

DEPARTAMENTO DE GENÉTICA EVOLUTIVA

EVOLUCIÓN DEL VIRUS DE LA HEPATITIS C EN
MUESTRAS HOSPITALARIAS DE LA COMUNIDAD
VALENCIANA

NURIA JIMÉNEZ HERNÁNDEZ

UNIVERSITAT DE VALENCIA
Servei de Publicacions
2005

Aquesta Tesi Doctoral va ser presentada a València el dia 27 de Julio de 2004 davant un tribunal format per:

- D. Juan del Olmo Puchalt
- D^a. Amparo Latorre Castillo
- D. Leandro Meandro Soria
- D. Julio Rozas Ciras
- D^a. Pedro Moreno Gómez

Va ser dirigida per:
D. Andrés Moya Simarro

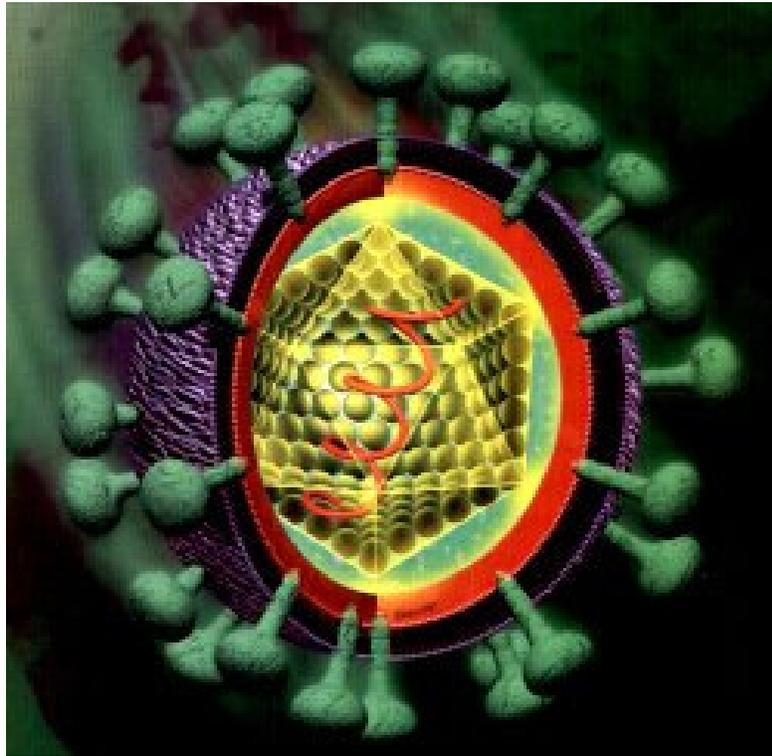
©Copyright: Servei de Publicacions
Nuria Jiménez Hernández

Depòsit legal:

I.S.B.N.:84-370-6129-6

Edita: Universitat de València
Servei de Publicacions
C/ Artes Gráficas, 13 bajo
46010 València
Spain
Telèfon: 963864115

Evolución del virus de la hepatitis C en muestras hospitalarias de la Comunidad Valenciana



Nuria Jiménez Hernández
Tesis doctoral 2004

VNIVERSITAT  VALÈNCIA

VNIVERSITAT  VALÈNCIA

FACULTAT DE CIÈNCIES BIOLÒGIQUES

DEPARTAMENT DE GENÈTICA

INSTITUT CAVANILLES DE BIODIVERSITAT I BIOLOGIA EVOLUTIVA



Evolución del virus de la hepatitis C en muestras hospitalarias de la Comunidad Valenciana

Memoria presentada por Nuria Jiménez Hernández para optar al grado de doctora en Ciencias Biológicas por la Universitat de València.

Valencia 2004

D. Andrés Moya Simarro, Doctor en Ciencias Biológicas y Catedrático del Departament de Genètica de la Universitat de València.

CERTIFICA: Que Nuria Jiménez Hernández, Licenciada en Ciencias Biológicas por la Universitat de València, ha realizado bajo su dirección el trabajo que lleva por título: “Evolución del virus de la hepatitis C en muestras hospitalarias de la Comunidad Valenciana”, para optar al Grado de doctora en Ciencias Biológicas por la Universitat de València.

Y para que conste, en el cumplimiento de la legislación vigente, firmo el presente certificado en Valencia, a 14 de Junio de 2004.

Fdo.: Dr. Andrés Moya Simarro

ÍNDICE

I. Introducción	1
1. Los virus	3
2. El origen de los virus	6
3. Los virus de RNA	7
4. Fuerzas evolutivas	9
5. Virus de RNA y la teoría de las cuasispecies	12
6. Filogenias moleculares y poblaciones virales	13
7. La hepatitis viral	15
8. El virus de la hepatitis C	18
8.1 Epidemiología y cuadro clínico	18
8.2 Estructura	20
8.3 Replicación	23
8.4 Heterogeneidad genética	25
8.5 Tratamiento frente al VHC	27
8.6 Mecanismos de persistencia viral	29
II. Objetivos	33
III. Material y métodos	37
1. Material	39
2. Esquema metodológico	43
3. Extracción viral de RNA y amplificación	44
3.1 Extracción de RNA	44
3.2 Transcripción inversa	44
3.3 PCR	45
3.4 Purificación de productos de PCR	47
4. Clonación y secuenciación	47
4.1 Ligación	48
4.2 Transformación	48
4.3 PCR de colonia	48
4.4 Secuenciación	51

5. Análisis de Secuencias	52
5.1 Alineamientos	52
5.2 Inferencia de la historia demográfica del VHC y cálculo del tamaño poblacional efectivo	53
5.3 Pruebas de neutralidad	58
5.3.1 Prueba de Tajima	59
5.3.2 Prueba de Fu y Li	61
5.3.3 Prueba de Fu	61
5.4 Análisis de polimorfismos	62
5.5 Análisis de la varianza y búsqueda de haplotipos compartidos	62
5.6 Detección de selección positiva	63
5.6.1 Búsqueda de selección positiva en grandes filogenias. Método aproximado II.	67
IV. Resultados	69
1. Análisis de polimorfismos de DNA	71
2. Análisis de la varianza y búsqueda de haplotipos compartidos	73
3. Historia demográfica	83
3.1 Historia demográfica de las poblaciones virales de los pacientes	86
3.2 Historia demográfica de los genotipos	89
3.3 Número reproductivo básico: R_0	92
4. Pruebas de neutralidad	97
5. Búsqueda de selección positiva	105
6. Cálculo del tamaño poblacional efectivo	120
V. Discusión	123
1. Indicadores de variación genética	125
2. La demografía de las poblaciones inferida por los genes	126
2.1 El problema de la historia demográfica a escala individual	126
2.2 Historia demográfica a nivel de genotipo	127
3. Las fuerzas que modulan la evolución de los virus de la hepatitis C	130
3.1. Detectando e interpretando la selección positiva	135
4. Tamaño poblacional efectivo	143

VI. Conclusiones	145
VII. Bibliografía	149
VIII. Anexos	167
Anexo I	169
- Estimaciones de Π	169
Anexo II	170
- Medidas de polimorfismo	170
Anexo III	173
- Espectro de frecuencias de sitios segregantes	173
Anexo IV	184
- Comparación de las medias de los parámetros empleados en el cálculo de la variación genética	184
Anexo V	186
- Descripción de los modelos demográficos implementados en GENIE V3.0	186
- Cálculo por máxima verosimilitud de los parámetros correspondientes a cada modelo demográfico. Estima paramétrica.	188
1. Análisis intrapaciente.	188
2. Análisis a nivel de genotipo	199
- Skylineplots. Estima no paramétrica de la historia demográfica a nivel de paciente	200
Anexo VI	208
- Árboles filogenéticos de las secuencias virales	208
Anexo VII	235
- Estimación de los parámetros correspondientes a los modelos de evolución de codones implementados por CODEML	235
- Comparación de modelos de evolución de codones	247
- Distribución de probabilidades a posteriori	259
IX. Apéndice	269
- Tipos de hepatitis	271
- Respuesta inmune	272

I. Introducción

1. LOS VIRUS

Los virus, con tamaños que apenas superan los 300 nm de tamaño y que necesitan de otros seres para reproducirse, han sido causantes de enfermedades de gran repercusión social y económica a lo largo de la historia, siendo el SIDA un ejemplo reciente de los estragos que pueden llegar a ocasionar. Ningún organismo queda fuera del alcance de estos agentes infecciosos, y su virulencia posee un rango bastante amplio de sintomatología, abarcando desde las del simple resfriado común hasta, por ejemplo, los terribles efectos causados por el virus ébola.



Dmitri Ivanovski,



Martinus Beijerinck

Desde su descubrimiento y posterior aislamiento entre 1892 y 1896, por Dmitri Ivanovski y Martinus Beijerinck, en experimentos con el virus del mosaico del tabaco, muchas han sido las especies de virus descubiertas hasta nuestros días.

La investigación del origen de los virus constituye una aventura difícil. A diferencia de lo que ocurre con plantas y animales, de los cuales se posee registro fósil más o menos abundante, la naturaleza de los virus los convierte prácticamente en invisibles en el registro geológico e histórico. Si además contamos con que tan solo disponemos de descripciones debidamente detalladas acerca de datos epidemiológicos y clínicos, relacionados con los virus de los últimos treinta años, parece difícil obtener evidencias sobre los tipos de virus que pudieron haber existido en el pasado. No obstante existen indicios que demuestran la existencia de los mismos a lo largo de la historia de la humanidad. En algunos esqueletos humanos del Neolítico y la Edad del bronce se han encontrado deformaciones que recuerdan a las producidas por el virus de la polio. En restos momificados del antiguo Egipto pueden reconocerse lesiones de la piel seguramente debidas al virus de la viruela, incluyendo la de Ramses V que probablemente murió en 1157 a.C. a causa de esta enfermedad. También Tucídides describe una severa epidemia en la Antigua Grecia, en el año 430 AC. Pero como

mencionábamos anteriormente, encontrar tales indicios en el registro histórico es bastante inusual.

En su forma más simple los virus constan fundamentalmente de dos componentes: El ácido nucleico y la cubierta proteica o cápsida. El material genético del virus puede consistir en DNA o RNA, que a su vez puede ser de cadena sencilla (sc) o doble (dc), lineal o circular. Además puede estar constituido por una sola molécula o por varios segmentos del ácido nucleico. El genoma de RNAsc puede ser de orientación positiva (+), lo cual quiere decir que sirve como RNA mensajero, o de orientación negativa (-), en cuyo caso necesita de una RNA transcriptasa viral para sintetizar la cadena complementaria que sirva como mensajero. La cápsida protege al genoma vírico de las nucleasas, y también está implicada en el reconocimiento celular durante la infección. El ensamblaje de la cápsida puede seguir dos patrones: helicoidal, donde las subunidades proteicas y el ácido nucleico se disponen en una hélice, o icosaédrica, donde las subunidades proteicas se ensamblan de forma simétrica formando un armazón que cubre al ácido nucleico. En algunos virus rodeando a esta estructura, también encontramos una envuelta, en ocasiones derivada de las membranas celulares modificadas del hospedador. Esta envuelta presenta proteínas glicosiladas transmembrana codificadas por el virus.

Las primeras clasificaciones virales estaban basadas en las enfermedades causadas por los mismos, teniendo en cuenta el tejido que infectaban, cómo lo infectaban y cómo se transmitían. En la actualidad los virus han sido clasificados dentro de jerarquías por el ICTV (*International Committee on Taxonomy of Viruses*) y grupos especializados (Tabla I.1). La jerarquía de clasificación de los virus procede inclusivamente de la forma siguiente:

Orden → Grupo → Familia → Subfamilia → Género → Especie →
Subespecie → Cepa → Variante → Aislado

Esta clasificación está basada, además de en propiedades físicas del virus, tales como el tamaño y la forma, en propiedades que pueden afectar al modo de replicación tales como si el genoma consiste en una sola molécula o es segmentado, o si el genoma es de DNA o RNA de cadena sencilla o doble. Además, también se considera el lugar

<p>Caudovirales</p> <p><i>Myoviridae</i> <i>Siphoviridae</i> <i>Podoviridae</i> <i>Tectiviridae</i> <i>Corticoviridae</i> <i>Plasmaviridae</i> <i>Lipothrixviridae</i> <i>Rudiviridae</i> <i>Fuselloviridae</i> <i>Poxviridae</i> <i>Chordopoxvirinae</i> <i>Entomopoxvirinae</i></p> <p><i>Asfarviridae</i> <i>Iridoviridae</i> <i>Phycodnaviridae</i> <i>Baculoviridae</i> <i>Herpesviridae</i> <i>Alphaherpesvirinae</i> <i>Betaherpesvirinae</i> <i>Gammaherpesvirinae</i></p> <p><i>Adenoviridae</i> <i>Polyomaviridae</i> <i>Papillomaviridae</i> <i>Polydnaviridae</i> <i>Ascoviridae</i></p>	<p>Mononegavirales</p> <p><i>Bornaviridae</i> <i>Filoviridae</i> <i>Paramyxoviridae</i> <i>Paramyxovirinae</i> <i>Pneumovirinae</i> <i>Rhabdoviridae</i> <i>Orthomyxoviridae</i> <i>Bunyaviridae</i> <i>Arenaviridae</i> Deltavirus Virus de la hepatitis D</p>
<p><i>Inoviridae</i> <i>Microviridae</i> <i>Geminiviridae</i> <i>Circoviridae</i> <i>Parvoviridae</i> <i>Parvovirinae</i> <i>Densovirinae</i></p>	<p><i>Narnaviridae</i> <i>Leviviridae</i> <i>Picornaviridae</i> Hepatovirus Virus de la hepatitis A <i>Sequiviridae</i> <i>Comoviridae</i> <i>Potyviridae</i> <i>Caliciviridae</i> "Hepatitis E – like viruses" Virus de la hepatitis E <i>Astroviridae</i> <i>Nodaviridae</i> <i>Tetraviridae</i> <i>Luteoviridae</i> <i>Tombusviridae</i></p> <p>Nidovirales</p> <p><i>Coronaviridae</i> <i>Arteriviridae</i> <i>Flaviviridae</i> Flavivirus Virus de la fiebre Amarilla Pestivirus Virus de la diarrea bovina Hepacivirus Virus de la hepatitis C <i>Togaviridae</i> <i>Bromoviridae</i> <i>Closteroviridae</i> <i>Barnaviridae</i></p>
<p><i>Hepadnaviridae</i> Orthohepadnavirus V. hepatitis B <i>Caulimoviridae</i> <i>Pseudoviridae</i> <i>Metaviridae</i> <i>Retroviridae</i></p>	
<p><i>Cystoviridae</i> <i>Reoviridae</i> <i>Birnaviridae</i> <i>Totiviridae</i> <i>Partitiviridae</i> <i>Hypoviridae</i></p>	

Tabla I.1. Clasificación de los virus según el ICTV. Por razones de espacio sólo se indica el nombre de: grupo (azul), familias (negro) y subfamilias. Sólo en caso de los virus de Hepatitis se señalan el género (morado oscuro) y especie (violeta claro)

- DNA cadena doble
- DNA cadena sencilla
- DNA y RNA RT virus
- RNA cadena doble
- RNA cadena sencilla (-)

- RNA cadena sencilla (+)

donde se produce la encapsidación y la cobertura por la envuelta en el caso de que la posean.

En la tabla I.1, podemos observar que está aceptado el uso de nombres latinizados terminados en –viridae para las familias virales, en –virinae para las subfamilias, y en –virus para los géneros. Los nombres vulgares siguen siendo utilizados para nombrar a los virus dentro de un género.

Actualmente existe un orden, 71 familias, 9 subfamilias, 164 géneros y más de 3600 especies de virus.

Como hemos mencionado anteriormente, las infecciones virales se encuentran entre las patologías de origen infeccioso más relevantes en nuestra época, bien porque su número absoluto y/o relativo se ha visto incrementado con respecto al total de infecciones, o por sus consecuencias sociales y económicas. Un virus que se ha hecho merecedor de una atención y seguimiento especial en los últimos años, es el virus de la Hepatitis C (VHC). Este virus será objeto de estudio en el presente trabajo.

2. EL ORIGEN DE LOS VIRUS

Tres teorías han sido propuestas para explicar el origen de los virus. Antes de pasar a comentarlas sería conveniente saber qué entendemos por origen. Para ello nos acogemos a la definición propuesta por Strauss *et al.* (1996), según la cual el origen de un virus o elemento genético se corresponde con el momento en el que su replicación y evolución comienzan a ser independientes de las macromoléculas de las cuales deriva. Cuando un virus adquiere la información genética para determinar su propia propagación, se dice entonces que ha alcanzado el estatus de un elemento genético nuevo e independiente.

Pasemos ahora a ver las tres teorías sobre el origen de los virus:

Teoría regresiva del origen de los virus. Esta teoría propone que los virus son formas degeneradas de parásitos intracelulares que han ido perdiendo progresivamente determinadas habilidades biosintéticas, manteniendo a su vez funciones esenciales tales como, un origen de replicación, una forma de regularla y un mecanismo de interacción con la maquinaria biosintética y replicativa del hospedador. Esta teoría es similar a las

teorías regresivas que se han propuesto para explicar el origen de ciertos orgánulos como mitocondrias y cloroplastos. No obstante, se considera que esta teoría no explica del todo bien el origen de todos los tipos virales, especialmente aquellos que pasan por intermediarios replicativos de RNA.

Teoría del origen de los virus a partir de componentes de DNA o RNA celular del hospedador. Esta teoría postula que los virus habrían surgido a partir de componentes celulares que adquirieron la habilidad de replicarse autónomamente y por tanto de evolucionar independientemente. Los únicos requerimientos de independencia serían contar con un origen de replicación y una fuente de proteína que envolviera al virión durante la fase extracelular del ciclo infeccioso. Esta teoría parece poseer mayor atractivo que la anterior ya que permite explicar el origen de todos los virus: los virus de DNA se habrían originado a partir de plásmidos o elementos transponibles, los retrovirus a partir de retrotransposones y los virus de RNA a partir de RNA mensajeros autoreplicativos.

Teoría del origen de los virus a partir de moléculas de RNA autoreplicativas. Esta teoría surge debido a la capacidad que presentan las moléculas de RNA de catalizar diversos tipos de reacciones bioquímicas. Así, propone que un sistema replicativo basado en RNA podría haber precedido al actual basado en DNA-RNA-proteína. Según esta teoría, los virus de RNA serían vestigios de un mundo ancestral basado en RNA.

3. LOS VIRUS DE RNA

Debido a que el rango de hospedador de los virus es muy amplio, y nos ocuparía mucho espacio el hablar de cada uno de ellos y sus peculiaridades, nos centraremos en esta sección en los virus de RNA solamente, ya que es en este grupo donde se encuentra el VHC.

Los virus de RNA son los parásitos moleculares que más frecuentemente afectan a animales y plantas. Diversos estudios comparativos de su estructura, organización genética y tipo de replicación, muestran que los virus de RNA utilizan estrategias muy dispares para asegurar su multiplicación en las células y establecerse como partículas libres.

Aparentemente la gran variabilidad genética de los virus de RNA puede estar en la base de la emergencia de nuevas enfermedades, o de la ineficacia de algunas vacunas o tratamientos antivirales. Entre los factores que pueden contribuir a su alta variabilidad tenemos los grandes tamaños poblacionales, tiempos de generación cortos y altas tasas de replicación. Esta alta variabilidad por tiempo de generación de cualquier virus de RNA, supera con creces la encontrada en organismos basados en DNA.

Mencionamos tres propiedades de los virus de RNA que, aunque no generales, pueden ayudar a entender su alto grado de variación genética. Se trata de la mutación, la reasociación y la recombinación.

Mutación. La mutación es un mecanismo clave en la generación de variación genética y nuevos tipos de virus de RNA. En la mayoría de casos se trata de mutaciones puntuales (una base es reemplazada por otra), sin embargo también se observan deleciones e inserciones. Pero ¿a qué se deben las altas tasa de mutación encontradas en los virus de RNA? Las RNA polimerasas de estos virus carecen de la prueba de lectura y su frecuencia de error es del orden de 10^{-4} sustituciones por nucleótido y ronda de replicación. Si a esto le añadimos el gran número de ciclos de replicación que un virus puede alcanzar en cortos períodos de tiempo, obtenemos una explicación de la alta tasa de divergencia que les caracteriza. La mayoría de errores cometidos por la replicasas son o bien neutros o bien deletéreos. No obstante, el hecho de que existan genomas mutados también ofrece al virus la capacidad de responder rápidamente a cambios que puedan producirse en el ambiente.

Reasociación. Este mecanismo también es capaz de generar diversidad. Actúa en aquellos virus que presentan su genoma en forma de segmentos discretos. Dos virus con genomas segmentados darían lugar a una progenie que poseería varias combinaciones de tales segmentos. Esta estrategia proporcionaría un método muy rápido de producir nuevas formas virales y se piensa que es responsable, por ejemplo, de los saltos antigénicos que se producen en el virus de la gripe y que dan lugar a pandemias globales. Un virus de la gripe humana podría reasociarse con algún otro, que afectase a otros animales, o de origen humano pero perteneciente a una cepa desconocida, formándose otro con genes esenciales para la replicación en humanos pero con nuevos antígenos de superficie para los cuales la población humana aún no tiene inmunidad ni vacuna para combatirla.

Recombinación. La recombinación homóloga no es un mecanismo muy extendido entre los virus de RNA. En el proceso el donador de secuencias reemplaza una región homóloga del aceptor dejando su estructura invariable. También se han descrito otros fenómenos de recombinación menos frecuentes (Lai 1992), como la recombinación homóloga aberrante (entre secuencias de virus similares no alineadas), o recombinación no homóloga (recombinación entre secuencias no relacionadas de RNA). El proceso por el que se produce este nuevo variante, implica elección de copia en vez de rotura y religamiento (mecanismo de recombinación homólogo en DNA). Bajo este modelo los RNAs híbridos se forman cuando la RNA polimerasa RNA-dependiente viral cambia de una molécula molde a otra durante la replicación. La recombinación puede jugar un papel fundamental en la evolución de los virus de RNA ya que permite la generación de variación genética, la reducción de carga mutacional y la producción de nuevos virus.

4. FUERZAS EVOLUTIVAS

Desde los años sesenta del siglo pasado se discute entre dos aproximaciones a la evolución molecular. Una sostiene que el proceso de evolución es dominado por la fijación de mutaciones neutras o casi neutras por efecto de la deriva genética, y la otra que es la selección natural de mutaciones ventajosas la fuerza más importante. El origen de esta controversia es más antiguo en el fondo. Corresponde al debate entre dos escuelas de la genética de poblaciones: la escuela clásica y la escuela del equilibrio, equilibradora. La disputa en este caso trataba de la cantidad de variación genética que existía dentro de las poblaciones. La escuela clásica creía que la selección natural era predominantemente negativa eliminando los alelos deletéreos, por lo que existiría poca variabilidad genética dentro de las poblaciones. La escuela equilibradora, por su parte, defendía que los niveles de variación genética en las poblaciones eran mas bien altos, debido a diferentes formas de selección estabilizadora, fundamentalmente aquéllas que favorecían a los heterocigotos. A partir de 1960, la electroforesis y secuenciación de proteínas mostraron que la variación genética es bastante alta tanto dentro como entre especies, lo cual parecía dar la razón a la escuela equilibradora. No obstante, con esta afirmación surgía una cuestión: Si toda esa diversidad era mantenida por la selección natural, aquellos individuos con alelos menos ventajosos, la mayoría, serían

selectivamente eliminados de la población, llegando la población a extinguirse debido a muerte selectiva. A este fenómeno se le conoce como el coste de la selección natural. Para resolver esta paradoja, Kimura (1968) y King y Jukes (1969) propusieron la teoría neutral de evolución molecular. Esta teoría explica los altos niveles de variación genética al admitir que la mayoría de mutaciones fijadas no tuvieran efectos sobre la eficacia. Este tipo de mutaciones, evidentemente, no conllevarían ningún coste selectivo. Son las llamadas mutaciones neutras y son perdidas (usualmente) o fijadas (muy ocasionalmente) por deriva genética. Esta teoría da un papel poco relevante a la selección natural en la evolución molecular.

Para los neutralistas la mayoría de mutaciones son o deletéreas y selectivamente eliminadas, o neutras, ambas con probabilidad determinada de llegar a ser fijadas. La selección se incorpora a la teoría neutral como una fuerza purificadora eliminando las mutaciones deletéreas, y en muy raras ocasiones juega un papel relevante en la fijación de mutaciones de efecto positivo. Aunque la mayoría de mutaciones se pierden por azar, una fracción muy pequeña podría finalmente llegar a fijarse.

Debido a que los neutralistas sostienen que la selección natural no juega un papel importante en la fijación de mutaciones, se puede afirmar que la teoría neutral considera que la evolución a largo plazo es producto de la deriva genética y la fijación por azar de las mutaciones neutras, y donde la mutación es clave para permitir el polimorfismo en las poblaciones naturales. Para muchos la teoría neutral representa un desafío frente a uno de los pilares del neodarwinismo: la evolución por selección natural. Sin embargo la teoría neutral no descarta por completo la fijación de mutaciones adaptativas; lo que afirma es que este proceso ocurre a tan bajas frecuencias que es irrelevante a gran escala. Es más, los neutralistas piensan que la mayoría de genes y proteínas están tan bien adaptados, por la acción en el pasado de la selección natural, que es muy improbable su mejora. Por ello, es muy posible que sean deletéreas o neutras.

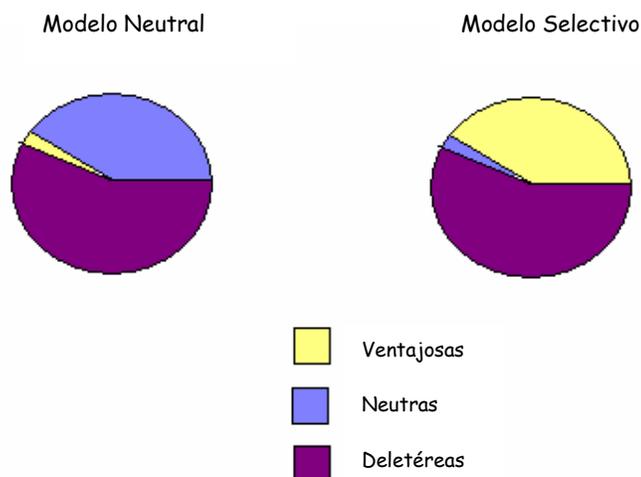


Figura I.1. Modelo neutral y modelo selectivo de evolución molecular. Los círculos representan el total de mutaciones en un gen. En el modelo neutral, la mayoría de las mutaciones son deletéreas y eliminadas por selección negativa, pero de aquellas que son fijadas la mayoría son neutras, y solo un número muy pequeño resultan ventajosas. Aunque el modelo selectivo coincide en que la mayoría de mutaciones son deletéreas y eliminadas, difiere en que la mayoría de aquellas mutaciones que son fijadas son selectivamente ventajosas y sólo un pequeño porcentaje es neutro. Las proporciones de las mutaciones neutras y ventajosas suponen la principal diferencia entre ambos modelos.

Por otra parte encontramos a los seleccionistas, que sostienen que las mutaciones (sustituciones) son fijadas por selección aportando una ventaja adaptativa, mientras que las mutaciones neutras son muy raras. Con otras palabras, según el punto de vista seleccionista, los diferentes mutantes que puedan aparecer no son selectivamente neutros ni fenotípicamente equivalentes, sino que suponen un repertorio de actividades biológicas diferentes en eficacia biológica sobre las cuales, la selección puede actuar. La mutación y la deriva genética supondrían fuerzas menores en perfilar el proceso evolutivo, siendo la selección la que tendría el papel determinante.

Ambos puntos de vista tienen en común que la mayoría de mutaciones tienen mayor probabilidad de ser deletéreas y, por tanto, de ser eliminadas rápidamente de la población, sin embargo lo que para los neutralistas se considera azar, para los seleccionistas constituye una necesidad (Figura I.1). Esta controversia afecta por supuesto, a la evolución de los virus de RNA. De hecho son excelentes modelos, por su rápida tasa de evolución, para confirmar una u otra teoría. Sin embargo no resulta fácil

determinar cual de las dos teorías da mejor cuenta del proceso de diversificación genética a largo plazo en un virus de RNA. Debemos considerar si la evolución está gobernada principalmente por fuerzas selectivas, por factores de azar, o incluso que ambos procesos actúen conjuntamente, o que estén implicados otros factores evolutivos.

5. VIRUS DE RNA Y LA TEORÍA DE LAS CUASISPECIES

Al hablar de virus de RNA hay un término que no podemos obviar debido a la gran repercusión y uso que ha tenido en los últimos años. El término al que nos referimos es el de las cuasispecies. Este apartado no obstante, no tiene la intención de entrar en el debate de si las cuasispecies existen o no, ya que no está dentro de los objetivos de esta tesis comprobar su existencia. Sin embargo, debido a la polémica suscitada, y al empleo de este término, en algunas ocasiones, para describir la variación genética del virus de la hepatitis C, hemos querido hacer referencia a ellas. La teoría de las cuasispecies surgió como un intento de describir el comportamiento de formas replicativas elementales. Tales formas se corresponderían con replicones de corta longitud y alta mutación. Estas dos propiedades ha llevado a algunos investigadores (Domingo 1992; Domingo y Holland 1997; Holland *et al.* 1992) a pensar que los virus de RNA podrían estar sujetos a las reglas evolutivas de las cuasispecies, diferentes a las formas de vida basadas en DNA (con tamaños genómicos mayores y tasas de mutación más bajas). El término de cuasispecies, como decimos fue acuñado por primera vez por Eigen, Schuster y colaboradores (Eigen y Schuster 1979; Eigen *et al.* 1981; Eigen y Biebricher 1988) para referirse a las primeras estructuras autorreplicativas. Las poblaciones de tales estructuras poseerían también una alta tasa de mutación que conduciría a la existencia de una mezcla compleja de diferentes, pero relacionados, genomas que se comportarían como una unidad de selección.

En términos sencillos una cuasispecie es una población en equilibrio entre mutación y selección natural. Las variantes se encuentran organizadas definiendo un genotipo canónico llamado “secuencia maestra”. Un elemento crítico en la teoría de las cuasispecies, aplicadas a los virus, es que la frecuencia de cualquier partícula individual es función de su tasa de replicación y de la probabilidad de que ésta aumente debido a los errores en la replicación de otros miembros de la población. Esto hace que los virus

no sean entidades independientes sino que se encuentran unidos por “acoplamiento mutacional”, por lo que la población entera forma una estructura cooperativa que evoluciona como una unidad. La consecuencia de esta estructura poblacional es que la selección natural no siempre es dirigida hacia la variante con mayor eficacia, como en la mayoría de modelos genético poblacionales, sino que actúa sobre toda la distribución de mutantes, que evolucionarán entonces a maximizar su tasa de replicación media.

6. FILOGENIAS MOLECULARES Y POBLACIONES VIRALES

Dada la importancia de la variación genética en las poblaciones virales, es crucial que esta diversidad sea estudiada y comprendida a fondo. Las reconstrucciones filogenéticas se han convertido en una herramienta muy importante en el estudio de esta variación ya que tienen la ventaja, sobre otros métodos, de que tienen en cuenta la información evolutiva contenida en las secuencias. Esto no quiere decir que otros métodos no deban de ser considerados, sino que la filogenia molecular contribuye a comprender la variación actual teniendo en cuenta la generación y evolución de la misma.

Como ya hemos dicho las filogenias no solo ofrecen un esquema visual de clasificación, si no que tienen multitud de otros usos. Existen factores evolutivos tales como la selección natural, la deriva genética y la historia demográfica de las poblaciones, que dejan diferentes huellas en las relaciones evolutivas entre las secuencias representadas en una filogenia. Por ello, un análisis exhaustivo de un árbol filogenético puede descifrar cuales de los procesos anteriores han actuado.

En particular, las reconstrucciones filogenéticas han resultado especialmente útiles en el estudio de la estructura genética de las poblaciones virales. A continuación nos centraremos en tres áreas donde el estudio de las filogenias puede ser realmente importante en el control y erradicación de las enfermedades virales.

¿Cuándo y cómo han surgido las infecciones virales? El estudio de los factores responsables de la aparición de las nuevas enfermedades virales es el principal objetivo de esta área de estudio. Un ejemplo de cómo las filogenias ayudan a desvelar cuando ha surgido una nueva infección, lo constituye el estudio filogenético realizado con VIH-1,

VIH-2 y los relacionados virus VIS aislados a partir de diferentes especies de primates. El análisis refleja cómo VIH-1 y VIH2 han pasado a la población humana, probablemente a partir de diferentes especies de simios, de forma independiente, y en más de una ocasión (Sharp *et al.* 1995). El uso de filogenias moleculares también puede resultar útil para averiguar que factores facilitan la propagación de la infección. Otro ejemplo; se ha sugerido que la revolución agrícola hace 10.000 años, supuso un aumento considerable del tamaño de las poblaciones humanas, favoreció el sedentarismo y por tanto el contacto más prolongado con animales. Estos tres factores pudieron favorecer la aparición y diseminación de enfermedades virales. Si el desarrollo de la agricultura ocasionó un aumento del número de enfermedades infecciosas, esperaríamos que los tiempos de emergencia para diferentes patógenos, estimados a partir de análisis filogenéticos, convergieran en este punto. Este mismo planteamiento ha permitido relacionar la propagación de determinados virus, como el virus del dengue, con grandes movimientos migratorios de las poblaciones humanas (Dekaban *et al.* 1995; Zanotto *et al.* 1996).

¿Qué fuerzas evolutivas dirigen las poblaciones virales? Las filogenias permiten averiguar cómo y por qué las poblaciones virales se estructuran a nivel geográfico. Estudios con el virus del SIDA o de la Hepatitis C han podido, aparte de determinar la existencia de diferentes genotipos y subtipos de los dos virus (Louwagie *et al.* 1993; Simmonds *et al.* 1993), relacionarlos con diferentes patrones de distribución geográfica. (Holmes *et al.* 1995; Simmonds *et al.* 1996). También se ha podido determinar las rutas de transmisión entre diferentes grupos de personas en una población (Holmes *et al.* 1995). Aparte de estos patrones de variación geográfica, también es importante la información que puede obtenerse de la dinámica viral a partir de la estructura de las ramas en una filogenia. Por ejemplo, alelos mantenidos por selección estabilizadora existirán como polimorfismos durante períodos de tiempo mayores produciendo ramas más largas que los alelos neutros, cuyo destino es determinado por deriva genética. Por otra parte, aquellas mutaciones ventajosas fijadas por arrastre selectivo existen como polimorfismos durante períodos de tiempo más cortos que las mutaciones neutras generando filogenias en estrella, con ramas muy cortas que representan la variación a partir del arrastre (Takahata y Nei 1990). Así mismo cambios en la tasa de crecimiento poblacional, también afectarán a la forma de las filogenias.

¿Cómo afecta la variabilidad viral al progreso de la enfermedad? Las altas tasas de mutación parecen conferir a los virus un alto potencial para adaptarse rápidamente a ambientes nuevos y variables. Debido a que esta adaptabilidad es el resultado de la variación, no sorprende que los virus de una misma población difieran en su capacidad para producir la enfermedad y sus síntomas. La asociación de ciertas mutaciones virales con diferentes progresos de la enfermedad es un aspecto importante de la medicina molecular y otra de las áreas donde las filogenias moleculares juegan un importante papel. Un ejemplo lo encontramos en ciertos trabajos con el virus de la hepatitis B. La infección producida por este virus puede causar desde una simple infección subclínica, hasta provocar severos daños en el hígado como la cirrosis o el hepatocarcinoma. Las causas que explican los diferentes desenlaces de la enfermedad están siendo estudiadas, pero parece que algunos de los casos donde la enfermedad es más grave se deben a una mutación (A₁₈₉₆) en la región precore del genoma del virus. Estudios filogenéticos han revelado que esta mutación está asociada a varios subtipos del virus que particularmente se encuentran en el sureste asiático.

7. LA HEPATITIS VIRAL

Se conocen 5 virus responsables de hepatitis vírica en humanos. Se nombran de la A a la E, y aparte de localizarse en el hígado y causar hepatitis en el hospedador, estos virus tienen muy poco en común, especialmente si atendemos a la caracterización molecular de sus genomas.

Puesto que el virus estudiado en esta tesis es el de la hepatitis C, dejaremos para apartados posteriores la explicación de las características más relevantes del mismo. En este apartado nos centraremos brevemente en los cuatro virus restantes.

Hepatitis A. El virus de la hepatitis A (VHA) ha sido clasificado como la única especie dentro de un nuevo género, *Hepatovirus*, perteneciente a la familia *Picornaviridae* (Gust *et al.* 1998). Es un virus de RNA circular de cadena simple positiva de una longitud de 7.5 Kb. La región codificadora del genoma constituye un fragmento de 6681 pb con una única pauta de lectura abierta, la cual está dividida en tres regiones: P1, que codifica para cuatro proteínas de la cápsida (VP1-VP4), y P2 y P3 que codifican para

proteínas no estructurales, implicadas en el procesado de la poliproteína y la replicación viral respectivamente (Rueckert y Wimmer 1984). El VHA, a diferencia del resto de virus causantes de hepatitis, ha sido adaptado para crecer en cultivo celular (Provost y Hilleman 1979), lo que ha conducido al desarrollo de una vacuna efectiva contra el mismo. Se transmite a través de una ruta fecal-oral, y causa una enfermedad aguda* de corta duración cuya severidad varía, pero nunca llega a hacerse crónica. (Battegay y Feistone 1997). La incidencia de esta enfermedad ha ido descendiendo en las últimas décadas en países industrializados de Europa y América del Norte (Hadler 1991), coincidiendo con el aumento de las medidas de higiene y mejora en el tratamiento de las aguas.

Hepatitis B. El virus de la hepatitis B (VHB) pertenece a la familia *Hepadnaviridae*, y está formado por una molécula de DNA circular de 3.2 Kb. El genoma posee cuatro pautas de lectura abierta, que se solapan parcialmente y dan lugar al menos a siete productos virales. El VHB se replica vía RNA en un paso catalizado por una transcriptasa inversa. En el proceso se distinguen cuatro pasos (Nassal y Schaller 1996): primero, el fragmento de DNA, de doble cadena parcial, es convertido en un fragmento circular covalentemente cerrado en el núcleo. Esta molécula circular de DNA sirve después como molde para la transcripción del pregenoma y varios mensajeros. En tercer lugar, se produce la traducción y el empaquetamiento del pregenoma de RNA y la polimerasa viral en la nucleocápsida. Tras la encapsidación se produce la síntesis de DNA negativo. Finalmente la polimerasa sintetiza la cadena complementaria de DNA, y los viriones, tras recubrirse con las proteínas de la envuelta, son liberados. Se distinguen hasta seis genotipos, los cuales se distribuyen geográficamente (Okamoto *et al.* 1988). Aunque el cultivo de VHB no ha sido posible, sí existe una vacuna contra el virus. Aparte de su transmisión por exposición a sangre o productos hemoderivados contagiados, es también común el contagio por vía sexual y perinatal. La infección aguda es clínicamente silenciosa y puede llegar a causar cronicidad (en el 10% de los casos) que a su vez puede desencadenar la aparición de cirrosis y más tarde desarrollarse un hepatocarcinoma.

* En el apéndice puede encontrarse una breve descripción de los términos médicos a los que nos referimos en esta introducción.

Hepatitis D. El virus de la hepatitis D (VHD) pertenece a la familia *Arenaviridae*, y es un virus de RNA circular de cadena simple negativa, el cual necesita de la presencia del VHB para poder establecer la infección. Es un virus con envuelta y con un genoma de 1,6 Kb (Wang *et al.* 1986). El VHD difiere de los otros virus de RNA animales en que no utiliza una RNA polimerasa viral, sino que hace uso de la RNA polimerasa II del hospedador. Posee una alta heterogeneidad genética que se pone de manifiesto por la presencia de dos formas diferentes del antígeno delta HDAg (Weiner *et al.* 1988; Luo *et al.* 1990; Wang *et al.* 1992). Estos dos HDAg (HDAg-p24 y HDAg-p27) difieren en su estructura en 19 aminoácidos (debidos a la sustitución de un codón de parada que permite la síntesis de 19 aminoácidos más en el HDAg-p27 (Bergmann y Gerin 1986). Ambas formas tienen funciones diferentes, mientras que la corta es requerida para la replicación del RNA, la larga suspende la replicación y es requerida para la encapsidación del virus. Existen tres genotipos del virus. La ruta de transmisión es similar a la del VHB. La infección casi siempre se establece durante el período crónico de la hepatitis B, lo que se conoce como superinfección delta, aunque en otros casos se produce durante la fase aguda (coinfección delta). En ambos casos puede llegar a producirse cronicidad. Debido a la necesidad de una infección previa o conjunta con el VHB, la vacuna contra este último es una forma de controlar la enfermedad. No obstante los pacientes con una infección crónica del VHB siguen presentando riesgo de contraer Hepatitis D.

Hepatitis E. El virus de la hepatitis E fue clasificado dentro de la familia *Caliciviridae*, aunque recientemente ha sido excluído de esta familia por el ICTV quedando pendiente su clasificación (Green *et al.* 2000) Es un virus sin envuelta con un RNA de cadena simple positiva de 7,6 Kb de longitud (Bradley *et al.* 1987; Reyes *et al.* 1990). Posee tres pautas de lectura abiertas. La primera codifica para la RNA polimerasa y otras proteínas no estructurales y la segunda y la tercera para las proteínas de la cápsida y una proteína inmunogénica de función desconocida respectivamente (Tam *et al.* 1991; Tsarev *et al.* 1992). Se transmite vía fecal-oral y no parece causar una enfermedad crónica.

8. EL VIRUS DE LA HEPATITIS C

El progreso realizado en el estudio del virus de la hepatitis C (VHC) es un ejemplo del éxito conseguido por la interacción de la ciencia básica y la investigación clínica. El VHC fue el primer virus que se descubrió por medio de técnicas moleculares. Aunque la hepatitis C, como enfermedad, fue reconocida por primera vez en 1975 (Feinstone *et al.* 1975), todos los medios disponibles para identificar el agente causante de la enfermedad fallaron hasta 1989, cuando el grupo de Choo y colaboradores (1989) clonaron y secuenciaron un fragmento del genoma del VHC. El uso de métodos moleculares ha sido un instrumento fundamental en el descubrimiento del virus así como en los avances realizados en el estudio de la hepatitis C. No obstante, a pesar del extraordinario progreso acontecido en la pasada década, el VHC continúa siendo uno de los mayores desafíos para virólogos y médicos.

8.1 Epidemiología y cuadro clínico

Aproximadamente 170 millones de personas en el mundo se encuentran infectadas por el VHC. Los estudios epidemiológicos muestran que la incidencia del virus es superior en los países menos desarrollados. Los datos más alarmantes se encuentran en varios países africanos, alcanzando su máximo en Egipto donde afecta al 14,5% de la población.

La principal vía de transmisión es la exposición directa o indirecta con sangre contaminada, bien sea a través de transfusiones de sangre, por consumo de drogas vía intravenosa, empleo de material quirúrgico mal esterilizado, etc. Desde su descubrimiento en 1989, el número de casos de hepatitis postransfusional se ha reducido considerablemente, no así el atribuible a los adictos a drogas por vía parenteral (ADVP), que ha aumentado considerablemente. Otra vía de transmisión menos prevalente es la vía materna (menos del 5% de las madres infectadas) (Di Bisceglie 1998). No se descarta que puedan existir otras rutas ya que un 10% de los casos aparecen sin evidencia de factores de riesgo, aunque es posible que estos individuos se hayan visto expuestos inadvertidamente a sangre o productos sanguíneos contaminados.

Una de las razones por las que la hepatitis C ha tardado tanto en salir a la luz, podría ser su frecuente infección silenciosa. La mayoría de pacientes con infección

aguda no padecen síntomas clínicos aparentes. Durante esta fase casi todos los pacientes desarrollan una fuerte respuesta inmune humoral y celular, que no consigue evitar la infección pero es una de las causas del daño hepático, debido a la destrucción masiva de células hepáticas por el continuo ciclo de replicación y liberación del virus y el ataque del sistema inmune contra las células infectadas. Los niveles de aminotransferasas (ALT) en suero, indicativos del daño a los hepatocitos, aumentan entre la segunda y la octava semana tras la exposición. La hepatitis C aguda puede ser grave y prolongada, pero raramente es fulminante (Farci et al. 1996). En algunos casos, la infección remite por sí sola; los síntomas duran varias semanas y desaparecen conforme caen los niveles de ALT y de RNA del VHC. No obstante el 85% de los pacientes infectados suelen desarrollar hepatitis crónica (Rosen y Gretch 1999). Esta cronicidad está marcada por la persistencia del RNA del VHC durante al menos seis meses después del establecimiento de la enfermedad. En este estadio tampoco se presentan síntomas específicos, suelen aparecer dolores abdominales deslocalizados, fatiga, dolor muscular y en las articulaciones. Un 20% de los enfermos crónicos llegan a padecer cirrosis y de un 1% a un 4% pueden sufrir hepatocarcinoma (Figura I.2).

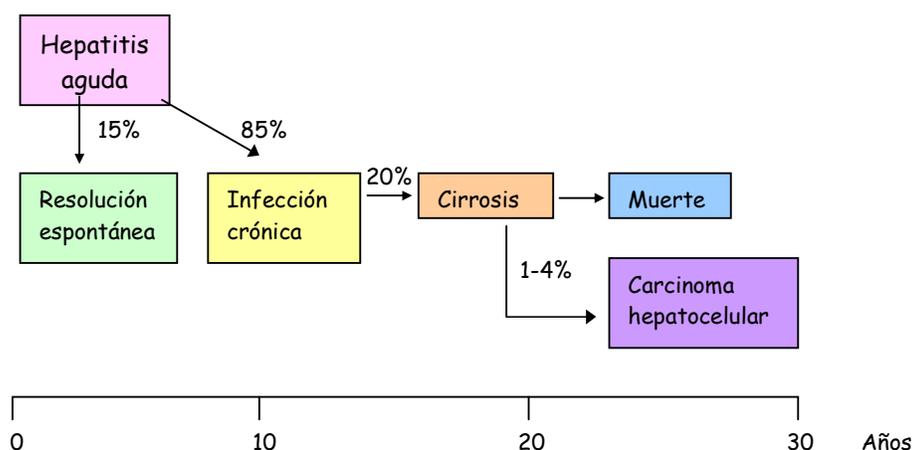


Figura I.2. Historia natural del VHC. La mayoría de individuos infectados desarrollan infección crónica. El intervalo medio de tiempo de duración de la enfermedad es entre 20 y 30 años.

Debido a que la cirrosis puede tardar varios años en desarrollarse, la infección con VHC puede permanecer silenciosa durante décadas. Por tanto los individuos infectados, en la mayor parte de casos, suelen desconocer su estado durante largos períodos de tiempo, y la enfermedad es detectada con frecuencia en análisis de sangre rutinarios (por niveles elevados de enzimas hepáticas) o donaciones de sangre (Di Bisceglie 1998). Además del daño hepático producido, que se ha convertido en una de las principales indicaciones para el trasplante de hígado, se han descrito otras manifestaciones extrahepáticas como crioglobulinemia y glomerulonefritis (Jonson *et al.* 1993)

8.2 Estructura

El VHC ha sido clasificado dentro del género *Hepacivirus* perteneciente a la familia *Flaviviridae*. (Reed y Rice 2000). Posee un parecido bastante cercano a otros miembros de esta familia en cuanto a secuencia y organización del genoma, modo de replicación, estructura y procesado de la poliproteína. La familia *Flaviviridae* contiene además a los géneros *Flavivirus* y *Pestivirus*.

EL VHC es un virus pequeño, recubierto por una envuelta, con una cadena de RNA lineal simple positiva de unas 9,6 Kb, y que posee una única pauta de lectura abierta (ORF) que es traducida a una poliproteína de unos 3100 aminoácidos (Figura I.3). Esta ORF es flanqueada en 5' y 3' por dos regiones no traducidas: 5'UTR y 3'UTR. La 5'UTR posee una longitud de 340 bases y es la más conservada del genoma. Posee una región interna de entrada ribosomal (IRES) que media la unión del RNA al ribosoma (Tsukiyama-Kohara *et al.* 1992). La función de 3'UTR aún no ha sido esclarecida, aunque se piensa que podría ser importante para la iniciación de la síntesis de la cadena negativa de RNA, o bien podría estar implicada en la regulación de la actividad NTPasa de la proteína NS3 (Tsuchihara *et al.* 1997; Spangberg *et al.* 2001). La poliproteína viral es fragmentada, *co* y *pos*-traduccionalmente por proteasas pertenecientes tanto al hospedador como al virus, en al menos 10 proteínas estructurales y no estructurales (NS) que siguen el siguiente orden: NH₂-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH (Schulze zur Wiesch *et al.* 2003). El primer corte lo realiza una proteasa del hospedador dando lugar a la única proteína de la cápsida, la proteína core, a las proteínas de la envuelta E1 y E2, y a una proteína fuertemente

hidrofóbica de función desconocida llamada p7. Las proteínas no estructurales restantes son procesadas por proteasas virales. El corte entre el sitio NS2 y NS3 es llevado a cabo por la proteinasa NS2-NS3 (dependiente de Zn^{+2}) de forma autocatalítica. La serinproteasa NS3 produce el corte entre NS3 y NS4A, NS4A y NS4B, NS4B y NS5A, NS5A y NS5B. El corte entre NS4B y NS5A es lento, por lo que suele encontrarse el intermediario NS4B/NS5A.

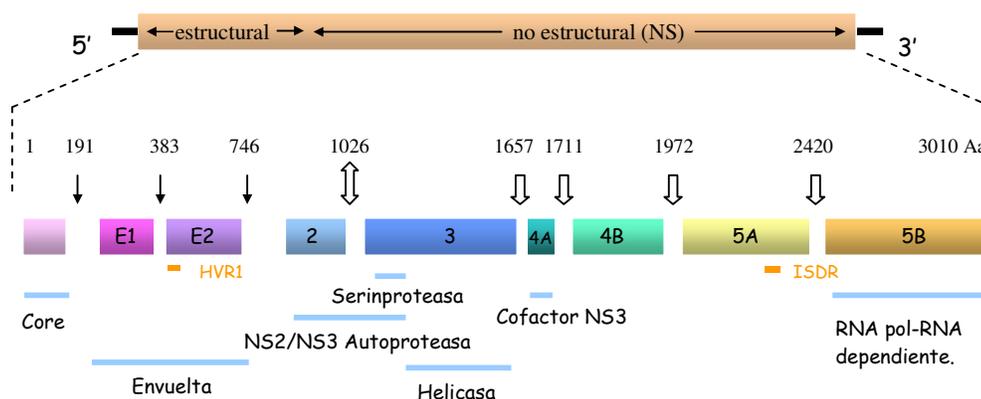


Figura I.3. Representación esquemática de la poliproteína del VHC. En la parte superior se muestra la ORF y las regiones no traducidas (UTR, representadas como una línea negra fina). En la parte intermedia del esquema se indican las diferentes proteínas que se obtienen tras el procesado de la poliproteína, así como los lugares de corte de las diferentes proteasas (Los cortes realizados con la proteasa del hospedador se representan con ↓, los realizados por la proteasa NS2-NS3 con ⇄, y los realizados con la proteasa NS3 con ⌋). También se apunta dónde están localizadas la región hipervariable 1 (HVR1) y la región determinante de la sensibilidad a interferón (ISDR). En la parte inferior del esquema se señalan las funciones de las proteínas obtenidas.

A continuación pasaremos a detallar las características más relevantes de cada una de las proteínas.

Core. Es una proteína fuertemente básica y evolutivamente muy conservada. Forma el principal constituyente de la nucleocápsida viral, y se ha observado que tiene capacidad multimerizadora e interactúa con las proteínas de la envuelta (Lo *et al.* 1996; Matsumoto *et al.* 1996). Aunque también se le atribuyen otras funciones como que podría jugar un papel importante en la regulación del crecimiento de las células infectadas con VHC y en el desarrollo de hepatocarcinoma (Moriya *et al.* 1998).

E1, E2, E2-p7. Como hemos visto las glicoproteínas E1 y E2 son liberadas de la poliproteína por una peptidasa del hospedador, pero en ocasiones esta rotura es incompleta o lenta dando lugar a los productos intermedios E2-p7 y p7-NS2. Estos intermediarios junto con p7 se piensa que están implicados en la gemación del virus. Las proteínas E1 y E2 forman un heterodímero covalentemente unido. Esta interacción entre las proteínas parece ser importante para el correcto ensamblaje de los viriones. Estas proteínas poseen su extremo C-terminal (altamente hidrofóbico) incluido en la membrana del retículo endoplásmico de la célula, mientras que su ectodominio permanece en el lumen (Schulze zur Wiesch *et al.* 2003). En el extremo N-terminal de la proteína E2 existe una región (residuos 493-519) que posee una alta variabilidad y es conocida como región hipervariable 1 (HVR1). La HVR1 parece ser la diana de anticuerpos neutralizantes y además hay quien apoya una correlación entre la heterogeneidad de esta zona y la respuesta al tratamiento con interferón (Boulestin *et al.* 2002). Otra función importante relacionada con la persistencia del VHC es la asociada a la región C-terminal de E2. Existe una secuencia de aminoácidos (residuos 659-670) que parece ser responsable de la inhibición de la protein-quinasa (PKR).

NS2. Es una proteína hidrofóbica transmembrana cuya función aún no ha sido determinada. Hasta el momento sólo se conoce el papel autocatalítico que desempeña junto a la NS3 en la rotura de la unión NS2/NS3 (Schulze zur Wiesch *et al.* 2003).

NS3. Proteína ligeramente hidrofílica a la que se le han atribuido diferentes actividades enzimáticas como serinproteasa, implicada en el procesamiento de las proteínas no estructurales, y NTPasa y helicasa que parecen ser esenciales en la replicación viral (Gallinari *et al.* 1998).

NS4A. Proteína hidrofóbica a la que se le han asociado dos funciones importantes en el procesamiento de la poliproteína del VHC: actúa, por un lado como cofactor de la NS3 (en su actividad como proteasa), y además como anclaje del complejo de replicación al retículo endoplásmico, mediante la unión de su extremo N-terminal a la NS3. Por otra parte también puede unirse directamente a la NS5B y mejorar su actividad polimerasa (Ishido *et al.* 1998; Kim *et al.* 1996).

NS4B. Existe muy poca información acerca de su función, pero su alta hidrofobicidad hace pensar que podría estar también implicada en el anclaje del complejo de replicación al retículo endoplásmico (Lin *et al.* 1997).

NS5A. Si bien el papel que desempeña para el VHC aún no está aclarado (aunque se piensa que podría estar relacionada con el complejo de replicación), existe una región en esta proteína que parece ser de especial interés por su posible relación con el tratamiento con interferón. Esta región llamada ISDR (región determinante de la sensibilidad a interferón), está localizada entre los residuos 2209 y 2248 de la poliproteína, y parece interactuar con el dominio de dimerización de la PKR. Esta interacción es diferente a la propuesta para la región de la proteína E2.

NS5B. Ha sido identificada como una RNA polimerasa-RNA dependiente (Behrens *et al.* 1996).

8.3 Replicación

Debido a que no ha podido conseguirse un cultivo celular eficiente del VHC, todo lo que podemos inferir acerca de su ciclo de replicación está basado en analogías con virus relacionados pertenecientes a la familia de los flavivirus y pestivirus, y por la caracterización de proteínas recombinantes del VHC. Con esta limitada información podemos resumir el ciclo replicativo como sigue (Figura I.4):

Reconocimiento y entrada. El primer paso es el reconocimiento de la célula que va a ser infectada. En este caso parece que por parte de la célula están implicados la proteína CD81 y el receptor de lipoproteínas de baja densidad (LDLR). CD81 forma parte de la familia de las tetraspaninas, unas moléculas de la superficie celular caracterizadas por presentar tres dominios transmembrana y dos lazos extracelulares (Reed *et al.* 1995; Pileri *et al.* 1998). Aunque se sabe que E2 interactúa con CD81, no parece estar implicada en la fusión celular (Takikawa *et al.* 2000). Por otra parte algunos estudios indican que el VHC utiliza a la LDLR para su unión y posterior entrada en la célula. El hecho de que CD81 y LDLR no sean exclusivas de los hepatocitos hace pensar en la existencia de un segundo receptor necesario para la especificidad celular (Wunschmann *et al.* 2000). Por parte del virus todo apunta a que es la glicoproteína E2 la responsable

del reconocimiento celular (Rosa *et al.* 1996; Zibert *et al.* 1995; Farci *et al.* 1996). Con respecto a la proteína E1 podría estar implicada en la fusión celular ya que presenta una serie de aminoácidos hidrofóbicos (péptido de fusión) similares a los responsables de la fusión celular en paramixovirus y flavivirus (Flint y McKeating 1999). Por último, parece que la E2 no puede establecer por sí sola la fusión celular sino que necesita de la E1 (Flint y McKeating 1999).

Traducción y procesamiento de la poliproteína. Una vez dentro del citoplasma celular, el RNA del virus es traducido directamente mediante la secuencia IRES en el retículo endoplásmico rugoso. Aquí también se lleva a cabo el procesamiento de la poliproteína por parte de las proteasas del hospedador y del virus.

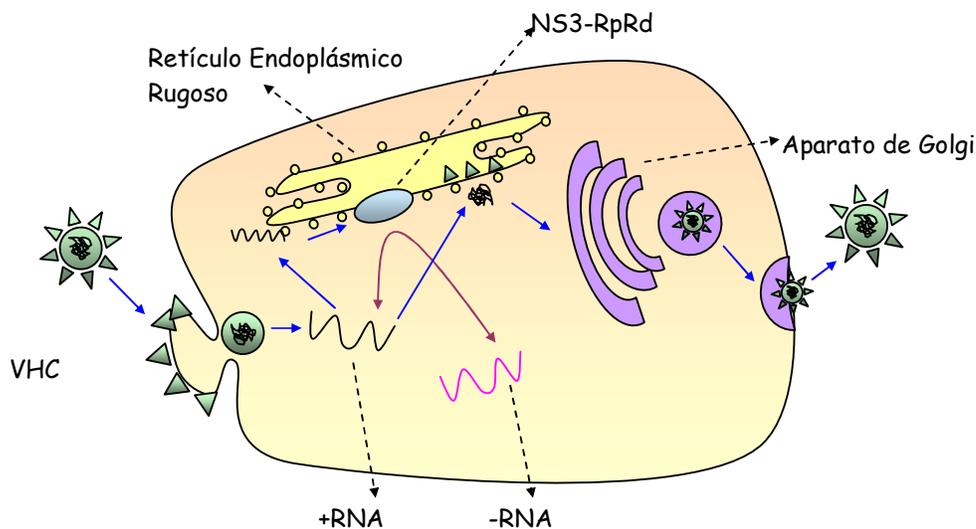


Figura I.4. Modelo hipotético del ciclo replicativo del VHC (Bartenschlager *et al.* 2000). Una vez el virus entra en la célula, su cadena de +RNA es traducida directamente. Después se produce el procesamiento de la poliproteína, dando lugar a las proteínas estructurales y no estructurales. La RpRd utiliza también el +RNA para la síntesis de -RNA que a la vez sirve como molde para fabricar más +RNA. Finalmente, una cadena de +RNA interactúa con las proteínas de la envuelta produciéndose la encapsidación y posterior liberación a través del aparato de Golgi.

Replicación del RNA. Tras el procesamiento de la poliproteína, se constituye el complejo de replicación formado en particular por la NS5B y la NS3 (aunque se piensa que alguna de las otras proteínas no estructurales también están implicadas). Este complejo se encuentra asociado a membranas intracelulares, de forma que la replicación del RNA tiene lugar en el citosol de la célula. La manera exacta de cómo ocurre la replicación del RNA aún resulta desconocida, pero sí se sabe seguro que la NS5B RNA polimerasa-RNA dependiente (RpRd), juega un papel clave en la síntesis de RNA de cadena positiva y negativa.

Ensamblaje viral y liberación. El ensamblaje del VHC tampoco ha sido estudiado en detalle, pero se piensa que debe comenzar con la interacción de la proteína core con el RNA. Esta unión parece que, de alguna forma, es un mecanismo que funcionaría como un interruptor: apagando el proceso de traducción/replicación y encendiendo el de ensamblaje (Shimoike *et al.* 1999). La nucleocápsida adquiere la envuelta mediante gemación al salir del retículo endoplásmico. Finalmente, el virión es liberado mediante la ruta de secreción mediada por el aparato de Golgi (Sato *et al.* 1993).

8.4 Heterogeneidad genética

Como muchos virus de RNA, el VHC muta continuamente (debido a la baja fidelidad de la replicación viral) y nunca se presenta como una población homogénea de genomas idénticos de RNA. La tasa de acumulación de mutaciones para el VHC se ha calculado que está alrededor de $10^{-3} - 10^{-4}$ sustituciones por sitio y por año (Ogata *et al.* 1991; Okamoto y Mishiro 1994). La heterogeneidad genética no está igualmente distribuida a lo largo del genoma, siendo los genes que codifican para las proteínas de la envuelta (E1-E2) los más heterogéneos. La región más variable de todo el genoma corresponde a la región hipervariable 1 (HVR1), localizada en el extremo N-terminal de la proteína E2 (Weiner *et al.* 1992).

La gran heterogeneidad presente en el VHC, ha llevado a su clasificación en 6 genotipos, los cuales difieren entre ellos en un 35% de los sitios nucleotídicos, y han sido nombrados del 1 al 6 por orden de descubrimiento. Cada genotipo, a su vez, es subdividido en un número determinado de subtipos (en total más de 50) nombrados

alfabéticamente también por orden de descubrimiento. Estos subtipos difieren entre ellos en un 20 – 25% de los nucleótidos (Robertson *et al.* 1998; Simmonds 1999).

Los diferentes genotipos presentan complejos patrones de distribución geográfica, prevalencia y rutas de transmisión, lo cual conduce a que podamos categorizarlos dentro de tres grupos (Smith *et al.* 1997):

Grupo endémico. En ciertas áreas geográficas encontramos un solo genotipo pero representado por multitud de subtipos, lo que es indicativo de un largo período de infección endémica. Estas regiones son: la costa de Guinea en el Este de África, para el genotipo 1, el Oeste y centro de África para el genotipo 2, el Norte de la India para el 3, África central para el 4 y el Sureste asiático para el 6.

Grupo Epidémico. En otras partes del mundo, sin embargo, encontramos varios genotipos pero representados por muy pocos subtipos. Esto se relaciona con una reciente y limitada introducción de los genotipos a partir de las zonas endémicas. Un ejemplo lo constituyen los genotipos 1a, 2b, 2a, 2b y 3a los cuales están distribuidos globalmente y son los responsables del mayor número de casos de hepatitis C. La rápida diseminación de este grupo a lo largo de todo el mundo parece que está relacionado con su transmisión a través de rutas que implican productos de sangre contaminados (como transfusiones de sangre, unidades de hemodiálisis...) y el consumo de drogas vía parenteral. Parece que los genotipos 1b y 2a están fuertemente relacionados con la primera vía de contagio, y su prevalencia podría estar decreciendo debido al exhaustivo control al que son sometidos los productos sanguíneos (Mondelli *et al.* 1999; Pybus *et al.* 2001). Por otra parte los genotipos 1a y 3a estarían más estrechamente relacionados con los adictos a drogas por vía parenteral (ADVP), y su prevalencia muestra un marcado crecimiento en los últimos años (Mondelli *et al.* 1999; Kalinina *et al.* 2001).

Grupo Local. Lo constituyen aquellos genotipos causantes de epidemias locales, caracterizados por tener una gran prevalencia en determinados grupos de riesgo y en áreas muy localizadas. El ejemplo más característico lo encontramos en Egipto. Estudios epidemiológicos muestran que más del 10% de la población está infectada con el genotipo 4a, a raíz del tratamiento utilizado contra la esquistosomiasis (Frank *et al.* 2000) en el que se produjo una inoculación masiva, mediante la reutilización de jeringuillas contaminadas.

Teniendo en cuenta las consideraciones anteriores, todo parece apuntar a que el diverso comportamiento epidemiológico encontrado en el VHC está determinado por las diferentes vías de contagio a través de las que se transmite. Si asumimos esto entonces podemos esperar que cualquiera de los 50 subtipos existentes, en cualquier momento, pueda causar una epidemia, tan pronto encuentre una ruta de transmisión apropiada y eficaz (Stumpf y Pybus 2002).

Aunque las rutas de transmisión podrían ser el factor determinante de la epidemiología global del VHC, la variación genética entre subtipos es responsable de diferencias en el progreso de la enfermedad y sus consecuencias clínicas (Mondelli *et al.* 1999). Así encontramos que, por ejemplo, los enfermos con genotipos 1a, 1b y 4a muestran una respuesta menos eficaz frente a tratamientos antivirales frente a los genotipos 2 y 3 (Bell *et al.* 1997), o que el subtipo 1b provoca daños más severos en el hígado que cualquier otro subtipo (Boothe *et al.* 1997; Mondelli *et al.* 1999). No obstante las causas por las que esto ocurre todavía no han sido del todo determinadas (Kleter *et al.* 1998; Takada *et al.* 1996).

8.5 Tratamiento frente al VHC.

El hecho de que la infección crónica con VHC puede llegar a causar cirrosis y hepatocarcinoma, crea la necesidad de desarrollar fármacos que erradiquen eficazmente la infección. El interferón- α sólo o en combinación con el nucleósido ribavirina son los únicos tratamientos que tienen cierta eficacia frente al VHC. En los tratamientos con interferón- α , tan sólo se consigue que menos de un 20% de los enfermos respondan al mismo, mientras que en combinación con ribavirina el porcentaje de respondedores aumenta a un 30% aproximadamente. Últimamente parece que se obtienen mejores resultados con el interferón- α pegilado en combinación con la ribavirina. El interferón- α pegilado presenta un residuo de polietilenglicol unido covalentemente al interferón- α que retarda la eliminación del fármaco, de forma que se consiguen unos niveles más constantes del mismo en sangre con dosis menos frecuentes. La respuesta sostenida al fármaco se consigue en 24-48 semanas en un 40-50% de los pacientes (Zeuzem *et al.* 2000; Zein y Zein 2002)

El interferón- α es una glicoproteína natural que es segregada por la célula en respuesta a una infección viral y produce simultáneamente una activación de la

respuesta inmunitaria del hospedador y la inhibición de la multiplicación del virus. Actúa uniéndose a su receptor de membrana celular, e induciendo la proteinquinasa celular PKR, que fosforila el factor eIF2 α , iniciando una cascada de reacciones (Figura I.5) que desembocan en el aumento de la expresión de ciertos genes, induciendo ciertas actividades celulares entre las que se incluye :

a) Disminución de la presencia del receptor CD81 en la membrana de las células hepáticas, responsables de la unión con la proteína de la envuelta viral E2, dificultando así el anclaje de las partículas víricas (Kronenberger *et al.* 2001).

b) Aumento de la capacidad de las células infectadas para presentar los antígenos víricos a los linfocitos T-CD8⁺ y activar así a las células asesinas naturales (NT) para la eliminación de las células diana.

c) Inhibición de la traducción dirigida por IRES (Chung *et al.* 2001; Shimazaki *et al.* 2002) y por tanto la replicación de los virus en las células infectadas.

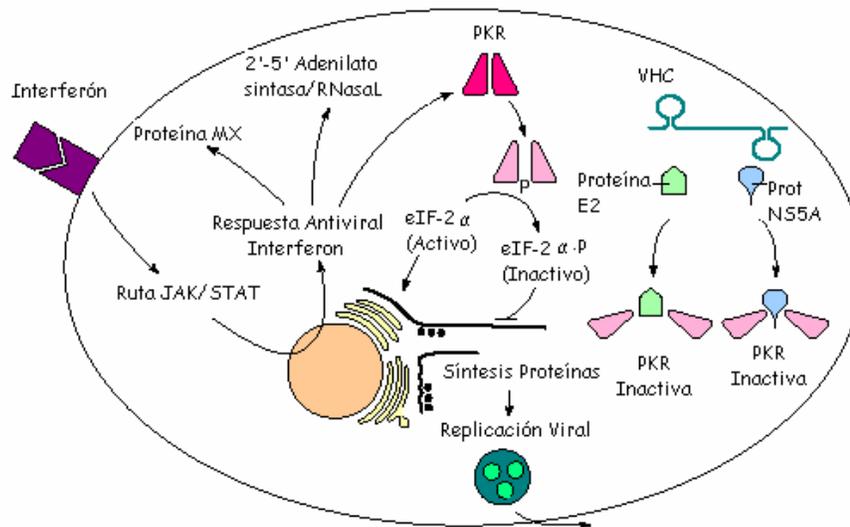


Figura I.5. Mecanismo de acción del interferón. Cuando el interferón se une a su receptor celular se activa la transcripción de RNAs que codifican para un rango de proteínas antivirales entre las que se encuentra la PKR. La PKR, mediante la fosforilación del factor eIF2 α , impide la replicación del virus. En la figura también se muestra la acción de las proteínas virales NS5A y E2, inactivando la acción de la PKR.

La acción antiviral de la ribavirina se debe a que la RpRd del virus cataliza la incorporación de ribavirina trifosfato frente a C y U, mutagenizando el genoma vírico y bloqueando además la elongación de la cadena de RNA, disminuyendo el número de virus infecciosos (Crotti *et al.* 2000; Maag *et al.* 2001). La ribavirina, además, promueve la respuesta inmune mediada por las células Tc tipo 1, potenciando la actividad del IFN- α .

Aunque no todos los autores parecen llegar a la misma conclusión, se ha sugerido que la heterogeneidad del virus está relacionada con la respuesta a tratamiento: a más heterogeneidad, menor respuesta. Asumiendo esta observación, podría llegar a predecirse si un paciente responderá o no al tratamiento, antes de someterse al mismo (Farci 2001). También se han observado cambios en la heterogeneidad del virus durante el tratamiento, sugiriendo que éste podría estar ejerciendo una presión selectiva sobre el virus (Mizokami *et al.* 1994, Sakuma *et al.* 1996, Yun *et al.* 1996). Análisis antes y después de la terapia muestran cómo se producen cambios en la HVR1 del virus en pacientes tratados, mientras que en los individuos no tratados no se observa tal cambio (Pawlotski *et al.* 1999; Polyak 1998).

8.6 Mecanismos de persistencia viral

En el 80 % de los casos el VHC establece una infección persistente (altos niveles de ALT durante más de 6 meses). Aunque los mecanismos por los cuales esto sucede aún no se han esclarecido, sí que se han sugerido algunos. El principal mecanismo de defensa que se postula es que la gran variabilidad del VHC permite a éste escapar de la respuesta inmune del hospedador. Tras la infección del VHC, se desencadena una respuesta inmune humoral y celular contra diferentes proteínas virales. La proteína E2 ha resultado ser una diana de la respuesta humoral y más en concreto la HVR1, la cual podría contener epítomos contra diferentes anticuerpos neutralizantes. Por tanto la presencia de mutaciones en esta región engañaría al sistema inmune permitiendo la persistencia del virus, ya que, además, la eliminación del virus por parte de estos anticuerpos es más lenta que la aparición de nuevos mutantes (Weiner *et al.* 1996). Aunque algunos autores no están del todo de acuerdo con esta correlación, hay que destacar que los no conformes llevaron a cabo experimentos con chimpancés, y se ha

encontrado que la infección del VHC en estos últimos es mucho más débil que en humanos, por lo que la comparación no sería muy significativa (Forns *et al.* 1999).

Aparte del escape de la respuesta humoral también se ha observado un escape a la respuesta celular. Varios estudios sugieren que en aquellos individuos que lograban superar la enfermedad se observaba una respuesta por parte de las células T-colaboradoras (CD4+) mucho más vigorosa que en aquellos en los que se cronificaba (Chang *et al.* 2001; Gerlach *et al.* 1999). Algo parecido también se encontraba con respecto a la respuesta por parte de las células T-citotóxicas (CTL) (Koziel *et al.* 1997; Rehermann 2000). El mecanismo por el que se producía este cambio también parece deberse a la heterogeneidad en las secuencias del VHC (Farci 2001).

La proteína core también parece tener un papel en la persistencia viral. Esta proteína podría inhibir la interleucina 12 y la producción de ácido nítrico, impidiendo así la respuesta inmune desencadenada por las células Th1 (Lee *et al.* 2001). Además se ha observado también que la proteína core podría unirse al receptor del complemento gC1qR (implicado en la respuesta temprana del hospedador contra la infección), inhibiendo la respuesta por parte de los linfocitos T (Kittlesen *et al.* 2000).

Otro de los mecanismos implicados en la persistencia viral es la resistencia a interferón, a través de la interacción por parte de las proteínas E2 y NS5A del virus, y la proteína celular con actividad antiviral PKR (Figura I.5).

La proteína E2, como hemos visto, posee una secuencia de aminoácidos (residuos 659-670) asociada a la región C-terminal, que parece ser responsable de la inhibición de la PKR. Esta secuencia de aminoácidos, llamada “*E2 phosphorylation homology domain*” (E2-PHD), es similar a la encontrada en el sitio de autofosforilación de la PKR y en el factor eIF2 α (diana de la PKR). Al unirse el E2-PHD a la PKR inhibiría su función (Taylor *et al.* 1999). No obstante, hay autores que no están del todo de acuerdo con este hecho, ya que si fuese cierto supondría que aquellos individuos que no respondieran al tratamiento con interferón poseerían una secuencia diferente en el E2-PHD, sin embargo hay estudios que revelan que apenas hay variación en esta región (Boulestin *et al.* 2002). Además otros estudios parecen indicar que la inhibición de la PKR parece ser un mecanismo más complejo que una simple interacción de un seudosustrato intermolecular (Taylor *et al.* 1999).

En cuanto a la proteína NS5A, como también hemos visto posee una región llamada ISDR, localizada entre los residuos 2209 y 2248 de la poliproteína. Esta región parece interactuar con el dominio de dimerización de la PKR (esta interacción es diferente a la propuesta para la región de la proteína E2), impidiendo la fosforilación de $eIF-2\alpha$, factor de inicio de la traducción en eucariotas, el cual, una vez fosforilado, inhibe la síntesis proteica celular y por tanto la del VHC (Figura I.6). De este modo el virus consigue salvar los efectos del interferón.

Aquellos mutantes del virus que presenten mutaciones en la ISDR serán susceptibles al interferón, ya que no podrá inhibir la actividad de la PKR (Enomoto *et al.* 1996; Gale *et al.* 1999; Kurosaki *et al.* 1997). Este hecho fue observado en experimentos realizados en Japón; no obstante estudios similares en Europa y Estados Unidos no parecen encontrar esta correlación, lo cual ha sugerido que la discrepancia podría encontrarse en diferencias genéticas entre los VHC de Japón, Europa y Estados Unidos (Nakano *et al.* 1999).

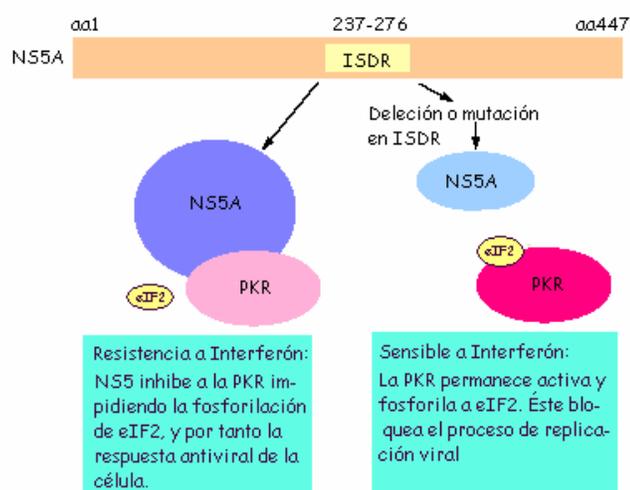


Figura I.6. Regulación de la PKR por la proteína viral NS5A. Este modelo sugiere que la región ISDR, que se encuentra en la NS5A, impide la respuesta frente al interferón, ya que la NS5A inhibe a la PKR, impidiendo la fosforilación de $eIF2\alpha$, y de este modo la inhibición de la replicación viral. Por el contrario, si la ISDR presenta alguna mutación, no puede inhibir a la PKR y de este modo se inhibe la replicación viral.

II. Objetivos

Debido al gran número de personas infectadas en todo el mundo con el VHC, ya hemos visto que éste virus se ha hecho merecedor de una atención y seguimiento especial por parte de las administraciones públicas. También hemos visto que, debido a la gran heterogeneidad del mismo, se distinguen hasta 6 genotipos y más de 50 subtipos, los cuales presentan diferencias en cuanto a su prevalencia, distribución geográfica y rutas de transmisión. Sin ir más lejos, en la Comunidad Valenciana, el VHC está presente en casi el 3% de la población (Mijas L. 1995). Los genotipos mayoritarios presentes, corresponden al 1b y 1a, siendo el 1b mucho más prevalente que el segundo (Zúñiga A. 2001; Ortega 2002). Así pues se plantea el principal objetivo de esta tesis: ¿Qué hace al genotipo 1b ser más prevalente en la Comunidad Valenciana que el 1a? Para ello nos proponemos cubrir tres objetivos adicionales encaminados a responder a esta pregunta:

1. Obtención, secuenciación y análisis de polimorfismos de un gran número de clones correspondientes a un fragmento de 472 nucleótidos entre las regiones pertenecientes a las proteínas E1 y E2, y otro gran número correspondiente a un fragmento de 743 nucleótidos perteneciente a la proteína NS5A.
2. Estudiar la historia demográfica en ambos genotipos, por medio del análisis filogenético aplicando métodos basados en máxima verosimilitud y en la teoría de la coalescencia.
3. Evaluar las fuerzas selectivas y/o estocásticas que pueden estar actuando en la evolución de ambos genotipos.

III. Material y Métodos

1. MATERIAL

Para la elaboración de este trabajo se dispuso de 50 muestras correspondientes a individuos infectados con VHC. Las tablas III.1 y III.2 muestran algunos datos. Los virus de estos pacientes pertenecían a dos genotipos de elevada prevalencia en la Comunidad Valenciana: el genotipo 1b y el genotipo 1a, disponiendo de 25 muestras de cada genotipo. Todas ellas fueron proporcionadas por 3 hospitales de la comunidad valenciana: Hospital Clínico de Valencia, Hospital. General de Valencia y Hospital. General de Alicante. Estos hospitales nos facilitaron, además, información clínica acerca de algunos de los pacientes en lo referente a prácticas de riesgo al que pertenecen, edad, sexo, fecha de contagio y si fueron sometidos o no a tratamiento con interferón. Con respecto a la fecha de contagio decir que la fecha exacta es desconocida para todos ellos. En algunos casos se hace referencia a un intervalo de tiempo de unos 10 años como fecha más probable de la infección (señalado con asterisco en la tabla III.2), y en el resto, tan sólo se indica la fecha en que fue diagnosticada la enfermedad. En cuanto a las prácticas de riesgo, se hace referencia a la posibilidad de que los pacientes estuvieran ligados de algún modo a las diferentes vías por las que puede transmitirse la enfermedad. Los individuos incluídos en este trabajo pertenecen a diferentes prácticas de riesgo entre los que se encuentran: adictos a drogas vía parenteral (ADVP), personas que alguna vez fueron sometidas a transfusiones de sangre, donantes de plasma, personas relacionadas con brotes de VHC anteriores y/o relación familiar con una persona ya infectada. Acerca de la edad de los pacientes, ésta oscila entre los 21 y los 73 años. Y por último, este trabajo se ha llevado a cabo sólo con muestras de pacientes antes de someterse a tratamiento con interferón.

Dos regiones del genoma del virus han sido utilizadas para llevar a cabo este estudio: la comprendida entre la proteína E1-E2, con una extensión de 472 nucleótidos y que contiene a la región hipervariable HVR1, y otra región de 743 nucleótidos de la proteína NS5A que contiene a la región ISDR (Figura III.1).

Las muestras se han nombrado mediante un número que corresponde al orden de llegada a nuestro laboratorio, y delante del número llevan una letra que hace referencia al hospital del que proceden, “A” para el H. General de Alicante, “C” para el H. Clínico de Valencia y “G” para el H. General de Valencia. Con posterioridad añadiremos una segunda letra minúscula (delante de la que hace referencia al hospital del que procede)

que indica con qué región estamos trabajando. Una “e” para la región E1-E2 y una “n” para la NS5A. Por ejemplo las secuencias “eC25”, corresponde al conjunto de clones obtenidos para la región E1-E2, y procede del individuo 25 del Hospital Clínico de Valencia.

En las tablas también se indica el número de secuencias o clones obtenidos para cada uno de los pacientes y para las dos regiones (E1-E2 y la NS5A) utilizadas en esta tesis. La obtención de estas secuencias se explica más detenidamente en el apartado de métodos.

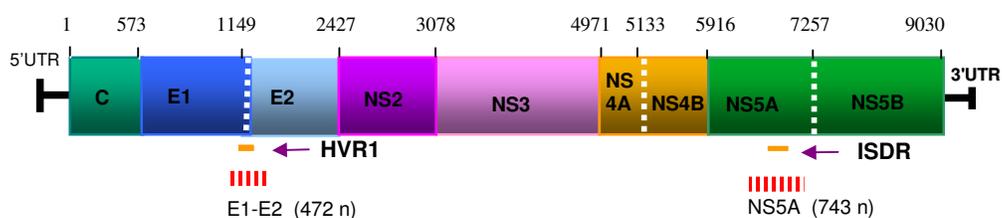


Figura III.1. Esquema del genoma del VHC. La longitud total de la poliproteína es de 9030 nucleótidos. Los fragmentos a rayas rojas indican las regiones clonadas y secuenciadas en esta tesis. Las rayas naranjas muestran, en relación al genoma y los fragmentos utilizados, donde se localizan las regiones HVR1 e ISDR.

Tabla III.1. Muestras de genotipo 1b empleadas en este estudio. Además, se indica el hospital de donde procede cada muestra (primera columna), el número de clones utilizados en nuestro análisis tanto para la región E1-E2 como para la NS5A (tercera y cuarta columna), la fecha de diagnóstico (quinta columna), la práctica de riesgo del paciente (sexta columna) y el sexo y la edad del paciente (séptima y octava columna respectivamente).

Hospital	Individuo Genot. 1b	Nº Clones E1-E2	Nº Clones NS5A	Fecha Diagnostico	Práctica Riesgo	Sexo	Edad
H. General Alicante	A03	100	38	??	??	??	??
	A06	100	96	??	??	??	??
	A14	100	58	2000	No	Mujer	35
	A16	104	52	??	??	??	??
	A35	110	27	??	??	??	??
H. Clínico Valencia	C03	100	86	1996	ADVP	Varón	28
	C04	100	92	1999	??	Varón	34
	C05	100	75	1999	??	Varón	51
	C08	101	77	1998	??	Varón	65
	C09	111	57	1993	??	Varón	54
	C10	100	63	1994	Donante plasma	Varón	38
	C12	101	92	1996	??	Varón	31
	C13	100	69	1999	??	Mujer	46
	C15	102	80	1998	Transfusión	Mujer	58
	C16	100	74	1992	??	Varón	45
	C19	100	73	1999	ADVP	Varón	51
H. General Valencia	C25	100	37	1998	Transfusión	Varón	57
	G02	100	65	??	??	Mujer	??
	G05	100	62	??	??	Varón	??
	G06	100	61	??	??	Varón	??
	G07	100	49	??	??	Varón	??
	G09	100	58	??	??	Varón	??
	G16	100	57	??	??	Varón	40
	G17	101	68	??	??	Mujer	53
	G26	100	84	??	??	Mujer	62
Total		2530	1650				

ADVP: Adictos a drogas por vía parenteral.

?: Dato desconocido.

No: No pertenece a ninguna práctica de riesgo.

Tabla III.2. Muestras de genotipo 1a empleadas en este estudio. Además, se indica el hospital de donde procede cada muestra (primera columna), el número de clones utilizados en nuestro análisis tanto para la región E1-E2 como para la NS5A (tercera y cuarta columna), la fecha de diagnóstico o en los casos en los que lleva asterisco la posible fecha de contagio (quinta columna), la práctica de riesgo del paciente (sexta columna) y el sexo y la edad del paciente (séptima y octava columna respectivamente).

Hospitales	Individuo. Genot. 1a	Nº Clones E1-E2	Nº Clones NS5A	Fecha Diagnóstico	Grupo Riesgo	Sexo	Edad
H. General Alicante	A05	100	25	2001	ADVP.	Varón	25
	A09	100	32	1995	No	Varón	35
	A20	113	67	2001	ADVP.	Varón	21
	A26	100	62	1999	ADVP.	Mujer	28
	A27	100	73	2001	No	Varón	27
	A34	100	29	??	??	??	??
H. Clínico Valencia	C01	100	65	1987	Donante plasma	Varón	35
	C02	100	96	1989	ADVP	Varón	36
	C06	100	76	1999	??	Varón	34
	C07	101	63	1997	ADVP.	Varón	41
	C11	110	92	1999	ADVP.	Varón	28
	C14	101	29	1999	Brote	Varón	26
	C17	100	94	1997	Brote	Mujer	57
	C18	100	63	1999	??	Mujer	44
	C20	100	91	1990	??	Varón	37
	C21	100	80	1995	ADVP.	Mujer	33
	C22	100	41	1980	ADVP.	Varón	40
	C24	100	51	1997	??	Varón	43
	C28	100	48	1990	ADVP.	Mujer	32
	C31	101	76	1990	??	Varón	42
C33	100	51	1997	??	Varón	23	
H. General Valencia	G10	100	28	1990-98*	ADVP.	Varón	
	G14	100	84	??	??	Varón	
	G19	100	42	1990-98*	ADVP.	Mujer	33
Total		2526	1548				

ADVP: Adictos a drogas por vía parenteral.

?: Dato desconocido.

No: No pertenece a ninguna práctica de riesgo.

Brote: Muestras que proceden de un brote epidemiológico.

*: Posible fecha de contagio.

2. ESQUEMA METODOLÓGICO

Con respecto a la metodología, podríamos dividirla en dos partes: la correspondiente a todos los procesos conducentes a la obtención de las secuencias de DNA y aquella que se encarga de los análisis genético-poblacionales de las mismas. La figura III.2 muestra un diagrama con los métodos empleados en esta tesis y su orden de aparición.

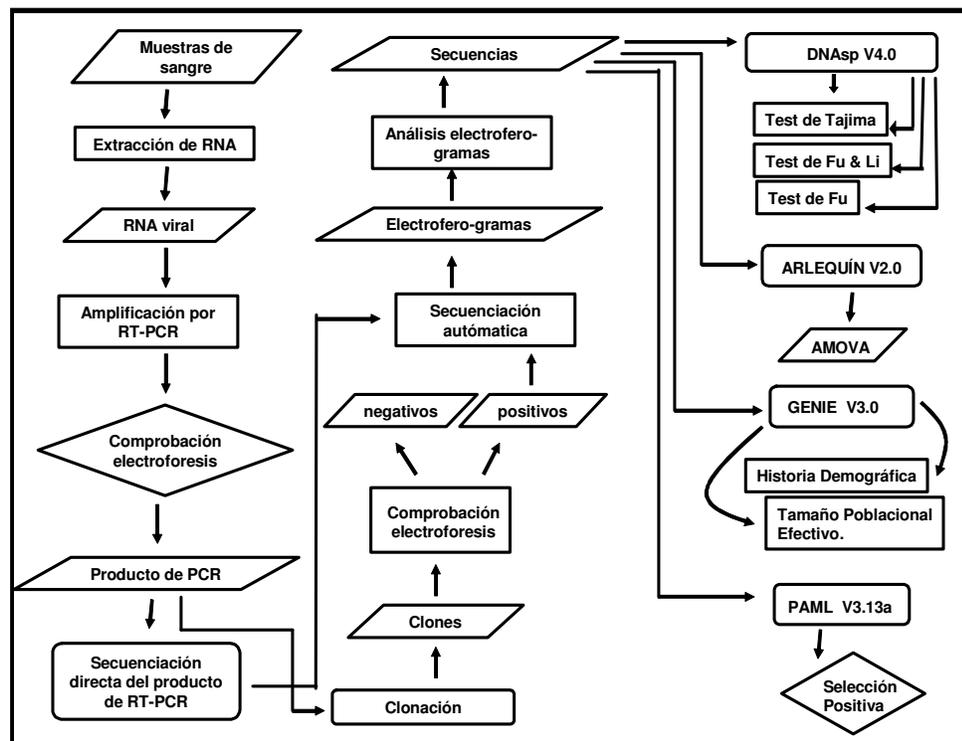


Figura III.2. Esquema de los métodos empleados para la obtención y análisis de las secuencias del VHC.

3. EXTRACCIÓN VIRAL DE RNA Y AMPLIFICACIÓN

Para la amplificación de las dos regiones se han seguido los pasos que se detallan a continuación.

3.1. Extracción de RNA

La extracción del RNA viral es un proceso sobre el cual hay que tener un especial cuidado para evitar la contaminación con DNA y ribonucleasas. Para ello se tomaron una serie de precauciones tales como: realizar la extracción en un lugar libre de ribonucleasas y alejado de amplificadores de DNA, cambiar frecuentemente de guantes durante el proceso, utilizar un juego de pipetas sólo para este fin, mantener los viales en hielo a 4°C durante la extracción, etc. Una forma de comprobar que en este paso no se ha producido ninguna contaminación es utilizar durante todo el proceso de extracción un control negativo, un vial con agua al que aplicamos los mismos pasos que al resto de viales.

El RNA viral fue extraído a partir de 140 µL de suero usando *High Pure Viral RNA Kit* (Roche) y siguiendo el protocolo indicado por el fabricante.

3.2. Transcripción inversa

A partir del RNA aislado, se obtuvo cDNA mediante transcripción inversa utilizando hexámeros. La reacción fue llevada a cabo en 20 µL, y contenía:

- 5 µL de RNA eluído.
- 4 µL de tampón de reacción 5x.
- 500 µM de cada dNTP (Promega).
- 0.5 µg de hexámeros (Promega).
- 100 unidades de MMLV Transcriptasa Reversa (USB™, Promega).
- 20 unidades de inhibidor de ribonucleasas (*RNasin Ribonuclease Inhibitor*, Promega).

La reacción fue incubada 45 min a 42°C y 3 min a 95°C.

3.3. PCR

En este paso también se presta una especial atención a las posibles contaminaciones por amplificadores de DNA. Por tanto la PCR también se prepara en un lugar diferente de donde se manipulan amplificadores de DNA, también cambiamos frecuentemente de guantes (uno para cada individuo), utilizamos un juego diferente de pipetas y también llevamos a cabo las reacciones a una temperatura de 4°C. Para comprobar que en este paso no se ha producido una contaminación utilizamos un segundo control negativo con agua.

El cDNA fue utilizado para la amplificación por PCR de las regiones de interés. Cuando en una primera ronda de PCR no se detectaba ningún amplicón, entonces se llevaban a cabo PCRs anidadas. La PCR se realizaba en un volumen de 100 µL que contenía:

- 10 µL de producto de transcripción reversa.
- 200 µM de cada dNTP.
- 400nM de cebador antígenómico.
- 400nM de cebador genómico.
- 2.5 unidades de *Pfu* DNA polimerasa (Promega).

En caso de no obtener ningún amplicón, la reacción de PCR secundaria se efectuaba igual que la primaria, pero añadiendo 1 µL del producto de la PCR primaria en lugar de los 10 µL del producto de retrotranscripción. Los cebadores empleados eran específicos para cada genotipo (1a o 1b) y región (E1-E2, NS5A). La nomenclatura de los cebadores indica con “1a-” o “1b-” el genotipo para los que son empleados, seguido de una letra que indica la región del genoma que amplifican (“E” para la región E1-E2 y “N” para la región NS5A), y otra letra que indica el sentido de elongación (“g” para el sentido genómico y “a” para el antígenómico). Por último se indicaba con “1” y “2” si se empleaban para PCR primaria o secundaria respectivamente.

