



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**ANÁLISIS EVOLUTIVO Y ESTRUCTURAL DEL
DOMINIO EXON DE LOS NIDOVIRUS**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGO

P R E S E N T A :

ADRIÁN CRUZ GONZÁLEZ



**DIRECTOR DE TESIS:
DR. ALEJANDRO RODRIGO JÁCOME RAMÍREZ**

Ciudad Universitaria, CD. MX., 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Hoja de datos del jurado

1. Datos del Tutor
Dr.
Alejandro Rodrigo
Jácome
Ramírez

2. Datos del sinodal 1
Dra.
Claudia Selene
Zárate
Guerra

3. Datos del sinodal 2
Dr.
Sigifredo
Pedraza
Sánchez

4. Datos del sinodal 3
Dr.
José Alberto
Campillo
Balderas

5. Datos del sinodal 4
Dr.
Santiago
Ávila
Ríos

Análisis evolutivo y estructural del dominio ExoN de los nidovirus

Agradecimientos	4
Resumen	5
Abstract	6
1. Introducción	7
1.1 Características generales de los virus de RNA	7
1.2 Origen y evolución de los virus de RNA	9
1.3 Características de los nidovirus	10
1.4 Exoribonucleasa 3'-5': importancia e implicaciones en virus de RNA	12
1.5 El análisis y la comparación de estructuras terciarias como una herramienta para conocer las relaciones evolutivas	15
2. Objetivos	17
2.1 Objetivo general	17
2.2 Objetivos particulares	17
2.3 Hipótesis	17
3. Métodos	17
3.1 Comparaciones estructurales y construcción del árbol basado en estructuras	17
3.2 Recuperación de las secuencias de los dominios ExoN y RdRp	18
3.3 Alineación de secuencias múltiples y estimación de las filogenias del dominio ExoN y RdRp	18
4. Resultados y discusión	18
4.1 Los eventos de secuestro polifilético de los dominios ExoN en virus	18
4.2 Análisis evolutivo del dominio ExoN en nidovirus	23
4.3 Los dedos de zinc están presentes en las ExoN de virus de RNA, pero no en las enzimas celulares y de fagos	27
4.4 El posible origen celular del dominio ExoN de los arenavirus	29
5. Limitaciones	31
6. Conclusiones	31
Referencias	32
Anexos	39

Agradecimientos

Agradezco a la UNAM por todas las facilidades que me brindó durante mi formación.

A mi familia, mis padres y hermano, quienes han sido los pilares de mi vida.

A mi tutor, el Dr. Alejandro Rodrigo Jácome Ramírez, por todo su apoyo y consejos, gracias a su guía pude culminar este proyecto satisfactoriamente. Le agradezco el haberme dado una increíble confianza, y el enseñarme la importancia de la constancia y el amor a lo que haces.

Al Dr. Antonio Lazcano, al Dr. Arturo Becerra y al Dr. José Campillo, de quienes tuve el privilegio de aprender, y quienes me contagiaron una inmensa pasión por la ciencia.

A todos los miembros del Laboratorio de Origen de la Vida, por sus consejos y aportaciones. En especial a mis colegas y amigos, el Dr. Israel Muñoz, el Dr. Alberto Vázquez, el Dr. Ricardo Hernández, el Dr. José Campillo y el M. en C. Wolfgang Cottom. Gracias por su apoyo incondicional e increíble trabajo.

Al Dr. Sigifredo Pedraza y al Dr. Heriberto Prado, por todo su apoyo y quienes han influido significativamente en mi formación.

A mis amigos de la facultad, Diana Aguilar, Emiliano Villarreal, Leonardo Novoa, Pamela Santiago, Yessica Chacón, Luis Antonio, Zahid Cosío, Bruno Gavin y Xóchitl Sánchez. Haberlos conocido y poder crecer juntos ha sido extraordinario.

A la DGAPA-UNAM (PAPIIT-IN214421 y PAPIME-PE204921) por el apoyo financiero durante el proyecto.

Al programa de Ayudantes de Investigador del Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo económico otorgado durante el desarrollo del proyecto (CVU-1002377).

Resumen

Los virus de RNA son entidades biológicas que poseen genomas pequeños y presentan altas tasas de mutación. La baja fidelidad en la replicación del genoma de los virus de RNA se atribuye, principalmente, a que carecen de mecanismos de edición que sí se encuentran presentes en células y algunos virus de DNA. Los nidovirus y los arenavirus son los únicos virus de RNA conocidos que codifican un dominio exoribonucleasa 3'-5' (ExoN). La actividad de edición del dominio ExoN ha jugado un papel clave en el crecimiento de los genomas en nidovirus, mientras que en los arenavirus este dominio participa en la supresión de la respuesta inmune innata del hospedero. Debido a la importancia del dominio ExoN para estos virus, inferir la posible historia evolutiva de esta proteína mediante el análisis de secuencias y homología estructural es de gran relevancia. Los resultados presentados en este trabajo sugieren que los genes que codifican para las exoribonucleasas virales han sido secuestrados en distintas ocasiones de sus hospederos celulares. Asimismo, los análisis realizados en este trabajo con las secuencias disponibles del dominio ExoN de nidovirus revela un alto nivel de conservación comparable al de las polimerasas de RNA dependientes de RNA virales (RdRp), las cuales son las proteínas virales más conservadas, lo que sugiere que las ExoN juegan un papel clave en el mantenimiento de la integridad y estabilidad del genoma. Por otra parte, los análisis de secuencia nos mostraron que dos dedos de zinc (ZF) están presentes en prácticamente todas las ExoN de nidovirus. Esto contrasta fuertemente con la reportada falta de ZF en las ExoN celulares, y abre la posibilidad para desarrollar estrategias terapéuticas específicas en contra de estos virus de RNA.

Abstract

RNA viruses are biological entities with small genomes and high mutation rates. The low fidelity in the replication of the genome of RNA viruses is mainly attributed to the lack of proofreading mechanisms that are present in cells and some DNA viruses. Nidoviruses and arenaviruses are the only known RNA viruses that encode a 3'-5' exoribonuclease domain (ExoN). The proofreading activity of the ExoN domain may have played a key role in the growth of nidoviral genomes, whereas in arenaviruses this domain partakes in the suppression of the host innate immune signaling. Due to the importance of the ExoN domain for these viruses, inferring the possible evolutionary history of this protein through sequence analysis and structural homology is of great relevance. The results presented in this work suggest that the genes coding for the viral ExoN domains have been hijacked from cellular hosts several times. Furthermore, analysis of the available nidoviral ExoN sequences revealed a high conservation level comparable to that of the viral RNA-dependent RNA polymerases (RdRp), which are the most conserved viral proteins, suggesting that ExoNs play a key role in maintaining the integrity and stability of their genomes. On the other hand, sequence and structural analyses of this work revealed that two highly conserved zinc fingers (ZF) are present in practically all ExoNs of nidoviruses. This is in sharp contrast to the lack of ZF in cellular ExoNs, opening the possibility of developing specific therapeutic strategies against these RNA viruses.

1. Introducción

1.1 Características generales de los virus de RNA

Los virus son simbioses obligados que dependen de una célula para desarrollar su ciclo replicativo. Sus genomas se componen de DNA o RNA, los cuales están rodeados por cubiertas proteicas (cápside y nucleocápside) y, a veces, por una envoltura lipídica adicional (López-García & Moreira, 2012). De acuerdo con Baltimore (1971), los virus pueden dividirse en clases, cada una de las cuales tiene su propia estrategia para llevar a cabo la replicación y transcripción de su genoma. Dentro de la clasificación de Baltimore, los virus de DNA están en los grupos I, dsDNA; II, ssDNA; y VII, dsDNA-RT, mientras que los virus de RNA se encuentran en los grupos III, dsRNA; IV, +ssRNA; V, -ssRNA; y VI, ssRNA-RT (Fig. 1).

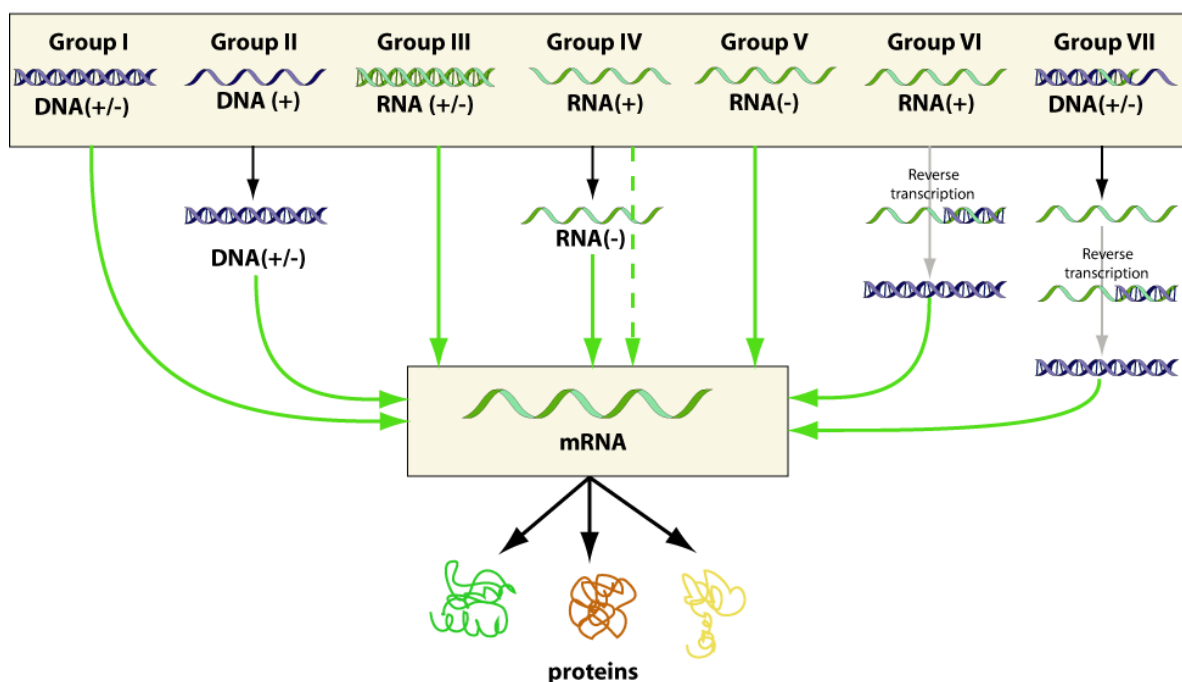


Figura 1. Tipos de genomas virales (DNA o RNA) y mecanismos de síntesis de RNA mensajero (mRNA). Azul, DNA; verde, RNA. Los virus de RNA están clasificados en los grupos III (bicatenario), IV (monocatenario positivo), V (monocatenario negativo) y VI (monocatenario positivo con DNA como intermediario en la replicación). Imagen tomada de ViralZone (<https://viralzone.expasy.org/254>).

Los virus de RNA son extremadamente abundantes y diversos en eucariotes; por el contrario, sólo dos familias infectan a bacterias (*Fiersviridae* y *Cystoviridae*), y aunque han sido detectados en análisis metagenómicos, no han sido aislados en arqueas (Campillo-Balderas *et al.*, 2015; Shi *et al.*, 2016; Wirth & Young, 2020). De hecho, los virus de RNA son entidades biológicas con características únicas. Esto debido a que poseen los genomas más pequeños, con un rango que va de 1.8 kb (*Saccharomyces cerevisiae killer virus M1*, dsRNA) a 41.1 kb (*Planarian secretory cell nidovirus*, +ssRNA) (Campillo-Balderas *et al.*, 2015; Saberi *et al.*, 2018).

Adicionalmente, los virus de RNA poseen altas tasas de mutación (10^{-6} - 10^{-4} sustituciones por nucleótido por infección celular), en comparación con los virus de DNA bicatenarios que poseen tasas de mutación más bajas por órdenes de magnitud (10^{-8} - 10^{-6} sustituciones por nucleótido por infección celular); y, comparado con tasas de mutación exhibidas por bacterias y algunos eucariontes, la diferencia es mucho mayor (10^{-10} - 10^{-11} sustituciones por nucleótido por replicación del genoma) (Drake, 1999; Duffy *et al.*, 2008; Sanjuán & Domingo, 2016; Kautz & Forrester, 2018). Estas características, aunado a los múltiples eventos de replicación viral por célula, favorecen que los virus de RNA alcancen grandes tamaños de poblaciones (Belshaw *et al.*, 2008). Las grandes poblaciones acompañadas por una vasta combinación de mutaciones (benéficas, neutrales y/o deletéreas) resultan en poblaciones de "cuasiespecies" virales, las cuales están estrechamente relacionadas y son sometidas a una selección natural continua (Fig. 2) (Duffy, 2018).

Finalmente, los virus de RNA, son entidades biológicas de vital importancia médica, epidemiológica y económica, puesto que causan muchas de las enfermedades emergentes conocidas en distintos grupos biológicos (Woolhouse *et al.*, 2013; Woolhouse & Brierley, 2018). Entre algunas de las enfermedades de importancia se incluyen la enfermedad por el virus del Ébola en humanos, la influenza aviar en varios grupos de aves, la enfermedad del virus del tabaco en plantas, la arteritis viral equina en caballos y la muy reciente enfermedad por coronavirus 2019 (COVID-19) en humanos.

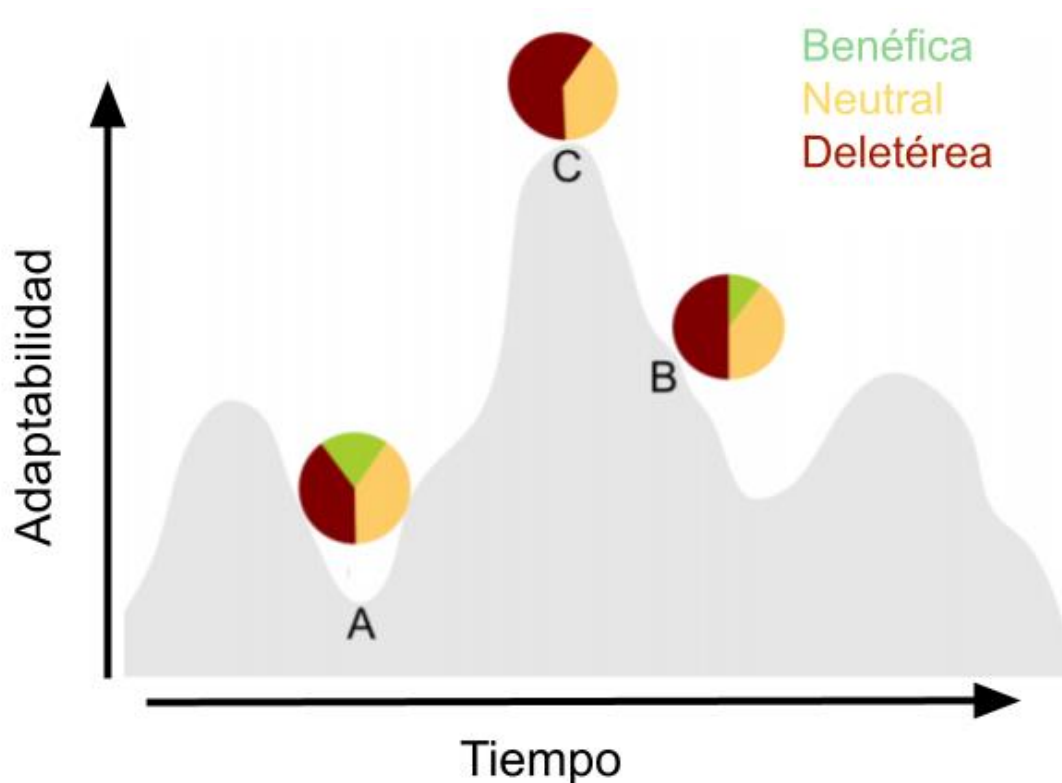


Figura 2. Mutaciones acumuladas en tres distintos tipos de genomas. El genotipo A no se encuentra bien adaptado al entorno, así que acumula más mutaciones que podrían ser benéficas. El genotipo B se encuentra mejor adaptado al ambiente que A, pero no se encuentra en el pico de aptitud, por lo que aún desarrolla mutaciones benéficas. El genotipo C se encuentra bien adaptado al ambiente por lo cual no desarrolla mutaciones benéficas. Imagen modificada de Duffy, 2018.

1.2 Origen y evolución de los virus de RNA

El descubrimiento de los RNAs catalíticos dio una credibilidad considerable a la propuesta de que las primeras entidades vivas se basaban en el RNA como material genético y como catalizador en una etapa hipotética llamada mundo del RNA (Gilbert, 1986). De hecho, las propiedades catalíticas, reguladoras y estructurales de las moléculas de RNA y los ribonucleótidos, combinadas con su ubicuidad en los procesos celulares, sugieren que desempeñaron un papel clave en la evolución temprana y quizás en el origen de la vida (Lazcano, 2010; Vázquez-Salazar & Lazcano, 2018). Hoy en día, las únicas entidades biológicas conocidas cuyos genomas consisten en RNA, se encuentran en la amplia gama de virus y viroides de RNA. Inferir los orígenes y las relaciones filogenéticas profundas de los virus de RNA ha demostrado ser una tarea difícil de resolver. Esto se debe principalmente a dos problemas. El primero consiste en la carencia de cualquier tipo de registro fósil, lo cual hace imposible utilizar tiempos de muestreo fósil para datar la evolución viral. El segundo se encuentra en sus altas tasas de mutación, que aseguran que la señal filogenética presente en sus secuencias se borre rápidamente (Holmes, 2009; Lazcano, 2010; Sanjuán & Bradwell, 2010).

