

UACM

Universidad Autónoma
de la Ciudad de México

Nada humano me es ajeno

COLEGIO DE CIENCIA Y TECNOLOGÍA

MAESTRÍA EN CIENCIAS GENÓMICAS

*Estudio comparativo de los procesos evolutivos en la proteína
de envoltura de los miembros de la familia Flaviviridae*

TESIS PARA OBTENER EL GRADO DE MAESTRO EN CIENCIAS GENÓMICAS

PRESENTA

HUGO GILDARDO CASTELÁN SÁNCHEZ

Directora de tesis

Dra. Claudia Selene Zárate Guerra

México, D.F. diciembre 2014.

SISTEMA BIBLIOTECARIO DE INFORMACIÓN Y DOCUMENTACIÓN



UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO COORDINACIÓN ACADÉMICA

RESTRICCIONES DE USO PARA LAS TESIS DIGITALES

DERECHOS RESERVADOS[©]

La presente obra y cada uno de sus elementos está protegido por la Ley Federal del Derecho de Autor; por la Ley de la Universidad Autónoma de la Ciudad de México, así como lo dispuesto por el Estatuto General Orgánico de la Universidad Autónoma de la Ciudad de México; del mismo modo por lo establecido en el Acuerdo por el cual se aprueba la Norma mediante la que se Modifican, Adicionan y Derogan Diversas Disposiciones del Estatuto Orgánico de la Universidad de la Ciudad de México, aprobado por el Consejo de Gobierno el 29 de enero de 2002, con el objeto de definir las atribuciones de las diferentes unidades que forman la estructura de la Universidad Autónoma de la Ciudad de México como organismo público autónomo y lo establecido en el Reglamento de Titulación de la Universidad Autónoma de la Ciudad de México.

Por lo que el uso de su contenido, así como cada una de las partes que lo integran y que están bajo la tutela de la Ley Federal de Derecho de Autor, obliga a quien haga uso de la presente obra a considerar que solo lo realizará si es para fines educativos, académicos, de investigación o informativos y se compromete a citar esta fuente, así como a su autor ó autores. Por lo tanto, queda prohibida su reproducción total o parcial y cualquier uso diferente a los ya mencionados, los cuales serán reclamados por el titular de los derechos y sancionados conforme a la legislación aplicable.



Este trabajo se llevó a cabo en el laboratorio de Bioinformática del Posgrado en Ciencias Genómicas de la Universidad Autónoma de la Ciudad de México bajo la tutoría de la Dra. Claudia Selene Zárate Guerra.

El alumno recibió una beca CONACYT durante la realización durante su proyecto de maestría (No. De apoyo 341797)

También se contó con el apoyo económico del UACM para asistir al 8 Congreso Nacional de Virología, por medio del apoyo para la participación de estudiantes en foros académicos o culturales nacionales.

El comité que asesoró este proyecto estuvo integrado por:

Dra. Martha Yocupicio Monroy UACM-PCG

Dra. Mariana Peimbert Torres UAM-Cuajimalpa.



Los lectores de tesis fueron:

Dra. Minerva Camacho Nuez
Posgrado en Ciencias Genómicas, UACM

Dra. Rosa Martha Yocupicio Monroy
Posgrado en Ciencias Genómicas, UACM

Dra. Mariana Peimbert Torres
Unidad Cuajimapa, UAM

AGRADECIMIENTOS

Mi sincero agradecimiento a la Dra. Selene Zárate, por permitirme desarrollar el proyecto de maestría bajo su tutela, por su apoyo, sus valiosas enseñanzas, por compartirme su conocimiento, por sus consejos y su interés en mi formación académica.

A la Dra. Mariana Peimbert, por su interés en el proyecto y en mi formación, además de sus comentarios tan acertados que enriquecieron el proyecto, y su dedicación en la revisión de este trabajo

A la Dra. Martha Yocupicio por sus valiosos comentarios y sugerencias en la realización del proyecto, y su dedicación en la en la revisión de este trabajo.

A la Dra. Minerva Camacho por su tiempo y se dedicación en la en la revisión de este trabajo.

A los Profesores-Investigares que integran la academia del Posgrado en Ciencias Genómicas, por la formación académica recibida.

A mi amigos y compañeros de la generación 2012 de la Maestría en Ciencias Genómicas, por todo lo compartido a lo largo de la maestría, Nayeli, Cecy, Mariana, Karla, Monse, Kenji, Edgar, David, Francisco y Ricardo.

A los alumnos que forman parte del posgrado en ciencias genómicas.

Al personal administrativo del posgrado en ciencias genómicas, por el apoyo en todos los procesos administravos.

A mis papás Guillermina y Juan, por su apoyo incondicional y que gracias a ellos logré este sueño.

Contenido

	Páginas
Abreviaturas	3
Resumen	4
Capítulo 1. Marco Teórico	5
1.1 Generalidades de los <i>Flavivirus</i>	5
1.1.1 Estructura viral	6
1.1.2 Organización del genoma	6
1.2 Género <i>Flavivirus</i>	9
1.3 Género <i>Pestivirus</i>	10
1.4 Género <i>Hepacivirus</i>	11
1.5 Género <i>Pegivirus</i>	11
1.6 Proteínas estructurales	12
1.7 Proteínas no estructurales	14
1.8 Proteína de envoltura de la familia <i>Flaviviridae</i>	16
Capítulo 2. Antecedentes Particulares	19
2. Procesos evolutivos en la proteína de envoltura de la familia <i>Flaviviridae</i>	19
2.1 Tasa de mutación y tasa de sustitución	19
2.2 Selección natural	21
2.3 Recombinación	23
2.4 Deriva Génica	25
Capítulo 3. Planteamiento del problema	27
3. Justificación	27
3.1 Hipótesis	27
3.2 Objetivo General	28
3.3 Objetivos Particulares	28
Capítulo 4. Materiales y métodos	29
4.1 Obtención de secuencias y alineamiento	29
4.2 Análisis de la variabilidad	30
4.3 Uso de codón	30
4.4 Análisis de la presencia de recombinación	33
4.5 Reconstrucción de filogenias	33
4.6 Métodos de Selección	34
4.6.1 <i>Evolutionary fingerprint</i> del gen de la proteína de envoltura	35
4.6.2 Determinación de selección episódica	35

4.7 Determinación de epítopes	36
4.8 Reconstrucción de caracteres ancestrales	36
4.8.1 Parsimonia	37
4.8.2 Máxima Verosimilitud	37
4.9 Análisis de cambios en las propiedades bioquímicas	37
4.10 Mapeo de las propiedades bioquímicas y sitios bajo selección en la estructura terciaria de la proteína	38
4.11 Modelado y análisis de las proteínas mutantes	39
Capítulo 5. Resultados	40
5.1 Análisis de la variabilidad	40
5.2 Uso de codón	41
5.3 Detección de la Selección	45
5.3.1 <i>Evolutionary Fingerprint</i>	48
5.3.2 Reconstrucción de caracteres ancestrales	49
5.3.3 Selección episódica	57
5.4 Análisis de las propiedades bioquímicas	64
Capítulo 6. Discusión	82
Capítulo 7. Conclusiones	91
Capítulo 8. Perspectivas	92
Complemento resultados	93
Capítulo 9. Referencias	110

Abreviaturas

DENV	Virus dengue
HCV	Virus de hepatitis C
BVDV	Virus de la diarrea viral bovina
CSFV	Virus de la fiebre porcina clásica
TBEV	Virus de la encefalitis transmitida por garrapatas
FEL	Probabilidad de efectos fijos
IFEL	Probabilidad de efectos fijos internos
MEME	Mixed Effects Model of Evolution
PRIME	PRoperty Informed Model of Evolution
dN/dS o ω	Tasa de sustituciones no sinónimas sobre sitio no sinónimo entre las sustituciones sinónimas sobre sitio sinónimo.
α	Sustituciones sinónimas
β	Sustituciones no sinónimas
pI	Punto Isoeléctrico
RMSD	Root Mean Square Deviation

Resumen

Las proteínas de envoltura de los virus de la familia *Flaviviridae*, se caracterizan por estar formada por tres dominios, los cuales son clave para la interacción con el receptor celular, definir el tropismo viral, permitir la entrada a la célula hospedera, participar en el ensamblaje del virión, además de ser el principal antígeno de superficie de los viriones. Estas funciones se encuentran conservadas en los géneros de la familia (*Flavivirus*, *Pestivirus* y *Hepacivirus*), pero las proteínas difieren en secuencia y conformación tridimensional. En este trabajo, investigamos los procesos evolutivos de las proteínas de envoltura de los virus de esta familia.

Dado que la función de las proteínas se conserva en los diferentes géneros, a pesar de las posibles diferencias en presiones selectivas específicas de sus respectivos hospederos, hipotetizamos que los procesos evolutivos que ocurren en éstas proteínas son similares, además, dado que es el principal antígeno de superficie es posible que las proteínas se encuentren evolucionando bajo fuerte selección positiva.

Este trabajo se analizaron los procesos evolutivos que han actuado sobre las proteínas de envoltura de los virus DENV, BVDV y HCV, como representantes de cada género, y encontramos que aunque estas proteínas presentan una alta variabilidad en la secuencia de nucleótidos y aminoácidos, la mayoría de los codones evolucionaron bajo selección negativa; esto resultaba incongruente, porque éste tipo de selección elimina la variabilidad. Por lo tanto analizamos si la variabilidad observada, era debida a episodios de selección actuando sobre las proteínas, y encontramos que la selección episódica, es responsable en parte de la variabilidad observada. Los pocos codones que evolucionaron bajo selección positiva parecen estar relacionados con virulencia y escape del reconocimiento de anticuerpos.

Por último, analizamos como estos procesos evolutivos actúan a nivel de proteína modificando las propiedades bioquímicas de éstas, encontramos que la propiedad más conservada fue el volumen, mientras que la polaridad fue la menos conservada. Así también se observó que proteína de envoltura de DENV resulto ser la más conservada en las propiedades bioquímicas, lo cual podría deberse a restricciones evolutivas derivadas de su ciclo de replicación, mientras la proteínas de HCV presentó un mayor grado de conservación, lo cual podría ser producto de la constante presión selectiva del hospedo. Con esto determinamos que los procesos evolutivos en las proteínas de envoltura de la familia *Flaviviridae* son similares en cuanto al grado de variabilidad observada, las tasas de mutación en codones sinónimos y no sinónimos, y el hecho de que la selección positiva tienda a presentarse de manera episódica. Sin embargo, parece haber diferencias en el grado de conservación de las propiedades bioquímicas de estas proteínas.

Capítulo 1

Marco Teórico

1.1 GENERALIDADES DE LOS *FLAVIVIRUS*

La familia *Flaviviridae* comprende a virus de RNA de cadena sencilla, de polaridad positiva, y consta de cuatro géneros: los *Flavivirus*, *Pestivirus*, *Hepacivirus* y *Pegivirus* (Twiddy *et al.*, 2003). Los virus pertenecientes a diferentes géneros tienen similitud en cuanto a la morfología del virión, la organización del genoma, y la estrategia de replicación (Leyssen *et al.*, 2000). Estos virus son responsables de una alta mortalidad y morbilidad, y pueden causar enfermedades como: encefalitis severa, dengue, fiebre hemorrágica, formas con choque, hepatitis, peste porcina clásica, enfermedades febriles (Figura 1) (Zanotto *et al.*, 1996; Acosta y Gómez, 2005).

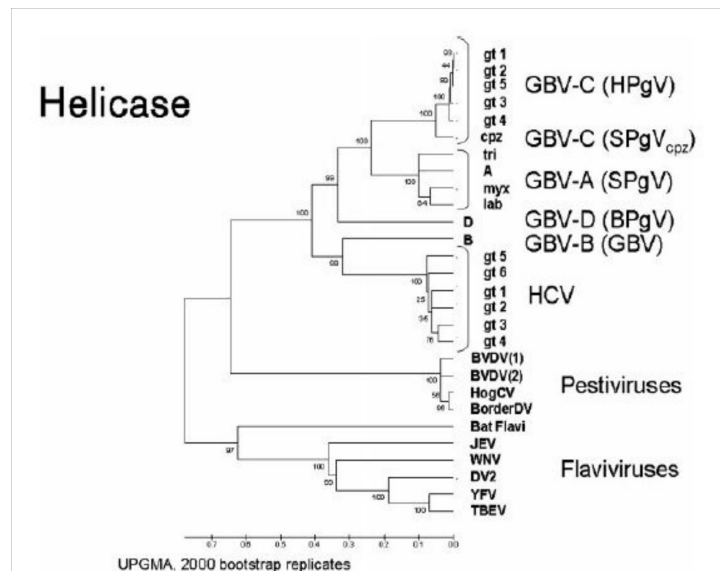


Figura 1. Árbol filogenético de la familia *Flaviviridae* construido con la proteína NS3 o Helicasa. El árbol filogenético muestra la relación entre los diferentes géneros de la familia, se puede observar que los géneros *Pegivirus*, *Hepacivirus* y *Pestivirus* se encuentran más cercanamente relacionados entre sí, mientras que el género *Flavivirus* es el más distante (Tomada de Stapleton *et al.*, 2011).

1.1.1 ESTRUCTURA VIRAL

Los virus de la familia *Flaviviridae*, son virus pequeños con un diámetro de aproximadamente 50 a 60 nm. La partícula viral es icosaédrica y está envuelta por una membrana lipídica proveniente del hospedero, sobre la cual se insertan dos proteínas de envoltura que forman homodímeros o heterodímeros. En el interior de la partícula viral está la proteína “core” que forma la nucleocápside o cápside, la cual se asocia con el genoma viral (Figura 2).

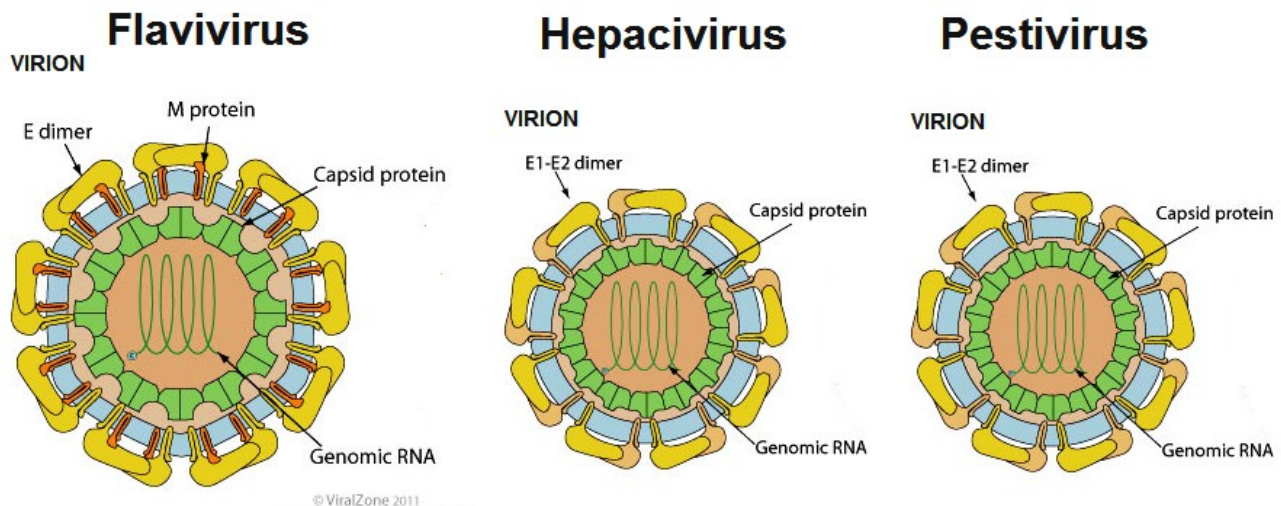


Figura 2. Estructura de los viriones de la familia *Flaviviridae*. A la izquierda el virión del género *Flavivirus*, donde la proteína E está formando homodímeros y se encuentra asociado a la proteína M. En el centro el virión de *Hepacivirus*, donde las proteínas E1-E2 forman heterodímeros. A la derecha el virión de *Pestivirus*, donde las proteínas E1-E2 forman heterodímeros. (Tomadas de <http://viralzone.expasy.org/2014>)

1.1.2 ORGANIZACIÓN DEL GENOMA

Los virus de la familia *Flaviviridae* poseen un genoma de RNA de cadena sencilla y polaridad positiva, con una longitud aproximada de 11 a 12.3 kb, dependiendo el género; En el género *Flavivirus*, el genoma tiene una longitud de 11 kb, con un solo marco de lectura (*Open Reading Frame*, ORF) que codifica para una poliproteína de 3386-3433 aminoácidos (aa), este único ORF está flanqueado por dos regiones no traducidas (*Untranslated Region*,

UTR), la primera en el extremo 5' con una longitud de 95-132 nucleótidos (nt), que posee varios dinucleótidos AG involucrado en la formación de estructuras secundarias, también aquí se localiza una estructura tipo CAP, 7-metil-guanosina. La segunda región es el extremo 3' que carece de un tracto de poli (A) tiene una longitud de 114 a 585 bases y posee secuencias para la circularización del genoma a través de interacciones con la región 5'UTR que facilitan su replicación y traducción (Leyssen *et al.*, 2000; Lindenbach *et al.*, 2007).

El genoma de los virus del género *Pestivirus*, tiene un tamaño de aproximadamente de 12.3 kb, de igual forma que los *Flavivirus*, su genoma posee un único ORF que codifica para una poliproteína de 3838-4000 aa. El extremo 5' UTR tiene una longitud aproximada de 385 bases y carece de una estructura tipo CAP, sin embargo poseen un Sitio Interno de Entrada al Ribosoma (*Internal Ribosome Entry Site*, IRES). En el extremo 3'UTR tiene una longitud de 225 bases y también carece de un tracto de poli (A) (Figura 3) (Rissatti *et al.*, 2007; Becher y Tautz, 2011).

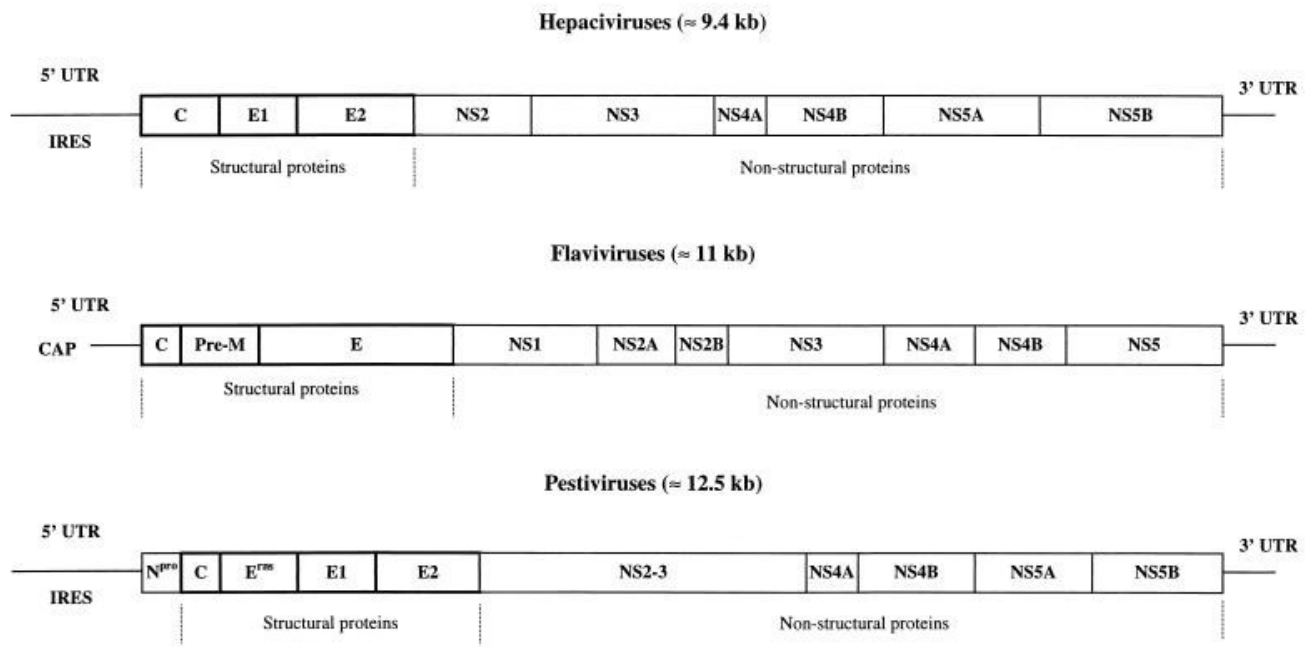
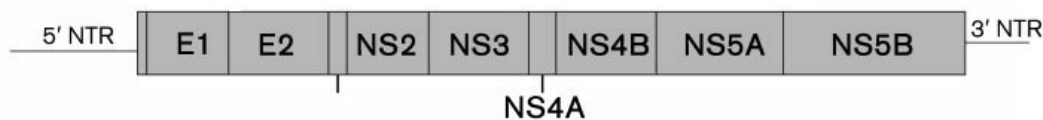


Figura 3. Organización del genoma de familia *Flaviviridae*. El genoma de los virus de esta familia es RNA de cadena sencilla y polaridad positiva, en el caso de los *Flavivirus* presentan CAP en el extremo 5' UTR, mientras que los *Pestivirus* y *Hepacivirus* tienen IRES. En el extremo 3' UTR carecen de poli (A). El genoma se codifica en forma de poliproteína, el cual es procesado postraduccionalmente (Tomada de Leyssen *et al.*, 2000).

El HVC, perteneciente al género *Hepacivirus*, tiene un genoma con una longitud de 9.4 kb que codifica para una poliproteína de 3000 aa, al igual que los *Pestivirus* su genoma carece de la estructura tipo CAP en el extremo 5' UTR, y posee un IRES, la longitud de este extremo es de 342 bases, la función de éste es permitir la unión del ribosoma de la célula hospedera. El extremo 3' UTR, carece de poli (A) y varía en longitud dependiendo del genotipo, sin embargo se ha descrito que ésta contiene regiones relacionadas que facilitan la traducción e iniciación de la síntesis del RNA viral (Li *et al.*, 2013).

Por último para el género *Pegivirus*, al igual que los otros géneros de la familia su genoma es de RNA de polaridad positiva que codifica para única poliproteína de 3484 aminoácidos que es procesada postraduccionalmente (Figura 4) (Ghai *et al.*, 2013).

GBV-A (SPgV), GBV-C (HPgV)



GBV-D (BPgV)

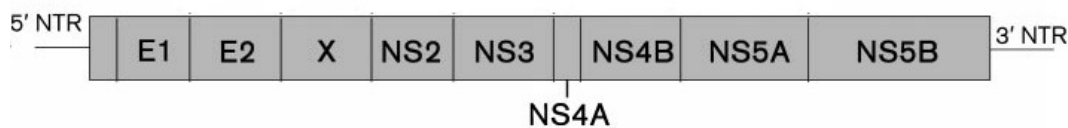


Figura 4: Organización del genoma de los virus del género *Pegivirus*. El genoma es de RNA cadena sencilla codificada para una poliproteína que es procesada co- y postraduccionalmente en proteínas virales individuales, la proteína del core, no ha sido identificada en estos virus como en los otros géneros de la familia. En la parte superior se muestra el genoma los *Pegivirus* que infectan a humanos y simios, en la parte inferior el de los que infecta a murciélagos (tomada y modificada de Ghail *et al.*, 2013).

1.2 Género *Flavivirus*

El género *Flavivirus* comprende más de 70 virus y se divide en 14 clados, que a su vez se agrupan en 3 grandes grupos, estos son: el grupo de virus transmitidos por mosquitos, el grupo de virus transmitidos por garrapatas, y el grupo de virus no transmitidos por vector. Los *Flavivirus* que causan enfermedades importantes en humanos pertenecen a los dos primeros grupos (Kuno *et al.*, 1998). Las enfermedades que provocan en sus hospederos mamíferos van desde una simple fiebre, hasta complicaciones graves como fiebres hemorrágicas, encefalitis, daño neurológico, y daño hepático. Los virus de este género tienen una amplia gama de hospederos tales como las aves, mosquitos, garrapatas, gatos, perros, vacas, caballos, monos, y los humanos (Lindenbach *et al.*, 2007).

El virus dengue (DENV), perteneciente a este género, es uno de los más relevantes, pues representa un problema de salud pública en regiones tropicales y subtropicales del mundo. Anualmente se registran entre 50 y 100 millones de casos de fiebre por dengue a nivel mundial (Zanotto *et al.*, 1996). En México, durante la última década se ha reportado un incremento en la incidencia de casos de dengue y fiebre hemorrágica por dengue en la población juvenil, por lo tanto es un problema de salud pública en nuestro país. (Torres *et al.*, 2014). El DENV se clasifica en cuatro serotipos, dengue 1, 2, 3, y 4 (DENV-1, DENV-2, DENV-3 y DENV-4) los cuales presentan una divergencia del 30 % a lo largo de la poliproteína. El virus se transmite a los humanos por la picadura de mosquitos del género *Aedes*, particularmente de las especies *Aedes aegypti* y *Aedes albopictus*.

La diversidad genética de estos cuatro serotipos ha sido estudiada en las proteínas estructurales del virión, lo que ha permitido definir diferentes subtipos o subgenotipos para cada serotipo, por ejemplo, el DENV 1 presenta cinco genotipos (Genotipo I, II, III, IV, y V), DENV 2, presenta cinco genotipos (Asiático I, Asia II, América-Asia, Cosmopolita, Americano), DENV 3 presenta cuatro genotipos (Genotipo I, II, III y IV) y finalmente DENV 4 presente cuatro genotipos (Genotipo I, II, III y IV). La divergencia entre genotipos es de aproximadamente de un 6-10 % a nivel de nucleótidos (Goncalvez *et al.*, 2002; Holmes, 2007; Weaver y Vasilakis, 2009).

1.3 Género *Pestivirus*

Los *Pestivirus* infectan a animales del orden *Artiodactyla*, es decir, animales rumiantes entre los que se encuentran animales de importancia económica, como el ganado bovino, porcino y ovino.

El género *Pestivirus* comprende a virus como: el virus de la diarrea viral bovina (*Bovine viral diarrhea virus*, BVDV), el virus de la fiebre porcina clásica (*Classical swine fever virus*, CSFV) y el virus de la enfermedad de frontera (*Border disease virus*, BDV). Estos *Pestivirus* se transmiten por tres vías que son la oral-fecal, la sexual y transplacentaria, ésta última es responsable de su persistencia en los hospederos, mientras que los virus BVDV y BDV, pueden cruzar barreras de especie entre hospederos del orden *Artiodactyla* (Lindenbach *et al.*, 2007; Leyssen *et al.*, 2000; Tang y Zhang 2007).

Los *Pestivirus* causan diarreas aguadas, síndromes hemorrágicos agudos y enfermedad de *wasting*, las cuales pueden convertirse en enfermedades persistentes. Los virus que causan persistencia en sus hospederos, son de biotipo no citopático, es decir, no producen efecto citopático en el cultivo celular. Esta última es una característica particular del BVDV, lo que ha permitido su clasificación de este virus en dos biotipos, el citopatogénico (cp) que producen efecto citopático en cultivo celular produciendo vacuolización citoplasmática, y apoptosis celular y el biotipo no citopatogénicos (nocp) que no inducen cambios morfológicos en las células infectadas (Becher y Tautz, 2011).

Por otro lado, los principales órganos blanco para la replicación de estos virus, son el tejido linfóide y epitelial, principalmente células linfocíticas, aunque estos virus también se pueden replicar en las células gastrointestinales, las glándulas y las neuronas (Leyssen *et al.*, 2000).

El BVDV, es uno de los agentes infecciosos más importantes de este género, ya que presenta una distribución mundial, y es endémico en la mayoría de las poblaciones bovinas, ocasionado pérdidas económicas de origen reproductivo. Los avances en la secuenciación han permitido clasificar a éste virus en 2 genotipos, (Genotipo I y II) adicionalmente cada genotipo presenta subgenotipos, que tienen una identidad entre sí del 80 al 85 %. Para el

genotipo I existen 11 subgenotipos de BVDV, y para el genotipo II existen 2 subgenotipos (Vargas *et al.*, 2009).

1.4 Género *Hepacivirus*

Dentro del género *Hepacivirus* se incluyen los virus de Hepatitis C, HCV (*Hepatitis C virus*, HCV) y el virus de hepatitis GB (GBV). Los *Hepacivirus*, comparten con los *Pestivirus* la organización del genoma y con ello un tamaño similar en la longitud de los genes (Lindenbach *et al.*, 2007).

De acuerdo con los análisis filogenéticos, el HCV se ha clasificado en seis genotipos, que presentan una divergencia del 30-35 % a nivel de nucleótidos a lo largo del genoma completo. Todos los genotipos de HCV se encuentran distribuidos por todo el mundo, siendo el genotipo 1 el más común y el que presenta mayor resistencia al tratamiento. Además dentro de cada genotipo existen subgenotipos o subtipos, por ejemplo, el genotipo 1 tiene 13 subtipos, estos se diferencian en un 20-25 % a nivel de nucleótidos (Simmonds *et al.*, 2004).

El HCV es el virus más importante del género. La Organización Mundial de la Salud (OMS), estima que el 3 % de población a nivel mundial está infectada con el HCV, lo que significa que más de 170 millones son portadores y están en riesgo de desarrollar cirrosis y/o cáncer de hígado (Leyssen *et al.*, 2000). En México se estima una prevalencia del 0.47-1.2 %, donde el genotipo predominante es el 1, y la principal vía de transmisión reportada es la transfusión de sangre (Vera *et al.*, 2005).

Las principales formas de transmisión del HCV, son la vía sexual, transplacentaria y por medio de agujas o sangre contaminada, provocando en su hospedero humano una hepatitis crónica, cirrosis hepática y hepatocarcinoma (Lindenbach *et al.*, 2007).

1.5 Género *Pegivirus*

Los miembros de este género se agruparon por mucho tiempo dentro del género de *Hepacivirus*, sin embargo el Comité Internacional de Taxonomía de Virus (*International Committee on Taxonomy of Viruses*, ICTV) en el año 2012 fue considerado como un género independiente. Los *Pegivirus* se descubrieron en el año 1995 en monos tamarinos, y actualmente los virus que forman parte de este género son el GBV-A o SPgV aislado de

simios, el GBV-C o HPgV aislado de humanos, y el GBV-D o BPgV aislado de murciélagos. El virus HPgV, que infecta a humanos, es un virus linfotrópico que establece infecciones subclínicas y no se asocia con hepatitis C. Este virus se encuentra ampliamente distribuido en países desarrollados, su prevalencia es alta y se transmite por vía sexual, vertical y a través del contacto con sangre contaminada. Actualmente este virus ha tomado importancia porque se ha asociado con el virus de inmunodeficiencia adquirida (VIH) provocando un aumento en la mortalidad en los pacientes (Kapoor *et al.*, 2013; Ghai *et al.*, 2013).

1.6 PROTEÍNAS ESTRUCTURALES

Las proteínas estructurales forman parte del virión, dependiendo del género el número de proteínas que componen a éste varían. En la tabla 1 se muestran las características y funciones principales de las proteínas estructurales

TABLA 1. PROTEÍNAS ESTRUCTURALES DE LA FAMILIA FLAVIVIRIDAE

Virus	Proteína	Peso molecular	Funciones y propiedades
VIRUS DENGUE	Core	12 kDa	Proteína altamente conservada y compuesta principalmente por aminoácidos básicos, lo que favorece la interacción con el genoma viral.
	E	53 kDa	Proteína glicosilada que se caracteriza por tener tres dominios funcionales, responsables de la entrada a la célula. Los receptores descritos son: DC-SIGN, (<i>Dentric Cell-Specific ICAM-3 Grabbing Nonintegrin</i>), heparán sulfato, CD14, las proteínas de choque térmico (HSP) de 70 y 90 kDa, el receptor de alta afinidad de laminina de 37/67 kDa y la proteína

			GRP78/BiP
	prM	18-19 kDa	Proteína presente en el virión cuando es inmaduro, evita que la fusión de la membrana viral y celular suceda antes de tiempo.
VIRUS DE LA DIARREA VIRAL BOVINA	Core	14 kDa	Proteína altamente conservada y rica en aminoácidos básicos, es la responsable de la formación de la cápside viral y unión del RNA viral.
	E ^{ms}	44-48 kDa	Proteína con actividad ribonucleasa, esencial para la producción de partículas virales infecciosas, también es indispensable para la unión y entrada a la célula hospedera.
	E1	33 kDa	Glicoproteína de envoltura que se asocia con E2 importante para el reconocimiento del receptor.
	E2	55 kDa	Glicoproteína integral de membrana que se asocia en homodímeros (E2-E2) o heterodímeros (E1-E2) permite el reconocimiento del receptor celular (CD45) y ensamblaje del virión.
VIRUS DE HEPATITIS C	Core	19-21 kDa	Proteína básica altamente conservada, además se propone que ésta posee la función de modulación de la transcripción de genes, es la responsable de la formación de la cápside viral y unión del RNA viral.
	E1	35 kDa	Proteína glicosilada importante para la fusión de membranas.
	E2	77 kDa	Glicoproteína responsable de la entrada del virión a la célula

			blanco de la neutralización de anticuerpos y unión al receptor. Los receptores reportados son: el CD81, SRBI (Scavenger receptor class B type I), heparán sulfato, y las lectinas de unión a manosa DC-SIGN (CD209) y L-SIGN (CD209-L).
--	--	--	---

1.7 PROTEÍNAS NO ESTRUCTURALES

Las proteínas no estructurales no forman parte del virión, pero son indispensables en la replicación y traducción del genoma. Los virus de esta familia tienen de 5 a 7 proteínas no estructurales, la función de algunas de ellas ha sido ampliamente descrita y de otras se desconoce (Tabla 2).

