



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

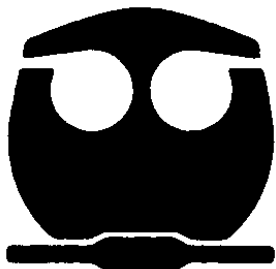
FACULTAD DE QUÍMICA

“Epidemiología Molecular del virus de la Inmunodeficiencia Humana: Estudio filogenético a partir de secuencias del gen env de virus circulantes en diferentes regiones geográficas de la República Mexicana”

TESIS

PARA OBTENER EL TÍTULO DE:
QUÍMICA FARMACÉUTICA BIÓLOGA

Presenta: Eréndira Rojas Ortega



México, D.F.

Año 2010



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

JURADO ASIGNADO:

PRESIDENTE: Profesora: Aida Navas Pérez

VOCAL: Profesor: Francisco Javier Plasencia de la Parra

SECRETARIO: Profesor: Santiago Ávila Ríos

1er. SUPLENTE: Profesora: Roció Gabriela Tirado Mendoza

2° SUPLENTE: Profesor: León Patricio Martínez Castilla

SITIO DONDE SE DESARROLLÓ EL TEMA: INSTITUTO NACIONAL DE ENFERMEDADES RESPIRATORIAS (INER), EN EL CENTRO DE INVESTIGACIÓN DE ENFERMEDADES INFECCIOSAS (CIENI).

ASESOR DEL TEMA: DR. SANTIAGO ÁVILA RÍOS

(Nombre Y Firma)

SUSTENTANTE : ERÉNDIRA ROJAS ORTEGA

(Nombre Y Firma)

Índice

1.0	Resumen:.....	5
2.0	Introducción:.....	7
2.1	Panorama general de la epidemia de VIH/SIDA en el mundo.....	7
2.2	Panorama general de la epidemia VIH/SIDA en México.....	10
2.3	Virología del VIH.....	16
2.3.1	Estructura del VIH.....	16
2.3.2	Organización del genoma viral.....	18
2.3.3	Análisis estructural de las proteínas codificadas por el gen <i>env</i>	22
2.3.4	Ciclo de replicación del VIH.....	25
2.4	Historia natural de la infección por VIH.....	28
2.5	Origen del VIH.....	32
2.6	Variabilidad del VIH.....	34
2.7	Clasificación del VIH.....	35
2.8	Epidemiología molecular del VIH en el mundo.....	37
2.9	Epidemiología molecular del VIH-1 en México.....	42
2.10	Filogenia.....	43
2.10.1	Alineamiento.....	47
2.10.2	Métodos filogenéticos.....	49
2.10.3	Modelos de evolución.....	54
2.10.4	Evaluación de la exactitud de un árbol.....	56
3.0	Planteamiento del problema.....	57
4.0	Hipótesis.....	58
5.0	Objetivos.....	59
5.1	Objetivo general.....	59
5.2	Objetivos particulares.....	59
6.0	Metodología.....	59
6.1	Selección de Pacientes.....	59
6.2	Aislamiento de ARN viral.....	60
6.3	RT-PCR anidada.....	61
6.4	Purificación de productos de PCR.....	64

6.5 Semicuantificación de productos de PCR	65
6.6 Secuenciación.....	66
6.5.1 PCR de secuenciación	67
6.5.2 Purificación de los productos de la reacción de secuenciación.....	68
6.5.3 Electroforesis capilar	69
6.6 Análisis filogenético	70
6.6.1 Alineamiento de secuencias	70
6.6.2 Construcción de árboles filogenéticos	70
6.7 Subtipificación	72
6.7.1 REGA.....	72
6.7.2 RIP	74
7.0 Resultados.....	82
7.1 Descripción de la cohorte en estudio.....	82
7.2 Subtipificación	89
7.3 Análisis Filogenético.....	96
7.3.1 Comparación entre arboles filogenéticos obtenidos a partir de los genes <i>pol</i> y <i>env</i>	96
7.3.2 Pruebas para determinar el modelo de nucleótidos mas optimo	100
7.3.3 Análisis filogenético de las secuencias del gen <i>env</i> y <i>pol</i> de VIH circulante en México mediante un modelo de distancias optimizado.....	107
8.0 Discusión de Resultados.....	118
9.0 Conclusiones.....	128
10.0 Referencias :	129

1.0 Resumen:

El conocimiento de la epidemiología molecular del VIH en México es notablemente escaso. Los trabajos publicados hasta la fecha muestran que el subtipo predominante en nuestro país es el subtipo B. A pesar de estos resultados los estudios publicados no cuentan con gran representatividad ya que se enfocan a ciertos grupos de riesgo y el tamaño de las cohortes es reducido.

En el Centro de Investigación de Enfermedades Infecciosas (CIENI) del Instituto Nacional de Enfermedades Respiratorias (INER) se llevo a cabo el estudio de epidemiología molecular del VIH más grande y representativo de México. En dicho estudio se analizaron más de 1500 secuencias del gen *pol* de virus provenientes de pacientes vírgenes a tratamiento antirretroviral de los diferentes estados de la República Mexicana. En dicho estudio se encontró que el 99.91 % de los virus analizados pertenecen al subtipo B y 0.09% a recombinantes BF (resultados no publicados). Al analizar las relaciones filogenéticas entre los diferentes virus mexicanos se encontró que no existía ningún agrupamiento por región geográfica, es decir, los virus se distribuían de forma homogénea en el árbol filogenético sin importar su origen. Estos resultados muestran una gran homogeneidad entre las secuencias de los virus circulantes en nuestro país. A pesar de esto, los resultados obtenidos en el análisis filogenético del gen *pol* no son completamente concluyentes ya que la alta homogeneidad se podría deber a la relativa baja variabilidad del gen *pol*.

En el presente estudio se analizaron 101 secuencias del gen *env* perteneciente a la región C2-V5. Se decidió analizar dicho gen ya que es el más variable del genoma del VIH. Los sueros de los cuales se extrajeron los RNAs virales se obtuvieron del banco de sueros de la cohorte del proyecto de Epidemiología Molecular del Centro de Investigación de Enfermedades Infecciosas (CIENI) del Instituto Nacional de Enfermedades Respiratorias (INER).

De los 101 pacientes participantes en el estudio, el 37% de los pacientes provenían del Centro, el 27% del Norte y el 36% del Sur. Los estados que participaron en el estudio fueron: Sonora, Baja California Norte, Nuevo León, Jalisco, D.F., Edo. de México, Morelos, Chiapas, Oaxaca, Veracruz y Guerrero.

Al realizar la subtipificación de las secuencias virales del gen *env* se encontró que el 100% de las secuencias pertenecían al subtipo B.

Con el fin de determinar si la variabilidad del gen viral analizado pudiera influenciar las conclusiones obtenidas a partir del análisis filogenético de VIH circulante mexicano, se realizó una comparación entre los árboles filogenéticos construidos con secuencias de los genes *env* y *pol* de los virus de individuos participantes en el presente estudio el método que se utilizó para construir los árboles fue el método Neighbor Joining, calculando las distancias evolutivas mediante el modelo de Kimura 2 parámetros usando como prueba bootstrap 1000 repeticiones.

Al analizar el árbol de *pol* encontramos que no existían agrupamientos por regiones geográficas mientras que en el árbol de *env* se logró distinguir algunos agrupamientos por región geográfica. A pesar de la existencia de estos agrupamientos no fue posible asegurar su existencia, ya que los valores de bootstrap fueron menores al 70%.

Encontramos que el modelo Kimura 2 parámetros no representó de forma adecuada las relaciones filogenéticas entre los virus mexicanos ya que presentaba relaciones filogenéticas erróneas en el árbol de *env*. Por tal motivo se probaron los diferentes modelos de nucleótidos incluidos en el programa MEGA 4.1 β , con lo cual se determinó que el modelo Maximum Composite Likelihood representaba con mayor exactitud las relaciones filogenéticas entre los virus mexicanos.

Al analizar las relaciones filogenéticas entre las secuencias del gen *env* de los virus mexicanos con el método Neighbor Joining y el modelo Maximum Composite likelihood, encontramos que en general no existió un agrupamiento por regiones geográficas, es decir los virus de las regiones Centro, Norte y Sur se distribuían de forma homogénea en el árbol filogenético. A pesar de esto se encontró un agrupamiento entre siete virus del Centro, los cuales no mantenían una línea de transmisión evidente. Sin embargo, no fue posible concluir la existencia de una diferenciación entre los virus del Centro, Norte y Sur, ya que el número de virus agrupados fue pequeño y por tanto no fue representativo.

2.0 Introducción:

2.1 Panorama general de la epidemia de VIH/SIDA en el mundo

La epidemia por el virus de inmunodeficiencia humana (VIH), causante del síndrome de inmunodeficiencia adquirida (SIDA), se ha convertido en un problema de salud mundial sin precedentes. Conocido desde hace 27 años, se estima que este virus ha causado alrededor de 25 millones de muertes en el mundo, generando cambios demográficos drásticos en los países mayormente afectados [1]. Sin embargo, los análisis epidemiológicos más recientes muestran resultados alentadores. En algunos países de América Latina y de África sub-Sahariana, el número anual de nuevas infecciones por VIH ha disminuido, al igual que el número de muertes causadas por SIDA. Estos avances se pueden atribuir a las campañas de información implementadas en varios países, al igual que a la ampliación del acceso al tratamiento antirretroviral. A pesar de esto es necesaria la ampliación de las políticas y programas de prevención y acceso al tratamiento antirretroviral a mayor número de regiones [1].

De acuerdo al reporte presentado por la Organización Mundial de la Salud (OMS) para el año 2007, existía un total de 32,2 millones de personas que vivían con VIH, de los cuales 30,8 millones eran adultos y 2,5 millones eran menores de 15 años. Del total de casos 15,4 millones eran mujeres. En este mismo año existieron 2,5 millones de nuevas infecciones por VIH, de las cuales 2,1 millones fueron en adultos y 420,000 en menores de 15 años. En este año se registraron 2,1 millones de defunciones, de las cuales 1,7 millones fueron en adultos y 330000 en menores de 15 años.[1] Sudáfrica continúa teniendo el mayor índice de casos de VIH y muertes por SIDA a nivel global. Para el 2007, el 35% de las personas infectadas por VIH y el 38% de las muertes causadas por SIDA se encontraban en esta área. De acuerdo al reporte del 2007, en África sub-sahariana se encontraban concentrados el 67% de personas que vivían con VIH (**ver figura 2.1.**)[1].

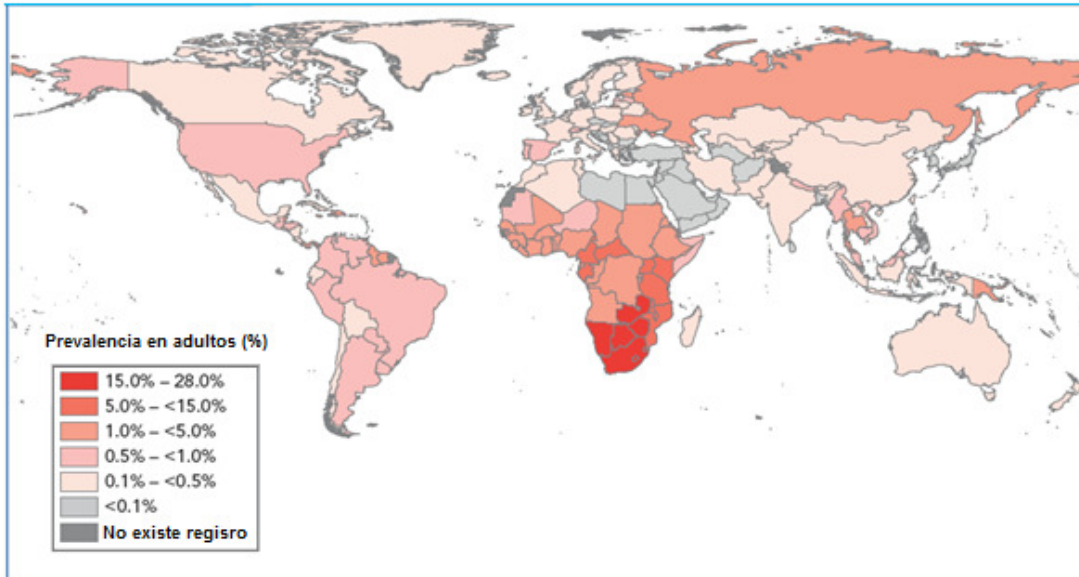


Figura 2.1. Panorama global de la epidemia de VIH, 2007[1].

La epidemia del VIH en el mundo sigue dos patrones:

- **Epidemias generalizadas en poblaciones generales de muchos países de África subsahariana**, en especial en la parte meridional del continente(ver figura 2.2).
- **Epidemias en el resto del mundo** que se concentran principalmente entre las poblaciones de mayor riesgo, como son los hombres que tienen sexo con hombres (HSH), usuarios de drogas intravenosas, profesionales del sexo y sus parejas sexuales [2].

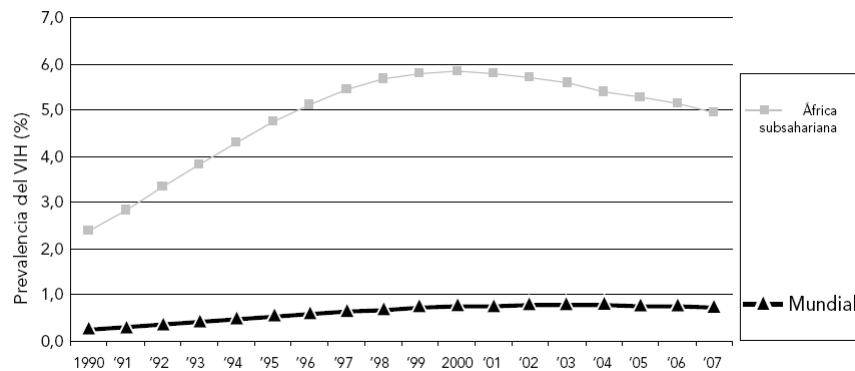


Figura 2.2. Prevalencia estimada del VIH en adultos (15-49)(%) en todo el mundo y en África Subsahariana ,1990-2007.

El 50% de personas que viven infectadas por VIH en el mundo son mujeres. La proporción de mujeres infectadas por VIH en los últimos 10 años se encuentra estable en la mayor parte del mundo. Sin embargo, en algunas regiones el número de mujeres infectadas ha mostrado un aumento debido principalmente a cuestiones asociadas a desigualdad de género, ya que las normas sociales restringen el acceso a la información básica de sexualidad y salud reproductiva a las mujeres, lo cual las coloca en una situación de vulnerabilidad.

(ver figura 2.3) [1].

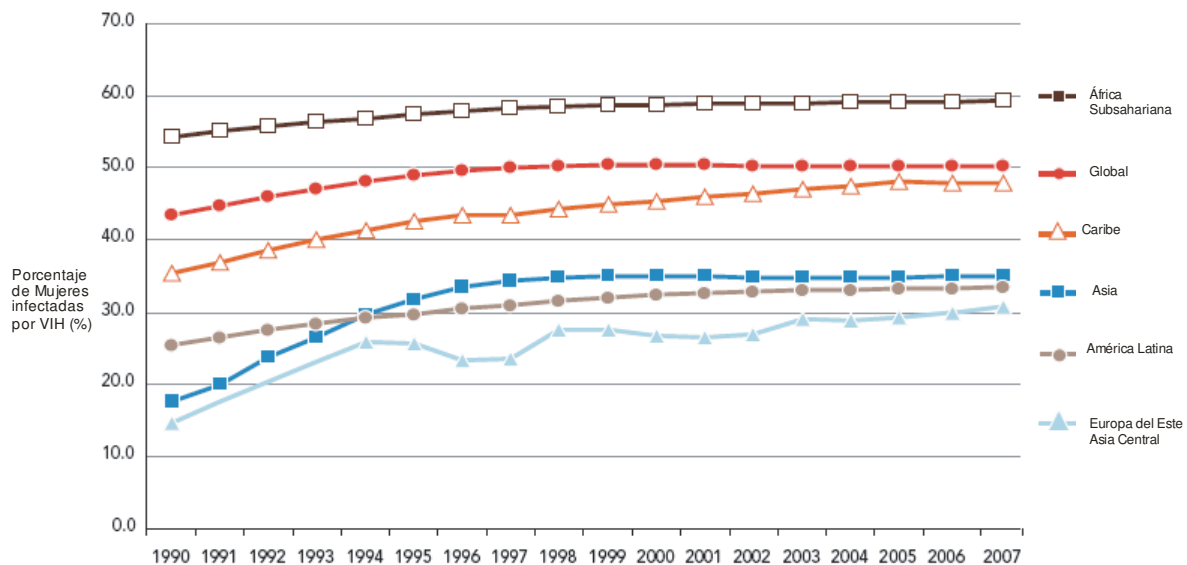


Figura 2.3 Porcentaje de individuos infectados con VIH mayores de 15 años que son mujeres (1990-2007) [1].

Entre el año 2001 y 2007, se produjeron aumentos similares en las cifras totales de nuevas infecciones en varones y mujeres, si bien la razón varón/mujer se mantuvo estable en todo el mundo. El número estimado de 15,4 millones [13,9-16,6 millones] de mujeres que viven con el VIH en 2007 representa 1,6 millones más que los 13,8 millones [12,7-15,2 millones] en 2001. En África Subsahariana, el 61% de los adultos infectados por VIH son mujeres, mientras que en el Caribe son el 43% [1].

Los porcentajes de mujeres que viven con el VIH en América Latina, Asia y Europa Oriental están aumentando lentamente, ya que el VIH se transmite a las parejas femeninas de varones que probablemente lo hayan contraído a través del consumo de drogas inyectables, las relaciones sexuales remuneradas o entre varones sin protección. En Europa Oriental y Asia Central, se estimó que las mujeres representaban el 26% de los adultos con el VIH en 2007 (en comparación con el 23% en 2001), mientras que en Asia este porcentaje alcanzó el 29% en 2007 (en comparación con el 26% en 2001)[2].

La vía más común de transmisión del VIH es por medio de vía sexual, seguida del uso de drogas intravenosas y finalmente la vía perinatal. El riesgo de contraer VIH por medio de transfusiones sanguíneas y sus derivados ha disminuido considerablemente ya que desde los años 1985-1986 se introdujeron pruebas rutinarias para la detección de VIH en la mayoría de los países [2].

2.2 Panorama general de la epidemia VIH/SIDA en México

De acuerdo al informe dado por el Centro Nacional para la Prevención y el Control del VIH/SIDA se han diagnosticado a partir de 1983 hasta el año 2008 un total de 26,200 personas infectadas por VIH en México. De acuerdo al Registro Nominal de Seropositivos a VIH (datos al 14 de noviembre del 2008), las entidades que cuentan con una mayor prevalencia de personas infectadas por VIH son Yucatán, Distrito Federal, Quintana Roo , Veracruz y Chihuahua con 0,05 casos por cada 100 habitantes (**ver tabla 2.1**). La entidad que cuenta con mayor número de infectados del total registrados a nivel nacional es el Distrito Federal (18,3%). La tabla 2.2.1 muestra la prevalencia del virus en cada entidad federativa[3].

Tabla 2.1. Prevalencia de VIH en las diferentes entidades de la República Mexicana*

Lugar por prevalencia registrada	Entidad Federativa	Personas con VIH registradas	Prevalencia de VIH registrada	%
1	Yucatán	1,037	0.05	4.0
2	Distrito Federal	4,786	0.05	18.3
3	Quintana Roo	654	0.05	2.5
4	Veracruz	3,499	0.05	13.4
5	Chihuahua	1,612	0.05	6.2
6	Baja California Sur	224	0.04	0.9
7	Baja California	1,211	0.04	4.6
8	Tamaulipas	1,027	0.03	3.9
9	Tabasco	615	0.03	2.3
10	Sonora	700	0.03	2.7
11	Sinaloa	697	0.03	2.7
12	Nuevo León	1,098	0.02	4.2
13	Chiapas	1,040	0.02	4.0
14	Oaxaca	800	0.02	3.1
15	Michoacán	752	0.02	2.9
16	Guerrero	576	0.02	2.2
17	Campeche	141	0.02	0.5
18	Puebla	917	0.02	3.5
19	Morelos	264	0.02	1.0
20	Nayarit	140	0.01	0.5
21	Coahuila	343	0.01	1.3
22	Colima	74	0.01	0.3
23	San Luis Potosí	297	0.01	1.1
24	Zacatecas	163	0.01	0.6
25	Jalisco	814	0.01	3.1
26	Durango	165	0.01	0.6
27	Aguascalientes	119	0.01	0.5
28	Querétaro	177	0.01	0.7
29	México	1,502	0.01	5.7
30	Guanajuato	416	0.01	1.6
31	Hidalgo	194	0.01	0.7
32	Tlaxcala	86	0.01	0.3
	Extranjeros	60		0.2
	Se desconoce	0		0.0
	Nacional *	26,200	0.02	100.0

Entidades con mayor prevalencia de personas infectadas con VIH

D.F. es la entidad con mayor números de casos de personas infectadas por VIH

*Datos al 14 de noviembre del 2008 del Registro Nominal de Seropositivos a VIH La prevalencia es por cada 100 habitantes[3].

De acuerdo al Registro Nominal de Seropositivos a VIH, del total de los casos registrados en México el 71% son hombres mientras que el 29% son mujeres. El grupo de edad que es más afectado son las personas entre 15 y 29 años (**ver figura 2.4**)[3].

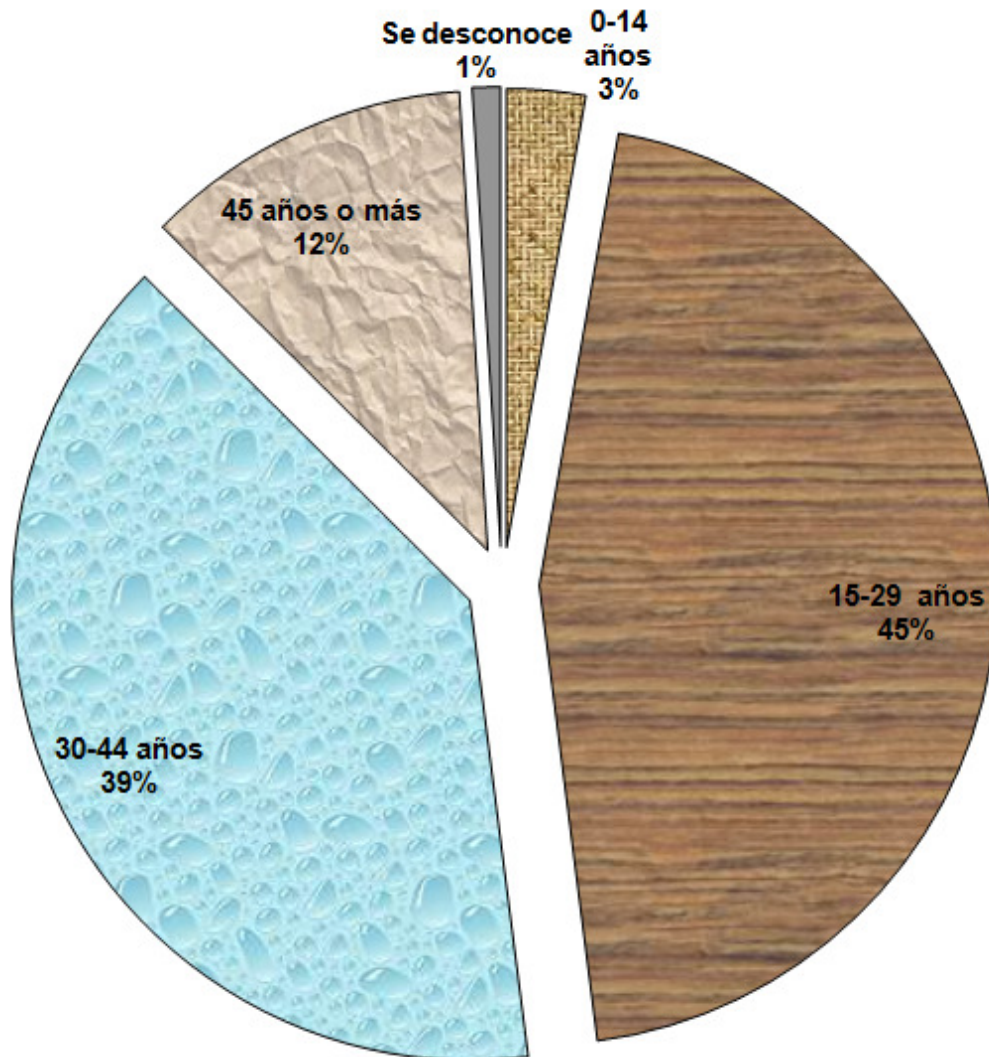
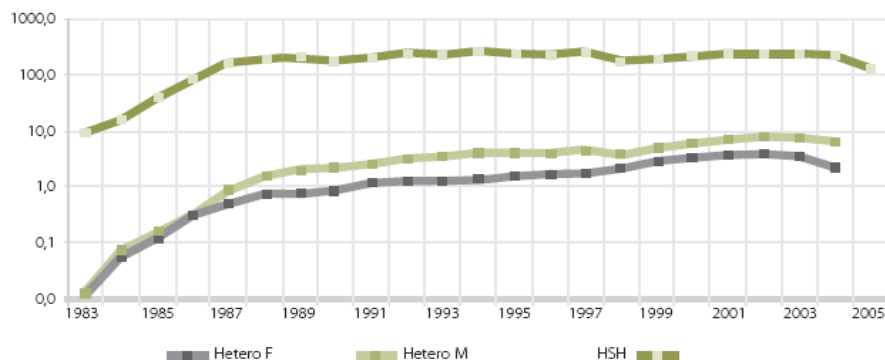


Figura2.4. Distribución por edades de las personas afectadas por VIH. El grupo de edad mayormente afectado es el de 15-29 años siendo este rango de edad el más prevalente en las personas infectadas con VIH [2].

Como se mencionó anteriormente, la epidemia de VIH/SIDA en México está centrada en hombres, dentro de los cuales los hombres que tienen sexo con otros

hombres (HSH) son el grupo más afectado de la población (**ver figura 2.5**). Sin embargo, investigaciones recientes demuestran que la infección ha comenzado a crecer más rápidamente en otros sectores de la población. Esto se debe a que el factor de riesgo no se encuentra en las preferencias sexuales, sino en conductas riesgosas como son la no utilización de condón, tener múltiples parejas sexuales y/o el uso de drogas intravenosas [4].



Tasas por 100.000 habitantes.

Fuentes: Censida. Cálculos propios utilizando datos al 15 de noviembre de 2005 de la Secretaría de Salud – Dirección General de Estadísticas – Registro Nacional de Casos de SIDA, Conapo, Proyecciones de Población 1980-2000 y 2000-2050.

Nota: Aún no se completan todos los casos diagnosticados entre 2001 y 2005 debido al retraso en la notificación de los mismos.

Figura 2.5. Tasa de afectación de casos acumulados de sida en adultos por categoría de transmisión sexual, 1983-2005. El grupo más afectado es el de hombres que tienen sexo con otros hombres (HSH) en comparación con los grupos de transmisión heterosexual en hombres (Hetero M) y mujeres (Hetero F). Tomado de [4].

De acuerdo al registro nacional de SIDA al día 14 de noviembre del año 2008 el 95,8% y 92,5% de casos de VIH en hombres (**ver figura 2.6**) y mujeres (**ver figura 2.7**) de 15 a 29 años respectivamente fueron contagiados por transmisión sexual, mientras que el 3,7% y 7,5% fueron contagiados por vía sanguínea en cada uno de los grupos. Para el 0,5% de los casos en hombres se desconoce la vía de transmisión (**ver tabla 2.2**)[3].

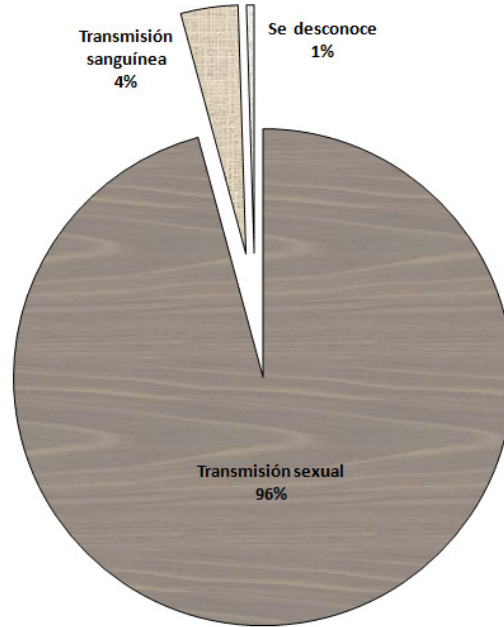


Figura 2.6. Vías de transmisión de la infección por VIH en hombres entre 15 y 29 años de edad en México. La vía de transmisión más frecuente entre hombres de 15 a 29 años es la sexual. Tomado de [3].



Figura 2.7 Vías de transmisión de la infección por VIH en mujeres entre 15 y 29 años de edad en México. La vía de transmisión más frecuente entre mujeres de 15 a 29 años es la vía sexual. Tomado de [3]

En la Tabla 2.2 se muestra la importancia de la vía de transmisión sexual dentro del marco epidemiológico del VIH/SIDA en México.

Tabla 2.2 Porcentaje de casos acumulados de SIDA por Categoría de Transmisión totales para la República mexicana al 15 de noviembre del 2007.*

VÍA SEXUAL	%	VÍA SANGUÍNEA	%	PERINATAL	%	COMBINADA	%	SE DESCONOCE (1)	%
HSH	45.9	Transfusión	3.0	Perinatal	2.3	UDI-HSH	0.3	Se Desconoce	35.0
Heterosexual	46.6	Hemofílicos	0.4						
		Donadores	0.5						
		UDI	1.1						
		Exp. Ocupacional	<0.1						
Total	92.5	Total	4.9	Total	2.3	Total	0.3		35.0

Fuente: Censida, con base en DGE- Registro Nacional de Casos de SIDA. Datos al 15 de noviembre del 2007.
 Nota: La categoría "Se Desconoce" se excluye del cálculo de las restantes categorías; la presentación se realiza a efectos de apreciar su magnitud.
 *Tomado de [4].

Desde la aparición del VIH/SIDA se comenzó a sospechar de la existencia de ciertos “grupos de riesgo”, idea que ha evolucionado hasta considerarse grupos de “vulnerabilidad” y “riesgo asociado”. El desarrollo conceptual de “grupo de vulnerabilidad” y “riesgo asociado”, responden a la necesidad de entender el fenómeno de la epidemia de VIH en el mundo. Los grupos vulnerables son grupos de riesgo socialmente desfavorecidos, lo cual aumenta la probabilidad de contraer VIH. En el caso de los grupos con riesgo asociado, son aquellos grupos que deliberadamente asumen el riesgo de infectarse con VIH[4].

De acuerdo a esta idea de *grupos vulnerables* se han distinguido los siguientes:

Migrantes: Se ha observado en estudios cuantitativos que hombres y mujeres que cruzan la frontera tienden a alterar sus hábitos sexuales, incrementando el número de parejas sexuales, además de que muestran un incremento significativo en el uso de drogas intravenosas, lo cual aumenta considerablemente el riesgo de contraer VIH [4].