Las parejas de cebadores utilizadas para la región E1-E2 en la primera ronda de PCR (Tabla III.3) fueron: 1a-Eg1 (o 1b-Eg1) y 1a-Ea (o 1b-Ea). Para la región de la NS5A, los cebadores utilizados fueron: 1b-Ng1 y 1b-Na para los 1b y B-1a-Ng1 (o L-1a-Ng1) y B-1a-Na (o L-1a-Na) para los 1a. En el caso de que fueran necesarias las

PCRs anidadas los cebadores utilizados fueron: 1a-Eg2 (o 1b-Eg2) y 1a-Ea (o 1b-Ea) para la región E1E2, y para la región NS5A 1b-Ng2 y 1b-Na para los 1b y B-1a-Ng2 (o L-1a-Ng2) y B-1a-Na (o L-1a-Na) para los 1a.

Tabla III.3. Cebadores utilizados en las diferentes rondas de PCR realizadas para amplificar las regiones E1-E2 y NS5A. La nomenclatura utilizada para los cebadores hace referencia al genotipo (1a o 1b), a la región que se va a amplificar (“E” para la región E1-E2 y “N” para la región NS5A) y al sentido de elongación (“g” para el sentido genómico, y “a” para el antígenómico). B y L diferencia dos conjuntos de cebadores alternativos, diseñados para una misma región.

Región	Cebador	Posición nucleotídica ^a	Secuencia 5'→ 3' ^b
E1 -E2	1a-Eg1	1278-1297	<i>CGCATGGCATGG</i> <u><i>RATATGAT</i></u>
	1a-Eg2	1288-1309	<u><i>GGRATATGATGATGAACTGGTC</i></u>
	1a-Ea	1861-1842	<i>GGAGTGAAGCARTATACTGG</i>
	1b-Eg1	1278-1297	<i>CGCATGGCYTGGG</i> <u><i>AYATGAT</i></u>
	1b-Eg2	1288-1309	<u><i>GGGATATGAT</i></u> <i>TRATGAA</i> <u><i>YTGTC</i></u>
	1b-Ea	1861-1842	<i>GGGGTGAARCARTAYACYGG</i>
NS5A	1b-Ng1	6703-6727	<i>TGGAYGGRGTRCGGYTGCAC</i> <u><i>AGGTA</i></u>
	1b-Ng2	6722-6741	<u><i>CAGGTACGCTCCRGYRTGCA</i></u>
	1b-Na	7507-7491	<i>CCYTCRAGGGGGGGCAT</i>
	B-1a-Ng1	6703-6727	<i>TGGAYGGGGTGC</i> <u><i>CGCTACATAGGTW</i></u>
	B-1a-Ng2	6722-6741	<u><i>TAGGTWYGC</i></u> <i>SCCCCCYTGCA</i>
	B-1a-Na	7507-7491	<i>CCCTCSAGRGGGGGCAT</i>
	L-1a-Ng1	6703-6727	<i>TGGACGGGGTGYGMCTRCAY</i> <u><i>AGGTT</i></u>
	L-1a-Ng2	6722-6741	<u><i>TAGGTTYGC</i></u> <i>CCCCCYTGCA</i>
	L-1a-Na	7507-7491	<i>CCYTCSARGGGRGGCAT</i>

^aPosiciones nucleotídicas referidas a la secuencia D50481

^bLos nucleótidos subrayados indican las posiciones solapantes de los cebadores anidados.

En todos los casos la PCR fue realizada en un termociclador (Perkin Elmer 2400) conforme al siguiente protocolo:

- Desnaturalización inicial a 94°C durante 3 min:
- 94°C durante 30 s. } x 5 ciclos
- 55°C durante 30 s. }
- 72°C durante 3 min. }
- 94°C durante 30 s. } x 35 ciclos
- 52°C durante 30 s. }
- 72°C durante 3 min. }
- Extensión final de 10 min. a 72°C

La banda amplificada se observó en un gel de agarosa de 1.4% con bromuro de etidio. En este gel también podíamos observar si habíamos tenido alguna contaminación, bien fuera en el paso de extracción o en el de PCR.

3.4. Purificación de productos de PCR

Los productos de DNA de la PCR para la región E1E2 y la región NS5A fueron purificados con *High Pure PCR product Purification Kit* (Roche), siguiendo el protocolo indicado en el manual del fabricante.

4. CLONACIÓN Y SECUENCIACIÓN

Los productos de DNA ya amplificados y purificados fueron posteriormente clonados y secuenciados, para lo que se siguieron los siguientes pasos.

4.1. Ligación

La ligación se llevó a cabo en un volumen de 10 μ L que contenía:

- 6 μ L de DNA.
- 1 μ L de ligasa (Promega).
- 2 μ L de tampón de reacción.
- 10 ng de plásmido *pBluescript II SK (+) phagemid* (Stratagene) digerido con *EcoRV*.

4.2. Transformación

De las mezclas de ligación anteriores se añadían 5 μ L a 75 μ L de células competentes *XLI-Blue MRF* de *E. coli*, obtenidas mediante el protocolo de CaCl_2 . Tras dejar en hielo 30 minutos se realizaba un choque térmico a 42°C durante 1 minuto y 30 segundos. Posteriormente se añadía a la mezcla de transformación 700 μ L de LB y se incubaba durante 1 hora. Después se sembraban 100 μ L en placas de LB que contenían Ampicilina, X-gal e IPTG y se incubaban toda la noche a 37°C.

4.3. PCR de colonia

Para identificar los clones que contenían nuestro fragmento, una vez realizada la transformación, utilizamos la PCR de colonia. Este tipo de PCR permite amplificar un fragmento de DNA clonado a partir de una colonia bacteriana sin necesidad de crecer en medio líquido y purificar plásmidos recombinantes. También en este paso se tomaron precauciones para evitar las contaminaciones, con este fin las reacciones se preparaban en una zona específica para ello, utilizando puntas con filtro y pipetas destinadas sólo para realizar las PCR. Una vez preparadas las reacciones y en otra zona del laboratorio, el proceso se realizaba como se explica a continuación. A la mezcla de reacción de la PCR, se le añadía, con un palillo estéril, una colonia bacteriana blanca de *E. coli* (contienen el fragmento que queremos secuenciar) obtenida en el paso 4.2 y se utilizaban los cebadores KS y SK (Stratagene), basados en las regiones flanqueantes del

vector que hemos utilizado (*pBluescript II SK (+)*). Finalmente la mezcla de reacción de un volumen final de 50 μL , contenía:

- 43 μL de H_2O .
- 5 μL de tampón de reacción 10x.
- 0.1 mM de dNTPs.
- 0.2 μM de cebador KS y SK.
- 0.05 u/ μL de *Taq* polimerasa (Promega).
- Una colonia bacteriana positiva.

La reacción de la PCR de colonia se efectuaba en placas de 96 pocillos y se llevaba a cabo en un termociclador (Perkin Elmer 9600) conforme al siguiente protocolo:

- 94°C durante 5 min.
- 94°C durante 30 s. } x 25 ciclos
- 55°C durante 30 s. }
- 72°C durante 30 s. }
- Extensión final de 72°C durante 7 min.

Este tipo de PCR ahorra un día de trabajo. De no hacerlo así, tras la transformación tendríamos que picar las colonias positivas (blancas) y ponerlas a crecer en LB durante toda la noche, y al día siguiente realizar la miniprep o purificación del plásmido recombinante. Como se puede apreciar en la figura III.3, los resultados obtenidos con este tipo de PCR son bastante buenos y tampoco supone problemas para los pasos sucesivos.

Las bandas correspondientes al fragmento que esperábamos fueron observadas en un gel de agarosa al 1,4% (Figura III.3).

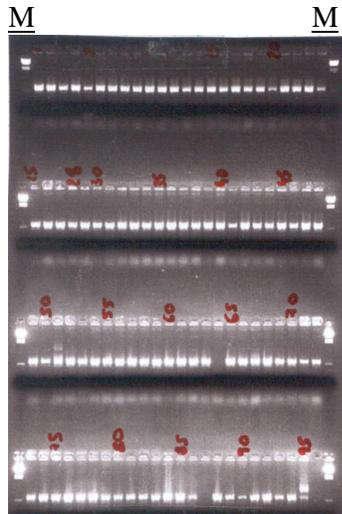


Figura III.3. Bandas correspondientes a las regiones amplificadas tras realizar la PCR de colonia. M: marcador 2 (fago lambda digerido con Hind III)

Los productos de PCR que contenían nuestro inserto, eran finalmente purificados en placas de 96 pocillos mediante un protocolo de precipitación de DNA basado en cloruro sódico, acetato amónico y etanol, tal como sigue: a cada uno de los pocillos se le añadían una mezcla que contenía 200 μL de etanol 100%, 32 μL de acetato amónico 5M, 2 μL de NaCl 5M y 8 μL de H_2O . Se mantiene a temperatura ambiente 30 minutos, y transcurrido este tiempo se centrifuga durante 30 minutos más a 4000 rpm. Después de la centrifugación se decanta y se añaden 100 μL de etanol 70% y se centrifuga 10 minutos a 4000 rpm. Seguidamente se seca el precipitado en una bomba de vacío y se resuspende en 15 μL de LTE (1mM Tris-HCl (pH 8,0), 0,1 mM EDTA (pH 8,0)).

Así, según la metodología anterior, obtuvimos finalmente alrededor de 100 clones de la región E1-E2 y un número que oscilaba entre 25 y 100 para la NS5A, ya que esta región corresponde a una zona del genoma menos variable.

4.4. Secuenciación

Los productos de PCR de colonia ya purificados fueron utilizados para realizar las reacciones de secuenciación. Estas reacciones se efectuaron sobre molde en dirección genómica y antígenómica con los cebadores KS y SK, basados en el vector plasmídico empleado. El reactivo utilizado para este paso fue el *ABI PRISM BigDye Terminator v3.0 Ready Reaction Cycle Sequencing KIT* (Applied Biosystems). Según el protocolo de este reactivo se aconseja utilizar 8 μL de la mezcla que contenía la *Taq* polimerasa y los terminadores. No obstante nosotros utilizamos una dilución 1:20 del volumen de la mezcla de *Byg Dye Terminator Ready Reaction*. Para realizar esta dilución, que debía prepararse cada vez que efectuáramos las reacciones, nos servimos del tampón rxn 5x (400mM TrisHCl pH9.0, 10 mM MgCl_2). Por ejemplo, para 5 reacciones preparábamos: 2 μL de *Big Dye mix* + 8 μL de tampón rxn 5x. Después para cada una de las reacciones tomábamos 2 μL del volumen final de la mezcla anterior. Desde el momento de su preparación, la mezcla se mantenía en hielo.

Finalmente, para cada reacción de secuenciación se preparaba la siguiente mezcla en un volumen final de 8 μL :

- 2 μL de *Big Dye Mix* (dilución 1:20 del stock ABI).
- 400 ng de DNA molde.
- 5 μM de cebador (KS o SK).
- dH₂O (cantidad suficiente).

Con este sistema ahorrábamos gran cantidad de reactivo, ya que con el volumen que el fabricante recomienda para realizar la reacción de secuenciación de una muestra nosotros preparábamos 20. La única diferencia con respecto al protocolo aconsejado se encuentra en el número de ciclos necesarios en la reacción de secuenciación. Según las condiciones sugeridas por Applied Biosystems, son necesarios 25 ciclos. Pues bien, utilizando la dilución 1:20 se necesitan 99 ciclos. Esto supone que la reacción de secuenciación tardará mucho más tiempo, no obstante el problema se soluciona dejando este paso para el final de la tarde, realizando la reacción durante la noche.

Las reacciones de secuenciación se preparaban en placas de 96 pocillos y se llevaban a cabo en un termociclador (Perkin Elmer 9600) conforme al siguiente protocolo:

- 95°C durante 2 min.
 - 95°C durante 10 s.
 - 50°C durante 5 s.
 - 60°C durante 4 min.
 - 4°C ∞
- } x 99 ciclos

Los productos de esta reacción se precipitaron con etanol y NaOAc siguiendo el protocolo propuesto por *Applied Biosystems*. Después fueron enviadas al Servicio de secuenciación de proteínas y ácidos nucleicos de la Universidad de Valencia, donde las muestras eran procesadas con el secuenciador automático ABI Prism 377.

Las secuencias obtenidas fueron revisadas y ambas cadenas ensambladas (sentido genómico y antígenómico) con el paquete de programas Staden (Staden *et al.* 1999).

La calidad de las secuencias obtenidas, tanto usando la PCR de colonia como la dilución 1:20 en las reacciones de secuenciación no difieren de las conseguidas por el método tradicional (aislamiento y purificación del plásmido tras la transformación y utilización del protocolo propuesto por *Applied Biosystems* para las reacciones de secuenciación).

5. ANÁLISIS DE SECUENCIAS

5.1. Alineamientos

Las secuencias ya revisadas y ensambladas fueron alineadas utilizando el programa CLUSTALX 1.81 (Thompson *et al.* 1997).

5.2. Inferencia de la historia demográfica del VHC y cálculo del tamaño poblacional efectivo

La historia demográfica de cada individuo fue estudiada con el programa GENIE V3.0. (de *Genealogy Interval Explorer*; Pybus y Rambaut 2002). Es un programa en C++ en el que los parámetros demográficos son calculados a partir de una reconstrucción filogenética cuya longitud de ramas es proporcional al tiempo. El método utilizado permite contrastar qué dinámica poblacional se ajusta mejor a los datos, entre hipótesis demográficas alternativas, por máxima verosimilitud. El método se basa en el modelo de tamaño poblacional variable de la teoría de la coalescencia (Griffiths y Tavaré 1994).

El modelo asume varios supuestos con respecto a la población objeto de análisis, concretamente: 1) muestreo al azar; 2) ausencia de recombinación, selección y subdivisión poblacional; 3) tasa de sustitución constante en todas las genealogías. GENIE V3.0 lleva a cabo dos aproximaciones para inferir la historia demográfica. La primera implica una estimación no paramétrica del modelo demográfico, mediante la representación gráfica del cambio con el tiempo del tamaño efectivo (*skylineplot*), es decir, para el intervalo de tiempo transcurrido en un internodo se calcula cuál sería la media armónica de $N_e(t)$ (siendo $N_e(t)$, el tamaño poblacional efectivo a un tiempo t antes del presente). La segunda implica el cálculo, mediante máxima verosimilitud, de una serie de parámetros demográficos (estimación paramétrica) para diferentes modelos de historia demográfica. Se contemplan 7 posibles modelos demográficos que pueden ser continuos o escalonados, a saber: tamaño poblacional constante, crecimiento exponencial, crecimiento con fase constante-exponencial-constante (con-exp-con), crecimiento en expansión, crecimiento logístico, crecimiento con fase con-exp-con escalonado, crecimiento en expansión escalonado y crecimiento logístico escalonado (Figura III.4). Los modelos son comparados mediante el criterio de información de

Akaike corregido (AICc) (Hurvich y Tsai 1989). El modelo que maximice el índice AICc será el que mejor se ajuste a nuestros datos.

Las estimaciones realizadas por ambos planteamientos (paramétrico y no paramétrico) deberían corresponderse estrechamente cuando el modelo correcto es escogido.

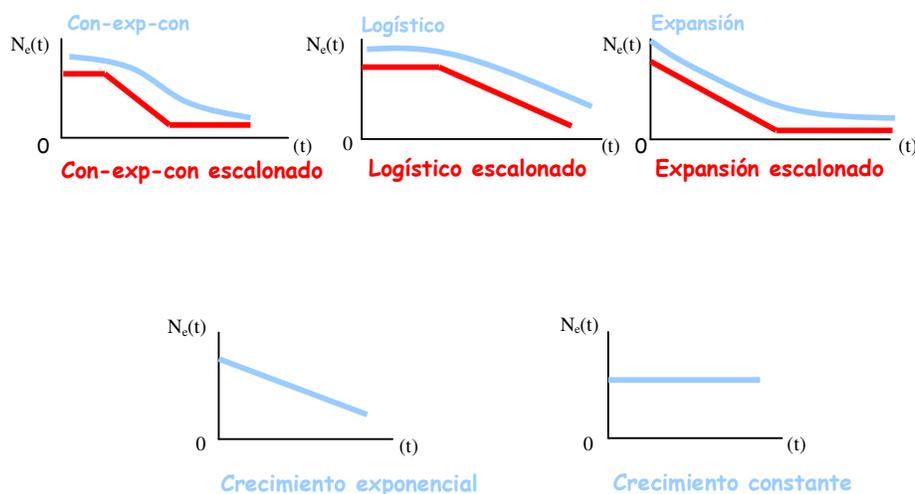


Figura III.4. Modelos demográficos implementados por GENIE. En el eje de ordenadas se representa el tamaño poblacional efectivo ($N_e(t)$) y en el de coordenadas el tiempo. Las curvas en azul representan los modelos continuos y las curvas en rojo los modelos escalonados. El tiempo presente corresponde a tiempo 0. Con-exp-con: Crecimiento con fase constante-exponencial-constante

Como entrada de datos para el programa GENIE V3.0 se utilizaron árboles de máxima verosimilitud (Ver ANEXO V). Estos árboles fueron inferidos con el programa PAUP*4b10 (de *Phylogenetic Analysis Using Parsimony and other methods*; Swofford 2001). Los algoritmos utilizados en este programa fueron: *Stepwise addition* y *nearest-neighbor interchange* bajo el modelo de sustitución de Hasegawa-Kishino-Yano 85 (HKY 85), asumiendo reloj molecular.

Puesto que el número de individuos y sobre todo el de secuencias de cada uno de ellos (teniendo en cuenta ambas regiones) era considerable, el factor tiempo de computación ha sido importante y determinante en nuestros análisis. Así, los algoritmos empleados en PAUP son los recomendados por el programa para acelerar la obtención de resultados. El modelo de evolución no ha sido calculado previamente para cada

individuo, sino que decidimos asumir HKY 85 en todos los individuos. Por lo que se refiere al reloj molecular, este también se asume ya que era uno de los requerimientos de GENIE V3.0 para llevar a cabo el análisis.

La historia demográfica fue inferida a dos niveles: las secuencias virales de los individuos y los genotipos. Para el primer nivel utilizamos todas las secuencias disponibles por individuo y se obtuvo la historia demográfica de cada uno de ellos. A nivel de genotipo utilizamos dos conjuntos de datos (Figura III.5), uno formado por las secuencias consenso de cada individuo, correspondientes a ambas regiones (E1-E2 y NS5A), y otro formado por las parejas de secuencias cuya distancia fuera máxima también atendiendo a ambas regiones. Es decir, ambos conjuntos estarían formados por 4 grupos de datos: el correspondiente a la región E1-E2 para el genotipo 1a, el correspondiente a la región E1-E2 para el genotipo 1b, el correspondiente a la región NS5A para el genotipo 1a y el correspondiente a la región NS5A para el genotipo 1b. Cada grupo de datos correspondientes al conjunto constituido por las secuencias consenso, estaría formado por 25 secuencias (una secuencia consenso por individuo), mientras que los grupos correspondientes al conjunto constituido por las secuencias cuya distancia fuera máxima estarían formados por 50 secuencias (dos secuencias por individuo).

La nomenclatura utilizada para nombrar a los 4 grupos del conjunto formado por las secuencias consenso es la siguiente: el término “Cons” seguido de una “E” o una “N”, según nos refiramos a la región E1-E2 o a la NS5A respectivamente, y esto a su vez seguido por 1a o 1b, dependiendo del genotipo del que se trate. Así por ejemplo el grupo ConsE1a, hace referencia al grupo de todas las secuencias consenso correspondientes a la región E1-E2 y pertenecientes a los individuos de genotipo 1a. En el caso de los grupos formados por las secuencias más divergentes utilizamos el término “Masdist” seguido de una E o una N, según la región que utilizásemos, y a su vez seguido de 1a o 1b que hacía referencia al genotipo correspondiente.

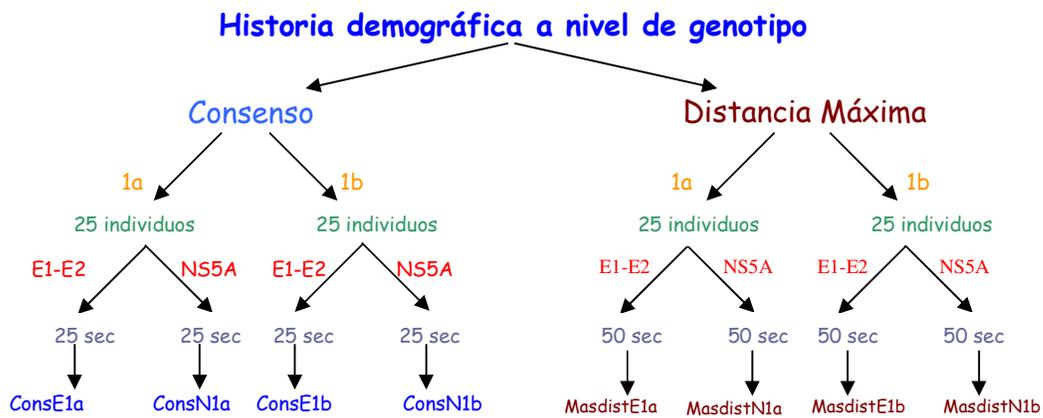


Figura III.5. Esquema de los conjuntos de datos empleados en la inferencia de la historia demográfica a nivel de genotipo.

Las secuencias consenso fueron calculadas empleando el paquete de programas EMBOSS (de *European Molecular Biology Open Software Suite*; Rice *et al.* 2000). Para obtener las parejas de secuencias cuya distancia fuera máxima, en primer lugar se obtuvo el modelo de sustitución nucleotídica que mejor se adecuaba a las secuencias obtenidas para cada paciente, mediante los programas PAUP y Modeltest 3.1. (Posada y Crandall 1998). Con el programa PAUP, se obtiene el valor del logaritmo de la verosimilitud de los datos respecto de 56 modelos de sustitución nucleotídica diferentes. Los valores de verosimilitud son empleados por el programa Modeltest para obtener el modelo que mejor se adecua a los datos, así como una estima de los parámetros que describen dicho modelo. Esto se lleva a cabo mediante la realización de pruebas de razón de verosimilitud o por el criterio de información de Akaike (o AIC; Akaike 1974). El mejor modelo y la estima de los parámetros del mismo fueron empleados para la obtención de una matriz de distancias con las secuencias de cada paciente. Esta matriz se calculó con el programa PAUP, obteniendo las distancias por máxima verosimilitud. A continuación, se buscó para cada paciente aquella pareja de secuencias cuya distancia fuera máxima en las matrices obtenidas.

En GENIE V3.0, para realizar la representación gráfica del cambio en el tamaño efectivo con respecto al tiempo, consideramos lo siguiente. El número de sustituciones que ocurren durante el intervalo de tiempo de un internodo sigue una distribución Poisson. Es decir, el número de sustituciones es proporcional al tiempo transcurrido,

siempre que el intervalo del internodo o la tasa de sustitución sean grandes. Sin embargo, si el producto de la tasa de sustitución por el intervalo de tiempo del internodo es pequeño no se puede aplicar esta aproximación. En ese caso se agrupan los intervalos que sean pequeños, es decir se agrupan varios sucesos de coalescencia, para que el producto de tiempo por tasa de sustitución sea lo suficientemente grande. No obstante, si el número de secuencias iguales que encontramos (intervalo del internodo = 0) para un individuo es muy elevado, también lo es el número de intervalos agrupados. La desventaja de este agrupamiento es que se pierde gran parte de la estructura temporal de los datos y el método puede no resultar apropiado.

Para evitar obtener resultados erróneos con el *Skylineplot*, analizamos la presencia de señal filogenética de los árboles correspondientes a cada paciente, y eliminamos aquellos cuya señal no fuera lo suficientemente grande como para obtener resultados óptimos con GENIE V3.0. Con este fin aplicamos análisis de mapeo por verosimilitud empleando el programa TREE-PUZZLE V5.0. (Schmidt *et al.* 2002) Este programa permite reconstruir árboles filogenéticos mediante máxima verosimilitud empleando el método de *quartet puzzling*. Pero además confecciona un mapa de verosimilitudes que permite visualizar el contenido filogenético de un conjunto de secuencias alineadas. El método se basa en un análisis por máxima verosimilitud de las tres topologías que pueden resolverse completamente a partir de cuatro secuencias (Strimmer y von Haeseler 1997). Las tres verosimilitudes, una por cada topología, se representan como puntos internos de un triángulo equilátero. El triángulo es dividido en diferentes regiones (Figura III.6). Una de las regiones representa evolución en forma de estrella (A^*), tres regiones representan una filogenia bien resuelta (A_1 , A_2 y A_3) y otras tres regiones reflejan la situación donde es difícil diferenciar entre dos de las tres topologías (A_{12} , A_{13} y A_{23}). La localización de las verosimilitudes en el triángulo define el modo de evolución. Si se analizan n secuencias entonces las verosimilitudes para cada conjunto de cuatro, se sitúan dentro del triángulo. La distribución de puntos resultante muestra si los datos presentan suficiente señal o no para realizar una reconstrucción.

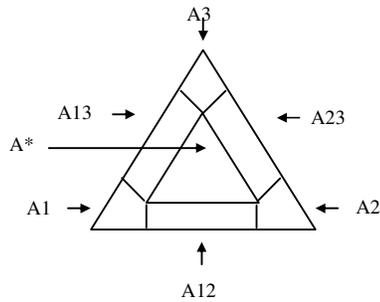


Figura III.6. Mapa de verosimilitudes obtenido con el programa TREE-PUZZLE V5.0. El triángulo está dividido en 7 regiones, donde se distribuyen las verosimilitudes de las posibles topologías por cada set de cuatro secuencias. Las regiones A1, A2 y A3 corresponden a árboles bien resueltos. A* a árboles en forma de estrella y A12, A13, y A23 a zonas donde los árboles no pueden ser resueltos

Finalmente las salidas del GENIE fueron representadas gráficamente utilizando el programa SIGMAPLOT 2001 V7.0.

El programa GENIE V3.0 ofrece además la posibilidad de realizar el cálculo del tamaño poblacional efectivo (N_e). Puesto que en la estima paramétrica de los modelos demográficos uno de los parámetros que calcula es θ , si sabemos la tasa de sustitución nucleotídica (μ) podemos averiguar N_e , ya que: $\theta = 2N_e\mu$. No obstante la estima no paramétrica también calcula N_e para cada intervalo entre los nodos de un árbol, por lo que si observamos las gráficas o *skylineplots* podemos evaluar cómo ha ido cambiando el tamaño poblacional efectivo con el tiempo.

5.3. Pruebas de neutralidad

Bajo la teoría neutral de la evolución todas las mutaciones en una región de DNA son selectivamente neutras, y el mantenimiento del polimorfismo en las poblaciones se debe principalmente a la acción de la deriva genética y la fijación al azar de esas mutaciones neutras o ligeramente deletéreas. Pero ¿cómo podemos contrastar la neutralidad a partir de un conjunto de secuencias? Hemos aplicado diferentes pruebas para contrastar la hipótesis neutra.

5.3.1 Prueba de Tajima

Bajo la teoría neutral, la cantidad esperada de variación genética por nucleótido es $\theta = 2N_e\mu$, donde N_e es el tamaño poblacional efectivo y μ la tasa de mutación. θ puede ser estimado a partir de dos parámetros Π y S . El parámetro Π , corresponde al número de diferencias nucleotídicas entre dos secuencias escogidas al azar de la población, y S corresponde al número de sitios segregantes o número de sitios que presentan variación entre las secuencias de la muestra. En condiciones de neutralidad las estimas de S y Π dan la misma estima de θ , y por tanto la diferencia entre Π y S no es significativamente distinta de 0. No obstante la presencia de la selección natural afecta a ambos parámetros de forma diferente, ya que Π evalúa la frecuencia de alelos mutantes mientras que S examina el número de sitios polimórficos. Por tanto, en presencia de selección natural obtendremos distintos valores de θ . Esto condujo a Tajima (1989) a proponer que la diferencia (D) entre las estimas de θ dadas por Π y S , servirían como prueba para detectar desviaciones de la neutralidad de las mutaciones ($D = 0$ bajo neutralidad). La significación de la prueba vendría dada por:

$$D = \frac{\Pi - S/a_n}{\sqrt{\text{Var}(\Pi - S/a_n)}}$$

donde n es el número de secuencias y $a = \sum (1/i)$ tomando i valores de: $i=1$ a $n-1$. Este último factor convierte a S en un estimador de la diversidad genética que no depende del número de secuencias.

Si nos fijamos en el espectro de frecuencias de polimorfismos (la frecuencia de alelos en cada sitio polimórfico) podemos obtener más información de D , concretamente los cambios debidos a la selección. El número de mutantes deletéreos se mantiene en la población a bajas frecuencias (por acción de la selección purificadora). Debido a que S ignora la frecuencia de mutantes, este parámetro está fuertemente afectado por la existencia de mutaciones deletéreas. No ocurre lo mismo para el parámetro Π que sí considera la frecuencia de mutantes (ANEXO I). Por tanto, si la población presenta mutaciones deletéreas, la estima de θ basada en S , será mayor que la basada en Π , y la diferencia entre ambos parámetros será significativamente menor que 0. Fenómenos como el arrastre selectivo o la selección de fondo, conducen a un exceso

de polimorfismos que segregan a bajas frecuencias, lo que disminuyen el valor de H con respecto al de S , obteniendo también valores de D significativamente menores que 0. No obstante, en el caso de selección estabilizadora los mutantes se mantienen en la población a frecuencias intermedias, aumentando en este caso el valor de H y dando lugar a D significativamente mayores que 0. Para comprender mejor que representa el espectro de frecuencias de polimorfismos nos ayudaremos de la figura III.7. En la gráfica se representa en el eje de ordenadas el número de sitios polimórficos y en el de coordenadas el número de mutaciones por sitio polimórfico. Así podemos ver que en 6 sitios polimórficos solo encontramos una mutación (polimorfismo que segrega a frecuencia única), y solo en un sitio encontramos 2 mutaciones (polimorfismo que segrega a frecuencias intermedias).

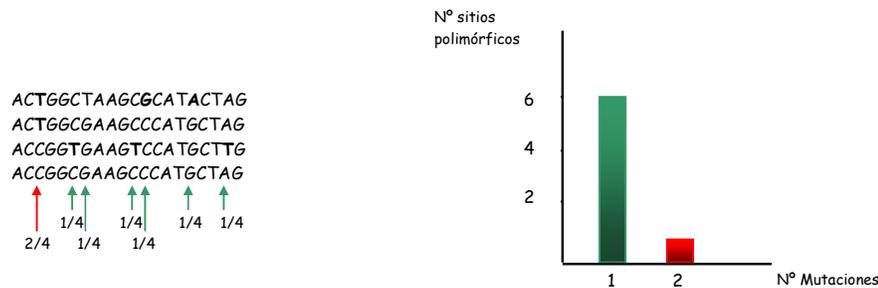


Figura III.7. Distribución de frecuencias de polimorfismos

Por otra parte el cambio en el tamaño de la población también afecta a ambos parámetros. Cuando el tamaño poblacional decrece, H disminuye más rápidamente que S ($D < 0$). Por el contrario cuando el tamaño poblacional aumenta, H aumenta más rápidamente que S ($D > 0$).

Así pues, aunque el modelo de Tajima sea una forma muy general de calcular cualquier desviación del modelo neutral, sí podemos determinar ciertas tendencias en la población. De esta manera, cuando D tiende a ser significativamente menor que 0, o bien indica que la selección purificadora, arrastre selectivo o selección de fondo están actuando, o bien indica que se ha producido un aumento en el tamaño de la población. Por el contrario si D es significativamente mayor que 0, indica que o bien está actuando la selección estabilizadora o que el tamaño de la población ha disminuido.

5.3.2 Prueba de Fu y Li

Una segunda prueba de neutralidad fue desarrollada por Fu y Li (1993), con un contraste muy parecido a la D de Tajima. Estas pruebas hacen estimas de θ a partir del número de mutaciones en las ramas internas y externas de una genealogía utilizando una estructura muy similar a la prueba de Tajima. El número de mutaciones de las ramas externas se estima a partir de los polimorfismos únicos. Así en estas pruebas, el cálculo de θ está basado en el número de mutaciones que aparecen una sola vez entre las secuencias o polimorfismos de frecuencia única, η_s , y en η , número total de mutaciones (estadístico D^*), o entre el número de mutaciones que aparecen una sola vez y Π (estadístico F^*). La significación de ambas pruebas viene dada por:

$$D^* = \frac{\left(\frac{n}{n+1}\right)\eta - a_n \eta_s}{\sqrt{\text{Var}(\eta - \eta_s)}} \qquad F^* = \frac{\Pi - \left(\frac{n}{n-1}\right)\eta_s}{\sqrt{\text{Var}(\Pi - \eta_s)}}$$

Las pruebas asumen que puesto que la selección purgará las mutaciones deletéreas, es más probable que las mutaciones presentes hayan aparecido recientemente y así se encontrarán en las ramas externas de una genealogía. De igual modo, las mutaciones en las ramas internas es más probable que sean antiguas y por lo tanto selectivamente neutras. Al igual que la prueba de Tajima valores iguales a 0 indican neutralidad, valores significativamente menores que 0 indican selección direccional (exceso de polimorfismos de frecuencia única) o aumento del tamaño poblacional, valores significativamente mayores que 0 indican la presencia de selección estabilizadora (exceso de polimorfismos a frecuencias intermedias) o una disminución del tamaño de la población.

5.3.3. Prueba de Fu:

La prueba de Fu evalúa la probabilidad de tener k alelos en una muestra de n secuencias dado θ .

$$F_s = \ln\left(\frac{S'}{1-S'}\right)$$

donde S' es la probabilidad de tener un número de alelos menor o igual que en una muestra al azar, y dado que $\theta = \pi$.

F_s tiende a ser negativo cuando se da arrastre selectivo (proceso mediante el cual una posición se encuentra ligada a otra que ha sido fijada por selección en la población) o crecimiento poblacional.

Según Fu (1997), la prueba de la F_s es muy potente, detectando crecimiento poblacional y arrastre selectivo, mientras que D^* y F^* lo son detectando selección de fondo.

Las pruebas anteriores se han implementado aplicando el programa DNAsp 4.0 (Rozas y Rozas 1999) para cada uno de los individuos por separado.

5.4 Análisis de polimorfismos

También con el programa DNAsp 4.0 hemos llevado a cabo el cálculo de distintos parámetros intrapoblacionales para una posterior comparación entre genotipos. Los parámetros calculados fueron: la diversidad nucleotídica (π), la diversidad haplotípica (Hd), el parámetro θ , número de sitios segregantes (S), número de mutaciones, el contenido en GC, y el número medio de diferencias nucleotídicas emparejadas (Π).

El cálculo de estos parámetros intrapoblacionales lo realizamos para cada individuo por separado y para cada una de las dos regiones empleadas.

5.5. Análisis de la varianza y búsqueda de haplotipos compartidos

La estructura genética de las poblaciones es estudiada por medio del análisis de la varianza molecular (AMOVA), implementado en el programa ARLEQUIN v2.0 (Schneider *et al.* 2000). AMOVA realiza estimas de los componentes de la varianza y los estadísticos F, que reflejan la correlación de la diversidad haplotípica a diferentes niveles de una subdivisión jerárquica (Excoffier *et al.* 1992). Realizamos dos tipos de AMOVA, una a nivel intrapoblacional, utilizando todas las secuencias a la vez pertenecientes a los individuos de un mismo genotipo y estableciendo dos niveles de jerarquía dentro de paciente y entre paciente. La otra fue a nivel interpoblacional, en el que utilizamos a la vez todas las secuencias de todos los individuos de ambos genotipos por separado, estableciendo los siguientes niveles de jerarquía: dentro de paciente, entre pacientes dentro de genotipos y entre genotipos. Estos análisis fueron llevados a cabo tanto para la región E1-E2 como para la NS5A.

La búsqueda de haplotipos compartidos fue llevada a cabo con el mismo programa, mediante el cálculo de distancias emparejadas. Se realizó la búsqueda entre todos los individuos de un mismo genotipo, y entre todos los individuos de ambos genotipos.

5.6. Detección de selección positiva

De todos los métodos existentes capaces de detectar la presencia de selección positiva utilizamos el desarrollado por Yang *et al.* (2000), implementado en el programa CODEML del paquete PAML V3.0 (de *Phylogenetic Analysis by Maximum Likelihood*; Yang 1997). Este método, basado en máxima verosimilitud, permite, mediante una prueba de razón de verosimilitudes, comparar entre distintos modelos de evolución de codones con diferente intensidad de la selección (reflejada en ω o razón entre la tasa de sustitución no sinónima/tasa de sustitución sinónima, dN/dS) la cual puede variar entre las diferentes posiciones aminoacídicas. Los parámetros en cada modelo son calculados por máxima verosimilitud. El programa también permite estimar las posiciones aminoacídicas que se encuentran bajo presión de selección positiva en los modelos pertinentes. A continuación haremos un repaso de los fundamentos de este método.

Modelo Markov de sustitución de codones. La sustitución entre codones se describe según el modelo Markov. Supongamos que existen n codones en la secuencia, y llamemos x_h al codón situado en el sitio h , $\omega^{(h)}$ será la tasa dN/dS en el sitio h (la cual puede variar entre sitios). Así la tasa de sustitución instantánea del codón i al j (i es distinto de j) en el sitio h vendrá dada por:

$$q_{ij}(h) = \begin{cases} 0, & \text{si } i \text{ y } j \text{ difieren en dos o tres posiciones nucleotídicas,} \\ \pi_j, & \text{si } i \text{ y } j \text{ difieren en una transversión sinónima,} \\ k\pi_j, & \text{si } i \text{ y } j \text{ difieren en una transición sinónima,} \\ \omega^{(h)}\pi_j, & \text{si } i \text{ y } j \text{ difieren en una transversión no sinónima,} \\ \omega^{(h)}k\pi_j, & \text{si } i \text{ y } j \text{ difieren en una transición no sinónima} \end{cases}$$

El parámetro κ es la tasa de transición/transversión, y π_j es la frecuencia de equilibrio del codón j , la cual es calculada usando las frecuencias nucleotídicas en las tres posiciones de codones. Dada la matriz $Q = \{q_{ij}\}$, la matriz de probabilidad de transición viene dada por $P(t) = \{p_{ij}(t)\} = e^{Qt}$ (Goldman y Yang 1994), donde $p_{ij}(t)$ es la probabilidad de que el codón i llegue a ser j después de un tiempo t , y donde el tiempo o longitud de las ramas, t , es medido como el número de sustituciones nucleotídicas por codón (sitio).

Cálculo por máxima verosimilitud, bajo un modelo de ω heterogéneo entre sitios.

El uso de un solo ω para cada sitio supondría el empleo de multitud de parámetros. Por tanto se hace uso de una distribución que toma en consideración el uso de ω heterogéneo entre sitios (Nielsen y Yang 1998). Supongamos que asumimos K categorías de sitios con proporciones y ω definidos como: p_0, p_1, \dots, p_{k-1} y $\omega_0, \omega_1, \dots, \omega_{k-1}$ respectivamente. Cómo a priori no sabemos a que categoría pertenece cada sitio, la probabilidad de encontrar el codón x_h en el sitio h con un ω dado, vendrá dada por:

$$P(x_h) = \sum_{k=0}^{k-1} p_k P(x_h | \omega_k)$$

Donde $P(x_h | \omega_k)$ es la probabilidad de observar x_h , dado que el sitio h es de categoría k con una tasa dN/dS ω_k . El logaritmo de la verosimilitud viene dado por la suma sobre todos los sitios n de la secuencia.

$$L = \sum_{h=1}^n \log\{P(x_h)\}$$

Los parámetros t (longitud de las ramas), κ (tasa de transición/transversión) y las proporciones de las categorías de sitios se calculan por máxima verosimilitud.

Tras el cálculo de los parámetros se lleva a cabo mediante inferencia bayesiana, el cálculo de la probabilidad de que un sitio pertenezca a una categoría de sitio dada (Nielsen y Yang 1998). La probabilidad posterior de que el sitio h con el codón x_h , pertenezca a la categoría k es:

$$P(\omega_k | x_h) = \frac{p_k P(x_h | \omega_k)}{P(x_h)} = \frac{p_k P(x_h | \omega_k)}{\sum_j p_j P(x_h | \omega_j)}$$

La categoría k que maximiza la probabilidad posterior, es la categoría más apropiada para el codón. Cuando los valores de ω (ω_k) para alguna categoría son >1 podremos identificar sitios bajo selección positiva. La probabilidad posterior correspondiente a las categorías con $\omega >1$ pueden ser sumadas para dar una probabilidad $P(\omega >1)$ para cada codón. Los codones en los cuales esta probabilidad sea mayor que un valor umbral (50%, 95%, 99%) serán identificados como potencialmente bajo selección positiva.

Modelos de ω variable entre sitios. Todos los modelos implican el cálculo (por máxima verosimilitud) de los siguientes parámetros: longitud de las ramas de la filogenia, tasa de transición/transversión y frecuencias nucleotídicas en las tres posiciones de codones. Hay un total de 13 modelos desarrollados (Yang *et al.* 2000), pero tan sólo vamos a detallar los utilizados en esta tesis: 0, 1, 2, 3, 7 y 8.

- *Modelo 0*: Asume un único valor $\omega = dN/dS$ para todo el alineamiento nucleotídico (Goldman y Yang 1994).

- *Modelo 1 (neutro)*: Este modelo considera dos categorías de sitios. La primera categoría incluye a los sitios conservados, donde las mutaciones no sinónimas son deletéreas y eliminadas por selección, y sólo las sustituciones sinónimas son posibles, por tanto se fija el valor $\omega_0 = 0$. La segunda categoría incluye sitios neutros, donde las

mutaciones no sinónimas son neutras y donde el valor de ω está fijado en 1 ($\omega_1 = 1$). Este modelo asume una proporción p_0 de sitios conservados con $\omega_0=0$, y una proporción $p_1=1-p_0$ de sitios neutros con $\omega_1=1$. Este modelo no permite detectar codones bajo selección positiva.

- *Modelo 2 (selección)*: Añade una categoría de sitios más, con respecto al modelo anterior. A parte de considerar y fijar dos categorías de codones con valores $\omega_0=0$ y $\omega_1=1$, permite la asignación de una tercera categoría donde se estima el valor de ω en lugar de fijarlo, y éste puede tomar valores mayores que uno ($\omega_2>1$), por lo que este modelo sí detecta codones bajo selección positiva. La proporción de codones pertenecientes a esta categoría sería $p_2=1-p_0-p_1$. Las mutaciones no sinónimas en estos codones tienen mayores probabilidades de fijación que las sinónimas.

- *Modelo 3 (discreto)*: Este modelo permite establecer tantas categorías de codones como consideremos. Para cada una de las categorías el programa estimará un valor de ω e incluirá en dichas categorías aquellos codones con valores de ω próximos al estimado para dicha categoría. Al igual que en los casos anteriores, para cada categoría de codones se estimarán las proporciones de los codones pertenecientes a la misma. Este modelo también detecta codones bajo selección positiva.

- *Modelo 7 (beta)*: Este modelo usa una distribución β de las sustituciones en los codones del alineamiento, que puede tomar diferentes formas en el intervalo (0,1). Este modelo no permite la estima de valores de ω por encima de 1 ($0<\omega<1$), por lo que sería equivalente a un modelo discreto neutral pero asumiendo una distribución β .

- *Modelo 8 (beta y ω)*: Asume también una distribución β de las sustituciones en los codones pero añade una categoría de sitios al modelo anterior, permitiendo que los valores de ω puedan ser mayores que 1. Considera una proporción de sitios p_0 con $0<\omega<1$, y una proporción $p_1=1-p_0$ con $\omega>1$.

Prueba de razón de verosimilitudes para la comparación de modelos. Para ver qué modelo refleja mejor la situación presente en nuestras secuencias, primero se contrasta si existen sitios donde $\omega>1$, comparándolo con un modelo que no permita la presencia de estos sitios, mediante una prueba de razón de verosimilitudes cuando los dos modelos están anidados. Este es el caso del Modelo 0 (M0) y el M1, el M1 y M2, M0 y M3, M1 y M3, M2 y M3, y M7 y M8. Si están anidados, $2\Delta L$ (dos veces la diferencia del logaritmo de las verosimilitudes de ambos modelos) será comparada con una χ^2 con un número de grados de libertad igual a la diferencia en el número de parámetros en

ambos modelos. Esto constituye una prueba de razón de verosimilitudes de neutralidad contra un modelo alternativo de selección positiva. En el caso de que los modelos no estén anidados utilizamos el Criterio de Información de Akaike (AIC) para hacer las comparaciones (Akaike 1974). El valor de AIC se puede calcular como $-2 \times (\log \text{verosimilitud del modelo}) + 2 \times (\text{n}^\circ \text{ de parámetros del modelo})$. El modelo que minimice el AIC, será considerado el más apropiado.

Si son los modelos de selección positiva los que proporcionan una mejor explicación para la evolución de estas regiones, se utiliza un método bayesiano para identificar los codones seleccionados positivamente cuando existan. Los codones que maximicen la probabilidad posterior para una categoría de sitios con $\omega > 1$, son potencialmente dianas de selección positiva.

Todo el desarrollo anterior es implementado por el paquete de programas PAML V3.0. Los seis modelos fueron contrastados en nuestro análisis para todos los individuos de ambos genotipos y para las dos regiones amplificadas (E1-E2 y NS5A). Como entrada de datos se utilizaba un árbol de máxima verosimilitud y el alineamiento de las secuencias en formato PHYLIP. El árbol utilizado es el mismo que obtuvimos con PAUP*4b10 para la inferencia de la historia demográfica (apartado 5.2).

Las tablas que muestran los diferentes cálculos y parámetros, así como las posiciones que presumiblemente se encuentran bajo selección positiva se han adjuntado en los ANEXOS IV y V.

5.6.1. Búsqueda de selección positiva en grandes filogenias. Método aproximado II.

Puesto que los análisis de máxima verosimilitud suponen mucho tiempo computacional y debido a que el número de secuencias por paciente es bastante elevado, utilizamos una aplicación del programa para grandes filogenias (Método aproximado II) desarrollado por Yang (2000). Si utilizásemos el método basado en codones, el mayor requerimiento de tiempo en la ejecución del programa lo encontraríamos en el cálculo de la longitud de las ramas, ya que en este caso la longitud de las mismas viene definida como el número de sustituciones nucleotídicas esperadas por codón y las matrices para llevar a cabo el cálculo son de 61×61 . En el método aproximado II, el problema de

calcular la longitud de las ramas se ha solucionado utilizando un análisis de verosimilitud basado en los nucleótidos (en vez de en codones). De este modo la matriz se reduce a una de 4 x 4. En esta nueva aplicación la longitud de las ramas viene definida como el número de sustituciones nucleotídicas esperado por sitio (no por codón). Así, la longitud de las ramas en el modelo de codones es calculada mediante la suma de las longitudes de las ramas en las tres posiciones de codones. El modelo que calcula las diferentes tasas de sustitución, frecuencia de bases y tasas de transiciones/transversiones en las tres posiciones de codones está implementado en el programa BASEML (BASEML, Mgene=4), también del paquete de programas de PAML.

Así pues, siguiendo este nuevo procedimiento para grandes filogenias propuesto por Yang (2000), los pasos que hemos seguido han sido los siguientes:

1. El árbol de máxima verosimilitud empleado como entrada de datos, primero es utilizado en BASEML para estimar la longitud de las ramas basándose en el modelo de nucleótidos.

2. Después, el árbol con la longitud de las ramas ya calculadas en BASEML, es utilizado como entrada de datos para CODEML especificando que fije la longitud de las ramas, para que no sean calculadas por el método basado en codones.

De este modo el tiempo de ejecución se reduce de manera considerable. Por ejemplo, para un individuo del que disponemos 100 secuencias víricas, el cálculo, utilizando el modelo de codones, de los diferentes parámetros para los 6 modelos de CODEML no había finalizado tras dos semanas. Con el modelo de nucleótidos para el cálculo de la longitud de las ramas, el tiempo de ejecución mayor no sobrepasaba las 18 horas.

IV. Resultados

1. ANÁLISIS DE POLIMORFISMOS DE DNA

En las poblaciones naturales existe una gran cantidad de polimorfismo almacenado en las secuencias de DNA. Este polimorfismo es mantenido por fuerzas tales como la mutación, la recombinación, la selección natural o la deriva genética. Por tanto, sería interesante conocer la cantidad de polimorfismo de DNA existente para así tratar de identificar los mecanismos implicados en mantenerlo. Existen diferentes parámetros para medir la variación genética en las secuencias de DNA, entre otros:

- Número de sitios segregantes (*S*): Número de posiciones con al menos 2 nucleótidos diferentes en la misma posición de la secuencia.
- Número medio de diferencias nucleotídicas emparejadas (*II*). Número medio de diferencias nucleotídicas entre cada par de secuencias de la muestra
- Diversidad nucleotídica (π): Probabilidad de que al coger dos secuencias al azar difieran en una posición determinada.
- Diversidad Haplotípica (*Hd*): Probabilidad de que al coger dos secuencias al azar, sean diferentes
- Número total de mutaciones (*Eta*).

Los parámetros anteriores fueron calculados para cada uno de los individuos por separado y para cada una de las dos regiones mediante el programa DNAsp 4.0 (Rozas y Rozas 1999). Los resultados se presentan en el ANEXO II. En la tabla IV.1 se presenta la comparación de las medias de los diferentes parámetros, entre genotipos en cada una de las regiones.

Tabla IV.1. Comparación de las medias de los diferentes parámetros, entre los dos genotipos, 1a y 1b, y para las dos regiones, E1-E2 y NS5A.

	E1a-1b		N1a-1b	
	Valor T	Valor U	Valor T	Valor U
Nº sitios variables	1,2558		2,7378*	
Nº total de mutaciones	1,2851		2,7441*	
Diversidad haplotípica		213		211,5
Diversidad nucleotídica	1,1237		2,2622*	
Nº dif nucleot. Emparej.	1,1245		2,2660*	

Nº dif nucleot. Emparej: Número medio de diferencias nucleotídicas emparejadas, E1a-1b, comparación de las medias entre los genotipos 1a y 1b para la región E1-E2. N1a-1b: Comparación de las medias entre los genotipos 1a y 1b para la región NS5A, T: Estadístico correspondiente a la prueba de la t, U: Estadístico correspondiente a la prueba de Mann-Whitney, *: $p < 0.05$.

La comparación de las medias se realizó mediante una prueba de la t, cuando los datos seguían una distribución normal (véase pruebas de Kolmogorov-Smirnov en las tablas 5-8 del ANEXO IV). Sólo en el caso de la diversidad haplotípica los datos no seguían una distribución normal por lo que la comparación de las medias se llevó a cabo mediante la prueba de Mann-Whitney.

En la figura IV.1 se muestra la media de los valores hallados en todos los individuos, para cada uno de los parámetros, y para cada genotipo y región. Podemos ver cómo los valores de variabilidad, en general, son mayores para los individuos de genotipo 1b que para los individuos de genotipo 1a para ambas regiones. No obstante, mientras que para la región E1-E2 las diferencias entre sus medias no son significativas, sí lo son para la región NS5A (Tabla IV.1).

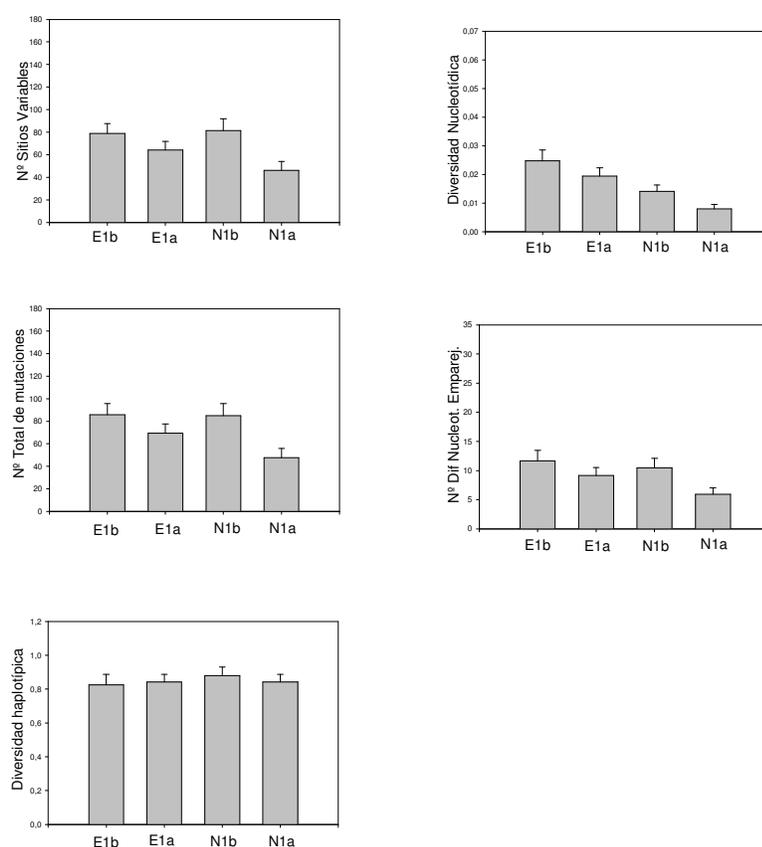


Figura IV.1. Representación gráfica de los valores medios para los diferentes parámetros en ambos genotipos (1a y 1b) y en ambas regiones (E1-E2 y NS5A).

2. ANÁLISIS DE LA VARIANZA Y BÚSQUEDA DE HAPLOTIPOS COMPARTIDOS

La estructura genética de las poblaciones (genotipos 1b y 1a) ha sido estudiada a través del análisis de la varianza molecular (AMOVA). Éste realiza estimas de los componentes de la varianza y los estadísticos F, y lleva a cabo una evaluación de su significación estadística (Excoffier *et al.* 1992).

Se han realizado diferentes AMOVAS atendiendo a diversos aspectos o categorías relacionados con la enfermedad, como diferenciación intra e intergenotípica, edad y sexo de los pacientes, prácticas de riesgo y número de tratamientos a los que ha sido sometido.

Para analizar el primer aspecto (diferenciación intra e intergenotípica), definimos dos tipos de estructura genética para ser contrastadas, mediante la asignación de diferentes grupos. Primero realizamos un estudio a nivel intrapoblacional, tomando como población los pacientes infectados con VHC del mismo genotipo y estableciendo el siguiente nivel de jerarquía: diferencias dentro de pacientes infectados con el mismo genotipo y diferencias entre los pacientes infectados con el mismo genotipo. Este análisis intrapoblacional se realizó para las dos regiones estudiadas (Tablas IV.2, 3, 4 y 5). Después se llevó a cabo un estudio a nivel interpoblacional, estableciendo los siguientes grupos: diferencias dentro de pacientes infectados con el mismo genotipo, diferencias entre los pacientes infectados con el mismo genotipo y diferencias entre genotipos. También se realizó para ambas regiones (Tablas IV.6 y 7).

Tabla IV.2. AMOVA intrapoblacional para la región E1-E2 genotipo 1a.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre pacientes	24	57882.092	23.75919 Va	83.87	0.83875*
Dentro de pacientes	2508	11456.048	4.56780 Vb	16.13	
Total	2532	69338.141	2.832.699		

F: Estadístico F

* p<0.05

Tabla IV.3. AMOVA intrapoblacional para la región E1-E2 genotipo1b

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre pacientes	24	78193.935	32.17625 Va	84.72	0.84717*
Dentro de pacientes	2502	14522.892	5.80451 Vb	15.28	
Total	2526	92716.826	37.98077		

F: Estadístico F

* p<0.05

Tabla IV.4. AMOVA intrapoblacional para la región NS5A genotipo1a

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre pacientes	24	37890.041	26.02151 Va	90.19	0.90195*
Dentro de pacientes	1498	4237.610	2.82885 Vb	9.81	
Total	1522	42127.651	2.885035		

F: Estadístico F

* p<0.05

Tabla IV.5. AMOVA intrapoblacional para la región NS5A genotipo1b

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre pacientes	24	49218.584	31.74051 Va	86.02	0.86025*
Dentro de pacientes	1591	8203.751	5.15635 Vb	13.98	
Total	1615	57422.335	36.89686		

F: Estadístico F

* p<0.05

Tabla IV.6. AMOVA interpoblacional para la región NS5A 1a frente a 1b

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre genotipos	1	109025.802	68.24209 Va	67.41	0.87794*
Entre pacientes dentro de genotipos	48	87108.625	28.96933 Vb	28.61	0.96022*
Dentro de pacientes	3089	12441.361	4.02763 Vc	3.98	0.67407*
Total	3138	208575.788	101.23906		

F: Estadístico F

* p<0.05

Tabla IV.7. AMOVA interpoblacional para la región E1-E2 1a vs 1b

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre genotipos	1	10601.955	41.15158 Va	55.39	0.84298*
Entre pacientes dentro de genotipos	48	135417.019	27.94285 Vb	37.61	0.92995*
Dentro de pacientes	4989	25966.800	5.20481 Vc	7.01	0.55386*
Total	5038	267885.774	74.29923		

F: Estadístico F

* p<0.05

En las tablas IV.2, IV.3, IV.4 y IV.5 podemos ver que a nivel intrapoblacional (población formada por los individuos de un mismo genotipo), y para ambas regiones, encontramos que un porcentaje muy alto de variación se encuentra entre pacientes infectados con un mismo genotipo (alrededor de un 80-90 %). No obstante, aunque este porcentaje sea alto no hay que destacar un fuerte efecto individuo (10-16 %) mostrando una vez más la gran variabilidad de este virus.

A nivel interpoblacional, en la región E1-E2, encontramos que es entre genotipos donde reside el mayor porcentaje de variación (Tabla IV.7). Éste es menor entre pacientes infectados con un mismo genotipo, y como cabría esperar la menor variabilidad la encontramos dentro de pacientes infectados con el mismo genotipo. Si observamos lo que ocurre en la región NS5A (Tabla IV.6), nos damos cuenta que tenemos la misma situación que en la E1-E2, sin embargo el porcentaje de variación entre genotipos es algo mayor en la NS5A, debido seguramente a que esta región es menos variable que la anterior. Estos resultados ponen de manifiesto una clara estructuración poblacional en los dos genotipos que muestran ser muy diferentes entre ellos. Es decir, la principal fuente de diversidad genética en VHC se encuentra entre los genotipos analizados, seguida de la observada entre los pacientes estudiados.

Para llevar a cabo los AMOVAs correspondientes al resto de categorías (edad, sexo, prácticas de riesgo y número de tratamientos a los que ha sido sometido), dividimos a los individuos en diferentes grupos, dentro de cada una de ellas como se muestra en la tabla IV.8:

Tabla IV.8. Grupos establecidos para realizar los AMOVAs atendiendo a las categorías de edad, sexo, práctica de riesgo y n° de tratamientos.

Nº GRUPO	EDAD	SEXO	PRÁCTICA DE RIESGO	Nº TRATAMIENTOS
1	21-30 años	Hombres	ADVP	1
2	31-40 años	Mujeres	Transf/Donante	2
3	41-50 años		Brote	
4	51-60 años			
5	61-70 años			

* p<0.05

ADVP: Adictos a drogas por vía parenteral.

Transf./Donante: Pacientes que sufrieron una transfusión de sangre, o fueron donantes de sangre.

Nº Tratamientos: Hace referencia al n° de tratamientos (sólo con interferón) a los que han sido sometidos, antes de tomar la muestra para realizar nuestro estudio.

Las tablas IV.9 a IV.12 muestran los resultados según edades.

Tabla IV.9. AMOVA para la categoría de edad correspondiente a la región E1-E2 y genotipo 1a.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	3	6055.872	-0.87390 Va	-3.11	-0.03106
Entre pacientes dentro de grupos	20	49813.324	24.53968 Vb	87.21	0.84585*
Dentro de pacientes	2409	10773.368	4.47213 Vc	15.89	0.84106*
Total	2432	66642.564	28.13792		

F: Estadístico F

* p<0.05

N° de grupos = 4

Grupo1: ea05, ea20, ea26, ea27, ec11, ec14, ec33, eg10

Grupo2: ea09, ec01, ec02, ec06, ec20, ec21, ec22, ec28, ec38, eg19

Grupo3: ec07, ec18, ec24, ec31

Grupo4: ec17, eg14

Tabla IV.10. AMOVA para la categoría de edad correspondiente a la región E1-E2 y genotipo 1b.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	4	11966.760	-0.70557 Va	-1.86	-0.01864
Entre pacientes dentro de grupos	16	52234.535	32.63898 Vb	86.25	0.84670*
Dentro de pacientes	2076	12268.171	5.90952 Vc	15.62	0.84384*
Total	2096	76469.466	37.84293		

F: Estadístico F

* p<0.05

N° de grupos = 5

Grupo1: ec03, eg06

Grupo2: ea14, ec04, ec09, ec10, eg16

Grupo3: ec12, ec15, eg05, eg09

Grupo4: ec05, ec13, ec16, ec19, ec25, eg02, eg07, eg17

Grupo5: ec08, eg26

Tabla IV.11. AMOVA para la categoría de edad correspondiente a la región NS5A y genotipo Ia.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	3	5022.840	-0.51857 Va	-1.79	-0.01791
Entre pacientes dentro de grupos	20	32538.953	26.68413 Vb	92.18	0.90562*
Dentro de pacientes	1471	4090.646	2.78086 Vc	9.61	0.90393*
Total	1494	41652.439	28.94642		

F: Estadístico F

* p<0.05

N° de grupos = 4

Grupo1: na05, na05, na20, na26, na27, nc11, nc14, nc33, ng10,

Grupo2: na09, nc01, nc02, nc06, nc20, nc21, nc22, nc28, nc38, ng19

Grupo3: nc07, nc18, nc24, nc31

Grupo4: nc17, ng14

Tabla IV.12. AMOVA para la categoría de edad correspondiente a la región NS5A y genotipo Ib.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	4	9184.357	0.03515 Va	0.10	0.00095
Entre pacientes dentro de grupos	16	32937.356	31.48337 Vb	85.28	0.85363*
Dentro de pacientes	1382	7460.490	5.39833 Vc	14.62	0.85377*
Total	1402	49582.204	36.91685		

F: Estadístico F

* p<0.05

N° de grupos = 5

Grupo1: nc03, ng06

Grupo2: na14, nc04, nc09, nc10, ng16

Grupo3: nc12, nc15, ng05, ng09

Grupo4: nc05, nc13, nc16, nc19, nc25, ng02, ng07, ng17

Grupo5: nc08, ng26

El mayor porcentaje de variación lo observamos entre pacientes que pertenecen a un mismo grupo de edad, siendo negativo y además no significativo el porcentaje correspondiente a diferencias entre los distintos grupos de edad. Este efecto es similar en ambos genotipos y en ambas regiones. Se refleja otra vez la gran variabilidad del virus, ya que aunque encontremos mayor variabilidad entre individuos el porcentaje de variación debido a las diferencias dentro de un mismo paciente no es despreciable. El hecho de que entre los diferentes grupos de edad encontremos un porcentaje de variación tan bajo reflejaría que las diferencias existentes en los virus no

son lo suficientemente grandes como para contemplar la posibilidad de que la edad de los pacientes influya en el progreso de la enfermedad.

En cuanto a la categoría sexo y número de tratamientos, los resultados son muy similares a los encontrados en la categoría anterior (Tablas IV.13-IV.20). Parece que ni el sexo ni el número de tratamientos muestran efectos significativos sobre la partición de la variación genética.

Tabla IV.13. AMOVA para la categoría de sexo correspondiente a la región E1-E2 y genotipo 1a.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	1	3399.131	1.01760 Va	3.51	0.03513*
Entre pacientes dentro de grupos	22	52470.064	23.48095 Vb	81.05	0.84001*
Dentro de pacientes	2409	10773.368	4.47213 Vc	15.44	0.84563*
Total	2432	66642.564	28.97068		

F: Estadístico F

* p<0.05

Nº de grupos: 2

Grupo1: eg19, ea26, ec21, ec38, ec28, ec17, ec18

Grupo2: eg14, eg10, ea09, ea20, ea05, ea27, ec22, ec24, ec20, ec33, ec31, ec02, ec06, ec07, ec01, ec14, ec11

Tabla IV.14. AMOVA para la categoría de sexo correspondiente a la región E1-E2 y genotipo 1b.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	1	3495.346	0.22540 Va	0.59	0.00589
Entre pacientes dentro de grupos	21	68638.870	32.37180 Vb	84.64	0.85141*
Dentro de pacientes	2295	12965.461	5.64944 Vc	14.77	0.85229*
Total	2317	85099.677	38.24664		

F: Estadístico F

* p<0.05

Nº de grupos: 2

Grupo1: eg02, eg17, eg26, ea06, ea14, ea03, ec12, ec13,

Grupo2: eg06, eg07, eg05, eg16, eg09, ec10, ec09, ec16, ec19, ec25, ec15, ec04, ec05, ec08, ec03

Tabla IV.15. AMOVA para la categoría de sexo correspondiente a la región NS5A y genotipo 1a.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	1	1787.483	-0.08146 Va	-0.28	-0.00280
Entre pacientes dentro de grupos	22	35774.310	26.35479 Vb	90.71	0.90455*
Dentro de pacientes	1471	4090.646	2.78086 Vc	9.57	0.90429*
Total	1.494	41652.439	29.05419		

F: Estadístico F

* p<0.05

Nº de grupos: 2

Grupo1: ng19, na26, nc21, nc38, nc28, nc17, nc18

Grupo2: ng14, ng10, na09, na20, na05, na27, nc22, nc24, nc20, nc33, nc31, nc02, nc06, nc07, nc01, nc14, nc11

Tabla IV.16. AMOVA para la categoría de sexo correspondiente a la región NS5A y genotipo 1b.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	1	2308.547	-0.02923 Va	-0.08	-0.00079
Entre pacientes dentro de grupos	21	44681.689	31.98144 Vb	86.27	0.86202*
Dentro de pacientes	1514	7750.543	5.11925 Vc	13.81	0.86191*
Total	1536	54740.779	37.07147		

F: Estadístico F

* p<0.05

Nº de grupos: 2

Grupo1: ng02, ng17, ng26, na06, na14, na03, nc12, nc13,

Grupo2: ng06, ng07, ng05, ng16, ng09, nc10, nc09, nc16, nc19, nc25, nc15, nc04, nc05, nc08, nc03

Tabla IV.17. AMOVA para la categoría de nº de tratamientos correspondiente a la región E1-E2 y genotipo 1a.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	1	2998.861	0.52261 Va	1.81	0.01812
Entre pacientes dentro de grupos	21	50504.041	23.72491 Vb	82.27	0.83789*
Dentro de pacientes	2304	10575.340	4.58999 Vc	15.92	0.84083*
Total	2326	64078.242	28.83751		

F: Estadístico F

* p<0.05

Nº de grupos: 2

Grupo1, ea05, ea09, ea20, ea26, ea27, ec06, ec11, ec14, ec18, ec38, eg10, eg14, eg19

Grupo2, ec01, ec02, ec07, ec20, ec21, ec22, ec24, ec28, ec31, ec33

Tabla IV.18. AMOVA para la categoría de n° de tratamientos correspondiente a la región E1-E2 y genotipo 1b.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	1	2952.837	-0.25610 Va	-0.68	-0.00676
Entre pacientes dentro de grupos	22	70770.897	32.14445 Vb	84.86	0.84295*
Dentro de pacientes	2373	14211.761	5.98894 Vc	15.81	0.84189*
Total	2396	87935.495	37.87730		

F: Estadístico F

* p<0.05

N° de grupos: 2

Grupo1: ea03, ea06, ea14, ea16, ec04, ec05, ec08, ec13, ec15, ec19, eg02, eg05, eg06, eg07, eg09, eg17, eg26

Grupo2: ec03, ec09, ec10, ec12, ec16, ec25, eg16

Tabla IV.19. AMOVA para la categoría de n° de tratamientos correspondiente a la región NS5A y genotipo 1a.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	1	1699.457	-0.26624 Va	-0.91	-0.00915
Entre pacientes dentro de grupos	21	35313.015	26.55218 Vb	91.22	0.90392*
Dentro de pacientes	1447	484.006	2.82240 Vc	9.70	0.90304*
Total	2396	41096.478	29.10834		

F: Estadístico F

* p<0.05

N° de grupos: 2

Grupo1, na05, na09, na20, na26, na27, nc06, nc11, nc14, nc18, nc38, ng10, ng14, ng19

Grupo2, nc01, nc02, nc07, nc20, nc21, nc22, nc24, nc28, nc31, nc33

Tabla IV.20. AMOVA para la categoría de n° de tratamientos correspondiente a la región NS5A y genotipo 1b.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	1	2313.764	0.08835 Va	0.24	0.00238
Entre pacientes dentro de grupos	22	46267.058	31.86425 Vb	85.76	0.85961*
Dentro de pacientes	1565	8144.158	5.20394 Vc	14.01	85995*
Total	1588	56724.980	37.15653		

F: Estadístico F

* p<0.05

N° de grupos: 2

Grupo1: na03, na06, na14, na16, nc04, nc05, nc08, nc13, nc15, nc19, ng02, ng05, ng06, ng07, ng09, ng17, ng26

Grupo2: nc03, nc09, nc10, nc12, nc16, nc25, ng16

Sin embargo, atendiendo a la categoría de prácticas de riesgo (Tablas IV.21-IV.24) aunque el porcentaje de variación encontrado es también mayor entre pacientes pertenecientes a un mismo grupo, el porcentaje encontrado entre prácticas de riesgo difiere en ambos genotipos. Mientras en los individuos de genotipo 1a el porcentaje es positivo, en los 1b este valor es mucho menor, llegando a ser negativo para la región E1-E2. Estos resultados podrían indicar una vinculación entre prácticas de riesgo y genotipo en el transcurso de la enfermedad.

Tabla IV.21. AMOVA para la categoría de prácticas de riesgo correspondiente a la región E1-E2 y genotipo 1a.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	2	8233.580	5.86352 Va	18.01	0.18006*
Entre pacientes dentro de grupos	12	27222.022	22.13840 Vb	67.98	0.82912*
Dentro de pacientes	1517	6921.557	4.56266 Vc	14.01	0.85989*
Total	1531	42377.159	32.56458		

F: Estadístico F

* p<0.05

Nº de grupos: 3

Grupo1: ea05, ea20, ea26, ec02, ec07, ec11, ec21, ec22, ec28, eg10, eg19

Grupo2: ec01

Grupo3: ec14, ec17, ec38

Tabla IV.22. AMOVA para la categoría de prácticas de riesgo correspondiente a la región E1-E2 y genotipo 1b.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	1	3083.642	-1.47405 Va	-3.91	-0.03915
Entre pacientes dentro de grupos	4	14063.076	35.61413 Vb	94.58	0.91020*
Dentro de pacientes	586	2059.033	3.51371 Vc	9.33	0.90668*
Total	591	19205.752	37.65378		

F: Estadístico F

* p<0.05

Nº de grupos: 2

Grupo1: ec03, ec16, ec19

Grupo2: ec09, ec13, ec25

Tabla IV.23. AMOVA para la categoría de prácticas de riesgo correspondiente a la región NSSA y genotipo 1a.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	2	4602.554	3.27613 Va	9.88	0.09883
Entre pacientes dentro de grupos	12	17665.047	26.40994 Vb	79.67	0.88405*
Dentro de pacientes	852	2951.169	3.46381 Vc	10.45	0.89551*
Total	866	25218.770	33.14988		

F: Estadístico F

* p<0.05

Nº de grupos: 3

Grupo1: na05, na20, na26, nc02, nc07, nc11, nc21, nc22, nc28, ng10, ng19

Grupo2: nc01

Grupo3: nc14, nc17, nc38

Tabla IV.24. AMOVA para la categoría de prácticas de riesgo correspondiente a la región NSSA y genotipo 1b.

Fuente de variación	Grados Libertad	Suma de cuadrados	Componentes de la varianza	Porcentaje de variación	F
Entre grupos	1	2549.366	0.39720 Va	0.91	0.00910
Entre pacientes dentro de grupos	4	9714.578	37.74284 Vb	86.43	0.87219*
Dentro de pacientes	385	2129.440	5.53101 Vc	12.67	0.87335*
Total	390	14393.384	43.67105		

F: Estadístico F

* p<0.05

Nº de grupos: 2

Grupo1: nc03, nc16, nc19

Grupo2: nc09, nc13, nc25

Una de las razones por las cuales el VHC pueda burlar las defensas del sistema inmune del hospedador y conseguir su permanencia en el mismo, podría residir en su alta variabilidad genética. Este hecho junto con que el genotipo 1b es más prevalente que el 1a en la Comunidad Valenciana, podría hacer pensar en la posibilidad de que los individuos 1b pudieran albergar alguna variante que les ofreciese alguna ventaja frente al ataque del sistema inmune. Esta variante la encontraríamos en todos o en la mayoría de los individuos 1b y no aparecería en los 1a. Procedimos a averiguar si esta variante podría localizarse en los genes que estamos analizando. Para ello, también con el programa ARLEQUIN v2.0 (Schneider *et al.* 2000), se procedió a la búsqueda de haplotipos compartidos entre pacientes de un mismo genotipo o entre pacientes de

diferentes genotipos. Sin embargo no se encontró ningún haplotipo común ni entre individuos de un mismo genotipo ni entre individuos de diferentes genotipos. Cada uno poseía haplotipos únicos que no se encontraban en otros individuos ya pertenecieran al mismo genotipo o no.

3. HISTORIA DEMOGRÁFICA

La filogenética y la genética de poblaciones proporcionan herramientas analíticas que permiten investigar sobre diferentes aspectos de la evolución de los virus, entre ellos la historia demográfica de las poblaciones. Éste último nos ofrece información acerca de si la población ha permanecido con un tamaño constante o ha experimentado cambios con el tiempo, y si es así, que patrón de crecimiento ha seguido.

Para inferir la historia demográfica a partir de una serie de secuencias hemos aplicado un método que realiza esta inferencia a partir de una genealogía (Pybus y Rambaut 2002) asumiendo “el modelo de tamaño poblacional variable” de la teoría de la coalescencia (Griffiths y Tavaré 1994). Esta teoría recoge un grupo de modelos que describen las propiedades estadísticas de las filogenias intrapoblacionales (Kingman 1982), y que permite estimar diversos parámetros genético-poblacionales. No vamos a desarrollar aquí esta teoría, pero sí como se infiere el tamaño poblacional efectivo (N_e) a partir de ella.

Una reconstrucción filogenética nos ofrece estimas acerca de g_i y t_i , donde g_i corresponde al tamaño del intervalo de un internodo (hace referencia al nº de linajes presentes en una genealogía durante cada intervalo), y t_i el tiempo al cual comienza un intervalo. A partir de la teoría de Griffiths y Tavaré (1994) sabemos que:

$$g_i^{(i)} = -\ln(U)Hi \quad \text{donde} \quad Hi = \left(\int_{x=t_i}^{g_i+t_i} \frac{1}{N_e(x)} dx / g_i \right)$$

Hi es la media armónica del tamaño poblacional efectivo en el rango $[t_i, g_i + t_i]$, siendo $[t_i, g_i + t_i]$ el período de tiempo delimitado por el intervalo de un internodo. $-\ln(U)$ representa el error cometido por azar. Debido a que las genealogías construídas están hechas bajo la suposición de reloj molecular, el intervalo de un internodo debe ser medido en unidades de sustitución esperadas por sitio y no en generaciones. Así

realizamos un cambio de variable: $\gamma_i = \mu g_i$ donde γ_i es el tamaño del intervalo en sustituciones por sitio y μ la tasa de mutación en sustituciones por sitio. De tal modo que la fórmula anterior queda reducida a:

$$\gamma_2^{(i)} = -\ln(U)H_i\mu$$

Debido a que $\ln(U)$ representa error por azar, el término $M_i = \gamma_2^{(i)}$ es una estimación de $H_i\mu$, que puede ser calculada a partir de una genealogía. Consecuentemente la representación de M_i frente al tiempo define una función escalonada, que constituye una estimación no paramétrica de la historia demográfica. M_i representa toda la información sobre $N_e(x)$ que podemos inferir a partir de un intervalo de un internodo. Es decir lo máximo que podemos inferir de un internodo es que define un intervalo de tiempo $[t_b, g_i + t_i]$ durante el cual la media armónica de $N_e(x)\mu$ se estima que es M_i .

A parte de la estimación no paramétrica de N_e también se estima paramétricamente la historia demográfica. Para ello se obtienen estimaciones de varios parámetros demográficos, por máxima verosimilitud, bajo una serie de modelos de historia demográfica, que describen la variación del tamaño poblacional efectivo a través del tiempo. Estos modelos son: 1) tamaño poblacional constante, 2) crecimiento exponencial, 3) crecimiento de expansión, 4) crecimiento logístico, 5) crecimiento con fase constante-exponencial-constante, 6) crecimiento de expansión escalonado y 7) crecimiento logístico escalonado. Finalmente, los modelos demográficos son contrastados mediante una prueba de razón de verosimilitudes, si los modelos están anidados, o mediante el criterio de información de Akaike corregido (AICc) si no lo están. En este caso el modelo que maximice AICc, será el que mejor explique la historia demográfica.

Los modelos de historia demográfica, explicados en material y métodos, e implementados en GENIE V3.0 (Pybus y Rambaut 2002) han sido aplicados a las secuencias virales de ambos genotipos.

El estudio de la historia demográfica del virus de la hepatitis C se ha llevado a cabo a dos escalas. Un primer análisis se realizó utilizando a los pacientes de forma individual, es decir, empleando todas las secuencias disponibles para cada uno de ellos por separado. La segunda escala ha consistido en la consideración de la historia del genotipo, utilizando sólo secuencias consenso y las secuencias más divergentes dentro de cada paciente, como detallamos más adelante.

En la tabla IV.25 se pueden observar los resultados obtenidos tras calcular la presencia de señal filogenética aplicando TREE-PUZZLE V.5. Aquellos individuos que presentaban árboles con menos de un 60% de señal filogenética fueron descartados para el análisis a nivel de paciente (para evitar resultados erróneos con el programa GENIE como ya se especificó en el apartado 5.2 de material y métodos), quedándonos solamente con algunos. Concretamente 14 individuos con genotipo 1a y 18 con genotipo 1b (marcados en rojo en la tabla IV.25). En la mayoría de casos coincidía que la señal filogenética era mayor o menor del 60% en ambas regiones. Es decir que cuando el árbol para la región E1-E2 era descartado, también lo era para la NS5A.

Tabla IV.25. Estudio de la señal filogenética.

Paciente 1b	% Inform. Región E1-E2	% Inform. Región NS5A	Paciente 1a	% Inform. Región E1-E2	% Inform. Región NS5A
A03	88.3	82.9	A05	7.2	27.9
A06	5.7	14.6	A09	64.5	64.9
A14	85.2	56	A20	79.3	63.6
A16	92.1	72.7	A26	87.9	81.3
A35	77.1	61.5	A27	75.5	84.2
C03	47.6	24.8	A34	88.9	87.6
C04	0	76	C01	71.4	75.5
C05	94.8	57.5	C02	14.6	29
C08	94.6	87.2	C06	0.5	10.7
C09	50.6	56.6	C07	89.2	82.8
C10	83.6	49.7	C11	12.2	9.8
C12	82.3	87.3	C14	45.5	63.4
C13	56.9	85.1	C17	67.1	2.3
C15	91.3	58	C18	72	0
C16	68.3	66	C20	91.5	67.7
C19	90.5	85.7	C21	90.1	48
C25	2.5	21.42	C22	94.7	86.2
G02	95.3	90.7	C24	62.8	58.2
G05	74.8	71.8	C28	89.5	71.4
G06	88.5	81.5	C31	26.5	5.7
G07	67.9	85	C33	61	0
G09	79.6	87.4	C38	18.5	0
G16	0	0	G10	0	14.9
G17	89.9	66.2	G14	58.2	61.4
G26	0	0	G19	84.3	69.6

En rojo se señalan aquellos individuos que presentan más de un 60 % de señal filogenética en ambas regiones.

% Inform: Porcentaje de señal filogenética.

3.1. Historia demográfica de las poblaciones virales de los pacientes

El estudio a esta escala, infería la historia demográfica de las poblaciones virales de cada uno de los pacientes, utilizando todas las secuencias virales disponibles para cada uno de ellos. En las tablas IV.26 y IV.27 se muestra el modelo de crecimiento que mejor ajusta para cada paciente y en el ANEXO V la estima de los parámetros para cada modelo demográfico (Tablas 9-12 del Anexo V), así como las gráficas *skylineplots* correspondientes a las estimas no paramétricas.

Si nos fijamos en los resultados de las tablas IV.26 y IV.27, en general se observa que los modelos de crecimiento dentro de cada paciente, no siguen un patrón característico que permita establecer diferencias entre genotipos. Más bien observamos que dentro de cada uno, el virus se comporta de manera diferente independientemente del genotipo al que pertenezca. Tampoco se observa una relación entre la historia demográfica de cada población con la diversidad nucleotídica y haplotípica, o con las prácticas de riesgo al que pertenece. Para un determinado modelo demográfico podemos encontrar diversidades nucleotídicas y haplotípicas variadas, y diversas prácticas de riesgo.

Por otra parte, si atendemos a lo que sucede en las dos regiones analizadas, observamos que en la mayoría de casos no coinciden las dinámicas obtenidas en la E1-E2 y en la NS5A. Así por ejemplo, en el caso de eA09, región E1-E2, estaríamos ante un modelo de expansión escalonado, mientras que según la NS5A estaríamos ante un modelo de crecimiento exponencial. Hasta ahora este tipo de análisis nunca se había llevado a un nivel tan específico como el individuo, si no que en trabajos similares como máximo se llega a genotipo. El no encontrar el mismo modelo en ambas regiones no tiene porqué interpretarse como una falta de peso en los resultados. Ambas regiones pertenecen a genes que codifican para proteínas con funciones y propiedades muy distintas. De hecho la región E1-E2 posee una tasa de sustitución nucleotídica mayor que la NS5A (Pybus *et al.* 2001). Estas diferencias, puesto que el análisis está hecho a un nivel tan específico, podrían ser responsables de que las dinámicas poblacionales no coincidan. No obstante hay que mencionar que aunque las dinámicas poblacionales sean diferentes, no apreciamos situaciones contradictorias, como una región bajo un modelo de crecimiento constante y en la otra un modelo de expansión. Dado que la tasa de sustitución es menor en la región NS5A, ambas regiones deberían encontrarse en momentos diferentes de la historia demográfica. En efecto, parece que la región NS5A

esté en una fase por la que ya ha pasado la región E1-E2. En esta cuestión profundizaremos más en el apartado de discusión.

Tabla IV.26. Modelos de crecimiento para cada paciente de genotipo 1a. La diversidad nucleotídica y haplotípica fue calculada con DNAsp V4.0. La E y la N situadas delante del nombre de cada muestra diferencian entre la región E1-E2 (e) y la región de la NSSA (n).

Individuo	Const	Expo	Expan	Logist	Con-Exp		Logist	Divers. Haplotípica	Divers. Nucleot.	Práctica de Riesgo
					-Con	esc				
eA09							•	0,976	0,012	¿?
eA20							•	0,988	0,024	ADVP
eA26		•						0,994	0,034	ADVP
eA27							•	0,970	0,021	¿?
eA34							•	0,997	0,029	¿?
eC01							•	0,955	0,002	Donante
eC07			•					0,995	0,025	ADVP
eC20							•	0,965	0,036	¿?
eC21							•	0,971	0,030	ADVP
eC22							•	0,993	0,043	ADVP
eC24							•	0,945	0,007	¿?
eC28		•						0,994	0,044	ADVP
eG14							•	0,974	0,026	¿?
eG19							•	0,975	0,028	ADVP
nA09		•						0,990	0,012	¿?
nA20				•				0,928	0,010	ADVP
nA26		•						0,99	0,019	ADVP
nA27							•	0,979	0,013	¿?
nA34		•						1,000	0,013	¿?
nC01							•	0,977	0,028	Donante
nC07							•	0,973	0,017	ADVP
nC20							•	0,863	0,005	¿?
nC21							•	0,983	0,006	ADVP
nC22		•						0,998	0,019	ADVP
nC24							•	0,94	0,003	¿?
nC28		•						0,991	0,015	ADVP
nG14							•	0,986	0,011	¿?
nG19		•						0,995	0,011	ADVP

Const: Tamaño poblacional constante, Expo: crecimiento exponencial, Con-exp-con: crecimiento con fase constante-exponencial-constante, Expan: crecimiento en expansión, Logist: crecimiento logístico, Expan esc: crecimiento en expansión escalonado, Logist esc: crecimiento logístico escalonado, Divers. Haplotípica: Diversidad. Haplotípica, Divers. Nucleot.: Diversidad nucleotídica.

Tabla IV.27. Modelos de crecimiento para cada paciente de genotipo 1b. La diversidad nucleotídica y haplotípica fue calculada con DNAsp V4.0. La E y la N situadas delante del nombre de cada muestra diferencian entre la región E1-E2 (e) y la región de la NS5A (n).

Individuo	Const	Expo	Expan	Logist	Con-Exp		Logist	Divers. Haplotípica	Divers. Nucleot.	Práctica de Riesgo
					-Con	esc				
eA03				●				0,993	0,026	¿?
eA14							●	0,997	0,033	¿?
eA16		●						0,998	0,054	¿?
eA35							●	0,996	0,012	¿?
eC05			●					0,998	0,047	¿?
eC08							●	0,996	0,046	¿?
eC10							●	0,990	0,027	¿?
eC12							●	0,997	0,025	¿?
eC13							●	0,992	0,012	Transfusión
eC15							●	0,998	0,058	¿?
eC16							●	0,969	0,251	ADVP
eC19		●						0,999	0,040	ADVP
eG02							●	0,986	0,045	¿?
eG05							●	0,952	0,012	¿?
eG06			●					0,999	0,029	¿?
eG07							●	0,996	0,040	¿?
eG09							●	0,998	0,023	¿?
eG17							●	0,997	0,051	¿?
nA03							●	0,984	0,013	¿?
nA14							●	0,990	0,008	¿?
nA16							●	0,998	0,021	¿?
nA35							●	0,963	0,006	¿?
nC05							●	0,995	0,010	¿?
nC08							●	0,980	0,019	¿?
nC10							●	0,947	0,008	¿?
nC12							●	0,993	0,015	¿?
nC13			●					0,997	0,042	Transfusión
nC15				●				0,846	0,012	¿?
nC16		●						0,998	0,012	ADVP
nC19							●	0,995	0,023	ADVP
nG02							●	0,994	0,022	¿?
nG05		●						0,991	0,025	¿?
nG06			●					0,999	0,026	¿?
nG07							●	0,996	0,036	¿?
nG09							●	0,984	0,023	¿?
nG17							●	0,993	0,015	¿?

Const: tamaño poblacional constante, Expo: crecimiento exponencial, Con-exp-con: crecimiento con fase constante-exponencial-constante, Expan: crecimiento en expansión, Logist: crecimiento logístico, Expan esc: crecimiento en expansión escalonado, Logist esc: crecimiento logístico escalonado, Divers. Haplotípica: Diversidad. Haplotípica, Divers. Nucleot.: Diversidad nucleotídica.

3.2. Historia demográfica de los genotipos

Para este análisis hemos utilizado todas las secuencias virales de cada uno de los individuos, incluidos los descartados en el estudio a escala intrapaciente, porque sus respectivos árboles no tenían suficiente señal filogenética. Los datos utilizados fueron separados en dos conjuntos (Ver figura III.5 del apartado de material y métodos). En el primero obtuvimos la secuencia consenso de todos los clones pertenecientes a cada uno de los individuos (en el apartado 5.2 de material y métodos se explica como se obtuvo). Así pues en este primer conjunto tendríamos 4 grupos formados por: 25 secuencias consenso correspondientes a la región E1-E2 (una secuencia consenso por cada paciente), 25 pertenecientes a la región NS5A y tanto para el genotipo 1b y como para el 1a.

Para cada genotipo y región por separado, se alinearon las secuencias consenso y tras realizar el árbol de máxima verosimilitud se ejecutaron los análisis correspondientes con el programa GENIE.

El mismo procedimiento fue seguido para el segundo conjunto de datos. Éste estaba formado por las dos secuencias más divergentes dentro de cada individuo (ver apartado 5.2 de material y métodos). Por tanto volvíamos a tener 4 grupos, pero ahora formados por 50 secuencias (dos secuencias por individuo) para la región E1-E2 y 50 para la NS5A, y tanto para el genotipo 1b como para el genotipo 1a.

De nuevo se alinearon las secuencias y se aplicó GENIE tras obtener el árbol de máxima verosimilitud.

Los resultados obtenidos muestran diferencias entre genotipos (Figura IV.2). Mientras los individuos de genotipo 1a parecen seguir dinámicas de expansión y crecimiento exponencial, los de genotipo 1b exhiben dinámicas que revelan un visible descenso en su tasa de crecimiento. No obstante, para el grupo ConsE1b (formado por las secuencias consenso correspondientes a la región E1-E2 de los individuos de genotipo 1b) la estima que obtenemos tanto por el método no paramétrico como por el paramétrico, refleja crecimiento exponencial y no una disminución de la tasa de crecimiento como en el resto de grupos formados por las secuencias de genotipo 1b. Además este resultado tampoco coincide con lo que obtenemos en el grupo también formado por las secuencias consenso de genotipo 1b, pero pertenecientes a la región NS5A. Sin embargo si nos fijamos en el árbol filogenético del grupo ConsE1b (Ver ANEXO VI) se observan ramas muy largas que podrían reflejar una estabilización en su

crecimiento. Así pues, el observar un crecimiento exponencial en este caso podría ser debido, o bien a que la dinámica de este grupo no se ajuste a ninguno de los modelos que sugiere el GENIE, o bien a que tal vez podría seguir una combinación de alguno de estos modelos.

La transformación del eje de tiempo, de nº de sustituciones a años, nos ayuda a interpretar lo que ha sucedido y lo que podría ocurrir en un futuro con respecto a la historia demográfica de ambos genotipos. GENIE permite hacer esta conversión si proporcionamos la tasa de sustitución por sitio y por año. Las tasas que utilizamos fueron 7.9×10^{-4} para la región E1-E2 y de 5.0×10^{-4} para la región correspondiente a la NS5A (Pybus *et al.* 2001). Podemos observar (Figura IV.2) que el genotipo 1b aparece antes en el tiempo, hace unos 200 años según este análisis, e inmediatamente experimenta un fuerte crecimiento hasta hace unos 50-60 años, cuando empieza a estabilizarse. El genotipo 1a aparece algo más tarde, sin embargo no se observa un crecimiento importante del mismo hasta hace unos 50-60 años aproximadamente. Si además unimos ambas gráficas podemos observar más claramente la coincidencia en el tiempo de los cambios en el crecimiento en ambos genotipos (Figura IV.3).