TABLA 2. PROTEÍNAS NO ESTRUCTURALES DE LA FAMILIA *FLAVIVIRIDAE*

Virus	Proteína	Peso molecular	Funciones y propiedades
VIRUS DENGUE	NS 1	42-50 kDa	Proteína que se encuentra en 3 formas distintas: asociada al retículo endoplásmico, como dímero en la membrana de la célula hospedera y en su forma secretada, induce inmunidad humoral, importante para el diagnóstico de la enfermedad.
	NS 2A	22 kDa	Proteína que forma parte del complejo de replicación y bloqueo de la síntesis de IFN.
	NS 2B	12 kDa	Proteína que es cofactor de NS3.
	NS 3	70 kDa	Proteína que tiene funciones

			de Proteasa/Helicasa.
	NS4A	16 kDa	Proteína que es Cofactor de NS3.
	NS4B	27 kDa	Forma parte del complejo de replicación.
	NS5	103 kDa	Proteína que es la RNA polimerasa viral y metiltransferasa.
VIRUS DE LA DIARREA VIRAL BOVINA	N pro		Proteína con actividad utoproteasa.
	NS 2	52 kDa	Serinproteasa capaz de formar rearreglos e inserciones en el genoma viral de fragmentos del genoma de la célula hospedera
	NS3		Proteína que tiene funciones de Proteasa/Helicasa.
	NS4A		Proteína con función desconocida.
	NS4B		Proteína de función desconocida
	NS5A	58 kDa	Proteína tiene actividad serin/treonina cinasas
	NS5B	75 kDa	Proteína que funciona como RNA polimerasa viral.
VIRUS DE HEPATITIS C	p7	7 kDa	Proteína con función desconocida, sin embargo en sistemas <i>invitro</i> funciona como canal iónico.
	NS2	23 kDa	Glicoproteína de membrana, que se desconoce su función madura, sin embargo antes de ser cortada de la poliproteína, ésta funciona como proteasa.

	NS3	70 kDa	Proteína con funciones de Proteasa/Helicasa.
	NS4A	8 kDa	Proteína que actúa como cofactor de NS3.
	NS4B	27 kDa	Proteína que forma parte del complejo de replicación.
	NS5A	58 kDa	Proteína con actividad serin/treonina cinasas y forma parte del complejo del replicación.
	NS5B	68 kDa	Proteína que funciona como RNA polimerasa viral.

1.8 PROTEÍNA DE ENVOLTURA DE LA FAMILIA *FLAVIVIRIDAE*

La familia *Flaviviridae* tiene proteínas de envoltura (E) que se han agrupado con las glicoproteínas de fusión clase II. Estas proteínas se sintetizan en forma de un complejo junto a otra proteína estructural (prM en los *Flavivirus*, proteína E1 en los *Pestivirus* y *Hepacivirus*) que luego se escinde para producir la forma activa. Generalmente, estas proteínas presentan secuencias internas de fusión o péptido de fusión. (Delos *et al.*, 2000; Li y Modis *et al.*, 2014).

La proteína E de DENV (*Flavivirus*), el virus de la diarrea viral bovina (BVDV) (*Pestivirus*) y el virus de hepatitis C (HCV) (*Hepacivirus*) se han cristalizado, por lo tanto en este proyecto utilizamos a estos virus como modelos de estudio.

En la proteína E (*Flavivirus*) y E2 (*Pesti-Hepacivirus*) se han descrito tres dominios: I, II y III. Estos dominios en todos los virus, presentan una conformación de hojas β y una o dos hélices α . Las características de estos dominios se describen a continuación:

El dominio I varía en longitud en los tres géneros, y aunque a nivel de secuencia no tiene similitud significativa, presentan una estructura de barril β conservada. En los *Flavivirus* y *Hepacivirus* se le domina dominio central, y la principal diferencia se debe a que la proteína E2 (*Hepacivirus*) tiene próximo al extremo carboxilo terminal una región llamada

hipervariable 1 (HVR *hypervariable región*). En el caso de los *Pestivirus* el dominio I se encuentra hacia el extremo N-terminal de la proteína, pero conserva la estructura de barril β . Dentro de este dominio se encuentran sitios de N-glicosilación, así por ejemplo para los *Flavivirus* hay un sitio de glicosilación, mientras que para los *Hepacivirus* hay 5 sitios de glicosilación, en este último caso son importantes para la unión a su receptor celular. Recientemente se demostró mediante estudios de mutagénesis que en el virus de hepatitis C el sitio de unión al receptor está dentro del dominio I (Nybekken *et al.*, 2006; Kong *et al.*, 2013).

El dominio II, en el caso de los *Flavivirus*, es el dominio de dimerización y además posee una secuencia interna de fusión o péptido de fusión, altamente hidrofóbica y rica en glicinas que es importante para la internalización del virus, también dentro del dominio II existe un sitio de N-glicosilación es (Asn67) que se propone que participa en la interacción del virión con el receptor. En el caso de los *Pestivirus* también se propone la presencia de un péptido de fusión en este dominio y también se ha descrito un sitio de glicosilación. Para el HCV (*Hepacivirus*) se propone la existencia de un posible péptido de fusión, en una región rica en aminoácidos no polares y glicina, además es una región conservada en todos los genotipos del HCV (McCaffrey *et al.*, 2012; Yagnik *et al.*, 2000).

El dominio III para los *Flavivirus* tiene un plegamiento tipo inmunoglobulina, y en el caso del DENV es responsable del contacto con el receptor celular, pues presenta aminoácidos cargados positivamente que se unen al heparán sulfato (receptor con carga negativa). Para los *Pestivirus* este dominio se compone por tres módulos de pequeñas láminas (IIIa, IIIb y IIIc) que poseen aminoácidos hidrofóbicos, que pudieran ser de anclaje a la membrana celular o membrana viral, además posee cuatro sitios de glicosilación importantes de igual forma para la posible unión al receptor. Para los *Hepacivirus* en el dominio III existen regiones de interacción con CD81 sugiriendo que es el posible dominio de unión al receptor (McCaffrey K., *et al.*, 2012). El dominio III entre CSFV (*Pestivirus*), Hepatitis C (*Hepacivirus*) y TBEV (*Flavivirus*) presentan una similitud importante a nivel de secuencia, dado que se obtienen alineamientos significativos con una $P < 0.005$ (Garry y Dash, 2002).

Debido a que estas proteínas comparten ciertas características, se esperaba que la estructura terciaria de las proteínas de envoltura en los géneros *Pestivirus* y *Hepacivirus* tuviera un plegamiento y características similares de una proteína fusión clase II como se había descrito en los *Flavivirus*. En el 2013 se lograron cristalizar las proteínas E2 del BVDV y del HCV, donde se descubrió que estas dos tienen un nuevo plegamiento, además que carecen de *hallmarks* característicos de una proteína de fusión de clase II, como el péptido interno de fusión, también se presume que posiblemente E2 de *Pestivirus* y *Hepacivirus* poseen un mecanismo diferente de entrada a la célula y éstas son un nueva clase de proteínas de fusión (Garry y Dash *et al.*, 2002; Krey *et al.*, 2010; El Omari *et al.*, 2013; Li *et al.*, 2013; Kong *et al.*, 2013; Li y Modis, 2014).

Capítulo 2

Antecedentes Particulares

2. PROCESOS EVOLUTIVOS EN LA PROTEÍNA DE ENVOLTURA DE LA FAMILIA *FLAVIVIRIDAE*

Los virus son entidades dinámicamente evolutivas, que responden a cambios del medio ambiente y del microambiente (la respuesta inmune del hospedero).

Los diferentes mecanismos que permiten la evolución de los virus, afectan a nivel genómico, por ejemplo se ha observado que las tasas de sustitución (el número de nuevas mutaciones que llegan a fijarse en la población por unidad de tiempo) son diferentes a lo largo el genoma de un virus; por ejemplo, en el virus de influenza A la tasa de sustitución en los ocho genes es diferente, encontrando tasas de sustitución más elevadas en los genes de las proteínas de la envoltura: la hemaglutinina y la neuraminidasa (Dimmock *et al*, 2007; Qu *et al.*, 2011).

2.1 TASA DE MUTACIÓN Y TASA DE SUSTITUCIÓN

La tasa de mutación, se define como todas aquellas mutaciones que ocurren por unidad de tiempo o por ciclo de replicación. Se ha estimado que la tasa de mutación en los virus de RNA está entre 10^{-3} a 10^{-5} errores por nucleótido copiado.

La alta tasa de error de las RNA polimerasas en los virus se atribuye a dos razones principalmente:

La primera, es bien sabido que las RNA polimerasas de los virus de RNA, en comparación con las DNA polimerasas de los virus de DNA, tiene deficiencia de la actividad correctora o exonucleasa.

La segunda, es la ausencia de mecanismos celulares conocidos de reparación posreplicativa que puedan actuar sobre los híbridos de RNA-DNA o intermediarios replicativos de RNA bicatenarios (Duffy *et al.*, 2008).

También se han propuesto diferentes teorías que explican el mantenimiento de la alta tasa de mutación en los virus de RNA, por ejemplo Belshaw *et al.*, 2008, proponen que se debe a la **historia de vida** y con ello se refieren a que los virus son parásitos obligados y el organismo infectado posee un sistema de defensa que tiene como objetivo eliminar a estos parásitos intracelulares, por lo tanto, la capacidad de los virus de generar un gran variabilidad fenotípica les permite evadir al sistema de defensa. La segunda hipótesis que plantean son las **limitaciones evolutivas**, a esto se refieren que los virus de RNA en su replicación no tienen la opción de utilizar la maquinaria de replicación de los hospederos y por lo tanto utilizan la RNA polimerasa viral. La tercera es la **compensación en la velocidad de replicación**, es decir, los virus pueden replicarse rápido o de forma fidedigna, pero no ambas, los virus al replicarse de forma rápida pueden generar una cantidad importante de progenie viral en un corto tiempo (Belshaw *et al.*, 2008).

Por otro lado Duffy *et al.*, (2008) atribuyen la alta tasa de mutación de los virus de RNA al **mecanismo de replicación**. En general existen dos mecanismos de replicación: la “replicación lineal reiterativa” (del inglés *stamping machine*) y la “replicación geométrica”. En la primera se supone que se utiliza solo una molécula de material genético como molde, a partir de la cual se genera toda la progenie, cuando una mutación ocurre en el molde ésta es transmitida a toda la progenie. En el segundo mecanismo, propone que algunos genomas virales producidos durante las primeras generaciones son utilizados como templados o moldes para generar genomas virales de las progenies futuras, y cuando ocurre una mutación, ésta se puede propagar de forma exponencial a parte de la progenie (Duffy *et al.*, 2008).

En general se sabe que los virus de RNA tienen una tasa de mutación entre 0.4 y 1.1 mutaciones por genoma por ciclo de replicación, (Duffy *et al.*, 2008), muchas de estas mutaciones son deletéreas o neutrales y otras son ventajosas, éstas tienen una mayor probabilidad de fijarse en la población, por lo tanto, la tasa de sustitución suele ser diferente a la tasa de mutación en determinadas regiones del genoma de un virus; por ejemplo, se ha

reportado que la mayoría de las mutaciones no sinónimas se fijan en los genes que codifican para proteínas estructurales, como en la proteína de envoltura del DENV, donde la tasa de sustitución es de 0.58×10^{-3} sustituciones/sitio/año, para el serotipo dengue 1 y el serotipo de dengue 2 tiene una tasa de sustitución de 0.5×10^{-3} sustituciones/sitio/año. En el caso del virus de hepatitis C, en estudios de evolución intra-hospedero, se ha reportado que la proteína de envoltura presenta tasas de sustitución más altas que en el caso del dengue, según Gray *et al.*; (2011) las regiones codificantes para E1 y E2 presentan una tasa de sustitución de 2.5×10^{-3} a 6.9×10^{-3} sustituciones/sitio/año, comparado con las proteínas no estructurales cuya tasa es de 1.0×10^{-3} sustituciones/sitio/año. Estas altas tasas de sustitución generan una gran diversidad en la población viral. (Gray *et al.*, 2011; Goncalvez *et al.*, 2002).

Para el género *Pestivirus* no se han reportado las tasas de sustitución, pero en el virus de la diarrea viral bovina, se sabe, mediante análisis de presiones selectivas, que existen sustituciones no sinónimas en las proteínas estructurales E1 y E2. (Neill *et al.*, 2012).

2.2 SELECCIÓN NATURAL

La selección natural es un mecanismo evolutivo que se define como la sobrevivencia diferencial de individuos con genotipos distintos dentro de una población. (Perfectti *et al.*, 2009; Graur y Li, 2000)

Las mutaciones que ocurren en el genoma pueden o no alterar el fenotipo de un organismo, y en consecuencia alterar su adecuación. La mayoría de las nuevas mutaciones que surgen en una población reducen la adecuación de sus portadores, dichas mutaciones se denominan deletéreas, y serán seleccionados en contra y finalmente retiradas de la población. Este tipo de selección se denomina **selección negativa o purificadora**. Ocasionalmente una mutación puede aumentar la adecuación de sus portadores, dicha mutación se considera ventajosa, y la fuerza que le permite aumentar su frecuencia en la población se denomina **selección positiva o diversificadora**. (Pybus y Shapiro, 2009; Graur y Li, 2000).

Existen muchos enfoques para analizar el efecto de la selección natural en un conjunto de secuencias codificantes de DNA van desde los métodos de calcular la selección de forma

indirecta a partir de las frecuencias de los polimorfismos encontrados en un juego de secuencias, hasta los que la determinan sitio por sitio, o rama de una reconstrucción filogenética.

Los métodos más comúnmente utilizados para determinar la presencia de selección se basan en la observación de la relación **dN/dS** o ω que compara las tasas de sustituciones no sinónimas por sitio no sinónimo (dN) con la tasa de sustituciones sinónimas por sitio sinónimo (dS), partiendo del supuesto de que las mutaciones sinónimas se comportan de manera neutral, es decir no afectan la adecuación del individuo que las porta.

Entonces los valores obtenidos de la relación dN/dS son interpretados de la siguiente forma:

$$dN / dS = 1 \quad \text{neutral}$$

$$dN/dS < 1 \quad \text{selección negativa}$$

$$dN/dS > 1 \quad \text{selección positiva}$$

Los métodos modernos de dN/dS utilizan la historia filogenética y la reconstrucción de caracteres ancestrales, además de modelos estadísticos para determinar el número y tipo de sustituciones que han ocurrido en los codones. También son capaces de calcular el dN/dS por separado para cada codón, y para cualquier conjunto de codones de un alineamiento, localizando con precisión la adaptación molecular de un gen. (Kosakovsky *et al*, 2005)

En virología la identificación de sitios-específicos bajo selección natural se han utilizado en varios casos para encontrar sitios funcionalmente importantes (Nielsen, 2005).

La selección natural es uno de los procesos evolutivos más importantes, ya que favorece la fijación de mutaciones ventajosas dentro de la población, así, se ha observado que mutantes de escape de la respuesta inmune del hospedero se encuentran frecuentemente bajo selección positiva (Yamaguchi y Gojobori, 2000).

En el caso de los *Flavivirus* se ha reportado que la mayoría de los aminoácidos de la proteína de envoltura evoluciona bajo selección negativa y pocos sitios bajo selección positiva. Así por ejemplo, en el caso del DENV, se ha demostrado que los codones E131 y

E83 están bajo selección positiva en el virus DENV 2, (Añez *et al.*, 2011; Clyde *et al.*, 2006; Goncalvez *et al.*, 2002). Twiddy *et al.*; (2002) al analizar la misma proteína en DENV 1 al 4, se encontró únicamente selección positiva en DENV 2 (Twiddy *et al.*, 2002). De hecho en estudios filogenéticos de arbovirus que analizan la proporción de mutaciones no sinónimas en el tiempo demostró que la selección purificadora es generalmente la fuerza de selección dominante en la evolución de este tipo de virus (Ciota y Kramer, 2010).

En perteneciente al género *Pestivirus*, de igual forma que en caso del DENV, la mayoría de los sitios de la proteína E2 están sujetos a una fuerte selección purificadora y se han descrito cinco sitios seleccionados positivamente (886, 888, 905, 944, y 946). El resultado reportado fue sorprendente, ya que estos sitios caen dentro del C-terminal de E2 y no en la N-terminal, ésta última región expuesta en la superficie y responsable de la unión de anticuerpos (Tang y Zhang, 2007).

En contraste, en el HCV se ha determinado una fuerte selección positiva en E2 específicamente en las regiones hipervariables, se hipotetiza que la razón por la que las regiones hipervariables están bajo selección positiva es que son blanco de unión de anticuerpos, sin embargo otras regiones de la proteína están bajo selección negativa. Resulta interesante que exista una asociación estadísticamente significativa entre la evolución de la enfermedad y el número de sitios seleccionados positivamente, es decir, el número de sustituciones no sinónimas dentro de la región hipervariable 1 de E2 fueron más altas en los pacientes con progresión a hepatitis C y más bajas en los pacientes diagnosticados con la enfermedad resuelta o fulminante. (Brown *et al.*, 2005; Sheridan *et al.*, 2004; Farci *et al.*, 2002).

2.3 RECOMBINACIÓN

La recombinación es el intercambio de material genético entre genomas diferentes dando lugar a la formación de moléculas quiméricas de un origen mixto de genomas parentales. (Worobey y Holmes 1999). El modelo de recombinación más aceptado para los virus de RNA es el mecanismo de “**copy choice**”. En este proceso una RNA polimerasa, una RNA polimerasa dependiente de RNA o una transcriptasa inversa, cambian de una

molécula de RNA (templado donador) a otra molécula de RNA (templado aceptor) durante la síntesis de la nueva cadena, por lo tanto generan una molécula de RNA híbrida. Hay varios factores que influyen en el cambio del templado, como son el grado de identidad entre los templados de RNA, la cinética de transcripción y la estructura secundaria del RNA. La estructura secundaria puede promover el cambio de templado de la RNA polimerasa por estancamiento de la misma durante la replicación, lo que facilita la transferencia de la polimerasa a una molécula RNA aceptora (Simon-Loriere y Holmes, 2011; Lemey y Posada, 2005).

Como se mencionó anteriormente la recombinación, es un proceso evolutivo importante en los virus de RNA, este proceso se da a nivel de genoma, dando como resultado un aumento en la variabilidad de un organismo y también tiene repercusiones a nivel estructural de las proteínas (Ramírez *et al.*, 1999). Se ha reportado la presencia de recombinación en poblaciones de DENV, así por ejemplo, Worobey *et al.*; (1999) identificaron recombinación intraserotipo en los genes E, M y C, es decir, genes que codifican para proteínas estructurales (Worobey *et al.*, 1999). Por su parte, Holmes *et al.*; (1999) analizaron la presencia de recombinación en los cuatro serotipos de dengue y reportaron la presencia de recombinación en los serotipos DENV 1 y DENV 2, en el gen E, y no así en los otros serotipos (Holmes *et al.*, 1999).

En un estudio de Twiddy *et al.*; (2003) en otros arbovirus como virus de la fiebre amarilla, virus del oeste del Nilo y virus transmitidos por garrapata no se encontró recombinación, ni tampoco en DENV 1 o DENV 3, únicamente los virus recombinantes, ya reportados por Holmes *et al.*; (1999), Worobey *et al.*; (1999), Toul *et al.*; (2001) (Twiddy *et al.*, 2003). Sin embargo Rico-Hesse en el 2010, afirma que los genomas recombinantes descritos hasta la fecha para dengue han sido obtenidos por técnicas de amplificación o reacción en cadena de la polimerasa y por lo tanto pueden ser productos artificiales del laboratorio, y hasta ahora no se han aislado y caracterizado completamente recombinantes de DENV que se están transmitiendo en la naturaleza y causen enfermedad (Rico-Hesse, 2010).

En el caso de HCV, se tiene evidencia de recombinación específicamente entre la intersección de los genes E1/E2, sin embargo el punto de ruptura se encuentra al final de la

región hipervariable 1. (Cristina J y Rodney C., 2006) Para los *Pestivirus*, específicamente en el virus de la diarrea viral bovina, se ha demostrado la presencia de recombinación en el gen NS2, pero no se ha reportado la presencia de recombinación en las proteínas de envoltura de este virus (Becher *et al.*, 2001).

Igualmente en el virus de la fiebre porcina clásica se ha demostrado la evidencia de recombinación, pero ninguno de estos dos estudios los ha demostrado en el gen de la proteína E (Qiang *et al.*, 2007).

Recientemente se han cristalizado las proteínas de envoltura de los BVDV y HCV revelando que estas proteínas no presentan un plegamiento clásico de proteína de envoltura clase II, algunos autores suponen que ésta posiblemente fue adquirida de otro microorganismo por medio de la recombinación y los procesos evolutivos que han ocurrido en ellas no se conocen del todo. Así Li y Modis (2014) dejan algunas preguntas que faltan por resolver: ¿Los *Pesti* y *Hepacivirus* evolucionan de un ancestro en común? ¿Cuáles son los procesos evolutivos de estas proteínas? ¿Cuál es el mecanismo molecular de fusión de estas proteínas? (Li y Modis 2014).

2.4 Deriva Génica

La deriva génica se define como las fluctuaciones estocásticas en las frecuencias de alelos en una población. En general, en las poblaciones grandes la fuerza de la deriva genética es menor que en las poblaciones más pequeñas.

Aunque los virus tienen generalmente tamaños grandes de población, lo que sugeriría que los procesos genéticos en estas poblaciones son en su mayoría deterministas, se ha observado que la deriva génica es un factor importante en la evolución del virus. Por ejemplo, durante los eventos de transmisión muchos virus se someten a severos cuellos de botella poblacionales, lo que produce efectos fundadores frecuentes en sus poblaciones.

También se ha propuesto una contribución significativa de la deriva génica para la evolución de los virus humanos, tales como norovirus, el sarampión, virus de la hepatitis B, coronavirus, el virus de la rabia, el virus del dengue y en hantavirus (Reuter *et al.*, 2008; Rota *et al.*, 1994; Vorinin *et al.*, 2009; Wang *et al.*, 2002).

Sin embargo, la determinación precisa de la función de la deriva génica en la evolución de los virus que infectan animales es complicado, debido a que es difícil discernir entre la deriva génica y las diferencias individuales de la respuesta del sistema inmune de hospederos infectados (Voronin *et al.*, 2009).

Capítulo 3

Planteamiento del Problema

3. JUSTIFICACIÓN

La proteína de envoltura tiene funciones indispensables en el ciclo de vida de los virus de la Familia *Flaviviridae*, como son: la de unión al receptor, mediar la fusión de las membranas viral y celular durante el proceso de entrada, contribuir al ensamble del virión, además de que inducen la producción de anticuerpos neutralizantes. El hecho de que estas funciones están conservadas a pesar de las diferencias en las secuencias entre los distintos géneros, sugieren que pueden haber evolucionado a partir de un ancestro en común, sin embargo los procesos evolutivos que han actuado en estas proteína pueden ser diferentes, lo que las han llevado a especializarse en diferentes hospederos.

El estudio comparativo de la evolución molecular de la proteína de envoltura permitirá entender la dinámica evolutiva de los miembros de la familia *Flaviviridae*, además de contribuir al desarrollo de estrategias de control, también será de utilidad para entender como las diferencias en el ciclo de vida de los distintos virus han modelado la evolución de esta proteína viral.

3.1 HIPÓTESIS

La proteína de envoltura de la familia *Flaviviridae* está sometida a las presiones selectivas específicas de sus hospederos y del ciclo replicativo del virión, sin embargo los procesos evolutivos que ocurren en esta proteína son similares porque su función está conservada.

La proteína de envoltura de los virus de la familia *Flaviviridae*, es un blanco importante de la respuesta inmune del hospedero, por lo tanto la fuerza que domina la dinámica evolutiva de esta proteína es la selección natural positiva.

3.2 OBJETIVO GENERAL

- ✓ Investigar los procesos evolutivos en la proteína de envoltura en diferentes miembros de la familia *Flaviviridae*.

3.3 OBJETIVOS PARTICULARES

- ✓ Analizar la variabilidad presente en la proteína de envoltura.
- ✓ Identificar las fuerzas evolutivas (selección natural) que han actuado sobre la proteína de envoltura.
- ✓ Analizar el efecto de la selección natural sobre las propiedades bioquímicas en la proteína de envoltura e investigar sus posibles implicaciones biológicas.

Capítulo 4

Materiales y Métodos

4.1 OBTENCIÓN DE LAS SECUENCIAS Y ALINEAMIENTO

En este proyecto se utilizaron como modelo de estudio a las proteínas de envoltura del DENV, el BVDV y el HCV, esto porque son los representantes más importantes de los géneros que conforman a la familia *Flaviviridae*, tanto por sus repercusiones a nivel mundial, como por el número de secuencias que se encuentran en las bases de datos públicas.

Como primer paso se hizo una búsqueda de las secuencias de nucleótidos del gen de la envoltura (E) del DENV y BVDV, en la base de datos *GenBank* del *National Center for Biotechnology Information* (NCBI) (Benson *et al.* 2012). Para el HCV se obtuvieron las secuencias de la base de datos pública *Database Hepatitis C*, de los Alamos (Kuiken *et al.* 2005) (Tabla 3).

TABLA 3. NÚMERO DE SECUENCIAS EN LAS BASES DE DATOS

Virus	Base de datos	Número de secuencias	Fecha de la descarga	Referencia
DENV	NCBI	3760	1 de julio del 2013	Benson <i>et al.</i> 2012
BVDV	NCBI	572	28 de septiembre del 2013	Benson <i>et al.</i> 2012
HCV	Alamos	5313	7 octubre del 2013	Kuiken <i>et al.</i> 2005

Posteriormente, se llevaron a cabo alineamientos de todas las secuencias mediante el algoritmo *MAFFT* (Katoh, 2009) implementado en *SeaView* (Gouy *et al.*, 2010), y seguido de esto se excluyeron las secuencias redundantes o duplicadas, es decir, las secuencias con una identidad mayor al 95%, esto mediante la herramienta *redundance* del programa *Jalview* (Waterhouse *et al.*, 2009). También se eliminaron aquellas secuencias que contenían

ambigüedades. Los números de acceso de las secuencias utilizadas en los análisis subsecuentes de encuentran en la tabla 30 en el apéndice.

4.2 ANÁLISIS DE VARIABILIDAD

A partir de los alineamientos de nucleótidos del gen de la proteína de envoltura, se analizó su variabilidad, lo que permitió observar el grado de conservación de este gen, para esto se usó el programa *HyPhy (Hypothesis testing using Phylogenies)* (Pond *et al.*, 2005), dividiendo los alineamientos en cada uno de los dominios (I, II y III) descritos para cada una de las proteínas, con lo que se obtuvo el grado de variabilidad para cada dominio. Se realizó este mismo análisis para las secuencias traducidas.

4.3 USO DE CODÓN

Dado que se ha comprobado que el uso de codón sinónimo varía entre genomas, entre genes y dentro de los genes, y que dicho sesgo puede ser el resultado de tendencias mutaciones o de la selección natural llevamos a cabo un análisis de este tipo. Para ello se utilizó la función *Codon usage bias* del programa DNA SP Versión 5.10, (Librado y Rozas, 2009). El programa utiliza el algoritmo de uso relativo de codones sinónimos (RSCU), los valores del RSCU de cada codón de un gen se calculan a partir de las frecuencias del uso de codón sinónimo.

Para ello se utilizaron las secuencia del gen E del virus DENV 1(D00505), DENV 2 (AB587932), DENV 3 (JX891665), DENV 4 (AB587932), HCV (EU255950), BVDV (DQ328323), también se incorporaron otros virus como Virus del Oeste del Nilo (*West nile virus*, WNV, 001563.2) y Virus de la fiebre amarilla (*Yellow virus* YVW, 002031.1) ambos del género *Flavivirus*, para el género *Pestivirus* se incorporó el virus de la fiebre porcina clásica (*Classical swine fever virus*, CSFV, 002671) y el virus de la hepatitis GB (HGB, 001837.1) del género *Hepacivirus*. En todos los casos se usó el gen de la proteína E.

También se obtuvo el uso relativo de los codones sinónimos de los hospederos de los virus mencionados arriba *Homo sapiens* (humano), *Bos primigenius taurus* (vaca), *Sus scrofa domestica* (puerco), *Ovis aries* (borrego), *A.aegypti*, *A.albopictus* (mosquitos) además

de que se adicionaron 2 organismos de referencia estos fueron la *Rana huanrenensis* (rana) y *Arabidopsis thaliana*. Los valores de frecuencia del uso de codón se obtuvieron de la base de datos *Codon Usage Database* (Nakamura *et al* 2000). Una vez obtenidos estos valores se procedió a obtener el valor de RSCU a partir de la siguiente formula:

$$RSCU = \frac{X_i}{\bar{x}}$$

Dónde:

X_i = Número observado de un determinado tipo de codón

$\bar{x}$ = media de x_i en el grupo sinónimo.

Es decir, el RSCU es la frecuencia observada de cada codón dividida por la frecuencia esperada, suponiendo un uso de codón aleatorio.

Para poder determinar si existía una relación entre el uso de codón de los virus con sus respectivos hospederos se compararon los valores de RSCU utilizando métodos de estadística multivariada, los cuales analizan, representan e interpretan los datos que resultan de observar más de una variable estadística sobre una muestra de individuos (Cuadras, 2014). Los métodos usados fueron *Clustering* Jerárquico o Agrupamiento Jerárquico, Escalamiento Multidimensional y *K-means* y cada uno de ellos se ejecutaron en el programa de estadística libre R (R Core Team, 2013).

Los métodos utilizados se describen a continuación:

AGRUPAMIENTO JERÁRQUICO

El objetivo de este método es agrupar un conjunto de datos multidimensionales en un conjunto homogéneo. El criterio que toma para el agrupamiento es la similitud entre los datos o casos. Existen diferentes tipos de *clustering* jerárquico o agrupamiento jerárquico los métodos de división jerárquica y los métodos aglomerativos. Los primeros asumen que en primer paso todos los datos conforman una sólo conglomerada o cluster, y posteriormente

éste se va dividiendo sucesivamente en cluster más pequeños de acuerdo a algún criterio seleccionado previamente. El resultado de este procedimiento se representa por medio de un dendrograma. Los segundos asumen que cada observación inicialmente como un cluster o grupo, y en cada paso se asocian los cluster más similares hasta llegar a formar un sólo cluster. (Venable y Rigley, 2002). En este caso se llevó a cabo un método de división jerárquica con el paquete `pvclust` del programa R (R Core Team, 2013). Éste algoritmo produce un dendrograma donde se representan los datos, al que se le asignan valores de soporte estadístico por medio de una probabilidad de *bootstrap*, donde el soporte es la frecuencia con la que aparece determinado cluster durante las réplicas de remuestreo. Además el dendrograma es soportado por una probabilidad no sesgada aproximada (*approximately unbiased*, AU), obtenida a partir de la probabilidad de *bootstrap* (Suzuki y Shimodaira, 2006). El *clustering* jerárquico fue llevado a cabo con 1000 réplicas de *bootstrap* y se utilizó el método *average* para determinar el número de clusters.

ESCALAMIENTO MULTIDIMENSIONAL.

Este método se basa en representar los datos en un espacio de pocas dimensiones que refleja la proximidad o similitud entre ellos. El criterio que toma para obtener la similitud entre dos datos es la distancia euclidiana, lo que permite que los datos sean representados de forma gráfica en un espacio euclidiano de baja dimensión.

A este método se le aplicó una prueba estadística llamada prueba de bondad de ajuste entre las distancias que informa que tan bien se ajustan los datos al representarlos en un espacio de únicamente 2 dimensiones.