Mujeres: La desigualdad social es uno de los factores más importantes en la transmisión de la infección por VIH. Esta desigualdad de género causa que las mujeres en ocasiones no cuenten con información básica de sexualidad y salud reproductiva. Existen también factores biológicos que favorecen la transmisión del

virus en determinados grupos. Al mantener relaciones heterosexuales, la mujer tiene de dos a cuatro veces mayor probabilidad de contraer VIH ya que la carga viral en el semen es mayor a la encontrada en los fluidos vaginales.[4]

Usuarios de drogas intravenosas (UDI): El crecimiento de la transmisión por esta vía fue lento hasta el año 2006, cuando los casos acumulados de SIDA asociados a esta ruta de transmisión eran apenas el 1.2% en el país, incluyéndose algunos casos de HSH que utilizaban drogas intravenosas [4].

Los grupos con riesgo asociado son:

Hombres que tienen sexo con otros Hombres. Este grupo es el más afectado de la población ya que un 45.9% de los casos de VIH en hombres, pertenece a hombres que tienen sexo con otros hombres. La razón por la cual tienen un riesgo asociado se debe a que en su mayoría sostiene relaciones sexuales a través de coito anal, esto se debe a que se lastima la mucosa réctalo debido a que es fina y puede desgarrarse fácilmente, con lo que las pequeñas lesiones en la mucosa son suficientes para permitir la entrada fácil del VIH. Una forma de disminuir el riesgo de contagio es el uso adecuado del condón. Sin embargo en muchos de los casos el contagio de VIH está asociado al uso de drogas intravenosas, a tener múltiples parejas sexuales y a la vulnerabilidad social [5].

Los(as) sexoservidores(as) El riesgo crece al aumentar el número de parejas sexuales. Muchos de los(as) sexoservidores(as) no utilizan condón, lo cual incrementa la probabilidad de que contraigan el virus[4].

2.3 Virología del VIH

2.3.1 Estructura del VIH

El virus de la inmunodeficiencia humana es un virus perteneciente al género de los lentivirus, dentro de la familia de los retrovirus. Estos virus se caracterizan por presentar enfermedad de curso crónico al igual que un periodo largo de latencia. Ejemplos de estos virus son los virus de inmunodeficiencia de los simios (SIV), los virus de inmunodeficiencia de los felinos (FIV) y el virus visna de las ovejas[6].

Existen dos tipos de VIH, VIH-1 y VIH-2, los cuales presentan un parecido sorprendente si se analizan con microscopía electrónica. A pesar de esto difieren en el peso molecular de las proteínas que los conforman y presentan diferencias en sus genes accesorios. A pesar de las diferencias existentes entre el VIH-1 y el VIH-2, ambos producen inmunodeficiencias similares causadas por la replicación del virus en las células T CD4⁺ del hospedero [6].

Debido a que en la mayor parte de los casos de infección por VIH se da por VIH-1, únicamente nos referimos a la estructura de este.

El VIH-1 es un virus de genoma de RNA con cubierta lipídica, con un diámetro de 100 nm. Cada partícula viral contiene en su envoltura 72 complejos de glicoproteínas, los cuales están integrados por trímeros de la glicoproteína externa gp120 y la proteína transmembranal gp41. La proteína de matriz p17 se encuentra anclada en el interior de la membrana lipoprotéica y la proteína p24 es aquella que forma la cápside viral, la cual contiene 2 copias del genoma viral de RNA. El RNA viral se encuentra en un complejo proteína-acido nucleico con p7, que funciona como nucleocápside. La partícula viral cuenta además con enzimas codificadas por el genoma viral que son fundamentales para el ciclo de replicación del VIH, las cuales incluyen a la transcriptasa reversa (p51/p66), integrasa (p32) y proteasa (p11) (**ver figura 2.7**)[6].

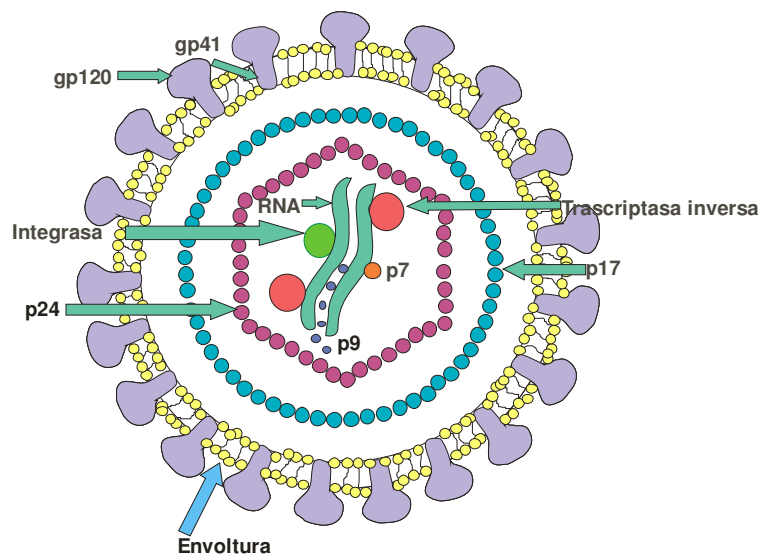


Figura 2.7. Estructura general del VIH. El VIH es un virus envuelto con genoma de RNA. Las proteínas gp120 y gp41 forman parte de la envoltura viral. La proteína p24 constituye la cápside viral, la proteína p7 constituye la nucleocápside y p17 se encuentra asociada a la membrana

constituyendo una matriz proteica. El virión contiene además enzimas codificadas por el genoma viral que son fundamentales para el ciclo de replicación viral, incluyendo la transcriptasa inversa, integrasa y proteasa.

2.3.2 Organización del genoma viral

La mayoría de los retrovirus incluyen tres genes estructurales principales: *gag* (antígeno específico de grupo), *pol* (polimerasa) y *env* (envoltura). Las repeticiones terminales largas (LTR, por sus siglas en inglés “*long terminal repeat*”) se encuentran en ambos extremos del genoma viral, y se conectan con el DNA de la célula infectada después de la integración (este fragmento no codifica para ninguna proteína). La función que cumplen los LTR’s es regulatoria ya que promueve la replicación y propagación del VIH cuando interactúa con las proteínas virales Tat y Vpr. Recientemente se ha descubierto que contiene un motivo llamado caja E. La caja E es un mediador importante de la regulación de la transcripción proviral, ya que la proteína bHLH interfiere con el complejo transcripcional [7].

En medio de los LTR’s se encuentran los genes *gag* que codifica para la cápside y nucleocápside, *pol* que codifica para las enzimas virales y *env* que codifica para las glicoproteínas de la envoltura. Por tanto, la estructura clásica de un retrovirus es 5’LTR-*gag-pol-env*-LTR3’. Además de los genes estructurales, el VIH contiene 6 genes adicionales (*vif*, *vpu*, *vpr*, *tat*, *rev* y *nef*), los cuales se encuentran sobrelapadas con otros marcos abiertos de lectura (ORF, por sus siglas en inglés)[5]. **(ver figura 2.8)**.

En el pasado *nef*, *vif*, *vpr* y *vpu* se clasificaron como genes accesorios ya que se creía que su función no era fundamental para el proceso de replicación. No obstante, ahora se sabe que sus funciones son indispensables para la replicación viral. Los genes *tat* y *rev* se producen en la parte inicial de la replicación viral y son importantes reguladores del ciclo de replicación viral [6].

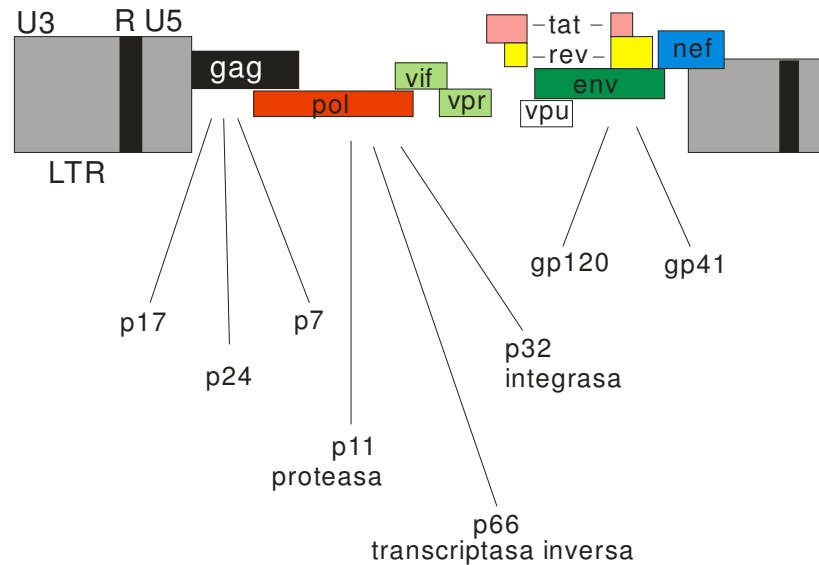


Figura 2.8. Organización del genoma del VIH. Se pueden observar los 3 genes estructurales principales *gag*, *pol* y *env*, y sus 6 genes accesorios/reguladores *vif*, *vpu*, *vpr*, *tat* y *nef*.

Al igual que en otros retrovirus competentes para la replicación, los 3 productos génicos principales de la traducción en el VIH son proteínas estructurales o enzimas que se requieren para su replicación. Estos productos en un principio son sintetizados como poliproteínas precursoras, las cuales son procesadas para su maduración por proteasas celulares o virales. El producto de *gag*, un precursor de 55KDa (Pr55^{gag}), tiene como productos finales la proteína de matriz (p17), la nucleocápside (p7) y la cápside (p24). La autocatálisis de la poliproteína Gag-Pol de 160 kDa da origen a la proteasa, a la retrotranscriptasa heterodimérica y a la integrasa. En el caso del precursor Env (*gp160*) la digestión proteolítica por enzimas celulares da como resultado la formación de la glicoproteína de superficie gp120 y la glicoproteína transmembranal gp41 [8].

Tat y Rev son proteínas reguladoras que se acumulan en el núcleo y se unen a regiones definidas del RNA viral. Tat y Rev estimulan la transcripción del DNA proviral de VIH-1 a RNA, promueven el alargamiento del RNA, incrementan el transporte de RNA de VIH del núcleo al citoplasma y son esenciales para la traducción. Rev es un factor de exportación nuclear que es importante para el cambio de la expresión temprana de proteínas reguladoras a las proteínas estructurales que se sintetizan posteriormente [6].

La proteína Tat (proteína trans-activadora), es una de las proteínas esenciales de la regulación de la expresión de los genes del VIH-1. Existen dos tipos de proteínas Tat, Exon Tat-1 (se encuentra en menor proporción) de 72 aminoácidos y Exon Tat-2 (es la forma más común) de 86 aminoácidos. La proteína Tat se encuentra localizada primordialmente en el núcleo. Esta proteína se une al TAR (elemento de respuesta de transactivación) de los RNAs y activa la iniciación de la transcripción y elongación de los promotores LTRs, esta proteína previene la poliadenilación del LTR 5' lo cual evita la terminación prematura del transcrito [9].

La proteína Rev es un factor regulatorio de la expresión del VIH-1, es una fosfoproteína de 19kD la cual se localiza en el núcleo. El VIH-1 ya integrado al genoma (estado proviral) genera tres tipos de mRNA según su procesamiento: de 9kb (no procesados) que codifican para las proteínas Gag y Pol; 4kb (semiprocados) que codifican para Env, Vif y Vpu y de 2kb (completamente procesados), que codifican para las proteínas Tat, Rev y Nef. Estos transcritos no procesados o semiprocados son exportados al citoplasma por medio de la proteína Rev. Esta proteína tiene la capacidad de interactuar con la estructura REE (Elemento de respuesta a Rev), presente en la región *env* de los transcritos virales y los exporta al citoplasma en asociación con ciertas proteínas celulares [10].

El producto de *nef* puede disminuir la expresión de CD4⁺ y de las moléculas de clase I del antígeno leucocitario humano (HLA) en la superficie de las células infectadas por VIH-1. Esto representa un mecanismo importante en el escape del virus de las respuestas de células T CD8⁺. Nef también está implicado en defectos en la activación de células, ya que se une a varias proteínas involucradas en las vías intracelulares de transducción de señales [6].

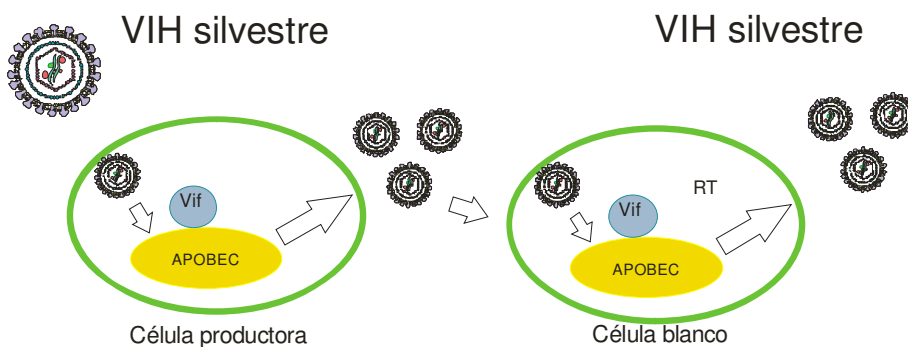
La proteína Vpr es una proteína que consta de 96 aminoácidos, que es importada en el virión. Esta proteína interactúa con la región p6 Gag del precursor Pr55 Gag. Esta proteína se encuentra en el núcleo. La función principal de Vpr es la inhibición de la expansión clonal al permitir la detención de la célula en la fase G2 del ciclo celular, etapa donde el LTR del VIH-1 es más activo [10,6].

Otra de las funciones que se ha encontrado para Vpr es su participación en la importación hacia el núcleo del Complejo de Preintegración del VIH. [10]

El gen *vpu* es importante para el proceso de gemación del virus puesto que mutaciones en este gen se asocian a la persistencia de las partículas virales en la superficie de la célula hospedera. Vpu también está involucrado en el proceso de degradación de los complejos CD4-gp160 en el retículo endoplásmico y por tanto permite el reciclaje de gp160 para la formación de viriones nuevos[6].

Estudios recientes han demostrado que VIH-1 que carece de Vif no se replica en células T CD4+ ni en macrófagos; ya que estos virus tienen la capacidad de entrar en la célula blanco e iniciar la retro transcripción, pero la síntesis del DNA proviral permanece incompleta. La fusión *in vitro* de las células permisivas y las no permisivas, generan un fenotipo no permisivo, lo cual sugiere que la replicación del virus depende de la presencia o ausencia de un inhibidor celular. Este factor inhibitorio endógeno se identificó como APOBEC3G (Apolipoproteína B de la enzima editora de mRNA, polipeptido catalítico similar a 3G)(**ver figura2.9**). APOBEC3G pertenece a la familia de las enzimas intracelulares que desaminan a la citosina, convirtiéndola en uracilo en el mRNA o en el DNA, lo que tiene como resultado la acumulación de mutaciones G por A que causan la degradación del DNA viral. El gen *vif* bloquea la actividad inhibitoria de APOBEC3G mediante la formación de un complejo APOBEC3G-Vif-ubiquitina, que es susceptible de ser degradado por el proteosoma de la célula infectada (**Figura 2.9**) .[6].

a)



b)

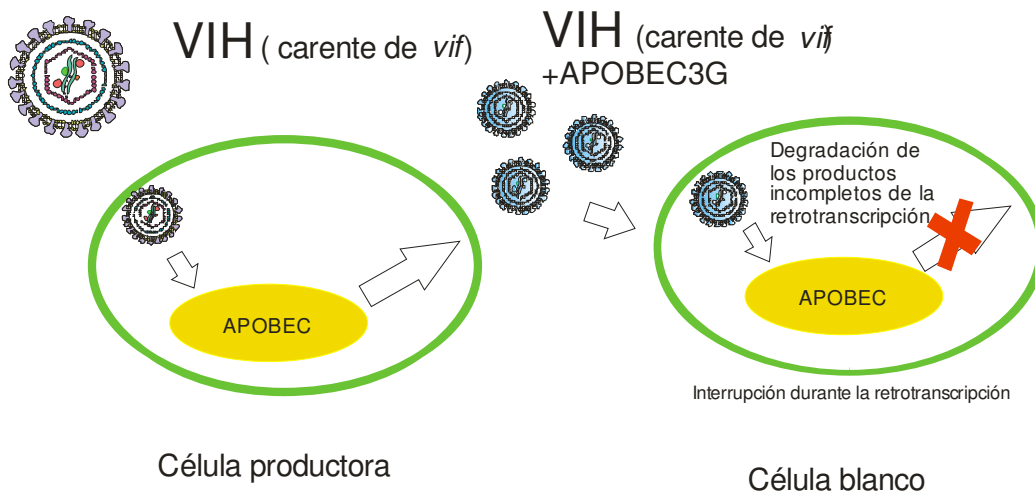


Figura 2.9. Papel de APOBEC3G como inhibidor retroviral en el ciclo de replicación del VIH.

La proteína viral Vif se une a APOBEC3G e impide su acción en los virus nuevos que se forman (figura a). Los virus en los que se suprime *vif* no pueden inhibir la APOBEC3G intracelular, que posteriormente se incorpora en los virus nuevos e interfiere con la retrotranscripción en la célula blanco como son los linfocitos y los macrófagos [6].

2.3.3 Análisis estructural de las proteínas codificadas por el gen *env*

La traducción del gen *env* da como primer producto el precursor gp160, el cual al ser digerido por enzimas proteolíticas celulares origina la producción de gp120 y gp41. Estas proteínas son glicosiladas por enzimas que se encuentran en el lumen del retículo endoplasmático rugoso (RER), las cuales agregan oligosacáridos ricos en manosas en los residuos de asparagina en motivos Asn-X-Ser o Asn-X-Thr. El VIH tiene 30 sitios de adición de oligosacáridos, los cuales se encuentran en su mayoría en la región gp120. El papel que cumplen los oligosacáridos es proveer regiones hidrofílicas, las cuales confieren estabilidad a la proteína plegada. Si se hace un cultivo de VIH en presencia de *tunicamicina*, fármaco inhibidor de la glicosilación, resulta en la inestabilidad de la proteína Gp120, lo cual causa que no se una eficazmente con la proteína CD4 [11].

Un análisis de las secuencias del gen *env* de diferentes aislados de VIH-1 demuestra la existencia de una gran variabilidad en este gen. Se han logrado definir 5 regiones variables y 4 regiones constantes en la región codificante de

gp120, las regiones variables se designan V1-V5 y las regiones constantes se designan C1-C4, estas regiones se alternan en la secuencia proteica (**Figura 2.10a**). Este patrón se encuentra presente también en el VIH-2 y VIS. Existen 18 cisteínas conservadas que forman 9 enlaces disulfuros. Estos enlaces segregan las regiones características de gp120. Una de estas regiones contiene el asa V1 y V2, mientras que el asa V3 y V4 también se definen por la formación de enlaces disulfuro (**Figura 2.10b**)[11].

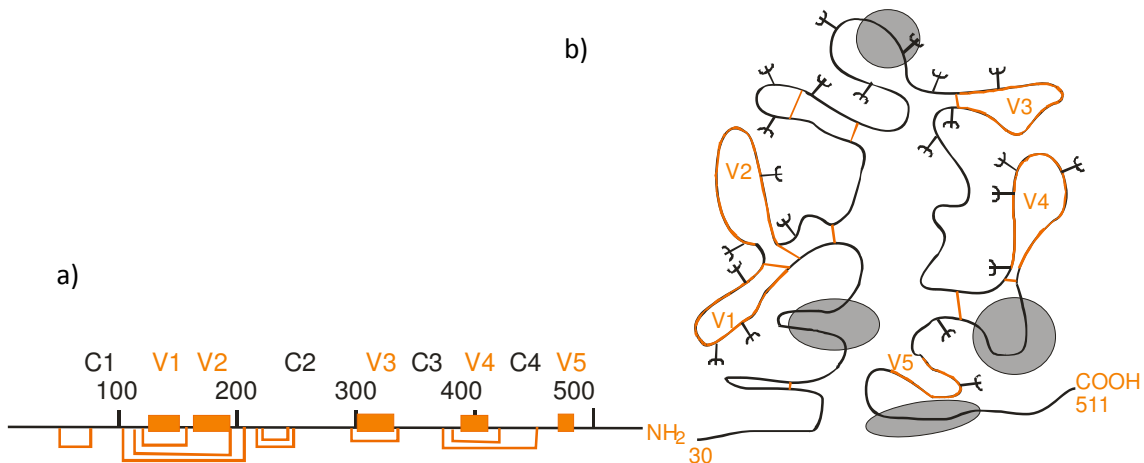


Figura 2.10. Representación esquemática de la proteína gp120 de VIH-1. (a) Representación lineal de la secuencia de aminoácidos de gp120. Se muestran los dominios constantes en negro y los dominios variables como rectángulos naranjas. Las líneas naranjas en la parte inferior representan enlaces disulfuros entre las cisteínas conservadas. (b) Representación plegada de la glicoproteína gp120 del VIH-1. Se muestran los dominios constantes en negro y los dominios variables en naranja. Los enlaces disulfuros se muestran como líneas delgadas color naranja, los oligosacáridos se muestran como estructuras ramificadas. Las zonas sombreadas son aquellas que tienen contacto con la molécula CD4. Tomado de [6].

La proteína gp41, es una proteína transmembranal, que se asocia en forma no covalente con la proteína gp120. La proteína gp41 con el péptido de fusión permite que la membrana celular y el VIH se fusionen. Es necesario antes de que exista la fusión una interacción entre la glicoproteína gp120 con CD4 y el correceptor de quimiocinas (**ver figura 2.11**) [12].

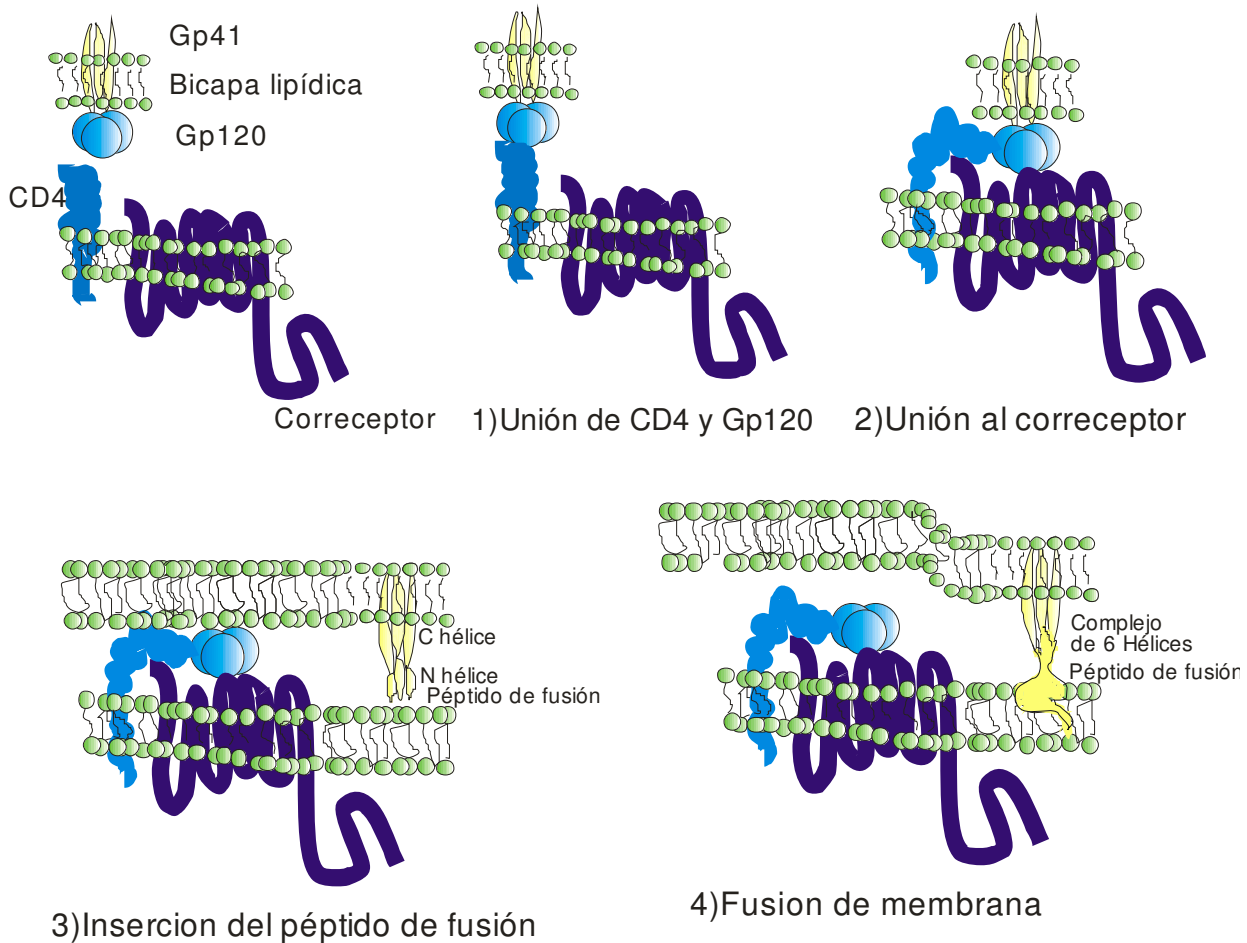


Figura 2.11. Representación esquemática de la fusión de membrana. La fusión de membranas inicia con la unión de CD4 y gp120(1), lo que facilita la interacción con el correceptor de quimiocinas (2). gp41 adopta un conformación que permite la inserción del péptido de fusión en la membrana (3), finalmente se fusiona la membrana celular con el virus (4) [12].

El producto proteico de gp41 tiene un dominio llamado péptido de fusión que se encuentra en el extremo N terminal (**ver figura 2.12**). La proteína gp41 tiene dos dominios el ectodominio y el dominio citoplasmático. En el ectodominio encontramos el péptido de fusión el cual es responsable de la fusión del virus con la membrana.[12].

Gp41(TM)

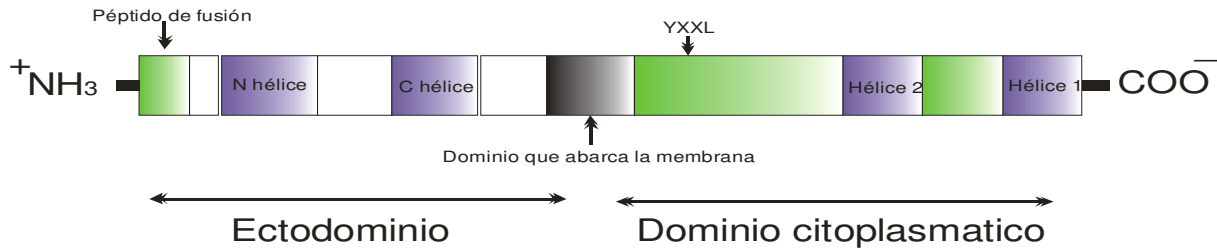


Figura 2.12 Representación lineal de la glicoproteína gp41. En el ectodominio se encuentra la región del péptido de fusión, así como dos dominios helicoidales (N hélice, C hélice). El dominio citoplasmático contiene una región Thr-X-X-Leu (YXXL) seguido por dos dominios helicoidales (Hélice2, Hélice1) [12].

2.3.4 Ciclo de replicación del VIH

El virus del SIDA puede infectar células T humanas en cultivo, multiplicarse por sí mismo y en muchos casos puede causar lisis de la célula hospedadora. Es mucho lo que se ha aprendido del ciclo de vida del VIH-1 a partir de los estudios *in vitro*. Se han caracterizado las diversas proteínas codificadas por el genoma viral y se conocen las funciones de la mayor parte de ellas [13].

La primera etapa de la infección por el VIH es la fijación y la entrada del virus en la célula blanco. El VIH-1 infecta las células que portan el antígeno CD4 en su superficie; por consiguiente el virus infecta preferentemente linfocitos T CD4+, además de que ciertas cepas del virus infectan también monocitos y otras células que expresan CD4 en su superficie. La preferencia por las células CD4+ se debe a una interacción de alta afinidad entre una proteína de envoltura gp120 del virus y el antígeno CD4 de la superficie celular. La interacción gp120-CD4 no basta para que el virus ingrese a la célula hospedadora y produzca infección, siendo necesaria la expresión adicional de otras moléculas de la superficie celular, que se han reconocido como correceptores virales y que están presentes en células como linfocitos T, monocitos/macrófagos y células dendríticas.[7] Cepas virales que utilizan preferentemente el correceptor CXCR4 infectan linfocitos T, mientras que cepas que utilizan preferentemente el correceptor CCR5 tienen tropismo preferencial hacia monocitos/macrófagos (aunque también pueden infectar linfocitos T).(Ver figura 2.13) [13].

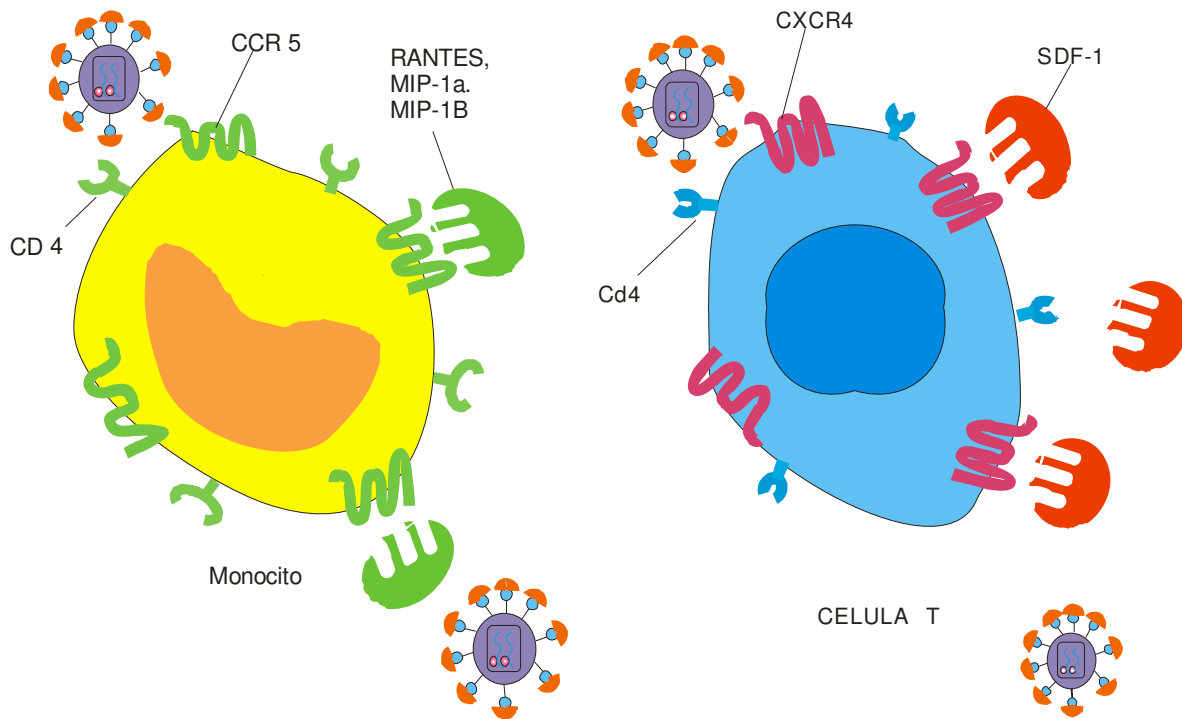


Figura 2.13 Tropismo de diferentes cepas de VIH. Algunas cepas tienen tropismo a los monocitos debido a que presentan el correceptor CCR5 y otras cepas presentan tropismo a las células T ya que presentan el correceptor de quimiocinas CXCR5.