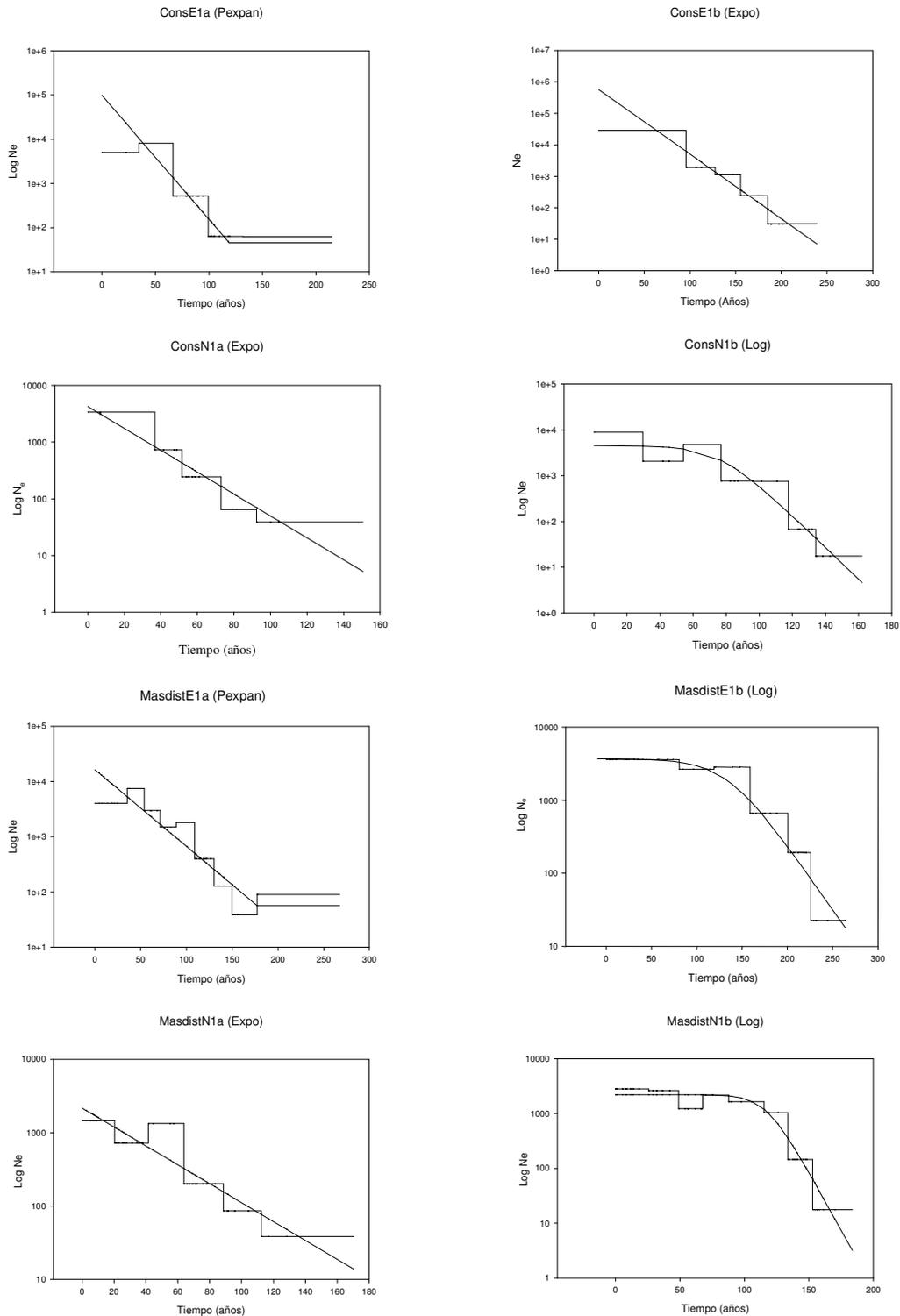


Figura IV.2. Skylineplots pertenecientes a los datos en el nivel de genotipo. La curva que representa los datos observados es la escalonada, mientras que la lineal corresponde a la esperada bajo el modelo correspondiente. Cons: Grupo de datos correspondiente a las secuencias consenso, Masdist: Grupo de datos formado por las dos secuencias más divergentes dentro de cada paciente. La E y la N señalan la región E1-E2 y NS5A respectivamente. Entre paréntesis se indica el modelo de crecimiento de cada grupo. Pexpan: Expansión escalonada, Expo: Exponencial, Log: Logístico. N_e : tamaño poblacional efetivo.

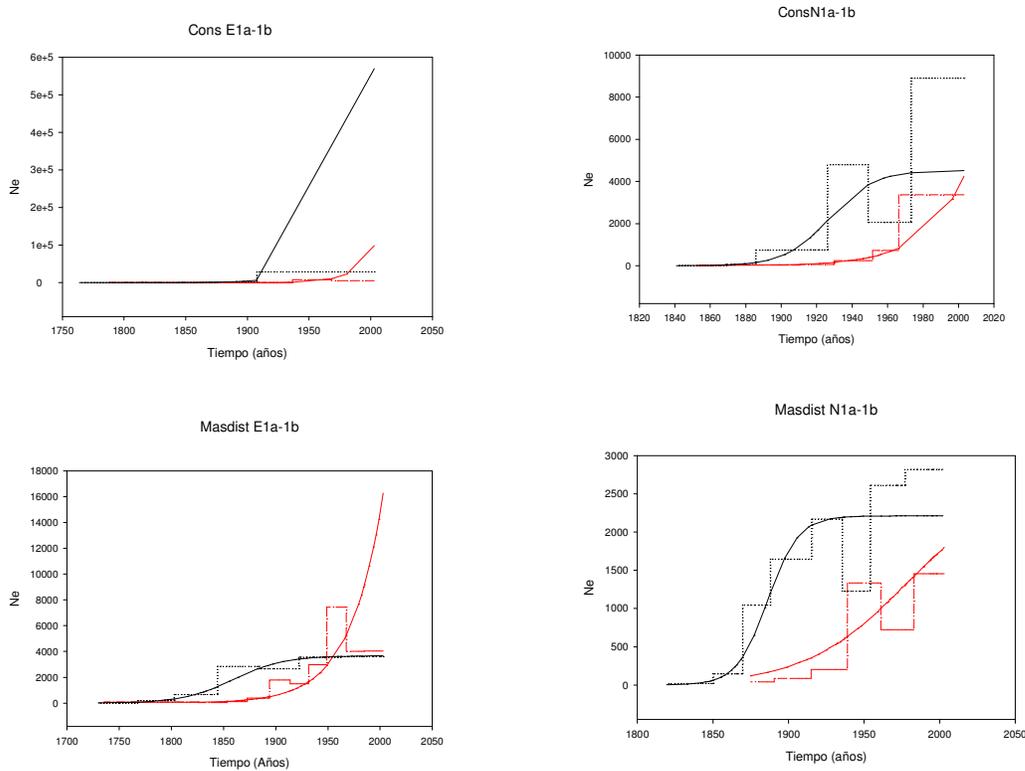


Figura IV.3. Representación conjunta de las curvas obtenidas para ambos genotipos, separadas por grupos de datos y regiones. Las curvas de color rojo representan al genotipo 1a, y las negras al 1b. Las líneas continuas corresponden a las curvas esperadas, mientras que las punteadas a las observadas. Cons E1a-1b: Curvas para las secuencias consenso de los individuos 1a y 1b y para la región E1-E2; Cons N1a-1b : Curvas para las secuencias consenso de los individuos 1a y 1b y para la región NS5A; Masdist E1a-1b: curvas para el conjunto de datos de las dos secuencias más divergentes dentro de cada paciente, para la región E1-E2; Masdist N1a-1b: curvas para el conjunto de datos de las dos secuencias más divergentes dentro de cada paciente, para la región NS5A.

3.3. Número Reproductivo Básico: R_0

Otro parámetro que puede ser estimado a partir de las secuencias génicas, y que ofrece también información acerca de las dinámicas de las poblaciones virales es R_0 o número reproductivo básico de un patógeno. R_0 representa el número medio de células infectadas que derivan de cualquier otra célula infectada al comienzo de la infección (Nowak *et al.* 2000). Pero quizás antes de empezar a hablar de este factor, sea conveniente ver de donde procede y los factores implicados en su cálculo. El modelo básico de dinámica viral tiene tres variables: el tamaño poblacional de células no

infectadas (x), de células infectadas (y) y de partículas virales libres (v). La figura IV.4 ilustra este modelo de dinámica viral.

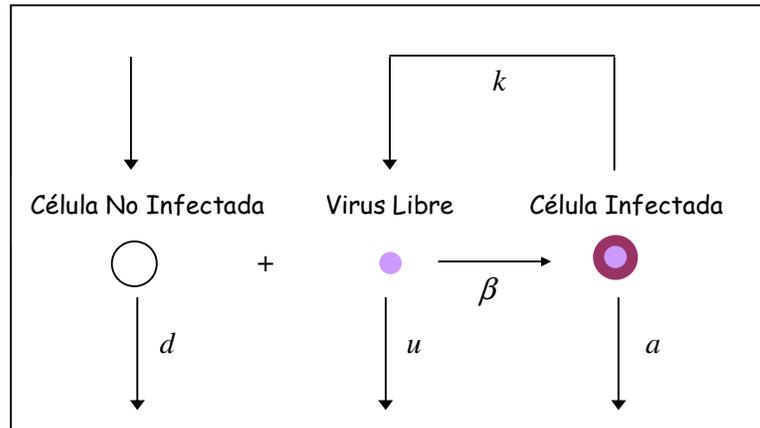


Figura IV.4: Esquema del modelo básico de dinámica viral. Las células no infectadas “reaccionan” con las partículas virales libres para dar lugar a células infectadas a una tasa constante β . Las células infectadas producen viriones a una tasa k . Las células no infectadas, virus libres y células infectadas mueren a unas tasas d , u , y a respectivamente. Las células son repuestas a una tasa constante λ . (Nowak *et al.* 2000)

Las partículas virales libres infectarán células no infectadas a una tasa proporcional al producto de sus abundancias, βxv . La tasa constante β , describe la eficacia de este proceso, incluyendo la tasa a la cual el virus encuentra células no infectadas, la tasa a la cual entra en la célula y la probabilidad de que la infección sea satisfactoria. Las células infectadas producen virus a una tasa proporcional a su abundancia, ky . Las células infectadas mueren a una tasa ay , y las partículas virales son extraídas del sistema a una tasa uv . Así la vida media de una célula infectada es $1/a$, mientras que la vida media de un virus es $1/u$. La cantidad total de partículas virales producidas a partir de una célula infectada es k/a . Además de describir la dinámica del proceso de infección también hay que tener en cuenta la dinámica de la población celular no infectada, ya que al fin y al cabo la infección depende en parte del número de células disponibles. Las células no infectadas se producen a una tasa constante λ , y mueren a una tasa dx . La vida media de una célula no infectada es por tanto $1/d$. En ausencia de infección, la dinámica poblacional de las células huésped viene dada por

$x = \lambda - dx$. Sin virus, la abundancia de células no infectadas converge al valor de equilibrio λ/d .

Combinando las dinámicas de la infección viral y la de las células no infectadas, obtenemos el modelo básico de dinámica viral.

$$\dot{x} = \lambda - dx - \beta xv$$

$$\dot{y} = \beta xv - ay$$

$$\dot{v} = ky - uv$$

Este es un sistema no lineal de ecuaciones diferenciales. El sistema no tiene una solución analítica del tiempo de desarrollo de las variables, pero a partir de él pueden derivarse una serie de aproximaciones que contribuyen al entendimiento del mismo.

Si un virus puede o no crecer y establecer una infección depende de R_0 o número reproductivo básico (Figura IV.5).

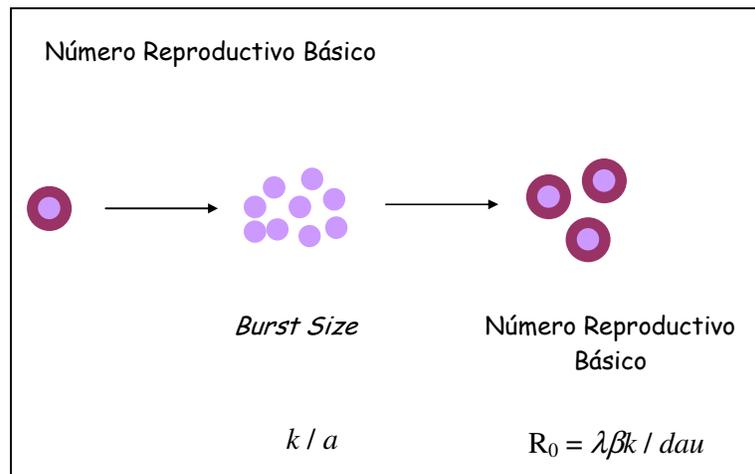


Figura IV.5: El *Burst Size* es el número total de viriones producidos a partir de cualquier célula. El número reproductivo básico es el número total de células infectadas a partir de una célula cualquiera en el principio de la infección. (Nowak *et al.* 2000)

La tasa a la cual una célula infectada da lugar a nuevas células infectadas viene dada por $\beta kx/u$. Si todas las células son no infectadas (individuo sano) entonces

$x = \lambda/d$. Y ya que la vida media de una célula no infectada es $1/a$ entonces tenemos que:

$$R_0 = \beta\lambda x/adu$$

Si cada célula infectada infecta como media a menos de una célula ($R_0 < 1$), entonces la infección no será viable, y el inóculo inicial del virus desaparecerá. Si cada célula infectada, por el contrario, infecta a más de una ($R_0 > 1$), entonces la infección sí tendrá lugar.

En el trabajo de Pybus *et al.* (2001), en el cual realizan un estudio acerca del comportamiento epidemiológico del virus de la hepatitis C, desarrollan un modelo epidemiológico básico teniendo en cuenta características propias del virus tales como que no existe una transmisión vertical y que existe poca protección contra una reinfección del virus. A partir de este modelo se deduce que R_0 vendría dado por la ecuación: $R_0 = BD$, donde B es una combinación de parámetros relacionados con la fuerza de la infección y D es la duración media de la enfermedad. A su vez la tasa de crecimiento exponencial del virus, r , (valor que obtenemos a partir del programa GENIE), viene relacionada con los parámetros B y D de la siguiente forma: $r = B - D^{-1}$. Por tanto podemos decir que:

$$R_0 = rD + 1$$

En su trabajo Pybus *et al.* 2001, realizaban, además, los análisis con GENIE y después comparaban los resultados con las estimas de R_0 obtenidas. Puesto que es difícil averiguar la fecha de infección de la hepatitis C, debido a que es una enfermedad asintomática, también es complejo saber cual es la duración exacta de la enfermedad. Así pues realizaron una estima de R_0 para varios valores de D (medido en años). En sus resultados encontraban cómo los individuos de genotipo 1a presentaban dinámicas de crecimiento exponencial, mientras que en los 1b se observaba una estabilidad en su crecimiento. Esto se correspondía con valores de R_0 algo mayores para 1a que para 1b, augurando un posible cambio en la prevalencia de los genotipos.

En nuestro caso hemos hecho una estima de R_0 para un valor de D correspondiente a 15 años. Para calcular r necesitamos una tasa de sustitución nucleotídica, μ , ya que el valor que obtenemos directamente a partir de GENIE es ρ , y

$\rho = r/\mu$. Puesto que dentro de los objetivos de esta tesis no estaba el cálculo de μ , realizamos estimas de R_0 con dos valores de esta tasa. En una de ellas utilizamos los mismos valores que en el trabajo de Pybus *et al.* 2001, y en la otra recurrimos a un valor de μ calculado por nuestro grupo de investigación para las mismas regiones (E1-E2 y NS5A) y los genotipos 1a y 1b. La decisión de utilizar las dos tasas para llevar a cabo las estimas de R_0 , fue para evitar en la medida de lo posible los sesgos, ya que el valor calculado por nuestro grupo se realizó utilizando solo a los pacientes de los que se poseía muestras tomadas a tres tiempos; a tiempo 0 (antes de someterse a tratamiento), tras 6 meses de tratamiento y tras 12 meses de tratamiento. Sin embargo las muestras utilizadas en esta tesis son todas las correspondientes a tiempo 0 y superan en número a las muestras de las que se posee los tres tiempos.

En las tablas IV.28-31 se presentan los resultados obtenidos utilizando ambas estimaciones de μ y un valor de D de 15 años. Las secuencias empleadas para este cálculo son las correspondientes al conjunto de datos que incluían las secuencias consenso (Cons E1a, ConsN1a, ConsE1b, ConsN1b) y las dos secuencias más divergentes dentro de cada individuo (MasdistE1a, MasdistN1a, MasdistE1b y MasdistN1b).

Tabla IV.28. R_0 calculado, para la región E1-E2, a partir de μ utilizado en Pybus *et al.* 2001.

	ρ	r	R_0
ConsE1a	83,619	0,066	1,991
MasdistE1a	41,595	0,033	1,493
R_0 Promedio			1,742
ConsE1b	45,989	0,036	1,545
MasdistE1b	50,475	0,040	1,598
R_0 Promedio			1,572

$\mu = 0.00079$ r : Tasa crecimiento exponencial
 $\rho: \rho/\mu$ R_0 : N° reproductivo básico
 R_0 Promedio: Promedio del valor de R_0 para ConsE1a y MasdistE1a, y ConsE1b y MasdistE1b

Tabla IV.29. R_0 calculado, para la región E1-E2, a partir de μ calculado en nuestro grupo.

	ρ	r	R_0
ConsE1a	83,619	0,276	5,139
MasdistE1a	41,595	0,137	3,059
R_0 Promedio			4,099
ConsE1b	45,989	0,152	3,276
MasdistE1b	50,475	0,167	3,498
R_0 Promedio			3,519

$\mu = 0.0033$ r : Tasa crecimiento exponencial
 $\rho: \rho/\mu$ R_0 : N° reproductivo básico
 R_0 Promedio: Promedio del valor de R_0 para ConsE1a y MasdistE1a, y ConsE1b y MasdistE1b

Tabla IV.30. R_0 calculado, para la región NS5A, a partir de μ utilizado en Pybus *et al.* 2001.

	ρ	r	R_0
ConsN1a	132,351	0,066	1,993
MasdistN1a	59,608	0,030	1,447
R_0 Promedio			1,720
ConsN1b	91,427	0,046	1,686
MasdiN1b	181,851	0,091	2,364
R_0 Promedio			2,025

$\mu = 0.0005$ r : Tasa crecimiento exponencial
 ρ : ρ/μ R_0 : N° reproductivo básico
 R_0 Promedio: Promedio del valor de R_0 para ConsN1a y MasdiN1a, y ConsN1b y MasdistN1b

Tabla IV.31. R_0 calculado, para la región NS5A, a partir de μ calculado en nuestro grupo.

	ρ	r	R_0
ConsN1a	132,35	0,232	4,474
MasdistN1a	59,608	0,104	2,565
R_0 Promedio			3,5195
ConsN1b	91,427	0,160	3,4
MasdiN1b	181,85	0,318	5,774
R_0 Promedio			4,5868

$\mu = 0.00175$ r : Tasa crecimiento exponencial
 ρ : ρ/μ R_0 : N° reproductivo básico
 R_0 Promedio: Promedio del valor de R_0 para ConsE1a y MasdistE1a, y ConsE1b y MasdistE1b

4 PRUEBAS DE NEUTRALIDAD

La observación de si un patrón de polimorfismos es consistente o no con el modelo de evolución neutral, resulta de gran interés en el estudio de la evolución. Existen diferentes pruebas estadísticas utilizadas con este fin, y que además permiten la identificación de que clase de fuerzas evolutivas, como la selección o los factores demográficos, están implicadas en la evolución de una región génica particular. La detección de la acción de la selección natural ha sido uno de los objetivos de los estudios genético poblacionales. Cuando actúa la selección natural no sólo afecta a la diana de selección, sino también a una serie de posiciones ligadas a estos sitios. Este tipo de selección produce una huella en la variación nucleotídica de estas posiciones ligadas, que nos permite su detección. No obstante, existen factores demográficos como cuellos de botella, subdivisión geográfica, que pueden dejar huellas muy similares a las producidas por la acción de la selección. Así, no es fácil discriminar entre factores selectivos o demográficos. La extensión de la huella a lo largo del genoma podría ayudar a diferenciarlos, ya que los factores demográficos suelen afectar a la totalidad del genoma, por lo que su huella sería detectable en casi la totalidad de los genes. Sin embargo los factores selectivos afectan a genes particulares, por lo que su marca sólo sería detectable en estas regiones génicas.

Así mismo en lo que respecta al rastro producido por la selección natural, existen a su vez diferentes modelos selectivos que dejan distintos tipos de huella (Figura IV.6):

- La selección estabilizadora es aquella que mantiene dos o más alelos en la población a frecuencias que se mantienen en equilibrio, favoreciendo de este modo un exceso de alelos a frecuencias intermedias y la permanencia de mutaciones neutras ligadas a ellas.
- El arrastre selectivo es un proceso por el cual cuando surge una mutación ventajosa, con el tiempo puede llegar a ser fijada por selección positiva junto con alguna mutación neutra asociada que ya existía en esa variante. De este modo se produce una disminución considerable de la variabilidad genética y se favorece el exceso de polimorfismos únicos.
- La selección de fondo produce un efecto similar al anterior, sólo que en este caso la selección actúa eliminando aquellas nuevas mutaciones que resultan deletéreas, eliminando también mutaciones neutras que pudieran encontrarse ligadas a esta variante.

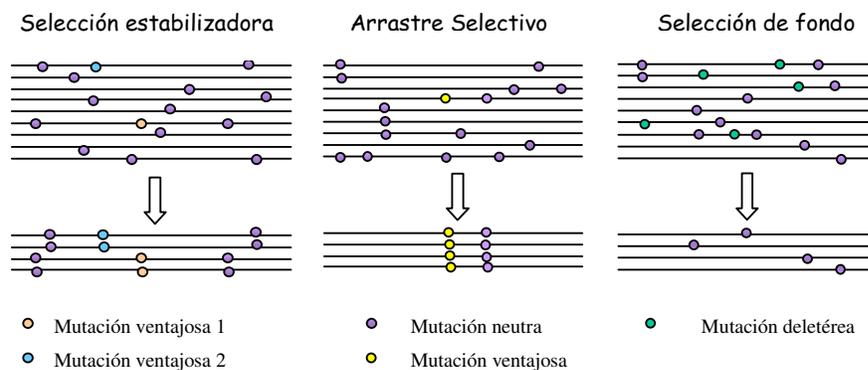


Figura IV.6. En el primer caso el mantenimiento de los dos alelos con las nuevas mutaciones, contribuye al mantenimiento de la variación genética. En el caso de arrastre selectivo se fija la nueva mutación junto con la mutación neutra asociada, aumentando su frecuencia pero disminuyendo la variabilidad. En el último caso las variantes que poseen mutaciones deletéreas son eliminadas de la población, por lo que se observa una disminución de la variabilidad ya que también han sido eliminadas las mutaciones neutras asociadas a la deletérea.

Así es posible inferir la acción de la selección a partir de su impacto sobre la variación genética. Algunas pruebas tales como la de Tajima (1989), Fu y Li (1993) o la de Fu (1997), han sido desarrolladas para determinar si el patrón de variación observado dentro de especies se comporta de acuerdo con el esperado bajo neutralidad (su base teórica fue explicada en el apartado de material y métodos). El rechazo de la hipótesis nula, supone que existen fuerzas evolutivas que provocan el alejamiento de la situación de neutralidad. No obstante, así como la obtención de los datos y la elaboración de las pruebas no suponen una tarea complicada, sí lo es por el contrario la interpretación de las mismas. Existen diferentes trabajos (Simonsen *et al.* 1995; Fu 1997; Ramos-Onsins y Rozas 2002), en los que se analiza el poder estadístico de estas pruebas, así como diversas consideraciones que hemos tenido en cuenta en la interpretación de nuestros resultados. A parte de prestar una especial atención a las consideraciones estadísticas anteriores, habría que tener también en cuenta, la función del gen al que corresponden las secuencias a la hora de postular nuestras hipótesis.

Como hemos visto anteriormente, cabría esperar que fenómenos como el arrastre selectivo y la selección de fondo, al disminuir la variabilidad genética debido a la fijación de una mutación ventajosa o eliminación de una deletérea (Figura IV.6), sean responsables de un exceso de alelos de frecuencia única. Esta situación provoca que los estadísticos D de Tajima y D^* y F^* de Fu y Li sean significativamente menores que cero (aunque esta situación también la encontramos cuando se ha producido un aumento en el tamaño de la población debido a eventos tales como un cuello de botella). Por otra parte fenómenos como la selección estabilizadora provocarían un exceso de alelos a frecuencias intermedias que se reflejará en valores significativamente mayores que cero para D , D^* y F^* , si bien también podemos obtener estos valores debido a una disminución en el tamaño poblacional de la población. Sin embargo como hemos comentado y como veremos en el apartado de discusión, las observaciones anteriores aún requieren ciertos matices para su interpretación. Así en este apartado sólo vamos a detenernos en comentar los resultados obtenidos

Las diferentes pruebas fueron llevadas a cabo para cada una de las muestras con el programa DNAsp v4.0. En las tablas IV.32-IV.35 quedan reflejados los resultados obtenidos.

Comenzaremos primero explicando qué sucede en la región E1-E2. Lo primero que tenemos que ver es en qué casos aceptamos o rechazamos la hipótesis nula de

neutralidad. Y aquí encontramos el primer problema. En algunos casos como el eA34 (genotipo 1a), aceptaríamos la hipótesis nula si sólo nos fijásemos en la D de Tajima, ya que el valor no es significativamente distinto de cero, pero la rechazaríamos si nos fijásemos en la D^* y F^* de Fu y Li. Así para aceptar o rechazar la situación de neutralidad, establecemos el siguiente criterio: consideraremos que aceptamos la hipótesis nula cuando ninguno de los estadísticos sea significativamente diferente de 0. Este es el caso de por ejemplo el eA26 (genotipo 1a) o el eA14 (genotipo 1b). En el resto de individuos, observamos que al menos uno de los estadísticos es significativo, y en la mayoría de casos los valores son menores que 0. Esto indicaría que se podrían estar produciendo fenómenos de selección purificadora, arrastre selectivo, selección de fondo o bien podría deberse a un aumento en el tamaño de la población. Sin embargo ¿cómo podemos averiguar cual de estos procesos es el que estamos observando realmente? En estos casos tendremos en cuenta que hay pruebas con mayor poder estadístico que otras, para detectar un determinado tipo de fuerza evolutiva. Según Fu 1997, y Simonsen *et al.* 1995, los estadísticos F_s y D de Tajima, son más sensibles en detectar arrastre selectivo; mientras que D^* y F^* los son para la selección de fondo (Fu 1997). ¿Pero que ocurre cuando tenemos valores significativos para los cuatro estadísticos o para de tres de ellos? Esto sucede, por ejemplo, en eA06 y eC09 (de genotipo 1b) o eA09 y eC06 (de genotipo 1a). En estos casos, y como discutiremos más adelante, (ver punto 3 del apartado de discusión) tendremos en cuenta las características funcionales y estructurales del gen correspondiente a la región E1-E2, para dar una interpretación lo más aproximada posible. En un número muy reducido de casos (por ejemplo eC15T0 y eG02 para el genotipo 1b, o eC20 y eC33 para 1a) encontramos valores de D , D^* y F^* mayores que 0. Puesto que no son significativamente distintos de cero, aplicaremos los criterios anteriores para su interpretación. Sólo uno de los casos, eC18, sí tiene valores significativamente mayores que 0, lo que estaría indicando la presencia de selección estabilizadora o una disminución del tamaño poblacional, no obstante lo trataremos más detenidamente en el apartado de discusión.

En cuanto a la región NS5A, la situación referida a los valores de D , D^* y F^* no difiere mucho de la encontrada para la E1-E2. Todos los valores de los estadísticos son menores que 0, excepto la D de Tajima para nA20 (1a). Así la discusión de los resultados se realizará aplicando los mismos criterios que en la región E1-E2.

Con respecto a si estas pruebas nos ofrecen algún dato que permita establecer diferencias entre los diferentes genotipos, la respuesta es que no. Encontramos la misma

situación tanto en el genotipo 1b como en el genotipo 1a, por lo que ambos deben estar sometidos a las mismas fuerzas evolutivas y tampoco aporta ninguna característica que permita explicar la mayor prevalencia de un genotipo sobre otro.

Tabla IV.32. Pruebas de neutralidad para la región E1-E2 en los individuos de genotipo 1b. El símbolo “*” representa el grado de significación.

1b	<i>D</i>	<i>D</i> *	<i>F</i> *	<i>F</i> _s	Div Nuc	<i>k</i> :	(<i>S</i>)
eA03	-0,9074	-0,7729	-0,996	-64,47 ***	0,0263	12,43	87
eA06	-2,5914 ***	-4,5647 **	-4,525 **	-49,92 ***	0,003	1,4111	40
eA14	-1,1315	-0,7462	-1,093		0,0329	15,518	107
eA16	-0,6918	-1,9268	-1,673		0,0539	25,423	150
eA35	-1,7169	-4,3496 **	-3,901 **	-34,19 ***	0,0117	5,5369	61
eC03	0,71403	-1,9677	-1,245	1,256	0,0042	2	8
eC04	-1,9088 *	-4,2773 **	-4,127 **	-9,082 ***	0,0002	0,1	5
eC05	-0,8373	-2,8816 *	-2,403 *		0,0469	22,131	134
eC08	-0,5121	-1,9226	-1,583	-34,41 ***	0,046	21,729	119
eC09	-2,2859 **	-3,8994 **	-3,871 **		0,007	3,317	61
eC10	-0,8853	-2,6926 *	-2,319 *		0,0266	12,546	83
eC12	-1,352	-2,2797	-2,264		0,0249	11,732	89
eC13	-2,2858 **	-4,1386 **	-4,029 **		0,0116	5,5188	87
eC15	0,4215	-1,3718	-0,724		0,0579	27,33	113
eC16	-0,9299	0,53098	-0,1	-41,38 ***	0,0251	11,863	78
eC19	-1,0199	-2,7724 *	-2,425 *	-34,18 ***	0,04	18,886	126
eC25	-1,9371 *	-2,4358 *	-2,684 *	-10,08 **	0,0008	0,3891	9
eG02	0,21588	-0,7121	-0,383	-24,74 ***	0,0449	21,198	93
eG05	-2,0612 *	-3,6812 **	-3,604 **	-62,86 ***	0,0124	5,8414	76
eG06	-1,4271	-3,499 **	-3,138 **		0,0287	13,528	112
eG07	-0,3493	-2,263	-1,74	-61,21 ***	0,0399	18,826	103
eG09	-1,9054 *	-3,6306 **	-3,472 **		0,0225	10,613	121
eG16	-1,3742	-2,8394 *	-2,792 *	-4,076 **	8E-05	0,0388	2
eG17	0,75428	-0,2876	0,1851	-34,22 ***	0,0507	23,935	97
eG26	-2,1439 *	-4,3879 **	-4,289 **	-13,46 ***	0,0004	0,1796	8

Div Nuc: Diversidad nucleotídica
 (*S*): número de sitios segregantes
 (*k*): número de diferencias emparejadas
D: Estadístico de la prueba de Tajima
*D** y *F**: Estadísticos de las pruebas de Fu y Li

*F*_s: Estadístico de la prueba de Fu
 *: p < 0.05
 **: p < 0.02
 ***: p < 0.01

Tabla IV.33. Prueba de neutralidad para la región NS5A en los individuos de genotipo 1b. El símbolo “**” representa el grado de significación.

1b	<i>D</i> :	<i>D</i> *	<i>F</i> *	<i>F</i> _s	Div Nuc	<i>k</i> :	(<i>S</i>)
nA03	-1,8895 *	-3,3594 **	-3,382 **	-20,64 ***	0,0132	9,8137	81
nA06	-2,6782 ***	-6,9392 **	-6,225 **		0,003	2,2018	63
nA14	-2,0624 *	-2,1742	-2,55 *	-44,86 ***	0,0078	5,7901	65
nA16	-1,6511	-3,0799 *	-3,038 *	-38,51 ***	0,0208	15,436	124
nA35	-1,436	-2,6562 *	-2,667 *	-12,17 ***	0,0062	4,5841	28
nC03	-1,9263 *	-0,9204	-1,549	-22,57 ***	0,0022	1,6438	23
nC04	-0,5779	-0,9277	-0,95	-49,71 ***	0,0058	4,2924	27
nC05	-1,7343	-0,4232	-1,134		0,0103	7,6137	72
nC08	-1,913 *	-4,9305 **	-4,43 **	-22,69 **	0,0187	13,913	144
nC09	-0,7469	-1,7569	-1,67	-12,89 ***	0,0033	2,4436	15
nC10	-1,4794	-1,7	-1,933	-22,51 ***	0,0076	5,6764	48
nC12	-1,778	-3,2177 *	-3,136 **		0,0147	10,919	114
nC13	-0,1676	-1,5352	-1,182	-28,27 ***	0,0419	31,099	153
nC15	-0,9381	-1,8326	-1,769	-12,65 **	0,0125	9,2522	62
nC16	-2,3082 **	-3,8847 **	-3,893 **		0,0117	8,6827	124
nC19	-1,656	-3,4791 **	-3,284 **	-43,43 ***	0,0234	17,419	153
nC25	-1,7433	-2,8149 *	-2,906 *	-8,495 ***	0,0015	1,1441	11
nG02	-1,3776	-1,8519	-1,995	-30,56 ***	0,0216	16,011	121
nG05	-0,9127	-2,9712 *	-2,596 *	-21,62 ***	0,0247	18,38	110
nG06	-1,5559	-3,4072 *	-3,212 *	-47,83	0,0263	19,53	155
nG07	-0,4125	-2,333	-1,934	-16,47 ***	0,0363	26,935	133
nG09	-0,7498	-1,3237	-1,321	-13,03 **	0,0226	16,792	97
nG16	-1,4497	-2,5924 *	-2,618 *	-3,215 **	9E-05	0,0702	2
nG17	-1,7496	-1,4592	-1,888	-37,85 ***	0,0155	11,49	107
nG26	-1,9325 *	-4,1458 **	-4,034 **	-8,49 ***	0,0002	0,1191	5

Div Nuc: Diversidad nucleotídica
(*S*): número de sitios segregantes
(*k*): número de diferencias emparejadas
D: Estadístico de la prueba de Tajima
*D** y *F**: Estadísticos de las pruebas de Fu y Li

*F*_s: Estadístico de la prueba de Fu
*: p < 0.05
***: p < 0.02
***: p < 0.01

Tabla IV.34. Prueba de neutralidad para la región E1-E2 en los individuos de genotipo 1a. El símbolo “*” representa el grado de significación.

1a	D:	D*	F*:	Fs	Div Nuc.	k:	(S)
eA05	-1,4591	-0,18329	-0,74	-7,302 **	0,0015	0,722	9
eA09	-2,3484 **	-2,78763 *	-3,122 **		0,0124	5,8368	96
eA20	-1,5171	-3,11531 *	-2,908 *		0,024	11,344	101
eA26	-1,0055	-1,21925	-1,354		0,0342	16,141	104
eA27	-0,7812	-0,33744	-0,626	-27,11 ***	0,0208	9,815	63
eA34	-1,4372	-4,06968 **	-3,534 **		0,0292	13,792	118
eC01	-1,4553	-2,97643 *	-2,798 *	-53,82 ***	0,0242	11,441	95
eC02	-1,5577	-1,47806	-1,794	-20,83 ***	0,0022	1,0528	13
eC06	-2,5796 ***	-5,45112 **	-5,213 **	-39,95 ***	0,0012	0,558	23
eC07	-0,8857	-2,23601	-1,994	..	0,0329	15,513	98
eC11	-2,4145 **	-3,84332 **	-3,88 **		0,0094	4,4345	85
eC14	-0,3624	-2,01993	-1,684	-5,066	0,0057	2,6844	16
eC17	-1,6515	-3,88819 **	-3,58 **	-37,48 ***	0,008	3,772	41
eC18	0,89128	1,8782 **	1,7593 *	-2,393	0,0407	19,187	76
eC20	0,24024	-1,14644	-0,677	-26,4 ***	0,0364	17,178	75
eC21	0,57893	0,79079	0,8441	-18 *	0,0304	14,354	58
eC22	0,0747	1,45854	1,0468	-50,42 ***	0,0431	20,345	91
eC24	-1,0486	-2,32583	-2,197	-38,12 ***	0,0066	3,1301	25
eC28	-0,5162	0,36072	-0,017	-43,56 ***	0,0435	20,549	115
eC31	-1,9715 *	-1,75625	-2,216	-24,51 ***	0,0123	5,7865	72
eC33	0,7323	-1,49359	-0,808	-9,123 **	0,0083	3,8992	16
eC38	-2,0181 *	-0,87217	-1,583	-36,3 ***	0,0045	2,1075	31
eG10	-1,6086	-3,41101 **	-3,334 **	-5,718 **	0,0001	0,06	3
eG14	-1,3529	-1,09555	-1,446	-30,27 **	0,0256	12,076	102
eG19	-0,5743	-0,14752	-0,392	-31,03 ***	0,0277	13,053	80

Div Nuc: Diversidad nucleotídica
(S): número de sitios segregantes
(k): número de diferencias emparejadas
D: Estadístico de la prueba de Tajima
D* y F*: Estadísticos de las pruebas de Fu y Li

Fs: Estadístico de la prueba de Fu
*: p < 0.05
***: p < 0.02
***: p < 0.01

Tabla IV.35. Prueba de neutralidad para la región NS5A en los individuos de genotipo 1a. El símbolo “*” representa el grado de significación.

1a	D:	D*	F*:	F _s	Div Nuc	k:	(S)
nA05	-1,6681	-1,05912	-1,454	-11,3***	0,0028	2,08	15
nA09	-1,1277	-1,14909	-1,351	-16,5***	0,0123	9,1472	52
nA20	0,04616	-0,20599	-0,133	-11,02**	0,0097	7,223	34
nA26	-1,317	-2,52934*	-2,464*	-25,7***	0,0192	14,274	106
nA27	-1,5982	-1,75871	-2,032	-32,26***	0,0126	9,3318	81
nA34	-1,4142	-2,05519	-2,178	-23,81***	0,0133	9,7241	59
nC01	-0,8558	-0,15846	-0,521	-23,77***	0,028	20,808	125
nC02	-1,6912	-3,77946**	-3,593**	-14,04***	0,0018	1,3393	17
nC06	-0,5733	-1,36873	-1,31	-0,157	0,0008	0,5916	4
nC07	-1,4	-3,43243**	-3,162*	-28,06***	0,017	12,627	97
nC11	-0,8547	0,26927	-0,173	-2,622	0,0034	2,4857	18
nC14	-0,6904	-1,77996	-1,69	-7,171***	0,0033	2,4138	12
nC17	-1,9495*	-2,48365	-2,702*	-3,854*	0,0007	0,5533	6
nC18	-2,1322*	-4,68171**	-4,526**	-2,979	0,0002	0,1795	7
nC20	-1,12	-3,89589**	-3,388**	-19,98***	0,0047	3,4886	28
nC21	-2,4312**	-5,2427**	-4,923**	-34,24***	0,0064	4,7563	84
nC22	-1,6016	-2,33099	-2,464	-30,23***	0,0186	13,849	98
nC24	-1,8648*	-2,88982*	-3,006*	-28,63***	0,0032	2,3984	25
nC28	-1,5906	-2,1246	-2,302	-30,27***	0,0152	11,273	87
nC31	-1,9639*	-1,93142	-2,322*	-5,15	0,0033	2,4269	30
nC33	-1,6497	-3,25036*	-3,22**	-5,082**	0,0001	0,0811	3
nC38	-1,7924*	-3,81386**	-3,721**	-4,71**	0,0001	0,0889	4
nG10	-1,1796	-1,46193	-1,605	-3,739*	0,0015	1,0926	7
nG14	-1,8833*	-1,16899	-1,754	-43,95***	0,0108	8,0215	91
nG19	-1,5595	-0,74594	-1,251	-47,61***	0,0107	7,9389	63

Div Nuc: Diversidad nucleotídica
(S): número de sitios segregantes
(k): número de diferencias emparejadas
D: Estadístico de la prueba de Tajima
D* y F*: Estadísticos de las pruebas de Fu y Li

F_s: Estadístico de la prueba de Fu
*: p < 0.05
***: p < 0.02
***: p < 0.01

5. BÚSQUEDA DE SELECCIÓN POSITIVA.

Todavía es desconocido el mecanismo por el cual el VHC presenta una alta tasa de persistencia viral y un variable desarrollo clínico de la enfermedad. Todo parece apuntar, no obstante, a una combinación entre el sistema inmune del hospedador y la variabilidad viral (Farci *et al.* 2000). En algunos estudios se ha observado que la región hipervariable situada en el gen E2 que codifica para una proteína de la envuelta del virus, podría ser uno de los determinantes antigénicos hacia donde van dirigidos algunos anticuerpos neutralizantes, y por tanto se sugiere que la gran variabilidad de esta región permitiría al virus escapar a esta respuesta inmune (Curran *et al.* 2002). Si esto es así, ¿sería debido a que en algunas posiciones con alta variabilidad está actuando la selección positiva, o bien se debe a otras fuerzas como la deriva genética y la selección purificadora?

La relación tasa de mutación no sinónima/sinónima ($\omega=dN/dS$) es un importante indicador de la presión selectiva a nivel proteico. Así cuando encontramos $\omega=1$ indicaría que estamos ante mutaciones neutras, $\omega>1$ ante selección positiva y $\omega<1$ ante selección purificadora. Los primeros estudios en la estima de esta tasa realizaban el cálculo promediando las tasas de sustitución sinónima (dS) y no sinónima (dN), sobre todos los sitios de la proteína. No obstante esta forma de cálculo puede no ser la más adecuada (Sharp 1997; Akashi 1999; Crandall *et al.* 1999), ya que en muchas proteínas una gran proporción de aminoácidos pueden ser muy invariables, debido a restricciones funcionales ($\omega=0$), y además la mayoría de proteínas parecen estar bajo selección purificadora la mayor parte del tiempo (Li 1997). De esta forma es muy difícil que ω alcance valores mayores que 1, incluso aunque se esté dando selección adaptativa en algunos aminoácidos. El método propuesto por Yang *et al.* 2000, consiste en desarrollar una serie de modelos estadísticos, donde se contempla la posibilidad de considerar ω variables en las diferentes posiciones aminoacídicas, es decir tener en cuenta que pueda haber sitios invariables, en los que cualquier mutación será eliminada por selección, sitios neutros y sitios en los que una mutación pueda resultar ventajosa y ser seleccionada positivamente. Después se aplican estos modelos a análisis filogenéticos de secuencias de DNA que codifican para proteínas. Encontrar $\omega>1$, significaría que las mutaciones no sinónimas ofrecen ventajas a la proteína y por tanto tienen mayor probabilidad de fijación que las sinónimas. Estos métodos están implementados en el programa CODEML del paquete de programas PAML.

Nosotros hemos aplicado el método para detectar si hay algunas posiciones que estén sometidas a selección positiva que puedan explicar el escape del sistema inmune y si es así, comprobar si tales posiciones coinciden en ambos genotipos. De lo contrario tendríamos otra causa potencial de la mayor prevalencia de un genotipo sobre otro.

En este caso llevamos a cabo el análisis para cada uno de los pacientes y para ambas regiones (E1-E2 y la NS5A). Puesto que los métodos de máxima verosimilitud requieren de mucho tiempo computacional, y además tenemos un gran número de secuencias para cada uno de los pacientes, nos hemos servido de una aplicación del programa para grandes filogenias (ver material y métodos apdo 5.6.1).

De los 13 modelos posibles de ω variables entre sitios utilizamos los modelos 0, 1, 2, 3, 7 y 8. El modelo 0 asume el mismo ω para todos los sitios, el 1 y el 7 solo asumen sitios con $\omega=0$ y $\omega<1$ (son los modelos neutros), y los modelos 2, 3, y 8 permiten además sitios con $\omega>1$. Para comparar tales modelos se utiliza una prueba de razón de verosimilitudes, cuando los dos modelos están anidados. En el caso de que los modelos no lo estén, hacemos la comparación utilizando el Criterio de Información de Akaike (AIC) ($AIC = -2 (\log \text{verosimilitud del modelo}) + 2 (\text{número de parámetros del modelo})$). El modelo que minimice el AIC, se considera el más apropiado.

Si son los modelos de selección positiva quienes proporcionan una mejor explicación para la evolución de estas regiones, se procede entonces a la identificación de los sitios seleccionados positivamente, mediante inferencia bayesiana (Nielsen y Yang 1998). De acuerdo con este método, la categoría de sitios (con una tasa dN/dS asociada) que maximice la probabilidad posterior, es la más probable para el sitio aminoacídico. La probabilidad posterior proporciona una medida de la precisión de la inferencia.

En el ANEXO VII se muestran las estimas de los parámetros para los seis modelos utilizados en cada uno de los pacientes, así como las distribuciones de las probabilidades posteriores. En la tabla IV.36 de este apartado se muestra para cada paciente el modelo más apropiado tras la comparación, según el Criterio de Información de Akaike.

Si observamos en el ANEXO VII (Tablas 14-17) nos damos cuenta que en general la ratio dN/dS del modelo 0 comprende valores entre 0 y 1, casi no hay casos donde encontramos valores mayores que uno, lo que estaría indicando que es la selección purificadora la que actúa mayoritariamente sobre las regiones E1-E2 y NS5A. No obstante al aplicar la prueba de razón de verosimilitudes o el AIC (ver tabla IV.36),

los modelos más apropiados para la mayoría de individuos, son aquellos que asumen que una proporción de sitios, muy pocos, están sometidos a una fuerte presión de selección positiva, éstos son los modelos 2, 3 y 8 (ver tabla IV.36).

Tabla IV.36. Modelo según el Criterio de Información de Akaike que más se adecua a cada individuo para ambas regiones.

E1-E2				NS5A			
Individuo 1a	Modelo	Individuo 1b	Modelo	Individuo 1a	Modelo	Individuo 1b	Modelo
eA03	M3	eA05	M1	nA03	M8	nA05	M0
eA06	M1	eA09	M3	nA06	M8	nA09	M1
eA14	M3	eA20	M3	nA14	M2	nA20	M2
eA16	M3	eA26	M3	nA16	M8	nA26	M3
eA35	M8	eA27	M3	nA35	M2	nA27	M8
eC03	M8	eA34	M3	nC03	M2	nA34	M2
eC04	¿?	eC01	M3	nC04	M7	nC01	M8
eC05	M3	eC02	M2	nC05	M8	nC02	M3
eC08	M3	eC06	M1	nC08	M2	nC06	M0
eC09	M3	eC07	M3	nC09	M0	nC07	M3
eC10	M8	eC11	M3	nC10	M2	nC11	M2
eC12	M3	eC14	M2	nC12	M8	nC14	M1
eC13	M8	eC17	M8	nC13	M2	nC17	M0
eC15	M8	eC18	M3	nC15	M3	nC18	M0
eC16	M3	eC20	M3	nC16	M2	nC20	M8
eC19	M3	eC21	M3	nC19	M3	nC21	M2
eC25	M2	eC22	M8	nC25	M2	nC22	M2
eG02	M3	eC24	M3	nG02	M2	nC24	M2
eG05	M8	eC28	M3	nG05	M3	nC28	M2
eG06	M3	eC31	M8	nG06	M3	nC31	M2
eG07	M3	eC33	M2	nG07	M8	nC33	M0
eG09	M8	eC38	M2	nG09	M3	nC38	M0
eG16	M0	eG10	M0	nG16	M0	nG10	M2
eG17	M3	eG14	M3	nG17	M8	nG14	M8
eG26	M2	eG19	M8	nG26	M0	nG19	M2

M0: modelo con una ratio dN/dS, M1: Modelo Neutral, M2: Modelo de Selección, M3 Modelo discreto, M7: Modelo beta, M8: Modelo beta y omega.

¿?: El programa daba error al ejecutar los cálculos.

También encontramos algunos individuos que presentan resultados diferentes a los anteriormente descritos, con respecto al valor de dN/dS , y que cabría detenerse a explicar. En algunos casos (eC21, eG26, nA06 y nA14) se observan valores dN/dS ligeramente mayores que uno pero muy cercanos a este valor, lo que implicaría que se encuentran bajo neutralidad, aunque también presentan una pequeña proporción de sitios bajo selección positiva. Los individuos eG09 y nG16 presentan valores dN/dS de infinito. Esto supone que el número de mutaciones sinónimas es igual a 0 y por tanto la mayoría de posiciones serían fijadas por selección positiva.

A simple vista no se observan diferencias entre genotipos, en ambos casos encontramos la misma situación: valores de dN/dS menores que uno indicando la actuación de la selección purificadora, pero a la vez también encontramos pequeñas proporciones de sitios bajo selección positiva. En las figuras IV.7-IV.10 se muestra un alineamiento formado por una secuencia de cada uno de los pacientes, en el que se han señalado las posiciones que se encuentran bajo selección positiva. En las secuencias pertenecientes a la región E1-E2 (Figuras IV.7 y IV.8) se puede observar la presencia de unas posiciones seleccionadas positivamente que se repiten en la mayoría de secuencias. Estos sitios se encuentran dentro de la HVR1, y además coinciden en ambos genotipos. Por tanto, parece que estas posiciones podrían ser relevantes para la evasión del mecanismo de defensa del sistema inmune por parte del virus. En la región correspondiente a la proteína NS5A (Figuras IV.9 y IV.10) aunque también encontramos posiciones bajo selección positiva, su número es menor que en la región E1-E2, y además las diferentes secuencias no parecen coincidir en las posiciones seleccionadas, como ocurría en la región anterior. En la región ISDR, otra de las regiones que parecen estar implicadas en la cronicidad de la enfermedad, apenas si encontramos posiciones fijadas en ambos genotipos.

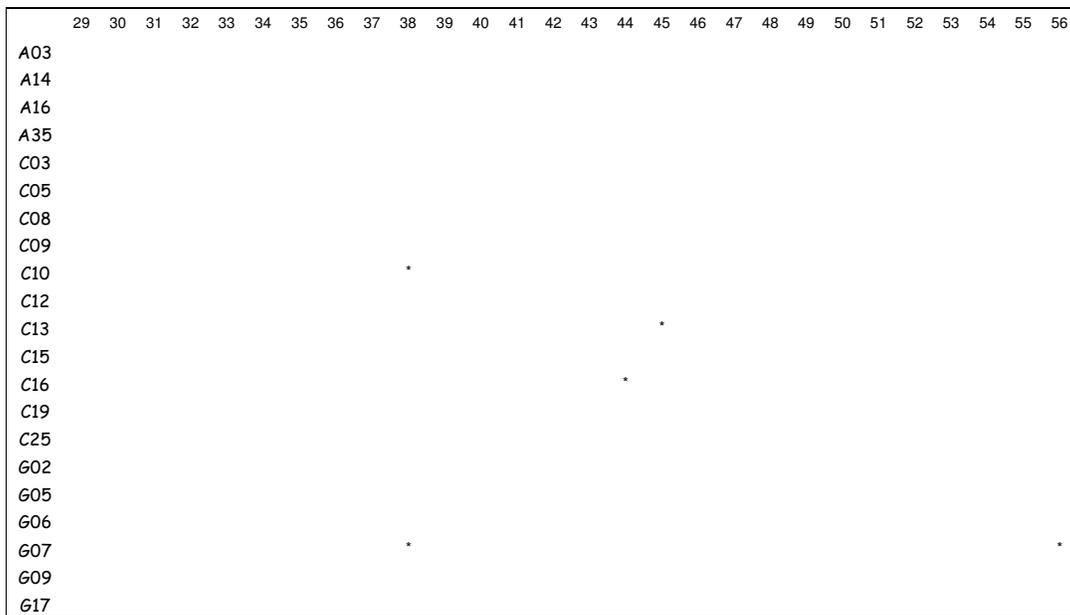
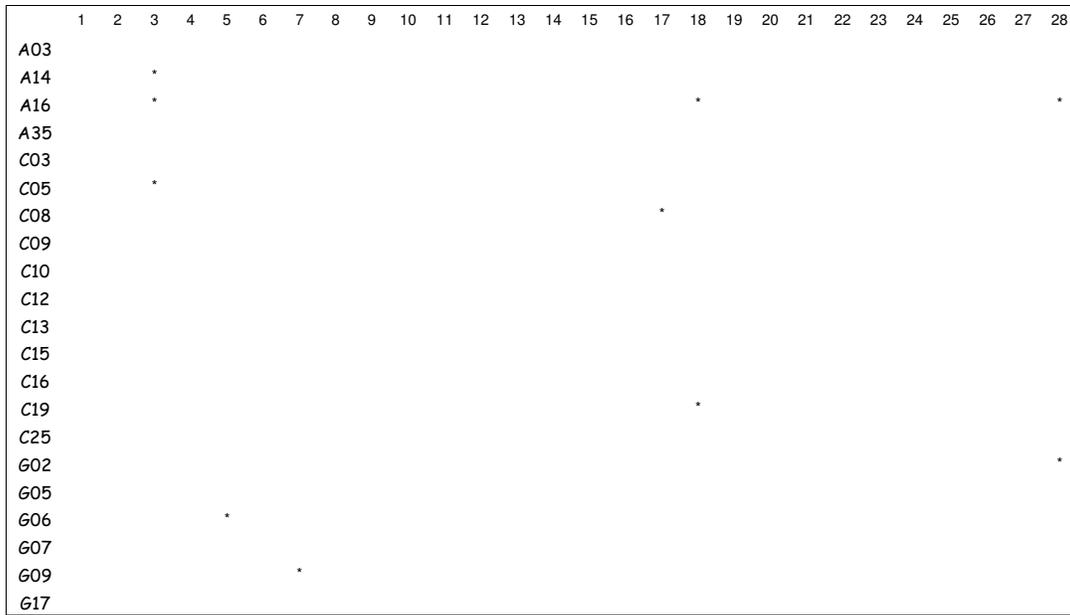


Figura IV.7: Con un asterisco se han señalado aquellos sitios aminoacídicos que se encuentran bajo selección positiva en la región E1-E2 de los individuos de genotipo 1b. En morado se ha señalado las posiciones correspondientes a la región hipervariable.

	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
A03			*				*							*			*				*	*		*				
A14	*		*		*			*	*	*	*	*	*	*	*	*			*	*	*	*		*		*		*
A16	*		*					*		*		*	*	*	*	*	*		*	*	*	*		*	*	*		*
A35			*																						*			
C03														*	*	*												
C05	*		*	*	*			*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
C08	*	*	*	*				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
C09	*		*					*	*					*							*				*			
C10				*													*					*					*	
C12	*		*	*				*	*	*	*	*	*	*	*	*	*		*							*		*
C13			*																									
C15	*		*	*			*				*			*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
C16			*	*	*	*		*		*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
C19	*		*		*						*		*			*	*	*	*	*	*	*	*	*	*	*	*	*
C25																												
G02	*		*					*	*			*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
G05								*	*					*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
G06			*											*														
G07			*					*	*				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
G09	*												*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
G17	*		*					*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112
A03																												
A14																							*				*	
A16																												*
A35																												*
C03																												
C05																												
C08			*																					*				*
C09																												
C10																												
C12																												
C13																										*		
C15													*		*								*			*		
C16																												
C19																												
C25																												
G02																								*				
G05																												
G06																												
G07																											*	
G09																								*				
G17																								*				

Figura IV.7. Continuación.

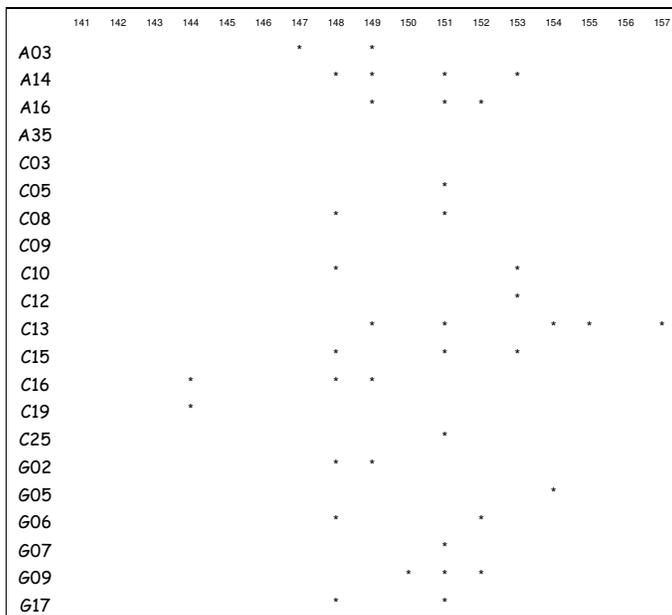
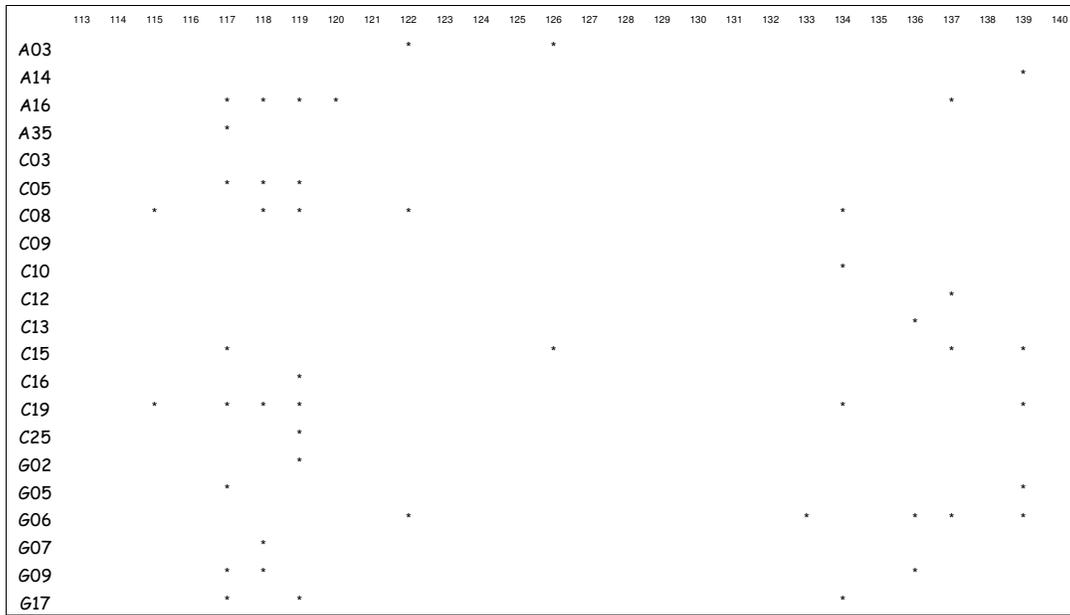


Figura IV.7. Continuación

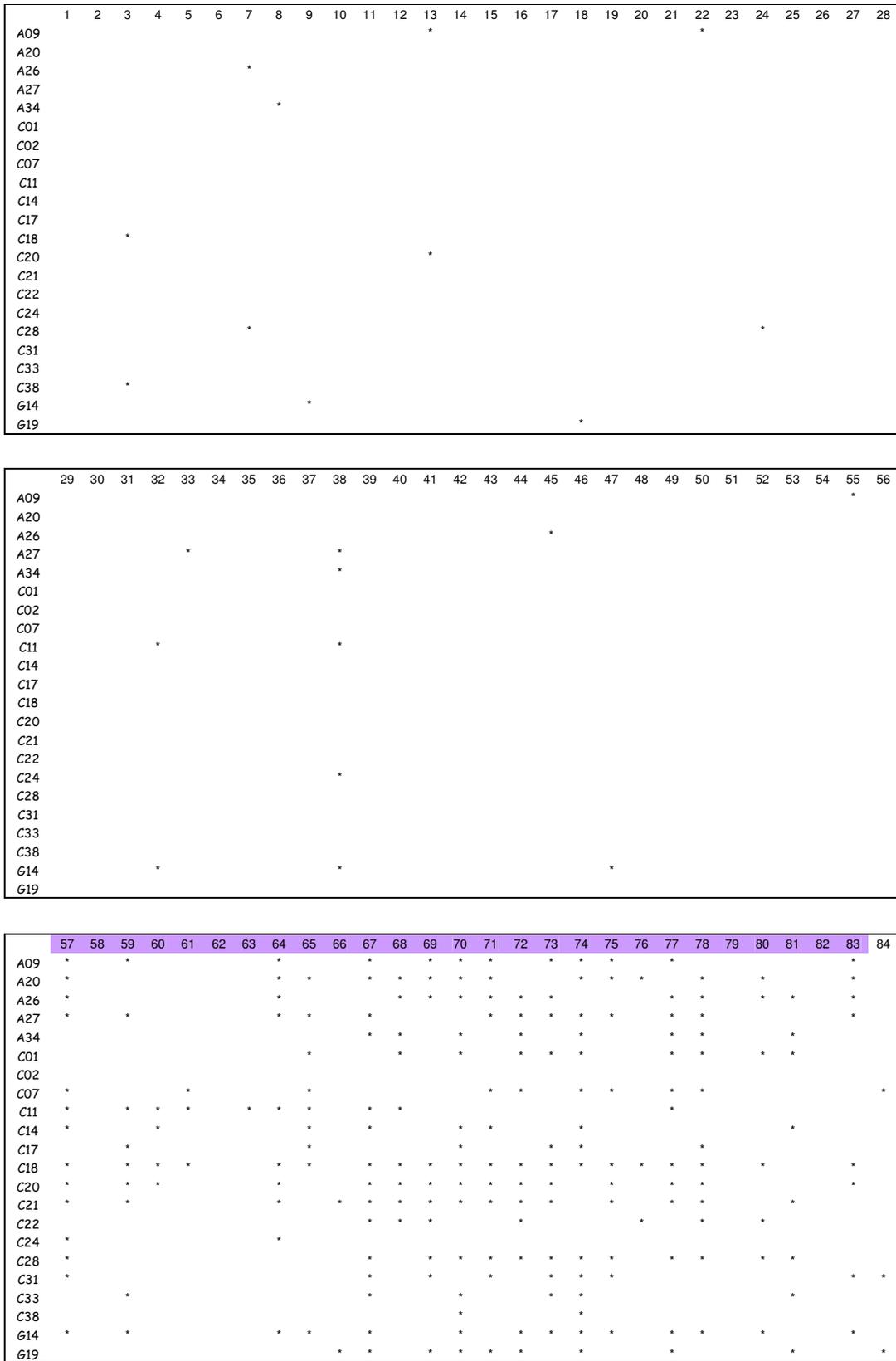


Figura IV.8. Posiciones aminoacídicas que se encuentran bajo selección positiva en la región E1-E2 de los individuos de genotipo 1a. En morado se ha señalado las posiciones correspondientes a la región hipervariable.

	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	
A09																						*						*	
A20																							*					*	
A26																*				*		*	*						*
A27																													*
A34												*																*	
C01			*			*					*																	*	
C02																												*	
C07					*	*																			*			*	
C11																												*	
C14																												*	
C17																												*	
C18																												*	
C20																								*				*	
C21																												*	
C22																								*				*	
C24																												*	
C28																			*				*					*	
C31																												*	
C33																							*					*	
C38																									*			*	
G14																							*	*				*	
G19																							*	*				*	

	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	
A09			*		*	*																				*			*
A20			*								*																		*
A26			*			*					*											*		*					*
A27			*	*	*	*	*				*											*			*				*
A34			*	*	*	*	*				*											*		*					*
C01			*		*	*	*				*															*			*
C02																													*
C07					*	*	*				*													*		*			*
C11					*	*	*				*										*								*
C14															*								*						*
C17															*														*
C18		*					*				*												*		*	*		*	*
C20		*									*												*		*		*		*
C21					*	*	*				*											*		*		*			*
C22					*	*	*			*											*				*		*		*
C24											*											*		*		*			*
C28		*		*	*	*	*				*			*	*	*					*		*		*		*		*
C31		*									*											*		*		*			*
C33											*											*		*		*			*
C38							*				*											*	*	*		*			*
G14							*				*											*	*	*		*			*
G19		*		*	*	*	*				*				*							*	*	*		*			*

	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	
A09						*	*										
A20						*	*		*	*	*						*
A26								*		*	*					*	*
A27																*	*
A34																*	*
C01														*		*	*
C02			*					*		*	*					*	*
C07																*	*
C11						*		*		*	*					*	*
C14								*		*	*					*	*
C17																*	*
C18						*	*	*		*	*					*	*
C20						*	*	*		*	*					*	*
C21							*	*		*	*					*	*
C22													*		*	*	*
C24																*	*
C28			*		*	*	*	*		*	*					*	*
C31																*	*
C33																*	*
C38			*				*									*	*
G14							*									*	*
G19							*									*	*

Figura IV.8. Continuación.

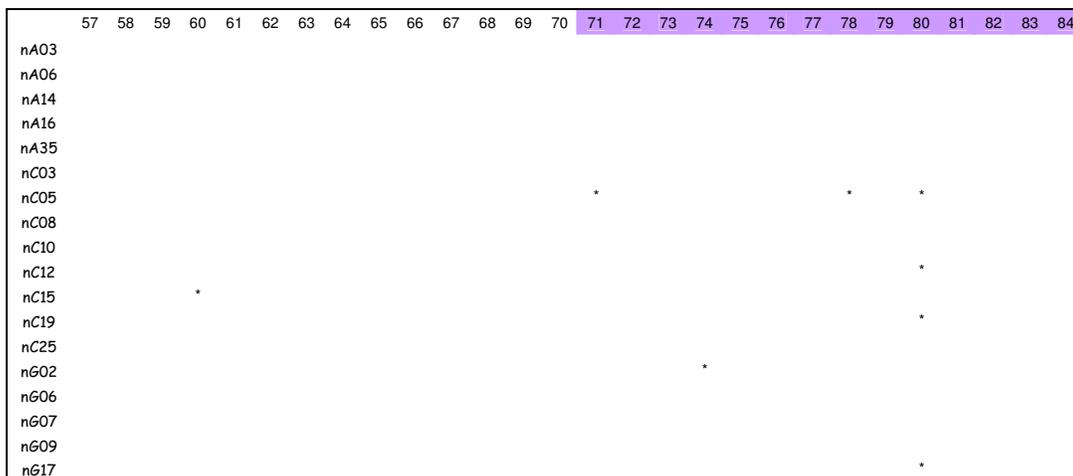
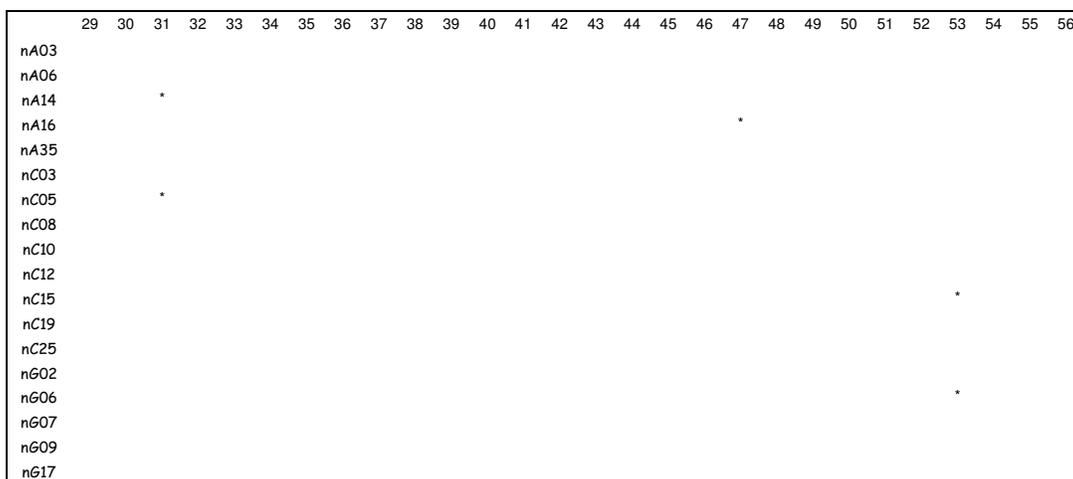
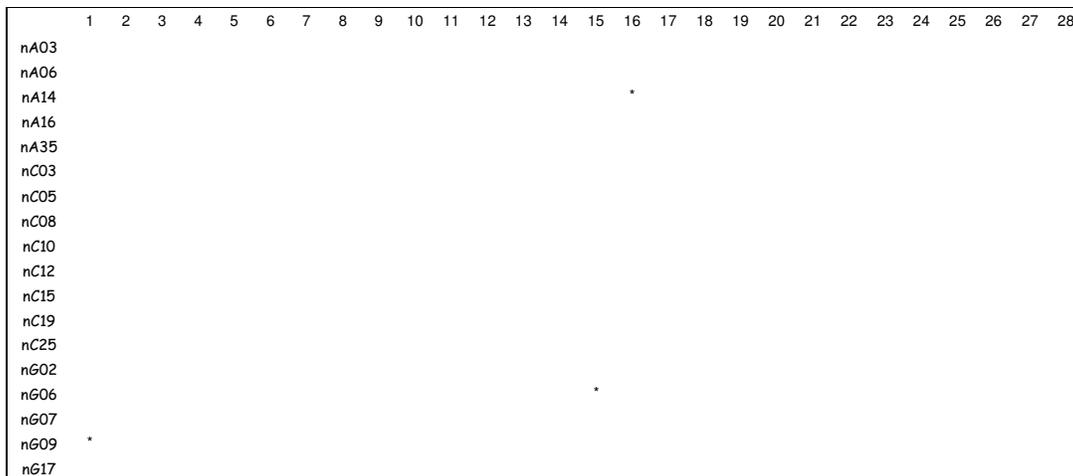


Figura IV.9. Posiciones aminoacídicas que se encuentran bajo selección positiva en la región NS5A de los individuos de genotipo 1b. En morado se ha señalado las posiciones correspondientes a la región ISDR.



Figura IV.9. Continuación

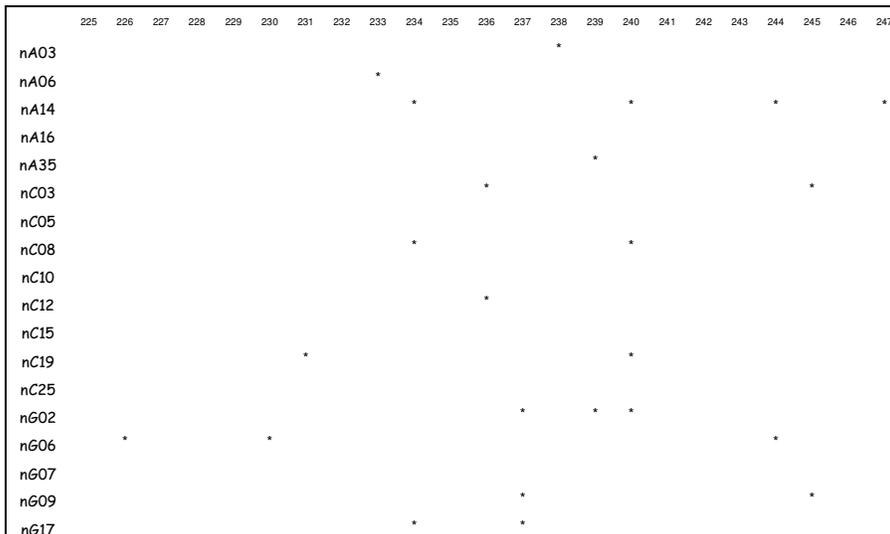
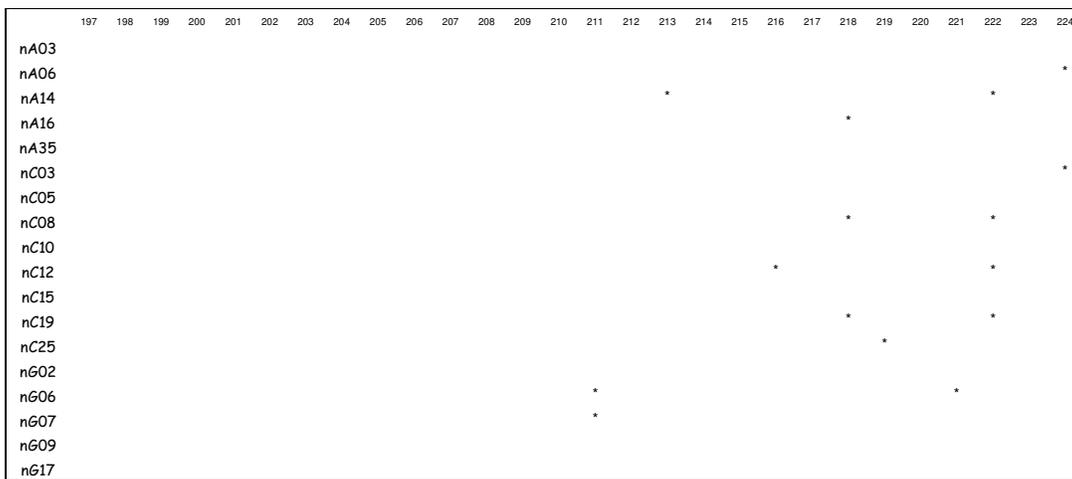
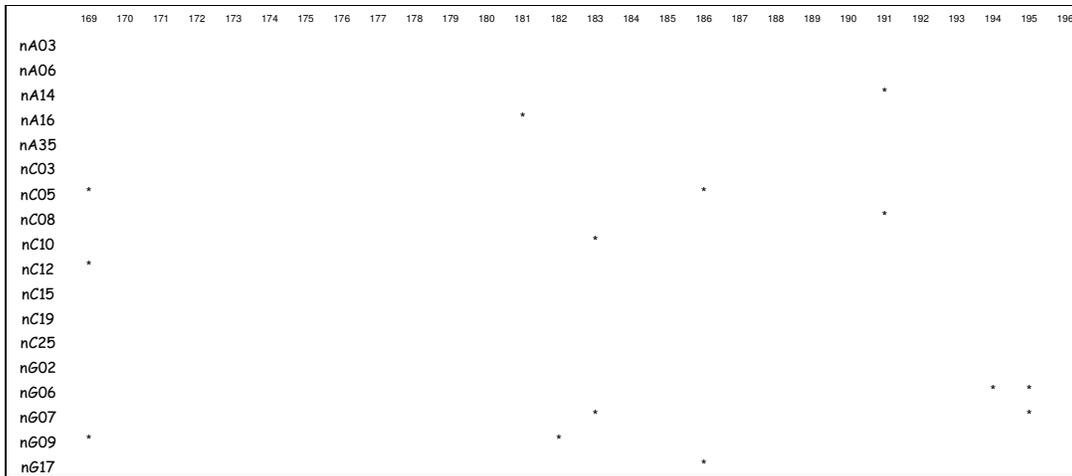


Figura IV.9. Continuación.

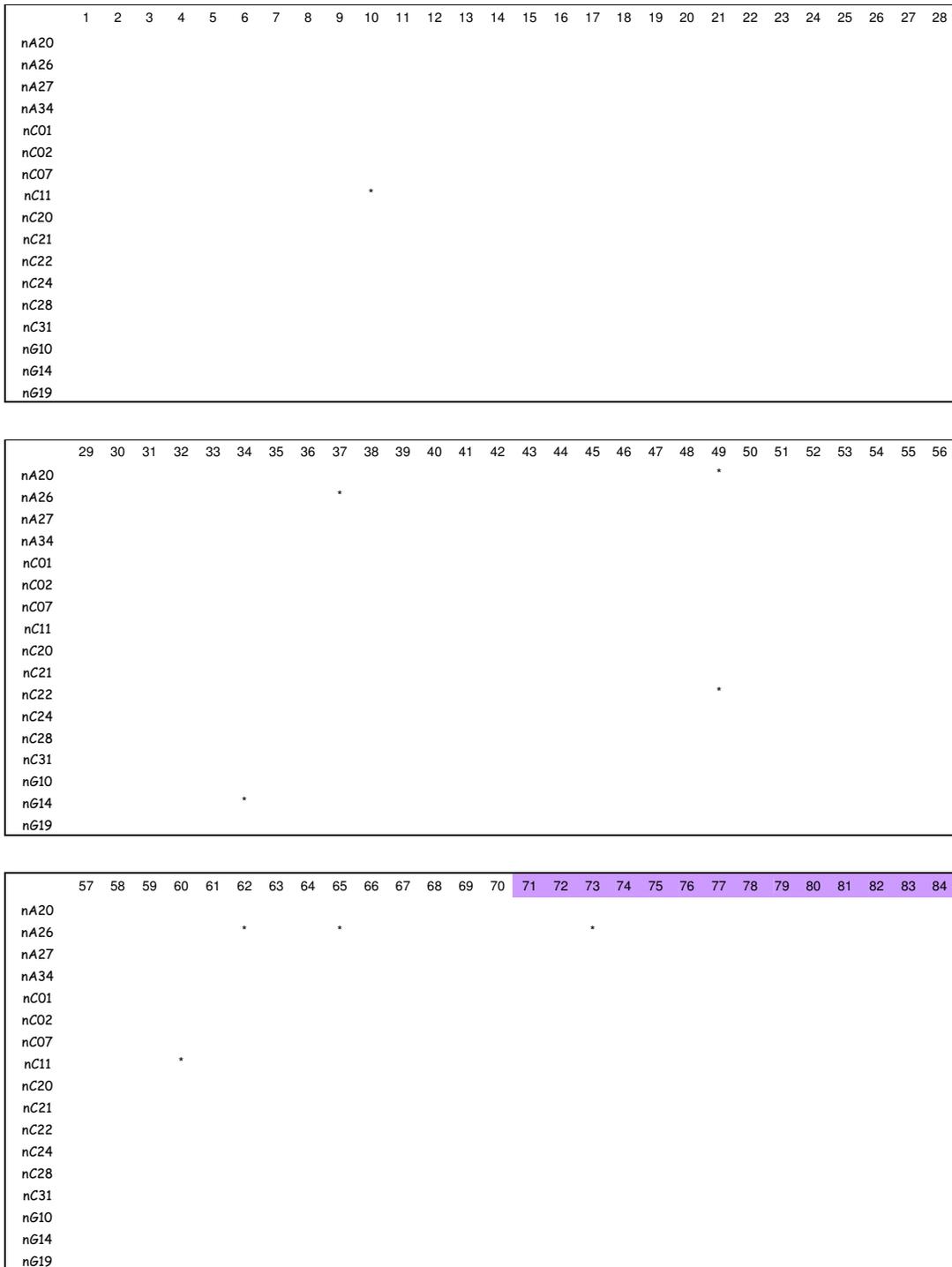


Figura IV.10. Posiciones aminoacídicas que se encuentran bajo selección positiva en la región NS5A de los individuos de genotipo 1a. En morado se han señalado las posiciones correspondientes a la región ISDR.

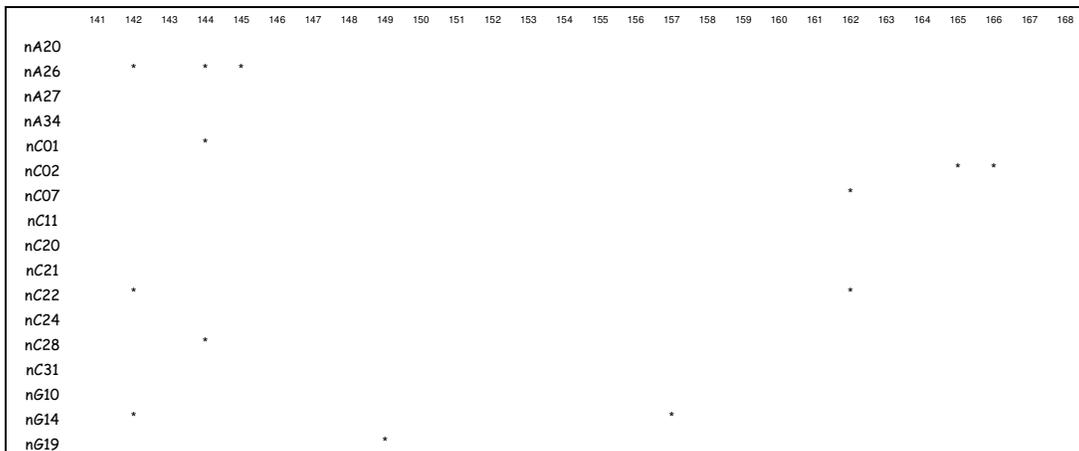
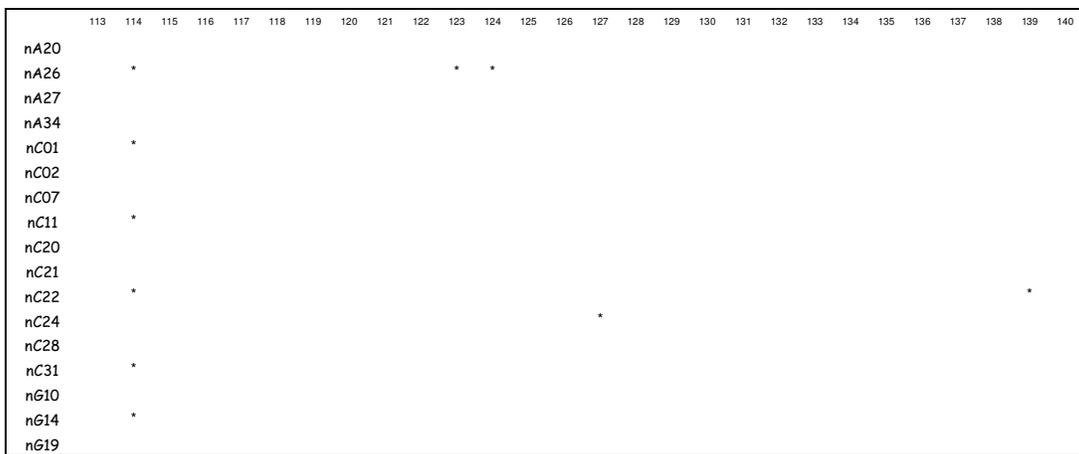
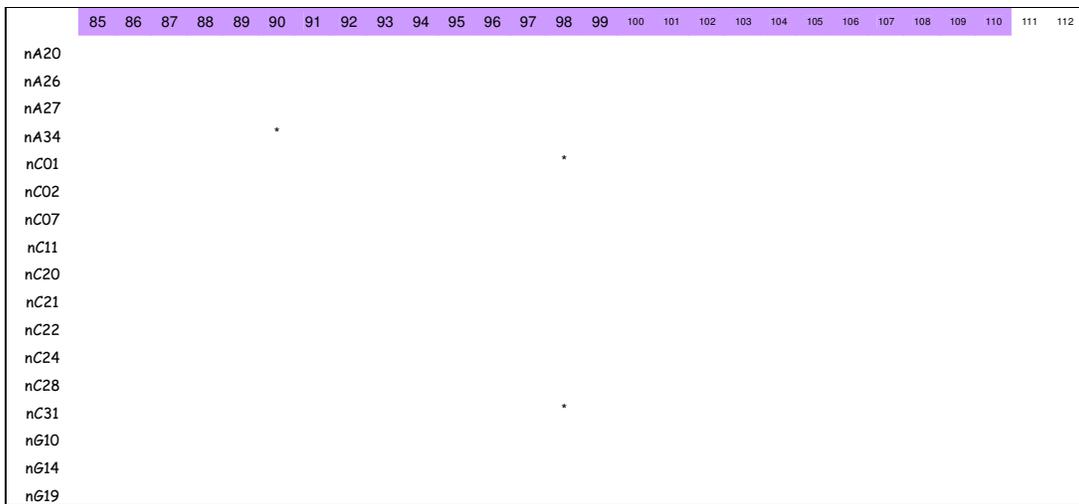


Figura IV.10. Continuación.

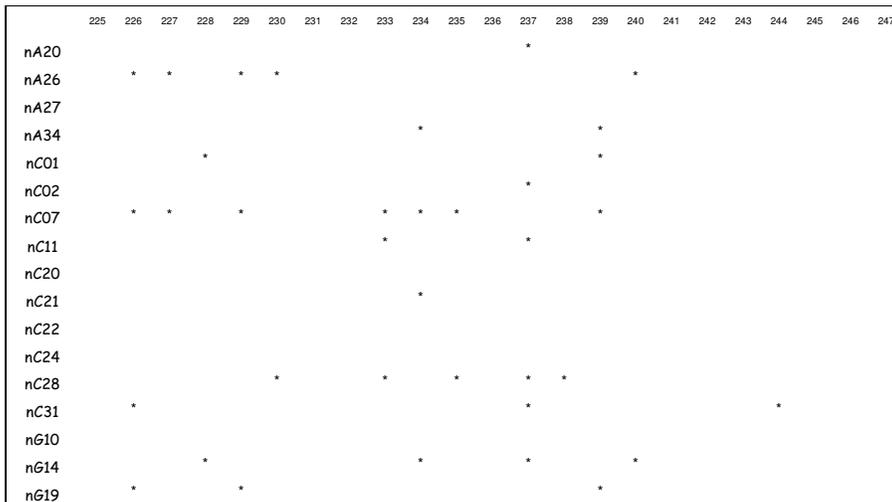
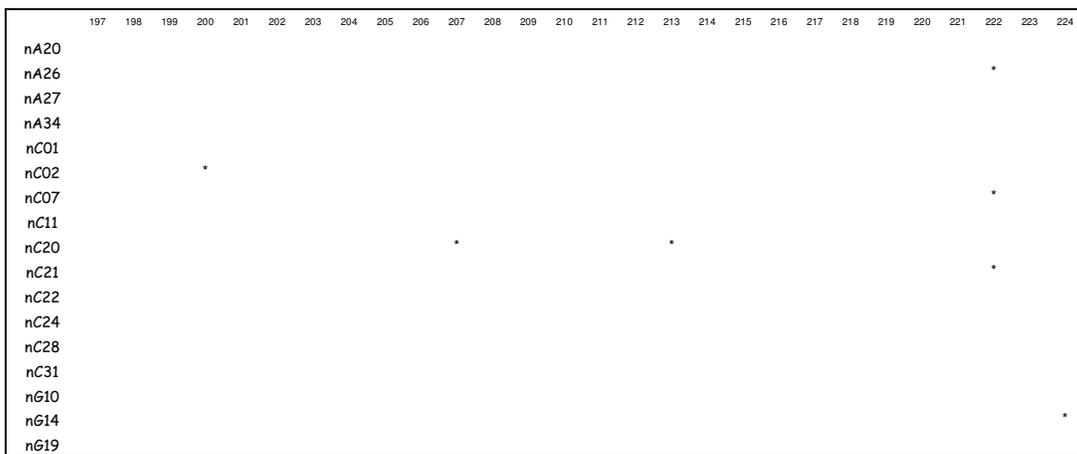
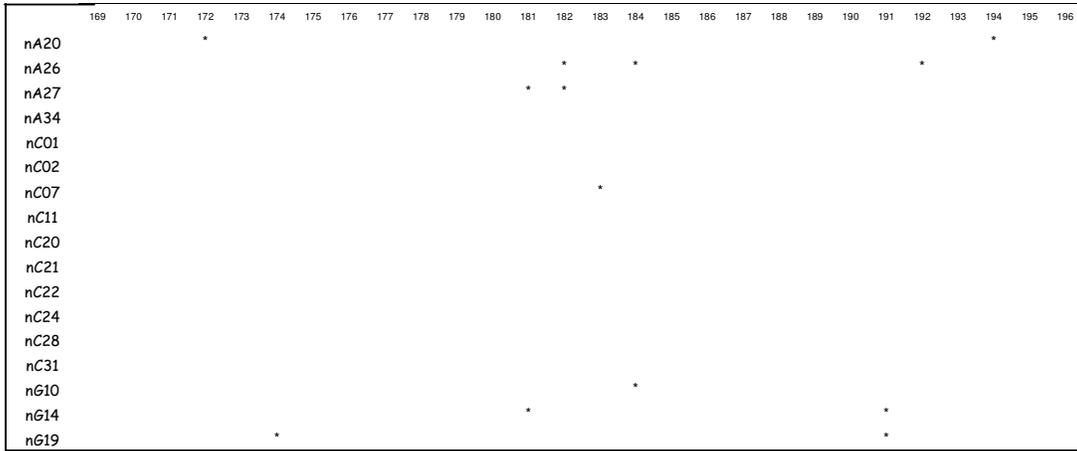


Figura IV.10. Continuación.

6. CÁLCULO DEL TAMAÑO POBLACIONAL EFECTIVO.

El tamaño poblacional efectivo (N_e) es uno de los parámetros clave en la genética de poblaciones, y al igual que otras medidas de variación genética dentro de poblaciones también es función de la tasa de mutación, flujo génico y tamaño poblacional (Ohta y Kimura 1971). N_e se define como una población ideal extraída de manera aleatoria de la población total, y posee las mismas características estadísticas que ésta última. En nuestro caso, puesto que la población es viral, podría ser descrito como el número medio de variantes virales que producen descendencia infecciosa (Nijhuis *et al.* 1998). Este parámetro proporciona una medida del efecto de la deriva genética sobre las poblaciones. La deriva genética es un importante determinante de la estabilidad de las frecuencias alélicas en las poblaciones a través de las generaciones, de la tasa de diferenciación entre poblaciones y de la diversidad genética dentro de las poblaciones (Lehman *et al.* 1998). Normalmente, las desviaciones de las condiciones ideales suele implicar que en la mayoría de ocasiones, N_e sea sustancialmente menor que el censo de individuos reproductores.

N_e se puede estimar a partir del parámetro θ (medida de la variación genética a nivel de DNA) ya que $\theta = 2N_e\mu$. Las estimas de N_e las hemos obtenido a partir de este parámetro mediante el programa GENIE, el cual a su vez, también tiene en cuenta las reconstrucciones filogenéticas. Este programa estimaba el cambio de N_e con respecto al tiempo. Los datos que se observan en la tabla IV.37 son la media de esos valores de N_e para cada uno de los individuos.

Las estimas de N_e ofrecen valores bastante bajos con respecto al tamaño censal real, oscilan entre 30 y 900. Aunque desconocemos la carga viral para cada individuo, en otros estudios se ha encontrado que este dato ronda entre 10^{10} y 10^{12} (Davis 2002). Este hecho (N_e bajos con respecto al tamaño censal) también se ha observado en otros virus de RNA como el VIH (Leigh Brown *et al.* 1997). No se observan diferencias en ningún orden de magnitud entre ambos genotipos para ninguna de las dos regiones, por lo que el tamaño efectivo no constituye una diferencia entre ambos. Existen diferentes factores que podrían explicar el por qué de estas bajas estimas, los cuales discutiremos más adelante.

Tabla IV.37. Tamaños poblacionales efectivos (N_e) para cada uno de los individuos de ambos genotipos y para ambas regiones. Tan sólo se muestran los individuos que se utilizaron para hallar la historia demográfica

E1b	N_e	E1a	N_e	N1b	N_e	N1a	N_e
eA03	286,275	eA09	86,882	nA03	120,056	nA09	72,57
eA14	310,431	eA20	167,349	nA14	171,424	nA20	74,145
eA16	963,659	eA26	386,001	nA16	41,412	nA26	106,381
eA35	88,009	eA27	73,825	nA35	27,696	nA27	75,653
eC05	603,287	eA34	360,875	nC05	157,695	nA34	140,766
eC08	424,929	eC01	149,909	nC08	87,144	nC01	142,271
eC10	161,357	eC07	203,034	nC10	46,58	nC07	129,977
eC12	287,839	eC20	184,615	nC12	180,916	nC20	26,777
eC13	124,301	eC21	113,207	nC13	197,3	nC21	104,08
eC15	556,788	eC22	459,059	nC15	73,365	nC22	284,194
eC16	154,503	eC24	37,451	nC16	237,754	nC24	26,508
eC19	574,073	eC28	292,975	nC19	219,831	nC28	165,565
eG02	232,366	eG14	97,37	nG02	185,543	nG14	5,857
eG05	93,7178	eG19	107,983	nG05	106,677	nG19	212,024
eG06	561,986			nG06	321,453		
eG07	270,781			nG07	189,601		
eG09	484,082			nG09	96,002		
eG17	306,778			nG17	140,984		

E1b, E1a: Región E1-E2 para los individuos de genotipo 1b y 1a respectivamente.

N1b, N1a: Región NS5A para los individuos de genotipo 1b y 1a respectivamente.

V. Discusión

1. INDICADORES DE VARIACIÓN GENÉTICA

La alta variabilidad genética presente en el VHC ha llevado a identificar hasta 6 genotipos distintos, nombrados del 1 al 6, los cuales difieren entre sí entre un 30 y un 50%. Cada genotipo a su vez se subdivide en multitud de subtipos (más de 50) nombrados alfabéticamente por orden de descubrimiento (Simmonds *et al.* 1994).

Diversos estudios sugieren que los genotipos y subtipos presentan un patrón epidemiológico muy variado en lo que respecta a distribución geográfica, prevalencia, rutas de transmisión, respuesta a tratamiento y curso de la enfermedad. El entendimiento de los factores que promueven estas diferencias supone un avance en el desarrollo de estrategias para acabar con la enfermedad.

Lo que aquí nos ocupa es dar con una explicación a por qué el genotipo 1b es más prevalente que el genotipo 1a en la Comunidad Valenciana (Zúñiga *et al.* López-Labrador *et al.* 1997). Para ello hemos analizado diferentes aspectos, comenzando por algo tan sencillo como el estudio de diferentes medidas de variación genética tales como el número de sitios segregantes, el número medio de diferencias nucleotídicas emparejadas, la diversidad nucleotídica, la diversidad haplotípica, y el número total de mutaciones. Cuando realizamos la comparación de las medias de los parámetros anteriores (Tabla IV.1 del apartado de resultados) no encontramos diferencias significativas entre ambos genotipos con respecto a la región E1-E2. No parece, por tanto, que la variación genética en esta región esté relacionada con la mayor prevalencia de un genotipo sobre otro. No obstante, veíamos que sí existían diferencias entre las medias de los diferentes parámetros en ambos genotipos para la región NS5A. A la vista de ello podemos hacernos algunas preguntas: ¿Están estas diferencias en la variación de la región NS5A, relacionadas con la diferencia en la prevalencia de los genotipos? ¿A qué se deben tales diferencias en la variación genética? ¿Qué fuerzas modelan esta variación? La respuesta a estas preguntas podrían dárnosla los análisis que se discuten a continuación.