K-MEANS

El algoritmo *K-means*, es un método de estadística multivariada de agrupamiento, aquí el método trabaja con un número de clusters pre especificado y minimiza la distancia intracluster. Dados los resultados observados en el escalamiento multidimensional y el Clustering jerárquico, donde se obtuvieron 2 clusters, se llevó a cabo el análisis de *K-means* para 2 clusters. Posteriormente el método calcula la distancia entre los datos o puntos dentro de cada cluster, seguido de estos, reasigna los datos a un cluster cuya distancia sea

la mínima de todos los clusters, este procedimiento se repite hasta que no se produzca ninguna resignación de los datos, es decir, hasta que todos los elementos se establezcan dentro de un cluster. Al igual que el escalamiento multidimensional toma como criterio para clasificación o agrupamiento la distancia euclidiana, además obtiene las medias de distancia intercluster por el método *average*. Este método de igual forma se desarrolló con los valores de RSCU obtenidos anteriormente (Veneble y Ripley, 2002).

4.4 ANÁLISIS DE LA PRESENCIA DE RECOMBINACIÓN.

La presencia de recombinación en las secuencias del gen E de cada uno de los virus se determinó utilizando los métodos SBP y GARD, implementado en el servidor *Datamonkey* (Pond y Frost, 2005). Ambos se basan en la detección de recombinación por medio de la comparación de filogenias de la partición de un alineamiento. Los métodos emplean el método Neighbor-Joining (NJ) con el modelo de sustitución de nucleótidos Tamura Nei 93, para reconstruir la topología de un árbol en cada fragmento no recombinante y compara el ajuste del modelo a los datos en presencia y ausencia de recombinación. (Pond *et al.*, 2006).

4.5 RECONSTRUCCIÓN DE FILOGENIAS.

La reconstrucción de árboles filogenéticos, se realizó mediante la inferencia bayesiana, usando el programa *Mr Bayes* 3.2 (Ronquist y Huelsenbeck, 2003) que estima la probabilidad de que determinadas topologías hayan sido producidas por el alineamiento, obteniendo una población de árboles filogenéticos a partir de la cual se construye un árbol consenso (Peña, 2011). Para construir las filogenias, se utilizaron los alineamientos en nucleótidos de los DENV, BVDV y HCV. Se corrió el programa 5, 000 000 de generaciones con muestreos cada 100 generaciones. Se destacaron las primeras 25 000 generaciones o 0.25 % de *burnin*. Con los árboles restantes se construyó un consenso con la regla de mayoría.

4.6 MÉTODOS DE SELECCIÓN.

La identificación de las fuerzas evolutivas que han actuado sobre la proteína E, como la selección natural, se usaron los alineamientos de codones del gen E de cada uno de los virus, para esto se ejecutaron dos algoritmos implementados en el servidor *Datamonkey* (Pond y Frost, 2005a) el de **probabilidad de efectos fijos (FEL)** y el de **probabilidad de efectos fijos internos (IFEL)**, estos utilizan los principios de máxima verosimilitud para estimar la proporción de tasas sinónimas y no sinónimas de cada sitio (Pond y Frost, 2005a). Estos métodos detectan la selección natural en un gen codificante, identificando las tasas de sustitución no sinónimas (dN cuando cambia un aminoácido) más altas en relación con las tasas de sustitución sinónimas (dS, mutaciones silenciosas) que se consideran neutrales, esto se representa como la relación $\omega = dN/dS$ (Tabla 4). Ambos ajustan la tasa de sustitución entre dN y dS sobre una base de sitio-por-sitio, la diferencia entre es que IFEL la selección natural es evaluada solamente a lo largo las ramas internas de las filogenias (Pond y Frost, 2005).

TABLA 4. INTERPRETACIÓN DE LOS RESULTADOS DE SELECCIÓN NATURAL

Tipo de selección	Tasa ω ó dN/dS esperada
Neutralidad	ω ó dN/dS=1
Selección positiva (diversificadora o adaptiva)	ω ó dN/dS > 1
Selección negativa (purificadora)	ω ó dN/dS < 1

4.6.1 *Evolutionary Fingerprinting* del gen de la proteína de envoltura.

También como parte de los análisis de selección natural, se efectuó un análisis en la herramienta ***evolutionary fingerprinting*** implementada en el servidor *Datamonkey* (Pond *et al.*, 2009), con ella se realizó comparación global de las tasas dN/dS de cada uno de los alineamientos de las proteínas de envoltura de los virus.

Este método plantea que los genes que codifican para proteínas con estructura y función similar tienen *fingerprints* evolutivos, que pueden ser utilizados para descubrir similitudes entre genes. El método realiza categorías de las tasas de dN/dS, estas son representadas en un gráfico donde se muestra la distribución de estas categorías. Para obtener la distribución de las categorías, asumen que la distribución de las tasas o dN/dS se explican mejor por medio de una distribución discreta bivariada. Dentro del gráfico donde se representan los resultados se muestra la distribución de las tasas o dN/dS como elipses o circunferencias, éstas varían en tamaño, el cual es proporcional a la cantidad de tasas dN/dS que se agrupan dentro de una determinada categoría. Por lo tanto las elipses reflejan si los valores de las tasas dN/dS son similares en distintos genes (Pond *et al.*, 2009).

4.6.2 Determinación de selección episódica.

Posteriormente, se analizó la presencia de selección episódica en los alineamientos de nucleótidos de la proteína de envoltura. Los métodos sitios por sitio para la detección de selección natural asumen que ω es constante a lo largo de la filogenia lo que no permite detectar momentos específicos en la historia de un gen en la que las fuerzas evolutivas actuaran de manera distinta. Para identificar procesos de selección episódica se utilizó la herramienta llamada ***Mixed effects model of evolution (MEME)*** (Murrell *et al.*, 2012), que permite que la distribución de ω puede variar sitio por sitio (efectos fijos) y también rama por rama en un sitio, además el método es capaz identificar *footprints* de selección episódica y penetrante (Murrell *et al.*, 2012).

4.7 DETERMINACIÓN DE EPÍTOPES.

El sistema inmune es una fuente de presión de selección que puede influir en la evolución de los virus, por ello se analizó si los sitios que han evolucionado bajo selección positiva forman parte de epítopes de reconocimiento de linfocitos B y T. Esta predicción se realizó únicamente para los virus HCV y DENV porque los servidores sólo pueden predecir secuencias de epítopes de linfocitos de humano, para ello se obtuvieron las secuencias consenso de los alineamientos de la proteína de envoltura. La predicción de epítopes de linfocitos B se hizo usando el servidor *BCPREDS server* (Chen *et al.*, 2007) con los criterios de 75% de especificidad y epítopes de longitud de 12 aa. Mientras que la predicción de linfocitos T se realizó en el servidor *ProPed-I* (Singh y Raghava, 2001) contiene alelos del MHC-I.

Para el BVDV se determinó la presencia de epítopes en las base de datos *ABC pred* (Saha y Raghava, 2007) que determina únicamente linfocitos B, sin embargo utiliza una base de datos de una amplia gama de organismos. Los criterios que se tomaron fueron los mismos epítopes de longitud de 12 aa y un umbral de 0.51.

4.8 RECONSTRUCCIÓN DE CARACTERES ANCESTRALES.

La reconstrucción de caracteres ancestrales es una reconstrucción de las secuencias en los nodos internos dentro una filogenia, realizar este tipo de análisis permite observar cómo ha cambiado el estado de un carácter en particular a lo largo de la filogenia.

La reconstrucciones ancestrales, se realizaron por medio de dos enfoques, uno por el método de parsimonia implementado en el programa *MacCLADE* (Maddison y Maddison, 1989) y otro método basado en máxima verosimilitud implementado en el servidor *FastML server* (Ashkenazy *et al.*, 2012), una diferencia importante entre estos es que el primero realiza la reconstrucción en aminoácidos y el segundo en codones.

4.8.1 Parsimonia.

La reconstrucción de caracteres ancestrales por parsimonia tiene la premisa de que el árbol correcto es aquel que tenga un menor número de cambios o pasos evolutivos (cambios de estado de un carácter) este método es efectivo para inferir caracteres ancestrales cuando las secuencias han divergido poco. Como se mencionó anteriormente la reconstrucción se llevaron acabo en el programa *MacClade* (Maddison y Maddison 1989) en general la capacidad para poder determinar la secuencia ancestral aminoácidos por este método depende del número de secuencias y la topología del árbol, el programa incluye dos algoritmos *ACCTRAN* y *DELTRAN*, el primero acelera la transformación de un carácter del árbol y el segundo retrasa la transformación de un carácter. Sí los dos algoritmos dan los mismo resultados, la inferencia de aminoácidos se tiene fuerte precisión en los resultados. Cuando ambos dan diferentes resultados, hay diferentes posibles caracteres ancestrales que puede considerar.

4.8.2 Máxima verosimilitud.

Las reconstrucciones de caracteres ancestrales por codón, se construyeron por el criterio de máxima verosimilitud en el servidor *FastML* (Ashkenazy *et al.*, 2012), el procedimiento de reconstrucción se puede dividir en dos partes: la primera es una reconstrucción de caracteres y la segunda en una reconstrucción de *indels*, posteriormente ambos resultados son integrados en la reconstrucción de caracteres más probable. Los cambios en los estados de carácter se infirieron bajo el modelo de sustitución JTT (Ashkenazy *et al.*, 2012).

4.9 ANÁLISIS DE CAMBIOS EN LAS PROPIEDADES BIOQUÍMICAS.

Los alineamientos de nucleótidos de cada uno de las proteínas de envoltura de los virus, fueron utilizados para evaluar el cambio en las en las propiedades de los aminoácidos de la proteína E a lo largo de la filogenia, y con ello se observó cómo han sido preservados o

modificados los estados de carácter en la evolución por ejemplo, sí un carácter es seleccionado positivamente, la evolución pudo haber actuado cambiando el volumen de la cadena lateral del aminoácido, mientras que la polaridad se mantenga conservada. Esto se llevó a cabo con la herramienta ***Property Informed Models of Evolution (PRIME)*** (Pond y Frost, 2005a), incluida dentro del servidor *Datamonkey* (Pond y Frost, 2005a), que se basa en los modelos conceptuales de IFEL y MEME (selección episódica); es decir, estima directamente la sustituciones sinónimas y no sinónimas en cada sitio, en cada sitio hay un número independiente de sustituciones y no supone que la tasa de sustitución sea igual en todos los sitios. Sin embargo las tasas de sustitución para cada codón en este programa dependen directamente de un conjunto de propiedades bioquímicas seleccionadas para cada residuo. Éste análisis se desarrolló bajo los criterios de Conant y Stedler (Conant *et al.*, 2007), que evaluó 5 propiedades bioquímicas, composición bioquímica, volumen, punto isoelectrico, e hidropatía.

De los resultados obtenidos se contabilizó cuál era la propiedad bioquímica que se conservaba o cambiaba de forma general en las proteínas de envoltura.

Por otra parte, de los sitios seleccionadas bajo selección positiva y selección negativa que presentaban variaciones a nivel de secuencia, se realizó un análisis de forma manual, en este análisis únicamente se evaluó el cambio de polaridad, hidropatía, punto isoelectrico y el volumen. Los resultados obtenidos se vaciaron en una tabla donde se contabilizó cual es la propiedad que tiene tendencia a ser conservada o cambiada por los procesos evolutivos.

4.10 MAPEO DE LAS PROPIEDADES BIOQUÍMICAS Y SITIOS BAJO SELECCIÓN EN LA ESTRUCTURA TERCIARIA DE LA PROTEÍNA.

El mapeo de los sitios bajo selección positiva y negativa, así como de los cambios en las propiedades bioquímicas obtenidos por medio de PRIME (Pond y Frost, 2005), en las proteínas de envoltura de DENV (1TG8) (Zhang *et al.*, 2004) y BVDV (4ILD) (Li *et al.*, 2013) y el modelo de la proteína de envoltura del HCV obtenido del servidor *Phyre 2* (Kelley y Sternberg *et al.*, 2009). La visualización de las estructuras y el mapeo se realizó con el programa *Chimera* (Pettersen EF *et al.*, 2004).

4.11 MODELADO Y ANÁLISIS DE LAS PROTEÍNAS MUTANTES.

Los sitios que mostraron cambios en la propiedad bioquímica del volumen en las proteínas de envoltura, se realizó un análisis de mutagénesis *in silico* con el objetivo de observar si dicho cambio generaba cambios importantes en la estructura terciaria de la proteína. Esto se llevó a cabo en el servidor *WHAT IF* (Vriend 1990), que es un paquete de modelado molecular versátil que se especializa en el trabajo con proteínas y las moléculas en su entorno, como el agua, ligandos, ácidos nucleicos, etc. Posteriormente se refinaron los modelos obtenidos por medio de minimización de la energía, seguido de esto se realizó un alineamiento estructural de las proteínas mutantes contra el PDB de la base de datos.

Una vez que se obtuvieron las proteínas mutantes, en el mismo servidor *WHAT IF* se repararon los modelos para *Docking* molecular, aquí se modelan las cadenas laterales, se corrigen las posiciones donde hay agua y se añaden hidrógenos. Posteriormente se evaluó la calidad de las proteínas en el servidor *Molprobit* (Chen *et al.*, 2009), donde además también se repara los PDB adicionando hidrógenos, corrigiendo ramas sobrelapadas y giros no permitidos en las proteínas, posteriormente se refinaron las proteínas por minimización de energía con el servidor *3Drefine* (Bhattacharya y Cheng, 2013), una vez refinadas nuevamente se procedió a evaluar las proteínas en el servidor *Molprobit*.

Las proteínas mutantes y proteínas nativas se alinearon estructuralmente en el programa *Chimera* (Pettersen EF *et al.*, 2004) y se obtuvo el valor RMSD de cada uno de estos alineamientos.

Capítulo 5

Resultados

5.1 ANÁLISIS DE LA VARIABILIDAD.

Para analizar el grado de variabilidad de la proteína E de DENV, HCV, y BVDV, que son los tres modelos de estudio de este trabajo, se obtuvieron las secuencias de las bases de datos públicas señaladas en la tabla 3. Posteriormente se construyó un alineamiento múltiple, el cual se dividió en tres segmentos, correspondientes a los tres dominios, como se indica en la sección de Métodos. Los resultados obtenidos se muestran en la tabla 5, donde se observa el número de sitios constantes y polimórficos, se observa que la variabilidad se encuentra distribuida de manera uniforme en los tres dominios, en un rango del 63 al 80 %.

Con el objetivo de determinar si este alto grado de variabilidad se reflejaba también en las secuencias de la proteína, los alineamientos en nucleótidos se tradujeron con el programa seaview (Gouy *et al.*, 2010).

En los alineamientos de aminoácidos también se obtuvo el porcentaje de variabilidad por dominio, y los resultados obtenidos se muestran en la tabla 6. Se puede observar que el grado de variabilidad a nivel de proteína es similar al de nucleótidos, lo que indica que muchos de los cambios a nivel de nucleótido corresponden a mutaciones no sinónimas.

TABLA 5. VARIABILIDAD A NIVEL DE NUCLEÓTIDOS EN EL GEN QUE CODIFICA PARA LA PROTEÍNA DE ENVOLTURA

Dominio	Sitos constantes y polimórficos	DENV		BVDV		HCV	
I	Constantes	144/393	36.64 %	53/263	20.25 %	85/314	27.07 %
	Polimórficos	175/393	63.30 %	210/263	79.84 %	256/314	72.92 %
II	Constantes	175/500	35.00 %	73/230	31.60 %	61/247	24.70 %
	Polimórficos	352/500	65.00 %	157/230	68.26 %	186/247	75.30 %
III	Constantes	103/296	37.79 %	88/320	27.41 %	73/214	34.11 %
	Polimórficos	193/296	65.20 %	232/320	72.50 %	141/214	67.28 %

TABLA 6. VARIABILIDAD A NIVEL DE AMINOÁCIDOS LA PROTEÍNA DE ENVOLTURA

Dominio	Sitos constantes y polimórficos	DENV		BVDV		HCV	
I	Constantes	49/130	37.69 %	64/167	38.82 %	34/98	34.69 %
	Variables	81/130	62.30 %	103/167	61.67 %	64/98	65.30 %
II	Constantes	15/87	17.24 %	22/76	28.57 %	27/106	23.23 %
	Variables	72/87	87.75 %	54/76	75.00 %	79/106	74.52 %
III	Constantes	26/104	25.00 %	16/82	19.51 %	26/71	36.36 %
	Variables	78/104	75.00 %	66/82	80.50 %	45/71	63.38 %

5.2 USO DE CODÓN.

Una manera de analizar la variabilidad de las secuencias de virus es determinar si existe un sesgo en el uso de codón y si este a su vez es similar al que presentan sus hospederos. Este tipo de análisis permite formular hipótesis respecto a las fuerzas evolutivas que han moldeado la variabilidad que observamos. El análisis de uso de codón de la proteína de envoltura de los virus con sus respectivos hospederos se llevó a cabo con el programa DNASP 5.10 (Librado y Rozas, 2009), con el cuál se calcularon los valores de

RSCU. Estos resultados se muestran en la tabla 7, donde en rojo se muestra los codones que tienen un $RSCU > 1$, que indica que ese codón está sobre-representado, y los que tienen un $RSCU < 0.6$ en verde, que corresponde a codones poco representados (Ma, MR. *et al.*, 2011). Por ejemplo, entre los codones con mayores sesgos se encontró que para el caso de la Arginina, el codón AGA fue preferido por todos los virus con excepción de HCV y mientras que el codón menos utilizado en todos los virus fue el CGT.

De los valores de RSCU obtenidos se puede observar que el uso de algunos codones es distinto al de sus hospederos, así por ejemplo en general el codón CTG que codifica para Leucina es preferido por todos los hospederos vertebrados.

TABLA 7. USO DE CODON DE LA PROTEÍNA DE ENVOLTURA DE 7 MIEMBROS DE LA FAMILIA FLAVIVIRIDAE Y SUS HOSPEDEROS.

Codón	Dengue 1	Dengue 4	Dengue 3	Dengue 2	YFV	WNV	Hepatitis C	HGB	VDVB	CSFV	Humano	Vaca	Borrego	Puerco	A.aegypti	A.albopictus	A.thaliana	Rana
AAA	1.43	1.1	1.03	1.28	1.13	0.77	0.86	0.55	0.89	0.67	0.87	0.78	0.68	0.76	0.8	0.58	0.98	1.6
AAC	1	1.25	1	1.18	1.1	1.5	1.41	0.73	1	1.11	1.06	1.19	1.51	1.21	1.2	1.36	0.96	1.34
AAG	0.57	0.9	0.97	0.72	0.87	1.23	1.14	1.45	1.11	1.33	1.17	1.22	1.32	1.24	1.2	1.42	1.02	0.4
AAT	1	0.75	1	0.82	0.9	0.5	0.59	1.27	1	0.89	0.94	0.81	0.49	0.79	0.8	0.64	1.04	1.32
ACA	1.53	2.13	2.22	2.46	1.33	1.33	0.57	0.32	1.42	1.44	1.14	1.01	0.78	0.92	0.72	0.6	1.24	1.44
ACC	0.95	0.71	0.89	0.72	0.72	1.44	2.14	1.92	1.03	1.33	1.42	1.55	2.05	1.68	1.48	1.8	0.8	0.48
ACG	0.73	0.71	0.53	0.62	0.1	0.62	0.86	0.64	0.39	0.41	0.46	0.56	0.38	0.57	1	1	0.6	0
ACT	0.8	0.44	0.36	0.21	1.85	0.62	0.43	1.12	1.16	0.82	0.99	0.89	0.78	0.83	0.8	0.64	1.36	2.08
AGA	2.8	3.79	3.9	3	3.38	2.84	0.32	2	3.88	3	1.29	1.14	1.12	1.12	0.65	0.59	2.09	1.79
AGC	1.45	0.58	0.64	0.6	1.35	1.02	2.08	0.48	0.75	0.86	1.12	1.53	1.58	1.62	1.07	1.07	0.78	1.79
AGG	0.8	1.26	0.6	1.5	1.5	0.95	2.84	3	1.06	2.33	1.27	1.2	1.16	1.23	0.54	0.42	1.2	0.59
AGT	0.21	0.58	0.43	1	0.77	0.88	0.69	1.44	1.5	1.71	1.22	0.87	1.48	0.77	0.96	0.77	0.96	1.14
ATA	1.56	1.1	1.2	1.32	0.58	1.05	0.43	0.33	1.5	1.95	0.51	0.45	0.63	0.42	0.39	0.39	0.72	0.48
ATC	0.72	0.9	0.6	1.15	1.06	0.9	1.71	2.33	0.75	0.45	1.41	1.57	1.74	1.67	1.6	1.88	0.51	0.36
ATT	0.72	1	1.2	0.53	1.35	1.05	0.86	0.33	0.75	0.6	1.08	0.98	0.63	0.91	1	0.75	1.32	1.94
CAA	0.84	0.83	1.17	1.16	0.38	0.92	0.44	0.2	1.43	1.14	0.53	0.46	0.57	0.44	0.82	0.6	1.12	1.5
CAC	1	1.08	0.77	0.73	0.46	1.08	1.78	1.11	0.86	1.11	1.16	1.25	0.92	1.3	1.16	1.24	1.36	0
CAG	1.16	1.08	0.83	0.84	1.63	1.08	1.56	1.8	0.57	0.86	1.47	1.54	1.43	1.56	1.16	1.4	0.88	0.5
CAT	1	0.92	1.23	1.27	1.54	0.92	0.22	0.89	1.14	0.89	0.84	0.75	1.08	0.7	0.84	0.76	1.22	2
CCA	2	1.87	1.33	2.22	1.23	1.65	0.8	1.29	1.11	1.47	1.11	1	1.03	0.94	1.2	1.08	1.32	2.32
CCC	0.29	1.6	0.64	0.67	0.92	1.18	1.28	0.86	0.89	1.05	1.29	1.39	1.29	1.46	0.84	1.12	0.44	0.44
CCG	0	0	0.27	0.44	0.31	0.47	0.8	0.86	0.44	0.63	0.45	0.53	0.42	0.56	1.32	1.44	1.07	0.2
CCT	1.71	0.53	2.57	0.67	1.54	0.71	1.12	1	1.56	0.84	1.15	1.08	1.26	1.05	0.68	0.36	1.52	1.56
CGA	2	0.32	0.6	0	0	0.63	0.32	0	0.35	0	0.65	0.68	0.89	0.6	1.2	0.96	0.72	1.19
CGC	0.4	0.63	0.3	0.75	0.75	0.32	1.89	0.8	0	0.33	1.1	1.17	1.15	1.31	1.26	1.31	0.42	0.59
CGG	0	0	0.3	0	0	0.95	0.63	1.6	0.71	0.33	1.21	1.32	0.86	1.29	0.96	1.2	0.54	0.59
CGT	0	0	0.3	0.75	0.38	0.32	0	0.4	0	0	0.48	0.49	0.82	0.44	1.38	1.49	1.02	1.19
CTA	1.17	1.09	1.41	0.75	1.03	0.44	0	0.58	0.86	0.77	0.43	0.36	0.24	0.33	0.53	0.53	0.66	0.84
CTC	0.13	0.91	0.88	1.2	0.69	1.02	1.71	0.97	1.2	0.89	1.17	1.26	1.83	1.35	0.83	0.84	1.01	0.54
CTG	2.61	1.09	1.41	1.95	1.37	2.34	1.54	1.35	1.2	1.4	2.73	2.59	2.46	2.68	2.27	2.75	0.66	0.84
CTT	0.39	0.73	0.35	0	0.86	0.73	1.2	1.16	0.51	0.51	0.79	0.7	0.74	0.65	0.65	0.48	1.56	1.85
GAA	1.29	1.59	1.31	1.54	0.92	1.17	1.09	0.4	1.12	0.83	0.84	0.78	0.75	0.72	1.16	1.1	1.04	1.58
GAC	1.11	1.1	1.24	1.29	1.19	1.2	1.17	1.08	1.06	1	1.07	1.16	1.34	1.2	0.88	1.04	0.64	1.06
GAG	0.71	0.41	0.69	0.46	1.08	0.83	0.91	1.6	0.88	1.18	1.16	1.22	1.25	1.28	0.84	0.9	0.96	0.42
GAT	0.89	0.9	0.76	0.71	0.81	0.8	0.83	0.92	0.94	1	0.93	0.84	0.66	0.8	1.12	0.96	1.36	0.94
GCA	1.79	1.85	1.43	1.27	0.56	0.37	0.73	0.51	0.53	2.1	0.91	0.8	0.9	0.74	0.76	0.6	1.08	0.76
GCC	0.41	0.92	1.29	1.09	1.78	1.3	1.82	1.44	2.13	0.95	1.6	1.71	1.55	1.8	1.48	1.8	0.64	1.36
GCG	0.41	0	0.14	0.73	0.56	0.65	1.27	0.62	0.27	0.19	0.42	0.48	0.37	0.5	0.68	0.6	0.56	0.36
GCT	1.38	1.23	1.14	0.91	1.11	1.67	0.18	1.44	1.07	0.76	1.06	1	1.18	0.96	1.08	1	1.72	1.52
GGA	2.4	2.14	2.25	2.61	1.44	2.49	0.77	0.57	1.42	0.76	1	0.95	1.05	0.91	1.48	1.2	1.48	2.12
GGC	0.4	0.57	0.44	0.57	1.28	0.45	1.68	0.98	0.77	0.65	1.35	1.43	1.33	1.46	1.04	1.08	0.56	0.49
GGG	0.8	0.86	0.65	0.65	1.12	0.53	0.9	0.9	1.29	1.41	1	0.99	0.71	1.05	0.36	0.48	0.64	0.28
GGT	0.4	0.43	0.65	0.16	0.16	0.53	0.65	1.55	0.52	1.19	0.65	0.64	0.92	0.57	1.12	1.24	1.36	1.12
GTA	0.39	0.65	0.77	0.72	0.53	0.27	0.35	0.08	2.08	0.98	0.47	0.4	0.36	0.34	0.6	0.52	0.6	1.04
GTC	0.98	0.84	0.77	1.13	0.62	0.45	1.74	1.44	0.64	1.24	0.95	1.01	0.91	1.07	1.08	1.32	0.76	0.24
GTG	1.56	1.67	1.53	1.33	1.69	1.64	1.57	1.76	1.12	1.24	1.85	1.95	2.27	2.03	1.28	1.32	1.04	0.96
GTT	1.07	0.84	0.94	0.82	1.16	1.64	0.35	0.72	0.16	0.53	0.73	0.64	0.46	0.57	1.04	0.88	1.6	1.76
TAC	1	0.91	0.91	0.89	1.33	1	1.33	1	1.16	1.27	1.11	1.21	1.28	1.27	1.36	0.72	0.96	0.86
TAT	1	1.09	1.09	1.11	0.67	1	0.67	1	0.84	0.73	0.89	0.79	0.72	0.73	0.64	0.56	1.04	1.12
TCA	2.69	2.71	2.57	2	1.74	1.61	0.69	0.72	1.5	2	0.9	0.79	0.48	0.73	0.66	0.48	1.19	1.32
TCC	0.41	0.97	0.64	1.2	1.16	0.59	1.62	0.72	0.38	0.86	1.31	1.37	1.28	1.5	1.19	1.37	0.78	0.42
TCG	0.83	0.39	0.64	0.2	0.19	0.73	0.23	1.44	0.38	0.29	0.33	0.39	0.28	0.39	1.43	1.67	0.59	0
TCT	0.41	0.77	1.07	1	0.77	1.17	0.69	1.2	1.5	0.29	1.13	1.04	0.91	0.99	0.66	0.54	1.67	1.32
TGC	1.38	1.54	1.23	1.5	1.33	0.5	1.1	0.84	1.18	1.2	1.09	1.15	1.28	1.21	1.16	1.3	0.8	0.58
TGT	0.62	0.46	0.77	0.5	0.67	1.5	0.9	1.16	0.82	0.8	0.91	0.85	0.72	0.79	0.84	0.7	1.2	1.42
TTA	0.65	0.91	1.24	0.75	0	0.15	0.34	0	1.37	0.64	0.46	0.38	0.24	0.32	0.36	0.24	0.84	0.42
TTC	0.89	0.21	0.9	1.22	0.74	0.87	1.82	1.44	1.07	1.47	1.07	1.15	1.06	1.21	1.44	1.52	0.98	0.76
TTG	1.04	1.27	0.71	1.35	2.06	1.32	1.2	1.94	0.86	1.79	0.77	0.71	0.49	0.67	1.31	1.14	1.32	1.43
TTT	1.11	1.79	1.1	0.78	1.26	1.13	0.18	0.56	0.93	0.53	0.93	0.85	0.94	0.79	0.56	0.48	1.02	1.24

Para analizar de una manera sistemática los patrones de variación del RSCU de los virus y sus hospederos realizamos análisis estadísticos multivariados para establecer, a partir de numerosos datos y variables, relaciones de similitud que permitieran clasificarlos. En este punto se utilizaron tres métodos de estadística multivariada: un agrupamiento jerárquico, un escalamiento multidimensional y el clasificador basado en *k-means*.

Primero los resultados del análisis de agrupamiento jerárquico mostraron dos grandes clústers o grupos, en el primero de estos se agrupan los hospederos y en el segundo a los virus. Sin embargo, se observó que el género *Hepacivirus* se encontró dentro del mismo clúster que su hospedero, lo que indica que el sesgo en el uso de codón es de estos virus es similar al de sus hospederos, mientras que para los virus del género *Flavivirus* y *Pestivirus* se agruparon aparte de sus hospederos (Figura 5).

Los resultados obtenidos del escalamiento multidimensional y el clasificador basado en *k-means* fueron similares a los obtenidos en el agrupamiento jerárquico, esto se muestran en el apartado de completo.

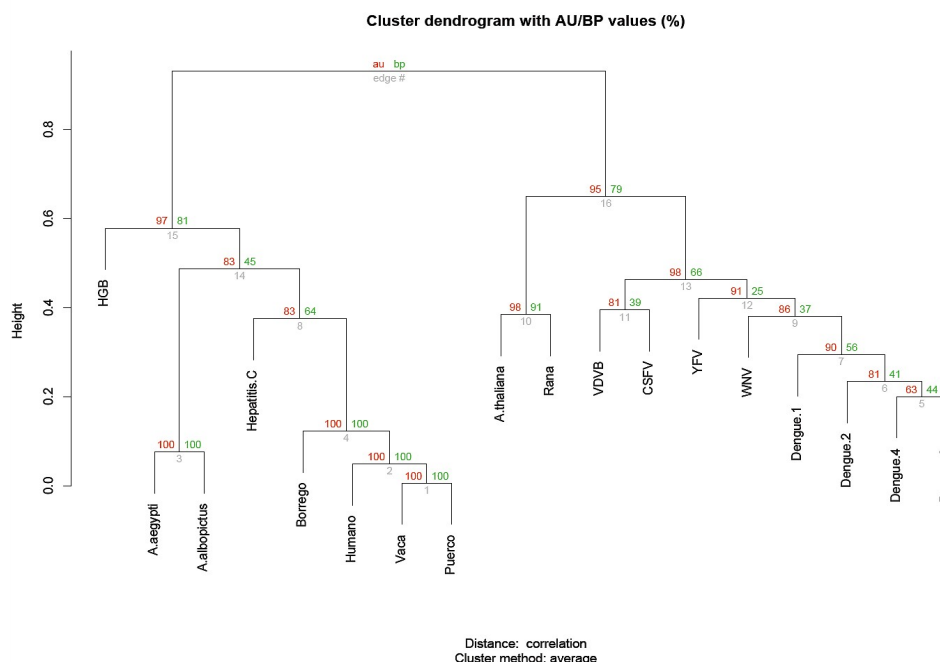


Figura 5. Dendrograma del agrupamiento jerárquico. Análisis de la similitud del uso de codón entre los virus y sus hospederos, mediante un agrupamiento jerárquico, el dendrograma fue realizado en R con el algoritmo pvclust, con 1000 réplicas de bootstrap. La escala de la izquierda representa la similitud entre los datos, este se observan dos grandes agrupamientos, el de los virus y sus hospederos.

Posteriormente, antes de realizar los análisis de selección natural se realizaron análisis de recombinación en las secuencias, ya que se sabe que este proceso sobreestima los codones que están evolucionado bajo selección natural, además se realizaron reconstrucciones filogenéticas que fueron utilizadas para la realizar un reconstrucción de caracteres ancestrales por codones, todos estos resultados se muestran en el complemento de resultados.

5.3 DETECCIÓN DE SELECCIÓN NATURAL.

La detección de la selección natural en el gen que codifica para la proteína de envoltura de los virus estudiados, se realizó mediante el método FEL (Pond *et al.*, 2005b) el cual es capaz de detectar la selección natural sitio por sitio; éste tipo de modelos pueden detectar la presencia de selección natural en algunos sitios específicos. En la tabla 9 se reportan los resultados obtenidos del análisis.