Al unirse gp120 a la molécula CD4 existe un cambio conformacional el cual permite que los correceptores se unan a gp120. Esta unión permite que el dominio de fusión de gp41 quede expuesto y se fusionen la membrana celular y viral, permitiendo el ingreso del virus a la célula.

La descapsidación del virus no ocurre inmediatamente después de la entrada. Se ha demostrado que la cápside completa viaja hacia el núcleo dirigida en un principio por microtubulos y posteriormente por filamentos de actina en las proximidades del núcleo. Movimientos lentos dirigidos por actina llevan las capsides a encajar en los poros nucleares. La transcripción inversa del genoma viral ocurre en el llamado complejo de retrotranscripción (RTC) dentro de la cápside, en el poro nuclear. Una vez retrotranscrito el genoma viral y el ámbito del poro nuclear ocurre la descapsidación, la cual es mediada por una estructura característica de los genomas se forma por la interacción en cis de dos secuencias lentivirales: El tracto central de polipurina (cPPT) y las secuencia central de

terminación (CTS). El ADN viral carente del ala central de ADN se acumula en los poros nucleares pero no se transporta al interior del núcleo la descapsidación favorecen la organización del complejo de preintegración (PIC) compuesto del ADNc de doble cadena, IN (integrasa), retrotranscriptasa (RT), p17 (MA), nucleocápside (NC) y Vpr. A través de un proceso poco caracterizado el PIC se transporta al interior del núcleo donde el ADNc de doble cadena puede integrarse en la cromatina de la célula hospedera o circularizarse formando estructuras episomales no funcionales. Se cree que las señales de localización nuclear de varias proteínas como IN, MA y Vpr son importantes para dirigir PIC al interior del núcleo. [4] El ADN integrado sirve como molde para la formación de mRNAs dependientes de la RNA polimerasa II. Para elevar la eficiencia de la transcripción, la RNA polimerasa II interactúa con la proteína viral Tat, y algunos transactivadores transcripcionales como son el NF- κ B y Sp1. Los transcritos son exportados hacia fuera del núcleo por Rev. Las poliproteínas de Gag y Gag-Pol son sintetizadas en los ribosomas libres del citoplasma y son transportados a la membrana citoplasmática. El transcrito del precursor Env (gp160) se traduce en el retículo endoplásmico donde es glicosilado. Del retículo endoplásmico pasa al aparato de Golgi en el cual se da una proteólisis con enzimas celulares que origina a los productos gp120 y gp41. Al igual que los productos de *gag* y *pol*, las glicoproteínas de envoltura son transportadas a la membrana plasmática. La poliproteína Pol/Gag y la proteína Gag se asocian al genoma viral (el cual se encuentra como un dímero de RNA) y finalmente se condensa formando una partícula esférica en la membrana citoplasmática. La maduración de la partícula vírica se da por la proteólisis de las poliproteínas Pol y Gag por la proteasa viral **(ver figura 2.14)** [12].

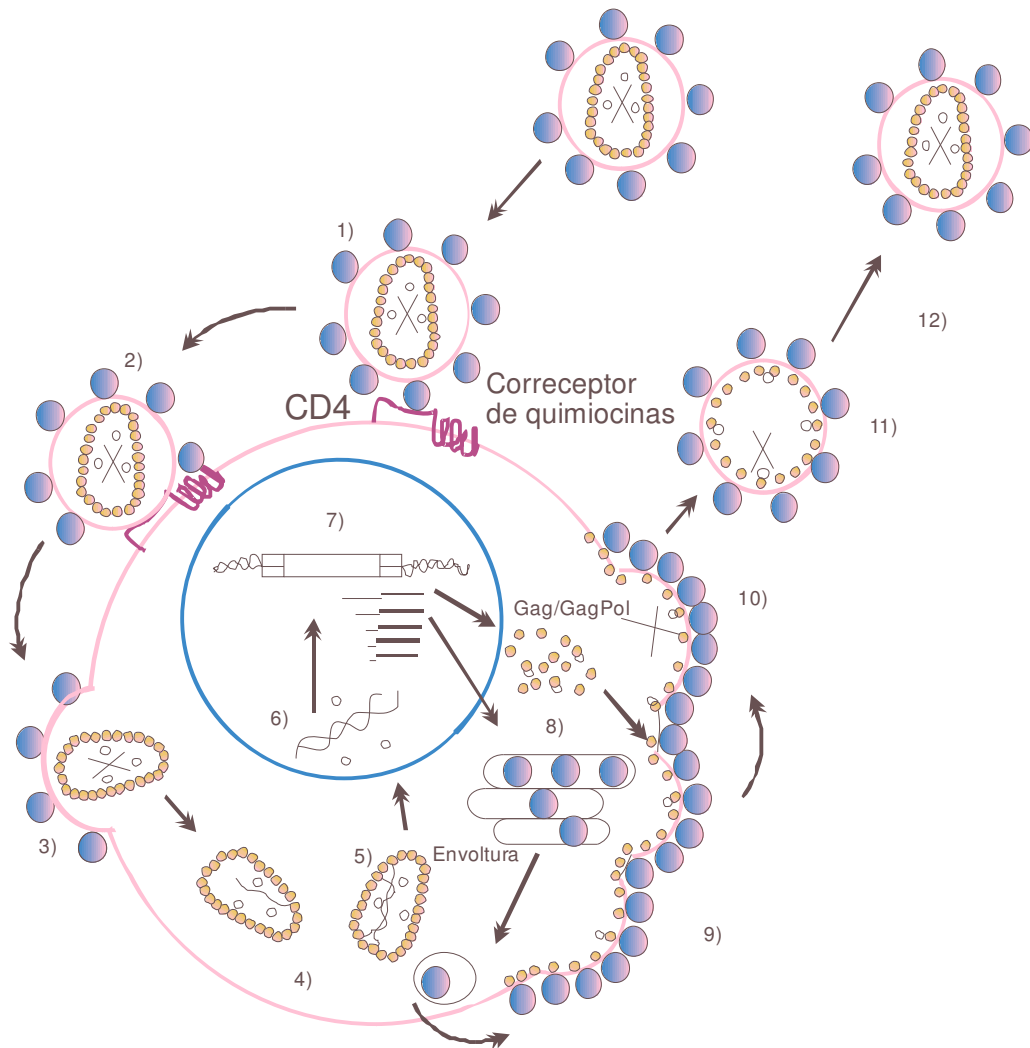


Figura 2.14 Esquema del ciclo de replicación del VIH. El primer paso es la interacción de gp120 con CD4 (1) lo cual causa un cambio conformacional que favorece la interacción del correceptor de quimiocinas con gp120 (1). La fusión del virus con la membrana celular se lleva a cabo entonces gracias a la interacción de gp41 (3). Al entrar el virus, la cápside se desintegra (4). En el citoplasma se lleva a cabo por medio de la retrotranscriptasa viral la síntesis de DNA de doble cadena(4) . El DNA de doble cadena se transporta al núcleo (5). Una vez en el núcleo, el DNA viral se integra al genoma de la célula gracias a la integrasa y se transcribe. Los mRNA son exportados del núcleo gracias a la proteína Rev, y se traducen en el citoplasma en el caso del precursor Gag Pol y en el retículo endoplásmico en el caso del precursor gp160 (8). Las proteínas virales se transportan a la membrana (9). Las poliproteínas Gag y Pol/Gag son asociadas a dímeros de RNA genómico viral (10), formando una partícula esférica en la membrana citoplasmática (11). La maduración de la partícula vírica se da a partir de que comienza la proteólisis de las poliproteínas Pol y Gag por la proteasa viral. Tomado de [12].

2.4 Historia natural de la infección por VIH

La prueba para la detección del VIH es por medio de la identificación de anticuerpos contra los antígenos de este virus, los cuales aparecen por lo general

en un periodo de tres meses después de la infección. Cuando los anticuerpos específicos contra el VIH aparecen se dice que el individuo ha seroconvertido o es seropositivo hacia el VIH. Aunque la velocidad de la evolución de la infección y el momento de inicio de los síntomas que caracterizan al SIDA varía considerablemente entre paciente y paciente, es posible trazar un esquema general para valorar el avance de la enfermedad por VIH (**ver figura 2.15**) [13].

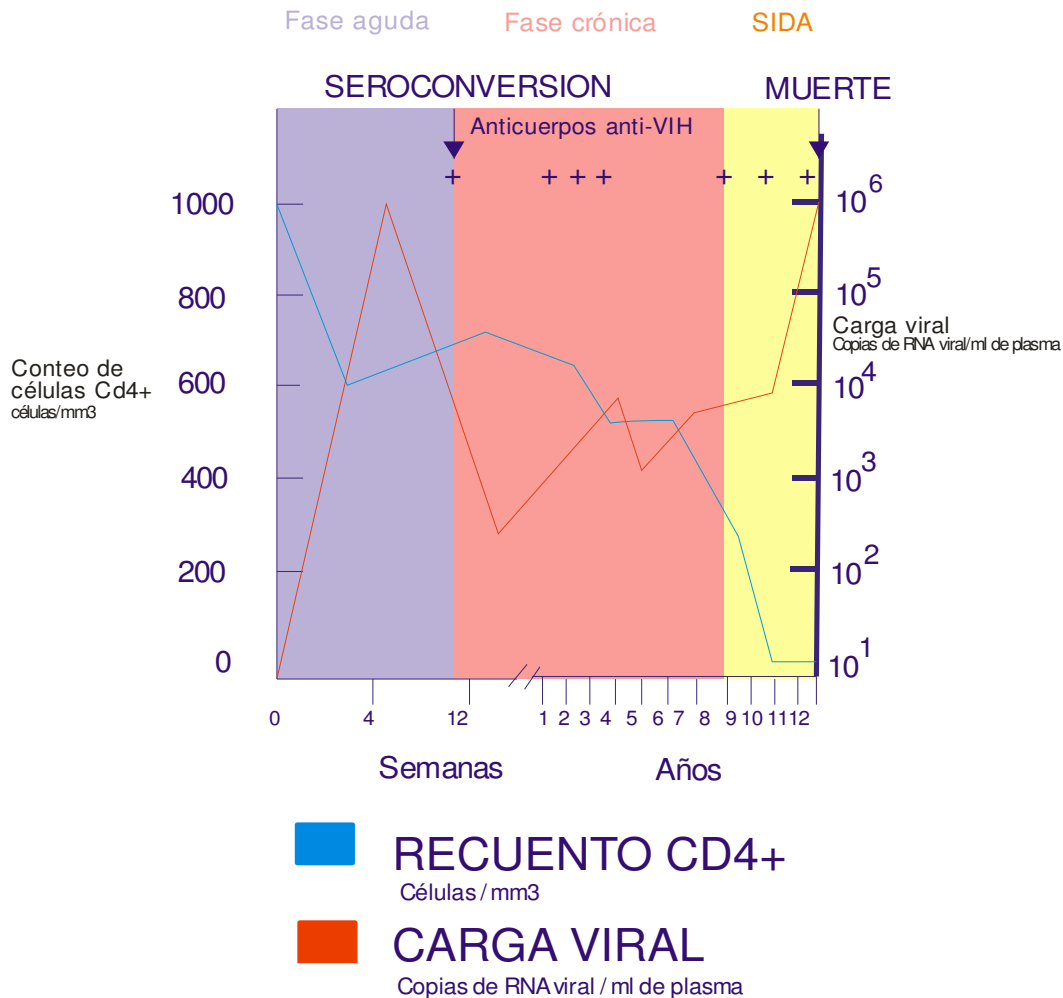


Figura 2.15 Historia natural de la infección por VIH que ilustra tres etapas del proceso infeccioso. Poco después de la infección es perceptible RNA en el suero. Sin embargo, la infección por VIH se identifica más a menudo por la presencia de anticuerpos anti-VIH después de la seroconversión, que ocurre en un plazo de unos cuatro meses después de contraer la infección. Por lo general no aparecen los síntomas clínicos indicadores de SIDA durante al menos ocho años

después de la infección, aunque este intervalo es variable. El inicio del SIDA clínico es anunciado con frecuencia con el descenso de las cifras de células T y aumento de la carga viral [13].

La fase aguda de la infección por VIH se caracteriza por una alta viremia en plasma y una disminución notoria de células CD4+ seguidas de una disminución y estabilización de la viremia y un aumento de células CD4+, normalmente a valores inferiores a los anteriores a la infección.[6] Durante la fase aguda, es frecuente la aparición del síndrome retroviral agudo, caracterizado por síntomas parecidos a la mononucleosis. Estos síntomas aparecen en los días o semanas posteriores a la infección.

Al finalizar la fase aguda se alcanza un equilibrio entre la replicación viral y la respuesta inmune del hospedero. La mayoría de los individuos infectados por el VIH no presentan manifestaciones clínicas durante esta fase que se extiende por un largo periodo de años, denominándose consecuentemente fase crónica o de latencia. Aún sin el uso de fármacos antirretrovirales, esta etapa se puede prolongar de 8 a 10 años o más. Es importante mencionar que en la fase crónica o de latencia existe una tasa de replicación del virus notable (produciéndose más de 1 000 000 000 de viriones nuevos al día) a pesar de que no exista sintomatología. Durante la fase crónica, las cuentas de células CD4+ tienden a disminuir y la viremia a incrementarse lentamente (**Figura 2.15**).

Al llegar la cuenta de células CD4+ a valores de 200 células/ μ l inicia la fase conocida como SIDA, caracterizada por la aparición de enfermedades oportunistas. Durante la etapa de SIDA pueden aparecer diversos síntomas inmunológicos, hematológicos, dermatológicos y neurológicos. Varios de estos síntomas se encuentran en la categoría B de la tabla de la CDC (*U.S. Centers for Disease Control and Prevention*) ver tabla 2.3. La etapa de SIDA también se caracteriza por el desarrollo de síntomas generales característicos como son pérdida de peso, sudoración nocturna y diarrea. En estos casos el valor de células CD4+ es muy importante ya que por debajo de las 200 células CD4+/ μ l se incrementa el riesgo de contraer varias enfermedades que definen al SIDA (**ver**

tabla 2.3). Por arriba de 200 células CD4+/ μ l es raro que aparezcan enfermedades que definan al SIDA[6].

La primera indicación de SIDA puede ser una infección oportunista por la levadura *Candida albicans*, que puede causar úlceras en la boca o en el caso de las mujeres puede causar infecciones vulvovaginales que no reaccionan con el tratamiento. Otro indicador temprano puede ser la aparición de neumonías por *Pneumocystis jirovecii*. El incremento de la concentración de VIH-1 circulante en el plasma y la reducción del número de células T CD4⁺ son casi siempre el anuncio de SIDA. Se ha establecido cierta relación entre el número de células T CD4⁺ y el tipo de infección que experimenta el paciente.(**ver tabla2.3**)[13].

El sistema de clasificación CDC clasifica a los individuos infectados por VIH basándose en criterios clínicos y de laboratorio. Este sistema define el SIDA como una reducción de leucocitos CD4+ a valores menores del 14% del total. Por tal motivo, un individuo con una cuenta menor a 200 células CD4+/ μ l se considera que se encuentra en la fase de SIDA.(**ver tabla 2.3**) [6].

Tabla 2.3. Clasificación de la infección por VIH por el sistema CDC

Recuento de células T CD4 ⁺	Categoría clínica		
	A	B	C
≥500/μL	A1	B1	C1
200-499/μL	A2	B2	C2
<200/μL	A3	B3	C3
Categoría A Pacientes asintomáticos al momento de la infección por VIH. Infección primaria aguda: enfermedad del tipo de la fiebre ganglionar que dura unas cuantas semanas a partir del momento de la infección. Linfadenopatía generalizada persistente (PGL): aumento del tamaño de los ganglios linfáticos que persiste durante tres meses o más sin datos de infección.		Categoría C Candidosis bronquial, traqueal o pulmonar Candidosis esofágica Cáncer cervical Coccidioidomicosis diseminada o extrapulmonar Criptosporidiosis intestinal crónica Enfermedad por citomegalovirus Rinitis por citomegalovirus Encefalopatía relacionada con VIH Herpes simple: úlceras en la boca duración mayor a un mes Neumonitis o esofagitis Histoplasmosis diseminada y extrapulmonar Sarcoma de Kaposi Linfoma de Burkitt Síndrome de emaciación por VIH Linfoma inmunoblástico Linfoma cerebral primario Septicemia por <i>Salmonella</i> (recurrente) Toxoplasmosis cerebral Neumonía por <i>Pneumocystis jirovecii</i> Enfermedad por <i>Mycobacterium</i> Infección por <i>Mycobacterium tuberculosis</i> extrapulmonar	
Categoría B Angiomatosis bacilar Candidosis bucofaríngeas Candidosis vulvovaginal Displasia cervical Síntomas generales como fiebre y diarrea que dura más que un mes. Leucoplaquia vellosa bucal Herpes zoster Púrpura trombocitopénica idiopática Listeriosis Enfermedad inflamatoria pélvica Neuropatía periférica			

2.5 Origen del VIH

Evidencias filogenéticas han demostrado una fuerte similitud entre los virus de inmunodeficiencia humana y los virus de inmunodeficiencia del simio, teniendo ambos ancestros comunes que sugieren la existencia de múltiples saltos zoonóticos de los virus simianos a los humanos. Ahora es claro que la epidemia global por VIH-1 es el resultado de un paso interespecie de un lentivirus que infectaba a chimpancés, el virus de la inmunodeficiencia simiana (SIV_{CPZ}, por sus siglas en inglés), lo cual se cree ocurrió en África oriental. La forma en la cual se cree que el SIV dio el paso interespecie de chimpancés a humanos, fue a raíz del

contacto directo continuo entre humanos y chimpancés, ya que en África oriental se acostumbra cazar chimpancés para usarlos como alimento [11]. Se cree que al momento de cazar al animal este le producía heridas al cazador las cuales posteriormente tenían contacto con la sangre infectada del animal. El caso del VIH-2 representa una zoonosis de SIV_{smm}, transmitido a los humanos por los simios mangabey ahumados (**ver figura2.16**) [8].

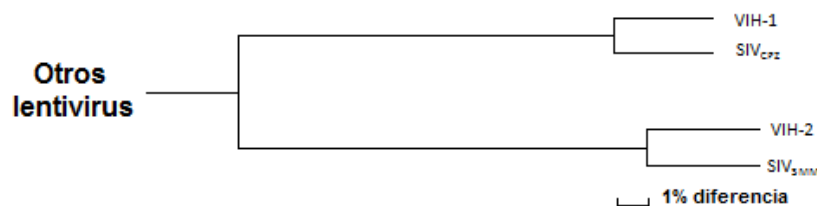


Figura 2.16. Relación filogenética entre el VIH-1 y VIH-2 basado en secuencias del gen pol. SIV_{CPZ} y SIV_{smm}, son lentivirus de primates no humanos, aislados de chimpancés y de monos sooty mangabey respectivamente.[8].

Se cree que los chimpancés adquirieron la infección por SIV después de la divergencia entre las múltiples subespecies. *Pan troglodytes troglodytes* y *Pan troglodytes schweinfurthii* son dos especies de chimpancé infectadas naturalmente a pesar de que otras especies que se encuentran estrechamente relacionadas no se infectan. Esto sugiere que es poco probable que el SIV_{CPZ} haya coevolucionado con su hospedero natural. No se sabe aun como los chimpancés se infectaron por primera vez por este virus, aunque es lógico suponer que la infección inicial ocurrió por contacto directo con otros monos, ya que los chimpancés cazan monos. Así, se cree que el SIV_{CPZ} se generó a partir de un evento de recombinación en el cual estaban involucrados dos SIV ancestrales. Es de notar que el SIV_{CPZ} infecta de forma asintomática a los chimpancés y como otros SIV no causa enfermedad en sus hospederos naturales (**ver figura2.17**). Si se infecta de forma experimental con este virus a animales susceptibles a este virus que no son los hospederos naturales, se tiene como consecuencia una enfermedad progresiva en la cual el animal presenta un síndrome de inmunodeficiencia

adquirida (SIDA). La diferencia entre la patogenicidad entre una especie y otra aun no se ha aclarado, pero se cree que se debe a una adaptación gradual del hospedero característico al virus a lo largo de los años[8].



Figura 2.17. Origen del VIH. arriba a la izquierda *P.t. schweinfurthii* (Chimpance), arriba derecha *P. t. troglodytes*(Chimpance),abajo Sooty mangabeys,.

2.6 Variabilidad del VIH

En el pasado se creía que el VIH presentaba una alta homogeneidad al igual que los retrovirus que se habían estudiado previamente, pero en los estudios recientes se ha encontrado que existe una gran heterogeneidad entre los virus circulantes en América, África, Europa y Asia. A partir de esto se sabe que no existen dos VIH idénticos. Por ejemplo, si se analizan las secuencias de los virus de un solo paciente se podrá observar que existe una gran cantidad de variantes que coexisten en el individuo en un punto determinado de la infección. Diversos factores causan cambios en los nucleótidos del genoma viral, algunos de los cuales pueden traducirse en cambios en la secuencia de aminoácidos de las proteínas virales. Estos cambios de los nucleótidos ocurren en todo el genoma, pero existe una mayor tasa de variación en el gen *env*. El término cuasiespecies describe la diversidad encontrada en la población de los virus en un solo individuo infectado con VIH-1[8].

Los factores que influyen en la gran heterogeneidad genética del virus son:

- La propensión al error de la transcriptasa inversa viral al transcribir el genoma viral de RNA a DNA (3×10^{-5} mutaciones/nucleótido/ciclo replicativo)[8].
- Alta tasa de recombinación que acompaña a la transcripción inversa del genoma viral (1.38×10^{-4} eventos de recombinación/sitios adyacentes)[14].
- La alta tasa de replicación viral (10^9 partículas virales/día)[8]
- La gran cantidad de individuos infectados [8].

2.7 Clasificación del VIH

El VIH se puede clasificar en dos tipos, VIH-1 y VIH-2. Estos dos virus difieren en su genoma en un 1% y presentan diferentes proteínas accesorias. Se sabe que el VIH-2 tiene una mayor relación con el virus SIV_{smm} el cual infecta a monos mangabey ahumados, mientras que el VIH-1 tiene mayor relación con el SIV_{CPZ} del chimpancé. El VIH-1 es el causante de la mayoría de los casos de infección por VIH a nivel mundial, en cambio el VIH-2 se encuentra localizado mayoritariamente en África oriental y central[15].

El VIH-1 se puede clasificar en grupos: M(Mayor), O(outlier) y N(no M/no O); El VIH-1 grupo M es el responsable de la pandemia de SIDA. El VIH-1 del grupo O es menos común y se encuentra localizado en África Occidental, principalmente en Cameroon, Gabon, Nigeria y La Guinea Ecuatorial. Se han encontrado casos de personas infectadas con VIH-1 grupo O también en Europa y Estados Unidos, aunque los individuos infectados con virus de este grupo han presentado una relación cercana con países de África Occidental. Menos del 10% de los casos de VIH-1 en el mundo son causados por el grupo O. Los casos de infección por VIH-1 grupo N son escasos y solo se han localizado en personas de África Occidental, en Cameroon[15].

El VIH-1 grupo M, causante de cerca del 90% del total de infecciones a nivel mundial, debido a la gran diversidad del VIH se puede dividir en subtipos o clados los cuales van de la A-D, F-H, J, K. La variación genérica que existe entre virus

pertenecientes al mismo subtipo es de un 15-20% mientras que el porcentaje de variación entre cada subtipo es de 25%-35%.[16]

En el caso del VIH-1 del grupo O no se han definido subtipos, pero se ha propuesto su clasificación en 5 variantes que se designan I-V. El VIH-2 se ha clasificado de en 5 subtipos que se designan de la A-E[15].

Adicionalmente, los subtipos A y F de VIH-1 del grupo M se pueden subclasificar en los sub-subtipos: A1-A4 y F1-F2. Existen además las llamadas formas recombinantes, las cuales incluyen virus híbridos originados por uno o más eventos de recombinación con virus de otros subtipos. Se han diferenciado dos tipos de formas recombinantes. Las formas recombinantes circulantes (CRF) y las formas recombinantes únicas (URF) (**ver figura2.18**) [17].

Si la forma recombinante se ha identificado en al menos 3 individuos no relacionados, esta se clasifica como forma recombinante circulante. Por el momento, se han descrito 35 formas recombinantes circulantes, las cuales se encuentran localizadas principalmente en África; solo 5 de estas formas recombinantes circulantes se han localizado fuera de África. Los virus pertenecientes al antes existente subtipo E, hoy en día se han redefinido como formas recombinantes circulantes. Existen casos aislados en que se encuentran formas recombinantes en ciertos individuos epidemiológicamente relacionados. Estas formas recombinantes se llaman formas recombinantes únicas (URFs), y se cree que son el resultado de un segundo fenómeno de recombinación entre formas recombinantes circulantes [17].

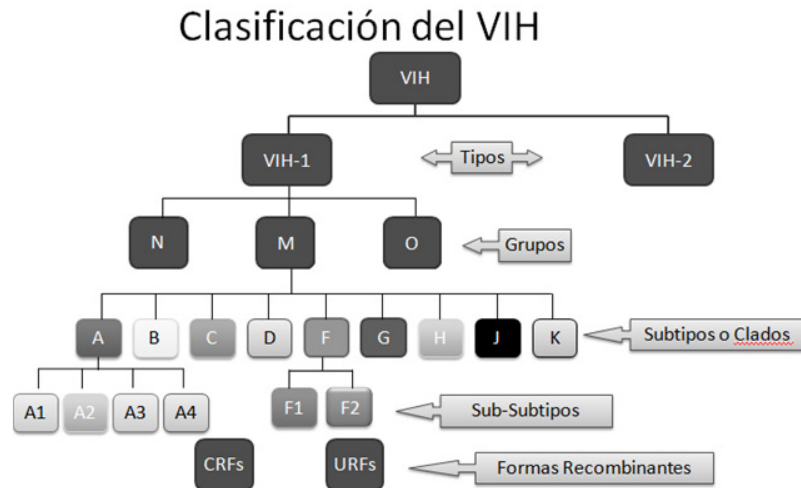


Figura 2.18 Clasificación del VIH. El VIH se clasifica en tipos: VIH-1 y VIH-2. El VIH-1 se ha diseminado mayoritariamente en el mundo, dividiéndose en los grupos M, N y O de los cuales el predominante es el grupo M. El grupo M a su vez se divide en subtipos: A-D,F-H,J,K. El subtipo A se clasifica en los sub-subtipos A1-A4, y el subtipo F se clasifica en los sub-subtipos F1 y F2. Existen además formas recombinantes, estas son las formas recombinantes circulantes (CRFs) y las formas recombinantes únicas (URFs).

2.8 Epidemiología molecular del VIH en el mundo

Se sabe que aproximadamente un 50% de los casos causados por VIH-1 a nivel mundial son causados por virus del subtipo C. Los subtipos A, B, D y G son los responsables del 12%,10%, 3% y 6% de las infecciones por VIH a nivel mundial. Los subtipos F, H, J y K juntos causan el 0.94% de las infecciones. Las dos formas recombinantes circulantes que causan mayor número de infecciones son la CRF01_AE y CRF02_AG las cuales son responsables del 5% de las infecciones a nivel global. Otra forma recombinante circulante con amplia distribución a nivel mundial es la CRF03_AB, la cual es responsable del 0.1% de las infecciones. El resto de las formas recombinantes causan el 8% de las infecciones. Todas las formas recombinantes en conjunto causan el 18% de las infecciones a nivel mundial (**ver figura 2.19**) [18].

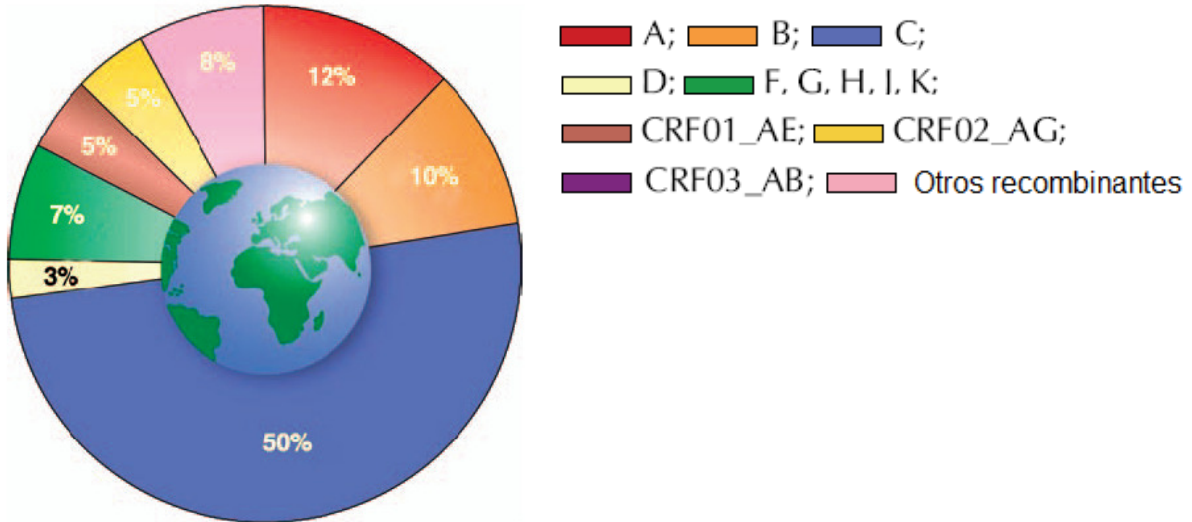


Figura 2.19 Prevalencia mundial de los subtipos de VIH. El 50% de los casos de VIH son causados por virus del subtipo C. Los subtipos A, B, D y G son los responsables del 12%,10%, 3% y 6% respectivamente [18].

Del 100% de casos de VIH en el mundo el 64% se localiza en África sub-Sahariana. Del total de personas infectadas en esta región por VIH-1 el 56% tiene infecciones por virus del subtipo C, y una pequeña proporción son infectados por virus del subtipo A (14%), G (10%), CRF02_AG (7%) y otros recombinantes (9%) (ver figura 2.20)[18].