2. LA DEMOGRAFÍA DE LAS POBLACIONES INFERIDA POR LOS GENES

Uno de los factores que parece influir en la diferente distribución geográfica y prevalencia de los genotipos es la ruta por la cual se transmite la enfermedad (Stumpf y Pybus 2002). La utilización de métodos basados en la teoría de la coalescencia puede ayudarnos a reconstruir la historia epidémica de diferentes secuencias del VHC, a partir de la diversidad genética viral presente (Pybus *et al.* 2001). Uno de estos métodos es el que hemos utilizado aquí para inferir la historia demográfica del VHC para los genotipos 1b y 1a. Este análisis se llevó a cabo a dos escalas: individual y genotípica.

2.1. El problema de la historia demográfica a escala individual

Como ya se comentó en la introducción, el VHC presenta una gran variabilidad, pudiendo encontrar dentro de un mismo paciente un número muy elevado de variantes genéticas tantas, muchas veces, como el número de secuencias analizadas. Tal variabilidad podría ser responsable, en primera instancia, de que la dinámica poblacional que encontramos dentro de pacientes no siga un patrón que nos permita establecer diferencias entre genotipos. De hecho, los virus dentro de cada paciente presentan una dinámica poblacional que difiere de la encontrada en otros individuos de su mismo genotipo. Por ello no podemos asociar una dinámica específica por genotipo. Stumpf y Pybus (2002) indican que tales niveles de variabilidad podrían depender de factores como el tejido del que se ha tomado la muestra, el tratamiento contra el VHC, el estadio de la enfermedad y la respuesta inmune-celular. Puesto que todas las muestras utilizadas en nuestro estudio proceden de sangre, y están tomadas antes de someterse a tratamiento con interferón y ribavirina, nos centraremos en los dos últimos puntos.

Atendiendo al estadio de la enfermedad, debido a que la hepatitis C es una enfermedad asintomática, es difícil saber la fecha de contagio de la misma. La variación genética del virus puede depender del tiempo que éste lleve en el interior del paciente, por tanto, algunas de las diferencias encontradas en las dinámicas poblacionales pueden deberse a esta causa.

Con respecto al factor referente a la respuesta inmune-celular, se ha observado en estudios donde se examinaba la respuesta humoral e inmunológica frente al VHC,

que la dinámica del virus difería mucho entre pacientes (Manzini *et al.* 1998; Herrmann *et al.* 2000; Lawal *et al.* 1997; Ni *et al.* 1999), bien por el propio estado del sistema inmune del enfermo, o bien por características propias del individuo, tales como edad, estilo de vida etc. Los datos disponibles en nuestro caso no nos permiten contrastar estos dos conjuntos de factores y asumimos que deben afectar a la variación genética detectada.

Así pues, parece que la dinámica poblacional del VHC, no depende totalmente de las características propias del virus, si no que está también bastante influenciando por las características propias de cada individuo, pudiendo ser esta otra de las causas por las que encontramos diferentes historias demográficas en los pacientes, imposibilitando establecer diferencias entre genotipos.

2.2. Historia demográfica a nivel de genotipo

Aunque en algunos trabajos se han encontrado algunas diferencias entre los genotipos 1a y 1b, que podrían explicar la mayor prevalencia de uno sobre otro, todavía no hay nada concluyente. Una de las hipótesis que se maneja es la diferencia en la patogenicidad de los mismos. Zein *et al.* (1996) muestran que el genotipo 1b produce una hepatitis más agresiva que el 1a, concretamente el primero conduce en mayor número de casos a importantes lesiones de hígado o incluso al desarrollo de un hepatocarcinoma. Sin embargo otros estudios no han encontrado tal asociación, llevando a sugerir que la diferencia en el progreso de la enfermedad no se debe a la mayor o menor patogenicidad de un genotipo, sino a que, puesto que el genotipo 1b es más frecuente en personas mayores, la mayor severidad de su enfermedad se debería más bien a que el virus lleva circulando por el hospedador un período de tiempo mayor (Zein *et al.* 2000; López-Labrador *et al.* 1997).

Otra de las hipótesis que podría explicar la prevalencia de un genotipo sobre otro es su ruta de transmisión. A lo largo de la historia se ha observado que la diferente distribución geográfica de los distintos genotipos ha estado ligada a una determinada ruta de transmisión. En ciertas regiones se ha encontrado un paralelismo entre la emergencia del virus y el tratamiento empleado en combatir diversas enfermedades, donde se utilizaban sueros humanos o glicerinados de linfa humana. (Smith *et al.* 1997; Mortimer 1995). También se ha encontrado una cierta correspondencia entre grupos de

personas con determinadas prácticas de riesgo, y un genotipo determinado Webster *et al.* 2000; Zein 2000). Este tipo de estudios ha permitido establecer una relación entre una ruta de transmisión determinada y un genotipo. De hecho, parece que los genotipos 1a y 3a están asociados a jóvenes consumidores de drogas vía intravenosa, mientras que el genotipo 1b y el 2 están asociados a pacientes que alguna vez han sido sometidos a una transfusión de sangre (Pawlotsky *et al.* 1995; Simmonds *et al.* 1996; Pol *et al.* 1995). De ser cierta esta hipótesis, el hecho de que el comportamiento epidemiológico del virus estuviera determinado por las rutas de transmisión más que por las diferencias propias de cada genotipo, llevaría a que en un futuro más o menos próximo podrían producirse cambios en su prevalencia, es decir que cualquiera de los genotipos que hoy presentan menor prevalencia puedan ser responsables de futuras epidemias (Kalinina *et al.* 2001; Chuang *et al.* 2001)

Las figuras IV.2 y IV.3 del apartado de resultados parecen dar apoyo a esta última hipótesis. Podemos observar cómo el genotipo 1b aparece antes en la Comunidad Valenciana que el 1a. Durante años el genotipo 1b presenta una tasa de crecimiento mucho mayor que la del 1a. Sin embargo, en los últimos 20 años aproximadamente se observa un cambio radical en el comportamiento del virus. Mientras el crecimiento para el genotipo 1b se estabiliza, en el genotipo 1a ha aumentado considerablemente. Tal como se muestra en la figura IV.3 del apartado de resultados, hasta hace aproximadamente unos 60 años (según esta reconstrucción) el genotipo 1b era el más prevalente con bastante diferencia sobre el 1a, y por tanto, la mayoría de individuos afectados estarían infectados por este genotipo, por lo que en los años siguientes lo más probable es que todos los individuos contagiados *de novo* lo fueran con genotipo 1b, hasta que poco a poco fue aumentando el número de infectados por 1a, coincidiendo con un supuesto aumento en el número de consumidores de drogas vía intravenosa (Pawlotsky *et al.* 1995, Webster *et al.* 2000). Esto explicaría que actualmente un gran porcentaje de individuos contagiados con 1b presenta edades entre 40 y 60 años (Webster *et al.* 2000) También hay que tener en cuenta el hecho de que alrededor de los años 80, aunque aún se desconocía la existencia del VHC, se sabía de la existencia de la hepatitis no-A no-B, y comenzaron a ponerse medios para hacer *screening* de sangre contra este tipo viral (Memon y Memon 2002), como por ejemplo, desestimar a donantes de sangre con alto contenido en ALT (alanina aminotransferasas) en sangre. Estas medidas bajaron considerablemente la incidencia de la hepatitis no-A no-B hasta su descubrimiento en 1989 (Choo *et al.* 1989), momento a partir del cual se pudo hacer

un *screening* específico contra el VHC. Esto también parece coincidir en el tiempo con los resultados presentados en nuestra figura, ya que podemos observar cómo la curva de crecimiento del genotipo 1a aumenta a medida que se estabiliza la de 1b.

Con respecto a los resultados obtenidos en el cálculo de R_0 , fijándonos sólo en la región E1-E2, los valores apoyarían lo encontrado en los *skylineplots*, ya que R_0 era mayor para los 1a que para los 1b, indicando que se esperaría un mayor riesgo de infección a partir del genotipo 1a. No obstante si nos fijamos en el caso de la región NS5A, los resultados se invierten. Ahora es el genotipo 1b el que presenta un mayor valor para R_0 , y esto no coincide con lo encontrado en los *skylineplots*, ya que en este caso los individuos 1b presentaban una estabilización en su crecimiento. Una posible explicación a esta disparidad sería la diferencia en la tasa de cambio entre las dos regiones. Como ya comentamos en el apartado de resultados, la región NS5A presenta una tasa de evolución menor que la región E1-E2. Por tanto la región NS5A podría encontrarse en una fase por la que ya ha pasado la E1-E2 (a lo largo de la misma historia demográfica), y esta situación podría estar alterando los resultados de R_0 .

De todas formas si tenemos en cuenta las observaciones de los trabajos mencionados anteriormente, (Smith *et al.* 1997; Mortimer 1995; Pawlotsky *et al.* 1995; Simmonds *et al.* 1996; Pol *et al.* 1995; Kalinina *et al.* 2001; Chuang *et al.* 2001) junto con lo encontrado en nuestros análisis con GENIE, podríamos apoyar la hipótesis de que la mayor prevalencia de un genotipo sobre otro se deba a las diferentes rutas de transmisión. Ya que por un lado, las vías de contagio muestran una clara relación entre la ruta de transmisión y un determinado genotipo, en este caso el 1b con transfusiones y el 1a con ADVP. Por otro lado la fecha en que el genotipo 1b comienza a disminuir su tasa de crecimiento parece coincidir con el aumento en 1a, y a su vez coincide con la primera introducción de métodos de *screening* de sangre contra no-A no-B en transfusiones. Por tanto, parece que en los últimos años el contagio vía transfusiones de sangre, debido a las fuertes medidas de control sobre los productos sanguíneos, está casi totalmente controlado, por lo que la infección del virus (genotipo 1b) por esta ruta es muy difícil, dejando vía libre al contagio por ADVP (genotipo 1a). Los resultados obtenidos con las AMOVAs también apoyarían esta suposición. Al analizar las categorías de edad, diferenciación intra e intergenotípica, sexo de los pacientes, prácticas de riesgo y número de tratamientos a los que ha sido sometido, sólo en la categoría de prácticas de riesgo encontramos que las diferencias entre los virus de los

diferentes grupos podrían ser lo suficientemente grandes como para contemplar la posibilidad de que la infección por un genotipo u otro pueda estar influenciada por las prácticas de riesgo de los pacientes.

El hecho de que actualmente el genotipo 1b sea más prevalente en la Comunidad Valenciana, posiblemente se deba a las huellas del pasado, y cabría la posibilidad de esperar que en los próximos años pudiera producirse un cambio de prevalencia como se ha observado en otros países (Kalinina *et al.* 2001; Chuang *et al.* 2001).

3. LAS FUERZAS QUE MODULAN LA EVOLUCIÓN DEL VIRUS DE LA HEPATITIS C

Es largo el debate sobre la fuerza dominante en evolución molecular: selección natural positiva o deriva genética (Kimura 1983; Gillespie 1991). Recientemente el debate seleccionismo/neutralismo apunta a casos, en número creciente, de presencia de selección positiva. La teoría neutral sigue siendo la mejor hipótesis nula para verificar o contrastar la presencia de la selección natural. La literatura genético-poblacional está plagada de este tipo de pruebas, algunas de los cuales hemos aplicado. Tienen por objeto rechazar o no la hipótesis nula de neutralidad evolutiva.

Debido a que son varios los factores o fuerzas naturales que pueden jugar un rol importante en la evolución de una población, es poco probable que una sola prueba estadística pueda detectar todas las posibles fuerzas evolutivas que puedan estar afectando el patrón de polimorfismo. Hay diferentes estudios que muestran que unas pruebas son más apropiadas que otras para detectar determinados factores evolutivos que otros (Simonsen *et al.* 1995; Fu 1997; Ramos-Onsins y Rozas 2002).

Como ya hemos visto en el apartado 4 de resultados, es complicado interpretar los valores obtenidos en las pruebas de neutralidad. Para simplificarlo, y puesto que no existían grandes diferencias entre los individuos pertenecientes a los genotipos 1a y 1b, con respecto a los valores de los estadísticos de las diferentes pruebas, trataremos los datos en conjunto. Es decir, hablaremos tan solo de lo que ocurre en la región E1-E2 y en la NS5A, sin hacer mención al genotipo al que pertenecen.

En la región E1-E2 se nos presentan tres situaciones: 1) Individuos en los que no rechazamos la hipótesis nula, ya que los estadísticos no son significativamente diferentes de 0 (cinco casos para el genotipo 1b, y tres para el 1a). 2) Individuos en los que al menos uno de los estadísticos presentan valores significativamente menores que 0 (21 casos para 1a, y 20 para 1b). 3) Por último, y en muy pocos casos, individuos en los que el valor de los estadísticos es mayor que 0. Empezaremos con la situación que más se repite. Son los casos donde encontramos desviaciones del modelo neutral, y por tanto parece que hay indicios de que podría estar actuando la selección natural. Pero qué clase de selección ¿estabilizadora o direccional? Si se tratase de selección estabilizadora, debido a que se mantienen varios alelos en la población, esto conduciría a un exceso de sitios polimórficos que segregan a frecuencias intermedias y por tanto encontraríamos valores positivos para la D de Tajima y la D^* y F^* de Fu y Li. Sin embargo la selección direccional, la cual favorece la fijación de mutaciones ventajosas o eliminación de las deletéreas y produce un exceso de mutaciones únicas y variantes con frecuencias muy bajas que conduce a valores negativos de la D de Tajima y la D^* y F^* de Fu y Li. Atendiendo a esto está claro que, puesto que los valores encontrados para los datos analizados poseen valores negativos, es la selección direccional la responsable de la variación encontrada en las secuencias analizadas. Pero aquí surge otra cuestión, la selección direccional ¿se da por la presencia de mutaciones ventajosas (arrastre selectivo) o por mutaciones deletéreas (selección de fondo)? Los valores de los estadísticos anteriores no nos dan más información para resolver esta pregunta. No obstante sí sabemos, por diferentes trabajos (Simonsen *et al.* 1995, Fu 1997; Ramos-Onsins y Rozas 2002), que la D de Tajima y la F_s de Fu, son más sensibles a la presencia de arrastre selectivo, mientras que la D^* y F^* de Fu y Li lo son para selección de fondo. Así, cuando sólo D y F_s sean significativos estaremos ante un caso de arrastre selectivo, mientras que si sólo lo son D^* y F^* estaremos ante selección de fondo. Según estas indicaciones encontramos selección de fondo en los individuos eC05, eC08 y eC10 para el genotipo 1b, y en eA20, eA34 y eC18 para 1a. En el resto encontramos arrastre selectivo. No obstante, quedan cuestiones por resolver, como: ¿qué pasa en los casos en que todos los estadísticos son significativos? Así pues parece importante para resolver todas las dudas tener en cuenta la región genómica que estamos analizando. Como ya hemos dicho en varias ocasiones, la región E1-E2 codifica para las proteínas de la envuelta y para la región hipervariable 1 (HVR1) podría ser uno de los determinantes hacia donde van dirigidos algunos anticuerpos neutralizantes. Se ha

sugerido que la gran variabilidad de esta región permitiría al virus escapar a esta respuesta inmune (Curran *et al.* 2002). Según esto, todo parece apuntar a que lo más probable es que en la mayoría de individuos se estén dando fenómenos de selección positiva como el arrastre selectivo. Puesto que esta región es diana de anticuerpos por parte del sistema inmune, resultaría beneficioso para el virus fijar mutaciones que le permitieran escapar del ataque del sistema inmune.

Con respecto a los individuos en los que no se rechaza la hipótesis nula de neutralidad, no resultaría muy beneficioso para el virus no cambiar su diana frente a los anticuerpos, por lo que la neutralidad es algo que no encaja bien en estos resultados. El hecho de que los valores de sus estadísticos no sean significativos podría, según Simonsen *et al.* (1995), en el caso de la D de Tajima, deberse además de a la neutralidad, a un arrastre selectivo más débil y más lejano en el tiempo. Como veíamos en la explicación de las pruebas de neutralidad, el valor de D depende de Π (número de diferencias nucleotídicas entre dos secuencias escogidas al azar en la población) y S (número de sitios que presentan variación entre las secuencias de la muestra). En presencia de arrastre selectivo aumenta el número de variantes que segregan a frecuencia única por lo que el valor de Π (que se ve afectado por la frecuencia de las mutaciones) será menor que el de S y la D de Tajima presentará valores negativos. Pero con el paso del tiempo pueden ir apareciendo nuevas mutaciones e ir aumentando el número de variantes que segregan a frecuencias intermedias aumentando así el valor de Π , por lo que los valores de D tenderán a acercarse a 0.

Por tanto, los individuos que presentan valores no significativos de D , apuntan hacia la posible actuación de arrastre selectivo antiguo. Si nos fijamos, además, en el espectro de frecuencias de variantes en estos individuos (Figura V.1), observamos que aunque algunas variantes se encuentran a frecuencias intermedias la mayoría segregan a frecuencias únicas, correspondiendo a la huella dejada por el arrastre selectivo.

Viendo los resultados anteriores podemos concretar que en la mayoría de individuos estamos ante casos en los que la selección positiva podría estar jugando un papel significativo.

En los individuos en que obteníamos valores de D mayores que 0 también asumimos que han sido sometidos a arrastre selectivo, puesto que estos valores no eran significativos, su espectro de frecuencias también mostraba exceso de variantes de frecuencia única (ANEXO III), y sólo los valores de F_s eran significativos. Sólo existen tres casos (eC18, eC21, eC22) en los que teniendo valores de D , D^* y F^* mayores que 0

no significativos, su espectro de frecuencias ofrece variantes que segregan a frecuencias intermedias. Estaríamos quizás ante ejemplos de selección estabilizadora, pero no lo podemos asegurar puesto que los valores no son significativos.

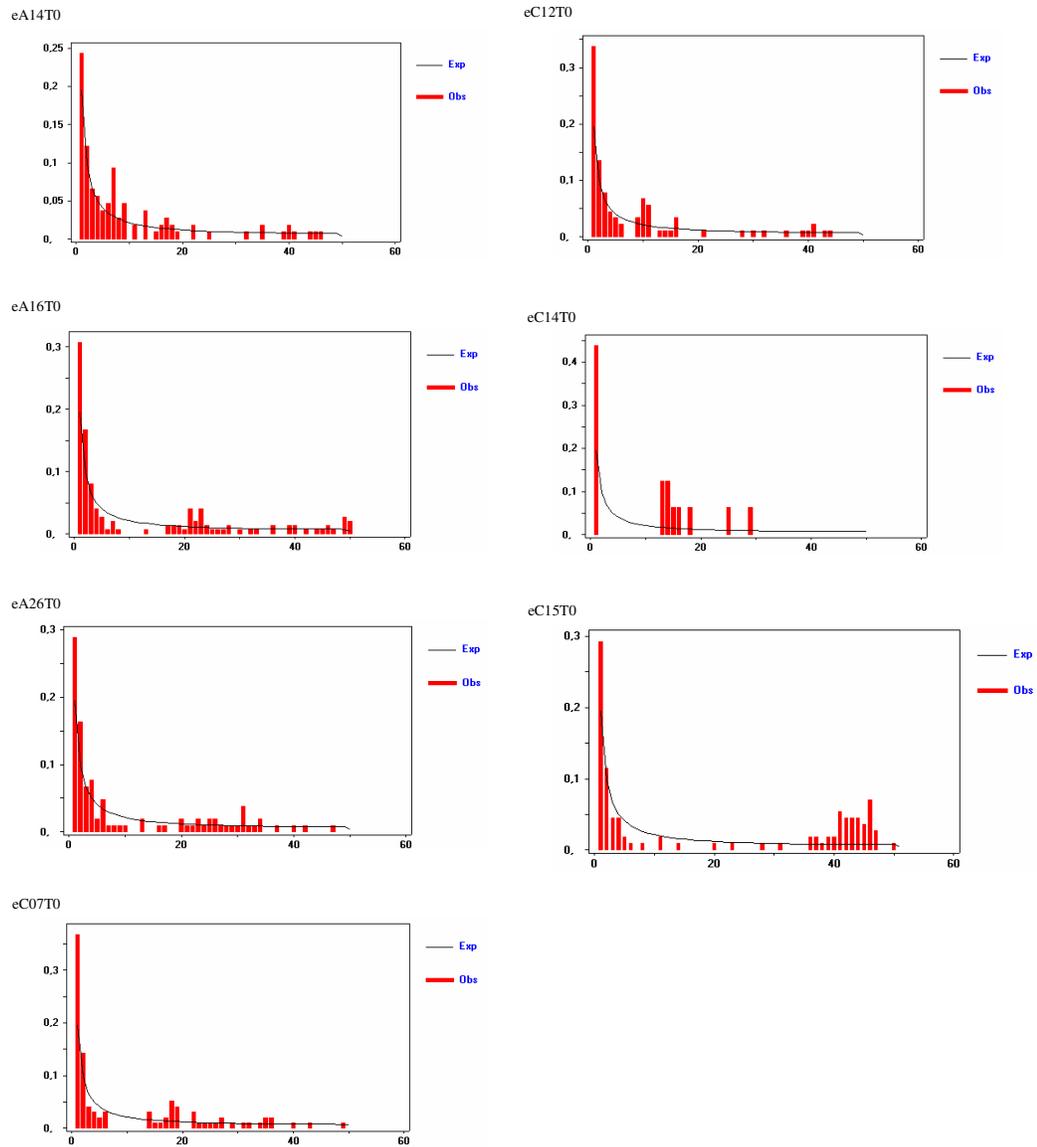


Figura V.1. Espectro de frecuencias de variantes para los individuos que presentan valores de D no significativos. En el eje de ordenadas se representa el porcentaje de sitios polimórficos, y en el de coordenadas el número de mutaciones por sitio polimórfico. Exp: Curva esperada bajo neutralidad. Obs: Curva observada

De todos modos, puesto que son sólo tres casos sobre cincuenta, seguimos pensando que en esta región se dan fenómenos de selección positiva que podría conferir al virus alguna ventaja frente al ataque de anticuerpos neutralizantes por parte del sistema inmune. Otro hecho que apoyaría esta hipótesis es que, como se ha mostrado en el apartado anterior, el tamaño poblacional efectivo es bastante bajo. Ante fenómenos de arrastre selectivo se da también este fenómeno ya que sólo los individuos que poseen la mutación ventajosa serán los ancestros de la muestra.

Con respecto a la región NS5A, salvo en dos casos para el genotipo 1a y uno en el genotipo 1b, al menos uno de los tres estadísticos (D , D^* y F^*) es significativamente menor que 0. Por tanto volvemos a encontrarnos con la posibilidad de estar ante fenómenos de arrastre selectivo o selección de fondo. Así pues, las propiedades de la proteína para la que codifica el gen al que corresponde nuestra región, de nuevo, pueden ayudarnos a encontrar la solución. Diversos trabajos (Chayama *et al.* 1997; Enomoto *et al.* 1995; 1996; Kurosaki *et al.* 1997; Saiz *et al.* 1998) hacen notar que la NS5A, y en concreto la región ISDR interacciona con la PKR inhibiendo su actividad (inhibición de la replicación viral), ofreciendo de esta forma resistencia a interferón. Además Polyak *et al.* (2001) y Gale *et al.* (1998) también han observado que la acumulación de mutaciones en la ISDR impide la unión a la PKR produciéndose la inactivación de la replicación viral.

Todo parece indicar, si tenemos en cuenta nuestros resultados y las observaciones anteriores, que lo que requiere el virus es acumular cuantos menos cambios mejor, eliminando aquellas mutaciones que pudieran suponerle alguna desventaja. Los valores negativos que encontramos en los estadísticos de las pruebas de neutralidad, lo que reflejan es la presencia de selección purificadora o selección de fondo.

Resumiendo, nos encontramos con desviaciones al modelo neutro en las dos regiones analizadas. En la región E1-E2 la desviación podría deberse a la actuación de la selección positiva o arrastre selectivo, confiriendo de este modo alguna ventaja frente al sistema inmune. En el caso de la región NS5A, la desviación del modelo neutral sería debido a la actuación de la selección purificadora o selección de fondo, eliminando aquellas mutaciones en la región ISDR que harían sensible al virus frente a la actividad de la PKR.

Aún queda por resolver, no obstante, otra cuestión que hasta ahora no hemos mencionado, y es que las desviaciones selectivas al modelo neutro, también pueden

deberse a factores demográficos tales como cambios en el tamaño de la población. Una forma de averiguar si las desviaciones del modelo neutral se debían a una causa u otra era ver si observábamos la misma huella en dos regiones distintas del genoma, ya que los factores demográficos suelen afectar a la totalidad del genoma, por lo que su rastro sería detectable en casi la totalidad de los genes. Los factores selectivos, por el contrario, afectan a genes particulares, por lo que su marca sólo sería detectable en estas regiones génicas.

Asimismo existen otros métodos que también nos permiten averiguar si ha actuado la selección en determinadas regiones génicas. Hemos de recordar que las pruebas de neutralidad, la idea fundamental que aportan es si las muestras se desvían o no del modelo neutral y especular sobre qué fuerzas evolutivas pueden estar provocando esto. Además, los estadísticos utilizados en estas pruebas no utilizan mucha de la información contenida en los datos, como por ejemplo cómo están relacionados los alelos unos con otros. De la misma forma, diferentes procesos evolutivos pueden generar distribuciones de frecuencias alélicas que son difíciles de diferenciar con estos estadísticos. Estas carencias pueden ser suplidas por métodos que tengan en cuenta reconstrucciones filogenéticas, y quizás puedan servirnos de más ayuda a la hora de asegurar con mayor certeza si ha actuado o no la selección. Con este fin hemos aplicado el método de máxima verosimilitud desarrollado por Yang *et al.* 2000.

3.1 Detectando e interpretando la selección positiva

Aunque la teoría de Darwin de evolución por selección natural es generalmente aceptada por los biólogos, la importancia de la selección natural en la evolución molecular ha sido objeto de debate durante bastante tiempo. La teoría neutral (Kimura 1983) mantiene que la mayor parte de la variación molecular observada se debe a la fijación al azar de mutaciones selectivamente neutras, siendo muy raros los casos de adaptación molecular. Varias han sido las pruebas de neutralidad desarrolladas; no obstante aunque son capaces de rechazar con cierto peso la hipótesis de neutralidad, raramente proporcionan una evidencia inequívoca de la presencia de la selección positiva.

La mayoría de casos donde parece demostrada la presencia de selección positiva, han sido identificados a partir de la comparación entre las tasas de sustitución sinónimas

(silenciosas dS) y no sinónimas (implican cambio aminoacídico dN). La relación $\omega=dN/dS$ mide la diferencia entre las dos tasas. Si un cambio aminoacídico es neutro, será fijado a la misma tasa que las mutaciones sinónimas con $\omega=1$. Si el cambio es deletéreo, la selección purificadora reducirá su tasa de fijación y de este modo $\omega<1$. Sólo cuando el cambio aminoacídico suponga una ventaja selectiva se fijará a una tasa de $\omega>1$. Por tanto una ω suficientemente mayor que uno es una evidencia que apoya la selección positiva. Se han desarrollado dos clases de métodos para la estima de las tasas dN y dS . La primera clase (Miyata y Yasunaga 1980; Nei y Gojobori 1986; Li *et al.* 1985, 1993; Pamilo y Bianchi 1993; Comeron 1995; Ina 1995) implica los siguientes pasos: Primero contar el número de sustituciones sinónimas, S , y no sinónimas, N , en las dos secuencias, luego contar las diferencias sinónimas y no sinónimas entre las dos secuencias y corregir para múltiples sustituciones para cada sitio. S y N son definidos como la longitud de las secuencias por la proporción de cambios sinónimos y no sinónimos antes de la selección de la proteína. La segunda clase hace referencia a los métodos de máxima verosimilitud basados en los modelos de sustitución de codones (Yang *et al.* 2000; Goldman y Yang 1994; Muse 1996; Yang 2000). Los parámetros de los modelos (divergencia de secuencias, tasa de transición/transversión, y ω son calculados por máxima verosimilitud y son utilizados para el cálculo de dN y dS . Después son utilizados tests estadísticos para verificar si dN es significativamente mayor que dS .

A la hora de decidirse por una de estas dos clases de métodos, hemos de tener en cuenta algunas consideraciones: La primera clase asume una tasa igual para transiciones y transversiones. Esto puede conducir a una infraestimación de S y una sobreestimación de N , ya que las transiciones en la tercera posición de codones tienen más probabilidad de ser sinónimas que las transiciones (Li *et al.* 1985). Otra consideración es que los primeros métodos asumen un uso de codones uniforme. Ignorar el sesgo en el uso de codones conduce a una sobreestimación de S y a una infraestimación de N (Yang y Nielsen y Yang 1998; Yang y Nielsen 2000). Por último, también asumen que todos los aminoácidos están bajo la misma presión de selección y con la misma ω . Esta tasa (ω), es calculada como una media de lo obtenido en la totalidad de codones. No obstante esta forma de cálculo puede no ser la más adecuada (Sharp 1997; Akashi 1999; Crandall *et al.* 1999), ya que en muchas proteínas una gran proporción de aminoácidos puede ser muy invariable, debido a constreñimientos funcionales ($\omega = 0$), y además la mayoría de proteínas parecen estar bajo selección purificadora la mayor parte del tiempo (Li 1997).

De esta forma es muy difícil que ω alcance valores mayores que 1, incluso aunque se esté dando selección adaptativa en algunos aminoácidos.

La segunda clase de métodos, a parte de tener en cuenta la tasa de transiciones/transversiones y el sesgo en el uso de codones, usa una distribución estadística para describir la variación de ω entre codones, y después aplica estos modelos a análisis filogenéticos de secuencias de DNA que codifican para proteínas. (Nielsen *et al.* 1994; Yang *et al.* 2000). Es decir, asume varias categorías de sitios en la proteína con diferentes ω . El test de selección positiva consistirá en lo siguiente: Primero se contrasta si existen sitios donde $\omega > 1$ mediante un test de razón de verosimilitudes, comparándolo con un modelo que no permita la presencia de estos sitios. Después se utiliza un método bayesiano para identificar los codones seleccionados positivamente cuando existan. Los codones que maximicen la probabilidad posterior para una categoría de sitios con $\omega > 1$, son potencialmente dianas de selección positiva.

Esta última clase de métodos es la que hemos utilizado en la búsqueda de selección positiva en nuestros datos (Tablas 14-17 del ANEXO VII). En este caso tampoco se observan diferencias entre genotipos tanto en la región E1-E2 como en la NS5A. Las posiciones que se encuentran bajo selección positiva coinciden tanto en los individuos de genotipo 1a como en los 1b para la región E1-E2 (Figura V.2). En la región NS5A hallamos una particularidad adicional, pues se observan posiciones bajo selección positiva pero que no parecen repetirse en el resto de individuos de un mismo genotipo, y las posiciones seleccionadas tampoco coinciden entre los dos genotipos. Es decir no encontramos un patrón como en el caso de la región E1-E2, donde los individuos de un genotipo dado tienen prácticamente las mismas posiciones seleccionadas.

Puesto que parece que la selección actúa por igual en ambos genotipos, tampoco esta fuerza evolutiva puede explicar la mayor prevalencia de un genotipo sobre otro. Así pues, también la discusión de los resultados en este apartado no hará distinción entre genotipos, sino que trataremos los datos en conjunto separados por regiones.

Comencemos por la región E1-E2. Al estudiar los resultados nos damos cuenta que la razón dN/dS (promediada en la longitud del fragmento) es menor que 0 en todos los caso excepto en los individuos eC04, eG16, eG26 para el genotipo 1b y en eC14, eC21, y eC33 para 1a. Esta situación nos indica que la mayoría de las posiciones

aminoacídicas de nuestro fragmento se encuentran bajo selección purificadora. No obstante, al aplicar el método de Yang, que admite la presencia de una distribución de ω heterogénea entre sitios, y realizar la prueba de razón de verosimilitudes, o aplicar AIC, entre los diferentes modelos que se proponen, observamos que en la mayoría de casos aparece un porcentaje de sitios que podrían estar sometidos a selección positiva. Debido a que son los modelos de selección positiva los que proporcionan una mejor explicación para la evolución de estas regiones, hemos procedido a la identificación de los sitios seleccionados positivamente, mediante inferencia bayesiana (Nielsen y Yang 1998). La probabilidad posterior proporciona una medida de la precisión de la inferencia. En el ANEXO VII, se muestran las distribuciones de probabilidad, a lo largo de la región E1-E2 para los diferentes individuos.

Al realizar alineamientos de las secuencias de los individuos de un mismo genotipo (Figura V.2) detectamos que hay una serie de posiciones (57, 59, 64, 65, 67, 68, 69, 70, 71, 72, 73, 74, 75, 77, 78, 80, 81, 83, 117, 119, 148) seleccionadas positivamente que coinciden en un gran número de individuos y que además se corresponden en ambos genotipos. Estas posiciones además recaen en la HVR1 situada en la proteína E2. A su vez también encontramos la presencia de algunos aminoácidos en la HVR1 que están bastante conservadas, las cuales podrían ser importantes para el mantenimiento de la estructura de la proteína.

Como ya se ha comentado, diversos estudios mencionan que esta región contiene epitopos contra los que actúan anticuerpos neutralizantes del sistema inmune (Suzuki y Gojobori 2001). También existen observaciones que apuntan a que el ataque por parte del sistema inmune induce de alguna manera la aparición de mutaciones en la HVR1 (Weiner *et al.* 1992; Farci 2001; Pawlotsky *et al.* 1998) que favorece el escape de este ataque. Esta hipótesis es apoyada, además, por estudios que demuestran ausencia de mutaciones en esta región en pacientes inmunodeprimidos (Gaud *et al.* 2003; Kumar *et al.* 1994; Booth *et al.* 1998).

Nuestros resultados también podrían explicarse a partir de la hipótesis anterior. Hemos encontrado mutaciones que suelen recaer, en su mayoría, en la HVR1, y todo parece indicar que se han producido por acción de la selección positiva. Estas mutaciones de algún modo enmascararían las dianas de los anticuerpos neutralizantes, permitiendo al virus seguir adelante con la infección. La emergencia de estas mutaciones en la HVR1 es el resultado de complejas interacciones entre la heterogeneidad de las variantes virales y la respuesta del sistema inmune, el cual está

genéticamente determinado y sujeto a factores externos (Pawlotsky *et al.* 1998). Este hecho puede explicar el por qué no encontramos en algunos individuos todas las posiciones seleccionadas que citábamos anteriormente (Figuras IV.7 y IV.8 del apartado de resultados), o por qué en individuos como eA05, eC06, eG10 (genotipo 1a), eA06, eG16 (genotipo 1b) es el modelo neutral el que mejor explica su situación. Como ya se vio en el análisis intrapaciente de las historias demográficas con el programa GENIE, en cada individuo el virus lleva dinámicas diferentes debidas posiblemente a diferentes factores propios del mismo, como quizás el estado del sistema inmune en determinados momentos, la ruta de contagio etc., lo que hace casi imposible encontrar el mismo patrón en todos ellos.

En resumen pensamos que la hipótesis de que el sistema inmune induce la fijación de mutaciones ventajosas en la HVR1, para escapar a ese ataque, es la mejor explicación de los resultados obtenidos.

Los resultados son diferentes con respecto a la región NS5A. También al observar la razón dN/dS , encontramos que los valores en casi todos los casos (excepto nA06, nA14, nG16 para genotipo 1a y nC33 para 1b) son menores que uno, lo que apuntaría la presencia mayoritaria de selección purificadora a lo largo del fragmento. No obstante también encontramos algunas posiciones bajo selección positiva (véase distribución de probabilidades a posteriori en el ANEXO VII), pero en menor número que en la región anterior, y además estas posiciones no se repiten en el resto de secuencias bien pertenezcan al mismo genotipo o no. Aún no está muy claro cual es la función de la proteína NS5A, aunque se piensa que podría formar parte del complejo de la replicasa viral, pero sí se conocen algunas de sus características y propiedades que podrían justificar lo que hemos encontrado. Se ha demostrado que la NS5A es capaz de interaccionar con la proteína kinasa (PKR) e inhibir su actividad. La PKR es una enzima intracelular inducida en presencia de interferones y con propiedades antivirales (Kauffman, 2000). La PKR puede ser activada por una variedad de señales entre las que se encuentra la presencia de RNA de doble cadena. La unión del RNA a la PKR induce un cambio conformacional permitiendo la fosforilación del factor de inicio de la traducción eIF2, impidiendo así la síntesis proteica y bloqueando la replicación viral. Pues bien, en algunos estudios (Tan y Katze 2001) se ha visto que la NS5A posee una región (ISDR o región determinante de la sensibilidad a interferón), capaz de interaccionar con la PKR impidiendo por un lado la fosforilación de eIF2 y permitiendo,

por otro, que pueda completarse el ciclo de replicación viral. Este mecanismo podría llevar al virus a ofrecer cierta resistencia al tratamiento con interferón.

Si nos fijamos en el alineamiento de las secuencias aminoacídicas (Figura V.3) podemos observar cómo las posiciones seleccionadas positivamente, no se encuentran en las regiones correspondientes a la ISDR, sino que se hayan esparcidas a lo largo del fragmento sin seguir un patrón aparente. Además, la región correspondiente a la ISDR está bastante conservada. Tales observaciones podrían explicar el hecho de que apenas encontremos posiciones bajo selección positiva. Para el virus sería beneficioso que la región ISDR permaneciese lo más invariable posible para así poder seguir interaccionando con la PKR e impedir el bloqueo de la replicación viral. Por tanto se espera encontrar fenómenos de selección purificadora que eliminen aquellas mutaciones que puedan alterar esta región. De hecho, en algunos estudios se ha visto que mutaciones en la ISDR impiden la unión a la PKR favoreciendo la disminución de la síntesis proteica (Enomoto *et al.* 1996; Gale *et al.* 1999; Kurosaki *et al.* 1997), dificultando de este modo la replicación del virus.

Por tanto parece que las desviaciones al modelo neutral que encontrábamos en las pruebas de neutralidad, se deben a la acción de la selección, positiva en el caso de la región E1-E2, favoreciendo la resistencia al ataque del sistema inmune y negativa en el caso de la región NS5A, impidiendo la acumulación de mutaciones en la ISDR y favoreciendo la replicación viral.

4. TAMAÑO POBLACIONAL EFECTIVO

El tamaño poblacional efectivo (N_e) de una población vendría definido por el tamaño de una población ideal que se vería afectado por la deriva genética a la misma tasa que la población actual (Wright 1931). Puesto que se trata de virus, N_e sería el número medio de variantes virales que producen descendencia infecciosa (Nijhuis *et al.* 1998). N_e , en general es sustancialmente menor que el censo de individuos reproductores. Es lo que hemos estimado en el VHC, y además también se ha observado en otros virus de RNA como el VIH (Leigh–Brown *et al.* 1997; Nijhuis *et al.* 1998; Grassly *et al.* 1999), en los cuales las estimas de N_e eran, varios órdenes de magnitud, menores que los tamaños censales reales.

Los tamaños poblacionales en VHC no difieren entre genotipos, por lo que el tamaño poblacional efectivo tampoco nos aporta una pista para poder explicar la mayor prevalencia de un genotipo frente a otro. Las estimas varían de 30 a 900. Este número es relativamente bajo si tenemos en cuenta que el tamaño censal puede alcanzar cifras del orden de 10^{10} a 10^{12} (Davis 2002). El por qué de este N_e tan pequeño podría ser explicado por la presencia de partículas no infecciosas debidas a mutaciones deletéreas producidas por la falta de la prueba de lectura de la polimerasa, o quizás por errores mecánicos durante el ensamblaje del virus, lo cual dificultaría el dar lugar a una progenie viable. Sin embargo carecemos de la información necesaria para poder discriminar entre esos factores. Pero sí existen otros factores que pueden influir en N_e y podemos evaluar, a saber: la ruta de transmisión y la selección natural. Los distintos modos de transmisión llevan asociadas diferentes tasas de producción de nuevas infecciones. Cuánto más rápido se produce la nueva infección mayor es la varianza asociada a esta tasa, lo cual implica un N_e bajo (Grassly *et al.* 1999). Las transmisiones del virus a través de por ejemplo la compartición de agujas en ADVP o mediante transfusiones de sangre, son mucho más rápidas que a través de la transmisión sexual. Puesto que en los genotipos del VHC que estamos analizando las principales vías de transmisión que se conocen son las dos primeras, esto podría ser una explicación de las bajas estimas en N_e , y de por qué no encontramos diferencias entre los dos genotipos.

La selección natural también puede jugar un papel importante. Sabemos que N_e está relacionado con la variabilidad genética mediante la expresión $\theta = 2N_e\mu$. Por tanto para tener valores bajos de N_e , la tasa de mutación debería ser baja. Sin embargo puesto que estos virus poseen unas tasas de mutación bastante altas, N_e no debería ser bajo. No

obstante existen otros factores que pueden alterar la variación genética manteniendo una alta tasa de mutación, como puede ser el arrastre selectivo. Este fenómeno, como ya vimos en apartados anteriores, produce una disminución de la variabilidad genética debido a la fijación de mutaciones ventajosas junto con otras mutaciones neutras ligadas. Esta disminución en la variabilidad, provocaría una disminución de N_e . Así el arrastre selectivo podría ser otra posible explicación a por qué encontramos estos bajos N_e . Recordemos que en el apartado de resultados correspondientes a las pruebas de neutralidad, observábamos que los valores de D de Tajima, D^* y F^* de Fu y Li y F_s de Fu que obteníamos podrían ser debidos a este fenómeno.

VI. Conclusiones

1. Las pruebas de neutralidad muestran en la mayoría de casos, sin distinción entre genotipos, una desviación de la hipótesis neutral hacia valores negativos que indican la presencia de fenómenos de selección direccional, bien favoreciendo la fijación de mutaciones ventajosas o eliminando las deletéreas.
2. Tras los análisis realizados con PAML, aunque todo parece apuntar a la actuación mayoritaria de la selección purificadora, se han encontrado posiciones aminoacídicas dentro de la región hipervariable 1 (HVR1) bajo selección positiva. Estas posiciones coinciden en la mayoría de individuos, pertenezcan o no al mismo genotipo. Este resultado apoya la hipótesis, ya formulada, de que el sistema inmune ejerce una presión selectiva sobre la región HVR1, permitiendo al virus escapar a los mecanismos de defensa del hospedador.
3. En comparación con el tamaño censal, el tamaño poblacional efectivo encontrado es muy bajo, independientemente de los genotipos. Tal valor puede ser debido a fenómenos de arrastre selectivo, probablemente en respuesta a la presión selectiva ejercida por el sistema inmune.
4. Los análisis de la historia demográfica del virus muestran que la dinámica poblacional seguida por ambos genotipos es diferente. Mientras en los individuos de genotipo 1b se observa una estabilización en su crecimiento, en los de genotipo 1a, el crecimiento es exponencial. Al realizar la comparación de ambas dinámicas en el tiempo, se observa una coincidencia en el momento de cambio de las mismas. Ello coincide con el empleo de medidas de control de productos sanguíneos más específicas contra el VHC, y el aumento del consumo de drogas vía intravenosa. Puesto que en el resto de análisis realizados no encontramos diferencias genotípicas, podemos concluir que la mayor prevalencia del genotipo 1b sobre el 1a, es debida a la diferente ruta de transmisión, que mayoritariamente siguen ambos genotipos.

VII. Bibliografía

- Akaike H. (1974). New look at statistical-model identification. *IEEE Trans. Automat. Control.* AC19:716-723.
- Akashi H. (1999). Within- and between-species DNA sequence variation and the 'footprint' of natural selection. *Gene* 238:9-51.
- Battegay M., Feinstone SM. (1997). Hepatitis A virus. In *Viral Hepatitis* (ed. Wilson RA.) pp. 35-84. Marcel Dekker, New York.
- Behrens SE., Tomei L., De Francesco R. (1996). Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J.* 15:12-22.
- Bell H., Hellum K., Harthug S., Maeland A., Ritland S., Myrvang B., von der Lippe B., Raknerud N., Skaug K., Gutigard BG., Skjaerven R., Prescott LE., Simmonds P. (1997). Genotype, viral load and age as independent predictors of treatment outcome of interferon-alpha 2a treatment in patients with chronic hepatitis C. Construct group. *Scand. J. Infect. Dis.* 29:17-22.
- Bergmann KF., Gerin JL. (1986). Antigens of hepatitis delta virus in the liver and serum of humans and animals. *J. Infect. Dis.* 154:702-706.
- Booth JC., Kumar U., Webster D., Monjardino J., Thomas HC. (1998). Comparison of the rate of sequence variation in the hypervariable region of E2/NS1 region of hepatitis C virus in normal and hypogammaglobulinemic patients. *Hepatology* 27:223-227.
- Boothe JC., Foster GR., Levine T., Thomas HC., Goldin RD. (1997). The relationship of histology in chronic HCV infection. *Liver* 17:144-51.
- Boulestin A., Sandres-Saune K., Payen JL., Alric L., Dubois M., Pasquier C., Vinel JP., Pascal JP., Puel J., Izopet J. (2002). Genetic heterogeneity of the envelope 2 gene and eradication of hepatitis C virus after a second course of interferon-alpha. *J. Med. Virol.* 68:221-228.
- Bradley DW., Krawczynski K., Cook EH Jr., McCaustland KA., Humphrey CD., Spelbring JE., Myint H., Maynard JE. (1987). Enterically transmitted non-A, non-B hepatitis: serial passage of disease in cynomolgus macaques and tamarins and recovery of disease-associated 27- to 34-nm viruslike particles. *Proc. Natl. Acad. Sci. U S A.* 84:6277-6281.
- Chang KM., Thimme R., Melpolder JJ., Oldach D., Pemberton J., Moorhead-Loudis J., McHutchison JG., Alter HJ., Chisari FV. (2001). Differential CD4(+) and CD8(+) T-cell responsiveness in hepatitis C virus infection. *Hepatology* 33:267-276.

-
- Chayama K., Tsubota A., Kobayashi M., Okamoto K., Hashimoto M., Miyano Y., Koike H., Kobayashi M., Koida I., Arase Y., Saitoh S., Suzuki Y., Murashima N., Ikeda K., Kumada H. (1997). Pretreatment virus load and multiple amino acid substitutions in the interferon sensitivity-determining region predict the outcome of interferon treatment in patients with chronic genotype 1b hepatitis C virus infection. *Hepatology* 25:745-749.
- Choo QL., Kuo G., Weiner AJ., Overby LR., Bradley DW. and Houghton M. (1989). Isolation of a cDNA derived from a blood-borne non-A., non-B hepatitis genome. *Science* 244:359-362.
- Chuang WL., Chen SC., Dai CY., Hou C., Wang JH., Lu SN., Huang JF., Lin ZY., Hsieh MY., Tsai JF., Wang LY., Chang WY., Yu ML. (2001) Changing prevalence of hepatitis C virus Genotypes Molecular Epidemiology and clinical Implications in the hepatitis C virus hyperendemic areas and a tertiary referral center in Taiwan. *J. Med. Virol.*65:58-65.
- Chung RT., He W., Saquib A., Contreras AM., Xavier RJ., Chawla A., Wang TC., Schmidt EV. (2001). Hepatitis C virus replication is directly inhibited by IFN-alpha in a full-length binary expression system. *Proc. Natl. Acad. Sci. U S A.* 98:9847-9852.
- Comeron JM. (1995). A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. *J. Mol. Evol.* 41:1152-1159.
- Crandall KA., Kelsey CR., Imamichi H., Lane HC., Salzman NP. (1999). Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol. Biol. Evol.* 16:372-382.
- Crotti S., Maag D., Arnold JJ., Zhong W., Lau JY., Hong Z., Andino R., Cameron CE. (2000). The broad-spectrum antiviral ribonucleoside rivabirin is an RNA virus mutagen. *Nat. Med.* 6:1375-1379.
- Curran R., Jameson CL., Craggs JK., Grabowska AM., Thomson BJ., Robins A., Irving WL., Ball JK. (2002). Evolutionary trends of the first hypervariable region of the hepatitis C virus E2 protein in individuals with differing liver disease severity. *J. Gen. Virol.* 83(Pt 1):11-23.
- Davis GL. (2002) Monitoring of viral levels during therapy of hepatitis C. *Hepatology.* 36:145-151.
- Dekaban GA., Digilio L., Franchini G. (1995). The natural history and evolution of human and simian T cell leukemia/lymphotropic viruses. *Curr. Opin. Genet. Dev.* 5:807-813.
- Di Bisceglie AM. (1998). Hepatitis C. *Lancet* 351:351-355.
- Domingo E. (1992). Genetic variation and quasi-species. *Curr. Opin. Genet. Dev.* 2:61-63.

- Domingo E., Holland JJ. (1997). RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51:151-178.
- Eigen M., Schuster P. (1981). Comments on "growth of a hypercycle" by King. *Biosystems.* 1981;13(4):235.
- Eigen M., Gardiner W., Schuster P., Winkler-Oswatitsch R. (1981). The origin of genetic information. *Sci. Am.* 244:88-92.
- Eigen M., Biebricher CK. (1988). Sequence space and quasispecies distribution. In: RNA Genetics Vol.3. Editors: Domingo E., Ahlquist P., Holland J. Ed. CRC Press., Boca Raton., Florida., EUA. 211-245.
- Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T., Yamamoto C., Izumi N., Marumo F., Sato C. (1995). Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J. Clin. Invest.* 96:224-230.
- Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T., Yamamoto C., Ogura Y., Izumi N., Marumo F., Sato C. (1996). Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N. Engl. J. Med.* 334:77-81.
- Excoffier L., Smouse PE., Quattro JM. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-91.
- Farci P., Alter HJ., Shimoda A., Govindarajan S., Cheung LC., Melpolder JC., Sacher RA., Shih JW., Purcell RH. (1996). Hepatitis C virus-associated fulminant hepatic failure. *N. Engl. J. Med.* 335:631-634.
- Farci P., Shimoda A., Wong D., Cabezon T., De Gioannis D., Strazzer A., Shimizu Y., Shapiro M., Alter HJ., Purcell RH. (1996). Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. *Proc. Natl. Acad. Sci. USA.* 93:15394-15399.
- Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder JC., Strazzer A., Chien DY., Munoz SJ., Balestrieri A., Purcell RH., Alter HJ. (2000). The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* 288:339-344.
- Farci P. (2001). Hepatitis C virus. The importance of viral heterogeneity. *Clin. Liver Dis.* 5:895-916. Review.
- Feinstone SM., Kapikian AZ., Purcell RH., Alter HJ., Holland PV. (1975). Transfusion-associated hepatitis not due to viral hepatitis type A or B. *N. Engl. J. Med.* 292:767-770.

-
- Flint M., McKeating JA. (1999). The C-terminal region of the hepatitis C virus E1 glycoprotein confers localization within the endoplasmic reticulum. *J. Gen. Virol.* 80:1943-1947.
- Forns X., Purcell RH., Bukh J. (1999). Quasispecies in viral persistence and pathogenesis of hepatitis C virus. *Trends Microbiol.* 7:402-410.
- Frank C., Mohamed MK., Strickland GT., Lavanchy D., Arthur RR., Magder LS., El Khoby T., Abdel-Wahab Y., Aly Ohn ES., Anwar W., Sallam I. (2000). The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet* 355:887-891.
- Fu YX., Li WH. (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693-709.
- Fu YX. (1997) Statistical tests of neutrality of mutations against population growth., hitchhiking and background selection. *Genetics* 147:915-925.
- Gale M Jr., Blakely CM., Kwieciszewski B., Tan SL., Dossett M., Tang NM., Korth MJ., Polyak SJ., Gretch DR., Katze MG. (1998). Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. *Mol Cell Biol.* 18(9):5208-5218.
- Gale M Jr., Kwieciszewski B., Dossett M., Nakao H., Katze MG. (1999). Antiapoptotic and oncogenic potentials of hepatitis C virus are linked to interferon resistance by viral repression of the PKR protein kinase. *J. Virol.* 73:6506-6516.
- Gallinari P., Brennan D., Nardi C., Brunetti M., Tomei L., Steinkuhler C., De Francesco R. (1998). Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus. *J. Virol.* 72:6758-6769.
- Gaud U., Langer B., Petropoulou T., Thomas HC., Karayiannis P. (2003). Changes in hypervariable region 1 of the envelope 2 glycoprotein of hepatitis C virus in children and adults with humoral immune defects. *J. Med. Virol.* 69:350-356.
- Gerlach JT., Diepolder HM., Jung MC., Gruener NH., Schraut WW., Zachoval R., Hoffmann R., Schirren CA., Santantonio T., Pape GR. (1999). Recurrence of hepatitis C virus after loss of virus-specific CD4(+) T-cell response in acute hepatitis C. *Gastroenterology* 117:933-941.
- Gillespie JH. (1991). The causes of molecular Evolution. Ed. Oxford Universtiy Press.
- Goldman N., Yang Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11:725-736.
- Grassly NC., Harvey PH., Holmes EC. (1999). Population dynamics of HIV-1 inferred from gene sequences. *Genetics* 151:427-438.

- Green KY., Ando T., Balayan MS., Berket T., Clarke IN., Estes MK., Matson DO., Nakatas S., Neill JD., Studdert MJ., Thiel HJ. (2000). Taxonomy of the caliciviruses. *J. Infect. Dis.* 181 Suppl2:S322-S330.
- Griffiths RC., Tavare S. (1994). Sampling theory for neutral alleles in a varying environment. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 344(1310):403-10.
- Gust J., Coulepis AG., Feinstone SM. et al. (1998). Taxonomic classification of hepatitis A virus. *Intervirology* 20:1-7.
- Hadler SC. (1991). Global impact of hepatitis A virus infection changing patterns. In *Viral Hepatitis and Liver disease* (Hollinger FB., Lemon SM., Margolis HS. Eds). Williams and Wilkins, Baltimore.
- Herrmann E., Neumann AU., Schmidt JM., Zeuzem S. (2000). Hepatitis C virus kinetics. *Antivir Ther*; 5(2):85-90. *J. Virol.* 72:6271-6276.
- Holland JJ., De La Torre JC., Steinhauer DA. (1992). RNA virus populations as quasispecies. *Curr. Top. Microbiol. Immunol.* 176:1-20.
- Holmes EC., Nee S., Rambaut A., Garnett GP., Harvey PH. (1995). Revealing the history of infectious disease epidemics through phylogenetic trees. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 29:33-40.
- Holmes EC., Zhang LQ., Robertson P., Cleland A., Harvey E., Simmonds P., Leigh Brown AJ. (1995). The molecular epidemiology of human immunodeficiency virus type 1 in Edinburgh. *J. Infect. Dis.* 171:45-53.
- Hurvich CM., Tsai CL. (1989). Regression and time series model selection in small samples. *Biometrika.* 76:297-307
- Ina Y. (1995). New methods for estimating the numbers of synonymous and nonsynonymous substitutions. *J. Mol. Evol.* 40:190-226.
- Ishido S., Fujita T., Hotta H. (1998). Complex formation of NS5B with NS3 and NS4A proteins of hepatitis C virus. *Biochem. Biophys. Res. Commun.* 244:35-40.
- Kalinina O., Norder H., Vetrov T., Zhdanov K., Barzunova M., Plotnikova V., Mukomolov S., and Magnus LO. (2001). Shift in predominating subtype of HCV from 1b to 3a in St. Petersburg mediated by increase in injecting drug use. *J. Med. Virol:* 65:517-524.
- Kauffman RJ. (2000). Double-stranded RNA-activated protein kinase PKR. In *Translational Control of Gene Expression* (N. Sonenberg., JWB. Hersey and MB Mathews Eds.) pp.503-528. Cold Spring Harbor Laboratory Press., Cold Spring Harbor., NY.

-
- Kim JL., Morgenstern KA., Lin C., Fox T., Dwyer MD., Landro JA., Chambers SP., Markland W., Lepre CA., O'Malley ET., Harbeson SL., Rice CM., Murcko MA., Caron PR., Thomson JA. (1996). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* 87:343-355.
- Kimura M. (1968). Evolutionary rate at the molecular level. *Nature* 17:624-626.
- Kimura M. (1983). The neutral theory of Molecular Evolution. Ed. Cambridge University Press.
- King JL., Jukes TH. (1969) Non-Darwinian evolution. *Science* 16:788-798.
- Kingman JFC. (1982). On the genealogy of large populations. *J. Appl. Prob.* 19:27-43.
- Kleter B., Brouwer JT., Nevens F., van Doorn LJ., Elewaut A., Versieck J., Michielsens PP., Hautekeete ML., Chamuleau RA., Brenard R., Bourgeois N., Adler M., Quint WG., Bronkhorst CM., Heijtkink RA., Hop WJ., Fevery J., Schalm SW. (1998) Hepatitis C virus genotypes: epidemiological and clinical associations. Benelux Study Group on Treatment of Chronic Hepatitis C. *Liver* 18:32-38.
- Kittlesen DJ., Chianes-Bullock KA., Yao ZQ., Braciale TJ., Hahn YS. (2000). Interaction between complement receptor gCIqR and hepatitis C virus core protein inhibits T-lymphocyte proliferation. *J. Clin. Invest.* 106: 1239-1246
- Koziel MJ., Walker BD. (1997). Characteristics of the intrahepatic cytotoxic T lymphocyte response in chronic hepatitis C virus infection. *Springer. Semin. Immunopathol.* 19:69-83.
- Kronenberger B., Ruster B., Elez R., Weber S., Piiper A., Lee JH., Roth WK., Zeuzem S. (2001). Interferon alfa down-regulates CD81 in patients with chronic hepatitis C. *Hepatology* 33:1518-1526.
- Kumar U., Monjardino J., Thomas HC. (1994). Hypervariable region of hepatitis C virus envelope glycoprotein (E2/NS1) in an agammaglobulinemic patient. *Gastroenterology* 106:1072-1075.
- Kurosaki M., Enomoto N., Murakami T., Sakuma I., Asahina Y., Yamamoto C., Ikeda T., Tozuka S., Izumi N., Marumo F., Sato C. (1997). Analysis of genotypes and amino acid residues 2209 to 2248 of the NS5A region of hepatitis C virus in relation to the response to interferon-beta therapy. *Hepatology* 25:750-753.
- Lai MM. (1992). RNA recombination in animal and plant viruses. *Microbiol. Rev.* 56:61-79. Review.
- Lawal Z., Petrik J., Wong VS., Alexander GJ., Allain JP. (1997). Hepatitis C virus genomic variability in untreated and immunosuppressed patients. *Virology* 228:107-111.

- Lee CH., Choi YH., Yang SH., Lee CW., Ha SJ., Sung YC. (2001) Hepatitis C virus core protein inhibits interleukin 12 and nitric oxide production from activated macrophages. *Virology*. 279:271-279.
- Lehman T., Hawley WA., Grebert H., Collins FH. (1998). The effective population size of *Anopheles gambiae* in Kenia: Implications for population structure. *Mol.Biol. Evol.* 15:264-276.
- Leigh Brown AJ. (1997). Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc. Natl. Acad. Sci. USA*. 94:1862-1865.
- Li WH., Wu CI., Luo CC. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* 2:150-174. Review.
- Li WH. (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* 36:96-99.
- Li WH. (1997). *Molecular Evolution*. Sinauer Associates., Sunderland., MA.
- Lin C., Wu JW., Hsiao K., Su MS. (1997). The hepatitis C virus NS4A protein: interactions with the NS4B and NS5A proteins. *J. Virol.* 71:6465-6471.
- Lo SY., Selby MJ., Ou JH. (1996). Interaction between hepatitis C virus core protein and E1 envelope protein. *J. Virol.* 70:5177-5182.
- Lopez-Labrador FX., Ampurdanes S., Forns X., Castells A., Saiz JC., Costa J., Bruix J., Sanchez Tapias JM., Jimenez de Anta MT., Rodes J. (1997). Hepatitis C virus (HCV) genotypes in Spanish patients with HCV infection: relationship between HCV genotype 1b., cirrhosis and hepatocellular carcinoma. *J. Hepatol.* 27:959-965.
- Louwagie J., McCutchan FE., Peeters M., Brennan TP., Sanders-Buell E., Eddy GA., van der Groen G., Fransen K., Gershby-Damet GM., Deleys R., et al. (1993). Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS*. 7:769-80.
- Luo GX., Chao M., Hsieh SY., Sureau C., Nishikura K., Taylor J. (1990). A specific base transition occurs on replicating hepatitis delta virus RNA. *J. Virol.* 64:1021-1027.
- Maag D., Castro C., Hong Z., Cameron CE. (2001). Hepatitis C virus RNA-dependent RNA polymerase (NS5B) as a mediator of the antiviral activity of ribavirin. *J. Biol. Chem.* 276:46094-46098.
- Manzin A., Solfrosi L., Petrelli E., Macarri G., Tosone G., Piazza M., Clementi M. (1998). Evolution of hypervariable region 1 of hepatitis C virus in primary infection. *J. Virol.* 72:6271-6276.

-
- Matsumoto M., Hwang SB., Jeng KS., Zhu N., Lai MM. (1996). Homotypic interaction and multimerization of hepatitis C virus core protein. *Virology* 218:43-51.
- McMichael AJ., Rowland-Jones SL. (2001). Cellular immune responses to HIV. *Nature* 410:980-987. Review
- Memon MI., Memon MA. (2002). Hepatitis C: an epidemiological review. *Viral. Hepat.* 9:84-100.
- Mijas Lafont L. (1995) Prevalencia del virus de la hepatitis C en la Comunidad Valenciana. En: HEPatitis Víricas. Ed. E. Ortega González, J. Aguilar S.L., Valencia
- Miyata T., Yasunaga T. (1980). Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *J. Mol. Evol.* 16:23-36.
- Mizokami M., Lau JY., Suzuki K., Nakano T., Gojobori T. (1994). Differential sensitivity of hepatitis C virus quasispecies to interferon-alpha therapy. *J. Hepatol.* 21:884-886.
- Mondelli MU., Silini E. (1999). Clinical significance of hepatitis C virus genotypes. *J. Hepatol.* 31:65-70.
- Moriya K., Fujie H., Shintani Y., Yotsuyanagi H., Tsutsumi T., Ishibashi K., Matsuura Y., Kimura S., Miyamura T., Koike K. (1998). The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. *Nat. Med.* 4:1065-1067.
- Mortimer PP. (1995). Arspenamine jaundice and the recognition of instrument-borne virus infection. *Genitourin. Med.* 71:109-119.
- Muse SV. (1996). Estimating synonymous and nonsynonymous substitution rates. *Mol. Biol. Evol.* 13:105-114.
- Nakano I., Fukuda Y., Katano Y., Nakano S., Kumada T., Hayakawa T. (1999). Why is the interferon sensitivity-determining region (ISDR) system useful in Japan? *J. Hepatol.* 30:1014-1022.
- Nassal M., and Schaller H. (1996). Hepatitis B virus replication. *J. Viral Hepat.* 3:217-226.
- Nei M., Gojobori T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3:418-426.
- Ni YH., Chang MH., Chen PJ., Hsu HY., Lu TW., Lin KH., Lin DT. (1999). Decreased diversity of hepatitis C virus quasispecies during bone marrow transplantation. *J. Med. Virol.* 58:132-138.

- Nielsen R., Yang Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929-936.
- Nijhuis M., Boucher CA., Schipper P., Leitner T., Schuurman R., Albert J. (1998). Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy. *Proc. Natl. Acad. Sci. U S A.* 95:14441-14446.
- Nowak MA., May RM. The basic model of virus dynamics. In: *Virus Dynamics. Mathematical principles of immunology and Virology.* Oxford University Press.13-26.
- Ogata N., Alter HJ., Miller RH., Purcell RH. (1991). Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc. Natl. Acad. Sci. USA.* 88:3392-3396.
- Ohta T., Kimura M. (1971). Behavior of neutral mutants influenced by associated overdominant loci in finite populations. *Genetics* 69:247-260.
- Okamoto H., Tsuda F., Sakugawa H., Sastrosoewignjo RI., Imai M., Miyakawa Y., Mayumi M. (1988). Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J. Gen. Virol.* 69:2575-2583.
- Okamoto H., Mishiro S. (1994). Genetic heterogeneity of hepatitis C virus. *Intervirology*37:68-76.
- Ortega E. (2002) Hepatitis víricas. Centro Francisco Tomás y Valiente UNCED Alzira-Valencia. Colección Interciencias.
- Pamilo P., Bianchi NO. (1993). Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol. Biol. Evol.* 10:271-281.
- Pawlotsky JM., Tsakiris L., Roudot-Thoraval F., Pellet C., Stuyver L., Duval J., Dhumeaux D. (1995). Relationship between hepatitis C virus genotypes and sources of infection in patients with chronic hepatitis C. *J. Infect. Dis.* 171:1607-1610.
- Pawlotsky JM., Pellerin M., Bouvier M., Roudot-Thoraval F., Germanidis G., Bastie A., Darthuy F., Remire J., Soussy CJ., Dhumeaux D. (1998). Genetic complexity of the hypervariable region 1 (HVR1) of hepatitis C virus (HCV): influence on the characteristics of the infection and responses to interferon alfa therapy in patients with chronic hepatitis C. *J. Med. Virol.* 54:256-264.
- Pileri P., Uematsu Y., Campagnoli S., Galli G., Falugi F., Petracca R., Weiner AJ., Houghton M., Rosa D., Grandi G., Abrignani S. (1998) Binding of hepatitis C virus to CD81. *Science* 282: 938-941.
- Pol S., Thiers V., Noursbaum JB., Legendre C., Berthelot P., Kreis H., Brechot C. (1995). The changing relative prevalence of hepatitis C virus genotypes: evidence in hemodialyzed patients and kidney recipients. *Gastroenterology* 108:581-583.

-
- Polyak SJ., Khabar KS., Paschal DM., Ezelle HJ., Duverlie G., Barber GN., Levy DE., Mukaida N., Gretch DR. (2001). Hepatitis C virus nonstructural 5A protein induces interleukin-8., leading to partial inhibition of the interferon-induced antiviral response. *J. Virol.* 75:6095-6106.
- Provost PJ., Hilleman MR. (1978). Inactivated hepatitis a-virus vaccine prepared from infected marmoset liver. *Proc. Soc. Exp. Biol. Med.* 159: 201-203.
- Pybus OG., Charleston MA., Gupta S., Rambaut A., Holmes EC., Harvey PH. (2001). The epidemic behavior of the hepatitis C virus. *Science* 292:2323-2325.
- Pybus OG., Rambaut A. (2002). GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics* 18:1404-1405.
- Ramos-Onsins SE., Rozas J. (2002). Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* 19:2092-2100.
- Reed KE., Grakoui A., Rice CM. (1995) Hepatitis C virus encoded NS2-3 protease: cleavage-site mutagenesis and requirements for biomolecular cleavage. *J. Virol.* 69:4127-4136
- Reed KE., Rice CM. (2000). Overview of hepatitis C virus genome structure., polyprotein processing., and protein properties. *Curr. Top. Microbiol. Immunol.* 242:55-84.
- Rehermann B. (2000). Interaction between the hepatitis C virus and the immune system. *Semin. Liver. Dis.* 20:127-141.
- Reyes GR., Purdy MA., Kim JP., Luk KC., Young LM., Fry KE., Bradley DW. (1990). Isolation of a cDNA from the virus responsible for enterically transmitted non-A., non-B hepatitis. *Science* 16:1335-1339.
- Rice P., Longden I., Bleasby A. (2000). EMBOSS: The European molecular biology open software suite. *Trends in Genetics.* 16:276-277.
- Robertson B., Myers G., Howard C., Brettin T., Bukh J., Gaschen B., Gojobori T., Maertens G., Mizokami M., Nainan O., Netesov S., Nishioka K., Shin i T., Simmonds P., Smith D., Stuyver L., Weiner A. (1998). Classification., nomenclature., and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Virus Taxonomy. *Arch. Virol.* 143:2493-2503.
- Rosa D., Campagnoli S., Moretto C., Guenzi E., Cousens L., Chin M., Dong C., Weiner AJ., Lau JY., Choo QL., Chien D., Pileri P., Houghton M., Abrignani S. (1996). A quantitative test to estimate neutralizing antibodies to the hepatitis C virus: cytofluorimetric assessment of envelope glycoprotein 2 binding to target cells. *Proc. Natl. Acad. Sci. U S A.* 93:1759-1763.
- Rosen HR., Gretch DR. (1999). Hepatitis C virus: current understanding and prospects for future therapies. *Mol. Med. Today.* 9:393-399.

- Rozas J., Rozas R. (1999). DnaSP version 3: an integrated program for molecular population *Genetics* and molecular evolution analysis. *Bioinformatics* 15:174-175.
- Rueckert RR., Wimmer E. (1984). Systematic nomenclature of picornavirus proteins *J.Virol.* 50: 957-959
- Saiz JC., Lopez-Labrador FX., Ampurdanes S., Dopazo J., Forns X., Sanchez-Tapias JM., Rodes J. (1998). The prognostic relevance of the nonstructural 5A gene interferon sensitivity determining region is different in infections with genotype 1b and 3a isolates of hepatitis C virus. *J. Infect. Dis.* 177:839-847.
- Sakuma I., Enomoto N., Kurosaki M., Marumo F., Sato C. (1996). Selection of hepatitis C virus quasispecies during interferon treatment. *Arch. Virol.* 141:1921-1932.
- Sato K., Okamoto H., Aihara S., Hoshi Y., Tanaka T., Mishiro S. (1993). Demonstration of sugar moiety on the surface of hepatitis C virions recovered from the circulation of infected humans. *Virology* 196:354-357.
- Schmidt HA., Strimmer K., Vingron M., von Haeseler A. (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502-504.
- Schneider S., Roessli D., Excoffier L. (2000) Arlequin ver. 2000: A software for population genetic data analysis. Genetics and Biometry Laboratory., University of Geneva., Switzerland.
- Schulze zur Wiesch J., Schmitz H., Borowski E., Borowski P. (2003). The proteins of the Hepatitis C virus: their features and interactions with intracellular protein phosphorylation. *Arch. Virol.* 148:1247-1267.
- Sharp PM., Robertson DL., Hahn BH. (1995). Cross-species transmission and recombination of 'AIDS' viruses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 29:41-47.
- Sharp PM. (1997). In search of molecular darwinism. *Nature* 385:111-112.
- Shimazaki T., Honda M., Kaneko S., Kobayashi K. (2002). Inhibition of internal ribosomal entry site-directed translation of HCV by recombinant IFN-alpha correlates with a reduced La protein. *Hepatology* 35:199-208.
- Shimoike T., Mimori S., Tani H., Matsuura Y., Miyamura T. (1999). Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. *J. Virol.* 73:9718-9725.
- Simmonds P., Holmes EC., Cha TA., Chan SW., McOmish F., Irvine B., Beall E., Yap PL., Kolberg J., Urdea MS. (1993). Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *J. Gen. Virol.* 74:2391-2399.

-
- Simmonds P., Alberti A., Alter HJ., Bonino F., Bradley DW., Brechot C., Brouwer JT., Chan SW., Chayama K., Chen DS., Choo QL., Colombo M., Cuypers HTM., Date T., Dusheiko GM., Esteban JL., Fay O., Hadziyannis SJ., Han J., Hatzakis A., Holmes EC., Hotta H., Houghton M., Irvine B., Kohara M., Kolberg JA., Kuo G., Lau JYN., Lelie PN., Maertens G., Mcomish F., Miyamura T., Mizokami M., Nomoto A., Prince AM., Reesink HW., Rice C., Roggendorf M., Schalm SW., Shikata T., Shimotohno K., Stuyver L., Trepo C., Weiner A., Yap PL., Urdea MS. (1994). A proposed system for the nomenclature of hepatitis C viral genotypes. *Hepatology* 19:1321-1324.
- Simmonds P., Mellor J., Craxi A., Sanchez-Tapias JM., Alberti A., Prieto J., Colombo M., Rumi MG., Lo Iacano O., Ampurdances-Mingall S., Forns-Bernhardt X., Chemello L., Civeira MP., Frost C., Dusheiko G. (1996). Epidemiological, clinical and therapeutic associations of hepatitis C types in western European patients. *J. Hepatol.* 24:517-524.
- Simmonds P., Mellor J., Sakuldamrongpanich T., Nuchaprayoon C., Tanprasert S., Holmes EC., Smith DB. (1996). Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: comparison with classifications based upon sequence similarity. *J. Gen. Virol.* 77:3013-3024.
- Simmonds P. (1999). Viral heterogeneity of the hepatitis C virus. *J. Hepatol.* 31 Suppl 1:54-60.
- Simonsen KL., Churchill GA., Aquadro CF. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141:413-429.
- Smith DB., Pathirana S., Davidson F., Lawlor E., Power J., Yap PL., and Simmonds P. (1997). The origin of hepatitis C virus genotypes. *J. Gen. Virol.* 78:321-328.
- Spangberg K., Wiklund L., Schwartz S. (2001). Binding of the La autoantigen to the hepatitis C virus 3' untranslated region protects the RNA from rapid degradation in vitro. *J. Gen. Virol.* 82:113-120.
- Staden R., Beal KF., Bonfield JK. (1998). The Staden package. *Methods Mol. Biol.* 2000 132:115-130.
- Strauss EG., Strauss JH., Levine AJ. (1996) Virus Evolution. In fields Virology. Ed B.N. Fields., Knipe DM. Howley PM. Et al. Lipincott-Raven Publishers., Philadelphia. Chapter 6., pp 153-171.
- Strimmer K., von Haeseler A. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. USA.* 24:6815-6819.
- Stumpf M., Pybus O. (2002). Genetic diversity and models of viral evolution for the hepatitis C virus. *FEMS Microbiol. Lett.* 214:143-152.
- Suzuki Y., Gojobori T. (2001). Positively selected amino acid sites in the entire coding region of hepatitis C virus subtype 1b. *Gene* 276:83-87.

- Swofford DL. (2001). PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). Sinauer Associates., Sunderland., MA.
- Tajima F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595
- Takada A., Tsutsumi M., Zhang SC., Okanoue T., Matsushima T., Fujiyama S., Komatsu M. (1996). Relationship between hepatocellular carcinoma and subtypes of hepatitis C virus: a nationwide analysis. *J. Gastroenterol. Hepatol.* 11:166-169.
- Takahata N., Nei M. (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 124:967-978.
- Takikawa S., Ishii K., Aizaki H., Suzuki T., Asakura H., Matsuura Y., Miyamura T. (200). Cell fusion activity of hepatitis C virus envelope proteins. *J. Virol.* 74: 5066-5074.
- Tam AW., Smith MM., Guerra ME., Huang CC., Bradley DW., Fry KE., Reyes GR. (1991). Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology* 185:120-131.
- Tan SL., Katze MG. (2001). How hepatitis C virus counteracts the interferon response: the jury is still out on NS5A. *Virology* 284:1-12. Review.
- Taylor DR., Shi ST., Romano PR., Barber GN., Lai MM. (1999). Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science* 285:107-10.
- Thompson JD., Gibson TJ., Plewniak F., Jeanmougin F., Higgins DG. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24: 4876-4882.
- Tsarev SA., Emerson SU., Reyes GR., Tsareva TS., Legters LJ., Malik IA., Iqbal M., Purcell RH. (1992). Characterization of a prototype strain of hepatitis E virus. *Proc. Natl. Acad. Sci. U S A.* 15:559-563.
- Tsuchihara K., Tanaka T., Hijikata M., Kuge S., Toyoda H., Nomoto A., Yamamoto N., Shimotohno K. (1997). Specific interaction of polypyrimidine tract-binding protein with the extreme 3'-terminal structure of the hepatitis C virus genome., the 3'X. *J. Virol.* 71:6720-6726.
- Tsukiyama-Kohara K., Iizuka N., Kohara M., Nomoto A. (1992). Internal ribosome entry site within hepatitis C virus RNA. *J. Virol.* 66:1476-1483.
- Wang KS., Choo QL., Weiner AJ., Ou JH., Najarian RC., Thayer RM., Mullenbach GT., Denniston KJ., Gerin JL., Houghton M. (1986). Structure., sequence and expression of the hepatitis delta (delta) viral genome. *Nature* 323:508-514.

-
- Weiner AJ., Geysen HM., Christopherson C., Hall JE., Mason TJ., Saracco G., Bonino F., Crawford K., Marion CD., Crawford KA., et al. (1992). Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc. Natl. Acad. Sci. U S A.* 89:3468-3472.
- Wright S. (1931). Evolution in mendelian populations. *Genetics* 16:97-159
- Wunschmann S., Medh JD., Klinzmann D., Schmidt WN., Stapleton JT. (2000). Characterization of hepatitis C virus (HCV) and HCV E2 interactions with CD81 and the low-density lipoprotein receptor. *J. Virol.* 74:10055-10062.
- Yang Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* 13:555-556.
- Yang Z., and Nielsen R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* 46:409-418.
- Yang Z. (2000). Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J. Mol. Evol.* 51:423-432.
- Yang Z., and Nielsen R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17:32-43.
- Yang Z., Nielsen R., Goldman N., Pedersen AM. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431-449.
- Yun ZB., Odeberg J., Lundeberg J., Weiland O., Uhlen M., Sonnerborg A. (1996) Restriction of hepatitis C virus heterogeneity during prolonged interferon-alpha therapy in relation to changes in virus load. *J. Infect. Dis.* 173:992-996.
- Zanotto PM., Gould EA., Gao GF., Harvey PH., Holmes EC. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc. Natl. Acad. Sci. U S A.* 23:548-553
- Zein NN., Poterucha JJ., Gross JB Jr., Wiesner RH., Therneau TM., Gossard AA., Wendt NK., Mitchell PS., Germer JJ., Persing DH. (1996). Increased risk of hepatocellular carcinoma in patients infected with hepatitis C genotype 1b. *Am J Gastroenterol* 91(12):2560-2562.
- Zein NN. (2000). Clinical significance of hepatitis C virus genotypes. *Clinical Microbiology Reviews* 13:223-234.
- Zein CO., Zein NN. (2002). Advances in therapy for hepatitis C infection. *Microbes Infect.* 4:1237-46.
- Zeuzem S., Feinman SV., Rasenack J., Heathcote EJ., Lai MY., Gane E., O'Grady J., Reichen J., Diago M., Lin A., Hoffman J., Brunda MJ. (2000). Peginterferon alfa-2a in patients with chronic hepatitis C. *N. Engl. J. Med.* 343:1666-1672.

- Zibert A., Schreier E., Roggendorf M. (1995). Antibodies in human sera specific to hypervariable region 1 of hepatitis C virus can block viral attachment. *Virology* 208:653-661.
- Zúñiga A., Colomina J., Fernández N., Cerón JA. (2001) Estudio de prevalencia de genotipos del virus de la hepatitis C en el área de salud 10 de la Comunidad Valenciana. Resumen P4.28 del VII Congreso Nacional de Virología celebrado en Valencia.

VIII. Anexos

ANEXO I:

Estimas de Π

Estimas del número medio de diferencias nucleotídicas entre pares de secuencias (Π).

$$\Pi = \frac{1}{[n(n-1)/2]} \sum \Pi_{ij}$$

Donde Π_{ij} es el número de diferencias nucleotídicas entre las secuencias i y j , y $n(n-1)/2$ es el número posible de pares de secuencias. Otra de manera de calcular Π , es a partir de las frecuencias de los diferentes alelos:

$$\Pi = \frac{n}{n-1} \sum X_i X_j \Pi_{ij}$$

Donde $X_i X_j$ son las frecuencias de los alelos i y j en la muestra, y $n/(n-1)$ es el factor de corrección de sesgo de muestreo. Ambas ecuaciones obtienen el mismo resultado.

ANEXO II

Medidas de polimorfismo.

En las tablas 1-4 se muestran diversas medidas de polimorfismo de DNA, para los individuos de genotipo 1a y 1b, y para las regiones E1-E2 (e) y NS5A (n). Las abreviaturas que se muestran en las tablas son las siguientes:

- **N° sec**: Número de secuencias.
- **Eta**: N° total de mutaciones.
- **S**: N° de sitios segregantes.
- **G+C**: Contenido en Guanina-Citosina.
- **Hd**: Diversidad haplotípica.
- **V(Hd)**: Varianza de la diversidad haplotípica.
- **Sd(Hd)**: Desviación estándar de la diversidad haplotípica.
- **π** : Diversidad nucleotídica.
- **Sd(π)**: Desviación estándar de la diversidad nucleotídica.
- **Π** : N° medio de diferencias nucleotídicas emparejadas.

Tabla 1. Medidas de polimorfismo correspondiente a la región E1-E2, para los individuos 1b.

Individuo 1b	N° Sec.	Eta.	S	G+C	Hd.	V(Hd).	Sd(Hd)	π	V(π)	Sd(π)	Π
eA03	100	89	87	0,605	0,993	0,000	0,003	0,026	0,000	0,001	12,430
eA06	100	42	40	0,583	0,641	0,003	0,058	0,003	0,000	0,001	1,411
eA14	100	122	107	0,588	0,997	0,000	0,002	0,033	0,000	0,002	15,518
eA16	100	166	150	0,595	0,998	0,000	0,002	0,054	0,000	0,002	25,423
eA35	109	63	61	0,586	0,966	0,000	0,012	0,012	0,000	0,000	5,537
eC03	100	8	8	0,586	0,607	0,001	0,032	0,004	0,000	0,000	2,000
eC04	100	5	5	0,602	0,098	0,002	0,041	0,000	0,000	0,000	0,100
eC05	100	153	134	0,599	0,998	0,000	0,002	0,047	0,000	0,001	22,131
eC08	100	133	119	0,588	0,996	0,000	0,002	0,046	0,000	0,001	21,729
eC09	111	62	61	0,587	0,95	0,000	0,013	0,007	0,000	0,001	3,317
eC10	100	89	83	0,587	0,99	0,000	0,004	0,027	0,000	0,001	12,546
eC12	100	103	89	0,599	0,997	0,000	0,002	0,025	0,000	0,002	11,732
eC13	100	94	87	0,615	0,992	0,000	0,004	0,012	0,000	0,001	5,519
eC15	102	126	113	0,586	0,998	0,000	0,002	0,058	0,000	0,002	27,330
eC16	101	86	78	0,574	0,969	0,000	0,011	0,025	0,000	0,002	11,863
eC19	100	141	126	0,601	0,999	0,000	0,002	0,040	0,000	0,001	18,886
eC25	100	9	9	0,595	0,277	0,004	0,059	0,001	0,000	0,000	0,389
eG02	100	103	93	0,602	0,986	0,000	0,005	0,045	0,000	0,001	21,198
eG05	100	82	76	0,589	0,952	0,000	0,016	0,012	0,000	0,001	5,841
eG06	100	123	112	0,587	0,999	0,000	0,002	0,029	0,000	0,001	13,528
eG07	100	109	103	0,599	0,996	0,000	0,002	0,040	0,000	0,002	18,826
eG09	100	129	121	0,594	0,998	0,000	0,002	0,023	0,000	0,001	10,613
eG16	103	2	2	0,609	0,039	0,001	0,026	0,000	0,000	0,000	0,039
eG17	101	101	97	0,599	0,997	0,000	0,002	0,051	0,000	0,001	23,935
eG26	100	8	8	0,595	0,173	0,003	0,051	0,000	0,000	0,000	0,180

Tabla 2. Medidas de polimorfismo correspondiente a la región NS5A, para los individuos 1b.

Individuo 1b	N° Sec.	Eta.	S	G+C	Hd.	V(Hd).	Sd(Hd)	π	V(π)	Sd(π)	Π
nA03	38	85	81	0,614	0,984	0,00019	0,014	0,013	0,0000005	0,001	9,814
nA06	96	65	63	0,595	0,810	0,00187	0,043	0,003	0,0000001	0,000	2,202
nA14	58	67	65	0,601	0,990	0,00003	0,006	0,008	0,0000004	0,001	5,790
nA16	52	130	124	0,586	0,998	0,00002	0,004	0,021	0,0000053	0,002	15,436
nA35	27	29	28	0,597	0,963	0,00063	0,025	0,006	0,0000004	0,001	4,584
nC03	86	23	23	0,611	0,833	0,00116	0,034	0,002	0	0,000	1,644
nC04	92	27	27	0,603	0,971	0,00006	0,007	0,006	0,0000001	0,000	4,292
nC05	75	77	72	0,608	0,995	0,00001	0,003	0,010	0,0000007	0,001	7,614
nC08	77	155	144	0,600	0,980	0,00005	0,007	0,019	0,0000047	0,002	13,913
nC09	57	15	15	0,596	0,835	0,002	0,045	0,003	0,0000001	0,000	2,444
nC10	63	48	48	0,594	0,947	0,00036	0,019	0,008	0,0000008	0,001	5,676
nC12	92	119	114	0,613	0,993	0,00001	0,004	0,015	0,0000005	0,001	10,919
nC13	69	157	153	0,595	0,997	0,00001	0,003	0,042	0,0000026	0,002	31,099
nC15	80	64	62	0,595	0,846	0,00161	0,040	0,012	0,0000028	0,002	9,252
nC16	74	130	124	0,605	0,998	0,00001	0,003	0,012	0,0000006	0,001	8,683
nC19	73	163	153	0,594	0,995	0,00001	0,004	0,023	0,0000004	0,001	17,419
nC25	37	11	11	0,620	0,725	0,0057	0,075	0,002	0,0000001	0,000	1,144
nG02	65	126	121	0,619	0,994	0,00002	0,004	0,022	0,0000008	0,001	16,011
nG05	62	117	110	0,603	0,991	0,00003	0,005	0,025	0,0000063	0,003	18,380
nG06	61	164	155	0,606	0,999	0,00001	0,003	0,026	0,0000019	0,001	19,530
nG07	48	135	133	0,602	0,996	0,00003	0,005	0,036	0,0000021	0,001	26,935
nG09	58	99	97	0,599	0,984	0,00005	0,007	0,023	0,0000022	0,001	16,792
nG16	57	2	2	0,593	0,070	0,00213	0,046	0,000	0	0,000	0,070
nG17	68	112	107	0,621	0,993	0,00001	0,004	0,015	0,0000016	0,001	11,490
nG26	84	5	5	0,614	0,116	0,0023	0,048	0,000	0	0,000	0,119

Tabla 3. Medidas de polimorfismo correspondiente a la región E1-E2, para los individuos 1a.

Individuo 1a	N° Sec.	Eta.	S	G+C	Hd.	V(Hd).	Sd(Hd)	π	V(π)	Sd(π)	Π
eA05	100	9	9	0,599	0,418	0,004	0,062	0,002	0,000	0,000	0,722
eA09	100	105	96	0,594	0,976	0,000	0,008	0,012	0,000	0,002	5,837
eA20	112	112	101	0,602	0,988	0,000	0,005	0,024	0,000	0,002	11,344
eA26	100	120	104	0,598	0,994	0,000	0,003	0,034	0,000	0,002	16,141
eA27	100	67	63	0,603	0,970	0,000	0,007	0,021	0,000	0,002	9,815
eA34	100	126	118	0,593	0,997	0,000	0,002	0,029	0,000	0,001	13,792
eC01	100	106	95	0,600	0,955	0,000	0,016	0,024	0,000	0,002	11,441
eC02	102	13	13	0,606	0,656	0,003	0,053	0,002	0,000	0,000	1,053
eC06	100	23	23	0,582	0,439	0,004	0,064	0,001	0,000	0,000	0,558
eC07	101	110	98	0,603	0,995	0,000	0,003	0,033	0,000	0,002	15,513
eC11	110	91	85	0,621	0,894	0,001	0,028	0,009	0,000	0,001	4,435
eC14	101	16	16	0,602	0,738	0,001	0,038	0,006	0,000	0,001	2,684
eC17	106	42	41	0,598	0,887	0,000	0,022	0,008	0,000	0,001	3,772
eC18	100	78	76	0,596	0,879	0,001	0,029	0,041	0,000	0,004	19,187
eC20	101	83	75	0,592	0,965	0,000	0,011	0,036	0,000	0,002	17,178
eC21	100	63	58	0,603	0,971	0,000	0,008	0,030	0,000	0,002	14,354
eC22	100	103	91	0,597	0,993	0,000	0,003	0,043	0,000	0,001	20,345
eC24	100	25	25	0,601	0,945	0,000	0,012	0,007	0,000	0,000	3,130
eC28	100	126	115	0,605	0,994	0,000	0,002	0,044	0,000	0,003	20,549
eC31	100	76	72	0,602	0,796	0,002	0,043	0,012	0,000	0,002	5,786
eC33	100	16	16	0,598	0,762	0,002	0,043	0,008	0,000	0,000	3,899
eC38	100	32	31	0,601	0,818	0,002	0,039	0,004	0,000	0,001	2,107
eG10	100	3	3	0,606	0,059	0,001	0,033	0,000	0,000	0,000	0,060
eG14	100	106	102	0,599	0,974	0,000	0,007	0,026	0,000	0,003	12,076
eG19	100	82	80	0,585	0,975	0,000	0,007	0,028	0,000	0,003	13,053

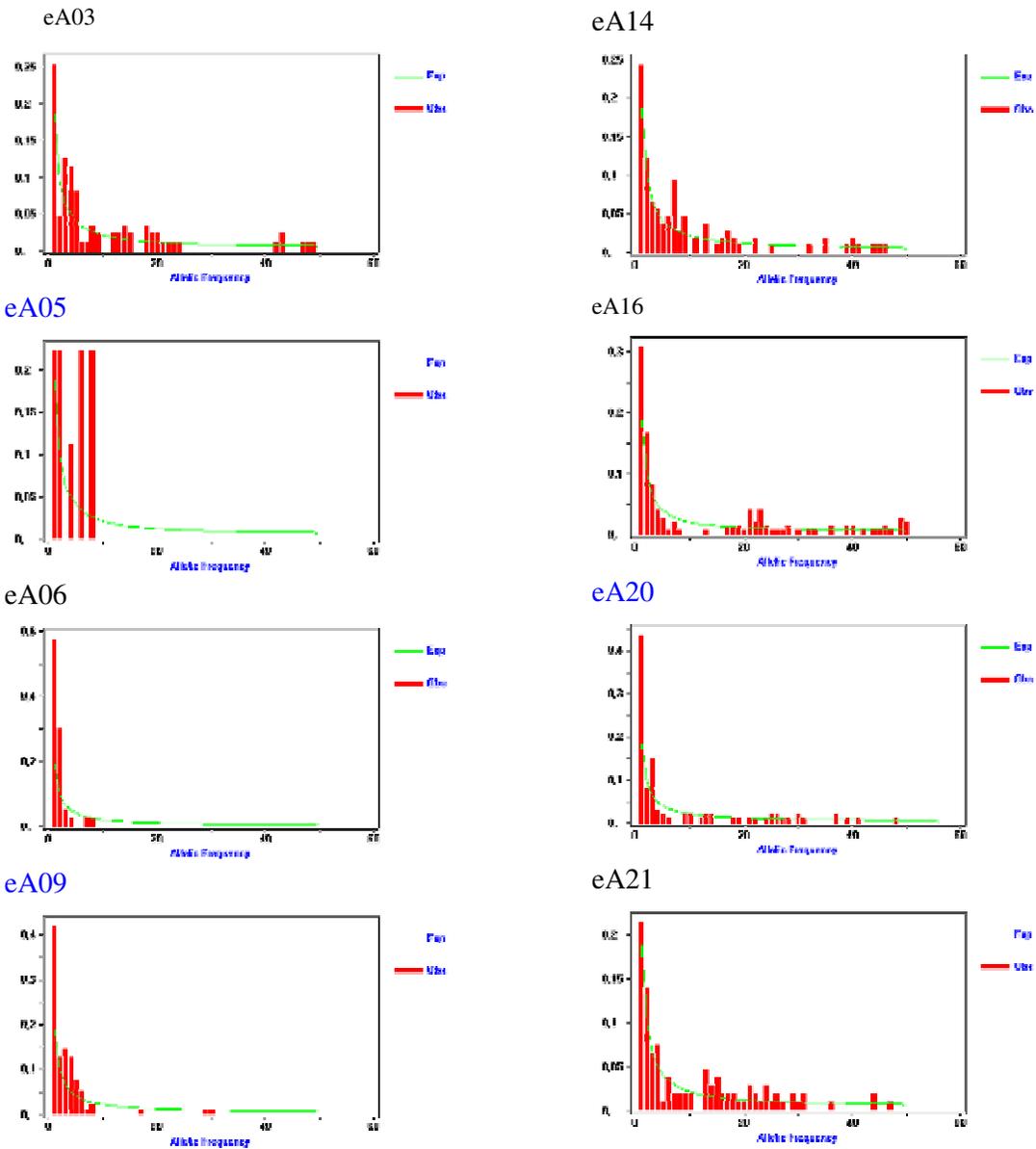
Tabla 4. Medidas de polimorfismo correspondiente a la región NS5A, para los individuos 1a.