Para los tres virus se observaron muy pocas posiciones bajo selección natural positiva, así por ejemplo, se obtuvo que para los virus DENV y BVDV, solo un sitio bajo selección positiva y para el HCV dos sitios. La mayoría de los sitios se encontró que han evolucionado bajo selección negativa entre el 72.40 al 84.47 % dentro de esta proteína, así por ejemplo, el virus con mayor número de sitios bajo selección negativa fue el DENV, seguido del BVDV y por último el HCV. Finalmente, entre un 25 y a un 35 % de los sitios han evolucionado bajo neutralidad (Tabla 8). Estos resultados nos indican que la selección negativa juega un papel muy importante en la evolución de estos virus.

TABLA 8. EVOLUCIÓN EN LA PROTEÍNA DE ENVOLTURA No. DE CODONES QUE EVOLUCIONARON BAJO SELECCIÓN CON $P \leq 0.1$ POR MEDIO DEL MÉTODO FEL

Procesos	DENV	BVDV	HCV
----------	------	------	-----

Resultados

evolutivo						
S. Positiva	1	0.25 %	1	0.28 %	2	0.74 %
S. Negativa	286	72.40 %	240	69.56 %	228	84.47 %
Sin selección	108	27.34 %	104	30.14 %	39	14.49 %
Long. Total	395	100 %	345	100 %	269	100 %

Las posiciones mostradas en la tabla 8 se mapearon sobre la estructura de listones de la proteína de envoltura de cada uno de los virus, para el HCV se utilizó una predicción basados en la proteína de envoltura del BVDV. Estos resultados se muestran en la figura 6A para el DENV, se señala el codón 132, que codifica para una valina, evoluciona bajo selección positiva y se encuentra dentro del dominio I (marcado en rojo), además se observó una distribución uniforme de los sitios bajo selección negativa a lo largo de los tres dominios de la proteína. En la figura 6B se muestra la proteína del BVDV y en color rojo se muestran los 2 sitios que están bajo selección positiva, el codón 832 que codifica para una serina encontrado por IFEL y el codón 975 que codifica para una tirosina, encontrado por IFEL y éstos están en los dominios II y III de la proteína. También en este caso la mayoría de los sitios que están evolucionando bajo selección negativa se distribuyen en toda la proteína. Por último en la figura 6C, se observa el modelo de la proteína de envoltura del virus de hepatitis C, y se encuentran marcados en rojo los codones que evolucionan bajo selección positiva los codones 364 y 444, que codifican para una treonina y tirosina, respectivamente, ambos están dentro del dominio I de la proteína, y los sitios bajo selección negativa se encuentran en los tres dominios.

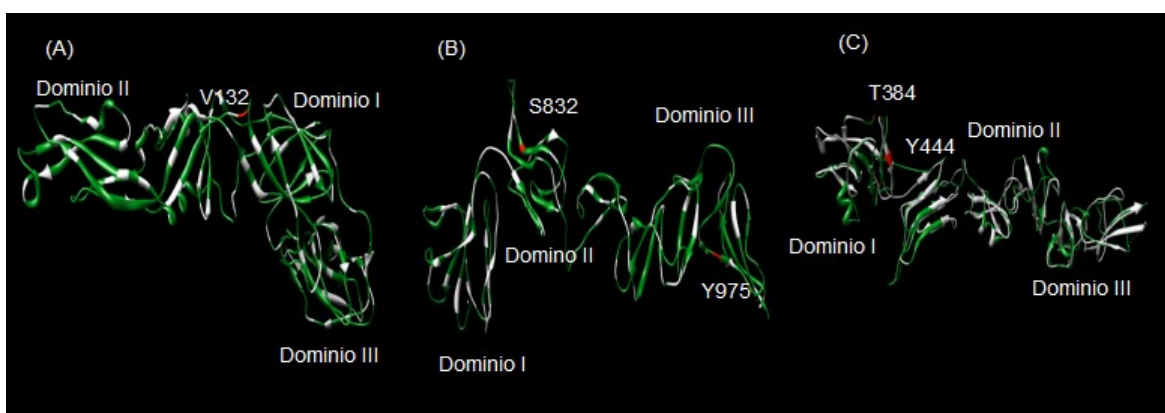


Figura 6. Distribución de los sitios bajo selección negativa y positiva. (A) Proteína de envoltura del DENV. (B) Proteína de envoltura del BVDV. (C) Proteína de envoltura del HCV. Los sitios que se encuentran marcados en color rojo está evolucionando bajo selección positiva y los que están en color verde están evolucionando bajo selección negativa.

Las pruebas de detección de sitios bajo selección natural (FEL e IFEL) identificaron pocos sitios bajo selección positiva, estos se observan en la tabla 9. Para DENV el codón 132, se encontró seleccionada positivamente por los métodos FEL e IFEL los valores de ω son muy altos, lo que indica que esta posición hay una gran cantidad de sustituciones no sinónimas. El análisis de predicción de epítopes encontró que ésta posición es un sitio de reconocimiento de linfocitos B y T (HLA-A1, HLA-B14, HLA-B3901, HLA-B40). Dado que es una región de reconocimiento de anticuerpos posiblemente éstos ejerzan cierta presión de selección que favorezca la presencia de alta variabilidad en esta posición.

TABLA 9. CODONES BAJO SELECCIÓN POSITIVA EN EL GEN E Y E2 USANDO LOS MÉTODOS FEL E IFEL.

Resultados

Codó n	DENV				Codó n	BVDV				Codó n	HCV			
	FEL		IFEL			FEL		IFEL			FEL		IFEL	
	P valu e	dN/ dS	P valu e	dN/ dS		P valu e	dN/ dS	P valu e	dN/ dS		P value	dN/dS	P valu e	dN/dS
132			0.01 04	2.3 x 10 ¹⁶	832			0.08 2	3.76 7	384	0.016 8	14.38 4		
132	0.01 4	200 . 035			975	0.02 95	4.53 8			444	0.000 49	3.387		

Un valor de $P < 0.1$ es estadísticamente significativo que muestra la evidencia de selección natural.

Por otro lado para el BVDV, se encontró el codón 975 bajo selección positiva con el método FEL con un $\omega = 4.538$ y el sitio 832 bajo selección positiva por método IFEL con $\omega = 3.767$ indicando más sustituciones no sinónimas que sustituciones sinónimas en estas posiciones (tabla 9). Se analizaron estas regiones para determinar si podría ser blanco del sistema inmune, sin embargo ninguna predicción arrojó información que apoyara esta hipótesis.

Por último para el HCV, se obtuvieron dos codones bajo selección positiva, el codón 364 $\omega = 14.38$ y la posición Y444 con $\omega = 3.38$ (tabla 9). El primer codón tiene un dN/dS muy alto posiblemente debido a que se encuentra en la región HRV1 que es una zona importante de reconocimiento de anticuerpos neutralizantes. El codón 444 está en una zona de reconocimiento de linfocitos B y hay linfocitos T (HLA-B5801 y MCH-Lid), por lo que esta posición puede estar sujeta a selección por parte del hospedero.

5.3.1 *Evolutionary Fingerprint.*

También como parte de los análisis realizados para la búsqueda de selección y con el objetivo de comparar los procesos evolutivos de las tres proteínas de estudio se realizó el

análisis llamado ***Evolutinary Fingerprint***. Los resultados (figura 7), muestran que para los tres virus los valores de ω indican que la mayoría de los codones o sitios de la proteína de envoltura evolucionan bajo el régimen de selección negativa. Esto confirma los resultados obtenidos por el método FEL, donde encontramos que la mayoría de los codones en las proteínas de envoltura evolucionan bajo selección negativa. También se pudo observar que la distribución de las categorías de ω entre los virus DENV y BVDV es similar con algunos sitios evolucionando cerca de la neutralidad mientras que la distribución de ω del HCV no es muy homogénea. Lo que indica que los valores ω son más diferentes entre sí, dentro del virus que se agruparon en más categorías.

Una característica importante que se observó en los gráficos de *evolutionary fingerprint* fue que las tasas de sustitución sinónimas (α), a lo largo de todos los sitios, son poco variables, especialmente en los casos de DENV y BVDV. Por otro lado, las tasas de sustitución no sinónima (β) son más variables pero en general tienen valores menores que los de α . Esto indica que los genes que codifican para las proteínas de envoltura evolucionan bajo el régimen de selección natural negativa

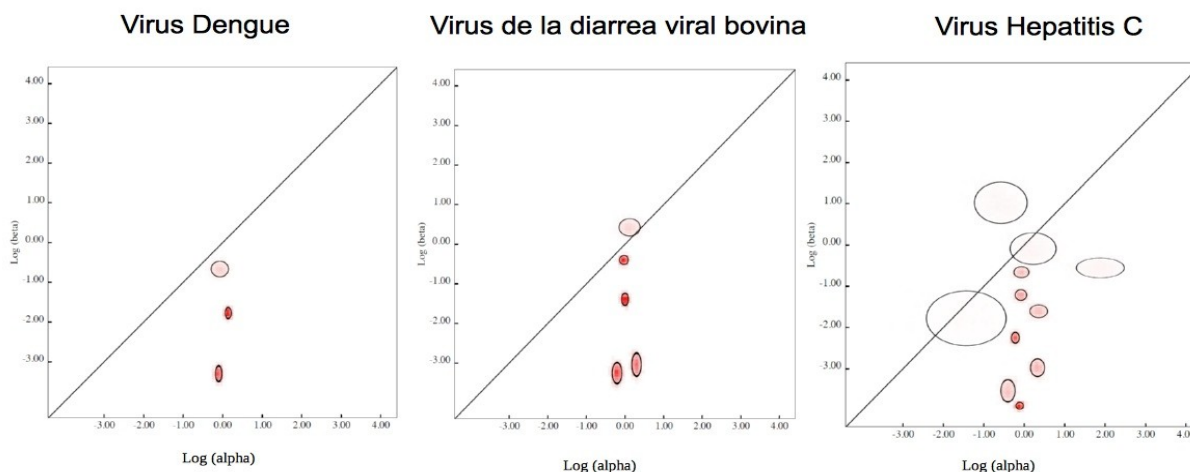


Figura 7. Evolutionary Fingerprint. Comparación de las tasas dN/dS en los gráficos se muestra la distribución estimada de los valores de ω (dN/dS), la diagonal indica el estado de neutralidad, por encima de esta indica selección positiva y por debajo selección negativa. Lado izquierdo el *fingerprint* del DENV. Al centro el *fingerprint* del BVDV y lado derecho el *fingerprint* del virus de HCV.

5.3.2 Reconstrucción de caracteres ancestrales.

Las reconstrucciones de caracteres ancestrales para la mayoría de las posiciones que resultaron estar bajo selección positiva o negativa (ver apéndice electrónico), esto con el objetivo de observar cómo fue la dinámica evolutiva de un codón en particular, además permitiría observar cómo han sido los cambios de carácter de estados en los codones que evolucionan bajo selección negativa.

En la figura 8 se aprecia la reconstrucción de caracteres ancestrales del aminoácido V132 de la proteína de envoltura del DENV, que se encontró bajo selección natural positiva. El aminoácido ancestral reconstruido fue la valina (V), y se observaron 10 cambios hacia otros aminoácidos a lo largo de la filogenia. El número de cambios no sinónimos observados es grande, por lo que se puede inferir que esta posición ha evolucionado bajo selección positiva. Así por ejemplo, en los serotipos DENV 1 y 3 esta posición evolucionó hacia tirosina (Y) e histidina (H) respectivamente, mientras que DENV 2 presentó un cambio a prolina (P) y DENV 4 cambio a isoleucina (I).

Entonces se pudo observar que los cuatro serotipos del DENV, presentaron cambios hacia aminoácidos diferentes, también parecieron cambios homoplásicos y apomorfícos. Este último tipo de cambios aislados en algunas ramas del árbol podrían ser eliminados por

medio de la selección negativa, además estos fueron cambios recientes que se encuentran en las puntas del árbol.

El aminoácido V132, no he repordo previamente bajo este tipo de selección, sin embargo entre las posiciones 127 a la 134 han sido reportadas como un péptido de reconocimiento de anticuerpos en pacientes convalecientes de una infección con el serotipo de DENV 2, además las posición entre 121 al 140 han sido reportados, como zona fuertemente reactiva para linfocitos B a pH ácidos, razones por las cuales está bajo fuerte selección natural (Twiddy *et al.*, 2002).

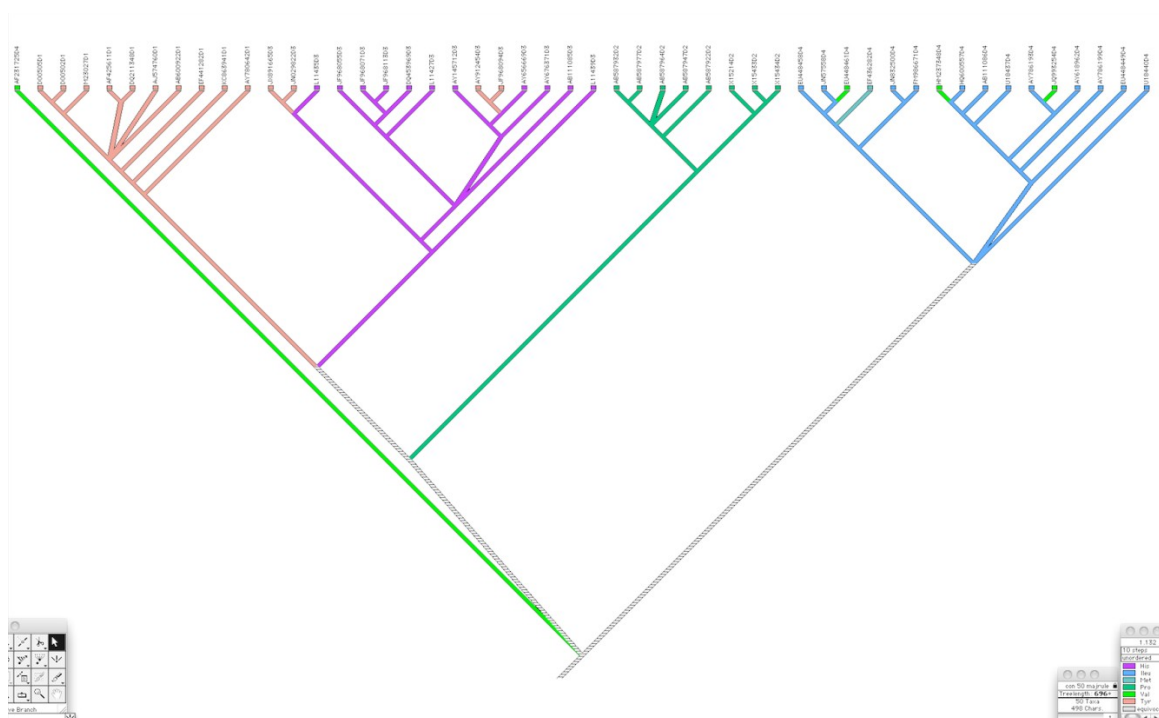


Figura 8. Reconstrucción de caracteres ancestrales de la posición V132 de la proteína de envoltura del DENV. El carácter ancestral valina se muestra en color verde, DENV1 presentó una tirosina en color salmón, DENV3 presentó una histidina color magenta en 4 puntas cambio hacia una tirosina, DENV2 presentó una prolina en color verde y DENV4 una isoleucina y en tres puntas cambio hacia valina.

Al igual que en el caso del DENV, se realizó una reconstrucción de caracteres ancestrales para los sitios bajo selección en el BVDV. En la figura 9 se muestra la reconstrucción para el aminoácido Y975, donde se encontró que el carácter ancestral fue una tirosina (Y), y que ocurrieron 21 cambios no sinónimos a lo largo de la filogenia. En el

genotipo 1 de este virus se observa que en algunos casos se conserva el aminoácido ancestral y en otros que ocurrieron cambios hacia histidina (H). Mientras que el genotipo II presentó una mayor cantidad de cambios hacia diversos aminoácidos, muchos de los cuales ocurrieron en las ramas terminales indicando que son cambios recientes. Los cambios observados en el genotipo II fueron del ancestral hacia serina (S), leucina (L), isoleucina (I), histidina (H), alanina (A), ácido aspártico (D), y asparagina (N).

El aminoácido Y975 se encuentra en el extremo carboxilo terminal y se ha reportado por Risatti *et al.*, (2007) este dominio carboxilo de proteína E2, está relacionado con la virulencia.

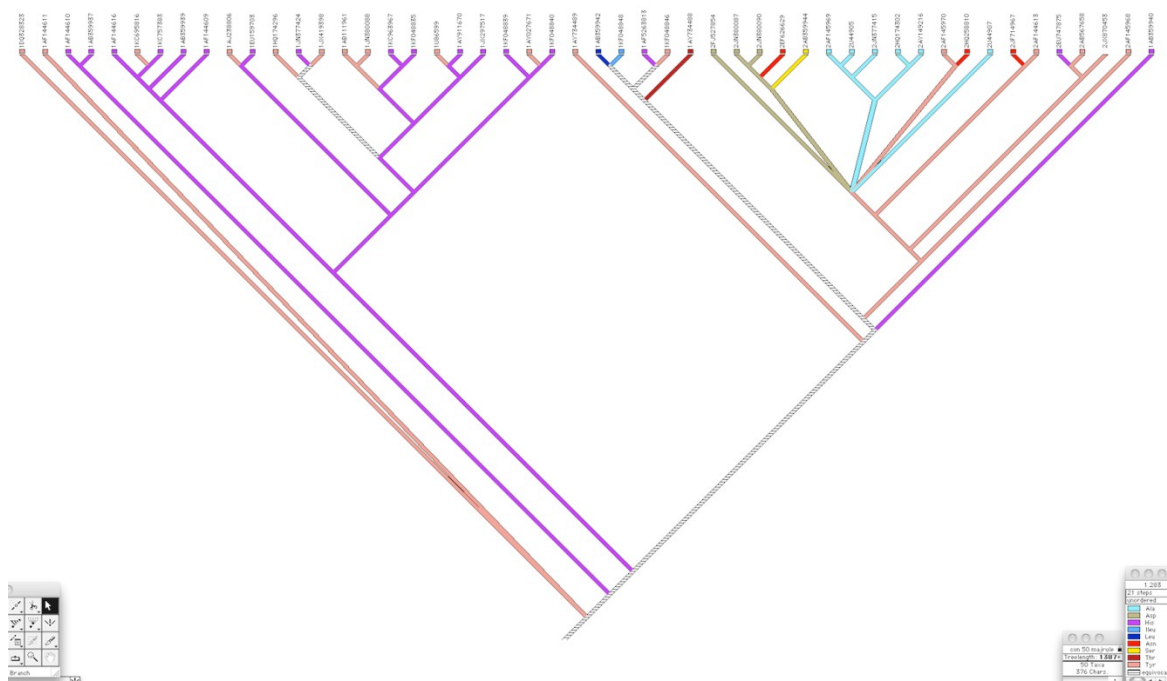


Figura 9. Reconstrucción de caracteres ancestrales de la posición Y975 de la proteína de envoltura del BVDV. El carácter ancestral fue una tirosina en color salmón, cambio a histidina en color lila para el genotipo I del BVDV. Para el genotipo dos se observaron cambios del ancestral hacia alanina color cian, ácido aspártico color verde olivo, serina en color amarillo, asparagina en color rojo, histidina en color lila, isoleucina en color azul y leucina en color azul marino.

La otra posición encontrada bajo selección positiva por el método IFEL fue el aminoácido S832, a la cual también se le realizó la reconstrucción de caracteres ancestrales.

En esta posición el codón ancestral fue la serina (S) se observaron 6 cambios hacia aminoácidos no sinónimos. Concretamente los cambios fueron del ancestral hacia una asparagina (N), en el genotipo 1, nuevamente el genotipo II presentó más cambios del aminoácido ancestral hacia alanina (A), lisina (L), glutamina (Q), cambios relativamente recientes y el cambio hacia ácido glutámico (D), el cual no es muy reciente.

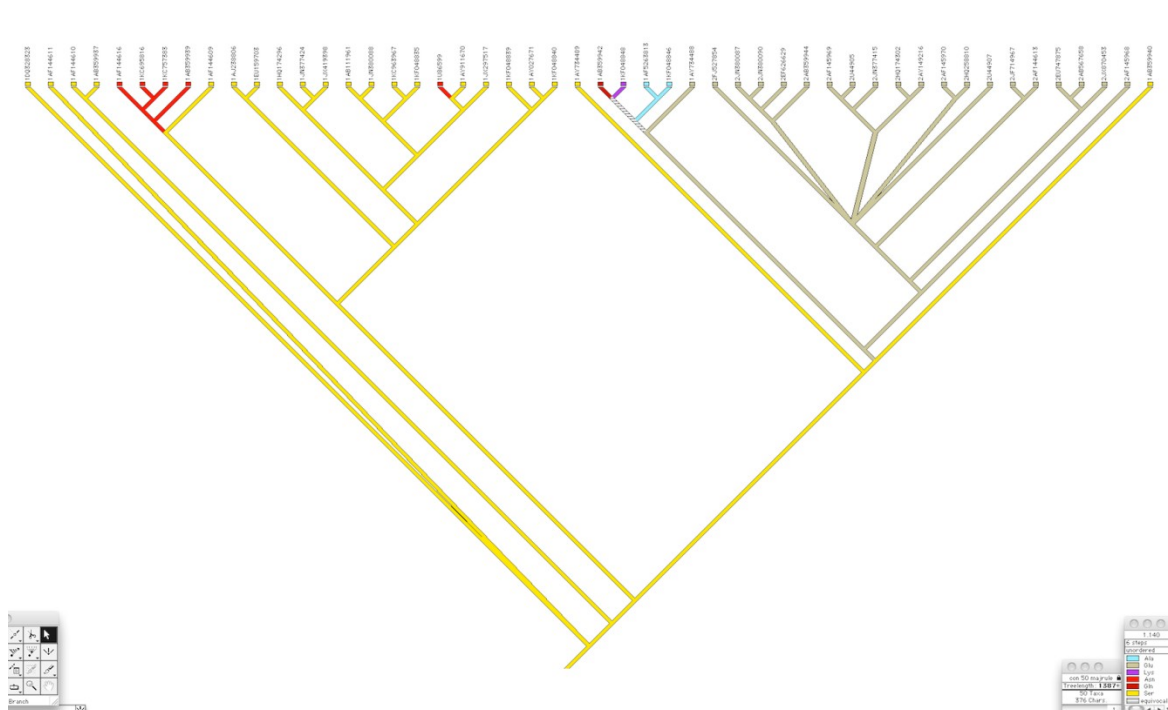


Figura 10. Reconstrucción de caracteres ancestrales de la posición S832 de la proteína de envoltura del BVDV. El carácter ancestral fue la serina, dentro del genotipo I se presentaron cambios hacia ácido aspártico en rojo. Par el genotipo II los cambios fueron del carácter ancestral hacia una glutamina en color rojo, ácido glutámico en color verde olivo), alanina en color cian.

Respecto al HCV se encontraron dos puntos bajo selección, estos fueron los aminoácidos T384 y Y444. Estos fueron altamente variables, cuando se realizó una reconstrucción de caracteres ancestrales para el HCV, los sitios bajo selección positiva mostraron una gran variedad de cambios a lo largo de la reconstrucción, por ejemplo en la posición T384 (Figura 11) se encontraron 31 cambios no sinónimos, el estado de carácter ancestral fue una asparagina, se observa que algunos de los cambios aparecen una solo ocasión y son recientes.

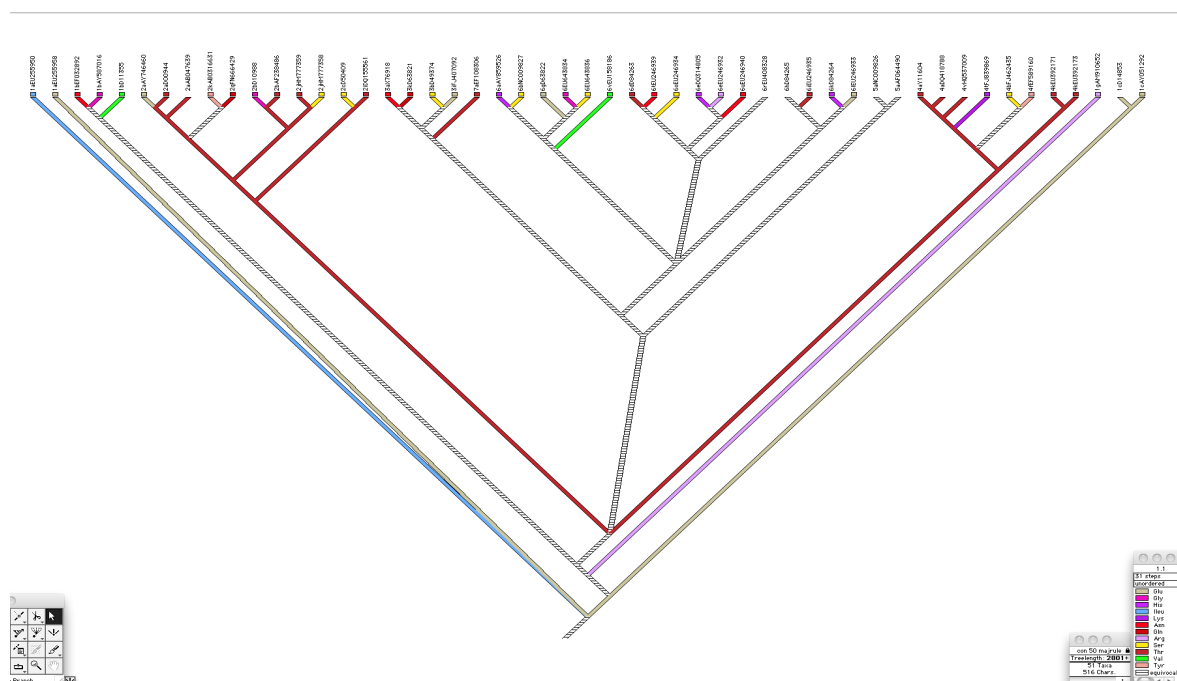


Figura 11. Reconstrucción de caracteres ancestrales del aminoácido T384 de la proteína de envoltura del HCV. El carácter ancestral asparagina en color rojo, que cambio hacia una glicina en color magenta, ácido glutámico en color dorado, histidina en color morado, isoleucina en color azul, lisina en color púrpura, asparagina en color rojo, glutamina en color rojo oscuro, arginina en color lila, serina en color amarillo, treonina en color rojo granate, valina en color verde, y tirosina en color rosa claro.

La gran variedad de cambios presentes a lo largo de la historia filogenética en estas dos posiciones T384 y Y444 podría indicar las fuertes presiones de selección a las que está sometido este virus, que permiten la fijación de estas mutaciones no sinónimas que posiblemente sean ventajosas para la eficiente evasión del sistema inmune, ya que las posiciones encontradas bajo selección son sitios de reconocimiento de anticuerpos.

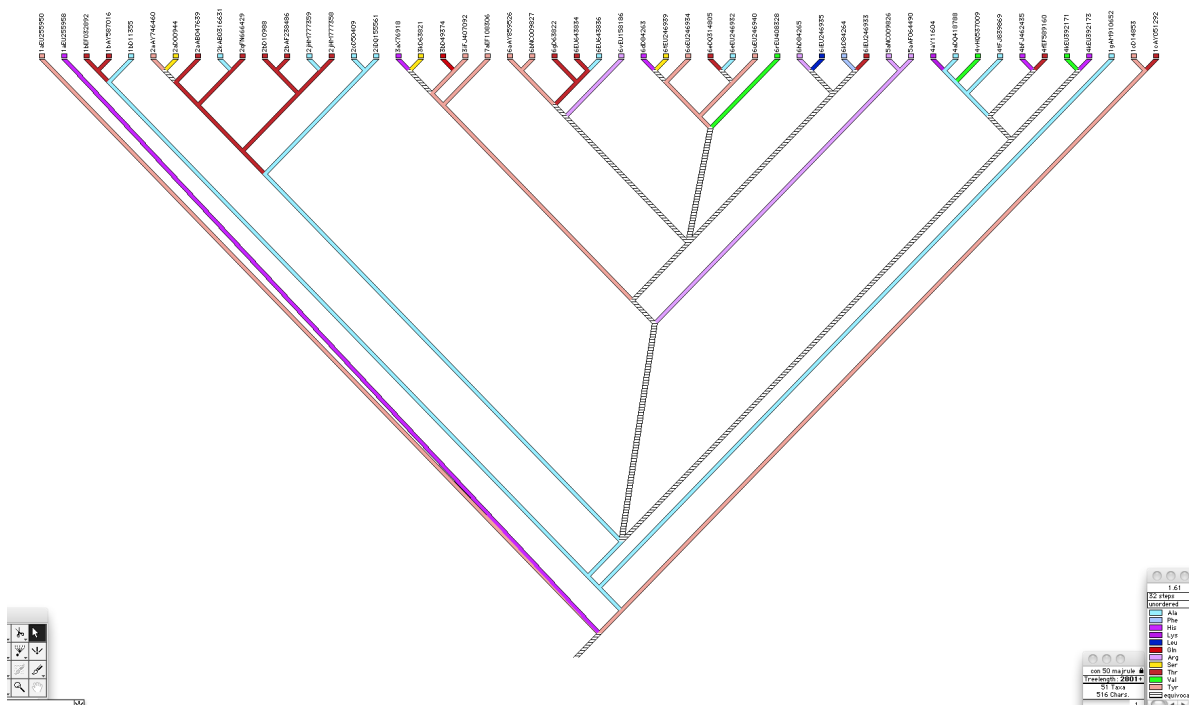


Figura 12. Reconstrucción de caracteres ancestrales de la posición Y444 de la proteína de envoltura del HCV. El carácter ancestral tirosina en color rosa claro, que cambio hacia alanina en color cian, fenilalanina en color azul, histidina en color morado, lisina en color purpura, leucina en color azul oscuro, glutamina en color rojo oscuro, treonina en color rojo granate, arginina en color lila, serina en color amarillo, y valina en color rosa.

De igual forma la reconstrucción de caracteres ancestrales para la posición Y444 (Figura 12) se observaron 32 cambios no sinónimos donde el carácter ancestral fue una tirosina.

Para el HCV en las reconstrucciones de los codones se observó muchos cambios no sinónimos, muchos de los cambios son recientes, los cuales puede ser eliminados por selección negativa.

En general se observaron una mayor cantidad de cambios, en la proteína de envoltura del HCV, comparada con las otras dos proteínas anteriores.

Por otra parte también se realizaron reconstrucciones para las posiciones bajo selección negativa donde se observó que en algunos codones no existen cambios dentro de la reconstrucción, sin embargo en otros codones bajo selección negativa mostraron cambios no sinónimos, resultando algo contradictorio, ya la selección negativa tiende a la conservación (ver figuras del apéndice electrónico). Por ejemplo el aminoácido V114 dentro

de la proteína de envoltura del DENV reportado que evoluciona bajo selección negativa, sin embargo se observó con la reconstrucción que este sitio era variable (Figura 13), aquí observamos que el carácter ancestral fue una valina en color verde, sin embargo en el clado correspondiente al serotipo DENV1 se observaron cambios hacia una leucina e Isoleucina. La presencia de estos cambios no sinónimos en algunas ramas del árbol podrían resultado de la selección episódica actuando en algunas ramas del árbol. Este interesante fenómeno también se pudo observar en las otras proteínas de envoltura de los virus BVDV y HCV.

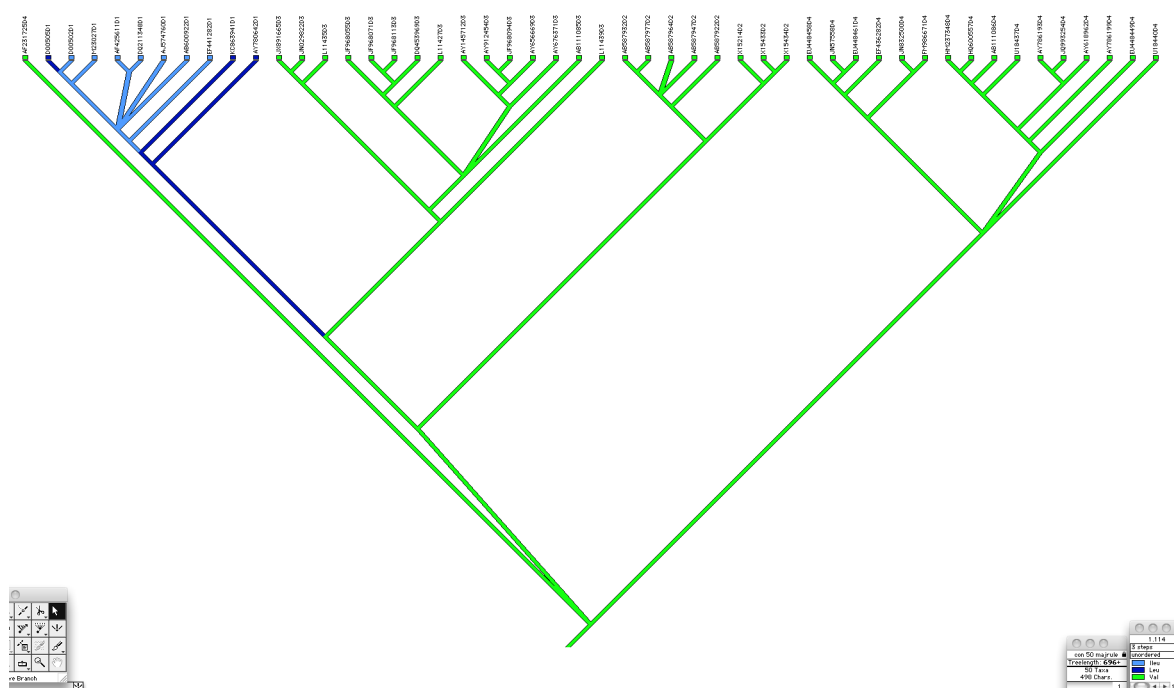


Figura 13. Reconstrucción de caracteres ancestrales aminoácido V114. Proteína de envoltura del DENV, en color verde valina codón ancestral, en color azul marino leucina, y en color azul la isoleucina.