A pesar de las dificultades en la reconstrucción de la historia evolutiva, se han propuesto tres importantes hipótesis sobre los orígenes de los virus de RNA (Holmes, 2009). La hipótesis de la regresión establece que estos virus provienen de pequeñas protocélulas de RNA, que perdieron sus elementos celulares en el curso de la evolución (Fig. 3A) (Fisher, 2010; Mughal *et al.* 2020). En la hipótesis del escape se postula que los virus de RNA son derivados de RNA celular y/o fragmentos de DNA tales como plásmidos y transposones, los cuales escaparon del hospedero adquiriendo capas proteicas protectoras y la capacidad de replicarse de forma autónoma (Fig. 3B) (Forterre & Prangishvili, 2009; Durzyńska & Goździcka-Józefiak, 2015). Por último, la hipótesis virocéntrica o precelular establece que los virus de RNA son descendientes de formas de vida precelulares de RNA, cuyos ancestros datan de miles de millones de años, por lo que adoptaron su estilo de vida parasitario más adelante en el tiempo evolutivo (Fig. 3C) (Koonin *et al.*, 2006; Holmes, 2009). Sin embargo, aunque cada hipótesis sobre el origen de los virus de RNA tiene puntos fuertes y débiles, su origen aún es tema de debate debido a que es difícil comprobar la validez de cada una de estas hipótesis (Holmes, 2009; Campillo-Balderas & Becerra, 2020).

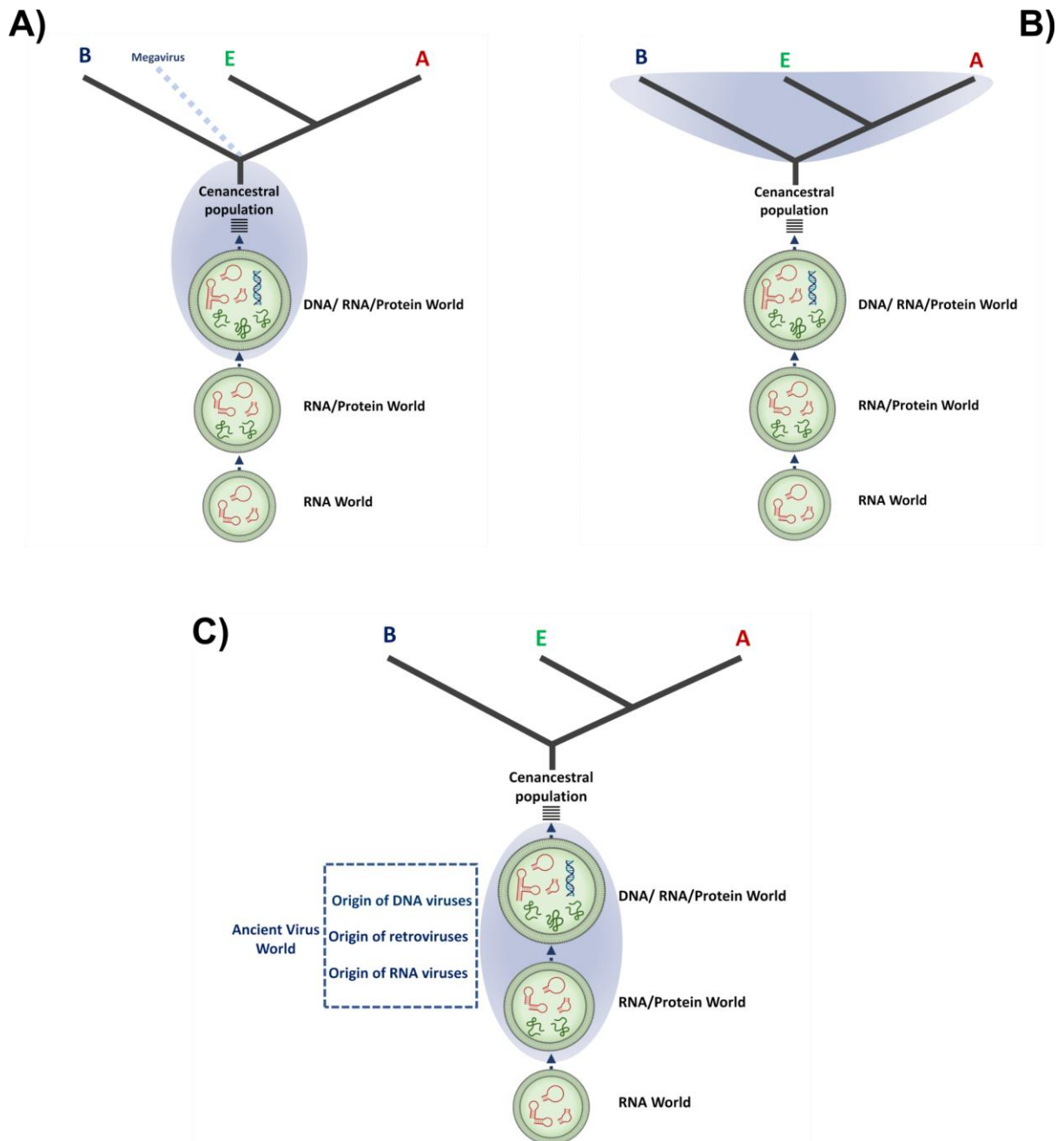


Figura 3. Teorías sobre el origen de los virus. (A) Hipótesis de la evolución regresiva; (B) hipótesis del escape; y (C) hipótesis del origen precelular. Las letras B, E y A representan a los dominios Bacteria Eukarya y Archaea. La marca azul indica el momento del posible origen de los virus en cada una de las hipótesis. Imagen modificada de Campillo-Balderas & Becerra, 2020.

1.3 Características de los nidovirus

El orden *Nidovirales* se conforma actualmente por ocho subórdenes distintos divididos en 14 familias, que a su vez contienen 109 especies virales (Tabla 1) (*International Committee on Taxonomy of Viruses*, <https://ictv.global/taxonomy/>; Siddell *et al.*, 2019). Los nidovirus pueden infectar una sorprendente variedad de hospederos vertebrados e invertebrados, incluidos mamíferos, aves, anfibios, peces, reptiles, artrópodos, moluscos y helmintos (Lauber *et al.*,

2013; Lefkowitz *et al.*, 2018). Este orden contiene a los virus de RNA con los genomas más grandes, los cuales van de menos de 13 kb para el virus de la arteritis equina (EAV, por sus siglas en inglés) hasta más de 41 kb para el nidovirus de las células secretoras de planaria (PSCNV, por sus siglas en inglés) (Saber *et al.*, 2018; Ogando *et al.*, 2019) (Tabla 1).

Tabla 1. Familias virales pertenecientes al orden *Nidovirales*.

Suborden	Familia	Tamaño del genoma (kb)	Hospederos
<i>Amidovirineae</i>	<i>Arteriviridae</i>	12.7 - 15.7	Caballos, cerdos, zarigüeyas, musarañas, primates, roedores
	<i>Cretegaviridae</i>	14.9	Tortuga
	<i>Gresnaviridae</i>	18.4	Serpiente
	<i>Olifoviridae</i>	15.3	Serpiente
<i>Nanidovirineae</i>	<i>Nanghoshaviridae</i>	13.1	Pez
	<i>Nanhypoviridae</i>	18.2	Pez
<i>Mesnidovirineae</i>	<i>Medioniviridae</i>	~ 20.2 - 25	Gasterópodos
	<i>Mesoniviridae</i>	~ 20	Artrópodos
<i>Cornidovirineae</i>	<i>Coronaviridae</i>	~ 27 - 32	Aves, bovinos, perros, gatos, cerdos, roedores, ballenas, humanos
<i>Tornidovirineae</i>	<i>Tobaniviridae</i>	~ 20 - 33	Mamíferos, peces, serpientes
<i>Ronidovirineae</i>	<i>Roniviridae</i>	~ 26	Peces, langostinos
	<i>Euroniviridae</i>	~ 24.5	Crustáceos
<i>Abnidovirineae</i>	<i>Abyssoviridae</i>	35.9	Gasterópodos
<i>Monidovirineae</i>	<i>Mononiviridae</i>	41.1	Helmintos

Azul, virus sin presencia del dominio ExoN; rojo, virus dotados con el dominio ExoN. Tabla tomada y modificada de Cruz-González *et al.*, 2021.

Los viriones poseen una envoltura lipídica y genomas de RNA lineales, monocatenarios, no segmentados y de polaridad positiva. El genoma de los nidovirus está compuesto de dos grandes ORFs (ORF1a y ORF1b), seguidos primero por un conjunto de cuatro genes de proteínas estructurales: espícula (S), membrana (M), envoltura (E) y nucleocápside (N), y por último un número variable de ORFs que codifican las llamadas proteínas accesorias (Masters,

2006; Snijder *et al.*, 2016). La traducción del ORF1a y el ORF1b, produce dos poliproteínas, la poliproteína 1a (pp1a) y la poliproteína 1ab (pp1ab), que dan como resultado un número variable de proteínas no estructurales (nsp's) entre los miembros de las distintas familias del orden, y que están implicadas en la replicación, transcripción, modificación de los organelos celulares, entre otras funciones (Fehr & Perlman, 2015). En coronavirus, la pp1a alberga de la nsp1 a la 11, mientras que la pp1ab, la cual resulta de un desplazamiento del marco ribosómico -1 en una superposición del ORF1a con el ORF1b, contiene todas las nsp's (nsp1–16) (Nakagawa *et al.*, 2016).

Los nidovirus presentan una diversidad considerable en cuanto al número y tamaño de las proteínas codificadas en sus genomas. Sin embargo, entre ellos se conservan cinco dominios centrales, incluida la proteasa principal (Mpro), la RNA polimerasa dependiente de RNA (RdRp) siempre unida a la nucleotidil-transferasa, el dominio helicasa de la superfamilia 1 (Hel) y un dominio de unión a zinc (Z), que siempre está asociado con la Hel (Fig. 4) (Ogando *et al.*, 2019).

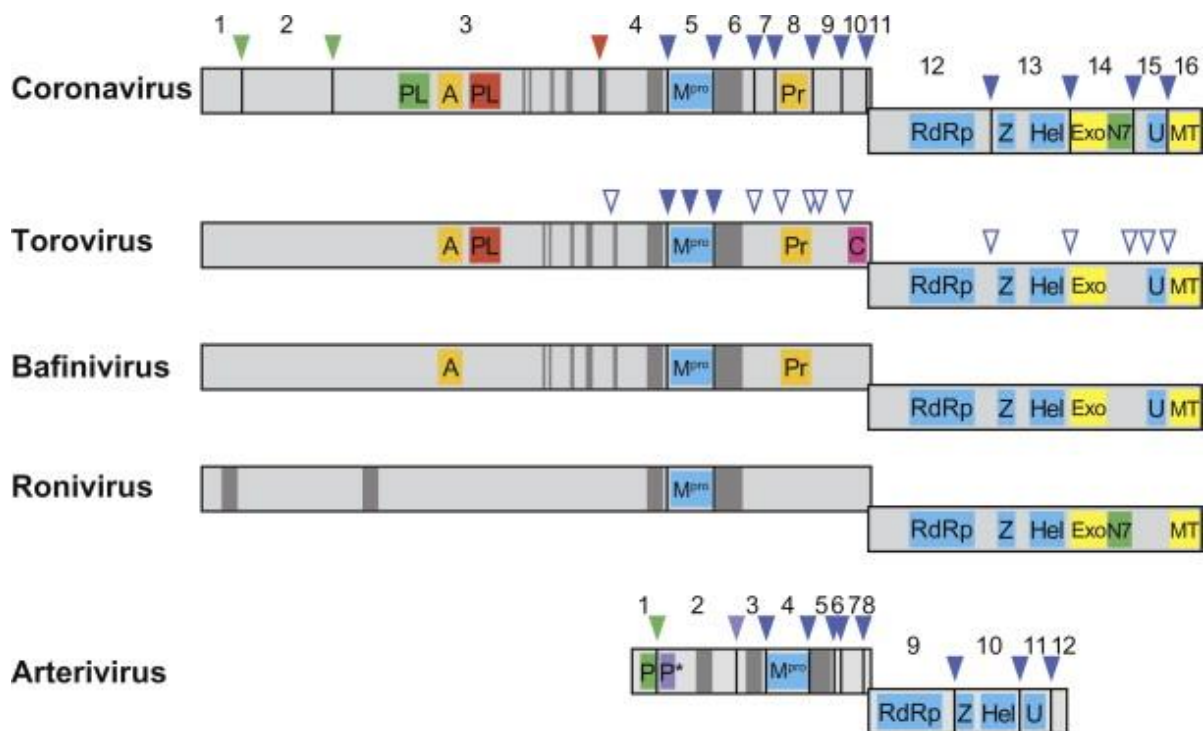


Figura 4. Representación esquemática de la organización general de las poliproteínas (pp1a y pp1ab) de algunos nidovirus. Abreviaturas de los dominios: PL/P, proteasa tipo papaína; Mpro, proteasa principal; RdRp, RNA polimerasa dependiente de RNA; Z, dominio de unión a zinc; Hel, helicasa de la superfamilia 1; Exo, exoribonucleasa 3'-5'; N7, N7-metiltransferasa; U, endorribonucleasa específica de uridilato; MT, 2'-O-metil transferasa. Los sitios de escisión en las pp1a y pp1ab por parte de la PL/P y de la Mpro se indican mediante puntas de flecha. Con la excepción del dominio U, los dominios resaltados en cuadros azules son característicos de nidovirus. Imagen modificada de King *et al.*, 2012.

1.4 Exoribonucleasa 3'-5': importancia e implicaciones en virus de RNA

Las exonucleasas 3'-5' (DnaQ-like) comparten ciertas características. Estas enzimas catalizan la escisión de nucleósidos monofosfato a partir de ácidos nucleicos en la dirección 3' a 5' utilizando un mecanismo que depende de dos iones Mg^{2+} y una molécula de agua reactiva (Yang, 2011). Estas enzimas las podemos encontrar asociadas a polimerasas o como proteínas autónomas, como la subunidad épsilon de la DNA Pol III y la proteína TREX2, respectivamente (Shevelev & Hübscher, 2002). Otra característica de estas exonucleasas es que encontramos los cuatro residuos conservados del sitio activo, aspartato (D) y glutamato (E), que se distribuyen en tres motivos canónicos (Exo I, D/E; Exo II, D; y Exo III, D) en la estructura primaria (Zuo & Deutscher, 2001). Sin embargo, dentro del motivo de secuencia, es posible identificar un quinto residuo conservado, el cual puede variar entre histidina (H) y tirosina (Y), conformando así dos grandes grupos entre estas proteínas: el grupo DEDDh y el grupo DEDDy (Zuo & Deutscher, 2001; Shevelev & Hübscher, 2002). Por último, las exonucleasas DnaQ-like, también comparten una topología conservada $\beta 1-\beta 2-\beta 3-\alpha A-\beta 4-\alpha B-\beta 5-\alpha C$ en la que cinco hebras forman una hoja β mixta central rodeada por al menos tres α -hélices (Yang, 2011; Cruz-González *et al.*, 2021). Esta familia de proteínas se encuentra clasificada en las bases de datos Pfam (Mistry *et al.*, 2021) y SCOP (Andreeva *et al.*, 2020) como RNase_T (ID: PF00929) y DnaQ-like 3'-5' exonuclease (ID: 4000547), respectivamente.

En coronavirus, la nsp14 (527 aa) es una proteína compuesta por dos dominios estructuralmente interconectados, pero funcionalmente distintos: el dominio N-terminal corresponde a la exoribonucleasa 3'-5' (ExoN, 1-287 aa) y el dominio C-terminal a la N7-metiltransferasa (N7-Mtase, 288-527 aa) (Fig. 5A) (Ferron *et al.*, 2018). El descubrimiento del dominio ExoN en coronavirus se basó en la similitud de la secuencia de la proteína nsp14 del coronavirus del síndrome respiratorio agudo grave (SARS-CoV, por sus siglas en inglés) con homólogos celulares, tales como el dominio exonucleasa de edición de la DNA polimerasa I de *Escherichia coli*, perteneciente a la familia DnaQ-like de exonucleasas (Snijder *et al.* 2003). Mediante ensayos *in vitro* con una nsp14 recombinante del SARS-CoV, se demostró que el dominio ExoN se desempeña como una exoribonucleasa 3'-5' (Minskaia *et al.*, 2006). Aunque el dominio ExoN del SARS-CoV posee estructuras adicionales como dos dedos de zinc (ZF) conformados por residuos de histidina y cisteína, y un sitio de unión a nsp10 (Fig. 5A), el núcleo catalítico se conforma por una hoja β central compuesta por cinco hebras β , las cuales están flanqueadas por α -hélices en cada lado (Fig. 5B) (Ma *et al.*, 2015). Tal disposición de los elementos estructurales centrales es similar a las estructuras de las exonucleasas DnaQ-like. La cristalización de la estructura terciaria de la nsp14 del SARS-CoV reveló que, de hecho, es una enzima DEED puesto que en el motivo Exo II, se identificó al glutamato (E) como el residuo catalítico en el sitio activo (Ma *et al.*, 2015; Ferron *et al.*, 2018). Sin embargo, este cambio no modifica la relación evolutiva entre las exonucleasas DnaQ-like y la ExoN viral (*vide infra*).

