ensayo antiviral. Cada gráfica representa un promedio de las mutaciones recogidas en cada replica tratadas con un mismo ensayo antiviral. Las primeras tres gráficas son las muestras sin análogo. Las segundas tres gráficas son las muestras tratadas con análogo a 300 µM. Las terceras tres gráficas son las muestras tratadas con análogo a 400 µM. Y las tres últimas gráficas son las muestras tratadas con análogo a 500 µM.

Cada columna representa en primer lugar las muestra sin péptido, en segundo lugar, las muestras con #14 y en tercer lugar las muestras con péptido #16.

### 3.2.2.2 Análisis de variabilidad y heterogeneidad

Haciendo uso de los reports proporcionados por el flujo de trabajo QuasiFlow se han extraído atributos como el número de lecturas, la cobertura, frecuencia de mutación, índice de Shannon, índice de diversidad genética y el ratio TsTv. Con ellos se han elaborado una serie de tablas con el fin de poder hacer un estudio de cada uno de valores para cada muestra de los ensayos. Estas tablas se encuentran recogidas en las tablas *Tabla A3.1* y *Tabla A3.2* del *Apéndice A*. Con el propósito de simplificar la comprensión de los resultados, se han confeccionado gráficos correspondientes a cada uno de los atributos previamente mencionados. Los datos utilizados para construir las gráficas representan un promedio obtenido de cada replica que compartía el mismo tratamiento antiviral. De esta forma se podrá visualizar de manera clara cómo han experimentado cambios los valores en las diversas muestras, *Figura 15*. Dado que cada barra de los diagramas representan un promedio de valores, se han incluido barras de error para poder observar la variabilidad de los datos de cada replica que forma cada ensayo antiviral.

En algunos de los atributos como el número de haplotipos, frecuencia de mutación, entropía de Shannon y diversidad nucleotídica en el Report, no se encontraron valores para la muestra que no fue sometida a ningún tratamiento antiviral. Esto es debido a que no se pudo estimar la variabilidad ya que presenta un número muy reducido de mutaciones en comparación con el resto de las lecturas, y el programa no lo detecta debido al filtro que se utiliza.

La cobertura mide el promedio del número de lecturas por posición de las 30Kb del virus. Por lo general se ha mantenido constante en todas las muestras, a excepción de las tratadas con el péptido #16, cuya cobertura disminuye en relación con las otras muestras. Se puede decir que los resultados presentan una fiabilidad elevada.

Con relación al índice de Shannon normalizado, establece una medida de la diversidad en cada cuasiespecie. En general estos valores se mantienen constantes y se encuentran más bien cercanos a 0, por lo que se puede concluir que en general los clones resultantes no varían mucho entre sí.

Además, es importante señalar que por lo general se han obtenido pocos haplotipos, pero aumenta al incrementar la concentración de análogo. Se ha observado que las muestras tratadas con péptido y sin análogo, son las que menos haplotipos presentan. Y la muestra que presenta mayor número de haplotipos con mucha diferencia de las otras muestras es la tratada con análogo a 500 y péptido #14.

Con respecto a la frecuencia de mutación en la mayoría de las muestras es significativamente baja. En la muestra de virus tratado a 300  $\mu$ M hay un aumento destacable en comparación con las otras muestras. En relación con el resto de muestras, por lo general, al aumentar la concentración de análogo, aumenta la frecuencia de mutación.

La diversidad nucleotídica es una medida de la heterogeneidad. En estas muestras la heterogeneidad no fue especialmente alta. Por lo tanto, hay un bajo número promedio de nucleótidos diferentes.

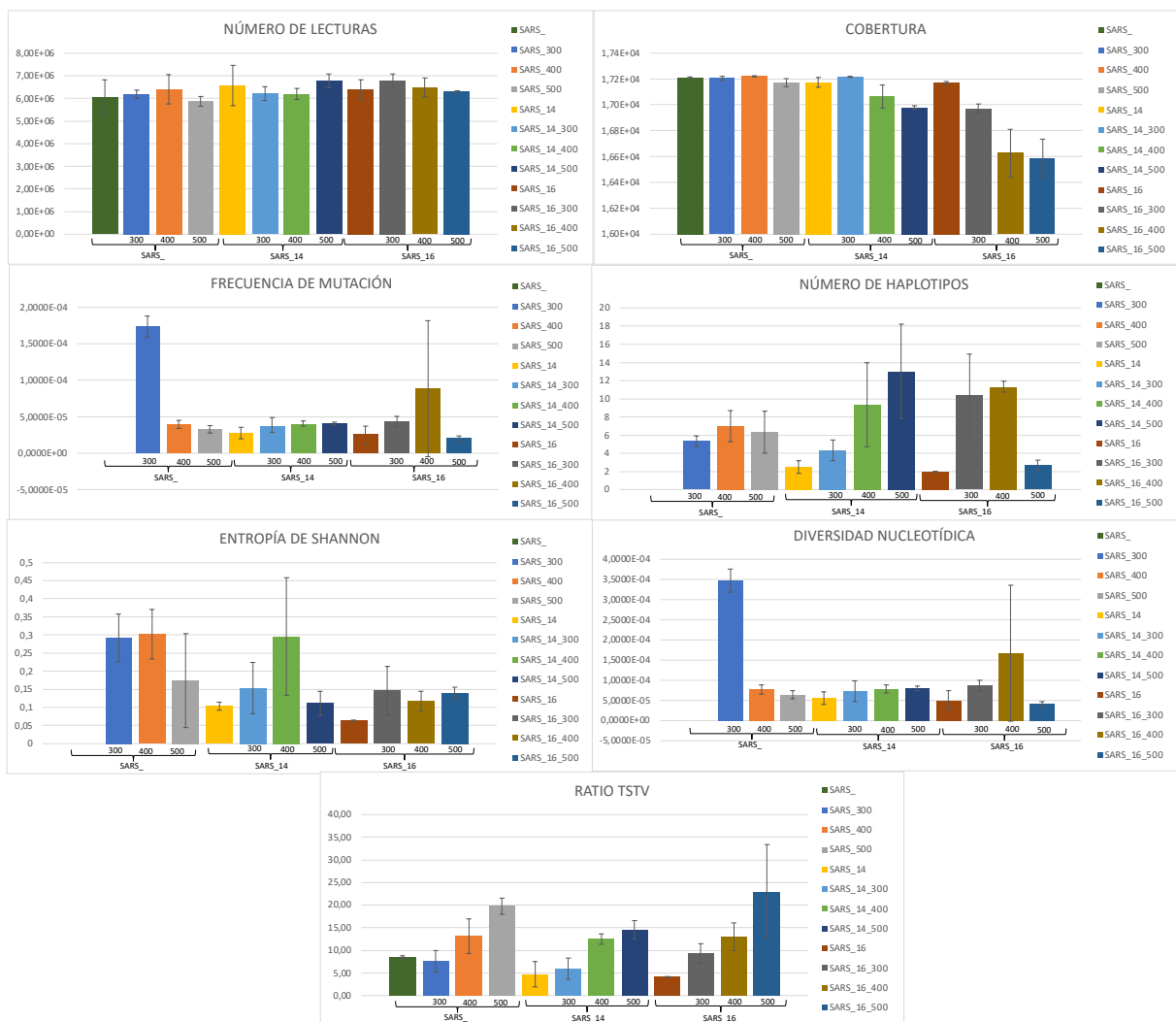
Con respecto al tipo de mutaciones, el ratio Ts/Tv, indica el número de transiciones con respecto al número de transversiones. Debido a que los valores obtenidos son generalmente altos, se puede concluir que por lo general hay muchas más transiciones. Esto queda reflejado en el Apéndice A1.

La frecuencia de mutación es un valor muy importante a la hora de estudiar la variabilidad del virus y también para ver como ha afectado los diferentes ensayos antivirales a las muestras. Por ellos, gracias a los datos proporcionados por el QuasiFlow, se ha podido estudiar la frecuencia de mutación de cada ORF en las diferentes muestras, *Figura 16* y *Figura 17*.

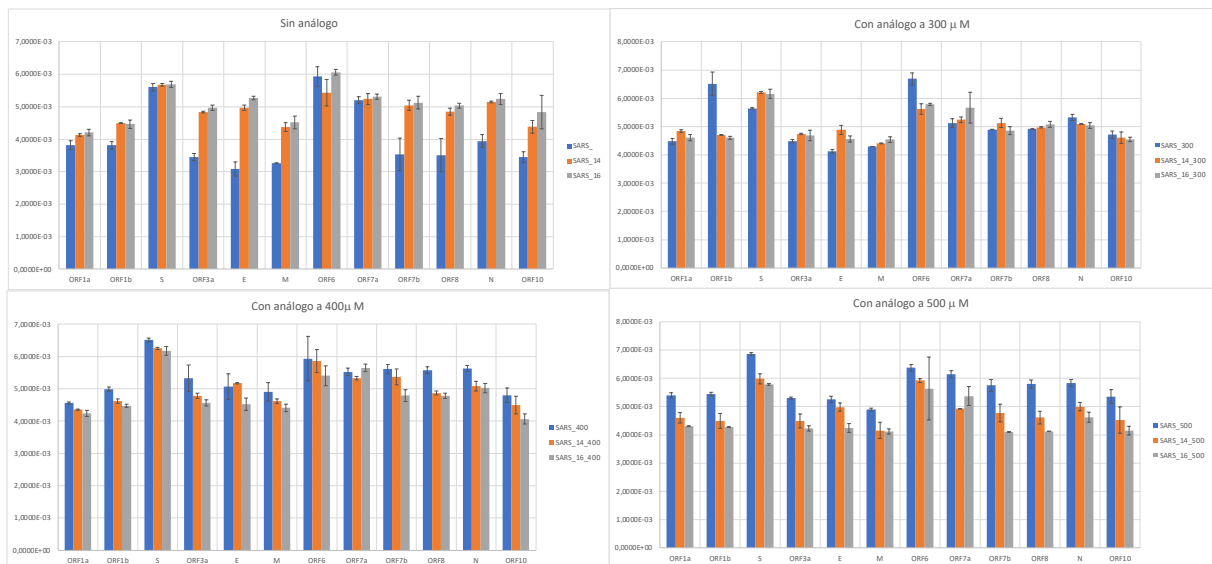
En la *Figura 16*, se puede observar la variación de la frecuencia de mutación en presencia de análogo a diferentes concentraciones. En el caso de las muestras sin análogo, la frecuencia de mutación aumenta en las muestras tratadas con péptido, sobre todo en presencia del péptido #16. Por lo general, en todas las regiones del genoma, los valores no varían significativamente. Pero se puede señalar la región ORF6 en la que la frecuencia de mutación es ligeramente más alta que en las demás regiones.

En las muestras tratadas con análogo, ocurre lo mismo que en las sin tratar, no hay variaciones exageradas en las diferentes regiones del genoma. En las muestras tratadas a 300  $\mu$ M, hay dos regiones relevantes, la ORF1b y la ORF6, en ambas las

frecuencia de mutación de las muestras tratadas sin péptido es relevante con respecto a las otras regiones y a las otras muestras. Por lo general en las muestras tratadas a 300  $\mu\text{M}$ , la frecuencia de mutación aumenta con la presencia de los péptidos, sobre todo del péptido #14. En las muestras tratadas a 400  $\mu\text{M}$ , la presencia de los péptidos disminuye la frecuencia de mutación en especial la presencia del péptido #16. Cabe destacar la región S y ORF6, en las que la frecuencia de mutación es ligeramente superior. En las muestras tratadas a 500  $\mu\text{M}$ , ocurre lo mismo que en las muestras tratadas a 400  $\mu\text{M}$ .



**Figura 15.** Análisis de variabilidad del SARS-CoV-2 en los ensayos antivirales. Las barras de error, que representan la desviación estandar de cada promedio de valores que forman cada barra del gráfico.

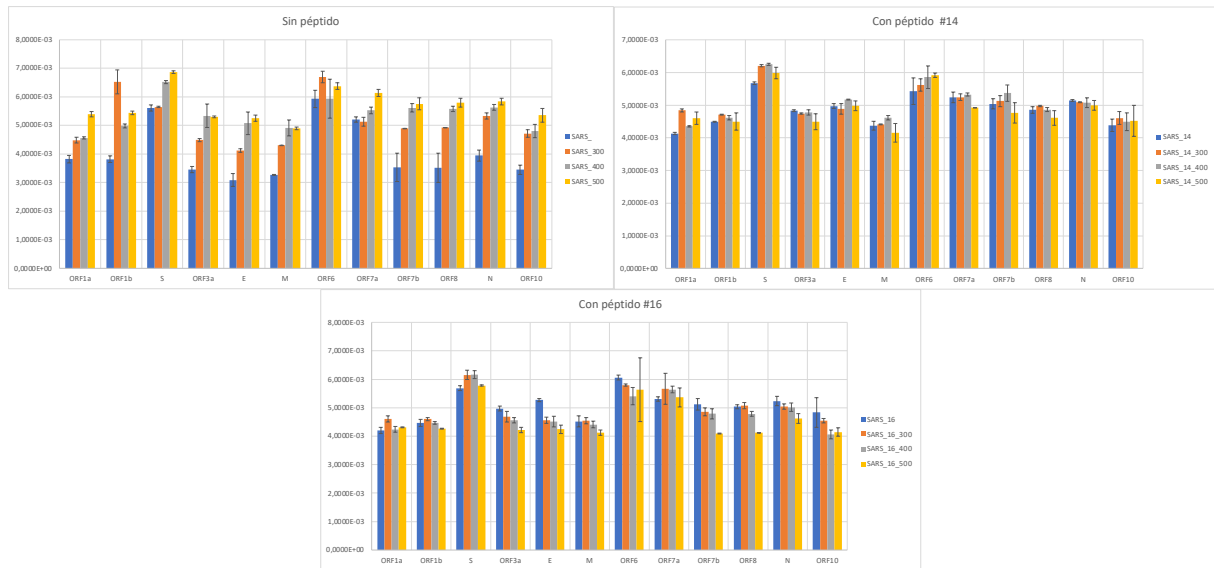


**Figura 16.** Gráficas de frecuencias de mutación en muestras sin péptido. En ellas se pueden observar las barras de error, que representan la desviación estandar de cada promedio de valores que forman cada barra del gráfico.

