

# **Análisis evolutivo y estructural de las proteínas nsp10, nsp14 y nsp16 para el diseño de una terapia antiviral pancoronavírica**

**Evolutionary and structural analysis of nsp10, nsp14 and nsp16 proteins for the design of a pan-coronaviral antiviral therapy**

**TRABAJO FIN DE GRADO EN BIOLOGÍA  
UNIVERSIDAD DE MÁLAGA**

Autor: Francisco José Villena González

Área de conocimiento y Departamento: Área de Genética. Dpto. de Biología Celular, Genética y Fisiología

Fecha de presentación: mayo de 2023

Tipo: Experimental



# **DECLARACIÓN DE ORIGINALIDAD DEL TFG**

D./Dña. Francisco José Villena González con DNI (NIE o pasaporte) XXXXXXXXN, estudiante del Grado en Biología de la Facultad de Ciencias de la Universidad de Málaga,

## **DECLARO:**

Que he realizado el Trabajo Fin de Grado titulado *"Análisis evolutivo y estructural de las proteínas nsp10, nsp14 y nsp16 para el diseño de una terapia antiviral pancoronavírica"* y que lo presento para su evaluación.

Dicho trabajo es original y todas las fuentes bibliográficas utilizadas para su realización han sido debidamente citadas en el mismo.

De no cumplir con este compromiso, soy consciente de que, de acuerdo con la Normativa reguladora de los procesos de evaluación de los aprendizajes del estudiantado de la Universidad de Málaga de 23 de julio de 2019, esto podrá conllevar la calificación de suspenso en la asignatura, sin perjuicio de las responsabilidades disciplinarias en las que pudiera incurrir en caso de plagio.

Para que así conste, firmo la presente en Málaga, el 28 de mayo de 2023

Fdo: Francisco José Villena González DNI: XXXXXXXXN



# **Análisis evolutivo y estructural de las proteínas nsp10, nsp14 y nsp16 para el diseño de una terapia antiviral pancoronavírica**

## **RESUMEN:**

Durante este siglo, se han registrado tres alertas sanitarias globales causadas por la aparición de nuevos coronavirus zoonóticos que pueden generar infecciones respiratorias agudas; SARS-CoV (2002), MERS-CoV (2012), SARS-CoV-2 (2019). En respuesta a la posible aparición de futuros coronavirus pandémicos, se destaca el desarrollo de múltiples terapias antivirales que se dirijan a características víricas conservadas y que puedan ser utilizadas en combinación, estableciéndose como la primera línea de defensa terapéutica más efectiva. Dentro de este contexto, en este Trabajo de Fin de Grado se realiza un análisis de la evolución y estructura de dos proteínas clave en el complejo de replicación y transcripción de los coronavirus, nsp14 y nsp16, así como del cofactor que ambas comparten, nsp10. Las regiones conservadas de esta última podrían ser claves para el diseño de pseudoligandos que ayuden a frenar el ciclo infectivo de coronavirus en humanos.

**Palabras clave:** coronavirus, evolución, estrategia antiviral, filogenia, análisis estructural, nsp10, nsp14, nsp16.



# **Evolutionary and structural analysis of nsp10, nsp14 and nsp16 proteins for the design of a pancoronaviral antiviral therapy**

## **ABSTRACT:**

During this century, there have been three global health alerts caused by the emergence of new zoonotic coronaviruses that can cause acute respiratory infections: SARS-CoV (2002), MERS-CoV (2012) and SARS-CoV-2 (2019). In response to the possible emergence of future pandemic coronaviruses, the development of multiple antiviral therapies that target conserved viral characteristics which can be used in combination is highlighted as the most effective first line of therapeutic defence. In this context, this Final Degree Project analyses the evolution and structure of two key proteins in the replication and transcription complex of coronaviruses, nsp14 and nsp16, as well as the cofactor they share, nsp10. The conserved regions of the latter could be key to the design of pseudoligands that help to stop the infectious cycle of coronaviruses in humans.

**Keywords:** coronavirus, evolution, antiviral strategy, phylogeny, structural analysis, nsp10, nsp14, nsp16.



# Índice

<b>1. Introducción</b>	<b>1</b>
1.1. Generalidades sobre la clasificación de los virus . . . . .	1
1.2. Taxonomía de los coronavirus . . . . .	1
1.3. Estructura y genoma de los coronavirus . . . . .	4
1.4. Bases de una posible terapia antiviral pancoronavírica . . . . .	5
<b>2. Objetivos</b>	<b>7</b>
<b>3. Material y métodos</b>	<b>9</b>
3.1. Análisis de secuencias aminoacídicas . . . . .	9
3.1.1. Búsqueda y selección de secuencias . . . . .	9
3.1.2. Formato y organización de las secuencias . . . . .	10
3.1.3. Alineamiento de secuencias . . . . .	10
3.1.4. Optimización de los alineamientos . . . . .	10
3.1.5. Construcción de árboles filogenéticos . . . . .	11
3.1.6. Visualización y edición de árboles filogenéticos . . . . .	12
3.2. Análisis estructural de complejos proteicos . . . . .	12
3.2.1. Búsqueda y selección de estructuras de referencia . . . . .	12
3.2.2. Análisis de interacciones globales entre proteínas . . . . .	13
3.2.3. Análisis de interacción de aminoácidos específicos . . . . .	13
3.3. Composición de figuras de resultados . . . . .	14
3.3.1. Renderizado de complejos proteicos . . . . .	14
3.3.2. Graficación de interacciones de nsp10 . . . . .	14
3.3.3. Construcción de logos de secuencias . . . . .	14
<b>4. Resultados</b>	<b>15</b>
4.1. Alineamientos de secuencias aminoacídicas de nsp10/14/16 . . . . .	15
4.2. Filogenias de nsp10/14/16 en <i>Orthocoronavirinae</i> . . . . .	15
4.3. Región de nsp10 con máxima interacción con nsp14 y nsp16 . . . . .	20
4.4. Logos de secuencias de la zona de máxima interacción de nsp10 y los dominios de nsp14 y nsp16 implicados. . . . .	20
<b>5. Discusión y futuras líneas</b>	<b>23</b>
5.1. Análisis de los alineamientos de nsp10/14/16 . . . . .	23
5.2. Historia evolutiva de nsp10/14/16 . . . . .	23
5.3. Región de nsp10 seleccionada . . . . .	24
<b>6. Conclusiones</b>	<b>27</b>
<b>Referencias</b>	<b>29</b>
<b>Anexos</b>	<b>39</b>



# 1. Introducción

## 1.1. Generalidades sobre la clasificación de los virus

Desde un enfoque evolutivo, un organismo puede ser definido como la unidad elemental de un linaje [genético] continuo y con una historia evolutiva individual (Luria y col., 1978). Los virus son organismos acelulares con genomas constituidos por ácidos nucleicos, que se replican obligadamente dentro de células hospedadoras utilizando los ribosomas y la maquinaria metabólica del huésped, sintetizando un conjunto de componentes que se autoensamblan formando viriones; partículas infectivas que protegen al genoma viral y permiten su transferencia a otras células (Claverie y Abergel, 2016; Koonin y col., 2021; Nasir y col., 2020; Rybicki, 2023).

En las últimas dos décadas, ha habido un avance significativo en la clasificación de los virus. Anteriormente, se utilizaba una taxonomía basada en características fenotípicas, como la estructura del virión, el rango de hospedación y la epidemiología. Sin embargo, actualmente la clasificación de los virus se ha transformado, adoptando una aproximación similar a la de los organismos celulares. Esta nueva clasificación se fundamenta principalmente en datos genómicos y filogenéticos, permitiendo conocer y reflejar sus relaciones evolutivas (Koonin y col., 2015; Koonin y col., 2020; Simmonds y col., 2017)

Los principios, procedimientos y nomenclatura empleados para nombrar y clasificar a los organismos, incluyendo los virus, están sujetos a la regulación de comités científicos internacionales. El *International Committee on Taxonomy of Viruses* (ICTV) es responsable de la taxonomía viral, de manera similar a cómo el *International Committee on Systematics of Prokaryotes* (ICSP) se encarga de la denominación de especies de bacterias y arqueas, o la *International Commission on Zoological Nomenclature* (ICZN) se encarga de las animales (ICZN, s.f.; Lefkowitz y col., 2018; Oren y col., 2023)

La aparición y estandarización de la nomenclatura binomial es un hecho bastante reciente en virología. Los nombres de los virus previos a la ratificación de la norma en 2021 han pasado a ser nombres comunes, siendo en la actualidad asunto del ICTV la designación de los nombres científicos de todas las especies reconocidas (Zerbini y col., 2022). A modo de ejemplo: el virus del salmón del Pacífico (PsNV) pertenece al género *Alphapironavirüs* y a la especie *Alphapironavirüs bona*.

Aunque la nomenclatura binomial se ha convertido en el formato oficial para nombrar a las especies, muchas de las definidas antes de la aprobación de esta regla aún no cumplen con este formato. Esta particularidad transitoria también afecta de forma significativa a las especies de la familia *Coronaviridae*. Por ello, en este trabajo se dará preferencia al uso de los nombres comunes, los cuales siguen siendo populares y ampliamente utilizados en la literatura científica.

## 1.2. Taxonomía de los coronavirus

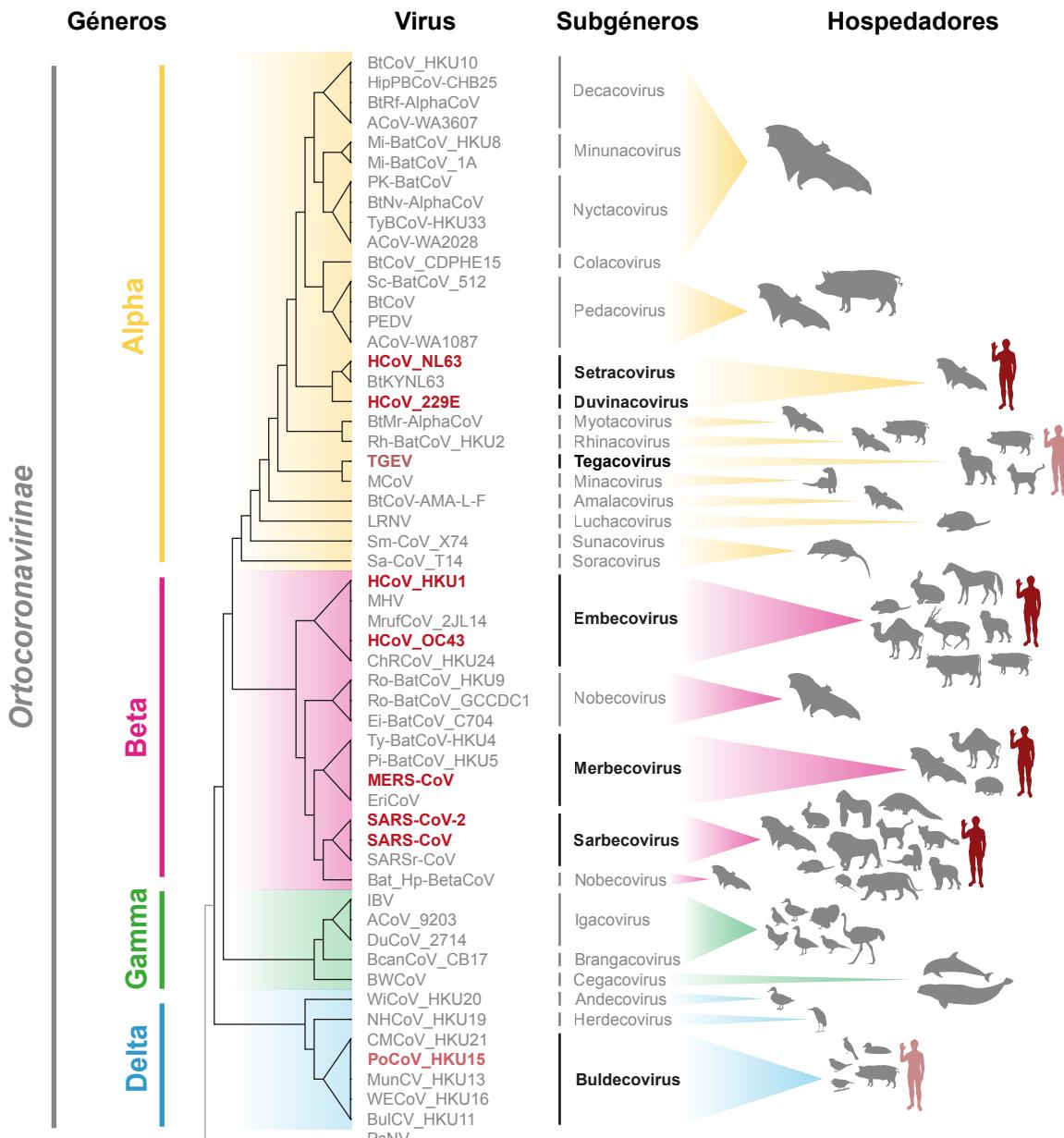
Dentro del orden *Nidovirales*, los miembros de la familia *Coronaviridae* se distribuyen en varias subfamilias que infectan a diversos grupos de vertebrados. Según el último reporte del ICTV sobre la taxonomía de estos virus, los tradicionalmente conocidos como coronavirus, que afectan a aves y mamíferos, se agrupan dentro de la subfamilia *Orthocoronavirinae* (Woo y col., 2023). Los estudios metagenómicos han ampliado el conocimiento

sobre el rango de hospedación de esta familia, añadiendo dos nuevas subfamilias: *Letovirinae* (Bukhari y col., 2018) (Bukhari et al., 2018), que afectan a anfibios, y *Pitovirinae*, presente en peces óseos (Mordecai y col., 2019). Además, se ha reconocido la existencia de secuencias de coronavirus encontradas en reptiles (Shi y col., 2018) y peces agnatos (Miller y col., 2021), las cuales están pendientes de revisión formal.

Hasta la fecha de realización de este trabajo, se ha reconocido una única especie para las subfamilias *Letovirinae* y *Pitovirinae*. La mayoría de los coronavirus conocidos pertenecen a la subfamilia *Orthocoronavirinae* y se distribuyen en cuatro géneros: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* y *Deltacoronavirus* (Woo y col., 2023).

Los coronavirus humanos que causan resfriados comunes (HCoVs) se clasifican en los géneros *Alphacoronavirus* y *Betacoronavirus*. En el género *Alphacoronavirus* se encuentran el HCoV-NL63 y el HCoV-229E, mientras que en el género *Betacoronavirus* se incluyen el HCoV-HKU1 y el HCoV-OC43 (Liu y col., 2021). Además, dentro del género *Betacoronavirus*, existen otros coronavirus zoonóticos pandémicos capaces de causar enfermedades graves, como el síndrome respiratorio agudo severo (SARS), causado por los virus SARS-CoV y SARS-CoV-2, y el síndrome respiratorio de Oriente Medio (MERS), causado por el MERS-CoV (Peiris y Poon, 2021). Es importante destacar que, a pesar de sus diferencias clínicas significativas, los virus del SARS se clasifican taxonómicamente dentro de la especie *Severe acute respiratory syndrome-related coronavirus* (Gorbalenya y col., 2020).