5.3.3 Selección Episódica.

Los resultados anteriores ponen en evidencia que la selección natural negativa juega un papel importante en la evolución en la familia *Flaviviridae*, sin embargo, en los análisis de variabilidad en la secuencia de aminoácidos en las proteínas de envoltura, encontramos que la mayoría de los sitios presentaban polimorfismos. Aunado a esto, las reconstrucciones de

caracteres ancestrales mostraron que algunas posiciones identificadas que evolucionan bajo selección negativa, presentando una acumulación de mutaciones sinónimas y no sinónimas. Esta aparente contradicción entre la variabilidad y la prevalencia de selección negativa encontrada en estos genes podría ser explicada por eventos de selección episódica.

La selección positiva episódica es aquella que actúa en ciertos momentos de la historia evolutiva de un gen. Los métodos sitio por sitio como FEL, asumen que los valores de dN/dS son constantes a lo largo de toda la filogenia, por lo que no son capaces de detectar eventos de selección episódica. Dado el patrón de variabilidad que mostraban muchas posiciones de los genes de estudio, en las que se acumulaban un gran número de mutaciones sinónimas y no sinónimas, quisimos probar la hipótesis de que este patrón era el resultado de eventos de selección positiva episódica y selección negativa. Por lo tanto, se realizó un análisis de selección episódica utilizando el programa MEME (***Mixed Effects Model of Evolution***). En el caso del DENV se obtuvieron 19 sitios con evidencia de selección episódica, de los cuales 14 se habían detectado bajo selección negativa, para el BVDV se obtuvieron 30 codones bajo selección episódica de los cuales 11 codones estaban bajo selección negativa, y para HCV se obtuvieron 20 codones de los cuales 15 estaban bajo selección negativa (tablas 10 y 11).

TABLA 10. CODONES QUE EVOLUCIONARON BAJO SELECCIÓN EPISÓDICA POSITIVA CON $P \leq 0.1$ EN LAS PROTEÍNAS DE ENVOLTURA.

	DENV	BVDV	HCV
	21, 23, 24, 25, 26, 27, 30, 114, 132, 164, 195, 205, 213, 217, 229, 235, 300, 301, 308	712, 716, 721, 723, 724, 734, 750, 760, 770, 832, 835, 844, 853, 854, 861, 862, 876, 887, 888, 908, 908, 925, 930, 938, 942, 945, 946, 948, 956, 974	387, 401, 412, 419, 432, 437, 444, 449, 475, 517, 527, 529, 536, 545, 547, 551, 562, 575, 578, 586
No. total de codones	19	30	20

TABLA 11. SITIOS BAJO SELECCIÓN EPISÓDICA

	DENV		BVDV		HCV	
	No.	S. Episódica	No.	S. Episódica	No.	S. Episódica
Selección positiva	1	1	1	1	2	1
Selección negativa	286	13	240	11	228	15
Sin selección	108	4	104	18	39	5

Posteriormente, se llevó a cabo una reconstrucción de codones ancestrales para los sitios que se detectaron que evolucionan bajo selección episódica. En los resultados se observa que el uso de más de un codón sinónimo lo que explica una tasa alta de mutaciones sinónimas, además de la existencia de cambios no sinónimos que ocurren en las ramas que se encuentran bajo selección episódica. Por ejemplo, el codón 114 que codifica para una valina en la proteína de envoltura del DENV (figura 21), en el que se detectó selección negativa con el método FEL, presenta alternativamente los codones GTA, GTG (Valina, V), el codón CTA (leucina, L) y el codón ATA (Isoleucina, I), con el método MEME, éste codón

se encontró que evoluciona bajo selección positiva, concretamente en la rama que está bajo selección positiva ocurre el cambio del codón GTA que codifica para la V hacia el codón CTA que codifica para una leucina, ocurriendo una mutación en la primera posición.

Estos tipos de mutaciones a lo largo de la historia filogenética no fueron encontrado por el método FEL, dado que éste contabiliza el número de cambios a largo de la filogenia, y como se puede observar en la figura 14, en la mayoría de las ramas las mutaciones que ocurren son sinónimas, por lo tanto el algoritmo considera que esta posición evoluciona bajo selección negativa, cuando realmente los procesos evolutivos que ocurrieron en este codón no fueron uniformes en el tiempo.

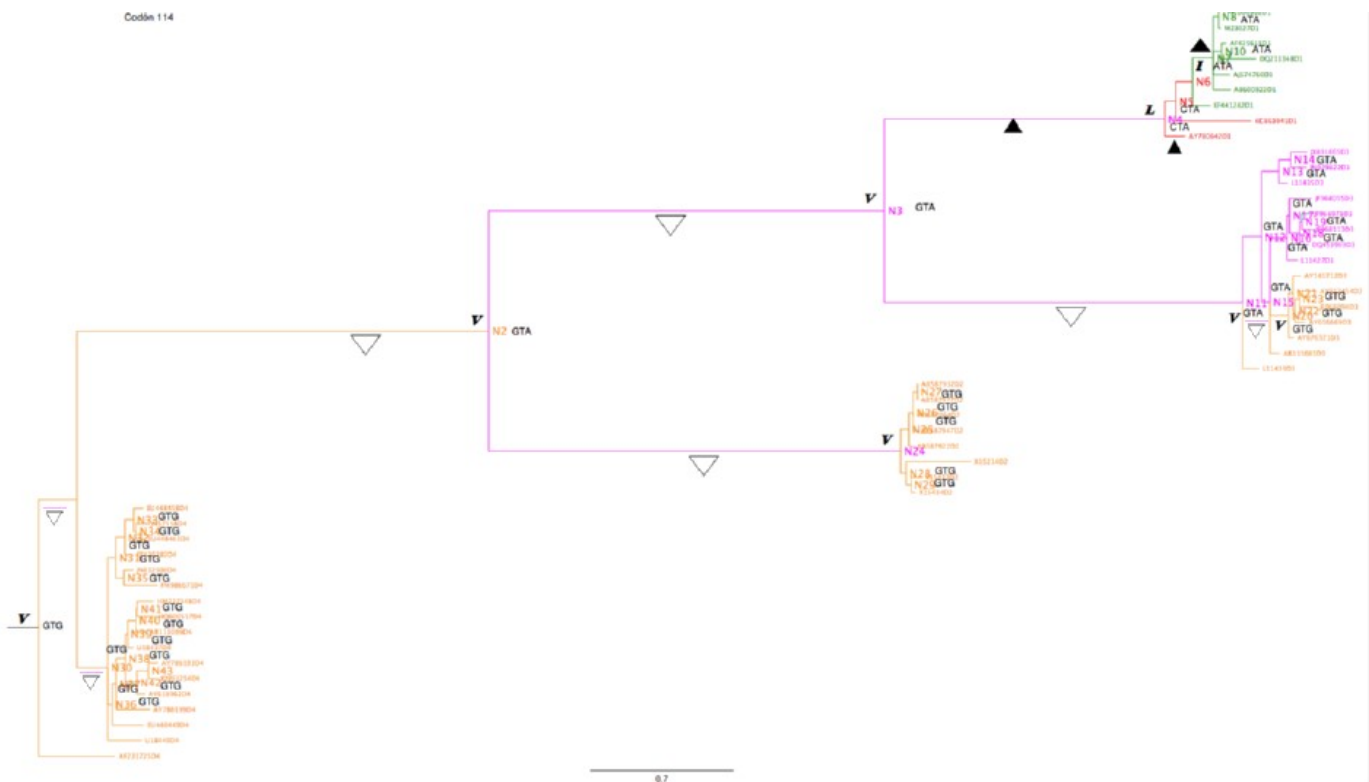


Figura 14. Reconstrucción de caracteres ancestrales del codón 114 que codifica para una Valina en la proteína E del DENV. La punta de flecha invertida indica que las ramas están bajo selección negativa y las puntas de flecha hacia arriba indican que evolucionan bajo selección positiva. El codón GTG que codifica para V codón ancestral las ramas de color anaranjado y el codón GTA que codifica para el mismo aminoácido en color rosa, mientras que las ramas en color rojo y verde son los cambios hacia Leucina e Isoleucina respectivamente.

En los casos de los virus BVDV y HCV ocurren eventos similares. Sólo por mencionar un ejemplo el codón 732 del BVDV (figura 15), que se encontró bajo selección negativa, es polimórfico y presenta los siguientes estados de carácter: el codón ancestral AGA que codifica para una Treonina (T), y dos codones sinónimos para este aminoácido ACC y ACG, también se observó otro codón no sinónimo TCC que codifica para una serina, él cual a su vez presentó otro codón sinónimo que fue el TCT, entonces hay acumulación de mutaciones sinónimos y no sinónimas.

En este caso el codón ancestral codifica para el aminoácido T, se encontraron cuatro ramas bajo selección positiva que corresponden a los momentos en los que ocurrieron los cambios hacia codones no sinónimos. La razón por la que el método FEL detecta selección negativa en ésta posición es porque en la mayoría de las ramas las mutaciones resultaron en codones sinónimos.

Para el HCV, por ejemplo del codón 386 (figura 16) que se encuentra bajo selección negativa, presentó el codón ancestral CAC que codifica para una histidina (H), y se encontraron una amplia variedad de codones no sinónimos por ejemplo el codón CGC que codifica para una arginina (R), el codón CAG que codifica para una glicina (G), el codón AAC que codifica para una treonina (T), y los codones TAC y TAT que codifican para una tirosina (Y). En las ramas donde presentaron cambios hacia codones no sinónimos resultaron estar bajo selección positiva, en general se encontraron 6 ramas en presencia de selección episódica.

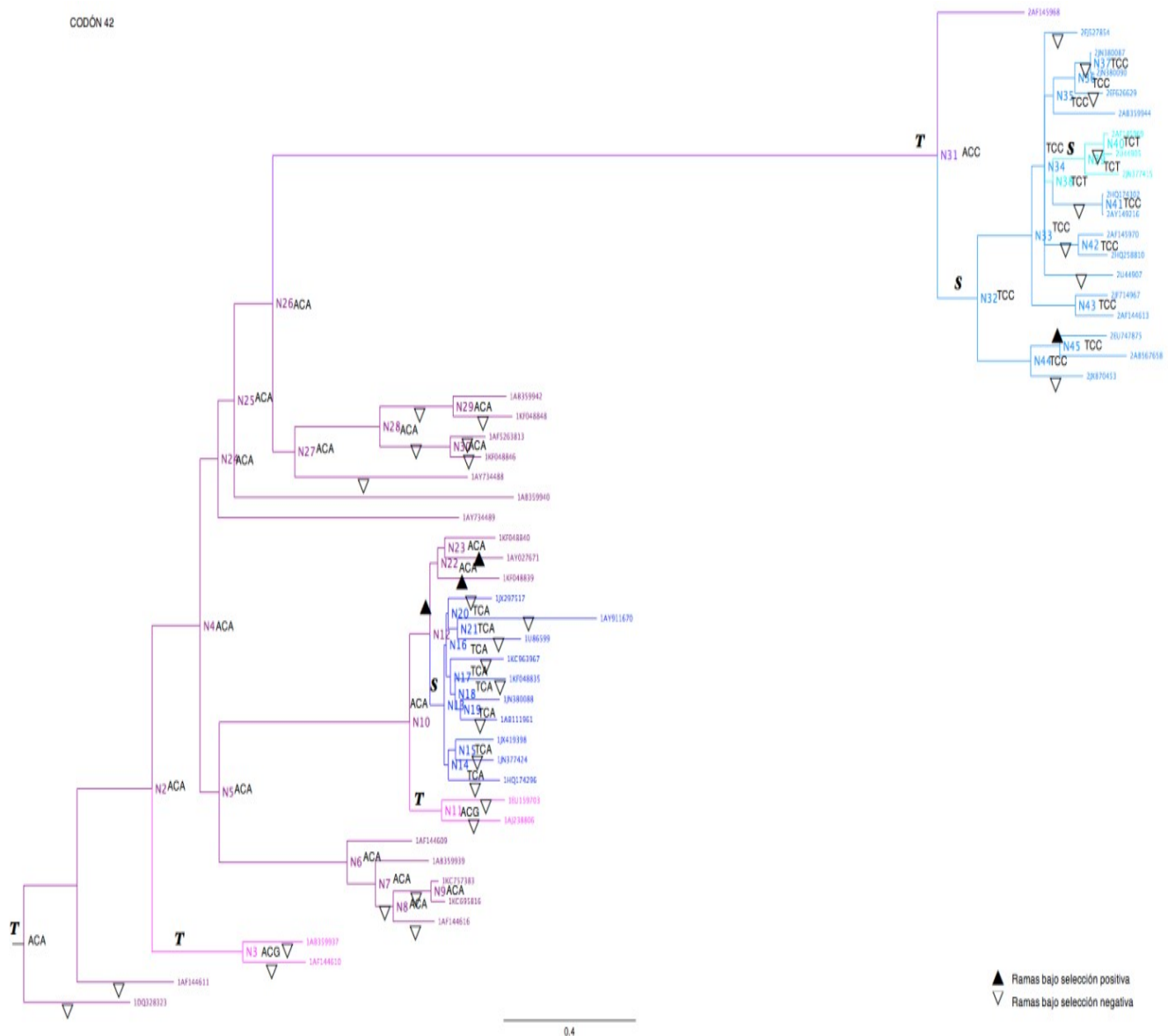


Figura 15. Reconstrucción de caracteres ancestrales del codón 732 que codifica para una Treonia dentro de la proteína de envoltura del BVDV. La punta de flecha invertida indica que las ramas están bajo selección negativa y las puntas de flecha indican que evolucionan bajo selección positiva. Los codones AGC, ACA, ACC que codifican para una T en tonalidades rosas, el codón TCC, TCA que codifica para una serina en tonalidades azules. Los cambios no sinónimos en las puntas del árbol.

Phylogenetic tree of the H gene under positive and negative selection. The tree shows various amino acid positions and their corresponding codons. Branches are color-coded: purple for positive selection and green for negative selection. A scale bar of 0.4 is shown at the bottom.

Legend:

- ▲ Rama bajo selección positiva
- ▽ Rama bajo selección negativa

62

Aunque relativamente pocos sitios se encontraron que evoluciona bajo selección positiva, la presencia de variabilidad y la presencia de mutaciones no sinónimas en algunas ramas de las filogenias fue un evento recurrente, lo que indica posiblemente que en la proteína de envoltura de estos virus, la selección episódica juega un papel importante en la fijación de ciertas variantes no sinónimas.

Con el objetivo de establecer el número de sitios que evolucionaron exclusivamente bajo selección negativa se contabilizaron, por un lado aquellos que no presentaban variabilidad a nivel de aminoácidos y por otro los sitios variables dentro de la proteína de envoltura, los resultados obtenidos se muestran en la tabla 12. Es importante resaltar que las posiciones que aparentemente evolucionaron bajo selección negativa pero que en realidad presentaban cierta variabilidad en la secuencia de aminoácidos fueron más predominantes, que aquellas que conservaron su carácter de estado.

Posteriormente se contabilizaron estos mismos sitios pero por dominio (Tabla 13), observando en el caso del DENV, la mayoría de los sitios variables bajo selección negativa se encontraron en el dominio II, para el BVDV la mayoría de estos sitios se encontraron en el dominio III y para el HCV en el dominio I. Por lo tanto existen diferencias en la fijación de mutaciones no sinónimas de cada uno de los dominios de estos virus, estas diferencias podrían deberse posiblemente que éstos están más accesibles al reconocimiento de la respuesta inmune.

TABLA 12. SITIOS NEGATIVOS VARIABLES Y CONSTANTES EN LA PROTEÍNA DE ENVOLTURA LOS VIRUS DE LA FAMILIA *FLAVIVIRIDAE*

Sitios	DENV	BVDV	HCV
S. Negativa sitios variables	146	154	154
S. Negativa sitios constantes	140	86	74
Total de sitios	286	240	228

TABLA 13. SITIOS NEGATIVOS VARIABLES Y CONSTANTES BAJO SELECCIÓN NEGATIVA EN LA PROTEÍNA DE ENVOLTURA DE LOS VIRUS DE LA FAMILIA *FLAVIVIRIDAE*

Sitios	DENV			BVDV			HCV		
	Dominios			Dominios			Dominios		
	I	II	III	I	II	III	I	II	III
S. Negativa sitios variables	56	62	37	37	36	82	59	55	40
S. Negativa sitios constantes	43	62	26	15	22	48	27	17	30
Total	286			240			228		

Para evaluar si los sitios bajo selección son accesibles al reconocimiento de la respuesta inmune o están expuestos en la superficie de la proteína, se mapearon sobre la estructura terciaria de la proteína. Este análisis sólo se realizó para las proteína de envoltura de los DENV y BVDV, ya que son las únicas que se encuentran cristalizadas. Los resultados obtenidos de este análisis se aprecian en la figura 17A y 17B. Los sitios bajo selección negativa, tanto variables como conservados, se encontraron sobre la superficie de la proteína de envoltura. Cabe destacar que algunas de las posiciones conservadas son estructuralmente importantes, concretamente en el caso de la proteína de envoltura del DENV, algunos de los sitios que resultaron bajo selección negativa constantes fueron las cisteínas que forman puentes disulfuro estos fueron los siguientes Cys3-30, Cys 60-121, Cys 74-105, Cys 92-116, Cys 185-286, Cys 305-336. También las posiciones de la 98 a la 111 son sitios bajo selección negativa constantes, y corresponden al *loop* que da lugar a la fusión las membranas viral y celular. También se encontró que los sitios de N-glicosilación 67 y 153 están bajo selección negativa y se encuentran conservados. Por lo tanto los sitios seleccionados negativamente podrían ayudar a mantener la estructura de la proteína y/o ser funcionalmente importantes. Para el caso de la proteína de envoltura del BVDV, también se encontraron que los sitios bajo selección negativa pueden ser funcionalmente importantes,

ejemplo de ello fue la Histidina 763, que participa en los cambios conformacionales en la proteína que permiten la entrada del virus a la célula. Los sitios de N-glicosilación N880, 810, 924, 992 también son altamente conservados y están bajo selección negativa.

La proteína de envoltura del HCV, los sitios variables y constantes que evolucionan bajo selección negativa, cabe resaltar que algunos de sitios bajo selección negativa que encontramos son funcionalmente importantes y esta ha sido previamente reportado, por ejemplo, los sitios encontrados que evolucionan bajo este tipo de selección fueron: los puentes disulfuro Los puentes de Cys 429-557, Cys 452-459, Cys 503-508, Cys 596-619. El loop putativo de fusión 508,509, 513, 516-519. Sitios de unión al receptor 421-423, 425, 427-429, 433-435, 436. Los sitios de N-glicosilación 423,540.

Por lo tanto parte de la variabilidad observada es debido a la presencia de episodios de selección positiva.

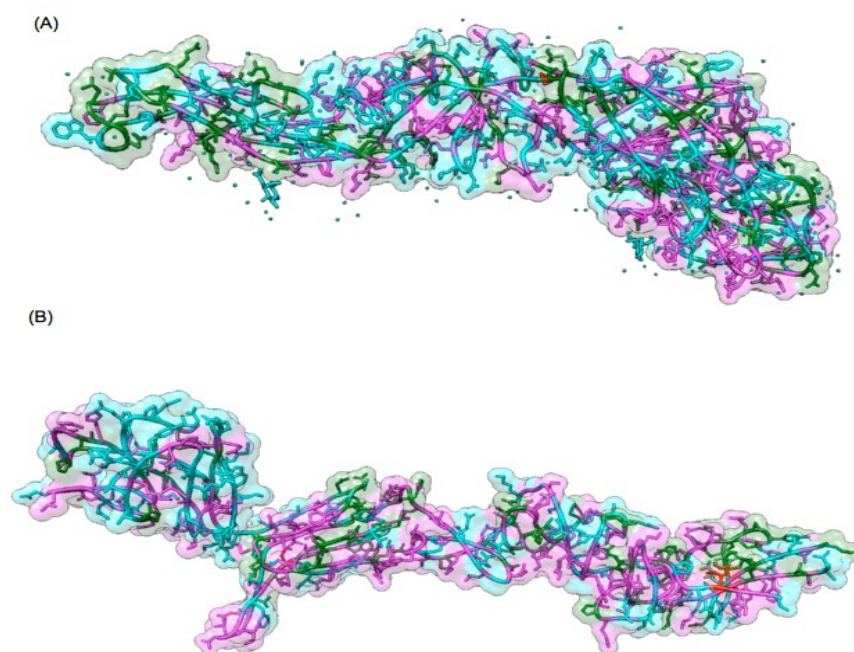


Figura 17. Sitios bajo selección positiva y negativa (variables y constantes). Las posiciones bajo selección positiva se mapearon en color rojo, mientras que los sitios bajo selección negativa que son constantes están en color verde, los sitios bajo selección negativa variables en color magenta, y los neutrales en color cian. (A) proteína de envoltura del DENV. (B) proteína de envoltura del virus de la diarrea viral bovina.

5.4 Análisis de las Propiedades Bioquímicas

Los resultados de la detección de la selección indicaron que la mayoría de los codones evolucionan bajo una selección negativa o purificadora que elimina mutaciones deletéreas, y favorece a la conservación de estado de carácter. Sin embargo, los resultados del análisis de variabilidad del objetivo 1, mostraron que la secuencia a nivel de proteína es altamente variable y posteriormente encontramos que estos polimorfismos se asocian la presencia de selección positiva episódica. Por lo tanto, quisimos investigar si esta variabilidad afectaba las propiedades bioquímicas de las proteínas. Para ello se llevó a cabo el análisis de cambios en las propiedades bioquímicas con la herramienta PRIME (Pond y Frost, 2005), como se indica en la sección de Métodos.

En el caso de la proteína de envoltura del DENV únicamente 5 aminoácidos cambiaron su polaridad (tabla 14), y sólo un aminoácido cambio la composición química, sin embargo la mayoría de los aminoácidos conservaron las propiedades bioquímicas. La propiedad bioquímica que se conservó mayor número de veces fue el punto isoeléctrico (tabla 14), seguido del volumen. Por lo tanto en esta proteína los cambios en los aminoácidos tienden a conservar las características bioquímicas y sólo en unos pocos casos resultan en un cambio en la polaridad. Notablemente, el aminoácido V132, que evoluciona bajo selección positiva, fue una posición que presentó cambios en la polaridad, encontrando así aminoácidos como H (Histidina), P (Prolina), Y (Tirosina) que son polares, y aminoácidos no polares en la posición (Isoelucina).

TABLA 14. CONSERVACIÓN Y CAMBIOS EN LAS PROPIEADES BIOQUIMICAS EN LA PROTEÍNA DE ENVOLTURA DEL DENV

Propiedad bioquímica	Aminoácidos que presentaron cambio en la propiedad bioquímica	No. de aminoácidos que presentaron cambios en la propiedad bioquímica	Aminoácidos que conservaron la propiedad bioquímica	No. de aminoácidos que conservaron la propiedad bioquímica
Polaridad	26, 132, 236, 274, 275	5	23, 24, 216	3
Hidropatía	0	0	26, 229, 274	3
Punto isoelectrico	0	0	26, 46, 155, 210, 274, 275, 300, 357, 380.	9
Composición química	83	1	26, 229, 275, 350	4
Volumen	83	1	51,58, 132	3

Para la proteína de envoltura del BVDV, 25 aminoácidos resultaron con cambios estadísticamente significativos en alguna propiedad bioquímica. Las propiedades bioquímicas polaridad e hidropatía tuvieron el mayor número de cambios, mientras que el punto isoelectrico y el volumen fueron las propiedades más conservadas (tabla 15). Uno de los aminoácidos que presentó cambio en la polaridad fue la posición Y975 que evoluciona bajo selección positiva, en esta posición se encuentran cuatro aminoácidos polares D, N, S, H y tres aminoácidos no polares A, I, L. Con respecto a la otra posición encontrada que evoluciona bajo selección positiva, la S832, no se encontró ningún cambio en las propiedades bioquímicas, aunque en esta posición existen muchos cambios de S por E, K, A, N, y el análisis de selección episódica presentó ramas bajo selección, sin embargo el método PRIME (Pond y Frost, 2005) no fue capaz de detectar cambios en las propiedades bioquímicas, esto indica posiblemente, que aunque el sitio estén seleccionados positivamente no producen cambios en la función de la proteína.

TABLA 15. CONSERVACIÓN Y CAMBIOS EN LAS PROPIEADES BIOQUIMICAS EN LA PROTEÍNA DE ENVOLTURA DEL BVDV

Propiedad bioquímica	Aminoácidos que presentaron cambio en la propiedad bioquímica	No. de aminoácidos que presentaron cambios en la propiedad bioquímica	Aminoácidos que conservaron la propiedad bioquímica	No. de aminoácidos que conservaron la propiedad bioquímica
Polaridad	737, 746, 800, 861, 872, 975, 1027	7	813, 836, 888, 955, 964, 744, 770, 804, 929	5
Hidropatía	748, 836, 888, 897, 938, 948, 961	7	716,770, 804, 926, 975	5
pl	732, 734, 735, 760, 938, 948	6	716, 724, 746, 782, 785, 795, 800, 802, 849, 887, 888, 931	12
Volumen	772, 825	2	725, 772, 825, 921, 935, 938, 946, 948, 975, 984	10
Composición química	944, 975	2	725, 778, 862, 938, 948, 1027	6

La proteína de envoltura del HCV, la polaridad y la composición química fueron las dos propiedades bioquímicas donde los aminoácidos presentaron más cambios, y donde los aminoácidos presentaron mayor conservación fue el volumen (tabla 16). Una posición que presentó cambios en las propiedades bioquímicas fue el aminoácido 348, que evoluciona bajo selección positiva (tabla 16), donde resulto con cambios estadísticamente significativos en las propiedades bioquímicas como punto isoeléctrico, polaridad, composición química, además el análisis de selección episódica indicó que muchas de las ramas están bajo selección positiva.

TABLA 16. CONSERVACIÓN Y CAMBIOS EN LAS PROPIEADES BIOQUIMICAS EN LA PROTEÍNA DE ENVOLTURA DEL HCV

Propiedad bioquímica	Aminoácidos que presentaron cambio en la propiedad bioquímica	No. de aminoácidos que presentaron cambios en la propiedad bioquímica	Aminoácidos que conservaron la propiedad bioquímica	No. de aminoácidos que conservaron la propiedad bioquímica
Polaridad	384, 386, 392, 453, 448, 529, 546, 589, 619, 589, 619, 638	10	394, 398, 402, 408, 446, 461, 457, 545	9
Hidropatía	394, 396, 402, 408, 446, 551	6	348, 385, 401, 404, 416, 419, 434, 439, 450, 457, 461, 465, 479, 480, 498, 517, 539, 562, 563, 589, 621, 636, 642, 648, 653	25
pl	384, 395, 404, 461, 562	5	387, 392, 396, 410, 412, 424, 440, 445, 456, 457, 529, 536, 546, 547, 558, 566, 571, 576, 596, 602, 619	21
Volumen	387, 396, 401	3	389, 392, 394, 395, 407, 408, 440, 443, 453, 456, 465, 477, 483, 487, 504, 520, 527, 529, 533, 587, 588, 619, 621, 633, 638, 653	26

Composición química	393, 397, 440, 443, 445, 446, 457, 461, 465, 527, 562, 587, 621, 649	14	384, 386, 387, 392, 401, 402, 405, 442, 444, 398, 529, 602, 619	13
---------------------	--	----	---	----

En resumen, en los resultados anteriores podemos observar que los aminoácidos de las tres proteínas de envoltura, existe una conservación en las siguientes propiedades bioquímicas: volumen, composición química y punto isoelectrico, mientras la polaridad, es la propiedad bioquímica menos conservada en las proteínas de envoltura. Esto nos indicaría que la selección natural como la variabilidad observada en las proteínas provoca un cambio en la polaridad de la proteína.

Sin embargo, posteriormente los cambios observados en algunas de estas propiedades bioquímicas fueron mapeados sobre la superficie de la proteína de los tres virus, con el objetivo de observar si estos cambios podrían repercutir en la función o estructura de la proteína y se comparó con las proteínas cristalizadas en las base de datos PDB. En esta parte se describirán lo observado en las cinco propiedades bioquímicas, para las tres proteínas.

En primer lugar se analizaron los cambios en la polaridad en los aminoácidos en las tres proteínas, anteriormente observamos que esta propiedad, fue la que menos conservaron los aminoácidos de las proteínas de envoltura. En la figura 18 se observan de arriba hacia abajo, las proteínas de envoltura del DENV, BVDV y HCV, en las proteínas de lado izquierdo se observan en color rojo los aminoácidos que presentaron cambios estadísticamente significativos, del lado derecho son las proteínas del PDB con los residuos con la cadena lateral cargada a pH neutro, es decir, en rojo se indica los residuos que poseen carga positiva, en azul los residuos con carga negativa, mientras los blancos son residuos sin carga o neutrales a pH 7. Los aminoácidos que resultaron con cambios significativos en esta propiedad, observamos que fueron cambios de aminoácidos polares cargados por aminoácidos polares sin carga, o de residuos no polares por residuos polares, esto sucedió en las tres proteínas. Por ejemplo, en la proteína de envoltura de DENV, 5 codones presentaron cambios estadísticamente significativos, estos fueron el residuo V132H, bajo selección positiva, que concretamente presentó cambios hacia un aminoácido

polar cargado (H); también los residuos E26S, D274Q, S275N, todos bajo selección negativa, presentaron cambios por aminoácidos polares y finalmente la posición Y236R que evoluciona bajo neutralidad, y también presentó cambio por un aminoácidos polar.

Para la proteína del BVDV, únicamente 6 residuos tienen cambios en las propiedades bioquímicas estadísticamente significativos, estos fueron Y975D bajo selección positiva, V737R, D800N, G861D, bajo selección negativa y T746R, L872 bajo neutralidad (este último no se muestra en la figura) (figura 18B).

Finalmente para la proteína del HCV, 10 residuos fueron reportados con cambios en la polaridad. En concreto los cambios encontrados fueron, las posiciones N384E bajo selección positiva, las posiciones H386Y, A392Y, P498E, A529N, N536S, L619M, F638H, R660Y, bajo selección negativa y la posición T589H bajo neutralidad. En general se observaron cambios de aminoácidos polares por aminoácidos polares con carga o viceversa, al igual que cambios por aminoácidos no polares por aminoácidos polares. Cabe resaltar de las posiciones N384E, H386Y, A392Y, se encuentran dentro la región hipervariable 1 (32 aminoácidos) de la proteína E2 de HCV, la cual se ha reportado ampliamente como responsable de evasión de la respuesta inmune; por lo tanto cambios en la polaridad en estas posiciones podrían estar implicadas con este posible mecanismo de evasión (figura 18C).

Por lo tanto en las tres proteínas observamos que los procesos evolutivos, como la selección natural positiva, episódica y procesos neutrales producen cambios en la polaridad de los aminoácidos.