En la *Figura 17*, se muestran las gráficas correspondientes a la frecuencia de mutación en las que se puede ver el efecto de cada péptido en las diferentes concentraciones del análogo. En las muestras que no presentan péptido, por lo general hay un aumento de la frecuencia de mutación al elevar la concentración de análogo. Las regiones ORF1b, S y ORF6, presentan las frecuencia de mutación más altas.

En el caso de las muestras tratadas con el péptido #14, la concentración de análogo no influye demasiado en la frecuencia de mutación, aun así, la región S y ORF6, presentan las frecuencias de mutación más elevadas en relación con las demás regiones.

Para las muestras tratadas con el péptido #16, se podría decir que la adición de análogo no influye significativamente como ocurre en las muestras sin péptido. Por lo general, la frecuencia de mutación tiende a disminuir ligeramente cuando aumenta la concentración de análogo. Las regiones S y ORF6, destacan por presentar las frecuencias de mutación más elevadas y no sigen el patrón de las demás, sino que en ellas la muestra que menor frecuencia de mutación presenta es en el caso de la región S, las muestras tratadas con análogo a 300  $\mu\text{M}$  y 400  $\mu\text{M}$ . Y en el caso de la ORF6 con mucha diferencia la muestra 500  $\mu\text{M}$ .



**Figura 17.** Gráficas de frecuencias de mutación en muestras con y sin péptido tratadas con diferentes concentraciones de análogo. En ellas se pueden observar las barras de error, que representan la desviación estandar de cada promedio de valores que forman cada barra del gráfico.



# 4

## Discusión y líneas futuras

Gran parte de la literatura científica sobre estudios genómicos de virus, se centra en el análisis de mutaciones presentes en las variantes más comunes, es decir, las mutaciones que se encuentran en las secuencias genéticas consenso. Sin embargo, las mutaciones menos frecuentes, no detectadas en estas secuencias, presentan gran importancia. Estas mutaciones minoritarias son estudiadas en el contexto de la cuasiespecie, pero debido a su complejidad, rara vez se les presta la debida atención. Es importante destacar que estas mutaciones minoritarias podrían tener un impacto significativo ya que a determinadas condiciones podrían conferir ventajas selectivas al virus. En consecuencia, en el futuro, estas mutaciones podrían volverse dominantes y desempeñar un papel fundamental en la evolución y comportamiento del virus (Domingo et al., 2006). Además, hasta el momento no se ha encontrado bibliografía en la que se haga secuenciación masiva del genoma completo de cuasiespecies, ni en la que se hayan identificado haplotipos, secuencias reconstruidas dentro de la cuasiespecie. Por ello este trabajo se centra a nivel de genoma completo en el estudio de la cuasiespecie y sus variables principales, aplicando a demás varias terapias antivirales. Por todo ello, he optado por elegir esta línea de investigación como enfoque para mi proyecto de Fin de Grado.

#### 4.1. Análisis filogenético

En cuanto al alineamiento de las secuencias, se pueden observar un gran número de secuencias altamente conservadas. Se ha podido corroborar que las secuencias obtenidas pertenecen a la variante Ómicron, comparando además con secuencias de otras variantes donde se pueden ver sus semejanzas o diferencias.

Cada variante del SARS-CoV-2, que se identifica mediante una letra griega, no se basa en una secuencia de referencia estándar. Estas se caracterizan por las mutaciones que comparten todas las secuencias pertenecientes a una misma variante. Estas mutaciones pueden conferir al virus ventajas en términos de cómo se manifiesta, su capacidad para adaptarse a diversos entornos, su resistencia a otros patógenos o una historia evolutiva que involucra una serie de cambios genéticos.

La variante Ómicron del SARS-CoV-2 se destaca por su extenso conjunto de mutaciones, superando en número a las observadas en las variantes previas en aproximadamente diez veces. Esta profusión de mutaciones ha suscitado preocupación en relación con la posible disminución de la efectividad de las vacunas (Cookson & Barnes, 2021; «Variante ómicron del SARS-CoV-2», 2023).

En este proyecto se ha usado como referencia, una secuencia de SARS-CoV-2 de la variante Ómicron. Esta secuencia fue de la cual partieron los países del ensayo antiviral. Tras haber hecho el alineamiento de las secuencias consenso de las diferentes muestras se han identificado las mutaciones que tienen con respecto a la secuencia de referencia. Se han detectado pocas variaciones, esto es lo esperado, debido a que, en la mutagénesis letal, prácticamente nunca cambia la secuencia consenso, debido a que hay un equilibrio mutación-selección. Por lo tanto, aunque en la cuasiespecie haya un elevado número de mutaciones, ninguna de ellas va a dominar como para aparecer estadísticamente en la secuencia consenso. Los cambios que aparezcan en las secuencias consenso son resultados importantes debido a que puede significar que se está seleccionando esa secuencia.

## 4.2. Análisis de composición

Con respecto al análisis de composición, analizando los datos obtenidos del QuasiFlow, se ha podido lograr una visión general de cuál es el porcentaje de cada mutación en cada uno de los diferentes tratamientos antivirales.

Aquí se ha podido observar una dominancia de la mutación T→C. Esta mutación contradice a lo visto en otros artículos, ya que estos afirman que la mutación más usual en el genoma del SARS-CoV-2 es la C→T, esto puede deberse a la acción de una proteína antiviral APOBEC (Li et al., 2020, 2020; Simmonds, 2020).

En ensayos anteriores a este proyecto, realizados con el análogo que se usa en las muestras de este ensayo, se observó una sobrerepresentación de sustituciones T→C y G→A. Estos datos teóricos coinciden con los obtenidos al realizar el estudio de la cuasi-especie de las muestras. En todas las muestras, en especial, las que no están tratadas con péptido, se observa un porcentaje muy alto de la sustitución T→C.

Las mutaciones detectadas en las distintas muestras podrían ser consecuencia de la terapia antiviral o servir como un mecanismo de adaptación del virus (Amicone et al., 2022; Geraghty et al., 2021). Es posible que hayan surgido mutaciones que proporcionen ventajas al virus durante el proceso de replicación en presencia del análogo antiviral, y que esta mutación haya prevalecido en toda la población viral, lo que explicaría su aparición en la secuencia consenso. Sin embargo, es interesante notar que esta mutación no se observa en la muestra de referencia del linaje Ómicron.

## 4.3. Análisis de variabilidad y heterogeneidad

Los virus de RNA presentan una frecuencia de mutación de  $10^{-3}$  a  $10^{-5}$  mut/nt (Holland et al., 1992) mucho mayores con relación a la frecuencia de mutación de las células humanas,  $10^{-8}$  mut/nt. Teniendo en cuenta estos datos, los resultados obtenidos indican que la frecuencia de mutación es más alta de lo establecido. La frecuencia de mutación media se encuentra  $5,2 \times 10^{-5}$  mut/nt. El motivo de este aumento puede ser debido a la presencia del análogo y también a que se está basando en los resultados de las variantes que componen las cuasi-especies y no en secuencias consenso.

Como se ha descrito anteriormente, el análogo, aumenta la frecuencia de mutación. Sin embargo, una dosis muy alta tanto de análogo como de péptido puede producir una inhibición de la replicación del virus debido a la mutagénesis letal. Esto haría que

aparezcan menos mutaciones y por consiguiente la variabilidad sea más baja. Por ello se puede observar una caída tanto de la frecuencia de mutación como de la diversidad de nucleótidos (Aiewsakun et al., 2023; Amicone et al., 2022).

Además, la influencia de los péptidos con sus funciones de bloquear la unión con el cofactor que activa la proteína nsp14, y, por lo tanto, inhibir la actividad exonucleasa correctora y con la función de la actividad metiltransferasa de la nsp16, relacionada con evitar el reconocimiento del genoma viral y la acción de las enzimas por parte del sistema inmunitario innato, influyen también en un aumento de la frecuencia de mutación y del incremento de número de haplotipos en las diferentes muestras.

Con respecto al número de haplotipos se ha observado que las muestras tratadas con péptido y sin análogo, son las que menos haplotipos presentan. Esto tiene sentido y quiere decir que parece que los péptidos disminuyen el número de haplotipos, es decir la variabilidad en heterogeneidad genética es menor cuando se aplica péptido que cuando se aplica el análogo porque introduce mutaciones. La muestra que presenta mayor número de haplotipos con mucha diferencia de las otras muestras es la tratada con análogo a 500  $\mu\text{M}$  y péptido #14 y en esto tiene que ver la función del péptido. Debido a que el péptido #14 interviene en la actividad ExoN correctora del SARS-CoV-2, debido a que se está inhibiendo a la exonucleasa e indica que está funcionando como debe.

Los datos obtenidos a lo largo de este trabajo, son **datos preliminares** de un proyecto de investigación. Hasta el momento los únicos estudios que se han hecho de la secuenciación de genoma completo son los descritos en esta memoria. Por lo tanto, con perspectivas a futuro, se pretende estudiar más en detalle todos los datos obtenidos a partir de este primer estudio.

Es importante hacer referencia a las limitaciones que este estudio conlleva. Principalmente, se relacionan con el hecho de que las células utilizadas para cultivar los experimentos son de una especie distinta a la humana, lo que implica características divergentes, especialmente en lo que respecta al sistema inmunológico. Dicha característica juega un papel crucial en el proceso de infección viral. Por consiguiente, se plantea la perspectiva de llevar a cabo estas terapias antivirales utilizando células que se asemejen más a las células humanas, o incluso utilizando células humanas en investigaciones futuras.

Las células VeroE6 fueron las seleccionadas para ser infectadas, estas células son muy útiles para la propagación y el asilamiento eficiente el virus, lo que proporciona ventajas en la evolución de terapias antivirales. Estas presentan muchos beneficios a la hora de hacer el estudio como puede ser su sensibilidad al virus, facilidad de cultivo, disponibilidad y ética. Pero las células Vero son deficientes de interferón. A diferencia de células normales de mamíferos, estas no secretan interferón alfa o beta cuando se ven infectadas por un virus. El interferón es una proteína producida por las células del sistema inmunológico en respuesta a infecciones virales. Tiene la capacidad de modular la respuesta inmunitaria y ayudar en la defensa contra las infecciones. Esto implica que el virus aumente su velocidad de propagación y que el proceso de infección no sea igual que en las células humanas 1.

A pesar de que ni el análogo ni los péptidos presentaron toxicidad para Vero E6 a las concentraciones ensayadas cabe la posibilidad de que este tratamiento presente cierta cierta citotoxicidad para células humanas. Por tanto, sería interesante estudiar el efecto de la estrategia antiviral en células humanas.



# 5

## Conclusión

A continuación, se enumerarán las conclusiones derivadas de los resultados obtenidos en este trabajo:

1. A lo largo del proceso de amplificación se han obtenido amplicones de 3 Kb, cada uno de ellos abarcaba una posición concreta del genoma del virus.
2. Se ha demostrado a través del análisis filogenético, que las muestras con las que se han desarrollado los ensayos son de la variante Ómicron de SARS-CoV-2.
3. Tras el análisis de las secuencias consenso obtenidas, se han identificado pocas mutaciones con respecto a la secuencia de referencia de Ómicron que se ha usado para este ensayo. Las únicas mutaciones que se pueden encontrar, están en la ORF1a y en concreto en las muestras tratadas con análogo a 300  $\mu$ M sin péptido.
4. Gracias al estudio de composición, se observa que la mutación que más predomina en las secuencias es la T $\rightarrow$ C, a diferencia de lo que aparece en la literatura científica que la mutación más observada es la C $\rightarrow$ T. Esto coincide con estudios anteriores que se han realizado con el análogo que se usa en este ensayo.
5. La frecuencia de mutación, la heterogeneidad genética y la diversidad nucleotídica aumenta con la concentración de análogo que se corresponde con disminuciones de infectividad del ensayo. El análogo induce mutaciones en el

genoma del virus provocando así la eliminación de la infección vírica, esto se conoce como mutagénesis letal.