Individuo 1a	N° Sec.	Eta.	S	G+C	Hd.	V(Hd).	Sd(Hd)	π	V(π)	Sd(π)	Π
nA05	25	15	15	0,607	0,870	0,004	0,062	0,003	0,000	0,000	2,080
nA09	32	53	52	0,607	0,990	0,000	0,011	0,012	0,000	0,002	9,147
nA20	67	34	34	0,603	0,928	0,000	0,019	0,010	0,000	0,001	7,223
nA26	62	108	106	0,596	0,990	0,000	0,005	0,019	0,000	0,001	14,274
nA27	73	86	81	0,607	0,979	0,000	0,009	0,013	0,000	0,001	9,332
nA34	29	61	59	0,601	1,000	0,000	0,009	0,013	0,000	0,001	9,724
nC01	65	131	125	0,599	0,977	0,000	0,013	0,028	0,000	0,005	20,808
nC02	96	17	17	0,597	0,657	0,002	0,045	0,002	0,000	0,000	1,339
nC06	76	4	4	0,600	0,295	0,004	0,061	0,001	0,000	0,000	0,592
nC07	63	100	97	0,602	0,973	0,000	0,013	0,017	0,000	0,001	12,627
nC11	92	18	18	0,600	0,318	0,004	0,064	0,003	0,000	0,001	2,486
nC14	29	12	12	0,615	0,857	0,002	0,048	0,003	0,000	0,000	2,414
nC17	25	6	6	0,616	0,367	0,015	0,122	0,001	0,000	0,000	0,553
nC18	78	7	7	0,602	0,076	0,002	0,041	0,000	0,000	0,000	0,179
nC20	91	28	28	0,598	0,863	0,001	0,031	0,005	0,000	0,000	3,489
nC21	80	86	84	0,598	0,983	0,000	0,009	0,006	0,000	0,000	4,756
nC22	44	108	98	0,597	0,998	0,000	0,005	0,019	0,000	0,001	13,849
nC24	50	25	25	0,600	0,940	0,000	0,020	0,003	0,000	0,000	2,398
nC28	49	91	87	0,602	0,991	0,000	0,007	0,015	0,000	0,001	11,273
nC31	73	31	30	0,602	0,413	0,006	0,074	0,003	0,000	0,001	2,427
nC33	74	3	3	0,609	0,080	0,002	0,043	0,000	0,000	0,000	0,081
nC38	90	4	4	0,615	0,066	0,001	0,036	0,000	0,000	0,000	0,089
nG10	28	7	7	0,599	0,542	0,012	0,112	0,001	0,000	0,000	1,093
nG14	84	92	91	0,600	0,986	0,000	0,005	0,011	0,000	0,000	8,022
nG19	52	65	63	0,610	0,995	0,000	0,005	0,011	0,000	0,001	7,939

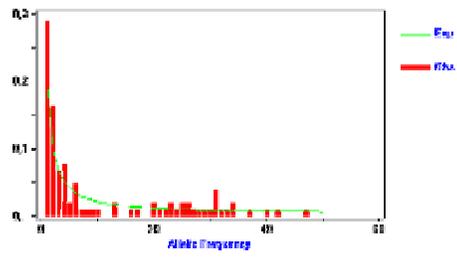
ANEXO III

Espectro de frecuencias de sitios segregantes.

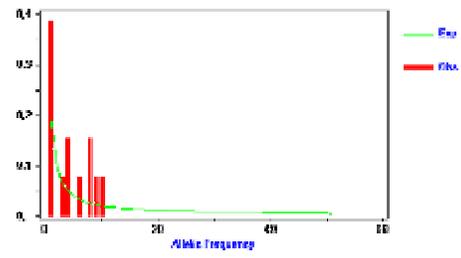
En este anexo se muestra el espectro de frecuencias de sitios segregantes (la frecuencia de alelos en cada sitio polimórfico), correspondiente a ambas regiones (E1-E2, NS5A) y para cada uno de los individuos de genotipo 1a (nombre en azul) y 1b. En el eje de coordenadas se representa el número de mutaciones por sitio polimórfico, y en el de ordenadas el porcentaje de sitios polimórficos. La línea continua corresponde a la curva esperada bajo neutralidad y las barras hacen referencia a los datos observados.



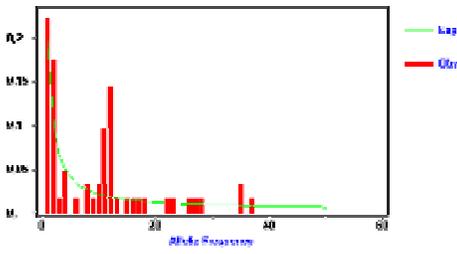
eA26



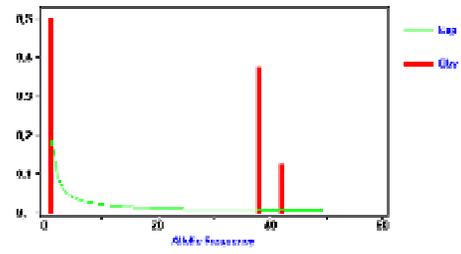
eC02



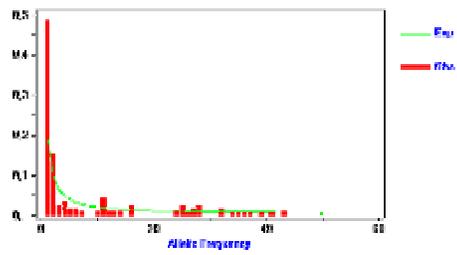
eA27



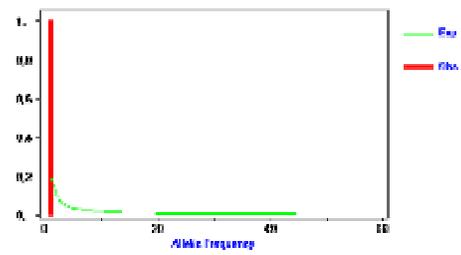
eC03



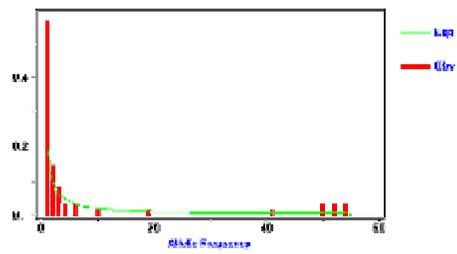
eA34



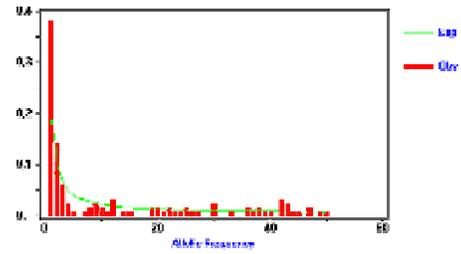
Ec04



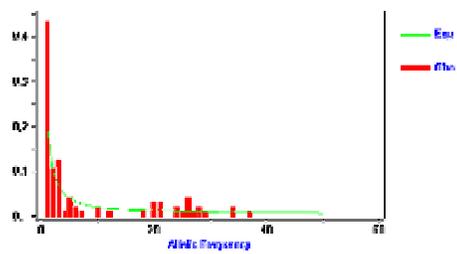
eA35



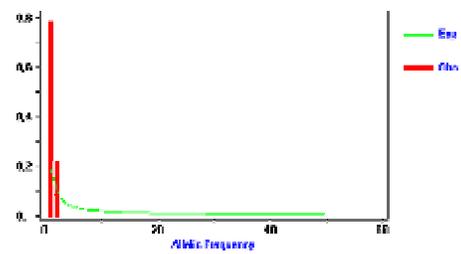
eC05



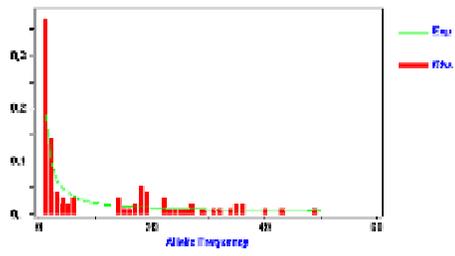
eC01



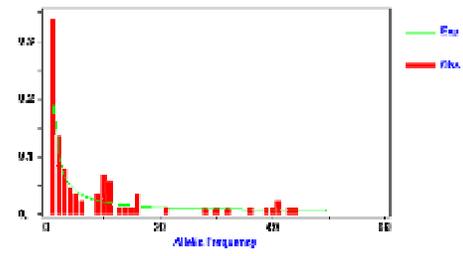
eC06



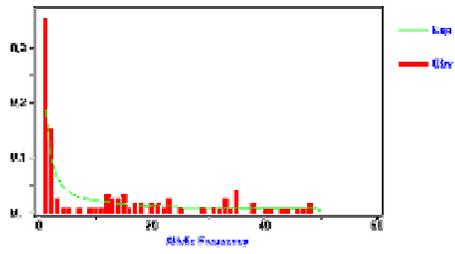
eC07



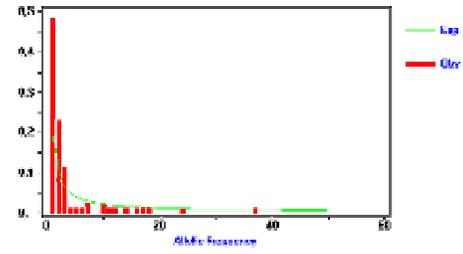
eC12



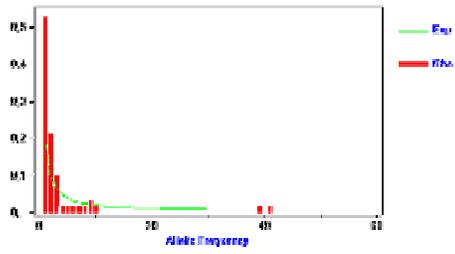
eC08



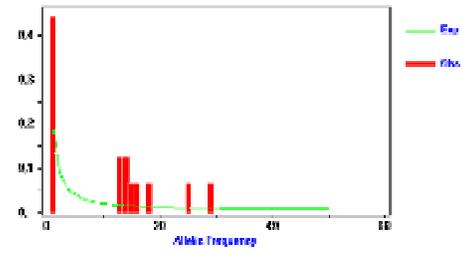
eC13



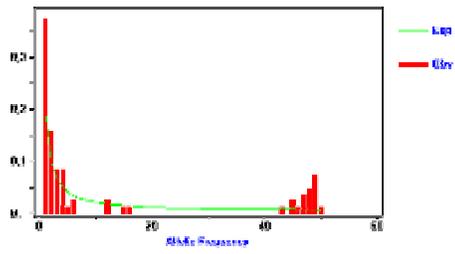
eC09



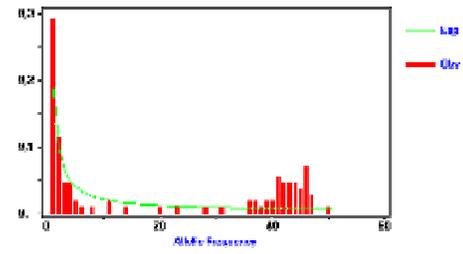
eC14



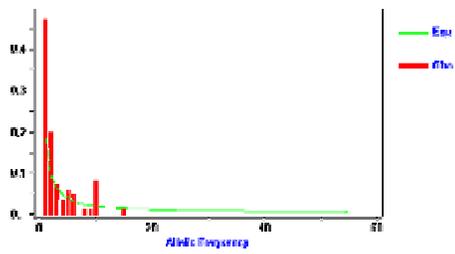
eC10



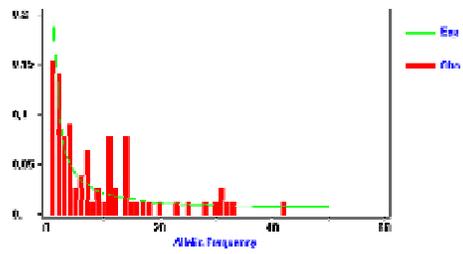
eC15



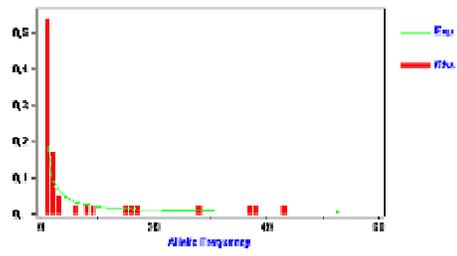
eC11



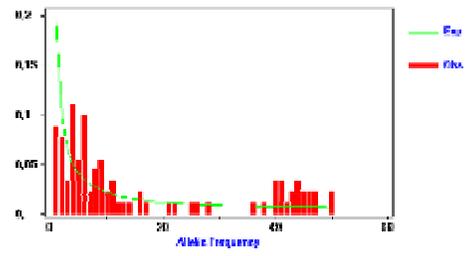
eC16



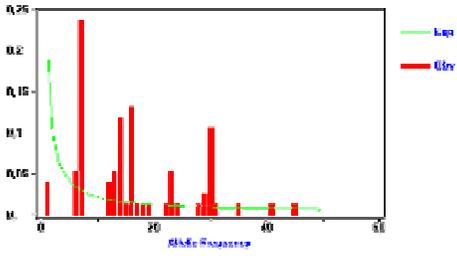
eC17



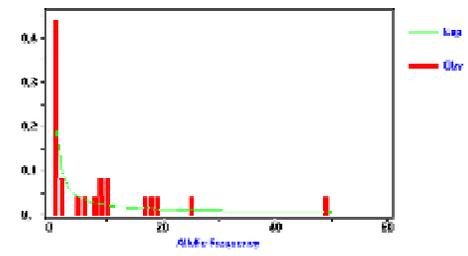
eC22



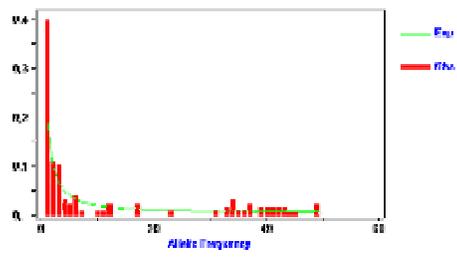
eC18



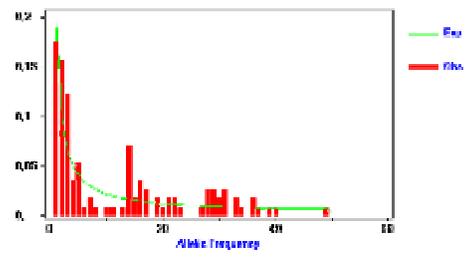
eC25



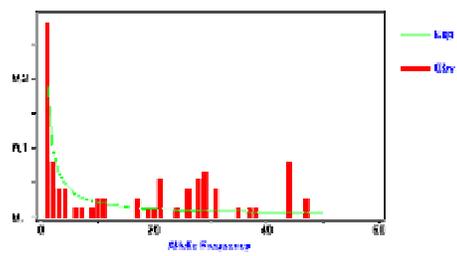
eC19



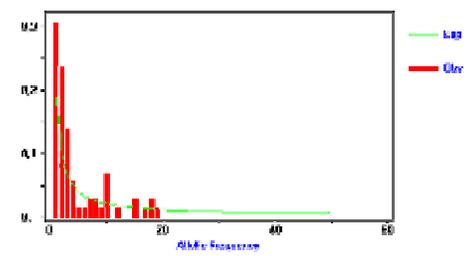
eC28



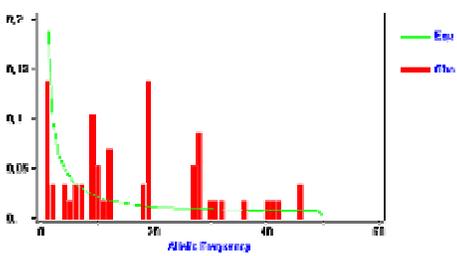
eC20



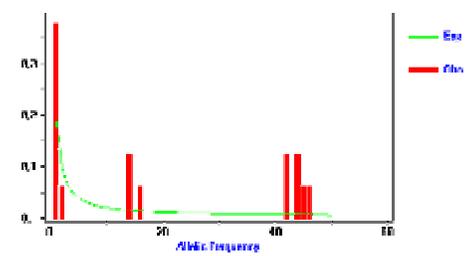
eC31



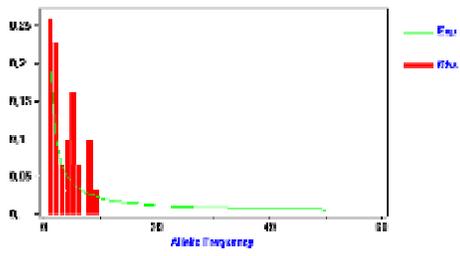
eC21



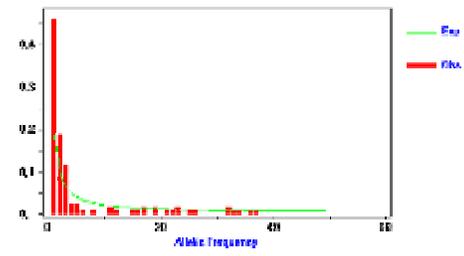
eC33



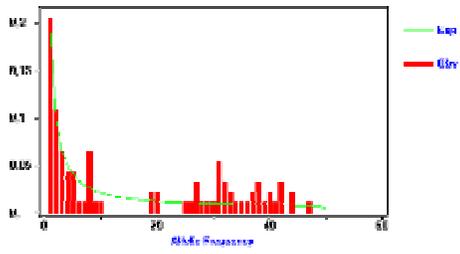
eC38



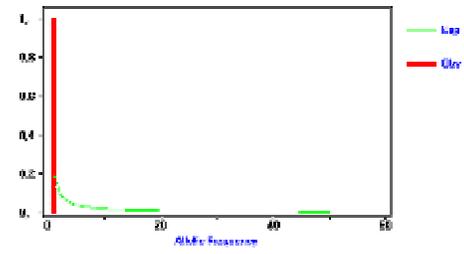
eG09



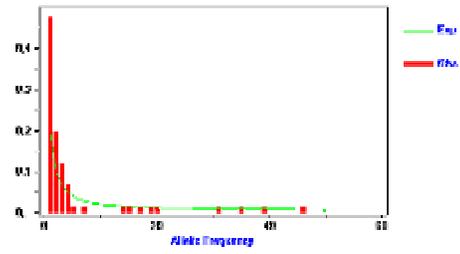
eG02



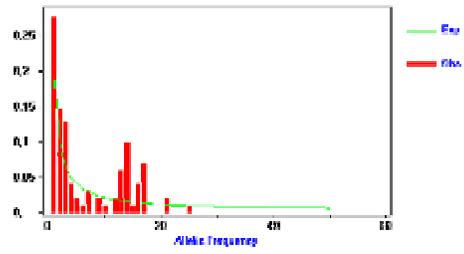
eG10



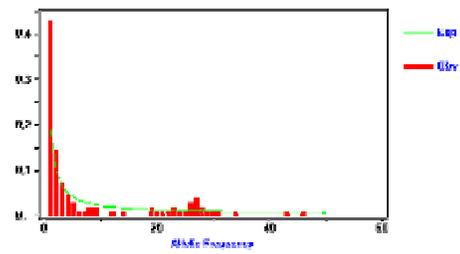
eG05



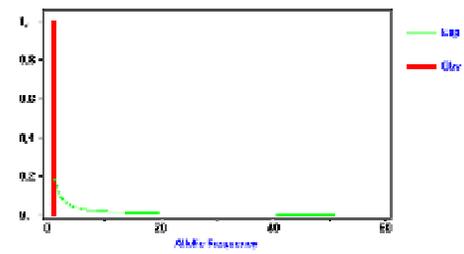
eG14



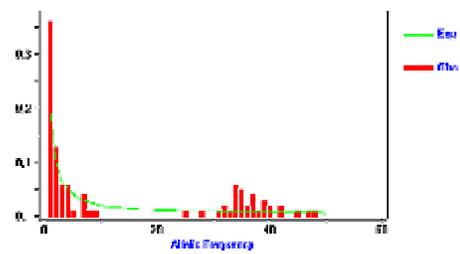
eG06



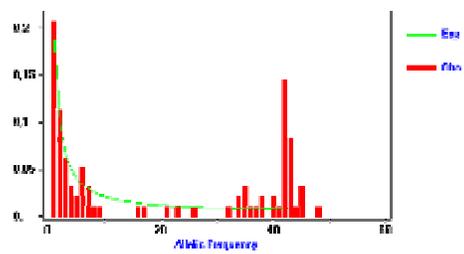
eG16



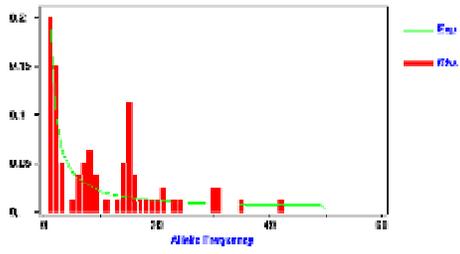
eG07



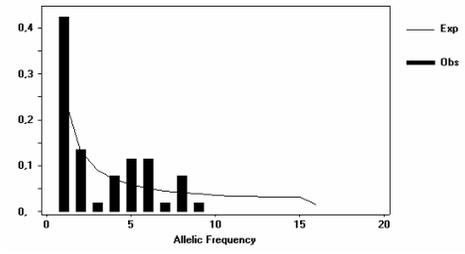
eG17



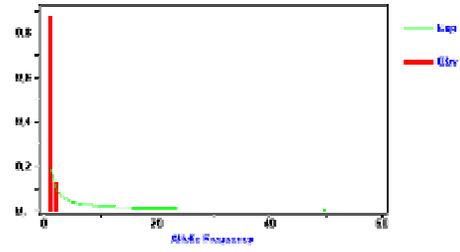
eG19



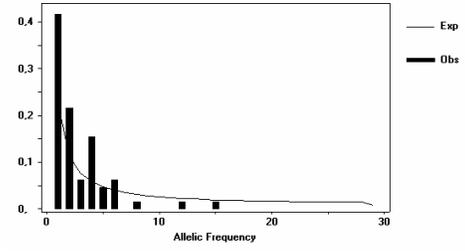
nA09



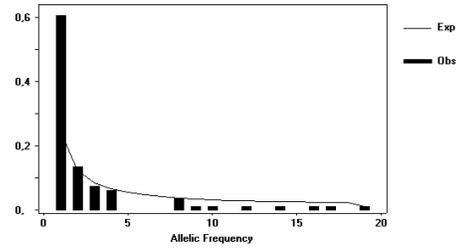
eG26



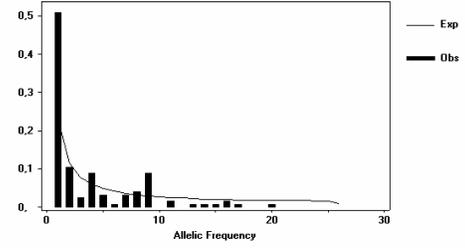
nA14



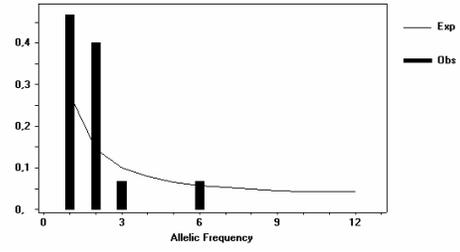
nA03



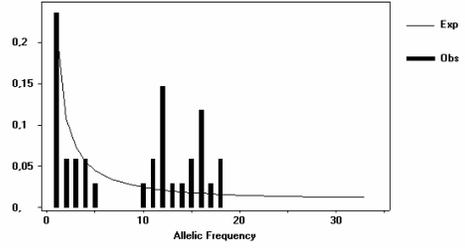
nA16



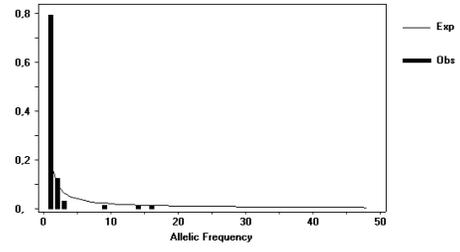
nA05



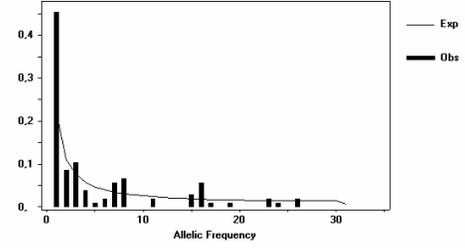
nA20



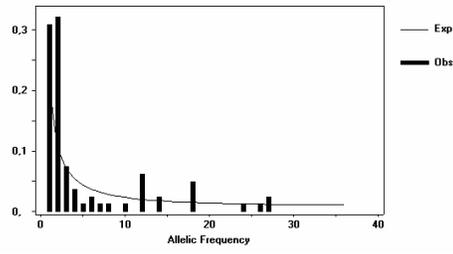
nA06



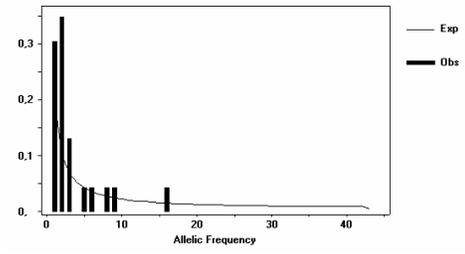
nA26



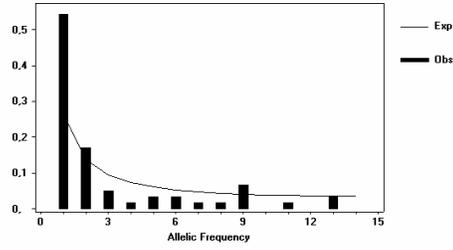
nA27



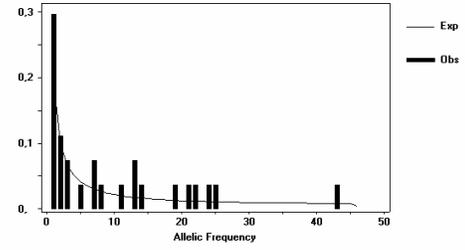
nC03



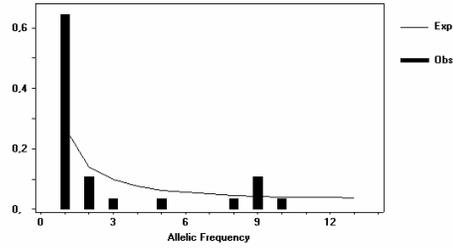
nA34



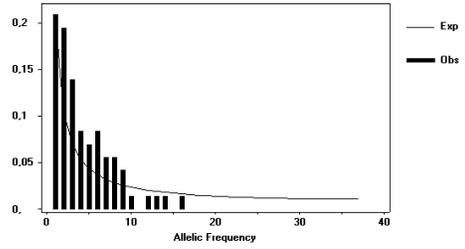
nC04



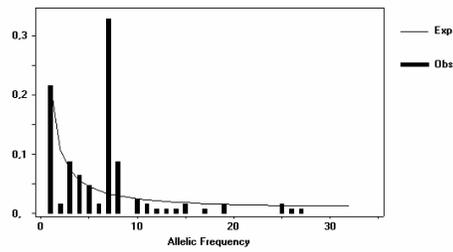
nA35



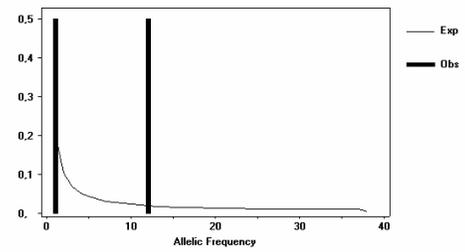
nC05



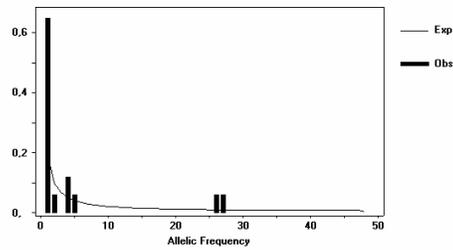
nC01



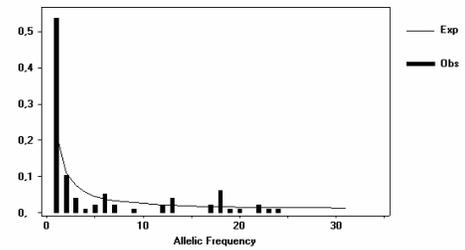
nC06



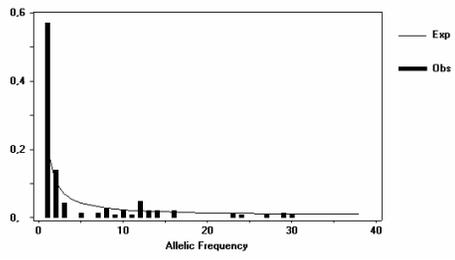
nC02



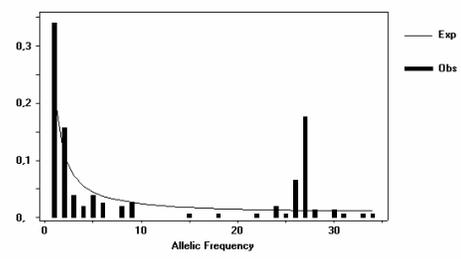
nC07



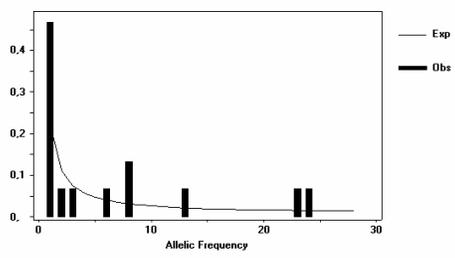
nC08



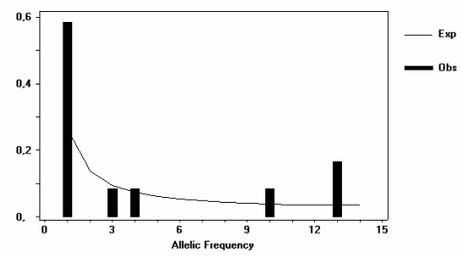
nC13



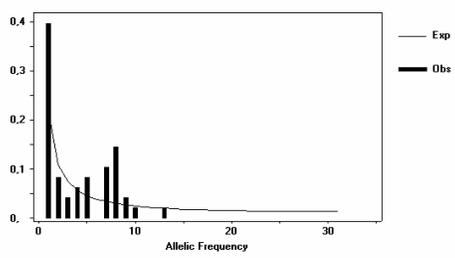
nC09



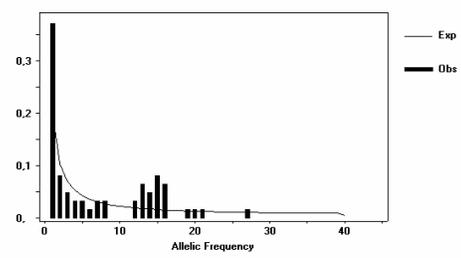
nC14



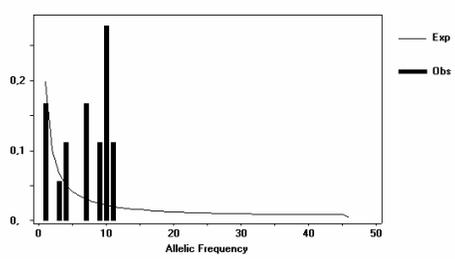
nC10



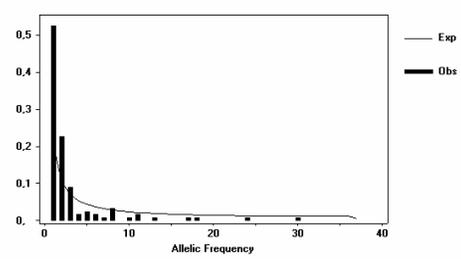
nC15



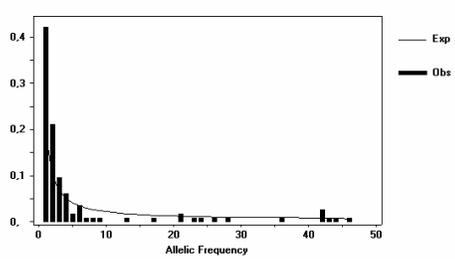
nC11



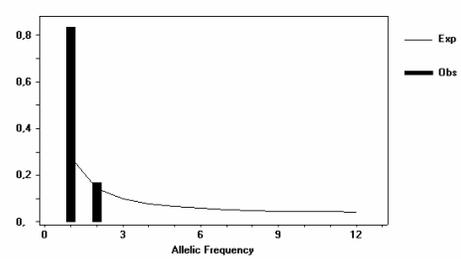
nC16



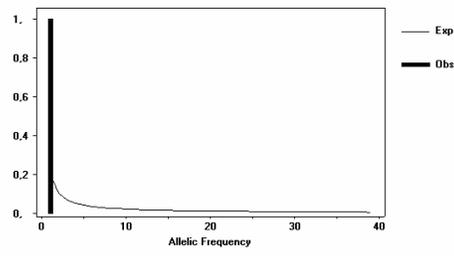
nC12



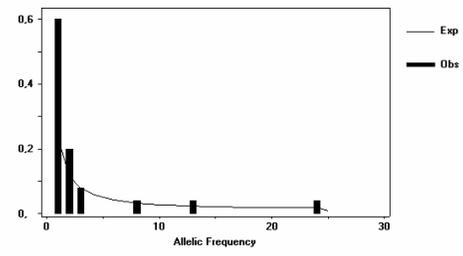
nC17



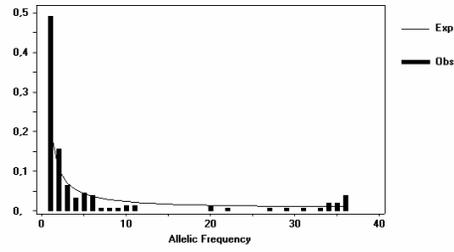
nC18



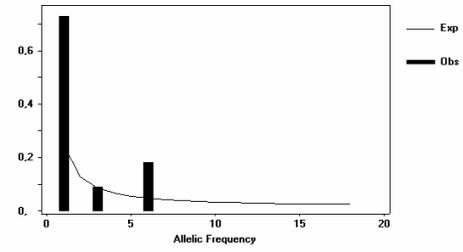
nC24



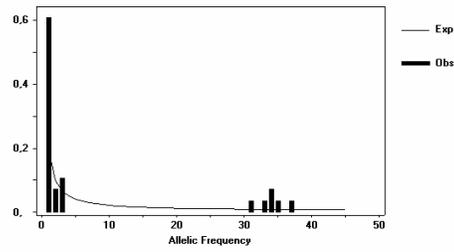
nC19



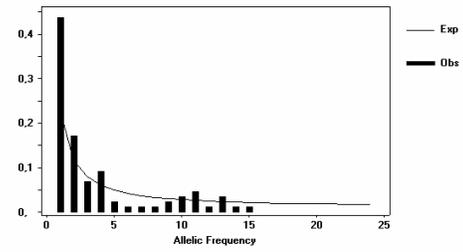
nC25



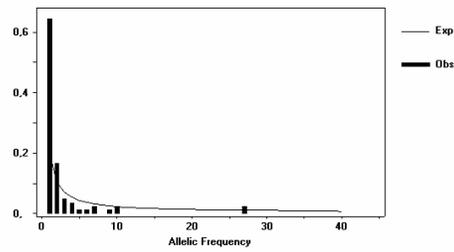
nC20



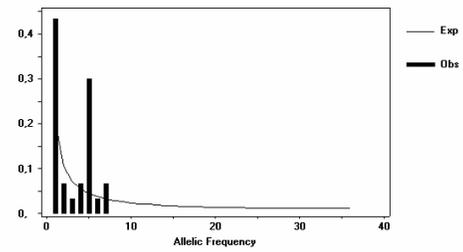
nC28



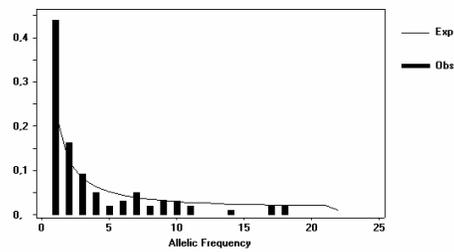
nC21



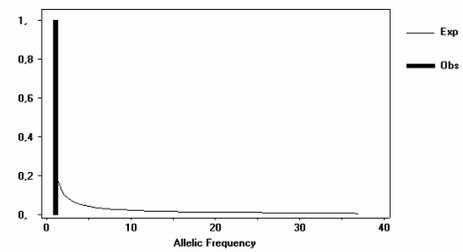
nC31



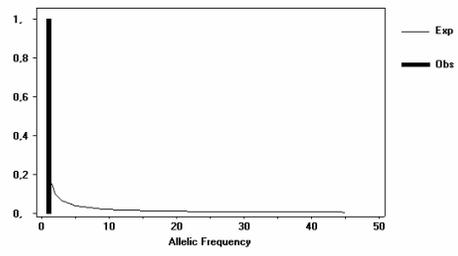
nC22



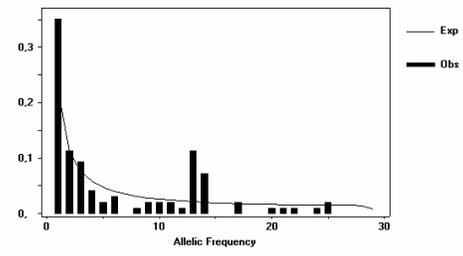
nC33



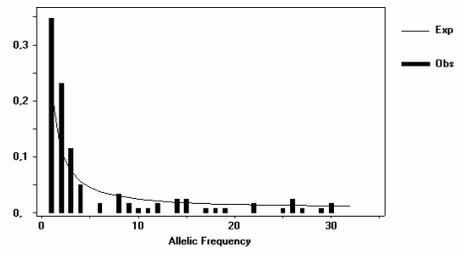
nC38



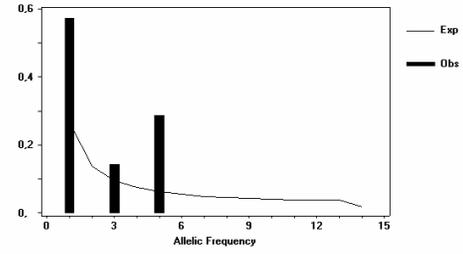
nG09



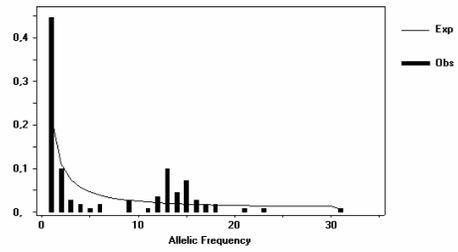
nG02



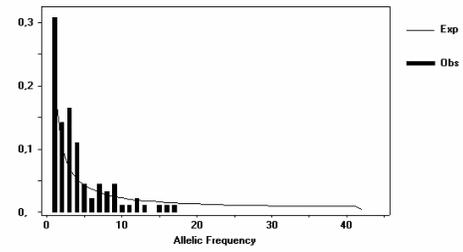
nG10



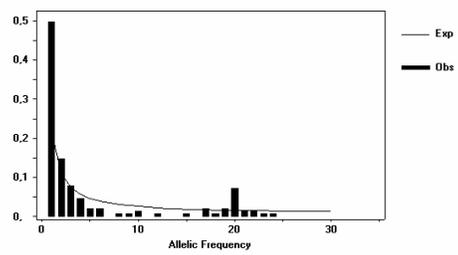
nG05



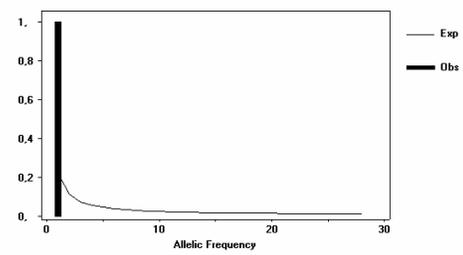
nG14



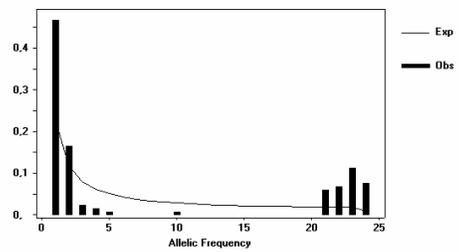
nG06



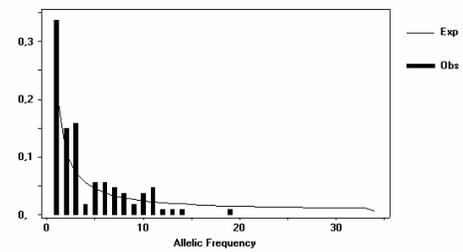
nG16



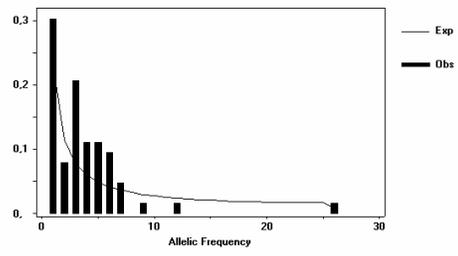
nG07



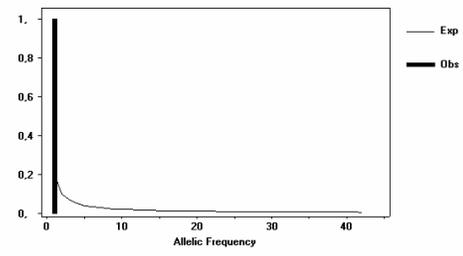
nG17



nG19



nG26



ANEXO IV

Comparación de las medias de los parámetros empleados en el cálculo de la variación genética.

La comparación de las medias de los diferentes parámetros empleados para calcular la variación genética, se realizó utilizando el test de la t siempre y cuando los datos se distribuyeran conforme a una normal. Para comprobar esta distribución llevamos a cabo previamente el test de Kolmogorov-Smirnov. Si la significación del estadístico es menor del 0.05 se rechaza la hipótesis nula (H_0 : Los datos se distribuyen conforme a una normal). En el caso de que se rechace H_0 , para la comparación de las medias se utilizará la prueba de Mann-Whitney.

Tabla 5. Prueba de Kolmogorov-Smirnov para los parámetros **S**: N° de sitios segregantes, **Eta**: N° total de mutaciones, **Hd**: Diversidad haplotípica, π : Diversidad nucleotídica, **II**: N° medio de diferencias nucleotídicas emparejadas. Datos pertenecientes a individuos de genotipo 1a, región E1-E2.

		S	ETA	HD	π	II
N		25	25	25	25	25
Parámetros normales	Media	64,24	69,32	0,8415	0,0194	9,1532
	Desviación típica	37,4369	41,6811	0,2308	0,0144	6,8143
Diferencias más extremas	Absoluta	0,1421	0,1505	0,2502	0,1667	0,1668
	Positiva	0,1327	0,1362	0,2502	0,1667	0,1668
	Negativa	-0,1421	-0,1505	-0,2445	-0,1060	-0,1061
Z de Kolmogorov-Smirnov		0,7105	0,7523	1,2512	0,8333	0,8338
Sig. asintót. (bilateral)		0,6936	0,6232	0,00873	0,4910	0,4903

N: N° de individuos

Sig. asintót. (bilateral): Significación del test al 95%.

Tabla 6. Prueba de Kolmogorov-Smirnov para los parámetros **S**: N° de sitios segregantes, **Eta**: N° total de mutaciones, **Hd**: Diversidad haplotípica, π : Diversidad nucleotídica, **II**: N° medio de diferencias nucleotídicas emparejadas. Datos pertenecientes a individuos de genotipo 1b, región E1-E2.

		S	ETA	HD	π	II
N		25	25	25	25	25
Parámetros normales	Media	78,76	85,92	0,8242	0,0247	11,6808
	Desviación típica	44,0514	49,3381	0,3203	0,0189	8,9364
Diferencias más extremas	Absoluta	0,1550	0,1483	0,4127	0,1431	0,1433
	Positiva	0,1434	0,1405	0,2927	0,1431	0,1433
	Negativa	-0,1550	-0,1483	-0,4127	-0,1081	-0,1080
Z de Kolmogorov-Smirnov		0,7751	0,7417	2,0635	0,7154	0,7163
Sig. asintót. (bilateral)		0,5851	0,6412	0,0004	0,6854	0,6839

N: N° de individuos

Sig. asintót. (bilateral): Significación del test al 95%.

Tabla 7. Prueba de Kolmogorov-Smirnov para los parámetros **S**: N° de sitios segregantes, **Eta**: N° total de mutaciones, **Hd**: Diversidad haplotípica, π : Diversidad nucleotídica, **II**: N° medio de diferencias nucleotídicas emparejadas. Datos pertenecientes a individuos de genotipo 1a, región NS5A.

		S	ETA	HD	π	II
N		25	25	25	25	25
Parámetros normales	Media	46,12	47,68	0,7254	0,0080	5,9277
	Desviación típica	39,2326	41,1448	0,3409	0,0075	5,5823
Diferencias más extremas	Absoluta	0,1813	0,1902	0,2903	0,2113	0,2113
	Positiva	0,1813	0,1902	0,2102	0,2113	0,2113
	Negativa	-0,1359	-0,1442	-0,2903	-0,1474	-0,1475
Z de Kolmogorov-Smirnov		0,9066	0,9512	1,4515	1,0564	1,0563
Sig. asintót. (bilateral)		0,3837	0,3260	0,0296	0,2144	0,2145

N: N° de individuos

Sig. asintót. (bilateral): Significación del test al 95%.

Tabla 8. Prueba de Kolmogorov-Smirnov para los parámetros **S**: N° de sitios segregantes, **Eta**: N° total de mutaciones, **Hd**: Diversidad haplotípica, π : Diversidad nucleotídica, **II**: N° medio de diferencias nucleotídicas emparejadas. Datos pertenecientes a individuos de genotipo 1b, región NS5A.

		S	ETA	HD	π	II
N		25	25	25	25	25
Parámetros normales	Media	81,48	85	0,8801	0,0141	10,4501
	Desviación típica	51,2934	54,1418	0,2487	0,0111	8,2713
Diferencias más extremas	Absoluta	0,1314	0,1310	0,3260	0,1135	0,1134
	Positiva	0,1314	0,1295	0,3163	0,1135	0,1134
	Negativa	-0,1306	-0,1310	-0,3260	-0,1047	-0,1048
Z de Kolmogorov-Smirnov		0,6572	0,6550	1,6301	0,5674	0,5671
Sig. asintót. (bilateral)		0,7808	0,7842	0,0098	0,9043	0,9046

N: N° de individuos

Sig. asintót. (bilateral): Significación del test al 95%.

ANEXO V:

Descripción de los modelos demográficos implementados en GENIE V3.0

Nombre del modelo: Const
 Descripción: Tamaño poblacional constante
 Ecuación: $N(t) = N(0)$
 N° de parámetros: 1

N° de parámetros	Definición	Escala de tiempo en generaciones	Escala de tiempo en sustituciones
0	Tamaño poblacional actual en el presente	$N(0)$	$\theta = N(0)\mu$

Nombre del modelo: Expo
 Descripción: Crecimiento exponencial
 Ecuación: $N(t) = N(0) \exp(-rt)$
 N° de parámetros: 2

N° de parámetros	Definición	Escala de tiempo en generaciones	Escala de tiempo en sustituciones
0	Tamaño poblacional en el presente	$N(0)$	$\theta = N(0)\mu$
1	Tasa de Crecimiento Exponencial	r	$\rho = r / \mu$

Nombre del modelo: Expan
 Descripción: Crecimiento de expansión
 Ecuación: $N(t) = N(0) [\alpha + (1-\alpha) \exp(-rt)]$
 N° de parámetros: 3

N° de parámetros	Definición	Escala de tiempo en generaciones	Escala de tiempo en sustituciones
0	Tamaño poblacional en el presente	$N(0)$	$\theta = N(0)\mu$
1	Tasa de crecimiento exponencial	r	$\rho = r / \mu$
2	Tamaño poblacional a $t=\infty$ como proporción de $N(0)$	α (alpha)	α

Nota: α no se ve afectado al cambiar la escala de tiempo

Nombre del modelo: Log
 Descripción: Crecimiento logístico
 Ecuación: $N(t) = N(0) [(1+c) / (1 + c \exp (rt))]$
 N° de parámetros: 3

N° de parámetros	Definición	Escala de tiempo en generaciones	Escala de tiempo en sustituciones
0	Tamaño poblacional en el presente	$N(0)$	$N(0)\mu$
1	Tasa de crecimiento exponencial	r	$\rho = r / \mu$
2	Parámetro de forma logística	c	c

Nota: c no se ve afectado al cambiar la escala de tiempo.

Nota: Siendo K la capacidad de carga de la población , entonces $c=[K/N(0)]-1$.

Nombre del modelo: Step
 Descripción: Crecimiento constante-exponencial-constante
 Ecuación: $N(t) = \begin{cases} N(0) & \text{si } t < x \\ N(0)f & \text{de lo contrario} \end{cases}$
 N° de parámetros: 3

N° de parámetros	Definición	Escala de tiempo en generaciones	Escala de tiempo en sustituciones
0	Tamaño poblacional en el presente	$N(0)$	$N(0)\mu$
1	Tiempo de cambio	x	$z = x / \mu$
2	Tamaño poblacional anterior al cambio, como proporción de $N(0)$	f	f

Nota: f no se ve afectado al cambiar la escala de tiempo

Nombre del modelo: Pexpan
 Descripción: Crecimiento expansión escalonado
 Ecuación: $N(t) = \begin{cases} N(0) \exp (-rt) & \text{si } t < -\log(\alpha)/r \\ N(0) \alpha & \text{de lo contrario} \end{cases}$

N° de parámetros: 3

N° de parámetros	Definición	Escala de tiempo en generaciones	Escala de tiempo en sustituciones
0	Tamaño poblacional en el presente	$N(0)$	$\theta = N(0)\mu$
1	Tasa de Crecimiento Exponencial	r	$\rho = r / \mu$
2	Tamaño poblacional anterior al cambio, como proporción de $N(0)$	f	f

Nota: α no se ve afectado al cambiar la escala de tiempo. El tiempo de cambio es $-\log(\alpha)/r$

Nombre del modelo:	PLog
Descripción:	Crecimiento logístico escalonado
Ecuación:	$N(t) = \begin{cases} N(0) & \text{si } t < x \\ N(0) \exp[-r(t-x)] & \text{de lo contrario} \end{cases}$
Nº de parámetros:	3

Nº de parámetros	Definición	Escala de tiempo en generaciones	Escala de tiempo en sustituciones
0	Tamaño poblacional en el presente: c	$N(0)$	$\theta = N(0)\mu$
1	Tasa de crecimiento exponencial	r	$\rho = r / \mu$
2	Tasa de cambio	x	x

Nota: x no se ve afectado al cambiar la escala de tiempo

Cálculo por máxima verosimilitud de los parámetros correspondientes a cada modelo demográfico. Estima paramétrica.

1. Análisis intrapaciente.

Tabla 9. Estima de los parámetros implicados en los diferentes modelos demográficos, para la región E1-E2 correspondiente a los individuos de genotipo 1b. También se detalla el valor umbral que agrupa los intervalos de cada internodo en el árbol filogenético (maxepsilon), el logaritmo de la verosimilitud (ln L), y la corrección del lnL según el criterio de información de Akaike corregido (AICc), el cual usamos para hacer la comparación entre los diferentes modelos. Con sombreado se señala el modelo que mejor se adapta a los datos según el AICc.

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f
eA03	Const	0,003	708,987	707,966	2,12E-01				
	Expo	0,003	730,249	728,186	5,07E-01	1,31E+02			
	Expan	0,003	730,249	727,122	5,07E-01	1,31E+02	2,75 E-16		
	Log	0,003	747,337	744,211	3,07E-01	4,02E+02		6,5E-03	
	Step								
	PExpan	0,003	730,249	727,122	5,07E-01	1,31E+02	1,2E-03		
	PLog	0,003	746,839	743,713	0,291014	318,427	1,1E-02		
eA14	Const	2,27E-03	706,562	705,541	0,217				
	Expo	2,27E-03	727,133	725,07	0,341	74,3114			
	Expan	2,27E-03	729,859	726,733	0,590	254,598	9,3E-02		
	Log	2,27E-03	727,133	724,006	0,341	74,3114		1,0E+04	
	Step	2,27E-03	706,562	703,435	0,217				1,4E-08 1,0E-01
	PExpan	2,27E-03	730,536	727,41	0,55454	189	1,2E-01		
	PLog	2,27E-03	727,510	724,384	0,322	74,9878	8,9E-04		
	PLog	8,51E-03	619,977	616,851	1,373	39,519	0,003		
eA16	Const	8,51E-03	588,773	587,752	0,713				
	Expo	8,51E-03	619,643	617,581	1,500	39,194			
	Expan	8,51E-03	619,643	616,517	1,500	39,1941	2,24e-15		
	Log	8,51E-03	588,773	586	0,713	9,87e-07		1,28e-25	
	Step	8,51E-03							
	PExpan	8,51E-03	619,643	616,517	1,500	39,194	0,003		
	PLog	8,51E-03	619,977	616,851	1,373	39,519	0,003		

θ : $N(0)\mu$; ρ : r/μ ; α : Tamaño poblacional anterior al cambio, como proporción de $N(0)$; c: Parámetro de forma logística; f: Tamaño poblacional anterior al cambio, como proporción de $N(0)$; x: Tiempo de cambio.

Tabla 9. (Continuación).

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
eC05	Const	0,005	633,474	632,453	0,454					
	Expo	0,005	660,826	658,763	0,706	44,2861				
	Expan	0,005	676,618	673,491	2,294	208,972	0,0375			
	Log	0,005	660,826	657,7	0,706	44,2861		1,00E+05		
	Step	0,005	633,474	630,348	0,454				5,81E-06	0,13993
	PExpan	0,005	675,713	672,586	2779,000	219,041	0,0440			
	PLog	0,005	660,952	657,825	0,679	44,3646	0,0009			
eC08	Const	0,007	669,187	668,167	0,316					
	Expo	0,007	690,897	688,834	0,595	64,686				
	Expan	0,007	690,897	687,771	0,595	64,686	3,85E-13			
	Log	0,007	691,900	688,774	0,501	86,723		0,560		
	Step	0,007	669,187	666,061	0,316				5,81E-06	4,14987
	PExpan	0,007	690,897	687,771	0,595	64,686	0,003			
	PLog	0,007	693,419	690,293	0,434	76,907	0,008			
eC10	Const	0,002	776,721	775,7	0,107					
	Expo	0,002	798,234	796,171	0,187	180,058				
	Expan	0,002	799,039	795,913	0,218	241,406	0,022			
	Log	0,002	798,234	795,107	0,187	180,058		7,00E+11		
	Step	0,002	776,721	773,594	0,107				1,40E-08	0,1
	PExpan	0,002	799,005	795,878	0,232	246,656	0,047			
	PLog	0,002	800,742	797,616	0,156	190,373	0,001			
eC12	Const	0,002	709,399	708,378	0,211					
	Expo	0,002	737,56	735,498	0,373	108,136				
	Expan	0,002	741,118	737,991	0,591	226,955	0,042			
	Log	0,002	737,56	734,434	0,373	108,136		1,00E+04		
	Step	0,002	709,399	706,272	0,211				5,81E-06	4,15E+00
	PExpan	0,002	741,168	738,042	0,699	231,006	0,057			
	PLog	0,002	738,273	735,146	0,338	101,398	0,0010565			
eC13	Const	2,27E-05	793,364	792,343	0,090					
	Expo	2,27E-05	831,144	829,081	0,308	512,996				
	Expan	2,27E-05	831,144	828,017	0,308	512,996	2,9e-17			
	Log	2,27E-05	841,552	838,426	0,164	962,379		0,044074		
	Step	2,27E-05	793,364	790,238	0,090				1,40E-08	0,0241
	PExpan	2,27E-05	831,144	828,017	0,308	512,996	0,0031677			
	PLog	2,27E-05	842,366	839,24	0,173	671,391	0,0018874			
eC15	Const	3,89E-03	652,245	651,225	0,444					
	Expo	3,89E-03	668,29	666,229	0,595	24,594				
	Expan	3,89E-03	679,135	676,011	1,490	170,601	0,086			
	Log	3,89E-03	668,29	665,166	0,595	24,594		1,00E+04		
	Step	3,89E-03	652,245	649,121	0,444				1,40E-08	17,2214
	PExpan	3,89E-03	679,567	676	1,351	122428,000	0,109			
	PLog	3,89E-03	668,369	665,245	0,583	24,641	0,001			
eC16	Const	8,10E-03	791,937	790,916	0,101					
	Expo	8,10E-03	795,606	793,544	0,125	42,146				
	Expan	8,10E-03	795,606	792,481	0,125	42,146	3,5e-15			
	Log	8,10E-03	798,193	795,068	0,110	52,767		5,15E-06		
	Step	8,10E-03	791,937	788,812	0,101				5,81E-06	4,150
	PExpan	8,10E-03	795,606	792,481	0,125	42,146	0,0267472			
	PLog	8,10E-03	798,532	795,407	0,114	63,944	0,0062749			

Tabla 9. (Continuación).

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
eG02	Const	0,0009	730,652	729,631	0,170					
	Expo	0,0009	734,513	732,451	0,208	23,973				
	Expan	0,0009	734,513	731,387	0,208	23,973	1,89E-12			
	Log	0,0009	735,465	732,339	0,189	53,530		0,151		
	Step	0,0009	730,652	727,525					0,170	1,40E-08
	PExpan	0,0009	734,513	731,387	0,208	23,973	0,007			
	PLog	0,0009	730,652	727,525	0,191	883,573	3,8E+12			
eC19	Const	2,99E-03	637,484	636,464	0,436					
	Expo	2,99E-03	686,19	684,127	1,185	103,458				
	Expan	2,99E-03	686,815	683,689	1,349	118,44	0,006			
	Log	2,99E-03	686,19	683,064	1,185	103,458		1,00E+05		
	Step	2,99E-03	637,484	634,358	0,436				1,40E-08	0,1
	PExpan	2,99E-03	668,338	683,212	127866,000	110,724	0,009			
	PLog	2,99E-03	686,508	683,382	1,059	103,981	0,001			
eG05	Const	0,0003	834,037	833,016	0,060					
	Expo	0,0003	842,433	840,370	0,089	158,570				
	Expan	0,0003	842,433	839,307	0,089	158,570	1,54E-13			
	Log	0,0003	845,791	842,664	0,074	342,940		0,115		
	Step	0,0003	834,037	830,910	0,060				0,060	5,81E-06
	PExpan	0,0003	842,433	839,307	0,089	158,570	0,042			
	PLog	0,0003	848,109	844,983	0,078	204,909	0,002			
eG06T0	Const	0,0020	638,690	637,669	0,431					
	Expo	0,0020	700,402	698,339	1,248	130,535				
	Expan	0,0020	708,856	705,730	2,694	232,742	0,010			
	Log	0,0020	700,402	697,276	1,248	130,535		1,36E+13		
	Step	0,0020	638,690	635,563	0,431				5,81E-06	4,150
	PExpan	0,0020	707,434	704,308	190,267	176,402	0,009			
	PLog	0,0020	700,550	697,424	1,138	130,725	0,001			
eG07	Const	0,0007	712,120	711,100	0,205					
	Expo	0,0007	733,674	731,611	0,320	82,079				
	Expan	0,0007	739,733	736,607	0,542	208,888	0,045			
	Log	0,0007	733,674	730,547	0,320	82,079		5,02E+12		
	Step	0,0007	712,120	708,940	0,205				3,37E-11	4,150
	PExpan	0,0007	740,755	737,656	0,545	189,964	0,067			
	PLog	0,0007	734,642	731,515	0,290	83,569	0,001			
eG09	Const	0,0036	658,145	657,125	0,354					
	Expo	0,0036	697,742	695,680	0,849	102,284				
	Expan	0,0036	697,860	694,734	0,901	110,839	0,0051			
	Log	0,0036	697,742	694,616	0,849	102,284		1,00E+04		
	Step	0,0036	658,145	655,019					0,354	5,81E-06
	PExpan	0,0036	697,742	694,616	0,849	102,284	0,0003			
	PLog	0,0036	698,943	695,817	0,704	104,414	0,0022			
eG17	Const	0,0015	707,792	706,771	0,235					
	Expo	0,0015	718,168	716,106	0,282	33,399				
	Expan	0,0015	747,721	744,596	0,946	278,736	0,022			
	Log	0,0015	718,168	715,043	0,282	33,399	1,00E+04			
	Step	0,0015	707,792	704,667	0,235				5,81E-06	1,40E-01
	PExpan	0,0015	750,538	747,413	0,940	259,421	0,038			
	PLog	0,0015	718,339	715,214	0,275	33,565	0,001			

Tabla 10. Estima de los parámetros implicados en los diferentes modelos demográficos, para la región E1-E2 correspondiente a los individuos de genotipo 1a. También se detalla el valor umbral que agrupa los intervalos de cada internodo en el árbol filogenético (maxepsilon), el logaritmo de la verosimilitud (ln L), y la corrección del lnL según el criterio de información de Akaike corregido (AICc), el cual usamos para hacer la comparación entre los diferentes modelos. Con sombreado se señala el modelo que mejor se adapta a los datos según el AICc.

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
eA09	Const	5,64E-06	834,880	833,859	0,059					
	Expo	5,64E-06	838,397	836,335	0,067	45,291				
	Expan	5,64E-06	840,288	837,161	0,092	401,008	2,7E-01			
	Log	5,64E-06	838,397	835,271	0,067	45,291		1,0E+04		
	Step	5,64E-06	834,880	831,753	0,059	1,40E-08	1,0E-01			
	PExpan	5,64E-06	842,616	839,489	0,105	355	2,8E-01			
	PLog	5,64E-06	838,913	835,787	0,065	47,9639	7,6E-04			
eA20	Const	0,0025	872,617	871,599	1,31E-01					
	Expo	0,0025	887,739	885,684	1,77E-01	87,796				
	Expan	0,0025	895,083	891,97	2,96E-01	290,975	0,063			
	Log	0,0025	887,739	884,627	0,177256	8,78E+01		1,00E+05		
	Step	0,0025	872,617	869,505	1,31E-01				5,81E-06	4,150
	PExpan	0,0025	897	893,991	0,294	252	0,089			
	PLog	0,0025	888,989	885,877	0,164214	90,3231	0,001			
eA26	Const	6,14E-03	680,152	679,131	0,283					
	Expo	6,14E-03	700,153	698,091	0,461	57,720				
	Expan	6,14E-03	700,153	697,027	0,462	57,679	2,45E-14			
	Log	6,14E-03	700,153	697,027	0,461	57,720		1,00E+05		
	Step	6,14E-03	680,152	677,025	0,283				5,81E-06	4,150
	PExpan	6,14E-03	700,153	697,027	0,461	57,720	0,0016			
	PLog	6,14E-03	701,028	697,902	0,420	59,065	0,0020			
eA27	Const	1,02E-03	873,665	872,644	0,040					
	Expo	1,02E-03	875,335	873,273	0,045	54,398				
	Expan	1,02E-03	875,335	872,209	0,045	54,398	1,69E-11			
	Log	1,02E-03	877	874	0,041	2228,230		1,00E-30		
	Step	1,02E-03	873,665	870,539	0,040				5,81E-06	4,150
	PExpan	1,02E-03	875,335	872,209	0,045	54,398	0,129			
	PLog	1,02E-03	874	870,539	0,044	4,8e-07	9,84e+10			
eA34	Const	0,00385547	685,815	684,795	2,67E-01					
	Expo	0,00385547	732,101	730,038	6,81E-01	159,338				
	Expan	0,00385547	736,432	733,306	1,12E+00	253,258	0,0121			
	Log	0,00385547	732,101	728,974	6,81E-01	159,338		1,00E+05		
	Step	0,00385547	685,815	682,689	0,26745				3,37E-11	4,14987
	PExpan	0,00385547	736,442	733,316	1,01E+05	225	0,0182			
	PLog	0,00385547	733,617	730,491	0,541593	162,519	0,0017			
eC01	Const	0,001	780,420	779,400	0,103					
	Expo	0,001	786,862	784,799	0,127	50,297				
	Expan	0,001	787,095	783,969	0,146	121,629	0,164			
	Log	0,001	786,862	783,735	0,127	50,297		1,05E+12		
	Step	0,001	780,420	777,294	0,103				3,37E-11	4,15
	PExpan	0,001	788,087	784,960	0,154	112,174	0,222			
	PLog	0,001	787,981	784,855	0,120	53,642	0,002			

θ : $N(0)\mu$; ρ : r/μ ; α : Tamaño poblacional anterior al cambio, como proporción de $N(0)$; c Parámetro de forma logística; f: Tamaño poblacional anterior al cambio, como proporción de $N(0)$; x: Tiempo de cambio.

Tabla 10. (Continuación).

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
eC07	Const	0,002	759,357	758,360	0,140					
	Expo	0,002	779,491	777,429	0,221	114,132				
	Expan	0,002	781,004	777,879	0,277	194,58	0,0490			
	Log	0,002	779,491	776,366	0,221	114,132	1,00E+04			
	Step	0,002	759,357	756,232	0,140				5,81E-06	4,15
	PExpan	0,002	780,693	777,568	0,257	191,823	0,096			
	PLog	0,002	780,507	777,382	0,200	116,843	0,001			
eC18	Const	0,0006	746,059	745,038	0,146					
	Expo	0,0006	746,059	743,996	0,146	1,00E-10				
	Expan	0,0006	746,059	742,932	0,145535	1,10E-10	0,9999			
	Log	0,0006	746,059	742,932	0,146	2,7E-10		3,74E-21		
	Step	0,0006								
	PExpan	0,0006	746,059	742,932	0,145535	1,00E-10	0,930773			
	PLog	0,0006	746,059	742,932	0,146	4,36E-21	1,70E+06			
eC20	Const	0,0013	766,457	765,437	0,130					
	Expo	0,0013	770,729	768,667	0,162	34,311				
	Expan	0,0013	770,729	767,604	0,162	34,311	1,25E-11			
	Log	0,0013	771,684	768,559	0,148	68,992		0,215998		
	Step	0,0013	766,457	763,332	0,130				1,40E-08	0,1
	PExpan	0,0013	770,729	767,604	0,162	34,311	0,004			
	PLog	0,0013	773,389	770,264	0,148	44,599	0,005669			
eC21	Const	0,0008	810,726	809,705	0,076					
	Expo	0,0008	811,012	808,95	0,080	9,187				
	Expan	0,0008	811,012	807,886	0,080	9,187	5,99e-13			
	Log	0,0008	817,035	813,908	0,077	1085,28		1,00E-30		
	Step	0,0008	810,726	807,599	0,076				5,81E-06	4,14987
	PExpan	0,0008	811,012	807,886	0,080	9,187	0,035			
	PLog	0,0008	817,077	813,951	0,080	1674,63	0,065			
eC22	Const	0,0120	670,714	669,693	0,312					
	Expo	0,0120	691,502	689,440	0,599	66,369				
	Expan	0,0120	691,502	688,376	0,599	66,369	1,54E-11			
	Log	0,0120	693,492	690,366	0,481	96,753		3,53E-01		
	Step	0,0120	670,714	667,588	0,312				1,40E-08	1,00E-01
	PExpan	0,0120	691,502	688,376	0,599	66,369	0,005			
	PLog	0,0120	696,147	693,021	0,439	80,540	0,008			
eC24	Const	8,36E-19	949,204	948,184	0,0187					
	Expo	8,36E-19	949,204	947,142	0,0187	1,04e-10				
	Expan	8,36E-19	949,204	946,078	0,0187	1,02e-11	7,844			
	Log	8,36E-19	949,204	946,078	0,0187	5618,87		5,46e-16		
	Step	8,36E-19	949,204	946,078	0,0187				5,81E-06	1
	PExpan	8,36E-19	949,204	946,078	0,0187	1,02e-10	0,629			
	PLog	8,36E-19	957,194	954,068	0,0216	766,402	0,002			
eC28	Const	0,0054	717,815	716,794	0,194					
	Expo	0,0054	725,264	723,201	0,247	31,181				
	Expan	0,0054	725,264	722,137	0,247	31,181	4,2e-15			
	Log	0,0054	725,266	722,139	0,245	32,665		9,212		
	Step	0,0054	717,815	714,689	0,194				1,40E-08	0,1
	PExpan	0,0054	725,264	722,137	0,247	31,181	0,066			
	PLog	0,0054	725,926	722,799	0,235	32,541	0,002			

Tabla 10. (Continuación).

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
eG14	Const	0,0025	837,169	836,149	0,058					
	Expo	0,0025	839,468	837,406	0,068	50,778				
	Expan	0,0025	839,468	836,342	0,068	50,778	1,96E-14			
	Log	0,0025	840,627	837,501	0,063	127,038		0,144		
	Step	0,0025	837,169	834,043	0,058				0,058	5,81E-06
	PExpan	0,0025	839,468	836,342	0,068	50,778	0,077			
	PLog	0,0025	842,390	839,263	0,063	79,984	0,004			
eG19	Const	0,0009	817,059	816,038	0,071					
	Expo	0,0009	819,542	817,479	0,082	37,911				
	Expan	0,0009	819,542	816,415	0,082	37,911	1,30E-11			
	Log	0,0009	819,882	816,755	0,078	68,514		0,383		
	Step	0,0009	817,059	813,932					0,071	1,40E-08
	PExpan	0,0009	819,542	816,415	0,082	37,911	0,028			
	PLog	0,0009	821,095	817,969	0,077	49,294	0,004			

Tabla 11. Estima de los parámetros implicados en los diferentes modelos demográficos, para la región NS5A correspondiente a los individuos de genotipo 1b. También se detalla el valor umbral que agrupa los intervalos de cada internodo en el árbol filogenético (maxepsilon), el logaritmo de la verosimilitud (ln L), y la corrección del lnL según el criterio de información de Akaike corregido (AICc), el cual usamos para hacer la comparación entre los diferentes modelos. Con sombreado se señala el modelo que mejor se adapta a los datos según el AICc

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
nA03	Const	1,88E-03	246,830	245,773	0,055					
	Expo	1,88E-03	271,635	269,459	0,358	594,551				
	Expan	1,88E-03	271,635	268,272	0,358	594,551	4,57E-11			
	Logistic	1,88E-03	273,939	270,575	0,172	777,634		0,172		
	Step	1,88E-03								
	PExpan	1,88E-03	271,635	268,272	0,358	594,551	0,002			
	PLog	1,88E-03	275,067	271,703	0,125	731,137	0,003			
nA14	Const	9,92E-07	450,297	449,260	0,0357					
	Expo	9,92E-07	468,219	466,108	0,1054	632,9350				
	Expan	9,92E-07	468,219	464,992	0,1054	632,9350	7,9e-17			
	Logistic	9,92E-07	473,202	469,976	0,0646	1081,7700		0,099		
	Step	9,92E-07	450,297	447,070	0,0357				3,00E-11	0,107
	PExpan	9,92E-07	468,219	464,992	0,1054	632,9350	0,010			
	PLog	9,92E-07	475,088	471,862	0,0572	911,8330	0,002			
nA16	Const	0,0004	346,790	345,749	0,087					
	Expo	0,0004	362,033	359,908	0,154	143,828				
	Expan	0,0004	365,861	362,605	0,828	1060,070	0,039			
	Logistic	0,0004	362,033	358,777	0,154	143,828		1,00E+05		
	Step	0,0004	346,790	343,535	0,087				5,81E-06	4,150
	PExpan	0,0004	366,159	362,904	0,550	666,250	0,060			
	PLog	0,0004	362,068	358,813	0,149	143,920	0,000			

θ : $N(0)\mu$; ρ : r/μ ; α : Tamaño poblacional anterior al cambio, como proporción de $N(0)$; c Parámetro de forma logística; f: Tamaño poblacional anterior al cambio, como proporción de $N(0)$; x: Tiempo de cambio.

Tabla 11. (Continuación)

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
nA35	Const	5,94E-08	197,765	196,682	0,012					
	Expo	5,94E-08	203,523	201,262	0,025	633,228				
	Expan	5,94E-08	203,523	199,978	0,025	633,228	2,6E-16			
	Logistic	5,94E-08	204,867	201,322	0,016	1573,29		0,019		
	Step	5,94E-08	197,765	194,22					5,81E-06	0,41499
	PExpan		203,523	200	0,025	633,228	0,038			
	PLog	5,94E-08	205,351	201,806	0,019	758,974	0,001			
nC04	Const	0,0006	851,784	850,761	0,020					
	Expo	0,0006	855,318	853,249	0,028	345,477				
	Expan	0,0006	855,318	852,180	0,028	345,478	2,40E-15			
	Logistic	0,0006	862,779	859,641	0,025	1164,44		0,058		
	Step	0,0006								
	PExpan	0,0006	855,318	852,180	0,028	345,477	0,045			
	PLog	0,0006	868,052	864,914	0,025	827,793	0,001			
nC05	Const	1,95E-08	567,157	566,129	0,074					
	Expo	1,95E-08	596,229	594,144	0,189	397,357				
	Expan	1,95E-08	597,117	593,946	0,233	494,043	0,008			
	Logistic	1,95E-08	596,258	593,087	0,181	408,454		6,836		
	Step	1,95E-08	567,157	563,986	0,074				5,81E-06	1
	PExpan	1,95E-08	596,935	593,764	0,210	439,216	0,011			
	PLog	1,95E-08	599,031	595,859	0,138	419,529	0,001			
nC08	Const	0,0011	648,875	647,848	0,032					
	Expo	0,0011	649,009	646,926	0,033	7,618				
	Expan	0,0011	649,009	645,842	0,033	242,817	7,17E-01			
	Logistic	0,0011	648,875	645,708	0,033	0,355E-06		9,37e-28		
	Step	0,0011	648,875	645,708	0,032				5,81E-06	4,150
	PExpan	0,0011	649,494	646,328	0,033	826,659	6,61E-01			
	PLog	0,0011	655,210	652,044	0,033	46863,5	6,64E-02			
nC10	Const	0,003	542,205	541,171	0,018					
	Expo	0,003	543,222	541,121	0,021	97,069				
	Expan	0,003	543,222	540,015	0,021	97,069	8,82e-15			
	Logistic	0,003	545,634	542,428	0,019	4586,65		1,00E-30		
	Step	0,003	542,205	538,998	0,018				5,81E-06	1
	PExpan	0,003	543,222	540,015	0,021	97,069	0,125			
	PLog	0,003	546,334	543,127	0,020	9006,48	0,015			
nC12	Const	0,0003	739,105	738,082	0,069					
	Expo	0,0003	765,464	763,396	0,130	336,071				
	Expan	0,0003	769,641	766,503	0,189	544,959	1,59E-02			
	Logistic	0,0003	765,464	762,326	0,130	336,071		4,80E+11		
	Step	0,0003	739,105	735,967					6,90E-02	1,40E-08
	PExpan	0,0003	769,967	766,829	0,174	473,814	0,019			
	PLog	0,0003	767,545	764,407	0,108	347,446	0,001			
nC13	Const	0,0014	497,292	496,262	0,089					
	Expo	0,0014	506,555	504,462	0,115	73,953				
	Expan	0,0014	518,288	515,101	0,330	584,832	0,0643237			
	Logistic	0,0014	506,555	503,367	0,115	73,953		1,00E+05		
	Step	0,0014	497,292	494,105	0,089				5,81E-01	414,987
	PExpan	0,0014	518,752	515,564	0,308	465,28	0,0938166			
	PLog	0,0014	506,632	503,444	0,112	741,755	0,000342871			

Tabla 11. (Continuación)

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
nC15	Const	0,0003	697,210	696,184	0,026					
	Expo	0,0003	697,285	695,206	0,027	17,693				
	Expan	0,0003	697,285	694,125	0,027	17,693	6,51E-12			
	Logistic	0,0003	700,370	697,577	0,028	3479,19		1,00E-30		
	Step	0,0003	697,210	694,05	0,026				1,40E-08	0,0241
	PExpan	0,0003	697,285	694,125	0,027	17,693	0,078			
	PLog	0,0003	700,703	697,543	0,028	184155,000	0,006			
nC16	Const	0,030	528,480	527,520	0,110					
	Expo	0,030	567,629	565,543	0,364	363,067				
	Expan	0,030	567,629	564,455	0,364	363,067	7,14E-13			
	Logistic	0,030	567,629	564,455	0,364	363,067		1,00E+05		
	Step	0,030	528,548	525,374	0,110				5,81E-06	1
	PExpan	0,030	567,629	564,455	0,364	363,067	0,00064			
	PLog	0,030	568,243	565,069	0,297	368,148	0,0006386			
nC19	Const	4,18E-19	525,429	524,400	0,101					
	Expo	4,18E-19	555,216	553,129	0,249	278,086				
	Expan	4,18E-19	555,791	552,615	0,304	352,749	0,013			
	Logistic	4,18E-19	555,216	552,040	0,249	278,086		1,00E+05		
	Step	4,18E-19	525,429	522,252	0,101				1,40E-08	0,024
	PExpan	4,18E-19	555,127	551,951	0,262	278,086	0,001			
	PLog	4,18E-19	556,608	553,431	0,195	287,042	0,001			
nG02	Const	0,003	465,79	464,758	0,083					
	Expo	0,003	480,419	478,321	0,159	173,740				
	Expan	0,003	480,419	477,219	0,159	173,740	6,22e-16			
	Logistic	0,003	480,966	477,766	0,135	227,156		0,6383		
	Step	0,003	465,79	462,59	0,083				1,40E-08	0,0241
	PExpan	0,003	480,419	477,219	0,159	173,740	0,012306			
	PLog	0,003	481,796	478,596	0,124	192,023	0,0021793			
nG05	Const	0,001	475,000	473,967	0,045					
	Expo	0,001	480,565	478,462	0,060	105,090				
	Expan	0,001	480,565	477,355	0,063	105,090	1,74E-13			
	Logistic	0,001	480,565	477,355	0,060	105,090	7,64E+10			
	Step	0,001	475,000	471,79	0,045				5,81E-06	4,14987
	PExpan	0,001	480,565	477,355	0,060	105,090	0,003025			
	PLog	0,001	480,936	477,726	0,057	108,611	0,0006605			
nG06T0	Const	0,002	391,312	390,278	0,156					
	Expo	0,002	432,839	430,734	0,451	253,878				
	Expan	0,002	453,855	450,641	3,575	720,450	0,003			
	Logistic	0,002	432,839	429,625	0,451	253,878		1,00E+04		
	Step	0,002	391,312	388,098					0,156	5,81E-06
	PExpan	0,002	432,870	430	0,424	253,992	0,000243362			
	PLog	0,002	432,839	429,625	0,424	253,878	0,000243			
nG07	Const	0,002	312,381	311,337	0,0875283					
	Expo	0,002	318,601	316,465	0,115057	54,7157				
	Expan	0,002	327,327	324,048	0,462114	622,928	0,0596378			
	Logistic	0,002	318,601	315,322	0,115057	54,7157		7,84E+12		
	Step	0,002	312,381	309,102					0,087528	5,81E-06
	PExpan	0,002	327,967	324,688	0,54728	606,136	0,0666556			
	PLog	0,002	318,654	315,375	0,112398	54,9158	0,0004714			

Tabla 11. (Continuación)

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
nG09	Const	0,002	446,265	445,229	0,0383196					
	Expo	0,002	450,938	448,827	0,0507965	117,28				
	Expan	0,002	451,353	448,126	0,0589599	212,194	0,0854399			
	Logistic	0,002	450,938	447,711	0,0507965	117,28		1,00E+04		
	Step	0,002	446,265	443,039	0,0383196				3,37E-11	4,14987
	PExpan	0,002	452,201	448,975	0,0618735	206,372	0,132058			
	PLog	0,002	451,986	448,76	0,0465265	128,052	0,0011633			
nG17	Const	0,003	483,835	482,804	0,0951805					
	Expo	0,003	502,916	500,822	0,20795	200,718				
	Expan	0,003	502,922	499,732	0,211215	205,474	1,85E-03			
	Logistic	0,003	503,373	500,183	0,178131	236,342		1,077		
	Step	0,003	483,835	480,645	0,0951805				1,40E-08	0,0241
	PExpan	0,003	502,916	499,725	0,20795	200,718	0,001			
	PLog	0,003	505,518	502,328	0,156205	219,125	0,002			

Tabla 12. Estima de los parámetros implicados en los diferentes modelos demográficos, para la región NS5A correspondiente a los individuos de genotipo 1a. También se detalla el valor umbral que agrupa los intervalos de cada internodo en el árbol filogenético (maxepsilon), el logaritmo de la verosimilitud (ln L), y la corrección del lnL según el criterio de información de Akaike corregido (AICc), el cual usamos para hacer la comparación entre los diferentes modelos. Con sombreado se señala el modelo que mejor se adapta a los datos según el AICc

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
nA09	Const	0,001	215,565	214,496	0,030					
	Expo	0,001	222,089	219,875	0,049	195,936				
	Expan	0,001	223,126	219,681	0,071	424,106	6,68E-02			
	Logistic	0,001	222,089	218,645	0,049	195,936		1,08E+240		
	Step	0,001	215,565	212,121					3,03E-02	1,40E-08
	PExpan	0,001	222,089	218,645	0,049	195,936	0,0072			
	PLog	0,001	222,284	218,839	0,046	198,815	0,0005			
nA20	Const	0,0006	598,624	597,592	0,0146					
	Expo	0,0006	598,680	596,584	0,0152	32,441				
	Expan	0,0006	598,680	595,486	0,0152	32,441	1,37E-11			
	Logistic	0,0006	602,158	598,964	0,0157	769,000		0,010		
	Step									
	PExpan	0,0006	598,680	595,486	0,0152	32,441	0,226			
	PLog	0,0006	598,624	595,430	0,0146	0,010	6,71 E+13			
nA26	Const	0,109	476,689	475,655	0,0441					
	Expo	0,109	487,464	485,360	0,0710	217,366				
	Expan	0,109	487,464	484,253	0,0710	217,366	4,85e-15			
	Logistic	0,109	487,464	484,253	0,0710	217,365		1,00E+05		
	Step	0,109	476,689	473,479	0,0441				5,81E-06	1
	PExpan	0,109	487,464	484,253	0,0710	217,366	0,005			
	PLog	0,109	488,521	485,310	0,0630	228,286	0,001			
	PExpan	0,0012	616,386	613,210	0,054	313,820	0,046			
	PLog	0,0012	620,301	617,125	0,044	392,325	0,001			

θ : $N(0)\mu$; ρ : r/μ ; α : Tamaño poblacional anterior al cambio, como proporción de $N(0)$; c Parámetro de forma logística; f: Tamaño poblacional anterior al cambio, como proporción de $N(0)$; x: Tiempo de cambio.

Tabla 12. (Continuación)

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
nA27	Const	0,0012	607,177	606,148	0,032					
	Expo	0,0012	616,386	614,299	0,054	313,819				
	Expan	0,0012	616,386	613,210	0,054	313,819	3.8e-16			
	Logistic	0,0012	619,116	615,939	0,041	705,229		0,068		
	Step	0,0012	607,177	604,001	0,032				5,81E-06	1
	PExpan	0,0012	616,386	613,210	0,054	313,820	0,046			
	PLog	0,0012	620,301	617,125	0,044	392,325	0,001			
nA34	Const	0,0024	164,955	163,878	0,073					
	Expo	0,0024	188,503	186,263	0,568	416,978				
	Expan	0,0024	188,503	185,003	0,568	416,978	3,77E-11			
	Logistic	0,0024	188,503	185,003	0,568	416,978		4,92 E+11		
	Step	0,0024	164,955	161,455	0,073				5,81E-06	1
	PExpan	0,0024	188,503	185,003	0,568	416,978	0,003			
	PLog	0,0024	188,503	185,003	0,568	416,978	1,24E-15			
nC01	Const	0,0025	483,060	482,028	0,063					
	Expo	0,0025	485,020	482,922	0,069	25,323				
	Expan	0,0025	491,036	487,836	0,117	308,089	1,50E-01			
	Logistic	0,0025	487,746	484,546	0,073	7576,240		1,00E-250		
	Step	0,0025	483,060	479,860	0,063				5,81E-06	4,150
	PExpan	0,0025	491,554	488,354	0,103	169,021	0,152			
	PLog	0,0025	485,094	481,894	0,068	25,663	0,001			
nC07	Const	0,0007	467,935	466,902	0,594					
	Expo	0,0007	485,339	483,297	0,142	324,298				
	Expan	0,0007	485,399	482,192	0,142	324,298	1,17E-13			
	Logistic	0,0007	487,243	484,036	0,106	451,747		0,347		
	Step	0,0007	467,935	464,728	0,059				5,81E-06	1
	PExpan	0,0007	485,399	482,192	0,142	324,298	0,010			
	PLog	0,0007	489,750	486,543	0,0989576	380,435	0,002			
nC20	Const	0,0007	930,538	929,515	0,007					
	Expo	0,0007	930,538	928,469	0,007	1,02e-10				
	Expan	0,0007	930,538	927,398	0,007	1,54e-10	7,8441			
	Logistic	0,0007	933,738	930,598	0,007	881,91		1,00E-30		
	Step	0,0007	930,538	927,398	0,007				1,40E-08	0,024
	PExpan	0,0007	930,538	927,398	0,007	1,05e-10	0,972			
	PLog	0,0007	937,887	934,748	0,008	594666	0,008			
nC21	Const	1,38E-06	649,598	648,572	0,048					
	Expo	1,38E-06	677,468	675,389	0,151	724,294				
	Expan	1,38E-06	677,468	674,308	0,151	724,294	5,67e-16			
	Logistic	1,38E-06	687,150	683,99	0,075	1790,01		0,008		
	Step	1,38E-06								
	PExpan	1,38E-06	677,468	674,308	0,151	724,294	0,002			
	PLog	1,38E-06	686,031	683	0,070	1513,22	0,002			
nC22	Const	0,003	258,888	257,840	0,139					
	Expo	0,003	295,001	292,851	0,898	308,420				
	Expan	0,003	295,448	292,141	1,285	371,041	0,002			
	Logistic	0,003	295,001	291,694	0,898	308,420		1,15E+11		
	Step	0,003	258,888	255,581	0,139				5,80E-06	1
	PExpan	0,003	295,153	291,846	0,134	371,559	0,006			
	PLog	0,003	295,453	292,146	0,537	315,141	0,002			

Tabla 12. (Continuación)

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
nC24	Const	3,46E-06	441,499	440,456	0,009					
	Expo	3,46E-06	444,231	442,101	0,014	659,343				
	Expan	3,46E-06	444,231	440,965	0,014	659,343	3.6e-15			
	Logistic	3,46E-06	453,418	450,151	0,012	3149,890		0,008		
	Step	3,46E-06	441,499	438,232	0,009				5,81E-06	0,140
	PExpan	3,46E-06	444,231	440,965	0,014	659,343	0,053			
	PLog	3,46E-06	441,499	438	0,009	471,427	1,57E+12			
nC28	Const	0,002	328,161	327,118	0,075					
	Expo	0,002	347,057	344,924	0,192	252,196				
	Expan	0,002	347,056	343,784	0,191	251,860	6,79E-14			
	Logistic	0,002	347,063	343,790	0,187	257,599		10,217		
	Step	0,002	328,161	324,889	0,075				5,81E-06	1
	PExpan	0,002	347,057	343,784	0,192	252,196	0,001			
	PLog	0,002	348,102	344,830	0,147	263,453	0,001			
nG14	Const	0,001	668,769	667,744	0,0617826					
	Expo	0,001	676,220	674,145	0,108526	205,568				
	Expan	0,001	676,220	673,069	0,108526	205,568	4,73E-16			
	Logistic	0,001	703,659	700,507	0,0760321	2748,86		4,68E-09		
	Step	0,001	668,769	665,617					0,061783	5,81E-06
	PExpan	0,001	676,220	673,069	0,108526	205,568	0,029			
	PLog	0,001	703,893	700,741	0,0736637	2886,180	0,007			
nG19	Const	0,002	365,629	364,588	0,060					
	Expo	0,002	382,833	380,708	0,141	283,945				
	Expan	0,002	382,833	379,577	0,141	283,886	1,68E-15			
	Logistic	0,002	382,944	379,689	0,124	327,012		1,350		
	Step	0,002	365,629	362,373	0,060				5,81E-06	1
	PExpan	0,002	382,833	379,577	0,141	283,945	0,002			
	PLog	0,002	383,258	380,003	0,123	289,107	0,001			

2. Análisis a nivel de genotipo

Tabla 13. Estima de los parámetros implicados en los diferentes modelos demográficos, para la región E1-E2 correspondiente a los análisis a nivel de genotipo. También se detalla el valor umbral que agrupa los intervalos de cada internodo en el árbol filogenético (maxepsilon), el logaritmo de la verosimilitud (ln L), y la corrección del lnL según el criterio de información de Akaike corregido (AICc), el cual usamos para hacer la comparación entre los diferentes modelos. Con sombreado se señala el modelo que mejor se adapta a los datos según el AICc.

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
ConsE1a	Const	0,022	78	77	0,795131					
	Expo	0,022	104.484	102.198	8	44				
	Expan	0,022	110.325	106.725	105.283	883.392	2,40E-04			
	Log	0,022	104.484	100.884	792.112	44		100000		
	Step	0,022	78	74	0,795131				5,81E-01	0,24097
	PEXpan	0,022	111	107	78	82	0,0005			
	PLog	0,022	104.484	100.884	8	44	1,27E-15			
ConsE1b	Const	0,019	64	63	1					
	Expo	0,019	107.901	105.615	449.984	59.845				
	Expan	0,019	107.935	104.335	571.862	62	4,85E-06			
	Log	0,019	107.901	104.301	449.984	59.845		100000		
	Step	0,019	64	60	1				5,81E-01	0,41499
	PEXpan	0,019	108.001	104.401	881.607	66	2,84E+00			
	PLog	0,019	108	104	450	60	1,40E-13			
ConsN1a	Const	0,005	98.907	98	0,328					
	Expo	0,005	120.475	118.189	2,227	80,920				
	Expan	0,005	121.759	118.159	4,941	115.447	2,43E-03			
	Log	0,005	120.475	116.875	2,227	80,920		100000		
	Step	0,005	98.907	95.307	0,328				1,40E-03	0,1
	PEXpan	0,005	122	118	3,656	99,955	0,004			
	PLog	0,005	121	117	1,027	80,958	1,00E-02			
ConsN1b	Const	0,010	84	83	0,599					
	Expo	0,010	117.346	115.061	282.619	90,39				
	Expan	0,010	117.346	113.746	28.334	90	1,04E-06			
	Log	0,010	117.348	113.748	20	91		2		
	Step	0,010	84	81	0,599				5,81E-01	414.987
	PEXpan	0,010	117	114	28	90,39	8,54E-05			
	PLog	0,010	117.346	113.746	28	90,39	1,24E-14			
MasdistE1a	Const	0,014	195.736	194.693	134.007					
	Expo	0,014	231.431	229,3	825.542	333.883				
	Expan	0,014	232.294	229.028	118.173	395.062	0,0013835			
	Log	0,014	231.919	228.652	547.689	366.663		0,789792		
	Step	0,014	195.736	192.469	134.007				5,81E-01	0,41499
	PEXpan	0,014	233,11	229.843	113.253	38.764	0,0044169			
	PLog	0,014	231.431	228.164	825.542	333.883	1,22E-67			
MasdistN1a	Const	0,010	259.039	257.997	0,36818					
	Expo	0,010	279.163	277.033	1	59				
	Expan	0,010	279.163	275.897	1	59	5,64E-17			
	Log	0,010	279,79	276.523	0,795159	750.876		0,48514		
	Step	0,010	259.039	255.773	0,36818				5,81E-01	414.987
	PEXpan	0,010	279.163	275.897	1	59	0,00448			
	PLog	0,010	279.163	275.897	1	59	3,6E-55/3,015e-15			

Tabla 13. (Continuación)

Individuo	Modelo	Maxepsilon	ln L	AICc	θ	ρ	α	c	f	x
MasdistE1b	Const	0,027	187,15	186.107	159.672					
	Expo	0,027	211.985	210	8	21				
	Expan	0,027	211.985	209	8	21	1,17E-06			
	Log	0,027	220.924	217.657	3	52	0,0027289			
	Step	0,027	187,15	183.883	159.672	5,81E-01	0,414987			
	PExpan	0,027	211.985	208.719	8	21	0,007815			
	PLog	0,027	220.829	217.563	3	46	0,105611			
MasdistN1b	Const	0,009	222.178	221.135	0,684296					
	Expo	0,009	245.037	242.903	3	46				
	Expan	0,009	245.037	241.764	316.378	46	2,34E-17			
	Log	0,009	259.785	256.512	1	195.818		1,08E+00		
	Step	0,009	222.178	218.905	0,684296				5,81E-01	414.987
	PExpan	0,009	245.037	241.764	316.378	46	0,0068			
	PLog	0,009	259.873	256,6	1	177.969	0,0562945			

θ : $N(0)\mu$; ρ : r/μ ; α : Tamaño poblacional anterior al cambio, como proporción de $N(0)$; c: Parámetro de forma logística; f: Tamaño poblacional anterior al cambio, como proporción de $N(0)$; x: Tiempo de cambio.