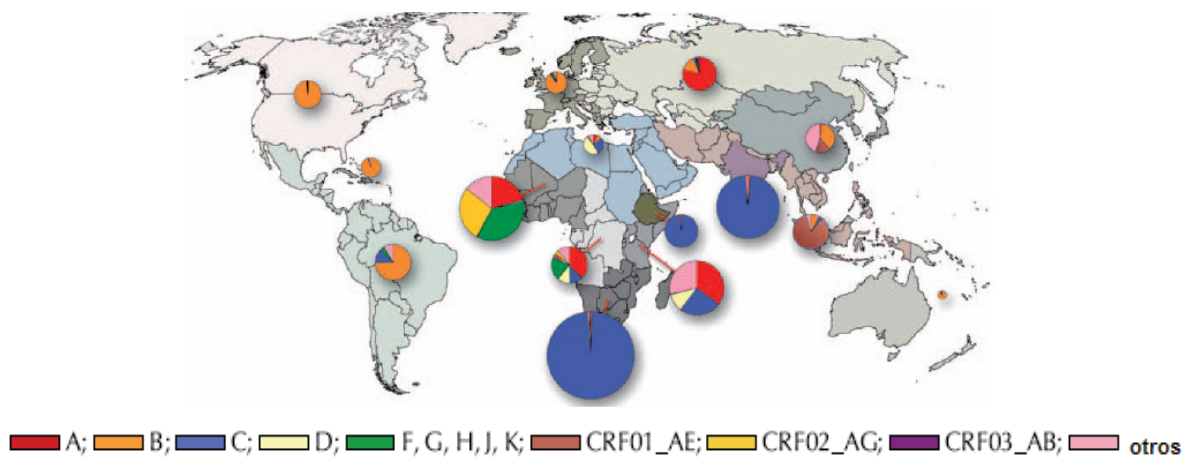


Figura 2.20. Distribución global de los subtipos de VIH-1. Se muestra la distribución de subtipos en diferentes regiones del mundo con datos de 2004. El tamaño de las gráficas de pastel

son proporcionales al número de individuos infectados en la región. Las regiones con mayor afectación son África y la India [18].

Un 30% de las personas infectadas por VIH-1 a nivel mundial se encuentran en Sudáfrica, mientras que un 4% de las personas infectadas por VIH-1 a nivel mundial se encuentran en Etiopía. En estas zonas, el 98% y el 99% de las infecciones son causadas por virus del subtipo C respectivamente [18].

En África occidental el 21% de las infecciones por VIH-1 son causadas por el subtipo A, 35% son causadas por el subtipo G, 28% son causadas por la forma recombinante circulante CRF02_AG y el resto por otras formas recombinantes.[18]

Nigeria, uno de los países más afectados por esta pandemia, tiene una prevalencia de 29% de subtipo A y 56% de subtipo G. Otros países de la región, como lo son Cameroon, Ghana, Côte d'Ivoire presentan una distribución de subtipos diferente, en la cual CRF02_AG predomina, existiendo además un pequeño porcentaje de infecciones causadas por VIH-1 subtipo A y otras cepas recombinantes[18].

En el este de África se encuentran el 10% de los individuos infectados por VIH-1. El subtipo más prevalente en esta región es el A con un 35% de los casos, C con 25%, D con el 11% y una gran proporción de formas recombinantes únicas, las cuales representan el 29% de los casos. Estos mismos subtipos son encontrados en todos los países de esta región[18].

En África central, donde se encuentra el 5% de los individuos infectados por VIH-1 en el mundo, se encuentra la mayor variedad de subtipos y recombinantes. En esta región, el subtipo que más prevalece es el subtipo A (38%), mientras que los subtipos G, C, D y otros recombinantes representan cerca del 11% del total de las infecciones en la región. Los subtipos F, H, CRF01_AE y CRF02_GA están entre el 3% y 4%. Los subtipos B, J, K representan menos del 2% de las infecciones. [18]

En la India se encuentran el 13% de los individuos infectados por VIH-1 en el planeta, y el subtipo que predomina en esta región es el C, causando un 97% de las infecciones. En el resto del Sur y Sureste de Asia se encuentran el 5% de los individuos infectados por VIH-1 a nivel mundial. La epidemia en esta región se

encuentra dominada por la forma recombinante circulante CRF01_AE (84%). Esta es la región en la cual se encuentra el mayor número de personas infectadas por formas recombinantes [18].

En el este de Asia, el cual representa el 3% de infectados por VIH-1, China es el país con mayor número de personas infectadas por VIH-1. En este país, el 38% de las infecciones están dadas por el subtipo B, el 15% de las infecciones está causada por la forma recombinante CRF01_AE y otros recombinantes representan el 45% de las infecciones, predominando CRF07_BC y CRF08_BC. Así, podemos decir que la mayor parte de las infecciones en esta región está dada por recombinantes (61%). En Hong Kong, el 50% de los casos es causado por virus del subtipo B y el 45% de los casos es causado por CRF01_AE. En Japón, el 81% de los casos es causado por virus del subtipo B y el resto es causado por virus pertenecientes al subtipo A, C y la forma recombinante CRF01_AE [18].

En Oceanía, la distribución que se encuentra en Nueva Zelandia y Australia muestra un predominio del subtipo B (88%), con un 5% de los casos asociados a virus del subtipo C y el recombinante CRF01_AE [18].

El Norte de África y medio Oriente incluyen el 1% de los individuos infectados por VIH-1 a nivel global. En esta región el subtipo que predomina es el D con un 47% de los casos, el subtipo C representa el 29%, los subtipos A, B y las formas recombinantes representan un 6%, 7% y 9% respectivamente de los casos de infección por VIH-1 en esta región [18].

Del total de casos de VIH-1 el 3% se encuentra en América del Norte. En estados Unidos, el 94% de las infecciones son causadas por el subtipo B. Desafortunadamente no hay datos disponibles de Haití, a pesar de que se trata de uno de los países mayormente afectados por el virus en esta región. En la república Dominicana y en Trinidad y Tobago más del 94% de las infecciones por VIH-1 son causadas por el subtipo B. En Cuba se encuentra una mayor diversidad

con un 48% de los casos causado por virus del subtipo B, un 41% causado por virus recombinantes y los casos restantes causados por subtipos C, D, F, G, H y J[18].

En Latinoamérica se encuentra el 4% de los individuos infectados por este virus, y el 74% de los casos es causado por virus del subtipo B. El 13% de los casos es causado por virus del subtipo C, el 4% es causado por virus del subtipo F y el 8% de los casos es causado por virus CRF12_BF y otros recombinantes B/F [18].

El subtipo B es el causante de más del 94% de los casos de infección por VIH-1 en Chile, Colombia, Ecuador, Honduras, México y Venezuela. En Brasil, Paraguay, Uruguay se encuentra el subtipo F. En Argentina, Bolivia, Paraguay y Uruguay se encuentran recombinantes de varios tipos en un 49%, 16%, 22% y 19% respectivamente. En Argentina, el 49% de los casos de infección por VIH-1 es causado por el subtipo B, mientras que en Brasil el 28% de los casos es causado por virus del subtipo C [18].

En Europa Occidental se encuentra el 1% de los casos de VIH-1 a nivel mundial, de los cuales el 88% es causado por el subtipo B; los casos restantes son causados por virus pertenecientes a los subtipos A, C, G y la forma recombinante CRF02_AG. En países como España, Francia e Italia la prevalencia del subtipo B es mayor al 92%. En los demás países de la región, la proporción del subtipo B es menor o igual al 82% con el mínimo de 39% en Portugal. Al mismo tiempo, en estos países se encuentran varios subtipos no-B que son importantes; por ejemplo, en Austria el subtipo A es altamente prevalente con un 14%, Dinamarca con un 8%, Reino Unido con 8% y Grecia con un 6%; el subtipo C es representativo en Dinamarca con un 18%, Suecia con un 30% y el Reino Unido con 17%. El subtipo G es representativo en Portugal con un 26%. En Portugal y Grecia existe un alto número de formas recombinantes, con un 31% y 12% respectivamente [18].

En Europa del Este y Asia Central se encuentra el 4% de los individuos infectados por VIH-1. En esta región el subtipo predominante es el A con un 79%. Un 15% está dado por el subtipo B y el resto está dado por los subtipos C (2%), F (1%) y CRF02_AG (3%) [18].

2.9 Epidemiología molecular del VIH-1 en México

El conocimiento de la epidemiología molecular del VIH en México es notablemente escaso. Los trabajos publicados hasta la fecha muestran que el subtipo predominante en México es el B. Sin embargo, las conclusiones de los estudios que se han realizado en el país son limitadas, al incluir cohortes pequeñas y enfocadas en grupos de riesgo particulares. Hasta la fecha solamente existen dos publicaciones que abordan el tema de la epidemiología molecular del VIH en el país. Rivera-Morales y colaboradores caracterizaron virus de una cohorte de 65 pacientes varones provenientes de Nuevo León, Jalisco, Puebla, Yucatán y Distrito Federal en el 2001[19]. Eyzaguirre y colaboradores estudiaron una cohorte de únicamente 11 individuos infectados por VIH de la frontera Norte de nuestro país, los cuales eran usuarios de drogas intravenosas (UDI) o sexoservidoras [20].

En el Centro de Investigación de Enfermedades Infecciosas (CIENI) del Instituto Nacional de Enfermedades Respiratorias (INER), lugar donde se llevó a cabo el presente trabajo de investigación, se encuentra en realización el estudio de epidemiología molecular de VIH más grande y con más alta representatividad hasta la fecha para México. Se ha reunido una cohorte de más de 1500 individuos infectados por VIH, sin tratamiento antirretroviral previo, provenientes de varios de los estados de la República Mexicana con mayor contribución al número total de infecciones por VIH en el país: Veracruz, Oaxaca, Estado de México, Chiapas, Puebla, D. F., Morelos, Jalisco, Guerrero, Nuevo León y Baja California (Jornadas CIENI enero 2009, datos no publicados). En este estudio, se analizaron secuencias del gen *pol* viral y se encontró que el 99.91% de los individuos incluidos en la cohorte mexicana estaban infectados por virus de subtipo B y un 0.09% por recombinantes BF (datos no publicados). Los análisis filogenéticos

derivados del gen *pol* de los virus de la cohorte mexicana mostraron una inesperadamente alta homogeneidad en el VIH circulante en el país [21], sin observarse agrupamientos de virus provenientes de las mismas áreas geográficas. Se cree que la Epidemia de VIH en México se encuentra altamente influenciada por la epidemia presente en los Estados Unidos. Sin embargo, la variabilidad del virus circulante en E. U. parece ser mayor que la observada en nuestro país. Un estudio reciente demostró que la prevalencia de subtipos no-B en E. U. está en aumento con un porcentaje general de subtipos no-B de 2.93% entre el 2004 y el 2008 y observándose la presencia de subtipos A1, A2, A3, C, D, F1 y G, así como 11 CRFs y 9 URFs [22] Estos datos contrastan con el estudio mexicano del CIENI donde solamente 0.09% de los virus circulantes corresponden a subtipos no-B. Sin embargo, los resultados obtenidos en el CIENI no son completamente concluyentes, ya que la alta homogeneidad de las secuencias de los virus circulantes analizados se podría deber a la relativamente baja variabilidad del gen *pol* dentro del genoma viral. Por esta razón, en el presente trabajo se presenta un análisis filogenético del VIH circulante en México, utilizando secuencias del altamente variable gen *env* viral, obtenidas a partir del banco de sueros del proyecto de epidemiología molecular del CIENI del INER. Los resultados de este estudio serán valiosos para apoyar las conclusiones del estudio de epidemiología molecular del VIH del CIENI, hasta ahora el más representativo de la epidemia de VIH en México, con el objeto de generar conocimiento a nivel molecular de la epidemia de VIH/SIDA en México, con aplicación en el diseño de posibles vacunas efectivas en nuestra población y la generación de políticas de salud pública que contribuyan al manejo de la epidemia en el país.

2.10 Filogenia

La filogenia es la ciencia que trata de reconstruir el flujo hereditario en diversos niveles evolutivos/tiempo. En un principio la filogenia se basaba en características morfológicas, posteriormente se introdujeron pruebas inmunológicas, finalmente la introducción de la Biología Molecular revolucionó esta ciencia [23].

En filogenia un carácter genealógico o heredable es aquél que nos proporciona información desde una perspectiva genealógica. De tal forma, un aminoácido o un nucleótido son caracteres genealógicos. La homología, es decir la relación entre caracteres sobre su ascendencia común, ayuda a reconstruir el flujo de herencia [24].

En el análisis filogenético se analizan las Unidades Taxonómicas Operativas (OTUs), las cuales pueden ser secuencias de ADN o de proteínas. En este tipo de análisis se trata de determinar el grado de homología entre las secuencias. Homología en filogenia se refiere a que dos caracteres han descendido del mismo ancestro. La homología se puede clasificar en diferentes tipos:

- **Ortología.** Es la relación entre OTUs en la que la divergencia acontece tras un evento de especiación. El ancestro común es el cenoancestro. La filogenia recuperada de estas OTUs refleja la filogenia de las especies[23].
- **Paralogía.** Es la condición evolutiva en la que la divergencia observada acontece tras un evento de duplicación genética. La mezcla de ortólogos y paralogos en un mismo análisis filogenético recupera la filogenia correcta de los genes pero no necesariamente la de los organismos o taxas[23].
- **Xenología.** Se refiere a la relación entre OTUs dada por un evento de transferencia horizontal entre linajes. Distorsiona fuertemente la filogenia de las especies[24].

Un árbol filogenético es una estructura matemática usada para representar la historia evolutiva (relación ancestro-descendencia) entre un grupo de OTUs. El patrón de relaciones históricas es la estima hecha de la filogenia o árbol evolutivo [24].

Cuando inferimos la filogenia se deducen las relaciones de ancestría y descendencia entre los OTUs, es decir la topología del árbol. Además de inferir la topología se puede determinar la longitud de las ramas que conectan a los OTUs, lo cual corresponde al número de cambios acumulados entre los OTUs; y en

ocasiones es posible determinar la posición de la raíz, la cual se considera el nodo hipotético más antiguo (**ver figura 2.21**) [24,25].

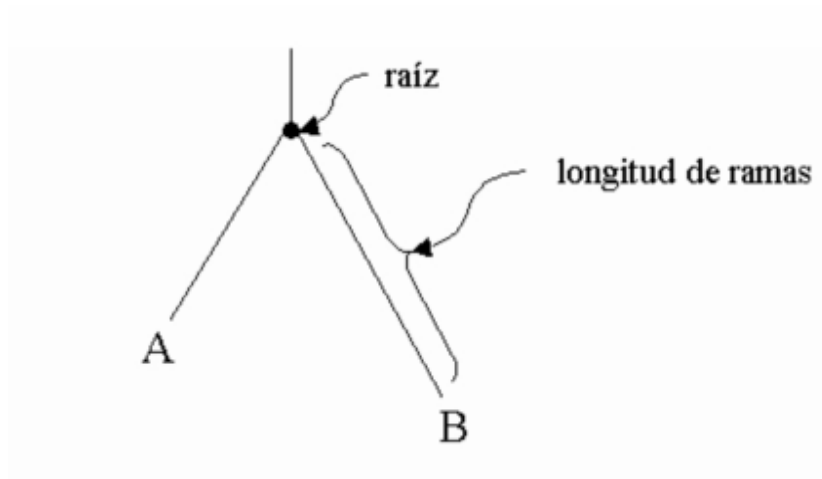


Figura 2.21 Inferencia filogenética . Tres inferencias importantes de un análisis filogenético son la topología, la longitud de las ramas y la posición de la raíz.

Cuando se representa la evolución de un árbol, se sugiere que la evolución ocurrió principalmente mediante ramificación. Los árboles tienen ramas y nodos y en ciertas ocasiones pueden tener raíz. Los nodos pueden ser externos o internos, los cuales representan a los ancestros hipotéticos. La raíz es el nodo hipotético ancestral más antiguo (**ver figura 2.22**) [24,25].

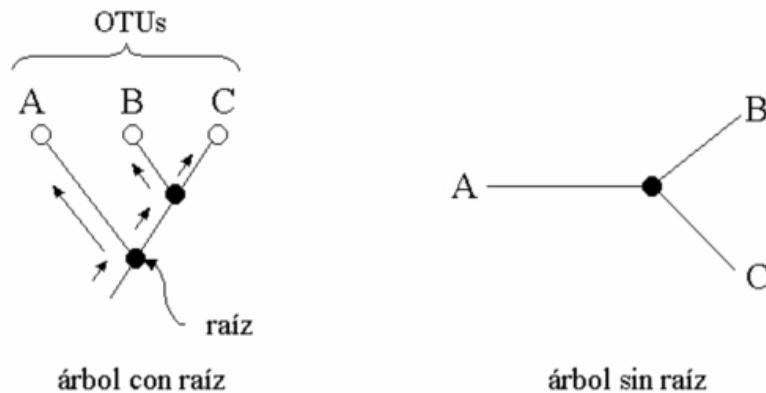


Figura 2.22 Anatomía de árboles con o sin raíz. Los nodos externos (OTUs) se dibujan en blanco, mientras que los nodos internos se muestran en negro. Las flechas en el árbol con raíz indican la dirección del cambio evolutivo [25].

Uno de los métodos más utilizados para enraizar un árbol filogenético es incluyendo en el alineamiento secuencias de algunos grupos externos[24]. Por ejemplo cuando construye un árbol filogenético con algún gen del VIH-1 del grupo M, se puede enraizar incluyendo una secuencia referencia de SIV y secuencias referencia de VIH-1 de los grupos N y O (ver figura 2.23):

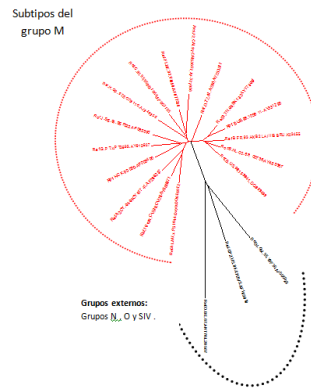


Figura 2.23 Árbol filogenético del gen *env* de VIH utilizando el método Neighbor Joining y el modelo Maximum Composite Likelihood. Debido a que el método Neighbor Joining da árboles filogenéticos desraizados, se enraizó introduciendo grupos externos, en este caso se incluyeron secuencias referencia del gen *env* de VIH-1 de los grupos N y O, así como una secuencia referencia del gen *env* de SIV.

Las ramas de los árboles tienen la propiedad de poder rotarse sobre sí mismas sin afectar a las relaciones entre los OTUs (ver figura 2.24).

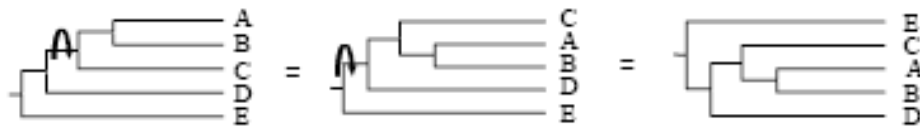


Figura 2.24 Propiedades de las ramas. Las ramas de los árboles pueden girar sin alterar las relaciones filogenéticas entre los OTUs[24].

Existen diversas topologías en los árboles filogenéticos (ver figura 2.25). Los cladogramas son árboles que sólo indican las relaciones de ancestros entre las OTUs. Una topología aditiva contiene la información sobre longitudes de ramas, que refleja la distancia genética de OTUs. Una topología ultramétrica, dendrograma o árbol linearizado, representa un tipo especial de árbol aditivo en el que los nodos

que conectan dichos dominios (**ver figura 2.27**). Estas mutaciones no interrumpen el marco de lectura ya que si lo hicieran inactivarían al gen. Se dice que a mayor número de mutaciones acumuladas existirá mayor distancia evolutiva[24].

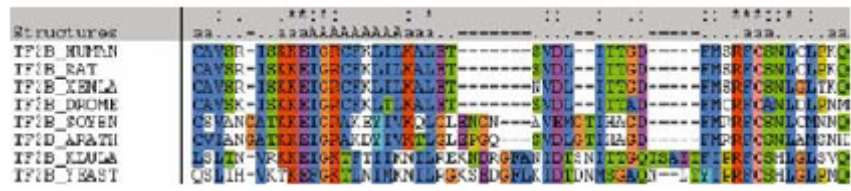


Figura 2.27 Ejemplo de alineamiento que contiene gaps. Los gaps ayudan a conservar las regiones homologas entre las secuencias de un alineamiento múltiple [24].

Existen diversos algoritmos que ayudan a determinar alineamientos múltiples, entre ellos los exhaustivos y los heurísticos o también llamados no exhaustivos. Los algoritmos exhaustivos garantizan determinar el alineamiento múltiple más óptimo. Mientras que los algoritmos heurísticos no garantizan encontrar el alineamiento múltiple más óptimo [24].

Cuando se desea realizar un alineamiento múltiple siguiendo la estrategia de alineamiento progresivo (**ver figura 2.28**) es necesario generar todos los posibles alineamientos pareados, usando los métodos heurísticos o exhaustivos con lo cual se calcula una puntuación con base en la matriz de sustitución. Posteriormente se calcula una matriz de distancia con base en las puntuaciones de los alineamientos pareados. Se estima un árbol de distancias (se realiza con el método Neighbor Joining o UPGMA, ver abajo), que representará de manera aproximada las relaciones entre las secuencias. Finalmente, el alineamiento riguroso y global se realiza siguiendo el orden de similitud indicado por el árbol guía (**ver figura 2.28**).[24]

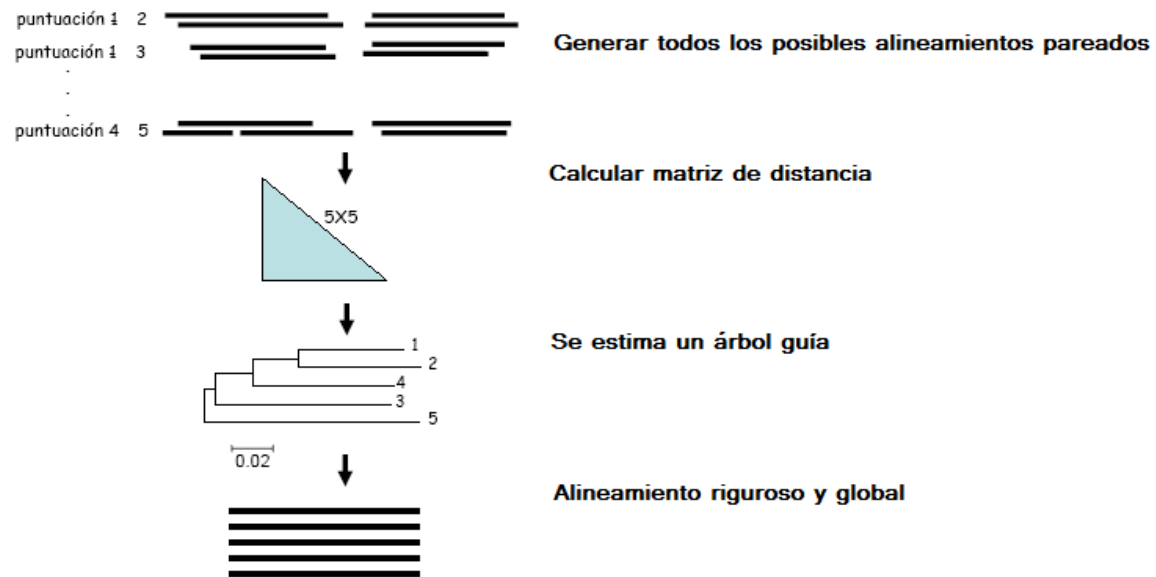


Figura 2.28 Pasos en la generación de un alineamiento múltiple siguiendo la estrategia de alineamiento progresivo [24].

2.10.2 Métodos filogenéticos

Una vez hecho el alineamiento, se procede a hacer un árbol filogenético. La diferencia entre cada método filogenético radica en cómo convierte los datos en forma numérica para poder ser analizados [23].

Los métodos de distancia convierten los alineamientos de secuencias en una matriz de distancia genética con base en el modelo evolutivo seleccionado. La matriz es usada por el método algorítmico de reconstrucción para calcular el árbol. Ejemplos de estos métodos son el Neighbor-Joining y el método no ponderado de par de grupos de medias aritméticas por sus siglas en inglés UPGMA (Unweighted Pair Group Method with Arithmetic Means) [24].

Los métodos discretos como son el de Máxima Parsimonia, Máxima Verosimilitud y los métodos Bayesianos, consideran cada sitio del alineamiento como una función probabilística [23].

2.10.2.1 Métodos de distancia

El método de neighbor-joining [23] es uno de los métodos más utilizados para la creación de árboles filogenéticos con base en matrices de distancia. Este método se basa en el principio de la evolución mínima: una vez definida la matriz de distancias entre secuencias, el principio dice que el mejor árbol es aquel que minimiza la longitud del árbol (la longitud del árbol = \sum longitud de todas las ramas) (ver figura 2.29) [26].

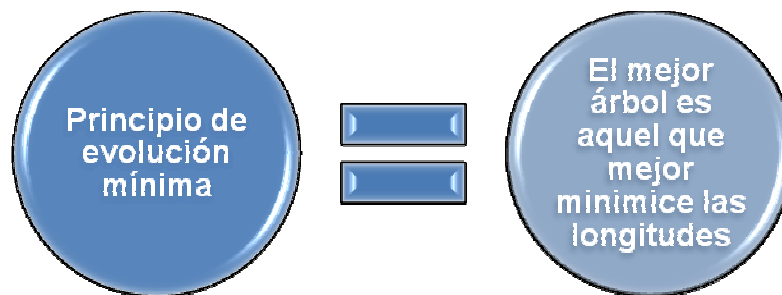


Figura 2.9 Principio de Evolución Mínima. El método Neighbor Joining se basa en el principio de la evolución mínima, el cual establece que el mejor árbol es aquel que minimice las distancias totales.

Para comenzar a construir un árbol filogenético con el método Neighbor Joining, se crea un árbol en forma de estrella en el cual todas las OTUs se encuentran enlazadas por un nodo central. Se toman al azar dos OTUs, se conectan a un segundo nodo interno y se mide la longitud total de la rama formada (ver figura 2.30). Esto se repite una y otra vez hasta que se agotan las posibilidades de combinación de pares. Las ramas que permanecen son aquellas que minimicen sus distancias. Posteriormente se calcula una nueva matriz considerando a las ramas formadas como nuevos OTUs compuestos. El árbol final es aquel que minimice el valor de la suma de sus ramas (ver figura 2.31). El problema que presenta este método es que parte de la información se pierde ya que se reduce a la expresión mínima (en particular en las identidades de los nucleótidos ancestrales y derivados), por tanto no se puede observar un nodo raíz, a menos que se inserte un grupo externo que nos permitirá ubicarnos [26].

$$\begin{aligned}
 S_{12} &= L_{XY} + (L_{1X} + L_{2X}) + \sum_{i=3}^N L_{iY} \\
 &= \frac{1}{2(N-2)} \sum_{k=3}^N (D_{1k} + D_{2k}) + \frac{1}{2} D_{12} + \frac{1}{N-2} \sum_{3 \leq i < j} D_{ij}.
 \end{aligned}$$

Figura 2.30. Ecuación para calcular las distancias entre OTUs [26].

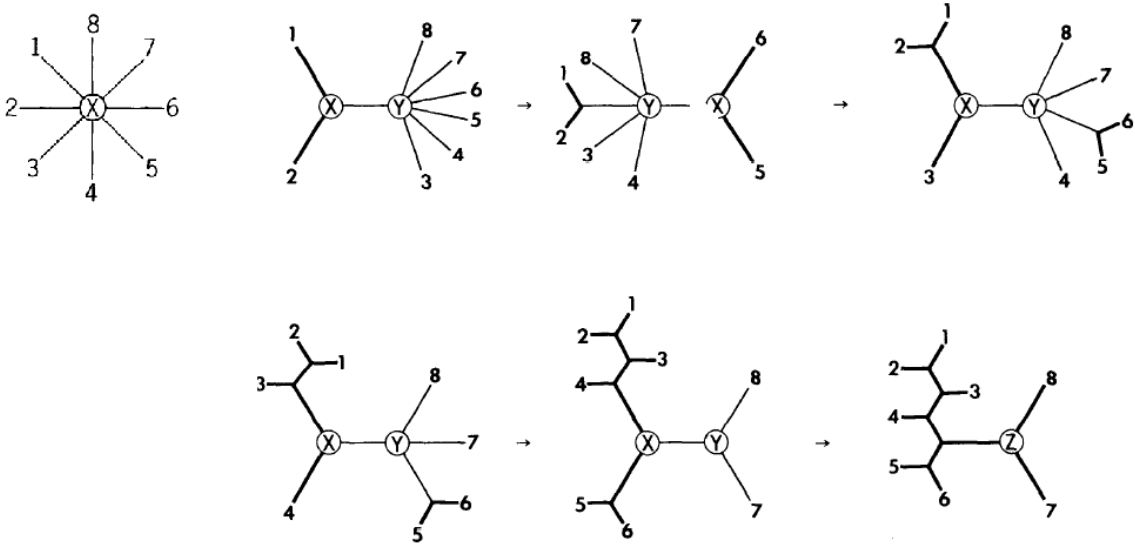


Figura 2.31. Aplicación del método Neighbor Joining. El primer paso es crear un árbol en forma de estrella a partir de todos los OTUs, todos conectados a un nodo central. Se toman dos OTUs al azar y se determina su distancia, esto se realiza hasta agotar todas las posibles combinaciones, las ramas que permanecen son aquellas que tengan menor distancia. Las nuevas ramas se consideran OTUs compuestos, con estos se calcula una nueva matriz de distancia. El árbol final es aquél que minimice mejor la suma de todas las ramas[26].

El método UPGMA es otro de los métodos de distancia. Este método permite construir un árbol ultramérico, lo cual resulta en una topología enraizada. Con el método UPGMA es posible obtener la longitud de las ramas al calcular la distancia y la topología[24].

En el método UPGMA se asume:

- La tasa de evolución para todos los OTUs es la misma, por esta razón la matriz de cambios que se forme debe tener distancias ultraméricas.

- Los nucleótidos son mutuamente independientes. Esta independencia significa que un cambio en un determinado sitio no altera la distribución en los demás.
- Después de que dos especies divergen continúan evolucionando independientemente.

Para poder utilizar el modelo UPGMA, es necesario partir de una matriz de distancias entre OTUs. El primer paso es identificar los OTUs similares entre sí (**figura 2.32**) y determinar cuales tienen menor distancia entre sí. Por ejemplo, si suponemos que los OTUs A y B tienen menor distancia entre sí, éstos se agrupan y se coloca un punto de ramificación a una distancia $d_{AB}/2$. Después de que se agrupan los OTUs A y B se consideran un OTU compuesto y se calcula de nuevo una matriz de distancia entre todos los OTUs agrupando aquéllos OTUs que tengan menor distancia y así sucesivamente hasta que todos los OTUs estén agrupados (**ver figura 2.32**) [25].