6. Los péptidos disminuyen el número de haplotipos, es decir la variabilidad en heterogeneidad genética es menor cuando se aplica péptido que cuando se aplica el análogo.
7. El mayor número de haplotipos corresponde con la muestra tratada con péptido #14 y análogo a 500  $\mu\text{M}$ . Esto indica que el péptido está ejerciendo su función de inhibición de las actividades ExoN correctora del SARS-CoV-2.

# Referencias

- Aiewsakun, P., Phumiphanjark, W., Ludowyke, N., Purwono, P. B., Manopwisedjaroen, S., Srisaowakarn, C., Ekronarongchai, S., Suksatu, A., Yuvaniyama, J., & Thitithanyanont, A. (2023). Systematic Exploration of SARS-CoV-2 Adaptation to Vero E6, Vero E6/TMPRSS2, and Calu-3 Cells. *Genome Biology and Evolution*, 15(4), evad035. <https://doi.org/10.1093/gbe/evad035>
- Aksamentov, I., Roemer, C., Hodcroft, E., & Neher, R. (2021). Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software*, 6(67), 3773. <https://doi.org/10.21105/joss.03773>
- Amicone, M., Borges, V., Alves, M. J., Isidro, J., Zé-Zé, L., Duarte, S., Vieira, L., Guiomar, R., Gomes, J. P., & Gordo, I. (2022). Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. *Evolution, Medicine, and Public Health*, 10(1), 142-155. <https://doi.org/10.1093/emph/eoac010>
- Argimón Pallàs, J. M., & Jiménez Villa, J. (1998). Artículo original (II): Material y métodos. *Atención Primaria*, 22(6), 387-390.
- BLAST: Basic Local Alignment Search Tool*. (s. f.). Recuperado 1 de septiembre de 2023, de <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- Cookson, C., & Barnes, O. (2021, diciembre 20). Omicron: What we know about Covid strain prompting fresh global restrictions. *Financial Times*. <https://www.ft.com/content/42c5ff3d-e676-4076-9b9f-7243a00cba5e>
- COVID - Coronavirus Statistics—Worldometer*. (s. f.). Recuperado 6 de septiembre de 2023, de <https://www.worldometers.info/coronavirus/>
- Cov-Lineages*. (s. f.). Recuperado 1 de septiembre de 2023, de [https://cov-lineages.org/lineage\\_list.html](https://cov-lineages.org/lineage_list.html)
- Current ICTV Taxonomy Release | ICTV*. (s. f.). Recuperado 1 de septiembre de 2023, de <https://ictv.global/taxonomy>
- Díaz Martínez, L. (2016). *Variabilidad genética de cuasiespecies de virus RNA y ssDNA in vivo ante cambios en la tasa de error*. Universidad de Málaga.
- Díaz-Martínez, L., Bricchette-Mieg, I., Pineño-Ramos, A., Domínguez-Huerta, G., &

- Grande-Pérez, A. (2018). Lethal mutagenesis of an RNA plant virus via lethal defection. *Scientific Reports*, 8(1), 1444. <https://doi.org/10.1038/s41598-018-19829-6>
- Documentation – Oracle VM VirtualBox*. (s. f.). Recuperado 1 de septiembre de 2023, de <https://www.virtualbox.org/wiki/Documentation>
- Domingo, E. (s. f.). *Composition, Complexity, Quasispecies, Dynamics, and Biological Implications*.
- Domingo, E., Martín, V., Perales, C., Grande-Pérez, A., García-Arriaza, J., & Arias, A. (2006). Viruses as Quasispecies: Biological Implications. *Quasispecies: Concept and Implications for Virology*, 299, 51-82. [https://doi.org/10.1007/3-540-26397-7\\_3](https://doi.org/10.1007/3-540-26397-7_3)
- Domingo, E., & Schuster, P. (Eds.). (2016). *Quasispecies: From Theory to Experimental Systems* (Vol. 392). Springer International Publishing. <https://doi.org/10.1007/978-3-319-23898-2>
- Domingo, E., Sheldon, J., & Perales, C. (2012). Viral Quasispecies Evolution. *Microbiology and Molecular Biology Reviews*, 76(2), 159-216. <https://doi.org/10.1128/MMBR.05023-11>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792-1797. <https://doi.org/10.1093/nar/gkh340>
- El SARS-CoV-2 y sus mutaciones | Vacunas / Asociación Española de Vacunología*. (s. f.). Recuperado 1 de septiembre de 2023, de <https://www.vacunas.org/el-sars-cov-2-y-sus-mutaciones/>
- Enfermedad del coronavirus 2019 (COVID-19)*. (2020, febrero 11). Centers for Disease Control and Prevention. <https://espanol.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html>
- Expósito Lara, A., Feria Díaz, G. E., González Benítez, S. N., Miguel Soca, P. E., Expósito Lara, A., Feria Díaz, G. E., González Benítez, S. N., & Miguel Soca, P. E. (2021). Variantes genéticas del SARS-CoV-2 y sus implicaciones clínicas. *MEDISAN*, 25(6), 1424-1446.
- Ferron, F., Subissi, L., Silveira De Moraes, A. T., Le, N. T. T., Sevajol, M., Gluais, L., Decroly, E., Vonrhein, C., Bricogne, G., Canard, B., & Imbert, I. (2018). Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proceedings of the National Academy of Sciences*,

115(2). <https://doi.org/10.1073/pnas.1718806115>

*GenBank Overview*. (s. f.). Recuperado 1 de septiembre de 2023, de <https://www.ncbi.nlm.nih.gov/genbank/>

*Genus: Betacoronavirus | ICTV*. (s. f.). Recuperado 23 de agosto de 2023, de <https://ictv.global/report/chapter/coronaviridae/coronaviridae/betacoronavirus>

Geraghty, R., Aliota, M., & Bonnac, L. (2021). Broad-Spectrum Antiviral Strategies and Nucleoside Analogues. *Viruses*, 13(4), 667. <https://doi.org/10.3390/v13040667>

González, F. J. V. (s. f.). *Evolutionary and structural analysis of nsp10, nsp14 and nsp16 proteins for the design of a pan-coronaviral antiviral therapy*.

González, M. S., & Rosales, J. M. L. (s. f.). *TÉCNICAS DE SECUENCIACIÓN MASIVA (NGS)*.

*GreenSafe DNA Gel Stain—Canvax Reagents SL*. (s. f.). Recuperado 1 de septiembre de 2023, de <https://www.canvaxbiotech.com/dna-electrophoresis/loading-buffers-staining/greensafe-dna-gel-stain/>

Gregori, J., Perales, C., Rodríguez-Frias, F., Esteban, J. I., Quer, J., & Domingo, E. (2016). Viral quasispecies complexity measures. *Virology*, 493, 227-237. <https://doi.org/10.1016/j.virol.2016.03.017>

*Helicos BioSciences Corporation*. (s. f.). Flagship Pioneering. Recuperado 1 de septiembre de 2023, de <https://www.flagship pioneering.com/companies/helicos>

Holland, J. J., De La Torre, J. C., & Steinhauer, D. A. (1992). RNA Virus Populations as Quasispecies. En J. J. Holland (Ed.), *Genetic Diversity of RNA Viruses* (Vol. 176, pp. 1-20). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-77011-1\\_1](https://doi.org/10.1007/978-3-642-77011-1_1)

*Illumina | Sequencing and array-based solutions for genetic research*. (s. f.). Recuperado 31 de agosto de 2023, de <https://emea.illumina.com/>

*Ion Torrent—ES*. (s. f.). Recuperado 31 de agosto de 2023, de <https://www.thermofisher.com/es/es/home/brands/ion-torrent.html>

Jiménez Villa, J., & Argimón Pallàs, J. M. (1998). Artículo original (III): Resultados. *Atención Primaria*, 22(8), 534-535.

Klempt, P., Brož, P., Kašný, M., Novotný, A., Kvapilová, K., & Kvapil, P. (2020). Performance of Targeted Library Preparation Solutions for SARS-CoV-2 Whole Genome Analysis. *Diagnostics*, 10(10), 769.

- <https://doi.org/10.3390/diagnostics10100769>
- Koetsier, G., Cantor, E., & New Englands Biolabs. (s. f.). *A Practical Guide to Analyzing Nucleic Acid Concentration and Purity with Microvolume Spectrophotometers*.
- Li, Y., Yang, X., Wang, N., Wang, H., Yin, B., Yang, X., & Jiang, W. (2020). Mutation profile of over 4500 SARS-CoV-2 isolations reveals prevalent cytosine-to-uridine deamination on viral RNAs. *Future Microbiology*, 15(14), 1343-1352. <https://doi.org/10.2217/fmb-2020-0149>
- Lobinska, G., Pilpel, Y., & Nowak, M. A. (2023). Evolutionary safety of lethal mutagenesis driven by antiviral treatment. *PLOS Biology*, 21(8), e3002214. <https://doi.org/10.1371/journal.pbio.3002214>
- Martínez, L. D. (s. f.). *Variabilidad genética de cuasiespecies de virus RNA y ssDNA in vivo ante cambios en la tasa de error*.
- NCBI Virus. (s. f.). Recuperado 1 de septiembre de 2023, de [https://www.ncbi.nlm.nih.gov/labs/virus/vssi#/virus?SeqType\\_s=Nucleotide&VirusLineage\\_ss=Severe%20acute%20respiratory%20syndrome%20coronaviruses%202,%20taxid:2697049](https://www.ncbi.nlm.nih.gov/labs/virus/vssi#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Severe%20acute%20respiratory%20syndrome%20coronaviruses%202,%20taxid:2697049)
- Oliva Marín, J. E. (2020). SARS-CoV-2: Origen, estructura, replicación y patogénesis [4]. *Alerta, Revista científica del Instituto Nacional de Salud*, 3(2). <https://doi.org/10.5377/alerta.v3i2.9619>
- Ortega Del Campo, S. (2020). *Variabilidad y evolución de digitaria streak virus en Digitaria setigera*.
- Pastrian-Soto, G. (2020). Bases Genéticas y Moleculares del COVID-19 (SARS-CoV-2). Mecanismos de Patogénesis y de Respuesta Inmune [1]. *International journal of odontostomatology*, 14(3), 331-337. <https://doi.org/10.4067/S0718-381X2020000300331>
- Picasso | RES - Red Española de Supercomputación. (s. f.). Recuperado 1 de septiembre de 2023, de <https://www.res.es/es/nodos-de-la-res/picasso>
- Primer designing tool. (s. f.). <https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?>
- Ramirez-Gonzalez, R. H., Bonnal, R., Caccamo, M., & MacLean, D. (2012). *Bio-samtools: Ruby bindings for SAMtools, a library for accessing BAM files containing high-throughput sequence alignments*.
- Roche España. (s. f.). Recuperado 1 de septiembre de 2023, de

<https://www.roche.es/>

- Seoane, P., Díaz-Martínez, L., Viguera, E., Claros, M. G., & Grande-Pérez, A. (2022). *QuasiFlow: A bioinformatic tool for genetic variability analysis from next generation sequencing data* [Preprint]. *Bioinformatics*.  
<https://doi.org/10.1101/2022.04.05.487169>
- Simmonds, P. (2020). Rampant C→U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories. *mSphere*, 5(3), e00408-20.  
<https://doi.org/10.1128/mSphere.00408-20>
- Stecher, G., Tamura, K., & Kumar, S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution*, 37(4), 1237-1239. <https://doi.org/10.1093/molbev/msz312>
- Swanstrom, R., & Schinazi, R. F. (2022). Lethal mutagenesis as an antiviral strategy. *Science*, 375(6580), 497-498. <https://doi.org/10.1126/science.abn0048>
- Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, 38(7), 3022-3027. <https://doi.org/10.1093/molbev/msab120>
- Variante ómicron del SARS-CoV-2. (2023). En *Wikipedia, la enciclopedia libre*.  
[https://es.wikipedia.org/w/index.php?title=Variante\\_%C3%B3micron\\_del\\_SARS-CoV-2&oldid=153225503#cite\\_note-Sin\\_nombre-pJ9x-1-13](https://es.wikipedia.org/w/index.php?title=Variante_%C3%B3micron_del_SARS-CoV-2&oldid=153225503#cite_note-Sin_nombre-pJ9x-1-13)
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189-1191.  
<https://doi.org/10.1093/bioinformatics/btp033>