ConsE1a: Grupo formado por las secuencias consenso, correspondiente a la región E1-E2, de todos los individuos de genotipo 1a.

ConsE1b: Grupo formado por las secuencias consenso, correspondiente a la región E1-E2, de todos los individuos de genotipo 1b.

ConsN1a: Grupo formado por las secuencias consenso, correspondiente a la región NS5A de todos los individuos de genotipo 1a.

ConsN1b: Grupo formado por las secuencias consenso, correspondiente a la región NS5A, de todos los individuos de genotipo 1b.

MasdistE1a: Grupo formado por las dos secuencias más divergentes de cada uno de los individuos de genotipo 1a, correspondiente a la región E1-E2.

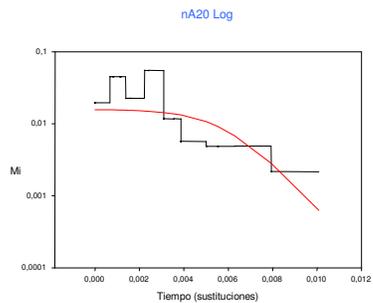
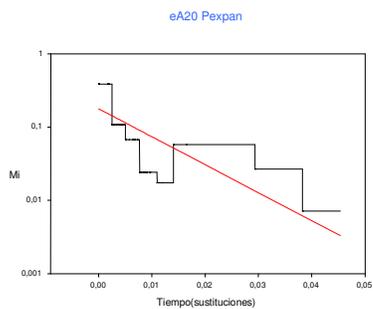
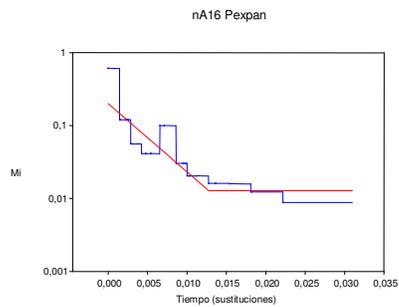
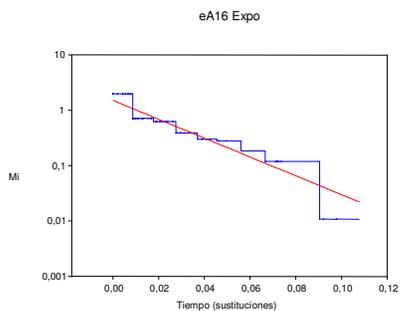
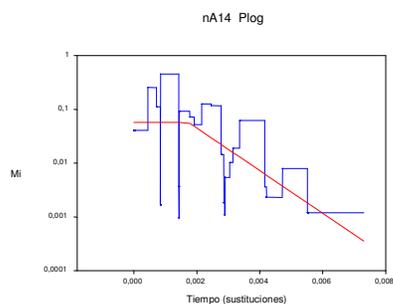
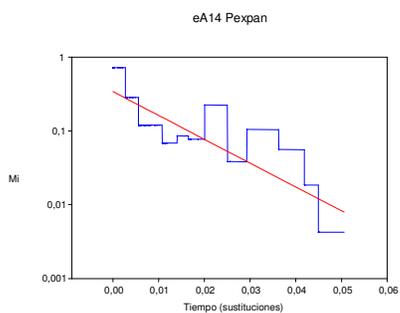
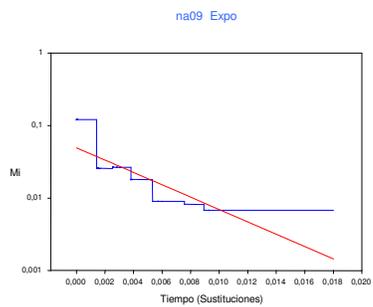
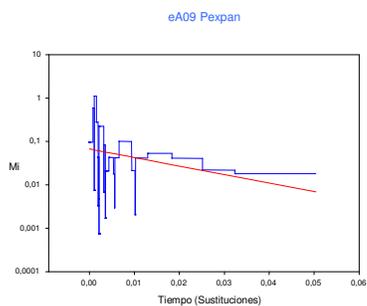
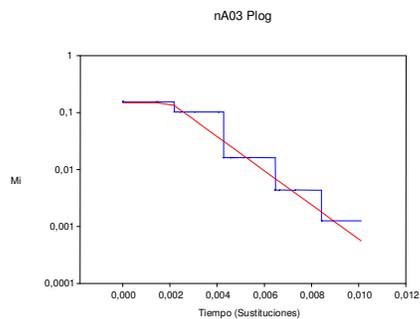
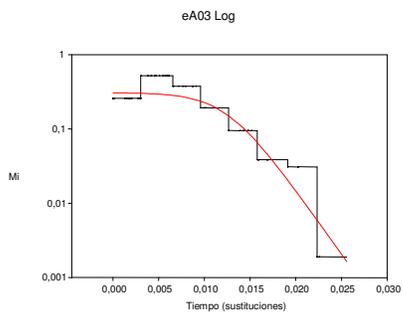
MasdistE1b: Grupo formado por las dos secuencias más divergentes de cada uno de los individuos de genotipo 1b, correspondiente a la región E1-E2.

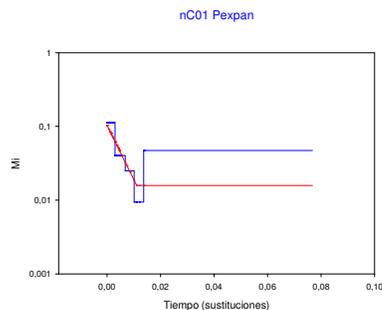
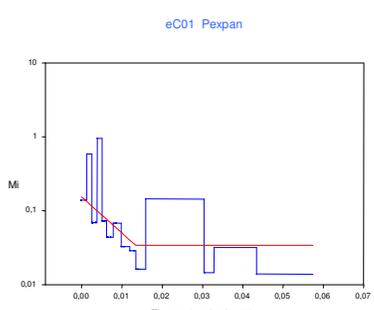
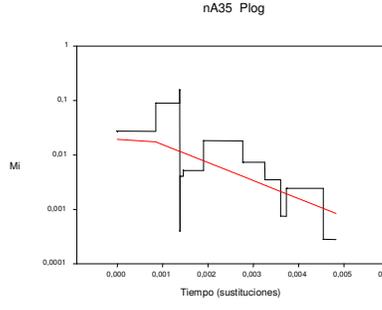
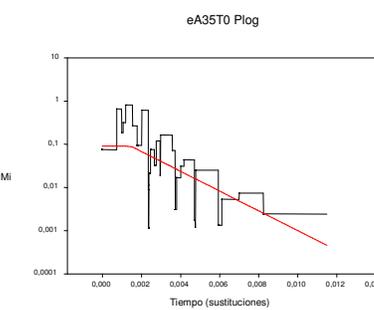
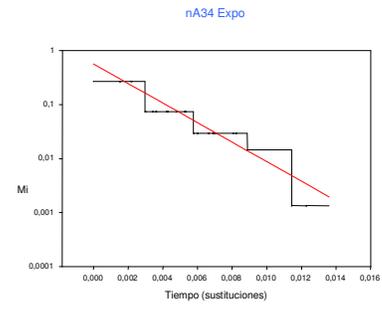
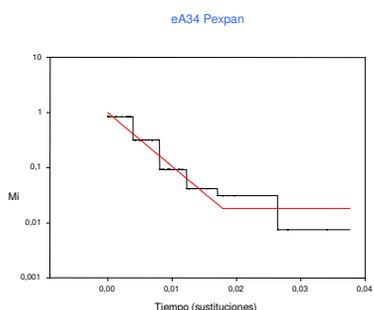
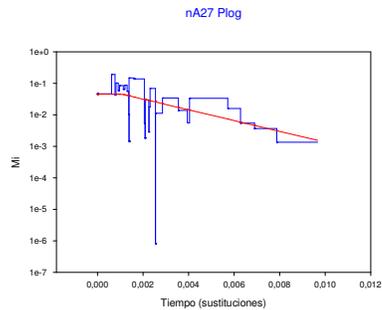
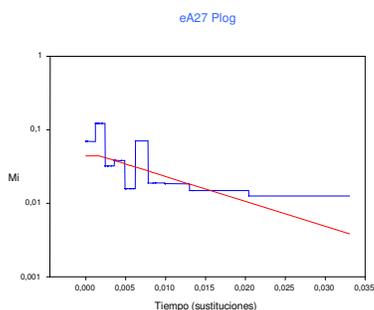
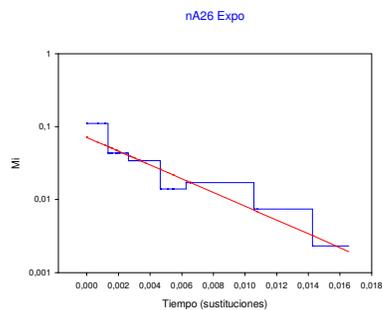
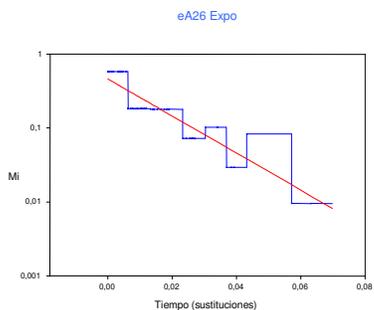
MasdistN1a: Grupo formado por las dos secuencias más divergentes de cada uno de los individuos de genotipo 1a, correspondiente a la región NS5A.

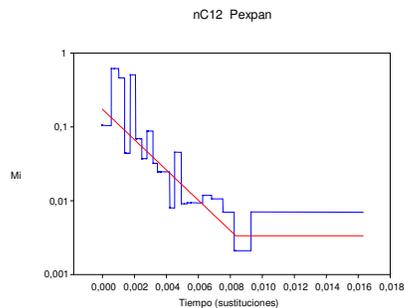
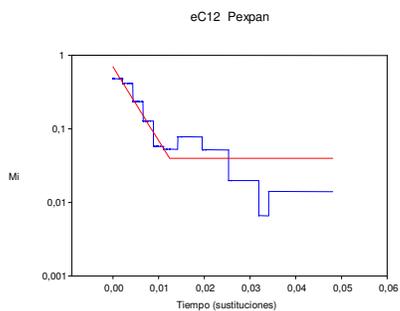
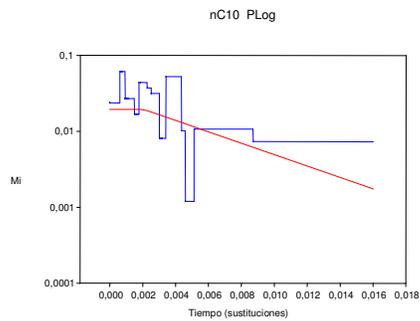
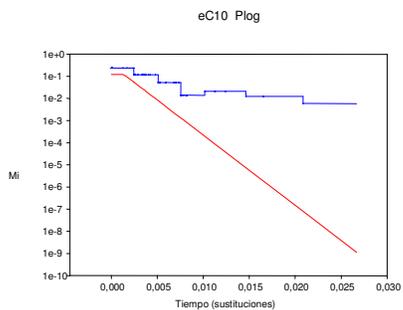
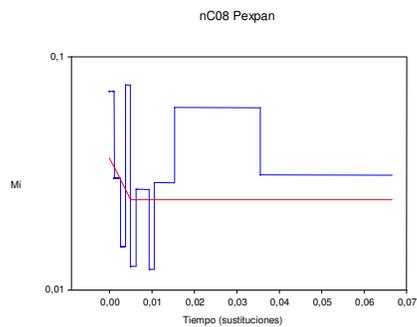
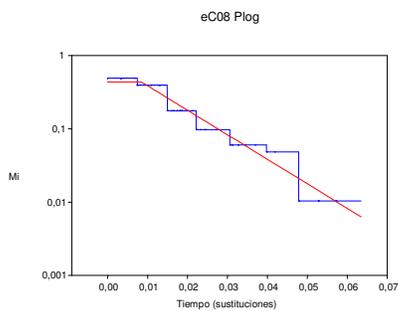
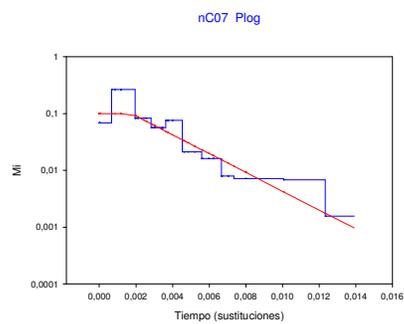
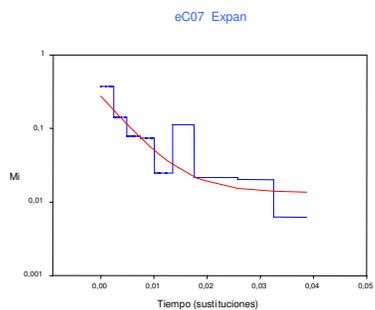
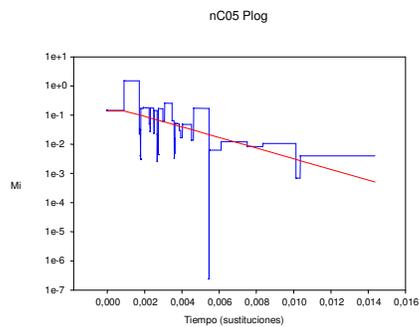
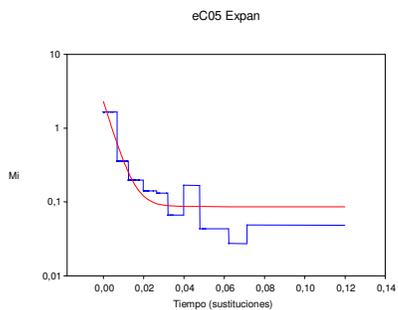
MasdistN1b: Grupo formado por las dos secuencias más divergentes de cada uno de los individuos de genotipo 1a, correspondiente a la región E1-E2.

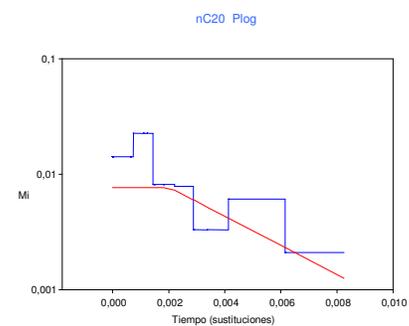
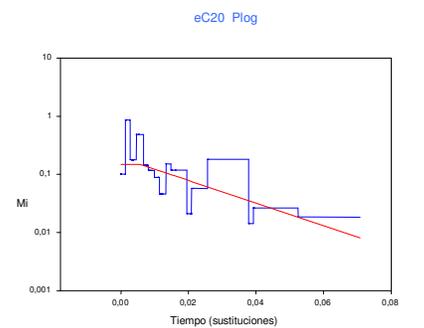
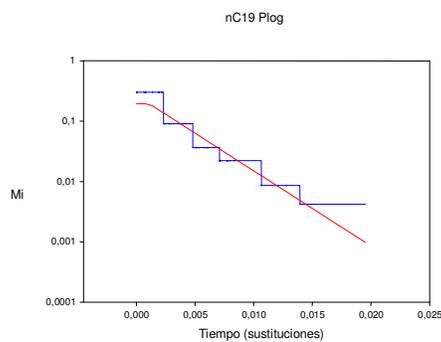
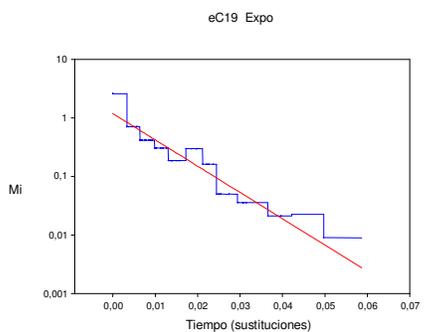
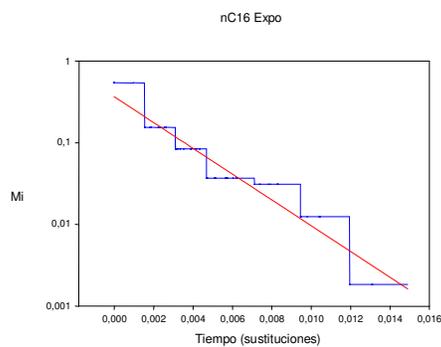
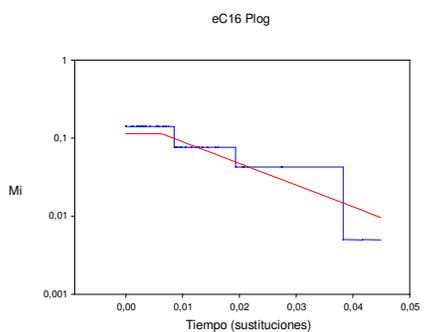
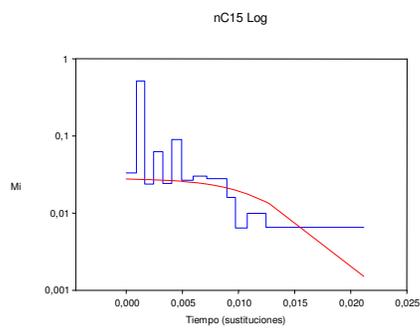
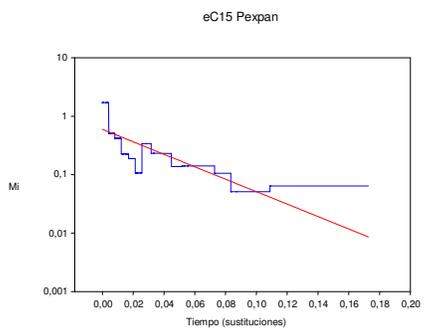
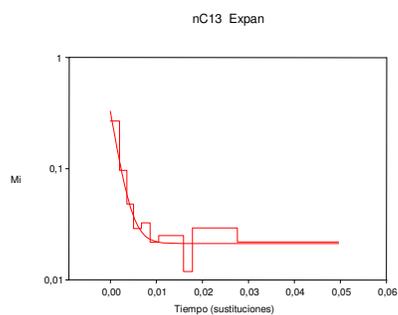
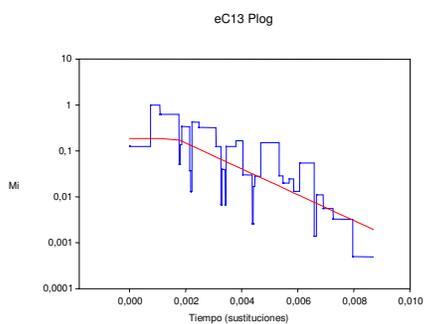
Skylineplots. Estima no paramétrica de la historia demográfica a nivel de paciente.

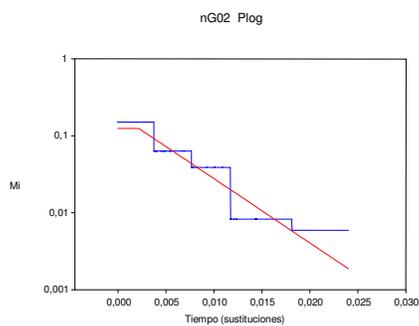
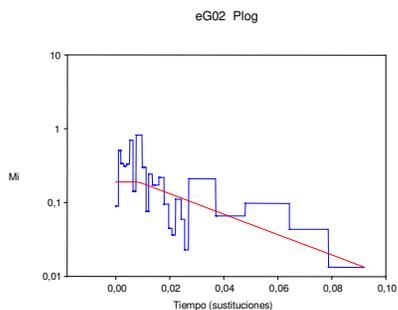
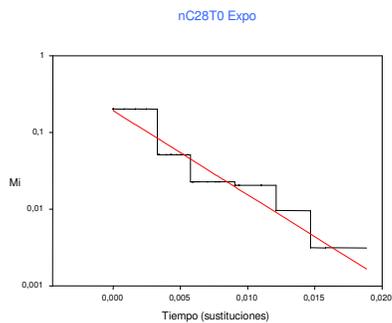
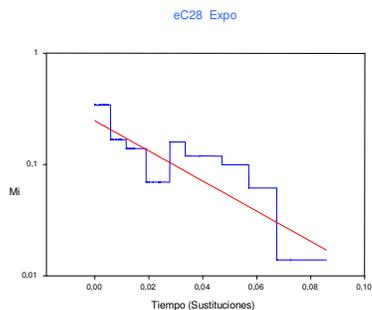
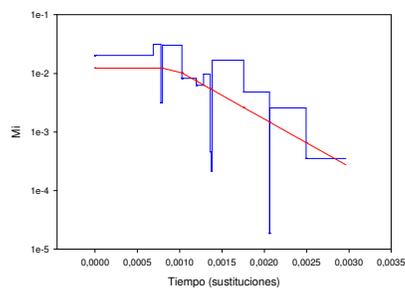
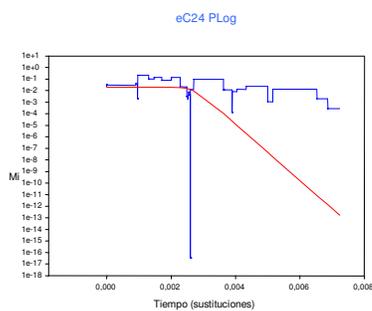
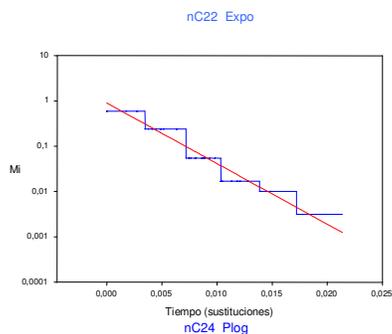
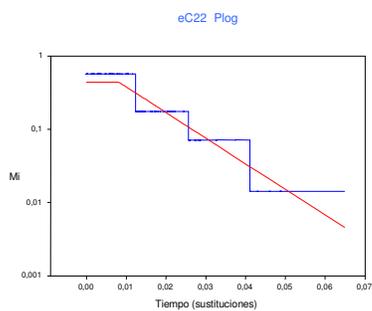
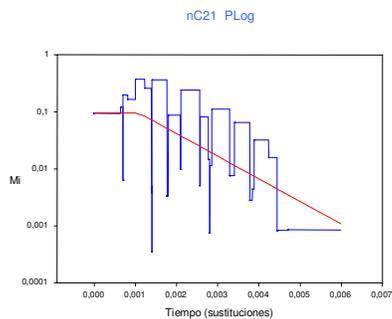
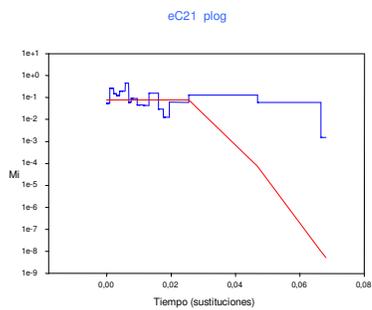
Los “skylineplots” que se presentan a continuación corresponden a la estima no paramétrica del modelo demográfico que mejor explica la dinámica poblacional de cada individuo tanto para la región E1-E2 (columna izquierda) como para la NS5A (columna derecha). Los “skylineplots” correspondientes a los análisis a nivel de genotipo no se muestran en este espacio porque ya aparecieron en el apartado de resultados. Los individuos están ordenados por orden alfabético, para ver a que genotipo corresponden mirar la tabla III.1 del apartado de material y métodos. Junto al nombre de la muestra se señala el nombre del modelo demográfico correspondiente. Expo: crecimiento exponencial, Expan: crecimiento en expansión, Log: crecimiento logístico, Pexpan: crecimiento en expansión escalonado, PLog: crecimiento logístico escalonado. MI: $N_e(x)\mu$ (donde $N_e(x)$ es el tamaño poblacional efectivo y μ la tasa de sustitución nucleotídica)

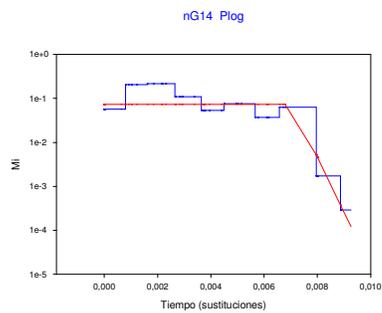
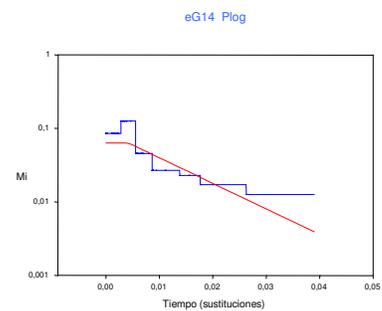
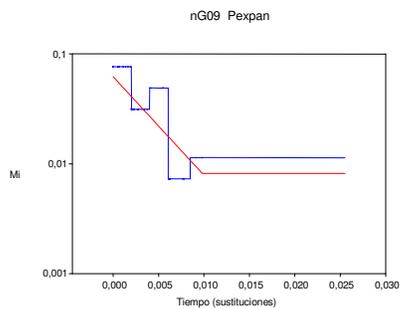
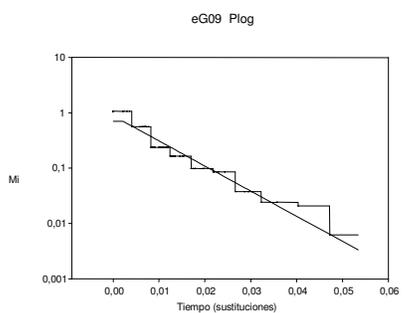
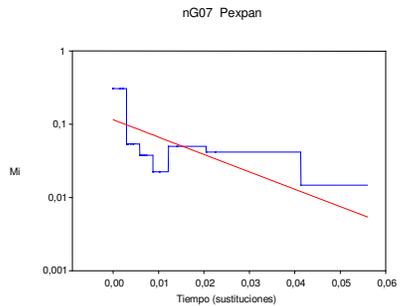
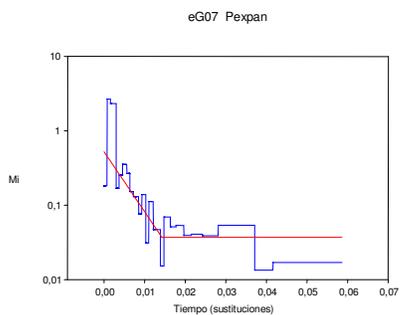
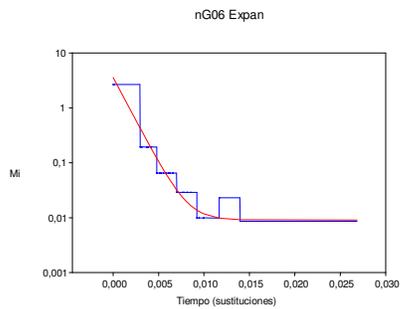
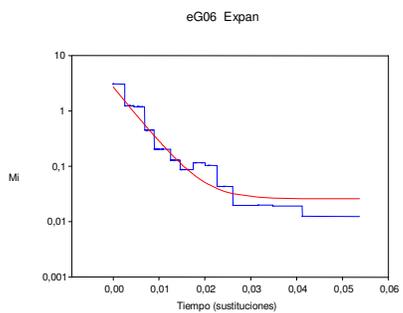
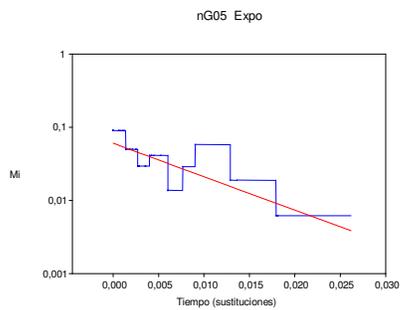
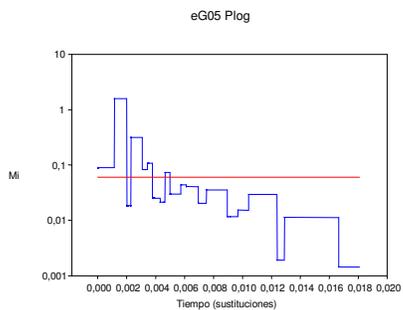


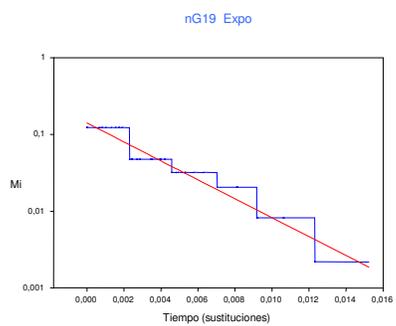
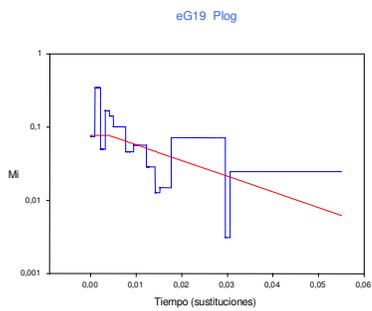
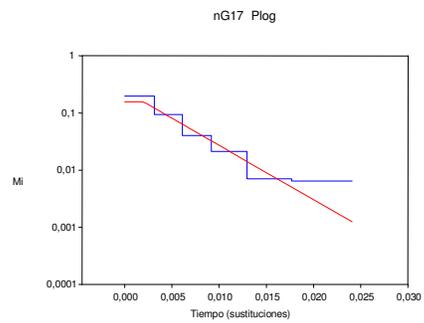
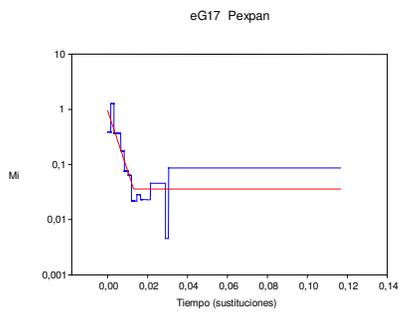










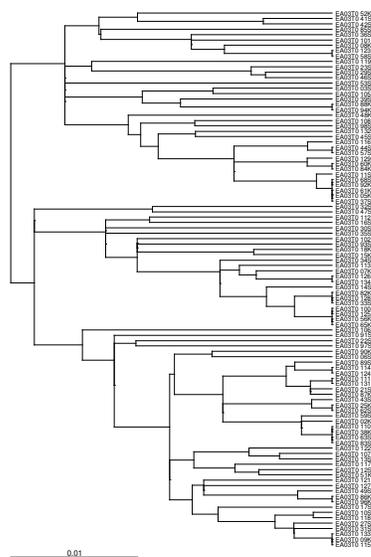


ANEXO VI

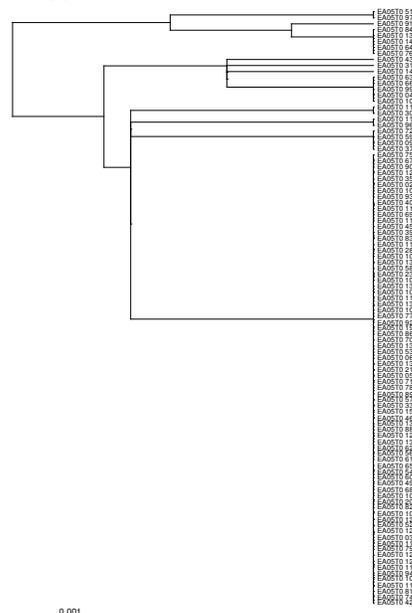
Árboles filogenéticos de las secuencias virales pertenecientes a cada individuo

Filogenias calculadas por máxima verosimilitud con el programa PAUP*4b10, correspondientes a los individuos utilizados en esta tesis. Los árboles están ordenados por orden alfabético-numérico, el genotipo al que pertenecen puede consultarse en la tabla III.1 del apartado de material y métodos.

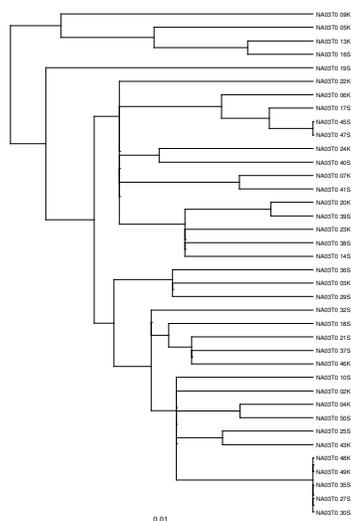
eA03



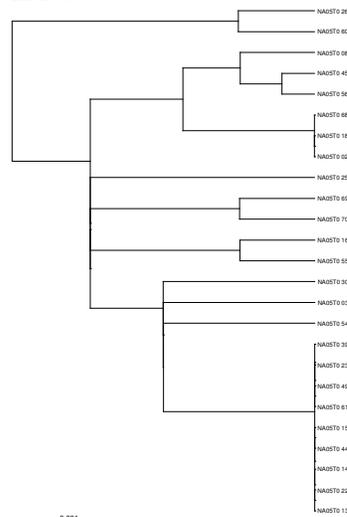
eA05



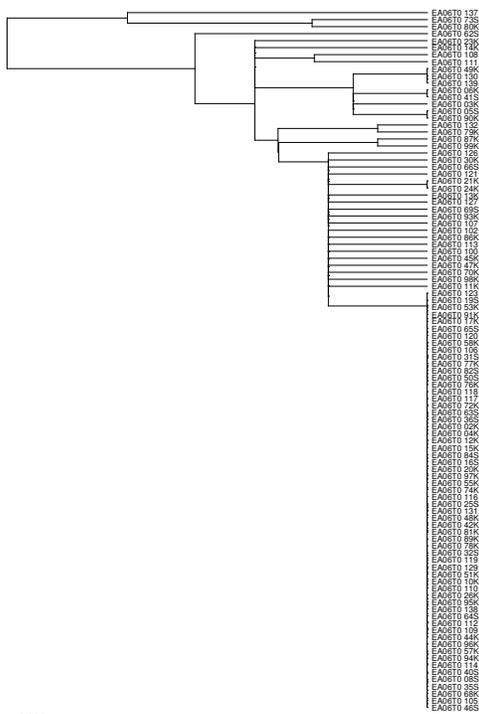
nA03



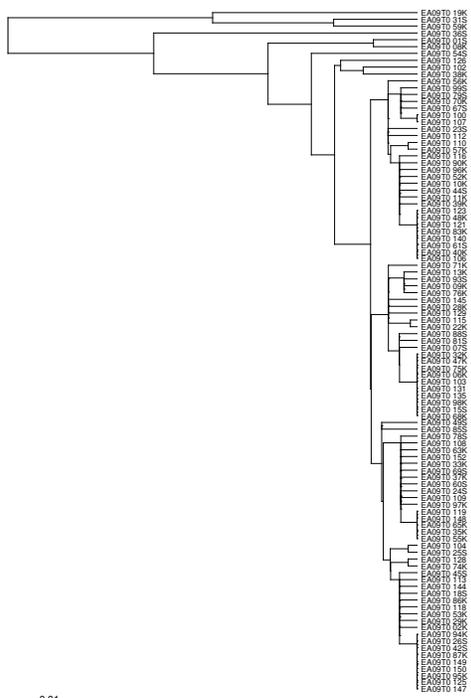
nA05



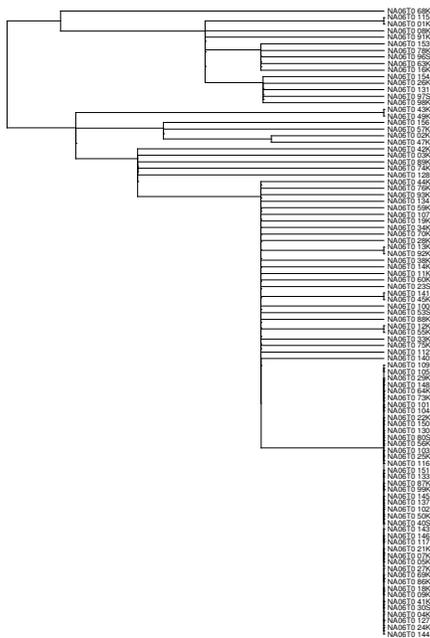
eA06



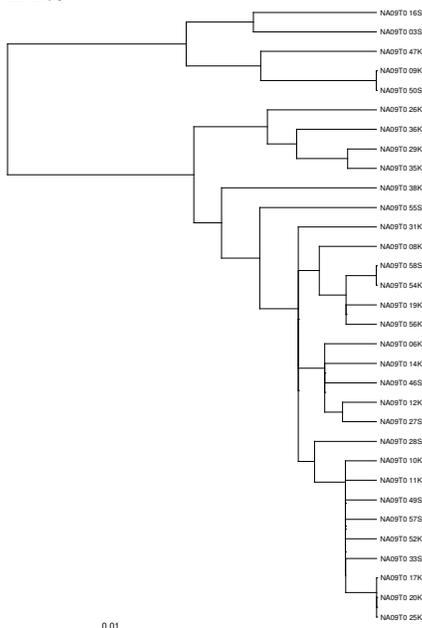
eA09



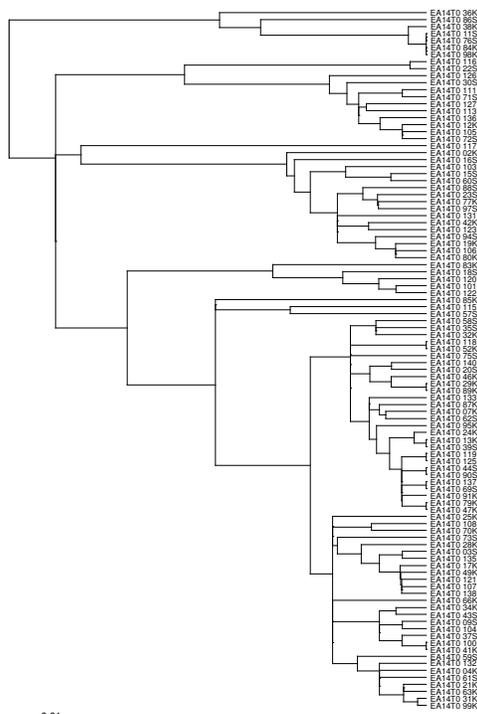
nA06



nA09

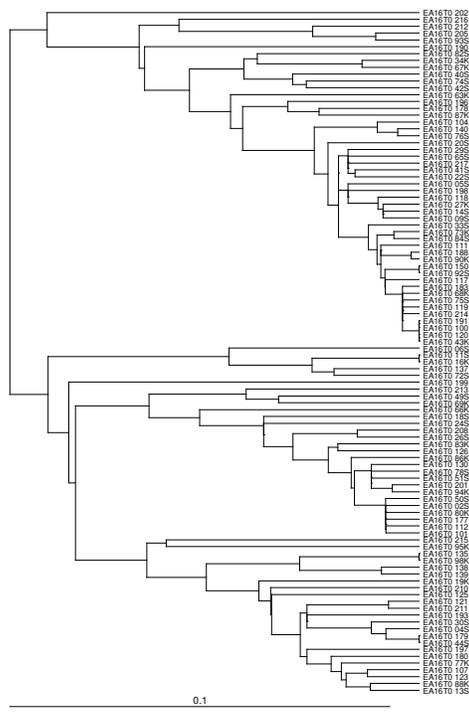


eA14



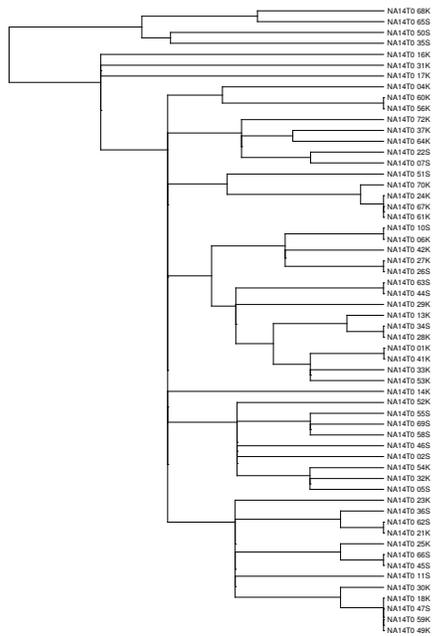
0.01

eA16



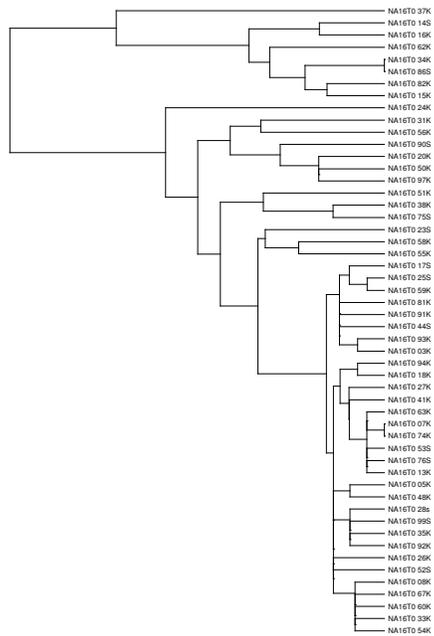
0.1

nA14



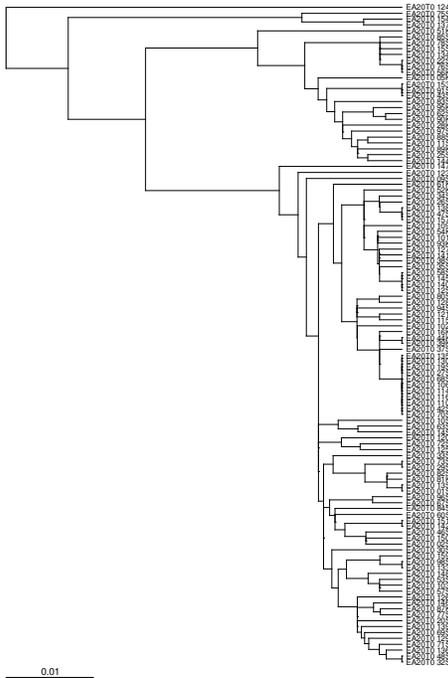
0.001

nA16

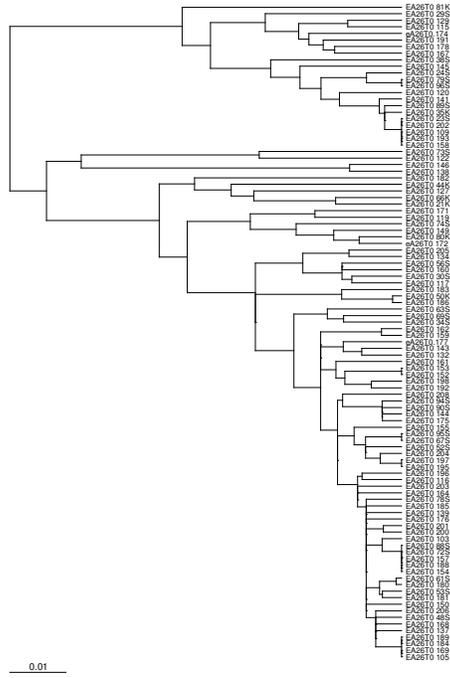


0.01

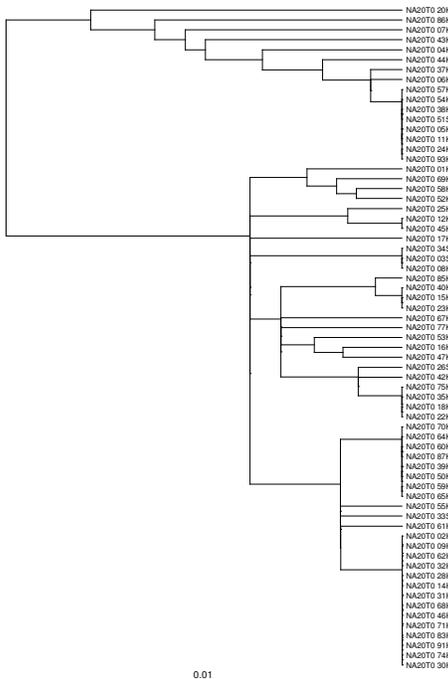
eA20



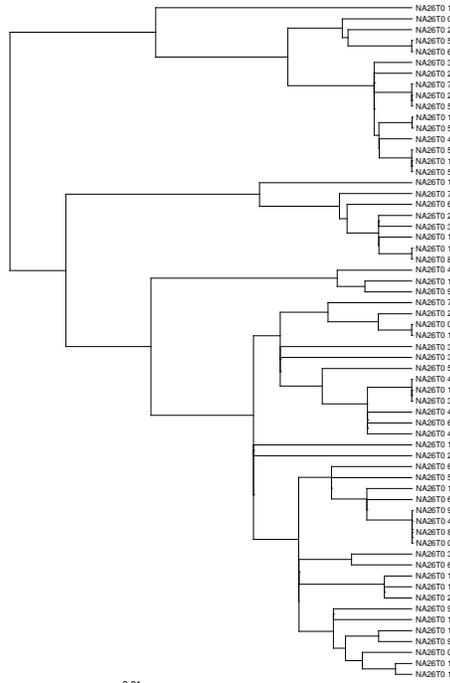
eA26



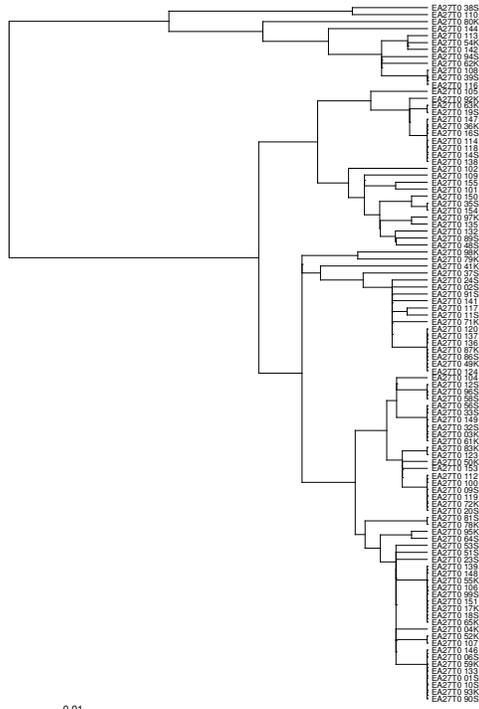
nA20



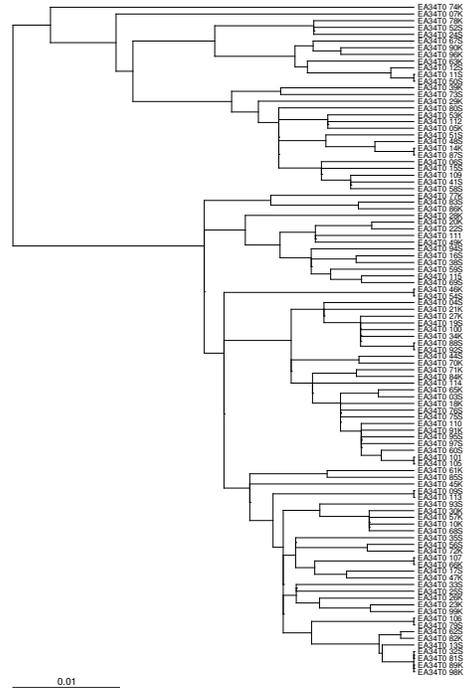
nA26



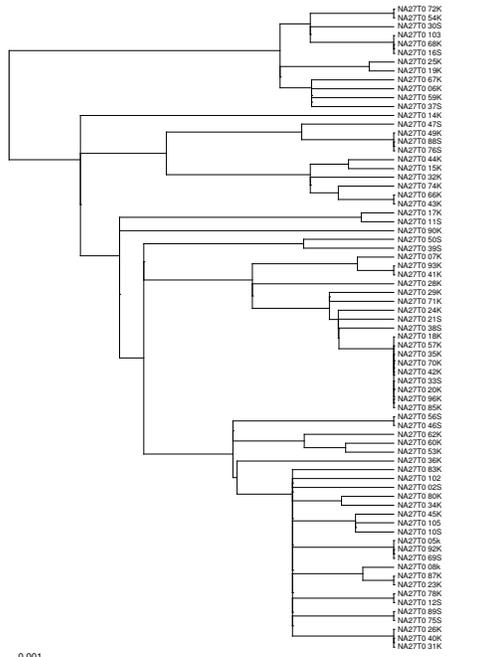
eA27



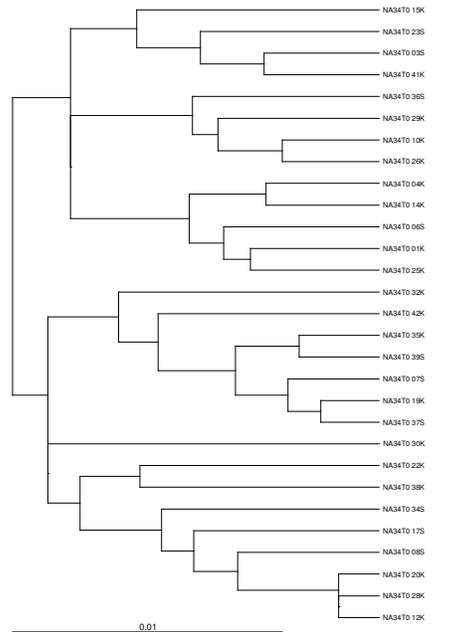
eA34



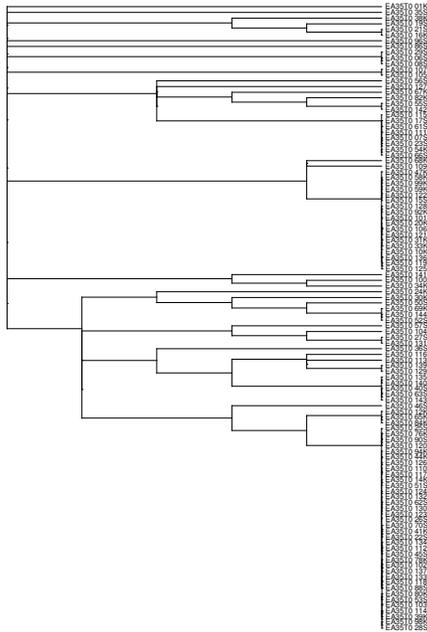
nA27



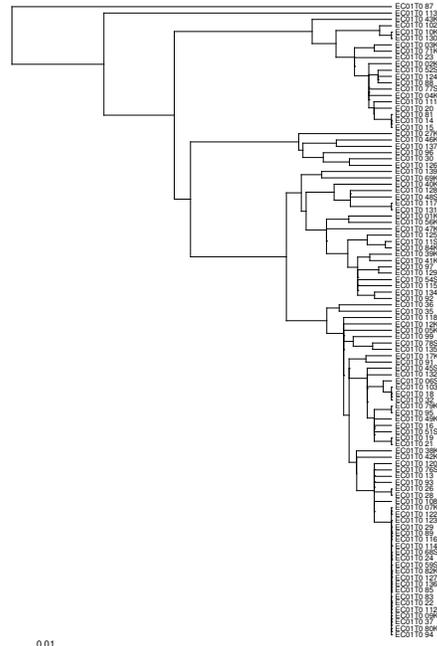
nA34



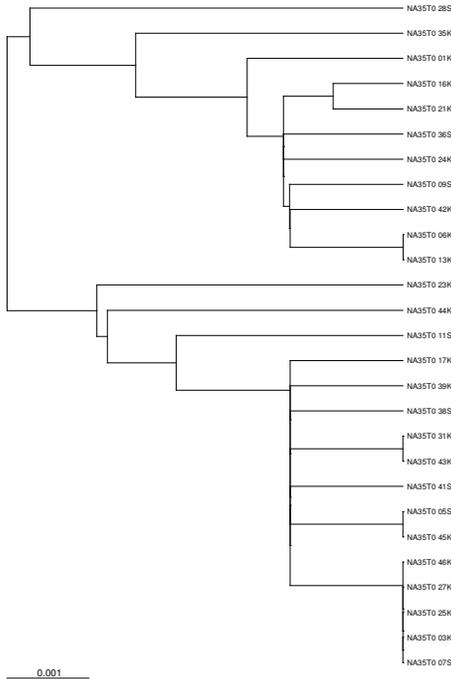
eA35



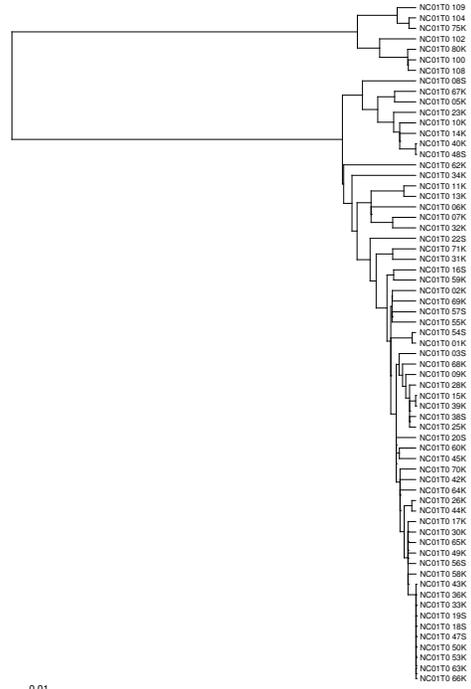
eC01



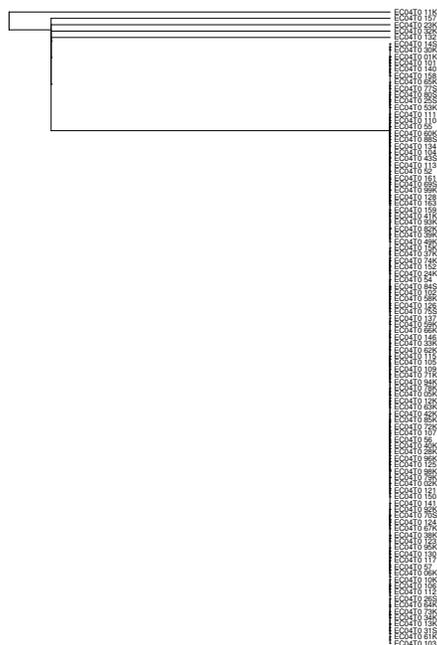
nA35



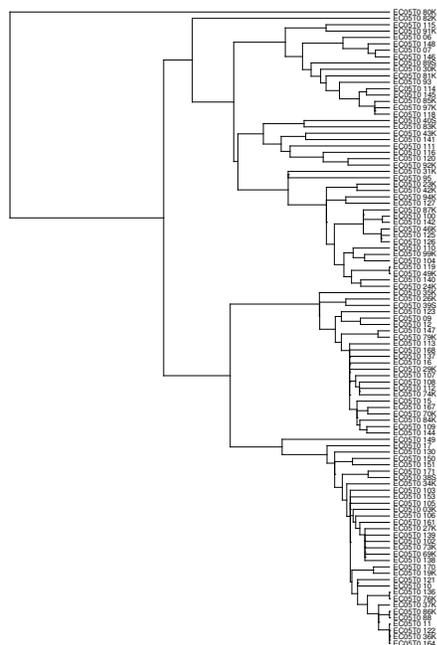
nC01



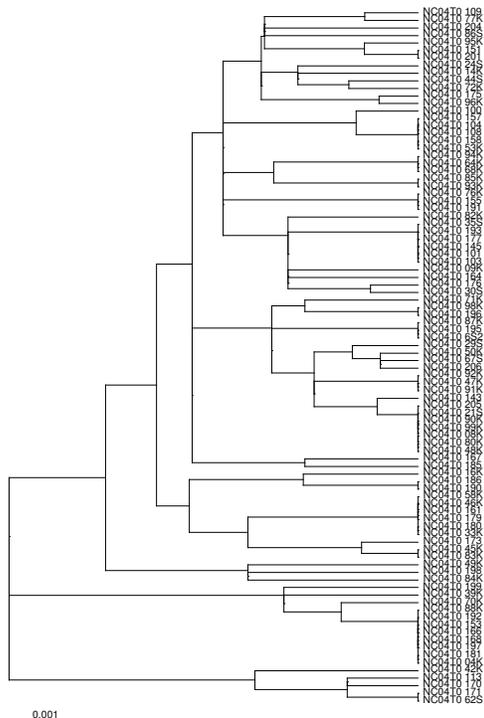
eC04



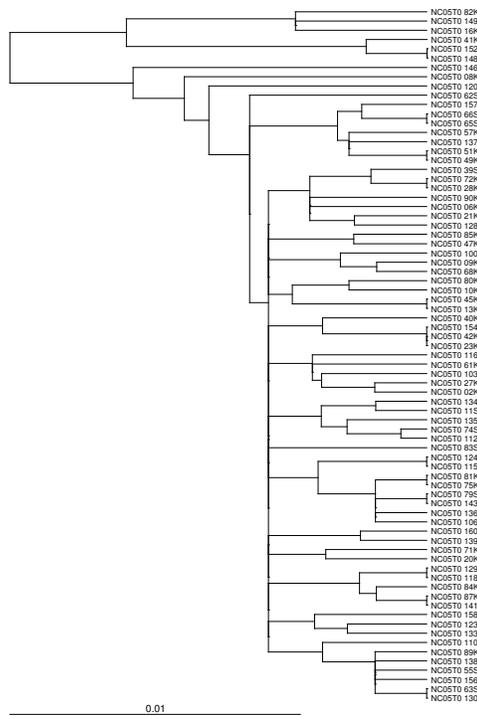
eC05



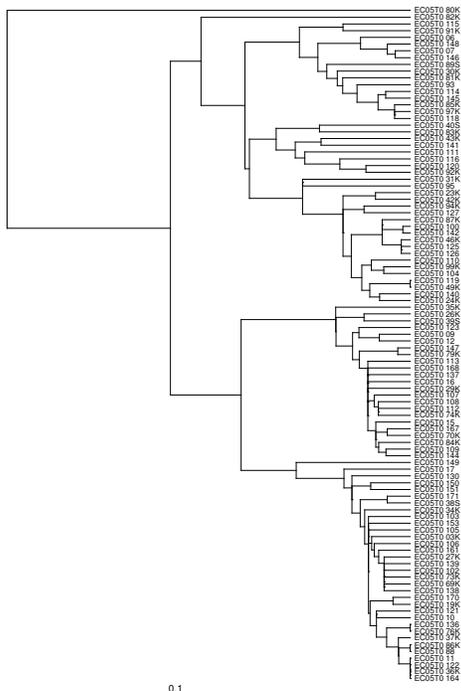
nC04



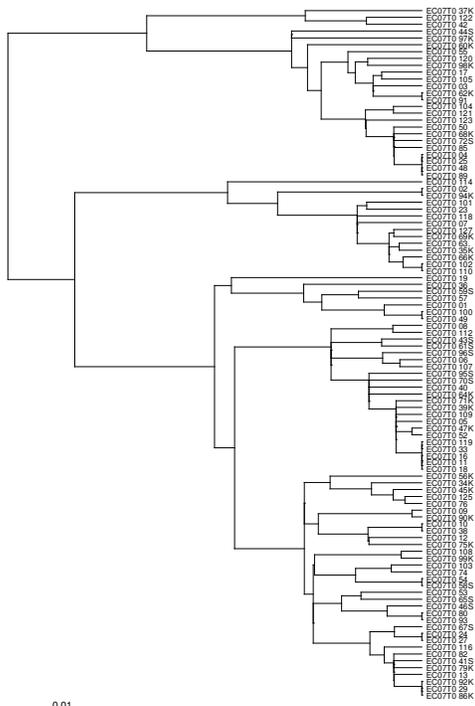
nC05



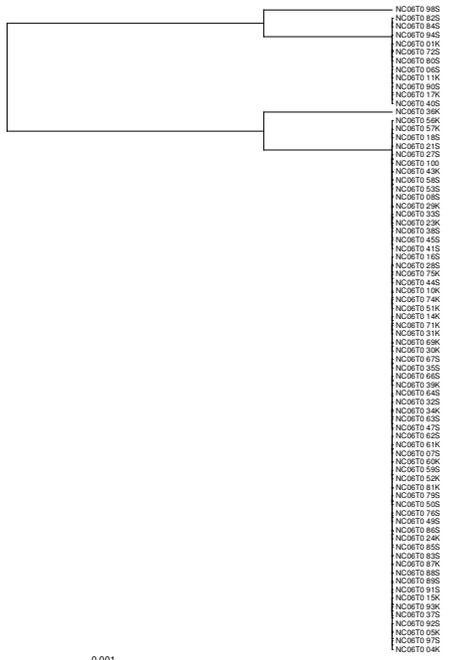
eC06



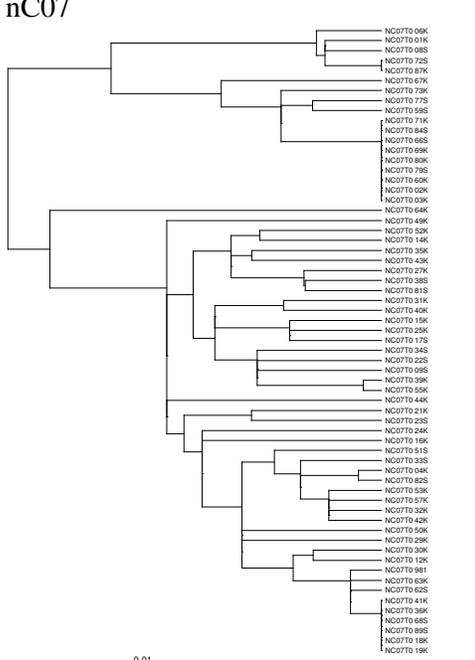
eC07



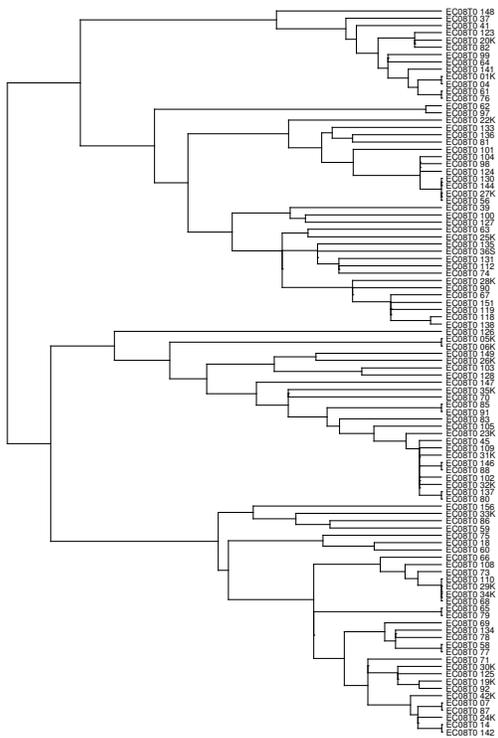
nC06



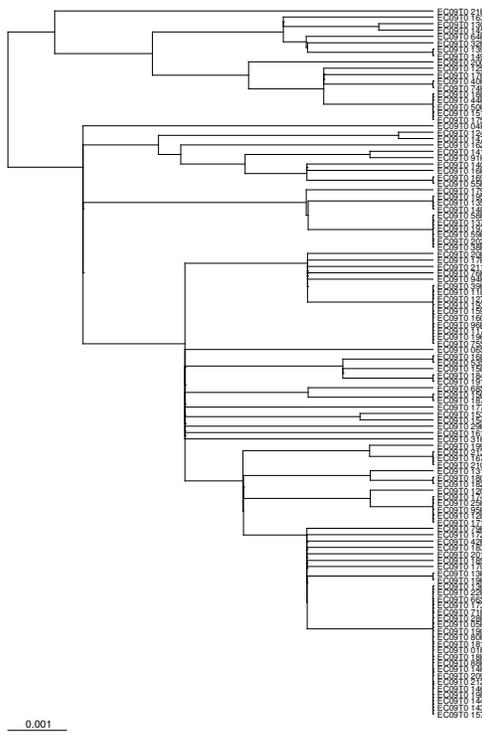
nC07



eC08



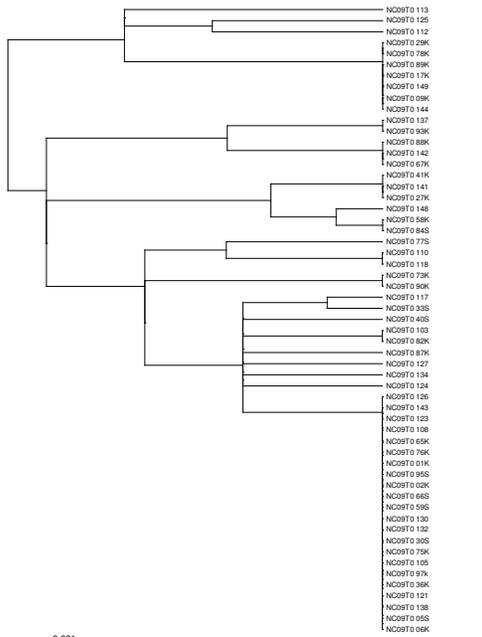
eC09



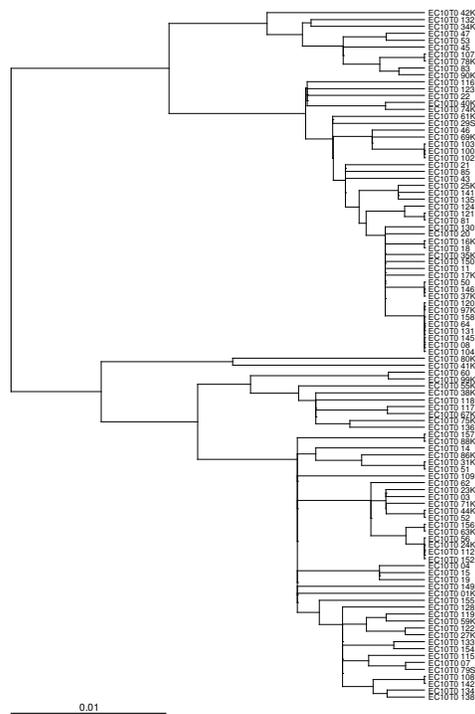
nC08



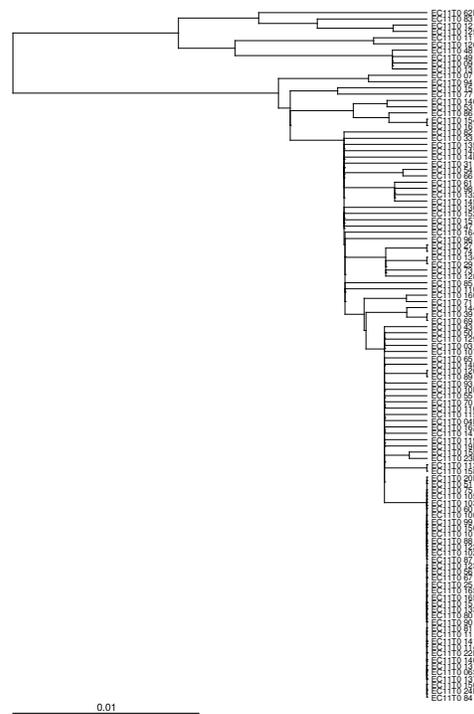
nC09



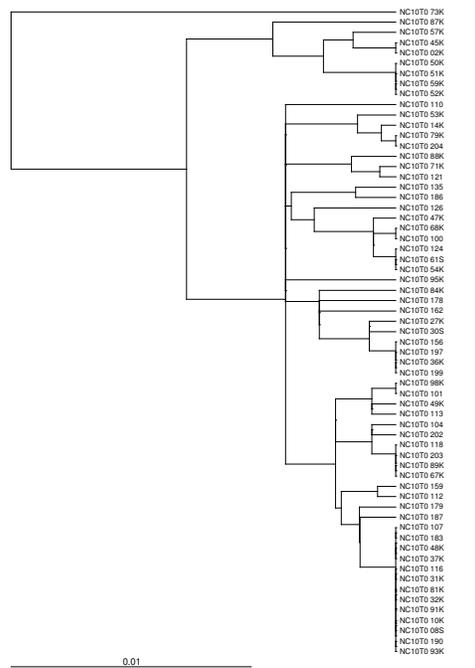
eC10



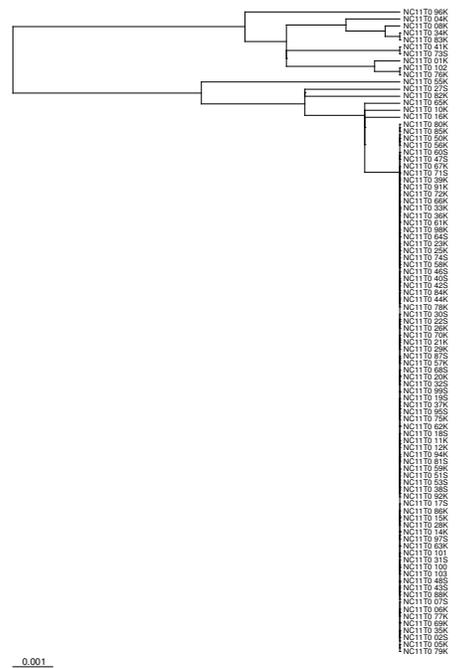
eC11



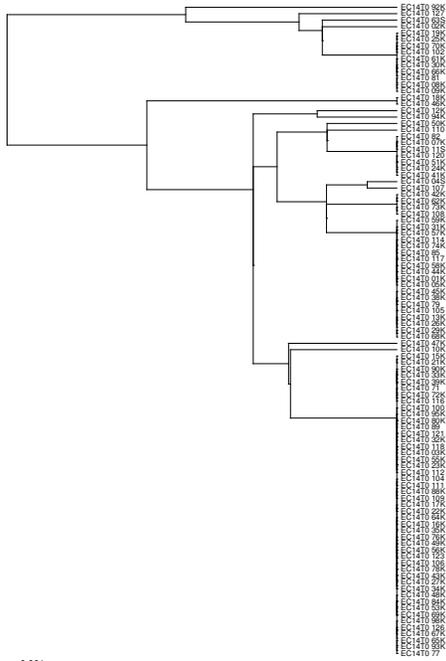
nC10



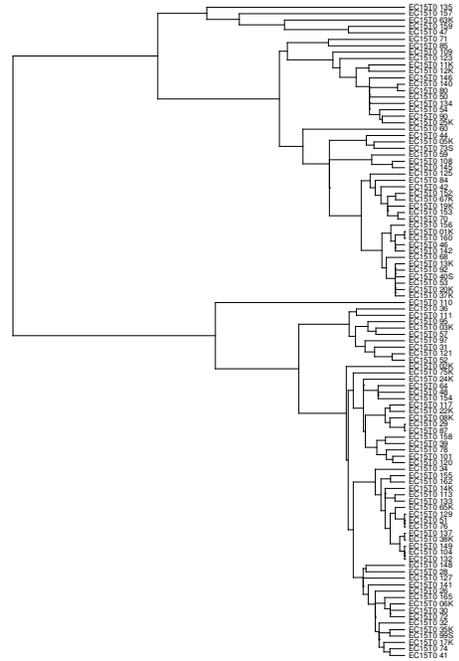
nC11



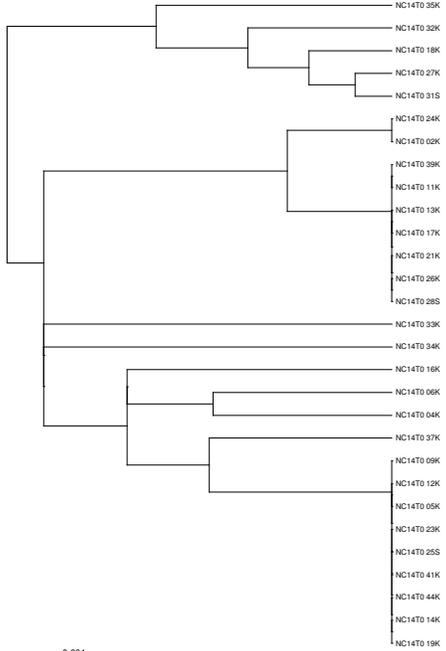
eC14



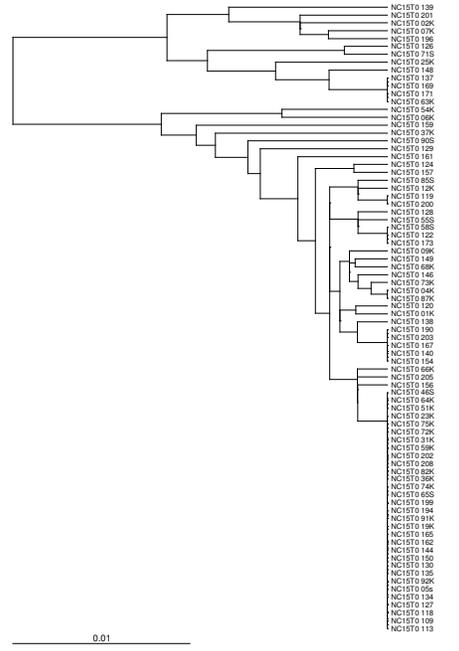
eC15



nC14



nC15



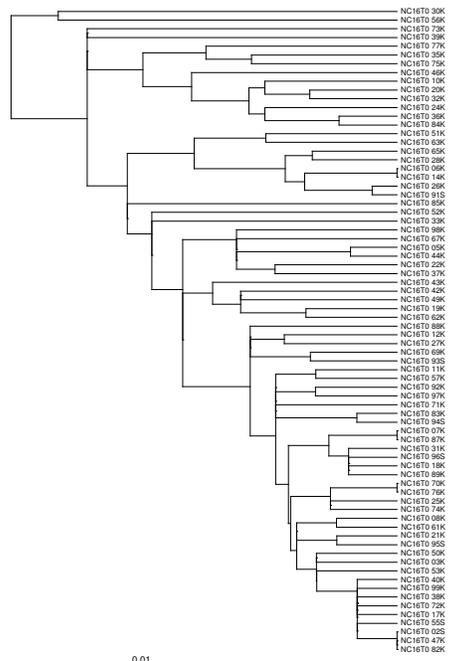
eC16



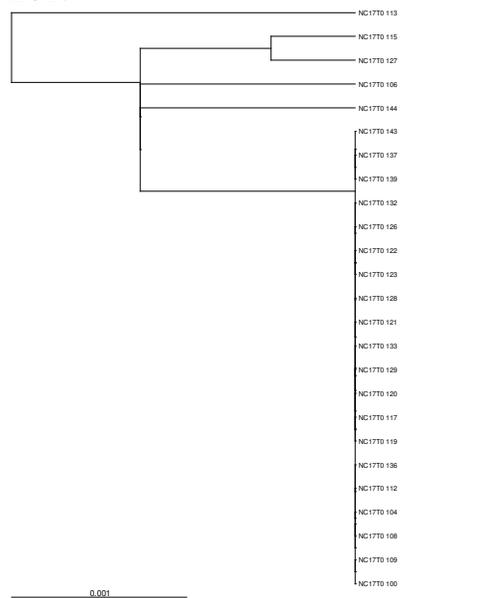
eC17



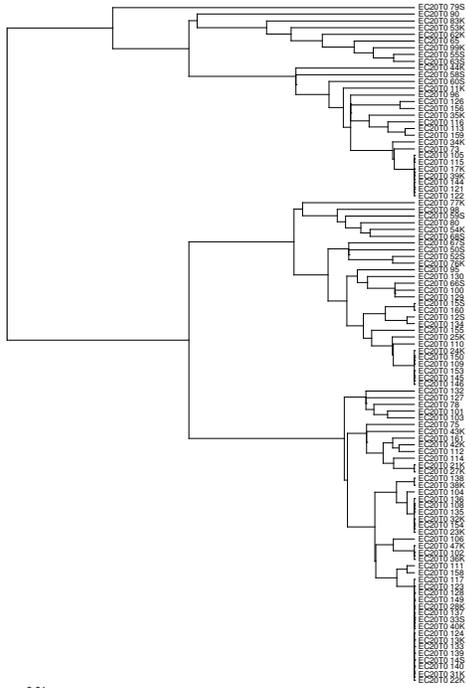
nC16



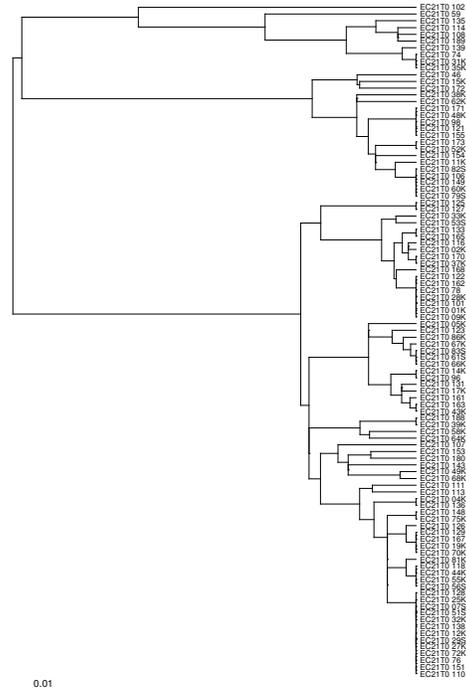
nC17



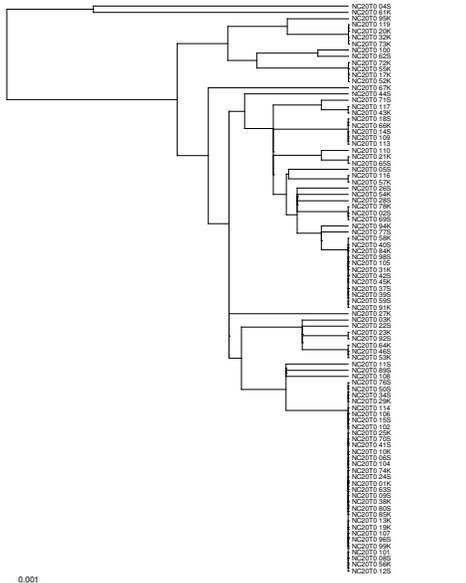
eC20



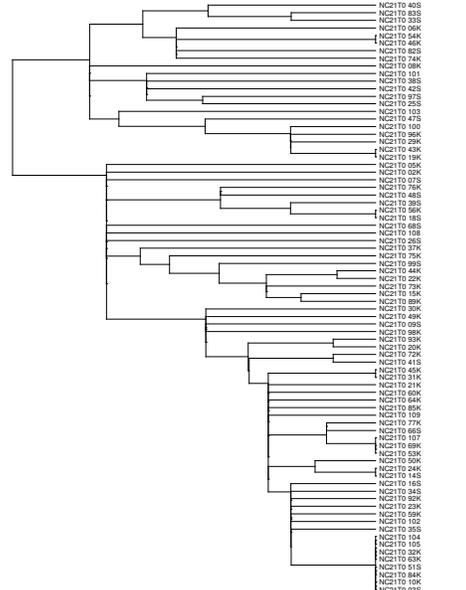
eC21



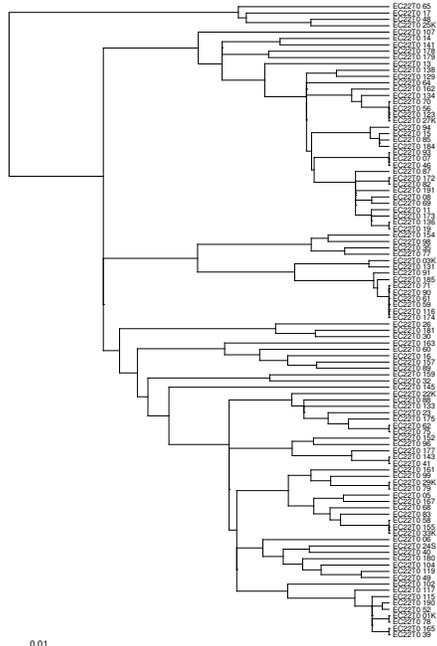
nC20



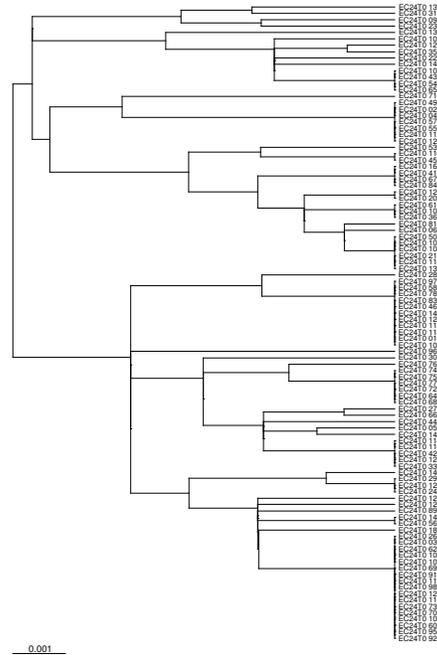
nC21



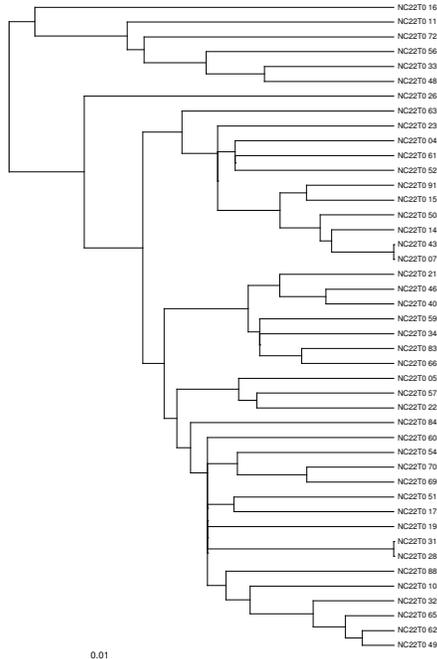
eC22



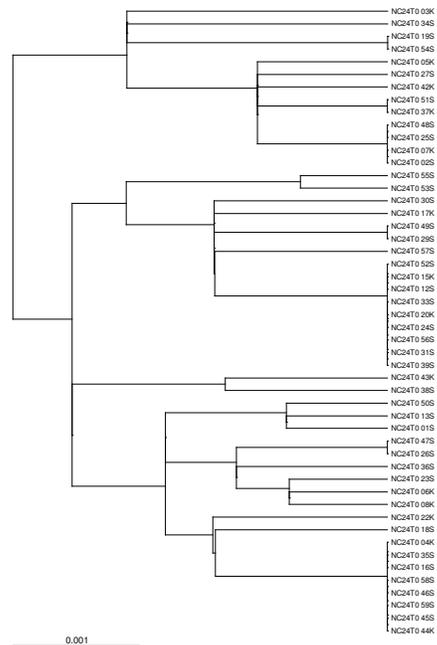
eC24

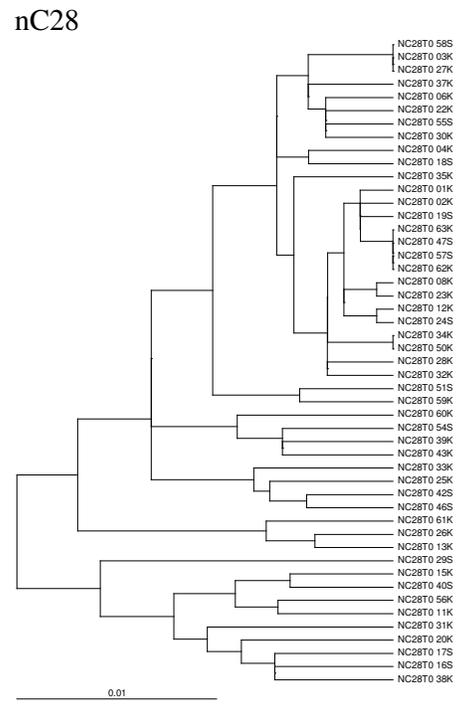
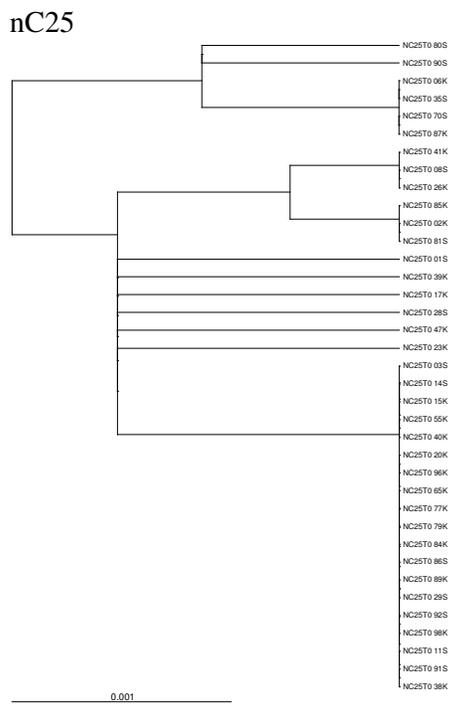
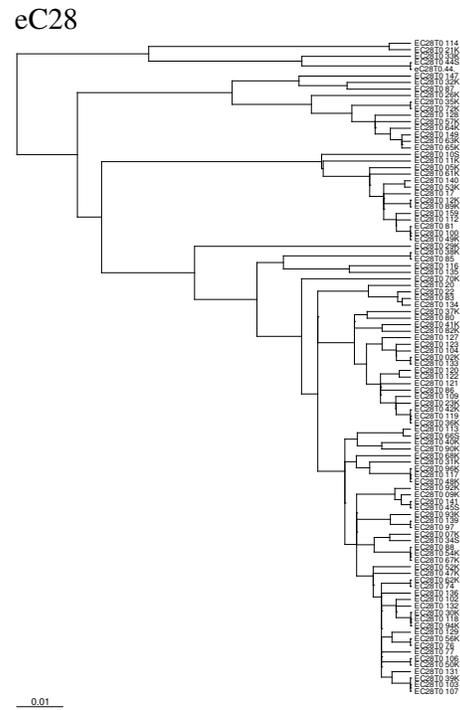
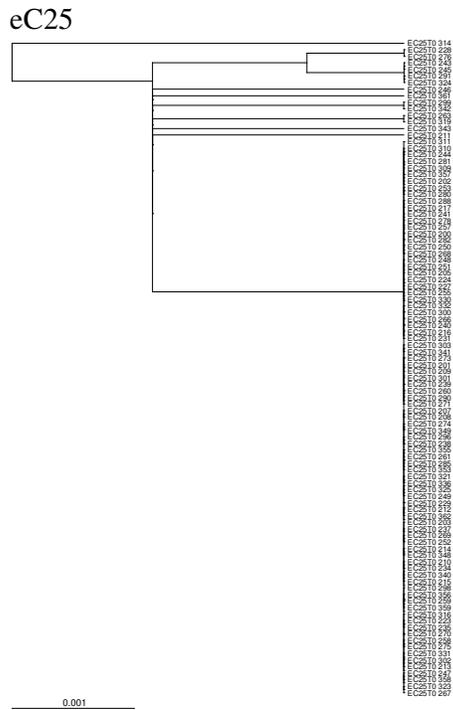


