



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**EVOLUCIÓN DEL DOMINIO *JELLY ROLL* EN
PROTEÍNAS DE CÁPSIDE ICOSAÉDRICA DE VIRUS
DE RNA**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGA

P R E S E N T A:

ANA KAREN MENDOZA MAYA



**DIRECTOR DE TESIS:
DR. JOSÉ ALBERTO CAMPILLO BALDERAS
CIUDAD DE MÉXICO 2018**



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
José Alberto
Campillo
Balderas

2. Datos del sinodal 1

Dr
Antonio Eusebio
Lazcano-Araujo
Reyes

3. Datos del sinodal 2

Dra
Claudia Selene
Zarate
Guerra

4. Datos del sinodal 3

Dr
Alejandro Rodrigo
Jácome
Ramírez

5. Datos del sinodal 4

Dr
Enrique
Rudiño
Piñera

Agradecimientos	3
Resumen	5
Abstract	6
1. Introducción	7
1.1 Características generales de los virus de RNA	7
1.2 Origen y evolución temprana de los virus de RNA	8
1.3 Estrategias para abordar el problema sobre el origen de los virus de RNA	9
1.3.1 Filogenias basadas en secuencias de residuos de aminoácidos	9
1.3.2 Filogenias basadas en estructuras terciarias	10
1.5 Estructura de las cápsides icosaédricas	11
1.5.1 El dominio <i>jelly roll</i> en las cápsides icosaédricas	12
2. Objetivos	14
2.1 Objetivo general	14
2.2 Objetivos particulares	14
2.3 Hipótesis	14
3. Métodos	15
3.1 Elección de estructuras cristalográficas de proteínas virales y celulares con el dominio <i>jelly roll</i> para la construcción de una base de datos	15
3.3 Comparación de los dominios <i>jelly roll</i> de proteínas virales y celulares	16
3.4 Construcción y edición de dendogramas	17
4. Resultados y discusión	18
4.1 La importancia de la definición operacional del dominio <i>jelly roll</i>	18
4.2 Descripción de la base de datos de dominios <i>jelly roll</i> virales y celulares	18
4.3 Análisis filogenético de los dominios <i>jelly roll</i> virales	19
4.3.1 Homología entre la proteína de cápside de Caliciviridae y la VP3 de Picornaviridae	22
4.3.2 El origen viral del subdominio S de VP2 de Birnaviridae	25
4.6 Las funciones de las proteínas celulares con el dominio <i>jelly roll</i>	30
4.6.1 La prevalencia de funciones relacionadas con carbohidratos de las proteínas celulares con dominios <i>jelly roll</i>	31
4.7 Los dominios <i>jelly roll</i> virales y celulares con similitud estructural	33
4.8 Homología entre los dominios <i>jelly roll</i> del mitógeno derivado de Yersinia y la CP de Bromoviridae	34
4.9 El origen celular del subdominio P de la proteína de cápside VP2 de Birnaviridae	36
5. Conclusiones	37
6. Literatura citada	39
Anexos	45

Agradecimientos

Agradezco al programa de Ayudantes de Investigador del Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico otorgado durante el desarrollo de este proyecto de investigación (expediente 18977-14225).

A la UNAM, por darme todos los recursos académicos y culturales necesarios para alcanzar mis metas.

A los profesores que me han inspirado y mostrado la gracia del aprendizaje constante.

A mi tutor, José Alberto Campillo Balderas, por todo su apoyo, amistad y consejos, sin quien la culminación de este proyecto no habría sido posible. Le agradezco el haberme enseñado lo que es la pasión y el amor por lo que haces, y el ser un ejemplo de superación constante.

Al Dr. Arturo Becerra y al Dr. Antonio Lazcano, de quienes ha sido un privilegio haber aprendido a lo largo de mi estancia en el laboratorio de Origen de la Vida.

Agradezco el apoyo proporcionado durante el desarrollo de los análisis de este trabajo a Claudia Alvarez, Alejandro Cisneros, Carolina Rocha y Alberto Vazquez. Así como los comentarios y sugerencias hechos por los sinodales de este proyecto, Dr. Antonio Lazcano, Dr. Enrique Rudiño, Dr. Rodrigo Jácome, y Dra. Selene Zarate.

A todos los miembros del laboratorio de Origen de la Vida por sus aportaciones. En particular, agradezco la amistad de Germán, Caro, Coral, Ra y Claudia. Gracias por acompañarme en este largo proceso.

A mis amigos de la facultad, Sara Lucero, Alfredo Rodriguez, Maria Peñavera, Natalia Chapela y Eric Grepe por su compañía y amistad sincera, con quienes crecer ha sido toda una aventura. Agradezco especialmente a Carolina y a Lucía, quienes han sido mis compañeras de viaje y han vuelto la vida en la CDMX más llevadera.

Al Dr. Josep Moreno y a mis queridos amigos del Museo de Paleontología por formarme y recibirme como un miembro más de la familia. En especial a Sergio y Daniel, por todas las conversaciones y experiencias compartidas.

A mis hermanos y padres, a quienes les debo todo.

Resumen

Existen tres hipótesis principales sobre el origen de los virus de RNA. La primera de ellas es la hipótesis del origen reductivo, que propone que los virus se originaron a partir de una célula parasitaria que redujo su material genético y su tamaño. La segunda es la hipótesis del origen primordial, que sugiere que los virus se originaron antes que las células. La tercera es la hipótesis del origen celular, que plantea que los virus se originaron a partir del escape de elementos celulares. Contrario a lo señalado por la hipótesis del origen primordial, la distribución taxonómica de los hospederos de los virus de RNA indica que éstos son de origen reciente, ya que infectan principalmente eucariontes. Además, se ha visto que algunos dominios proteínicos como el *jelly roll* forman parte tanto de proteínas celulares como de proteínas virales, lo que apoya la hipótesis del origen celular de los virus.

El dominio *jelly roll* es un plegamiento estructural que se compone por ocho hebras beta antiparalelas que se acomodan en dos láminas de cuatro hebras cada una, dándole una forma de sándwich o de cuña. Con el fin de determinar si los dominios *jelly roll* celulares y los virales son homólogos es necesario un análisis filogenético. Debido a que la conservación a nivel de secuencia de residuos de aminoácidos es muy baja, los análisis filogenéticos tradicionales son poco informativos. Dado que la estructura terciaria suele conservarse más que la estructura primaria, se optó por realizar análisis basados en la primera. Esto permitió determinar que los dominios *jelly roll* de virus son homólogos a los de proteínas celulares, lo cual apoyaría la hipótesis del origen celular. Adicionalmente, los dendogramas resultantes de las comparaciones de los dominios *jelly roll* de proteínas celulares y virales, permitieron hacer inferencias sobre la historia evolutiva de los subdominios S y P de la proteína VP2 de Birnaviridae, así como de la familia de proteínas de factores de necrosis tumoral y el mitógeno derivado de *Yersinia*.

Abstract

There are three hypotheses about the origin of RNA virus. The first one is the Reduction hypothesis, which proposes that virus originated from parasitic cells that reduced its genetic material and its size. The second one is the Virus-first hypothesis, which suggests that virus originated before cells. The third one is the Escape hypothesis, which states that viruses originated from the escape of cellular elements. Contrary to the hypothesis of Virus-first, the taxonomic distribution of RNA virus hosts indicates that they are of recent origin since they mainly infect eukaryotes. In addition, it has been seen that some protein domains such as *jelly roll* are part of both cellular and viral proteins, which supports the hypothesis of the cellular origin of viruses or escape.

The *jelly roll* domain is a structural folding that consists of eight antiparallel beta strands that fit into two sheets of four strands each, giving it a sandwich or wedge shape. In order to determine whether the cellular and viral *jelly roll* domains are homologous, phylogenetic analysis is necessary. Because, the conservation at the sequence level of amino acid residues is very low, traditional phylogenetic analyzes are not very informative. Given that the tertiary structure is usually preserved more than the primary structure, it was decided to perform analyzes based on the first one. This allowed determining that *jelly roll* domains of viruses are homologous to those of cellular proteins, which would support the hypothesis of escape origin. Additionally, the dendrograms resulting from the comparisons of *jelly roll* domains of cellular and viral proteins, allowed us to make inferences about the evolutionary history of the S and P subdomains of the VP2 protein of Birnaviridae, as well as the protein family of tumor necrosis factors and the mitogen derived from *Yersinia*.

1. Introducción

1.1 Características generales de los virus de RNA

Los virus son entidades submicroscópicas capaces de replicarse utilizando la maquinaria molecular de su hospedero (Lazcano, 2005). Sus genomas están conformados o por DNA o por RNA y, al mismo tiempo, pueden contener proteínas que conforman la nucleocápside y la cápside. De acuerdo con la clasificación propuesta por Baltimore (1971), los virus se agrupan de acuerdo al tipo de genoma y al mecanismo de replicación (Fig. 1).

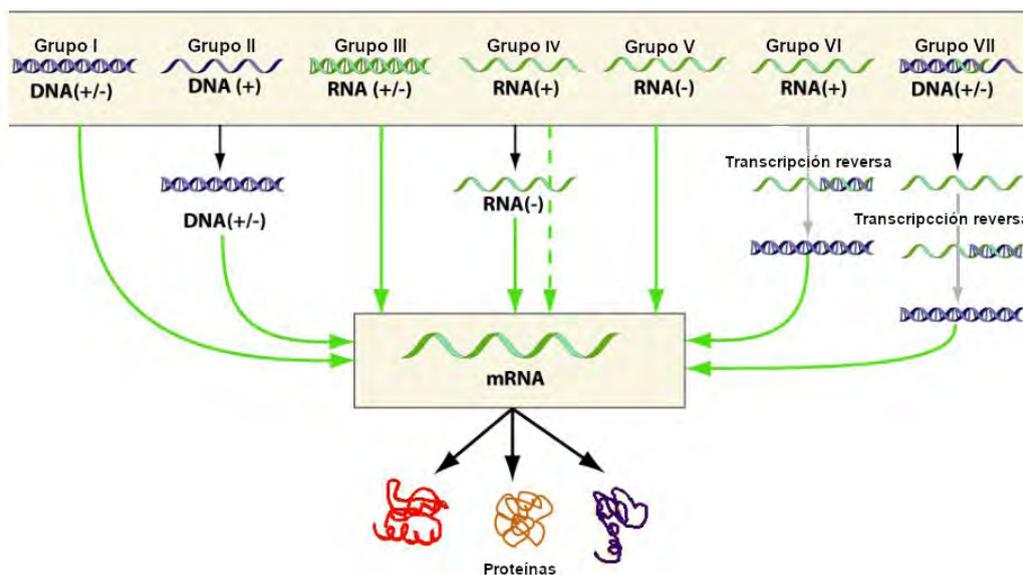


Figura 1. Tipo de genoma (DNA o RNA; monocatenario o bicatenario) y mecanismos de síntesis de RNA mensajero. Azul, DNA; verde, RNA. Los virus de RNA están clasificados en los grupos III (bicatenario), IV (monocatenario con polaridad positiva), V (monocatenario con polaridad negativa) y VI (monocatenario con polaridad positiva y con un DNA intermediario en su ciclo de replicación). Imagen modificada de ViralZone.

Los virus de RNA, al igual que los de DNA monocatenario, a diferencia de los de DNA de doble cadena, tienen una alta tasa de mutación (de 10^{-3} a 10^{-7} nucleótidos por genoma por replicación, n/g/r; Fig. 2) (Duffy et al, 2008), un genoma muy pequeño (menor a 35 kilopares de bases, kbp), un tamaño de población grande, poca transferencia horizontal y duplicación de genes, comúnmente marcos de lectura sobrelapados, y menos recombinación de genes (Holmes, 2009). Además, ocasionan la mayoría de las enfermedades emergentes conocidas en humanos, debido a su gran tasa de mutación, y provocan frecuentemente zoonosis (Woolhouse et al, 2013).

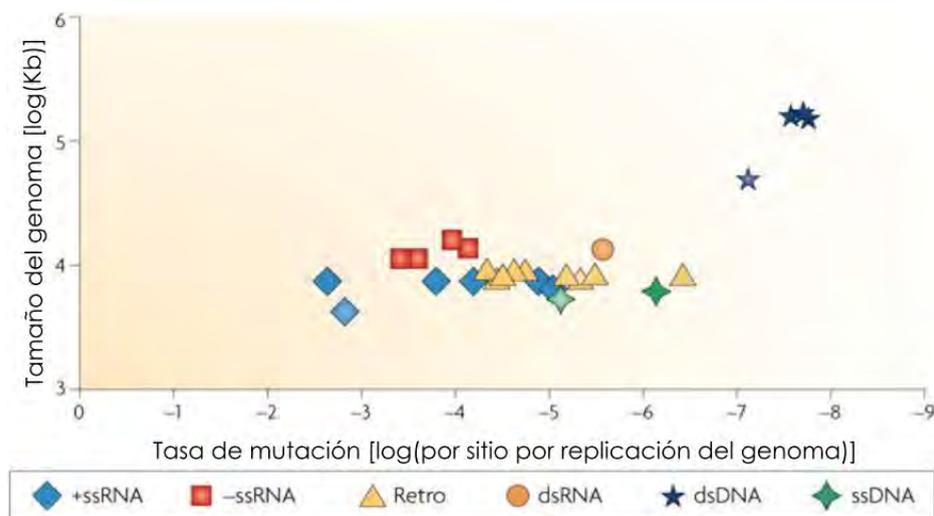


Figura 2. Tasa de mutación promedio de los virus por replicación del genoma. Virus de RNA presentan la mayor tasa de mutación promedio. Imagen de Duffy et al, 2008.

1.2 Origen y evolución temprana de los virus de RNA

Para inferir el ancestro común de todos los seres vivos (LCA), Carl Woese y colaboradores, 1970, propusieron la comparación del RNA ribosomal de Bacteria, Archaea y Eukarya. Sin embargo, determinar las relaciones filogenéticas profundas de los virus, así como su propio origen, ha sido todo un reto debido a la alta tasa de mutación de sus secuencias, la inexistencia de un registro fósil y sus características politéticas.

Se distinguen tres hipótesis principales sobre el origen de los virus (Fig. 3). La hipótesis virocéntrica propone que los virus se originaron a partir de elementos genéticos primordiales, los cuales son más antiguos que las células. En la hipótesis de la evolución regresiva o modelo de degeneración celular se argumenta que los virus descienden de células parásitas intracelulares que desarrollaron una creciente dependencia de su hospedero, lo cual derivó en la pérdida de la maquinaria necesaria para replicarse y en la reducción de su genoma, conservando, sin embargo, la capacidad de infectar. En la hipótesis del origen celular o de escape de genes se sugiere que los virus se originaron en las células a partir de moléculas de RNA o DNA que adquirieron algunas proteínas necesarias para replicarse (como las polimerasas) y para la formación de un virión (Bubanovic *et al.*, 2005; Forterre, 2006; Koonin *et al.*, 2006).

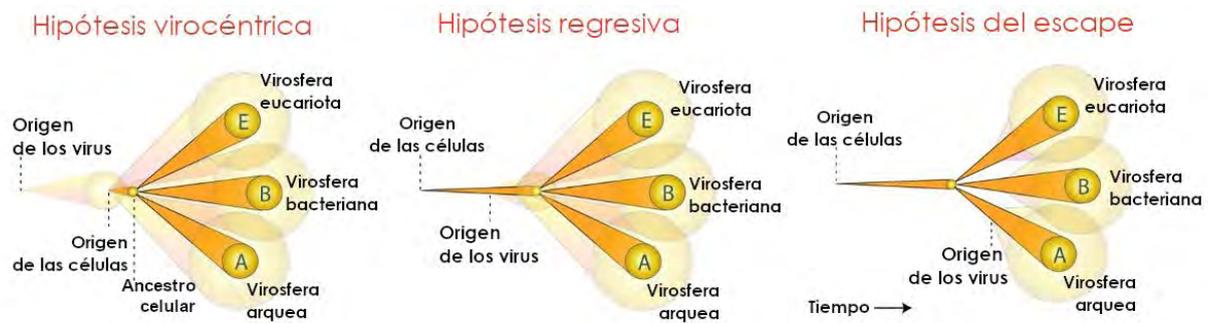


Figura 3. Hipótesis sobre el origen de los virus. A, Archaea; B, Bacteria; E, Eukarya. Imagen modificada de Nasir et al, 2012.