# Apéndice A

## Tablas

### A1. Tablas de sesgos de nucleótidos

A1 SARS_A				A3 SARS_B				A5 SARS_C				Media SARS		
Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%
Transversiones	3	8.57%		Transversiones	3	8.33%		Transversiones	1	3.57%		Transversiones	2,333333333	7.07%
A→C	0	0.00%	A→C	0	0.00%	A→C	0	0.00%	A→C	0	0.00%	A→C	0	0.00%
A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%
C→A	0	0.00%	C→A	0	0.00%	C→A	0	0.00%	C→A	0	0.00%	C→A	0	0.00%
C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%
G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%
G→T	1	2.86%	G→T	1	2.78%	G→T	1	2.78%	G→T	0	0.00%	G→T	0,666666667	2.02%
T→A	0	0.00%	T→A	1	2.78%	T→A	0	0.00%	T→A	0	0.00%	T→A	0,333333333	1.01%
T→G	2	5.71%	T→G	1	2.78%	T→G	1	2.78%	T→G	1	3.57%	T→G	1,333333333	4.04%
Transiciones	24	68.57%	Transiciones	25	69.44%	Transiciones	20	71.43%	Transiciones	23	69.70%	Transiciones	23	69.70%
A→G	3	8.57%	A→G	4	11.11%	A→G	1	3.57%	A→G	2	6.666666667	A→G	2	6.666666667
C→T	5	14.29%	C→T	4	11.11%	C→T	5	17.86%	C→T	5	17.86%	C→T	4,666666667	14.14%
G→A	1	2.86%	G→A	2	5.56%	G→A	2	7.14%	G→A	1,666666667	5.05%	G→A	1,666666667	5.05%
T→C	15	42.86%	T→C	15	41.67%	T→C	12	42.86%	T→C	14	42.42%	T→C	14	42.42%
Sustituciones totales	27	77.14%	Sustituciones totales	28	77.78%	Sustituciones totales	21	75.00%	Sustituciones totales	25,333333333	76.77%	Sustituciones totales	25,333333333	76.77%
Inserciones	0	0.00%	Inserciones	0	0.00%	Inserciones	0	0.00%	Inserciones	0	0.00%	Inserciones	0	0.00%
Delecciones	8	22.86%	Delecciones	8	22.22%	Delecciones	7	25.00%	Delecciones	7,666666667	23.23%	Delecciones	7,666666667	23.23%
Mutaciones totales	35	100.00%	Mutaciones totales	36	100.00%	Mutaciones totales	28	100.00%	Mutaciones totales	33	100.00%	Mutaciones totales	33	100.00%
A7 SARS_300_A				A9 SARS_300_B				B2 SARS_300_C				Media SARS_300		
Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%
Transversiones	1	2.70%		Transversiones	4	11.43%		Transversiones	3	7.32%		Transversiones	2	5.41%
A→C	0	0.00%	A→C	1	2.86%	A→C	1	2.44%	A→C	1	2.44%	A→C	0,666666667	1.80%
A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%
C→A	0	0.00%	C→A	0	0.00%	C→A	0	0.00%	C→A	0	0.00%	C→A	0	0.00%
C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%
G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%
G→T	0	0.00%	G→T	0	0.00%	G→T	0	0.00%	G→T	0	0.00%	G→T	0	0.00%
T→A	1	2.70%	T→A	1	2.86%	T→A	1	2.44%	T→A	1	2.70%	T→A	1	2.70%
T→G	0	0.00%	T→G	2	5.71%	T→G	1	2.44%	T→G	1	2.70%	T→G	1	2.70%
Transiciones	36	97.30%	Transiciones	25	71.43%	Transiciones	31	75.61%	Transiciones	30,666666667	82.88%	Transiciones	30,666666667	82.88%
A→G	7	18.92%	A→G	2	5.71%	A→G	6	14.63%	A→G	5	13.51%	A→G	5	13.51%
C→T	5	13.51%	C→T	4	11.43%	C→T	5	12.20%	C→T	4,666666667	12.61%	C→T	4,666666667	12.61%
G→A	1	2.70%	G→A	0	0.00%	G→A	0	0.00%	G→A	0,333333333	0.90%	G→A	0,333333333	0.90%
T→C	23	62.16%	T→C	19	54.29%	T→C	20	48.78%	T→C	20,666666667	55.86%	T→C	20,666666667	55.86%
Sustituciones totales	37	100.00%	Sustituciones totales	29	82.86%	Sustituciones totales	34	82.93%	Sustituciones totales	32,666666667	88.29%	Sustituciones totales	32,666666667	88.29%
Inserciones	0	0.00%	Inserciones	2	5.71%	Inserciones	2	4.88%	Inserciones	1,333333333	3.60%	Inserciones	1,333333333	3.60%
Delecciones	0	0.00%	Delecciones	4	11.43%	Delecciones	5	12.20%	Delecciones	3	8.11%	Delecciones	3	8.11%
Mutaciones totales	37	100.00%	Mutaciones totales	35	100.00%	Mutaciones totales	41	100.00%	Mutaciones totales	37	100.00%	Mutaciones totales	37	100.00%
B4 SARS_400_A				B6 SARS_400_B				B8 SARS_400_C				Media SARS_400		
Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%
Transversiones	5	7.81%		Transversiones	4	6.78%		Transversiones	4	4.88%		Transversiones	4	5.88%
A→C	1	1.56%	A→C	0	0.00%	A→C	0	0.00%	A→C	0	0.00%	A→C	0,333333333	0.49%
A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%
C→A	1	1.56%	C→A	1	1.69%	C→A	1	1.22%	C→A	1	1.47%	C→A	1	1.47%
C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%
G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%
G→T	0	0.00%	G→T	0	0.00%	G→T	0	0.00%	G→T	0	0.00%	G→T	0	0.00%
T→A	1	1.56%	T→A	1	1.69%	T→A	1	1.22%	T→A	1	1.47%	T→A	1	1.47%
T→G	2	3.13%	T→G	2	3.39%	T→G	2	2.44%	T→G	2	2.94%	T→G	2	2.94%
Transiciones	51	79.69%	Transiciones	47	79.66%	Transiciones	70	85.37%	Transiciones	56	82.35%	Transiciones	56	82.35%
A→G	8	12.50%	A→G	4	6.78%	A→G	11	13.41%	A→G	7,666666667	11.27%	A→G	7,666666667	11.27%
C→T	10	15.63%	C→T	9	15.25%	C→T	11	13.41%	C→T	10	14.71%	C→T	10	14.71%
G→A	3	4.69%	G→A	4	6.78%	G→A	4	4.88%	G→A	3,666666667	5.39%	G→A	3,666666667	5.39%
T→C	30	46.88%	T→C	30	50.85%	T→C	44	53.66%	T→C	34,666666667	50.98%	T→C	34,666666667	50.98%
Sustituciones totales	56	87.50%	Sustituciones totales	51	86.44%	Sustituciones totales	74	90.24%	Sustituciones totales	60	88.24%	Sustituciones totales	60	88.24%
Inserciones	2	3.13%	Inserciones	2	3.39%	Inserciones	2	2.44%	Inserciones	2	2.94%	Inserciones	2	2.94%
Delecciones	6	9.38%	Delecciones	6	10.17%	Delecciones	6	7.32%	Delecciones	6	8.82%	Delecciones	6	8.82%
Mutaciones totales	64	100.00%	Mutaciones totales	59	100.00%	Mutaciones totales	82	100.00%	Mutaciones totales	68	100.00%	Mutaciones totales	68	100.00%
C1 SARS_500_A				C3 SARS_500_B				C5 SARS_500_C				Media SARS_500		
Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%		Tipos de mutaciones	Nº de mutación	%
Transversiones	4	4.88%		Transversiones	4	4.26%		Transversiones	4	4.26%		Transversiones	4	4.44%
A→C	0	0.00%	A→C	0	0.00%	A→C	0	0.00%	A→C	0	0.00%	A→C	0	0.00%
A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%	A→T	0	0.00%
C→A	1	1.22%	C→A	1	1.06%	C→A	1	1.06%	C→A	1	1.11%	C→A	1	1.11%
C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%	C→G	0	0.00%
G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%	G→C	0	0.00%
G→T	0	0.00%	G→T	0	0.00%	G→T	0	0.00%	G→T	0	0.00%	G→T	0	0.00%
T→A	1	1.22%	T→A	1	1.06%	T→A	1	1.06%	T→A	1	1.11%	T→A	1	1.11%
T→G	2	2.44%	T→G	2	2.13%	T→G	2	2.13%	T→G	2	2.22%	T→G	2	2.22%
Transiciones	71	86.59%	Transiciones	84	89.36%	Transiciones	82	87.23%	Transiciones	79	87.78%	Transiciones	79	87.78%
A→G	6	7.32%	A→G	8	8.51%	A→G	7	7.45%	A→G	7	7.78%	A→G	7	7.78%
C→T	5	6.10%	C→T	6	6.38%	C→T	7	7.45%	C→T	6	6.67%	C→T	6	6.67%
G→A	3	3.66%	G→A	1	1.06%	G→A	2	2.13%	G→A	2	2.22%	G→A	2	2.22%
T→C	57	69.51%	T→C	69	73.40%	T→C	66	70.21%	T→C	64	71.11%	T→C	64	71.11%
Sustituciones totales	75	91.46%	Sustituciones totales	88	93.62%	Sustituciones totales	86	91.49%	Sustituciones totales	83	92.22%	Sustituciones totales	83	92.22%
Inserciones	1	1.22%	Inserciones	1	1.06%	Inserciones	1	1.06%	Inserciones	1	1.11%	Inserciones	1	1.11%
Delecciones	6	7.32%	Delecciones	5	5.32%	Delecciones	7	7.45%	Delecciones	6	6.67%	Delecciones	6	6.67%
Mutaciones totales	82	100.00%	Mutaciones totales	94	100.00%	Mutaciones totales	94	100.00%	Mutaciones totales	90	100.00%	Mutaciones totales	90	100.00%

Tabla A1.1 Tablas de sesgos de nucleótidos de las muestras tratadas sin péptido.