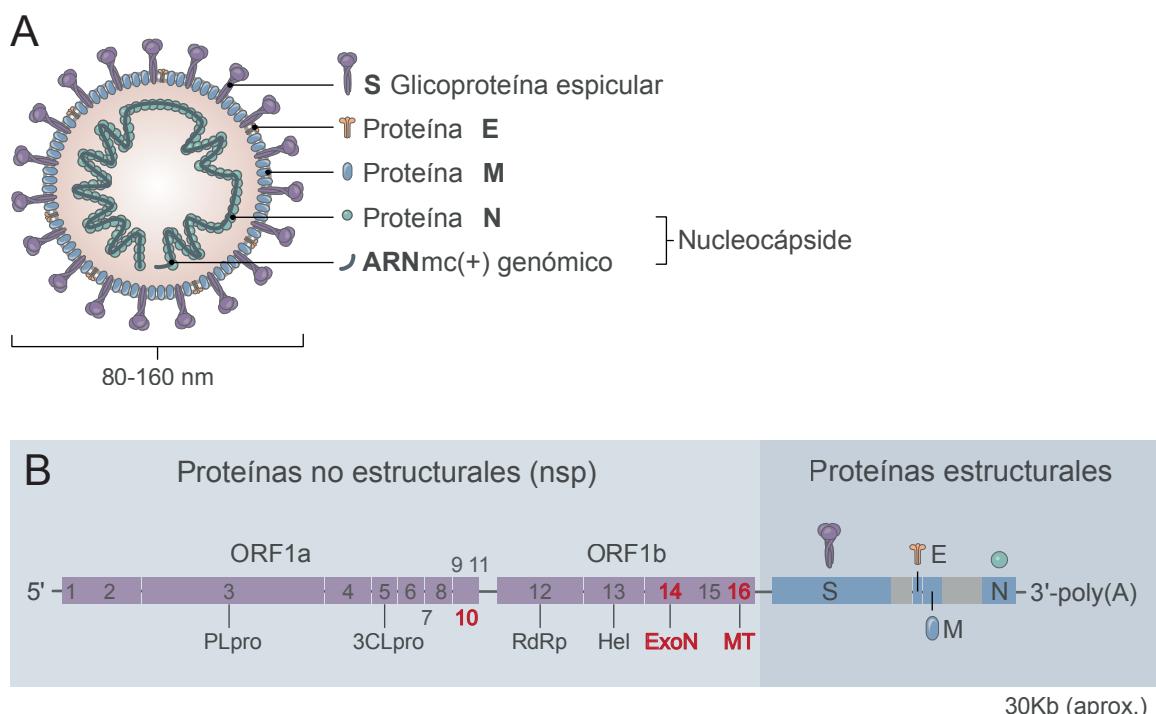
Hasta el año 2021, los reportes de las infecciones humanas por coronavirus se limitaban a especies de alfa- y betacoronavirus; hasta la identificación de secuencias del deltacoronavirus porcino (PoCoV-HKU15) en muestras de sangre de tres niños haitianos con un cuadro febril agudo inespecífico (Lednický y col., 2021). Un año más tarde se reportó en Malasia el salto de virus de la gastroenteritis transmisible (TGEV) caninos a 8 pacientes ingresados por neumonía durante 2017–2018 (Vlasova y col., 2022). Estos hallazgos, dos años después de la pandemia mundial causada por el SARS-CoV-2, resaltan nuevamente la posibilidad de que nuevos coronavirus, previamente no asociados a los humanos, puedan realizar saltos zoonóticos que vuelvan a poner en riesgo nuestra salud (**Figura 1**).



**Figura 1:** Cladograma de la subfamilia Orthocoronavirinae enraizado a partir de la secuencia del PsNV (subfamilia Pitovirinae), con sus relativos hospedadores agrupados por subgéneros según el último informe del ICTV (Woo y col., 2023). En términos generales, los virus de los géneros *Alphacoronavirus* y *Betacoronavirus* se han detectado en una amplia gama de mamíferos, mientras que los géneros *Deltacoronavirus* y *Gammacoronavirus* se han encontrado principalmente en aves. Los hospedadores se indican a través de siluetas tomadas de PhyloPic (Keesey, 2023). Los virus que infectan a los seres humanos se marcan en un color rojo intenso, mientras que aquellos cuyo reporte ha sido ocasional se resaltan en un tono más suave. Los nombres completos de los virus y sus especies asociadas quedan recogidos en la (**Tabla A1**), disponible en anexos. La figura es de elaboración propia para este trabajo.

### 1.3. Estructura y genoma de los coronavirus

En líneas generales, los viriones de la familia *Coronaviridae* son partículas esféricas con un diámetro aproximado de 80–160 nm. Presentan una envoltura constituida por una bicapa lipídica sobre la que se anclan proteínas estructurales de membrana (M), envoltura (E) y espiga (S) (**Figura 2A**). Las proteínas M y E comparten papel en la morfogénesis del virión, siendo esta última también un factor de virulencia para algunas especies (Jimenez-Guardeño y col., 2014; Neuman y col., 2011; Ye y Hogue, 2007). Por otro lado, la proteína S media en el proceso de reconocimiento, fijación y entrada a la célula hospedadora (Bosch y col., 2003). Dentro del virión se encuentra la nucleocápside, la cual a su vez está constituida por un genoma viral de ARN monocatenario (ARNmc) positivo de unas 30 kilobases (Kb) encapsulado por nucleoproteínas (N) que participan en su síntesis y traducción (Enjuanes y col., 2006) (**Figura 2B**).



**Figura 2:** Estructura (**A**) y genoma (**B**) de Coronaviridae. El material genético de los coronavirus posee un primer marco abierto de lectura (ORF1a) que codifica proteínas relacionadas con la evasión inmunitaria, mientras el segundo (ORF1b) contiene a la ARN polimerasa viral dependiente de ARN (RdRp) y otras proteínas no estructurales asociadas con la replicación. Ambos ORFs codifican poliproteínas que procesadas por las proteinasas virales (nsp3 y nsp5) dan lugar a cada nsp individual. El otro tercio del genoma contiene los ORFs de las proteínas estructurales del virión. En color rojo se destacan en la figura las actividades enzimáticas clave para este trabajo: exonucleasa (ExoN) y metiltransferasa (MT) desarrolladas por nsp14 y nsp16 respectivamente, y nsp10 (cofactor de ambas). Figura tomada y modificada de (Asensio, 2020a).

El tamaño de los genomas de la mayoría de los virus de ARN varía dentro de un estrecho rango de aproximadamente 1–15 Kb. Esto se considera consecuencia de la baja fidelidad en sus procesos de replicación, debido a la falta de actividad correctora de errores (Duffy, 2018; Peck y Lauring, 2018; Sanjuán y col., 2010). Los coronavirus sobrepasan ampliamente este límite con sus ~30 kb, siendo de los genomas virales de ARN más grandes conocidos. La expansión de sus genomas, así como en otras especies del orden

Nidovirales, viene justificada por la aparición de la capacidad de corrección de errores 3'→5' exoribonucleasa (ExoN) en sus complejos de replicación y transcripción (RTC). Dicha actividad enzimática fue descubierta en virus de ARN por primera vez en 2003, gracias al análisis bioinformático del genoma de SARS-CoV. La identidad de secuencia de exonucleasas celulares compartida con un fragmento de la nsp14 de aquel coronavirus evidenciaron la funcionalidad de esta última (Snijder y col., 2003). Tres años más tarde, esta predicción fue confirmada *in vitro* (Minskaia y col., 2006).

#### 1.4. Bases de una posible terapia antiviral pancoronavírica

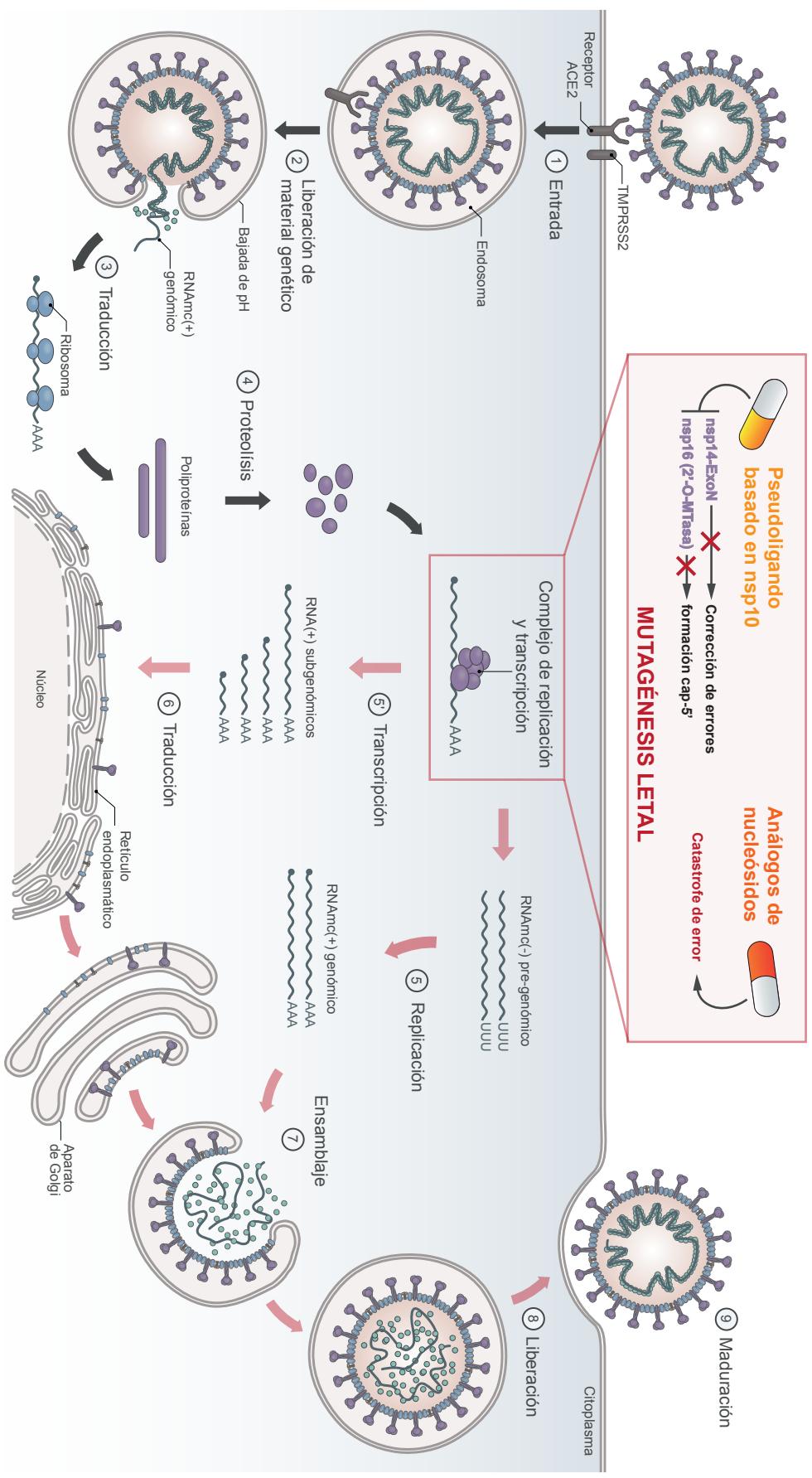
Los virus de ARN operan con tasas de error cercanas a los límites compatibles con el mantenimiento de su información genética. Existe una estrategia antiviral conocida como mutagénesis letal, que implica la administración de análogos de nucleósidos, con el objetivo de aumentar las tasas de mutación viral y así reducir y eliminar la carga viral (Díaz-Martínez y col., 2018; Swanstrom y Schinazi, 2022). Algunos ejemplos de la efectividad clínica de estas terapias contra enfermedades humanas causadas por virus de ARN son el uso de ribavirina contra el dengue y la hepatitis C, o el remdesivir contra el ébola (Huggins y col., 1991; Mulangu y col., 2019; Ortega-Prieto y col., 2013). Sin embargo, los intentos de aplicar estos y otros análogos de nucleósidos para mitigar la pandemia coronavírica del 2019 (COVID-19) causada por el SARS-CoV-2 fracasaron, principalmente por la actividad ExoN de la nsp14 (Stevaert y col., 2022).

La nsp14 de los coronavirus es una proteína bifuncional, constituida por el dominio aminoterminal con actividad ExoN (nsp14-ExoN) y otro con actividad Guanina-N7-metiltransferasa (nsp14-MT) en su extremo carboxilo (Chen y col., 2009). Para la estabilización y correcto funcionamiento de su sitio activo, la nsp14-ExoN requiere interactuar con una unidad de nsp10, mientras que la nsp14-MT lleva a cabo su actividad sin la necesidad de un cofactor proteico adicional (Ferron y col., 2018).

Además, la nsp10 también actúa como cofactor para la actividad de la nsp16. Esta última es una enzima con actividad 2'-O-metiltransferasa que, junto con la actividad de la nsp14-MT, desempeña un papel crucial en la formación de la caperuza 5' del genoma viral; estructura protectora que evita el reconocimiento del genoma viral por parte del sistema inmunitario innato (Pan y col., 2022).

Las enzimas eucariotas que realizan actividades semejantes a las de la nsp14-ExoN y la nsp16 de los coronavirus presentan estructuras considerablemente diferentes y no requieren de cofactores proteicos para desempeñar sus funciones (Galloway y Cowling, 2019; Rona y col., 2022). Estas diferencias estructurales abren la posibilidad de aprovechar las interacciones con la nsp10 como diana antiviral pancoronavírica.

Desde 2020, el equipo de investigación liderado por la Dra. Ana Grande en el campo de “Evolución de virus y terapias antivirales” está trabajando en el diseño de posibles pseudoligandos, basados en estructuras del SARS-CoV-2, que compitan con los dominios de unión para nsp10 de la nsp14-ExoN y la nsp16. Además, se pretende aumentar la efectividad de dicha terapia mediante su uso combinado con análogos de bases, que dirijan a los coronavirus más rápidamente hacia la catástrofe de error por mutagénesis letal (**Figura 3**). Los estudios en relación a este proyecto se llevan a cabo en la Universidad de Málaga, en colaboración con el Centro de Biología Molecular ‘Severo Ochoa’ y la Universidad de Elche (Grande-Pérez, 2021).



**Figura 3:** Ciclo replicativo del SARS-CoV-2 y la estrategia antiviral teórica que fundamenta los objetivos de este trabajo. La adición del pseudoligando basado en nspl10 competiría por el sitio de interacción de este mismo cofactor e impediría la correcta estabilización de las actividades de nspl4-ExoN y nspl6. La supresión de estas enzimas reduciría la fidelidad del complejo de replicación y transcripción (CRT) e inhibiría la formación del cap-5', exponiendo al genoma viral a su degradación por la respuesta innata celular. Además, gracias a la inactivación de la actividad correctora de la nspl4-ExoN, la adición de análogos de bases aceleraría la tasa de mutación durante la replicación y transcripción del genoma viral. Esto provocaría en el genoma del virus una pérdida acumulativa de su información (catástrofe de error), que aceleraría la extinción del virus por acumulación de mutaciones letales en su genoma. Figura adaptada de (Asensio, 2020b).