La idea de que los virus de RNA son tan antiguos como la vida misma (hipótesis virocéntrica) fue fácilmente aceptada, puesto que fue relacionada con la teoría del mundo del RNA (Forterre, 2006). Sin embargo, dado que ningún virus puede proliferar sin su hospedero y los virus de RNA infectan principalmente eucariontes, la idea de que son tan antiguos debe ser cuestionada (Campillo-Balderas, 2015).

La distribución taxonómica de los hospederos de los virus de RNA parece indicar que estos virus son de origen reciente, dado que infectan principalmente eucariontes, lo cual concuerda con la hipótesis del origen celular o de escape de genes (Campillo-Balderas, 2018). La presencia de virus de RNA en procariontes es rara. Y únicamente se conocen dos familias de virus de RNA que infectan bacterias. Éstas, sin embargo, son proteobacterias, organismos comunes en la microbiota de muchos animales, lo que sugiere que pueden ser resultado de transferencia (Campillo-Balderas, 2018).

1.3 Estrategias para abordar el problema sobre el origen de los virus de RNA

1.3.1 Filogenias basadas en secuencias de residuos de aminoácidos

A partir del desarrollo de las tecnologías de secuenciación (Sanger & Tuppy, 1951) y el rápido incremento de secuencias disponibles en bases de datos, el interés por el desarrollo de herramientas para el análisis de éstas creció considerablemente y fue de esta manera como se

crearon diversos programas de alineamiento de secuencias que siguen filosofías similares (Tobergte & Curtis, 2013).

En los estudios evolutivos, los análisis de secuencias de aminoácidos permiten determinar si dos o más proteínas son homólogas, e incluso inferir qué cambios sucedieron a partir del último ancestro común de ambas (Tobergte & Curtis, 2013). No obstante, en algunas proteínas la conservación a nivel de secuencia de residuos de aminoácidos es muy baja, como es el caso de las proteínas de cápside de virus de RNA, por lo que los análisis filogenéticos tradicionales basados en estructura primaria son poco informativos cuando se intentan estudiar homologías profundas (Cerny *et al.*, 2015).

1.3.2 Filogenias basadas en estructuras terciarias

En las proteínas, la estructura terciaria se conserva más que la estructura primaria, unido a la creciente cantidad de estructuras terciarias determinadas y depositadas en bases de datos como el *Protein Data Bank* (PDB, <http://www.rcsb.org>; Berman *et al.*, 2000), ello permite aproximarnos a la historia evolutiva de estas moléculas (Holm & Sander, 1993; Cerny *et al.*, 2014).

Un claro ejemplo de lo anterior son las proteínas denominadas RNA polimerasas dependientes de RNA, enzimas que participan en la replicación de los ácidos nucleicos y que a pesar de ser una de las moléculas compartidas por todos los virus de RNA, a excepción de los retrovirus, su secuencia no está lo suficientemente conservada para análisis filogenéticos. Sin embargo, se conserva su estructura terciaria, particularmente el subdominio *palm* (Jácome *et al.*, 2015). Esto ha permitido llevar a cabo estudios evolutivos basados en la comparación de estructuras terciarias de esta molécula, y se ha establecido que es un buen marcador filogenético (Cerny *et al.*, 2014; Jácome *et al.*, 2015).

1.4 Cápsides de virus de RNA

Se le llama cápside a la cubierta proteínica del material genético viral (Flint *et al.*, 2013). Algunos virus tienen una cubierta adicional que rodea a la cápside y está compuesta por

lípidos de la membrana de su hospedero. Este tipo de virus se les conoce como virus envueltos, mientras que a los que carecen de esta envoltura lipídica se les denomina virus desnudos (Rossmann & Johnson, 1989; Flint *et al.*, 2013).

De acuerdo con su forma, éstas se clasifican en tres tipos: helicoidal, icosaédrica y compleja (que está ausente en virus de RNA). En las cápsides helicoidales, las proteínas se acomodan alrededor del material genético y forman una especie de espiral, mientras que en las icosaédricas adoptan la forma de un poliedro de 20 caras triangulares (Flint *et al.*, 2013).

1.5 Estructura de las cápsides icosaédricas

Las cápsides se conforman por subunidades repetitivas que se autoensamblan y adquieren formas geométricas (Rossmann & Johnson, 1989). En algunos casos una cápside se constituye de la misma subunidad repetida m veces, mientras que otras pueden estar compuestas por varios tipos de subunidades repetidas n veces (Flint *et al.*, 2013). Para ilustrar esto último se puede citar el caso de la cápside del Virus del Mosaico del Cauquí, que se compone de tres tipos de proteínas: la VP1, que forma pentámeros; las VP2 y VP3, que se acomodan en hexámeros. A estos dos tipos de arreglos se les denomina capsómeros (Fig. 4; Chatterji *et al.*; 2004; Flint *et al.*, 2013).

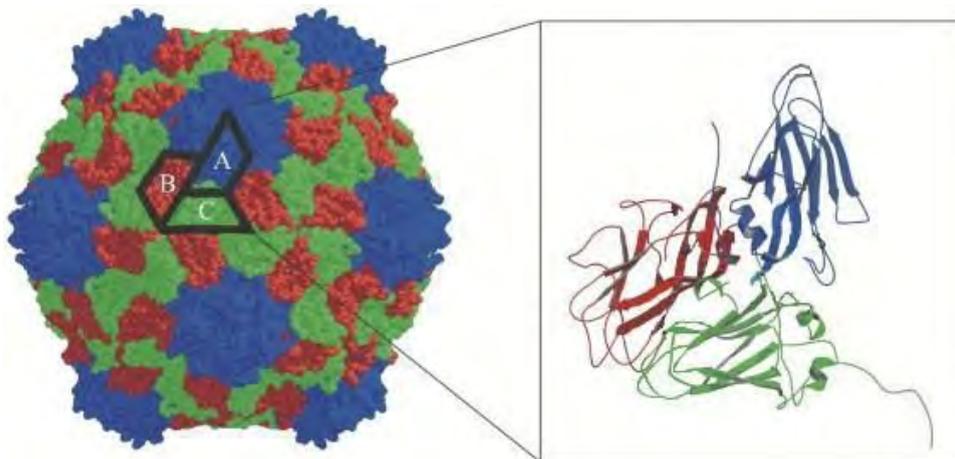


Figura 4. Cápside del virus del mosaico del cauquí [ssRNA(+)]. Esta cápside se conforma de tres subunidades, la VP1 (A), la VP2 (B) y la VP3 (C). Las subunidades se ensamblan formando dos tipos de capsómeros: pentámeros (azul) y hexámeros (rojo-verde). Imagen modificada de Chatterji *et al.*, 2004

Una forma de medir el tamaño de una cápside es mediante su triangulación (T), la cual es una estimación de la distancia entre sus capsómeros, es decir, una mayor triangulación indica un mayor tamaño de cápside (Prasad & Schmid, 2013).

1.5.1 El dominio *jelly roll* en las cápsides icosaédricas

Un dominio proteínico se define como una región funcional y/o estructural reconocible en una proteína (Lodish *et al.*, 2000). Un ejemplo de lo anterior es el llamado dominio *jelly roll*.

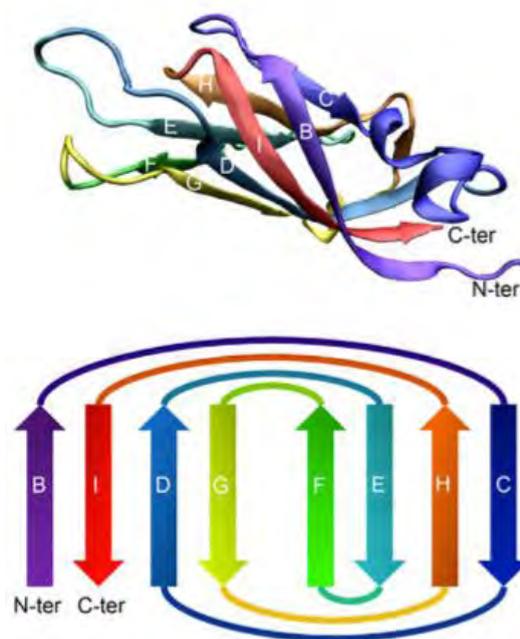


Figura 5. Dominio *jelly roll*. El dominio *jelly roll* se compone de ocho hebras beta antiparalelas, que se acomodan en dos láminas (BIDG y CHEF). El amino terminal y el carboxilo terminal siempre son adyacentes. Imagen modificada de Cheng & Brooks, 2013.

El término *jelly roll* fue usado por primera vez por Richardson (1981) para describir un plegamiento estructural de hebras beta antiparalelas en la proteína de cápside del virus del mosaico del tomate, en el que las ocho hebras se acomodan en dos láminas de cuatro hebras cada una, adquiriendo una forma de sándwich o de cuña (Fig. 5). Las ocho hebras del *jelly roll* son etiquetadas de la B a la I, y se acomodan de modo que las hebras B-I-D-G forman una lámina y las hebras C-H-E-F forman otra (Chelvanayagam *et al.*, 1992).

En el trabajo realizado por Chelvanayagam *et al.* (1992), en el que se compararon 25 secuencias correspondientes a dominios *jelly roll* tanto virales (n= 22) como celulares (n= 3),

se observó que los pares de residuos de aminoácidos V-L, V-I y L-L son recurrentes. Sin embargo, no se reporta homología a nivel de secuencia. Además, se detectó una tendencia a presentar residuos de aminoácidos hidrofóbicos en las “asas” (estructuras que conectan las hebras beta), y se propuso que esta característica, junto con la capacidad del *jelly roll* de adoptar una estructura cuaternaria, ha favorecido su prevalencia en moléculas multiméricas como las cápsides (Chelvanayagam *et al.*, 1992; Donadini *et al.*, 2004).

Krupovic y Koonin, 2017, compararon las estructuras terciarias de dominios *jelly roll* virales y celulares, y concluyeron que son homólogos. Además, detectaron que el dominio *jelly roll* se encuentra aproximadamente en el 30% de las cápsides virales conocidas, principalmente en las icosaédricas. Basándose en lo anterior, se ha propuesto que el *jelly roll* se originó en células y posteriormente, fue tomado por los virus en múltiples ocasiones a partir de sus hospederos (Kuprovic & Bamford, 2011; Kuprovic & Koonin, 2017). No obstante, no se han construido dendogramas basados en comparaciones de la estructura terciaria de este dominio.

2. Objetivos

2.1 Objetivo general

Inferir la historia evolutiva del dominio *jelly roll* en proteínas de virus de RNA con cápside icosaédrica.

2.2 Objetivos particulares

- Elaborar una base de datos con la información biológica y experimental de las estructuras tridimensionales de proteínas con el dominio *jelly roll* de las familias de virus de RNA utilizadas por Jácome *et al*, 2015, así como de proteínas celulares que también presentan este dominio.
- Editar y comparar la estructura tridimensional de los dominios *jelly roll* de las distintas proteínas virales y celulares.
- Construir dendrogramas basados en la comparación de estructuras terciarias de los dominios *jelly roll* de proteínas virales y los *jelly roll* de las proteínas celulares.

2.3 Hipótesis

Los dominios *jelly roll* de cápsides virales y de proteínas celulares comparten un origen común, que se infiere a partir de la comparación de sus estructuras terciarias.

3. Métodos

3.1 Elección de estructuras cristalográficas de proteínas virales y celulares con el dominio *jelly roll* para la construcción de una base de datos

Las proteínas celulares con el dominio *jelly roll* fueron buscadas en el *Protein Data Bank* (PDB, <http://www.rcsb.org>; Berman *et al.*, 2000); se usaron las palabras “*jelly roll*” como texto de búsqueda. La búsqueda de proteínas virales se llevó a cabo en la base de datos *ViralZone*. (<https://viralzone.expasy.org>; Hulo *et al.*, 2011) y se limitó sólo a las familias virales con genoma de RNA y cápsides icosaédricas usadas por Jácome *et al.* (2015), aunque únicamente se encontraron reportadas estructuras terciarias de proteínas de cápside con dominio *jelly roll* en las familias Picornaviridae, Birnaviridae, Caliciviridae, Alphatetraviridae, Bromoviridae y Reoviridae. El catálogo completo fue depositado en [línea](#).

Con el fin de evitar la comparación de estructuras determinadas por métodos distintos, la búsqueda se limitó a las estructuras cristalográficas determinadas por medio de difracción de rayos X. Como criterios adicionales, se buscó que la resolución fuera mayor a 3 Ångströms ($A = 1 \times 10^{-10} \text{ m} = 0,1 \text{ nm}$) y se eliminó la redundancia. Por último, se eligieron aquellas proteínas con la mayor longitud de secuencia, con un máximo de tres mutaciones y un máximo de cuatro ligandos (con excepción de la glucanotransferasa de *Bacillus circulans* [pdb:3wnk], que presenta cinco ligandos).

Se agregaron a mano los datos experimentales y biológicos correspondientes, como el nombre de la molécula, la familia proteínica (dato verificado en Pfam, *Protein family database*, <http://pfam.xfam.org>; Finn, 2006), su clasificación, el ID de cuatro letras de la proteína proporcionado por el PDB, su estequiometría, el número de cadenas, la longitud de la secuencia anotada en el Uniprot (*Universal Protein*, <http://www.uniprot.org/>; Bateman *et al.*, 2017) y en el PDB, el rango (localización en la estructura tridimensional de la proteína) del dominio *jelly roll* (obtenido del PDBsum, <http://www.ebi.ac.uk/pdbsum/>; Laskowski, 2009), los ligandos de la proteína, el número de mutaciones, los dominios adyacentes al *jelly roll*, la especie a la que pertenece la proteína, el dominio biológico (Bacteria, Archaea o Eucaria) y su tipo de metabolismo; en el caso de las proteínas virales también se anotó la

especie y familia viral, el tipo de simetría, la clasificación de Baltimore, el tamaño del genoma, la triangulación de la cápside, el diámetro del virión, el tropismo, el hospedero al que infecta, el vector y el receptor celular que utiliza al iniciar la infección.

3.2 Edición de los archivos PDB de las estructuras cristalográficas de proteínas de virus de RNA y de células con el dominio jelly roll

Algunas de las proteínas tenían dominios adicionales al *jelly roll*. En estos casos fue necesaria la edición de los archivos PDB con el fin de aislar el dominio de interés para este trabajo, ya que la presencia de los dominio adyacentes dificulta más el reconocimiento de las regiones conservadas. La edición fue realizada con el programa Chimera 1.12 (<http://www.cgl.ucsf.edu/chimera>; Huang *et al.*, 1996) y fue basada en los rangos registrados en el *PDBsum* para el *jelly roll* de cada una de las proteínas utilizadas en esta tesis.

3.3 Comparación de los dominios jelly roll de proteínas virales y celulares

Debido a las modificaciones conformacionales que los ligandos pueden causar en las proteínas (Bustos-Jaimes *et al.*, 2002) y el efecto que éste conlleva en las topologías de los dendogramas (Jácome *et al.*, en prep), las proteínas fueron divididas y comparadas de acuerdo con los siguientes subgrupos: (1) proteínas virales con ligandos, (2) Proteínas celulares con ligandos, (3) proteínas virales sin ligandos, (4) proteínas virales con ligandos, (5) proteínas virales y proteínas celulares sin ligandos, (6) proteínas virales y proteínas celulares con ligandos. El último dendograma no se incluye en este escrito ya que el análisis de las más de seis mil combinaciones continúa en proceso.

Se usó la herramienta de comparación múltiple del servidor en línea de *Dali: Protein Structure Comparison Server* (<http://ekhidna2.biocenter.helsinki.fi/dali/>; Holm & Sander, 1995; Holm & Laakso, 2016), que construye una matriz de distancia entre los carbonos alfa de cada una de las proteínas utilizadas en el análisis. Posteriormente el algoritmo compara estas matrices, de las cuales se obtiene el valor Z:

$$Z = \frac{\text{valor observado} - \text{media}}{\text{desviación estándar}}$$

3.4 Construcción y edición de dendogramas

El servidor de *Dali* construye dendogramas basados en el valor *Z*. Éstos fueron descargados en formato *Newick* y cargados en el visualizador y editor de dendogramas *iTOL* (Interactive Tree Of Life, <https://itol.embl.de>; Letunic & Bork, 2006; Letunic & Bork, 2016).