C7 SARS_14_A			C9 SARS_14_B			D2 SARS_14_C			Media SARS_6		
Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%
Transversiones	10	20,83%	Transversiones	4	9,09%	Transversiones	7	18,42%	Transversiones	6,33333333	14,84%
A→C	1	2,08%	A→C	0	0,00%	A→C	1	2,63%	A→C	0,66666667	1,56%
A→T	1	2,08%	A→T	0	0,00%	A→T	0	0,00%	A→T	0,33333333	0,78%
C→A	2	4,17%	C→A	1	2,27%	C→A	1	2,63%	C→A	1,33333333	3,13%
C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%
G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%
G→T	0	0,00%	G→T	0	0,00%	G→T	0	0,00%	G→T	0	0,00%
T→A	1	2,08%	T→A	1	2,27%	T→A	1	2,63%	T→A	1	2,34%
T→G	5	10,42%	T→G	2	4,55%	T→G	4	10,53%	T→G	3,66666667	8,59%
Transiciones	28	58,33%	Transiciones	32	72,73%	Transiciones	24	63,16%	Transiciones	28	65,63%
A→G	4	8,33%	A→G	9	20,45%	A→G	5	13,16%	A→G	6	14,06%
C→T	6	12,50%	C→T	10	22,73%	C→T	7	18,42%	C→T	7,66666667	17,97%
G→A	2	4,17%	G→A	1	2,27%	G→A	1	2,63%	G→A	1,33333333	3,13%
T→C	16	33,33%	T→C	12	27,27%	T→C	11	28,95%	T→C	13	30,47%
Sustituciones totales	38	79,17%	Sustituciones totales	36	81,82%	Sustituciones totales	31	81,58%	Sustituciones totales	34,33333333	80,47%
Inserciones	1	2,08%	Inserciones	0	0,00%	Inserciones	0	0,00%	Inserciones	0,33333333	0,78%
Delecciones	9	18,75%	Delecciones	8	18,18%	Delecciones	7	18,42%	Delecciones	8	18,75%
Mutaciones totales	48	100,00%	Mutaciones totales	44	100,00%	Mutaciones totales	38	100,00%	Mutaciones totales	42,66666667	100,00%
D4 SARS_14_300_A			D6 SARS_14_300_B			D8 SARS_14_300_C			Media SARS_14_300		
Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%
Transversiones	8	10,00%	Transversiones	8	11,59%	Transversiones	13	19,70%	Transversiones	8,33333333	11,85%
A→C	1	1,25%	A→C	2	2,90%	A→C	1	1,52%	A→C	1,33333333	1,90%
A→T	0	0,00%	A→T	0	0,00%	A→T	1	1,52%	A→T	0,33333333	0,47%
C→A	2	2,50%	C→A	1	1,45%	C→A	1	1,52%	C→A	1,33333333	1,90%
C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%
G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%
G→T	0	0,00%	G→T	0	0,00%	G→T	2	3,03%	G→T	0,66666667	0,95%
T→A	1	1,25%	T→A	2	2,90%	T→A	2	3,03%	T→A	1,66666667	2,37%
T→G	4	5,00%	T→G	3	4,35%	T→G	6	9,09%	T→G	4,33333333	6,16%
Transiciones	64	80,00%	Transiciones	52	75,36%	Transiciones	43	65,15%	Transiciones	53	75,36%
A→G	13	16,25%	A→G	14	20,29%	A→G	9	13,64%	A→G	12	17,06%
C→T	11	13,75%	C→T	9	13,04%	C→T	10	15,15%	C→T	10	14,22%
G→A	2	2,50%	G→A	2	2,90%	G→A	0	0,00%	G→A	1,33333333	1,90%
T→C	38	47,50%	T→C	27	39,13%	T→C	24	36,36%	T→C	29,66666667	42,18%
Sustituciones totales	72	90,00%	Sustituciones totales	60	86,96%	Sustituciones totales	56	84,85%	Sustituciones totales	61,33333333	87,20%
Inserciones	1	1,25%	Inserciones	1	1,45%	Inserciones	1	1,52%	Inserciones	1	1,42%
Delecciones	7	8,75%	Delecciones	8	11,59%	Delecciones	9	13,64%	Delecciones	8	11,37%
Mutaciones totales	80	100,00%	Mutaciones totales	69	100,00%	Mutaciones totales	66	100,00%	Mutaciones totales	70,33333333	100,00%
E1 SARS_14_400_A			E3 SARS_14_400_B			E5 SARS_14_400_C			Media SARS_14_400		
Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%
Transversiones	7	7,53%	Transversiones	6	5,00%	Transversiones	7	6,42%	Transversiones	6	5,63%
A→C	1	1,08%	A→C	0	0,00%	A→C	1	0,92%	A→C	0,66666667	0,63%
A→T	1	1,08%	A→T	1	0,83%	A→T	0	0,00%	A→T	0,66666667	0,63%
C→A	1	1,08%	C→A	1	0,83%	C→A	1	0,92%	C→A	1	0,94%
C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%
G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%
G→T	1	1,08%	G→T	0	0,00%	G→T	2	1,83%	G→T	1	0,94%
T→A	1	1,08%	T→A	2	1,67%	T→A	1	0,92%	T→A	1,33333333	1,25%
T→G	2	2,15%	T→G	2	1,67%	T→G	2	1,83%	T→G	2	1,88%
Transiciones	78	83,87%	Transiciones	106	88,33%	Transiciones	92	84,40%	Transiciones	92	86,25%
A→G	25	26,88%	A→G	32	26,67%	A→G	21	19,27%	A→G	26	24,38%
C→T	10	10,75%	C→T	15	12,50%	C→T	10	9,17%	C→T	11,66666667	10,94%
G→A	3	3,23%	G→A	9	7,50%	G→A	6	5,50%	G→A	6	5,63%
T→C	40	43,01%	T→C	50	41,67%	T→C	55	50,46%	T→C	48,33333333	45,31%
Sustituciones totales	85	91,40%	Sustituciones totales	112	93,33%	Sustituciones totales	99	90,83%	Sustituciones totales	98	91,88%
Inserciones	1	1,08%	Inserciones	2	1,67%	Inserciones	1	0,92%	Inserciones	1,33333333	1,25%
Delecciones	7	7,53%	Delecciones	6	5,00%	Delecciones	9	8,26%	Delecciones	7,33333333	6,88%
Mutaciones totales	93	100,00%	Mutaciones totales	120	100,00%	Mutaciones totales	109	100,00%	Mutaciones totales	106,666667	100,00%
E7 SARS_14_500_A			E9 SARS_14_500_B			F2 SARS_14_500_C			Media SARS_14_500		
Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%
Transversiones	5	4,90%	Transversiones	6	5,56%	Transversiones	6	6,74%	Transversiones	5	5,05%
A→C	1	0,98%	A→C	0	0,00%	A→C	1	1,12%	A→C	0,66666667	0,67%
A→T	0	0,00%	A→T	0	0,00%	A→T	0	0,00%	A→T	0	0,00%
C→A	1	0,98%	C→A	1	0,93%	C→A	1	1,12%	C→A	1	1,01%
C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%
G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%
G→T	0	0,00%	G→T	0	0,00%	G→T	1	1,12%	G→T	0,33333333	0,34%
T→A	1	0,98%	T→A	1	0,93%	T→A	1	1,12%	T→A	1	1,01%
T→G	2	1,96%	T→G	4	3,70%	T→G	2	2,25%	T→G	2,66666667	2,69%
Transiciones	89	87,25%	Transiciones	94	87,04%	Transiciones	73	82,02%	Transiciones	85,33333333	86,20%
A→G	18	17,65%	A→G	27	25,00%	A→G	13	14,61%	A→G	19,33333333	19,53%
C→T	12	11,76%	C→T	16	14,81%	C→T	14	15,73%	C→T	14	14,14%
G→A	5	4,90%	G→A	6	5,56%	G→A	1	1,12%	G→A	4	4,04%
T→C	54	52,94%	T→C	45	41,67%	T→C	45	50,56%	T→C	48	48,48%
Sustituciones totales	94	92,16%	Sustituciones totales	100	92,59%	Sustituciones totales	79	88,76%	Sustituciones totales	90,33333333	91,25%
Inserciones	1	0,98%	Inserciones	0	0,00%	Inserciones	1	1,12%	Inserciones	0,66666667	0,67%
Delecciones	7	6,86%	Delecciones	8	7,41%	Delecciones	9	10,11%	Delecciones	8	8,08%
Mutaciones totales	102	100,00%	Mutaciones totales	108	100,00%	Mutaciones totales	89	100,00%	Mutaciones totales	99	100,00%

Tabla A1.2 Tablas de sesgos de nucleótidos de las muestras tratadas con péptido #14.

L4 SARS_16_A				L6 SARS_16_B				L8 SARS_16_C				Media SARS_1			
Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	
Transversiones	7	15,22%	Transversiones	7	16,28%	Transversiones	7	16,67%	Transversiones	5,333333333	12,70%	Transversiones	5,333333333	12,70%	
A→C	2	4,35%	A→C	2	4,65%	A→C	1	2,38%	A→C	1,666666667	3,97%	A→C	1,666666667	3,97%	
A→T	0	0,00%	A→T	0	0,00%	A→T	0	0,00%	A→T	0	0,00%	A→T	0	0,00%	
C→A	1	2,17%	C→A	0	0,00%	C→A	1	2,38%	C→A	0,666666667	1,59%	C→A	0,666666667	1,59%	
C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	
G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	
G→T	1	2,17%	G→T	0	0,00%	G→T	0	0,00%	G→T	0,333333333	0,79%	G→T	0,333333333	0,79%	
T→A	1	2,17%	T→A	2	4,65%	T→A	2	4,76%	T→A	1,666666667	3,97%	T→A	1,666666667	3,97%	
T→G	2	4,35%	T→G	3	6,98%	T→G	3	7,14%	T→G	2,666666667	6,35%	T→G	2,666666667	6,35%	
Transiciones	30	65,22%	Transiciones	28	65,12%	Transiciones	28	66,67%	Transiciones	28,666666667	68,25%	Transiciones	28,666666667	68,25%	
A→G	5	10,87%	A→G	8	18,60%	A→G	3	7,14%	A→G	5,333333333	12,70%	A→G	5,333333333	12,70%	
C→T	8	17,39%	C→T	7	16,28%	C→T	10	23,81%	C→T	8,333333333	19,84%	C→T	8,333333333	19,84%	
G→A	2	4,35%	G→A	0	0,00%	G→A	1	2,38%	G→A	1	2,38%	G→A	1	2,38%	
T→C	15	32,61%	T→C	13	30,23%	T→C	14	33,33%	T→C	14	33,33%	T→C	14	33,33%	
Sustituciones totales	37	80,43%	Sustituciones totales	35	81,40%	Sustituciones totales	35	83,33%	Sustituciones totales	34	80,95%	Sustituciones totales	34	80,95%	
Inserciones	0	0,00%	Inserciones	0	0,00%	Inserciones	0	0,00%	Inserciones	0	0,00%	Inserciones	0	0,00%	
Delecciones	9	19,57%	Delecciones	8	18,60%	Delecciones	7	16,67%	Delecciones	8	19,05%	Delecciones	8	19,05%	
Mutaciones totales	46	100,00%	Mutaciones totales	43	100,00%	Mutaciones totales	42	100,00%	Mutaciones totales	42	100,00%	Mutaciones totales	42	100,00%	
M1 SARS_16_300_A				M3 SARS_16_300_B				M5 SARS_16_300_C				Media SARS_16_300			
Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	
Transversiones	9	11,11%	Transversiones	9	7,89%	Transversiones	7	7,95%	Transversiones	7	7,53%	Transversiones	7	7,53%	
A→C	2	2,47%	A→C	1	0,88%	A→C	1	1,14%	A→C	1,333333333	1,43%	A→C	1,333333333	1,43%	
A→T	0	0,00%	A→T	1	0,88%	A→T	0	0,00%	A→T	0,333333333	0,36%	A→T	0,333333333	0,36%	
C→A	2	2,47%	C→A	2	1,75%	C→A	1	1,14%	C→A	1,666666667	1,79%	C→A	1,666666667	1,79%	
C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	
G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	
G→T	1	1,23%	G→T	2	1,75%	G→T	1	1,14%	G→T	1,333333333	1,43%	G→T	1,333333333	1,43%	
T→A	1	1,23%	T→A	1	0,88%	T→A	3	3,41%	T→A	1,666666667	1,79%	T→A	1,666666667	1,79%	
T→G	3	3,70%	T→G	2	1,75%	T→G	1	1,14%	T→G	2	2,15%	T→G	2	2,15%	
Transiciones	62	76,54%	Transiciones	98	85,96%	Transiciones	72	81,82%	Transiciones	77,333333333	83,15%	Transiciones	77,333333333	83,15%	
A→G	14	17,28%	A→G	24	21,05%	A→G	17	19,32%	A→G	18,333333333	19,71%	A→G	18,333333333	19,71%	
C→T	12	14,81%	C→T	13	11,40%	C→T	8	9,09%	C→T	11	11,83%	C→T	11	11,83%	
G→A	3	3,70%	G→A	3	2,63%	G→A	6	6,82%	G→A	4	4,30%	G→A	4	4,30%	
T→C	33	40,74%	T→C	58	50,88%	T→C	41	46,59%	T→C	44	47,31%	T→C	44	47,31%	
Sustituciones totales	71	87,65%	Sustituciones totales	107	93,86%	Sustituciones totales	79	89,77%	Sustituciones totales	84,333333333	90,68%	Sustituciones totales	84,333333333	90,68%	
Inserciones	0	0,00%	Inserciones	1	0,88%	Inserciones	1	1,14%	Inserciones	0,666666667	0,72%	Inserciones	0,666666667	0,72%	
Delecciones	10	12,35%	Delecciones	6	5,26%	Delecciones	8	9,09%	Delecciones	8	8,60%	Delecciones	8	8,60%	
Mutaciones totales	81	100,00%	Mutaciones totales	114	100,00%	Mutaciones totales	88	100,00%	Mutaciones totales	93	100,00%	Mutaciones totales	93	100,00%	
M7 SARS_16_400_A				M9 SARS_16_400_B				M2 SARS_16_400_C				Media SARS_16_400			
Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	
Transversiones	7	8,64%	Transversiones	5	5,38%	Transversiones	4	5,97%	Transversiones	5,333333333	12,70%	Transversiones	5,333333333	12,70%	
A→C	0	0,00%	A→C	0	0,00%	A→C	1	1,49%	A→C	0,333333333	0,41%	A→C	0,333333333	0,41%	
A→T	1	1,23%	A→T	0	0,00%	A→T	0	0,00%	A→T	0,333333333	0,41%	A→T	0,333333333	0,41%	
C→A	1	1,23%	C→A	1	1,08%	C→A	1	1,49%	C→A	1	1,24%	C→A	1	1,24%	
C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	
G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	
G→T	0	0,00%	G→T	1	1,08%	G→T	0	0,00%	G→T	0,333333333	0,41%	G→T	0,333333333	0,41%	
T→A	1	1,23%	T→A	1	1,08%	T→A	1	1,49%	T→A	1	1,24%	T→A	1	1,24%	
T→G	4	4,94%	T→G	2	2,15%	T→G	1	1,49%	T→G	2,333333333	2,90%	T→G	2,333333333	2,90%	
Transiciones	68	83,95%	Transiciones	78	83,87%	Transiciones	55	82,09%	Transiciones	67	83,40%	Transiciones	67	83,40%	
A→G	19	23,46%	A→G	23	24,73%	A→G	13	19,40%	A→G	18,333333333	22,82%	A→G	18,333333333	22,82%	
C→T	5	6,17%	C→T	11	11,83%	C→T	9	13,43%	C→T	8,333333333	10,37%	C→T	8,333333333	10,37%	
G→A	3	3,70%	G→A	1	1,08%	G→A	3	4,48%	G→A	2,333333333	2,90%	G→A	2,333333333	2,90%	
T→C	41	50,62%	T→C	43	46,24%	T→C	30	44,78%	T→C	38	47,30%	T→C	38	47,30%	
Sustituciones totales	75	92,59%	Sustituciones totales	83	89,25%	Sustituciones totales	59	88,06%	Sustituciones totales	72,333333333	90,04%	Sustituciones totales	72,333333333	90,04%	
Inserciones	1	1,23%	Inserciones	0	0,00%	Inserciones	1	1,49%	Inserciones	0,666666667	0,83%	Inserciones	0,666666667	0,83%	
Delecciones	5	6,17%	Delecciones	10	10,75%	Delecciones	7	10,45%	Delecciones	7,333333333	9,13%	Delecciones	7,333333333	9,13%	
Mutaciones totales	81	100,00%	Mutaciones totales	93	100,00%	Mutaciones totales	67	100,00%	Mutaciones totales	80,333333333	100,00%	Mutaciones totales	80,333333333	100,00%	
N4 SARS_16_500_A				N6 SARS_16_500_B				N8 SARS_16_500_C				Media SARS_16_500			
Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	Tipos de mutaciones	Nº de mutación	%	
Transversiones	1	2,38%	Transversiones	2	4,65%	Transversiones	1	2,33%	Transversiones	1,333333333	20,00%	Transversiones	1,333333333	20,00%	
A→C	0	0,00%	A→C	0	0,00%	A→C	0	0,00%	A→C	0	0,00%	A→C	0	0,00%	
A→T	0	0,00%	A→T	0	0,00%	A→T	0	0,00%	A→T	0	0,00%	A→T	0	0,00%	
C→A	0	0,00%	C→A	1	2,33%	C→A	0	0,00%	C→A	0,333333333	0,78%	C→A	0,333333333	0,78%	
C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	C→G	0	0,00%	
G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	G→C	0	0,00%	
G→T	0	0,00%	G→T	0	0,00%	G→T	0	0,00%	G→T	0	0,00%	G→T	0	0,00%	
T→A	0	0,00%	T→A	0	0,00%	T→A	0	0,00%	T→A	0	0,00%	T→A	0	0,00%	
T→G	1	2,38%	T→G	1	2,33%	T→G	1	2,33%	T→G	1	2,34%	T→G	1	2,34%	
Transiciones	35	83,33%	Transiciones	34	79,07%	Transiciones	35	81,40%	Transiciones	34,666666667	81,25%	Transiciones	34,666666667	81,25%	
A→G	6	14,29%	A→G	8	18,60%	A→G	6	13,95%	A→G	6,666666667	15,63%	A→G	6,666666667	15,63%	
C→T	4	9,52%	C→T	2	4,65%	C→T	4	9,30%	C→T	3,333333333	7,81%	C→T	3,333333333	7,81%	
G→A	0	0,00%	G→A	2	4,65%	G→A	0	0,00%	G→A	0,666666667	1,56%	G→A	0,666666667	1,56%	
T→C	25	59,52%	T→C	22	51,16%	T→C	25	58,14%	T→C	24	56,25%	T→C	24	56,25%	
Sustituciones totales	36	85,71%	Sustituciones totales	36	83,72%	Sustituciones totales	36	83,72%	Sustituciones totales	36	84,38%	Sustituciones totales	36	84,38%	
Inserciones	0	0,00%	Inserciones	0	0,00%	Inserciones	0	0,00%	Inserciones	0	0,00%	Inserciones	0	0,00%	
Delecciones	6	14,29%	Delecciones	7	16,28%	Delecciones	7	16,28%	Delecciones	6,666666667	15,63%	Delecciones	6,666666667	15,63%	
Mutaciones totales	42	100,00%	Mutaciones totales	43	100,00%	Mutaciones totales	43	100,00%	Mutaciones totales	42,666666667	100,00%	Mutaciones totales	42,666666667	100,00%	