nC22

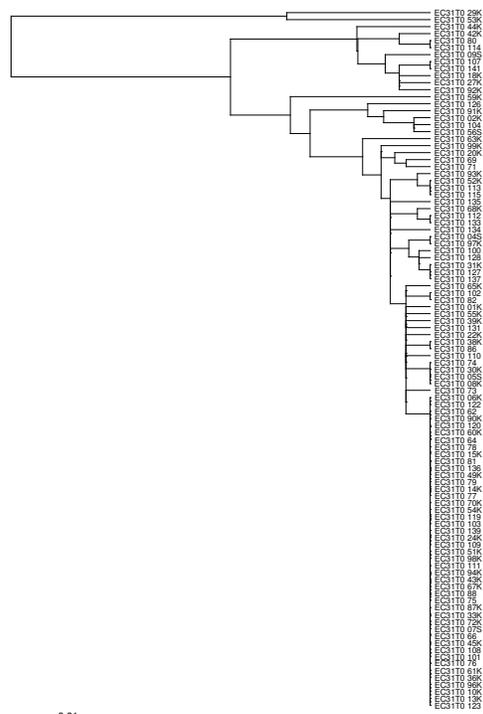


nC24



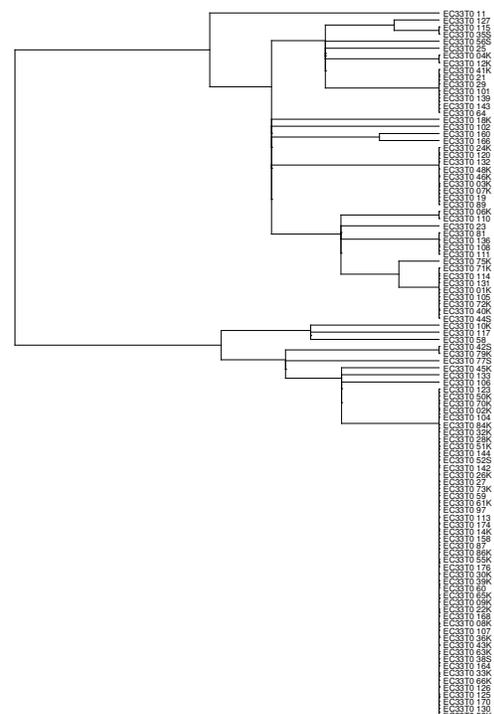


eC31



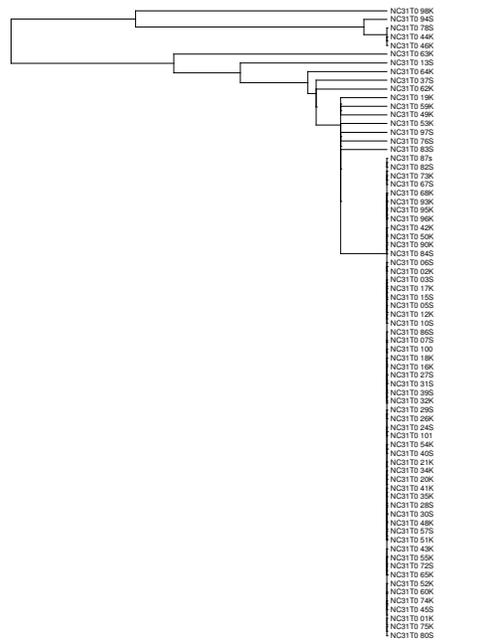
0.01

eC33



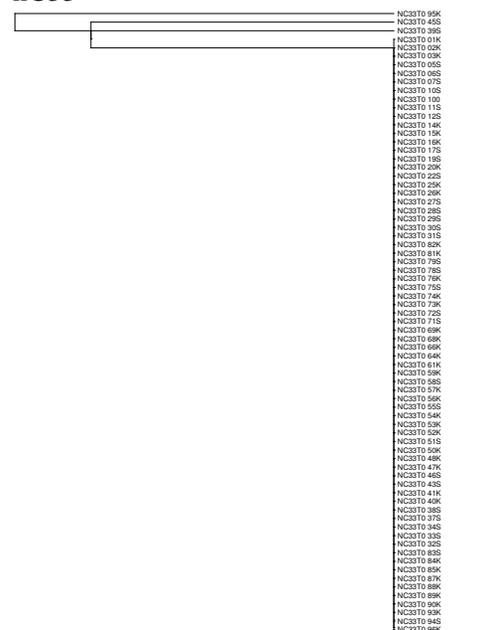
0.001

nC31



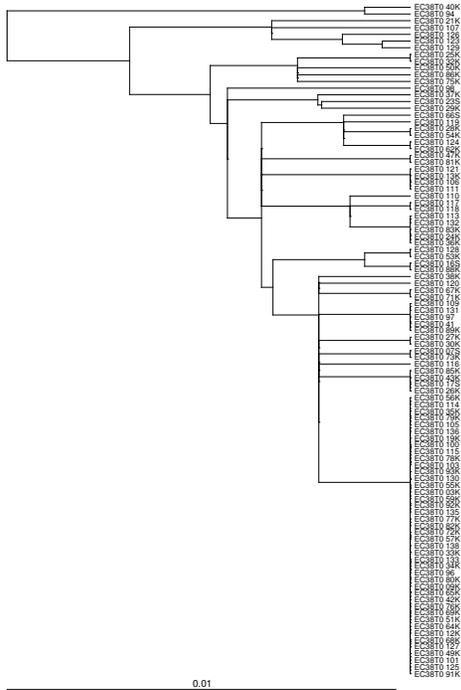
0.01

nC33

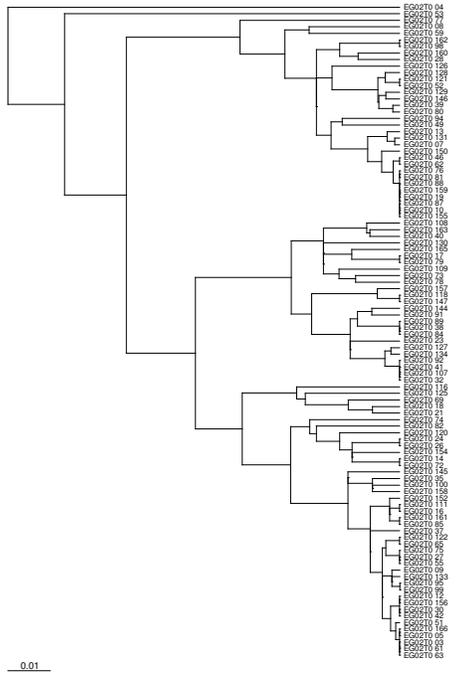


0.001

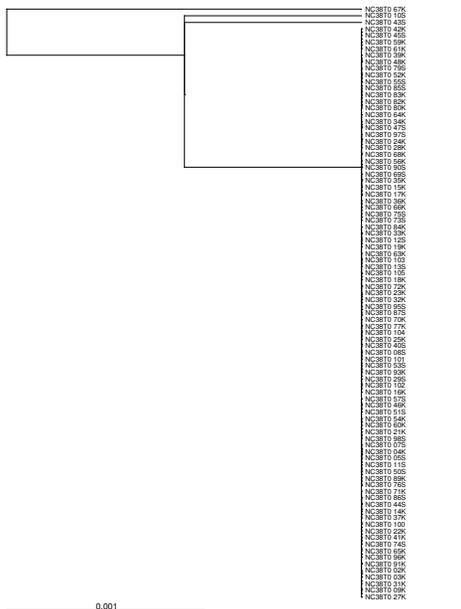
eC38



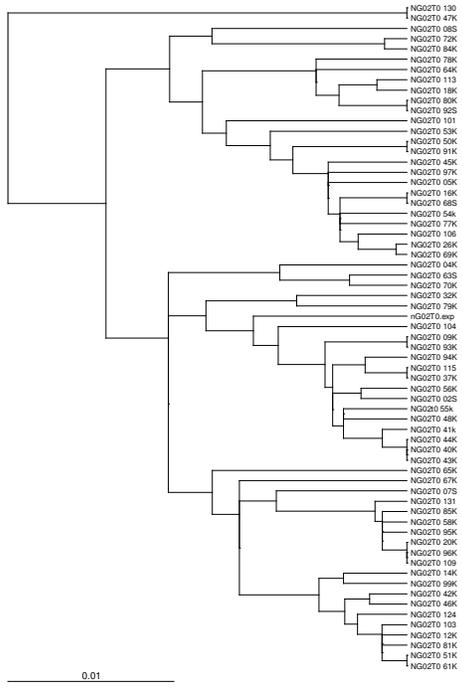
eG02



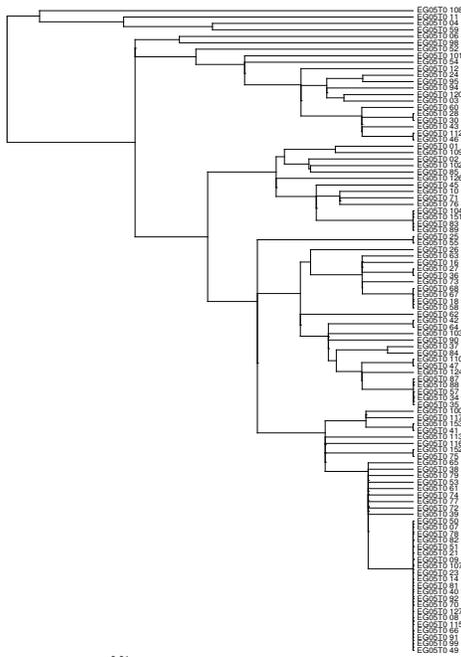
nC38



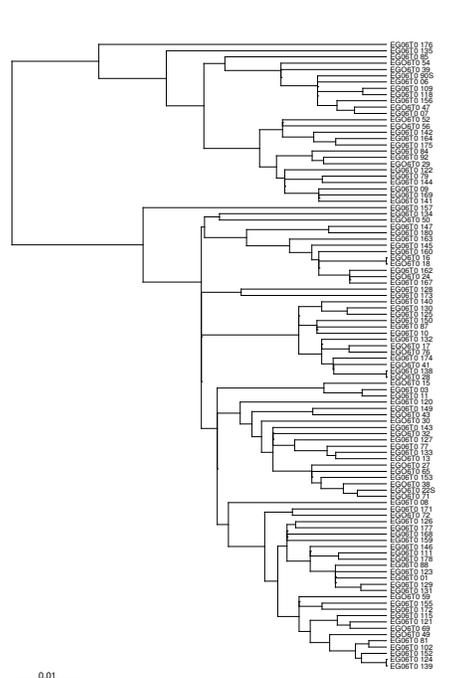
nG02



eG05



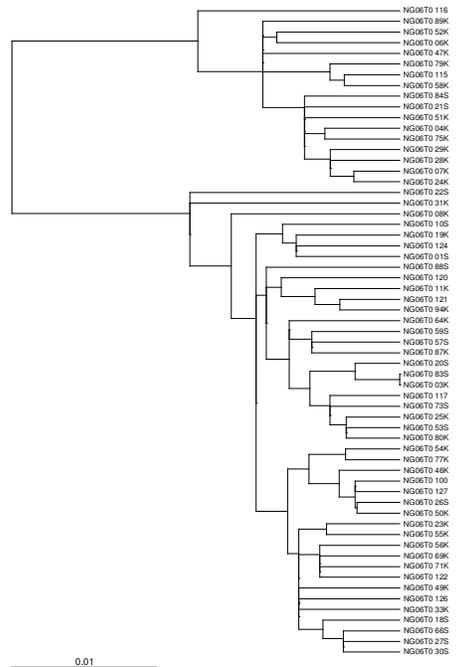
eG06



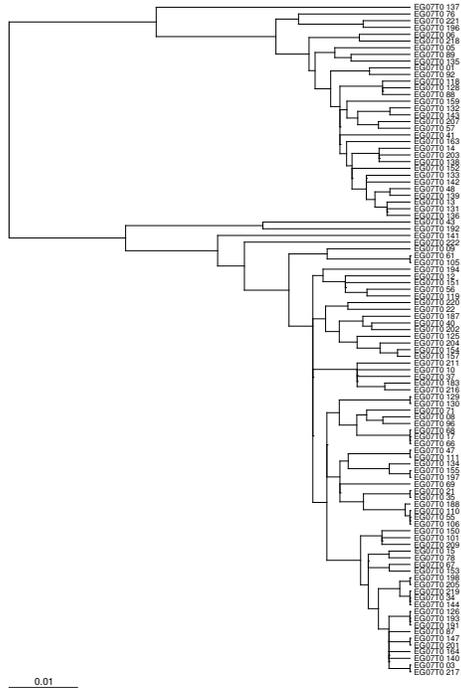
nG05



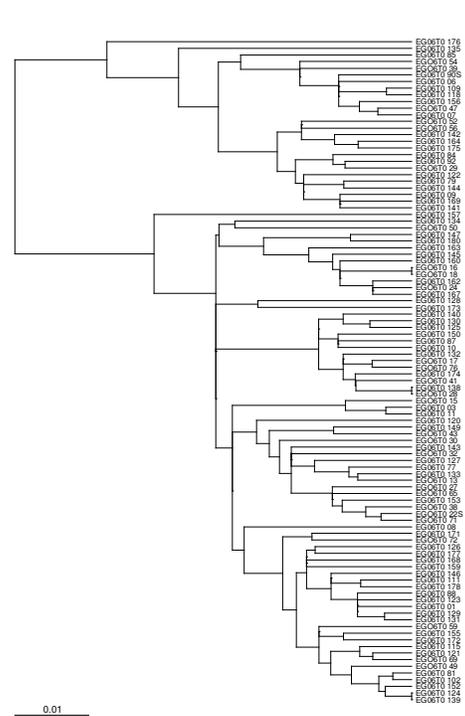
nG06



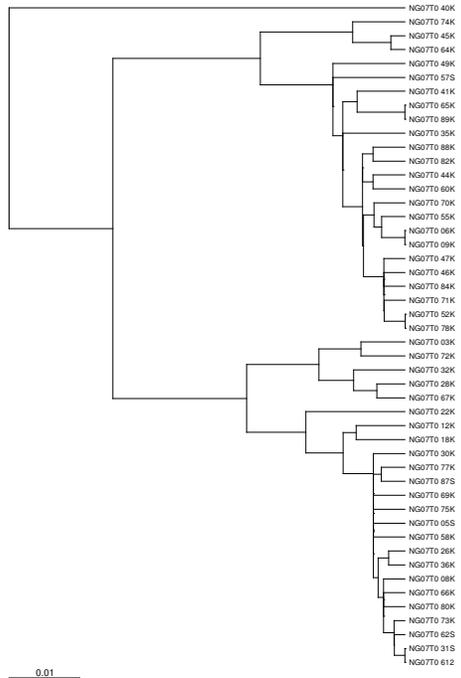
eG07



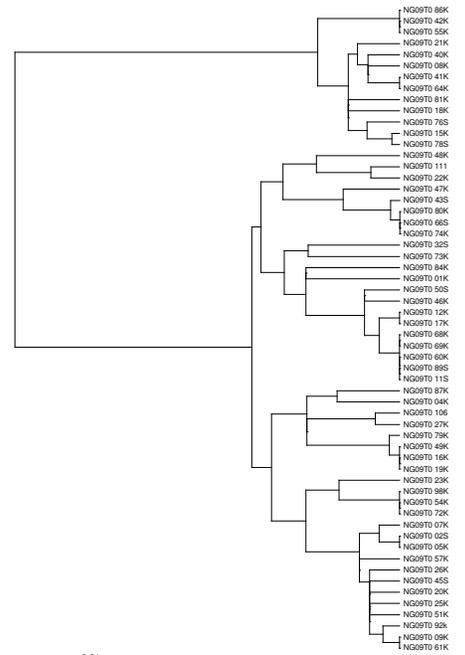
eG09



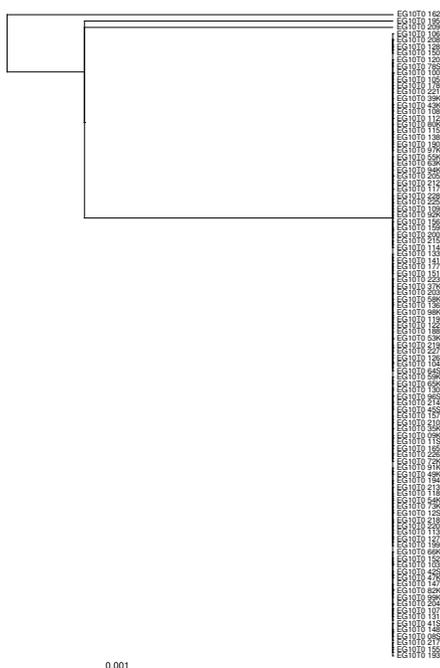
nG07



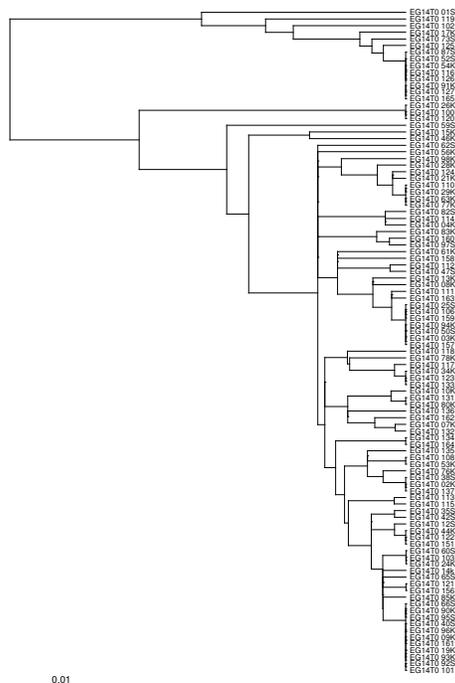
nG09



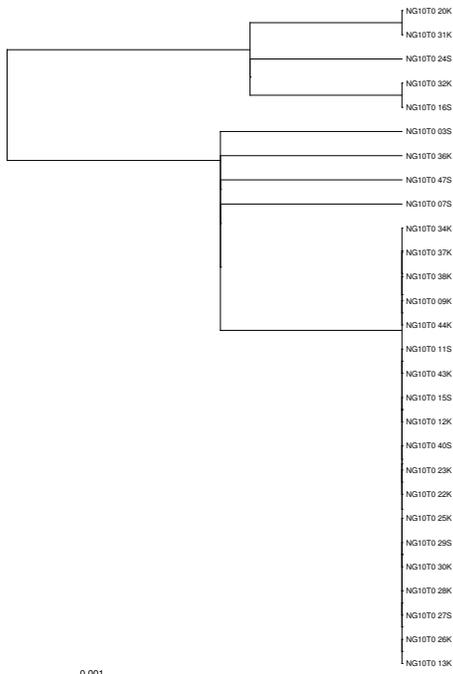
eG10



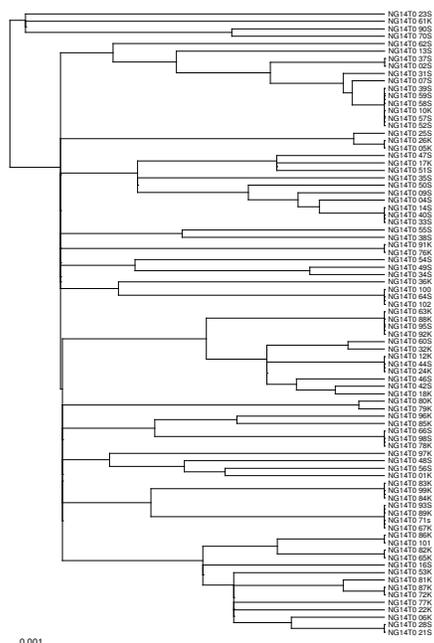
eG14



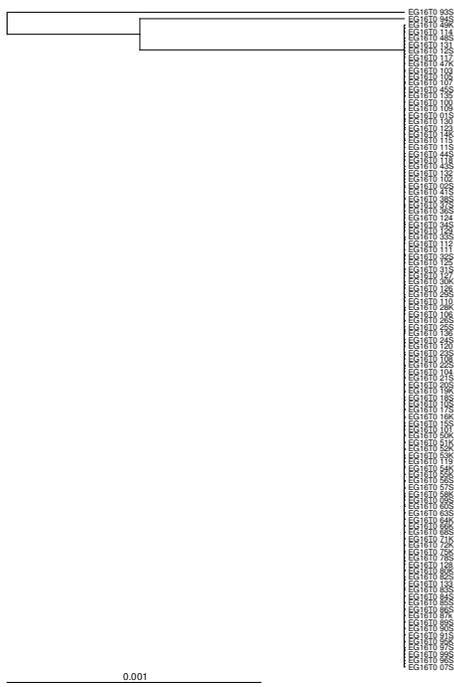
nG10



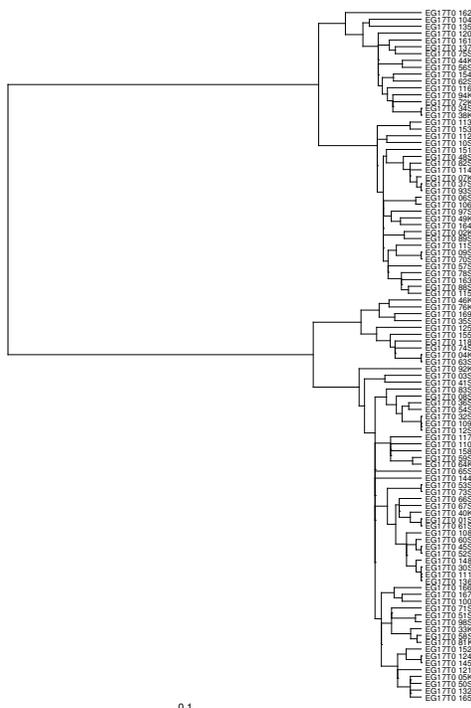
nG14



eG16



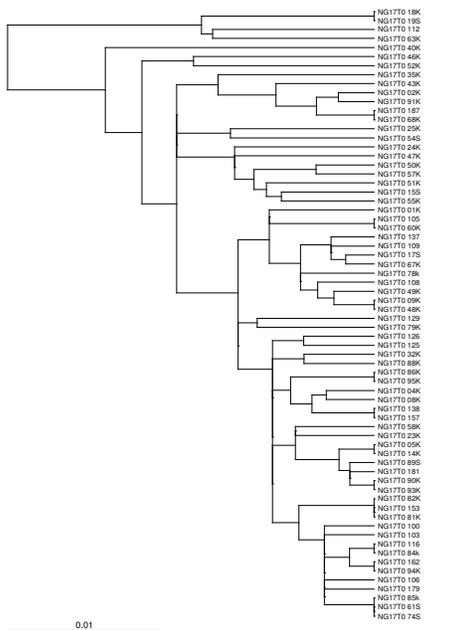
eG17



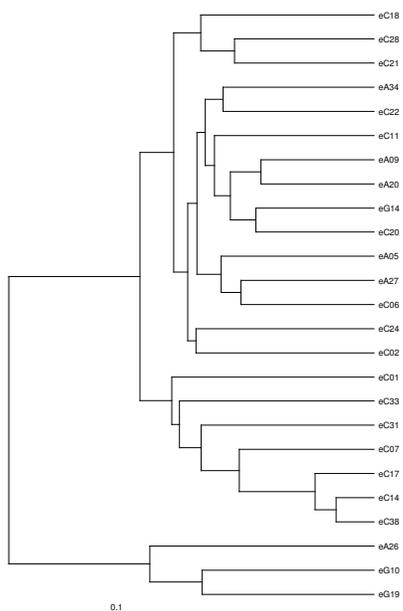
nG16



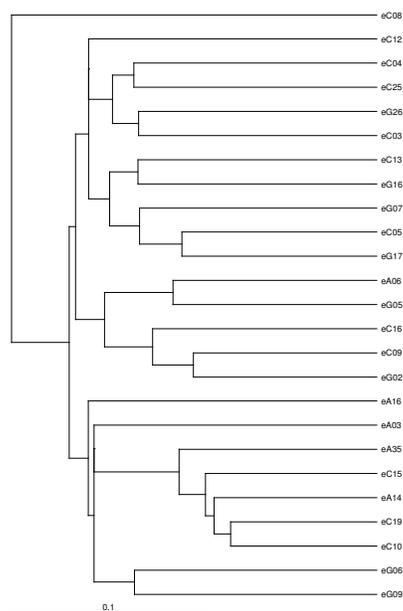
nG17



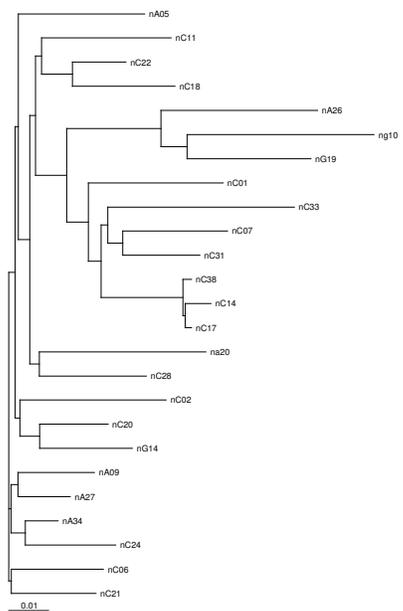
ConsE1a



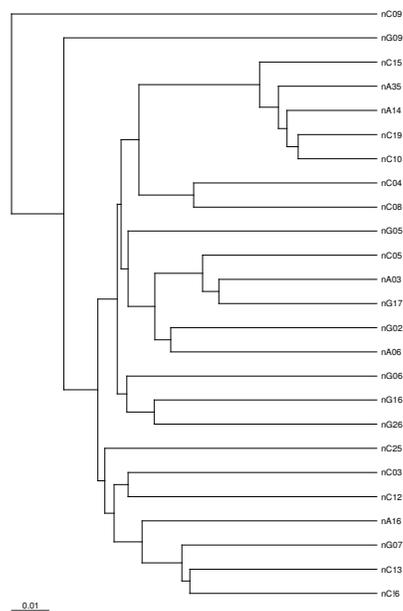
ConsE1b



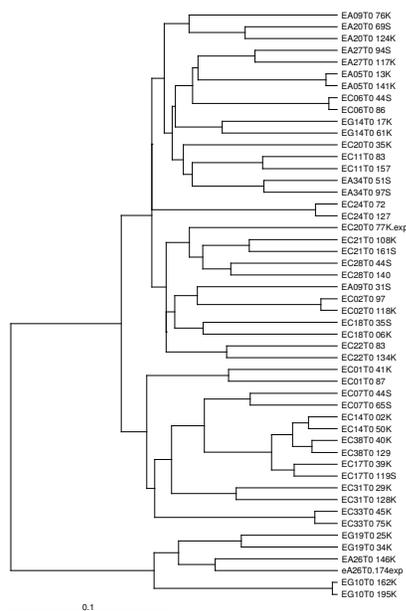
ConsN1a



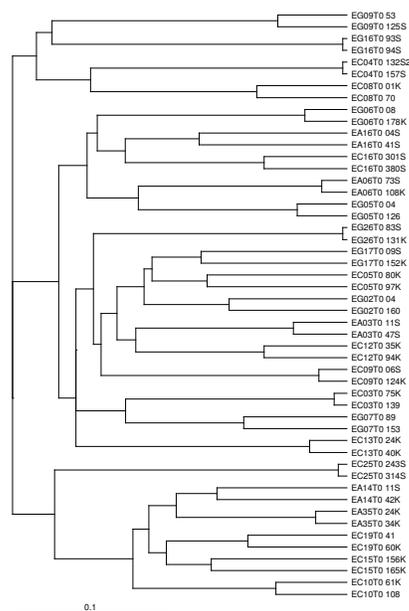
ConsN1b



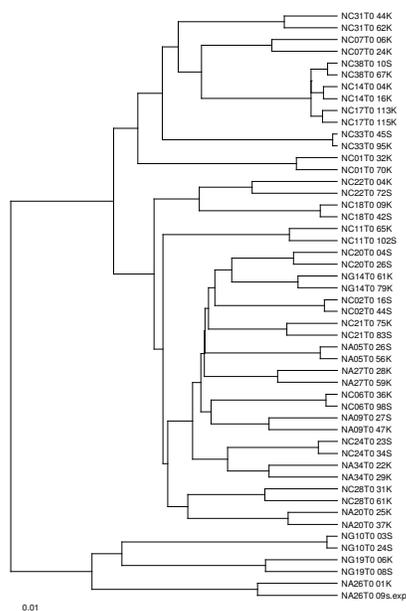
MasdistE1a



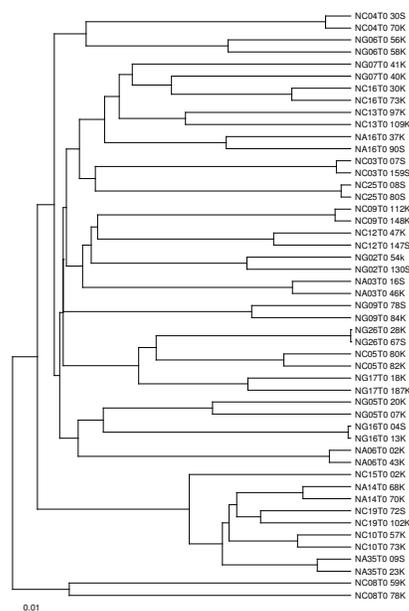
MasdistE1b



MasdistN1a



MasdistN1b



ANEXO VII

Estima de los parámetros correspondientes a los modelos de evolución de codones implementados por CODEML.

Tabla 14. Parámetros correspondientes a los modelos 0, 1, 2, 3, 7 y 8 calculados para la región E1-E2 de los individuos 1b.

individuo	modelo	Ln L	dN/dS	$w_0/w_1/w_2$	$p_0/p_1/p_2$	p-q
eA03	0	-4715,153	0,208			
	1	-4241,055	0,150	w: 0,00000 1,00000	p: 0,85043 0,14957	
	2	-4154,592	0,288	w: 0,00000 1,00000 3,64392	p: 0,85055 0,09689 0,05256	
	3	-4121,176	0,205	w: 0,00460 0,58730 3,10936	p: 0,88438 0,06306 0,05255	
	7	-4196,865	0,110			p= 0,00951 q= 0,06239
8	-4122,824	0,207	w: 3,11825	p: 0,95183 0,04817	p= 0,03679 q= 0,54623	
eA06	0	-1102,617	0,291			
	1	-1090,612	0,241	w: 0,00000 1,00000	p: 0,75910 0,24090	
	2	-1090,181	0,283	w: 0,00000 1,00000 1,37808	p: 0,78760 0,02501 0,18739	
	3	-1090,178	0,283	w: 0,00001 0,00711 1,37780	p: 0,121640 0,67619 0,20217	
	7	-1090,907	0,209		p: 0,00806 q: 0,02604	p= 0,00100 q= 2,22670
8	-1090,181	0,283	w: 1,34050	p: 0,78877 0,21123		
eA14	0	-5642,178	0,598			
	1	-4914,722	0,192	w: 0,00000 1,00000	p: 0,80778 0,19222	
	2	-4582,394	0,643	w: 0,00000 1,00000 7,31252	p: 0,811130 0,116930 0,07194	
	3	-4502,058	0,623	w: 0,01329 2,96170 14,58821	p: 0,87349 0,10617 0,02034	
	7	-4854,395	0,105			p= 0,01091 q= 0,07703
8	-4550,739	0,431	w: 7,56422	p: 0,95661 0,04339	p= 0,00823 q= 0,05391	
eA16	0	-7941,129	0,322			
	1	-7091,121	0,268	w: 0,00000 1,00000	p: 0,73163 0,26837	
	2	-6833,010	0,494	w: 0,00000 1,00000 4,94146	p: 0,75256 0,18492 0,06253	
	3	-6672,032	0,343	w: 0,01384 1,39995 6,90841	p: 0,85628 0,12005 0,02367	
	7	-6884,067	0,109			p= 0,01075 q= 0,07222
8	-6674,191	0,261	w: 5,77729	p: 0,97347 0,02653	p= 0,00810 q= 0,05170	
eA35	0	-5146,905	0,329			
	1	-4679,611	0,229	w: 0,00000 1,00000	p: 0,77055 0,22945	
	2	-4430,986	0,548	w: 0,00000 1,00000 6,98781	p: 0,78822 0,15560 0,05618	
	3	-4305,576	0,334	w: 0,03042 2,91871 12,66837	p: 0,93246 0,05646 0,01108	
	7	-4538,665	0,109			p= 0,01009 q= 0,06743
8	-4362,946	0,330	w: 4,54613	p: 0,93065 0,06935	p= 0,00834 q= 0,20563	
eC03	0	-1499,238	0,910			
	1	-1351,208	0,043	w: 0,00000 1,00000	p: 0,95681 0,04319	
	2	-1138,218	0,985	w: 0,00000 1,00000 72,82880	p: 0,94484 0,04222 0,01294	
	3	-1136,770	0,928	w: 0,02894 48,31103 69,68123	p: 0,98709 0,00000 0,01291	
	7	-1355,133	0,101			p= 0,00623 q= 0,04475
8	-1136,781	0,928	w: 69,70262	p: 0,98709 0,01291	p= 2,99757 q= 99,00000	
eC04	0	-664,941	999,000			
	1	-666,273	1,000	w: 0,00000 1,00000	p: 0,00000 1,00000	
	2	-664,941	999,000	w: 0,00000 1,00000 999,00000	p: 0,00000 0,00000 1,00000	
	3	-664,941	999,000	w: 0,00001 0,00001 999,00000	p: 0,00000 0,00000 1,00000	
	7	-666,273	1,000			p= 4,24926 q= 0,00101
8	-664,941	998,999	w: 999,00000	p: 0,00000 1,00000	p= 0,00100 q= 2,06743	

Ln L: Logaritmo de la verosimilitud

dN/dS: relación tasa de sustitución no sinónima/sinónima calculada como media en la longitud del fragmento.

w0: dN/dS correspondiente a los sitios neutros; w1: dN/dS correspondiente a los sitios bajo selección purificadora; w2: dN/dS correspondiente a los sitios bajo selección positiva.

p0: Proporción de sitios con w0; p1: Proporción de sitios con w1; p2: Proporción de sitios con w2

p y q: Parámetros propios de la distribución beta.

Tabla 14. (Continuación)

individuo modelo	Ln L	dN/dS	$w_0/w_1/w_2$	$p_0/p_1/p_2$	p-q	
eC05	0	-8258,034	0,471			
	1	-7282,908	0,260	w: 0,00000 1,00000	p: 0,74041 0,25959	
	2	-6762,580	0,717	w: 0,00000 1,00000 7,24457	p: 0,74873 0,17673 0,07454	
	3	-6597,759	0,487	w: 0,01836 2,10931 8,77930	p: 0,88402 0,08214 0,03384	
	7	-7030,472	0,105			p= 0,00799 q= 0,05311
8	-7024,582	0,114	w: 0,47971	p: 0,97591 0,02409	p= 0,00713 q= 0,04679	
eC08	0	-7884,332	0,755			
	1	-6774,132	0,203	w: 0,00000 1,00000	p: 0,79706 0,20294	
	2	-6189,886	0,718	w: 0,00000 1,00000 8,82324	p: 0,80663 0,12633 0,06704	
	3	-6091,214	0,877	w: 0,01413 3,24558 18,81623	p: 0,86012 0,11349 0,02639	
	7	-6712,744	0,122			p= 0,01379 q= 0,08749
8	-6170,017	0,628	w: 8,03387	p: 0,93476 0,06524	p= 0,01175 q= 0,07860	
eC09	0	-2074,947	0,537			
	1	-1928,642	0,146	w: 0,00000 1,00000	p: 0,85398 0,14602	
	2	-1766,366	0,541	w: 0,00000 1,00000 25,69311	p: 0,83131 0,15362 0,01507	
	3	-1759,964	0,540	w: 0,09590 5,77705 35,06210	p: 0,96574 0,02574 0,00852	
	7	-1923,644	0,119			p= 0,00940 q= 0,05826
8	-1923,060	0,107	w: 0,14195	p: 0,82612 0,17388	p= 0,00191 q= 0,01368	
eC10	0	-5225,828	0,450			
	1	-4086,868	0,295	w: 0,00000 1,00000	p: 0,70499 0,29501	
	2	-3833,674	0,699	w: 0,00000 1,00000 15,03470	p: 0,88619 0,07209 0,04172	
	3	-4234,221	0,842	w: 0,00001 0,96686 1,87805	p: 0,35929 0,39633 0,24438	
	7	-4071,133	0,305			p= 0,00681 q= 0,01342
8	-3824,382	0,638	w: 13,91805	p: 0,96154 0,03846	p= 0,01373 q= 0,10064	
eC12	0	-4946,205	0,241			
	1	-4153,543	0,182	w: 0,00000 1,00000	p: 0,81789 0,18211	
	2	-3854,566	0,361	w: 0,00000 1,00000 23,81692	p: 0,83372 0,15774 0,00854	
	3	-3790,593	0,289	w: 0,00097 1,81100 44,65528	p: 0,91599 0,08084 0,00317	
	7	-4073,099	0,111			p= 0,01813 q= 0,13225
8	-3795,511	0,204	w: 37,40068	p: 0,99757 0,00243	p= 0,01070 q= 0,06937	
eC13	0	-2570,038	0,237			
	1	-2427,210	0,193	w: 0,00000 1,00000	p: 0,80718 0,19282	
	2	-2356,425	0,332	w: 0,00000 1,00000 8,97284	p: 0,80153 0,18172 0,01675	
	3	-2341,230	0,241	w: 0,03261 1,18202 7,78943	p: 0,90632 0,07840 0,01527	
	7	-2406,227	0,118			p= 0,01560 q= 0,10340
8	-2342,925	0,231	w: 7,53719	p: 0,98511 0,01489	p= 0,06426 q= 0,47043	
eC15	0	-9694,093	0,333			
	1	-8263,008	0,183	w: 0,00000 1,00000	p: 0,81689 0,18311	
	2	-7685,599	0,548	w: 0,00000 1,00000 7,44171	p: 0,81642 0,12709 0,05649	
	3	-7516,726	0,357	w: 0,00335 0,11629 4,43701	p: 0,85069 0,07144 0,07787	
	7	-7883,800	0,105			p= 0,00748 q= 0,04946
8	-7516,852	0,357	w: 4,43723	p: 0,92248 0,07752	p= 0,09397 q= 5,30139	
eC16	0	-4044,847	0,345		0,01489	
	1	-3585,683	0,161	w: 0,00000 1,00000	p: 0,83951 0,16049	
	2	-3391,522	0,432	w: 0,00000 1,00000 8,48402	p: 0,85356 0,10835 0,03809	
	3	-3375,187	0,413	w: 0,01201 1,41834 8,28661	p: 0,89547 0,06761 0,03692	
	7	-3551,331	0,107			p= 0,01084 q= 0,07489
8	-3375,723	0,359	w: 7,61188	p: 0,96668 0,03332	p= 0,01265 q= 0,08787	
eC19	0	-7154,518	0,484			
	1	-6347,076	0,256	w: 0,00000 1,00000	p: 0,74377 0,25623	
	2	-5885,599	0,839	w: 0,00000 1,00000 7,46106	p: 0,75808 0,14951 0,09241	
	3	-5792,102	0,503	w: 0,01472 0,50582 5,18308	p: 0,85711 0,05348 0,08941	
	7	-6106,251	0,107			p= 0,00886 q= 0,05864
8	-5793,770	0,503	w: 4,98600	p: 0,90567 0,09433	p= 0,12734 q= 2,98650	

Tabla 14. (Continuación)

individuo modelo	Ln L	dN/dS	w ₀ /w ₁ /w ₂	p ₀ /p ₁ /p ₂	p-q	
eC25	0	-767,331	0,277			
	1	-758,714	0,104	w: 0,00000 1,00000	p: 0,89648 0,10352	
	2	-750,329	0,274	w: 0,00000 1,00000 35,87674	p: 0,89737 0,09773 0,00490	
	3	-750,070	0,303	w: 0,00001 2,31854 38,70140	p: 0,94431 0,05091 0,00478	
	7	-758,716	0,100			p= 0,00138 q= 0,01330
8	-752,654	0,283	w: 8,79249	p: 0,96786 0,03214	p= 0,00100 q= 1,98920	
eG02	0	-7619,816	0,301			
	1	-6538,301	0,176	w: 0,00000 1,00000	p: 0,82370 0,17630	
	2	-6241,782	0,384	w: 0,00000 1,00000 5,54450	p: 0,84600 0,10348 0,05053	
	3	-6158,667	0,332	w: 0,00752 1,80248 8,16290	p: 0,90135 0,07546 0,02319	
	7	-6415,928	0,111			p= 0,01107 q= 0,07348
8	-6169,319	0,281	w: 5,22575	p: 0,96611 0,03389	p= 0,01195 q= 0,08368	
eG05	0	-3764,709	0,010			
	1	-2958,792	0,149	w: 0,00000 1,00000	p: 0,85137 0,14863	
	2	-2740,817	0,468	w: 0,00000 1,00000 21,58610	p: 0,85267 0,13173 0,01560	
	3	-2737,849	0,624	w: 0,00086 0,04712 13,51189	p: 0,87746 0,07672 0,04582	
	7	-2932,794	0,137			p= 0,02661 q= 0,16217
8	-2737,964	0,624	w: 13,42162	p: 0,95386 0,04614	p= 0,06623 q= 10,03342	
eG06	0	-5030,698	0,310			
	1	-4490,605	0,174	w: 0,00000 1,00000	p: 0,82569 0,17431	
	2	-4236,768	0,475	w: 0,00000 1,00000 7,12655	p: 0,83470 0,11470 0,05059	
	3	-4181,666	0,311	w: 0,00988 0,56434 5,63712	p: 0,89972 0,05188 0,04840	
	7	-4395,159	0,107			p= 0,01150 q= 0,07980
8	-4185,116	0,312	w: 5,63100	p: 0,95464 0,04536	p= 0,05773 q= 0,83765	
eG07	0	-7304,205	0,149			
	1	-5784,789	0,138	w: 0,00000 1,00000	p: 0,86182 0,13818	
	2	-5662,015	0,176	w: 0,00000 1,00000 2,25238	p: 0,89593 0,04663 0,05744	
	3	-5443,774	0,179	w: 0,00014 0,12845 3,23818	p: 0,93240 0,01278 0,05481	
	7	-5589,141	0,111			p= 0,02266 q= 0,17542
8	-5448,400	0,178	w: 3,22892	p: 0,94486 0,05514	p= 0,00924 q= 1,86284	
eG09gp	0	-5107,825	0,263			
	1	-4099,659	0,220	w: 0,00000 1,00000	p: 0,78007 0,21993	
	2	-3978,432	0,447	w: 0,00000 1,00000 5,47472	p: 0,78590 0,16204 0,05206	
	3	-3945,781	0,259	w: 0,00001 0,00129 2,94553	p: 0,64128 0,27101 0,08771	
	7	-4006,283	0,122			p= 0,02188 q= 0,14768
8	-3910,128	0,260	w: 4,14803	p: 0,94969 0,05031	p= 0,01718 q= 0,29631	
eG16gp	0	-648,966	999,000			
	1	-649,454	1,000	w: 0,00000 1,00000	p: 0,00000 1,00000	
	2	-648,966	999,000	w: 0,000 1,000 999,000	p: 0,00000 0,00000 1,00000	
	3	-648,966	999,000	w: 0,00001 0,00001 999,000	p: 0,00000 0,00000 1,00000	
	7	-649,454	1,000			p= 4,05914 q= 0,00100
8	-648,966	998,999	w: 999,00000	p: 0,00000 1,00000	p= 0,00100 q= 1,79758	
eG17gp	0	-7883,050	0,240			
	1	-6693,093	0,140	w: 0,00000 1,00000	p: 0,86000 0,14000	
	2	-6284,197	0,349	w: 0,00000 1,00000 7,58953	p: 0,86370 0,10406 0,03224	
	3	-6191,110	0,290	w: 0,00271 0,23523 4,68565	p: 0,90073 0,03995 0,05931	
	7	-6510,521	0,106			p= 0,00896 q= 0,06010
8	-6194,426	0,290	w: 4,68555	p: 0,94211 0,05789	p= 0,05107 q= 1,80542	
eG26gp	0	-688,650	1,129			
	1	-688,546	0,713	w: 0,00000 1,00000	p: 0,28728 0,71272	
	2	-687,507	1,202	w: 0,00000 1,00000 37,75316	p: 0,22750 0,76081 0,01169	
	3	-687,502	1,189	w: 0,00001 0,74747 37,07515	p: 0,00000 0,98785 0,01215	
	7	-688,662	1,000			p= 3,70622 q= 0,00102
8	-687,619	1,128	w: 6,78061	p: 0,83366 0,16634	p= 0,00100 q= 1,96334	

Tabla 15. Parámetros correspondientes a los modelos 0, 1, 2, 3, 7 y 8 para la región E1-E2 de los individuos 1a.

individuo modelo	Ln L	dN/dS	w ₀ /w ₁ /w ₂	p ₀ /p ₁ /p ₂	p-q	
eA05	0	-851,240	0,036			
	1	-845,179	0,028	w: 0,00000 1,00000	p: 0,97209 0,02791	
	2	-845,045	0,034	w: 0,00000 1,00000 1,46621	p: 0,97693 0,00000 0,02307	
	3	-845,046	0,034	w: 0,00001 0,00001 1,46625	p: 0,14300 0,83392 0,02307	
	7	-846,730				p= 0,00926 q= 0,18308
8	-845,045	0,034	w= 1,46621	p: 0,97693 0,02307	p= 0,00100 q= 2,11265	
eA09	0	-2359,022	0,223			
	1	-2214,958	0,194	w: 0,00000 1,00000	p: 0,80560 0,19440	
	2	-2180,508	0,283	w: 0,00000 1,00000 6,98492	p: 0,80741 0,17741 0,01518	
	3	-2174,603	0,247	w: 0,01315 1,07795 6,54670	p: 0,85397 0,13172 0,01431	
	7	-2208,458	0,117			p= 0,01425 q= 0,09339
8	-2175,753	0,222	w= 6,21849	p: 0,98255 0,01745	p= 0,01682 q= 0,11548	
eA20	0	-4866,911	0,609			
	1	-4334,374	0,228	w: 0,00000 1,00000	p: 0,77246 0,22754	
	2	-3898,876	0,858	w: 0,00000 1,00000 15,22540	p: 0,78183 0,17320 0,04497	
	3	-3855,219	0,903	w: 0,05878 4,33509 26,69243	p: 0,90500 0,07541 0,01959	
	7	-4255,448	0,113			p= 0,01283 q= 0,08558
8	-3883,081	0,691	w= 12,94137	p: 0,95525 0,04475	p= 0,01611 q= 0,10763	
eA26	0	-6196,784	0,676			
	1	-5501,584	0,278	w: 0,00000 1,00000	p: 0,72244 0,27756	
	2	-5042,326	0,831	w: 0,00000 1,00000 9,68742	p: 0,73880 0,19559 0,06561	
	3	-4965,727	0,813	w: 0,04247 3,33952 17,09934	p: 0,86598 0,11018 0,02385	
	7	-5374,212	0,154			p= 0,04128 q= 0,22556
8	-4997,887	0,595	w= 8,05876	p: 0,93913 0,06087	p= 0,01297 q= 0,08819	
eA27	0	-3926,368	0,889			
	1	-3382,987	0,140	w: 0,00000 1,00000	p: 0,86032 0,13968	
	2	-3053,874	0,623	w: 0,00000 1,00000 12,67668	p: 0,87116 0,08652 0,04233	
	3	-3001,478	0,980	w: 0,01499 5,30023 28,24888	p: 0,90086 0,07994 0,01920	
	7	-3371,199	0,105			p= 0,00463 q= 0,02888
8	-3053,456	0,606	w= 12,42907	p: 0,95918 0,04082	p= 0,00122 q= 0,00714	
eA34	0	-5146,905	0,329			
	1	-4679,611	0,229	w: 0,00000 1,00000	p: 0,77055 0,22945	
	2	-4430,986	0,548	w: 0,00000 1,00000 6,98781	p: 0,78822 0,15560 0,05618	
	3	-4305,576	0,334	w: 0,03042 2,91871 12,66837	p: 0,93246 0,05646 0,01108	
	7	-4538,665	0,109			p= 0,01009 q= 0,06743
8	-4362,946	0,330	w= 4,54613	p: 0,93065 0,06935	p= 0,00834 q= 0,20563	
eC01	0	-4361,919	0,750			
	1	-3912,368	0,266	w: 0,00000 1,00000	p: 0,73439 0,26561	
	2	-3551,950	0,885	w: 0,00000 1,00000 11,70864	p: 0,74893 0,19192 0,05915	
	3	-3475,608	0,776	w: 0,05115 4,56486 23,84124	p: 0,90077 0,08487 0,01436	
	7	-3808,441	0,109			p= 0,00961 q= 0,06351
8	-3523,112	0,599	w= 9,12793	p: 0,94646 0,05354	p= 0,01888 q= 0,12999	
eC02	0	-987,794	0,139			
	1	-944,330	0,029	w: 0,00000 1,00000	p: 0,97059 0,02941	
	2	-922,938	0,147	w: 0,00000 1,00000 12,46875	p: 0,97717 0,01205 0,01078	
	3	-922,552	0,142	w: 0,00739 0,00739 12,26400	p: 0,48106 0,50800 0,01094	
	7	-948,361	0,100			p= 0,00100 q= 0,00742
8	-922,555	0,142	w= 12,26475	p: 0,98906 0,01094	p= 0,77258 q= 99,00000	

Ln L: Logaritmo de la verosimilitud

dN/dS: relación tasa de sustitución no sinónima/sinónima calculada como media en la longitud del fragmento.

w₀: dN/dS correspondiente a los sitios neutros; w₁: dN/dS correspondiente a los sitios bajo selección purificadora; w₂: dN/dS correspondiente a los sitios bajo selección positiva.

p₀: Proporción de sitios con w₀; p₁: Proporción de sitios con w₁; p₂: Proporción de sitios con w₂

p y q: Parámetros propios de la distribución beta.

Tabla 15. (Continuación).

individuo modelo	Ln L	dN/dS	w ₀ /w ₁ /w ₂	p ₀ /p ₁ /p ₂	p-q	
eC06	0	-832,779	0,120			
	1	-831,497	0,126	w: 0,00000 1,00000	p: 0,87435 0,12565	
	2	-831,495	0,124	w: 0,00000 1,00000 0,93916	p: 0,86811 0,00000 0,13189	
	3	-831,495	0,124	w: 0,00001 0,00001 0,93922	p: 0,38162 0,48651 0,13187	
	7	-831,536	0,120			p= 0,00843 q= 0,05167
8	-831,495	0,124	w= 0,93916	p: 0,86811 0,13189	p= 0,00100 q= 0,17934	
eC07	0	-5737,173	0,270			
	1	-4995,837	0,150	w: 0,00000 1,00000	p: 0,85021 0,14979	
	2	-4759,350	0,307	w: 0,00000 1,00000 8,12767	p: 0,86036 0,11616 0,02348	
	3	-4698,188	0,302	w: 0,00792 1,63716 8,18932	p: 0,91714 0,05864 0,02422	
	7	-4915,934	0,107			p= 0,00641 q= 0,04064
8	-4709,487	0,233	w= 7,04259	p: 0,98211 0,01789	p= 0,00996 q= 0,06591	
eC11	0	-2168,180	0,468			
	1	-2013,535	0,215	w: 0,00000 1,00000	p: 0,78539 0,21461	
	2	-1934,156	0,543	w: 0,00000 1,00000 7,77189	p: 0,78504 0,16659 0,04837	
	3	-1926,429	0,479	w: 0,04273 2,17253 13,06067	p: 0,88113 0,10203 0,01684	
	7	-2005,420	0,123			p= 0,01625 q= 0,10414
8	-1931,228	0,464	w= 6,84902	p: 0,95041 0,04959	p= 0,04944 q= 0,32914	
eC14	0	-1538,854	23,168			
	1	-1424,804	0,091	w: 0,00000 1,00000	p: 0,90931 0,09069	
	2	-1146,868	23,735	w: 0,00000 1,000006 3,80257	p: 0,32221 0,63890 0,03889	
	3	-1146,856	23,344	w: 0,00001 0,64483584,15862	p: 0,00000 0,96110 0,03890	
	7	-1424,902	0,100			p= 0,00270 q= 0,02564
8	-1146,901	19,811	w=496,80568	p: 0,96109 0,03891	p= 99,000 q=99,000	
eC17	0	-2057,564	0,675			
	1	-1896,033	0,183	w: 0,00000 1,00000	p: 0,81660 0,18340	
	2	-1714,049	0,806	w: 0,00000 1,00000 20,67738	p: 0,78203 0,18810 0,02988	
	3	-1711,076	0,677	w: 0,03424 0,78047 18,29384	p: 0,83820 0,13198 0,02982	
	7	-1883,600	0,117			p= 0,01258 q= 0,08106
8	-1711,212	0,676	w= 18,27604	p: 0,97010 0,02990	p= 0,12113 q= 0,77392	
eC18	0	-5903,244	0,410			
	1	-4940,179	0,148	w: 0,00000 1,00000	p: 0,85228 0,14772	
	2	-4558,215	0,477	w: 0,00000 1,00000 8,87681	p: 0,86691 0,08945 0,04364	
	3	-4548,525	0,478	w: 0,00147 1,25961 9,36202	p: 0,87739 0,08284 0,03977	
	7	-4925,988	0,101			p= 0,00618 q= 0,04333
8	-4558,477	0,441	w= 8,46425	p: 0,96610 0,03390	p= 0,02411 q= 0,12842	
eC20	0	-6040,630	0,540			
	1	-5102,998	0,130	w: 0,00000 1,00000	p: 0,86993 0,13007	
	2	-4534,609	0,573	w: 0,00000 1,00000 13,29601	p: 0,88200 0,08104 0,03696	
	3	-4498,898	0,600	w: 0,01062 2,16156 13,79748	p: 0,92445 0,03887 0,03668	
	7	-5035,259	0,106			p= 0,00834 q= 0,05541
8	-4513,900	0,492	w= 12,07456	p: 0,96785 0,03215	p= 0,00945 q= 0,06288	
eC21	0	-5389,860	1,325			
	1	-4658,498	0,200	w: 0,00000 1,00000	p: 0,80041 0,19959	
	2	-4144,971	0,930	w: 0,00000 1,00000 13,55813	p: 0,81890 0,12149 0,05961	
	3	-3972,513	1,707	w: 0,01123 5,98524 87,09611	p: 0,85656 0,13310 0,01035	
	7	-4632,454	0,146			p= 0,02825 q= 0,16243
8	-4139,604	0,873	w= 12,92178	p: 0,93996 0,06004	p= 0,00768 q= 0,05200	
eC22gp	0	-9525,933	0,387			
	1	-7321,015	0,176	w: 0,00000 1,00000	p: 0,82412 0,17588	
	2	-6942,961	0,578	w: 0,00000 1,00000 6,22222	p: 0,83444 0,08657 0,07899	
	3	-6907,606	0,487	w: 0,00017 0,72267 5,54303	p: 0,84771 0,07405 0,07824	
	7	-7228,001	0,103			p= 0,00835 q= 0,05854
8	-6897,578	0,464	w= 5,54905	p: 0,93365 0,06635	p= 0,00790 q= 0,05521	

Tabla 15. (Continuación).

individuo modelo	Ln L	dN/dS	$w_0/w_1/w_2$	$p_0/p_1/p_2$	p-q	
eC24	0	-1906,429	0,361			
	1	-1730,485	0,066	w: 0,00000 1,00000	p: 0,93439 0,06561	
	2	-1569,651	0,508	w: 0,00000 1,00000 25,45589	p: 0,92842 0,05371 0,01786	
	3	-1555,320	0,442	w: 0,02538 16,81916 50,85158	p: 0,98195 0,01473 0,00333	
	7	-1732,201	0,100			p= 0,00194 q= 0,02117
	8	-1723,095	0,038	w= 0,01270	p: 0,28918 0,71082	p= 0,00703 q= 0,04965
eC28	0	-8209,900	0,535			
	1	-6375,292	0,235	w: 0,00000 1,00000	p: 0,76536 0,23464	
	2	-6041,609	0,582	w: 0,00000 1,00000 7,26266	p: 0,78391 0,15768 0,05840	
	3	-5962,800	0,607	w: 0,00020 1,60311 7,65892	p: 0,83543 0,10797 0,05660	
	7	-6286,880	0,100			p= 0,00683 q= 0,05534
	8	-5981,496	0,463	w= 6,27961	p: 0,94633 0,05367	p= 0,01976 q= 0,12126
eC31gp	0	-2238,894	0,451			
	1	-2011,620	0,162	w: 0,00000 1,00000	p: 0,83792 0,16208	
	2	-1902,683	0,514	w: 0,00000 1,00000 8,51538	p: 0,83005 0,12420 0,04575	
	3	-1901,292	0,463	w: 0,00001 0,67167 7,91281	p: 0,81557 0,13759 0,04684	
	7	-2011,587	0,205			p= 0,01015 q= 0,03436
	8	-1901,533	0,475	w= 8,06223	p: 0,95578 0,04422	p= 0,00604 q= 0,03582
eC33gp	0	-2028,352	4,159			
	1	-1805,231	0,059	w: 0,00000 1,00000	p: 0,94116 0,05884	
	2	-1403,316	3,997	w: 0,00000 1,00000 128,29119	p: 0,92143 0,04779 0,03078	
	3	-1408,233	4,357	w: 0,08814 0,19007 139,09737	p: 0,96929 0,00000 0,03071	
	7	-1807,454	0,101			p= 0,00296 q= 0,01843
	8	-1403,892	4,357	w= 139,07192	p: 0,96932 0,03068	p= 0,00765 q= 0,10665
eC38gp	0	-1303,342	0,196			
	1	-1249,379	0,098	w: 0,00000 1,00000	p: 0,90234 0,09766	
	2	-1237,625	0,187	w: 0,00000 1,00000 5,66161	p: 0,90909 0,07039 0,02052	
	3	-1239,693	0,211	w: 0,00001 0,00001 2,79298	p: 0,28819 0,63642 0,07539	
	7	-1249,384	0,100			p= 0,00100 q= 0,01165
	8	-1239,693	0,211	w= 2,79267	p: 0,92460 0,07540	p= 0,00100 q= 2,29299
eG10	0	-633,813	0,249			
	1	-633,827	0,262	w: 0,00000 1,00000	p: 0,73805 0,26195	
	2	-633,813	0,249	w: 0,00000 1,00000 0,24891	p: 0,00000 0,00000 1,00000	
	3	-633,813	0,249	w: 0,24891 0,24891 0,42481	p: 0,88171 0,11829 0,00000	
	7	-633,813	0,253			p= 33,49339 q= 99,00000
	8	-633,813	0,249	w= 1,03859	p: 1,00000 0,00000)	p= 32,85126 q= 99,00000
eG14gp	0	-3874,776	0,252			
	1	-3451,194	0,159	w: 0,00000 1,00000	p: 0,84088 0,15912	
	2	-3325,095	0,285	w: 0,00000 1,00000 8,11832	p: 0,84947 0,13168 0,01885	
	3	-3308,842	0,274	w: 0,00623 1,25431 8,12748	p: 0,88606 0,09562 0,01832	
	7	-3429,013	0,108			p= 0,00777 q= 0,05009
	8	-3318,013	0,282	w= 8,03921	p: 0,99004 0,00996	p= 0,00705 q= 0,02318
eG19gp	0	-5015,478	0,433			
	1	-4160,128	0,193	w: 0,00000 1,00000	p: 0,80673 0,19327	
	2	-4016,966	0,477	w: 0,00000 1,00000 4,84955	p: 0,82511 0,09653 0,07835	
	3	-4023,534	0,431	w: 0,00002 0,00219 3,21276	p: 0,68763 0,17823 0,13414	
	7	-4130,505	0,100			p= 0,00840 q= 0,07036
	8	-3999,451	0,393	w= 4,49319	p: 0,93349 0,06651	p= 0,00897 q= 0,07583

Tabla 16. Parámetros correspondientes a los modelos 0, 1, 2, 3, 7 y 8 para la región NS5A de los individuos 1b.

individuo modelo	Ln L	dN/dS	w ₀ /w ₁ /w ₂	p ₀ /p ₁ /p ₂	p-q	
nA03	0	-2112,455	0,098			
	1	-2079,212	0,099	w: 0,00000 1,00000	p: 0,90090 0,09910	
	2	-2053,941	0,170	w: 0,00000 1,00000 15,42345	p: 0,90592 0,08879 0,00529	
	3	-2043,550	0,116	w: 0,00001 0,12625 13,85022	p: 0,61901 0,37601 0,00497	
	7	-2076,675	0,114			p= 0,01444 q= 0,09746
	8	-2043,613	0,116	w= 13,85077	p: 0,99502 0,00498	p= 0,40077 q= 7,54688
nA06	0	-3589,898	0,970			
	1	-3530,782	0,891	w: 0,00000 1,00000	p: 0,10866 0,89134	
	2	-3473,041	1,287	w: 0,00000 1,00000 56,02725	p: 0,10820 0,88462 0,00719	
	3	-3519,658	1,009	w: 0,00001 0,09688 1,29770	p: 0,10225 0,12968 0,76807	
	7	-3520,692	0,788			p= 0,14079 q= 0,03809
	8	-3466,517	1,015	w= 37,19691	p: 0,99433 0,00567	p= 0,02149 q= 0,00402
nA14	0	-2916,816	0,001			
	1	-2617,018	0,117	0,1173	p: 0,00000 1,00000	p: 0,88271 0,11729
	2	-2543,355	1,153	w: 0,00000 1,00000 15,71308	p: 0,90019 0,02821 0,07160	
	3	-2543,329	1,137	w: 0,00001 0,80917 15,18586	p: 0,90013 0,02645 0,07343	
	7	-2622,348	0,200			p= 0,00214 q= 0,00863
	8	-2556,742	1,093	w= 9,60385	p: 0,88617 0,11383	p= 0,00100 q= 1,94034
nA16	0	-6327,782	0,000			
	1	-6449,250	0,133	w: 0,00000 1,00000	p: 0,86719 0,13281	
	2	-6529,757	0,550	w: 0,00000 1,00000 1,99870	p: 0,61478 0,22047 0,16475	
	3	-6115,057	0,153	w: 0,00002 2,20946 4,86988	p: 0,96627 0,00442 0,02931	
	7	-6140,240	0,107			p= 0,01915 q= 0,15275
	8	-6115,772	0,066	w= 2,66221	p: 0,97629 0,02371	p= 0,02034 q= 2,11455
nA35	0	-1497,601	0,217			
	1	-1459,084	0,061	w: 0,00000 1,00000	p: 0,93899 0,06101	
	2	-1427,237	0,228	w: 0,00000 1,00000 21,66579	p: 0,94633 0,04525 0,00842	
	3	-1427,097	0,229	w: 0,03387 0,03389 20,01148	p: 0,00121 0,98904 0,00975	
	7	-1460,086	0,100			p= 0,00184 q= 0,01844
	8	-1427,099	0,229	w= 20,01256	p: 0,99025 0,00975	p= 3,51656 q= 99,00000
nC03	0	-1557,923	0,296			
	1	-1524,722	0,092	w: 0,00000 1,00000	p: 0,90772 0,09228	
	2	-1493,019	0,336	w: 0,00000 1,00000 19,02342	p: 0,92754 0,05784 0,01462	
	3	-1492,274	0,329	w: 0,04307 0,04309 17,97199	p: 0,98223 0,00184 0,01592	
	7	-1524,752	0,100			p= 0,00100 q= 0,01111
	8	-1502,924	0,303	w= 7,85019	p: 0,96142 0,03858	p= 0,00100 q= 2,02827
nC04	0	-2369,493	0,017			
	1	-2366,518	0,029	w: 0,00000 1,00000	p: 0,97098 0,02902	
	2	-2362,944	0,016	w: 0,00000 1,00000 0,31252	p: 0,94896 0,00000 0,05104	
	3	-2362,873	0,016	w: 0,00343 0,00343 0,42761	p: 0,43448 0,53601 0,02952	
	7	-2363,578	0,018			p= 0,01154 q= 0,30204
	8	-2363,578	0,018	w= 3,77603	p: 1,00000 0,00000	p= 0,01150 q= 0,30069
nC05	0	-2700,627	0,086			
	1	-2620,392	0,068	w: 0,00000 1,00000	p: 0,93155 0,06845	
	2	-2604,217	0,109	w: 0,00000 1,00000 8,00778	p: 0,93597 0,05767 0,00636	
	3	-2596,564	-2596,564	w: 0,01436 0,01436 3,44040	p: 0,29062 0,68521 0,02417	
	7	-2620,711	0,105			p= 0,01066 q= 0,07542
	8	-2596,650	0,097	w= 3,44536	p: 0,97587 0,02413	p= 1,49505 q= 99,00000

Ln L: Logaritmo de la verosimilitud

dN/dS: relación tasa de sustitución no sinónima/sinónima calculada como media en la longitud del fragmento.

w₀: dN/dS correspondiente a los sitios neutros; w₁: dN/dS correspondiente a los sitios bajo selección purificadora; w₂: dN/dS correspondiente a los sitios bajo selección positiva.

p₀: Proporción de sitios con w₀; p₁: Proporción de sitios con w₁; p₂: Proporción de sitios con w₂

p y q: Parámetros propios de la distribución beta.

Tabla 16. (Continuación)

individuo	modelo	Ln L	dN/dS	$w_0/w_1/w_2$		$p_0/p_1/p_2$		p-q
nC08	0	-5009,025	0,074					
	1	-4428,850	0,187	w:	0,00000 1,00000	p:	0,81320 0,18680	
	2	-4353,938	0,313	w:	0,00000 1,00000 6,29698	p:	0,81974 0,15516 0,02509	
	3	-4322,139	0,186	w:	0,00001 0,31137 4,52418	p:	0,81709 0,15230 0,03061	
	7	-4395,604	0,123					p= 0,01062 q= 0,06534
	8	-4325,701	0,182	w=	4,46913	p:	0,96613 0,03387	p= 0,01541 q= 0,34254
nC09	0	-1540,288	0,252					
	1	-1497,717	0,063	w:	0,00000 1,00000	p:	0,93698 0,06302	
	2	-1450,436	0,275	w:	0,00000 1,00000 34,88716	p:	0,93643 0,05734 0,00623	
	3	-1449,706	0,262	w:	0,04514 0,04518 34,33226	p:	0,99365 0,00002 0,00633	
	7	-1498,603	0,100					p= 0,00104 q= 0,01183
	8	-1449,711	0,262	w=	34,33587	p:	0,99367 0,00633	p= 4,72770 q= 99,00000
nC10	0	-2060,962	0,123					
	1	-2026,542	0,099	w:	0,00000 1,00000	p:	0,90086 0,09914	
	2	-2018,347	0,146	w:	0,00000 1,00000 8,85421	p:	0,90455 0,08901 0,00644	
	3	-2017,263	0,126	w:	0,00001 0,32428 4,71875	p:	0,82027 0,16442 0,01530	
	7	-2026,178	0,107					p= 0,00762 q= 0,04944
	8	-2017,296	0,126	w=	4,79921	p:	0,98519 0,01481	p= 0,07671 q= 1,17204
nC12	0	-4352,494	0,187					
	1	-4105,178	0,123	w:	0,00000 1,00000	p:	0,87682 0,12318	
	2	-4000,257	0,255	w:	0,00000 1,00000 7,52159	p:	0,87740 0,10229 0,02031	
	3	-3981,538	0,184	w:	0,00278 0,37397 6,56116	p:	0,86022 0,11890 0,02088	
	7	-4089,163	0,116					p= 0,00871 q= 0,05430
	8	-3981,586	0,184	w=	6,56348	p:	0,97903 0,02097	p= 0,05417 q= 0,95313
nC13	0	-3390,071	0,190					
	1	-3366,813	0,614	w:	0,00000 1,00000	p:	0,38599 0,61401	
	2	-3340,369	0,265	w:	0,00000 1,00000 0,39894	p:	0,35697 0,01460 0,62843	
	3	-3340,447	0,266	w:	0,00001 0,41521 0,41521	p:	0,35925 0,24805 0,39269	
	7	-3346,707	0,275					p= 0,42401 q= 1,11248
	8	-3346,669	0,275	w=	4,83327	p:	0,99934 0,00066	p= 0,42999 q= 1,14189
nC15	0	-3139,415	0,085					
	1	-2992,171	0,060	w:	0,00000 1,00000	p:	0,93961 0,06039	
	2	-2944,874	0,124	w:	0,00000 1,00000 8,13311	p:	0,94268 0,04792 0,00939	
	3	-2932,697	0,091	w:	0,01011 0,58366 7,56557	p:	0,97372 0,01690 0,00937	
	7	-2993,097	0,104					p= 0,00976 q= 0,06871
	8	-2933,828	0,091	w=	7,55788	p:	0,99069 0,00931	p= 0,03572 q= 1,14455
nC16	0	-2922,207	0,053					
	1	-2912,434	0,099	w:	0,00000 1,00000	p:	0,90100 0,09900	
	2	-2898,441	0,050	w:	0,00000 1,00000 0,09168	p:	0,66371 0,02092 0,31537	
	3	-2898,362	0,051	w:	0,00001 0,10459 1,17093	p:	0,68871 0,29400 0,01729	
	7	-2901,297	0,058					p= 0,06583 q= 0,96574
	8	-2898,371	0,051	w=	1,20549	p:	0,98424 0,01576	p= 0,22157 q= 5,97774
nC19	0	-5183,133	0,145					
	1	-4856,219	0,116	w:	0,00000 1,00000	p:	0,88444 0,11556	
	2	-4726,737	0,212	w:	0,00000 1,00000 8,11893	p:	0,88231 0,10441 0,01328	
	3	-4689,162	0,142	w:	0,01216 0,60532 6,65261	p:	0,93875 0,04587 0,01537	
	7	-4820,568	0,110					p= 0,01188 q= 0,08016
	8	-4691,819	0,143	w=	6,64947	p:	0,98540 0,01460	p= 0,06124 q= 1,08547
nC25	0	-1139,918	0,541					
	1	-1134,679	0,217	w:	0,00000 1,00000	p:	0,78255 0,21745	
	2	-1122,239	0,616	w:	0,00000 1,00000 60,63307	p:	0,74305 0,25092 0,00603	
	3	-1122,102	0,596	w:	0,00001 0,23633 59,55051	p:	0,00000 0,99393 0,00607	
	7	-1134,695	0,200					p= 0,00237 q= 0,00932
	8	-1122,103	0,596	w=	59,55843	p:	0,99393 0,00607	p= 30,68495 q= 99,000

Tabla 16. (Continuación)

individuo modelo	Ln L	dN/dS	$w_0/w_1/w_2$	$p_0/p_1/p_2$	p-q	
nG02	0	-4100,311	0,113			
	1	-3914,524	0,127	w: 0,00000 1,00000	p: 0,87301 0,12699	
	2	-3877,550	0,190	w: 0,00000 1,00000 4,49745	p: 0,87631 0,10463 0,01906	
	3	-3842,111	0,121	w: 0,02012 1,19935 3,91722	p: 0,95808 0,02311 0,01881	
	7	-3884,097	0,112			p= 0,01191 q= 0,07935
	8	-3843,652	0,120	w= 2,96944	p: 0,96889 0,03111	p= 0,11354 q= 3,34892
nG05	0	-4182,568	0,289			
	1	-3793,464	0,134	w: 0,00000 1,00000	p: 0,86615 0,13385	
	2	-3649,904	0,383	w: 0,00000 1,00000 6,02585	p: 0,87021 0,07941 0,05038	
	3	-3616,742	0,300	w: 0,02052 4,25775 10,6556	p: 0,94471 0,04831 0,00699	
	7	-3766,345	0,109			p= 0,00872 q= 0,05682
	8	-3622,035	0,299	w= 5,14618	p: 0,94595 0,05405	p= 0,42796 q= 17,69713
nG06	0	-4610,681	0,170			
	1	-4321,047	0,141	w: 0,00000 1,00000	p: 0,85952 0,14048	
	2	-4210,842	0,259	w: 0,00000 1,00000 7,84070	p: 0,86968 0,11146 0,01886	
	3	-4173,011	0,182	w: 0,01953 1,16717 6,93103	p: 0,94873 0,03329 0,01798	
	7	-4281,883	0,111			p= 0,01254 q= 0,08458
	8	-4178,206	0,184	w= 5,47425	p: 0,97370 0,02630	p= 0,08106 q= 1,63265
nG07	0	-4361,547	0,092			
	1	-4165,222	0,119	w: 0,00000 1,00000	p: 0,88061 0,11939	
	2	-4109,702	0,182	w: 0,00000 1,00000 6,09225	p: 0,88407 0,10305 0,01289	
	3	-4034,507	0,093	w: 0,00001 0,04949 3,96855	p: 0,60899 0,37219 0,01883	
	7	-4108,386	0,110			p= 0,01232 q= 0,08385
	8	-4034,637	0,093	w= 3,96856	p: 0,98118 0,01882	p= 0,40580 q= 19,78552
nG09	0	-3691,247	0,112			
	1	-3484,776	0,071	w: 0,00000 1,00000	p: 0,92868 0,07132	
	2	-3441,790	0,135	w: 0,00000 1,00000 6,03891	p: 0,93291 0,05361 0,01348	
	3	-3427,371	0,123	w: 0,00666 1,55234 5,93789	p: 0,96030 0,02724 0,01246	
	7	-3479,405	0,104			p= 0,00939 q= 0,06587
	8	-3436,995	0,120	w= 3,52573	p: 0,97003 0,02997	p= 0,01332 q= 0,39184
nG16	0	-1000,975	999,000			
	1	-1001,790	1,000	w: 0,00000 1,00000	p: 0,00000 1,00000	
	2	-1000,975	705,103	w: 0,00000 1,00 705,10319	p: 0,00000 0,00000 1,00000	
	3	-1000,975	999,000	w: 0,00001 0,00001, 999,000	p: 0,00000 0,00000 1,00000	
	7	-1001,790	1,000			p= 4,21342 q= 0,00100
	8	-1000,975	998,999	w=999,00000	p: 0,00000 1,00000	p= 0,00100 q= 2,26206
nG17	0	-3299,198	0,057			
	1	-3193,222	0,057	w: 0,00000 1,00000	p: 0,94290 0,05710	
	2	-3183,724	0,084	w: 0,00000 1,00000 4,07548	p: 0,94735 0,04248 0,01017	
	3	-3175,557	0,065	w: 0,00253 0,22559 2,59445	p: 0,93910 0,04016 0,02074	
	7	-3195,676	0,103			p= 0,00975 q= 0,07065
	8	-3175,810	0,066	w= 2,58099	p: 0,97923 0,02077	p= 0,01204 q= 0,36384
nG26	0	-1015,154	0,137			
	1	-1015,172	0,142	w: 0,00000 1,00000	p: 0,85833 0,14167	
	2	-1015,154	0,137	w: 0,00000 1,00000 0,13712	p: 0,00000 0,00000 1,00000	
	3	-1015,154	0,137	w: 0,00001 0,13712 0,29908	p: 0,00000 1,00000 0,00000	
	7	-1015,154	0,137			p= 15,76978 q= 99,0000
	8	-1015,154	0,137	w= 12,47109	p: 1,00000 0,00000	p= 15,76779 q= 99,0000

Tabla 17. Parámetros correspondientes a los modelos 0, 1, 2, 3, 7 y 8 para la región E1-E2 de los individuos 1a.

individuo modelo	Ln L	dN/dS	$w_0/w_1/w_2$	$p_0/p_1/p_2$	p-q	
nA05	0	-1173,805	0,051			
	1	-1174,319	0,059	w: 0,00000 1,00000	p: 0,94095 0,05905	
	2	-1173,805	0,051	w: 0,00000 1,00000 0,05060	p: 0,00000 0,00000 1,00000	
	3	-1173,805	0,051	w: 0,00393 0,05060 0,29573	p: 0,00000 1,00000 0,00000	
	7	-1173,810	0,051			p= 5,32064 q= 99,000
	8	-1173,810	0,051	w= 0,00001	p: 1,00000 0,00000	p= 5,32065 q= 99,000
nA09	0	-1831,504	0,064			
	1	-1807,495	0,058	w: 0,00000 1,00000	p: 0,94229 0,05771	
	2	-1806,192	0,075	w: 0,00000 1,00000 3,81606	p: 0,94492 0,04800 0,00708	
	3	-1805,246	0,066	w: 0,00001 0,35389 3,23927	p: 0,90673 0,08185 0,01143	
	7	-1809,588	0,100			p= 0,00260 q= 0,02751
	8	-1805,275	0,066	w= 3,18883	p: 0,98824 0,01176	p= 0,01119 q= 0,24849
nA20	0	-3053,572	0,001			
	1	-2588,738	0,048	w: 0,00000 1,00000	p: 0,95163 0,04837	
	2	-2544,541	0,168	w: 0,00000 1,00000 7,31009	p: 0,95255 0,02837 0,01908	
	3	-2558,404	0,155	w: 0,00001 0,00007 5,25835	p: 0,40178 0,56873 0,02949	
	7	-2594,232	0,100			p= 0,00220 q= 0,02191
	8	-2561,440	0,162	w= 3,45100	p: 0,95296 0,04704	p= 0,00100 q= 1,54703
nA26	0	-3828,970	0,451			
	1	-3491,725	0,121	w: 0,00000 1,00000	p: 0,87910 0,12090	
	2	-3320,853	0,485	w: 0,00000 1,00000 7,41714	p: 0,88274 0,06002 0,05723	
	3	-3315,227	0,454	w: 0,01712 2,16379 7,76541	p: 0,92481 0,02597 0,04922	
	7	-3487,164	0,106			p= 0,00724 q= 0,04715
	8	-3317,281	0,454	w= 6,91392	p: 0,93924 0,06076	p= 0,03191 q= 0,67538
nA27	0	-3149,948	0,081			
	1	-3074,529	0,092	w: 0,00000 1,00000 11,63417	p: 0,91315 0,08193 0,00492	
	2	-3041,589	0,139	w: 0,00000 1,00000 11,63417	p: 0,91315 0,08193 0,00492	
	3	-3030,097	0,092	w: 0,00001 0,34643 10,67157	p: 0,87249 0,12286 0,00465	
	7	-3070,317	0,109			p= 0,00873 q= 0,05662
	8	-3030,186	0,092	w= 10,67105	p: 0,99531 0,00469	p= 0,04026 q= 0,77348
nA34	0	-2107,501	0,181			
	1	-2023,610	0,071	w: 0,00000 1,00000	p: 0,92900 0,07100	
	2	-1981,994	0,193	w: 0,00000 1,00000 12,56989	p: 0,92909 0,06034 0,01057	
	3	-1981,764	0,185	w: 0,00489 0,92259 12,36345	p: 0,93531 0,05417 0,01052	
	7	-2024,644	0,100			p= 0,00130 q= 0,01624
	8	-1982,220	0,186	w= 12,35032	p: 0,98975 0,01025	p= 0,01045 q= 0,18110
nC01	0	-2348,898	0,224			
	1	-2315,179	0,230	w: 0,00000 1,00000	p: 0,77005 0,22995	
	2	-2309,287	0,292	w: 0,00000 1,00000 5,36907	p: 0,76966 0,21617 0,01417	
	3	-2305,323	0,235	w: 0,03710 0,66164 3,98030	p: 0,80565 0,17130 0,02304	
	7	-2313,628	0,190			p= 0,07010 q= 0,29919
	8	-2305,357	0,235	w= 3,89790	p: 0,97583 0,02417	p= 0,17327 q= 1,00661
nC02	0	-1522,567	0,590			
	1	-1488,963	0,131	w: 0,00000 1,00000	p: 0,86933 0,13067	
	2	-1408,045	0,492	w: 0,00000 1,00000 198,32067	p: 0,82424 0,17415 0,00160	
	3	-1403,739	0,612	w: 0,11270 8,17043 223,85517	p: 0,98124 0,01715 0,00161	
	7	-1489,349	0,100			p= 0,00100 q= 0,00958
	8	-1408,162	0,513	w= 197,10081	p: 0,99841 0,00159	p= 0,00100 q= 0,00378

Ln L: Logaritmo de la verosimilitud

dN/dS: relación tasa de sustitución no sinónima/sinónima calculada como media en la longitud del fragmento.

w0: dN/dS correspondiente a los sitios neutros; w1: dN/dS correspondiente a los sitios bajo selección purificadora; w2: dN/dS correspondiente a los sitios bajo selección positiva.

p0: Proporción de sitios con w0; p1: Proporción de sitios con w1; p2: Proporción de sitios con w2

p y q: Parámetros propios de la distribución beta.

Tabla 17. (Continuación)

individuo modelo	Ln L	dN/dS	w ₀ /w ₁ /w ₂	p ₀ /p ₁ /p ₂	p-q	
nC06	0	-1135,975	0,022			
	1	-1136,144	0,025	w: 0,00000 1,00000	p: 0,97528 0,02472	
	2	-1135,975	0,022	w: 0,00000 1,00000 0,02164	p: 0,00000 0,00000 1,00000	
	3	-1135,975	0,022	w: 0,00001 0,01660 0,02164	p: 0,00000 0,00000 1,00000	
	7	-1135,976	0,022			p= 2,22839 q= 99,000
	8	-1135,976	0,022	w= 0,00001	p: 1,00000 0,00000	p= 2,22839 q= 99,000
nC07	0	-3424,476	0,311			
	1	-3184,591	0,118	w: 0,00000 1,00000	p: 0,88240 0,11760	
	2	-3081,805	0,343	w: 0,00000 1,00000 6,84848	p: 0,87815 0,08406 0,03778	
	3	-3077,492	0,309	w: 0,01350 0,60808 6,39259	p: 0,89921 0,06004 0,04074	
	7	-3181,862	0,109			p= 0,00724 q= 0,04605
	8	-3080,987	0,339	w= 6,80939	p: 0,96557 0,03443	p= 0,00507 q= 0,03123
nC11	0	-1685,689	0,237			
	1	-1608,821	0,045	w: 0,00000 1,00000	p: 0,95537 0,04463	
	2	-1550,958	0,269	w: 0,00000 1,00000 20,00147	p: 0,97133 0,01603 0,01264	
	3	-1550,369	0,263	w: 0,01150 0,01151 19,82156	p: 0,02457 0,96275 0,01268	
	7	-1612,037	0,100			p= 0,00177 q= 0,01880
	8	-1557,914	0,264	w= 13,11444	p: 0,97990 0,02010	p= 0,00100 q= 2,21031
nC14	0	-1256,142	0,000			
	1	-1256,132	0,000	w: 0,00000 1,00000	p: 1,00000 0,00000	
	2	-1256,132	0,000	w: 0,00000 1,00000 0,00001	p: 0,99992 0,00000 0,00008	
	3	-1256,133	0,000	w: 0,00001 0,00001 0,00001	p: 0,26703 0,69393 0,03905	
	7	-1256,132	0,000			p= 0,00100 q= 1,38619
	8	-1256,132	0,000	w= 0,22151	p: 1,00000 0,00000	p= 0,00100 q= 2,24449
nC17	0	-1037,600	0,955			
	1	-1037,601	0,980	w: 0,00000 1,00000	p: 0,02047 0,97953	
	2	-1037,600	0,955	w: 0,00000 1,00000 0,95538	p: 0,00000 0,00018 0,99982	
	3	-1037,600	0,955	w: 0,00001 0,95539 1,58398	p: 0,00000 1,00000 0,00000	
	7	-1037,601	1,000			p= 4,11686 q= 0,00100
	8	-1037,600	0,955	w= 0,95529	p: 0,00000 1,00000	p= 1,90808 q= 0,00100
nC18	0	-1028,865	0,187			
	1	-1028,881	0,193	w: 0,00000 1,00000	p: 0,80742 0,19258	
	2	-1028,865	0,187	w: 0,00000 1,00000 0,18661	p: 0,00000 0,00000 1,00000	
	3	-1028,865	0,187	w: 0,00001 0,00001 0,18661	p: 0,00000 0,00000 1,00000	
	7	-1028,865	0,189			p= 23,06144 q= 99,0000
	8	-1028,865	0,187	w=611,47210	p: 1,00000 0,00000	p= 22,77358 q= 99,0000
nC20	0	-2266,013	0,262			
	1	-2142,000	0,060	w: 0,00000 1,00000	p: 0,93959 0,06041	
	2	-2014,204	0,295	w: 0,00000 1,00000 34,92490	p: 0,93552 0,05769 0,00679	
	3	-2013,882	0,263	w: 0,02110 0,03977 33,45812	p: 0,00000 0,99332 0,00668	
	7	-2143,902	0,103			p= 0,00191 q= 0,01131
	8	-2011,346	0,263	w= 33,45779	p: 0,99330 0,00670	p= 0,07311 q= 1,51278
nC21	0	-2902,194	0,001			
	1	-2613,201	0,097	w: 0,00000 1,00000	p: 0,90331 0,09669	
	2	-2571,286	0,188	w: 0,00000 1,00000 15,03902	p: 0,90367 0,08981 0,00653	
	3	-2606,270	0,183	w: 0,00001 0,00001 1,96457	p: 0,37158 0,53525 0,09317	
	7	-2613,220	0,100			p= 0,00203 q= 0,02132
	8	-2605,762	0,182	w= 1,94094	p: 0,90609 0,09391	p= 0,00100 q= 2,85909
nC22	0	-3572,023	0,001			
	1	-3185,027	0,100	w: 0,00000 1,00000	p: 0,89981 0,10019	
	2	-3176,289	0,147	w: 0,00000 1,00000 3,67060	p: 0,90159 0,08032 0,01809	
	3	-3184,082	0,117	w: 0,00001 0,00001 1,22280	p: 0,56177 0,34252 0,09571	
	7	-3196,444	0,142			p= 0,00861 q= 0,04966
	8	-3189,109	0,180	w= 3,87043	p: 0,99063 0,00937	p= 0,00879 q= 0,05023

Tabla 17. (Continuación)

individuo modelo	Ln L	dN/dS	$w_0/w_1/w_2$	$p_0/p_1/p_2$	p-q	
nC24	0	-1454,991	0,365			
	1	-1431,506	0,145	w: 0,00000 1,00000	p: 0,85464 0,14536	
	2	-1385,224	0,369	w: 0,00 1,0000011 2,14208	p: 0,82468 0,17358 0,00174	
	3	-1384,934	0,391	w: 0,08119 3,6884511 4,99045	p: 0,96454 0,03383 0,00163	
	7	-1432,175	0,200			p= 0,00272 q= 0,01143
	8	-1385,482	0,369	w=106,19229	p: 0,99825 0,00175	p= 0,00250 q= 0,01379
nC28	0	-2585,798	0,204			
	1	-2462,364	0,107	w: 0,00000 1,00000	p: 0,89341 0,10659	
	2	-2432,529	0,215	w: 0,00000 1,00000 5,24557	p: 0,89403 0,08033 0,02564	
	3	-2432,001	0,209	w: 0,01020 1,14851 5,18685	p: 0,91738 0,05662 0,02601	
	7	-2462,267	0,103			p= 0,00160 q= 0,00943
	8	-2432,660	0,212	w= 5,17181	p: 0,97494 0,02506	p= 0,01079 q= 0,15444
nC31	0	-1551,662	0,210			
	1	-1514,712	0,092	w: 0,00000 1,00000	p: 0,90843 0,09157	
	2	-1493,682	0,252	w: 0,00000 1,00000 11,25482	p: 0,92507 0,05767 0,01726	
	3	-1492,939	0,241	w: 0,03939 0,03939 10,22648	p: 0,21834 0,76189 0,01977	
	7	-1514,761	0,100			p= 0,00100 q= 0,01060
	8	-1499,348	0,235	w= 5,09224	p: 0,95381 0,04619	p= 0,00100 q= 2,13719
nC33	0	-1009,742	999,000			
	1	-1010,830	1,000	w: 0,00000 1,00000	p: 0,00000 1,00000	
	2	-1009,742	999,000	w: 0,00000 1,00000 999,000	p: 0,00000 0,00000 1,00000	
	3	-1009,742	999,000	W: 0,00001 0,00001999,0000	p: 0,00000 0,00000 1,00000	
	7	-1010,830	1,000			p= 4,41207 q= 0,00100
	8	-1009,742	998,999	w=999,00000	p: 0,00000 1,00000	p= 0,00100 q= 1,83596
nC38	0	-1013,176	0,417			
	1	-1013,188	0,426	w: 0,00000 1,00000	p: 0,57368 0,42632	
	2	-1013,176	0,417	w: 0,00000 1,00000 0,41714	p: 0,00000 0,00000 1,00000	
	3	-1013,176	0,417	w: 0,00001 0,00001 0,41714	p: 0,00000 0,00000 1,00000	
	7	-1013,176	0,423			p= 72,52834 q= 99,0000
	8	-1013,176	0,424	w= 3,76503	p: 1,00000 0,00000	p= 72,79552 q= 99,0000
nG10	0	-1103,214	0,525			
	1	-1098,934	0,210	w: 0,00000 1,00000	p: 0,79027 0,20973	
	2	-1086,375	0,596	w: 0,00000 1,00000126,49128	p: 0,74971 0,24753 0,00275	
	3	-1086,234	0,572	w: 0,23342 0,23351123,54769	p: 0,00042 0,99684 0,00274	
	7	-1098,938	0,200			p= 0,00142 q= 0,00644
	8	-1091,731	0,533	w= 18,40813	p: 0,97102 0,02898	p= 0,00100 q= 2,09010
nG14	0	-3050,273	0,128			
	1	-2968,560	0,114	w: 0,00000 1,00000	p: 0,88621 0,11379	
	2	-2964,025	0,148	w: 0,00000 1,00000 2,84544	p: 0,88774 0,09299 0,01926	
	3	-2956,498	0,128	w: 0,01548 0,52957 2,15901	p: 0,92450 0,03029 0,04521	
	7	-2964,718	0,110			p= 0,00897 q= 0,05827
	8	-2956,735	0,128	w= 2,13003	p: 0,95285 0,04715	p= 0,09332 q= 2,60107
nG19	0	-2882,028	0,000			
	1	-2514,451	0,053	w: 0,00000 1,00000	p: 0,94728 0,05272	
	2	-2490,334	0,144	w: 0,00000 1,00000 5,60402	p: 0,95081 0,02854 0,02065	
	3	-2498,052	0,144	w: 0,00001 0,00002 3,27223	p: 0,28927 0,66664 0,04409	
	7	-2518,959	0,100			p= 0,00109 q= 0,01270
	8	-2498,084	0,142	w= 3,39466	p: 0,95827 0,04173	p= 0,00494 q= 0,41067

Comparación de modelos de evolución de codones**Tabla 18.** Comparación de modelos según el test de razón de verosimilitudes y AIC (Criterio de información de Akaike) para la región E1-E2 en los individuos 1b. En rojo se señalan los valores de AIC más bajos.

Individuo	Modelo	2ΔL	df	P valor	Modelos	(-2)Ln L	2K	AIC
eA03	M0 vs M2	1121,122	2	3,5594E-244	M0	9430,306	2	9432,306
	M1 vs M2	172,927	2	2,81497E-38	M1	8482,111	2	8484,111
	M0 vs M3	1187,954	4	6,5085E-256	M2	8309,184	6	8315,184
	M1 vs M3	239,759	4	1,04576E-50	M3	8242,352	10	8252,352
	M2 vs M3	66,832	2	3,0733E-15	M7	8393,730	4	8397,730
	M7 vs M8	148,081	2	6,99086E-33	M8	8245,649	8	8253,649
eA06	M0 vs M2	24,872	2	3,97372E-06	M0	2205,233	2	2207,233
	M1 vs M2	0,862	2	0,649740648	M1	2181,224	2	2183,224
	M0 vs M3	24,877	4	5,32585E-05	M2	2180,362	6	2186,362
	M1 vs M3	0,868	4	0,929148631	M3	2180,357	10	2190,357
	M2 vs M3	0,005	2	0,997336553	M7	2181,813	4	2185,813
	M7 vs M8	1,451	2	0,484128947	M8	2180,362	8	2188,362
eA14	M0 vs M2	2119,569	2	0	M0	11284,356	2	11286,356
	M1 vs M2	664,657	2	4,6946E-145	M1	9829,444	2	9831,444
	M0 vs M3	2280,240	4	0	M2	9164,787	6	9170,787
	M1 vs M3	825,328	4	2,5053E-177	M3	9004,116	10	9014,116
	M2 vs M3	160,672	2	1,29007E-35	M7	9708,790	4	9712,790
	M7 vs M8	607,312	2	1,3302E-132	M8	9101,478	8	9109,478
eA16	M0 vs M2	2216,238	2	0	M0	15882,258	2	15884,258
	M1 vs M2	516,222	2	8,0135E-113	M1	14182,242	2	14184,242
	M0 vs M3	2538,193	4	0	M2	13666,020	6	13672,020
	M1 vs M3	838,177	4	4,125E-180	M3	13344,064	10	13354,064
	M2 vs M3	321,955	2	1,22534E-70	M7	13768,133	4	13772,133
	M7 vs M8	419,750	2	7,11835E-92	M8	13348,383	8	13356,383
eA35	M0 vs M2	1048,721	2	1,8757E-228	M0	6097,309	2	6099,309
	M1 vs M2	360,161	2	6,19642E-79	M1	5408,748	2	5410,748
	M0 vs M3	1076,880	4	7,7709E-232	M2	5048,588	6	5054,588
	M1 vs M3	388,319	4	9,28741E-83	M3	5020,429	10	5030,429
	M2 vs M3	28,159	2	7,68004E-07	M7	5394,322	4	5398,322
	M7 vs M8	373,131	2	9,45495E-82	M8	5021,191	8	5029,191
eC03	M0 vs M2	722,040	2	1,6255E-157	M0	2998,476	2	3000,476
	M1 vs M2	425,980	2	3,15907E-93	M1	2702,416	2	2704,416
	M0 vs M3	724,936	4	1,3888E-155	M2	2276,436	6	2282,436
	M1 vs M3	428,876	4	1,59984E-91	M3	2273,540	10	2283,540
	M2 vs M3	2,896	2	0,23506881	M7	2710,267	4	2714,267
	M7 vs M8	436,704	2	1,48252E-95	M8	2273,563	8	2281,563
eC04	M0 vs M2	0,000	2	1	M0	1329,882	2	1331,882
	M1 vs M2	2,663	2	0,264028824	M1	1332,546	2	1334,546
	M0 vs M3	0,000	4	1	M2	1329,882	6	1335,882
	M1 vs M3	2,663	4	0,615635219	M3	1329,882	10	1339,882
	M2 vs M3	0,000	2	1	M7	1332,546	4	1336,546
	M7 vs M8	2,663	2	0,264029088	M8	1329,882	8	1337,882
eC05	M0 vs M2	2990,910	2	0	M0	16516,069	2	16518,069
	M1 vs M2	1040,657	2	1,0576E-226	M1	14565,816	2	14567,816
	M0 vs M3	3320,551	4	0	M2	13525,159	6	13531,159
	M1 vs M3	1370,298	4	1,9051E-295	M3	13195,517	10	13205,517
	M2 vs M3	329,642	2	2,62543E-72	M7	14060,944	4	14064,944
	M7 vs M8	11,780	2	0,002766968	M8	14049,164	8	14057,164

2ΔL: 2 veces la diferencia del logaritmo de las verosimilitudes.

df: grados de libertad.

L: valor de verosimilitud.

K: n° de parámetros del modelo.