4. Resultados y discusión

4.1 La importancia de la definición operacional del dominio *jelly roll*

Cuando se busca en las bases de datos como el *PDB* a todas aquellas estructuras que presentan el dominio *jelly roll* es muy común encontrarse con falsos positivos. Esto podría deberse a una falta de exactitud en la definición, lo cual se traduce en problemas de anotación, por ello, se vuelve necesaria una definición más amplia. Stirk y colaboradores (1992) definieron el dominio *jelly roll* como un motivo estructural tipo llave griega, constituido por ocho cadenas beta. El patrón de puentes de hidrógeno formados entre cadenas contiguas se rompe en dos sitios, en consecuencia, la estructura se acomoda en dos láminas beta exclusivamente antiparalelas y de cuatro hebras cada una. Es importante subrayar que las hebras adyacentes en secuencia aparecen en diferentes láminas, con excepción de la cadena E y F, que se encuentran en la misma lámina, por lo que el dominio tiene únicamente una conexión tipo horquilla mientras que el resto son tipo arco.

La definición de Stirk y colaboradores (1992) proporciona una lista detallada de las características que una estructura debe tener para ser llamada *jelly roll*, dándole especial importancia a la conectividad que siguen las hebras. Por esta razón, se verificó que todas las estructuras tridimensionales utilizadas en este trabajo siguieran esta definición. Con el objetivo de acelerar el proceso, únicamente se prestó atención a la pauta seguida por las hebras, es decir, a su conectividad. Ello permite el reconocimiento del dominio a pesar de las inserciones que presenta en las distintas proteínas, así como la identificación y eliminación de los falsos positivos.

4.2 Descripción de la base de datos de dominios *jelly roll* virales y celulares

En total se obtuvieron 26 estructuras terciarias virales del *Protein Data Bank*, el promedio de longitud del dominio *jelly roll* en esta muestra es de 221 residuos de aminoácidos. En cuanto a la biología de los virus de la muestra, en su mayoría presentan tropismos hacia epitelios tanto intestinales como respiratorios (Tabla S1), con excepción de el virus de la enfermedad bursal infecciosa (Birnaviridae) que infecta linfocitos B inmaduros (Delgui *et al.*, 2009) y el virus del mosaico amarillo de Turnip (Tymoviridae) que infecta el floema de su hospedero

(ViralZone, <https://viralzone.expasy.org/54>). Solamente cuatro proteínas de la muestra presentan dominios adicionales al *jelly roll*, entre los que se encuentran el *V-type ATP synthase subunit C fold* (n= 1), el “*Bluetongue Virus 10*” (n= 2) y el *Nucleoplasmin-like* (n=1).

La muestra de proteínas celulares con dominio *jelly roll* comprende 78 estructuras terciarias. En esta muestra el promedio de longitud del dominio *jelly roll* fue de 165.8 residuos de aminoácidos. En cuanto a los dominios adicionales al *jelly roll*, 19 de las proteínas de la muestra presentan uno o más. Entre estos dominios adyacentes se encuentran el *Calnexin luminal domain* (n= 1), *non-globular proline- rich hairpin domain* (n= 1), *Neuraminidase* (n=2), *Rossmann fold* (n=3), *Immunoglobulin-like* (n= 4), *TIM Barrel* (n= 4), *Tumor Necrosis Factor Receptor* (n= 4), *Beta-galactosidase* (n= 2), *duf1285 like fold* (n=1), *Cathepsin D* (n= 1), *Delta-Endotoxin* (n= 1) y *Trefoil (Acidic Fibroblast Growth Factor)* (n= 1).

4.3 Análisis filogenético de los dominios *jelly roll* virales

En trabajos previos, la RNA polimerasa dependiente de RNA mostró ser un buen marcador filogenético (Jácome *et al.*, 2015; Fig. 6). Al igual que los dendogramas construidos a partir de las comparaciones estructurales de RNA polimerasas, los dendogramas basados en la comparación de estructuras de dominios *jelly roll* muestran a las proteínas agrupadas de acuerdo a la familia viral a la que pertenecen, que a su vez se agrupan de acuerdo con el grupo de la clasificación de Baltimore al que corresponden (Fig. 7 y 8).

Cabe mencionar que como prueba, se realizó un dendograma utilizando los dominios *jelly roll* con y sin ligandos que muestra clados definidos en los que las proteínas se agrupan de acuerdo a su tipo y a la familia viral a la que corresponden, lo que indica que en este caso los ligandos no afectan la topología del dendograma (Fig. S1).

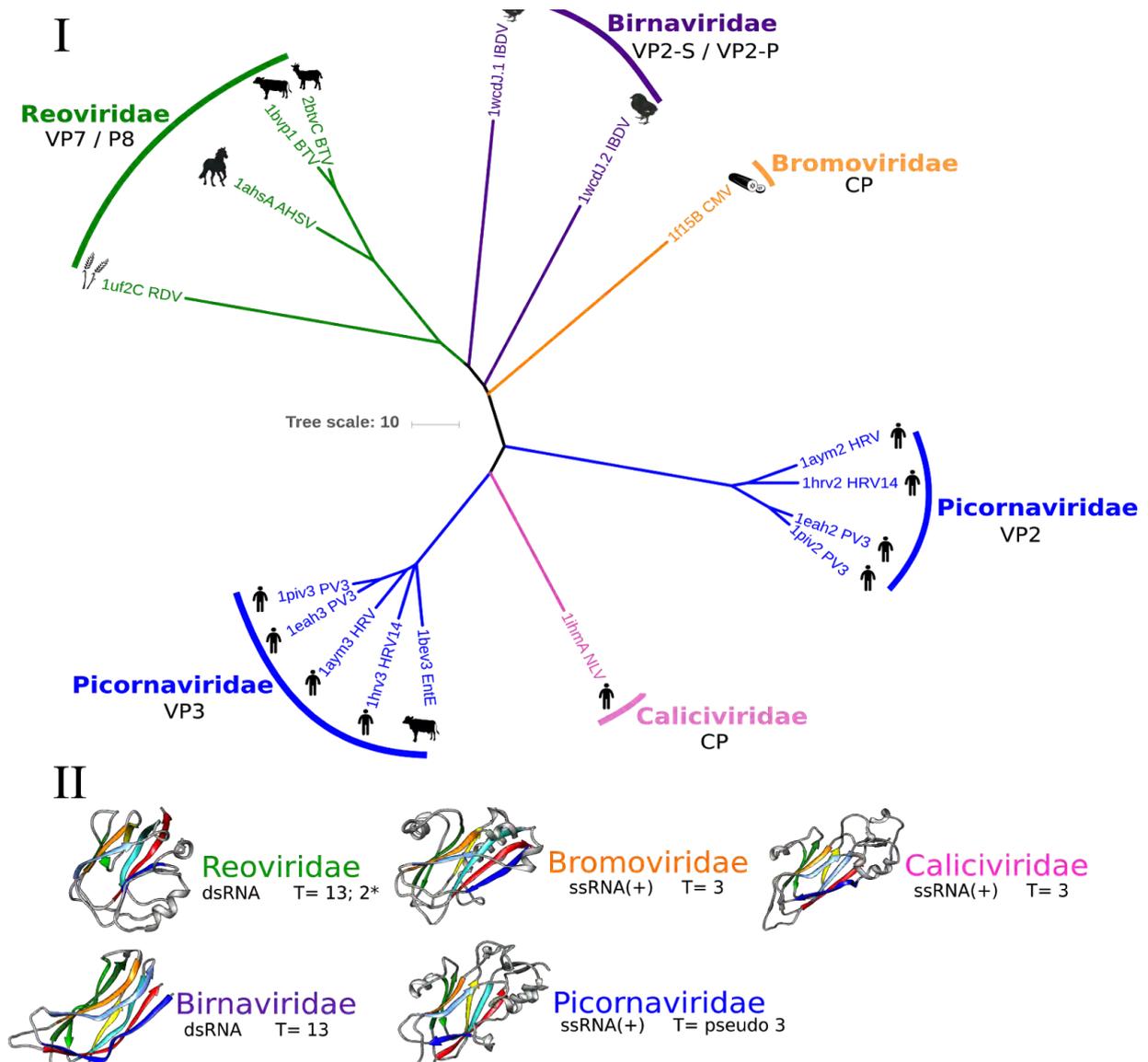


Figura 7. Dendrograma de dominios jelly roll de proteínas de cápside sin ligandos. (I) Se muestran la familia viral y la proteína correspondiente en cada clado. En cada rama se señala el código PDB y el virus al que corresponde. La figura negra indica el hospedero al cual infecta cada uno de los virus. En la parte inferior (II) se muestran los dominios jelly roll de las cápsides de las familias virales incluidas en el dendrograma, además se señalan el grupo de Baltimore y la triangulación de la cápside correspondiente. RDV: virus del enanismo del arroz; AHSV: virus de la enfermedad equina africana; BTV: virus de la lengua azul; IBDV: virus de la enfermedad de bursitis Infecciosa; CMV: virus del mosaico del pepino; HRV: rhinovirus humano; HRV14: rhinovirus humano 14; PV3: poliovirus 3; NLV: virus de norwalk; EntE: enterovirus E. *Triangulación de la cápside interna.

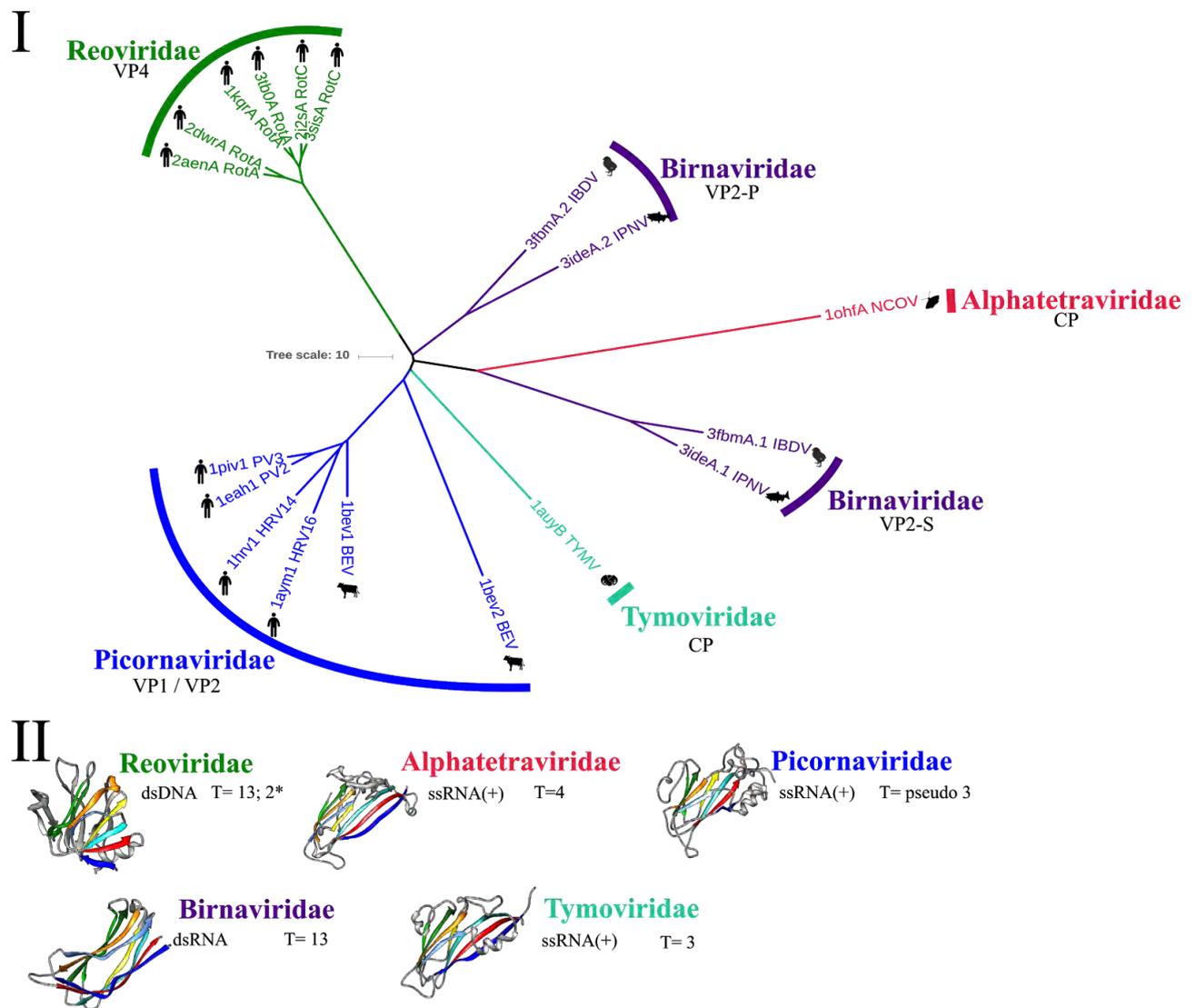


Figura 8. Dendrograma de dominios jelly roll de proteínas de cápside con ligandos. (I) Se muestran la familia viral y la proteína correspondiente en cada clado. En cada rama se señala el código PDB y el nombre de la proteína, así como el virus al que corresponde. El icono negro indica el hospedero al cual infecta cada uno de los virus. En la parte inferior (II) se muestran los dominios jelly roll de las cápsides de las familias virales incluidas en el dendrograma, además se señalan el grupo de Baltimore y la triangulación de la cápside correspondiente. RotA: rotavirus A; RotC: rotavirus C; IBDV: virus de la bursitis infecciosa; IPNV: virus de necrosis pancreática infecciosa; NCOV: virus omega de Nudaurelia capensis; TYMV: virus del mosaico amarillo del nabo; BEV: enterovirus bovino; HRV16: rinovirus humano 16; HRV14: rinovirus humano 14; PV2: poliovirus 2; PV3: poliovirus 3. *Triangulación de la cápside interna.

4.3.1 Homología entre la proteína de cápside de Caliciviridae y la VP3 de Picornaviridae

En el dendrograma basado en comparaciones de estructuras de dominios *jelly roll* virales, la proteína de cápside (CP) de Caliciviridae se agrupa con las proteínas VP2 y VP3 de

Picornaviridae (Fig. 7). Anteriormente se había señalado que hay una relación filogenética entre ambas familias (Le Gall *et al.*, 2008). No obstante, la CP de Caliciviridae forma un clado con la VP3 de Picornaviridae, lo que sugiere que ambas proteínas son homólogas.

La CP de Caliciviridae se divide en dos dominios denominados S y P (Fig. 9), el dominio S es el que forma la cápside icosaédrica y presenta un plegamiento tipo *jelly roll*, mientras que el dominio P se ubica en el exterior de la cápside y perpendicularmente al dominio S (Venkataram Prasad *et al.*, 1999). Al realizar una comparación estructural del dominio S de la CP de Caliciviridae con la VP3 de Picornaviridae, éstas resultan significativamente similares ($P= 2.40e-13$; Fig. 10).

Sin embargo, la similitud de la CP con la VP3 también es reconocible a nivel de estructura primaria, ya que además de compartir una secuencia prolina-prolina-glicina en la “asa” que une las hebras E y F de sus dominios *jelly roll* (Venkataram Prasad *et al.*, 1999), es posible reconocer una mayor semejanza entre ambas, en comparación con las VP2 y VP1 de Picornaviridae (Koonin & Gorbalenya, 1992).

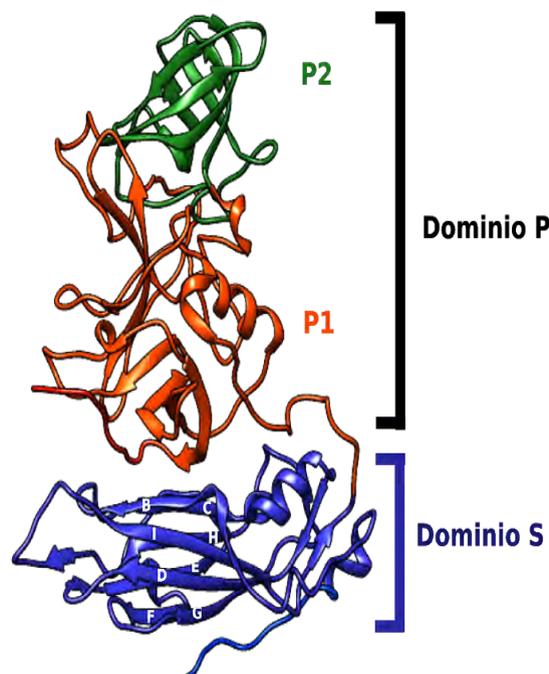


Figura 9. Proteína de cápside de Caliciviridae. La proteína de cápside de Caliciviridae se divide en dos dominios, el S y el P. El dominio S forma la cápside icosaédrica y presenta un plegamiento tipo *jelly roll*. El dominio P se divide en P1 y P2, forma una especie de proyección en la cápside. Es decir, es perpendicular al dominio S. PDB: 1ihm. Imagen producida con el programa UCSF Chimera (<http://www.rbvi.ucsf.edu/chimera/>).