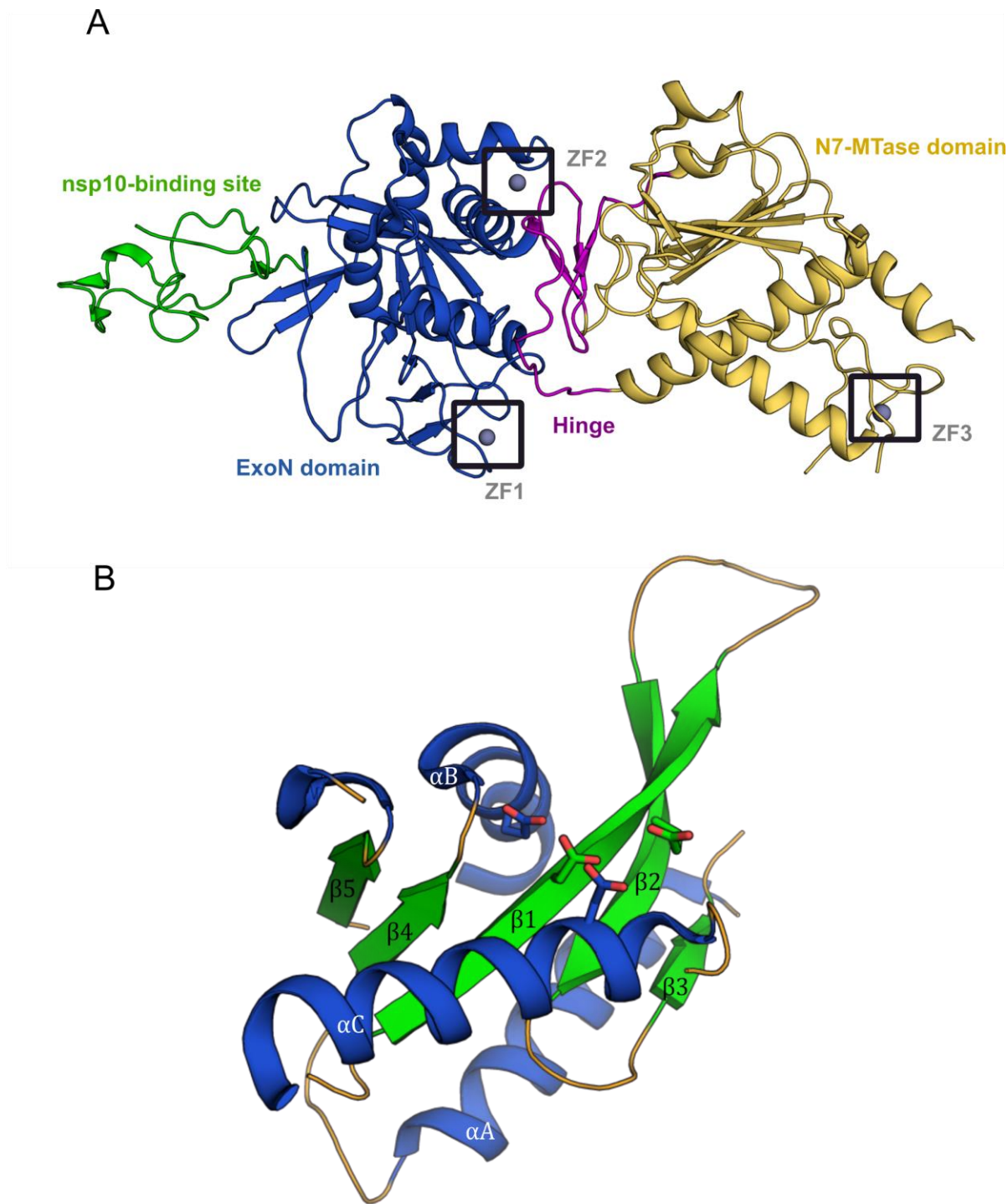


Figura 5. Estructura de la nsp14 del SARS-CoV y su dominio ExoN. (A) Representación de la estructura terciaria del heterodímero nsp14. En color verde se resalta el sitio de unión nsp10-nsp14; en azul el dominio ExoN de la nsp14; en púrpura la región de bisagra; y en amarillo el dominio N7-MTase de la nsp14. Los iones de zinc están representados como esferas grises y resaltados en recuadros. (B) Se muestra la topología característica β 1- β 2- β 3- α A- β 4- α B- β 5- α C del núcleo catalítico de las exonucleasas DnaQ-like (las cadenas laterales de los residuos del sitio activo se representan con bastones) utilizando el dominio ExoN del SARS-CoV (ID de PDB: 5C8U). α -hélices en azul; hojas β en verde; y elementos de conexión en naranja. Imagen producida con PyMOL (The PyMOL Molecular Graphics System, Version 1.8.2.1 Schrödinger, LLC).

Mediante la comparación de variantes de virus mutantes del dominio ExoN y variantes silvestres del SARS-CoV y del virus de la hepatitis del ratón (MHV, por sus siglas en inglés),

se confirmó que la inactivación del dominio ExoN confiere un "fenotipo mutador", en el cual aumenta la frecuencia de mutación en relación con el fenotipo silvestre (Eckerle *et al.*, 2007, 2010). De igual forma, la capacidad del dominio ExoN para escindir nucleótidos mal apareados en el extremo 3' de un sustrato de dsRNA se demostró *in vitro* utilizando la nsp14 del SARS-CoV, lo cual confirmó su actividad como una enzima de edición (Bouvet *et al.*, 2012). Asimismo, se ha reportado que el dominio ExoN puede estar relacionado con funciones adicionales, como la evasión del sistema inmune en el virus de gastroenteritis transmisible (TGEV, por sus siglas en inglés) (Becares *et al.*, 2016), replicación y síntesis de RNA en MERS-CoV y SARS-CoV-2 (Ogando *et al.*, 2020), y recombinación en MHV, MERS-CoV y SARS-CoV-2 (Gribble *et al.*, 2021).

La baja fidelidad en la replicación del genoma de los virus de RNA se atribuye, principalmente, al hecho de que tanto la RdRp como la transcriptasa reversa (RT, por sus siglas en inglés) carecen de mecanismos de edición, los cuales sí se encuentran presentes en polimerasas celulares (Lloyd *et al.*, 2013; Ganai & Johansson, 2016; Kautz & Forrester, 2018). Es por ello que, el hallazgo del dominio ExoN de edición en virus de RNA ha sido una aportación importante al estudio de la evolución de los virus. El dominio ExoN es de vital importancia en el estudio del origen y evolución de los virus de RNA, no sólo por su capacidad de edición, sino porque únicamente está presente en virus del orden *Nidovirales* con genomas de más de 20 kb (Gorbalenya *et al.*, 2006; Ogando *et al.*, 2019) (Tabla 1). Esto podría sugerir que la adquisición del dominio ExoN por los nidovirus, probablemente permitió la expansión del genoma más allá del límite observado para otros virus de RNA (Nga *et al.*, 2011; Lauber *et al.*, 2013). También, se han identificado homólogos de la familia de exonucleasas DnaQ-like en otros virus, incluidos los arenavirus (virus de RNA monocatenario con polaridad negativa), así como en los fagos ϕ 29, RB69, T4 y T7 de DNA bicatenario (Wang *et al.*, 1996; Doublie *et al.*, 1998; Kamtekar *et al.*, 2006; Hastie *et al.*, 2012; Xia *et al.*, 2014).

El estudio de la nsp14 en coronavirus, específicamente del dominio ExoN, es importante en un ámbito clínico. La familia *Coronaviridae*, incluye a virus como el SARS-CoV, el MERS-CoV, el más reciente SARS-CoV-2, y algunos otros similares que son una de las principales amenazas para la salud pública y del ganado (Fan *et al.*, 2019). La inhibición de la función del dominio ExoN en los coronavirus afectaría la estabilidad e integridad del genoma viral, lo cual puede conducir a copias resultantes que no sean viables, y por lo tanto no patógenas (Smith *et al.*, 2013). Indudablemente, el dominio ExoN viral es un blanco de investigación de gran importancia, tanto en un ámbito clínico, como biológico y evolutivo.

1.5 El análisis y la comparación de estructuras terciarias como una herramienta para conocer las relaciones evolutivas

El hacer comparaciones siempre ha sido un aspecto fundamental de la biología, y en específico, de la biología evolutiva. De hecho, la motivación de generar una cantidad enorme de datos biológicos (secuencias, estructuras, etc.) es obtener nuevos conocimientos biológicos haciendo

comparaciones; comparaciones entre organismos, entre secuencias, entre estructuras, etc. (Sierk & Kleywegt, 2004). Desde los 80's, con el trabajo de Lesk & Chothia (1980) sobre la similitud estructural entre las hemoglobinas, se sabe que el grado de divergencia en las estructuras terciarias de proteínas homólogas en una familia está relacionado con el grado de similitud en sus secuencias de aminoácidos. También, es bien sabido que las estructuras terciarias de las proteínas homólogas se conservan mejor que sus secuencias de aminoácidos (Flores *et al.*, 1993). Por ejemplo, en varios casos en los que se observa una similitud de secuencia baja, las proteínas retienen el plegamiento y conservan las características bioquímicas y/o propiedades funcionales, lo cual sugiere una conexión evolutiva (Russell & Sternberg, 1997; Garau *et al.*, 2005; Jácome *et al.*, 2015; Cruz-González *et al.*, 2021). Es decir, el nivel de similitud entre dos proteínas aumenta a medida en que las vemos desde "secuencias de bases en genes", a "secuencias de aminoácidos en proteínas" y hasta "estructuras terciarias de proteínas" (Balaji & Srinivasan, 2007). Por lo tanto, utilizar la comparación de estructuras terciarias, y la construcción de árboles evolutivos basados en estas, como una herramienta para conocer las relaciones evolutivas entre proteínas muy divergentes, es una alternativa interesante.

Actualmente existen varios programas que nos permiten realizar comparaciones estructurales. En general, estos métodos comparan la geometría de los átomos de la cadena de carbonos alfa ($C\alpha$), pero se basan en diferentes algoritmos y han sido diseñados para diversas aplicaciones (Sierk & Kleywegt, 2004; Mayr *et al.*, 2007). Las comparaciones de las estructuras terciarias entre proteínas homólogas nos proporcionan dos valores importantes. En primer lugar, una distancia cuadrática promedio en Angstroms (RMSD, por sus siglas en inglés) del número de $C\alpha$ de los aminoácidos que se alinean en el espacio tridimensional de las estructuras comparadas. Y, en segundo lugar, el número de residuos superpuestos entre las dos proteínas homólogas. En sí, el RMSD no puede ser utilizado por sí solo como un "proxy" de distancia evolutiva, ya que los valores de RMSD y el número de residuos superpuestos no son independientes; si se emparejan menos residuos, el valor de RMSD será menor, y viceversa, a mayor número de residuos superpuestos el RMSD tenderá a ser mayor. Para normalizar los resultados, se utiliza una distancia geométrica denominada "Structural alignment score" (SAS), que se calcula mediante la fórmula $(RMSD \times 100) / \text{número de residuos superpuestos}$ (Subbiah *et al.*, 1993), la cual permite construir una matriz de distancias. Esta matriz es utilizada por programas de análisis evolutivo para obtener un árbol filogenético de las proteínas comparadas. Estudios previos con las proteínas Metallo-B-lactamasas (Garau *et al.*, 2005) y lipocalinas (Lakshmi *et al.*, 2015) han demostrado la utilidad de métodos similares al comprobar las relaciones evolutivas entre proteínas celulares cuyos niveles de identidad son bastante bajos. En el caso de proteínas virales, la comparación de las estructuras terciarias, y los correspondientes árboles filogenéticos, han permitido aclarar las relaciones evolutivas de proteínas de la cápside de virus de DNA (Krupovic & Bamford 2008), y más recientemente las relaciones entre las RdRp's virales (Jácome *et al.*, 2015).

En este trabajo, se realizó un análisis de la posible historia evolutiva del dominio ExoN viral. Debido a la disparidad en la tasa de mutación entre las entidades biológicas de DNA y RNA, se construyeron filogenias basadas en estructuras terciarias, en las que se pueden reconocer

varios eventos de secuestro del dominio ExoN viral. Este análisis evolutivo de las ExoN virales, en el que se han incluido sus contrapartes celulares, demuestra el secuestro viral polifilético del dominio ExoN celular. De igual forma, el análisis filogenético de las secuencias del dominio ExoN en nidovirus reveló un nivel de conservación de secuencia similar al de las RdRp. Esto y su gran importancia en el ciclo viral sugieren que el dominio ExoN de los nidovirus debe considerarse como un blanco terapéutico importante contra estos virus.

2. Objetivos

2.1 Objetivo general

- Inferir la posible historia evolutiva del dominio exoribonucleasa 3'-5' de los nidovirus.

2.2 Objetivos particulares

- Identificar la distribución del dominio exoribonucleasa 3'-5' en virus del Orden *Nidovirales*.
- Construir un árbol evolutivo basado en la comparación de estructuras terciarias del dominio exoribonucleasa 3'-5' en el que se incluyen enzimas virales y celulares.

2.3 Hipótesis

- Las exoribonucleasas de los nidovirus tienen un origen celular.

3. Métodos

3.1 Comparaciones estructurales y construcción del árbol basado en estructuras

Se realizó una búsqueda de homólogos estructurales del dominio ExoN del SARS-CoV (complejo nsp14-nsp10, cadena B, residuos 1-287; ID de PDB: 5C8U) en la base de datos del Protein Data Bank (PDB) utilizando el servidor DALI (Holm & Rosenström, 2010). Se recuperaron treinta y tres estructuras cristalográficas (Tabla S1) no redundantes con una puntuación $Z > 4$ de la base de datos del PDB (Burley *et al.*, 2019). Para construir el árbol evolutivo basado en estructuras, fue necesaria la edición de los archivos PDB con el fin de aislar el dominio de interés, ya que la presencia de dominios adyacentes podría dificultar el reconocimiento de las regiones conservadas. Posteriormente, se realizaron comparaciones por pares entre las estructuras seleccionadas con el servidor en línea PDBe Fold (Mir *et al.*, 2018), el cual se basa en el alineamiento y comparación de elementos estructurales secundarios. De cada comparación, se recopiló el RMSD y el número de residuos superpuestos para calcular la puntuación de alineación estructural SAS (Subbiah *et al.*, 1993), que se define como:

$$(100 \times RMSD) \div \# \text{ de residuos superpuestos}$$

y se construyó una matriz de distancia geométrica usando los valores SAS. Finalmente, se utilizó el programa Fitch, incluido en el paquete PHYLIP (Felsenstein, 1993), para calcular el árbol con la matriz de distancias de los valores SAS.

3.2 Recuperación de las secuencias de los dominios ExoN y RdRp

Las secuencias de la poliproteína 1ab de los nidovirus se obtuvieron manualmente de la base de datos NCBI-RefSeq. Para cada secuencia, se ejecutó un algoritmo de alineación local (Smith-Waterman con los parámetros predeterminados), utilizando el dominio ExoN (YP_009725309.1) y la secuencia de la nsp12 (ID de PDB: 6M71) del SARS-CoV-2 como referencias.

3.3 Alineación de secuencias múltiples y estimación de las filogenias del dominio ExoN y RdRp

Se construyó una alineación de secuencias múltiples para cada una de las dos proteínas con el servidor PROMALS3D (Pei & Grishin, 2014) cargando el dominio ExoN de SARS-CoV (ID de PDB: 5C8U), y la estructura terciaria de la nsp12 (ID de PDB: 6M71) del SARS-CoV-2 como referencias. Para las secuencias del dominio ExoN, el mejor modelo bajo el criterio de Akaike fue LG + F + R5 calculado con ModelFinder (Kalyaanamoorthy *et al.*, 2017), y se infirió una filogenia con máxima verosimilitud y 100 bootstraps no paramétricos implementados en IQ-TREE2 (Minh *et al.*, 2020). Las ramas con valores <50% se colapsaron utilizando TreeCollapserCL4 (Hodcroft, <http://emmahodcroft.com/TreeCollapseCL.html>). Para las secuencias de la RdRp, el mejor modelo bajo el criterio de Akaike fue LG + F + R6 calculado con ModelFinder (Kalyaanamoorthy *et al.*, 2017), y se infirió una filogenia con máxima verosimilitud y 100 bootstraps no paramétricos implementados en IQ-TREE2 (Minh *et al.*, 2020). Las ramas con valores <50% se colapsaron utilizando TreeCollapserCL4 (Hodcroft, <http://emmahodcroft.com/TreeCollapseCL.html>). Las filogenias se editaron y visualizaron con Figtree (Rambaut, <http://tree.bio.ed.ac.uk/software/figtree/>).