## **2. Objetivos**

El objetivo general de este trabajo es efectuar un análisis de las secuencias de nsp10/14/16 en *Orthocoronavirinae*, de cara al posible diseño de un antiviral panco-coronavírico que permita frenar en humanos el ciclo biológico de coronavirus actuales y/o emergentes. Para la consecución de esta tarea, se plantearon los siguientes objetivos específicos:

1. Evaluar la conservación y evolución de las proteínas nsp10/14/16 dentro del subfamilia *Orthocoronavirinae*, mediante el análisis de secuencias aminoacídicas disponibles en bases de datos.
2. Identificar la región de nsp10 con máxima interacción con nsp14-ExoN y nsp16, en base al análisis de sus estructuras resueltas por cristalografía de rayos X para SARS-CoV-2.
3. Analizar la conservación de los residuos de la región de nsp10 identificada, así como de los residuos de nsp14-ExoN y nsp16 involucrados en su interacción, a partir de los análisis de secuencias previamente realizados.



### 3. Material y métodos

Para llevar a cabo este trabajo, se utilizó un Macbook Pro (13 pulgadas, 2020) equipado con el chip Apple M1 y memoria RAM de 16 GB, operando bajo el sistema operativo macOS Ventura versión 13.3.1.

Los programas y bases de datos empleados en este trabajo quedan detallados en las **Tablas 1 y 2**.

**Tabla 1:** Programas utilizados en sus versiones para macOS

Nombre	Referencia	Acceso
MAFFT v7.490	Katoh y Standley, 2013	<a href="https://mafft.cbrc.jp/alignment/software/">https://mafft.cbrc.jp/alignment/software/</a>
trimAl v1.4	Capella-Gutiérrez y col., 2009	<a href="https://github.com/inab/trimal">https://github.com/inab/trimal</a>
IQ-TREE v2.2.0	Minh y col., 2020	<a href="http://www.iqtree.org/">http://www.iqtree.org/</a>
FigTree v1.4.4	Rambaut, 2007	<a href="https://github.com/rambaut/figtree/releases">https://github.com/rambaut/figtree/releases</a>
Jalview v2.11.2.6	Waterhouse y col., 2009	<a href="https://www.jalview.org/">https://www.jalview.org/</a>
WebLogo v3.7.12	Crooks y col., 2004	<a href="https://weblogo.threethreeplusone.com/create.cgi">https://weblogo.threethreeplusone.com/create.cgi</a>
ChimeraX v1.6	Pettersen y col., 2021	<a href="https://www.rbvi.ucsf.edu/chimerax/">https://www.rbvi.ucsf.edu/chimerax/</a>
Adobe Illustrator v27.0	Inc., 2023	<a href="https://www.adobe.com/es/products">https://www.adobe.com/es/products</a>

**Tabla 2:** Bases de datos en línea

Nombre	Referencia	Acceso
National Center for Biotechnology Information (NCBI)	Sayers y col., 2023	<a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a>
GenBank	Benson y col., 2013	<a href="https://www.ncbi.nlm.nih.gov/genbank/">https://www.ncbi.nlm.nih.gov/genbank/</a>
Protein Data Bank (PDB)	Berman y col., 2000	<a href="https://www.rcsb.org/">https://www.rcsb.org/</a>
The database of the International Committee on Taxonomy of Viruses (ICTV)	Lefkowitz y col., 2018	<a href="https://ictv.global/">https://ictv.global/</a>

#### 3.1. Análisis de secuencias aminoacídicas

##### 3.1.1. Búsqueda y selección de secuencias

El último informe del grupo de estudio del ICTV responsable de la familia *Coronaviridae* incluye los listados de especies reconocidas dentro de cada género, así como el acceso a un genoma de referencia para cada una dentro del GenBank (Woo y col., 2023). En base a ello, se procedió a la descarga manual de las secuencias aminoacídicas correspondientes a las proteínas nsp10, nsp14 y nsp16 de todas las especies pertenecientes a la subfamilia Orthocoronavirinae. Adicionalmente, fueron incluidas las del único representante de la subfamilia *Pitovirinae* para su uso como especie externa en alineamientos y filogenias. El representante de la subfamilia *Letovirinae* fue descartado debido a que solo constaba una secuencia parcial de su genoma.

Cabe recordar que dentro de la especie Severe acute respiratory syndrome-related coronavirus existen virus diferenciados en base a su elevada relevancia clínica, como es el caso del SARS-CoV y el SARS-CoV-2, y es por lo que sus secuencias fueron añadidas al informe del ICTV y se incluyen en este trabajo. El listado completo de las especies utilizadas se adjunta en la **Tabla A1**.

Los métodos utilizados en este trabajo para el análisis de secuencias se ejemplificaron a partir de nsp10, pero también se aplicaron de manera similar a nsp14 y nsp16.

### 3.1.2. Formato y organización de las secuencias

Todas las secuencias aminoacídicas fueron descargadas como ficheros de texto formato FASTA (.fasta). Con el objetivo de trabajar con las secuencias obtenidas utilizando otras herramientas, fue necesario concatenarlas en un único archivo. Para ello, se utilizó el siguiente comando en la terminal de macOS:

```
1 cat * >> ALL_NSP10.fasta
```

El comando significa concatenar (`cat`) el contenido de todos los archivos (\*) de la carpeta en la que se trabaja y redirigirlos (`>>`) a uno nuevo (`ALL_NSP10.fasta`). Este procedimiento permite agrupar las secuencias, generando un único documento FASTA para cada una de las familias de proteínas nsp10.

### 3.1.3. Alineamiento de secuencias

Se realizó la alineación múltiple de las secuencias biológicas mediante MAFFT en la terminal de macOS, el cual ofrece múltiples estrategias para llevar a cabo el proceso de alineación. En este caso se consideró que el dominio completo de cada familia de proteínas puede alinearse globalmente, por lo que se ejecutó el siguiente comando:

```
1 Mafft --globalpair --maxiterate 1000 ALL_NSP10.fasta >
ALN_ALL_NSP10.fasta
```

Este comando ejecuta el programa (`Mafft`) y establece que las secuencias deben ser alineadas globalmente en toda su extensión (`--globalpair`), utilizando 1000 iteraciones (`--maxiterate 1000`) para refinar el alineamiento. El archivo de entrada utilizado contiene las secuencias a alinear (`ALL_NSP10.fasta`) y el resultado final se exporta (`>`) a un nuevo archivo (`ALN_ALL_NSP10.fasta`).

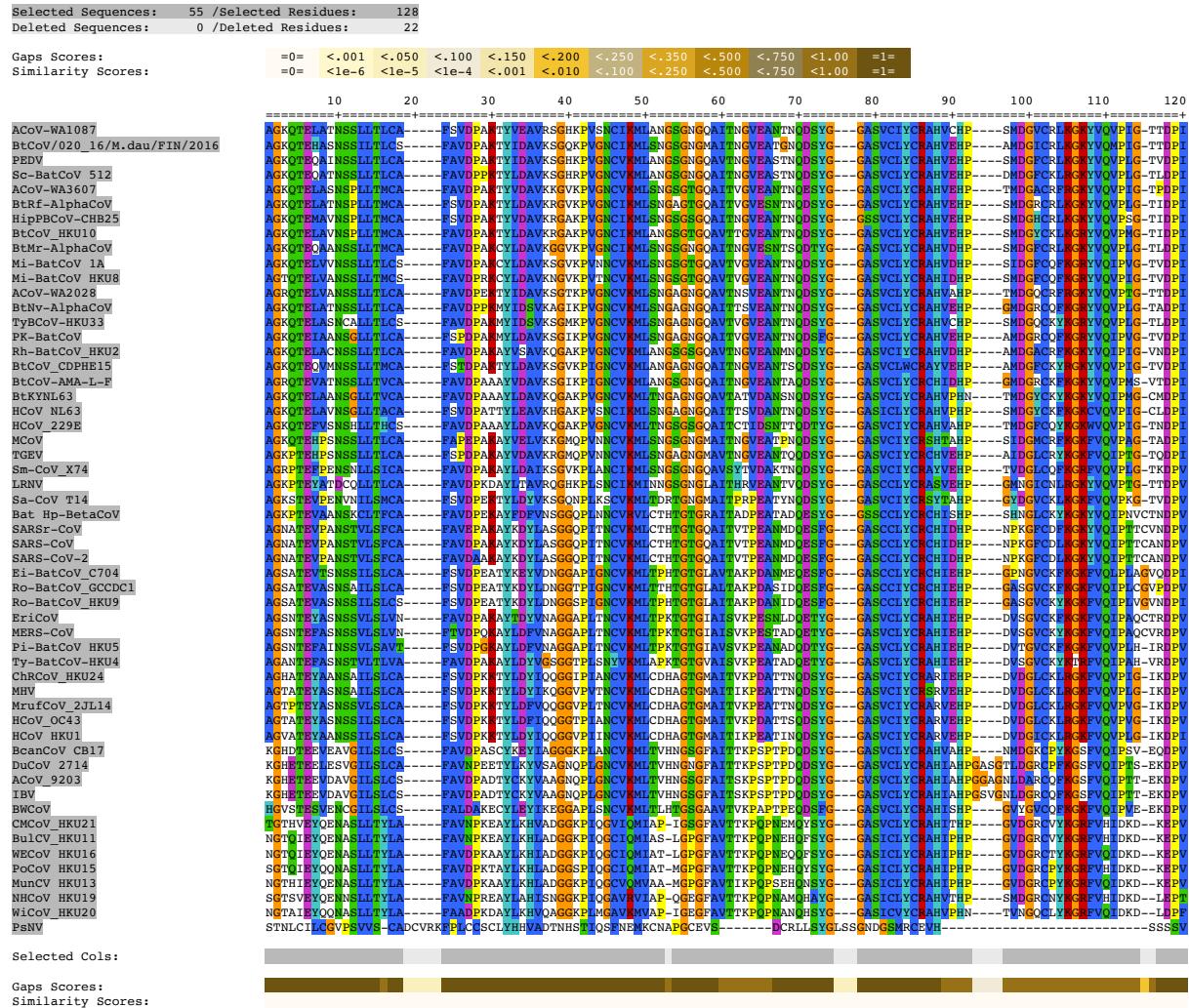
### 3.1.4. Optimización de los alineamientos

Los alineamientos obtenidos fueron depurados con TrimAl, una herramienta bioinformática que permite mejorar la calidad de los alineamientos múltiples y aumentar de este modo la confiabilidad de los análisis filogenéticos. Se utilizó a través de la terminal de macOS mediante el siguiente comando:

```
1 trimal -automated1 -in ALN_ALL_NSP10.fasta -out
TRIM_ALN_ALL_NSP10.fasta -htmlout ALN_ALL_NSP10.html
```

Este comando ejecuta el programa (`trimal`) usando un método heurístico enfocado en optimización de alineamientos que serán utilizados para la inferencia de árboles filogenéticos por métodos de máxima verosimilitud (`-automated1`). A partir de un archivo de entrada

(-in ALN\_ALL\_NSP10.fasta) se obtiene otro de salida con el alineamiento ya optimizado (-out TRIM\_ALN\_ALL\_NSP10.fasta) así como un informe en formato HTML (*HyperText Markup Language*) del proceso de depuración (-htmlout ALN\_ALL\_NSP10.html). El informe utiliza el esquema de colores del programa Clustal X (Thompson y col., 1997) (**Tabla A2**), el cual facilita poder evaluar los alineamientos y detectar posibles problemas (**Figura 4**).



**Figura 4:** Ejemplo de informe de los resultados de depuración de TrimAl. Se muestra parcialmente el alineamiento de entrada de las proteínas nsp10 con esquema de color Clustal X. Debajo, la fila de columnas seleccionadas (Selected Cols) señala en gris oscuro aquellas que se mantendrán en el alineamiento optimizado, y en gris claro las que habrán sido eliminadas..

### 3.1.5. Construcción de árboles filogenéticos

Los árboles filogenéticos fueron inferidos mediante IQ-TREE. Este programa emplea el método de máxima verosimilitud para obtener la representación más probable en función de los parámetros de algún modelo de evolución molecular previamente propuesto. Se utilizó a través de la terminal de macOS mediante el siguiente comando:

```
1 iqtree2 -s TRIM_ALN_ALL_NSP10.fasta --alrt 1000 -B 1000 -m
MFP -T 4
```

El comando ejecuta al programa (*iqtree2*) para inferir un árbol filogenético a partir del alineamiento de secuencias depurado para esta finalidad (*TRIM\_ALN\_ALL\_NSP10.fasta*).

Se aplica una prueba de razón de verosimilitud aproximada de 1000 réplicas (`--alrt 1000`) para evaluar internamente la calidad de las ramificaciones resultantes (Guindon y col., 2010), así como una aproximación de bootstrap ultrarrápido de 1000 réplicas (Hoang y col., 2018). Esta última es un test estadístico ampliamente utilizado para cuantificar porcentualmente la robustez de los nodos de los árboles filogenéticos; aquellos marcados con valores más cercanos a 100 tienen alta probabilidad de ser correctos, mientras que aquellos que se vayan alejando de dicho valor tienen más probabilidad de haber sido producidos por azar. Para seleccionar el modelo de evolución molecular se ejecuta una búsqueda automatizada (`-m MFP`) basada en el criterio de información bayesiana, es decir, la selección del modelo que mejor se ajuste a los datos con la menor complejidad posible (Kalyaanamoorthy y col., 2017). Para acelerar la ejecución de todas estas tareas se destinaron 4 núcleos de procesamiento (`-t 4`).

El comando descrito anteriormente genera varios archivos de salida que llevan el mismo nombre que el archivo de entrada, diferenciables por su extensión. Para este caso, es relevante el que contiene la estructura del árbol consenso generado por IQ-TREE y los valores de bootstrap para cada uno de sus nodos (`TRIM_ALN_ALL_NSP10.fasta.contree`).

### 3.1.6. Visualización y edición de árboles filogenéticos

El programa IQ-TREE produce árboles no enraizados, es decir, no señalan un nodo basal que oriente la dirección del proceso evolutivo en el tiempo. Especialmente en virología, suele sacrificarse la idoneidad teórica de este formato forzando un nodo basal que permita visualizar resultados con una mayor claridad (Gulyaeva y Gorbatenya, 2021; Hu y col., 2021; Ruiz-Aravena y col., 2022).

La visualización de los árboles fue realizada con FigTree. Este programa permite también editarlos mediante una selección de ajustes disponibles en su interfaz gráfica. A partir de los árboles no enraizados, se seleccionó al grupo externo para presentarlos en formato rectangular. Además, se incluyeron los valores de Bootstrap para cada bifurcación y se exportaron las figuras en formato de gráficos vectoriales escalables (SVG) para su posterior edición en Adobe Illustrator.

## 3.2. Análisis estructural de complejos proteicos

### 3.2.1. Búsqueda y selección de estructuras de referencia

Para obtener las estructuras experimentales de los complejos proteicos necesarios para el análisis, se realizaron dos búsquedas en el RCSB PDB utilizando los términos “SARS-CoV-2 nsp10 nsp14 complex” y “SARS-CoV-2 nsp10 nsp16 complex”. Los resultados se ordenan por defecto en base a una puntuación de relevancia basada en la coincidencia con los términos de búsqueda y otros factores, como la resolución de refinamiento (Rose y col., 2021).

En la **Tabla 3** se indican las estructuras tridimensionales de los complejos nsp10/nsp14-ExoN y nsp10/nsp16 resueltas por cristalográfia de rayos X para SARS-CoV-2, con sus respectivos códigos de identificación dentro del PDB (PDB ID).