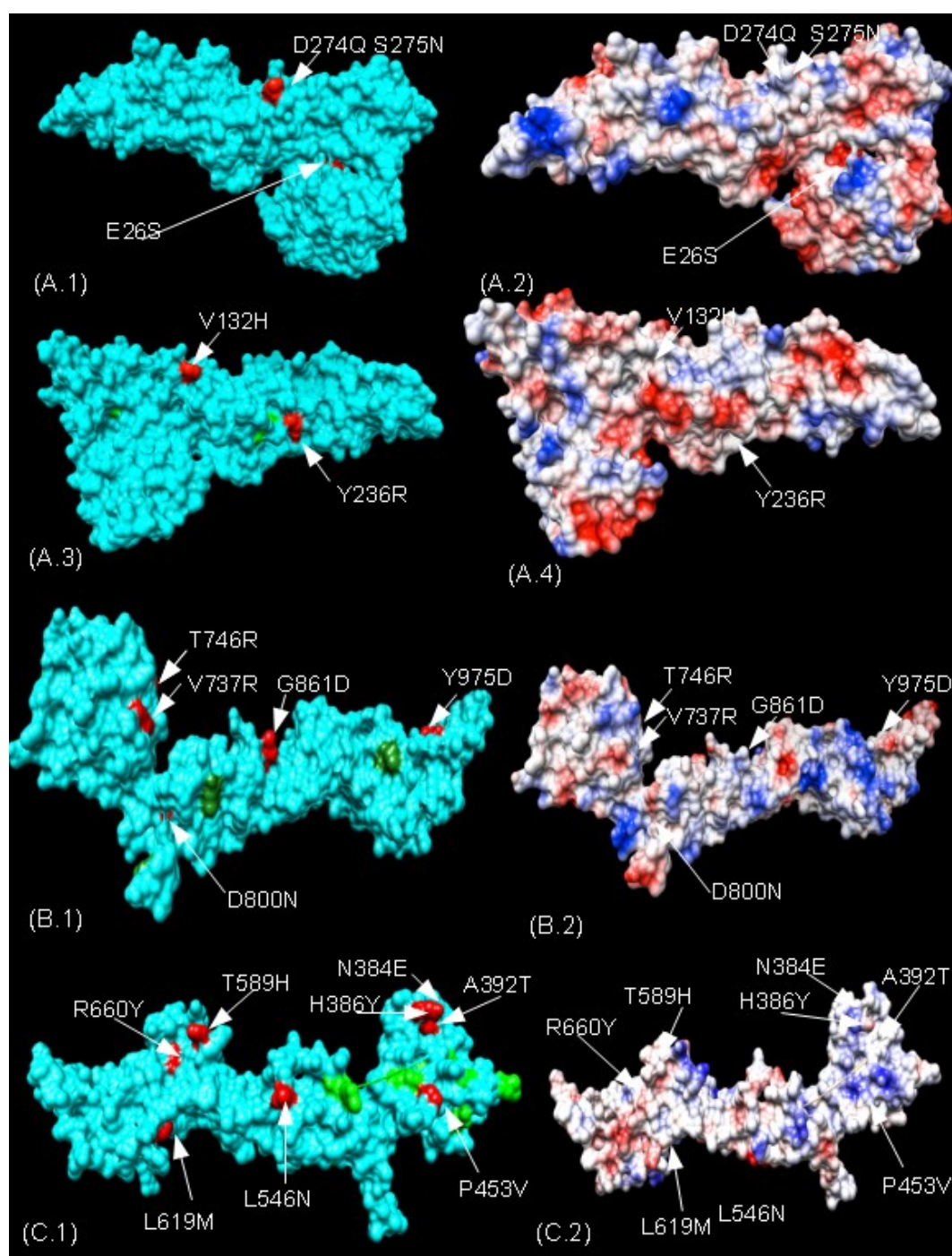


Figura 18. Comparación de los cambios y conservación de la polaridad en la proteína de envoltura. En las figuras A.1, A.3 (proteína E de DENV), B.1 (proteína E2 BVDV) y C.1 (modelo E2 HCV) se muestran en color rojo las posiciones donde existe cambio de las propiedades bioquímicas de los residuos y en verde los residuos donde se conserva. Las figuras A.2, A.3 PDB de la proteína E dengue. La figura B.2 PDB de la proteína E2 de BVDV y la figura C.2 modelo de la proteína E2 del HCV.

La hidropatía es otra propiedad bioquímica que fue analizada por PRIME (Pond y Frost, 2005), los resultados se muestran en la figura 19, en primer lugar se observa la proteína de envoltura del DENV, seguida a las proteínas de envoltura de BVDV y HCV. Del lado derecho se muestra la proteína de envoltura del virus.

En el caso de la proteína E del DENV no se encontraron ningún cambio en esta propiedad. Para la proteína del E2 del BVDV se encontraron 7 cambios estadísticamente significativos (tabla 15), en la figura 19B se observan únicamente 4 de los 7 aminoácidos.

En general los aminoácidos que presentaron el cambio fueron 4 bajo selección negativa, S836W, Q897E, I938R, I961M y 3 codones bajo neutralidad, L748R, M888E, T948A.

Para la proteína E2 de HCV se encontraron 6 codones que cambian, R394A, T396A, K408A, L409M, estos evolucionan bajo selección negativa, mientras que las posiciones K446N, A551T, evolucionan bajo neutralidad (figura 19C).

En resumen los cambios observados en los aminoácidos de las proteínas E2 de los virus BVDV y HCV, son cambios por aminoácidos hidrofílicos o aminoácidos hidrófobos en la superficie; sin embargo se esperaría tener cambios hacia aminoácidos hidrofílicos, así también como cambios por los aminoácidos que presentan carga en la en la cadena lateral, ya que se ha reportado que estos mayoritariamente se encuentran sobre la superficie, porque son responsables de ciertas interacciones. Así entonces observamos de igual forma que los procesos evolutivos actúan modificando la hidropatía de los aminoácidos presentes en estas proteínas.

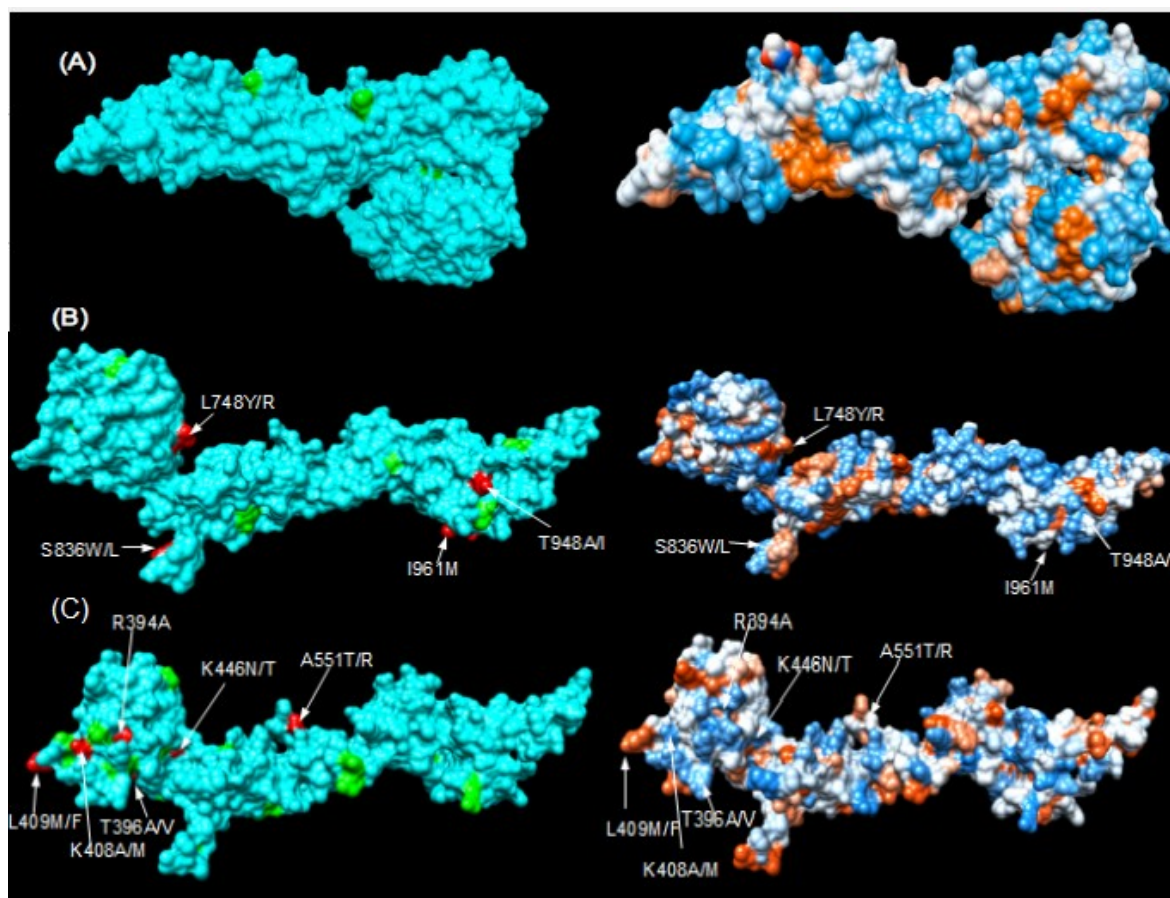


Figura 19. Comparación de las proteínas de envoltura mostrando los sitios hidropáticos y cambios en ésta propiedad. (A) Proteína de envoltura del DENV. (B) Proteína de envoltura del BVDV. (C) Proteína de envoltura del HCV.

El punto isoeléctrico (pI) es otra propiedad que evalúa PRIME, que se define como el pH al cual una sustancia anfótera tiene carga neta igual a cero. La proteína que presentó mayor número de residuos con cambios significativos en esta propiedad fue la proteína E2 del BVDV con 6 residuos, en la figura únicamente se muestran 3 residuos que presentaron cambios (Figura 20B), seguida de la proteína del HCV que presentó 5 residuos con cambios (Figura 20C), mientras que la proteína E del DENV no presentó ningún residuo con cambios en el punto isoeléctrico (Figura 20A). En la figura 20B se puede ver que la proteína E2 del BVDV, existe una tendencia de aminoácidos polares por aminoácidos polares con carga, un ejemplo son los residuos Q732E donde el cambio es por un aminoácido polar carga negativa. Otro ejemplo es el residuo T498R donde el cambio es por un aminoácido polar cargado. Estos cambios observados podrían modificar el punto isoeléctrico de la proteína, ya que estos aminoácidos cargados son los responsables de causar el cambio de pH durante el proceso de entrada del virus a la célula hospedera.

Por otro lado en la figura 20C, se observan 3 residuos de los 5 residuos que presentaron cambios en el punto isoeléctrico de la proteína E2 del HCV. En esta proteína se encontraron resultados similares a la proteína E2 del BVDV, donde los aminoácidos polares sin carga presentan cambio por aminoácidos polares cargados, donde estos últimos podrían conferir cambio en el punto isoeléctrico. En general es importante el mantenimiento del punto isoeléctrico en los aminoácidos de estas proteínas, esto porque se ha demostrado que la proteína de envoltura sufre cambios conformacionales dependientes de pH que permiten la entrada del virus a la célula, por lo tanto los procesos evolutivos podrían estar conservando las propiedades isoeléctricas de los aminoácidos de las proteínas, dado que se observaron pocos cambios en esta propiedad.

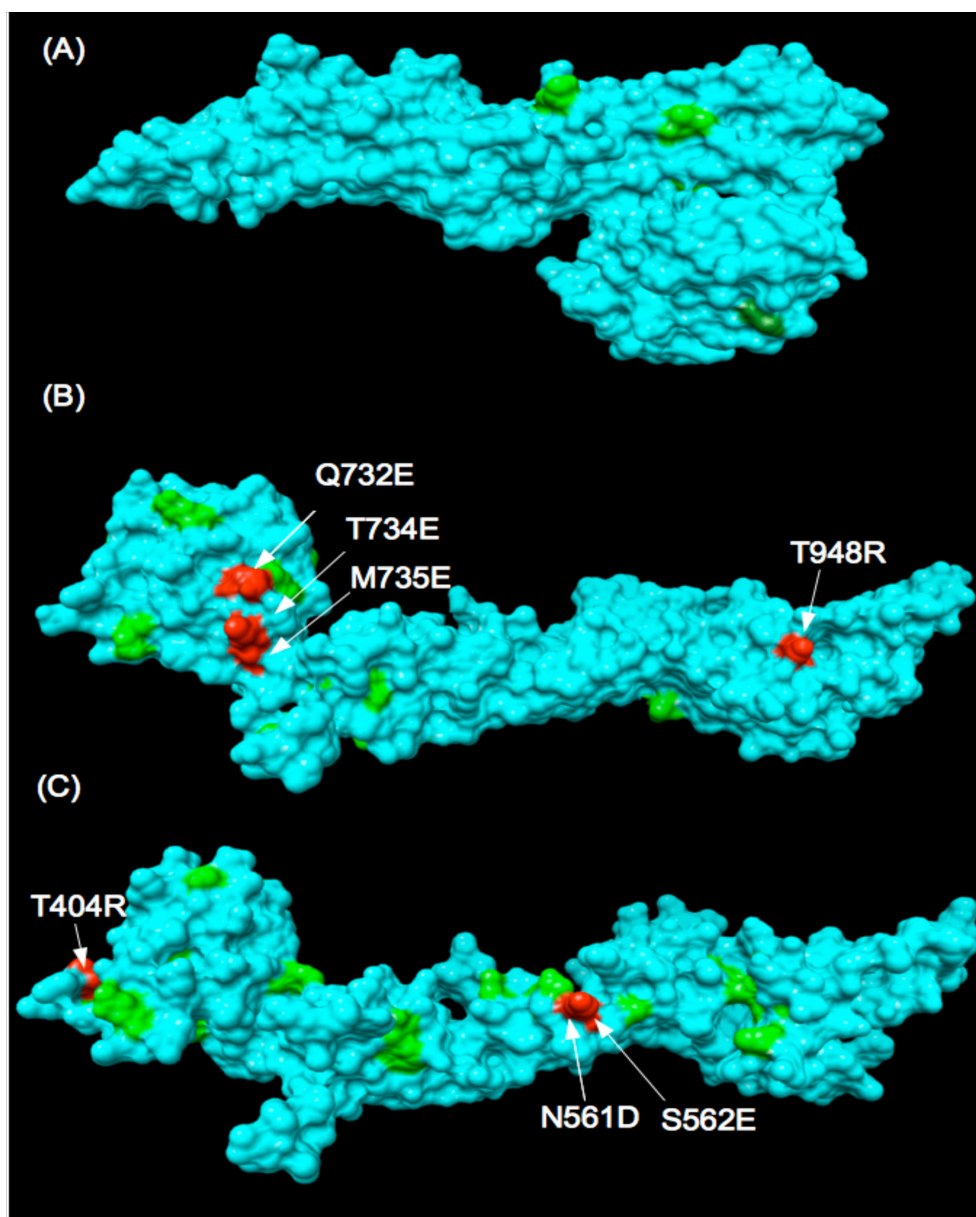


Figura 20. Cambios y conservación del punto isoeléctrico de las proteínas de envoltura. En color rojo se muestran los aminoácidos donde se presenta un cambio en el punto isoeléctrico, en color verde se muestran los codones donde se conserva el punto isoeléctrico.

Por último se evaluó el volumen, como observamos anteriormente se tiene cierta tendencia a la conservación en el volumen en los aminoácidos que conforman estas proteínas, por lo tanto quisimos ver que ocurría en aquellas posiciones donde se producía un cambio en el volumen de la cadena lateral de los aminoácidos estadísticamente significativos.

Para corroborar si en estos aminoácidos provocan un cambio significativo en la estructura de la proteína se modelaron mutantes basado en la homología, para el DENV y para el BVDV, las posiciones que se mutaron fueron para el DENV K83V, P, N y para el BVDV los residuos Y725P, H, S y C862V. Estas mutantes a su vez se refinaron mediante minimización de la energía y se alinearon estructuralmente con el PDB nativo. En la figura 21 se muestran los resultados obtenidos. Las mutantes K83V y K83P (figura 21A.1 y 21.A.2) no permiten que se dé un alineamiento completo de la mutante y el PDB nativo concretamente el alineamiento del dominio III entre estas dos proteínas no es completo. La mutante K83N mostró un alineamiento similar al PDB nativo a lo largo de toda la proteína; posiblemente el cambio de arginina por prolina, únicamente cambia el volumen en la posición y no como las otras mutantes que producen un cambio o rotamiento del dominio III. También se obtuvo el RMSD de estos alineamientos, éste es la desviación estándar de la distancia promedio entre los átomos de una molécula, e indica cuanto difieren dos moléculas entre sí, entre mayor sea el RMSD las moléculas alineadas son más diferentes. Los valores de RMSD para las mutantes K83V y K83P fueron mucho más altos que el RMSD de la mutante K83N (Tabla 17). Esto indica que en general las mutantes K83V y K83P, producen cambios que repercuten en la estructura de la proteína, mientras que la mutante K83N no lo afecta el volumen de forma global.

TABLA 17. PROTEÍNA E DEL DENV PDB REFINADO CONTRA LAS MUTANTES DEL CODÓN K83

Proteína E dengue contra mutante	RMSD
Proteína E dengue refinada contra K83V refinada	1.927
Proteína E dengue refinada contra K83P refinada	1.935
Proteína E dengue refinada contra K83N refinada	0.062

Por otro lado para la proteína E2 del BVDV, dos aminoácidos presentaron cambio en el volumen con estadística significativa. Los residuos fueron la Y725 y T826 de igual forma que la proteína anterior se modelaron en base a homología, se refinaron por medio de la minimización de la energía y se realizaron alineamientos estructurales entre el PDB de la proteína E2 del BVDV y sus mutantes. El primer alineamiento muestra las mutantes del codón Y725 (tabla 18), donde a simple vista no se observaron diferencias en los alineamientos al parecer, estas mutantes no provocan cambios importantes en el volumen. Los valores de RMSD obtenidos muestran que la mutante Y725H tiene valores más altos mientras que la mutante Y725P tiene el valor más pequeño de RMSD, por lo tanto la mutante Y725H podría tener una repercusión más grande en el volumen (Figura 22A).

El otro residuo en la proteína que presentó cambios en el volumen fue el T862 en la proteína E2 (Tabla 19), al igual que la posición anterior en los alineamientos estructurales no se observan diferencias entre la proteína E2 nativa y las mutantes. Pero los valores RMSD indican que la mutante T862C tiene el valor de RMSD más alto que la mutante T862V. Por lo tanto al igual que las mutantes del codón Y725 estos cambios en el volumen no podrían afectar de forma global el volumen de la proteína de envoltura, sino solo de la posición donde ocurre el cambio (Figura 22B).

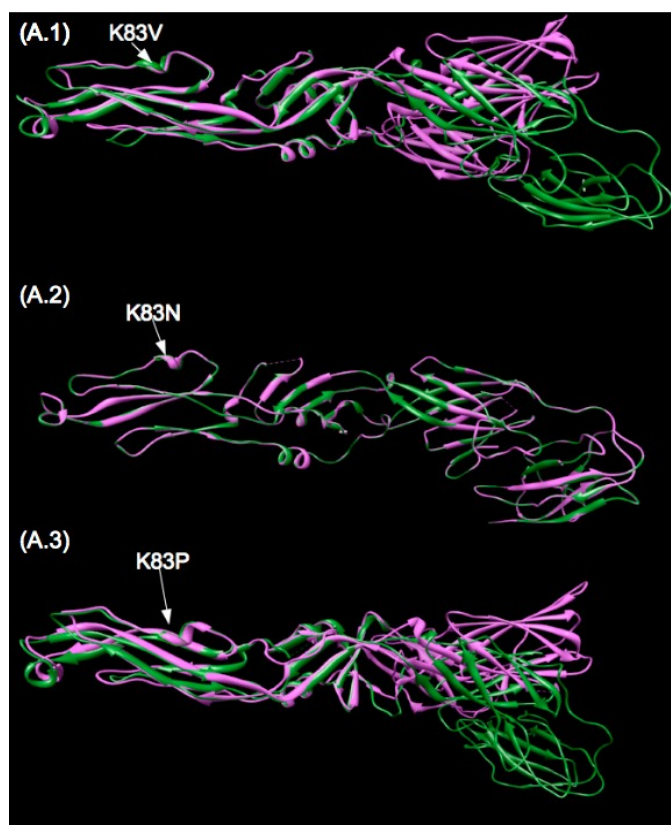


Figura 21. Alineamiento estructural de las proteínas de envoltura del DENV. Figura 21A.1 Alineamiento de la proteína de envoltura con la mutante K38V. Figura 21A.2 Alineamiento de la proteína de envoltura con la mutante K38N. Figura 21A.3 Alineamiento de la proteína de envoltura con la mutante K38P. En color verde es la proteína nativa, y en color magenta es la proteína mutante.

TABLA 18. PROTEÍNA E2 DEL BVDV PDB REFINADO CONTRA LAS MUTANTES DEL CODÓN Y725

Proteína E2 BVDV contra mutantes	RMSD
Proteína E2 refinado contra Y725S	0.146
Proteína E2 refinado contra Y725H	0.168
Proteína E2 refinado contra Y725P	0.134

TABLA 19. PROTEÍNA E2 DEL BVDV PDB REFINADO CONTRA LAS MUTANTES DEL CODÓN T862

Proteína E2 BVDV contra mutantes	RMSD
Proteína E2 contra T862V	0.188
Proteína E2 contra T862C	0.202

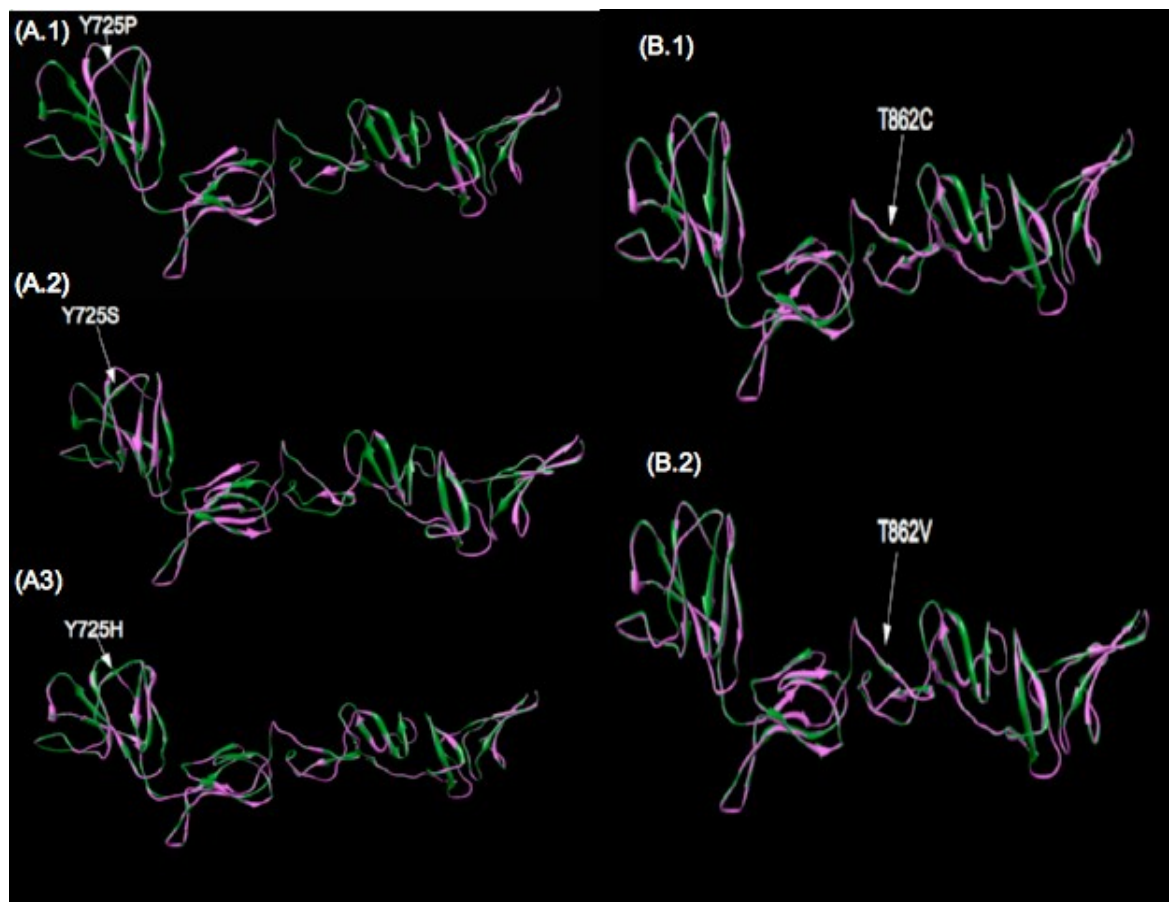


Figura 22. Alineamiento estructural de las proteínas de envoltura del BVDV. Figura 22A.1 Alineamiento de la proteína de E2 nativa con la mutante Y725P. Figura 22A.2 Alineamiento de la proteína de envoltura con la mutante Y725S. Figura 22A.3 Alineamiento de la proteína de envoltura con la mutante Y725H. Figura 22B.1 Alineamiento estructural de la proteína de envoltura con la mutante T862C. Figura 22B.1 Alineamiento estructural de la proteína de envoltura con la mutante T862V. En color verde es la proteína nativa, y en color magenta es la proteína mutante.

La propiedades bioquímica con un mayor número de cambios fue la polaridad, por lo tanto, los procesos evolutivos (selección natural negativa, positiva, episódica) que actúan sobre la proteína podría propiciar estos cambios en la polaridad de la cadena lateral de los aminoácidos, y estos procesos por otro lado tienen a la conservación del volumen, composición química y punto isoeléctrico.

Dado que en muchas posiciones los cambios en las propiedades no resultaron estadísticamente significativos, se llevó a cabo un análisis manual de las propiedades polaridad, hidropatía, volumen y punto isoeléctrico, para determinar que ocurría en cada proteína desde un punto de vista global. Los resultados se muestran en la tabla 20, cabe destacar que para este análisis se tomaron en cuenta los sitios bajo selección negativa que fueron variables y los sitios bajo selección positiva. En el DENV se observó que en su mayoría se conservan todas las propiedades bioquímicas estudiadas. La propiedad bioquímica más conservada fue el volumen, mientras la polaridad fue la propiedad bioquímica donde mayor número de codones presentaron cambio. Para el BVDV, la polaridad y la hidropatía fueron las propiedades donde más los sitios cambiaron la propiedad, y la más conservada fue el volumen. Por último, para el HCV se obtuvieron resultados similares a los del BVDV las propiedades con más cambios fueron polaridad e hidropatía y la más conservada fue el volumen.

Por otra parte en los sitios bajo selección negativa que eran variables, se puede observar un fenómeno interesante, mediante el análisis manual, observamos que aunque existían cambios por otros aminoácidos, éstos poseían las mismas propiedades bioquímicas, es decir, ninguna de las cuatro propiedades analizadas cambiaba, esto se muestra en la tabla 21.

TABLA 20. CAMBIOS Y CONSERVACIÓN EN LAS PROPIEDADES BIOQUIMICAS DE LOS SITIOS BAJO SELECCION POSITIVA Y NEGATIVA VARIABLES (análisis manual)

Propiedad bioquímica		DENV	BVDV	HCV
Polaridad	Conservada	87	66	55
	No conservada	59	88	101
Hidropatía	Conservada	98	73	63
	No conservada	49	82	93
pI	Conservada	98	85	100
	No conservada	49	70	56
Volumen	Conservada	133	126	124
	No conservada	14	29	32

TABLA 21. CONSERVACIÓN EN LAS PROPIEDADES BIOQUIMICAS DE LOS SITIOS BAJO NEGATIVA A PESAR DE SER SITIOS VARIABLES (Análisis manual)

DENV	BVDV	HCV
Codones		
4, 6, 11, 15, 43, 45, 50, 53, 61, 65, 66, 72, 80, 91, 93, 94, 108, 113, 114, 125, 129, 138, 139, 141, 143, 173, 181, 183, 189, 191, 199, 207, 208, 218, 219, 220, 272, 273, 275, 279, 281, 286, 292, 323, 324, 325, 340, 360, 361, 367, 372, 382, 383, 392	694, 695, 706, 710, 715, 731, 736, 738, 745, 759, 762, 786, 793, 796, 805, 809, 817, 824, 831, 868, 875, 877, 882, 884, 913, 917, 926, 931, 941, 944, 950, 953, 956, 961, 963, 970, 1022, 1029, 1031, 1034	399, 402, 403, 413, 414, 417, 418, 422, 428, 437, 442, 447, 450, 456, 462, 473, 488, 495, 496, 507, 515, 519, 520, 521, 523, 524, 539, 541, 542, 543, 544, 546, 554, 555, 563, 572, 573, 614, 615, 617, 619, 620, 632, 636, 640, 644
Total de codones= 54	Total de codones= 40	Total de codones= 46

En resumen, 54 codones, codificantes dentro de la proteína de envoltura del DENV que fueron variables pero se detectaron bajo selección negativa (tabla 21), no presentaron cambios, en ninguna propiedad bioquímica. Lo mismo ocurrió en el BVDV y HCV, con 40 y 46 posiciones respectivamente. Por lo tanto estos sitios, aunque existan mutaciones que codifiquen para otro aminoácido distinto no producen cambios en las propiedades bioquímicas, conservando la función de la proteína.

Capítulo 6

Discusión

La proteína de envoltura de los virus de la familia *Flaviviridae* posee funciones indispensables para el ciclo de vida de éstos, sin embargo esta proteína difiere tanto en secuencia como estructuralmente en los diferentes géneros de la familia. Dado que muchas de las funciones de esta proteína se encuentran conservadas en los diferentes géneros, es posible que los procesos evolutivos que actúan sobre ellas sean similares entre sí, de tal manera que se conserve la integridad funcional de la proteína. Por lo tanto, este proyecto se llevó acabo con la finalidad de comparar los procesos evolutivos que ocurren en la proteína de envoltura de los virus DENV, BVDV y HCV de la familia *Flaviviridae*.

Lo primero que evaluamos fue la variabilidad de las proteínas de envoltura de los tres modelos de estudio. Como se mostró en los resultados, esta variabilidad se distribuye uniformemente a lo largo de los tres dominios, presentando valores en un rango del 60 al 70 % de los sitios. Esto es probablemente un reflejo del proceso mutacional que ocurre durante el proceso de replicación, ya se conoce que los virus de RNA presentan una tasa de mutación entre los 0.4 y 1.1 errores por nucleótido por ronda de replicación, debido en gran medida a que la RNA polimerasa, que se encarga de copiar su genoma, es deficiente de la actividad exonucleasa, y como consecuencia se generan poblaciones polimórficas y con gran plasticidad (Lauring *et al.*, 2013; Belsthaw *et al.*, 2008; Zárate y Novella, 2004).

Por otra parte, un análisis de uso de codones -realizado con el fin de determinar si el uso de codones de los virus se asemejaba al de sus hospederos- reveló una agrupación en dos *clusters*: el primero constituido por la mayoría de los virus (DENV, BVDV, CSFV, YFV, WNV) y el segundo formado por *Hepacivirus* (HCV, HGB) y los hospederos (mosquito, borrego, vaca, puerco y humano). Esta asociación del uso de codón en los géneros de la familia *Flaviviridae* ya ha sido reportada previamente por Lobo *et al.* (2009); además, los

autores mostraron que los virus del género *Flavivirus* que no se transmiten por medio de un vector también presentan un uso de codón similar al de sus hospederos.

La similitud entre el uso de codón del HCV y su hospedero es probablemente debida al ciclo de vida de este virus, que causa una infección crónica; es decir, posiblemente durante este tiempo el virus ha optimizado su uso de codón con el objetivo de generar una alta progenie durante las primeras etapas de la infección, permitiendo así establecer la cronicidad (Lobo *et al.*, 2009). Esto puede deberse a un proceso de adaptación donde la abundancia relativa de tRNAs hace posible una optimización en el proceso de traducción, o bien a un proceso mutacional en el que las limitaciones en la composición de los nucleótidos en el medio celular influyen en la probabilidad del tipo de mutación (Sharp y Li, 1986). En el caso específico de los virus del género *Flavivirus*, éstos mantienen un ciclo de replicación, lo que implica una restricción evolutiva puesto que un cambio que favorezca la replicación y traducción dentro de uno de sus hospederos podría perjudicar la replicación y traducción en el otro, lo que explicaría que su uso de codones se diferencie del de sus hospederos, por lo que observamos una diferente agrupación.

La variabilidad observada a nivel de gen no necesariamente tendría que reflejarse en la variabilidad a nivel de aminoácidos, porque es posible que muchas de las mutaciones que se fijan en la población sean sinónimas. Encontramos que los sitios variables en la secuencia de aminoácidos se distribuyeron de forma uniforme a lo largo de los tres dominios, presentando porcentajes altos de alrededor del 70-80 %; estos resultados indican que muchas de las mutaciones que ocurren a nivel de nucleótidos no son sinónimas, revelando posiblemente la presencia de selección positiva o de los efectos de la deriva en mutaciones casi-neutras.