4. Resultados y discusión

4.1 Los eventos de secuestro polifilético de los dominios ExoN en virus

Dado que la estructura terciaria de las proteínas se encuentra más conservada que la secuencia de aminoácidos (Černý *et al.*, 2015; Jácome *et al.*, 2015), construimos un árbol filogenético basado en las superposiciones espaciales de las estructuras terciarias disponibles de exonucleasas de la familia DnaQ-like. En total, se recuperaron 34 estructuras terciarias del dominio exonucleasa. De ellas, dieciséis pertenecen a eucariontes, nueve a bacterias, una a arqueas, cuatro a virus de DNA de doble cadena, tres a virus de RNA de una cadena negativa, y una a virus de RNA de una cadena positiva (Tabla S1). En general, algunas de las funciones que desempeñan estas exonucleasas son: edición, procesamiento y maduración del RNA,

degradación de RNA y/o DNA citoplasmático, actividad inmune y reparación del DNA. Como se muestra en el árbol de estructuras (Fig. 6), existe una distribución no aleatoria de los dominios ExoN, con las exonucleasas DEDDh y DEDDy ubicadas en clados distintos y claramente definidos. Las exonucleasas tienen una amplia gama de funciones, y el hecho de que diferentes funciones estén ubicadas en un mismo clado resalta su versatilidad funcional y falta de especificidad absoluta.

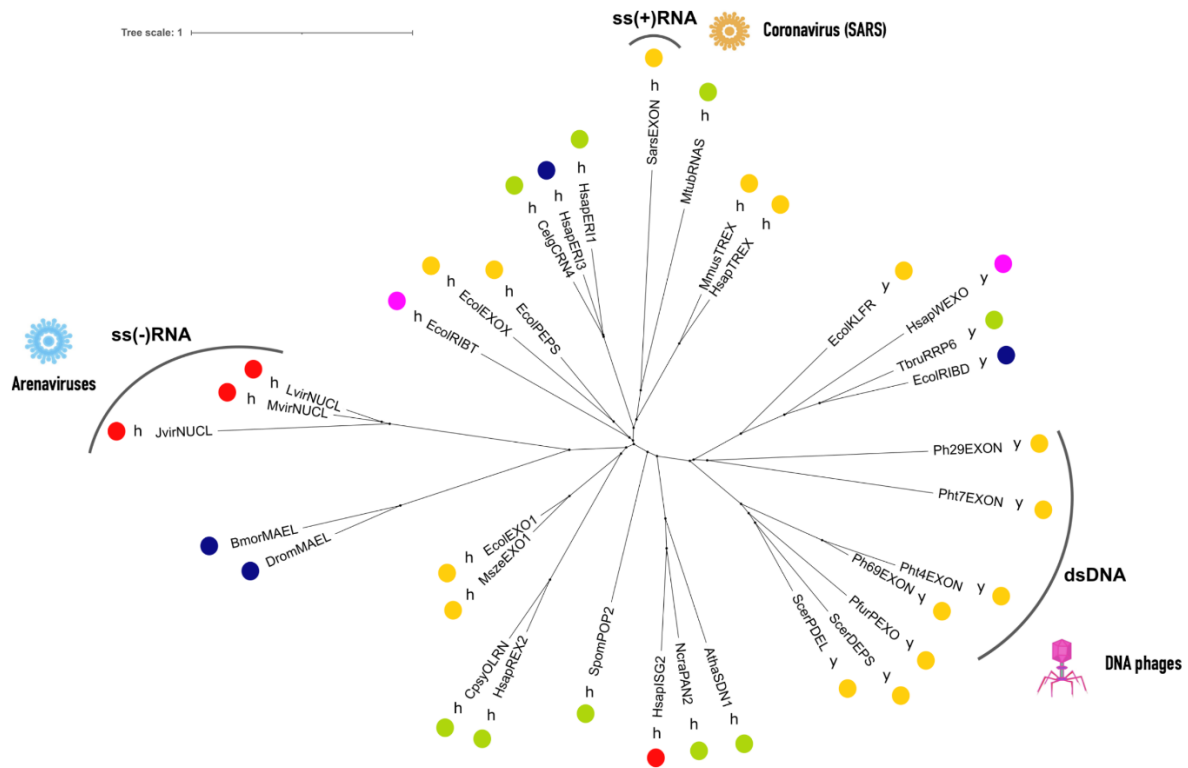


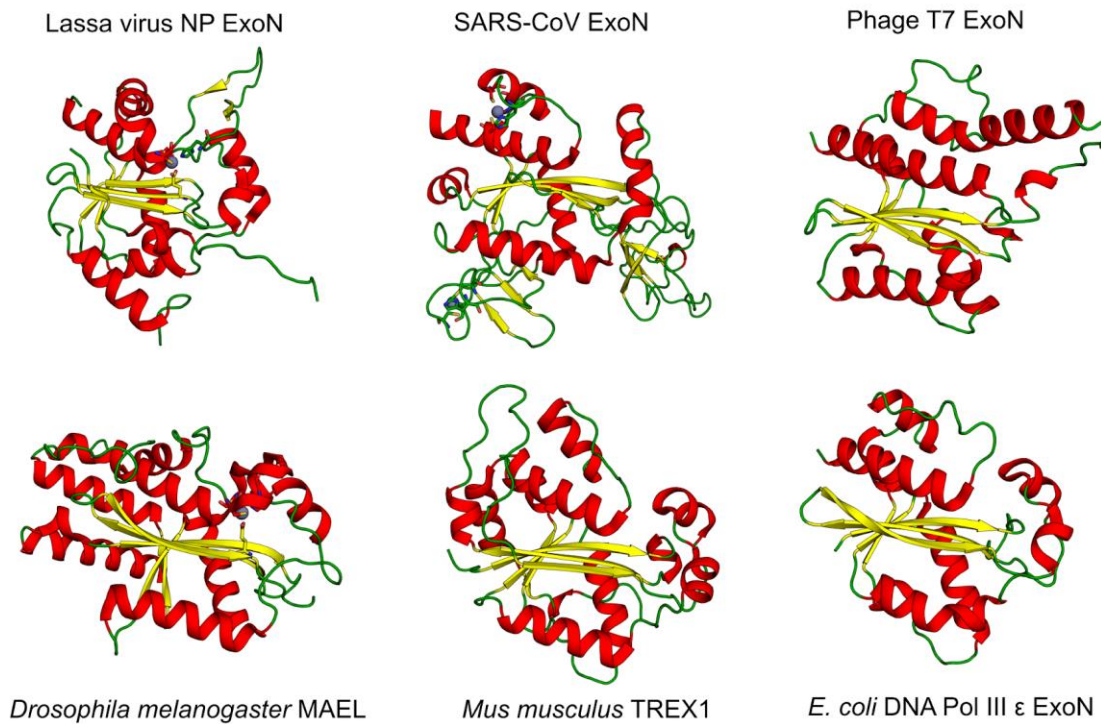
Figura 6. Árbol no enraizado basado en las comparaciones estructurales de las exonucleasas DnaQ-like. Los puntos indican la función de la exonucleasa, cuyos colores representan: amarillo, edición; verde, degradación de RNA y/o DNA citoplasmático; rojo, actividad inmune; azul oscuro, procesamiento y maduración de RNA; fucsia, reparación del DNA. Las letras simples indican el subgrupo de exonucleasa DEDD según el quinto residuo más conservado en el núcleo catalítico: y, subgrupo DEDDy; h, subgrupo DEDDh. Ver Tabla S1 para los nombres de cada OTU. Imagen modificada de Cruz-González *et al.*, 2021.

Actualmente, aparte de los nidovirus, los otros únicos virus de RNA dotados con un dominio ExoN 3'-5' son los arenavirus, los cuales son virus de RNA de polaridad negativa, segmentados y envueltos que pertenecen a la familia *Arenaviridae* (orden *Bunyavirales*). Sin embargo, se ha sugerido que el dominio ExoN en arenavirus está involucrado en la supresión de la señalización inmune innata, al degradar el dsRNA y evitar su detección por el sistema inmune, y no en la actividad de edición (Reynard *et al.*, 2014); lo cual es consistente con el papel del dominio ExoN de edición en el aumento del tamaño del genoma de los nidovirus (~25 kb) ya que los genomas de los arenavirus son pequeños (~10.5 kb). En el árbol de estructuras (Fig. 6), el dominio ExoN del SARS-CoV se encuentra dentro de una rama que incluye las ExoN de tipo DEDDh de eucariontes y procariontes con múltiples funciones, incluida la edición, la degradación del ácido nucleico citoplasmático y las actividades de maduración, procesamiento y unión al RNA. Por otro lado, el dominio ExoN de los arenavirus (*Lassa virus*, *Mopeia virus*

y *Junin virus*) está ubicado en una rama diferente de las ExoN de tipo DEDDh, y su rama hermana incluye las enzimas MAEL, las cuales son un grupo atípico de nucleasas que carecen de la característica pentada catalítica DEDDh/y.

Aunque las ExoN del SARS-CoV (*Coronaviridae*) y del Lassa virus (LASV, *Arenaviridae*) son homólogas y poseen características de la familia de exonucleasas DnaQ-like (Fig. 7A), las diferencias estructurales entre las proteínas apoyan la idea de eventos de adquisición independientes por parte de estos virus, que en el caso de los arenavirus llevaron a una función diferente. A continuación, se describen algunas de estas diferencias (Fig. 7B). En primer lugar, el dominio ExoN del LASV tiene un motivo “loop” básico (K516, K517, K518 y R519) en forma de un brazo saliente ubicado a la izquierda del sitio activo, que interactúa con el RNA, y el cual está ausente en el dominio ExoN del SARS-CoV. En segundo lugar, el dominio ExoN del SARS-CoV está dotado de un sitio de unión a cofactor adicional (sitio de unión a nsp10) que está ausente en el dominio ExoN del LASV. En tercer lugar, el dominio ExoN del SARS-CoV está interconectado estructuralmente con un dominio N7-MTasa, mientras que el dominio ExoN del LASV está vinculado al dominio central de la nucleoproteína (NP) (Fig. 8). Finalmente, como se observa en la Figura 7B, ninguno de los ZFs presentes en el dominio ExoN del SARS-CoV corresponde espacialmente al del LASV (Ma *et al.*, 2015; Hastie *et al.*, 2011; Hastie *et al.*, 2012). Por lo tanto, las diferencias estructurales entre los dominios ExoN virales, aunado a sus distintas posiciones en el árbol de estructuras (Fig. 6) y las distintas funciones que desempeñan, nos sugieren que estas proteínas tienen orígenes distintos. El secuestro del dominio ExoN por parte de los virus de RNA, aunado a la amplia distribución de virus de RNA en eucariontes y su falta de representación en bacterias y arqueas (Campillo-Balderas *et al.*, 2015), podría ser una evidencia más de que los virus de RNA tienen un origen reciente.

A



B

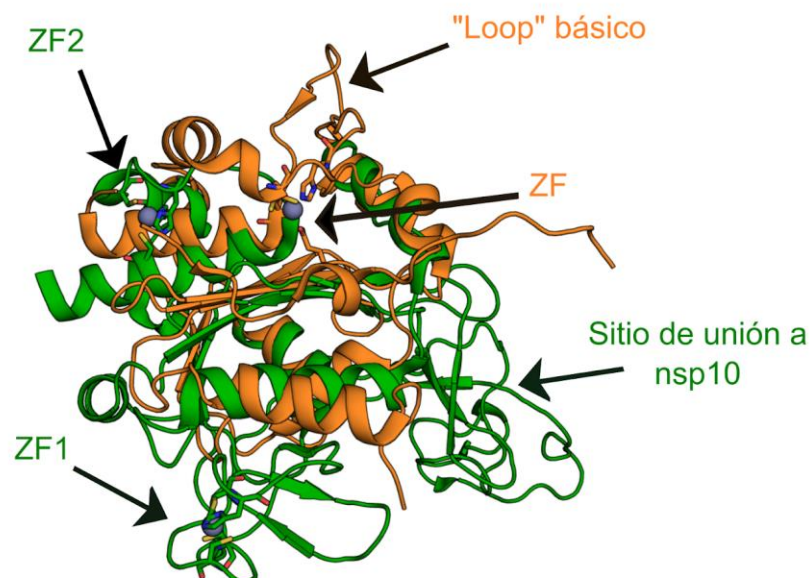


Figura 7. Comparaciones estructurales de exonucleasas de la familia DnaQ-like. (A) Conservación estructural entre las exonucleasas de la familia DnaQ-like. α -hélices resaltadas en rojo; hojas β en amarillo; y elementos de conexión en verde. (B) Superposición estructural entre la ExoN del LASV (ID de PDB: 3Q7V), naranja; y el dominio ExoN del SARS-CoV (ID de PDB: 5C8U), verde. Imagen tomada de Cruz-González *et al.*, 2021.

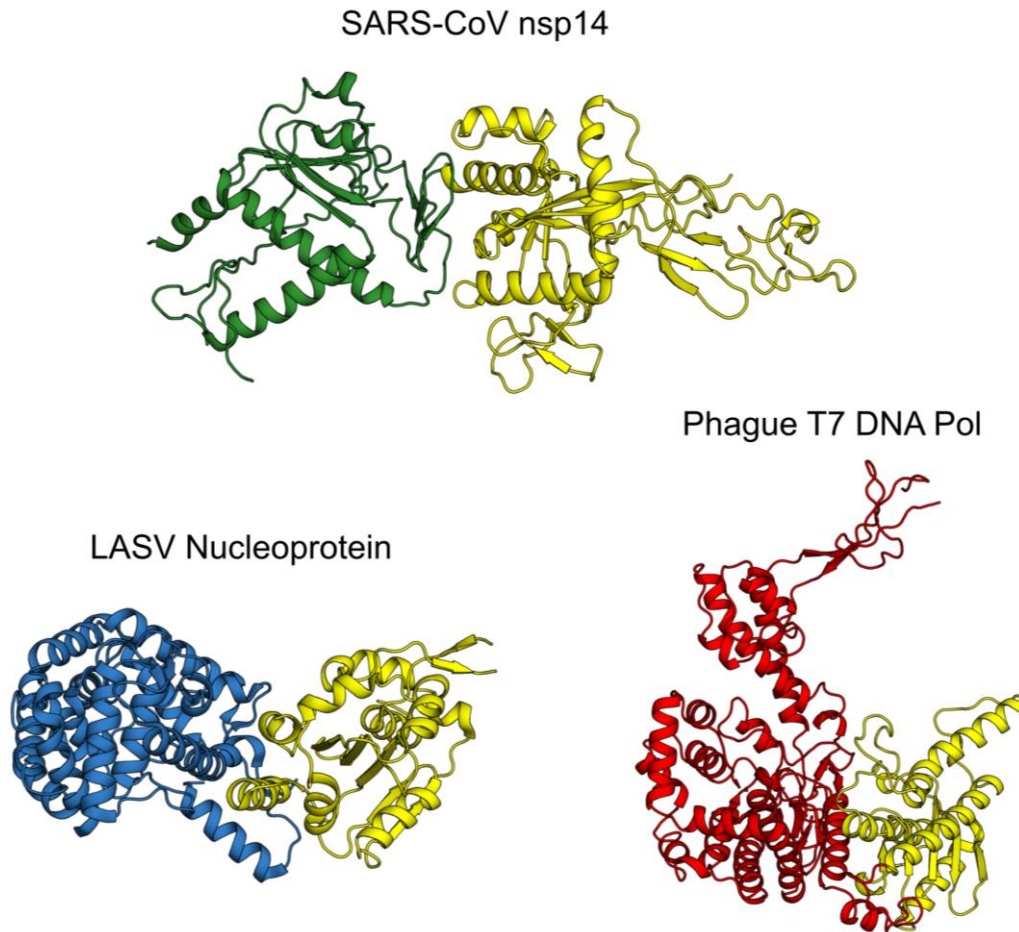


Figura 8. Estructuras de las exonucleasas virales con sus dominios aledaños. En amarillo, se resalta al dominio ExoN de las distintas proteínas virales. Para la nsp14 del SARS-CoV se resalta en verde el dominio N7-MTase, para la nucleoproteína del LASV se resalta en azul al “core” de la NP y para el fago T7 se resalta en rojo a la DNA polimerasa. Imagen producida con PyMOL (The PyMOL Molecular Graphics System, Version 1.8.2.1 Schrödinger, LLC).

Por otra parte, en el árbol de estructuras (Fig. 6), las ExoNs de los fagos de dsDNA se agrupan junto con otras ExoN de tipo DEDDy celulares, las cuales participan en procesos de edición de DNA. En estos fagos, el dominio ExoN está unido covalentemente a su polimerasa (Fig. 8), al igual que en la DNA polimerasa I y III de *Escherichia coli*, lo que sugiere que tanto la polimerasa como la exonucleasa fueron tomadas en un solo evento de secuestro por el ancestro de estos fagos de dsDNA. Esto indica un evento de secuestro viral independiente al de las ExoNs de arenavirus y nidovirus, lo cual es consistente con sus diferentes ubicaciones en el árbol de estructuras. Esto se ve reforzado por el hecho de que pertenecen al subgrupo DEDDy en lugar del subgrupo DEDDh (Zuo & Deutscher, 2001), donde se encuentran las ExoNs del SARS-CoV y de arenavirus. Este árbol sugiere fuertemente que estas exonucleasas virales han desarrollado diversas funciones después de tres eventos de secuestro viral claramente independientes.

Curiosamente, la peculiar Maelstrom (MAEL; que tiene la topología típica $\beta 1$ - $\beta 2$ - $\beta 3$ - αA - $\beta 4$ - αB - $\beta 5$ - αC del núcleo catalítico de las exonucleasas DnaQ-like, pero carece del motivo de secuencia DEDD) posee un ZF diferente cuyos residuos de unión son E213, C312, H234 y

C444, y que está presente dentro del dominio endoribonucleasa, al igual que en el dominio ExoN de los arenavirus (Hastie *et al.*, 2011; Matsumoto *et al.*, 2015). Estas similitudes estructurales y sus posiciones en el árbol de estructuras (Fig. 6) sugieren una relación evolutiva bastante estrecha entre el dominio ExoN de arenavirus y Maelstrom (*vide infra*).