**Tabla 3:** Bases de datos en línea

Recurso	Fuente	PDB ID
Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-exoribonuclease domain	(Lin et al., 2021)	7DIY
1.80 Angstrom Resolution Crystal Structure of NSP16-NSP10 Complex from SARS-CoV-2	(Rosas-Lemus et al., 2020)	6W4H

Cada estructura seleccionada fue descargada en formato de archivos del PDB (.pdb); un fichero de texto plano que contiene las coordenadas tridimensionales de cada átomo y otros datos relevantes para análisis estructurales (Green y Zardecki, 2019).

### 3.2.2. Análisis de interacciones globales entre proteínas

La visualización y los análisis de las estructuras seleccionadas fueron llevados a cabo con el programa ChimeraX. Este programa dispone de una interfaz gráfica que permite acceder a un subconjunto limitado de funcionalidades, así como una consola de texto que permite el manejo completo y preciso de la herramienta.

Para localizar los aminoácidos involucrados en la interacción nsp10/nsp14-ExoN de la estructura (PDB ID:7DIY), se ejecutaron tres comandos separados secuencialmente por punto y coma:

```
1 sequence chain #1/A; sequence chain #1/B; interfaces #1  
    interfaceResidueAreaCutoff 4.5
```

Al aplicar los comandos sequence chain en la ventana principal de ChimeraX, aparecen las secuencias aminoacídicas en formato texto de nsp10 (#1/A) y nsp14-ExoN (#1/B) en ventanas secundarias. El comando interfaces genera en la esquina inferior derecha un diagrama de interacción del complejo (#1) seleccionando como distancia límite máxima entre residuos 4.5 ángstroms (interfaceResidueAreaCutoff 4.5) en base a las indicaciones del material suplementario de la publicación asociada al complejo (Lin y col., 2021). Al clicar sobre el diagrama la opción (Select Contact Residues of B and A) quedan marcados en la ventana principal y secundarias de color verde los aminoácidos que intervienen en la formación del complejo.

Estas mismas instrucciones fueron ejecutadas para la localización de los aminoácidos involucrados en la interacción nsp10/nsp16 de la correspondiente estructura seleccionada (PDB ID: 6W4H).

### 3.2.3. Análisis de interacción de aminoácidos específicos

Dentro del paquete de herramientas de ChimeraX, la herramienta `contacts` permite identificar y visualizar los residuos implicados en las interacciones no covalentes de regiones concretas. Esta función ha sido utilizada en cada estructura de referencia con el fin de identificar los aminoácidos de nsp14-ExoN (PDB ID: 7DIY) y nsp16 (PDB ID: 6W4H) que participan en la región de máxima interacción de nsp10 compartida para ambos complejos.

### **3.3. Composición de figuras de resultados**

#### **3.3.1. Renderizado de complejos proteicos**

La herramienta save de Chimera X es utilizada para renderizar las visualizaciones moleculares generadas en el programa y permitir la exportación de los resultados en formato de imagen PNG. Posteriormente, estas imágenes fueron montadas utilizando Adobe Illustrator para su presentación y análisis.

#### **3.3.2. Graficación de interacciones de nsp10**

Para representar en un único esquema los aminoácidos de nsp10 involucrados en las interacciones con nsp14-ExoN y nsp16, deducidos a partir de las respectivas estructuras experimentales tomadas como referencia, se utilizó el programa de visualización y edición de alineamientos Jalview. Bajo los aminoácidos de la secuencia de nsp10 se añadieron dos filas de anotaciones, en las que en base a viñetas de texto se señala la interacción con nsp14 y/o nsp16. Finalmente se exportó en formato SVG para su montaje con las imágenes de estructuras moleculares en Adobe Illustrator.

#### **3.3.3. Construcción de logos de secuencias**

Un logo de secuencias es una representación gráfica que permite visualizar la conservación de una secuencia aminoacídica (o ácidos nucleicos) a partir de los datos de un alineamiento (Schneider y Stephens, 1990). Para ello, se separaron las secuencias en formato FASTA de cada familia de proteínas en base a los cuatro géneros de la familia *Orthocoronavirinae*. Posteriormente, se realizaron nuevos alineamientos con MAFFT para cada conjunto de secuencias, y se aislaron con Jalview de cada alineamiento las regiones de nsp14-ExoN, nsp16 y nsp10 previamente identificadas con ChimeraX. Los alineamientos de las regiones seleccionadas se exportaron a un nuevo archivo FASTA para la construcción de los logos de secuencias, a través de la herramienta en línea WebLogo. Los gráficos obtenidos fueron descargados en formato SVG para su montaje en una figura conjunta con Adobe Illustrator.

## 4. Resultados

### 4.1. Alineamientos de secuencias aminoacídicas de nsp10/14/16

En relación al [alineamiento de nsp10](#), la baja identidad y menor longitud relativa a la secuencia del virus del salmón del Pacífico (PsNV) redundó en la aparición de tres regiones huecas sobre sí misma, así como otras dos en el resto de secuencias. Tras el proceso de depuración, 128 de las 150 columnas (~85 %) fueron identificadas como posiciones informativas comparables.

Respecto al [alineamiento de nsp14](#), se observó que la secuencia del PsNV conserva bastantes posiciones homólogas. Sin embargo, también se evidenció una gran delección que abarca las primeras 83 posiciones del alineamiento (relativas al dominio nsp14-ExoN). Tras el proceso de depuración, 428 de las 567 (~75 %) se identificaron como columnas informativas comparables.

En cuanto al [alineamiento de nsp16](#), la secuencia del PsNV presentó una baja cobertura respecto a las de *Orthocoronavirinae*; queda flanqueada por dos grandes huecos que abarcan las regiones delimitadas por las posiciones 1–41 y 248–313. A pesar de ello, cabe destacar que la región abarcada por la secuencia está bastante conservada respecto a sus homólogas. Tras el proceso de depuración, 235 de las 313 (~85 %) se identificaron como columnas informativas comparables.

Las secuencias del PsNV, seleccionado inicialmente como especie externa, fueron descartadas para la reconstrucción de filogenias debido a la significativa ausencia de posiciones homólogas respecto a sus secuencias ortólogas de la subfamilia *Orthocoronavirinae*. No obstante, dentro de dicho grupo de estudio los alineamientos de nsp10, nsp14 y nsp16 muestran por lo general una elevada identidad de secuencia.

Los informes de depuración en HTML y los alineamientos previos en FASTA, con las secuencias del PsNV, están para su consulta en [https://github.com/villena-francis/bachelors\\_thesis/tree/main/data/align\\_ext\\_sp](https://github.com/villena-francis/bachelors_thesis/tree/main/data/align_ext_sp).

Los alineamientos realizados para la reconstrucción de los árboles filogenéticos, sin las secuencias del PsNV, están disponibles para ser consultados en .

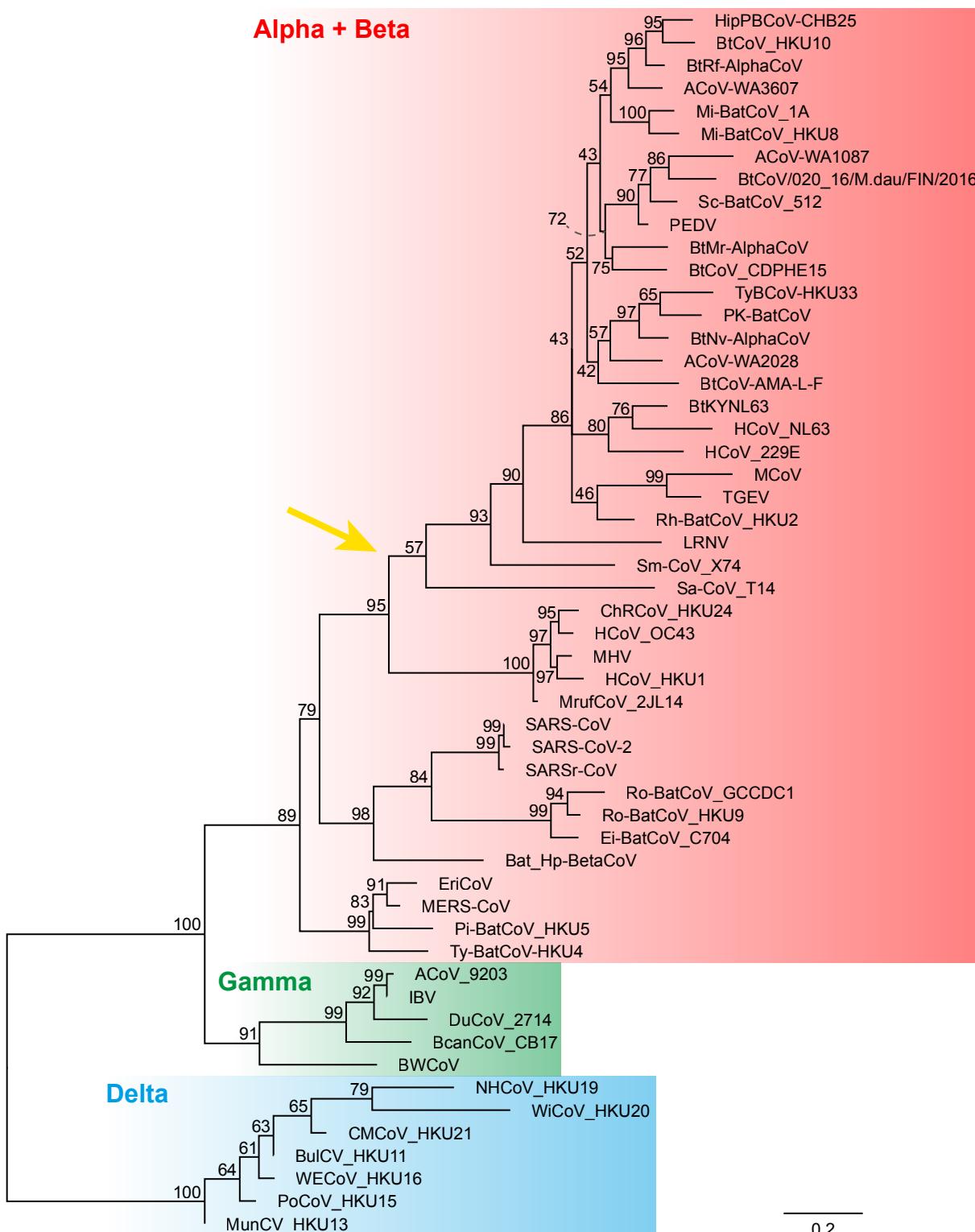
### 4.2. Filogenias de nsp10/14/16 en *Orthocoronavirinae*

Los modelos evolutivos seleccionados como más apropiados para la reconstrucción de las filogenias fueron Q.insect+I+G4, para nsp10 y nsp14, y Q.insect+G4 para nsp16 (Cuongbb y col., 2022). Los árboles filogenéticos obtenidos fueron enraizados manualmente seleccionando como grupo externo al género *Deltacoronavirus*, en base al conocimiento actual de las relaciones filogenéticas de los coronavirus dentro del orden *Nidovirales* (Zhou y col., 2021).

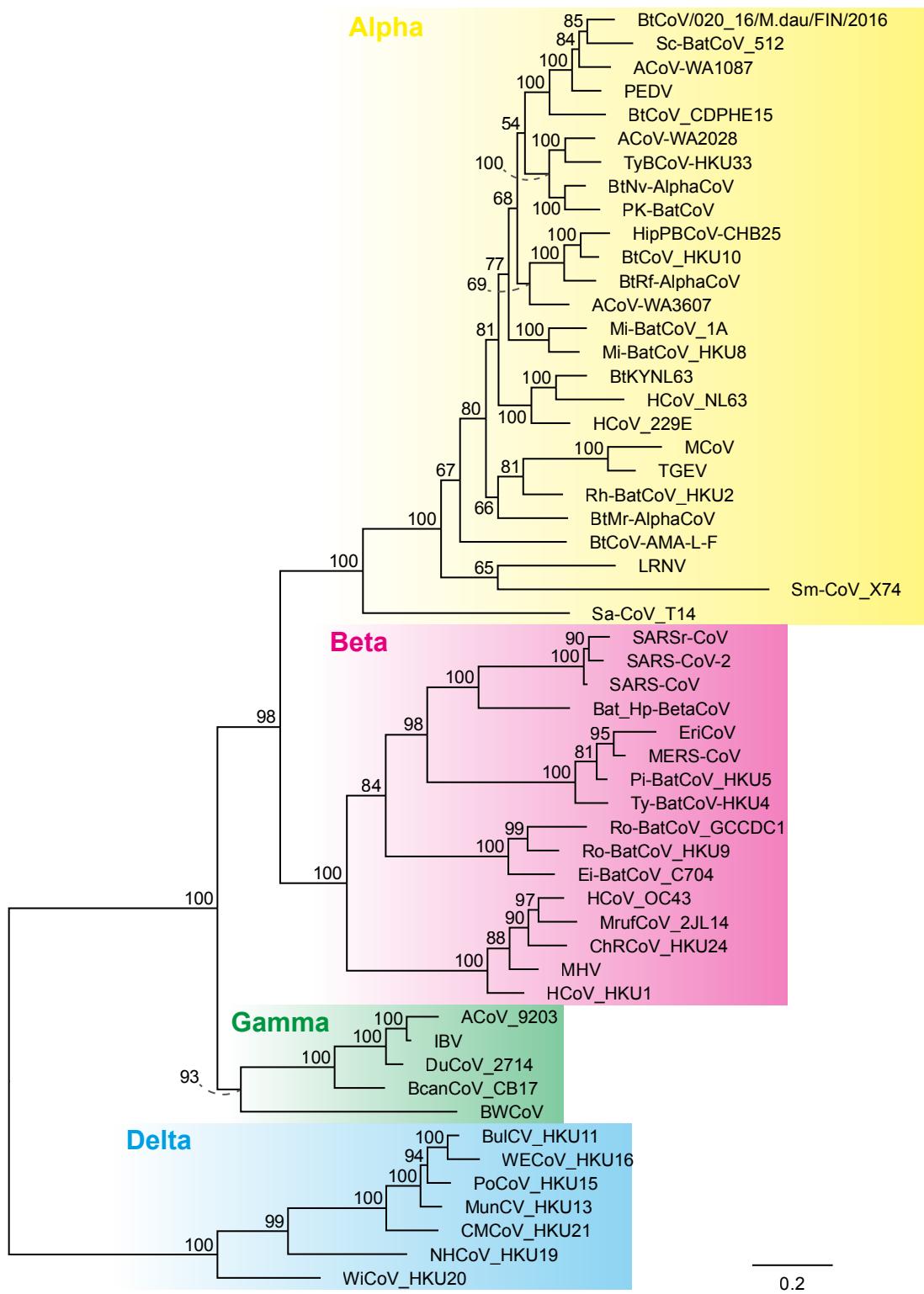
El árbol filogenético de la familia nsp10 (**Figura 5**) muestra tres grupos con nodos basales bien soportados según los valores de bootstrap. Dos de estos grupos se corresponden con los géneros *Gammacoronavirus* y *Deltacoronavirus*, mientras que en el tercero las secuencias pertenecientes al género Betacoronavirus forman un grupo parafilético con *Alphacoronavirus* como grupo monofilético más reciente.