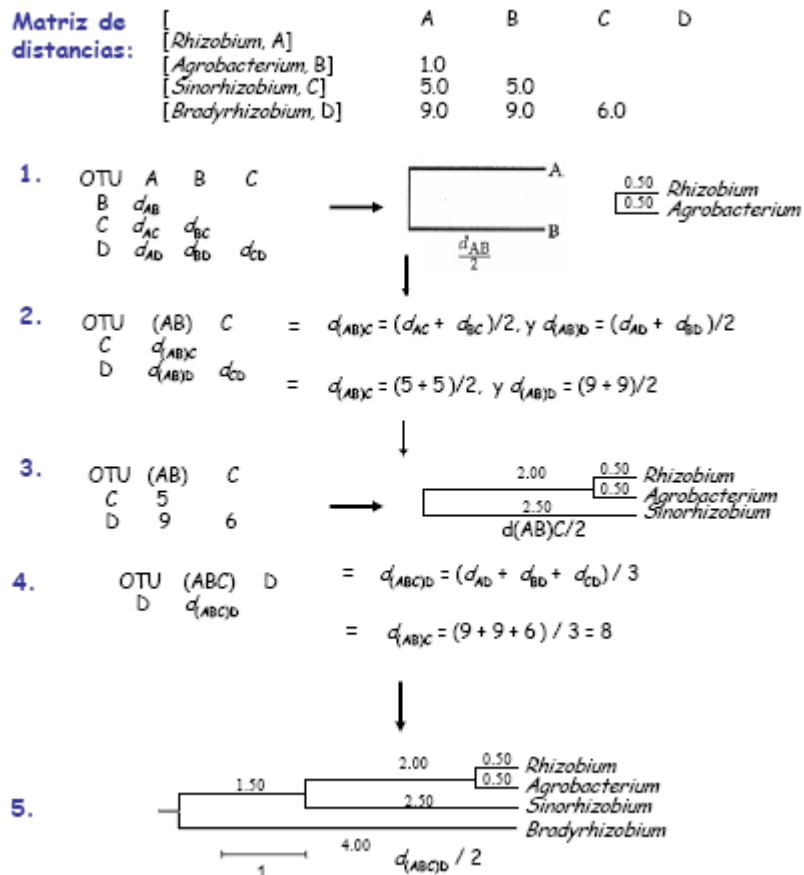


Figura 2.32 Ejemplo de inferencia filogenética por medio del método UPGMA. (1) Se identifican las OTUs similares entre sí y determinar cuales tienen menor distancia entre sí. (2) Las OTUs A y B tienen menor distancia entre sí, éstos se agrupan y se coloca un punto de ramificación a una distancia $d_{AB}/2$. (3) Los OTUs A y B agrupados se consideran un OTU compuesto y se calcula de nuevo una matriz de distancia entre todos los OTUs agrupando aquellos OTUs que tengan menor distancia (4) y así sucesivamente hasta que todos los OTUs estén agrupados (5) [24].

2.10.2.2 Métodos discretos

Los métodos discretos, también llamados métodos basados en caracteres, están basados directamente en el análisis de los caracteres que forman las secuencias. Los métodos discretos cuentan los eventos de mutación acumulados en las secuencias y pueden por lo tanto eliminar la pérdida de información cuando se transforman a distancias. La preservación de la información de los caracteres significa que la dinámica evolutiva se puede estudiar [27].

Los métodos discretos más utilizados son:

- Máxima Parsimonia
- Máxima Verosimilitud

El método Máxima Parsimonia selecciona el árbol que tiene el mínimo número de cambios evolutivos, es decir aquel cuyas ramas tengan en promedio la mínima longitud. Este método está basado en el principio conocido como Navaja de Occam, el cual se basa en una premisa muy simple: en igualdad de condiciones la solución más sencilla es probablemente la más correcta, esto se debe a que requiere el menor número de suposiciones y operaciones lógicas [27].

Siguiendo el principio de parsimonia, el árbol con el menor número de sustituciones es probablemente la mejor opción para explicar las diferencias entre los taxones estudiados. Este principio sólo es válido cuando los cambios evolutivos se dan dentro de un gran lapso de tiempo.

El método de Máxima Verosimilitud emplea modelos probabilísticos para seleccionar el mejor árbol. Este método selecciona aquel árbol que tenga la más alta probabilidad (verosimilitud) de reflejar el proceso evolutivo real [28].

El método máxima verosimilitud es un método exhaustivo que busca todas las posibles topologías y considera el valor de ln Likelihood o de versosimilitud, el cual es igual a la probabilidad de los datos observados dadas una topología particular (τ), set de longitudes de rama (u) y modelo de sustitución (φ). Se trata de encontrar las estimas de los valores de cada parámetro del modelo y luego comparar las verosimilitudes de los distintos modelos, escogiendo la mejor (topología), es decir se trata de seleccionar modelos y parámetros que maximicen la función de optimización[24].

Dicho método calcula la versosimilitud total de todas las secuencias ancestro que evolucionaron en nodos internos y eventualmente a las secuencias existentes. En ocasiones este método incorpora parámetros que consideran las tasas de variación entre sitios [28].

2.10.3 Modelos de evolución

Los modelos de evolución representan una declaración de probabilidad para el cambio de un nucleótido a otro (**ver figura 2.33**). El primer modelo de la evolución desarrollado fue el modelo de Jukes y Cantor, el cual representa los cambios múltiples en un solo sitio con los parámetros de velocidad igual para todos los tipos de cambio. Más tarde, apareció el modelo Kimura-2 parámetros, el cual cuenta las transiciones (cambio de purina a purina o pirimidina a pirimidina) y las transversiones (cambios de purina a pirimidina o viceversa), reflejando las diferencias en las tasas de transición y transversión[29].

El modelo de Tamura- 3 parámetros es una extensión del modelo Kimura- 2 parámetros pero ahora se consideraba la proporción de GC y AT.[17] En el caso del modelo Tajima- Nei se asume que las frecuencias de los nucleótidos son estacionarias, con lo cual se estima el número de sustituciones. Este modelo plantea el uso de cuatro parámetros y le asigna uno a cada nucleótido (**ver figura 2.33**) [30].

Mutaciones

Nucleótido original	A	T	C	G
1. Jukes-Cantor				
A	...	λ	λ	λ
T	λ	...	λ	λ
C	λ	λ	...	λ
G	λ	λ	λ	...
2. Kimura 2 parámetro				
A	...	β	β	α
T	β	...	α	β
C	β	α	...	β
G	α	β	β	...
3. Tamura 3 parámetros				
A	...	$(1-\theta)\beta$	$\theta\beta$	$\theta\alpha$
T	$(1-\theta)\beta$...	$\theta\alpha$	$\theta\beta$
C	$(1-\theta)\beta$	$(1-\theta)\alpha$...	$\theta\beta$
G	$(1-\theta)\alpha$	$(1-\theta)\beta$	$\theta\beta$...
4. Tajima Nei				
A	...	$g_T\lambda$	$g_C\lambda$	$g_G\lambda$
T	$g_A\lambda$...	$g_C\lambda$	$g_G\lambda$
C	$g_A\lambda$	$g_T\lambda$...	$g_G\lambda$
G	$g_A\lambda$	$g_T\lambda$	$g_C\lambda$...

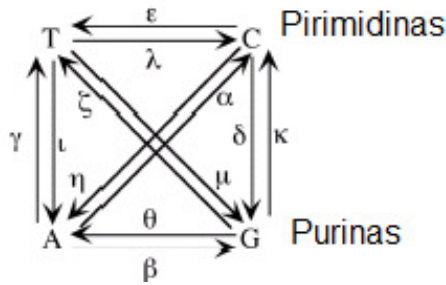


Figura 2.33 Diferentes modelos de evolución . Los modelos de evolución representan una tasa de cambio de un nucleótido a otro. A la izquierda se muestran las matrices de tasa de mutación con los diferentes modelos [29,31].

Modelos posteriores incorporan las diferencias de los índices relativos del modelo como son el tiempo de reversibilidad, las diferencias en las frecuencias de nucleótidos entre pares de bases, la tasa de heterogeneidad entre los sitios, los sitios invariables y la posición del codón.

En la mayoría de los modelos al momento de estimar divergencias entre dos secuencias se asume que éstas han evolucionado siguiendo el mismo patrón de sustitución de nucleótidos después de la divergencia de su ancestro común (homogeneity assumption). Si esta suposición se viola, la distancia evolutiva estimada será parcial, lo cual dará como resultado un árbol filogenético con errores. El modelo de Tamura Kumar (log det) permite calcular las distancias evolutivas entre pares de secuencias que violan esta suposición [32].

Uno de los modelos más recientes es el Maximum Composite Likelihood (2004), el cual es utilizado en la creación de árboles filogenéticos con el método Neighbor Joining, este modelo minimiza en un 60% los errores causados por el método, minimizando el error, puesto que calcula las distancias evolutivas de forma simultánea a diferencia de los demás modelos de nucleótidos. La ventaja que

confiere la construcción de un árbol con el método Neighbor-Joining y el modelo máximo composite likelihood con respecto a utilizar un método más complicado como es Máxima Verosimilitud (ML) es que se utiliza un método más simple como es el Neighbor – Joining y se obtienen resultados similares, sin que el análisis computacional lleve tanto tiempo como podría ser el caso al utilizar ML [33].

Por el momento el modelo más utilizado es Kimura- 2 parámetros tal vez por razones históricas o simplemente porque es el modelo que se encuentra en los paquetes de software más utilizados. A pesar de esto, este modelo ha demostrado ser demasiado simplista [29].

Dada la gran cantidad de modelos a elegir, resulta complicado hacer una elección razonable del modelo adecuado. Podríamos elegir el modelo más complejo que se encuentre disponible sabiendo que por definición los modelos son simplificaciones de la realidad, lo cual nos podría indicar que es el más cercano a ésta comparado con otros modelos menos complejos. El problema con estos modelos es que cuentan con alta complejidad, lo cual requiere un mayor número de parámetros. Todos estos parámetros deben ser estimados a partir de los datos, por tanto uno tiene que tener datos suficientes para calcularlos con precisión. El mejor modelo es aquél más simple que represente con mayor exactitud la realidad. [33] Por lo tanto, una alternativa es el ajuste de los datos utilizando un criterio de máxima verosimilitud, a través de pruebas de verosimilitud, aplicando el criterio de Akaike, o el criterio Bayesiano, los cuales permiten determinar estadísticamente la ganancia relativa en el riesgo de añadir más parámetros del modelo[29].

2.10.4 Evaluación de la exactitud de un árbol

Las pruebas estadísticas que se utilizan para determinar la exactitud de un árbol filogenético son necesariamente complejas porque un árbol es geométrico en lugar de numérico, y la exactitud de una parte de la topología puede ser mayor o menor que la exactitud de las otras partes[23].

Para asignar los límites de confianza a los diferentes puntos de un árbol se lleva a cabo un análisis llamado “bootstrap”. Éste se puede aplicar a cualquier método. En

esta prueba se muestrea la propia muestra repetidas veces, esto puede ser de 100 a 1000 veces. Para interpretar los resultados, se calcula un árbol consenso y en él se indica en cuántas ocasiones se ha obtenido dicha topología. De forma que si en el 100% de las ocasiones las secuencias A y B se han separado de C y D, podemos confiar que las relaciones son correctas. Un valor de “bootstrap” mayor al 70% indica que el resultado es confiable [23] .

3.0 Planteamiento del problema

La información existente de la epidemiología molecular del VIH en México es escasa. Los trabajos que se han publicado indican que el subtipo predominante es el B. A pesar de ello las conclusiones de los estudios publicados son bastante limitadas ya que las cohortes son pequeñas y están enfocados en ciertos grupos de la población [19,20].

En el Centro de Investigación en Enfermedades Infecciosas (CIENI) del Instituto Nacional de Enfermedades Respiratorias (INER), lugar donde se desarrolló el presente proyecto de tesis se encuentra en realización el estudio de epidemiología molecular de VIH más grande hasta la fecha en México. Se han incluido más de 1500 pacientes infectados por VIH, sin tratamiento antirretroviral previo, provenientes de varios de los estados con mayor prevalencia de VIH en el país: Veracruz, Oaxaca, Estado de México, Chiapas, Puebla, D. F., Morelos, Jalisco, Guerrero, Nuevo León y Baja California (Jornadas CIENI enero 2009, datos no publicados). En este estudio se analizaron secuencias del gen viral *pol* y se encontró que el 99.91% de los individuos estaban infectados por virus del subtipo B altamente homogéneos entre sí y un 0.09% por recombinantes BF (Datos no publicados). Se cree que la Epidemia de VIH en México se encuentra altamente influenciada por la epidemia presente en los Estados Unidos debido al fenómeno migratorio. Sin embargo, la variabilidad del virus circulante en E. U. parece ser mayor que la observada en nuestro país. Un estudio reciente demostró que la prevalencia de subtipos no-B en E. U. está en aumento con un porcentaje general de subtipos no-B de 2.93% entre el 2004 y el 2008 y observándose la presencia de subtipos A1, A2, A3, C, D, F1 y G, así como 11 CRFs y 9 URFs diferentes [22.] Estos datos contrastan con el estudio mexicano del CIENI donde solamente 0.09% de los virus circulantes corresponden a subtipos no-B y muestran la importancia de estudiar las características específicas de la epidemiología molecular del virus

en el territorio nacional. Sin embargo, la alta homogeneidad entre los virus circulantes de la República Mexicana observada en el estudio del CIENI del INER, se podría deber a la relativamente baja variabilidad del gen *pol*, a partir del cual se realizó el análisis filogenético. Por tal motivo, en el presente trabajo se realizó un análisis filogenético del gen más variable del VIH, el gen *env*, con el objetivo de estudiar más a fondo las relaciones filogenéticas de los virus circulantes en el país y confirmar observaciones previas que muestran una alta homogeneidad entre ellos. Para realizar este análisis se obtuvieron las secuencias del gen *env* a partir del banco de sueros del proyecto de epidemiología molecular del CIENI del INER. Los resultados de este estudio serán valiosos para apoyar las conclusiones del estudio de epidemiología molecular del VIH en México, con aplicación en el diseño de posibles vacunas efectivas para nuestra población y la generación de nuevas políticas de salud pública que contribuyan al manejo de la epidemia en el país.

4.0 Hipótesis

Dada la alta variabilidad del gen viral *env*, estudios filogenéticos basados en secuencias de este gen de VIH circulante en diferentes regiones de la República Mexicana, nos permitirán observar posibles agrupamientos geográficos y clusters de diferenciación que aporten información sobre la historia epidemiológica del VIH/SIDA en México.

5.0 Objetivos

5.1 Objetivo general

- Desarrollar un estudio filogenético comparativo de VIH-1 circulante en diferentes regiones de la República Mexicana utilizando el gen *env*.

5.2 Objetivos particulares

- Extraer RNA viral a partir de virus circulantes en muestras de plasma de pacientes VIH positivos de diferentes regiones de la República Mexicana.
- Amplificar el fragmento C2-V5 del gen viral *env*, de los virus provenientes de pacientes infectados por VIH-1 de las diferentes regiones de la República Mexicana, para posteriormente ser secuenciados y analizados.
- Generar un banco de secuencias del gen viral *env* de virus circulantes en México.
- Establecer las relaciones filogenéticas entre los virus circulantes en el país, comparándolas con virus provenientes de países vecinos.
- Establecer los subtipos virales circulantes en el país, a través de análisis filogenéticos del gen viral *env*.

6.0 Metodología

6.1 Selección de Pacientes

Se reunió una cohorte de 101 individuos infectados por VIH, sin tratamiento antirretroviral previo, provenientes del Norte, Centro y Sur de la República Mexicana, incluyendo los siguientes estados: Sonora, Baja California Norte, Nuevo León, Jalisco, D. F., Estado de México, Morelos, Chiapas, Oaxaca, Veracruz y Guerrero. Los individuos participantes fueron seleccionados de la cohorte de Epidemiología Molecular y Evolución del VIH en México del CIENI del INER, buscando una representación equitativa de individuos de las regiones Norte, Centro y Sur del país. Para cada individuo seleccionado, se obtuvieron muestras de plasma del banco de muestras sanguíneas del CIENI del INER, las cuales fueron tomadas de febrero 2008 a febrero 2009.

El estudio de Epidemiología Molecular del VIH en México del CIENI del INER fue evaluado y aprobado por el Comité de Ética Institucional. Todos los pacientes que participaron en este estudio firmaron previamente una carta de consentimiento informado que incluye los detalles del estudio como objetivo, duración, procedimientos, riesgos, molestias, beneficios potenciales y la información de contacto de los responsables del estudio. En todo momento se respetó la confidencialidad de los individuos participantes mediante el uso de claves asociadas a las muestras biológicas empleadas en el presente estudio.

6.2 Aislamiento de ARN viral

El RNA viral se aisló por medio del estuche de reactivos “**QIAamp Viral RNA Mini Kit**” (Qiagen, Valencia, CA) de acuerdo a las especificaciones del fabricante [35]. A partir de 1 ml de plasma se concentraron las partículas de VIH por medio de centrifugación a 17000 rpm, por 2 horas a 4°C hasta obtener un volumen final de 140 µl de plasma. Este concentrado se lisó por medio de un buffer de lisis que

libera el ARN genómico del virus, el cual se hizo pasar por una columna de sílica que retiene el ARN viral. Se eliminaron los restos de plasma por medio de buffers de lavados y finalmente se cambió la fuerza iónica de la matriz de la columna con un buffer de elución dando como resultado la liberación del RNA viral purificado (**ver figura 6.1**) El RNA purificado se almacenó a -80°C hasta su uso.

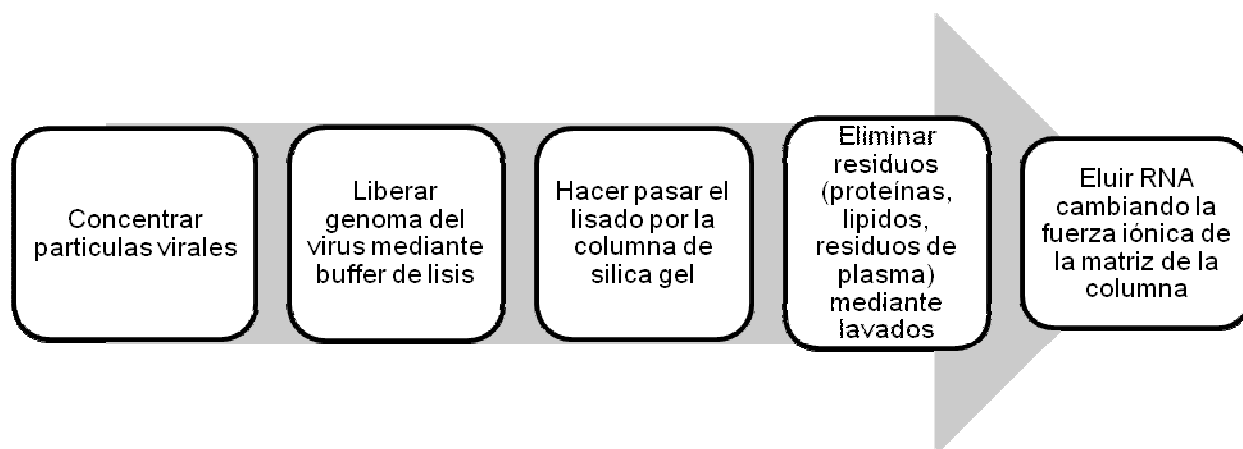


Figura 6.1 Aislamiento de RNA viral. Se muestran los pasos generales para el aislamiento de ARN viral a partir de muestras de plasma utilizando el estuche de reactivos “QIAamp Viral RNA Mini Kit” (Qiagen, Valencia, CA).

6.3 RT-PCR anidada

Se amplificó la región C2-V5 del gen *env* viral, mediante reacción en cadena de la polimerasa (PCR) anidada, utilizando como primers ED3/ED12 y S7/ES8 (**ver figura 6.2**) en la primera y segunda vuelta de amplificación respectivamente. Para la primera vuelta de amplificación se utilizó el estuche de reactivos “**SuperScript™ One-Step RT-PCR with Platinum® Taq**” (Invitrogen, Brown Deer, WI) de acuerdo a las especificaciones del fabricante [36]. Las condiciones de la reacción de amplificación fueron: Una concentración final de 0.4 mM de cada primer, 1.4 mM como concentración final de MgSO_4 , una concentración final de dNTPs de 200 mM (éstos se encuentran en el Mix de reacción 2x), agregando 0.5 μl de RT/ Platinum® Taq Mix” y 5 μl del RNA proveniente del virus en plasma de cada uno de individuos del estudio en un volumen final de reacción de 25 μl . Esta mezcla de reacción fue sometida a: 1 ciclo de 65°C , 30s, y 45°C 35 min; seguido

de 1 ciclo de 94°C 2 min; 3 ciclos de 94°C 1 min , 55°C 1 min y 72°C 1 min; 32 ciclos de 94°C 15s, 55°C 45s y 72°C 1 min y 1 ciclo de 72°C 5 min y 4°C ∞ (**ver figura 6.3**).

a) Primers primera vuelta

ED3 5´-TTAGGCATCTCCTATGGCAGGAAGAAGCGG (5956-5985)

ED12 5´-AGTGCTTCCTGCTGCTCCCAAGAACCCAAG (7822-7792)

b) Primers segunda vuelta

ES7 5´-tgtaaacgacggccagtCTGTTAAATGGCAGTCAGC(7001-7020)

ES8 5´-caggaaacagctatgaccCACTTCTCCAATTGTCCCTCA(7667-7647)

Figura 6.2 Primers de amplificación. a) Primes utilizados en la primera vuelta. b) Primers utilizados en la segunda vuelta de la RT-PCR anidada [37]

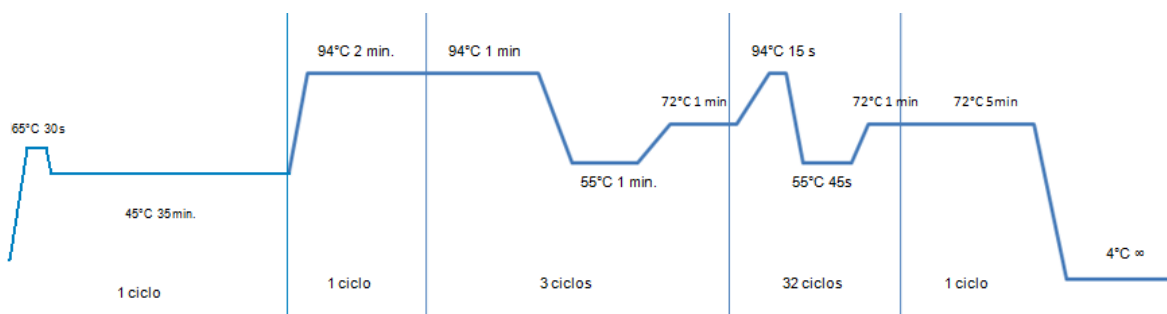


Figura 6.3 Condiciones para la RT-PCR de la primera vuelta de amplificación. Se muestra el programa utilizado para la primera vuelta de amplificación del gen *env* a partir de RNA de virus en plasma utilizando los primers externos ED3 y ED12..

En la segunda vuelta de amplificación se utilizó el estuche de reactivos “**Platinum® Taq DNA Polymerase**” (Invitrogen, Brown Deer, WI), de acuerdo a las especificaciones del fabricante [38]. Las condiciones de reacción fueron: Una

concentración final de 0.4 mM de cada primer (ES7/ES8), una concentración final de 1.6 mM de MgSO₄, una concentración final de 200 mM dNTPs, una unidad de Taq DNA polimerasa y 2.5 µl de los amplicones obtenidos en la primera vuelta de amplificación en un volumen final de 25µl, a un pH de 8. Esta mezcla de reacción fue sometida a: 1 ciclo de 2 min 94°C; 3 ciclos de 94°C 1 min, 55°C 1 min y 72°C 1 min; 32 ciclos de 94°C 15s, 55°C 45s y 72°C 1 min y 1 ciclo 72°C 5 min y 4°C ∞ (ver figura 6.4).

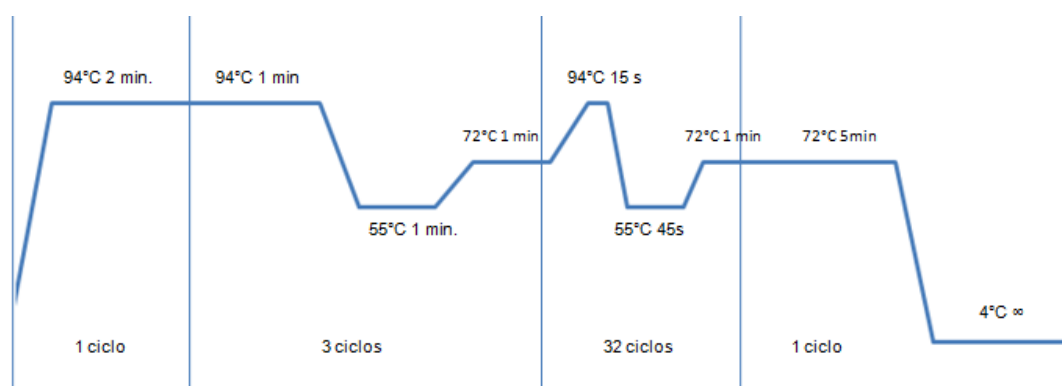


Figura 6.4 Condiciones para la PCR de la segunda vuelta de amplificación. Se muestra el programa utilizado para la segunda vuelta de amplificación del gen *env* a partir de RNA de virus en plasma utilizando los primers internos ES7 y ES8.

El producto de la primera reacción originó un fragmento de 1981 pb el cual se utilizó para realizar la segunda reacción de PCR la cual produjo un producto de 700 pb (ver figura 6.5). El producto final de amplificación se identificó por medio de una electroforesis en gel de agarosa al 1.5% con bromuro de etidio (ver figura 6.6). En todos los casos se incluyó un control de contaminación agregando agua en lugar de ARN viral a la reacción.

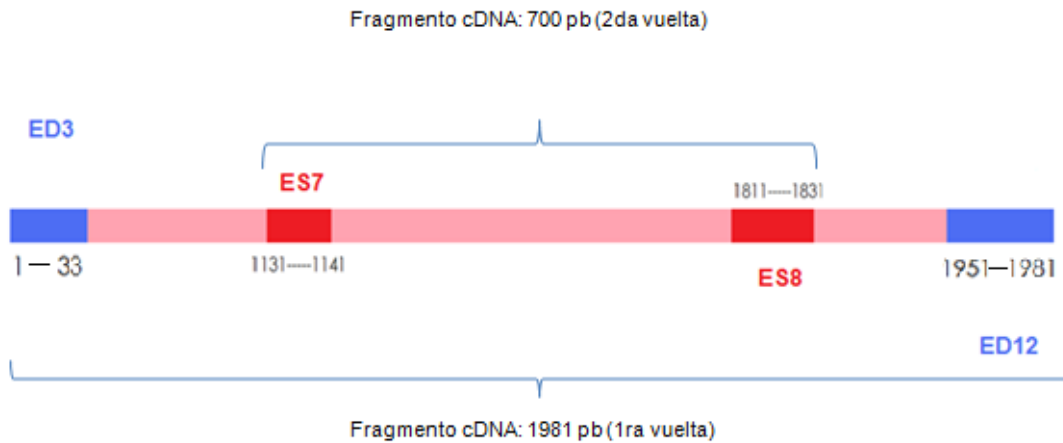


Figura 6.5 Esquema de la amplificación anidada de la región C2-V3 del gen *env*. Los primers externos fueron ED3 y ED12 (marcados en azul) y los primers internos fueron ES7 y ES8 (marcados en rojo).

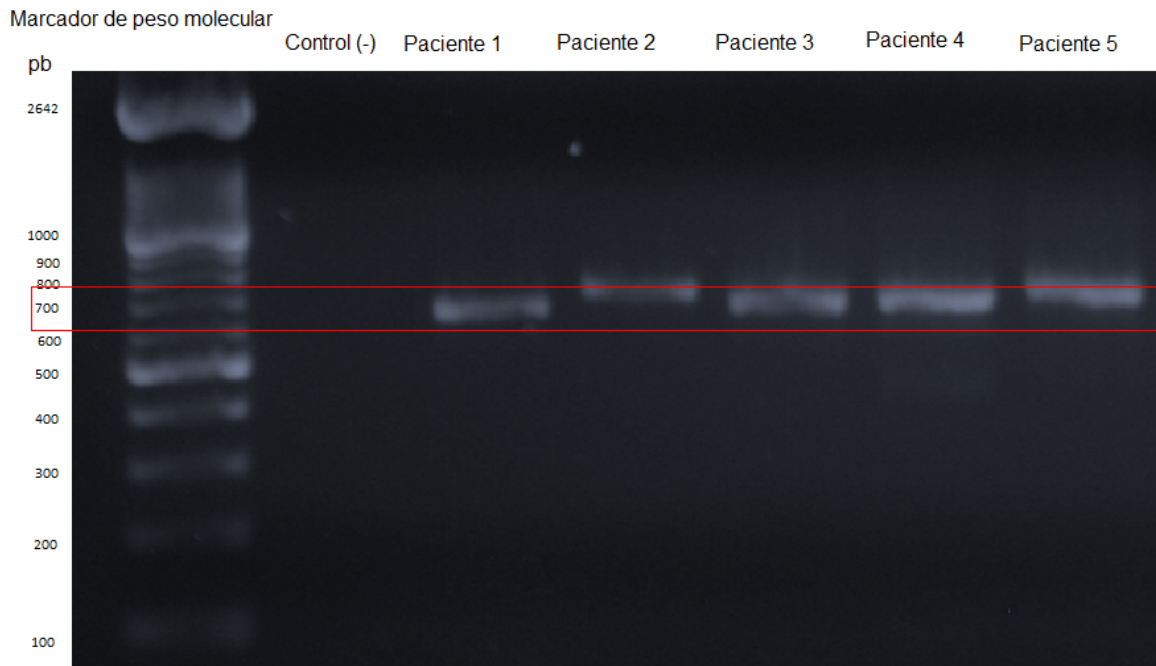


Figura 6.6 Amplificación de fragmento de la región C2-V5 del gen *env* (700pb). Gel de agarosa al 1.5% teñido con bromuro de etidio representativo de la amplificación de un fragmento de 700 pb abarcando la región C2-V5 del gen *env* viral. La primera columna muestra el marcador de peso molecular, la segunda corresponde al control negativo, y las columnas 3 a 7 corresponden a muestras de pacientes.

6.4 Purificación de productos de PCR

Los productos de PCR se purificaron por medio del estuche de reactivos “**QIAquick PCR Purification Kit**” (Qiagen, Valencia, CA), de acuerdo a las especificaciones del fabricante [39]. Este kit consta de columnas de sílica gel las cuales retienen al DNA (100pb-10kb), dejando pasar por la columna todo aquello que no sea DNA (sales, buffers, enzimas etc.). Para eliminar las trazas (restos de sales, buffers, enzimas de la reacción de PCR) se realizó un lavado con un buffer de fuerza iónica alta y finalmente se eluyó el DNA que se encontraba capturado en la columna con un buffer de fuerza iónica baja. Los productos de amplificación se almacenaron a -20°C hasta su uso (**ver figura 6.7**).

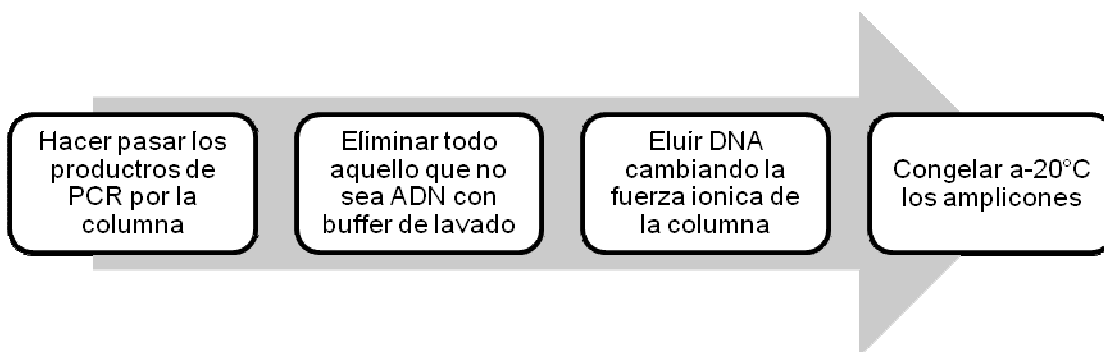
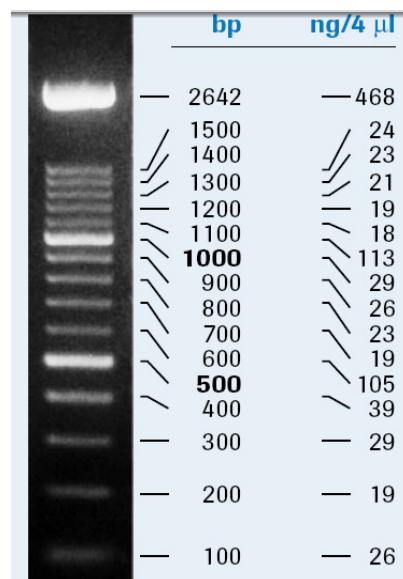


Figura 6.7 Purificación de productos de PCR. Diagrama que ilustra el proceso de purificación de productos de amplificación con el estuche de reactivos “QIAquick PCR Purification Kit” (Qiagen, Valencia, CA).