4.2 Análisis evolutivo del dominio ExoN en nidovirus

La capacidad viral para cruzar las barreras taxonómicas e infectar nuevas especies está bien establecida, pero toda la evidencia disponible indica que los nidovirus están restringidos a *Animalia* (Tabla 1), lo que sugiere que el orden *Nidovirales* se originó tarde en la historia de la biósfera. Dado que la aparición de los animales ocurrió en algún momento hace unos 750 millones de años, los datos disponibles indican que el secuestro del dominio ExoN en nidovirus tuvo lugar durante el Proterozoico tardío, antes de la diversificación de estos virus, y pudo haber jugado un papel clave en su éxito evolutivo.

Como se muestra en la Figura 9 y la Tabla 1, todos los nidovirus con genomas mayores a 20 kb están dotados del dominio ExoN, en el que la pentada catalítica está altamente conservada. De igual forma, la presencia de residuos de histidina y cisteína altamente conservados, indica que todos estos virus de RNA poseen dos ZF en el dominio ExoN. La única excepción conocida es el PSCNV, que infecta a planarias y pertenece a la familia *Mononidoviridae*, en el que sólo se puede identificar el ZF1 en el dominio ExoN, mientras que los residuos que podrían corresponder al ZF2 no se pueden asignar con seguridad (Fig. 9). Esto es bastante interesante, puesto que se ha demostrado que el ZF2 es esencial para la replicación y la correcta función del dominio ExoN, al menos para los betacoronavirus SARS-CoV y MERS-CoV (Ma *et al.*, 2015; Ogando *et al.*, 2020). De hecho, el PSCNV tiene varias diferencias genómicas y moleculares importantes con otros nidovirus (Tabla 2), lo cual podría explicar la ausencia del ZF2.

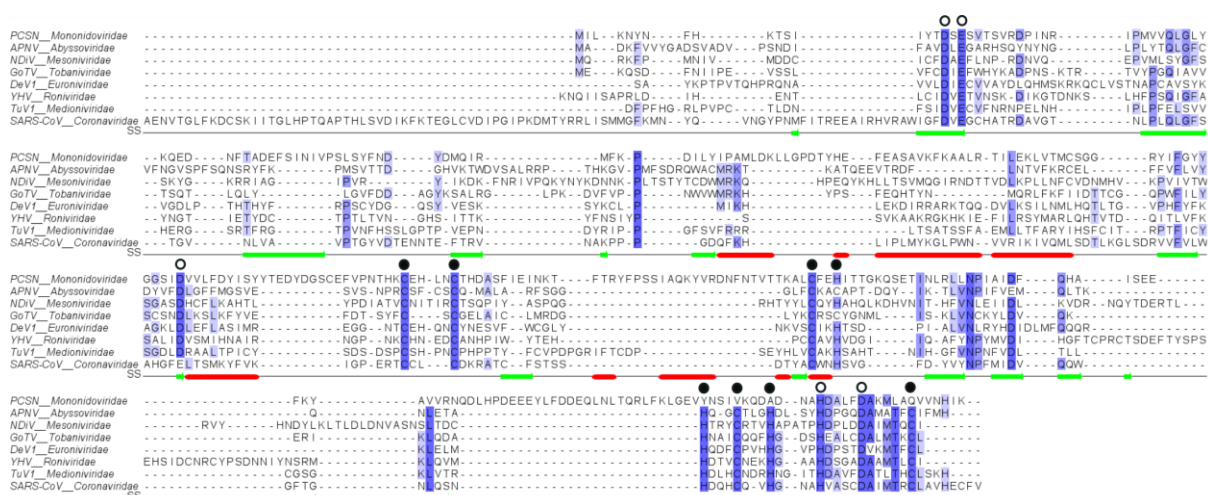


Figura 9. Alineación múltiple de secuencias del dominio ExoN de nidovirus. Los motivos de secuencia conservados Exo I (DE), Exo II (D/E) y Exo III (Dh) se señalan con círculos sin relleno. El motivo de unión de zinc 1 (ZF1, CCCH/C) y el motivo de unión de zinc 2 (ZF2, HCHC) se señalan con círculos rellenos. La estructura secundaria (SS) se indica como: rojo, α-hélices; verde, hojas β. El espectro en azules indica la conservación en las secuencias. Abreviatura de secuencias virales: PCSN, Nidovirus de las células secretoras de planaria; APNV,

Nidovirus de *Aplysia californica*; NDiV, Nam Dinh virus; GoTV, Torovirus de cabra; DeV1, Decronivirus 1; YHV, Yellow head virus; y TuV1, Turrinivirus 1. Imagen tomada de Cruz-González *et al.*, 2021.

Tabla 2. Diferencias entre el nidovirus de las células secretoras de planaria y otros nidovirus, basado en Saberi *et al.*, 2018.

	Otros nidovirus	PSCNV
Tamaño del genoma	12.7 a 35.9 kb	41.1 kb
Exonucleasa	Presente, con la excepción de los subórdenes <i>Amidovirineae</i> y <i>Nanidovirineae</i> . El dominio ExoN contiene dos dedos de zinc (ZF1 y ZF2).	Presente, pero en lugar del motivo 2H2C, el dominio ExoN tiene un motivo ES2Q en el ZF1, y probablemente carece del ZF2.
Tamaño del ORF1b	Variación mínima y pocos dominios sin anotar.	Desproporcionalmente largo e incluye varios dominios sin anotar.
Organización de los ORFs	Un arreglo de múltiples ORFs. ORF1a y ORF1b sobrelapados y múltiples ORFs al extremo 3' (3'ORFS)	Un solo ORF sobrelapado con otros ORFs pequeños en distintos marcos de lectura.
Genes adicionales	Carece de genes que codifican para ribonucleasa T2, anquirinas y fibronectinas de tipo 2.	Contiene genes que codifican para ribonucleasa T2, anquirinas y fibronectinas de tipo 2.
Diferencias en la 3CLpro	Cisteína como nucleófilo catalítico en la 3CLpro. “Pocket” de unión a sustrato con Histidina.	Ser-His-Asp como la triada catalítica en la 3CLpro. “Pocket” de unión a sustrato con Valina.
Motivo C de la RdRp	Serina en la firma SDD específica de nidovirus.	Serina sustituida por Glicina en la firma (GDD).
Dominio NiRAN	Todos los nidovirus retienen los siete residuos invariables.	Sólo seis de los siete residuos invariables se conservan.

Sin duda, el PSCNV es un caso interesante y debe ser motivo de análisis experimentales adicionales. Por un lado, es el virus de RNA con el genoma lineal más grande caracterizado al día de hoy, con una longitud de 41.1 kilobases (Saberi *et al.* 2018), lo que es notable para las entidades biológicas basadas en RNA con una tasa de mutación tan elevada (Duffy, 2018). Por lo anterior, este virus no parece tener las características de un nidovirus típico, y plantea la incógnita de cuál podría ser el mecanismo por el cual este virus de RNA inusualmente grande puede mantener la integridad y estabilidad de su genoma en la replicación viral.

La conservación en secuencia del dominio ExoN entre la mayoría de los nidovirus destaca su importancia. Cabe resaltar que en los coronavirus la secuencia del motivo catalítico en el dominio ExoN es DEED, en lugar del típico motivo DEDD que se encuentra en las enzimas celulares y otros dominios ExoN en nidovirus (Figura 10) (Ma *et al.*, 2015). Sin embargo, los estudios de mutagénesis con MERS-CoV mostraron que la mutante E191D (DEDD) es estable en cultivo celular y que la proteína mutada exhibe un nivel de actividad similar al tipo salvaje (DEED) (Ogando *et al.*, 2020). Aunque se necesitan más estudios que atiendan este tema, el cambio conservativo en el motivo de secuencia (Glu a Asp, y viceversa) sugiere que la actividad enzimática podría no verse afectada y, en consecuencia, mantener la estabilidad y fidelidad viral.

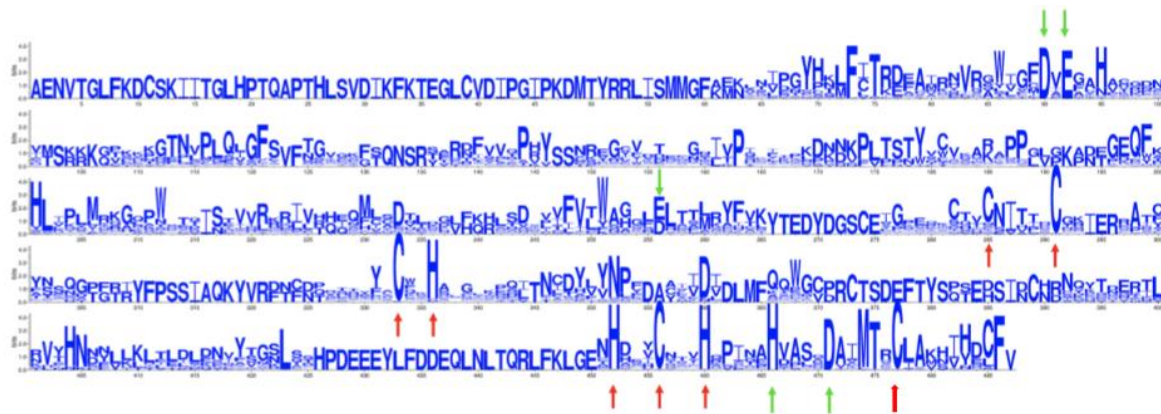


Figura 10. Conservación de la secuencia entre las ExoN de los nidovirus. Los motivos Exo I (DE), Exo II (D/E) y Exo III (Dh) se señalan con flechas verdes, mientras que los aminoácidos que forman el ZF1 (CCCH/C) y el ZF2 (HCHC) se señalan con flechas rojas. El logo se hizo con WebLogo 3 (<http://weblogo.threeplusone.com/>).

Debido al alto nivel de conservación de la RdRp, construimos un árbol filogenético con las secuencias de la RdRp de nidovirus con el objetivo de compararlo con el árbol filogenético del dominio ExoN. El árbol filogenético del dominio ExoN en nidovirus (Fig. 11) muestra a cada una de las familias formando su propio clado, con valores altos de bootstrap cerca de los bordes (ver Métodos). Solo las familias *Roniviridae* y *Euroniviridae* están agrupadas con valores altos más cercanos a la raíz del árbol. El *Nidovirus de Aplysia californica* (APNV, Familia *Abyssoviridae*) se ubica como grupo hermano de la familia *Coronaviridae*, mientras que el PSCNV se encuentra en la raíz del árbol formando su propio clado. El árbol filogenético de RdRp (Fig. 11) exhibe una topología similar al árbol de ExoN, con valores altos de bootstrap desde la raíz hasta los bordes (ver Métodos). La mayoría de las familias virales forman su propio clado; sin embargo, algunas familias se agrupan como grupos hermanos, i. e., *Coronaviridae* con *Tobaniviridae*, *Mesoniviridae* con *Medioniviridae* y *Roniviridae* con *Euroniviridae*. En este árbol, tanto el APNV como el PSCNV se derivan como clados independientes. Al comparar los árboles, las únicas familias que siempre se encuentran agrupadas son *Mesoniviridae* y *Medioniviridae*, pertenecientes al suborden *Mesnidovirineae* y las familias *Roniviridae* y *Euroniviridae*, pertenecientes al suborden *Ronidovirineae* (Siddell *et al.*, 2019). Cabe resaltar que, aunque no observamos las mismas agrupaciones entre la mayoría de las familias al comparar los árboles (Fig. 11), la topología general es similar en ambos árboles, con las diferentes familias virales agrupadas consistentemente.

ExoN



RdRp

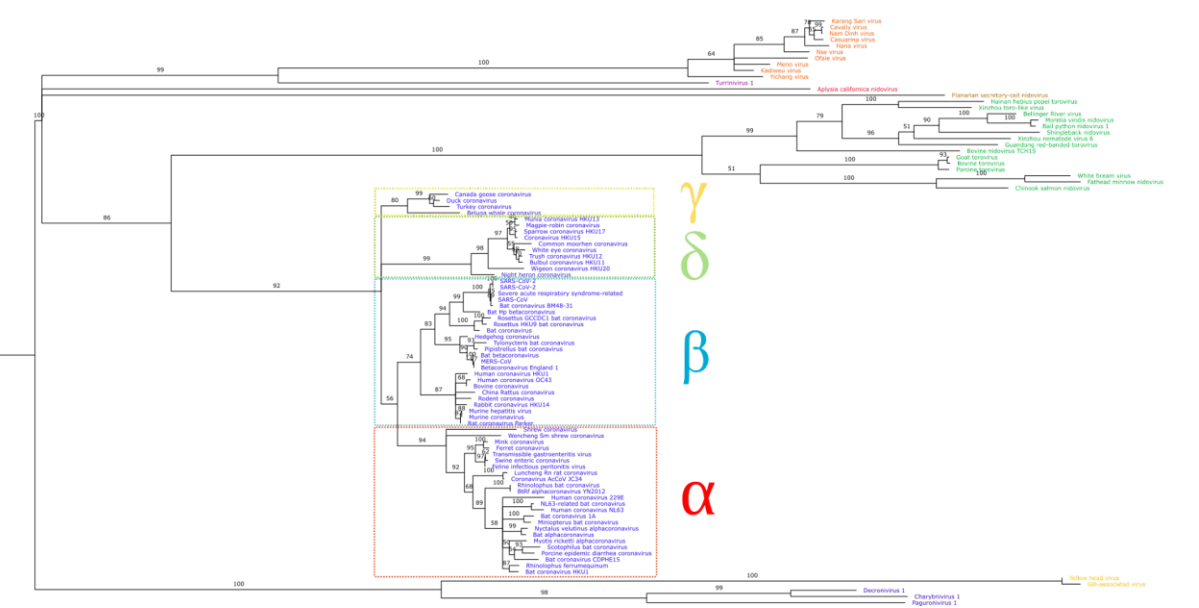


Figura 11. Filogenias de máxima probabilidad para la RdRp y la ExoN en nidovirus. Los colores en los nombres de las especies virales corresponden a las familias a las que pertenecen. La familia *Coronaviridae* se ha dividido en sus correspondientes géneros: alphacoronavirus (α), betacoronavirus (β), deltacoronavirus (δ), and gammacoronavirus (γ). Las ramas con valores de bootstrap <50% se han colapsado. Imagen tomada de Cruz-González *et al.*, 2021.

Los patrones de relación evolutiva entre los grupos virales en las filogenias de RdRps y ExoNs exhiben una topología similar. El alto nivel de conservación a nivel de secuencia (95% de identidad) entre el dominio ExoN del SARS-CoV y del SARS-CoV-2, es similar al observado cuando se comparan las secuencias de las proteínas involucradas en el complejo RdRp (nsp8, nsp9 y nsp12) (96% de identidad) (Yoshimoto, 2020). La gran similitud entre la topología del árbol de ExoN con el árbol de la altamente conservada RdRp indica que el dominio ExoN también es de suma importancia para estos virus. Esto es un indicio de que en entidades basadas en RNA, como los nidovirus, los procesos de edición juegan un papel clave en el mantenimiento de la integridad y estabilidad del genoma.

4.3 Los dedos de zinc están presentes en las ExoN de virus de RNA, pero no en las enzimas celulares y de fagos

Los ZF pueden describirse como un grupo de “andamios” estables cuya estructura es mantenida por un ión de zinc. Varían en secuencia y estructura, lo cual se ve reflejado en la coordinación del ión Zn^{2+} con residuos de cisteína y/o histidina y la forma en la que el ZF interactúa con otras moléculas (Malgieri *et al.*, 2015; Eom *et al.*, 2016; Cassandri *et al.*, 2017). Normalmente, los ZF actúan como módulos de interacción y se unen a varias moléculas, incluidos ácidos nucleicos, proteínas, lípidos y compuestos pequeños (Andreini *et al.*, 2011; Cassandri *et al.*, 2017).

Un análisis de las estructuras y la literatura de la familia de exonucleasas DnaQ-like mostró que las únicas dos exonucleasas celulares con dominios de unión a zinc (ZBD) son las enzimas “cell death-related nuclease 4” (CRN-4) (Hsiao *et al.*, 2009) y “target of Egr1” (Deng *et al.*, 2019). Aunque es importante mencionar que el ZBD de la enzima CRN-4 no se encuentra en la misma posición (Fig. 12), y que aún no se tiene una estructura terciaria de la enzima “target of Egr1”. De igual forma, no se identificaron ZF en las estructuras de los dominios ExoN de edición de *Escherichia phage RB69* (*Myoviridae*), *Bacillus virus phi29* (*Podoviridae*), *Escherichia virus T7* (*Autographiviridae*), ni de *Escherichia virus T4* (*Myoviridae*). Asimismo, la alineación múltiple de secuencias de exonucleasas virales y celulares pertenecientes a la familia DnaQ-like, demuestra la falta de ZF tanto en los fagos de DNA como en los dominios ExoN celulares (Fig. 13). Esto confirma y amplía las conclusiones de Snijder *et al.*, (2003) de que tanto las exonucleasas de los fagos de DNA y las exonucleasas celulares pertenecientes a la familia DnaQ-like carecen de ZF dentro del dominio exonucleasa.