Los árboles correspondientes a las familias nsp14 y nsp16 (**Figuras 6 y 7**) muestran

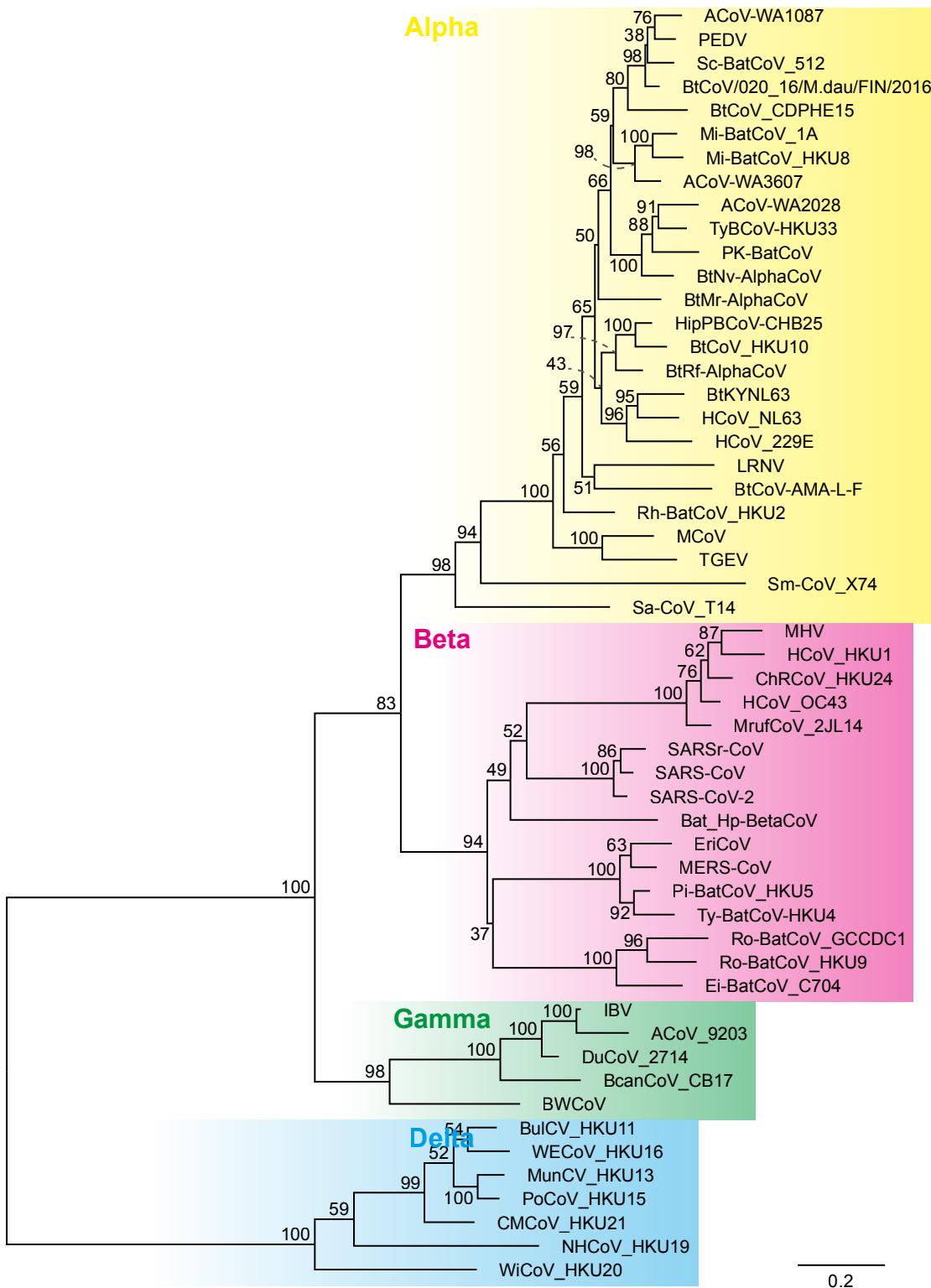
una notable similitud. En ambos casos, las secuencias se agrupan en cuatro clados correspondientes a cada uno de los géneros de la subfamilia *Orthocoronavirinae*. Los nodos basales que los separan muestran una considerable robustez, respaldados por valores de bootstrap significativos.



**Figura 5:** Árbol filogenético de máxima verosimilitud de las relaciones entre las proteínas nsp10 de *Orthocoronavirinae*. Se destacan tres grupos principales, correspondientes a los géneros *Alphacoronavirus + Betacoronavirus* (rojo), *Gammacoronavirus* (verde) y *Deltacoronavirus* (cian). La flecha amarilla señala la rama que aisla a todas las secuencias de *Alphacoronavirus* dentro de los *Betacoronavirus*, y los números sobre cada nodo corresponden a los valores de bootstrap. La barra de la esquina inferior derecha representa en escala el número de sustituciones de aminoácidos por sitio.



**Figura 6:** Árbol filogenético de máxima verosimilitud de las relaciones entre las proteínas nsp14 de *Orthocoronavirinae*. Se destacan cuatro grupos principales, correspondientes a los géneros *Alphacoronavirus* (amarillo), *Betacoronavirus* (magenta), *Gammacoronavirus* (verde) y *Deltacoronavirus* (cian). Los números sobre cada nodo corresponden a los valores de bootstrap y la barra de la esquina inferior derecha representa en escala el número de sustituciones de aminoácidos por sitio.



**Figura 7:** Árbol filogenético de máxima verosimilitud que muestra las relaciones entre las proteínas nsp16 de *Orthocoronavirinae*. Se destacan cuatro grupos principales, correspondientes a los géneros *Alphacoronavirus* (amarillo), *Betacoronavirus* (magenta), *Gammacoronavirus* (verde) y *Deltacoronavirus* (cian). Los números sobre cada nodo corresponden a los valores de bootstrap, y la barra de la esquina inferior derecha representa en escala el número de sustituciones de aminoácidos por sitio.

### **4.3. Región de nsp10 con máxima interacción con nsp14 y nsp16**

Utilizando la estructura cristalográfica de referencia para la interacción nsp10/nsp14-ExoN (PDB ID: 7DIY), se localizaron un total de 47 residuos de nsp10 que interactúan con nsp14-ExoN. Por otro lado, la estructura utilizada como base para el estudio de la interacción (PDB ID: 6W4H) reveló 22 residuos de nsp10 interactúan con nsp16. Se encontró que de los residuos identificados en nsp10, 19 fueron comunes a ambas interacciones.

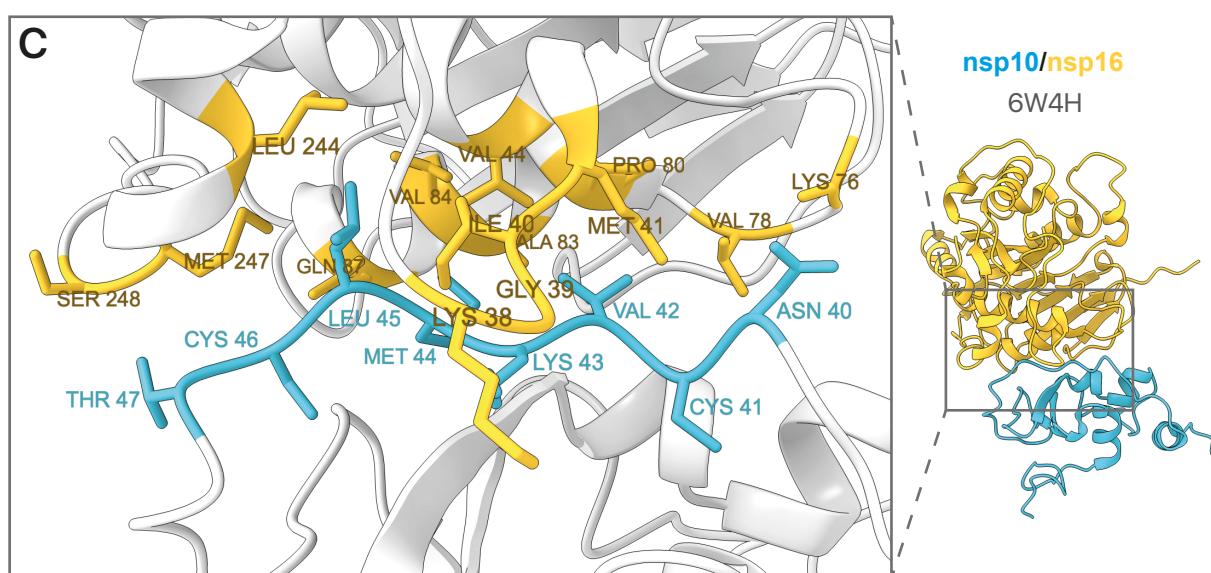
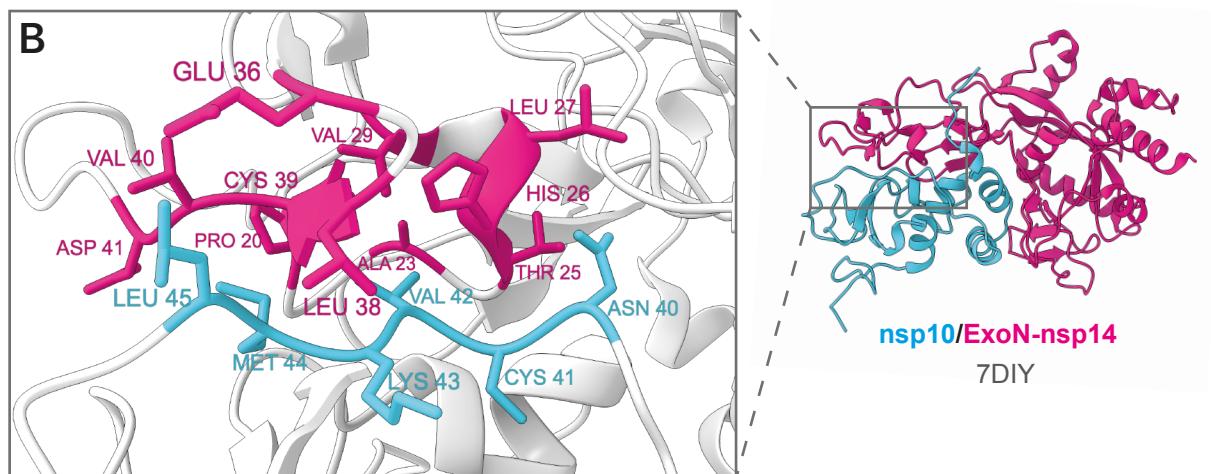
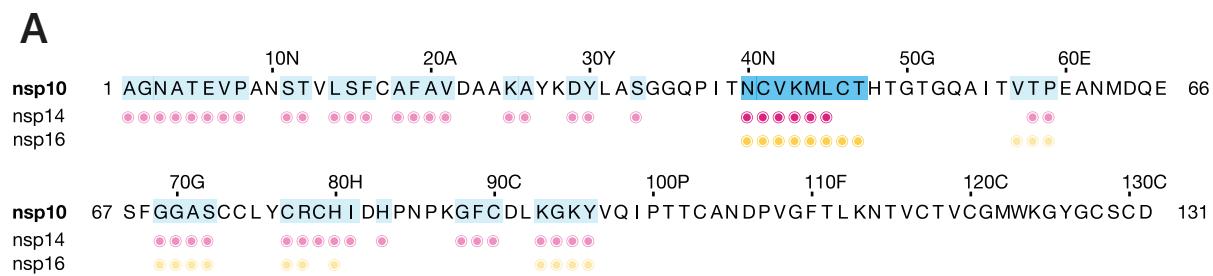
La región de nsp10 que muestra la mayor cantidad consecutiva de residuos involucrados en interacciones tanto con nsp14 como con nsp16 se encuentra en las posiciones 40–45, relativas a la secuencia aminoacídica del SARS-CoV-2. Sin embargo, con el objetivo de mejorar la interacción con nsp16 sin afectar de manera significativa la interacción con nsp14, se decidió extender la región de interés hasta la posición 47. Estos hallazgos se presentan de manera gráfica en la **Figura 8A**.

La región específica de nsp10 que ha sido seleccionada se trata de una cadena polipeptídica extendida, es decir, que no presenta conformación secundaria de hélice alfa ni lámina beta. Dicho domino establece interacciones no covalentes con 11 residuos de nsp14-ExoN (**Figura 8B**), y con otros 14 residuos de nsp16 (**Figura 8C**).

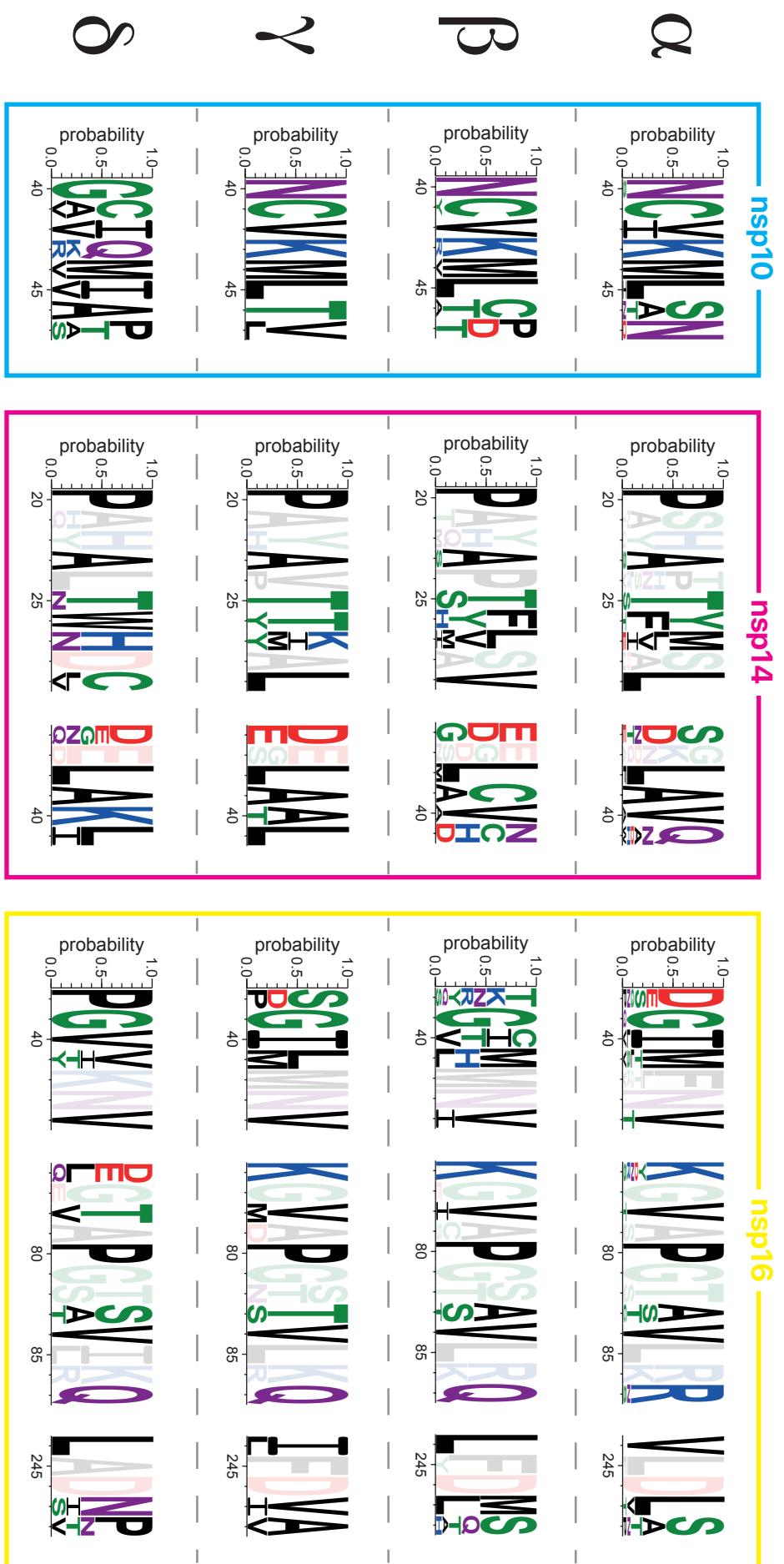
Las estructuras del PDB empleadas para este los análisis de este trabajo, con las configuraciones de visualización y anotaciones aplicadas, están disponibles para su consulta en formato de sesión de ChimeraX (.cxs) en el siguiente enlace: [https://github.com/villena-francis/bachelors\\_thesis/tree/main/data/structures](https://github.com/villena-francis/bachelors_thesis/tree/main/data/structures).