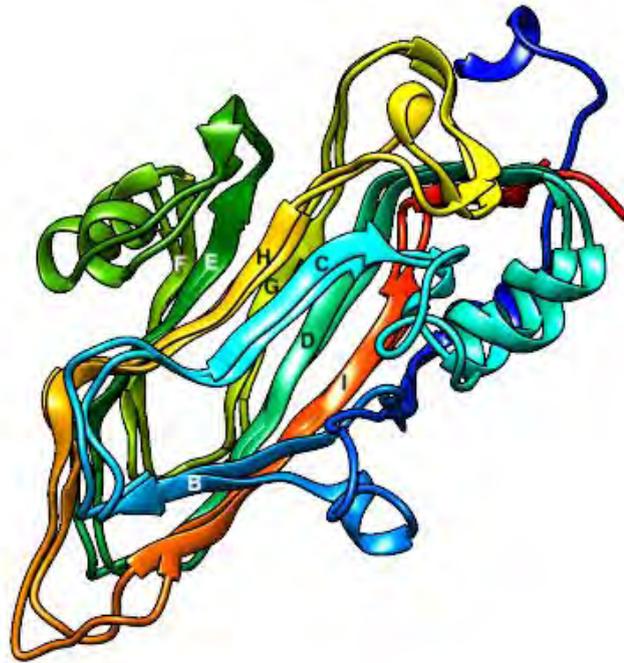


Figura 10. Comparación estructural del dominio jelly roll de la proteína de cápside de Caliciviridae (PDB: 1ihm) con el de la VP3 de Picornaviridae (PDB: 1piv3). La comparación se llevó a cabo en el servidor en línea de FATCAT (<http://fatcat.sanfordburnham.org>). RMSD= 2.49 Å, $P < 0.05$, $P = 2.40e-13$ y 163 posiciones alineadas (no. residuos de aminoácidos jelly roll: 177-CP y 184-VP3). Imagen producida con el programa UCSF Chimera (<http://www.rbvi.ucsf.edu/chimera/>).

4.3.2 El origen viral del subdominio S de VP2 de Birnaviridae

La proteína de cápside VP2 de la familia Birnaviridae comprende tres subdominios, llamados B (de base), S (de *shell*) y P (de proyección). En el virión, el dominio B se ubica al interior de la cápside, el S forma la cápside icosaédrica y el P se orienta perpendicularmente a la cápside, formando una especie de protuberancia (Fig. 11). Los subdominios S y P tienen una topología tipo *jelly roll* (Coulibaly *et al.*, 2005).

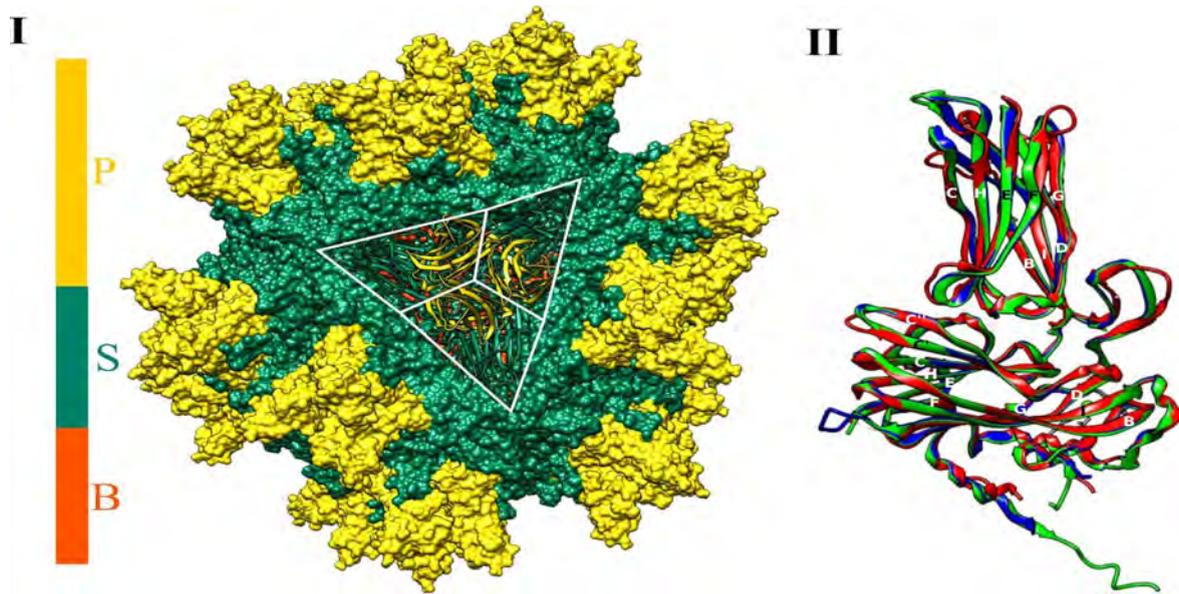


Figura 11. La proteína VP2 de la cápside de Birnaviridae. (I) Las proteínas VP2 de Birnavirus (PDB: 1wcd) se acomodan en trímeros para formar la cápside. La VP2 consta de tres dominios, el dominio B (naranja) compuesto por alfas, el dominio S (verde) que se ubica en la cápside y el dominio P (amarillo) que se proyecta hacia el exterior de la cápside. Tanto el S como el P adoptan la topología del jelly roll. (II) Comparación de las estructuras terciarias de tres VP2 de Birnavirus, realizada con FATCAT (<http://fatcat.sanfordburnham.org/>). Las estructuras azul (PDB: 1wcd) y verde (PDB: 3fbm) corresponden a la VP2 del Virus de la enfermedad de la bursitis infecciosa, la estructura coloreada de rojo (PDB: 3ide) es una VP2 del Virus necrosis pancreática infecciosa. Imagen producida con el programa UCSF Chimera (<http://www.rbvi.ucsf.edu/chimera/>).

Algo que llama la atención en el dendograma basado en la comparación de los dominios *jelly roll* virales, es la agrupación de la familia Alphatetraviridae [ssRNA(+)], que tiene una T=3 e infecta vertebrados, con la familia Birnaviridae (dsRNA), que tiene, en cambio, una T=13 e infecta invertebrados (Fig. 8).

Se ha visto que el subdominio S de la proteína VP2 de Birnaviridae (dsRNA) es más parecido a las proteínas de cápside de virus de RNA de cadena positiva que a los virus de RNA de doble cadena (Zeddman *et al.*, 2010). Esto coincide con lo que señalan los dendogramas basados en las comparaciones de dominios *jelly roll*, puesto que la proteína de cápside de Alphatetraviridae es agrupada con el subdominio S (Fig. 12).

Esta similitud del subdominio S de la VP2 de Birnaviridae con la CP de Alphatetraviridae ha llevado a proponer un origen común para las dos familias virales a partir de un ancestro con genoma de RNA de cadena sencilla positiva, que posteriormente dio origen a algunas de las familias de virus de RNA de doble cadena como Birnaviridae (Coulibaly *et al.*, 2005).

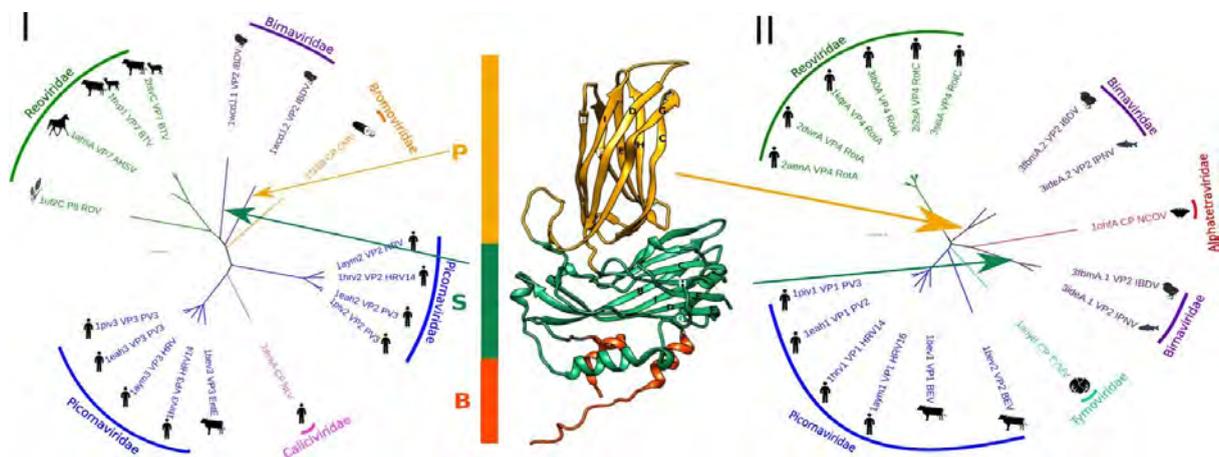


Figura 12. Los dominios S y P de la VP2 de Birnaviridae en los dendogramas basados en la comparación de estructuras de proteínas virales que presentan el dominio jelly roll. Se muestran los dendogramas de las proteínas virales sin ligandos (I) y con ligandos (II). En el centro se encuentra la estructura de la VP2 de Birnaviridae y se señalan sus tres dominios: B en naranja, S en verde y P en amarillo. Las flechas indican la posición de los dominios S (flechas verdes) y los dominios P (flechas amarillas) en los dendogramas. En cada rama se señala el código PDB y el nombre de la proteína, así como el virus al que corresponde. En negro se indica el hospedero al cual infecta cada uno de los virus. RotA: rotavirus A; RotC: rotavirus C; IBDV: virus de la enfermedad de bursitis infecciosa; IPNV: virus de necrosis pancreática infecciosa; NCOV: virus omega de nudaurelia capensis; TYMV: virus del mosaico amarillo del nabo; BEV: enterovirus bovino; HRV16: rhinovirus humano 16; HRV14: rhinovirus humano 14; PV2: poliovirus 2; PV3: poliovirus 3; RDV: virus del enanismo del arroz; AHSV: virus de la enfermedad equina africana; BTV: virus de la lengua azul; CMV: virus del mosaico del pepino; NLV: virus de Norwalk; EntE: enterovirus E.

4.4 Análisis filogenéticos de los dominios jelly roll celulares

La topología de los dendogramas de dominios *jelly roll* celulares mostró ser sensible al tamaño de la muestra. Ya que el dendograma de estructuras tridimensionales con ligands (n= 59, Fig. 13) no agrupa a las familias proteínicas del mismo modo que el dendograma de estructuras terciarias sin ligandos (n= 19, Fig. 14). En una prueba realizada en la que se compararon 63 de los 78 estructuras terciarias, el dendograma solamente agrupa a algunas familias como las TNF, Glyco_hydro_11 y 12, Lectinas y algunas oxigenasas (Fig. S2), lo que podría deberse al efecto del tamaño de la muestra o bien a un efecto de los ligandos.

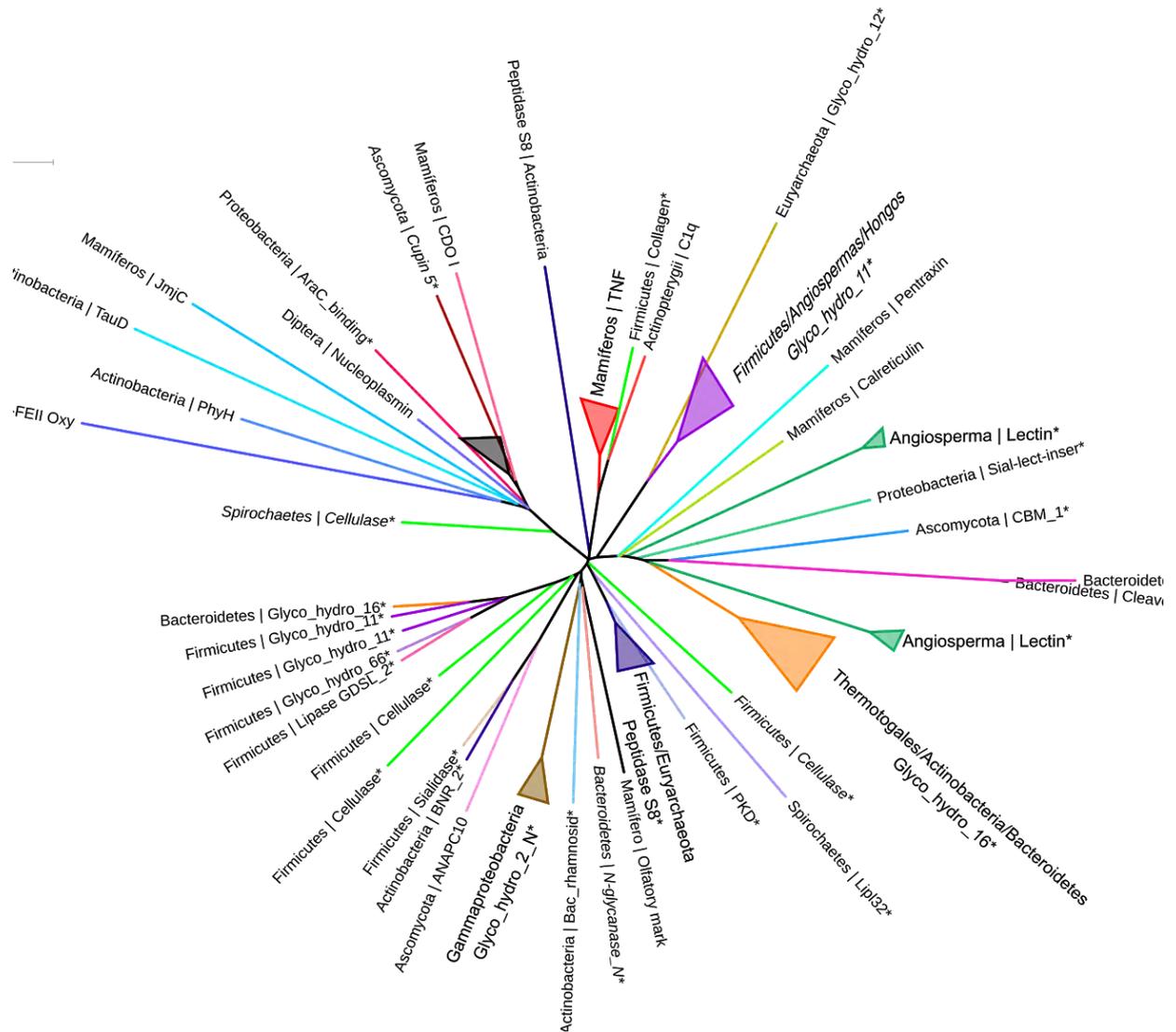


Figura 13. Dendograma basado en la comparación de estructuras terciarias de proteínas celulares con el dominio jelly roll con ligandos. Cada familia de proteínas es señalada con un color específico. Al final de cada rama se indica el grupo taxonómico al que corresponden las proteínas usadas, seguido del nombre de la familia proteínica. Las proteínas celulares cuya función se relaciona con los carbohidratos están marcadas con un asterisco (*).