Al menos parte de la variabilidad observada en la proteína de envoltura del DENV se ha atribuido a las fuerzas selectivas; así por ejemplo, Chao *et al.* mencionan que los dominios I y II son variables mientras que el dominio III es más conservado, y esto lo atribuyen a restricciones funcionales en ese dominio. En su trabajo ellos demostraron que algunos de los sitios dentro de los dominios I y II evolucionan bajo selección positiva, mientras que el dominio III evoluciona bajo selección negativa. (Chao *et al.*, 2005). Esto se contrapone a los resultados que obtuvimos con respecto a la variabilidad en la proteína de E del DENV, donde la distribución de los sitios polimórficos es homogénea en los tres

dominios. Lo mismo ocurre con las proteínas de envoltura en BVDV y HCV, pues éstas son altamente variables en los tres dominios.

En estudios previos donde se evaluaron las fuerzas evolutivas que actuaban sobre las proteínas de envoltura de diferentes virus como el virus influenza (proteína Hemaglutinina, HA), el coronavirus SARS (proteína Spike, S), y el virus de la inmunodeficiencia humana (gp160), se mostró que estas proteínas fueron altamente variables y que la mayoría de posiciones evolucionaron bajo selección natural positiva, concretamente los epítopes de reconocimiento de anticuerpos (Tang *et al.*, 2009; Qu *et al.*, 2011; Yamaguchi y Gojobori, 2000).

En este trabajo quisimos investigar si las proteínas de envoltura de tres miembros de la familia *Flaviviridae* (DENV, BVDV y HCV) también debían su variabilidad a la selección positiva. Sin embargo, en nuestro caso los resultados obtenidos del análisis de selección natural indicaron que la mayoría de los sitios de las tres proteínas de envoltura evolucionan bajo selección negativa y pocos sitios evolucionan bajo selección positiva, contrario a lo reportado en los virus de RNA antes mencionados. No obstante, para cada una de las proteínas de envoltura de los virus que estudiamos existen reportes de las particularidades en su evolución; concretamente, diferentes fuerzas actúan en las proteínas de envoltura de estos virus:

En el caso de la proteína E del DENV -y en general de los arbovirus- está reportado ampliamente que la mayoría de los residuos dentro de las proteínas de envoltura evolucionan bajo fuerte selección negativa debido a limitaciones evolutivas, pues estos virus realizan un ciclo de replicación entre dos hospederos, el mosquito y el humano (Twiddy *et al.*, 2002; Woelk y Holmes, 2002).

Holmes *et al* muestran que la selección positiva es rara en la proteína de envoltura del DENV y estiman que más del 90% de las mutaciones no sinónimas son deletéreas, por lo que son eliminadas de la población por medio de la selección purificadora. A nivel intra-paciente la proteína de envoltura presenta muchas mutaciones no sinónimas que están bajo selección positiva, sin embargo muchas de éstas se pueden perder por los cuellos de botella durante la transmisión, ya que a nivel intra-hospedero la mayoría de las mutaciones observadas son sinónimas (Holmes, 2003; Behura y Severson *et al.*, 2013).

Por otro lado, Rodthong y Auewarakul reportan 12 sitios bajo selección positiva en la proteína de envoltura del DENV, afirmando que éstos se encuentran sobre la superficie, que son accesibles a anticuerpos y que posiblemente son sitios responsables de diversificación; sin embargo presumen que esta proteína puede estar bajo fuerte selección positiva dentro de un paciente (Rodthong y Auewarakul, 2012).

Los estudios anteriores ponen en evidencia que el virus DENV generalmente evoluciona bajo selección negativa y algunas cuantas posiciones evolucionan bajo selección positiva, habiendo así correlación con nuestros resultados, ya que a lo largo de la proteína E la mayoría de las posiciones evolucionan bajo selección negativa y solamente la posición V132 bajo selección positiva. No se ha reportado previamente que esta posición esté bajo selección positiva, pero las posiciones 127 a la 134 sí se encuentran reportadas como un péptido de reconocimiento de anticuerpos en pacientes convalecientes de una infección con el serotipo 2 del DENV (Twiddy *et al.*, 2002). Ahora bien, en el análisis de variabilidad encontramos en la proteína E una gran cantidad de sitios variables que no son explicados ni congruentes con la selección negativa observada, por lo que posiblemente exista otro proceso por medio del cual evoluciona la proteína E del DENV.

Para el BVDV, Tang y Zhang han reportado que en el genotipo 1 de este virus la presencia de selección positiva se restringe a pocas posiciones, y que en general la mayoría de los aminoácidos de esta proteína están evolucionando bajo selección negativa (Tang y Zhang, 2008), lo cual coincide con los resultados obtenidos en este trabajo. Sin embargo, de igual forma que la proteína E del DENV, la proteína E2 del BVDV es variable en la secuencia de aminoácidos, y en este caso la variabilidad no ha sido asociada a selección positiva.

Parte de la variabilidad observada en las posiciones que evolucionan bajo selección positiva está relacionada con la disminución notable de la virulencia. Por ejemplo, Rissati *et al.* realizaron diferentes mutaciones y al menos 13 sustituciones en la región Y975 -que abarca las posiciones 882 aa a la 1046 aa- y cuando en el aminoácido 975 se cambió alanina por ácido aspártico (A975E), observaron que este único cambio causaba la muerte en cerdos; al asociarse a otras mutaciones (D958G y R979S) de igual forma se provocaba la muerte (Rissati *et al.*, 2007). Por lo tanto, se sabe que el extremo carboxilo de la proteína E2 está relacionado con la virulencia.

El otro sitio bajo selección fue el S832 en la proteína E2 del BVDV; se ha reportado que este sitio está dentro de un epítoto relacionado con la virulencia (TAV**S**PTTLR, en negrita la posición 832). El epítoto se encuentra dentro de la proteína de envoltura E2 del virus de fiebre porcina clásica cepa Brescia, que es altamente virulenta. Las mutaciones en este epítoto producen cepas que no son capaces de replicar en células SK6, y cuando se inocularon cerdos con mutantes de virus de fiebre porcina en este epítoto, no se produjo la muerte a los animales de experimentación, mientras que la cepa silvestre con el epítoto TAV**S**PTTLR sí lo hace, siendo altamente virulenta (Risatti *et al.*, 2006).

En el caso del HCV se reporta que la proteína E2 se encuentra bajo fuertes presiones de selección del sistema inmune -principalmente en las regiones hipervariables- y que la selección purificadora domina la evolución de esta proteína en pacientes con hepatitis crónica (Brown *et al.*, 2005; Sheridan *et al.*, 2004; Farci *et al.*, 2002). En nuestro caso, como se mencionó en los resultados, únicamente se encontraron 2 sitios bajo selección positiva dentro del dominio I de la proteína, pero en general la mayoría de los sitios evolucionan bajo selección negativa; esto coincide en parte con los estudios anteriores, ya que las posiciones E384 y E444 evolucionan bajo selección positiva, encontrándose la primera de éstas dentro de la región hipervariable 1, y la segunda en el sitio de reconocimiento del receptor.

En conjunto nuestros resultados y los reportes de la literatura señalan que la selección negativa domina la evolución de estas proteínas, mas esto se contrapone con la alta variabilidad que presentan. La distribución de los sitios bajo selección negativa se encontró de forma uniforme a lo largo de los tres dominios. Sin embargo Chao *et al.* (2005) mencionan que el dominio III de la proteína E de DENV es más conservado y se encuentra bajo selección negativa, puesto que es el dominio de unión al receptor.

El hecho de que la mayoría de los codones de los tres genes han evolucionado bajo selección negativa se pudo confirmar en el análisis de *Evolutionary Fingerprint*, que parte del supuesto de que proteínas con funciones similares habrían evolucionado bajo los mismos procesos evolutivos, independientemente de su relación de homología. En conjunto los análisis de detección de selección sitio por sitio nos indicaron que, a lo largo de la filogenia, en la mayoría de los codones se acumula un mayor número de mutaciones sinónimas que no sinónimas, lo que se contradice con el grado de variabilidad encontrada en las proteínas. De manera convencional se asume que el exceso de mutaciones sinónimas es un reflejo de

la acción de la selección negativa, que limita el número de mutaciones no sinónimas debido a que éstas darían lugar a variantes con una menor adecuación.

Con el objetivo de analizar la dinámica evolutiva de los codones bajo selección negativa y positiva, se realizaron reconstrucciones de caracteres ancestrales. En las reconstrucciones de caracteres ancestrales de los codones bajo selección positiva se mostró la historia de los cambios de aminoácidos, como era de esperarse. Por otro lado, en los sitios bajo selección negativa existían dos tipos de reconstrucciones: en algunos casos se veía que de manera casi exclusiva se acumulaban mutaciones sinónimas (de acuerdo con lo esperado en este tipo de selección), mientras que en otros la mayoría de las ramas presentaban mutaciones sinónimas pero en ciertas regiones de la filogenia presentaban cambios no sinónimos. Esta observación, aunada al gran número de polimorfismos en estas proteínas, nos llevó a plantear la posibilidad de que algunos de los sitios estuvieran evolucionando bajo selección episódica.

La selección episódica se refiere al hecho de que no existe una razón para pensar que la selección natural actúe de forma constante a lo largo de la historia evolutiva de un gen; en muchos casos es posible que la selección favorezca el aumento en frecuencia de aquellas variantes que confieran una ventaja en la adecuación de un organismo, y que posteriormente estas variantes sean mantenidas por selección negativa (Pond *et al.*, 2009).

Los análisis de selección episódica realizados sobre las secuencias de estudio revelaron que algunos de los sitios que se habían clasificado como seleccionados negativamente en realidad evolucionaron bajo selección episódica, aunque de hecho se encontraron pocas posiciones bajo selección episódica. Por ejemplo, para DENV únicamente 13 posiciones evolucionan bajo selección episódica, mientras que para BVDV son 11 y para HCV son 15 posiciones. Aunque son pocas las posiciones donde se encontró la presencia hacia cambios no sinónimos, esto era un evento recurrente en algunas de las posiciones, por lo tanto, la selección episódica permite la fijación de estas variables.

La selección natural episódica es muy importante en la evolución de las proteínas; Studer y Rechavi mencionan que la evolución funcional de las proteínas parece implicar períodos donde se alterna la selección conservadora y periodos de cambios relativamente rápidos. Durante los periodos de cambios rápidos se da una acumulación de mutaciones no sinónimas, y muchas de éstas son fijadas por la selección episódica. Asimismo se han

desarrollado varios modelos para evaluar la evolución episódica en las secuencias de proteínas y éstos se han implementado para detectar la divergencia funcional entre los grupos de secuencias de proteínas. Con estos modelos se ha observado, por ejemplo, que un residuo de serina se puede conservar plenamente en todos los genes ortólogos dentro de un clado, mientras que una mezcla de alanina, ácido aspártico y lisina se encuentra en los genes ortólogos de otro clado. Este tipo de asimetría se ha observado en varios genes duplicados, como el receptor tirosina quinasa en el pez teleósteo, o la familia groEL en *Clamidia*, lo que sugiere la evolución de una nueva función; por lo tanto, se hipotetiza que la selección episódica es responsable de la neo-funcionalización o incluso de la subfuncionalización, que resulta de la duplicación de genes. (Studer y Rechavi, 2009).

Sin embargo, cuando se mapearon sobre la proteína los codones bajo selección negativa que codifican para diferentes aminoácidos, se observó que muchos de éstos se encuentran en la superficie de la proteína, por lo que pueden ser reconocidos por anticuerpos o interactuar con otras proteínas celulares o virales, motivos por los cuales podrían favorecer la variabilidad en estos sitios.

Para evaluar el efecto de la variabilidad observada en la secuencia de aminoácidos de las proteínas de envoltura, y la selección natural actuando sobre éstos, se realizó el análisis de las propiedades bioquímicas de las proteínas, que reconstruye la historia de estos cambios sitio por sitio. Este análisis complementa los anteriores, pues no todos los cambios no sinónimos tienen el mismo impacto en la función; además, en este análisis asumimos que los cambios que conservan las propiedades bioquímicas serán menos importantes (en términos de la información que nos proporcionan) que aquellos que las modifican.

Los resultados que se obtuvieron mostraron cambios en la polaridad e hidropatía, y sin embargo se encontró una tendencia particular a la conservación en el volumen, pl y composición química de las proteínas.

La polaridad en las proteínas de envoltura virales es importante, ya que permite las interacciones proteína-proteína, interacciones con el receptor, con el co-receptor y la unión de anticuerpos con componentes de la inmunidad celular (Sowmya *et al.*, 2011). Por ejemplo, Kumar *et al.* (2010) mostraron que en la proteína de envoltura gp160 del VIH la mayoría de los aminoácidos evolucionan bajo selección positiva, y que a su vez la mayoría

de estos aminoácidos son polares, lo que sugiere que la selección natural podría conservar la polaridad de los residuos bajo selección. En nuestro caso, los sitios bajo selección positiva o episódica sufren cambios en la polaridad, lo que podría ser relevante para la evasión del sistema inmune.

Los aminoácidos polares también son responsables del proceso de entrada viral –específicamente en el DENV- ya que se sabe que las His-282 e His-317 (que están muy conservadas) son sensores del pH, y su protonación debida a la reducción del pH en el endosoma genera el cambio conformación de la proteína E, exponiendo el péptido fusogénico del dominio II (Nayak *et al.*, 2009; Gadkari y Srinivasan, 2010).

Como se mencionó en los resultados, encontramos conservación en el punto isoeléctrico, el cual se encuentra relacionado con la polaridad de los aminoácidos, ya que los aminoácidos polares con carga como las histidinas censan cambios de pH. Además, el punto isoeléctrico es importante en las proteínas porque permite su solubilidad en un ambiente acuoso. El punto isoeléctrico depende de los aminoácidos cargados como la histidina -cadenas laterales de imidazol- además de otros aminoácidos como el ácido glutámico, ácido aspártico, cisteína -grupo tiol-, tirosina -grupo fenol-, lisina -grupo ϵ -amonio- y arginina -grupo guanidinio- (Nehete *et al.*, 2013). Como observamos en los resultados, esta propiedad mostró pocos cambios, sin embargo los cambios que se reportaron en todas las proteínas fueron hacia aminoácidos cargados como ácido aspártico, glutámico, y arginina, es decir, hacia aminoácidos polares con carga y responsables de censar el cambio del pH.

El mantenimiento del punto isoeléctrico durante el proceso de entrada viral es importante, como se ha reportado para el caso del DENV: La acidificación del virión dentro del endosoma celular produce una disociación de los dímeros de la proteína E, exponiendo así el péptido de fusión; posteriormente se da una reorientación de dominios que permite la formación irreversible de trímeros del dominio II, y la inserción del péptido de fusión en la membrana endosomal (Modis *et al.*, 2004). En HCV y BVDV no se ha reportado con precisión cómo es el proceso de fusión de membranas, pero se sabe que ambos entran a la célula por una endocitosis mediada por clatrina, y entonces entran por medio de un endosoma cuyo microambiente interno es ácido (Codran *et al.*, 2006; Lecot *et al.*, 2005).

Por otro lado, como se expuso en los resultados, la propiedad bioquímica que más se conservó en las proteínas de envoltura fue el volumen. Los sitios que se detectaron con

cambio en el volumen de la cadena lateral de los aminoácidos se analizaron con más detalle en las proteínas E del DENV y E2 del BVDV. En la primera, las mutaciones K83V y K83P provocan un desplazamiento espacial del dominio III, que no permite la correcta alineación del dominio III de las proteínas mutantes con la proteína nativa; este resultado es sumamente interesante pues podría tener implicaciones para la función de la proteína y su empaquetamiento en el virión. En el caso de las mutantes de E2 del BVDV se observó un RMSD no muy alto, indicando que estas mutantes no producen cambios muy drásticos en el alineamiento de la proteína. La razón de encontrar pocos cambios en el volumen se debe probablemente a que los cambios en esta propiedad pueden provocar rotamientos en los dominios de la estructura de la proteína. Específicamente, se sabe que en la proteína E del DENV los ectodominios son muy importantes durante el ensamble del virión -los cambios dentro del dominio II de la proteína provocarían una alteración en el ensamble del virión- (Klein *et al.*, 2013). Sin embargo los cambios en el volumen podrían dar un incorrecto arreglo de dímeros sobre la superficie del virión (Perera y Kuhn, 2008).

Dado que muchos de los aminoácidos no resultaron con cambios estadísticamente significativos en las propiedades bioquímicas, se realizó un análisis manual de las propiedades bioquímicas, a partir de los alineamientos múltiples y sin tomar en cuenta una reconstrucción de caracteres ancestrales; por lo anterior, este tipo de análisis pudiera estar subestimando o sobreestimando los cambios en las propiedades bioquímicas, pero nos permitió observar de forma global los cambios que ocurrieron. En general, observamos que la propiedad bioquímica que menos se conservó en las proteínas de envoltura de los virus fue la polaridad, y esto -como se mencionó anteriormente- puede ser un mecanismo de evasión de la respuesta inmune; la propiedad bioquímica que más se conservó fue el volumen, porque cualquier cambio en el volumen de la cadena lateral de los aminoácidos puede repercutir en la formación de homodímeros o en el ensamble del virión, lo que daría como resultado una partícula viral no viable. Ahora bien, al comparar las tres proteínas de envoltura observamos que la proteína de envoltura del DENV presentó los porcentajes más altos de conservación (a pesar de que las tres proteínas presentaban porcentajes similares de variabilidad a nivel de secuencia) mientras que la proteína de HCV presentó menor grado de conservación. Este patrón de variabilidad y conservación en la proteína de DENV podría deberse -al menos en parte- a que el ciclo de vida del DENV requiere que sea capaz de

replicarse en dos hospederos, y esto es consistente con lo que se ha observado para otros arbovirus. En el caso de HCV la variabilidad puede ser el resultado de la constante presión del sistema inmune actuando sobre el virus, o de una mayor robustez en estas proteínas, lo que correlaciona con el gran número de genotipos reportados.

En conjunto nuestros resultados muestran que en los virus estudiados de la familia *Flaviviridae* la evolución de sus proteínas de envoltura es en parte similar, ya que éstas son altamente variables en la secuencia de aminoácidos; muchas de estas variantes son fijadas por la selección episódica, lo que indica por un lado que la selección natural no actúa de forma constante en la evolución de las proteínas -permitiendo la fijación de variantes no sinónimas- y por el otro que estos procesos selectivos favorecen la conservación del volumen de la cadena lateral de los aminoácidos, mientras que cambian la polaridad de éstos. También observamos que la proteína de envoltura del DENV es más conservada a pesar de la variabilidad observada y que la proteína de envoltura del HCV es menos conservada. Todos estos hallazgos aportan información relevante sobre la dinámica evolutiva de los virus de esta familia.

Capítulo 7

Conclusiones

Los procesos evolutivos tales como la tasa de mutación y el efecto de la selección natural, son similares en la proteína de envoltura *Flaviviridae*.

Pocas posiciones del gen que codifica para de las proteínas de envoltura de la familia *Flaviviridae* evolucionaron bajo selección natural positiva y parecen estar relacionados con el escape del reconocimiento por anticuerpos.

Estas proteínas son altamente variables, sin embargo la selección episódica, es en buena medida responsable de la variabilidad observada. Estos eventos de selección episódica se alternan con períodos de selección negativa que da lugar a la acumulación de mutaciones no sinónimas.

En cuanto a las propiedades bioquímicas de los aminoácidos presentes en estas proteínas, se observó que el volumen es la propiedad más conservada, mientras que la polaridad es la menos conservada. Por otro lado, se encontró que la proteína de envoltura del DENV es más conservada que las proteínas del BVDV y HCV.

Capítulo 8

Perspectivas

Los resultados encontrados en el desarrollo de esta tesis revelaron similitudes en los procesos evolutivos de las proteínas estudiadas, por lo que sería interesante llevar a cabo la caracterización de procesos evolutivos en las proteínas de envoltura de otros virus, y determinar si estas proteínas que llevan a cabo funciones semejantes han evolucionado bajo las mismas fuerzas selectivas. En particular sería interesante analizar a proteína de envoltura de los miembros del género *Pegivirus*, para completar en estudio de la familia *Flavivirus*.

En otro estudio analizar si los cambios en las propiedades bioquímicas como en el volumen y la polaridad producen cambios funcionales relevantes en el virión, esto mediante la generación de proteínas mutantes en el laboratorio y posteriormente determinar efectos en la unión al receptor, el reconocimiento y neutralización por anticuerpos, o si se producen partículas virales viables a partir de estas proteínas mutantes.

Complemento Resultados

USO DE CODÓN ANÁLISIS DE ESTADÍSTICA MULTIVARIADA.

El análisis de similitud de uso de codón entre los virus y sus hospederos, también se analizó mediante un escalamiento multidimensional, que permite observar gráficamente y en pocas dimensiones la distancia existente entre los diferentes genomas de acuerdo con su uso de codón, así por ejemplo si dos puntos observados se encuentran próximos entre sí indica que ambos tienen un patrón de uso de codón semejante y por lo tanto existe una distancia menor entre ellos; en el caso contrario, cuando dos puntos se encuentran distantes entre sí, indica que tienen un sesgo en el uso de codón muy distinto. Los resultados obtenidos del escalamiento multidimensional realizado se observan en la figura 23, donde los valores de distancia se representan en dos dimensiones para poder ser observados; se obtuvieron dos grupos: por un lado los hospederos se encuentran hacia lado derecho del gráfico en el eje de la dimensión 1 dentro de los valores positivos, y cercanos a éstos se agruparon los virus del género *Hepacivirus*; Por otro lado se agruparon los virus de los géneros *Flavivirus* y *Pestivirus* (Figura 23). Es decir, el uso de codón entre los géneros *Flavivirus* y *Pestivirus*, es más parecido entre sí y éste es distante, al uso de codón de sus hospederos, en concordancia con lo observado en el análisis de clustering jerárquico. El resultado de la prueba de bondad de ajuste fue de 55.932 %, que indica que al representar los datos en dos dimensiones se conserva el 56 % de la variabilidad real.

Finalmente, se utilizó el método de **k-means**, el cual clasifica los datos en un número de clusters predeterminados, en este caso se hizo el análisis para 2 clusters, dados los resultados observados previamente.

Los resultados obtenidos de k-means se observan en la tabla 22, donde en el cluster 1 se agrupan los virus de los géneros *Flavivirus* y *Pestivirus*, y los organismos *Rana huanrenensis* y *Aradopsis Thaliana*, mientras que dentro del cluster número 2 se agrupan los virus de HCV y HGB, junto con los hospederos.

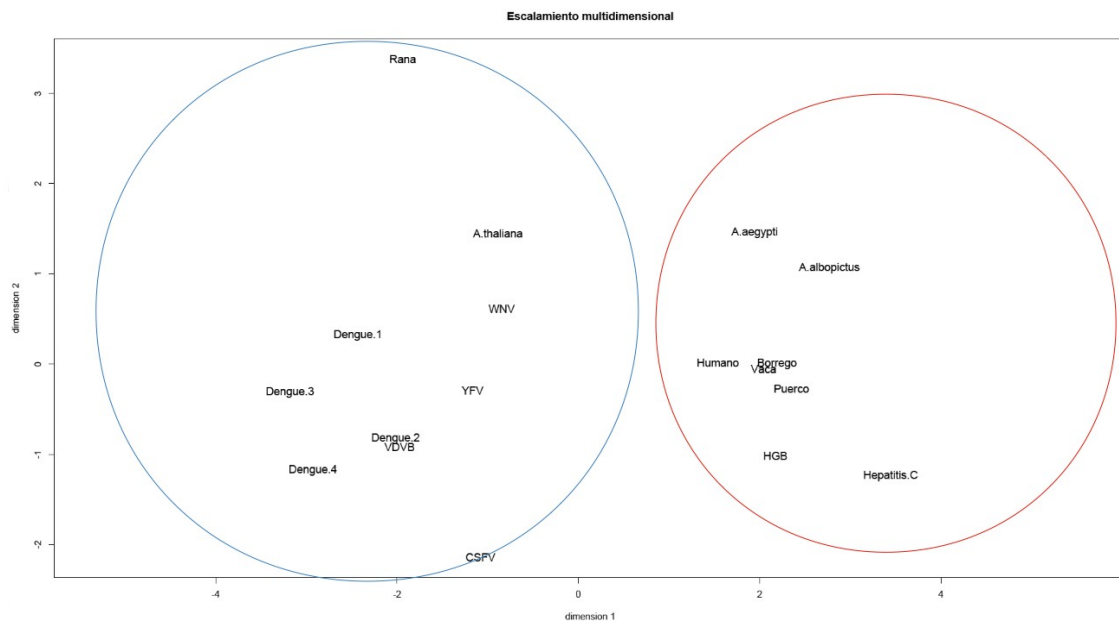


Figura 23. Escalamiento multidimensional. Análisis del uso de codón mediante un escalamiento multidimensional realizado en R, donde se observan dos cluster, dentro del círculo azul el cluster que agrupo a los virus junto con *A. Thaliana* y la rana, mientras que en color rojo el clúster que agrupo a los hospedero.

Los resultados obtenidos confirman lo obtenido por los métodos de clustering jerárquico y escalamiento multidimensional. El hecho de que el uso de codón entre los virus del genero *Hepacivirus* sea parecido al de su hospedero podría indicar que la selección natural ha sido responsable, al menos en parte, del sesgo en el uso de codón que presentan estos virus.

La herramienta k-means también arrojó las medias del uso de codón entre los dos clusters, estas medias se graficaron mediante el programa R (www.r-project.org/) donde se pudieron observar diferencias en las presencias del uso de codón entre los dos clusters (gráfico 1). Por ejemplo los codones AGA (R), ACA (T), CCA (Q), CAT (H), GGA (G) y TCA (S), son utilizados en mayor grado por el cluster de los virus, mientras que en sus hospederos estos codones son poco utilizados. Por otro lado en el cluster 2 los codones que tienen mayor preferencia a ser utilizados fueron ACC (T), AGC (S), ATC (L), CAC (H), CGG (R), CGC (R), CTG (L), GCC (A), GTG (V) y TCC (S).

TABLA 22. CLUSTER OBTENIDOS POR MEDIO DEL ALGORITMO K-MEANS

Cluster 1	Cluster 2
DENV 1	HCV
DENV 2	HGB
DENV 3	<i>Homo sapiens (humano)</i>
DENV 4	<i>Bos primigenius taurus (vaca)</i>
YFV	<i>Ovis aries (borrego)</i>
WNV	<i>Sus scrofa domestica (puerco)</i>
BVDV	<i>A.aegypti</i>
CSFV	<i>A.albopictus</i>
<i>Arabidopsis thaliana</i>	
<i>Rana huanrenensis (Rana)</i>	

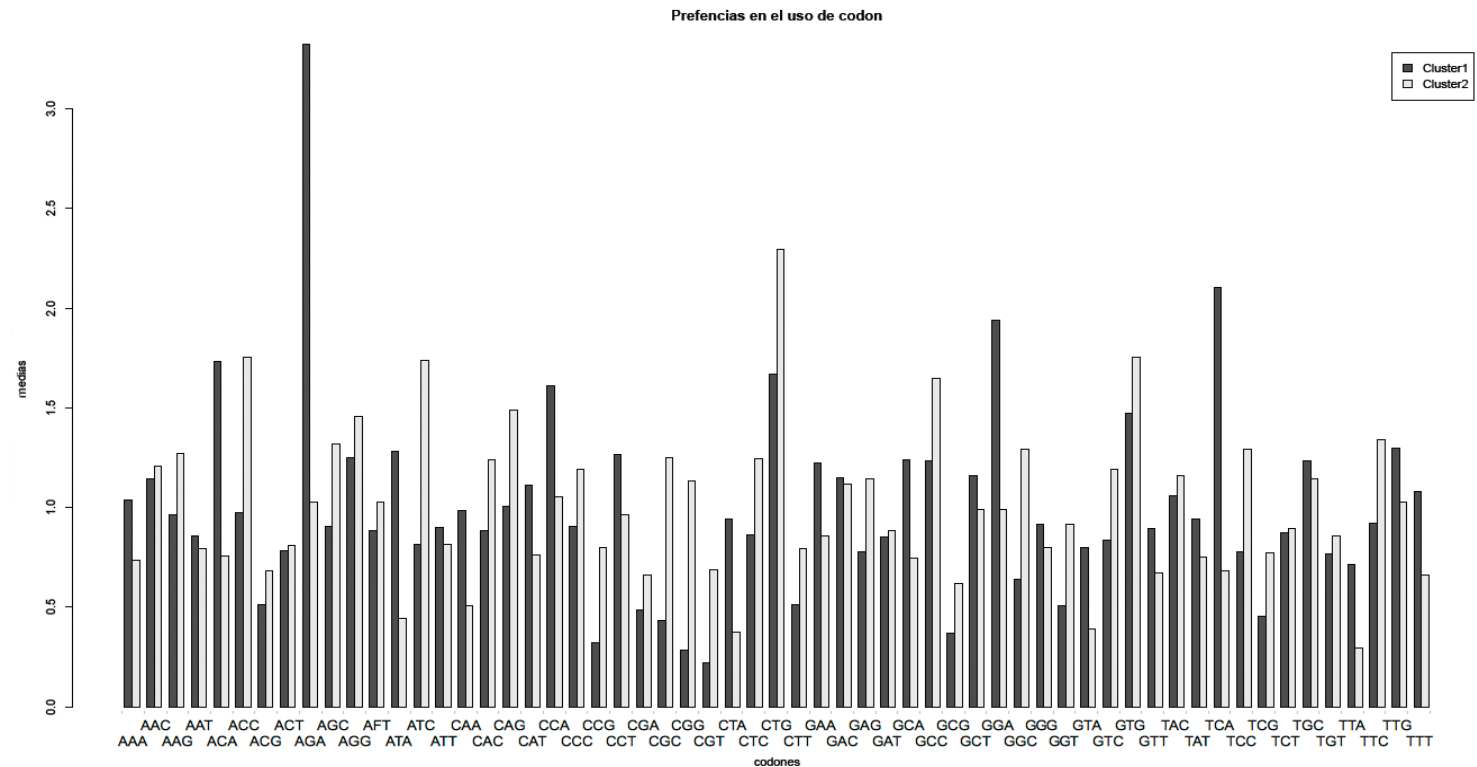


Gráfico 1. Medias para cada codón obtenidas por el método k-means. En el eje de las y son las medias de los codones, mientras que el eje de las x los codones analizados. En color gris oscuro el cluster 1, mientras que el gris claro el cluster 2.

ANÁLISIS DE RECOMBINACIÓN.

Los análisis de recombinación en los alineamientos, se realizaron con el objetivo de determinar la presencia de este evento evolutivo, el cual tiene implicaciones importantes en los análisis de evolución molecular. Posada y Crandall (2002) compararon diferentes métodos para la detección de recombinación y encontraron que la recombinación genera incongruencias en la reconstrucción filogenética, por ello las secuencias recombinantes deben ser removidas del análisis. Por otro lado Shriner *et al.*, (2003) mencionan que la presencia de recombinación en las secuencias sobreestima el número de sitios que evolucionan bajo selección positiva.

El análisis para detectar recombinación se llevó a cabo por medio de dos herramientas: GARD y SBP, las cuales están implementados en el servidor Datamonkey (www.datamonkey.org). Estos métodos detectan la presencia de recombinación al revelar

incongruencias en las ramas de árboles contruidos a partir de segmentos de un alineamiento.

Los resultados obtenidos del análisis se muestran en la tabla 8, donde se reportan los puntos de recombinación detectados por ambos métodos; así entonces para el DENV el método SBP determinó el punto 1832 dentro del dominio III, el otro punto encontrado por GARD fue el 1328 dentro del dominio I. Para el BVDV únicamente se encontró un punto por el método SBP el 3079 al final del dominio IIIa. Finalmente para el HCV se encontraron tres puntos, la posición 1596 por el método SBP que se encuentra próximo al final de la región hipervariable y las posiciones 1794 dentro del dominio I. Estos sitios también fueron mapeados en un representación lineal en nucleótidos de la proteína de envoltura, para como ser su localización física (figura 8) podemos observar que los puntos encontrados se encuentran próximos a las intersecciones entre dominios.

TABLA 23. RECOMBINACIÓN EN EL GEN DE LA PROTEÍNA DE ENVOLTURA.