Tabla A1.3 Tablas de sesgos de nucleótidos de las muestras tratadas con péptido #16.



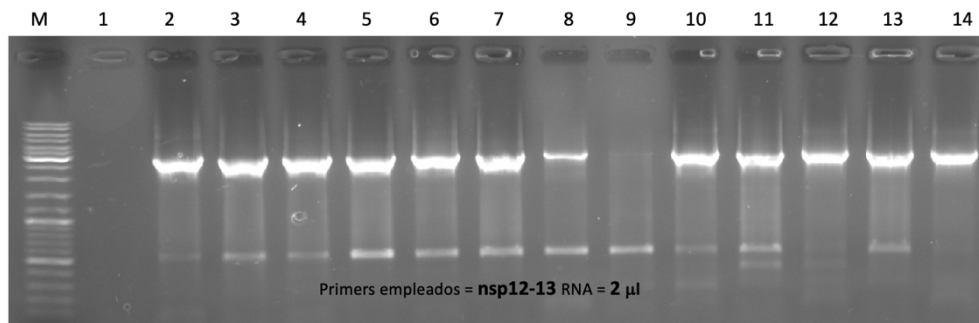




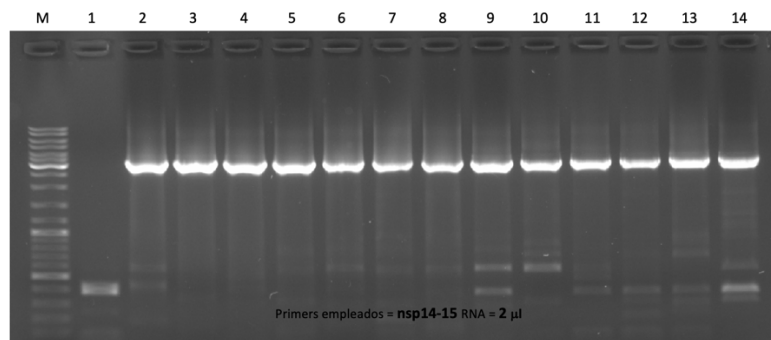
# Apéndice B

## Figuras

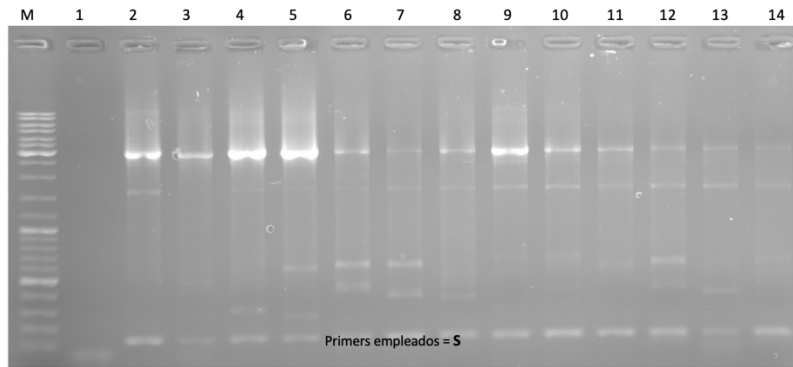
### B.1. Imágenes geles



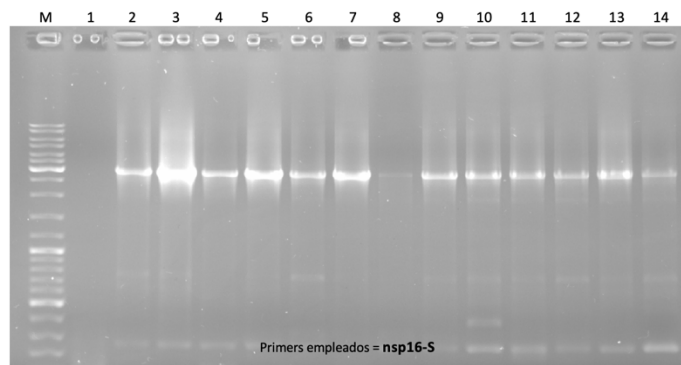
**Figura B1. 1** Gel con cebador Nsp12-13, añadiendo 2µl de RNA. 1.C- 2.C+ 3. A1(1:1), 4. A1(1:10), 5. A3(1:1), 6. A3(1:10), 7. A5(1:1), 8. A5(1:10), 9. A7(1:1), 10. A7(1:10), 11. A9(1:1), 12. A9(1:10), 13. B2(1:1), 14. B2(1:10).



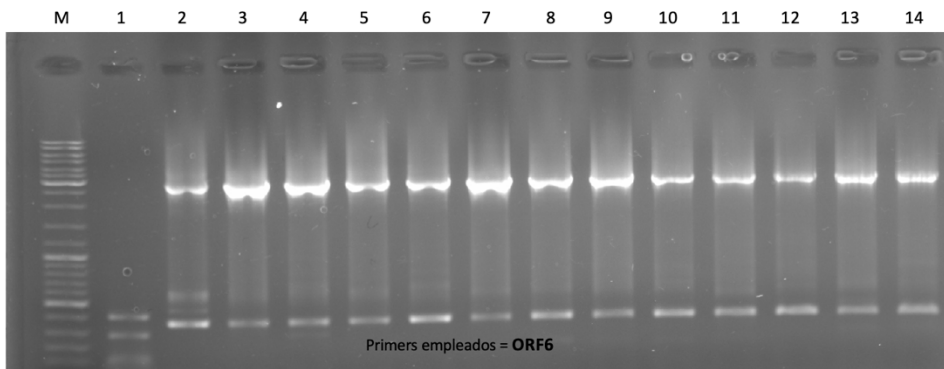
**Figura B1. 2** Gel con cebador Nsp14-15 añadiendo 2µl de RNA. 1.C- 2.C+ 3. A1(1:1), 4. A1(1:10), 5. A3(1:1), 6. A3(1:10), 7. A5(1:1), 8. A5(1:10), 9. A7(1:1), 10. A7(1:10), 11. A9(1:1), 12. A9(1:10), 13. B2(1:1), 14. B2(1:10).



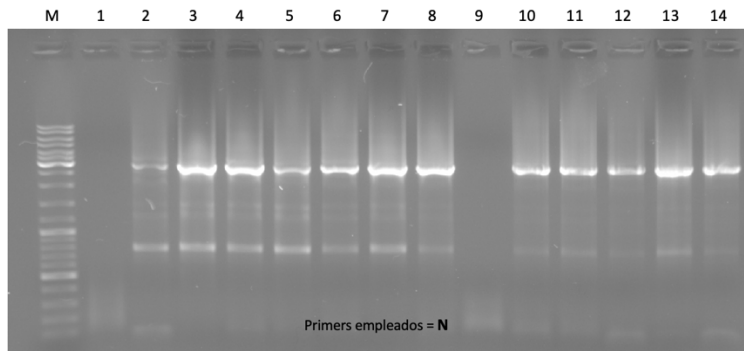
**Figura B1. 3** Gel con cebador S añadiendo 2 $\mu$ l de RNA. 1.C- 2.C+ 3. A1(1:1), 4. A3(1:1), 5. A5(1:1), 6. A7(1:1), 7. A9(1:1), 8. B2(1:1), 9. A1(1:10), 10. A3(1:10), 11. A5(1:10), 12. A7(1:10), 13. A9(1:10), 14. B2(1:10).



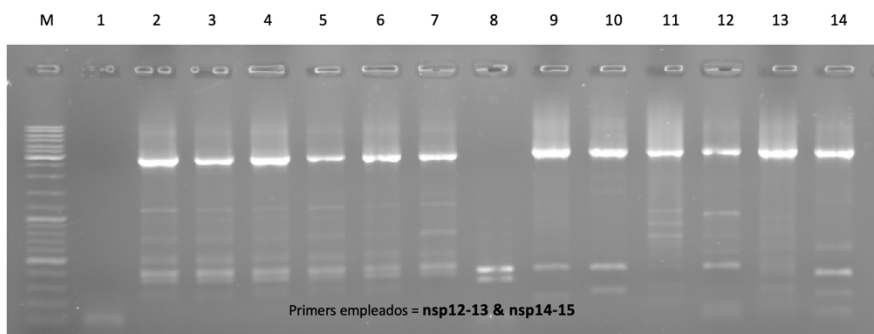
**Figura B1. 4** Gel con cebador Nsp16-S añadiendo 2 $\mu$ l de RNA. 1.C- 2.C+ 3. A1(1:1), 4. A1(1:10), 5. A3(1:1), 6. A3(1:10), 7. A5(1:1), 8. A5(1:10), 9. A7(1:1), 10. A7(1:10), 11. A9(1:1), 12. A9(1:10), 13. B2(1:1), 14. B2(1:10).



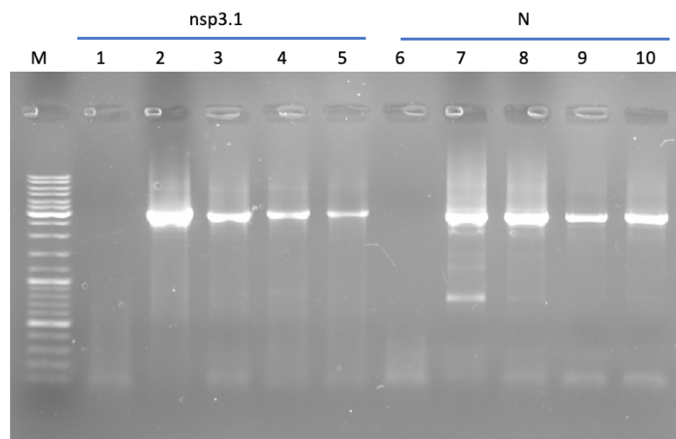
**Figura B1. 5** Gel con cebador ORF-6 añadiendo 2 $\mu$ l de RNA. 1.C- 2.C+ 3. A1(1:1), 4. A1(1:10), 5. A3(1:1), 6. A3(1:10), 7. A5(1:1), 8. A5(1:10), 9. A7(1:1), 10. A7(1:10), 11. A9(1:1), 12. A9(1:10), 13. B2(1:1), 14. B2(1:10).



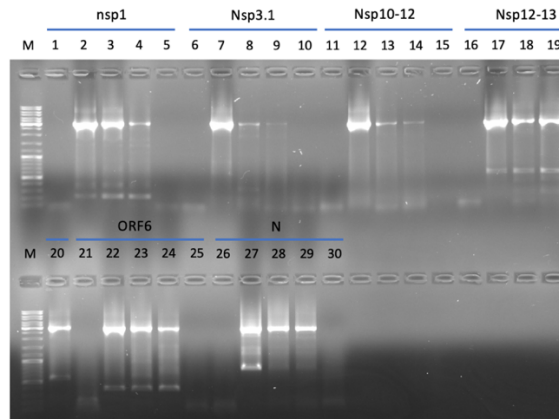
**Figura B1. 6** Gel con cebador N añadiendo 2 $\mu$ l de RNA. 1.C- 2.C+ 3. A1(1:1), 4. A1(1:10), 5. A3(1:1), 6. A3(1:10), 7. A5(1:1), 8. A5(1:10), 9. A7(1:1), 10. A7(1:10), 11. A9(1:1), 12. A9(1:10), 13. B2(1:1), 14. B2(1:10).