6.5 Semicuantificación de productos de PCR

Los productos de PCR se semicuantificaron en el gel de agarosa, comparando la intensidad de la banda del producto contra la intensidad de las bandas del marcador (ver tabla 6.1). Al tener las concentraciones se realizaron diluciones para obtener una concentración aproximada de $2.5 \text{ ng DNA}/\mu\text{l}$. Las diluciones de producto de PCR se utilizaron en las reacciones de PCR de secuenciación.

Tabla 6.1 Nanogramos de DNA en cada banda del marcador de peso molecular XIV de Roche (*Molecular Biochemicals*)[40]



6.6 Secuenciación

La secuenciación se basa en el empleo de los cuatro didesoxirribonucleótidos (ddNTPs) marcados con fluoróforos específicos, los cuales carecen del grupo hidroxilo de la posición 3' de la molécula, de manera que cuando uno de estos nucleótidos se incorpora a una cadena de DNA en crecimiento, esta cadena no puede continuar elongándose ya que la DNA polimerasa necesita un extremo 3' OH para añadir el siguiente nucleótido. El secuenciador realiza una electroforesis capilar con lo que acomoda los fragmentos obtenidos en la reacción de secuenciación con los ddNTPs por peso molecular. Los productos de reacción generados en la PCR de secuenciación salen uno a uno del capilar, ordenados por peso molecular, emitiendo una señal diferente que corresponde al fluoróforo del ddNTP terminador de la reacción de secuencia. Esta señal es producida al excitarse el fluoróforo del ddNTP terminal de cada fragmento mediante un láser, para ser captada por un sistema óptico en el equipo. Una computadora interpreta la señal óptica y la secuencia es finalmente mostrada como un electroferograma.

6.5.1 PCR de secuenciación

La reacción de secuenciación se realizó utilizando el reactivo “BigDye® Terminator v1.1 Cycle Sequencing Kit” de Applied Biosystems (Foster City, CA) de acuerdo a las recomendaciones del fabricante [41]. Este reactivo contiene a cada uno de los didesoxirribonucleótidos trifosfatados marcados con un fluoróforo distinto, los cuales funcionan como terminadores de la extensión de la cadena en la reacción de secuenciación. La técnica se realizó de acuerdo a la tabla 6.2, con un volumen final de reacción de 10µl.

Los primers que se utilizaron para la reacción fueron ES7 (primer forward) y ES8 (primer reverse), y como opción alternativa a ES8 se utilizó ED33 en los casos de secuencias problemáticas en las que la variabilidad de la región cercana al sitio de unión de ES8 impidió ver una señal limpia (**ver figura 6.8**). La concentración final de los primers en la reacción fue de 0.4mM (**ver tabla 6.2**).



★ ED33 solo se utilizó en las repeticiones de la secuenciación

Figura 6.8 Diagrama que ilustra los primers de secuenciación. ES7 se usó como primer forward o sentido, mientras que ES8 se usó como primer reverse o antisentido. ED33 se usó como primer reverse alternativo en secuencias problemáticas debido a la variabilidad del gen *env*.

Tabla 6.2 Reacción de secuenciación.

Reacción de secuenciación (1x)	Volumen (μl)
"Terminator Ready Reaction Mix"	4μl
Producto de reacción de PCR purificado (dilución con una concentración aproximada a 2.5ngDNA/μL).	2μl
Primer (sentido o antisentido) (2Mm)	2μl
Agua desionizada	2μl
Volumen total	10μl

La reacción de secuenciación se sometió a las siguientes condiciones en el termociclador: 1 ciclo de 5 min a 95 °C; 30 Ciclos de 30 s a 95 °C, 10 s a 50 °C y 4 min a 60 °C; 1 ciclo de 4 °C a tiempo ∞ (ver figura 6.9).

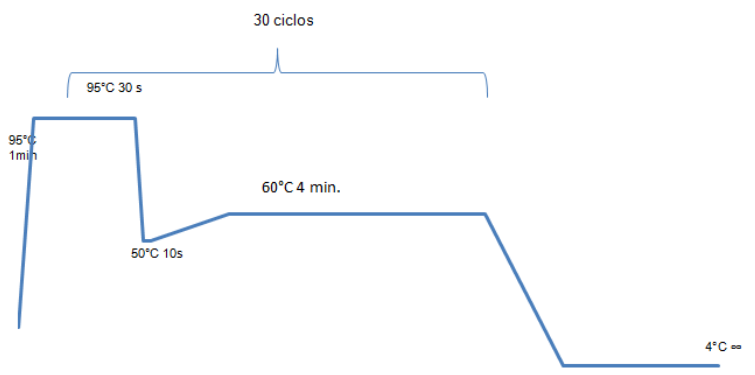


Figura.6.9 Condiciones para la PCR de secuenciación. Programa usado para la reacción de secuenciación de los productos de amplificación del gen *env* empleando el reactivo “BigDye® Terminator v1.1 Cycle Sequencing Kit” (Applied Biosystems, Foster City, CA).

6.5.2 Purificación de los productos de la reacción de secuenciación

Los productos de la reacción de secuenciación se precipitaron con etanol absoluto (2.5 volúmenes) y acetato de sodio 3 Mm (10%v/v), seguido de una centrifugación a 2000 x g por 30 min. El etanol remanente se eliminó volteando la placa y centrifugando de nueva cuenta a 150 x g por 1 minuto. Posteriormente se realizó un lavado con etanol al 80% centrifugando por 5 minutos a 2000 x g. Al término de la centrifugación se volteó la placa y se centrifugó a 150 x g por 3 minutos para eliminar el etanol remanente. Se repitió este último paso para eliminar completamente el etanol. Ya purificados los productos de la reacción de secuenciación, éstos se disolvieron en 20 µl de formamida (ver figura 6.10).

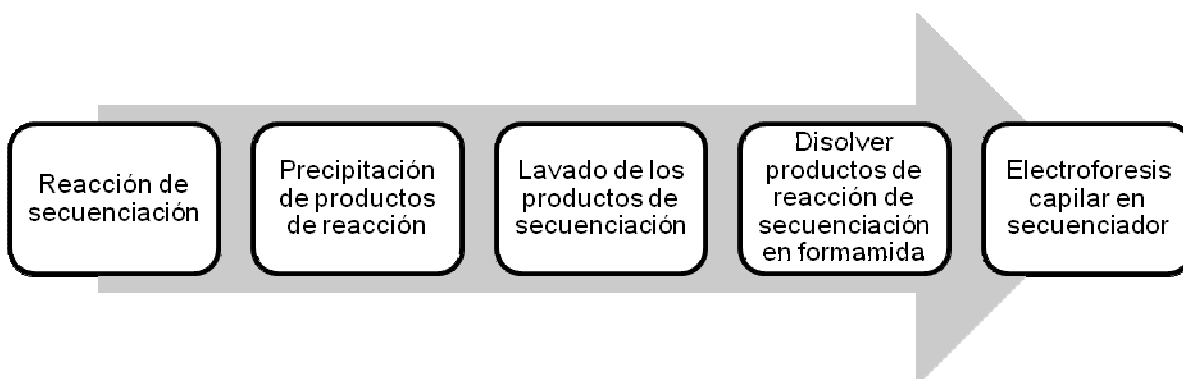


Figura.6.10 Secuenciación. Se muestran los pasos generales del proceso de secuenciación del gen viral *env*.

6.5.3 Electroforesis capilar

Las secuencias del gen *env* se obtuvieron por electroforesis capilar en un secuenciador ABI-Prism 3100 Avant Genetic Analyzer (Applied Biosystems, Foster City, CA) [42]. Los datos fueron recolectados por el software ViroSeq. Las condiciones a las cuales se llevaron la electroforesis capilar se muestran en la **tabla 6.3**.

Tabla 6.3 Condiciones de electroforesis capilar

Temperatura	50°C
Tiempo de inyección de la muestra	22 segundos
Voltaje de inyección	1 kV
Tiempo de corrida	6500s
Voltaje de corrida	12.2kV
Grosor del capilar	50mm

El polímero que se utilizó fue POP-6 (Applied Biosystems, Foster City, CA), el tiempo de inyección de la muestra fue de 15s, el voltaje fue arriba de 20 kV y el tiempo de corrida fue de 2 horas.

6.6 Análisis filogenético

6.6.1 Alineamiento de secuencias

Las secuencias forward y reverse se ensamblaron y editaron en el programa **Geneious pro 4.6.4** con el cual se obtuvo una secuencia consenso de la región C2-V5 del gen *env* para cada individuo participante en el estudio. Las secuencias de todos los individuos de la cohorte fueron alineadas con el programa ClustalW incluido en el programa **Mega 4.1 B** incluyendo todas las secuencias de referencia de VIH-1 y SIV_{CPZ} obtenidas de la base de datos de los Alamos [43]. Una vez obtenido el alineamiento y editado este de manera manual, se construyeron árboles filogenéticos con el programa **Mega 4.1 B** (ver figura 6.11).

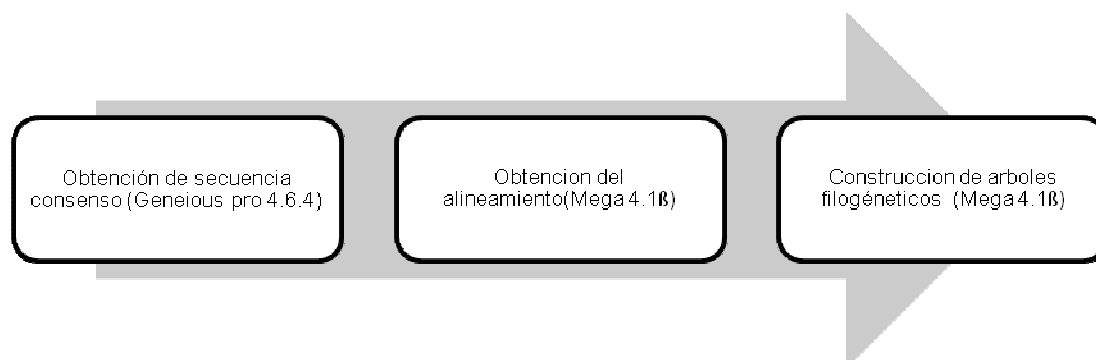


Figura.6.11 Análisis filogenético. Se muestran los pasos generales a través de los cuales se realizó el análisis filogenético de las secuencias del gen *env* de los individuos de la cohorte mexicana, incluyendo los programas computacionales utilizados en el proceso.

6.6.2 Construcción de árboles filogenéticos

Se construyeron árboles filogenéticos basados en matrices de distancias para las secuencias mexicanas a partir de un fragmento de de 700 pb de la región C2-V5 del gen *env* del VIH, utilizando el programa MEGA 4.1β. Se empleó el método de Neighbor-Joining (NJ). NJ es un método algorítmico que proporciona una buena aproximación heurística para encontrar el árbol de evolución mínima más corto en distancia genética. Este método secuencialmente encuentra vecinos que

minimizan la longitud total del árbol. Las distancias evolutivas se calcularon con diversos modelos de nucleótidos:

- Jukes Cantor
- Kimura 2 parámetros
- Tamura 3 parámetros
- Tajima Nei
- Tamura Kumar (Log Det)
- Maximum Composite Likelihood

En todos los casos se utilizó la opción de **pairwise deletion**, es decir las inserciones o deleciones que se encontraban en el alineamiento fueron eliminadas al comparar con las demás secuencias. Para evaluar el porcentaje de réplica de cada árbol se aplicó la prueba de bootstrap, realizando 1000 pseudorréplicas de muestreo aleatorio para estimar el árbol consenso (ver figura 6.12).

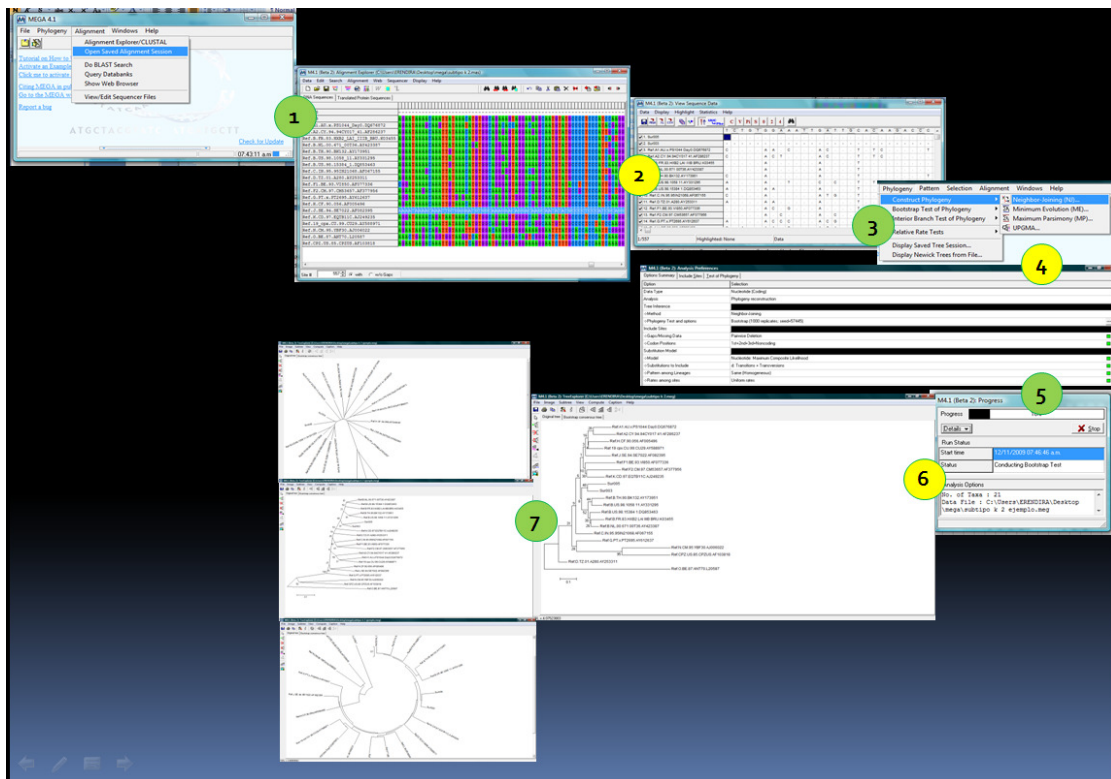


Figura 6.12 Construcción de Árboles filogenéticos con Mega 4.1β .Se construyó un alineamiento con secuencias *env* de VIH circulante en diversas regiones de la República

Mexicana (1), y éste se convirtió de formato MAS a MEG (2). El método que se utilizó en todos los casos fue Neighbor Joining (3) y se seleccionó uno de los modelos de nucleótidos existentes en el programa (Jukes Cantor, Kimura 2 parametros, Tamura 3 parámetros, Tajima Nei, Tamura Kumar (Log Det) y Maximum Composite Likelihood). La resolución del árbol filogenético construido se probó mediante la prueba de bootstrap con 1000 repeticiones (4). Se analizaron los datos (5) y finalmente se obtuvo el árbol filogenético (6). Este árbol puede tomar diversas topologías(7).

6.7 Subtipificación

La subtipificación de VIH basada en secuencias del gen *env* se realizó mediante dos métodos independientes disponibles en línea que se describirán en las secciones siguientes.

6.7.1 REGA

REGA es una herramienta para la subtipificación de secuencias de VIH-1 que se encuentra en la Base de Datos de la Universidad de Standford [44]. La subtipificación se llevó a cabo por medio de métodos filogenéticos, en un proceso que involucra cuatro pasos:

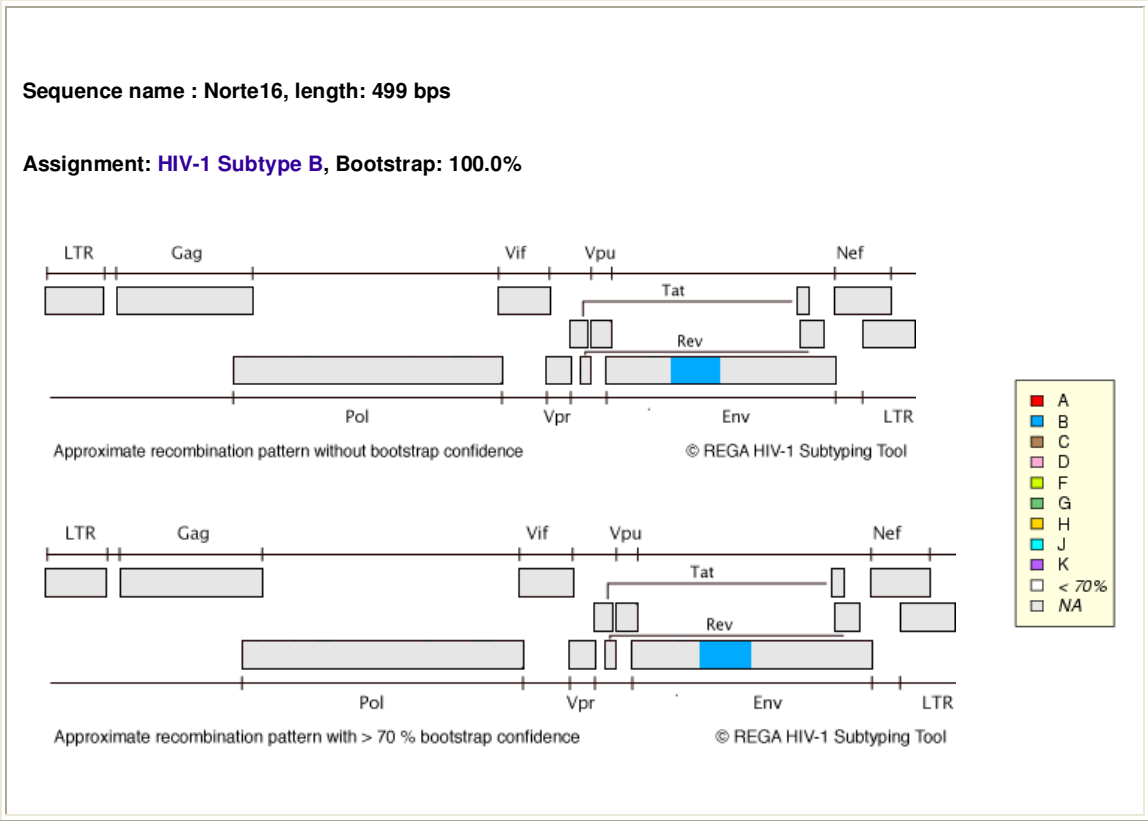
- Alineación de las secuencias con el programa Clustal W.
- Construcción de árboles filogenéticos a partir del método HKY con el software PAUP. El alineamiento contiene la secuencia que se desea subtipificar y secuencias referencia de todos los subtipos del grupo M (**ver tabla 6.4**).
- Construcción de un segundo árbol con la secuencia que se desea subtipificar y las secuencias referencia de los subtipos de VIH y CRFs.
- La secuencia que se desea subtipificar se divide a la mitad y se analiza con secuencias de recombinantes.
- Finalmente se analiza la secuencia con el software Treepuzzle para encontrar similitudes con los diferentes subtipos.

Tabla 6.4 Referencias utilizadas en la subtipificación de secuencias de VIH-1

Subtype or CRFs	
A1	A1_UG85U455 A1_UG9292UG037
A2	A2_CD9797CDKTB48 A2_CY9494CY01741
B	B_NL0067100T36 B_US90WEAU160
C	C_BR9292BR025 C_BW9696BW0502
D	D_CM0101CM4412HAL D_TZ01A280
F1	F1_BE93V1850 F1_FR96MP411
F2	F2_CM195MP255 F2_CM95MP257
G	G_FT93HH8793121 G_NG9292NG083
H	H_BE93V1991 H_CF9090CF056
J	J_SE93SE7887 J_SE94SE7022
K	K_CD97EQTB11C K_CM96MP535
CRF01_AE	01_AECF19090CF11697 01_AETH90CM240
CRF02_AG	02_AGCM9797CMP807 02_AGFR91DJ264
CRF03_AB	03_ABRU97KAL153 03_ABRU98RU98001
CRF04_CPX	04_CPXG1R9197PVCH

La secuencia que se desea subtipificar se introduce en formato FASTA, pudiéndose analizar hasta 1000 secuencias a la vez. El análisis se muestra en una tabla la cual contiene el nombre de la secuencia, el tamaño el subtipo asignado por el programa y un diagrama en el cual se ilustra que región se subtipificó (**ver figura6.13**).

a)



b)

Name	Length	Report	Assignment	Support	Genome
Norte16	499bp	Report	HIV-1 Subtype B	100.0	

Figura 6.13. Análisis de subtipificación de VIH con REGA. a) Esquema que muestra el subtipo de la secuencia de VIH-1. b) En el reporte aparece el nombre de la secuencia, el tamaño, el subtipo asignado, el porcentaje que soporta el resultado y un esquema de la región subtipificada.

6.7.2 RIP

RIP (Recombinant Identification Program), disponible en la base de datos de los Álamos [45], es una herramienta de subtipificación e identificación de eventos de recombinación basada en métodos filogenéticos. Se realiza el alineamiento incluyendo la secuencia que se desea subtipificar y alrededor de 400 secuencias de referencia de todos los subtipos del grupo M y recombinantes.

Cada secuencia que se desea subtipificar se alinea con secuencias de referencia de todos los subtipos y formas recombinantes por medio del programa HMMRE, el cual genera un alineamiento de forma automática.

A continuación se muestra un ejemplo del alineamiento obtenido en la subtipificación:

```

1: 60 [ 1: 60]
query: 594      : TATCTGAAACTGTAAACATTACTTGTACAAGACCCAACAACAATACAAGAAAAGGGATAC
a : CON_A1      : -TA-C--GC-----A---A-----C----T-----A-T----
b : A2.CY.94CY017_41 : -TA--A-GC----CTA-----TC-----A-T----
c : CON_B       : -GAA---T----G-A---A-----A-T---
d : CON_C       : -TAA---T----G-A--GTG-----T-----A-T--A
e : CON_D       : -TAA--GT----MA---A--C---G--T-----C--A-T---
f : CON_F1      : -TAA---T----C-G---A-----A-T---
g : CON_F2      : -TAA-AG-T----HA---A-----G-A-T---
h : CON_G       : -TAA-----A-G-A---A---TC-----A-T--A
i : CON_H       : -TAA-AV-C----V-A---C---C---T---T---G----A-T---
j : J.SE.94.SE7022 : -TAA--C----G-A--GTG----C---T--T-----T----
k : K.CD.97.EQTB11C : -TAA---G---G-A---A-----G-A-G-----A-T----
l : CON_01_AE   : -TAA-A-T----G-A--C-A----C-----TC-----C-A-T--A
      best match      :
      significant     :

```

```

61: 120 [ 61: 120]
query: 594      : ATATAGGACCAGGGAGAACATTGTATACAGGAATAATAGGAGATATAAGACAAGCACATT
a : CON_A1      : G-----ACA-G---C-----T-----G-----Y---
b : A2.CY.94CY017_41 : GCT-T-----ACA-G-C--C---AAT-----G-C-----
c : CON_B       : -----G---T-----
d : CON_C       : GG-----ACA----C-----C-----
e : CON_D       : -----CA-G--C-C---AA-----
f : CON_F1      : --T-----AC-G---T-----C--C--A-G-----
g : CON_F2      : G-----AC-G---C-----T-----A-G--T---
h : CON_G       : SAW-C-----ACA-G-G--C-----T-----C-----

```

```

i : CON_H      : --T---G---ACA-G---C-----T--C-----
j : J.SE.94.SE7022 : -C-G-----ACA-GTGC-C-C---G-----GA----T---
k : K.CD.97.EQTB11C : -----A--G---C-----T-----G--G--G---
l : CON_01_AE   : C-----ACA-GT---C-----A----T---
  best match   :
  significant  :

```

121: 180 [121: 180]

```

query: 594      : GTAATMTTAGTAGAMCAGCATGGAACAATACTTTAAAACAGGTAGCTGGAAAATTACTAA
a : CON_A1      : ----G-C-----T--A----T--A----C--A-----AA-C----AG--
b : A2.CY.94CY017_41 : ----A-C-AC-A-A--TT-----TG-C-----C--A-----A-C----AG-G
c : CON_B       : ---CA-----G--AA-----T--C-----A--T-AA-----AG-G
d : CON_C       : ---CA----GA-GATAA-----T--A----C--A----AG-AA-----GA-G
e : CON_D       : ---CA----G--G--A----T--A----C-----AA-----GG-G
f : CON_F1      : ---CG----G--A--CA-----T--A--W---G---R--AAG-C--G--AAGT
g : CON_F2      : ---CA--A-G--A--CTG----TG-A-----A-D-----AV-G-G--CAA--
h : CON_G       : ----G-----AY-AA-----T--G-TGW---GA-T--CA-W-C-M--C--AAR-
i : CON_H       : ----A----G--AA-VA-----T--G-----C-C-----T-AC-C---GG--
j : J.SE.94.SE7022 : ---CA-----AA-AT----T--C-----CGTAGA-----AA---C-AG-G
k : K.CD.97.EQTB11C : ---CA---CG--GG-CAG----T--A--G---C--A---AAGAA-G----GG--
l : CON_01_AE   : --G-GA--A-G--A--AA-----TG-A-----A--A-----AA-G
  best match   :
  significant  :

```

181: 240 [181: 240]

```

query: 594      : AACAGTTTACAACAACAATAGCCTTTAATCYCTCCTCAGGAGGGGATCCAGAAATTGTAA
a : CON_A1      : --T-C---AC-A-----AT---C-AA-----T-----AC--
b : A2.CY.94CY017_41 : --GA-A-C-AG-A--C--AT---C-AA-----C-----AC--
c : CON_B       : ---A---AT-A-----T-----AA-----C-----
d : CON_C       : ---C--C-AT-A-----AAA--G-A-CA-----C-T-----AC--
e : CON_D       : -C-TTC--AG-----ATT---A-CA---G-----C-----AC--
f : CON_F1      : CT--T--C-AT-----AAA---CTCA-----C-T-----AC--
g : CON_F2      : -C--C---AT-A-T--G---V---V-D-CA-A-----T-----AC--
h : CON_G       : --ATC---AC-AG-AC--A-----CTCA--TG-----C-T-----AC--
i : CON_H       : D-T-C---AT-G-----ATT---A-CACA-----CATG--G--AB--
j : J.SE.94.SE7022 : ---C---AT-A-----A---CATCAC-----CAT-----AC--
k : K.CD.97.EQTB11C : ---C---AC-A-----ATA--C-A-CA-----C--C--G-CAC--
l : CON_01_AE   : -G--C---AT-AG-----AT---C-A-CAC-----A---T-----AC--
  best match   :          cccccccccccccccccccccccccccc
  significant  :          ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

```

241: 300 [241: 300]

```

query: 594      : TGCACAGTTTAAATTGTGGAGGGGAATTCTTCTACTGTAATACAACAGAACTGTTTAATA
a : CON_A1      : CA-T-----A----T----T-----T--GC-----
b : A2.CY.94CY017_41 : CA-TT-----C--A----T----T--C-----GC-----G

```

c: CON_B : -----T-----C-----
 d: CON_C : CA-T-C-----A--A---T---T-C---T-A-----
 e: CON_D : CA---C-----T-----C---T--G-----
 f: CON_F1 : ---T-----A--A---T---C---T--G-----G
 g: CON_F2 : CA-T-----A--A---T---C-C--D--BT-----G
 h: CON_G : CA-T---C-----A--A---T---T-----T--G-----
 i: CON_H : CA-T-----A--A---T---T-C---T--GG-----
 j: J.SE.94.SE7022 : CA-T-----A---T---T-----T-AC---C---
 k: K.CD.97.EQTB11C : G---T---C---A---A---T-C---G---T---CAC-G-G-G
 l: CON_01_AE : ---TCA-----A-----T---T-C-----A-----
 best match :
 significant :

301: 360 [301: 343]

query: 594 : GTACTAGKARAWAATGAWAWYACCMCCCYCCCMWGCAGAATMCAACTWMTTAWCATCAGG
 a: CON_A1 : -C---T-G-TCA-----C-CT-TAA-T-T---AT-----AA-G-AAA---TA-AT-T-
 b: A2.CY.94CY017_41 : ---GT-G-ACC---GA-GT-TAAT--T---AT-----AA---AAA---TA-A--T-
 c: CON_B : ----T-G-G-A-----A-CT-T-A-A-T---AT-----AA---AAA---TA-A--T-
 d: CON_C : ----ATAC-ACA---CA-CC-T-A-A-T---AT-----AA---AAA---TA-A--T-
 e: CON_D : ----AT-G-AMA-----K-CA-T-A-A-T---AT-----AA---AAA---TA-A--T-
 f: CON_F1 : AC--AG-ACTMC---GC-CT-T-A-T-T---AT-TC---AA---AAA--GTA-A--T-
 g: CON_F2 : AA--A-AVGDCA----C-HT-T-A-T-TT--AT-T-----AAG--AAA--GTA-A--T-
 h: CON_G : A--G-MTBGA-T----K-CT-T-A-A-T---AT-T-A---AA---AAA--GTG-GA-T-
 i: CON_H : --G-T-G-ABA-VA--A-AT-TTA-A-TG--AT-----AA---AAA--GTA-A--T-
 j: J.SE.94.SE7022 : --G-T-G-G-C-CACA-CT-T-A-AATA--AT-T-A---AA---AAA--GTG-GA-T-
 k: K.CD.97.EQTB11C : AC--AGAG-G-GG--ACC-CA-T-A-AAT---AT-T-G--AA---AGA-A-TA-A--T-
 l: CON_01_AE : ----T-GGTGT-----C-CT-T-ATA-TT--AT---AG--AA-G-AAA---TA-A--T-
 best match :
 significant :

361: 420 [344: 403]

query: 594 : CGGCAGGWAGKAGRRRCGRRCRAWGTCYCCCCCYCATCGGACGACTAACATCATCCTCC
 a: CON_A1 : T-----AG--T--GACAAG-A-T--ATG---T-C---CA-G-G---T-AGG-GTGAA
 b: A2.CY.94CY017_41 : T-----AG--T--GAAGAG-A-T--ATG---T-C---T-C-G--A---T-AAG-GTA-A
 c: CON_B : T-----A--T--GAAAAG-A-T--ATG---T-C---A-G---A--TTAG--GT--A
 d: CON_C : T-----AG--T--GACGAG-A-T--ATG---T-C---T-C-G--AAC-T-A---GTAAA
 e: CON_D : T-----G--T--GAAAAG-A-T--ATG---T-C---T-A-G-----TCAA--GT--A
 f: CON_F1 : T-----A--T--GACGAG-A-T--ATG--G-T-C---T-C-G--AAC-TTA-C-GTAA-
 g: CON_F2 : T----AAG--T--GACGAG-A-T--ATG---T-C---T-CHG--VA--TTCAG-GTAA-
 h: CON_G : T----AG--TG-GACAAG-A-T--ATG---T-C-----C-G--AAC-TTA---GTAAA
 i: CON_H : T----AG--T--GACAAG-A-T--ATG---T-C---AA-G--AAC-TTA---GTGTA
 j: J.SE.94.SE7022 : T----AAG-AC--GACAAG-A-TA-ATG---T-C-----C-G--AAC-TTA---G-AAA
 k: K.CD.97.EQTB11C : T----AA--TG-GGCAAG-A-TA-ATG---T-C--C--C-G--AAC-TTA-C-GTAGA
 l: CON_01_AE : T-----G--T--GACAAG-A-T--ATG-T-T-C---A-TG--AG--TTAAT-GTGTA
 best match :


```

significant      :

421: 428 [ 404: 411]
query: 594      : TCTATTAC
a : CON_A1      : --C-----
b : A2.CY.94CY017_41 : --C-----
c : CON_B       : -----
d : CON_C       : ----C--
e : CON_D       : -----
f : CON_F1      : -----
g : CON_F2      : ----C--
h : CON_G       : --C-----
i : CON_H       : -----
j : J.SE.94.SE7022 : -----
k : K.CD.97.EQTB11C : -----
l : CON_01_AE   : -----
  best match    :
  significant    :

```

Con el alineamiento se determinan la distancia *s* (similitud), distancia *d* (Jukes Cantor) y la distancia *p* (Hamming) .Los resultados del análisis se expresan en gráficos que muestran la distancia entre el virus problema y los diferentes virus de referencia a lo largo de la secuencia analizada.