### **4.4. Logos de secuencias de la zona de máxima interacción de nsp10 y los dominios de nsp14 y nsp16 implicados.**

Como se muestra en la (**Figura 9**), las regiones seleccionadas de nsp10, nsp14-ExoN y nsp16 presentan de manera general una mayor conservación en el conjunto formado por *Alphacoronavirus*, *Betacoronavirus* y *Gammacoronavirus*, excluyendo a *Deltacoronavirus*. En esos tres géneros, los residuos de nsp10 involucrados tanto en la interacción con nsp14 como nsp16 (40–45) están muy conservados. Sin embargo, aquellos que interactúan únicamente con nsp16 (46–47) exhiben gran variabilidad en los cuatro géneros de *Orthocoronavirinae*.



**Figura 8:** Esquema de residuos de nsp10 (cian) involucrados en las interacciones con nsp14-ExoN (magenta) y nsp16 (amarillo), resaltando la región seleccionada como candidata en tonos más intensos (**A**). Los aminoácidos de nsp14-ExoN y nsp16 que interactúan con esta región de nsp10 se resaltan en las respectivas estructuras experimentales PDB ID: 7DIY (**B**) y PDB ID: 6W4H (**C**). La secuencia y estructuras utilizadas para esta figura corresponden al SARS-CoV-2.



**Figura 9:** Logos de secuencias aminoacídicas del dominio de nspl0 (cian) seleccionado en este trabajo y las regiones de nspl4 (magenta) y nspl6 (amarillo), relativos a los géneros *Alphacoronavirus*, ( $\alpha$ ), *Betacoronavirus* ( $\beta$ ), *Gammacoronavirus* ( $\gamma$ ) y *Deltacoronavirus* ( $\delta$ ). Los residuos se muestran en diferentes colores: hidrófobo = negro, polar = verde, básico = azul, ácido = rojo, neutro = morado. Las posiciones de los residuos indicadas en el eje X se corresponden a las ocupadas en las estructuras de referencia de nspl0/nspl4-ExoN (PDB ID: 7DIY) y nspl0/nspl6 (PDB ID: 6W4H) empleados en este trabajo. Las posiciones de nspl4 y nspl6 coloreadas débilmente no intervienen en la interacción directa con residuos de nspl0.

## 5. Discusión y futuras líneas

### 5.1. Análisis de los alineamientos de nsp10/14/16

En cuanto a las secuencias de nsp10/14/16 de las especies de *Orthocoronavirinae*, los alineamientos revelan en general gran número de posiciones altamente conservadas. Sin embargo, es importante tener en cuenta que esta observación debe ser interpretada en el contexto de la amplia variabilidad genética de los virus en comparación con los organismos celulares. A simple vista los patrones observados no son tan canónicos como los que se pudiesen encontrar, por ejemplo, en alineamientos de proteínas implicadas en el desarrollo de vertebrados (Domingo y col., 2021; Duffy, 2018; Duffy y Holmes, 2008; Holmes, 2009).

La secuencia de la región seleccionada de nsp10, posiciones 45–52 del [alineamiento de nsp10](#), presenta una muy baja identidad de secuencia en *Alphapironavirus bona* (*Pitovirinae*) respecto a todas sus ortólogas de *Orthocoronavirinae* con las que fue alineada. Además, para esta misma especie, en el [alineamiento de nsp14](#), la secuencia comienza a partir de la posición 42, lo cual evidencia la ausencia de la región amino terminal que para el resto de coronavirus contiene todos los dominios de interacción con la región de nsp10 seleccionada para el objetivo principal de este trabajo, localizados a partir de la estructura la estructura de referencia (**Figura 8**).

Las significativas disparidades observadas entre las secuencias de nsp10/14/16 de la especie perteneciente a la subfamilia *Pitovirinae* y las de coronavirus de aves y mamíferos (*Orthocoronavirinae*) contrastan con el conocimiento existente sobre los dominios de interacción para estos últimos (Baddock y col., 2022; Vithani y col., 2021). Se ha demostrado que mutaciones puntuales en residuos pertenecientes a la interfaz de interacción de nsp14 y nsp10 producen una disminución de la fidelidad de la replicación viral, con el consiguiente aumento de la tasa de error y la inestabilidad genómica (Bouvet y col., 2012; Chen y col., 2011; Ogando y col., 2019; Rosas-Lemus y col., 2020; Takada y col., 2023). A pesar de ello, las notables diferencias y huecos en las secuencias de nsp10 y nsp14 de *Alphapironavirus bona* no parecen haber impedido que el tamaño de su genoma alcance las 36,6 kilobases (Mordecai y col., 2019). Si estos hechos se mantuviesen para nuevas especies reconocidas de coronavirus externos a *Orthocoronavirinae*, sería indicio de notables diferencias entre los modelos de interacción nsp10-nsp14 de coronavirus de aves y mamíferos respecto a los de otros vertebrados.

### 5.2. Historia evolutiva de nsp10/14/16

Como se indica en resultados, las secuencias relativas al virus del salmón del Pacífico fueron descartadas para la reconstrucción de las filogenias. Las distancias evolutivas inferidas para sus secuencias de nsp10/14/16 vendrían dadas principalmente por la ausencia de posiciones comparables en esta candidata a especie externa. Además, esta falta de información aumentaría a su vez la probabilidad de posicionar a las secuencias del PsNV en otros nodos de las filogenias, reduciendo sus valores de bootstrap y con ello la fiabilidad de las ramificaciones generadas.

Las relaciones filogenéticas de las proteínas nsp14 y nsp16 de los coronavirus que infectan aves y mamíferos (**Figuras 6 y 7**) concuerdan con las expectativas establecidas por la taxonomía filogenética asignada a sus respectivas especies (Gulyaeva y Gorbalenya, 2021; Woo y col., 2023; Zhou y col., 2021). En ambos casos, se presume que las

secuencias de ambas proteínas se diferenciaron desde el ancestro común de la subfamilia *Orthocoronavirinae* en los linajes correspondientes a los cuatro géneros existentes.

La filogenia reconstruida utilizando las secuencias de nsp10 (**Figura 5**) difiere en cierta medida de las relativas a nsp14 y nsp16. La agrupación de las secuencias de *Alphacoronavirus* como un grupo monofilético dentro de *Betacoronavirus* puede atribuirse a un resultado artefactual. Esto puede ser causado por la diferencia notable en el número de residuos entre las proteínas nsp10/14/16, siendo el número de residuos en nsp10 aproximadamente cuatro y dos veces menor que los de nsp14 y nsp16, respectivamente. La reducción en el número de posiciones informativas y la excesiva conservación de las secuencias contribuyen a reducir de forma generalizada los valores de bootstrap de los nodos que sitúan a las secuencias de *Alphacoronavirus* dentro de *Betacoronavirus* en la filogenia obtenida para nsp10.

Estos hallazgos respaldarían la idoneidad de desarrollar un pseudoligando basado en este cofactor, ya que su conservación evolutiva le conferiría potencial para ser eficaz contra todos los coronavirus humanos actuales.

### 5.3. Región de nsp10 seleccionada

En base a los resultados obtenidos, la región de nsp10 con máxima interacción con nsp14-ExoN y nsp16 fue evidente; con un total de 6 residuos comunes en ambas interacciones. Además, se identificaron y añadieron otros 2 residuos contiguos adicionales que interactúan exclusivamente con nsp16 (**Figura 8A**). Con relación al diseño de un pseudoligando, esta selección no solo sería relevante para una mayor selectividad de unión a dicha proteína diana, sino que también podría ayudar a reducir la probabilidad de agregaciones inespecíficas con otros componentes celulares.

Como se puede apreciar en los logos de secuencias (**Figura 9**), los residuos de la zona de máxima interacción de nsp10 que interactúan tanto con nsp14-ExoN como con nsp16 (40–45) muestran una alta conservación en los géneros *Alphacoronavirus*, *Betacoronavirus* y *Gammacoronavirus*. En esta misma región, los *Deltacoronavirus* comparten con los géneros mencionados anteriormente solo tres posiciones con residuos hidrofóbicos (42, 44, 45). La preservación de estas tres posiciones en todo *Orthocoronavirinae* sugiere su papel esencial en la activación de las funciones de nsp14-ExoN y nsp16. En las estructuras presentadas como referencia (**Figura 8A** y **8B**) puede observarse que dichas posiciones hidrofóbicas están involucradas en la formación de agrupaciones con otros residuos similares en los dominios de interacción de estas enzimas, lo cual reduciría sus exposiciones al agua del citosol y haría estas uniones termodinámicamente favorables. Las interacciones del resto de posiciones de la región seleccionada (40, 41, 43, 46, 47) estarían mediadas por interacciones polares. Sin embargo, para validar consistentemente esta última afirmación se requerirían análisis adicionales.

En el caso hipotético de utilizar un pseudoligando basado en los mismos residuos de un betacoronavirus, como el SARS-CoV-2, podría esperarse una efectividad similar contra la nsp14-ExoN de otros beta- y alfa-coronavirus humanos debido a la alta identidad de secuencia que existe entre estos grupos. No obstante, y siguiendo con este ejemplo, su selectividad contra nsp16 sería más variable debido a las posiciones cambiantes del séptimo y octavo residuo del pseudoligando (47, 48), incluso dentro de los betacoronavirus.

Es importante mencionar que este estudio presenta limitaciones relacionadas princi-

palmente con la capacidad de computo del dispositivo doméstico utilizado. El objetivo principal de este trabajo es proporcionar indicios de conservación evolutiva que puedan ser utilizados como estrategia antiviral. En este sentido, se justifica la necesidad de seguir profundizando en diversos aspectos, comenzando por el análisis cuantitativo de las energías de interacción entre los residuos de la región seleccionada y las proteínas diana correspondientes. Mediante estos análisis, se podría avanzar en el diseño del pseudoligando al inferir los residuos más apropiados para las posiciones que en las zonas diana de nsp14-ExoN y nsp16 deberían establecer interacciones polares, asegurando su eficacia frente a un amplio espectro de coronavirus.



## 6. Conclusiones

Los alineamientos de secuencias y las regiones seleccionadas según modelos cristalográficos de las proteínas correspondientes proporcionan una visión acerca de la conservación de una posible diana contra la replicación de coronavirus humanos. En base a los resultados de este trabajo, podemos concluir lo siguiente:

1. Las nsp10/14/16 de la subfamilia *Orthocoronavirinae* están bastante conservadas, especialmente las relativas a los géneros que reúnen a casi todos los coronavirus capaces de infectar a humanos: *Alphacoronavirus* y *Betacoronavirus*.
2. Según las estructuras experimentales de SARS-CoV-2 tomadas como referencia, la región de nsp10 con máxima interacción con nsp14-ExoN y nsp16 está constituida por 6 aminoácidos que configuran una cadena extendida sin estructura secundaria.
3. Las posiciones de la región de nsp10 citada en el apartado anterior están bastante conservadas dentro del conjunto de los géneros *Alphacoronavirus*, *Betacoronavirus* y *Gammacoronavirus*. Sin embargo, *Deltacoronavirus* solo preserva 3 posiciones hidrofóbicas comunes para toda la subfamilia *Orthocoronavirinae*, concentrándose la mayor variabilidad de esta región en las posiciones hidrofílicas.

Por todo ello, los análisis *in silico* de este trabajo apoyan la idoneidad de la región de nsp10 con máxima interacción con nsp14-ExoN y nsp16 para su uso como molde en el diseño de posibles pseudoligandos con actividad antiviral pancoronavírica.



## Referencias

- Asensio, V. (2020a). Estructura y genoma del coronavirus SARS-CoV-2. Consultado el 24 de mayo de 2023, desde [https://commons.wikimedia.org/wiki/File:Estructura\\_y\\_genoma\\_del\\_coronavirus\\_SARS-CoV-2.svg](https://commons.wikimedia.org/wiki/File:Estructura_y_genoma_del_coronavirus_SARS-CoV-2.svg?uselang=es#/media/File:Estructura_y_genoma_del_coronavirus_SARS-CoV-2.svg)
- Asensio, V. (2020b). Ciclo de multiplicación celular del SARS-CoV-2. Consultado el 28 de mayo de 2023, desde [https://commons.wikimedia.org/wiki/File:SARS-CoV-2\\_ciclo.png](https://commons.wikimedia.org/wiki/File:SARS-CoV-2_ciclo.png)
- Baddock, H. T., Brolih, S., Yosaatmadja, Y., Ratnaweera, M., Bielinski, M., Swift, L. P., Cruz-Migoni, A., Fan, H., Keown, J. R., Walker, A. P., Morris, G. M., Grimes, J. M., Fodor, E., Schofield, C. J., Gileadi, O., & McHugh, P. J. (2022). Characterization of the SARS-CoV-2 ExoN (nsp14ExoN–nsp10) complex: implications for its role in viral genome stability and inhibitor identification. *Nucleic Acids Research*, 50(3), 1484-1500. <https://doi.org/10.1093/nar/gkab1303>
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, 41(Database issue), D36-D42. <https://doi.org/10.1093/nar/gks1195>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235-242. <https://doi.org/10.1093/nar/28.1.235>
- Bosch, B. J., van der Zee, R., de Haan, C. A. M., & Rottier, P. J. M. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *Journal of Virology*, 77(16), 8801-8811. <https://doi.org/10.1128/jvi.77.16.8801-8811.2003>
- Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B., & Decroly, E. (2012). RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, 109(24), 9372-9377. <https://doi.org/10.1073/pnas.1201130109>
- Bukhari, K., Mulley, G., Gulyaeva, A. A., Zhao, L., Shu, G., Jiang, J., & Neuman, B. W. (2018). Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abyssoviridae, and from a sister group to the Coronavirinae, the proposed genus Alphaletovirus. *Virology*, 524, 160-171. <https://doi.org/10.1016/j.virol.2018.08.010>
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972-1973. <https://doi.org/10.1093/bioinformatics/btp348>

Chen, Y., Cai, H., Pan, J., Xiang, N., Tien, P., Ahola, T., & Guo, D. (2009). Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, 106(9), 3484-3489. <https://doi.org/10.1073/pnas.0808790106>

Chen, Y., Su, C., Ke, M., Jin, X., Xu, L., Zhang, Z., Wu, A., Sun, Y., Yang, Z., Tien, P., Ahola, T., Liang, Y., Liu, X., & Guo, D. (2011). Biochemical and Structural Insights into the Mechanisms of SARS Coronavirus RNA Ribose 2'-O-Methylation by nsp16/nsp10 Protein Complex [Publisher: Public Library of Science]. *PLOS Pathogens*, 7(10), e1002294. <https://doi.org/10.1371/journal.ppat.1002294>