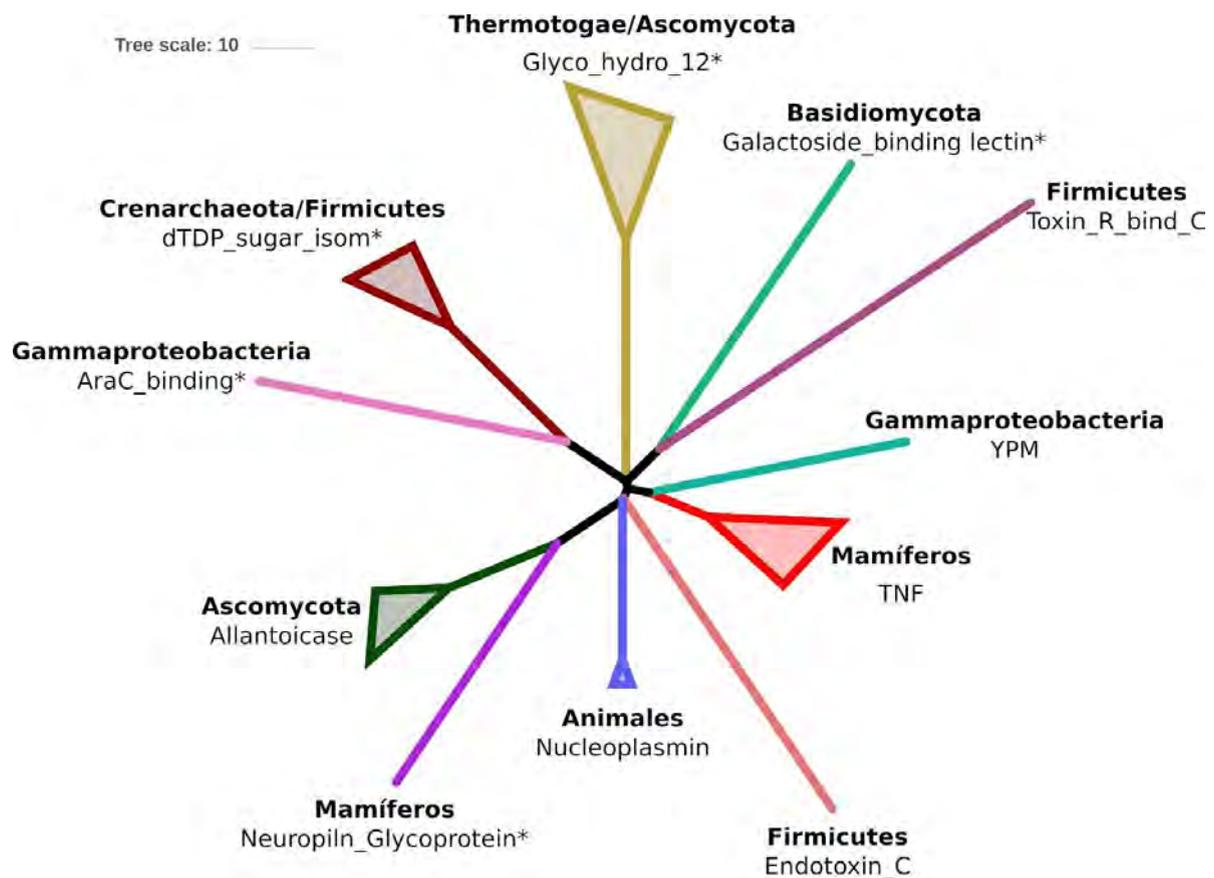


Figura 14. Dendograma basado en la comparación de estructuras terciarias del dominio *jelly roll* de proteínas celulares *sin ligandos*. Cada familia de proteínas es señalada con un color específico. Al final de cada rama se indica el grupo taxonómico al que corresponden las proteínas usadas, seguido del nombre de la familia proteínica. Las proteínas celulares cuya función se relaciona con los carbohidratos son marcadas con un asterisco (*).

4.5 Las inserciones del dominio *jelly roll* son más variables en las proteínas celulares

El dominio *jelly roll* está ampliamente distribuido tanto taxonómicamente como funcionalmente (Krupovic & Koonin, 2017). Por ello no resulta del todo sorprendente encontrarse con un amplio número de variaciones estructurales dentro de éste, que a pesar del diferente grado de inserciones continúan presentando las ocho cadenas beta conectadas de la forma específica del *jelly roll* (Fig. 15).

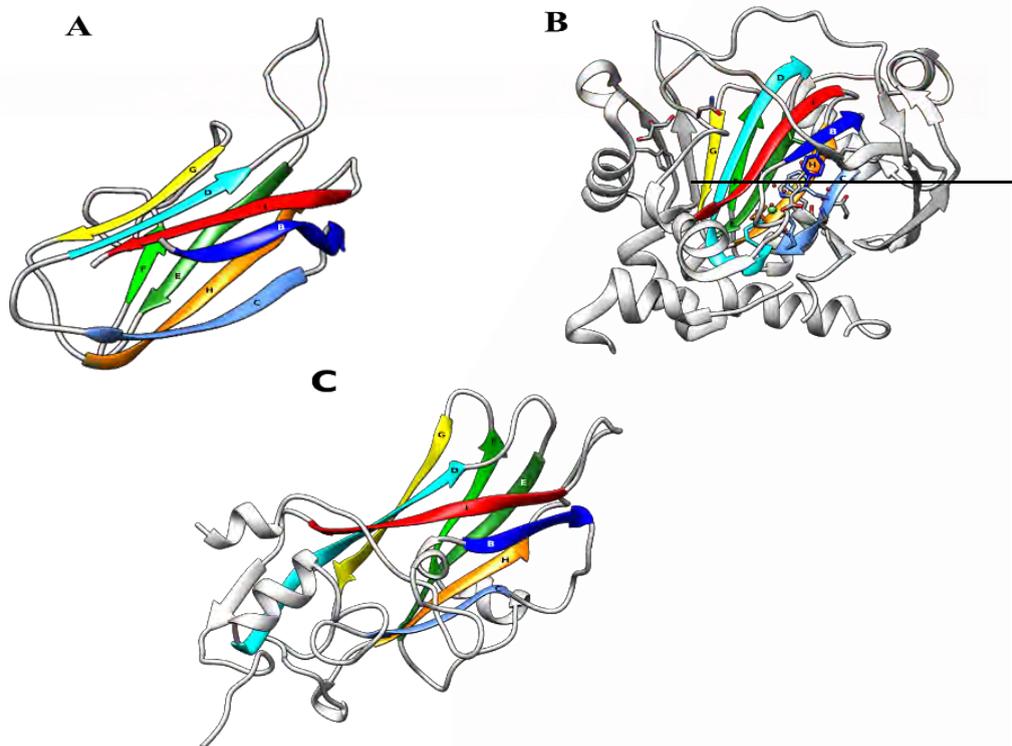


Figura 15. Inserciones en el dominio jelly roll. Dentro de las proteínas que presentan el dominio jelly roll hay una gran variedad de inserciones. Por ejemplo, (A) la proteasa de *Bacillus sp.* (PDB:1wmd) que mantiene la forma canónica del dominio jelly roll. Está el caso extremo de la (B) 2-Oxoglutarato oxigenasa de *Homo sapiens* (PDB: 2w2i) que tiene una gran cantidad de inserciones, y la (C) VP2 de Poliovirus tipo 3 (PDB:1piv) en la que la cantidad de inserciones no es excesiva pero sí notoria. Imagen producida con el programa UCSF Chimera (<http://www.rbvi.ucsf.edu/chimera/>). En las imágenes se muestran tres ejemplos de dominios jelly roll, las betas que conforman la estructura canónica son etiquetadas como BIDG-CHEF y las inserciones son señladas en color gris.

A diferencia de los alineamientos de secuencias, las herramientas de comparación de estructuras terciarias aún no implementan el uso de *gaps*, por lo que resulta problemático para el programa detectar las regiones conservadas cuando hay una gran cantidad de inserciones. Por ejemplo, cuando se intentaron las comparaciones de las proteínas con la herramienta *Multiseq* de *VMD* (*Visual Molecular Dynamics*, <http://www.scs.illinois.edu/schulten/multiseq/>), un paquete de visualización y análisis de moléculas que permite la comparación de estructuras terciarias y construcción de dendogramas, no fue posible llevar a cabo dicho análisis debido a que el programa solamente puede comparar moléculas estrictamente similares (Roberts *et al.*, 2006), y aunque todas las proteínas tienen en común el dominio *jelly roll*, *Multiseq* no fue capaz de reconocerlo como una región conservada. Por el contrario, cuando se realizaron comparaciones del mismo tipo

de proteína en una familia viral (la VP4 de Reoviridae, por ejemplo), el programa fue capaz de realizar las comparaciones.

Se encontró que las variaciones en la longitud de las hebras y de las “asas”, así como las inserciones, son de distinta longitud y naturaleza. En ocasiones, las proteínas presentan un *jelly roll* canónico y sin inserciones, como es el caso de una proteasa de *Bacillus* sp. (Fig. 15A). El ejemplo extremo a éste es el de algunas oxigenasas como la 2-oxoglutarato de *Homo sapiens* (Fig. 15B), que conserva el *jelly roll* en el centro de la estructura, pero éste presenta varias inserciones entre sus conexiones.

Aunque las proteínas celulares poseen la mayor variedad de inserciones en el dominio *jelly roll*, también se encuentran inserciones en los *jelly rolls* de proteínas virales. Por ejemplo, en la VP2 de Poliovirus tipo 3 (Fig. 15C) el número de inserciones parece estar entre las que presentan la proteasa y la oxigenasa mencionadas arriba.

4.6 Las funciones de las proteínas celulares con el dominio *jelly roll*

Las proteínas celulares con dominio *jelly roll* están involucradas en procesos celulares como la defensa inmunológica. Por ejemplo, tenemos tres familias de proteínas, las Pentraxinas, las C1q y las TNF's que llevan a cabo la activación de reacciones de defensa como la respuesta inmunológica aguda, la activación del sistema complementario del suero, inician la reacción de inflamación y la unión a proteínas, respectivamente (Anexos: Tabla S2; Fig. 13 y 14; Emsley *et al.*, 1994; Gravestain & Borst, 1998; Miao *et al.*, 2014).

En lo que se refiere a la unión a receptores, algunas proteínas como la endotoxina C de *Bacillus thuringiensis* y la toxina R de *Clostridium botulinum* desempeñan un papel importante durante la infección, dado que llevan a cabo el reconocimiento y la unión al receptor de sus hospederos (Morse *et al.*, 2001; Zhang *et al.*, 2010). Otras proteínas, como el mitógeno YPM de *Yersinia pseudotuberculosis* y el marcador olfativo de *Mus musculus*, participan en procesos que involucran la unión a receptores, tales como la activación de células T y en la transducción de señales olfativas (Smith *et al.*, 2001; Donadini *et al.*, 2004).

Otros procesos celulares en los que participan las proteínas con *jelly roll* son la regulación de la anafase de la mitosis por una familia de proteínas denominada ANAPC10 (Au *et al.*, 2002), la formación del centrómero llevada a cabo por la familia de proteínas llamada CENP-C_C (Cohen *et al.*, 2008), la desmetilación de histonas en la que participa la familia JmjC (Hillrinhaus *et al.*, 2011), así como la unión a histonas y la transferencia de DNA en la que participan las nucleoplasminas (Frehlick *et al.*, 2007).

Las 2-oxoglutarato oxigenasas y las oxigenasas dependientes de hierro que llevan a cabo reacciones óxido-reducción, presentan un plegamiento tipo *jelly roll* con una gran cantidad de inserciones (Fig. 15B) en el que no solamente se da la interacción con el Fe sino que también se encuentra el sitio catalítico (Aik *et al.*, 2012).

Algunos procesos metabólicos en los que participan las proteínas celulares con *jelly roll* son el metabolismo de purinas por las *Allantoicase* (Leulliot *et al.*, 2004), la actividad endopeptidasa tipo serina llevada a cabo por la *Peptidase S8* (Foophow *et al.*, 2010), y la conversión de cisteína en ácido cisteinesulfínico por la *Cysteine dioxygenase type I* ó *CDO_I* (Simmons *et al.*, 2006; Anexos: Tabla S2).

4.6.1 La prevalencia de funciones relacionadas con carbohidratos de las proteínas celulares con dominios *jelly roll*

Muchas de las familias de proteínas de la muestra llevan a cabo funciones relacionadas con carbohidratos, entre éstas están la hidrólisis de enlaces glicosídicos (Henrissat *et al.*, 1995; Boraston *et al.*, 2007), la formación de uniones celulares mediante proteínas glicosiladas (Li *et al.*, 2011), la regulación de la apoptosis mediante unión a carbohidratos (Yang *et al.*, 2009), la biosíntesis de dTDP-I-rhamnosa (Dong *et al.*, 2003), la catálisis de la alfa oxidación del ácido fitánico (McCulloch *et al.*, 2015), la hidrólisis de lípidos (Montanier *et al.*, 2009), el almacenamiento de carbohidratos en plantas (Dunwell *et al.*, 2004), y la unión a carbohidratos (Jacobson *et al.*, 2013) como el ácido siálico (Gaskell *et al.*, 1995), la arabinosa, la galactosa y la fructosa (Thamotharan *et al.*, 2011).

Los oligosacáridos son muy abundantes y diversos en la naturaleza. Éstos median muchos fenómenos biológicos importantes que incluyen funciones estructurales, la provisión de energía al organismo, y el reconocimiento específico tanto endógeno como exógeno, es decir, la diferenciación entre receptores del mismo organismo y receptores no pertenecientes al organismo, sino de patógenos o simbioses (Gagneux & Varki, 1999).

Muchos patógenos que infectan organismos multicelulares, como virus y bacterias, inician la infección mediante una interacción con las glicoproteínas de membrana localizadas en algunos tejidos expuestos y en células con funciones endógenas (Gagneux & Varki, 1999), como los epitelios gastrointestinales y nasales en vertebrados (Kagnoff & Eckmann, 1997; Philpott *et al.*, 2001; Borza *et al.*, 2004), y los linfocitos T CD4 en la infección del VIH (Sattentau & Moore, 1993).

Lo mismo ocurre con los virus de RNA icosaédricos, que en su mayoría infectan invertebrados y plantas (Campillo-Balderas *et al.*, 2015), puesto que muchos de ellos utilizan sacáridos como receptores al iniciar la infección de sus hospederos, además de usar para esto a células del tracto respiratorio o intestinal, así como algunas del sistema inmune (Tabla 2S).

Uno de los sacáridos más comunes entre los glicanos de vertebrados es el ácido siálico. Está presente sobre todo en los deuterostomados, aunque, también se encuentra en algunos de sus patógenos bacterianos. La presencia de esta molécula en dos linajes tan distantes sugiere que se trata de un mimetismo generado por convergencia evolutiva. Esta hipótesis se sostiene en que los genes de los patógenos que codifican para la síntesis y transferencia del ácido siálico no son homólogos a sus contrapartes en los animales (Gagneux & Varki, 1999). Además, un patógeno que expresa ácido siálico reduce mucho la respuesta inmune generada en el hospedero (Moxon *et al.*, 1994).

Entre las proteínas celulares con dominio *jelly roll* se encuentran tres sialidasas, la BNR_2 de *Micromonospora viridifaciens* (PDB: 1eut), la sialidasa de *Clostridium perfringens* (PDB: 2v72) y la sialidasa con inserción de lectina de *Vibrio cholerae* (PDB: 1kit). Estas moléculas se encargan de remover los residuos terminales de ácido siálico en glicoconjugados, se las ha relacionado con la patogenicidad de varias enfermedades, aunque también son producidas por

organismos no patógenos, así como con la nutrición celular y con las interacciones celulares (Taylor, 1996; Severi *et al.*, 2007).

Previamente se había reportado la recurrencia del *jelly roll* en las glicoproteínas. Incluso se propuso que los virus habían reclutado las proteínas de cápside de sus hospederos, lo que permitió la protección de su material genético y la capacidad de reconocer los azúcares en las superficies de sus hospederos (Krupovic y Koonin, 2017).

4.7 Los dominios jelly roll virales y celulares con similitud estructural

En el dendograma de dominios *jelly roll* celulares y virales sin ligandos (Fig. 16), únicamente se agruparon tres familias de proteínas celulares con proteínas virales: la familia 12 de glucósido hidrolasas, la familia de factores de necrosis tumoral (TNF) y el mitógeno derivado de *Yersinia* (YPM).