En el gráfico de la distancia *s* (similitud) en el eje de las “y” se muestra la proporción de similitud y en el eje de las “x” la región de la secuencia. La referencia con mayor similitud a la secuencia problema tendrá el mayor valor de distancia *s* (**ver figura 5.14**).

Gráfico de distancia *s* (similitud)

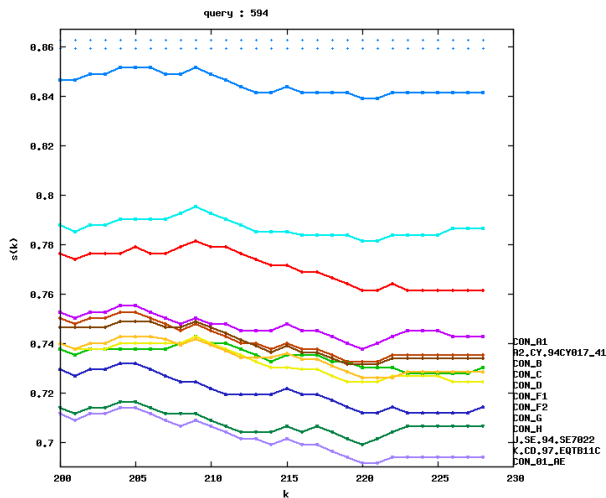


Figura 6.14 Gráfica de distancia s obtenida como resultado del análisis de subtipificación por RIP. Se muestra un ejemplo representativo del análisis de subtipificación por RIP. La secuencia de referencia que cuenta con mayor similitud a la secuencia problema es la que presenta valores mayores de distancia s a lo largo de la secuencia. En este ejemplo la secuencia problema correspondería al subtipo B.

La distancia p llamada distancia de Hamming, es la forma más sencilla de calcular la distancia entre dos secuencias. Para obtener esta distancia únicamente se cuentan el número de diferencias existentes entre las dos secuencias. A pesar de lo anterior, la distancia p no refleja las relaciones evolutivas.

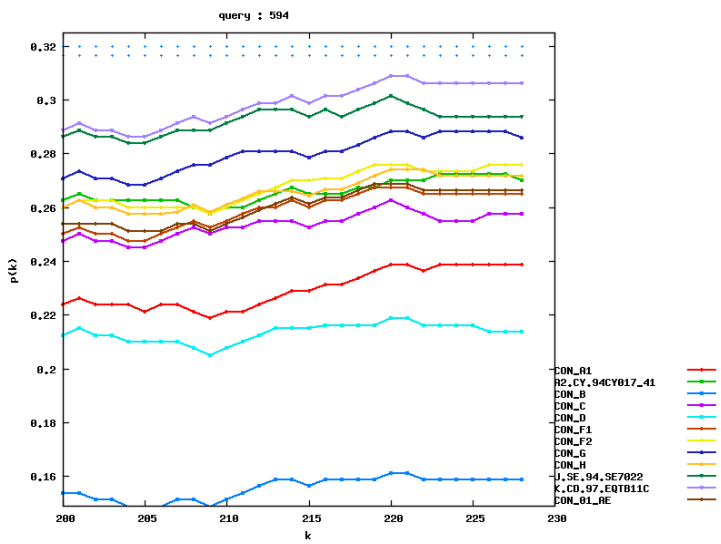
La distancia d es un intento por ajustar el número de diferencias entre dos secuencias de las posibles mutaciones en paralelo o de las reversiones. Este parámetro fue propuesto por Jukes y Cantor en 1969, con la siguiente ecuación:

$$d = \frac{3}{4} \ln(1 - \frac{4p}{3})$$

En el caso de los gráficos de distancia p y d, la referencia con mayor similitud es aquella que cuenta con una menor distancia. Al igual que en el gráfico distancia s (similitud) en el eje de las "x" se encuentra la región de la secuencia y en el eje de las "y" la distancia correspondiente. Las secuencias de referencia que presenten mayor similitud con la secuencia problema tendrán menor valor de distancia p y d (ver figura 6.15).

a)

[Grafico distancia p \(Hamming\)](#)



b) Grafico distancia d (Jukes Cantor)

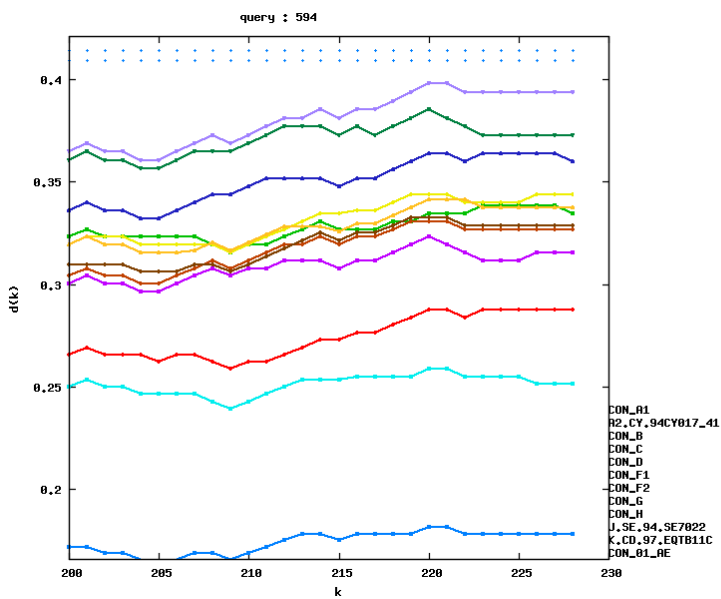
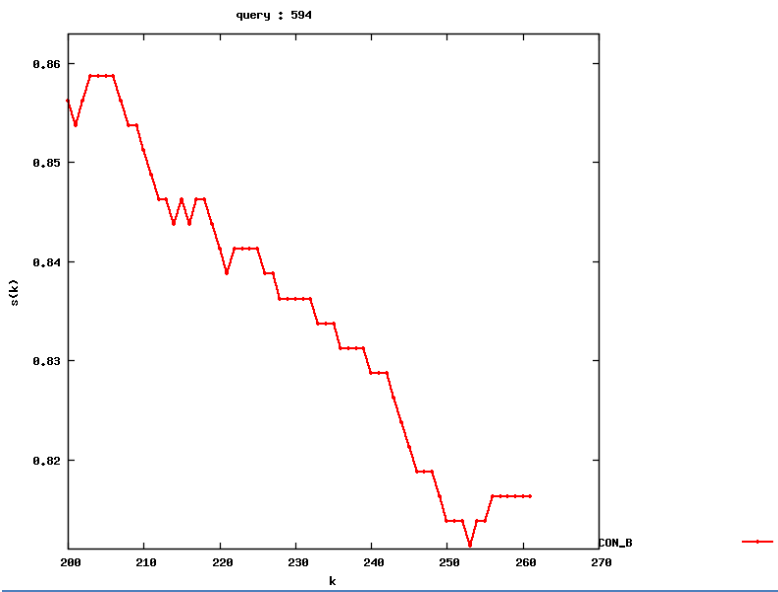


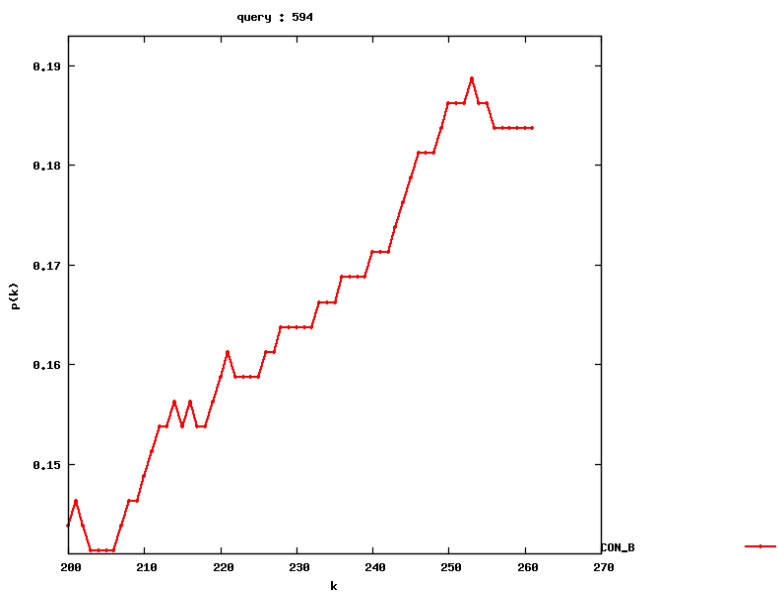
Figura 6.15 Gráficos de distancia p y d obtenidos como resultado del análisis de subtipificación de VIH por RIP. El subtipo que presenta mayor similitud es aquél que cuente con un valor menor valor de distancia p y d. En este caso la secuencia problema corresponde al subtipo B.

RIP cuenta con una opción en la cual se simplifican las gráficas de distancia s, p y d; es decir únicamente muestran los subtipos con mayor similitud a la secuencia problema (**ver figura 6.16**).

a) [Grafico de distancia s \(similitud\)](#)



b) Grafico de distancia p(Hamming)



c) Grafico de distancia d(Jukes Cantor)

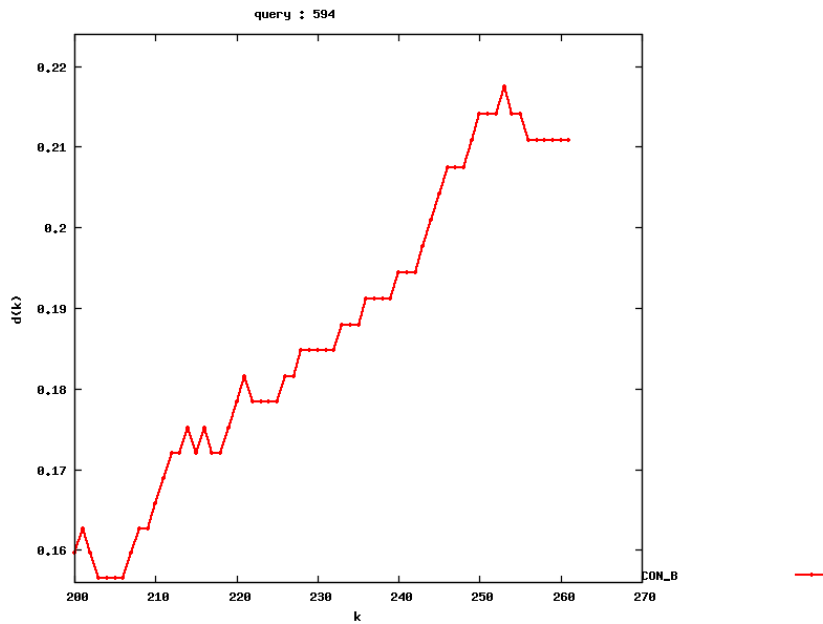


Figura 6.16 Ejemplo de los gráficos simplificados. En todos los ejemplos, se muestra que el subtipo que es más similar a la secuencia problema es el subtipo B.

7.0 Resultados

7.1 Descripción de la cohorte en estudio.

Se analizaron 101 secuencias del gen *env* de pacientes infectados por VIH-1 provenientes del Norte, Sur y Centro de la República Mexicana. Los estados que participaron en este estudio fueron: Sonora, Baja California Norte, Nuevo León, Jalisco, D.F., Edo. de México, Morelos, Chiapas, Oaxaca, Veracruz y Guerrero (ver figura 7.1).



Figura 7.1 Procedencia de los individuos infectados por VIH-1 que participaron en el estudio. Se clasificaron los individuos infectados por VIH participantes en el estudio según su región de procedencia: Norte, Sur y Centro de la República Mexicana. Los estados que participaron en el estudio fueron: Sonora, Baja California Norte, Nuevo León, Jalisco, D.F., Edo. De México, Morelos, Chiapas, Oaxaca, Veracruz y Guerrero. En rosa se muestran los estados participantes incluidos en la región Norte, en azul los incluidos en la región Centro y en verde los incluidos en la región Sur.

De las 101 secuencias analizadas del gen *env* de VIH de individuos mexicanos, el 37% (38 pacientes) provenía del Centro, el 36% (36 pacientes) provenía del Sur y el 27% (27 pacientes) del Norte (**Figura 7.2**). El 67.33% de estos pacientes eran hombres, mientras que un 32.67% eran mujeres, la edad promedio fue de 34.7 años. Todos los pacientes incluidos en el presente estudio formaban parte de la cohorte del proyecto de epidemiología molecular del Centro de Investigación en Enfermedades Infecciosas (CIENI) del Instituto Nacional de Enfermedades Infecciosas (INER).

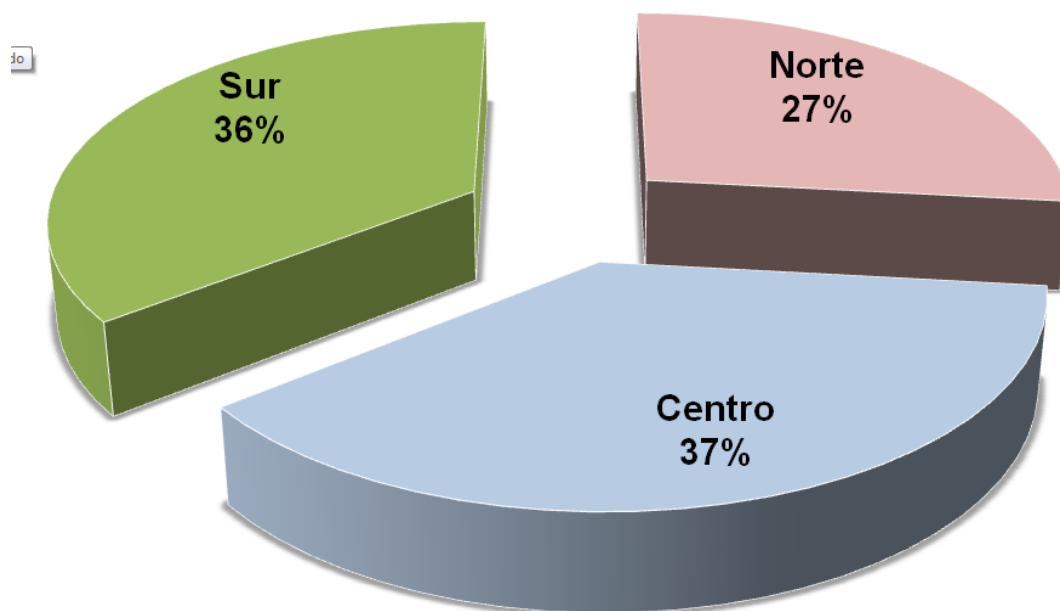


Figura 7.2 Proporción de pacientes provenientes de las regiones Norte, Centro y Sur de la República Mexicana. El 37% (38) de los pacientes del estudio provenía del Centro, el 36% (36) provenía del Sur y el 27%(27) del Norte. Todos los pacientes participantes en el estudio pertenecían a la cohorte de Epidemiología Molecular del CIENI del INER.

El promedio de la carga viral de todos los pacientes analizados fue de 882,456 copias de RNA/ml, mientras que el promedio de células CD4+ fue de 271 células/ μ l. Mostrándose rangos amplios en la distribución de valores tanto de carga viral como de cuentas de células CD4+. En la **Tabla 7.1** se hace una

descripción estadística de la cohorte entera y por región: Centro, Norte y Sur. En general, las medianas de los valores de carga viral fueron similares en las tres regiones en estudio y con respecto a la cohorte considerada como un todo. Las medianas de cuentas de células CD4+ fueron también similares, con una ligera tendencia a valores más altos en la región Norte. La mediana de 212 células/ μ l de la cohorte completa refleja el retraso en el diagnóstico de la infección por VIH característico de las cohortes latinoamericanas, presentándose la mayoría de los pacientes en los centros de salud hasta etapas tardías de la infección.

Tabla.7.1 Descripción estadística de la cohorte					
Todos los pacientes de la corte (n=101)			Pacientes del Centro (n=38)		
	CD4⁺ (Células/ μ l)	Carga viral (copias RNA /ml)		CD4⁺ (Células/ μ l)	Carga viral (copias RNA/ml)
Promedio	271	882,456	Promedio	256	195,735
Desviación estándar	224	7,054,895	Desviación estándar	258	255,081
Máximo	1058	71,000,000	Máximo	1058	988,565
Mínimo	1	82	Mínimo	14	82
Rango	1057	70,999,918	Rango	1044	988,483
Mediana	212	52,846	Mediana	191	74,038
Pacientes del Norte (n=27)			Pacientes del Sur (n=36)		
	CD4⁺ (Células/ μ l)	Carga viral (copias RNA/ml)		CD4⁺ (Células/ μ l)	Carga viral (copias RNA/ml)
Promedio	315	115,888	Promedio	254	2,182,255
Desviación estándar	221	151,228	Desviación estándar	188	11,807,436
Máximo	850	614,592	Máximo	826	71,000,000
Mínimo	24	174	Mínimo	1	136
Rango	826	614,418	Rango	825	70,999,864
Mediana	300	50,319	Mediana	201	54,173

Se realizó una comparación entre el Log_{10} (carga viral) de pacientes infectados por VIH de las regiones Centro, Norte y Sur de la República Mexicana. En la **figura 7.3** se muestran unos diagramas de Cajas y bigotes que demuestran la similitud de las distribuciones de los valores de carga viral en las tres regiones en estudio.

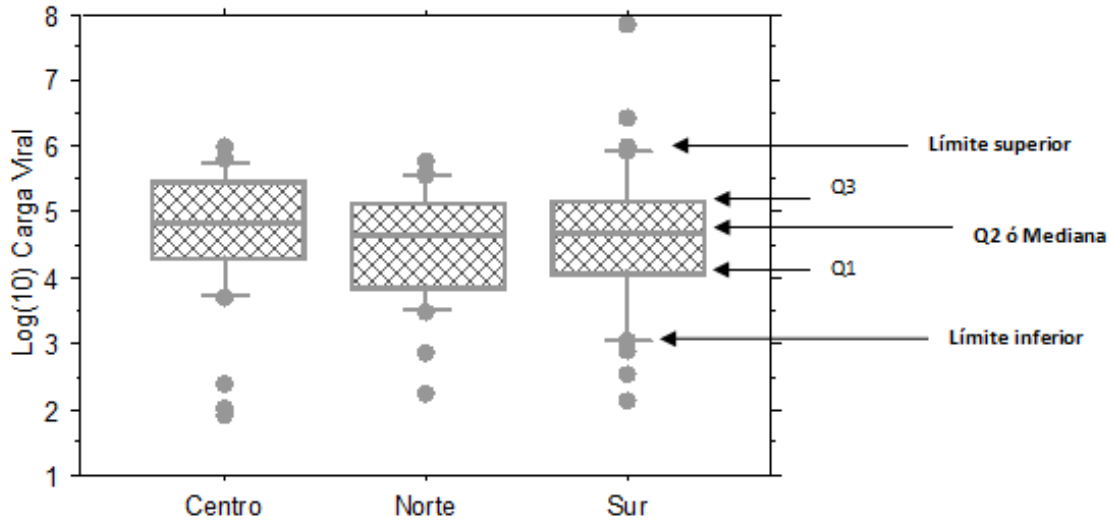


Figura.7.3 Diagrama de cajas y bigotes del \log_{10} de la carga viral en las tres regiones de la República Mexicana. El rectángulo representa los cuartiles Q1 y Q3. La línea horizontal que divide al rectángulo representa el cuartil Q2 el cual es la mediana. Los bigotes representan los límites inferiores y superiores. Los puntos que se salen de los límites superiores e inferiores son datos atípicos.

Por medio de la prueba Kruskal Wallis se determinó que no existían diferencias significativas entre las medianas de carga viral de pacientes provenientes de la región Centro, Norte y Sur de la República Mexicana, ya que el valor de P de los rangos (0.5748) era mayor a $\alpha=0.05$. Por tanto se aceptó la hipótesis nula (H_0), la cual nos indicó que las medianas del \log_{10} (carga viral) eran similares en las regiones Centro, Norte, Sur de la República Mexicana (ver tabla 7.2 y Tabla 7.3).

Tabla 7.2 Valores de los parámetros obtenidos por la prueba Kruskal Wallis aplicada a los valores de Log_{10} (carga viral) de pacientes mexicanos provenientes de las regiones Centro, Norte y Sur de la República Mexicana

Prueba Kruskal-Wallis Log_{10} (Carga viral)	
Grupo variable: Región	
Grados de libertad	2
# Grupos	3
#Rangos	16
H	1.107
Valor de P	0.5749
H corregida para los rangos	1.107
Correccion del valor de P de los rangos	0.5748

Tabla 7.3 Información de los rangos obtenidos por la prueba Kruskal Wallis aplicada a los valores de células CD4^+ de pacientes mexicanos provenientes de las regiones Centro Norte y Sur de la República Mexicana

Región	#	Σ Rango	Rango promedio
Centro	38	2072.000	54.526
Norte	27	1265.000	46.852
Sur	36	1814.000	50.389

Al igual que en el caso de las cargas virales, se comparó la cuenta de células CD4^+ en las regiones Centro, Norte y Sur de la República Mexicana. Se observó que las medianas de células CD4^+ en las regiones Centro y Sur eran similares, mientras que la mediana de células CD4^+ en el Norte era ligeramente elevada (**ver figura7.4**)

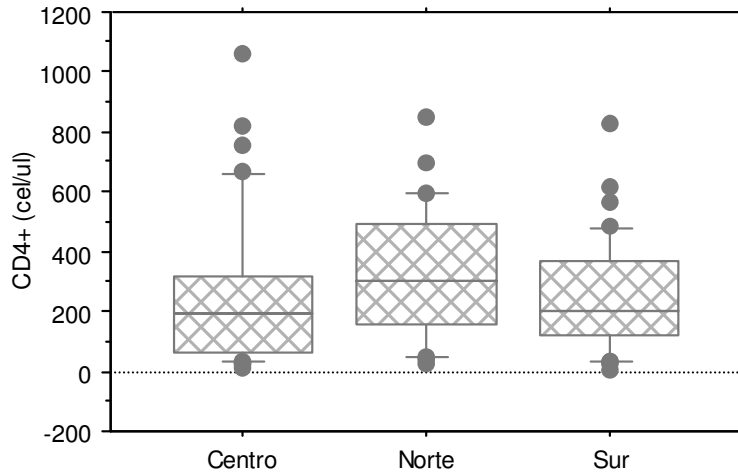


Figura.7.4 Diagrama de cajas y bigotes del conteo de células CD4+ en las diversas regiones de la República Mexicana. El rectángulo caja representa los cuartiles Q1 y Q3. La línea horizontal que divide al rectángulo representa el cuartil Q2 el cual es la mediana. Los bigotes representan los límites inferiores y superiores. Los puntos que se salen de los límites superiores e inferiores son datos atípicos.

Se realizó la prueba Kruskal Wallis con lo cual se determinó que no existían diferencia significativa entre los valores de células CD4⁺ de pacientes provenientes de las regiones Centro, Norte y Sur de la República Mexicana, ya que el valor corregido de P de los rangos (0.2803) era mayor a $\alpha=0.05$. Por tanto se aceptó la hipótesis nula (H_0), la cual nos indicaba que las medianas del conteo de células CD4⁺ eran similares en las regiones Centro, Norte, Sur de la República Mexicana (ver tabla 7.4 y Tabla 7.5).

Tabla 7.4 Valores de los parámetros obtenidos por la prueba Kruskal Wallis aplicada a los valores de células CD4⁺ de pacientes mexicanos provenientes de las diversas regiones de la República Mexicana

Prueba Kruskal-Wallis CD4 ⁺ (cel/ml)	
Grupo variable: Región	
Grados de libertad	2
# Grupos	3
#Rangos	1
H	2.543
Valor de P	0.2803
H corregida para los rangos	2.544
Corrección del valor de P de los rangos	0.2803

Tabla 7.5 Información de los rangos obtenidos por la prueba Kruskal Wallis aplicada a los valores de células CD4⁺ de pacientes mexicanos provenientes de las diversas regiones de la Republica Mexicana

Región	#	Σ Rango	Rango promedio
Centro	38	1759.000	46.286
Norte	27	1567.000	58.037
Sur	36	1825.000	50.694

7.2 Subtipificación de virus circulantes en la República Mexicana

Con el objetivo de determinar el subtipo viral de las secuencias del gen *env* incluidas en el presente estudio se realizó un análisis de subtipificación por medio de los programas REGA [44] y RIP [43], en un principio se realizó la subtipificación por medio de REGA, con este programa únicamente fue posible subtipificar el 12%(12) de las secuencias debido a que la subtipificación se ve limitada ya que cuenta únicamente con 29 secuencias referencia, por tal motivo se decidió utilizar el programa RIP con el cual fue posible subtipificar la mayoría de las secuencias (**ver tabla 7.6**). Después de la aplicación de ambos programas se determinó que el 87% de las secuencias eran del subtipo B (**ver figura 7.5**). Sin embargo, el 13% restante de las secuencias no se pudieron subtipificar por medio de estos programas ya que el tamaño de las secuencias no era adecuado.

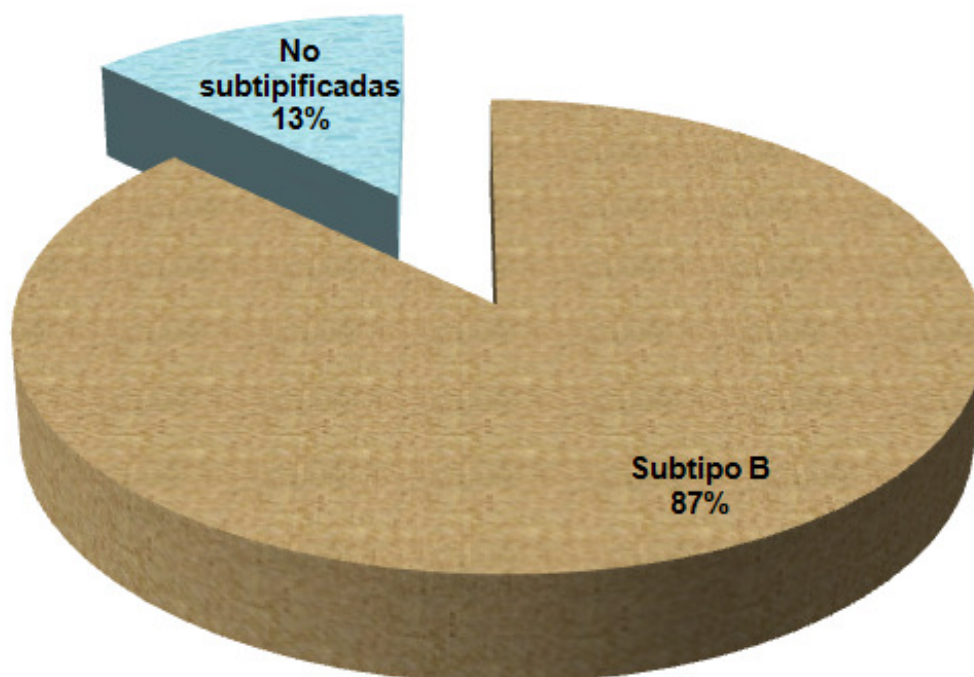


Figura 7.5 subtipificación por medio de RIP y REGA. El 87% de las secuencias analizadas eran del subtipo B mientras el 13% no se pudo determinar el subtipo al cual pertenecían ya que la longitud de las secuencias era menor a la requerida por los programas.

Tabla 7.6 Subtipificación de los virus circulantes en la cohorte mexicana por análisis de secuencias del gen *env*, mediante las herramientas RIP y REGA.