Claverie, J.-M., & Abergel, C. (2016). Giant viruses: The difficult breaking of multiple epistemological barriers. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 59, 89-99. <https://doi.org/10.1016/j.shpsc.2016.02.015>

Clustal Colour Scheme. (s.f.). Consultado el 23 de abril de 2023, desde <https://www.jalview.org/help/html/colourSchemes/clustal.html>

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator: Figure 1. *Genome Research*, 14(6), 1188-1190. <https://doi.org/10.1101/gr.849004>

Cuongbb, Schmidt, H., Trifinopoulos, J., Bui, M., & Uit, T. (2022). All common substitution models and usages. <http://www.iqtree.org/doc/Substitution-Models>

Díaz-Martínez, L., Brichette-Mieg, I., Pineño-Ramos, A., Domínguez-Huerta, G., & Grande-Pérez, A. (2018). Lethal mutagenesis of an RNA plant virus via lethal defection [Number: 1 Publisher: Nature Publishing Group]. *Scientific Reports*, 8(1), 1444. <https://doi.org/10.1038/s41598-018-19829-6>

Domingo, E., García-Crespo, C., Lobo-Vega, R., & Perales, C. (2021). Mutation Rates, Mutation Frequencies, and Proofreading-Repair Activities in RNA Virus Genetics. *Viruses*, 13(9), 1882. <https://doi.org/10.3390/v13091882>

Duffy, S. (2018). Why are RNA virus mutation rates so damn high? *PLOS Biology*, 16(8), e3000003. <https://doi.org/10.1371/journal.pbio.3000003>

Duffy, S., & Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. *Journal of Virology*, 82(2), 957-965. <https://doi.org/10.1128/JVI.01929-07>

Enjuanes, L., Almazán, F., Sola, I., & Zuñiga, S. (2006). Biochemical aspects of coronavirus replication and virus-host interaction. *Annual Review of Microbiology*, 60, 211-230. <https://doi.org/10.1146/annurev.micro.60.080805.142157>

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 [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, 115(2), E162-E171. <https://doi.org/10.1073/pnas.1718806115>

Galloway, A., & Cowling, V. H. (2019). mRNA cap regulation in mammalian cell function and fate. *Biochimica et Biophysica Acta. Gene Regulatory Mechanisms*, 1862(3), 270-279. <https://doi.org/10.1016/j.bbagr.2018.09.011>

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I., Ziebuhr, J., & Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2 [Number: 4 Publisher: Nature Publishing Group]. *Nature Microbiology*, 5(4), 536-544. <https://doi.org/10.1038/s41564-020-0695-z>

Grande-Pérez, A. (2021). Nueva terapia antiviral de combinación frente al coronavirus para el desarrollo de futuros tratamientos. Consultado el 25 de mayo de 2023, desde [https://www.uma.es/sala-de-prensa/noticias/nueva-terapia-antiviral-de-combinacion-frente-al-coronavirus-para-el-desarrollo-de-futuros-tratamientos/?set\\_language=es](https://www.uma.es/sala-de-prensa/noticias/nueva-terapia-antiviral-de-combinacion-frente-al-coronavirus-para-el-desarrollo-de-futuros-tratamientos/?set_language=es)

Green, R., & Zardecki, C. (2019). PDB101: Learn: Guide to Understanding PDB Data: Dealing with Coordinates. Consultado el 18 de abril de 2023, desde <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/dealing-with-coordinates>

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3), 307-321. <https://doi.org/10.1093/sysbio/syq010>

Gulyaeva, A. A., & Gorbalenya, A. E. (2021). A nidovirus perspective on SARS-CoV-2. *Biochemical and Biophysical Research Communications*, 538, 24-34. <https://doi.org/10.1016/j.bbrc.2020.11.015>

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, 35(2), 518-522. <https://doi.org/10.1093/molbev/msx281>

Holmes, E. C. (2009). The Evolutionary Genetics of Emerging Viruses. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 353-372. <https://doi.org/10.1146/annurev.ecolsys.110308.120248>

Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, 19(3), 141-154. <https://doi.org/10.1038/s41579-020-00459-7>

Huggins, J. W., Hsiang, C. M., Cosgriff, T. M., Guang, M. Y., Smith, J. I., Wu, Z. O., LeDuc, J. W., Zheng, Z. M., Meegan, J. M., & Wang, Q. N. (1991). Prospective, double-blind, concurrent, placebo-controlled clinical trial of intravenous ribavirin therapy of hemorrhagic fever with renal syndrome. *The Journal of Infectious Diseases*, 164(6), 1119-1127. <https://doi.org/10.1093/infdis/164.6.1119>

ICZN. (s.f.). About the ICZN | International Commission on Zoological Nomenclature. Consultado el 20 de abril de 2023, desde <https://www.iczn.org/about-the-iczn/>

Inc., A. (2023). Adobe Illustrator. Consultado el 11 de marzo de 2023, desde <https://www.adobe.com/es/products/illustrator.html>

Jimenez-Guardeño, J. M., Nieto-Torres, J. L., DeDiego, M. L., Regla-Nava, J. A., Fernandez-Delgado, R., Castaño-Rodriguez, C., & Enjuanes, L. (2014). The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis. *PLoS pathogens*, 10(8), e1004320. <https://doi.org/10.1371/journal.ppat.1004320>

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates [Number: 6 Publisher: Nature Publishing Group]. *Nature Methods*, 14(6), 587-589. <https://doi.org/10.1038/nmeth.4285>

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4), 772-780. <https://doi.org/10.1093/molbev/mst010>

Keesey, T. M. (2023). PhyloPic. Consultado el 16 de abril de 2023, desde <https://www.phylopic.org>

Koonin, E. V., Dolja, V. V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*, 479-480, 2-25. <https://doi.org/10.1016/j.virol.2015.02.039>

Koonin, E. V., Dolja, V. V., Krupovic, M., & Kuhn, J. H. (2021). Viruses Defined by the Position of the Virosphere within the Replicator Space. *Microbiology and Molecular Biology Reviews*, 85(4), e00193-20. <https://doi.org/10.1128/MMBR.00193-20>

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., Zerbini, F. M., & Kuhn, J. H. (2020). Global Organization and Proposed Megataxonomy of the Virus World. *Microbiology and Molecular Biology Reviews*, 84(2), e00061-19. <https://doi.org/10.1128/MMBR.00061-19>

Lednicky, J. A., Tagliamonte, M. S., White, S. K., Elbadry, M. A., Alam, M. M., Stephen-  
son, C. J., Bonny, T. S., Loeb, J. C., Telisma, T., Chavannes, S., Ostrov, D. A., Ma-  
vian, C., Beau De Rochars, V. M., Salemi, M., & Morris, J. G. (2021). Independent  
infections of porcine deltacoronavirus among Haitian children [Publisher: Nature  
Publishing Group]. *Nature*, 600(7887), 133-137. <https://doi.org/10.1038/s41586-021-04111-z>

Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., &  
Smith, D. B. (2018). Virus taxonomy: the database of the International Committee  
on Taxonomy of Viruses (ICTV). *Nucleic Acids Research*, 46(D1), D708-D717.  
<https://doi.org/10.1093/nar/gkx932>

Lin, S., Chen, H., Chen, Z., Yang, F., Ye, F., Zheng, Y., Yang, J., Lin, X., Sun, H.,  
Wang, L., Wen, A., Dong, H., Xiao, Q., Deng, D., Cao, Y., & Lu, G. (2021).  
Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals  
an exoribonuclease with both structural and functional integrity. *Nucleic Acids  
Research*, 49(9), 5382-5392. <https://doi.org/10.1093/nar/gkab320>

Liu, D. X., Liang, J. Q., & Fung, T. S. (2021). Human Coronavirus-229E, -OC43, -NL63,  
and -HKU1 (Coronaviridae). *Encyclopedia of Virology*, 428-440. <https://doi.org/10.1016/B978-0-12-809633-8.21501-X>

Luria, S., Darnell, J. E., Baltimore, D., & Campbell, A. (1978). *General virology* (3rd ed).  
New York : Wiley. Consultado el 14 de abril de 2023, desde <http://archive.org/details/generalvirology00luri>

Miller, A. K., Mifsud, J. C. O., Costa, V. A., Grimwood, R. M., Kitson, J., Baker, C.,  
Brosnahan, C. L., Pande, A., Holmes, E. C., Gemmell, N. J., & Geoghegan, J. L.  
(2021). Slippery when wet: cross-species transmission of divergent coronaviruses  
in bony and jawless fish and the evolutionary history of the Coronaviridae. *Virus  
Evolution*, 7(2), veab050. <https://doi.org/10.1093/ve/veab050>

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von  
Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods  
for Phylogenetic Inference in the Genomic Era (E. Teeling, Ed.). *Molecular Biology  
and Evolution*, 37(5), 1530-1534. <https://doi.org/10.1093/molbev/msaa015>

Minskaia, E., Hertzig, T., Gorbalyena, A. E., Campanacci, V., Cambillau, C., Canard,  
B., & Ziebuhr, J. (2006). Discovery of an RNA virus 3 →5 exoribonuclease that  
is critically involved in coronavirus RNA synthesis [Publisher: Proceedings of the  
National Academy of Sciences]. *Proceedings of the National Academy of Sciences*,  
103(13), 5108-5113. <https://doi.org/10.1073/pnas.0508200103>

Mordecai, G. J., Miller, K. M., Di Cicco, E., Schulze, A. D., Kaukinen, K. H., Ming, T. J.,  
Li, S., Tabata, A., Teffer, A., Patterson, D. A., Ferguson, H. W., & Suttle, C. A.  
(2019). Endangered wild salmon infected by newly discovered viruses. *eLife*, 8,  
e47615. <https://doi.org/10.7554/eLife.47615>

Mulangu, S., Dodd, L. E., Davey, R. T., Tshiani Mbaya, O., Proschan, M., Mukadi, D., Lusakibanza Manzo, M., Nzolo, D., Tshomba Oloma, A., Ibanda, A., Ali, R., Coulibaly, S., Levine, A. C., Grais, R., Diaz, J., Lane, H. C., Muyembe-Tamfum, J.-J., PALM Writing Group, Sivahera, B., ... PALM Consortium Study Team. (2019). A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics. *The New England Journal of Medicine*, 381(24), 2293-2303. <https://doi.org/10.1056/NEJMoa1910993>

Nasir, A., Romero-Severson, E., & Claverie, J.-M. (2020). Investigating the Concept and Origin of Viruses. *Trends in Microbiology*, 28(12), 959-967. <https://doi.org/10.1016/j.tim.2020.08.003>

Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., Droege, B., Klaus, J. P., Makino, S., Sawicki, S. G., Siddell, S. G., Stamou, D. G., Wilson, I. A., Kuhn, P., & Buchmeier, M. J. (2011). A structural analysis of M protein in coronavirus assembly and morphology. *Journal of Structural Biology*, 174(1), 11-22. <https://doi.org/10.1016/j.jsb.2010.11.021>

Ogando, N. S., Ferron, F., Decroly, E., Canard, B., Posthuma, C. C., & Snijder, E. J. (2019). The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity. *Frontiers in Microbiology*, 10, 1813. <https://doi.org/10.3389/fmicb.2019.01813>

Oren, A., Arahal, D. R., Göker, M., Moore, E. R., Rossello-Mora, R., & Sutcliffe, I. C. (2023). *International Code of Nomenclature of Prokaryotes (2022 Revision) Preprint* (inf. téc.). Zenodo. <https://doi.org/10.5281/zenodo.7770135>

Note that the typeset version will improve the formatting and presentation of the published version, notably in the Tables. Page numbers will differ in the typeset version of the ICNP.

Ortega-Prieto, A. M., Sheldon, J., Grande-Pérez, A., Tejero, H., Gregori, J., Quer, J., Esteban, J. I., Domingo, E., & Perales, C. (2013). Extinction of hepatitis C virus by ribavirin in hepatoma cells involves lethal mutagenesis. *PloS One*, 8(8), e71039. <https://doi.org/10.1371/journal.pone.0071039>

Pan, R., Kindler, E., Cao, L., Zhou, Y., Zhang, Z., Liu, Q., Ebert, N., Züst, R., Sun, Y., Gorbatenya, A. E., Perlman, S., Thiel, V., Chen, Y., & Guo, D. (2022). N7-Methylation of the Coronavirus RNA Cap Is Required for Maximal Virulence by Preventing Innate Immune Recognition [Publisher: American Society for Microbiology]. *mBio*, 13(1), e03662-21. <https://doi.org/10.1128/mbio.03662-21>

Peck, K. M., & Lauring, A. S. (2018). Complexities of Viral Mutation Rates [Publisher: American Society for Microbiology]. *Journal of Virology*, 92(14). <https://doi.org/10.1128/JVI.01031-17>

Peiris, M., & Poon, L. L. M. (2021). Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (Coronaviridae). En D. H. Bamford

& M. Zuckerman (Eds.), *Encyclopedia of Virology (Fourth Edition)* (pp. 814-824). Academic Press. <https://doi.org/10.1016/B978-0-12-814515-9.00138-7>

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1), 70-82. <https://doi.org/10.1002/pro.3943>

Rambaut, A. (2007). FigTree. Consultado el 21 de abril de 2023, desde <http://tree.bio.ed.ac.uk/software/Figtree/>

Rona, G., Zeke, A., Miwatani-Minter, B., de Vries, M., Kaur, R., Schinlever, A., Garcia, S. F., Goldberg, H. V., Wang, H., Hinds, T. R., Bailly, F., Zheng, N., Cottelle, P., Desmaële, D., Landau, N. R., Dittmann, M., & Pagano, M. (2022). The NSP14/NSP10 RNA repair complex as a Pan-coronavirus therapeutic target. *Cell Death and Differentiation*, 29(2), 285-292. <https://doi.org/10.1038/s41418-021-00900-1>

Rosas-Lemus, M., Minasov, G., Shuvalova, L., Inniss, N. L., Kiryukhina, O., Wiersum, G., Kim, Y., Jedrzejczak, R., Maltseva, N. I., Endres, M., Jaroszewski, L., Godzik, A., Joachimiak, A., & Satchell, K. J. F. (2020). The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with Sadenosylmethionine. *bioRxiv*, 2020.04.17.047498. <https://doi.org/10.1101/2020.04.17.047498>

Rose, Y., Duarte, J. M., Lowe, R., Segura, J., Bi, C., Bhikadiya, C., Chen, L., Rose, A. S., Bittrich, S., Burley, S. K., & Westbrook, J. D. (2021). RCSB Protein Data Bank: Architectural Advances Towards Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive. *Journal of Molecular Biology*, 433(11), 166704. <https://doi.org/10.1016/j.jmb.2020.11.003>

Ruiz-Aravena, M., McKee, C., Gamble, A., Lunn, T., Morris, A., Snedden, C. E., Yinda, C. K., Port, J. R., Buchholz, D. W., Yeo, Y. Y., Faust, C., Jax, E., Dee, L., Jones, D. N., Kessler, M. K., Falvo, C., Crowley, D., Bharti, N., Brook, C. E., ... Plowright, R. K. (2022). Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews. Microbiology*, 20(5), 299-314. <https://doi.org/10.1038/s41579-021-00652-2>

Rybicki, E. P. (2023). Chapter 1 - What are viruses, and how were they discovered? En E. P. Rybicki (Ed.), *Cann's Principles of Molecular Virology (Seventh Edition)* (pp. 1-53). Academic Press. <https://doi.org/10.1016/B978-0-12-822784-8.00001-5>

Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M., & Belshaw, R. (2010). Viral Mutation Rates. *Journal of Virology*, 84(19), 9733-9748. <https://doi.org/10.1128/JVI.00694-10>

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Farrell, C. M., Feldgarden, M., Fine, A. M., Funk, K., Hatcher, E., Kannan, S., Kelly, C., Kim, S., Klimke, W., Landrum, M. J., Lathrop, S., Lu, Z., Madden, T. L., ...