Tabla 18. (Continuación)

Individuo	Modelos	2ΔL	df	P valor	Modelos	(-) $2\ln L$	2K	AIC
eC08	M0 vs M2	3388,893	2	0	M0	15768,664	2	15770,664
	M1 vs M2	1168,492	2	1,8417E-254	M1	13548,263	2	13550,263
	M0 vs M3	3586,235	4	0	M2	12379,771	6	12385,771
	M1 vs M3	1365,835	4	1,7693E-294	M3	12182,429	10	12192,429
	M2 vs M3	197,343	2	1,40467E-43	M7	13425,488	4	13429,488
	M7 vs M8	1085,455	2	1,9793E-236	M8	12340,033	8	12348,033
eC09	M0 vs M2	617,162	2	9,6608E-135	M0	4149,894	2	4151,894
	M1 vs M2	324,552	2	3,34516E-71	M1	3857,284	2	3859,284
	M0 vs M3	629,965	4	5,0635E-135	M2	3532,732	6	3538,732
	M1 vs M3	337,355	4	9,41496E-72	M3	3519,929	10	3529,929
	M2 vs M3	12,803	2	0,001658733	M7	3847,287	4	3851,287
	M7 vs M8	1,168	2	0,55773519	M8	3846,119	8	3854,119
eC10	M0 vs M2	2784,307	2	0	M0	10451,656	2	10453,656
	M1 vs M2	506,388	2	1,0944E-110	M1	8173,737	2	8175,737
	M0 vs M3	1983,213	4	0	M2	7667,349	6	7673,349
	M1 vs M3	294,706	4	1,50247E-62	M3	8468,442	10	8478,442
	M2 vs M3	801,094	2	1,1084E-174	M7	8142,265	4	8146,265
	M7 vs M8	493,501	2	6,8806E-108	M8	7648,764	8	7656,764
eC12	M0 vs M2	2183,276	2	0	M0	9892,409	2	9894,409
	M1 vs M2	597,954	2	1,4322E-130	M1	8307,086	2	8309,086
	M0 vs M3	2311,223	4	0	M2	7709,133	6	7715,133
	M1 vs M3	725,900	4	8,5869E-156	M3	7581,186	10	7591,186
	M2 vs M3	127,946	2	1,64734E-28	M7	8146,198	4	8150,198
	M7 vs M8	555,176	2	2,7868E-121	M8	7591,022	8	7599,022
eC13	M0 vs M2	427,226	2	1,69474E-93	M0	5140,076	2	5142,076
	M1 vs M2	141,570	2	1,81314E-31	M1	4854,421	2	4856,421
	M0 vs M3	457,617	4	9,79615E-98	M2	4712,850	6	4718,850
	M1 vs M3	171,962	4	3,96682E-36	M3	4682,459	10	4692,459
	M2 vs M3	30,391	2	2,51529E-07	M7	4812,455	4	4816,455
	M7 vs M8	126,604	2	3,22328E-28	M8	4685,851	8	4693,851
eC15	M0 vs M2	4016,988	2	0	M0	19388,186	2	19390,186
	M1 vs M2	1154,819	2	1,7149E-251	M1	16526,017	2	16528,017
	M0 vs M3	4354,734	4	0	M2	15371,198	6	15377,198
	M1 vs M3	1492,564	4	0	M3	15033,453	10	15043,453
	M2 vs M3	337,745	2	4,56574E-74	M7	15767,601	4	15771,601
	M7 vs M8	733,897	2	4,328E-160	M8	15033,704	8	15041,704
eC16	M0 vs M2	1306,651	2	1,8376E-284	M0	8089,694	2	8091,694
	M1 vs M2	388,323	2	4,75004E-85	M1	7171,366	2	7173,366
	M0 vs M3	1339,320	4	9,9297E-289	M2	6783,043	6	6789,043
	M1 vs M3	420,991	4	8,09461E-90	M3	6750,375	10	6760,375
	M2 vs M3	32,668	2	8,05743E-08	M7	7102,662	4	7106,662
	M7 vs M8	351,217	2	5,42291E-77	M8	6751,445	8	6759,445
eC19	M0 vs M2	2537,838	2	0	M0	14309,035	2	14311,035
	M1 vs M2	922,954	2	3,8295E-201	M1	12694,151	2	12696,151
	M0 vs M3	2724,832	4	0	M2	11771,197	6	11777,197
	M1 vs M3	1109,947	4	5,285E-239	M3	11584,204	10	11594,204
	M2 vs M3	186,994	2	2,48222E-41	M7	12212,502	4	12216,502
	M7 vs M8	624,962	2	1,9555E-136	M8	11587,540	8	11595,540
eC25	M0 vs M2	34,003	2	4,13413E-08	M0	1534,662	2	1536,662
	M1 vs M2	16,769	2	0,000228378	M1	1517,428	2	1519,428
	M0 vs M3	34,521	4	5,8254E-07	M2	1500,659	6	1506,659
	M1 vs M3	17,287	4	0,001699517	M3	1500,140	10	1510,140
	M2 vs M3	0,518	2	0,771660958	M7	1517,432	4	1521,432
	M7 vs M8	12,124	2	0,002329718	M8	1505,308	8	1513,308

Tabla 18. (Continuación)

Individuo	Modelos	2ΔL	df	P valor	Modelos	(-2)Ln L	2K	AIC
eG02	M0 vs M2	2756,069	2	0	M0	15239,633	2	15241,633
	M1 vs M2	593,039	2	1,6721E-129	M1	13076,602	2	13078,602
	M0 vs M3	2922,298	4	0	M2	12483,564	6	12489,564
	M1 vs M3	759,267	4	5,1017E-163	M3	12317,335	10	12327,335
	M2 vs M3	166,229	2	8,01555E-37	M7	12831,855	4	12835,855
	M7 vs M8	493,218	2	7,9283E-108	M8	12338,637	8	12346,637
eG05	M0 vs M2	2047,784	2	0	M0	7529,418	2	7531,418
	M1 vs M2	435,951	2	2,16033E-95	M1	5917,584	2	5919,584
	M0 vs M3	2053,720	4	0	M2	5481,634	6	5487,634
	M1 vs M3	441,886	4	2,46562E-94	M3	5475,698	10	5485,698
	M2 vs M3	5,935	2	0,051423809	M7	5865,588	4	5869,588
	M7 vs M8	389,659	2	2,43548E-85	M8	5475,928	8	5483,928
eG06	M0 vs M2	1587,861	2	0	M0	10061,396	2	10063,396
	M1 vs M2	507,675	2	5,7523E-111	M1	8981,210	2	8983,210
	M0 vs M3	1698,064	4	0	M2	8473,535	6	8479,535
	M1 vs M3	617,878	4	2,0934E-132	M3	8363,332	10	8373,332
	M2 vs M3	110,203	2	1,17418E-24	M7	8790,318	4	8794,318
	M7 vs M8	420,086	2	6,01818E-92	M8	8370,232	8	8378,232
eG07	M0 vs M2	3284,379	2	0	M0	14608,409	2	14610,409
	M1 vs M2	245,547	2	4,78672E-54	M1	11569,577	2	11571,577
	M0 vs M3	3720,860	4	0	M2	11324,030	6	11330,030
	M1 vs M3	682,028	4	2,7129E-146	M3	10887,549	10	10897,549
	M2 vs M3	436,481	2	1,65713E-95	M7	11178,281	4	11182,281
	M7 vs M8	281,481	2	7,53652E-62	M8	10896,800	8	10904,800
eG09gp	M0 vs M2	2258,786	2	0	M0	10215,650	2	10217,650
	M1 vs M2	242,454	2	2,24804E-53	M1	8199,318	2	8201,318
	M0 vs M3	2324,089	4	0	M2	7956,864	6	7962,864
	M1 vs M3	307,757	4	2,29859E-65	M3	7891,561	10	7901,561
	M2 vs M3	65,303	2	6,60185E-15	M7	8012,567	4	8016,567
	M7 vs M8	192,311	2	1,73829E-42	M8	7820,255	8	7828,255
eG16gp	M0 vs M2	0,000	2	1	M0	1297,932	2	1299,932
	M1 vs M2	0,977	2	0,613651565	M1	1298,909	2	1300,909
	M0 vs M3	0,000	4	1	M2	1297,932	6	1303,932
	M1 vs M3	0,977	4	0,913314803	M3	1297,932	10	1307,932
	M2 vs M3	0,000	2	1	M7	1298,909	4	1302,909
	M7 vs M8	0,977	2	0,613652179	M8	1297,932	8	1305,932
eG17gp	M0 vs M2	3197,706	2	0	M0	15766,101	2	15768,101
	M1 vs M2	817,792	2	2,6221E-178	M1	13386,187	2	13388,187
	M0 vs M3	3383,881	4	0	M2	12568,394	6	12574,394
	M1 vs M3	1003,967	4	4,9309E-216	M3	12382,220	10	12392,220
	M2 vs M3	186,174	2	3,73874E-41	M7	13021,041	4	13025,041
	M7 vs M8	632,189	2	5,272E-138	M8	12388,853	8	12396,853
eG26gp	M0 vs M2	2,286	2	0,318853669	M0	1377,300	2	1379,300
	M1 vs M2	2,078	2	0,353893238	M1	1377,091	2	1379,091
	M0 vs M3	2,296	4	0,681544403	M2	1375,014	6	1381,014
	M1 vs M3	2,087	4	0,719721372	M3	1375,004	10	1385,004
	M2 vs M3	0,010	2	0,995163733	M7	1377,323	4	1381,323
	M7 vs M8	2,085	2	0,352545002	M8	1375,238	8	1383,238

Tabla 19. Comparación de modelos según el test de razón de verosimilitudes y AIC para la región E1-E2 en los individuos 1a. En rojo se señalan los valores de AIC más bajos.

Individuo	Modelos	2ΔL	df	P valor	Modelos	(-2)Ln L	2K	AIC
eA05	M0 vs M2	12,390	2	0,00203915	M0	1702,479944	2	1704,480
	M1 vs M2	0,268	2	0,874425657	M1	1690,358	2	1692,358
	M0 vs M3	12,388	4	0,014685904	M2	1690,090	6	1696,090
	M1 vs M3	0,266	4	0,991890418	M3	1690,092	10	1700,092
	M2 vs M3	0,002	2	0,998911593	M7	1693,461	4	1697,461
	M7 vs M8	3,371	2	0,185307254	M8	1690,090	8	1698,090
eA09	M0 vs M2	357,028	2	2,96787E-78	M0	4718,044	2	4720,044
	M1 vs M2	68,900	2	1,09271E-15	M1	4429,917	2	4431,917
	M0 vs M3	368,839	4	1,49906E-78	M2	4361,017	6	4367,017
	M1 vs M3	80,711	4	1,23101E-16	M3	4349,206	10	4359,206
	M2 vs M3	11,811	2	0,002724081	M7	4416,917	4	4420,917
	M7 vs M8	65,412	2	6,25207E-15	M8	4351,505	8	4359,505
eA20	M0 vs M2	1936,070	2	0	M0	9733,823	2	9735,823
	M1 vs M2	870,996	2	7,3387E-190	M1	8668,748	2	8670,748
	M0 vs M3	2023,384	4	0	M2	7797,752	6	7803,752
	M1 vs M3	958,310	4	3,8643E-206	M3	7710,439	10	7720,439
	M2 vs M3	87,314	2	1,09666E-19	M7	8510,895	4	8514,895
	M7 vs M8	744,734	2	1,9187E-162	M8	7766,161	8	7774,161
	eA26	M0 vs M2	2308,916	2	0	M0	12393,568	2
M1 vs M2		918,517	2	3,5198E-200	M1	11003,169	2	11005,169
M0 vs M3		2462,113	4	0	M2	10084,651	6	10090,651
M1 vs M3		1071,714	4	1,0234E-230	M3	9931,454	10	9941,454
M2 vs M3		153,197	2	5,41574E-34	M7	10748,425	4	10752,425
M7 vs M8		752,652	2	3,6622E-164	M8	9995,773	8	10003,773
eA27	M0 vs M2	1744,989	2	0	M0	7852,737	2	7854,737
	M1 vs M2	658,226	2	1,1696E-143	M1	6765,974	2	6767,974
	M0 vs M3	1849,781	4	0	M2	6107,748	6	6113,748
	M1 vs M3	763,018	4	7,8594E-164	M3	6002,956	10	6012,956
	M2 vs M3	104,792	2	1,75672E-23	M7	6742,397	4	6746,397
	M7 vs M8	635,485	2	1,0146E-138	M8	6106,913	8	6114,913
eA34	M0 vs M2	1431,838	2	0	M0	10293,811	2	10295,811
	M1 vs M2	497,250	2	1,0556E-108	M1	9359,223	2	9361,223
	M0 vs M3	1682,658	4	0	M2	8861,973	6	8867,973
	M1 vs M3	748,071	4	1,3571E-160	M3	8611,152	10	8621,152
	M2 vs M3	250,821	2	3,42784E-55	M7	9077,329	4	9081,329
	M7 vs M8	351,436	2	4,85893E-77	M8	8725,893	8	8733,893
eC01	M0 vs M2	1619,938	2	0	M0	8723,838	2	8725,838
	M1 vs M2	720,835	2	2,9694E-157	M1	7824,735	2	7826,735
	M0 vs M3	1772,621	4	0	M2	7103,900	6	7109,900
	M1 vs M3	873,519	4	9,1E-188	M3	6951,217	10	6961,217
	M2 vs M3	152,684	2	7,00051E-34	M7	7616,882	4	7620,882
	M7 vs M8	570,658	2	1,2112E-124	M8	7046,224	8	7054,224
eC02	M0 vs M2	129,712	2	6,81524E-29	M0	1975,588	2	1977,588
	M1 vs M2	42,784	2	5,12264E-10	M1	1888,661	2	1890,661
	M0 vs M3	130,484	4	3,06804E-27	M2	1845,877	6	1851,877
	M1 vs M3	43,557	4	7,92983E-09	M3	1845,104	10	1855,104
	M2 vs M3	0,773	2	0,679588372	M7	1896,723	4	1900,723
	M7 vs M8	51,614	2	6,19701E-12	M8	1845,109	8	1853,109

2ΔL: 2 veces la diferencia del logaritmo de las verosimilitudes.

df: grados de libertad.

L: valor de verosimilitud.

K: nº de parámetros del modelo.

Tabla 19. (Continuación)

Individuo	Modelos	2ΔL	df	P valor	Modelos	(-) $2\ln L$	2K	AIC
eC06	M0 vs M2	2,568	2	0,276873654	M0	1665,558	2	1667,558
	M1 vs M2	0,004	2	0,997903201	M1	1662,994	2	1664,994
	M0 vs M3	2,568	4	0,632442041	M2	1662,990	6	1668,990
	M1 vs M3	0,004	4	0,999997852	M3	1662,990	10	1672,990
	M2 vs M3	0,000	2	0,9999975	M7	1663,072	4	1667,072
	M7 vs M8	0,082	2	0,959874243	M8	1662,990	8	1670,990
	eC07	M0 vs M2	1955,646	2	0	M0	11474,345	2
M1 vs M2		472,974	2	1,9728E-103	M1	9991,673	2	9993,673
M0 vs M3		2077,969	4	0	M2	9518,699	6	9524,699
M1 vs M3		595,297	4	1,6145E-127	M3	9396,376	10	9406,376
M2 vs M3		122,323	2	2,74039E-27	M7	9831,869	4	9835,869
M7 vs M8		412,894	2	2,19371E-90	M8	9418,974	8	9426,974
eC11		M0 vs M2	468,047	2	2,3168E-102	M0	4336,360	2
	M1 vs M2	158,756	2	3,36126E-35	M1	4027,069	2	4029,069
	M0 vs M3	483,502	4	2,4779E-103	M2	3868,313	6	3874,313
	M1 vs M3	174,211	4	1,3048E-36	M3	3852,858	10	3862,858
	M2 vs M3	15,455	2	0,000440594	M7	4010,841	4	4014,841
	M7 vs M8	148,385	2	6,0064E-33	M8	3862,456	8	3870,456
	eC14	M0 vs M2	783,973	2	5,7879E-171	M0	3077,708	2
M1 vs M2		555,872	2	1,9672E-121	M1	2849,608	2	2851,608
M0 vs M3		783,995	4	2,2491E-168	M2	2293,735	6	2299,735
M1 vs M3		555,895	4	5,4258E-119	M3	2293,713	10	2303,713
M2 vs M3		0,023	2	0,988761628	M7	2849,805	4	2853,805
M7 vs M8		556,003	2	1,8424E-121	M8	2293,801	8	2301,801
eC17		M0 vs M2	687,032	2	6,5012E-150	M0	4115,129	2
	M1 vs M2	363,970	2	9,2255E-80	M1	3792,067	2	3794,067
	M0 vs M3	692,978	4	1,1555E-148	M2	3428,097	6	3434,097
	M1 vs M3	369,916	4	8,7748E-79	M3	3422,151	10	3432,151
	M2 vs M3	5,946	2	0,051148608	M7	3767,200	4	3771,200
	M7 vs M8	344,776	2	1,35749E-75	M8	3422,423	8	3430,423
	eC18	M0 vs M2	2690,058	2	0	M0	11806,487	2
M1 vs M2		763,929	2	1,3029E-166	M1	9880,358	2	9882,358
M0 vs M3		2709,437	4	0	M2	9116,429	6	9122,429
M1 vs M3		783,308	4	3,168E-168	M3	9097,050	10	9107,050
M2 vs M3		19,379	2	6,19234E-05	M7	9851,977	4	9855,977
M7 vs M8		735,024	2	2,464E-160	M8	9116,953	8	9124,953
eC20		M0 vs M2	3012,040	2	0	M0	12081,259	2
	M1 vs M2	1136,778	2	1,4185E-247	M1	10205,997	2	10207,997
	M0 vs M3	3083,463	4	0	M2	9069,219	6	9075,219
	M1 vs M3	1208,200	4	2,6573E-260	M3	8997,796	10	9007,796
	M2 vs M3	71,423	2	3,09587E-16	M7	10070,517	4	10074,517
	M7 vs M8	1042,717	2	3,7748E-227	M8	9027,800	8	9035,800
	eC21	M0 vs M2	2489,778	2	0	M0	10779,720	2
M1 vs M2		1027,054	2	9,5072E-224	M1	9316,996	2	9318,996
M0 vs M3		2834,695	4	0	M2	8289,942	6	8295,942
M1 vs M3		1371,971	4	8,2654E-296	M3	7945,025	10	7955,025
M2 vs M3		344,917	2	1,26551E-75	M7	9264,908	4	9268,908
M7 vs M8		985,701	2	9,0743E-215	M8	8279,207	8	8287,207
eC22gp		M0 vs M2	5165,944	2	0	M0	19051,867	2
	M1 vs M2	756,108	2	6,5054E-165	M1	14642,031	2	14644,031
	M0 vs M3	5236,655	4	0	M2	13885,923	6	13891,923
	M1 vs M3	826,819	4	1,1909E-177	M3	13815,211	10	13825,211
	M2 vs M3	70,712	2	4,41742E-16	M7	14456,003	4	14460,003
	M7 vs M8	660,846	2	3,1561E-144	M8	13795,157	8	13803,157

Tabla 19. (Continuación)

Individuo	Modelos	2ΔL	df	P valor	Modelos	(-)2Ln L	2K	AIC
eC24	M0 vs M2	673,555	2	5,4879E-147	M0	3812,858	2	3814,858
	M1 vs M2	321,667	2	1,4153E-70	M1	3460,970	2	3462,970
	M0 vs M3	702,218	4	1,1533E-150	M2	3139,303	6	3145,303
	M1 vs M3	350,331	4	1,48804E-74	M3	3110,639	10	3120,639
	M2 vs M3	28,663	2	5,96824E-07	M7	3464,402	4	3468,402
	M7 vs M8	18,212	2	0,000110991	M8	3446,189	8	3454,189
eC28	M0 vs M2	4336,580	2	0	M0	16419,799	2	16421,799
	M1 vs M2	667,366	2	1,2116E-145	M1	12750,585	2	12752,585
	M0 vs M3	4494,198	4	0	M2	12083,219	6	12089,219
	M1 vs M3	824,984	4	2,9748E-177	M3	11925,601	10	11935,601
	M2 vs M3	157,618	2	5,93798E-35	M7	12573,760	4	12577,760
	M7 vs M8	610,768	2	2,3622E-133	M8	11962,991	8	11970,991
eC31gp	M0 vs M2	672,422	2	9,6676E-147	M0	4477,788	2	4479,788
	M1 vs M2	217,875	2	4,88755E-48	M1	4023,241	2	4025,241
	M0 vs M3	675,203	4	8,1494E-145	M2	3805,366	6	3811,366
	M1 vs M3	220,656	4	1,3546E-46	M3	3802,585	10	3812,585
	M2 vs M3	2,781	2	0,248952292	M7	4023,174	4	4027,174
	M7 vs M8	220,107	2	1,60056E-48	M8	3803,066	8	3811,066
eC33gp	M0 vs M2	1250,071	2	3,5518E-272	M0	4056,704	2	4058,704
	M1 vs M2	803,830	2	2,822E-175	M1	3610,462	2	3612,462
	M0 vs M3	1240,237	4	3,0132E-267	M2	2806,633	6	2812,633
	M1 vs M3	793,996	4	1,5341E-170	M3	2816,466	10	2826,466
	M2 vs M3	9,834	2	0,007321449	M7	3614,908	4	3618,908
	M7 vs M8	807,125	2	5,4332E-176	M8	2807,784	8	2815,784
eC38gp	M0 vs M2	131,435	2	2,87964E-29	M0	2606,685	2	2608,685
	M1 vs M2	23,509	2	7,85566E-06	M1	2498,759	2	2500,759
	M0 vs M3	127,298	4	1,47268E-26	M2	2475,250	6	2481,250
	M1 vs M3	19,372	4	0,000664061	M3	2479,387	10	2489,387
	M2 vs M3	4,136	2	0,126413373	M7	2498,768	4	2502,768
	M7 vs M8	19,381	2	6,18597E-05	M8	2479,386	8	2487,386
eG10	M0 vs M2	0,000	2	1	M0	1267,627	2	1269,627
	M1 vs M2	0,028	2	0,985980206	M1	1267,655	2	1269,655
	M0 vs M3	0,000	4	1	M2	1267,627	6	1273,627
	M1 vs M3	0,028	4	0,99990126	M3	1267,627	10	1277,627
	M2 vs M3	0,000	2	1	M7	1267,627	4	1271,627
	M7 vs M8	0,000	2	0,999947001	M8	1267,627	8	1275,627
eG14gp	M0 vs M2	1099,362	2	1,8903E-239	M0	7749,552	2	7751,552
	M1 vs M2	252,199	2	1,72084E-55	M1	6902,389	2	6904,389
	M0 vs M3	1131,868	4	9,3646E-244	M2	6650,190	6	6656,190
	M1 vs M3	284,705	4	2,15562E-60	M3	6617,684	10	6627,684
	M2 vs M3	32,506	2	8,73833E-08	M7	6858,025	4	6862,025
	M7 vs M8	222,000	2	6,21275E-49	M8	6636,025	8	6644,025
eG19gp	M0 vs M2	1997,022	2	0	M0	10030,955	2	10032,955
	M1 vs M2	286,322	2	6,69755E-63	M1	8320,255	2	8322,255
	M0 vs M3	1983,888	4	0	M2	8033,933	6	8039,933
	M1 vs M3	273,188	4	6,55522E-58	M3	8047,067	10	8057,067
	M2 vs M3	13,134	2	0,001405815	M7	8261,010	4	8265,010
	M7 vs M8	262,109	2	1,21297E-57	M8	7998,902	8	8006,902

Tabla 20. Comparación de modelos según el test de razón de verosimilitudes y AIC para la región NS5A en los individuos 1b. En rojo se señalan los valores de AIC más bajos.

Individuo	Modelos	2ΔL	df	P valor	Modelos	(-)2Ln L	2K	AIC
nA03	M0 vs M2	117,029674	2	3,8666E-26	M0	4224,910826	2	4226,91083
	M1 vs M2	50,543562	2	1,05829E-11	M1	4158,424714	2	4160,42471
	M0 vs M3	137,811218	4	8,30207E-29	M2	4107,881152	6	4113,88115
	M1 vs M3	71,325106	4	1,19172E-14	M3	4087,099608	10	4097,09961
	M2 vs M3	20,781544	2	3,07146E-05	M7	4153,350954	4	4157,35095
	M7 vs M8	66,124754	2	4,37716E-15	M8	4087,2262	8	4095,2262
nA06	M0 vs M3	233,71385	2	1,77698E-51	M0	7179,795858	2	7181,79586
	M1 vs M3	115,48103	2	8,38712E-26	M1	7061,563038	2	7063,56304
	M0 vs M4	140,479616	4	2,22824E-29	M2	6946,082008	6	6952,08201
	M1 vs M4	22,246796	4	0,000178976	M3	7039,316242	10	7049,31624
	M2 vs M4	93,234234	2	5,68124E-21	M7	7041,383218	4	7045,38322
	M7 vs M9	108,349872	2	2,96568E-24	M8	6933,033346	8	6941,03335
nA14	M0 vs M4	746,922484	2	6,4246E-163	M0	5833,631906	2	5835,63191
	M1 vs M4	147,327468	2	1,01917E-32	M1	5234,03689	2	5236,03689
	M0 vs M5	746,973638	4	2,3452E-160	M2	5086,709422	6	5092,70942
	M1 vs M5	147,378622	4	7,41989E-31	M3	5086,658268	10	5096,65827
	M2 vs M5	0,051154	2	0,974747321	M7	5244,696544	4	5248,69654
	M7 vs M10	131,213144	2	3,21683E-29	M8	5113,4834	8	5121,4834
nA16	M0 vs M5	403,950142	2	1,92018E-88	M0	12655,56467	2	12657,5647
	M1 vs M5	161,013944	2	1,08709E-35	M1	12898,50087	2	12900,5009
	M0 vs M6	425,449812	4	8,80242E-91	M2	13059,51482	6	13065,5148
	M1 vs M6	668,38601	4	2,4385E-143	M3	12230,11486	10	12240,1149
	M2 vs M6	829,399954	2	7,9084E-181	M7	12280,48028	4	12284,4803
	M7 vs M11	48,937208	2	2,36276E-11	M8	12231,54307	8	12239,5431
nA35	M0 vs M6	140,72875	2	2,76147E-31	M0	2995,202864	2	2997,20286
	M1 vs M6	63,693034	2	1,4765E-14	M1	2918,167148	2	2920,16715
	M0 vs M7	141,009352	4	1,7161E-29	M2	2854,474114	6	2860,47411
	M1 vs M7	63,973636	4	4,23294E-13	M3	2854,193512	10	2864,19351
	M2 vs M7	0,280602	2	0,869096598	M7	2920,172668	4	2924,17267
	M7 vs M12	65,974242	2	4,71928E-15	M8	2854,198426	8	2862,19843
nC03	M0 vs M7	129,807714	2	6,49551E-29	M0	3115,846476	2	3117,84648
	M1 vs M7	63,405946	2	1,70441E-14	M1	3049,444708	2	3051,44471
	M0 vs M8	131,29837	4	2,05455E-27	M2	2986,038762	6	2992,03876
	M1 vs M8	64,896602	4	2,70555E-13	M3	2984,548106	10	2994,54811
	M2 vs M8	1,490656	2	0,47457862	M7	3049,504098	4	3053,5041
	M7 vs M13	43,656284	2	3,31252E-10	M8	3005,847814	8	3013,84781
nC04	M0 vs M8	13,097104	2	0,001432188	M0	4738,985464	2	4740,98546
	M1 vs M8	7,147096	2	0,028056134	M1	4733,035456	2	4735,03546
	M0 vs M9	13,240402	4	0,01015897	M2	4725,88836	6	4731,88836
	M1 vs M9	7,290394	4	0,121315193	M3	4725,745062	10	4735,74506
	M2 vs M9	0,143298	2	0,93085757	M7	4727,15573	4	4731,15573
	M7 vs M14	0,000502	2	0,999749031	M8	4727,156232	8	4735,15623
nC05	M0 vs M9	192,81979	2	1,34808E-42	M0	5401,253676	2	5403,25368
	M1 vs M9	32,350462	2	9,44466E-08	M1	5240,784348	2	5242,78435
	M0 vs M10	208,125726	4	6,72236E-44	M2	5208,433886	6	5214,43389
	M1 vs M10	47,656398	4	1,11299E-09	M3	5193,12795	10	5203,12795
	M2 vs M10	15,305936	2	0,000474633	M7	5241,421192	4	5245,42119
	M7 vs M15	48,120982	2	3,55354E-11	M8	5193,30021	8	5201,30021

2ΔL: 2 veces la diferencia del logaritmo de las verosimilitudes.

df: grados de libertad.

L: valor de verosimilitud.

K: nº de parámetros del modelo.

Tabla 20. (Continuación)

Individuo	Modelos	$2\Delta L$	df	P valor	Modelos	$(-2)\ln L$	2K	AIC
nC08	M0 vs M10	1310,173706	2	3,1579E-285	M0	10018,04909	2	10020,0491
	M1 vs M10	149,825582	2	2,92273E-33	M1	8857,700966	2	8859,70097
	M0 vs M11	1373,771326	4	3,3641E-296	M2	8707,875384	6	8713,87538
	M1 vs M11	213,423202	4	4,87532E-45	M3	8644,277764	10	8654,27776
	M2 vs M11	63,59762	2	1,54865E-14	M7	8791,208138	4	8795,20814
	M7 vs M16	139,806276	2	4,37979E-31	M8	8651,401862	8	8659,40186
nC09	M0 vs M2	0,046	2	0,97726	M0	-0,5032	2	1,497
	M1 vs M2	0,423	2	0,80929	M1	-0,126	2	1,874
	M0 vs M3	0,021	4	0,99995	M2	-0,5492	6	5,451
	M1 vs M3	0,398	4	0,98262	M3	-0,5242	10	9,476
	M2 vs M3	0,025	2	0,98758	M7	-0,2	4	3,800
	M7 vs M8	0,324	2	0,85027	M8	-0,5244	8	7,476
nC10	M0 vs M3	85,230	2	3,10773E-19	M0	4121,92426	2	4123,924
	M1 vs M3	16,390	2	0,000276	M1	4053,08362	2	4055,084
	M0 vs M4	87,399	4	4,69696E-18	M2	4036,69381	6	4042,694
	M1 vs M4	18,559	4	0,00096	M3	4034,525098	10	4044,525
	M2 vs M4	2,169	2	0,33812	M7	4052,356796	4	4056,357
	M7 vs M9	17,764	2	0,00014	M8	4034,592656	8	4042,593
nC12	M0 vs M4	704,475	2	1,0598E-153	M0	8704,98836	2	8706,988
	M1 vs M4	209,843	2	2,71084E-46	M1	8210,35672	2	8212,357
	M0 vs M5	741,912	4	2,927E-159	M2	8000,513412	6	8006,513
	M1 vs M5	247,280	4	2,50894E-52	M3	7963,076732	10	7973,077
	M2 vs M5	37,437	2	7,42554E-09	M7	8178,325638	4	8182,326
	M7 vs M10	215,154	2	1,90472E-47	M8	7963,171306	8	7971,171
nC13	M0 vs M5	99,403	2	2,60005E-22	M0	6780,141284	2	6782,141
	M1 vs M5	52,888	2	3,27753E-12	M1	6733,62646	2	6735,626
	M0 vs M6	99,248	4	1,42238E-20	M2	6680,7386	6	6686,739
	M1 vs M6	52,733	4	9,69266E-11	M3	6680,893696	10	6690,894
	M2 vs M6	0,155	2	0,925382605	M7	6693,413692	4	6697,414
	M7 vs M11	0,076	2	0,962880468	M8	6693,33804	8	6701,338
nC15	M0 vs M6	389,082	2	3,24957E-85	M0	6278,829424	2	6280,829
	M1 vs M6	94,595	2	2,87775E-21	M1	5984,34156	2	5986,342
	M0 vs M7	413,435	4	3,47731E-88	M2	5889,747004	6	5895,747
	M1 vs M7	118,947	4	8,96524E-25	M3	5865,394606	10	5875,395
	M2 vs M7	24,352	2	5,15162E-06	M7	5986,194722	4	5990,195
	M7 vs M12	118,539	2	1,81762E-26	M8	5867,65535	8	5875,655
nC16	M0 vs M7	47,531	2	4,77273E-11	M0	5844,41312	2	5846,413
	M1 vs M7	27,987	2	8,37091E-07	M1	5824,868752	2	5826,869
	M0 vs M8	47,689	4	1,09553E-09	M2	5796,882086	6	5802,882
	M1 vs M8	28,145	4	1,16569E-05	M3	5796,723784	10	5806,724
	M2 vs M8	0,158	2	0,923900405	M7	5802,594916	4	5806,595
	M7 vs M13	5,853	2	0,053585381	M8	5796,741958	8	5804,742
nC19	M0 vs M7	912,792	2	6,161E-199	M0	10366,26559	2	10368,266
	M1 vs M7	258,966	2	5,8385E-57	M1	9712,438954	2	9714,439
	M0 vs M8	987,941	4	1,4651E-212	M2	9453,4732	6	9459,473
	M1 vs M8	334,115	4	4,71418E-71	M3	9378,324412	10	9388,324
	M2 vs M8	75,149	2	4,8045E-17	M7	9641,135484	4	9645,135
	M7 vs M13	257,498	2	1,21634E-56	M8	9383,63764	8	9391,638
nC25	M0 vs M8	35,357	2	2,10041E-08	M0	2279,835068	2	2281,835
	M1 vs M8	24,880	2	3,95761E-06	M1	2269,357718	2	2271,358
	M0 vs M9	35,630	4	3,4471E-07	M2	2244,477976	6	2250,478
	M1 vs M9	25,153	4	4,68664E-05	M3	2244,204608	10	2254,205
	M2 vs M9	0,273	2	0,872245812	M7	2269,389612	4	2273,390
	M7 vs M14	25,183	2	3,40053E-06	M8	2244,206452	8	2252,206

Tabla 20. (Continuación)

Individuo	Modelos	2ΔL	df	P valor	Modelos	(-)2Ln L	2K	AIC
nG02	M0 vs M9	445,521	2	1,80442E-97	M0	8200,621866	2	8202,622
	M1 vs M9	73,946	2	8,7657E-17	M1	7829,047038	2	7831,047
	M0 vs M10	516,400	4	1,8999E-110	M2	7755,100838	6	7761,101
	M1 vs M10	144,825	4	2,61407E-30	M3	7684,221558	10	7694,222
	M2 vs M10	70,879	2	4,06219E-16	M7	7768,193674	4	7772,194
	M7 vs M15	80,889	2	2,72341E-18	M8	7687,304382	8	7695,304
nG05	M0 vs M9	1065,328	2	4,6457E-232	M0	8365,135	2	8367,135
	M1 vs M9	287,121	2	4,4916E-63	M1	7586,929	2	7588,929
	M0 vs M10	1131,651	4	1,0434E-243	M2	7299,808	6	7305,808
	M1 vs M10	353,445	4	3,16304E-75	M3	7233,484	10	7243,484
	M2 vs M10	66,324	2	3,96243E-15	M7	7532,690	4	7536,690
	M7 vs M15	288,620	2	2,12284E-63	M8	7244,070	8	7252,070
nG06	M0 vs M9	799,679	2	2,2488E-174	M0	9221,362	2	9223,362
	M1 vs M9	220,410	2	1,37598E-48	M1	8642,093	2	8644,093
	M0 vs M10	875,339	4	3,6696E-188	M2	8421,683	6	8427,683
	M1 vs M10	296,070	4	7,62835E-63	M3	8346,023	10	8356,023
	M2 vs M10	75,661	2	3,71989E-17	M7	8563,765	4	8567,765
	M7 vs M15	207,353	2	9,41674E-46	M8	8356,412	8	8364,412
nG07	M0 vs M10	503,691	2	4,2162E-110	M0	8723,094	2	8725,094
	M1 vs M10	111,040	2	7,72546E-25	M1	8330,444	2	8332,444
	M0 vs M11	654,081	4	3,0476E-140	M2	8219,404	6	8225,404
	M1 vs M11	261,431	4	2,24219E-55	M3	8069,013	10	8079,013
	M2 vs M11	150,391	2	2,20349E-33	M7	8216,772	4	8220,772
	M7 vs M16	147,497	2	9,36383E-33	M8	8069,275	8	8077,275
nG09	M0 vs M10	498,914	2	4,5951E-109	M0	7382,493	2	7384,493
	M1 vs M10	85,971	2	2,14551E-19	M1	6969,551	2	6971,551
	M0 vs M11	527,751	4	6,6568E-113	M2	6883,580	6	6889,580
	M1 vs M11	114,809	4	6,85337E-24	M3	6854,742	10	6864,742
	M2 vs M11	28,838	2	5,46922E-07	M7	6958,810	4	6962,810
	M7 vs M16	84,820	2	3,81553E-19	M8	6873,990	8	6881,990
nG16	M0 vs M11	0,001	2	0,999579089	M0	2001,950	2	2003,950
	M1 vs M11	1,629	2	0,44293201	M1	2003,579	2	2005,579
	M0 vs M12	0,000	4	1	M2	2001,951	6	2007,951
	M1 vs M12	1,630	4	0,803476958	M3	2001,950	10	2011,950
	M2 vs M12	0,001	2	0,999579089	M7	2003,579	4	2007,579
	M7 vs M17	1,630	2	0,442746018	M8	2001,950	8	2009,950
nG17	M0 vs M12	230,947	2	7,08596E-51	M0	6598,395	2	6600,395
	M1 vs M12	18,995	2	7,50341E-05	M1	6386,443	2	6388,443
	M0 vs M13	247,281	4	2,50739E-52	M2	6367,448	6	6373,448
	M1 vs M13	35,329	4	3,97591E-07	M3	6351,114	10	6361,114
	M2 vs M13	16,334	2	0,000283898	M7	6391,353	4	6395,353
	M7 vs M18	39,733	2	2,35558E-09	M8	6351,620	8	6359,620
nG26	M0 vs M13	0,000	2	1	M0	2030,308	2	2032,308
	M1 vs M13	0,035	2	0,982593278	M1	2030,343	2	2032,343
	M0 vs M14	0,000	4	1	M2	2030,308	6	2036,308
	M1 vs M14	0,035	4	0,999847616	M3	2030,308	10	2040,308
	M2 vs M14	0,000	2	1	M7	2030,308	4	2034,308
	M7 vs M19	3E-05	2	0,999985	M8	2030,308	8	2038,308

Tabla 21. Comparación de modelos según el test de razón de verosimilitudes y AIC para la región NS5A en los individuos 1a. En rojo se señalan los valores de AIC más bajos.

Individuo	Modelos	2ΔL	df	P valor	Modelos	(-)2Ln L	2K	AIC
nA05	M0 vs M2	0,000	2	1	M0	2347,611	2	2349,611
	M1 vs M2	1,027	2	0,59825871	M1	2348,638	2	2350,638
	M0 vs M3	0,000	4	1	M2	2347,611	6	2353,611
	M1 vs M3	1,027	4	0,905603348	M3	2347,611	10	2357,611
	M2 vs M3	0,000	2	1	M7	2347,619	4	2351,619
	M7 vs M8	0,000	2	1	M8	2347,619	8	2355,619
nA09	M0 vs M2	50,623	2	1,01733E-11	M0	3663,007	2	3665,007
	M1 vs M2	2,606	2	0,27176189	M1	3614,990	2	3616,990
	M0 vs M3	52,515	4	1,07648E-10	M2	3612,385	6	3618,385
	M1 vs M3	4,498	4	0,342771407	M3	3610,492	10	3620,492
	M2 vs M3	1,892	2	0,388202959	M7	3619,175	4	3623,175
	M7 vs M8	8,625	2	0,013402286	M8	3610,551	8	3618,551
nA20	M0 vs M2	1018,062	2	8,5245E-222	M0	6107,143	2	6109,143
	M1 vs M2	88,394	2	6,38866E-20	M1	5177,476	2	5179,476
	M0 vs M3	990,334	4	4,4387E-213	M2	5089,081	6	5095,081
	M1 vs M3	60,667	4	2,10077E-12	M3	5116,809	10	5126,809
	M2 vs M3	27,728	2	9,52884E-07	M7	5188,464	4	5192,464
	M7 vs M8	65,583	2	5,73915E-15	M8	5122,881	8	5130,881
nA26	M0 vs M2	1016,235	2	2,1251E-221	M0	7657,941	2	7659,941
	M1 vs M2	341,745	2	6,18063E-75	M1	6983,451	2	6985,451
	M0 vs M3	1027,487	4	3,9413E-221	M2	6641,706	6	6647,706
	M1 vs M3	352,997	4	3,95272E-75	M3	6630,454	10	6640,454
	M2 vs M3	11,252	2	0,00360303	M7	6974,327	4	6978,327
	M7 vs M8	339,766	2	1,66259E-74	M8	6634,562	8	6642,562
nA27	M0 vs M2	216,718	2	8,71517E-48	M0	6299,896	2	6301,896
	M1 vs M2	65,880	2	4,94587E-15	M1	6149,058	2	6151,058
	M0 vs M3	239,702	4	1,07568E-50	M2	6083,178	6	6089,178
	M1 vs M3	88,864	4	2,29488E-18	M3	6060,194	10	6070,194
	M2 vs M3	22,984	2	1,02131E-05	M7	6140,635	4	6144,635
	M7 vs M8	80,264	2	3,7238E-18	M8	6060,371	8	6068,371
nA34	M0 vs M2	251,014	2	3,11209E-55	M0	4215,001	2	4217,001
	M1 vs M2	83,233	2	8,43556E-19	M1	4047,221	2	4049,221
	M0 vs M3	251,474	4	3,13324E-53	M2	3963,987	6	3969,987
	M1 vs M3	83,694	4	2,87124E-17	M3	3963,527	10	3973,527
	M2 vs M3	0,460	2	0,794395366	M7	4049,289	4	4053,289
	M7 vs M8	84,848	2	3,76289E-19	M8	3964,441	8	3972,441
nC01	M0 vs M2	79,222	2	6,26891E-18	M0	4697,795	2	4699,795
	M1 vs M2	11,785	2	0,002759624	M1	4630,359	2	4632,359
	M0 vs M3	87,150	4	5,30466E-18	M2	4618,574	6	4624,574
	M1 vs M3	19,714	4	0,000568757	M3	4610,645	10	4620,645
	M2 vs M3	7,928	2	0,01898335	M7	4627,257	4	4631,257
	M7 vs M8	16,543	2	0,00025575	M8	4610,714	8	4618,714
nC02	M0 vs M2	229,043	2	1,83616E-50	M0	3045,133	2	3047,133
	M1 vs M2	161,837	2	7,20429E-36	M1	2977,927	2	2979,927
	M0 vs M3	237,656	4	2,96601E-50	M2	2816,090	6	2822,090
	M1 vs M3	170,450	4	8,37388E-36	M3	2807,477	10	2817,477
	M2 vs M3	8,613	2	0,013480407	M7	2978,697	4	2982,697
	M7 vs M8	162,373	2	5,51069E-36	M8	2816,324	8	2824,324

2ΔL: 2 veces la diferencia del logaritmo de las verosimilitudes.

df: grados de libertad.

L: valor de verosimilitud.

K: n° de parámetros del modelo.

Tabla 21. (Continuación)

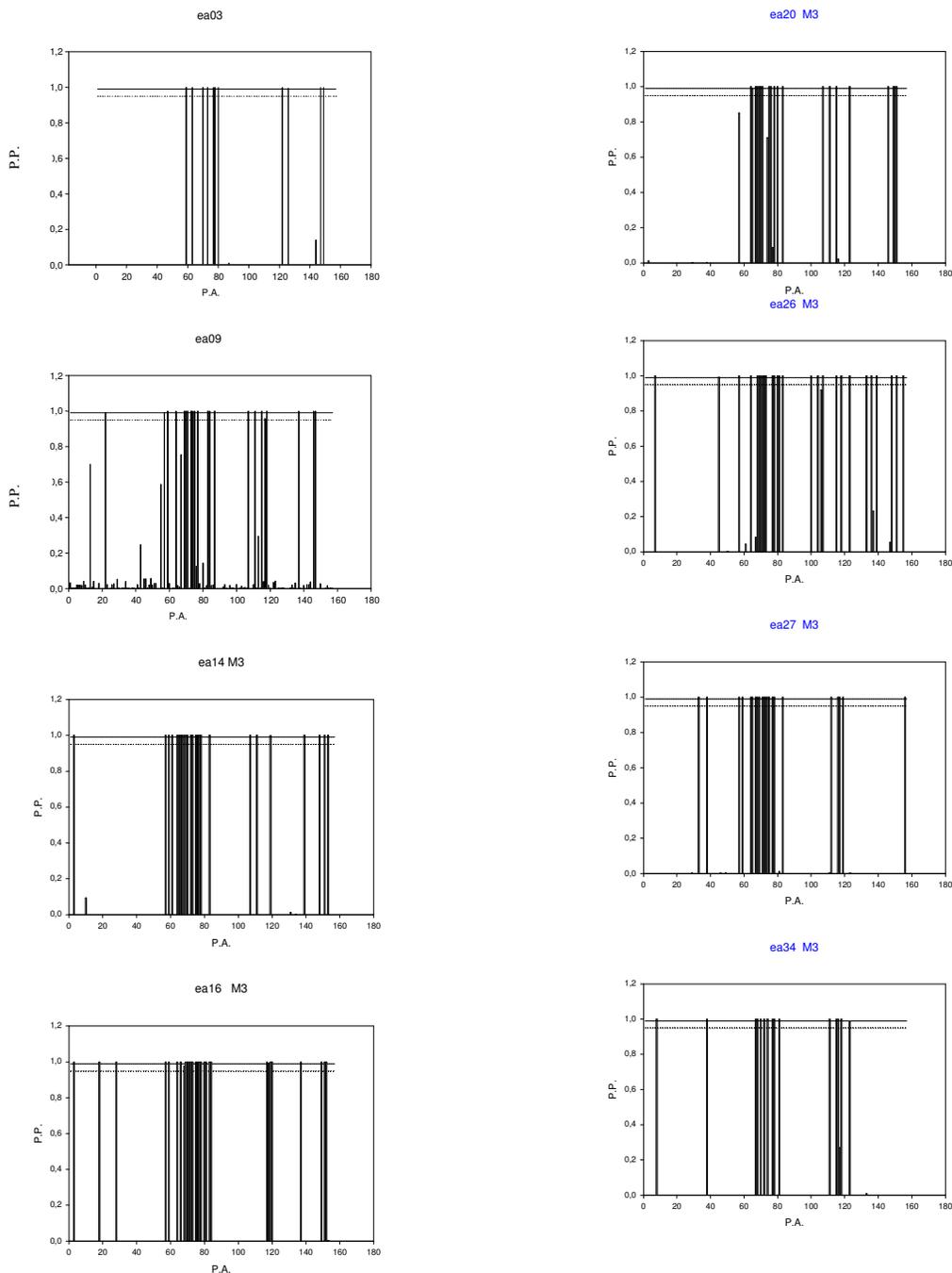
Individuo	Modelos	ΔL	df	P valor	Modelos	(-) 2Ln Lnl	2K	AIC
nC06	M0 vs M2	0,000	2	1	M0	2271,950	2	2273,950
	M1 vs M2	0,338	2	0,844513971	M1	2272,288	2	2274,288
	M0 vs M3	0,000	4	1	M2	2271,950	6	2277,950
	M1 vs M3	0,338	4	0,987231764	M3	2271,950	10	2281,950
	M2 vs M3	0,000	2	1	M7	2271,952	4	2275,952
	M7 vs M8	0,000	2	1	M8	2271,952	8	2279,952
nC07	M0 vs M2	685,343	2	1,5127E-149	M0	6848,953	2	6850,953
	M1 vs M2	205,572	2	2,29419E-45	M1	6369,182	2	6371,182
	M0 vs M3	693,969	4	7,0484E-149	M2	6163,610	6	6169,610
	M1 vs M3	214,198	4	3,32061E-45	M3	6154,984	10	6164,984
	M2 vs M3	8,627	2	0,01338952	M7	6363,725	4	6367,725
	M7 vs M8	201,752	2	1,54938E-44	M8	6161,973	8	6169,973
nC11	M0 vs M2	269,461	2	3,07144E-59	M0	3371,377	2	3373,377
	M1 vs M2	115,726	2	7,4212E-26	M1	3217,642	2	3219,642
	M0 vs M3	270,639	4	2,32356E-57	M2	3101,917	6	3107,917
	M1 vs M3	116,903	4	2,44846E-24	M3	3100,739	10	3110,739
	M2 vs M3	1,178	2	0,55494959	M7	3224,075	4	3228,075
	M7 vs M8	108,246	2	3,1237E-24	M8	3115,829	8	3123,829
nC14	M0 vs M2	0,021	2	0,989709315	M0	2512,285	2	2514,285
	M1 vs M2	0,000	2	0,999918003	M1	2512,264	2	2514,264
	M0 vs M3	0,019	4	0,99995693	M2	2512,264	6	2518,264
	M1 vs M3	0,002	4	0,999999547	M3	2512,266	10	2522,266
	M2 vs M3	0,002	2	0,998966534	M7	2512,264	4	2516,264
	M7 vs M8	0,000	2	0,999978	M8	2512,264	8	2520,264
nC17	M0 vs M2	0,000	2	1	M0	2075,199	2	2077,199
	M1 vs M2	0,002	2	0,998840673	M1	2075,202	2	2077,202
	M0 vs M3	0,000	4	1	M2	2075,199	6	2081,199
	M1 vs M3	0,002	4	0,999999328	M3	2075,199	10	2085,199
	M2 vs M3	0,000	2	1	M7	2075,202	4	2079,202
	M7 vs M8	0,003	2	0,998530081	M8	2075,199	8	2083,199
nC18	M0 vs M2	0,000	2	1	M0	2057,730	2	2059,730
	M1 vs M2	0,032	2	0,984042689	M1	2057,762	2	2059,762
	M0 vs M3	0,000	4	1	M2	2057,730	6	2063,730
	M1 vs M3	0,032	4	0,999871999	M3	2057,730	10	2067,730
	M2 vs M3	0,000	2	1	M7	2057,731	4	2061,731
	M7 vs M8	0,000	2	0,999895006	M8	2057,731	8	2065,731
nC20	M0 vs M2	503,617	2	4,3757E-110	M0	4532,025	2	4534,025
	M1 vs M2	255,592	2	3,1539E-56	M1	4284,001	2	4286,001
	M0 vs M3	504,261	4	8,0268E-108	M2	4028,409	6	4034,409
	M1 vs M3	256,236	4	2,95111E-54	M3	4027,765	10	4037,765
	M2 vs M3	0,644	2	0,724686597	M7	4287,803	4	4291,803
	M7 vs M8	265,112	2	2,70234E-58	M8	4022,692	8	4030,692
nC21	M0 vs M2	661,817	2	1,9425E-144	M0	5804,389	2	5806,389
	M1 vs M2	83,831	2	6,2568E-19	M1	5226,403	2	5228,403
	M0 vs M3	591,848	4	9,0064E-127	M2	5142,572	6	5148,572
	M1 vs M3	13,862	4	0,007748522	M3	5212,541	10	5222,541
	M2 vs M3	69,969	2	6,40418E-16	M7	5226,439	4	5230,439
	M7 vs M8	14,915	2	0,000577233	M8	5211,525	8	5219,525
nC22	M0 vs M2	791,466	2	1,3654E-172	M0	7144,045	2	7146,045
	M1 vs M2	17,476	2	0,00016041	M1	6370,055	2	6372,055
	M0 vs M3	775,882	4	1,2862E-166	M2	6352,579	6	6358,579
	M1 vs M3	1,891	4	0,755830808	M3	6368,164	10	6378,164
	M2 vs M3	15,585	2	0,000412874	M7	6392,889	4	6396,889
	M7 vs M8	14,670	2	0,000652219	M8	6378,218	8	6386,218

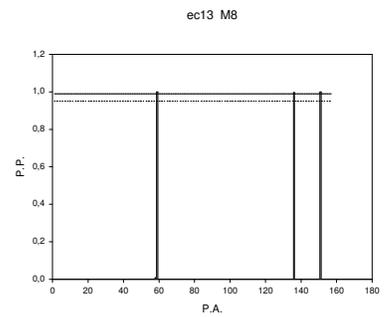
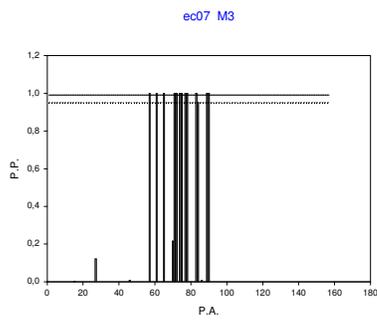
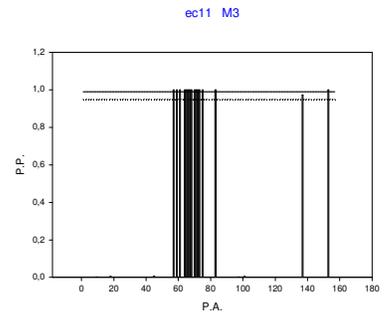
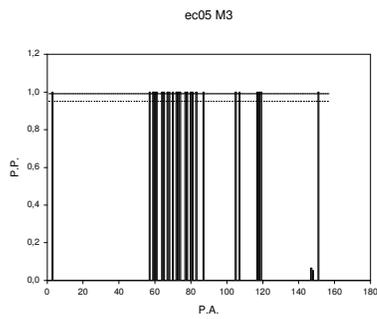
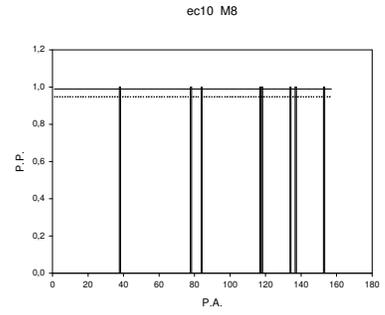
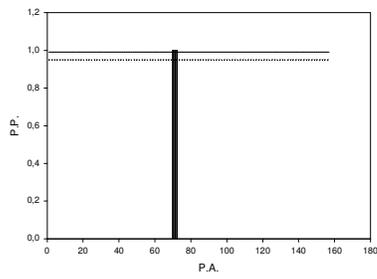
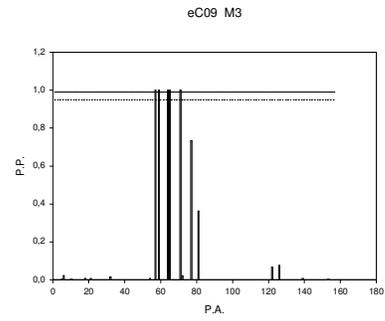
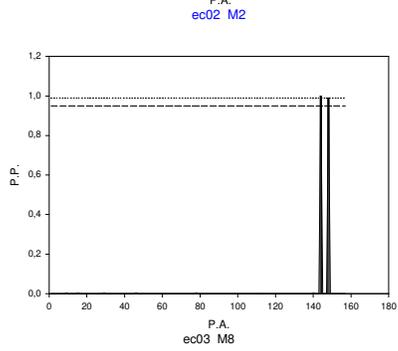
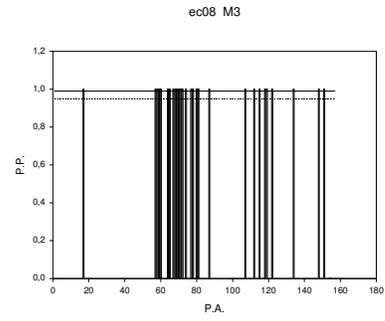
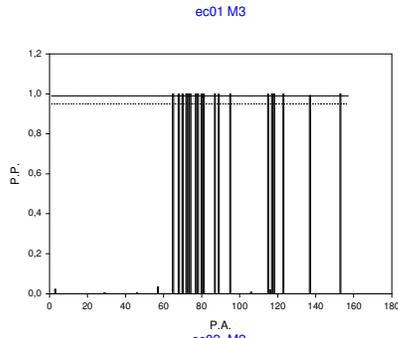
Tabla 21. (Continuación)

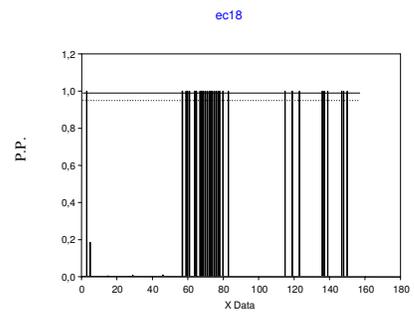
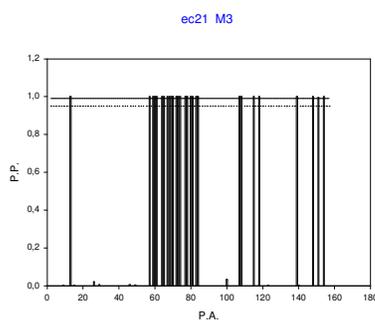
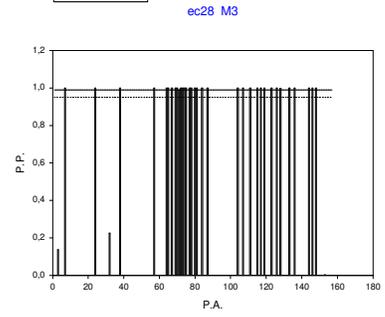
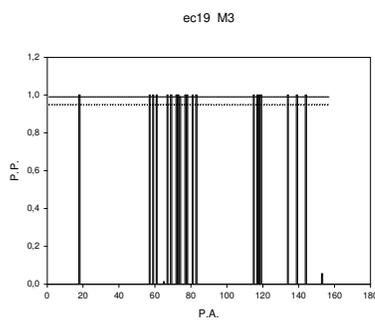
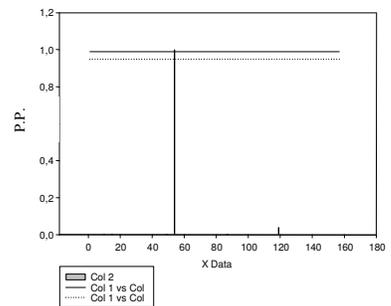
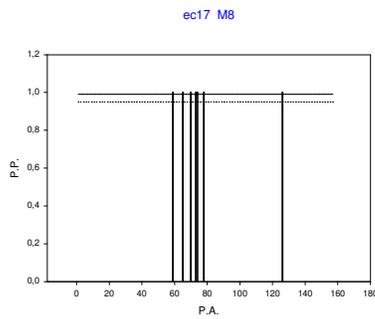
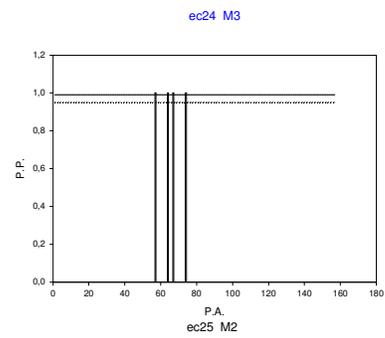
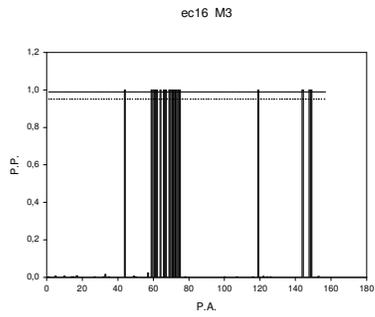
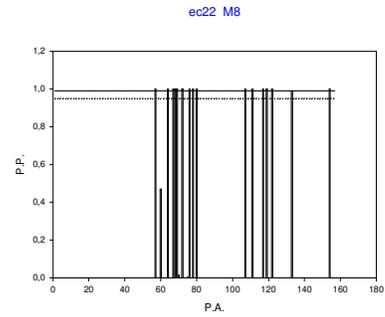
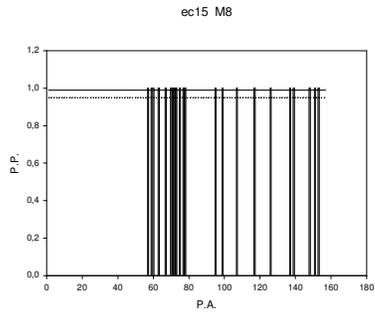
Individuo	Modelos	2ΔL	df	P valor	Modelos	(-)2Lnl	2K	AIC
nC24	M0 vs M2	139,533	2	5,02012E-31	M0	2909,982	2	2911,982
	M1 vs M2	92,563	2	7,94533E-21	M1	2863,012	2	2865,012
	M0 vs M3	140,115	4	2,6674E-29	M2	2770,449	6	2776,449
	M1 vs M3	93,145	4	2,82639E-19	M3	2769,868	10	2779,868
	M2 vs M3	0,581	2	0,747765391	M7	2864,350	4	2868,350
	M7 vs M8	93,386	2	5,26655E-21	M8	2770,964	8	2778,964
nC24	M0 vs M2	139,533	2	5,02012E-31	M0	2909,982	2	2911,982
	M1 vs M2	92,563	2	7,94533E-21	M1	2863,012	2	2865,012
	M0 vs M3	140,115	4	2,6674E-29	M2	2770,449	6	2776,449
	M1 vs M3	93,145	4	2,82639E-19	M3	2769,868	10	2779,868
	M2 vs M3	0,581	2	0,747765391	M7	2864,350	4	2868,350
	M7 vs M8	93,386	2	5,26655E-21	M8	2770,964	8	2778,964
nC28	M0 vs M2	306,537	2	2,73049E-67	M0	5171,596	2	5173,596
	M1 vs M2	59,671	2	1,10335E-13	M1	4924,729	2	4926,729
	M0 vs M3	307,595	4	2,49132E-65	M2	4865,058	6	4871,058
	M1 vs M3	60,728	4	2,03971E-12	M3	4864,001	10	4874,001
	M2 vs M3	1,057	2	0,589420168	M7	4924,533	4	4928,533
	M7 vs M8	59,213	2	1,38676E-13	M8	4865,320	8	4873,320
nC31	M0 vs M2	115,960	2	6,60246E-26	M0	3103,324	2	3105,324
	M1 vs M2	42,060	2	7,35755E-10	M1	3029,425	2	3031,425
	M0 vs M3	117,447	4	1,87472E-24	M2	2987,364	6	2993,364
	M1 vs M3	43,547	4	7,96622E-09	M3	2985,877	10	2995,877
	M2 vs M3	1,487	2	0,475429352	M7	3029,522	4	3033,522
	M7 vs M8	30,826	2	2,02385E-07	M8	2998,696	8	3006,696
nC33	M0 vs M2	0,000	2	1	M0	2019,484	2	2021,484
	M1 vs M2	2,177	2	0,336781812	M1	2021,661	2	2023,661
	M0 vs M3	0,000	4	1	M2	2019,484	6	2025,484
	M1 vs M3	2,177	4	0,703308196	M3	2019,484	10	2029,484
	M2 vs M3	0,000	2	1	M7	2021,661	4	2025,661
	M7 vs M8	2,177	2	0,336782149	M8	2019,484	8	2027,484
nC38	M0 vs M2	0,000	2	1	M0	2026,353	2	2028,353
	M1 vs M2	0,024	2	0,988095427	M1	2026,377	2	2028,377
	M0 vs M3	0,000	4	1	M2	2026,353	6	2032,353
	M1 vs M3	0,024	4	0,999928858	M3	2026,353	10	2036,353
	M2 vs M3	0,000	2	1	M7	2026,353	4	2030,353
	M7 vs M8	0,000	2	0,999967001	M8	2026,353	8	2034,353
nG10	M0 vs M2	33,678	2	4,86257E-08	M0	2206,427	2	2208,427
	M1 vs M2	25,119	2	3,51075E-06	M1	2197,869	2	2199,869
	M0 vs M3	33,959	4	7,59668E-07	M2	2172,749	6	2178,749
	M1 vs M3	25,400	4	4,1793E-05	M3	2172,468	10	2182,468
	M2 vs M3	0,281	2	0,868914976	M7	2197,877	4	2201,877
	M7 vs M8	14,415	2	0,000740947	M8	2183,462	8	2191,462
nG14	M0 vs M2	172,496	2	3,49127E-38	M0	6100,547	2	6102,547
	M1 vs M2	9,069	2	0,010733259	M1	5937,120	2	5939,120
	M0 vs M3	187,550	4	1,78089E-39	M2	5928,051	6	5934,051
	M1 vs M3	24,123	4	7,54549E-05	M3	5912,996	10	5922,996
	M2 vs M3	15,054	2	0,000538217	M7	5929,436	4	5933,436
	M7 vs M8	15,967	2	0,000340997	M8	5913,469	8	5921,469
nG19	M0 vs M2	783,388	2	7,7517E-171	M0	5764,056	2	5766,056
	M1 vs M2	48,235	2	3,35645E-11	M1	5028,903	2	5030,903
	M0 vs M3	767,951	4	6,7132E-165	M2	4980,668	6	4986,668
	M1 vs M3	32,798	4	1,31373E-06	M3	4996,105	10	5006,105
	M2 vs M3	15,437	2	0,00044453	M7	5037,919	4	5041,919
	M7 vs M8	41,751	2	8,5884E-10	M8	4996,168	8	5004,168

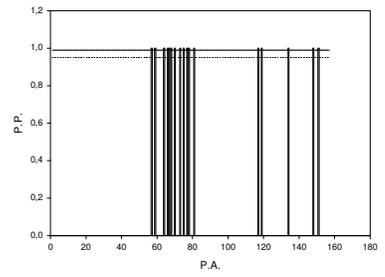
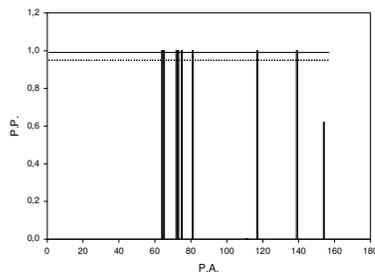
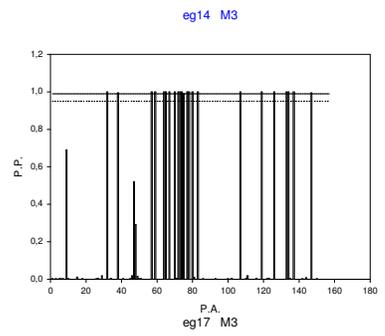
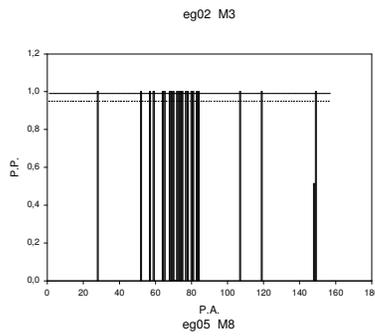
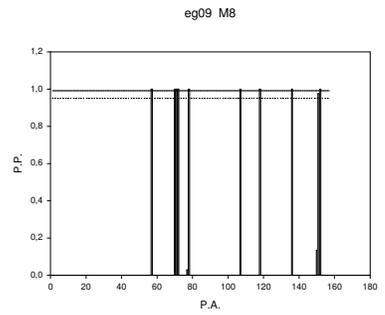
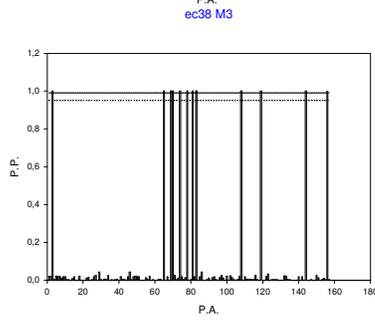
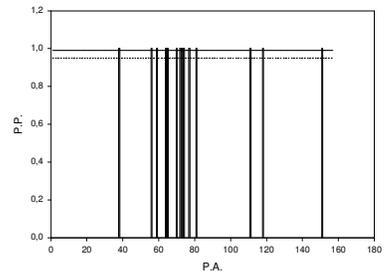
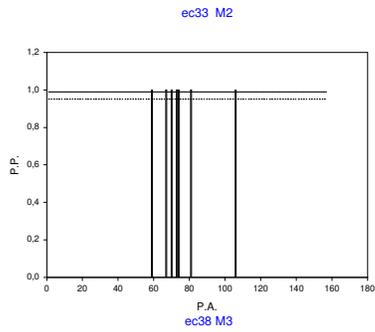
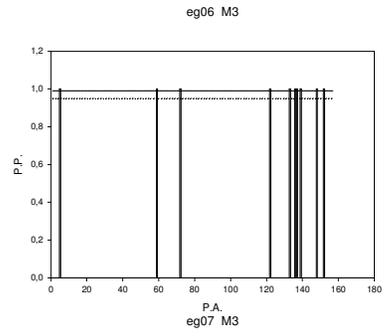
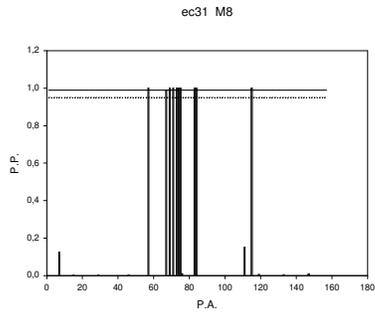
Distribución de probabilidades a posteriori

A continuación se presenta la distribución de las probabilidades a posteriori (P.P.) de las posiciones aminoacídicas (P.A.) bajo selección positiva. La línea continua muestra un nivel de significación al 99% y la discontinua al 95 %.

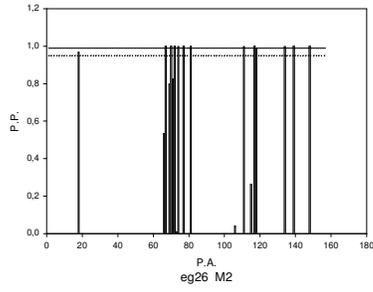




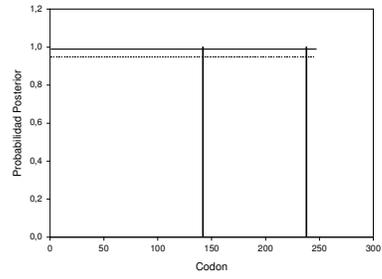




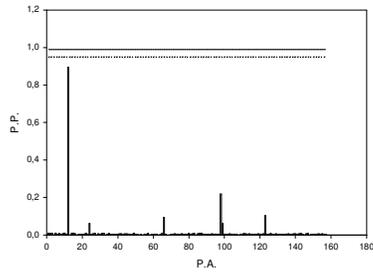
eg19 M8



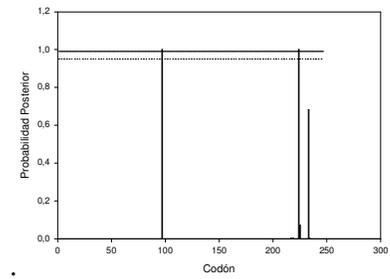
nA03 M8



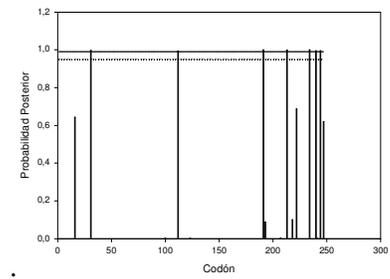
eg26 M2



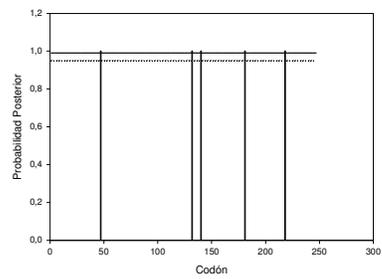
nA06



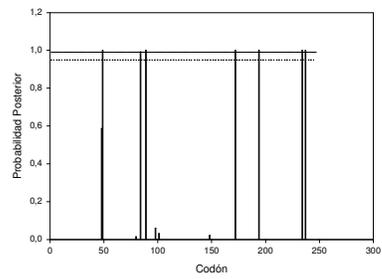
nA14 M2

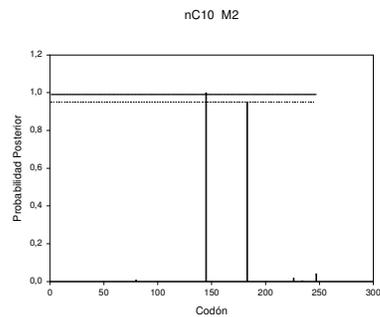
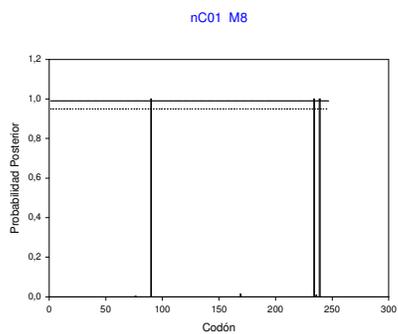
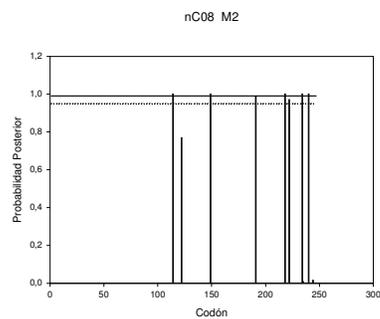
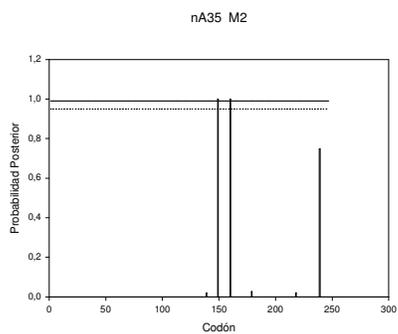
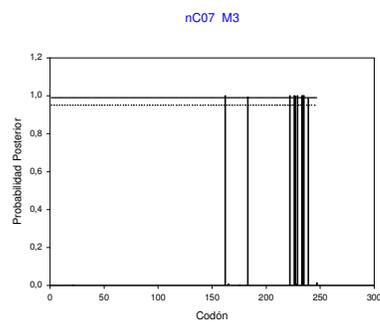
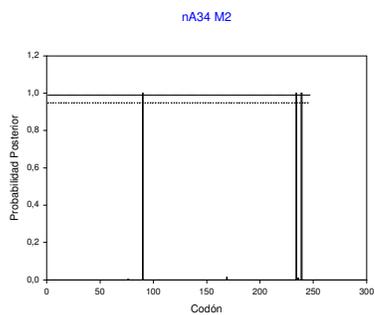
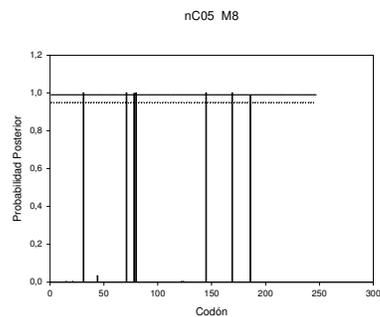
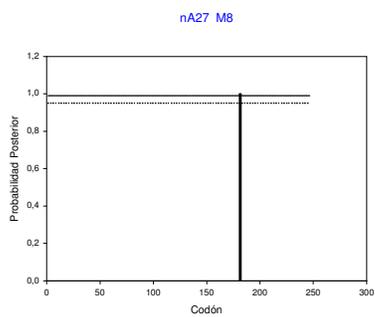
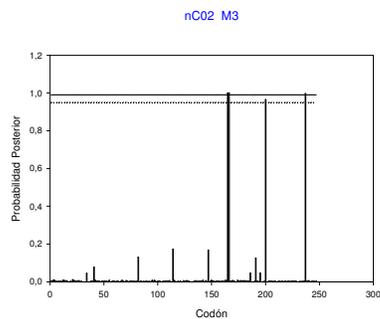
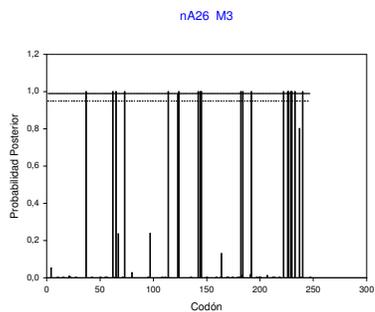


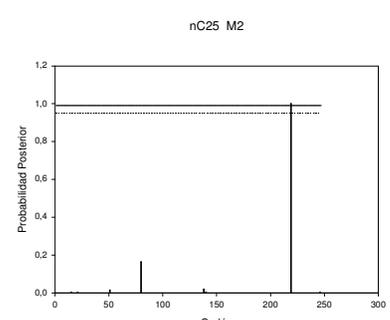
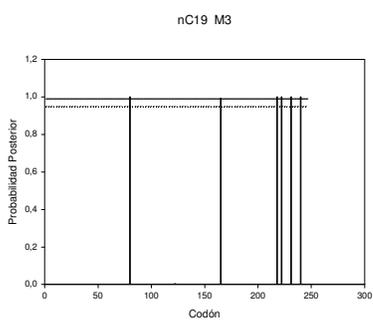
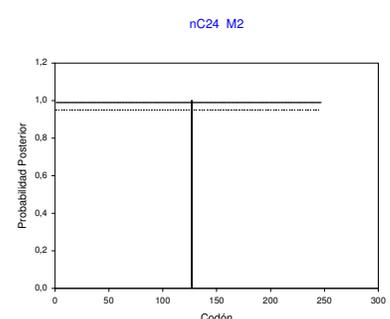
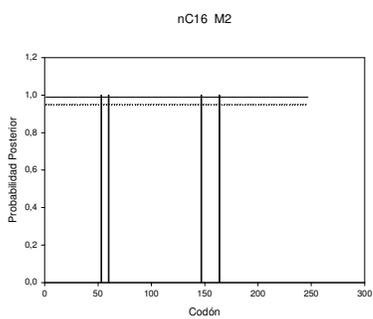
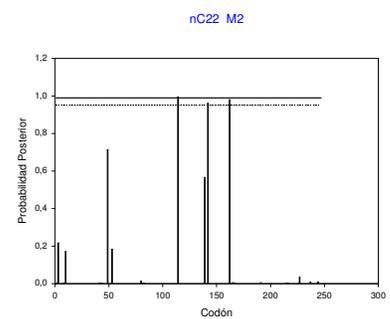
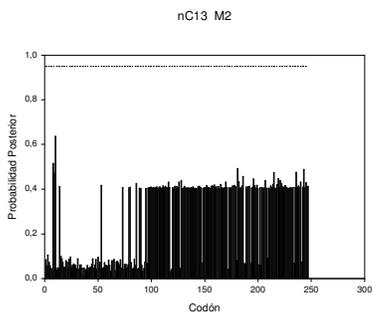
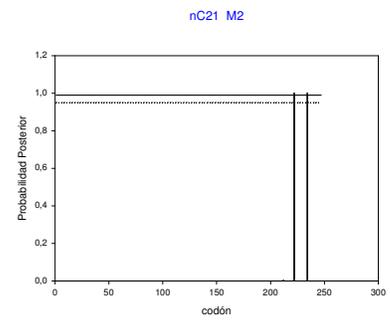
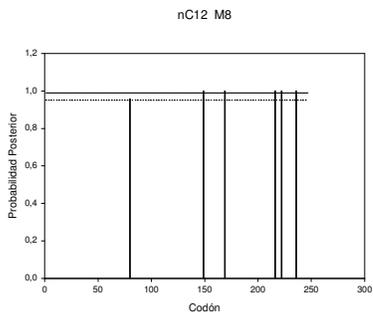
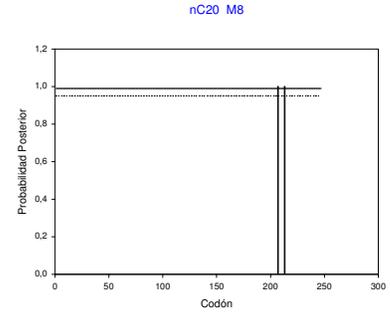
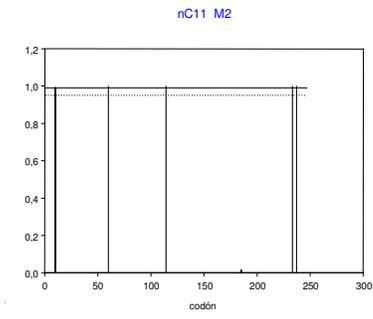
nA16 M8

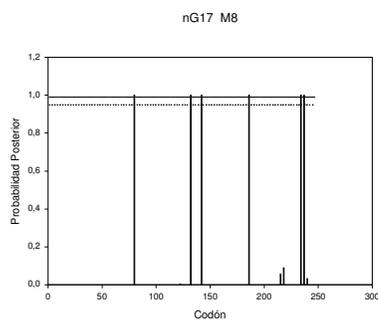
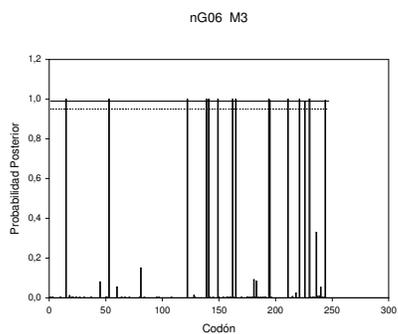
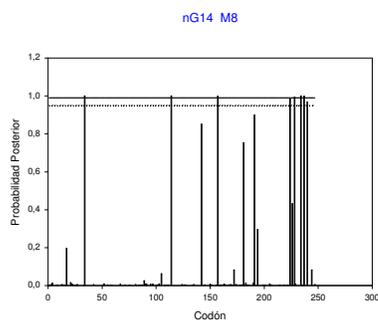
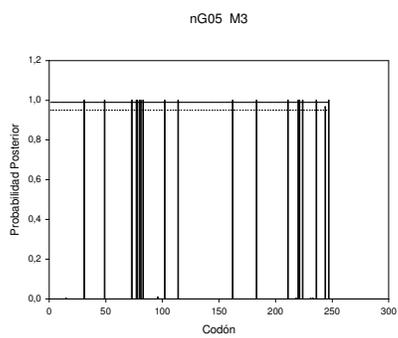
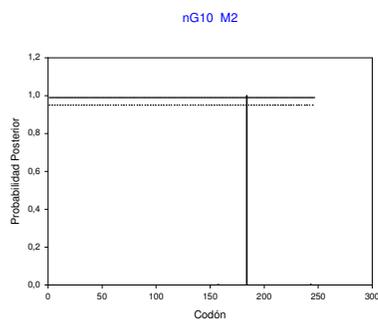
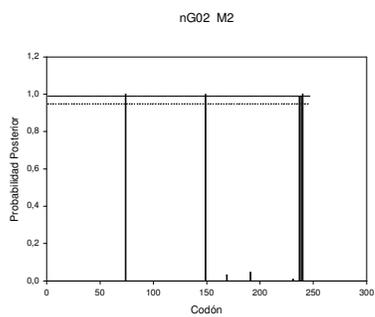
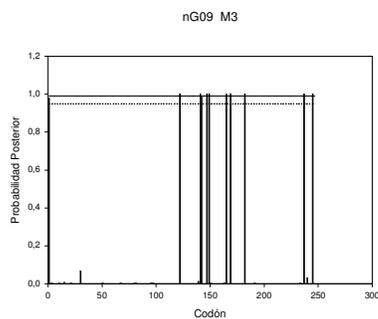
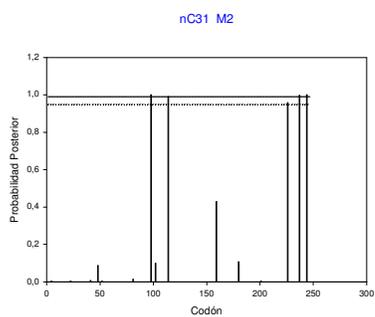
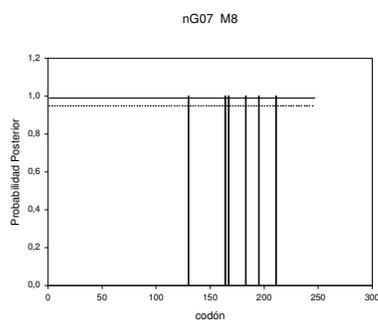
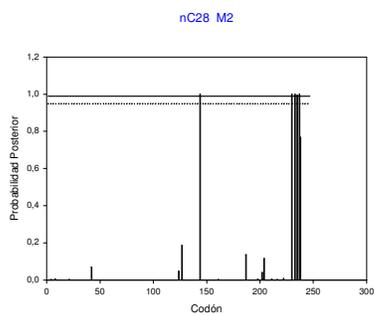


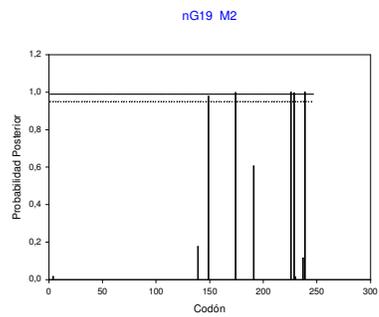
nA20 M2











IX. Apéndice

Apéndice

1. TIPOS DE HEPATITIS

Hepatitis Vírica Aguda. Inflamación difusa del hígado causada por virus hepatotrópicos específicos. En la mayoría de los pacientes la inflamación comienza repentinamente y dura unas pocas semanas. Una persona que padezca hepatitis vírica aguda puede convertirse en un portador crónico del virus. El portador no presenta síntomas, pero está aún infectado.

Síntomas. La hepatitis varía desde una afección menor análoga a un resfriado hasta una insuficiencia hepática fulminante y mortal, en función de la respuesta inmune del paciente y otros factores virus-huésped insuficientemente conocidos.

Datos de Laboratorio. Las espectaculares elevaciones de las aminotransferasas son el signo característico de la enfermedad. La ALT está característicamente más elevada que la AST.

Variantes de la hepatitis aguda:

Hepatitis anictérica: Afección leve sin ictericia semejante a un resfriado, puede ser la única manifestación clínica especialmente en la infección con VHC.

Hepatitis recidivante: Hepatitis recurrente durante la fase de recuperación.

Hepatitis fulminante: Síndrome raro. Se inicia con una encefalopatía hepática, tras unas horas puede aparecer el coma acompañado de edema cerebral. Tiene lugar una necrosis masiva del hígado y una disminución del tamaño del órgano. En adultos la supervivencia es infrecuente.

Hepatitis Crónica. Se define como una inflamación del hígado que dure más de 6 meses. La hepatitis crónica, aunque mucho menos frecuente que la hepatitis aguda, puede durar años e incluso décadas. En algunos casos, la continua inflamación afecta lentamente al hígado, produciendo en ocasiones cirrosis e insuficiencia hepática.

Síntomas. Alrededor de un tercio de los casos de hepatitis crónica se desarrolla después de una hepatitis vírica aguda. En el resto, se desarrolla gradualmente sin ninguna enfermedad previa evidente. Son muchas las personas que padecen hepatitis crónica sin presentar ningún síntoma, pero en las que los presentan, éstos a menudo consisten en una sensación de enfermedad, falta de apetito y cansancio y, en algunas ocasiones, algo de fiebre y un ligero malestar en la parte superior del abdomen. La ictericia puede aparecer o no.

Diagnóstico de laboratorio. Evidencia de una inflamación hepatocelular activa con elevaciones de las aminotransferasas y valores variables de la bilirrubina y fosfatasa alcalina.

2. RESPUESTA INMUNE

La *respuesta inmune inespecífica* es la primera barrera defensiva del organismo y no requiere sensibilización previa. Este tipo de respuesta es mediada por células con capacidad fagocítica y células asesinas naturales.

La *respuesta específica o adquirida* se desarrolla solo frente a la sustancia que indujo su iniciación y en ella participan prioritariamente los *linfocitos* y los elementos solubles liberadas por los mismos, *anticuerpos* y *linfocinas*. La respuesta inmune específica se considera que puede ser de dos tipos: *humoral* y *celular*. Aunque la separación de ambos tipos de respuesta es más de tipo didáctico que real, en general se considera que cuando el elemento efector final son las inmunoglobulinas formadas por los **linfocitos B** se trata de una *respuesta tipo humoral* (Figura 1), mientras que cuando participan los linfocitos T tanto colaboradores (**Th**) como citotóxicos (**Tc**), se trata de una *respuesta tipo celular*.

La respuesta inmune humoral. Es mediatizada por los linfocitos B, que reconocen al antígeno a través de las inmunoglobulinas de membrana. Sin embargo este estímulo no es suficiente para que se inicien los procesos de proliferación de estas células. Para ello es necesario que los linfocitos B, además del estímulo antigénico, reciban el de ciertas interleucinas. Las inmunoglobulinas, al detectar al antígeno y unirse a él, actúan como transductores de la información de la presencia de los mismos, que serán posteriormente destruidos por el mecanismo más idóneo, en el que colaborarán, además del propio anticuerpo, el *sistema del complemento*, *macrófagos*, *los polimorfonucleares o células K*. El término **complemento (C')** engloba una gran variedad de proteínas, que interactúan en un determinado orden. Cuando se produce la activación del C' se pone en marcha una serie de reacciones, en forma de "cascada", de tal forma que se van generando productos activos que además de influir en que la reacción prosiga tienen diferentes acciones biológicas importantes en la defensa del organismo.

La respuesta inmune de tipo celular. Actúa principalmente frente a bacterias y virus, así como evitando la aparición y desarrollo de células tumorales. En ella participan esencialmente los linfocitos T colaboradores y citotóxicos (Figura 1). Los linfocitos reconocen el antígeno mediante el *receptor T (TcR)* y lo hacen solo cuando el antígeno es degradado y

procesado en el interior de las *células presentadoras de antígeno (APC)* y sus determinantes antigénicos son expuestos en la superficie de estas células en el seno de una *molécula del complejo principal de histocompatibilidad*. Las *moléculas del Complejo Mayor de Histocompatibilidad (MHC)* son una serie de glicoproteínas presentes en las membranas de todas las células nucleadas, entre las que se encuentran las células inmunocompetentes. Estas moléculas son esencialmente de dos tipos o clases, *clase I* y *clase II*. La separación de las funciones de los linfocitos T colaboradores $CD4^+$ y $CD8^+$ viene dada por el origen de los antígenos que reconocen y, en último término, por donde han sido procesados por vía exógena en el sistema endosomal de las células presentadoras de antígeno y expresados en superficie por el producto de los genes MHC de clase II. Los linfocitos citolíticos $CD8^+$ reconocen a los antígenos que han sido procesados endógenamente en el citosol de la célula infectada y presentados en superficie por moléculas MHC de clase I, mientras que los linfocitos $CD4^+$ interactúan con el antígeno en el contexto de moléculas de clase II

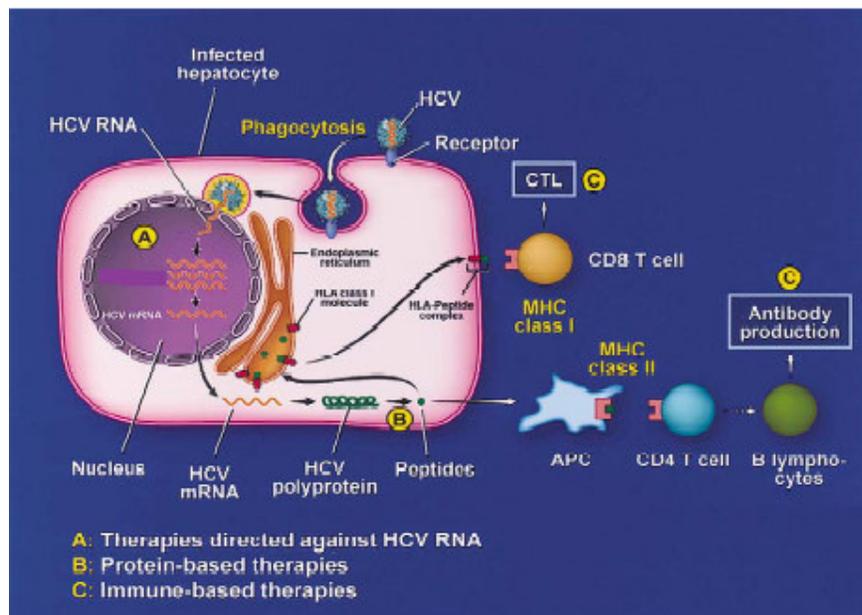


Figura 1. Respuesta humoral y celular desencadenada tras la entrada del VHC en un hepatocito (Zein et al. 2002).

Cuando tiene lugar el reconocimiento antigénico entre el TcR y la molécula MHC que porta el antígeno, se desencadena una cascada de reacciones bioquímicas en el citoplasma de la célula T, dando así lugar al proceso de activación, proliferación y diferenciación celular. Como consecuencia de estos eventos se producirá finalmente la activación de la transcripción de los

genes implicados en la síntesis de la proteína y factor implicado en una determinada función, tal como la síntesis de interleucina 2 u otros factores. La consecuencia final de este tipo de respuesta es la formación de células Th activas productoras de **interleucinas** y células citotóxicas (**CTL**) que poseen capacidad de lisar a las células que portan el antígeno que indujo su activación.

Citocinas. Las *citocinas* son una serie de sustancias producidas por células en respuesta a una gran variedad de estímulos y que son capaces de regular el funcionamiento de otras células. Estas sustancias pueden ser de diversos tipos entre los que se encuentran los denominados *factores de crecimiento*, polipéptidos que estimulan la proliferación de diferentes tipos celulares; las *linfocinas*, producidas por linfocitos (sobre todo Th) y de gran importancia en la regulación del sistema inmune. La familia, globalmente denominada *interferones*, fueron originalmente identificadas como agentes capaces de proteger a las células frente infecciones virales. Hoy se sabe que los interferones tienen otras muchas funciones, tales como actuar en los procesos de diferenciación y proliferación celular así como en la modulación del sistema inmunológico. Los interferones se clasifican en dos tipos: Tipo I que comprende el IFN α , sintetizado por leucocitos, y el IFN β , sintetizado por fibroblastos, ambos producidos en respuesta directa al virus; y el Tipo II, al que pertenece el IFN γ , sintetizado en respuesta al reconocimiento de células infectadas por los linfocitos T y células NK. La unión de ambos tipos de interferón a sus receptores celulares desencadena una serie de procesos que conlleva la activación de la transcripción de genes (normalmente expresados a bajos niveles) que impiden la replicación del virus o interfieren con procesos virales. A esta serie de pasos se le conoce como la ruta JACK/STAT (JACK: *Janus tyrosine kinases*, STAT: *signal transducer and activator of transcription*). Además los interferones pueden enlentecer el proceso de crecimiento celular, haciéndolas más susceptibles a la apoptosis y limitando así la infección viral.