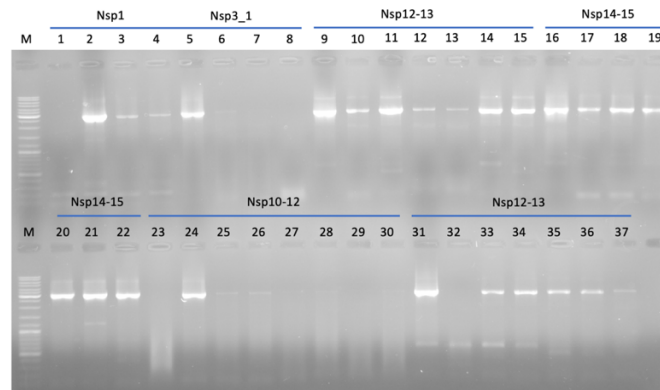
**Figura B1. 7** Gel con cebador Nsp12-13 añadiendo 2 $\mu$ l de RNA. 1. C-, 2. C2(1:10), 3. C2(1:100), 4. C4(1:10), 5. C4(1:100), 6. C6(1:10), 7. C6(1:100). Gel con cebador Nsp14-15 añadiendo 2 $\mu$ l de RNA. 8. C-, 9. C2(1:10), 10. C2(1:100), 11. C4(1:10), 12. C4(1:100), 13. C6(1:10), 14. B2(1:100).



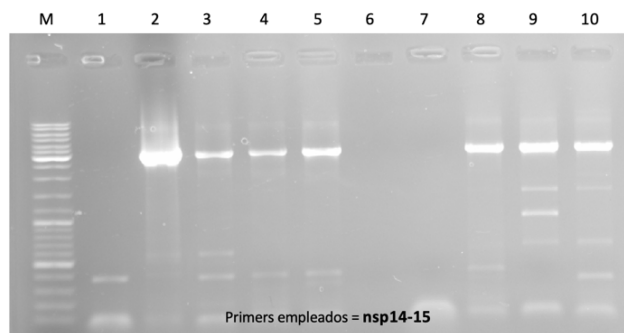
**Figura B1. 8** Gel con cebador N añadiendo 4 $\mu$ l de RNA. 6. C-, 7. C+, 8. B4(1:10), 9. B6(1:100), 10. B8(1:10). Los demás carriles están incluidos en el trabajo de María Jeréz.



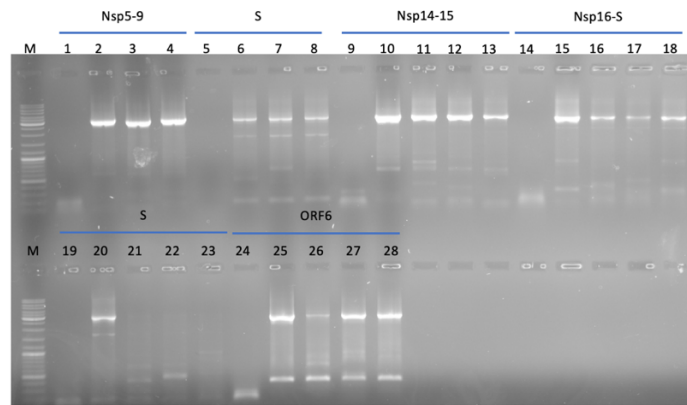
**Figura B1. 9** Gel con cebador Nsp12-13 añadiendo 8µl de RNA. 16. C-, 17. C+, 18. C1(1:10), 19. C3(1:10), 20. C5(1:10). Gel con cebador ORF6 añadiendo 4 µl de RNA. 21. C-, 22. C+, 23. C1(1:1), 24. C3(1:1), 25. C5(1:1). Gel con cebador N añadiendo 4 µl de RNA. 26. C-, 27. C+28. C1(1:1), 29. C3(1:1), 30. C5(1:1). Los demás carriles están incluidos en el trabajo de María Jeréz.



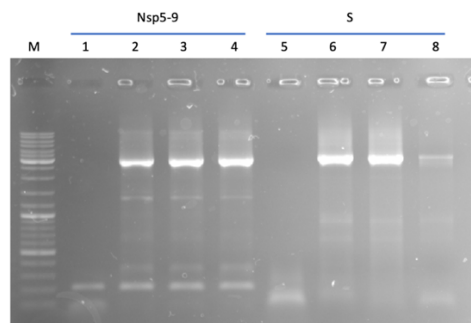
**Figura B1. 10** Gel con cebador Nsp12-13 añadiendo 8µl de RNA. 9. C+, 10. L4(1:1), 11.L4(1:10), 12. L6(1:1), 13. L6(1:10), 14. L8(1:1), 15. L8(1:10). Gel con cebador Nsp14-15 añadiendo 8µl de RNA. 16. C+, 17. L4 (1:1), 18. L4(1:10), 19. L6(1:1), 20. L6(1:10), 21. L8(1:1), 22. L8(1:1). Gel con cebador Nsp12-13 añadiendo 8µl de RNA. 31. C+, 32. C7(1:1), 33. C9(1:1), 34. D2(1:1), 35. C7(1:10), 36. C9(1:10), 37. D2(1:10). Los demás carriles están incluidos en el trabajo de María Jeréz.



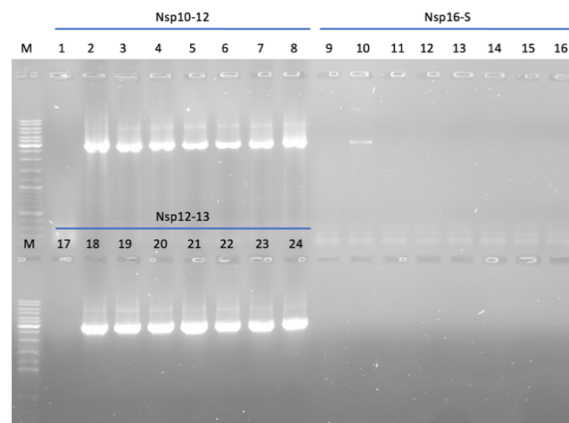
**Figura B1. 11** Gel con cebador Nsp14-15 añadiendo 8µl de RNA. 1. C-, 2. C+, 3.L4(1:10), 4. L6(1:10), 5. L8(1:10). Gel con cebador Nsp14-15 añadiendo 14µl de RNA. 6. C-, 7. C+ 8. L4(1:10), 9. L6(1:10), 10. L8(1:10).



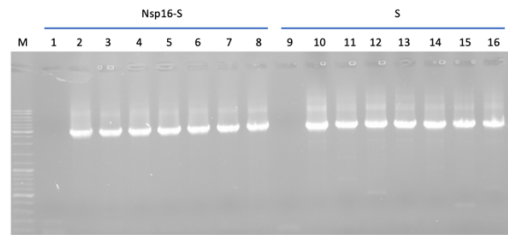
**Figura B1. 12** Gel con cebador S añadiendo 8 $\mu$ l de RNA. 5. C-, 6. A7(1:10), 6. A9(1:10), 8. B2(1:10). Gel con cebador nsp14-15 añadiendo 8 $\mu$ l de RNA. 9. C-, 10. C+, 11. B4(1:10), 12. B6(1:10), 13. B8(1:10). Gel con cebador nsp16-s añadiendo 8 $\mu$ l de RNA. 14. C-, 15. C+, 16. B4(1:10), 17. B6(1:10), 18. B8(1:10). Gel con cebador S añadiendo 8 $\mu$ l de RNA. 19. C-, 20. C+, 21. B4(1:10), 22. B6(1:10), 23. B8(1:10). Gel con cebador ORF6 añadiendo 8 $\mu$ l de RNA. 24. C-, 25. C+, 26. B4(1:10), 27. B6(1:10), 28. B8(1:10). Los demás carriles estan incluidos en el trabajo de Marıa Jerez.



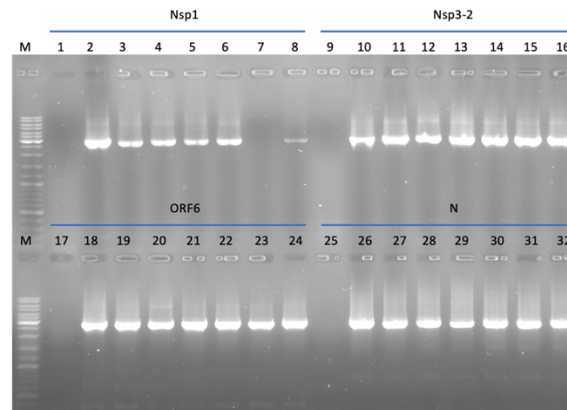
**Figura B1. 13** Gel con cebador S añadiendo 8 $\mu$ l de RNA. 5. C-, 6. B4(1:10), 7. B6(1:10), 8. B8(1:10). Los demas carriles estan incluidos en el trabajo de Marıa Jerez.



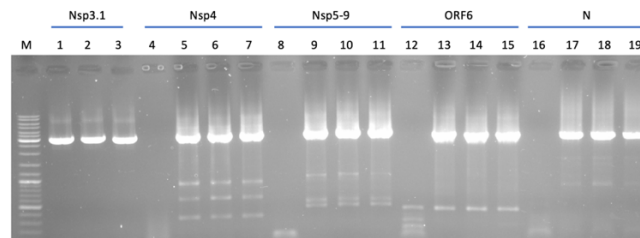
**Figura B1. 14** Gel con cebador nsp16-S añadiendo 2 $\mu$ l de RNA. 9. C-, 10. C+, 11. C7(1:1), 12. C9(1:1), 13. D2(1:1), 14. L4(1:1), 15. L6(1:1), 16. L8(1:1). Gel con cebador nsp12-13 añadiendo 2 $\mu$ l de RNA. 17. C-, 18. C+, 19. C7(1:1), 20. C9(1:1), 21. D2(1:1), 22. L4(1:1), 23. L6(1:1), 24. L8(1:1). Los demas carriles estan incluidos en el trabajo de Marıa Jerez.



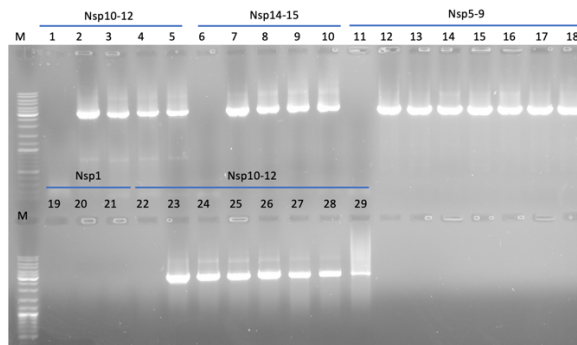
**Figura B1. 15** Gel con cebador nsp16-S añadiendo 2 $\mu$ l de RNA. 1. C-, 2. C+, 3. C7(1:1), 4. C9(1:1), 5. D2(1:1), 6. L4(1:1), 7. L6(1:1), 8. L8(1:1). Gel con cebador S añadiendo 2 $\mu$ l de RNA. 9. C-, 10. C+, 11. C7(1:1), 12. C9(1:1), 13. D2(1:1), 14. L4(1:1), 15. L6(1:1), 16. L8(1:1).



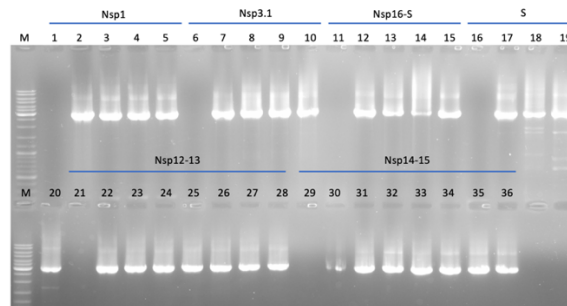
**Figura B1. 16** Gel con cebador ORF6 añadiendo 2 $\mu$ l de RNA. 17. C-, 18. C+, 19. L4(1:1), 20. L6(1:1), 21. L8(1:1), 22. C7(1:1), 23. C9(1:1), 24. D2(1:1). Gel con cebador N añadiendo 2 $\mu$ l de RNA. 25. C-, 26. C+, 27. L4(1:1), 28. L6(1:1), 29. L8(1:1), 30. C7(1:1), 31. C9(1:1), 32. D2(1:1). Los demás carriles están incluidos en el trabajo de María Jeréz.