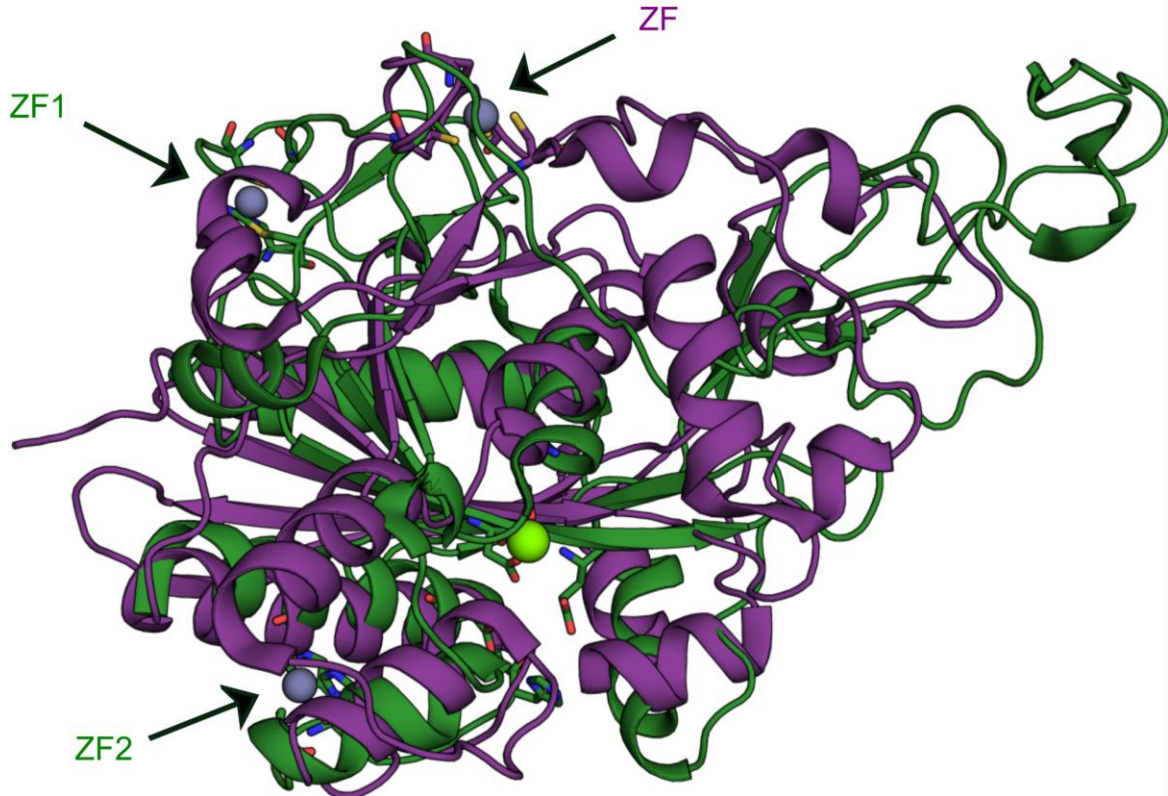


Figura 12. Comparación estructural del dominio ExoN en CoVs y la enzima CRN-4. Superposición estructural entre la enzima cell death-related nuclease 4 (ID de PDB: 3CG7), violeta; y el dominio ExoN del SARS-CoV (ID de PDB: 5C8U), verde. La localización de los sitios de unión a Zinc se indica con flechas. Imagen producida con PyMOL (The PyMOL Molecular Graphics System, Version 1.8.2.1 Schrödinger, LLC).

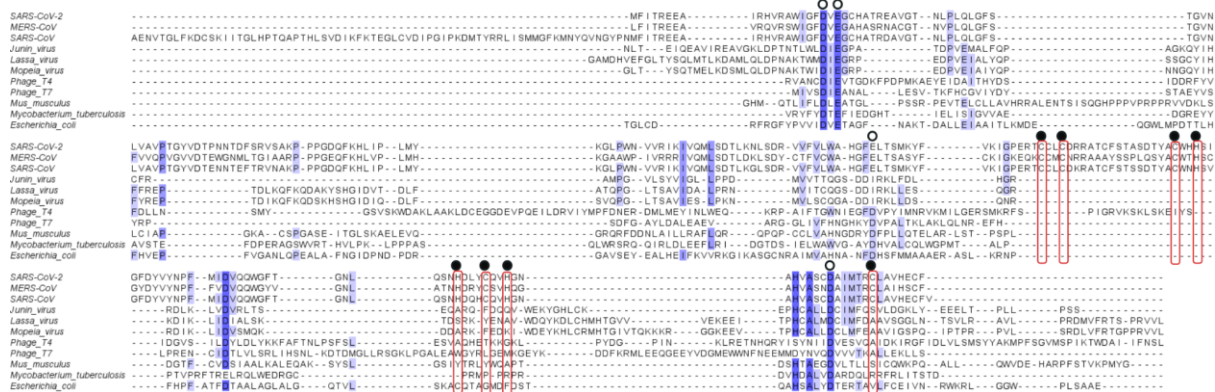


Figura 13. Alineación de secuencia guiada por estructura de algunos dominios ExoN virales y celulares. Los aminoácidos de los motivos de secuencia conservados Exo I (DE), Exo II (D/E) y Exo III (D) se señalan con círculos sin relleno. Los aminoácidos del motivo de unión a zinc 1 (ZF1, CCCH/C) y del motivo de unión a zinc 2 (ZF2, HCHC) presentes sólo en estructuras virales se señalan con círculos rellenos y enmarcados en rojo. El espectro en azules indica la conservación en las secuencias.

En contraste, las exonucleasas de virus de RNA como la del SARS-CoV y la del LASV, están dotadas con dos y un ZF, respectivamente (Fig. 7B) (Ma *et al.*, 2015; Hastie *et al.*, 2012). En los nidovirus, los ZF parecen desempeñar un papel esencial en la estabilidad de la estructura y quizás también en la catálisis. Se sabe que los estudios de mutagénesis dirigidos al ZF1 del MHV (C206A y C209A), el TGEV (C210H) y el MERS-CoV (C210H) afectan la replicación

del genoma (Eckerle *et al.*, 2007; Becares *et al.*, 2016; Ogando *et al.*, 2020). Además, se encontró que las mutantes del ZF1 para el *White bream virus* (C6101A, C6104A, C6122A y C6125A) y el SARS-CoV, carecen de actividad nucleolítica y no pueden expresarse como proteínas solubles, respectivamente (Ma *et al.*, 2015; Durzynska *et al.*, 2018). También, las mutantes del ZF2 para el SARS-CoV (C261A y H264R) y el MERS-CoV (C261A y H264R) abolieron la actividad enzimática y la replicación, respectivamente (Ma *et al.*, 2015; Ogando *et al.*, 2020). La conservación de los ZF en nidovirus (Fig. 9) y los estudios de mutagénesis mencionados anteriormente, indican que desempeñan un papel estructural clave en la función de la ExoN viral. La conservación de la secuencia y los rasgos de los ZF sugieren una adquisición monofilética del dominio ExoN, que tuvo lugar en la población de nidovirus ancestrales antes de su división en varias familias.

La presencia de los dos ZF en prácticamente todas las ExoN de nidovirus (Fig. 9), contrasta fuertemente con la falta de ZF en las ExoN celulares, y abre la posibilidad para desarrollar estrategias terapéuticas específicas en contra de estos virus de RNA. Además de la vital función de edición del dominio ExoN, se han reportado posibles actividades adicionales al dominio ExoN, como la evasión del sistema inmune en el TGEV (Becares *et al.*, 2016), replicación y síntesis de RNA en MERS-CoV y SARS-CoV-2 (Ogando *et al.*, 2020), y recombinación en MHV, MERS-CoV y SARS-CoV-2 (Gribble *et al.*, 2021). Esto indica que el dominio ExoN juega un papel muy importante al asegurar la replicación y viabilidad de estos virus. Incluso, trabajos basados en similitud estructural y de secuencia con homólogos celulares (Cruz-González *et al.*, 2020), y ensayos celulares *in vitro* (Ngwe Tun *et al.*, 2021); han señalado al dominio ExoN como un blanco terapéutico. En consecuencia, se propone considerar al dominio ExoN de la nsp14 en futuros estudios para el desarrollo y pruebas de fármacos antivirales, tomando en cuenta a los ZF presentes en los dominios virales.

4.4 El posible origen celular del dominio ExoN de los arenavirus

Maelstrom (MAEL) es una endoribonucleasa conservada presente en animales y protistas. Esta enzima se encuentra relacionada con la regulación de ciertos elementos genéticos endógenos como los retrotransposones (Zhang *et al.*, 2008; Pek *et al.*, 2012; Matsumoto *et al.*, 2015). Los ensayos de mutación que reducen la actividad entre los ortólogos de MAEL indican que está involucrada en la unión al ssRNA celular y no en la catálisis. En particular, se ha descrito a MAEL como una proteína de unión a RNA que interactúa con RNAs asociados a Piwi (piRNA), protegiendo el genoma de los transposones al reprimirlos en las gónadas de animales (Matsumoto *et al.*, 2015; Sato & Siomi, 2015). MAEL parece ser muy similar en estructura al dominio ExoN de los arenavirus, pero carece del motivo DEDD. Sin embargo, estudios evolutivos realizados por Zhang *et al.* (2008) demostraron que el dominio MAEL en protistas como *Entamoeba histolytica*, *Trypanosoma brucei* y *Leishmania braziliensis* está dotado del motivo DEDD. Interesantemente, estudios por Chen *et al.* (2015) han informado de la existencia de un dominio MAEL en la ameba con el motivo de secuencia DEDD, una tetrada ECHC característica de MAEL, y una posible actividad exoribonucleasa.

La similitud estructural entre el dominio ExoN de arenavirus y MAEL, apoya la posibilidad de un origen celular del dominio ExoN de arenavirus (Fig. 14). Como sugirieron Sato y Siomi (2015), MAEL pudo haber evolucionado de una exonucleasa DEDDh a una tétrada ECHC al cambiar los residuos catalíticos. Por lo tanto, los arenavirus podrían haber secuestrado la enzima antes de la aparición de la función de unión al RNA y la pérdida de la actividad enzimática como exoribonucleasa. Un posible mecanismo de evolución para las pseudonucleasas Ssd1 sugerido por Ballou *et al.*, (2020) y siguiendo las ideas de Jeffery (2019), en las cuales las pseudoenzimas surgen de enzimas que perdieron su capacidad catalítica, pero se ven implicadas en otra función; lo anterior podría explicar la transición evolutiva de MAEL, reforzando la hipótesis del origen celular del dominio ExoN en arenavirus.

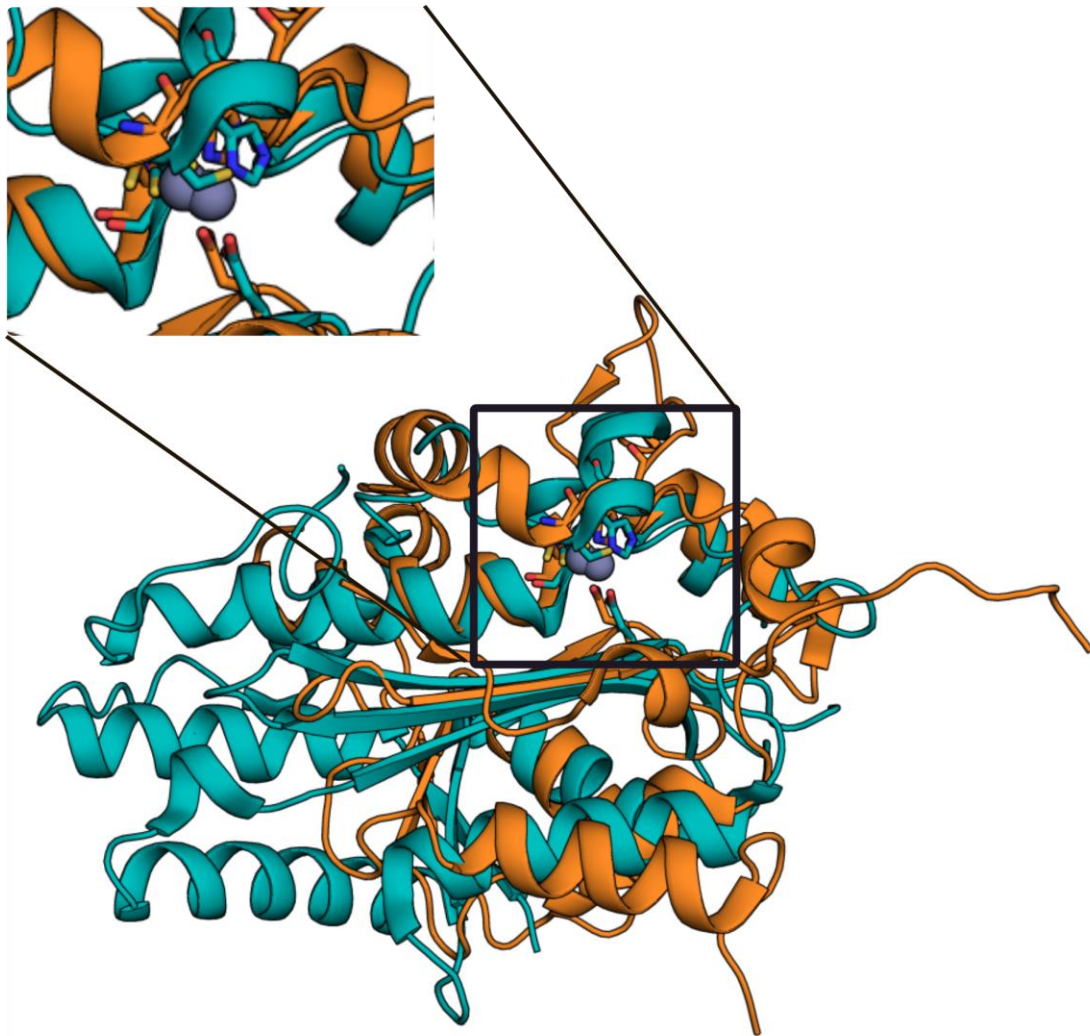


Figura 14. Superposición estructural entre el dominio ExoN del LASV y el dominio Maelstrom de *Drosophila melanogaster*. ExoN del LASV (ID de PDB: 3Q7V) en naranja; MAEL de *Drosophila melanogaster* (ID de PDB: 4YBG) en azul. En el acercamiento, se destaca la presencia del ZF (ECHC) presente en ambas proteínas. Imagen tomada de Cruz-González *et al.*, 2021.

5. Limitaciones

Al realizar este trabajo se encontraron algunas limitaciones. En primer lugar, es importante señalar el escaso número de estructuras terciarias de proteínas, lo cual es una limitante en el número de la muestra. Esto se debe principalmente a que los métodos para obtener las estructuras terciarias de las proteínas son demandantes y requieren de tiempo. En segundo lugar, los métodos como la elaboración del árbol basado en la comparación de estructuras terciarias han comenzado a utilizarse más frecuentemente en los últimos años, y su desarrollo e implementación siguen en discusión, por ejemplo, aún se carece de pruebas estadísticas como Bootstrap o Jackknife para medir qué tan robustos son los resultados que arrojan. Por último, cabe resaltar que, la elaboración de árboles filogenéticos basados en secuencias de estructura primaria de virus de RNA (como el elaborado aquí para la RdRp y ExoN) tienen sus limitaciones al ser secuencias que divergen mucho debido a la alta tasa de mutación presentada por estos virus.

6. Conclusiones

Las propiedades catalíticas, reguladoras y estructurales de las moléculas de RNA y los ribonucleótidos, combinadas con su ubicuidad en los procesos celulares, sugieren que desempeñaron un papel clave en la evolución temprana y quizás en el origen de la vida misma. Hoy en día, las únicas entidades biológicas conocidas con genomas de RNA se encuentran en la amplia gama de virus y viroides de RNA. Aunque los virus de RNA pueden proporcionar información sobre la estructura y evolución de los genomas celulares tempranos antes de la aparición del DNA, es poco probable que sean descendientes directos de formas de vida primitivas basadas en RNA. Los resultados presentados aquí sugieren que los genes que codifican para el dominio ExoN han sido secuestrados por virus al menos tres veces: una por fagos de DNA (RB69, ϕ 29, T4, T7) e, independientemente, por arenavirus y por nidovirus, los cuales son virus de RNA.

Como lo demuestran los análisis estructurales realizados en este trabajo las ExoN virales son homólogas, sin embargo, tienen un origen distinto y funciones diferentes. La actividad de edición de RNA por parte del dominio ExoN en nidovirus y la función de evasión inmune en arenavirus resalta la versatilidad de estas enzimas, en las que algunos cambios estructurales pueden conducir a una función distinta. El análisis de secuencias realizado aquí nos señala la conservación de los residuos que coordinan los iones metálicos y forman los ZF. Esto demuestran la importancia de los ZF en la estabilización estructural de las exonucleasas en entidades basadas en RNA como arenavirus y nidovirus. El caso de la enzima celular MAEL es, por otro lado, bastante interesante. Los análisis de homología estructural del presente estudio nos sugieren que el plegamiento del dominio de tipo exonucleasa cambió su función a lo largo de la evolución, desde la edición del DNA hasta la unión al RNA, perdiendo su actividad catalítica en el proceso.