Sherry, S. T. (2023). Database resources of the National Center for Biotechnology Information in 2023. *Nucleic Acids Research*, 51(D1), D29-D38. <https://doi.org/10.1093/nar/gkac1032>

Schneider, T. D., & Stephens, R. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*, 18(20), 6097-6100. <https://doi.org/10.1093/nar/18.20.6097>

Shi, M., Lin, X.-D., Chen, X., Tian, J.-H., Chen, L.-J., Li, K., Wang, W., Eden, J.-S., Shen, J.-J., Liu, L., Holmes, E. C., & Zhang, Y.-Z. (2018). The evolutionary history of vertebrate RNA viruses [Number: 7700 Publisher: Nature Publishing Group]. *Nature*, 556(7700), 197-202. <https://doi.org/10.1038/s41586-018-0012-7>

Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., Davison, A. J., Delwart, E., Gorbunova, A. E., Harrach, B., Hull, R., King, A. M., Koonin, E. V., Krupovic, M., Kuhn, J. H., Lefkowitz, E. J., Nibert, M. L., Orton, R., Roossinck, M. J., ... Zerbini, F. M. (2017). Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*, 15(3), 161-168. <https://doi.org/10.1038/nrmicro.2016.177>

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L. M., Guan, Y., Rozanov, M., Spaan, W. J. M., & Gorbunova, A. E. (2003). Unique and Conserved Features of Genome and Proteome of SARS-coronavirus, an Early Split-off From the Coronavirus Group 2 Lineage. *Journal of Molecular Biology*, 331(5), 991-1004. [https://doi.org/10.1016/S0022-2836\(03\)00865-9](https://doi.org/10.1016/S0022-2836(03)00865-9)

Stevaert, A., Groaz, E., & Naesens, L. (2022). Nucleoside analogs for management of respiratory virus infections: mechanism of action and clinical efficacy. *Current Opinion in Virology*, 57, 101279. <https://doi.org/10.1016/j.coviro.2022.101279>

Swanstrom, R., & Schinazi, R. F. (2022). Lethal mutagenesis as an antiviral strategy. *Science*, 375(6580), 497-498. <https://doi.org/10.1126/science.abn0048>

Takada, K., Ueda, M. T., Shichinohe, S., Kida, Y., Ono, C., Matsuura, Y., Watanabe, T., & Nakagawa, S. (2023). Genomic diversity of SARS-CoV-2 can be accelerated by mutations in the nsp14 gene. *iScience*, 26(3), 106210. <https://doi.org/10.1016/j.isci.2023.106210>

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL\_X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. *Nucleic Acids Research*, 25(24), 4876-4882. <https://doi.org/10.1093/nar/25.24.4876>

Vithani, N., Ward, M. D., Zimmerman, M. I., Novak, B., Borowsky, J. H., Singh, S., & Bowman, G. R. (2021). SARS-CoV-2 Nsp16 activation mechanism and a cryptic pocket with pan-coronavirus antiviral potential. *Biophysical Journal*, 120(14), 2880-2889. <https://doi.org/10.1016/j.bpj.2021.03.024>

Vlasova, A. N., Diaz, A., Damtie, D., Xiu, L., Toh, T.-H., Lee, J. S.-Y., Saif, L. J., & Gray, G. C. (2022). Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia. *Clinical Infectious Diseases*, 74(3), 446-454. <https://doi.org/10.1093/cid/ciab456>

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>

Woo, P. C., J. de Groot, R., Haagmans, B., K.P. Lau, S., W. Neuman, B., Perlman, S., Sola, I., van der Hoek, L., C.P. Wong, A., & Yeh, S.-H. (2023). Family: Coronaviridae | ICTV. Consultado el 20 de abril de 2023, desde <https://ictv.global/report/chapter/coronaviridae/coronaviridae>

Ye, Y., & Hogue, B. G. (2007). Role of the coronavirus E viroporin protein transmembrane domain in virus assembly. *Journal of Virology*, 81(7), 3597-3607. <https://doi.org/10.1128/JVI.01472-06>

Zerbini, F. M., Siddell, S. G., Mushegian, A. R., Walker, P. J., Lefkowitz, E. J., Adriaenssens, E. M., Alfenas-Zerbini, P., Dutilh, B. E., García, M. L., Junglen, S., Krupovic, M., Kuhn, J. H., Lambert, A. J., Łobocka, M., Oksanen, H. M., Robertson, D. L., Rubino, L., Sabanadzovic, S., Simmonds, P., ... Varsani, A. (2022). Differentiating between viruses and virus species by writing their names correctly. *Archives of Virology*, 167(4), 1231-1234. <https://doi.org/10.1007/s00705-021-05323-4>

Zhou, Z., Qiu, Y., & Ge, X. (2021). The taxonomy, host range and pathogenicity of coronaviruses and other viruses in the Nidovirales order. *Animal Diseases*, 1(1), 5. <https://doi.org/10.1186/s44149-021-00005-9>



## Anexos

**Tabla A1:** Genomas de referencia de los que se obtuvieron las secuencias de nsp10/14/16.

Especie	Nombre común	Abreviatura	ID GenBank
<b>Género <i>Alphacoronavirus</i></b>			
<i>Alphacoronavirus AMALF</i>	bat alphacoronavirus isolate AMALF	BtCoV-AMA-L-F	MT663548
<i>Bat coronavirus CDPHE15</i>	bat coronavirus CDPHE15	BtCoV CDPHE15	KF430219
<i>Alphacoronavirus CHB25</i>	Hipposideros pomona bat coronavirus CHB25	HipPBCoV-CHB25	MN611525
<i>Alphacoronavirus WA3607</i>	alphacoronavirus sp. WA3607	ACoV-WA3607	MK472070
<i>Bat coronavirus HKU10</i>	Rousettus bat coronavirus HKU10	BtCoV HKU10	JQ989270
<i>Rhinolophus ferrumequinum alphacoronavirus HuB-2013</i>	BtRf-AlphaCoV/HuB2013	BtRf-AlphaCoV	KJ473807
<i>Human coronavirus 229E</i>	human coronavirus 229E	<i>HCoV_229E</i>	AF304460
<i>Human coronavirus 229E</i>	human coronavirus 229E	<i>HCeV_229E</i>	JX503061
<i>Lucheng Rn rat coronavirus</i>	Lucheng Rn rat coronavirus	LRNV	KF294380
<i>Mink coronavirus 1</i>	mink coronavirus	MCoV	HM245925
<i>Miniopterus bat coronavirus 1</i>	Miniopterus bat coronavirus 1	<i>Mi – BatCoV_1A</i>	EU420138
<i>Miniopterus bat coronavirus HKU8</i>	Miniopterus bat coronavirus	<i>HKU8Mi – BatCoV_HKU8</i>	EU420139
<i>Myotis ricketti alphacoronavirus Sax-2011</i>	BtMr-AlphaCoV/SAX2011	BtMr-AlphaCoV	KJ473806
<i>Alphacoronavirus HKU33</i>	Tylonycteris bat coronavirus HKU33	TyBCoV-HKU33	MK720944
<i>Alphacoronavirus WA2028</i>	alphacoronavirus sp. WA2028	ACoV-WA2028	MK472068
<i>Nyctalus velutinus alphacoronavirus SC-2013</i>	BtNv-AlphaCoV/SC2013	BtNv-AlphaCoV	KJ473809
<i>Pipistrellus kuhlii coronavirus 3398</i>	alphacoronavirus bat coronavirus	PK-BatCoV	MH938449
<i>Alphacoronavirus BT020</i>	bat alphacoronavirus <i>BtCoV/020_16/M.dau/FIN/2016</i>	<i>BtCoV/020_16/M.dau/FIN/2016</i>	MG923574
<i>Alphacoronavirus WA1087</i>	alphacoronavirus sp. WA1087	ACoV-WA1087	MK472067
<i>Porcine epidemic diarrhea virus</i>	porcine epidemic diarrhea virus	PEDV	AF353511
<i>Scotophilus bat coronavirus 512</i>	Scotophilus bat coronavirus 512	<i>Sc – BatCoV_512</i>	DQ648858
<i>Rhinolophus bat coronavirus HKU2</i>	Rhinolophus bat coronavirus HKU2	<i>Rh – BatCoV_HKU2</i>	EF203064
<i>Human coronavirus NL63</i>	human coronavirus NL63	<i>HCeV_NL63</i>	AY567487
<i>NL63-related bat coronavirus strain BtKYNL63-9b</i>	NL63-related bat coronavirus	BtKYNL63	KY073745
<i>Sorex araneus coronavirus T14</i>	common shrew coronavirus Tibet-2014	<i>Sa – CoV_T14</i>	KY370053
<i>Suncus murinus coronavirus X74</i>	Suncus murinus coronavirus Xingguo-74	<i>Sm – CoV_X74</i>	KY967715
<i>Alphacoronavirus 1</i>	transmissible gastroenteritis virus	TGEV	AJ271965

---

**Género Betacoronavirus**

---

<i>Betacoronavirus 1</i>	human coronavirus OC43	<i>HCoV_OC43</i>	AY585228
<i>China Rattus coronavirus HKU24</i>	betacoronavirus HKU24	<i>ChRCoV_HKU24</i>	KM349742
<i>Human coronavirus HKU1</i>	human coronavirus HKU1	<i>HCoV_HKU1</i>	AY597011
<i>Murine coronavirus</i>	murine hepatitis virus	MHV	AY700211
<i>Myodes coronaviruses 2JL14</i>	Myodes rufocanus vole coronavirus 2/JL2014	<i>MrufCoV_2JL14</i>	KY370046
<i>Murine coronavirus</i>	murine hepatitis virus	MHV	AY700211
<i>Myodes coronaviruses 2JL14</i>	Myodes rufocanus vole coronavirus 2/JL2014	<i>MrufCoV_2JL14</i>	KY370046
<i>Bat Hp-betacoronavirus</i>	bat Hp-betacoronavirus/Zhejiang2013	<i>Bat_Hp – BetaCoV</i>	KF636752
<i>Hedgehog coronavirus 1</i>	hedgehog coronavirus 1	EriCoV	KC545383
<i>Middle East respiratory syndrome-related coronavirus</i>	Middle East respiratory syndrome-related coronavirus	MERS-CoV	JX869059
<i>Pipistrellus bat coronavirus HKU5</i>	Pipistrellus bat coronavirus HKU5	<i>Pi – BatCoV_HKU5</i>	EF065509
<i>Tylonycteris bat coronavirus HKU4</i>	Tylonycteris bat coronavirus HKU4	<i>Ty – BatCoV_HKU4</i>	EF065505
<i>Eidolon bat coronavirus C704</i>	Eidolon helvum bat coronavirus CMR704-P12	<i>Ei – BatCoV_C704</i>	MG693168
<i>Rousettus bat coronavirus GCCDC1</i>	Rousettus bat coronavirus	<i>Ro – BatCoV_GCCDC1</i>	KU762338
<i>Rousettus bat coronavirus HKU9</i>	Rousettus bat coronavirus HKU9	<i>Ro – BatCoV_HKU9</i>	EF065513
<i>Severe acute respiratory syndrome-related coronavirus</i>	severe acute respiratory syndrome coronavirus	SARS-CoV	AY274119
<i>Severe acute respiratory syndrome-related coronavirus</i>	severe acute respiratory syndrome-related coronavirus	SARS-CoV	KY352407
<i>Severe acute respiratory syndrome-related coronavirus</i>	severe acute respiratory syndrome coronavirus 2	SARS-CoV-2	MN908947

---

**Género Deltacoronavirus**

---

<i>Wigeon coronavirus HKU20</i>	wigeon coronavirus HKU20	<i>WiCoV_HKU20</i>	JQ065048
<i>Bulbul coronavirus HKU11</i>	bulbul coronavirus HKU11	<i>BulCV_HKU11</i>	FJ376619
<i>Common moorhen coronavirus HKU21</i>	common moorhen coronavirus HKU21	<i>CMCoV_HKU21</i>	JQ065049
<i>Coronavirus HKU15</i>	porcine coronavirus HKU15	<i>PoCoV_HKU15</i>	JQ065043
<i>Munia coronavirus HKU13</i>	munia coronavirus HKU13	<i>MunCV_HKU13</i>	FJ376622
<i>White-eye coronavirus HKU16</i>	white-eye coronavirus HKU16	<i>WECoV_HKU16</i>	JQ065044
<i>Night heron coronavirus HKU19</i>	night heron coronavirus HKU19	<i>NHCoV_HKU19</i>	JQ065047

---

**Género Gammacoronavirus**

---

<i>Goose coronavirus CB17</i>	Canada goose-Branta canadensis-coronavirus-Cambridge Bay 2017	<i>BcanCoV_CB17</i>	MK359255
<i>Beluga whale coronavirus SW1</i>	beluga whale coronavirus	BWCoV	EU111742

<i>Avian coronavirus</i>	infectious bronchitis virus	IBV	M95169
<i>Avian coronavirus 9203</i>	infectious bronchitis virus Ind-TN92-03	<i>ACoV_9203</i>	KR902510
<i>Duck coronavirus 2714</i>	duck coronavirus DK/GD/27/2014	<i>DuCoV_2714</i>	KM454473

---

**Tabla A2:** Combinación de colores de Clustal, tomada de la web de Jalview («Clustal Colour Scheme», s.f.). A cada residuo en el alineamiento se le asigna un color según los criterios siguientes.

Clustal X Default Colouring			
Category	Colour	Residue at position	{ Threshold, Residue group }*
Hydrophobic	BLUE	A,I,L,M,F,W,V	{>60%, WLVIMAFCHP}
		C	{>60%, WLVIMAFCHP}
Positive charge	RED	K,R	{>60%,KR},{>80%, K,R,Q}
Negative charge	MAGENTA	E	{>60%,KR},{>50%,QE},{>85%,E,Q,D}
		D	{>60%,KR}, {>85%, K,R,Q}, {>50%,ED}
Polar	GREEN	N	{>50%, N}, {>85%, N,Y}
		Q	{>60%,KR},{>50%,QE},{>85%,Q,E,K,R}
		S,T	{>60%, WLVIMAFCHP}, {>50%, TS}, {>85%,S,T}
Cysteines	PINK	C	{>85%, C}
Glycines	ORANGE	G	{>0%, G}
Prolines	YELLOW	P	{>0%, P}
Aromatic	CYAN	H,Y	{>60%, WLVIMAFCHP}, {>85%, W,Y,A,C,P,Q,F,H,I,L,M,V}
Unconserved	WHITE	any / gap	If none of the above criteria are met

\*Los criterios de coloración se indican de la forma condicional {>X %, y, z}, siendo coloreadas aquellas posiciones del alineamiento con la representación porcentual mínima (X) de determinados aminoácidos combinables (y, z). De este modo, una posición con K o R se coloreará de rojo si la columna contiene más de un 60 % de residuos K y/o R, o más de un 80 % de residuos K, R y Q.