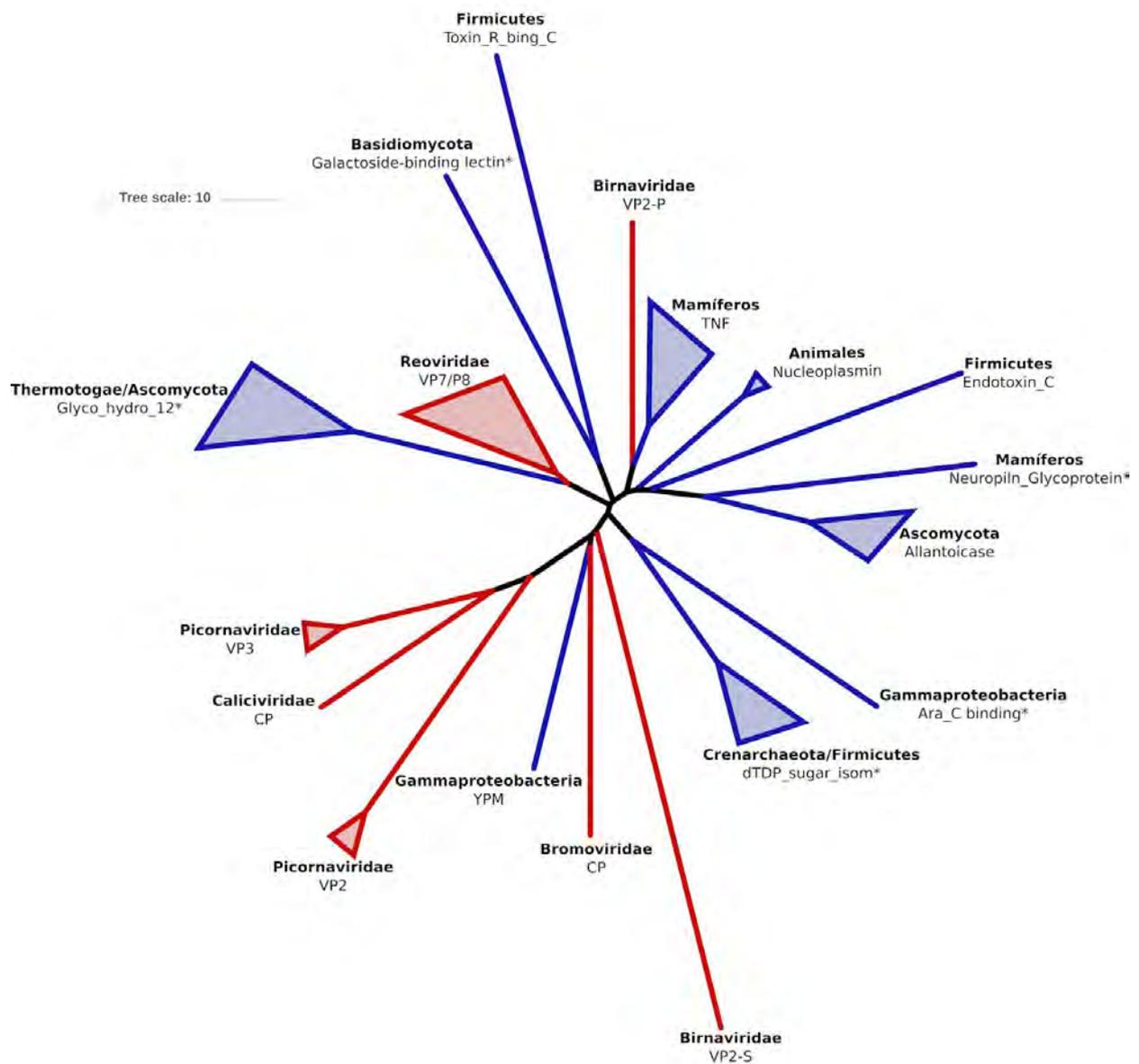


Figura 16. Dendrograma de los jelly rolls de proteínas virales y celulares sin ligando. Los dominios jelly roll de las proteínas virales se muestran en rojo y los de las proteínas celulares en azul. Las proteínas celulares cuya función se relaciona con los carbohidratos son marcadas con un asterisco (*).

4.8 Homología entre los dominios jelly roll del mitógeno derivado de *Yersinia* y la CP de *Bromoviridae*

Cuando se comparan solamente *jelly rolls* celulares, el mitógeno derivado de *Yersinia* (YPM) se agrupa con los factores de necrosis tumoral (TNF) (Fig. 14), pero cuando se comparan *jelly rolls* virales con celulares, YPM se asocia con la proteína de cápside de *Bromoviridae* (Fig. 16).

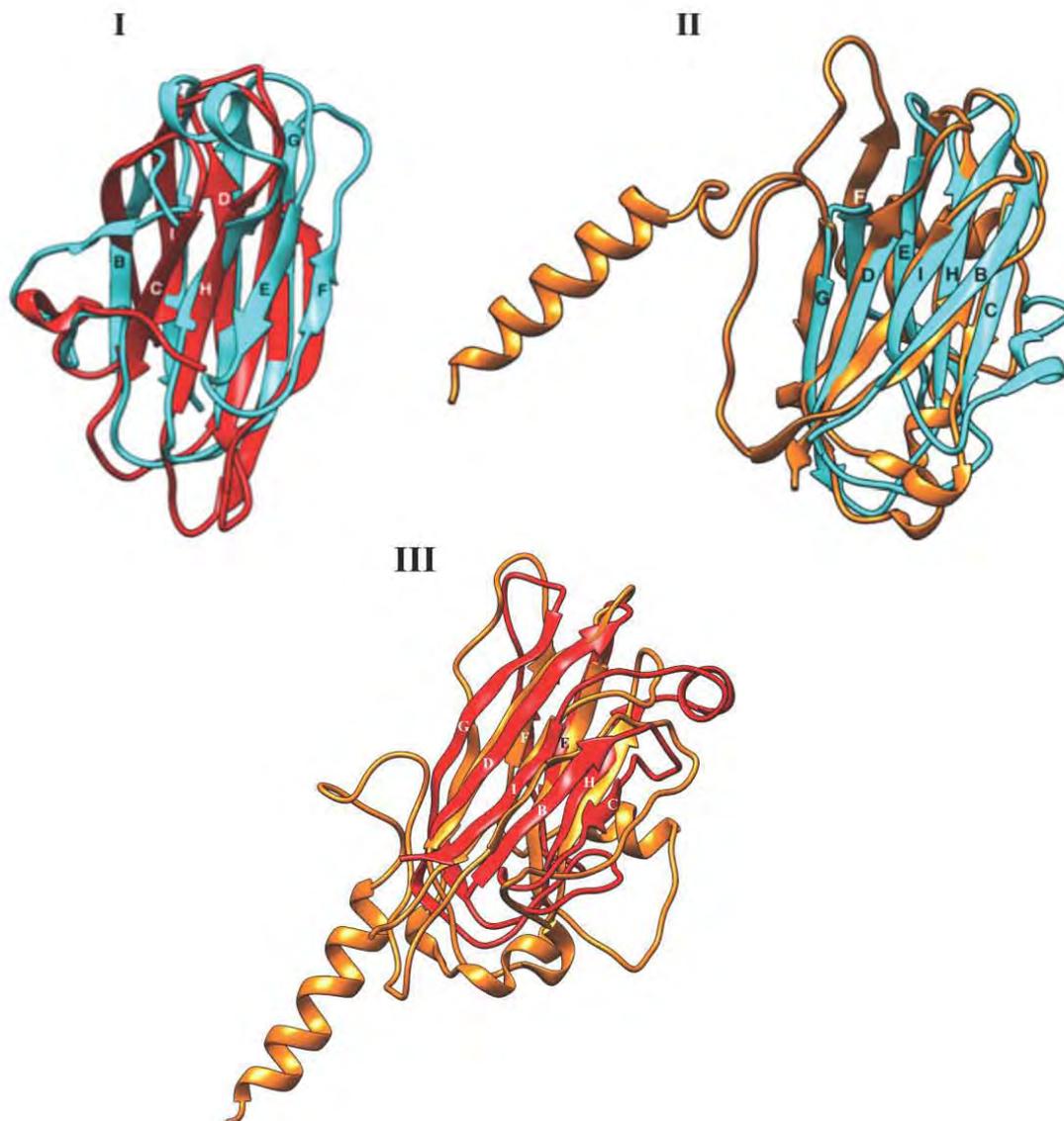


Figura 17. Comparación estructural de el mitógeno derivado de Yersinia con un factor de necrosis tumoral (I) y con la proteína de cápside (CP) de Bromoviridae (II). YPM (PDB= 1pm4; 116 residuos de aminoácidos; color cyan), y TNF (PDB= 2hey; 117 residuos de aminoácidos; color rojo), resultaron significativamente similares (RMSD= 3.06 Å; $P= 1.04e-02$; y 89 posiciones equivalentes), al igual que la CP de Bromoviridae (PDB= 1j15; 178 residuos de aminoácidos; color naranja) y YPM (RMSD= 3.19 Å; $P= 7.98e-04$; y 99 posiciones equivalentes). Mientras que la comparación TNF-CP obtuvo un valor no significativo (RMSD de 3.30 Å; $P= 3.12e-01$; y 97 posiciones equivalente con dos giros). Las comparaciones fueron realizadas con ASH Structural Superposition (<https://sysimm.org/ash/>). Imagen producida con el programa UCSF Chimera (<http://www.rbvi.ucsf.edu/chimera/>).

Estos tres tipos de proteínas no sólo comparten el tipo de plegamiento *jelly roll*, sino que también forman trímeros, además de que YPM tiene un gran parecido estructural con la

proteína de cápside del virus del mosaico del tabaco. Ésto se reportó cuando se determinó la estructura terciaria de YPM (Donadini *et al.*, 2004). El mismo grupo de investigadores menciona el parecido que este mitógeno guarda con TNF y otras proteínas virales, como la de cápside de Bromoviridae. Lo cual concuerda con lo observado en los dendogramas y los resultados obtenidos al comparar el YPM con un TNF y con la proteína de cápside de Bromoviridae (Fig. 17).

4.9 El origen celular del subdominio P de la proteína de cápside VP2 de Birnaviridae

La similitud entre el subdominio P de la proteína VP2 de Birnaviridae con las proteínas TNF es evidenciado en el dendograma de dominios *jelly roll* virales y celulares (Fig. 16) y la comparación en FatCat arrojó que son significativamente similares (Fig. 18). Por lo que sabemos, se ha observado una relación biológica entre ambos, ya que la producción de proteínas TNF aumenta cuando un Birnaviridae infecta a su hospedero (Brieland *et al.*, 2001; Rautenschlein *et al.*, 2002), aunque esta relación no deja clara la homología entre ambas, la similitud estructural apoya la hipótesis del origen común de TNF y el subdominio P de VP2 de Birnaviridae.

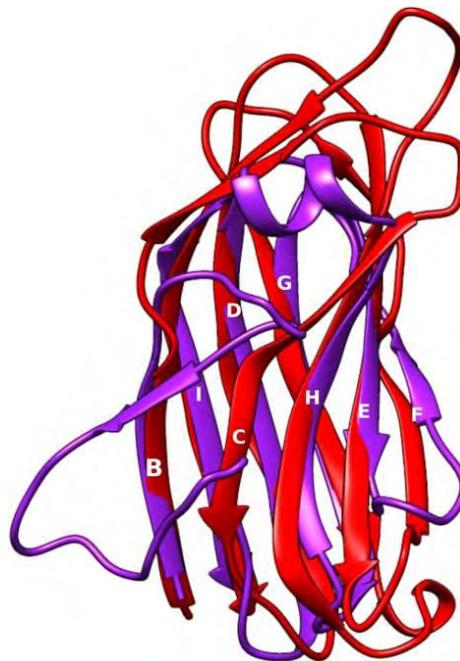


Figura 18. El dominio P de la VP2 de Birnaviridae y TNF. Se muestra un TNF de *Mus musculus* (PDB: 2hey; 117 residuos de aminoácidos; rojo) y el dominio P de la VP2 de Birnaviridae (PDB: 1wed; 135 residuos de aminoácidos; violeta) que son significativamente similares (RMSD= 3.04; P= 4.43e-03; y 95 posiciones equivalentes). Imagen producida con el programa UCSF Chimera (<http://www.rbvi.ucsf.edu/chimera/>). La comparación se realizó en el servidor en línea de FATCAT (<http://fatcat.sanfordburnham.org>).

5. Conclusiones

La compilación de un catálogo de proteínas virales y celulares que presentan el dominio *jelly roll* más la información biológica disponible sobre las proteínas de la muestra permitieron realizar un análisis general sobre la evolución del dominio. La construcción de los dendrogramas basados en estructuras terciarias mostró que el *jelly roll* es un buen marcador filogenético viral a nivel de familia. Además permitieron reconocer la homología entre las proteínas CP de Caliciviridae y la VP3 de Picornaviridae.

Los dendrogramas basados en comparaciones de estructuras terciarias de dominios *jelly roll* celulares mostraron sensibilidad al tamaño de la muestra y posiblemente a la presencia de ligandos. No obstante, al ser comparados con los dominios virales se evidenciaron algunas homologías entre proteínas como YPM y CP de Bromoviridae, así como la CP de Alphetetraviridae y el subdominio S de la VP2 de Birnaviridae.

El caso de la VP2 de Birnaviridae resulta interesante debido a la diferencia estructural que muestran sus subdominios S y P, que parecen tener historias evolutivas distintas. El subdominio S por un lado muestra haberse originado a partir de una proteína ancestral que también dio origen a CP de Alphetetraviridae. Lo que concuerda con la hipótesis de Coulibaly y colaboradores, 2005, del origen común de Alphetetraviridae y Birnaviridae a partir de un ancestro con genoma de RNA de cadena sencilla positiva, que posteriormente dio origen a algunas de las familias de virus de RNA de doble cadena como Birnaviridae (Coulibaly *et al.*, 2005). Por otro lado, la similitud del *jelly roll* del subdominio P con el *jelly roll* de la familia de factores de necrosis tumoral (TNF) indica que este tuvo un origen celular probablemente a partir del secuestro de genes de su hospedero.

Los análisis realizados en este trabajo indican que los dominios *jelly roll* de virus de RNA tienen un origen celular, lo que concuerda con la hipótesis del origen celular de los virus o hipótesis del escape de genes. Muestra de esto es la distribución filogenética de los hospederos de los virus de RNA, siendo principalmente el grupo celular de origen más reciente, los eucariontes, con excepción de algunas proteobacterias comunes en la microbiota de muchos animales, por lo que podría deberse a una transferencia.

Cabe destacar que aunque esperábamos una mayor diversidad entre los dominios *jelly roll* virales, puesto que las RNA polimerasas dependientes de RNA tienen una alta tasa de mutación, comparadas con las DNA polimerasas, encontramos que el grado de inserciones de los dominios *jelly roll* celulares es mucho mayor. Algunas posibles explicaciones son la relevancia de la estabilidad y la forma de las proteínas de cápside, que incluyen a aquellas presentan el *jelly roll*, en el ensamblaje de las cápsides; y importancia de la estructura de las proteínas de cápside, en la óptima infección del hospedero. Hacen falta estudios experimentales que permitan responder ésto.

Lo anterior se relaciona con la recurrencia de funciones relacionadas con carbohidratos en las proteínas celulares, debido a la importancia de estas biomoléculas en el reconocimiento celular, en procesos inmunológicos y su prevalencia en epitelios (a los que los virus muestran un fuerte tropismo), resulta sencillo pensar que el secuestro de proteínas celulares glicosiladas (como son la mayoría de las proteínas que presentan el dominio *jelly roll*) resultara beneficioso para el virus.

6. Literatura citada

- Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R., & Schofield, C. J. (2012). Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Current Opinion in Structural Biology*, 22(6), 691–700. <https://doi.org/10.1016/j.sbi.2012.10.001>
- Arshan Nasir, Kyung Mo Kim & Gustavo Caetano-Anollés (2012) Viral evolution, Mobile Genetic Elements, 2:5, 247-252, DOI: 10.4161/mge.22797
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews*, 35(3), 235–241.
- Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., ... Zhang, J. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 45(D1), D158–D169. <https://doi.org/10.1093/nar/gkw1099>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28(1), 235–242. <https://doi.org/10.1093/nar/28.1.235>
- Borza, C. M., Morgan, A. J., Turk, S. M., & Hutt-Fletcher, L. M. (2004). Use of gHgL for Attachment of Epstein-Barr Virus to Epithelial Cells Compromises Infection. *Journal of Virology*, 78(10), 5007–5014. <https://doi.org/10.1128/JVI.78.10.5007-5014.2004>
- Brieland, J., Essig, D., Jackson, C., Frank, D., Loebenberg, D., Menzel, F., ... Domenico, B. D. I. (2001). Comparison of Pathogenesis and Host Immune Responses to *Candida glabrata* and *Candida albicans* in Systemically Infected Immunocompetent Mice Comparison of Pathogenesis and Host Immune Responses to *Candida glabrata* and *Candida albicans* in Systemically Infec. *Infection and Immunity*, 69(8), 5046–5055. <https://doi.org/10.1128/IAI.69.8.5046>
- Bubanovic, I., Najman, S., & Andjelkovic, Z. (2005). Origin and evolution of viruses: Escaped DNA/RNA sequences as evolutionary accelerators and natural biological weapons. *Medical Hypotheses*, 65(5), 868–872. <https://doi.org/10.1016/j.mehy.2005.05.038>
- Büchen-Osmond, C. (2003). The universal virus database ICTVdB. *Computing in Science and Engineering*, 5(3), 16–25. <https://doi.org/10.1109/MCISE.2003.1196303>
- 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(December). <https://doi.org/10.3389/fevo.2015.00143>
- Campillo-Balderas, J. A. (2018). Origen y evolución temprana de los virus y su relación con el último ancestro común de los seres vivos. Universidad Nacional Autónoma de México.