Finalmente, cabe resaltar el uso de estructuras terciarias en este tipo de estudios, puesto que, debido a su alto grado de conservación, nos permite comparar y comprender de mejor forma las relaciones evolutivas entre entidades biológicas basadas en DNA y aquellas basadas en RNA, lo cual no sería posible utilizando la estructura primaria.

Referencias

- Andreini, C., Bertini, I., & Cavallaro, G. (2011). Minimal functional sites allow a classification of zinc sites in proteins. *PLoS ONE*, 6(10), e26325. <https://doi.org/10.1371/journal.pone.0026325>
- Andreeva, A., Kulesha, E., Gough, J., & Murzin, A. G. (2020). The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic acids research*, 48(D1), D376–D382. <https://doi.org/10.1093/nar/gkz1064>
- Ballou, E. R., Cook, A. G., & Wallace, E. (2020). Repeated evolution of inactive pseudonucleases in a fungal branch of the Dis3/RNase II family of nucleases. *Molecular biology and evolution*, msaa324. Advance online publication. <https://doi.org/10.1093/molbev/msaa324>
- Baltimore D. (1971). Expression of animal virus genomes. *Bacteriol Rev.* 35(3):235–241.
- Balaji, S., & Srinivasan, N. (2007). Comparison of sequence-based and structure-based phylogenetic trees of homologous proteins: Inferences on protein evolution. *Journal of biosciences*, 32(1), 83–96. <https://doi.org/10.1007/s12038-007-0008-1>
- Becares, M., Pascual-Iglesias, A., Nogales, A., Sola, I., Enjuanes, L., & Zuñiga, S. (2016). Mutagenesis of Coronavirus nsp14 Reveals Its Potential Role in Modulation of the Innate Immune Response. *Journal of virology*, 90(11), 5399–5414. <https://doi.org/10.1128/JVI.03259-15>
- Belshaw, R., Gardner, A., Rambaut, A., & Pybus, O. G. (2008). Pacing a small cage: mutation and RNA viruses. *Trends in ecology & evolution*, 23(4), 188–193. <https://doi.org/10.1016/j.tree.2007.11.010>
- 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. *Proceedings of the National Academy of Sciences of the United States of America*, 109(24), 9372–9377. <https://doi.org/10.1073/pnas.1201130109>
- Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Dalenberg, K., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Kalro, T., Liang, Y., Lowe, R., ... Zardecki, C. (2019). RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic acids research*, 47(D1), D464–D474. <https://doi.org/10.1093/nar/gky1004>
- Campillo-Balderas, J. A., Lazcano, A. & Becerra, A (2015). Viral Genome Size Distribution Does not Correlate with the Antiquity of the Host Lineages. *Frontiers in Ecology and Evolution*, 3, 143. <https://doi.org/10.3389/fevo.2015.00143>
- Campillo-Balderas J.A. & Becerra A. (2020). The Origin and Early Evolution of Life on Earth: A Laboratory in the School of Science. In: Souza V., Segura A., Foster J. (eds) *Astrobiology and Cuatro Ciénegas Basin as an Analog of Early Earth*. Cuatro Ciénegas Basin: An Endangered Hyperdiverse Oasis. Springer, Cham. https://doi.org/10.1007/978-3-030-46087-7_10

- Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., Melino, G., & Raschellà, G. (2017). Zinc-finger proteins in health and disease. *Cell death discovery*, 3, 17071. <https://doi.org/10.1038/cddiscovery.2017.71>
- Černý, J., Černá Bolfiková, B., de A Zanotto, P. M., Grubhoffer, L., & Růžek, D. (2015). A deep phylogeny of viral and cellular right-hand polymerases. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 36, 275–286. <https://doi.org/10.1016/j.meegid.2015.09.026>
- Chen, K. M., Campbell, E., Pandey, R. R., Yang, Z., McCarthy, A. A., & Pillai, R. S. (2015). Metazoan Maelstrom is an RNA-binding protein that has evolved from an ancient nuclease active in protists. *RNA (New York, N.Y.)*, 21(5), 833–839. <https://doi.org/10.1261/rna.049437.114>
- Cruz-González, A., Aguilar-Cámara, A., Álvarez-Carreño, C., Becerra, A., Campillo-Balderas, J. A., Cisneros-Martínez, A., Córdova-Villalba, G., Cottom-Salas, W., Cruz-González, C., Hernández-Morales, R., Miranda-Pérez, I., Muñoz-Velasco, I., Quintero-Navarro, L. A., Rivas, M., Cabañas, N., Sánchez-Rocha, A. C., Vázquez-Salazar, A., Jácome, R., & Lazcano, A. (2020). SARS-CoV-2 Exoribonuclease (NSP14) As a Potential Therapeutic Target to Treat COVID-19. Available at SSRN: <https://ssrn.com/abstract=3632018>
- Cruz-González, A., Muñoz-Velasco, I., Cottom-Salas, W., Becerra, A., Campillo-Balderas, J., A., Hernández-Morales, R., Vázquez-Salazar, A., Jácome, R., & Lazcano A. (2021). Structural analysis of viral ExoN domains reveals polyphyletic hijacking events. *PLoS ONE*, 16(3), e0246981. <https://doi.org/10.1371/journal.pone.0246981>
- Deng, T., Huang, Y., Weng, K., Lin, S., Li, Y., Shi, G., Chen, Y., Huang, J., Liu, D., Ma, W., & Songyang, Z. (2019). TOE1 acts as a 3' exonuclease for telomerase RNA and regulates telomere maintenance. *Nucleic acids research*, 47(1), 391–405. <https://doi.org/10.1093/nar/gky1019>
- Doublíé, S., Tabor, S., Long, A. M., Richardson, C. C., & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*, 391(6664), 251–258. <https://doi.org/10.1038/34593>
- Drake J. W. (1999). The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. *Annals of the New York Academy of Sciences*, 870, 100–107. <https://doi.org/10.1111/j.1749-6632.1999.tb08870.x>
- Duffy, S., Shackelton, L. A., & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nature reviews. Genetics*, 9(4), 267–276. <https://doi.org/10.1038/nrg2323>
- 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>
- Durzyńska, J., & Goździcka-Józefiak, A. (2015). Viruses and cells intertwined since the dawn of evolution. *Virology journal*, 12, 169. <https://doi.org/10.1186/s12985-015-0400-7>
- Durzynska, I., Sauerwald, M., Karl, N., Madhugiri, R., & Ziebuhr, J. (2018). Characterization of a bafinivirus exoribonuclease activity. *The Journal of general virology*, 99(9), 1253–1260. <https://doi.org/10.1099/jgv.0.001120>
- Eckerle, L. D., Lu, X., Sperry, S. M., Choi, L., & Denison, M. R. (2007). High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *Journal of virology*, 81(22), 12135–12144. <https://doi.org/10.1128/JVI.01296-07>
- Eckerle, L. D., Becker, M. M., Halpin, R. A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R. L., Baric, R. S., Stockwell, T. B., Spiro, D. J., & Denison, M. R. (2010). Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS pathogens*, 6(5), e1000896. <https://doi.org/10.1371/journal.ppat.1000896>

- Eom, K. S., Cheong, J. S., & Lee, S. J. (2016). Structural Analyses of Zinc Finger Domains for Specific Interactions with DNA. *Journal of microbiology and biotechnology*, 26(12), 2019–2029. <https://doi.org/10.4014/jmb.1609.09021>
- Fan, Y., Zhao, K., Shi, Z. L., & Zhou, P. (2019). Bat Coronaviruses in China. *Viruses*, 11(3), 210. <https://doi.org/10.3390/v11030210>
- Fehr, A. R., & Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. *Methods in molecular biology (Clifton, N.J.)*, 1282, 1–23. https://doi.org/10.1007/978-1-4939-2438-7_1
- Felsenstein J. (1993). PHYLIP: Phylogeny Inference Package. Seattle, WA: University of Washington.
- Ferron, F., Subissi, L., Silveira De Morais, A. T., Le, N., 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 of the United States of America*, 115(2), E162–E171. <https://doi.org/10.1073/pnas.1718806115>
- Fisher, S. (2010). Are RNA Viruses Vestiges of an RNA World?. *Journal for General Philosophy of Science*, 41, 121–141. <https://doi.org/10.1007/s10838-010-9119-8>
- Flores, T. P., Orengo, C. A., Moss, D. S., & Thornton, J. M. (1993). Comparison of conformational characteristics in structurally similar protein pairs. *Protein science : a publication of the Protein Society*, 2(11), 1811–1826. <https://doi.org/10.1002/pro.5560021104>
- Forterre, P., & Prangishvili, D. (2009). The origin of viruses. *Research in microbiology*, 160(7), 466–472. <https://doi.org/10.1016/j.resmic.2009.07.008>
- Ganai, R. A., & Johansson, E. (2016). DNA Replication-A Matter of Fidelity. *Molecular cell*, 62(5), 745–755. <https://doi.org/10.1016/j.molcel.2016.05.003>
- Garau, G., Di Guilmi, A. M., & Hall, B. G. (2005). Structure-based phylogeny of the metallo-beta-lactamases. *Antimicrobial agents and chemotherapy*, 49(7), 2778–2784. <https://doi.org/10.1128/AAC.49.7.2778-2784.2005>
- Gilbert, W. (1986). Origin of life: The RNA world. *Nature* 319, 618. <https://doi.org/10.1038/319618a0>
- Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J., & Snijder, E. J. (2006). Nidovirales: evolving the largest RNA virus genome. *Virus research*, 117(1), 17–37. <https://doi.org/10.1016/j.virusres.2006.01.017>
- Gribble, J., Stevens, L. J., Agostini, M. L., Anderson-Daniels, J., Chappell, J. D., Lu, X., Pruijssers, A. J., Routh, A. L., & Denison, M. R. (2021). The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS pathogens*, 17(1), e1009226. <https://doi.org/10.1371/journal.ppat.1009226>
- Hastie, K. M., Kimberlin, C. R., Zandonatti, M. A., MacRae, I. J., & Saphire, E. O. (2011). Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3' to 5' exonuclease activity essential for immune suppression. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6), 2396–2401. <https://doi.org/10.1073/pnas.1016404108>
- Hastie, K. M., King, L. B., Zandonatti, M. A., & Saphire, E. O. (2012). Structural basis for the dsRNA specificity of the Lassa virus NP exonuclease. *PloS one*, 7(8), e44211. <https://doi.org/10.1371/journal.pone.0044211>
- Hodcroft, E. TreeCollapser CL4. Available online: <http://emmahodcroft.com/TreeCollapseCL.html> (accessed in December 2020).
- Holm, L., & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic acids research*, 38(Web Server issue), W545–W549. <https://doi.org/10.1093/nar/gkq366>

- Holmes C. E. (2009). The evolution and emergence of RNA viruses. Oxford University Press, Chapter 1-2.
- Hsiao, Y. Y., Nakagawa, A., Shi, Z., Mitani, S., Xue, D., & Yuan, H. S. (2009). Crystal structure of CRN-4: implications for domain function in apoptotic DNA degradation. *Molecular and cellular biology*, 29(2), 448–457. <https://doi.org/10.1128/MCB.01006-08>
- Jácome, R., Becerra, A., Ponce de León, S., & Lazcano, A. (2015). Structural Analysis of Monomeric RNA-Dependent Polymerases: Evolutionary and Therapeutic Implications. *PloS one*, 10(9), e0139001. <https://doi.org/10.1371/journal.pone.0139001>
- Jeffery C. J. (2019). The demise of catalysis, but new functions arise: pseudoenzymes as the phoenixes of the protein world. *Biochemical Society transactions*, 47(1), 371–379. <https://doi.org/10.1042/BST20180473>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T., von Haeseler, A., & Jermin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Kamtekar, S., Berman, A. J., Wang, J., Lázaro, J. M., de Vega, M., Blanco, L., Salas, M., & Steitz, T. A. (2006). The phi29 DNA polymerase:protein-primer structure suggests a model for the initiation to elongation transition. *The EMBO journal*, 25(6), 1335–1343. <https://doi.org/10.1038/sj.emboj.7601027>
- Kautz, T. F., & Forrester, N. L. (2018). RNA Virus Fidelity Mutants: A Useful Tool for Evolutionary Biology or a Complex Challenge?. *Viruses*, 10(11), 600. <https://doi.org/10.3390/v10110600>
- King, A. M. Q., Adams, M. J., Carstens, E. B., & Lefkowitz, E. J. (2012) Order - Nidovirales. *Virus Taxonomy*, Elsevier, pages 784-794, ISBN: 9780123846846. <https://doi.org/10.1016/B978-0-12-384684-6.00066-5>
- Koonin, E. V., Senkevich, T. G., & Dolja, V. V. (2006). The ancient Virus World and evolution of cells. *Biology direct*, 1, 29. <https://doi.org/10.1186/1745-6150-1-29>
- Krupovic, M., & Bamford, D. H. (2008). Virus evolution: how far does the double beta-barrel viral lineage extend?. *Nature reviews. Microbiology*, 6(12), 941–948. <https://doi.org/10.1038/nrmicro2033>
- Lakshmi, B., Mishra, M., Srinivasan, N., & Archunan, G. (2015). Structure-Based Phylogenetic Analysis of the Lipocalin Superfamily. *PloS one*, 10(8), e0135507. <https://doi.org/10.1371/journal.pone.0135507>
- Lauber, C., Goeman, J. J., Parquet, M., Nga, P. T., Snijder, E. J., Morita, K., & Gorbalenya, A. E. (2013). The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS pathogens*, 9(7), e1003500. <https://doi.org/10.1371/journal.ppat.1003500>
- Lazcano, A. (2010). Origen y evolución de los virus: ¿Genes errantes o parásitos primitivos?. *Mensaje Bioquímico*, 34, 73-84.
- 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>
- Lesk, A. M., & Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *Journal of molecular biology*, 136(3), 225–270. [https://doi.org/10.1016/0022-2836\(80\)90373-3](https://doi.org/10.1016/0022-2836(80)90373-3)
- López-García, P. & Moreira D. (2012). Viruses in Biology. *Evolution: Education and Outreach*, 5, 389–398. <https://doi.org/10.1007/s12052-012-0441-y>
- Lloyd, S. B., Kent, S. J., & Winnall, W. R. (2014). The high cost of fidelity. *AIDS research and human retroviruses*, 30(1), 8–16. <https://doi.org/10.1089/AID.2013.0153>

- Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., & Rao, Z. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proceedings of the National Academy of Sciences of the United States of America*, 112(30), 9436–9441. <https://doi.org/10.1073/pnas.1508686112>
- Malgieri, G., Palmieri, M., Russo, L., Fattorusso, R., Pedone, P. V., & Isernia, C. (2015). The prokaryotic zinc-finger: structure, function and comparison with the eukaryotic counterpart. *The FEBS journal*, 282(23), 4480–4496. <https://doi.org/10.1111/febs.13503>
- Masters P. S. (2006). The molecular biology of coronaviruses. *Advances in virus research*, 66, 193–292. [https://doi.org/10.1016/S0065-3527\(06\)66005-3](https://doi.org/10.1016/S0065-3527(06)66005-3)
- Matsumoto, N., Sato, K., Nishimasu, H., Namba, Y., Miyakubi, K., Dohmae, N., Ishitani, R., Siomi, H., Siomi, M. C., & Nureki, O. (2015). Crystal Structure and Activity of the Endoribonuclease Domain of the piRNA Pathway Factor Maelstrom. *Cell reports*, 11(3), 366–375. <https://doi.org/10.1016/j.celrep.2015.03.030>
- Mayr, G., Domingues, F. S., & Lackner, P. (2007). Comparative analysis of protein structure alignments. *BMC structural biology*, 7, 50. <https://doi.org/10.1186/1472-6807-7-50>
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular biology and evolution*, 37(8), 2461. <https://doi.org/10.1093/molbev/msaa131>
- Minskaia, E., Hertzog, T., Gorbalenya, 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. *Proceedings of the National Academy of Sciences of the United States of America*, 103(13), 5108–5113. <https://doi.org/10.1073/pnas.0508200103>
- Mir, S., Alhroub, Y., Anyango, S., Armstrong, D. R., Berrisford, J. M., Clark, A. R., Conroy, M. J., Dana, J. M., Deshpande, M., Gupta, D., Gutmanas, A., Haslam, P., Mak, L., Mukhopadhyay, A., Nadzirin, N., Paysan-Lafosse, T., Sehnal, D., Sen, S., Smart, O. S., Varadi, M., ... Velankar, S. (2018). PDBe: towards reusable data delivery infrastructure at protein data bank in Europe. *Nucleic acids research*, 46(D1), D486–D492. <https://doi.org/10.1093/nar/gkx1070>
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E., Tosatto, S., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic acids research*, 49(D1), D412–D419. <https://doi.org/10.1093/nar/gkaa913>
- Mughal, F., Nasir, A., & Caetano-Anollés, G. (2020). The origin and evolution of viruses inferred from fold family structure. *Archives of virology*, 165(10), 2177–2191. <https://doi.org/10.1007/s00705-020-04724-1>
- Nakagawa, K., Lokugamage, K. G., & Makino, S. (2016). Viral and Cellular mRNA Translation in Coronavirus-Infected Cells. *Advances in virus research*, 96, 165–192. <https://doi.org/10.1016/bs.aivir.2016.08.001>
- Nga, P. T., Parquet, M., Lauber, C., Parida, M., Nabeshima, T., Yu, F., Thuy, N. T., Inoue, S., Ito, T., Okamoto, K., Ichinose, A., Snijder, E. J., Morita, K., & Gorbalenya, A. E. (2011). Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. *PLoS pathogens*, 7(9), e1002215. <https://doi.org/10.1371/journal.ppat.1002215>
- Ngwe Tun, M. M., Morita, K., Ishikawa, T., Urata, S. (2021). The Antiviral Effect of the Chemical Compounds Targeting DED/EDh Motifs of the Viral Proteins on Lymphocytic Choriomeningitis Virus and SARS-CoV-2. *Viruses*, 13(7):1220. <https://doi.org/10.3390/v13071220>