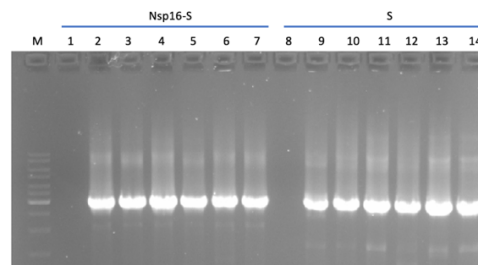
**Figura B1. 17** Gel con cebador ORF6 añadiendo 4 $\mu$ l de RNA. 12. C-, 13. C1(1:1), 14. C3(1:1), 15. C5(1:1). Gel con cebador N añadiendo 4 $\mu$ l de RNA. 16. C-, 17. C1(1:1), 18. C3(1:1), 19. C5(1:1). Los demás carriles están incluidos en el trabajo de María Jeréz.



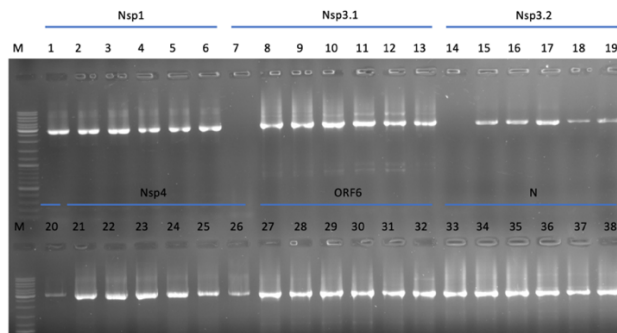
**Figura B1. 18** Gel con cebador nsp14-15 añadiendo 4µl de RNA. 6. C-, 7. C+, 8. C1(1:1), 9. C3(1:1), 10. C5(1:1). Los demás carriles están incluidos en el trabajo de María Jeréz.



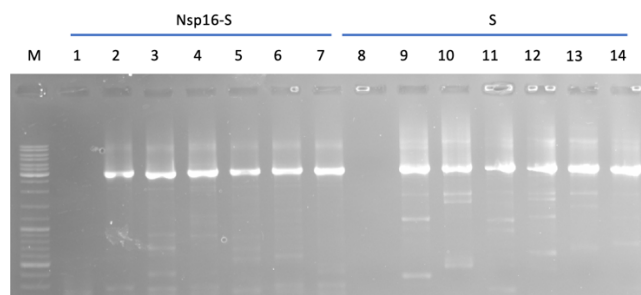
**Figura B1. 19** Gel con cebador nsp16-S añadiendo 4µl de RNA. 11. C-, 12. C+, 13. C1(1:1), 14. C3(1:1), 15. C5(1:1). Gel con cebador S añadiendo 4µl de RNA. 16. C-, 17. C+, 18. C1(1:1), 19. C3(1:1), 20. C5(1:1). Gel con cebador nsp12-13 añadiendo 4µl de RNA. 21. C-, 22. C+, 23. D4(1:1), 24. D6(1:1), 25. D8(1:1), 26. M1(1:1), 27. M3(1:1), 28. M5(1:1). Gel con cebador nsp14-15 añadiendo 4µl de RNA. 29. C-, 30. C+, 31. D4(1:1), 32. D6(1:1), 33. D8(1:1), 34. M1(1:1), 35. M3(1:1), 36. M5(1:1). Los demás carriles están incluidos en el trabajo de María Jeréz.



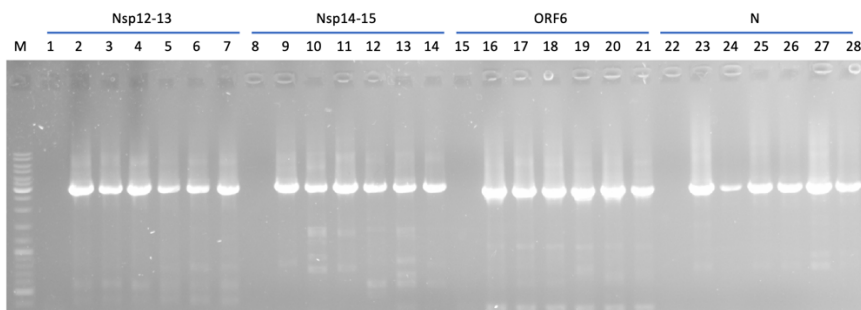
**Figura B1. 20** Gel con cebador nsp16-S añadiendo 4µl de RNA. 1. C-, 2. D4(1:1), 3. D6(1:1), 4. D8(1:1), 5. M1(1:1), 6. M3(1:1), 7. M5(1:1). Gel con cebador S añadiendo 4µl de RNA. 8. C-, 9. D4(1:1), 10. D6(1:1), 11. D8(1:1), 12. M1(1:1), 13. M3(1:1), 14. M5(1:1).



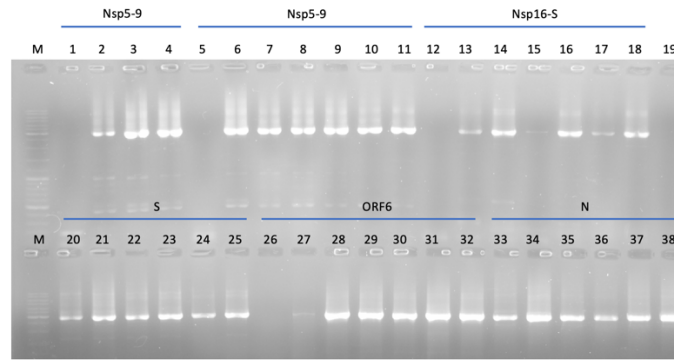
**Figura B1. 21** Gel con cebador ORF6 añadiendo 4 $\mu$ l de RNA. 27. D4(1:1), 28. D6(1:1), 29. D8(1:1), 30. M1(1:1), 31. M3(1:1), 32. M5(1:1). Gel con cebador N añadiendo 4 $\mu$ l de RNA. 33. D4(1:1), 34. D6(1:1), 35. D8(1:1), 36. M1(1:1), 37. M3(1:1), 38. M5(1:1). Los demás carriles están incluidos en el trabajo de María Jeréz.



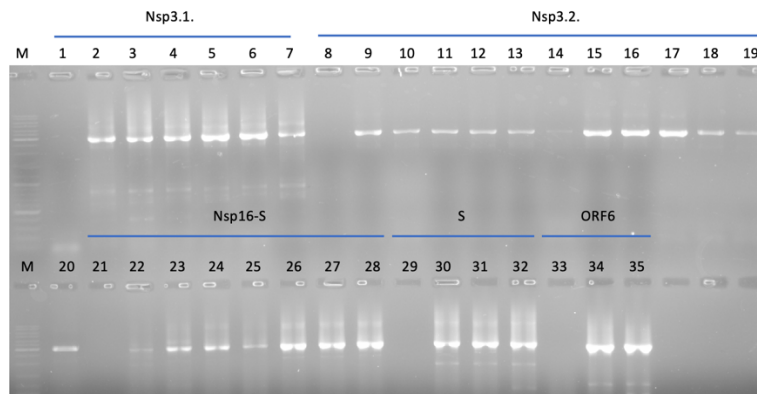
**Figura B1. 22** Gel con cebador nsp16-S añadiendo 5 $\mu$ l de RNA. 1. C-, 2. E1(1:1), 3. E3(1:1), 4. E3(1:1), 5. M7(1:1), 6. M9(1:1), 7. N2(1:1). Gel con cebador S añadiendo 4 $\mu$ l de RNA. 8. C-, 9. E1(1:1), 10. E3(1:1), 11. E5(1:1), 12. M7(1:1), 13. M9(1:1), 14. N2(1:1).



**Figura B1. 23** Gel con cebador nsp12-13 añadiendo 5 $\mu$ l de RNA. 1. C-, 2. E1(1:1), 3. E3(1:1), 4. E3(1:1), 5. M7(1:1), 6. M9(1:1), 7. N2(1:1). Gel con cebador nsp14-15 añadiendo 4 $\mu$ l de RNA. 8. C-, 9. E1(1:1), 10. E3(1:1), 11. E5(1:1), 12. M7(1:1), 13. M9(1:1), 14. N2(1:1). Gel con cebador ORF6 añadiendo 5 $\mu$ l de RNA. 15. C-, 16. E1(1:1), 17. E3(1:1), 18. E3(1:1), 19. M7(1:1), 20. M9(1:1), 21. N2(1:1). Gel con cebador nsp14-15 añadiendo 4 $\mu$ l de RNA. 22. C-, 23. E1(1:1), 24. E3(1:1), 25. E5(1:1), 26. M7(1:1), 27. M9(1:1), 28. N2(1:1).



**Figura B1. 24** Gel con cebador nsp16-S añadiendo 6µl de RNA. 12. C-, 13. E7(1:1), 14. E9(1:1), 15. F2(1:1), 16. N4(1:1), 17. N6(1:1), 18. N8(1:1). Gel con cebador S añadiendo 6µl de RNA. 19. C-, 20. E7(1:1), 21. E9(1:1), 22. F2(1:1), 23. N4(1:1), 24. N6(1:1), 25. N8(1:1). Gel con cebador ORF6 añadiendo 6µl de RNA. 26. C-, 27. E7(1:1), 28. E9(1:1), 29. F2(1:1), 30. N4(1:1), 31. N6(1:1), 32. N8(1:1). Gel con cebador S añadiendo 6µl de RNA. 33. E7(1:1), 34. E9(1:1), 35. F2(1:1), 36. N4(1:1), 37. N6(1:1), 38. N8(1:1). Los demás carriles están incluidos en el trabajo de María Jeréz.



- |   |   |   |
|---|---|---|
| 1. Control negativo                             | 13. 5ªFU #1 FU 500 µM réplica B (N6) 1:1 (6 µl) | 25. 4ªFU #6 FU 500 µM réplica C (F2) 1:1 (6 µl) |
| 2. 4ªFU #6 FU 500 µM réplica B (E7) 1:1 (6 µl)  | 14. 5ªFU #1 FU 500 µM réplica C (N8) 1:1 (6 µl) | 26. 4ªFU 400 µM réplica A (D6) 1:1 (4 µl)       |
| 3. 4ªFU #6 FU 500 µM réplica B (E9) 1:1 (6 µl)  | 15. 4ªFU #6 FU 400 µM réplica B (E1) 1:1 (5 µl) | 27. 4ªFU 400 µM réplica B (D8) 1:1 (4 µl)       |
| 4. 4ªFU #6 FU 500 µM réplica C (F2) 1:1 (6 µl)  | 16. 4ªFU #6 FU 400 µM réplica B (E3) 1:1 (5 µl) | 28. 4ªFU 400 µM réplica C (M1) 1:1 (4 µl)       |
| 5. 5ªFU #1 FU 500 µM réplica A (N4) 1:1 (6 µl)  | 17. 4ªFU #6 FU 400 µM réplica C (E5) 1:1 (5 µl) | 29. Control negativo                            |
| 6. 5ªFU #1 FU 500 µM réplica B (N6) 1:1 (6 µl)  | 18. 5ªFU #1 FU 400 µM réplica A (M7) 1:1 (5 µl) | 30. 4ªFU 400 µM réplica A (B4) 1:1 (4 µl)       |
| 7. 5ªFU #1 FU 500 µM réplica C (N8) 1:1 (6 µl)  | 19. 5ªFU #1 FU 400 µM réplica B (M9) 1:1 (5 µl) | 31. 4ªFU 400 µM réplica B (B6) 1:1 (4 µl)       |
| 8. Control negativo                             | 20. 5ªFU #1 FU 400 µM réplica C (N2) 1:1 (5 µl) | 32. 4ªFU 400 µM réplica C (B8) 1:1 (4 µl)       |
| 9. 4ªFU #6 FU 500 µM réplica B (E7) 1:1 (6 µl)  | 21. Control negativo                            | 33. Control negativo                            |
| 10. 4ªFU #6 FU 500 µM réplica B (E9) 1:1 (6 µl) | 22. 5ªFU #1 FU 500 µM réplica B (N6) 1:1 (6 µl) | 34. 4ªFU 400 µM réplica A (B4) 1:1 (4 µl)       |
| 11. 4ªFU #6 FU 500 µM réplica C (F2) 1:1 (6 µl) | 23. 4ªFU #6 FU 500 µM réplica B (E7) 1:1 (6 µl) | 35. 4ªFU #6 FU 500 µM réplica A (E7) 1:1 (6 µl) |
| 12. 5ªFU #1 FU 500 µM réplica A (N4) 1:1 (6 µl) | 24. 4ªFU #6 FU 500 µM réplica B (E9) 1:1 (6 µl) |   |

**Figura B1. 25** Gel con cebador nsp16-S añadiendo 6µl de RNA. 21. C-, 22. N6(1:1), 23. E7(1:1), 24. E9(1:1), 25. F2(1:1), 26. D6(1:1), 27. D8(1:1), 28. M1(1:1). Gel con cebador nsp16-S añadiendo 4µl de RNA. 29. C-, 30. B4(1:1), 31. B6(1:1), 32. B8(1:1). Gel con cebador ORF6 añadiendo 4µl de RNA. 33. C-, 34. B4(1:1), 35. E7(1:1). Los demás carriles están incluidos en el trabajo de María Jeréz.



UNIVERSIDAD  
DE MÁLAGA

| [uma.es](http://uma.es)

E.T.S de Ingeniería Informática  
Bulevar Louis Pasteur, 35  
Campus de Teatinos  
29071 Málaga

E.T.S. DE INGENIERÍA INFORMÁTICA