- Černý, J., Bolfíková, B. Č., Valdés, J. J., Grubhoffer, L., & Růžek, D. (2014). Evolution of tertiary structure of viral RNA dependent polymerases. *PLoS ONE*, *9*(5). <https://doi.org/10.1371/journal.pone.0096070>
- Černý, J., Černá Bolfíková, B., Paolo, P. M., Grubhoffer, L., & Růžek, D. (2015). A deep phylogeny of viral and cellular right-hand polymerases. *Infection, Genetics and Evolution*, *36*, 275–286. <https://doi.org/10.1016/j.meegid.2015.09.026>
- Chelvanayagam, G., Heringa, J., & Argos, P. (1992). Anatomy and evolution of proteins displaying the viral capsid jellyroll topology. *Journal of Molecular Biology*, *228*(1), 220–242. [https://doi.org/10.1016/0022-2836\(92\)90502-B](https://doi.org/10.1016/0022-2836(92)90502-B)
- Cheng, S., & Brooks, C. L. (2013). Viral Capsid Proteins Are Segregated in Structural Fold Space. *PLoS Computational Biology*, *9*(2). <https://doi.org/10.1371/journal.pcbi.1002905>
- Cheng, Y. S., Ko, T. P., Huang, J. W., Wu, T. H., Lin, C. Y., Luo, W., Li, Q., Ma, Y., Huang, C., Wang, A., Liu, A., & Guo, R. T. (2012). Enhanced activity of *Thermotoga maritima* cellulase 12A by mutating a unique surface loop. *Applied Microbiology and Biotechnology*, *95*(3), 661–669. <https://doi.org/10.1007/s00253-011-3791-4>
- Coulibaly, F., Chevalier, C., Gutsche, I., Pous, J., Navaza, J., Bressanelli, S., Delmas, B., & Rey, F. A. (2005). The birnavirus crystal structure reveals structural relationships among icosahedral viruses. *Cell*, *120*(6), 761–772. <https://doi.org/10.1016/j.cell.2005.01.009>
- Delgui, L., González, D., & Rodríguez, J. F. (2009). Infectious bursal disease virus persistently infects bursal B-lymphoid DT40 cells. *Journal of General Virology*, *90*(5), 1148–1152. <https://doi.org/10.1099/vir.0.008870-0>
- Duffy, S., Shackelton, L.A., and Holmes, E.C. (2008). Rates of evolutionary change in viruses: Patterns and determinants. *Nature Reviews Genetics* *9*, 267–276.
- Donadini, R., Liew, C. W., Kwan, A. H. Y., Mackay, J. P., & Fields, B. A. (2004). Crystal and Solution Structures of a Superantigen from *Yersinia pseudotuberculosis* Reveal a Jelly-Roll Fold. *Structure*, *12*(1), 145–156. <https://doi.org/10.1016/j.str.2003.12.002>
- Emsley, J., White, H., O'Hara, B., Oliva, G., Srinivasan, N., Tickle, I., Blundell, T., Pepys, & M., Wood, S. (1994). Structure of pentameric human serum amyloid P component. *Nature*, *367*(6461), 338–345. <https://doi.org/10.1038/367338a0>
- Finn, R. D. (2006). Pfam: clans, web tools and services. *Nucleic Acids Research*, *34*(90001), D247–D251. <https://doi.org/10.1093/nar/gkj149>
- Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., ... Mitchell, A. L. (2017). InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Research*, *45*(D1), D190–D199. <https://doi.org/10.1093/nar/gkw1107>

- Flint, J. (Princeton U., Racaniello, V. R. (Columbia U., Rall, G. F. (Fox C. C. C., & Shalka, A. M. (Fox C. C. C. (2015). *Principles of Virology* (4rd ed.), Chapter 4. ASM Press.
- Foophow, T., Tanaka, S. ichi, Angkawidjaja, C., Koga, Y., Takano, K., & Kanaya, S. (2010). Crystal structure of a subtilisin homologue, Tk-SP, from thermococcus kodakaraensis: Requirement of a C-terminal ??-Jelly Roll domain for hyperstability. *Journal of Molecular Biology*, 400(4), 865–877. <http://doi.org/10.1016/j.jmb.2010.05.064>
- Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Research*, 117(1), 5–16. <https://doi.org/10.1016/j.virusres.2006.01.010>
- Gagneux, P., & Varki, A. (1999). Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology*, 9(8), 747–755. <https://doi.org/10.1093/glycob/9.8.747>
- Green, K. Y., Ando, T., Balayan, M. S., Berke, T., Clarke, I. N., Estes, M. K., Matson, D. O., Nakata, S., Neill, J. D., Studdert, M. J., Thiel, H. (2000). Taxonomy of the Caliciviruses. *The Journal of Infectious Diseases*, 181(s2), S322–S330. <https://doi.org/10.1086/315591>
- Holm, L., & Sander, C. (1995). Dali: a network tool for prtoein structure comparison. *Trends Biochem Sci*, 20(June), 478–480.
- Holm, L., & Laakso, L. M. (2016). Dali server update. *Nucleic Acids Research*, 44(W1), W351–W355. <https://doi.org/10.1093/nar/gkw357>
- Holmes, E. C. (2009). *The Evolution and Emergence of RNA Viruses*. Oxford University Press, Chapter 1 y 2, 1-36.
- Huang, C. C., Couch, G. S., Pettersen, E. F., & Ferrin, T. E. (1996). Chimera: An Extensible Molecular Modeling Application Constructed Using Standard Components. *Pacific Symposium on Biocomputing*, 1, 724. Retrieved from <http://www.cgl.ucsf.edu/chimera>
- Hulo, C., De Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., & Le Mercier, P. (2011). ViralZone: A knowledge resource to understand virus diversity. *Nucleic Acids Research*, 39(SUPPL. 1), 576–582. <https://doi.org/10.1093/nar/gkq901>
- Jácome R, Campillo-Balderas JA, Becerra A, López-Vidal Y, Lazcano A, Ponce de León S. A critical assessment of tertiary structure-based phylogenies: the case of hepatitis C virus and human immunodeficiency virus 1 polymerases (en preparación).
- Jácome, R., Becerra, A., De León, S. P., & Lazcano, A. (2015). Structural analysis of monomeric RNA- dependent polymerases: Evolutionary and therapeutic implications. *PLoS ONE*, 10(9), 1–26. <https://doi.org/10.1371/journal.pone.0139001>

- Kagnoff, M. F., & Eckmann, L. (1997). Perspectives Series : Host / Pathogen Interactions Epithelial Cells as Sensors for Microbial Infection. *The Journal of Clinical Investigations*, 100(1), 6–10. <https://doi.org/10.1172/JCI119522>
- Khademi, S., Zhang, D., Swanson, S. M., Wartenberg, A., Witte, K., & Meyer, E. F. (2002). Determination of the structure of an endoglucanase from *Aspergillus niger* and its mode of inhibition by palladium chloride. *Acta Crystallographica Section D: Biological Crystallography*, 58(4), 660–667. <https://doi.org/10.1107/S0907444902003360>
- Koonin, E. V., & Gorbalenya, A. E. (1992). An insect picornavirus may have genome organization similar to that of caliciviruses. *FEBS Lett.*, 297(12), 81–86.
- 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>
- Koonin, E. V., Senkevich, T. G., & Dolja, V. V. (2006). The ancient virus world and evolution of cells. *Biology Direct*, 1. <https://doi.org/10.1186/1745-6150-1-29>
- Krupovic, M., & Bamford, D. H. (2011). Protein Conservation in Virus Evolution. *Encyclopedia of Life Sciences (ELS)*, 1–10. <https://doi.org/10.1002/9780470015902.a0023265>
- Krupovic, M., & Koonin, E. V. (2017). Multiple origins of viral capsid proteins from cellular ancestors. *Proceedings of the National Academy of Sciences*, 114(12), E2401–E2410. <https://doi.org/10.1073/pnas.1621061114>
- Laskowski, R. A. (2009). PDBsum new things. *Nucleic Acids Research*, 37(SUPPL. 1), 355–359. <https://doi.org/10.1093/nar/gkn860>
- Lazcano, A. (2010). Origen Y Evolución De Los Virus: ¿Genes Errantes O Parásitos Primitivos? *Facultad de Ciencias, UNAM*, 34, 73–82.
- O. Le Gall, P. Christian, C.M. Fauquet, A.M.Q. King, N.J. Knowles, N. Nakashima, G. Stanway, A.E. Gorbalenya (2008) *Picornavirales*, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T=3 virion architecture. *Arch. Virol.* 153, 715–727.
- Leulliot, N., Quevillon-Cheruel, S., Sorel, I., Graille, M., Meyer, P., Liger, D., ... Van Tilbeurgh, H. (2004). Crystal structure of yeast allantoicase reveals a repeated jelly roll motif. *Journal of Biological Chemistry*, 279(22), 23447–23452. <http://doi.org/10.1074/jbc.M401336200>

- Lodish, H. F., Zhou, B., Liu, G., & Chen, C. Z. (2008, February). Micromanagement of the immune system by microRNAs. *Nature Reviews Immunology*. <https://doi.org/10.1038/nri2252>
- Moxon, E. R., Rainey, P. B., Nowak, M. A., & Lenski, R. E. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. *Current Biology*, *4*(1), 24–33. [https://doi.org/10.1016/S0960-9822\(00\)00005-1](https://doi.org/10.1016/S0960-9822(00)00005-1)
- Nguyen, A., & Van Nguyen, A. (2014). Eccrine porocarcinoma: A report of 2 cases and review of the literature. *Cutis*, *93*(1), 43–46. <https://doi.org/10.1016/j>
- Philpott, D. J., Girardin, S. E., & Sansonetti, P. J. (2001). Innate immune responses of epithelial cells following infection with bacterial pathogens. *Current Opinion in Immunology*, *13*(4), 410–416. [https://doi.org/10.1016/S0952-7915\(00\)00235-1](https://doi.org/10.1016/S0952-7915(00)00235-1)
- Prasad, B. V. V. (1999). X-ray Crystallographic Structure of the Norwalk Virus Capsid. *Science*, *286*(5438), 287–290. <https://doi.org/10.1126/science.286.5438.287>
- Prasad, B. V. V. (College of M. H., & Schmid, M. F. (College of M. H. (2012). *Principles of Virus Structural Organization* (Vol. 726). <https://doi.org/10.1007/978-1-4614-0980-9>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J Comput Chem*. <https://doi.org/10.1002/jcc.20084>
- Rautenschlein, S., Yeh, H. Y., Njenga, M. K., & Sharma, J. M. (2002). Role of intrabursal T cells in infectious bursal disease virus (IBDV) infection: T cells promote viral clearance but delay follicular recovery. *Archives of Virology*, *147*(2), 285–304. <https://doi.org/10.1007/s705-002-8320-2>
- Richardson, J. (1981). The anatomy and taxonomy of protein structure. *Advances in Protein Chemistry*, *Vol. 34*, 301-303, ISBN 0-12-034234-0
- Roberts, E., Eargle, J., Wright, D., & Luthey-Schulten, Z. (2006). MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics*, *7*, 1–11. <https://doi.org/10.1186/1471-2105-7-382>
- Rossmann, M. (Purdue U., & Johnson, J. E. (Purdue U. (1989). Icosahedral RNA virus structure. *Annual Review of Biochemistry*, *58*, 533–73.
- Sanger, F., & Tuppy, H. (1951). The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *The Biochemical Journal*, *49*(4), 463–81. [http://doi.org/10.1016/0006-3002\(53\)90071-7](http://doi.org/10.1016/0006-3002(53)90071-7)

- Sattentau, Q. J., & Moore, J. P. (1993). The role of CD4 in HIV binding and entry. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 342(1299), 59–66. <https://doi.org/10.1098/rstb.1993.0136>
- Severi, E., Hood, D. W., & Thomas, G. H. (2007). Sialic acid utilization by bacterial pathogens. *Microbiology*, 153(9), 2817–2822. <https://doi.org/10.1099/mic.0.2007/009480-0>
- Simmons, C. R., Liu, Q., Huang, Q., Hao, Q., Begley, T. P., Karplus, P. A., & Stipanuk, M. H. (2006). Crystal structure of mammalian cysteine dioxygenase: A novel mononuclear iron center for cysteine thiol oxidation. *Journal of Biological Chemistry*, 281(27), 18723–18733. <http://doi.org/10.1074/jbc.M601555200>
- Standley, D. M., Toh, H., & Nakamura, H. (2007). ASH structure alignment package: Sensitivity and selectivity in domain classification. *BMC Bioinformatics*, 8, 1–8. <https://doi.org/10.1186/1471-2105-8-116>
- Stirk, H. J., Woolfson, D. N., Hutchinson, E. G., & Thornton, J. M. (1992). Depicting topology and handedness in jellyroll structures. *FEBS Letters*, 308(1), 1–3. [https://doi.org/10.1016/0014-5793\(92\)81036-L](https://doi.org/10.1016/0014-5793(92)81036-L)
- Taylor, G. (1996). Sialidases: Structures, biological significance and therapeutic potential. *Current Opinion in Structural Biology*, 6(6), 830–837. [https://doi.org/10.1016/S0959-440X\(96\)80014-5](https://doi.org/10.1016/S0959-440X(96)80014-5)
- Tobergte, D. R., & Curtis, S. (2013). Bioinformatics (Sequence and Genome Analysis). *Journal of Chemical Information and Modeling*, 53(9), 1689–1699. <http://doi.org/10.1017/CBO9781107415324.004>
- Woolhouse MEJ, Adair K, Brierley L. 2013. RNA viruses: a case study of the biology of emerging infectious diseases. *Microbiol Spectrum* 1(1):OH-0001-2012. doi:10.1128/microbiolspec.OH-0001-2012.
- Ye, Y., & Godzik, A. (2004). FATCAT: A web server for flexible structure comparison and structure similarity searching. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkh430>
- Zeddam, J. L., Gordon, K. H. J., Lauber, C., Alves, C. A. F., Luke, B. T., Hanzlik, T. N., ... Gorbalenya, A. E. (2010). Euprosteria elaeasa virus genome sequence and evolution of the Tetraviridae family: Emergence of bipartite genomes and conservation of the VPg signal with the dsRNA Birnaviridae family. *Virology*, 397(1), 145–154. <https://doi.org/10.1016/j.virol.2009.10.042>