- 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>
- Ogando, N. S., Zevenhoven-Dobbe, J. C., van der Meer, Y., Bredenbeek, P. J., Posthuma, C. C., & Snijder, E. J. (2020). The Enzymatic Activity of the nsp14 Exoribonuclease Is Critical for Replication of MERS-CoV and SARS-CoV-2. *Journal of virology*, 94(23), e01246-20. <https://doi.org/10.1128/JVI.01246-20>
- Pei, J., & Grishin, N. V. (2014). PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods in molecular biology* (Clifton, N.J.), 1079, 263–271. https://doi.org/10.1007/978-1-62703-646-7_17
- Pek, J. W., Ng, B. F., & Kai, T. (2012). Polo-mediated phosphorylation of Maelstrom regulates oocyte determination during oogenesis in *Drosophila*. *Development* (Cambridge, England), 139(24), 4505–4513. <https://doi.org/10.1242/dev.082867>
- Rambaut A. FigTree. Available from: <http://tree.bio.ed.ac.uk/software/figtree/>
- Reaney D. C. (1982). The evolution of RNA viruses. *Annual review of microbiology*, 36, 47–73. <https://doi.org/10.1146/annurev.mi.36.100182.000403>
- Reynard, S., Russier, M., Fizet, A., Carnec, X., & Baize, S. (2014). Exonuclease domain of the Lassa virus nucleoprotein is critical to avoid RIG-I signaling and to inhibit the innate immune response. *Journal of virology*, 88(23), 13923–13927. <https://doi.org/10.1128/JVI.01923-14>
- Russell, R. B., & Sternberg, M. J. (1997). Two new examples of protein structural similarities within the structure-function twilight zone. *Protein engineering*, 10(4), 333–338. <https://doi.org/10.1093/protein/10.4.333>
- Saberi, A., Gulyaeva, A. A., Brubacher, J. L., Newmark, P. A., & Gorbalenya, A. E. (2018). A planarian nidovirus expands the limits of RNA genome size. *PLoS pathogens*, 14(11), e1007314. <https://doi.org/10.1371/journal.ppat.1007314>
- Sanjuán, R., & Domingo-Calap, P. (2016). Mechanisms of viral mutation. *Cellular and molecular life sciences : CMLS*, 73(23), 4433–4448. <https://doi.org/10.1007/s00018-016-2299-6>
- Sanjuán R. & Bradwell K. (2010). The Evolution and Emergence of RNA Viruses. *Systematic Biology*, 59 (5), 610–612. <https://doi.org/10.1093/sysbio/syq049>
- Sato, K., & Siomi, M. C. (2015). Functional and structural insights into the piRNA factor Maelstrom. *FEBS letters*, 589(14), 1688–1693. <https://doi.org/10.1016/j.febslet.2015.03.023>
- Shevelev, I. V., & Hübscher, U. (2002). The 3' 5' exonucleases. *Nature reviews. Molecular cell biology*, 3(5), 364–376. <https://doi.org/10.1038/nrm804>
- Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., Qin, X. C., Li, J., Cao, J. P., Eden, J. S., Buchmann, J., Wang, W., Xu, J., Holmes, E. C., & Zhang, Y. Z. (2016). Redefining the invertebrate RNA virosphere. *Nature*, 540(7634), 539–543. <https://doi.org/10.1038/nature20167>
- Siddell, S. G., Walker, P. J., Lefkowitz, E. J., Mushegian, A. R., Adams, M. J., Dutilh, B. E., Gorbalenya, A. E., Harrach, B., Harrison, R. L., Junglen, S., Knowles, N. J., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Nibert, M., Rubino, L., Sabanadzovic, S., Sanfaçon, H., Simmonds, P., Varsani, A., ... Davison, A. J. (2019). Additional changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of Viruses (October 2018). *Archives of virology*, 164(3), 943–946. <https://doi.org/10.1007/s00705-018-04136-2>
- Sierk, M. L., & Kleywegt, G. J. (2004). Déjà vu all over again: finding and analyzing protein structure similarities. *Structure* (London, England : 1993), 12(12), 2103–2111. <https://doi.org/10.1016/j.str.2004.09.016>

- Smith, E. C., Blanc, H., Surdel, M. C., Vignuzzi, M., & Denison, M. R. (2013). Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS pathogens*, 9(8), e1003565. <https://doi.org/10.1371/journal.ppat.1003565>
- Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J., & Gorbalenya, 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)
- Snijder, E. J., Decroly, E., & Ziebuhr, J. (2016). The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Advances in virus research*, 96, 59–126. <https://doi.org/10.1016/bs.aivir.2016.08.008>
- Subbiah, S., Laurents, D. V., & Levitt, M. (1993). Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core. *Current biology : CB*, 3(3), 141–148. [https://doi.org/10.1016/0960-9822\(93\)90255-m](https://doi.org/10.1016/0960-9822(93)90255-m)
- Vázquez-Salazar, A., & Lazcano, A. (2018). Early Life: Embracing the RNA World. *Current biology : CB*, 28(5), R220–R222. <https://doi.org/10.1016/j.cub.2018.01.055>
- Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H., & Steitz, T. A. (1996). Crystal structures of an NH₂-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. *Biochemistry*, 35(25), 8110–8119. <https://doi.org/10.1021/bi960178r>
- Wirth, J., & Young, M. (2020). The intriguing world of archaeal viruses. *PLoS pathogens*, 16(8), e1008574. <https://doi.org/10.1371/journal.ppat.1008574>
- Woolhouse, M. E., Adair, K., & Brierley, L. (2013). RNA Viruses: A Case Study of the Biology of Emerging Infectious Diseases. *Microbiology spectrum*, 1(1), 10.1128/microbiolspec.OH-0001-2012. <https://doi.org/10.1128/microbiolspec.OH-0001-2012>
- Woolhouse, M., & Brierley, L. (2018). Epidemiological characteristics of human-infective RNA viruses. *Scientific data*, 5, 180017. <https://doi.org/10.1038/sdata.2018.17>
- Xia, S., & Konigsberg, W. H. (2014). RB69 DNA polymerase structure, kinetics, and fidelity. *Biochemistry*, 53(17), 2752–2767. <https://doi.org/10.1021/bi4014215>
- Yang W. (2011). Nucleases: diversity of structure, function and mechanism. *Quarterly reviews of biophysics*, 44(1), 1–93. <https://doi.org/10.1017/S0033583510000181>
- Yoshimoto F. K. (2020). The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *The protein journal*, 39(3), 198–216. <https://doi.org/10.1007/s10930-020-09901-4>
- Zhang, D., Xiong, H., Shan, J., Xia, X., & Trudeau, V. L. (2008). Functional insight into Maelstrom in the germline piRNA pathway: a unique domain homologous to the DnaQ-H 3'-5' exonuclease, its lineage-specific expansion/loss and evolutionarily active site switch. *Biology direct*, 3, 48. <https://doi.org/10.1186/1745-6150-3-48>
- Zuo, Y. & Deutscher, M. P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.* 29, 1017–1026. <https://doi.org/10.1093/nar/29.5.1017>

Anexos

Tabla S1. PDBs de las exonucleasas DnaQ-like utilizadas en este trabajo.

ID de PDB	Abreviación	Proteína	Organismo
2JGU	PfurPEXO	DNA pol Exonucleasa	<i>Pyrococcus furiosus</i>
1D9D	EcolKLFR	Klenow fragment DNA pol1	<i>Escherichia coli</i>
1FXX	EcolEXO1	Exonucleasa 1	<i>Escherichia coli</i>
1YT3	EcolRIBD	Ribonucleasa D	<i>Escherichia coli</i>
2IDO	EcolPEPS	Subunidad Epsilon (Pol III)	<i>Escherichia coli</i>
4FZX	EcolEXO	Exonucleasa X	<i>Escherichia coli</i>
4KAZ	EcolRIBT	Ribonucleasa T	<i>Escherichia coli</i>
4OKE	MtubRNAS	RNase AS	<i>Mycobacterium tuberculosis</i>
4RG8	MszeEXO1	Exonucleasa I	<i>Methylocaldum szegediense</i>
6A4A	CpsyOLRN	CpsORN Oligoribonucleasa	<i>Colwellia psychrerythraea</i>
1W0H	HsapERI1	Exorribonucleasa ERI1	<i>Homo sapiens</i>
1WLJ	HsapISG2	ISG20	<i>Homo sapiens</i>
1Y97	HsapTREX	Exonucleasa TREX2	<i>Homo sapiens</i>
2FBX	HsapWEXO	Werner syndrome helicase-exo	<i>Homo sapiens</i>
2P51	SpomPOP2	POP2 (deadenylation)	<i>Schizosaccharomyces pombe</i>
2XRI	HsapERI3	ERI1 Exorribonucleasa 3	<i>Homo sapiens</i>
3B6O	MmusTREX	TREX1 Exonucleasa	<i>Mus musculus</i>
3CG7	CelgCRN4	Cell-death related nuclease 4	<i>Caenorhabditis elegans</i>
3IAY	ScerPDEL	Subunidad Delta de la DNA pol	<i>Saccharomyces cerevisiae</i>
4CZW	NcraPAN2	Subunidad PAN2 (Poly(A)-Specific Ribonuclease)	<i>Neurospora crassa</i>
4NLB	TbruRRP6	Ribosomal RNA processing protein 6	<i>Trypanosoma brucei</i>
4PTF	ScerDEPS	Subunidad Epsilon de la DNA pol	<i>Saccharomyces cerevisiae</i>
4YBG	DromMAEL	Maelstrom	<i>Drosophila melanogaster</i>
5AF0	BmorMAEL	Maelstrom	<i>Bombyx mori</i>
5Z9X	AthaSDN1	Arabidopsis SDN1	<i>Arabidopsis thaliana</i>
6RCL	HsapREX2	Oligoribonucleasa REXO2	<i>Homo sapiens</i>
1Q9X	EphaDPOL	DNA pol RB69 (gp43)	<i>Escherichia phage RB69</i>
1T7P	Pht7EXON	DNA pol ExoN	<i>Phage T7</i>
1NOZ	Pht4EXON	DNA pol ExoN	<i>Phage T4</i>
2EX3	Ph29EXON	DNA pol ExoN	<i>Phage phi29</i>
4GV6	LvirNUCL	Nucleoproteína	<i>Lassa virus</i>
4K7E	JvirNUCL	Nucleoproteína	<i>Junin virus</i>
5LRP	MvirNUCL	Nucleoproteína	<i>Mopeia virus</i>
5C8U	SarsEXON	NSP14	<i>SARS-CoV</i>

En colores, se indica a qué grupo biológico pertenecen las proteínas: amarillo, arqueas; rojo, bacterias; azul, eucariontes; y verde, virus.

Como resultado de este trabajo de investigación se publicó el artículo “**Structural analysis of viral ExoN domains reveals polyphyletic hijacking events**” en la revista PLoS ONE (<https://doi.org/10.1371/journal.pone.0246981>), del cual fuí el primer autor.

PLOS ONE

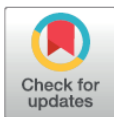
RESEARCH ARTICLE

Structural analysis of viral ExoN domains reveals polyphyletic hijacking events

Adrián Cruz-González¹, Israel Muñoz-Velasco¹, Wolfgang Cottom-Salas^{1,2}, Arturo Becerra¹, José A. Campillo-Balderas¹, Ricardo Hernández-Morales¹, Alberto Vázquez-Salazar³, Rodrigo Jácome^{1*}, Antonio Lazcano^{1,4*}

1 Facultad de Ciencias, Universidad Nacional Autónoma de México, México City, México, **2** Escuela Nacional Preparatoria, Plantel 8 Miguel E. Schulz, Universidad Nacional Autónoma de México, México City, México, **3** Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, California, United States of America, **4** El Colegio Nacional, México City, México

* alar@ciencias.unam.mx (AL); rodrigo.jacome@ciencias.unam.mx (RJ)



Abstract

Nidoviruses and arenaviruses are the only known RNA viruses encoding a 3'-5' exonuclease domain (ExoN). The proofreading activity of the ExoN domain has played a key role in the growth of nidoviral genomes, while in arenaviruses this domain partakes in the suppression of the host innate immune signaling. Sequence and structural homology analyses suggest that these proteins have been hijacked from cellular hosts many times. Analysis of the available nidoviral ExoN sequences reveals a high conservation level comparable to that of the viral RNA-dependent RNA polymerases (RdRp), which are the most conserved viral proteins. Two highly preserved zinc fingers are present in all nidoviral exonucleases, while in the arenaviral protein only one zinc finger can be identified. This is in sharp contrast with the reported lack of zinc fingers in cellular ExoNs, and opens the possibility of therapeutic strategies in the struggle against COVID-19.

OPEN ACCESS

Citation: Cruz-González A, Muñoz-Velasco I, Cottom-Salas W, Becerra A, Campillo-Balderas JA, Hernández-Morales R, et al. (2021) Structural analysis of viral ExoN domains reveals polyphyletic hijacking events. PLoS ONE 16(3): e0246981. <https://doi.org/10.1371/journal.pone.0246981>

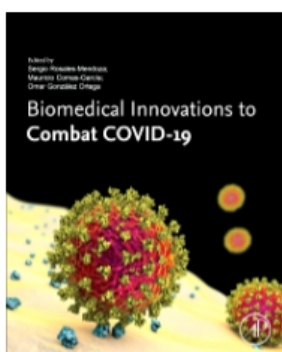
Editor: Jean-Luc EPH Darlix, "INSERM", FRANCE

Received: January 27, 2021

Accepted: February 24, 2021

Published: March 17, 2021

De igual forma, participé en la elaboración del capítulo “**Structural biology of the SARS-CoV-2 replisome: evolutionary and therapeutic implications**” para el libro “Biomedical Innovations to Combat COVID-19” (<https://www.elsevier.com/books/biomedical-innovations-to-combat-covid-19/rosales-mendoza/978-0-323-90248-9>).



Biomedical Innovations to Combat COVID-19 1st Edition

☆☆☆☆☆ Write a review

Editors: Sergio Rosales-Mendoza, Mauricio Comas-García, Omar González-Ortega

Paperback ISBN: 9780323902489

Imprint: Academic Press

Published Date: 15th October 2021

Page Count: 410

Chapter 4

Structural biology of the SARS-CoV-2 replisome: evolutionary and therapeutic implications

R. Hernández-Morales¹, A. Becerra¹, J.A. Campillo-Balderas¹, W.F. Cottom-Salas^{1,2}, A. Cruz-González¹,
R. Jácome¹, A. Lazcano^{1,3}, I. Muñoz-Velasco¹ and A. Vázquez-Salazar⁴

¹Facultad de Ciencias, Universidad Nacional Autónoma de México, CDMX, Mexico City, Mexico, ²Escuela Nacional Preparatoria, Plantel 8 Miguel E. Schulz, UNAM, CDMX, Mexico City, Mexico, ³El Colegio Nacional, CDMX, Mexico City, Mexico, ⁴Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, United States

Por último, participé como coautor en la elaboración del artículo “**Two Short Low Complexity Regions (LCRs) are Hallmark Sequences of the Delta SARS-CoV- 2 Variant Spike Protein**” publicado en la revista Scientific Reports (<https://doi.org/10.1038/s41598-022-04976-8>).

www.nature.com/scientificreports

scientific reports



OPEN **Two short low complexity regions (LCRs) are hallmark sequences of the Delta SARS-CoV-2 variant spike protein**

Arturo Becerra¹, Israel Muñoz-Velasco¹, Abelardo Aguilar-Cámara¹, Wolfgang Cottom-Salas^{1,2}, Adrián Cruz-González¹, Alberto Vázquez-Salazar³, Ricardo Hernández-Morales¹, Rodrigo Jácome¹, José Alberto Campillo-Balderas¹ & Antonio Lazcano^{1,4}✉

Low complexity regions (LCRs) are protein sequences formed by a set of compositionally biased residues. LCRs are extremely abundant in cellular proteins and have also been reported in viruses, where they may partake in evasion of the host immune system. Analyses of 28,231 SARS-CoV-2 whole proteomes and of 261,051 spike protein sequences revealed the presence of four extremely conserved LCRs in the spike protein of several SARS-CoV-2 variants. With the exception of Iota, where it is absent, the Spike LCR-1 is present in the signal peptide of 80.57% of the Delta variant sequences, and in other variants of concern and interest. The Spike LCR-2 is highly prevalent (79.87%) in Iota. Two distinctive LCRs are present in the Delta spike protein. The Delta Spike LCR-3 is present in 99.19% of the analyzed sequences, and the Delta Spike LCR-4 in 98.3% of the same set of proteins. These two LCRs are located in the furin cleavage site and HR1 domain, respectively, and may be considered hallmark traits of the Delta variant. The presence of the medically-important point mutations P681R and D950N in these LCRs, combined with the ubiquity of these regions in the highly contagious Delta variant opens the possibility that they may play a role in its rapid spread.