Método	DENV	BVDV	HCV
SBP	1832	3079	1596
GARD	1328	No hay recombinación	1794

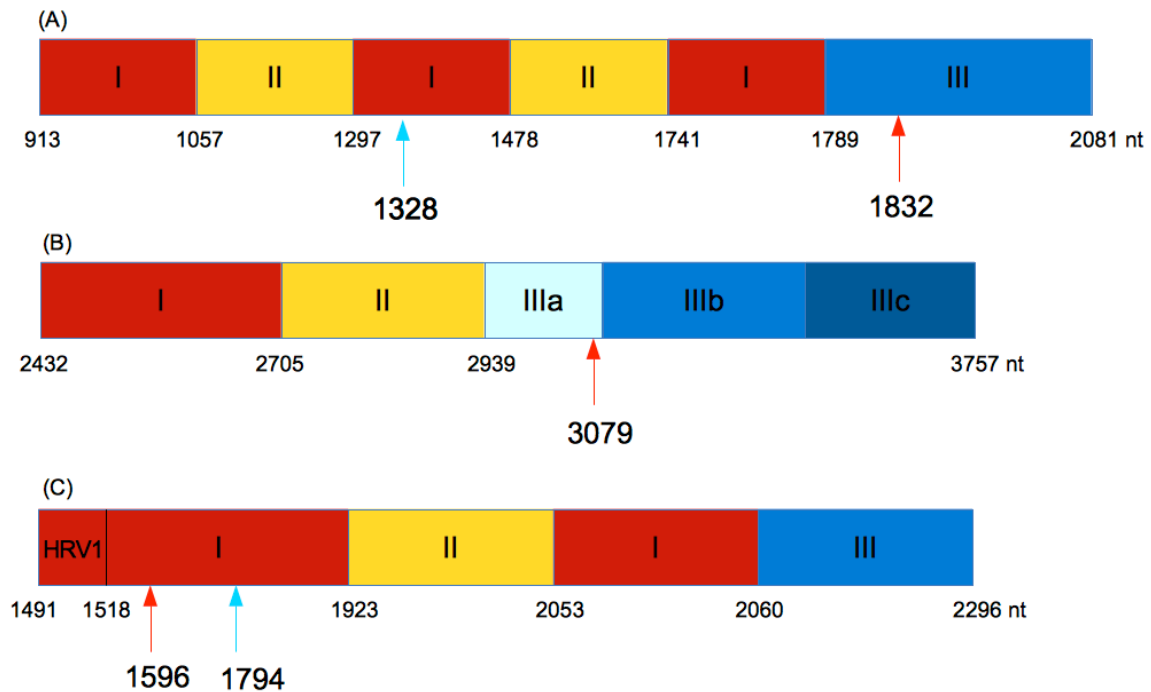


Figura 24. Sitios de recombinación dentro de la proteína de envoltura. Representación lineal de la proteína de envoltura de los virus DENV (A), BVDV (B), HCV(C), las flechas en rojo marcan la posición recombinante encontrada por el método SBP y la flecha en azul la posición recombinante encontrada por el método GARD.

ARBOLES FILOGENÉTICOS.

Árboles filogenéticos fueron construidos, por el método Bayesiano el objetivo de realizar estas filogenias fue utilizarlas posteriormente para llevar a cabo una reconstrucción de caracteres ancestrales por aminoácidos y por codones. En la figura 25 se muestra el árbol filogenético del DENV, donde se observan 4 clados que corresponden a los cuatro serotipos del virus, en las ramas internas del árbol se muestran los valores de probabilidad posterior, los cuales fueron iguales a 1, indicando que el árbol tiene buen soporte estadístico.

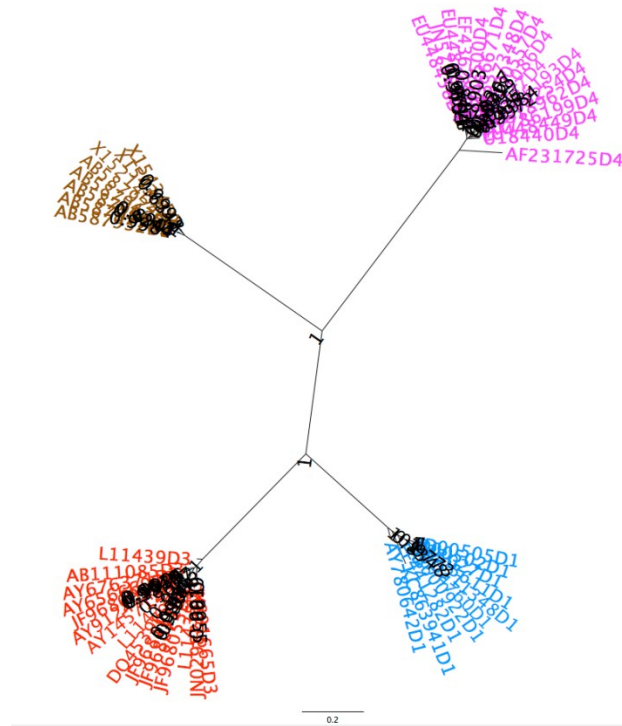


Figura 25. Árbol filogenético del DENV. Reconstrucción filogenética por el método Bayesiano, utilizado la secuencia completa de la proteína E de DENV en el programa Mr Bayes; las ramas que corresponden al serotipo 1 de dengue en color azul, las ramas que corresponden al serotipo dengue 2 en color café, el serotipo dengue 3 en color rojo y el serotipo dengue 4 en color rosa.

Para el BVDV, también se realizó la reconstrucción filogenética, en la figura 26 se observan dos grandes clados, estos corresponden a los dos genotipos descritos para este virus en color verde el genotipo 1 y en color naranja el genotipo 2, al interior de las ramas se muestra la probabilidad posterior con valores mayores a 0.8 indicando que el árbol tiene buen soporte estadístico.

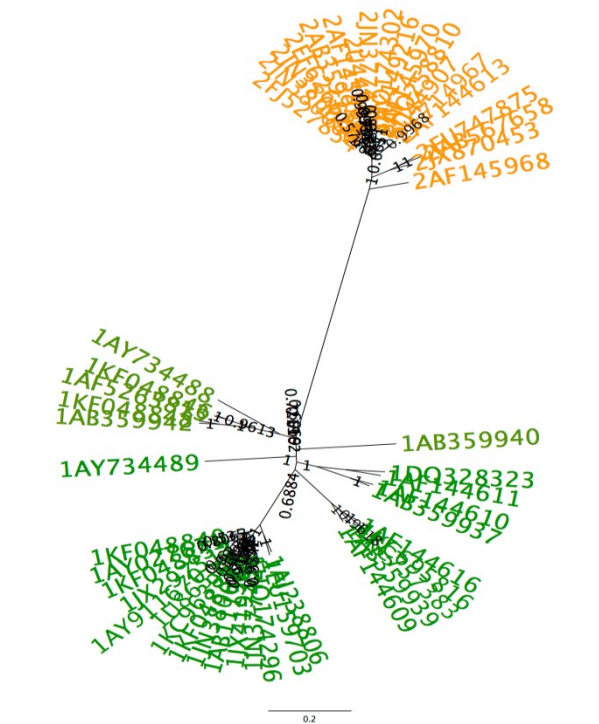


Figura 26. Árbol filogenético del BVDV. Reconstrucción filogenética por el método Bayesiano, utilizado la secuencia completa de la proteína E2 de BVDV en el programa Mr Bayes; las ramas en color verde corresponde al genotipo 1 del BVDV y el color naranja el genotipo 2 del BVDV.

Finalmente se realizó el árbol filogenético del HCV, en este virus se han descrito 7 genotipos, y dentro de cada genotipo existen diferentes subgenotipos. El árbol obtenido se observa en la figura 27, donde las secuencias se clasificaron en 7 clados igual a lo reportado en la literatura. En las ramas internas del árbol se muestran los valores obtenidos de probabilidad posterior cuales son mayores de 0.58 que tiene un soporte estadístico aceptable.



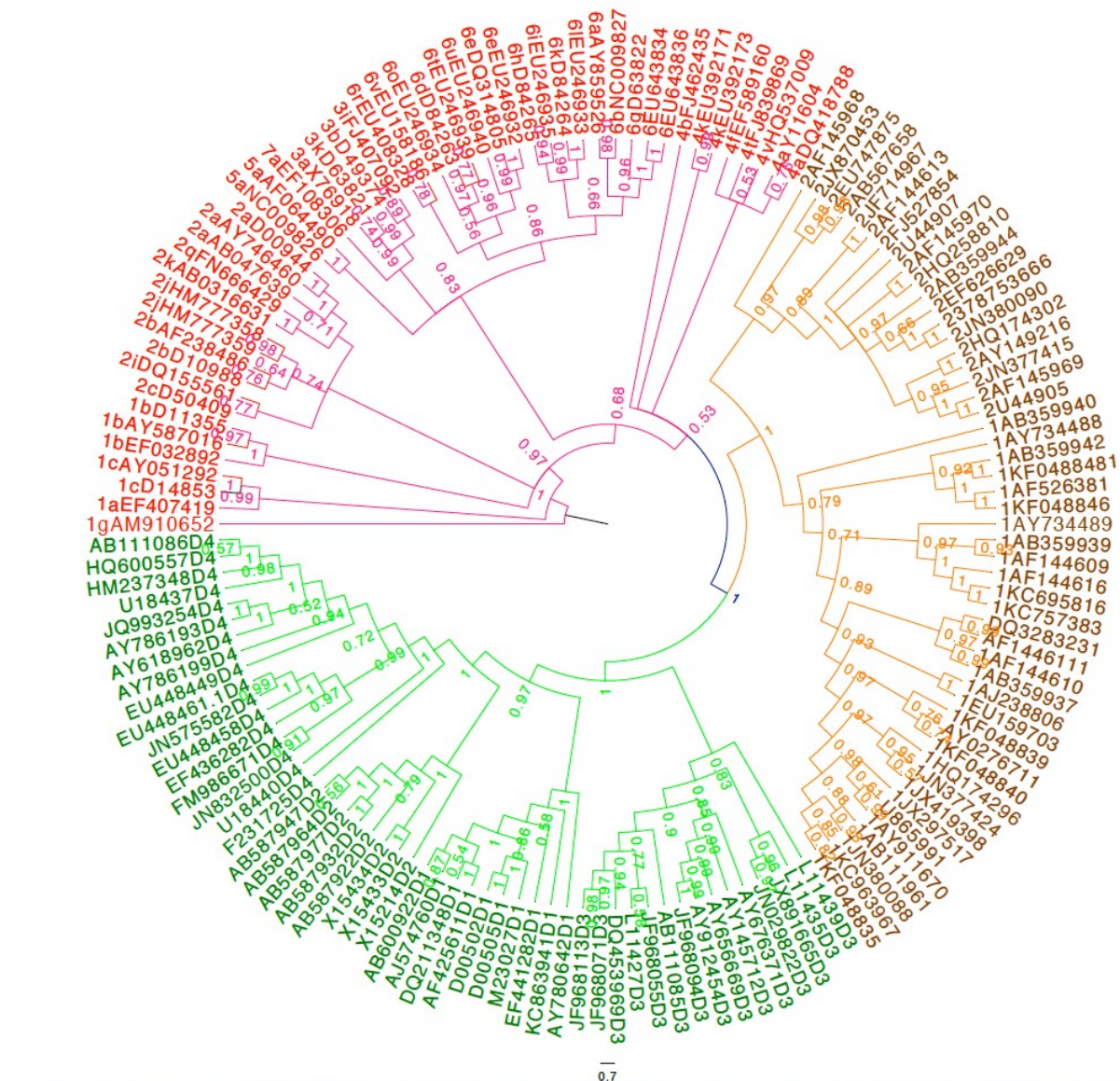


Figura 28. Árbol filogenético de la familia *Flaviviridae*. Reconstrucción filogenética por el método Bayesiano, utilizado la secuencia completa de las proteínas de envoltura de los virus DENV, BVDV, y HCV en el programa Mr Bayes. El HCV en color rojo, el BVDV virus en color café y el DENV en color verde.

Capítulo 9

Referencias

- Acosta C., y Gómez I. (2005). Biología y métodos de diagnóstico del dengue. *Revista de biomedicina*, 16:1134-137.
- Ashkenazy, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., y Pupko, T. (2012). FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic acids research*, 40(W1), W580-W584.
- Becher P., Orlich M., y Thiel, HJ. (2001). RNA recombination between persisting pestivirus and a vaccine strain: generation of cytopathogenic virus and induction of lethal disease. *Journal Virology* 75:6356-6264.
- Becher, P., y Tautz, N. (2011). RNA recombination in pestiviruses. *RNA biology*, 8(2), 216-224.
- Behura, S. K., y Severson, D. W. (2013). Nucleotide substitutions in dengue virus serotypes from Asian and American countries: insights into intracodon recombination and purifying selection. *BMC microbiology*, 13(1), 37.
- 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.
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic acids research*, gks1195
- Bhattacharya, D., & Cheng, J. (2013). 3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization. *Proteins: Structure, Function, and Bioinformatics*, 81(1), 119-131
- Brown, R. J., Juttla, V. S., Tarr, A. W., Finnis, R., Irving, W. L., Hemsley, S., y Ball, J. K. (2005). Evolutionary dynamics of hepatitis C virus envelope genes during chronic infection. *Journal of general virology*, 86(7), 1931-1942.
- Chao, D. Y., King, C. C., Wang, W. K., Chen, W. J., Wu, H. L., y Chang, G. J. (2005). Strategically examining the full-genome of dengue virus type 3 in clinical isolates reveals its mutation spectra. *Virology Journal*, 2, 72.
- Chen J, Liu H, Yang J, Chou K (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* 33: 423-428.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J. Richardson, D. C. (2009). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1), 12-21.

- Ciota A., y Kramer L. 2010. Insights into Arbovirus Evolution and Adaptation from Experimental Studies. *Viruses*. 2: 2594-2617.
- Clyde K., Kyle J., y Harris E. (2005). Recent Advances in Deciphering Viral and Host Determinants of Dengue Virus Replication and Pathogenesis. *Journal of Virology*. 80 (23): 114178-11431.
- Codran, A., Royer, C., Jaeck, D., Bastien-Valle, M., Baumert, T. F., Kieny, M. P., y Martin, J. P. (2006). Entry of hepatitis C virus pseudotypes into primary human hepatocytes by clathrin-dependent endocytosis. *Journal of General Virology*, 87(9), 2583-2593.
- Conant, G. C., Wagner, G. P., & Stadler, P. F. (2007). Modeling amino acid substitution patterns in orthologous and paralogous genes. *Molecular phylogenetics and evolution*, 42(2), 298-307.
- Cristina J., y Rodney C., (2006). Evidence of structural genomic region recombination in Hepatitis C. *Virology Journal*. 3:1-8.
- Cuadras C., (2014) Nuevos métodos de Análisis multivariante. CMC Editions. Barcelona.
- Curran, R., Jameson, C. L., Craggs, J. K., Grabowska, A. M., Thomson, B. J., Robins, A., y Ball, J. K. (2002). Evolutionary trends of the first hypervariable region of the hepatitis C virus E2 protein in individuals with differing liver disease severity. *Journal of general virology*, 83(1), 11-23.
- Delos, S.E., Gilbert, J.M., y White, J.M. (2000). The central proline of an internal viral fusion peptide serves two important roles. *Journal Virology*., 74, 1686-1693.
- Dimmock N.J. Earton A.J., Leppard KN. (2007). Introduction to Modern Virology. Sexta edición. Blacwell Publishing. 2714-273.
- Duffy, S., Shackelton, L. A., y Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics*, 9(4), 267-276.
- El Omari, K., Iourin, O., Harlos, K., Grimes, J.M., Stuart, D.I., (2013). Structure of a pestivirus envelope glycoprotein E2 clarifies its role in cell entry. *Cell Rep*. 3, 3035. doi:10.1016/j.celrep.2012.12.001.
- Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J. C., y Alter, H. J. (2000). The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science*, 288(5464), 339-344.
- Gadkari, R. A., y Srinivasan, N. (2010). Prediction of protein-protein interactions in dengue virus coat proteins guided by low resolution cryoEM structures. *BMC structural biology*, 10(1), 17.
- Garry, R.F., Dash, S., (2003). Proteomics computational analyses suggest that hepatitis C virus E1 and pestivirus E2 envelope glycoproteins are truncated class II fusion proteins. *Virology* 307, 255-265.
- Ghai, R. R., Sibley, S. D., Lauck, M., Dinis, J. M., Bailey, A. L., Chapman, C. A., y Goldberg, T. L. (2013). Deep sequencing identifies two genotypes and high viral genetic diversity of human pegivirus (GB virus C) in rural Ugandan patients. *Journal of General Virology*, 94(Pt 12), 2670-2678.

- Goncalvez, A. P., Escalante, A. A., Pujol, F. H., Ludert, J. E., Tovar, D., Salas, R. A., y Liprandi, F. (2002). Diversity and evolution of the envelope gene of dengue virus type 1. *Virology*, 303(1), 110-119.
- Gouy, M., Guindon, S., y Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*, 27(2), 221-224.
- Graur, D., y W-H. Li, (2000). Fundamentals of Molecular Evolution. USA Second Edition.
- Gray, R. R., Parker, J., Lemey, P., Salemi, M., Katzourakis, A., y Pybus, O. G. (2011). The mode and tempo of hepatitis C virus evolution within and among hosts. *BMC evolutionary biology*, 11(1), 131.
- Holmes, E. C. (2003). Patterns of intra-and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. *Journal of virology*, 77(20), 11296-11298.
- Holmes, E. C. (2007). The evolution and epidemiology of dengue virus. *Salud Pública de México*, 49, 296-300.
- Holmes, E. C., Worobey, M., y Rambaut, A. (1999). Phylogenetic evidence for recombination in dengue virus. *Molecular biology and evolution*, 16(3), 405-409.
- Kapoor, A., Simmonds, P., Cullen, J.M., Scheel, T.K.H., Medina, J.L., Giannitti, F., Nishiuchi, E., Brock, K.V., Burbelo, P.D., Rice, C.M., Lipkin, W.I., 2013. Identification of a Pegivirus (GB VirusLike Virus) That Infects Horses. *Journal Virology*. 87, 7185-7190.
- Karlin, S. A. M. U. E. L., Blaisdell, B. E., y Schachtel, G. A. (1990). Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: data and hypotheses. *Journal of virology*, 64(9), 4264-4273.
- Katoh, K., Asimenos, G., y Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 39-64). Humana Press.
- Kelley, L. A., & Sternberg, M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols*, 4(3), 363-371.
- Klein, D. E., Choi, J. L., y Harrison, S. C. (2013). Structure of a dengue virus envelope protein late-stage fusion intermediate. *Journal of virology*, 87(4), 2287-2293.
- Kong, L., Giang, E., Nieusma, T., Kadam, R. U., Cogburn, K. E., Hua, Y., y Law, M. (2013). Hepatitis C virus E2 envelope glycoprotein core structure. *Science*, 342(6162), 1090-1094.
- Krey, T., d'Alayer, J., Kikuti, C.M., Saulnier, A., Damier-Piolle, L., Petitpas, I., Johansson, D.X., Tawar, R.G., Baron, B., Robert, B., England, P., Persson, M.A.A., Martin, A., Rey, F.A., 2010. The Disulfide Bonds in Glycoprotein E2 of Hepatitis C Virus Reveal the Tertiary Organization of the Molecule. *PLoS Pathog* 6, e1000762.
- Kuiken, C., Yusim, K., Boykin, L., & Richardson, R. (2005). The Los Alamos hepatitis C sequence database. *Bioinformatics*, 21(3), 379-384.
- Kumar, S. R., Patil, J. A., Cecilia, D., Cherian, S. S., Barde, P. V., Walimbe, A. M., y

- Mourya, D. T. (2010). Evolution, dispersal and replacement of American genotype dengue type 2 viruses in India (1956–2005): selection pressure and molecular clock analyses. *Journal of General Virology*, 91(3), 707-720.
- Kuno, G., Chang, G. J. J., Tsuchiya, K. R., Karabatsos, N., y Cropp, C. B. (1998). Phylogeny of the genus *Flavivirus*. *Journal of virology*, 72(1), 73-83.
- Lauring, A. S., Frydman, J., y Andino, R. (2013). The role of mutational robustness in RNA virus evolution. *Nature Reviews Microbiology*, 11(5), 327-336.
- Lecot, S., Belouzard, S., Dubuisson, J., y Rouillé, Y. (2005). Bovine viral diarrhea virus entry is dependent on clathrin-mediated endocytosis. *Journal of virology*, 79(16), 10826-10829.
- Lemey P., y Posada D. 2005. Introduction to recombination detection. En Lemey P., Salemi M., y Vandamme A. (Eds.) *The Phylogenetic Handbook*. Camdridge. United Kingdoom. pp. 407-4418.
- Leyssen, P., De Clercq, E., Neyts, J., (2000). Perspectives for the treatment of infections with Flaviviridae. *Clin. Microbiol. Rev.* 13, 67-82.
- Li, Y., Modis, Y. 2014. A novel membrane fusion protein family in Flaviviridae?. *Trends Microbiol.*
- Li, Y., Wang, J., Kanai, R., & Modis, Y. (2013). Crystal structure of glycoprotein E2 from bovine viral diarrhea virus. *Proceedings of the National Academy of Sciences*, 110(17), 6805-6810.
- Li, Y., Wang, J., Kanai, R., y Modis, Y. (2013). Crystal structure of glycoprotein E2 from bovine viral diarrhea virus. *Proceedings of the National Academy of Sciences*, 110(17), 6805-6810.
- Librado, P., y Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451-1452.
- Lindenbach B D, Thiel H J, and Rice C M. 2007. *Flaviviridae: The Virus and Their Replication*. Fourth ed. Lippincott William Wilkins.
- Lobo, F. P., Mota, B. E., Pena, S. D., Azevedo, V., Macedo, A. M., Tauch, A., y Franco, G. R. (2009). Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PloS one*, 4(7), e6282.
- Ma, M. R., Ha, X. Q., Ling, H., Wang, M. L., Zhang, F. X., Zhang, S. D. Yan, W. (2011). The characteristics of the synonymous codon usage in hepatitis B virus and the effects of host on the virus in codon usage pattern. *Virology journal*, 8(1), 1-10.
- Maddison, W. P., y Maddison, D. R. (1989). MacClade: analysis of phylogeny and character evolution. *Evolution (PMBD)*, 18590847
- McAllister, J., Casino, C., Davidson, F., Power, J., Lawlor, E., Yap, P. L., y Smith, D. (1998). Long-term evolution of the hypervariable region of hepatitis C virus in a common-source-infected cohort. *Journal of virology*, 72(6), 4893-4905.
- McCaffrey, K., Boo, I., Tewierek, K., Edmunds, M. L., Pournbourios, P., & Drummer, H.

- E. (2012). Role of conserved cysteine residues in hepatitis C virus glycoprotein e2 folding and function. *Journal of virology*, 86(7), 3961-3974.
- Modis, Y., (2014). Relating structure to evolution in class II viral membrane fusion proteins. *Current. Opinion. Virology*. 5C: 3441.
- Modis, Y., Ogata, S., Clements, D., & Harrison, S. C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature*, 427(6972), 313-319.
- Moelbert, S., Emberly, E., y Tang, C. (2004). Correlation between sequence hydrophobicity and surface-exposure pattern of database proteins. *Protein Science*, 13(3), 752-762.
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Pond, S. L. K. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS genetics*, 8(7), e1002764.
- Nakamura, Y., Gojobori, T., & Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic acids research*, 28(1), 292-292.
- Nayak, V., Dessau, M., Kucera, K., Anthony, K., Ledizet, M., & Modis, Y. (2009). Crystal structure of dengue virus type 1 envelope protein in the postfusion conformation and its implications for membrane fusion. *Journal of virology*, 83(9), 4338-4344.
- Nehete, J. Y., Bhambar, R. S., Narkhede, M. R., & Gawali, S. R. (2013). Natural proteins: Sources, isolation, characterization and applications. *Pharmacognosy reviews*, 7(14), 107.
- Neill, J. D., Newcomer, B. W., Marley, S. D., Ridpath, J. F., y Givens, M. D. (2012). Greater numbers of nucleotide substitutions are introduced into the genomic RNA of bovine viral diarrhea virus during acute infections of pregnant cattle than of non-pregnant cattle. *Virology Journal*, 9, 150.
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review Genetics*, 39, 197-218.
- Nybekken G., Nelson C., Chen B., Diamond M., y Fremont D. 2006. Crystal Structure of the West Nile Virus Envelope Glycoprotein. *Journal of Virology*. 80: 11467-11474.
- Owsianka, A., Clayton, R. F., Loomis-Price, L. D., McKeating, J. A., y Patel, A. H. (2001). Functional analysis of hepatitis C virus E2 glycoproteins and virus-like particles reveals structural dissimilarities between different forms of E2. *Journal of General Virology*, 82(8), 1877-1883.
- Peña, C. (2011). Métodos de inferencia filogenética. *Revista Peruana de Biología*, 18(2), 265-267.
- Perera, R., y Kuhn, R. J. (2008). Structural proteomics of dengue virus. *Current opinion in microbiology*, 11(4), 369-377.
- Perfectti, F., Picó, F. X., y Gómez, J. M. (2009). La huella genética de la selección natural. *Revista Ecosistemas*, 18(1).

- 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. *Journal of computational chemistry*, 25(13), 1605-1612.
- Pond, S. L. K., Frost, S. D., Grossman, Z., Gravenor, M. B., Richman, D. D., y Brown, A. J. L. (2006). Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS computational biology*, 2(6), e62.
- Pond, S. L. K., Posada, D., Gravenor, M. B., Woelk, C. H., y Frost, S. D. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular biology and evolution*, 23(10), 1891-1901.
- Pond, S. L. K., Scheffler, K., Gravenor, M. B., Poon, A. F., y Frost, S. D. (2010). Evolutionary fingerprinting of genes. *Molecular biology and evolution*, 27(3), 520-536.
- Pond, S. L. K., y Frost, S. D. (2005a). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution*, 22(5), 1208-1222.
- Pond, S. L. K., y Frost, S. D. (2005b). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*, 21(10), 2531-2533.
- Pond, S. L.K, A.F.Y. Poon, y W. Frost, S.D. (2009) Estimating selection pressures on alignments of coding sequences, second edition ed. New York: Cambridge University Press, pp 419-490.
- Pybus O. y Shapiro B. 2009. Natural selection and adaptation of molecular sequences. En Lamey P., Salemi M., y Vandamme A. (Eds.) The Phylogenetic Handbook. Cambridge. United Kingdom. pp. 407-418.
- Qiang C., Zheng N., Guo J., y Long Yon. 2007. Evidence of natural recombination in classical swine fever virus. *Virus Research*. 1(126)179-185.
- Qu, Y., Zhang, R., Cui, P., Song, G., Duan, Z., y Lei, F. (2011). Evolutionary genomics of the pandemic 2009 H1N1 influenza viruses (pH1N 1v). *Virology Journal*, 8, 250.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Ramirez G., Díaz A., Camacho M., Cisneros A., y Muñoz M.L. 2009. Multiple recombinants in two dengue virus, serotype-2 isolates from patients from Oaxaca, Mexico. *BMC Microbiology*. 9:14-12.
- Reuter G, Pankovics P, Szucs G (2008) Genetic drift of norovirus genotype GII- 4 in seven consecutive epidemic seasons in Hungary. *Journal Clinic Virology*, 42: 135–140.
- Risatti, G. R., Holinka, L. G., Carrillo, C., Kutish, G. F., Lu, Z., Tulman, E. R., y Borca, M. V. (2006). Identification of a novel virulence determinant within the E2 structural glycoprotein of classical swine fever virus. *Virology*, 355(1), 94-101.
- Risatti, G. R., L. G. Holinka, I. Fernandez Sainz, C. Carrillo, G. F. Kutish, Z. Lu, J. Zhu, D. L. Rock, and M. V. Borca. (2007). Mutations in the carboxyl terminal region of E2

- glycoprotein of classical swine fever virus are responsible for viral attenuation in swine. *Virology*, 364(2), 371-382.
- Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), 1572-1574.
- Rota PA, Bloom AE, Vanchiere JA, Bellini WJ (1994) Evolution of the nucleoprotein and matrix genes of wild-type strains of measles virus isolated from recent epidemics. *Virology* 198: 724–730.
- Saha, S., & Raghava, G. P. (2007). Prediction methods for B-cell epitopes. In *Immunoinformatics* (pp. 387-394). Humana Press.
- Salemi, M., Lemey, P., y Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press.
- Sharp, P. M., y Li, W. H. (1987). The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Molecular Biology and Evolution*, 4(3), 222-230.
- Sheridan, I., Pybus, O. G., Holmes, E. C., y Klennerman, P. (2004). High-resolution phylogenetic analysis of hepatitis C virus adaptation and its relationship to disease progression. *Journal of virology*, 78(7), 3447-3454.
- Simmonds, P. (2004). Genetic diversity and evolution of hepatitis C virus—15 years on. *Journal of General Virology*, 85(11), 3173-3188.
- Simon-Loriere E. y E.Holmes. (2012). Why do RNA viruses recombine?. *Natural Review Microbiology*. 8:617-626.
- Singh, H., & Raghava, G. P. S. (2001). ProPred: prediction of HLA-DR binding sites. *Bioinformatics*, 17(12), 1236-1237.
- Sowmya, G., Shamini, G., Anita, S., Sakharkar, M., Mathura, V., Rodriguez, H., & Shapshak, P. (2011). HIV-1 envelope accessible surface and polarity: clade, blood, and brain. *Bioinformation*, 6(2), 48.
- Stapleton, J.T., Fong, S., Muerhoff, A.S., Bukh, J., Simmonds, P., 2011. The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae. *Journal General Virology*. 92, 233-246.
- Studer, R., y Robinson-Rechavi, M. (2009). Evidence for an episodic model of protein sequence evolution. *Biochemical Society Transactions*, 37(4), 783.
- Suzuki, R., y Shimodaira, H. (2006). Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), 1540-1542.
- Tang F., y C. Zhang. 2007. Evidence for positive selection on the E2 gene of bovine viral diarrhea virus type 1. *Virus Genes*.35:629-634.
- Tolou HJG, Couissinier-Paris GP, Durand JP, Mercier V, dePina JJ, de-Micco P, Billoir F, Charrel RN, de-Lamballerie X. (2001) Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. *Journal Gen* 82:1283-1290.

- Torres-Galicia, I., Cortés-Poza, D., y Becker, I. (2014). Dengue en México: análisis de dos décadas. *Gaceta médica de Mexico*, 150, 122-7.
- Twiddy SS, Holmes EC, Rambaut A. (2003). Inferring the rate and time-scale of dengue virus evolution. *Molecular Biology and Evolution*. 20: 122-129.
- Twiddy, S. S., Farrar, J. J., Vinh Chau, N., Wills, B., Gould, E. A., Gritsun, T., y Holmes, E. C. (2002). Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology*, 298(1), 63-72.
- Vargas, D. S., Jaime, J., y Vera, V. J. (2009). Perspectivas para el control del Virus de la Diarrea Viral Bovina (BVDV). *Revista Colombiana de Ciencias Pecuarias*, 22(4), 677-688.
- Venables, W., y Ripley, B. D. (2000). *S programming*. Springer.301-330.
- Vera, D. L. L., Juárez, N. J., Díaz, G. M., Méndez, N. J., Chirino, S. R., Dehesa, V. M., y Bazán, P. C. (2005). Epidemiologic and situational panorama of hepatitis C in Mexico. *Rev Gastroenterology Mexico*, 70, 25-32.
- Voronin Y., Holte S., Overbaugh., y Emerman. (2009). Genetic Drift of HIV populations in culture. *PLoS Genetics*. 5(3): e1000431.
- Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *Journal of molecular graphics*, 8(1), 52-56
- Wang WK, Sung TL, Lee CN, Lin TY, King CC (2002) Sequence diversity of the capsid gene and the nonstructural gene NS2B of dengue-3 virus in vivo.
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., y Barton, G. J. (2009). Jalview Version 2a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189-1191.
- Weaver, S. C., y Vasilakis, N. (2009). Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. *Infection, Genetics and Evolution*, 9(4), 523-540.
- Woelk, C. H., y Holmes, E. C. (2002). Reduced positive selection in vector-borne RNA viruses. *Molecular biology and evolution*, 19(12), 2333-2336.
- Worobey, M., A. Rambaut, y E. C. Holmes. (1999). Widespread intraserotype recombination in natural populations of dengue virus. *Proc. Natl. Acad. Sci. USA* 96:7352-7357.
- Worobey, M., y Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *Journal of General Virology*, 80(10), 2535-2543.
- Yagnik, A. T., Lahm, A., Meola, A., Roccasecca, R. M., Ercole, B. B., Nicosia, A., y Tramontano, A. (2000). A model for the hepatitis C virus envelope glycoprotein E2. *Proteins: Structure, Function, and Bioinformatics*, 40(3), 355-366.
- Yamaguchi-Kabata, Y., & Gojobori, T. (2000). Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes. *Journal of Virology*, 74(9), 4335-4350.

- Zanotto PM, Gould EA, Gao GF, Harvey PH, Holmes EC. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc Natl Acad Sci USA* 93: 548-553.
- Zárata, S., y Novella, I. S. (2004). Vesicular stomatitis virus evolution during alternation between persistent infection in insect cells and acute infection in mammalian cells is dominated by the persistence phase. *Journal of virology*, 78(22), 12236-1224.
- Zhang, Y., Zhang, W., Ogata, S., Clements, D., Strauss, J. H., Baker, T. S., Rossmann, M. G. (2004). Conformational changes of the flavivirus E glycoprotein. *Structure*, 12(9), 1607-1618.