Anexos

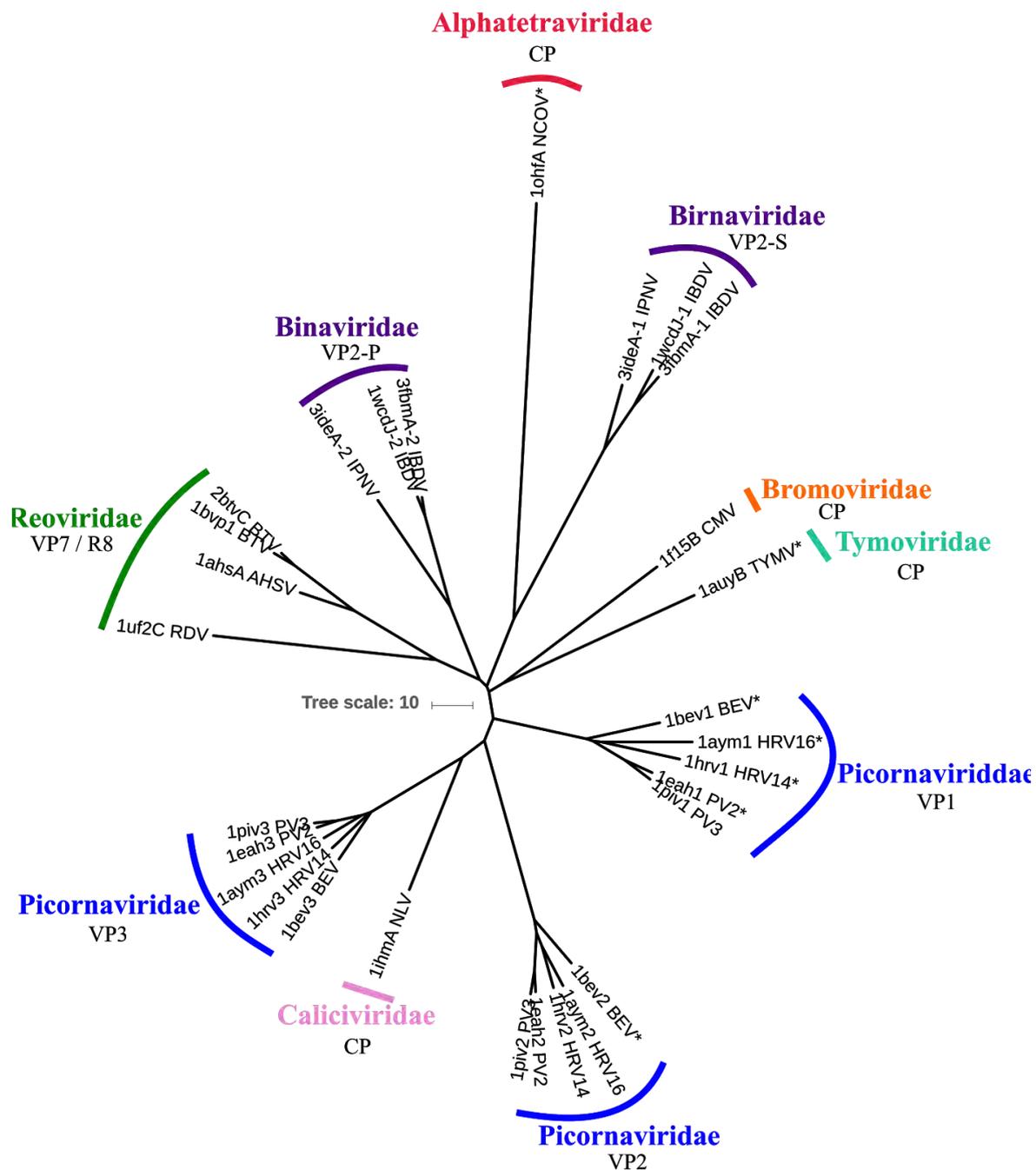


Fig. S1. Dendrograma de dominios *jelly roll* virales con (*) y sin ligandos. Se muestra el nombre de la familia viral y el nombre de la proteína a la que corresponde cada clado. Se señalan con asterisco (*) las estructuras con ligandos. En cada rama se presentan el código PDB de la estructura tridimensional y el nombre del virus. La topología es consistente con las topologías de los dendrogramas con y sin ligandos (Figuras 7 y 8), lo que indica que en este caso los ligandos no afectan la topología. Los dominios *jelly roll* se agrupan de acuerdo con el tipo de proteína y familia viral a la que pertenecen. *RDV*: virus del enanismo del arroz; *AHSV*: virus de la enfermedad equina africana; *BTV*: virus de la lengua azul; *IBDV*: virus de la enfermedad de bursitis Infecciosa; *CMV*: virus del mosaico del pepino; *HRV*: rinovirus humano; *HRV14*: rinovirus humano 14; *PV3*: poliovirus 3; *NLV*: virus de norwalk; *EntE*: enterovirus E. *RotA*: rotavirus A; *RotC*: rotavirus C; *IPNV*: virus de necrosis pancreatica infecciosa; *NCOV*: virus omega de *Nudaurelia capensis*; *TYMV*: virus del mosaico amarillo del nabo; *BEV*: enterovirus bovino; *HRV16*: rinovirus humano 16; *HRV14*: rinovirus humano 14; *PV2*: poliovirus 2.

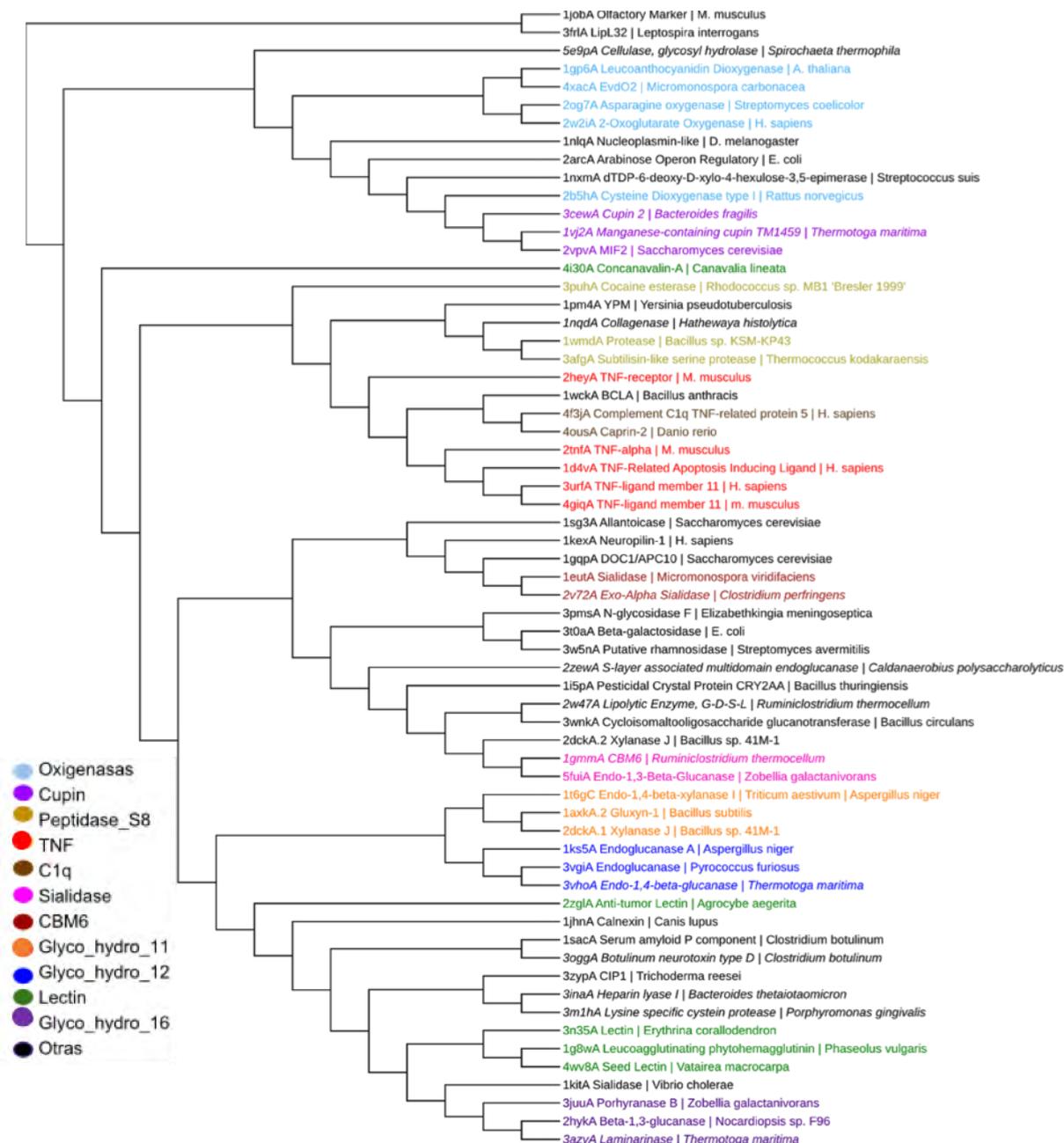


Fig. S2. Dendrograma basado en la comparación de dominios *jelly roll* celulares con y sin ligandos. Dendrograma resultante de las primeras pruebas realizadas, se compararon 63 de las 78 (debido a limitaciones del servidor en línea de *Dali*) estructuras tridimensionales celulares. La topología del dendrograma no muestra ninguna pauta, salvo la agrupación de algunas familias de proteínas como algunas TNF (rojo), algunas oxigenasas (azul claro), lectinas (verde), endoglucanasas (azul oscuro; familia *Glyco_hydro_12*) y la familia *Glyco_hydro_11* (naranja).

Tabla S1. Receptores celulares y tropismos de virus de RNA con proteínas de cápside que presentan el dominio *jelly roll*. La mayoría de los virus de la muestra utilizan glicoproteínas de epitelios celulares como receptores al iniciar la infección en sus hospederos. NCOV: virus omega de *Nudaurelia capensis* ; IPNV: virus de necrosis pancreática infecciosa; IBDV: virus de la enfermedad de bursitis infecciosa; CMV: virus del mosaico del pepino; NLV: virus de la gastroenteritis; BEV: enterovirus bovino; HRV16: rinovirus humano; PV3: poliovirus tipo 3; AHSV: virus de la enfermedad equina africana; BTV: virus de la lengua azul; RDV: virus del enanismo del arroz; RotA: rotavirus A; RotC: rotavirus C; TYMV: virus del mosaico amarillo del nabo. Se dejaron en blanco los recuadros correspondientes a los receptores/tropismos de los cuales no encontramos información. *Receptores de Enterovirus en general.

Virus	Familia	Receptor celular	Tropismo
NCOV	Alphatetraviridae		Células del intestino medio
IPNV	Birnaviridae		Precusores de las células B productoras de anticuerpos
IBDV	Birnaviridae	Proteína N-glicosilada	Precusores de las células B productoras de anticuerpos
CMV	Bromoviridae		
NLV	Caliciviridae	Antígenos del grupo histo-sanguíneo (<i>SugarBindDB</i>)	Epitelio intestinal
BEV	Picornaviridae	alpha2,3-linked sialic, CD55 & Integrin alpha2 – beta1 *	Tracto gastrointestinal
HRV16	Picornaviridae	ICAM1 (molécula de adhesión intracelular)	Tracto respiratorio superior
PV3	Picornaviridae	PVR (<i>Piconavirus receptor</i>)	Tracto gastrointestinal
AHSV	Reoviridae	Ácido siálico, JAM (<i>Junction adhesion molecule</i>)	
BTV	Reoviridae	Ácido siálico, JAM (<i>Junction adhesion molecule</i>)	
RDV	Reoviridae	Endocitosis mediada por clatrin	Tejidos neoplásicos del floema
RotA	Reoviridae	Integrinas	Tracto gastrointestinal
RotC	Reoviridae	Ácido siálico	Tracto gastrointestinal
TYMV	Tymoviridae		Floema

Tabla S2. Funciones de las proteínas celulares con el dominio *jelly roll*. De izquierda a derecha se listan las familias de *pfam*, el número de proteínas en la muestra que pertenecen a cada familia, los códigos PDB (se señalan en negritas los códigos de las proteínas sin ligandos) y las funciones correspondientes a cada familia. Cian claro: funciones relacionadas con carbohidratos; naranja: funciones de reconocimiento y unión a receptores; amarillo: funciones inmunológicas; verde: otras funciones metabólicas; azul claro: otros procesos celulares. Para la realización de esta tabla se tomaron como referencia las funciones reportadas en *Pfam* e *INTERPRO* para cada familia de proteínas y se clasificaron artificialmente para las condiciones de esta tesis.

<i>AraC_binding</i>	2	2arc;2ara	Dominio de unión a arabinosa y dimerización de la proteína reguladora del gen bacteriano AraD
<i>Bac_rhamnosid</i>	1	3w5n	Hidrolizar el enlace glucosídico entre dos o más carbohidratos, o entre un carbohidrato y un no carbohidratado
<i>BNR_2</i>	1	1eut	Sialidasa, se une al ácido siálico
<i>CBM_1</i>	1	3zyp	Unión a carbohidratos, hidrolización de compuestos O-glicosilados
<i>Cellulase</i>	4	2zew;1j83;4ba6;5e9p	Hidrolización de compuestos O-glicosilados
<i>Cleaved_Adhesin</i>	1	3m1h	Uniones celulares, lisis y unión a carbohidratos
<i>Cupin</i>	4	1vj2;3fjs;3cew;1xe7	Germinación y almacenamiento en plantas
<i>dTDP_sugar_isom</i>	2	1nxm;1wlt	Biosíntesis de dTDP-I-rhamnosa. componente esencial de la pared bacteriana

<i>Galactoside-binding lectin</i>	1	2zgl	Unión a carbohidratos: regulación extracelular e intracelular de la apoptosis
<i>Glyco_hydro_11</i>	6	1axk;1c1f;1t6g;2dck;1te1;1gmm	Hidrólisis del enlace glicosídico
<i>Glyco_hydro_12</i>	3	1ks5;3vho;3vgi	Hidrólisis del enlace glicosídico
<i>Glyco_hydro_16</i>	4	5fui;3juu;3azy;2hyk	Hidrólisis del enlace glicosídico
<i>Glyco_hydro_2_N</i>	2	3bga; 3t0A	Hidrólisis del enlace glicosídico
<i>Glyco_hydro_66</i>	1	3wnk	Hidrólisis del enlace glicosídico
<i>Lectin_legB</i>	6	3n35;4i30;4wv8;1g8w;3ax4;3u4x	Unión a glucosa/manosa o galactosa. Pueden estar involucrados en la unión de las bacterias fijadoras de nitrógeno a las leguminosas y en la protección contra los patógenos
<i>Lipase_GDSL_2</i>	1	2w47	Esterasas y lipasas
<i>Lip32</i>	1	3frl	Lipoproteína de membrana
<i>N-glycanase_N</i>	1	3pms	Unión a azúcares y cataliza la eliminación completa de las cadenas de N-oligosacáridos unidos glicoproteínas
<i>Neuropilin_Glycoprotein</i>	1	1kex	Glicoproteínas transmembranales, señalización
<i>PhyH</i>	1	4xac	Catálisis de la alfa oxidación del ácido fitánico

<i>PKD</i>	1	1nqd	Glicoproteína de membrana involucrada en las interacciones proteína-proteína y proteína-carbohidrato
<i>Polysacc_lyase</i>	1	3ina	Despolimeriza la heparina mediante la escisión del enlace glucosídico
<i>Sial-lect-inser</i>	1	1kit	Se unen a las lectinas con alta afinidad ayudando a dirigir la proteína a los ambientes ricos en ácido siálico
<i>Sialidase</i>	1	2v72	Hidrólisis del enlace glicosídico
<i>Endotoxin_C</i>	1	1i5p	Unión a receptor celular durante la infección
<i>Olfactory_mark</i>	1	1job	Transducción de señales olfativas
<i>Toxin_R_bind_C</i>	1	3ogg	Proceso de infección mediante la unión al nervio presináptico
<i>YPM</i>	1	1pm4	Activación de células T mediante la unión al receptor
<i>C1q</i>	2	4ous, 4f3j	Activa el sistema complementario de suero
<i>Pentraxin</i>	1	1sac	Respuesta inmunológica aguda
<i>TNF</i>	9	4giq;3v56;2tnf;1wck; 1d4v;1jh5;1rj8;3urf;2hey	Procesos inmunológicos
<i>Allantoicase</i>	1	1sg3	Metabolismo de purinas
<i>CDO_I</i>	1	2b5h	Convertir cisteína en ácido cisteinesulfínico
<i>Peptidase_S8</i>	3	3afg;3puh;1wmd	Actividad de endopeptidasa tipo serina

<i>2OG-Fell_Oxy</i>	1	1gp6	Actividad oxidoreductasa
<i>Calreticulin</i>	1	1jhn	Unión a calcio, almacenamiento de calcio en retículo endoplasmático
<i>TauD</i>	1	2og7	Procesos oxido-reducción,
<i>ANAPC10</i>	1	1gqp	Regulación de la anafase de la mitosis
<i>CENP-C_C</i>	1	2vpv	Formación de centrómero
<i>JmjC</i>	1	2w2i	Desmetilación de histonas
<i>Nucleoplasmin</i>	3	1nlq; 1k5j	Unión a histonas y transferencia de DNA