Paciente	Paciente	Estado de Procedencia	Región de procedencia	SUPTIPO REGA*	SUBTIPO RIP*
Clave CIENI	Clave en el presente estudio				
0327-RRN	S01	Chiapas	Sur	INDEFINIDO	B
345-MACN	S02	Chiapas	Sur	INDEFINIDO	B
0348-JLC	S03	Chiapas	Sur	INDEFINIDO	INDEFINIDO
0357-RRB	N01	Sonora	Norte	INDEFINIDO	INDEFINIDO
0358-MGA	S04	Chiapas	Sur	INDEFINIDO	INDEFINIDO
0369-LCQL	N02	Sonora	Norte	INDEFINIDO	B
0372-JCH	N03	Sonora	Norte	B	INDEFINIDO
0379-LDLSS	S05	Chiapas	Sur	INDEFINIDO	INDEFINIDO
0380-MCFO	S06	Chiapas	Sur	INDEFINIDO	B
382-RGM	S07	Chiapas	Sur	INDEFINIDO	B
0431-RTG	N04	Tijuana	Norte	INDEFINIDO	B

0469-JHJ	S08	Chiapas	Sur	B	INDEFINIDO
0554-MGG	S09	Chiapas	Sur	INDEFINIDO	INDEFINIDO
0556-EHL	S10	Chiapas	Sur	INDEFINIDO	B
0591-GTS	N05	Tijuana	Norte	INDEFINIDO	B
0592-MMG	N06	Tijuana	Norte	INDEFINIDO	B
0594-ICR	N07	Tijuana	Norte	INDEFINIDO	B
0712-EGR	S11	Oaxaca	Sur	INDEFINIDO	B
0713-MJAO	S12	Oaxaca	Sur	B	B
0740-ACM	N08	Sonora	Norte	B	B
0964-REPG	S13	Veracruz	Sur	INDEFINIDO	B
0967-JACC	S14	Guerrero	Sur	INDEFINIDO	B
0968-MTL	S15	Veracruz	Sur	INDEFINIDO	INDEFINIDO
0971-MGGM	S36	Guerrero	Sur	INDEFINIDO	B
1029-MMB	C01	Guadalajara	Centro	INDEFINIDO	B
1030-PPD	C02	Guadalajara	Centro	INDEFINIDO	B
1041-MIHC	S16	Veracruz	Sur	INDEFINIDO	INDEFINIDO
1046-JMUR	C03	Guadalajara	Centro	INDEFINIDO	INDEFINIDO
1061-MOMF	N09	Nuevo León	Norte	INDEFINIDO	B
1064-FPO	S17	Veracruz	Sur	INDEFINIDO	B
1073-MMG	S18	Veracruz	Sur	INDEFINIDO	B
1077-OBL	C04	Guadalajara	Centro	INDEFINIDO	INDEFINIDO
1082-PARA	S19	Veracruz	Sur	INDEFINIDO	B
1092-EHG	C05	D.F.	Centro	INDEFINIDO	B
1097-BJC	C06	Estado de México	Centro	INDEFINIDO	INDEFINIDO
1103-MGL	C07	Morelos	Centro	INDEFINIDO	B
1106-IMR	S20	Oaxaca	Sur	INDEFINIDO	INDEFINIDO
1108-AM	C08	Morelos	Centro	INDEFINIDO	B
1110-JFMA	N10	Nuevo León	Norte	INDEFINIDO	B
1116-JGM	N11	Nuevo León	Norte	INDEFINIDO	B
1118-JACV	C09	Estado de México	Centro	INDEFINIDO	INDEFINIDO
1121-ADGM	N12	Nuevo León	Norte	INDEFINIDO	INDEFINIDO
1130-MMG	N13	Nuevo León	Norte	INDEFINIDO	B
1132-ACG	N14	Nuevo León	Norte	INDEFINIDO	B
1133-ESM	C10	Estado de México	Centro	INDEFINIDO	B
1135-JASV	N15	Nuevo León	Norte	B	B
1138-HZP	N16	Nuevo León	Norte	INDEFINIDO	B
1148-RGP	S21	Oaxaca	Sur	B	B
1151-FAC	S22	Oaxaca	Sur	INDEFINIDO	B
1152-JMS	S23	Oaxaca	Sur	INDEFINIDO	B
1156-JPR	C11	Guadalajara	Centro	INDEFINIDO	B

1158-JMS	N17	Nuevo León	Norte	INDEFINIDO	B
1159-YCO	C12	Morelos	Centro	INDEFINIDO	B
1160-JTG	C13	Morelos	Centro	INDEFINIDO	B
1161-GUA	C14	Morelos	Centro	INDEFINIDO	B
1164-VRC	N18	Baja California	Norte	INDEFINIDO	B
1168-REPM	N19	Nuevo León	Norte	INDEFINIDO	B
1169-JECG	N20	Nuevo León	Norte	INDEFINIDO	B
1172-IRR	N21	Nuevo León	Norte	B	B
1176-JRS	C15	Guadalajara	Centro	INDEFINIDO	B
1179-CARC	S24	Veracruz	Sur	INDEFINIDO	B
1185-BHE	C16	Estado de México	Centro	INDEFINIDO	B
1187-JIVE	C17	D.F.	Centro	INDEFINIDO	B
1188-LAZE	C18	D.F.	Centro	INDEFINIDO	B
1189-AMGC	C19	D.F.	Centro	INDEFINIDO	B
1190-MGA	N22	Nuevo León	Norte	INDEFINIDO	B
1191-JM	C20	Morelos	Centro	INDEFINIDO	B
1192-ASM	C21	Morelos	Centro	INDEFINIDO	B
1194-CGMC	S25	Veracruz	Sur	B	B
1196-JLG	S26	Veracruz	Sur	INDEFINIDO	B
1197-RAR	N23	Baja California	Norte	INDEFINIDO	B
1198-DMMO	N24	Baja California	Norte	INDEFINIDO	B
1200-PIN	C22	Estado de México	Centro	INDEFINIDO	B
1201-EES	C23	D.F.	Centro	B	B
1202-LGR	C24	D.F.	Centro	INDEFINIDO	B
1203-YEMC	C25	D.F.	Centro	INDEFINIDO	B
1209-RAM	S27	Oaxaca	Sur	INDEFINIDO	B
1210-TVL	S28	Oaxaca	Sur	INDEFINIDO	B
1213-MMF	S29	Oaxaca	Sur	INDEFINIDO	B
1215-MRM	S30	Oaxaca	Sur	INDEFINIDO	B
1218-AAT	C26	Morelos	Centro	B	B
1220-OAM	C27	Morelos	Centro	B	B
1222-JTT	C28	Morelos	Centro	INDEFINIDO	B
1223-JLMS	C29	Morelos	Centro	INDEFINIDO	B
1226-JNSP	C30	Guadalajara	Centro	INDEFINIDO	B
1227-CSP	C31	Guadalajara	Centro	INDEFINIDO	B
1229-MACR	C32	D.F.	Centro	INDEFINIDO	B
1232-LC	N25	Baja California	Norte	INDEFINIDO	B
1234-JLOA	S31	Veracruz	Sur	INDEFINIDO	B
1235-SFDP	S32	Veracruz	Sur	INDEFINIDO	B
1236 -GA	C33	Morelos	Centro	INDEFINIDO	B

1244-EVR	C38	Estado de México	Centro	INDEFINIDO	B
1247-JLFF	S33	Veracruz	Sur	INDEFINIDO	B
1250-EMD	C37	D.F.	Centro	INDEFINIDO	B
1251-AGF	C34	Estado de México	Centro	INDEFINIDO	B
1261-GCA	C35	Jalisco	Centro	INDEFINIDO	B
1272-GPM	C36	Jalisco	Centro	B	B
1298-MLBL	N26	Nuevo León	Norte	INDEFINIDO	B
1315-JVL	S34	Oaxaca	Sur	INDEFINIDO	B
1317-FRA	S35	Oaxaca	Sur	INDEFINIDO	B
1335-MRVC	N27	Sonora	Norte	INDEFINIDO	B

*Las Secuencias en un principio fueron subtipificadas por medio REGA [44] .Para determinar el subtipo de las secuencias indeterminadas y confirmar el subtipo de las secuencias subtipificadas por medio de REGA se utilizo RIP[45].

Para determinar el subtipo de las secuencias que no se pudieron analizar por medio de las herramientas REGA y RIP se realizó un análisis filogenético manual utilizando el programa MEGA 4.1 β. En este análisis se construyó un árbol filogenético con el método Neighbor Joining y el modelo Maximum Composite Likelihood, probando la resolución del árbol con el método de Bootstrap, realizando 1000 repeticiones. Se incluyeron las 13 secuencias problemáticas, al menos una referencia de cada subtipo del grupo M y una referencia del grupo N y O así como una referencia de SIV. Las secuencias de referencia se obtuvieron de la base de datos de los Álamos [43]. En este análisis fue posible observar que las secuencias que no se pudieron subtipificar por medio de los programas REGA y RIP (OTUs rojas) se agruparon con las secuencias referencia del subtipo B (OTUs verdes). Lo cual nos indica que mantienen una estrecha relación filogenética con dicho subtipo (**ver figura 7.6**).

Subtipo B

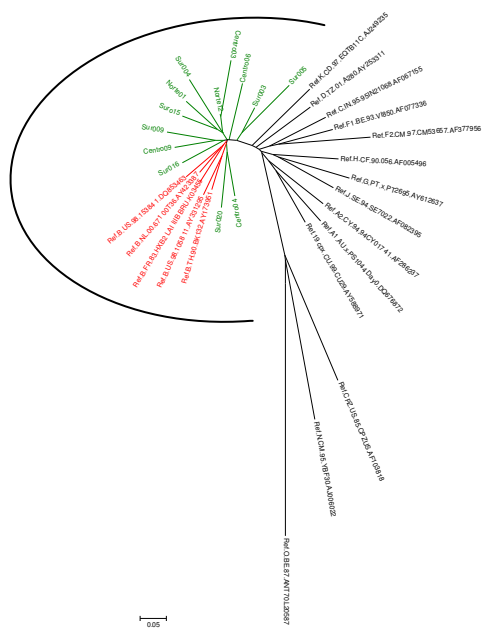


Figura.7.6 **Árbol filogenético para la subtipificación de secuencias del gen *env* problemáticas.** Se incluyeron las secuencias que no fue posible subtipificar por medio de las herramientas RIP y REGA, así como 14 secuencias de referencia obtenidas de la Base de Datos de Los Alamos [43]. Se muestra la relación evolutiva de 27 taxas. La historia evolutiva se obtuvo usando el método Neighbor-Joining. El árbol que se muestra es el árbol consenso después de 1000 repeticiones de bootstrap. La suma de las longitudes de las ramas fue de 4.55549827. Las distancias evolutivas fueron computadas por medio del modelo Maximum Composite Likelihood. Todas las inserciones o deleciones que se encontraban en el alineamiento fueron eliminadas al comparar con las demás secuencias (opción: pairwise deletion). El análisis filogenético se realizó en el programa Mega 4.1β. Las OTUs en negro son referencias de subtipos no B, las OTUs verdes representan las secuencias problemáticas y las rojas representan las referencias del subtipo B.

Es posible observar que las OTUs Sur003 y Sur005 se alejan un poco de las demás OTUs de pacientes mexicanos por tal motivo se decidió construir un árbol filogenético con todas las secuencias de los pacientes mexicanos, incluyendo cuatro referencias del subtipo B y una referencia de SIV_{CPZ}, esta última referencia se incluyo para enraizar el árbol. Esto nos ayudo a determinar la relación filogenética de las OTUs Sur003 y Sur005 con las OTUs de virus mexicanos (**ver figura 7.7**). Observando la topología del árbol encontramos que existe los dos virus (Sur003 y Sur005) se encuentran relacionados con virus mexicanos que pertenecen al subtipo B, esto nos indica que pertenecen a este subtipo.

B, las OTUs rojas son las secuencias Sur003 y Sur005 , en azul se encuentra la referencia de CPZ, Todas las referencias se obtuvieron de la Base de Datos de Los Alamos [43]

Al conjuntar los resultados obtenidos por los programas RIP, REGA; y el análisis filogenético realizado con las secuencias no subtipificadas por estos programas, observamos que el 100% de las secuencias pertenecen al subtipo B.

7.3 Análisis Filogenético

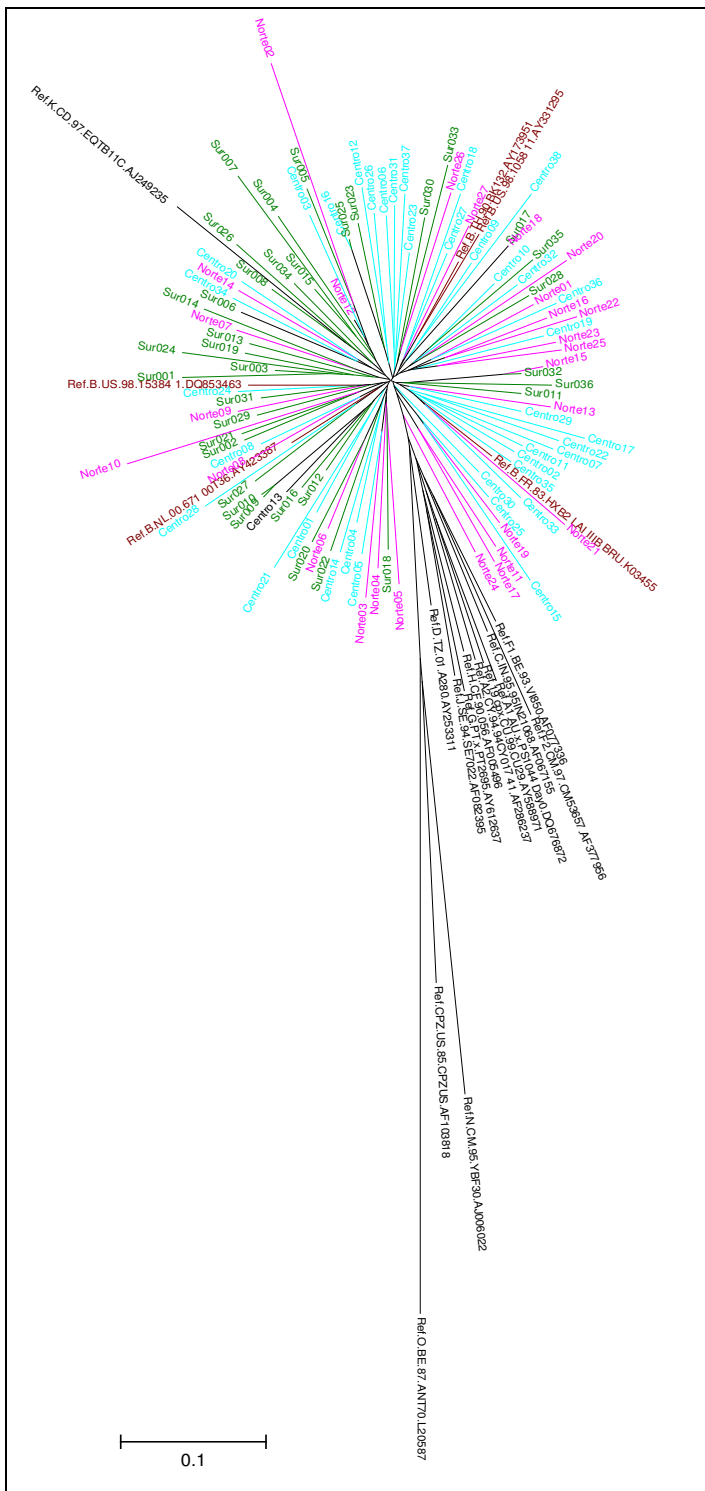
7.3.1 Comparación entre árboles filogenéticos obtenidos a partir de los genes *pol* y *env*

Con el objeto de determinar si la variabilidad del gen viral analizado pudiera influenciar las conclusiones obtenidas a partir del análisis filogenético de VIH circulante mexicano, se realizó una comparación entre los árboles filogenéticos construidos con secuencias de los genes *env* y *pol* de los virus de individuos participantes en el presente estudio. Para realizar esta comparación se construyó un árbol con secuencias el gen *pol* y un árbol con secuencias del gen *env* de los mismos virus, utilizando el método Neighbor Joining, calculando las distancias evolutivas mediante el modelo de Kimura 2 parámetros. La resolución de ambos árboles se probó mediante el método de bootstrap con 1000 repeticiones. Cada OTU se coloreó de acuerdo a su región de procedencia con el objetivo de identificar posibles agrupamientos por zona geográfica.

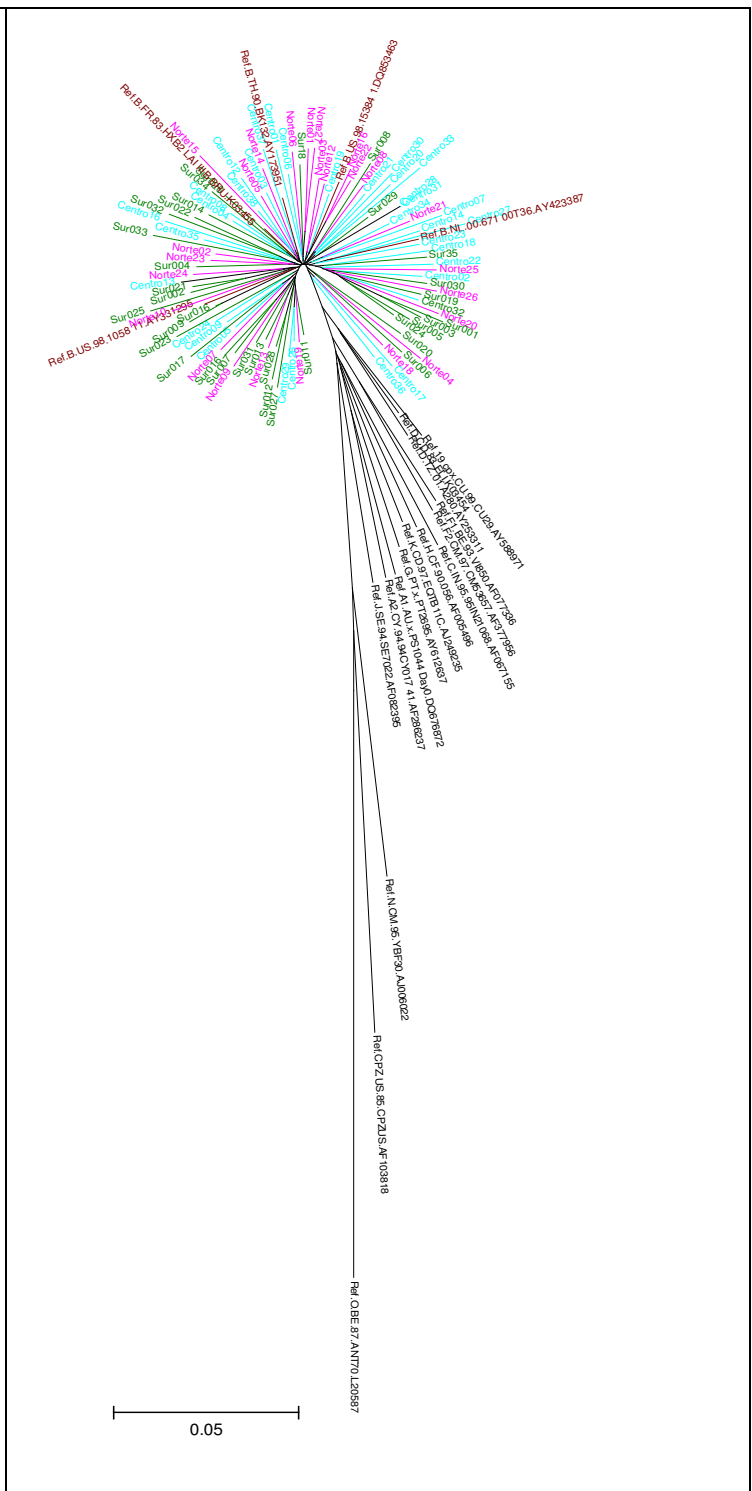
Al observar el árbol que se construyó a partir de secuencias del gen *pol* es posible notar que las OTUs se distribuyen de forma homogénea (**ver figura 7.8**), es decir no es posible observar agrupamientos entre OTUs de la misma región, esto lo podemos atribuir a que las secuencias del gen *pol* son muy homogéneas entre los virus circulantes en el país.

Si observamos el árbol que se construyó a partir de secuencias del gen *env* de virus provenientes de pacientes mexicanos infectados por VIH, en general encontramos que las OTUs se distribuyen de forma homogénea, a pesar de ello

también es posible distinguir algunos agrupamientos entre OTUs de la misma región (ver figura 7.8).



Árbol filogenético del gen *env*



Árbol filogenético del gen *pol*

Figura.7.8 Comparación entre árboles filogenéticos obtenidos a partir de los genes *env* y *pol* de VIH de los individuos incluidos en la cohorte de estudio. Se muestra la relación filogenética de 120 taxas . La historia evolutiva se obtuvo usando el método Neighbor-Joining. El árbol que se muestra es el árbol consenso después de 1000 repeticiones de bootstrap. La suma de las longitudes de las ramas del árbol de *pol* fue de 3.56976778, en el caso del árbol de *env* la suma de las longitudes de las ramas fue de 13.27077729. Las distancias evolutivas fueron computadas por medio del modelo Kimura-2 parámetros. Todas las inserciones o deleciones que se encontraban en el alineamiento fueron eliminadas exclusivamente después de comparar con las demás secuencias del alineamiento (opción: pairwise deletion). El análisis filogenético se realizó en el programa Mega 4.1β. Se incluyó al menos una secuencia de referencia de cada subtipo del grupo M y secuencias de referencia de los grupos O y N. Todas las secuencias de referencia se obtuvieron de la base de datos de los Álamos [42]. Las referencias del subtipo B se muestran en color vino, las referencias de subtipos NO B se muestran en negro. Las OTUs provenientes del Norte se muestran de color rosa, las OTUs provenientes del Centro en verde agua y las OTUs provenientes del Sur en verde. En el árbol de *env* se señala la referencia del subtipo K con una línea punteada.

A pesar de que sea posible observar algunos agrupamientos entre algunas OTUs pertenecientes a la misma región geográfica en el árbol de *env*, no podemos asegurar que estos sean reales ya que los valores de bootstrap son menores a 70% (ver figura 7.9).

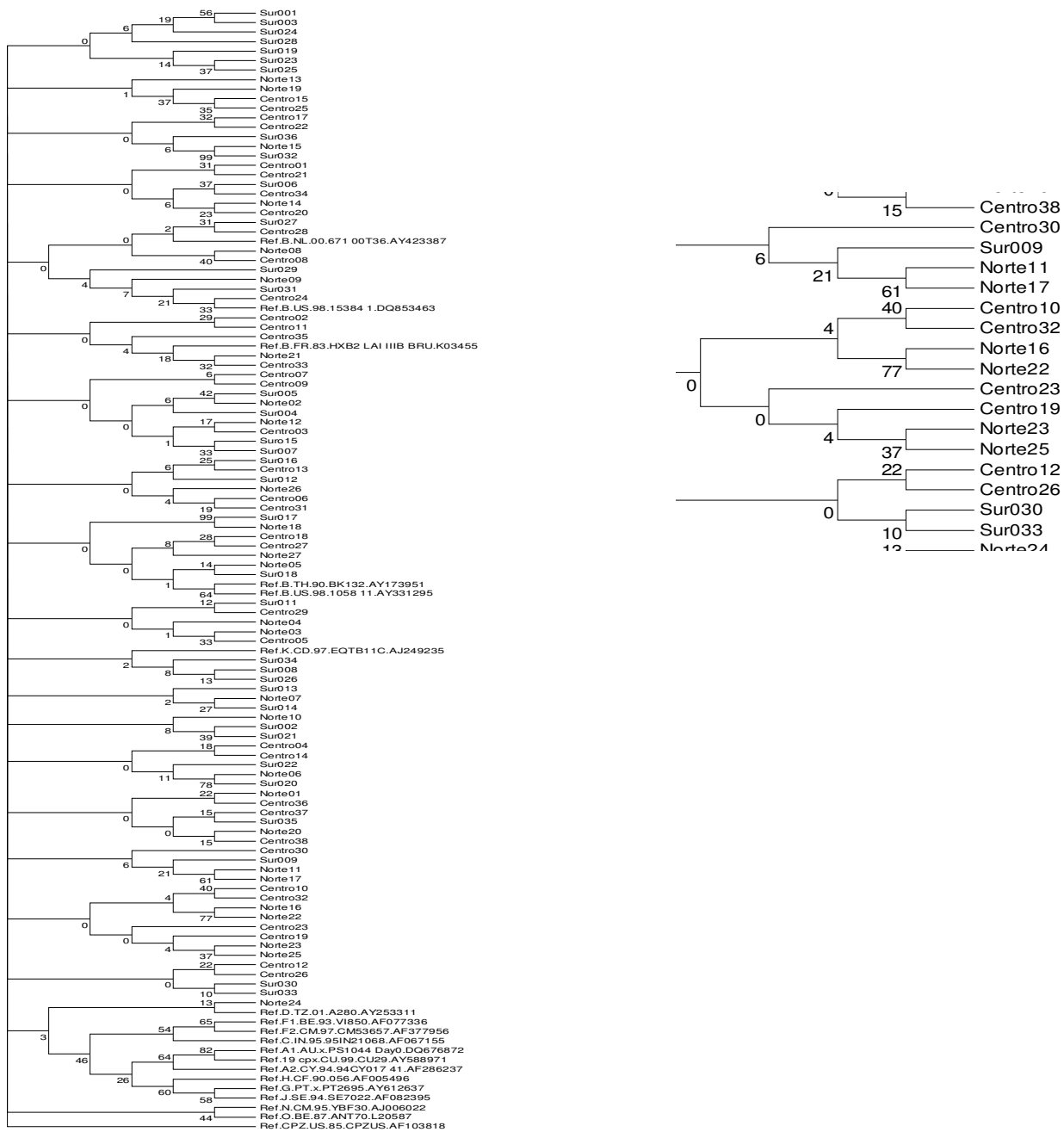


Figura 7.9 Se probó el método filogenético por bootstrap 1000 repeticiones. Relación filogenética de 120 taxos del gen *env* de VIH de individuos mexicanos. La historia evolutiva se obtuvo usando el método Neighbor-Joining. El árbol que se muestra es el consenso después de 1000 repeticiones de bootstrap ya que la suma de las longitudes de las ramas del árbol de *env* suma fue de 13.27077729. Las distancias evolutivas fueron computadas por medio del modelo Kimura-2 parámetros. Todas las inserciones o deleciones que se encontraban en el alineamiento fueron eliminadas exclusivamente después de comparar con las demás secuencias del alineamiento (opción: pairwise deletion). El análisis filogenético se realizó en el programa Mega 4.1β. Los valores de bootstrap se muestran al lado de las ramas, a la derecha se muestra un acercamiento a los valores de bootstrap de una de las ramas. En casi todos los casos el valor de bootstrap fue menor al 70%.

Si realizamos un acercamiento a la cabeza del árbol de *env* que se construyó con el método Neighbor Joining y el modelo Kimura 2 parámetros ,se observó un artefacto del modelo, ya que entre las OTUs que representan a los diferentes virus de pacientes mexicanos infectados por VIH se intercaló una referencia del subtipo K (**ver figura 7.10**). Esto último representa un error en la topología del árbol, ya que ningún virus analizado guarda una relación filogenética cercana con el subtipo K y el virus que mantenía una relación cercana en el árbol filogenético que se construyó con el método Neighbor Joining y el modelo Kimura 2 parámetros, es decir el virus Sur008 pertenecía al subtipo B de acuerdo al análisis de subtipificación de REGA.

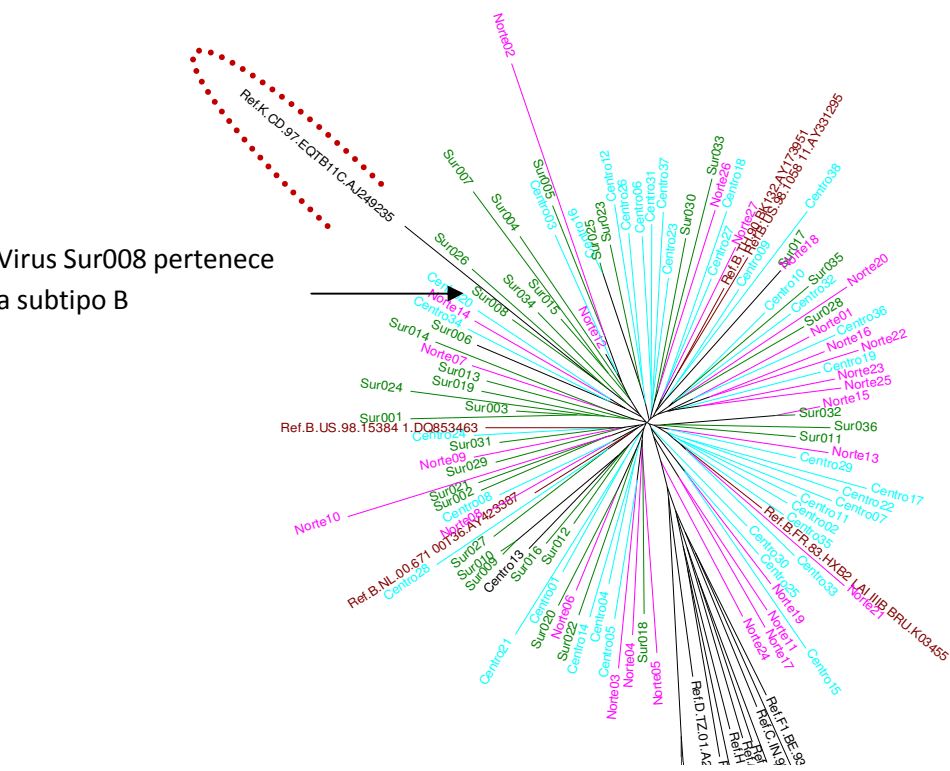
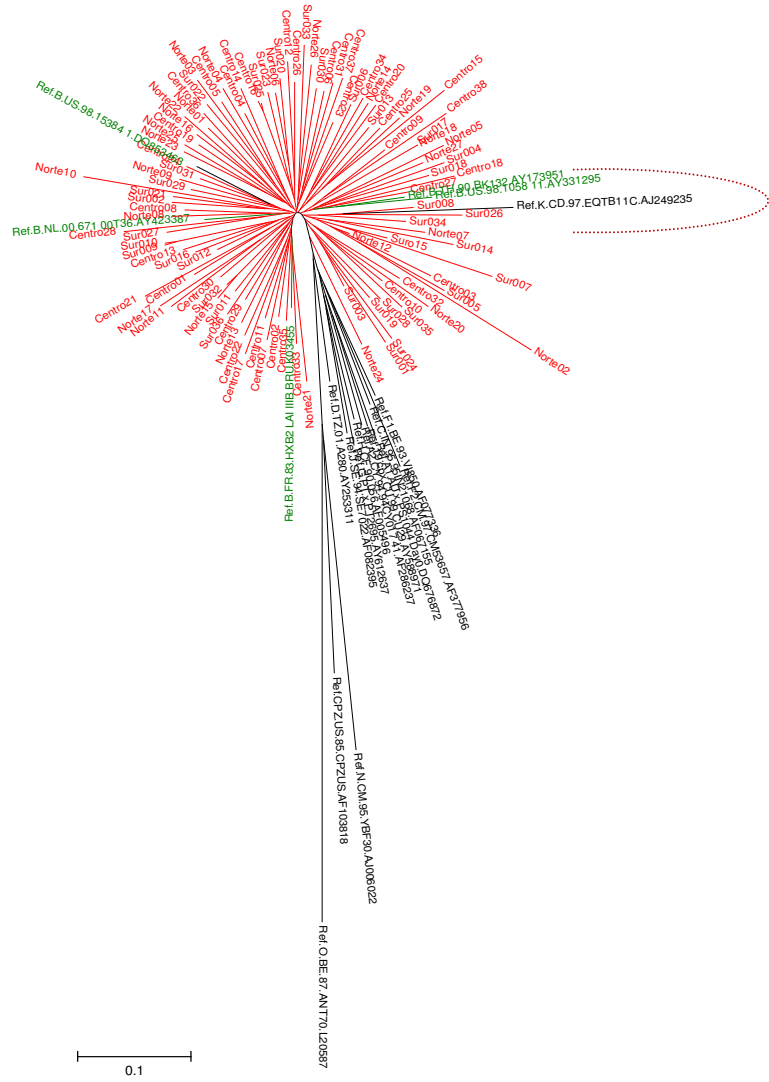


Figura 7.10 Acercamiento a la cabeza del árbol de *env* que se construyo con el método Neighbor Joining y el modelo Kimura 2 parametros. Entre las OTUs que representa a virus de pacientes mexicanos infectados por VIH se observa una referencia del subtipo K , lo cual muestra un error del modelo, ya que el 100% de las secuencias pertenecen al subtipo B y ningún virus analizado presenta relación con el subtipo K. Con una flecha se señala la OTU que presenta mayor relación con la referencia K.

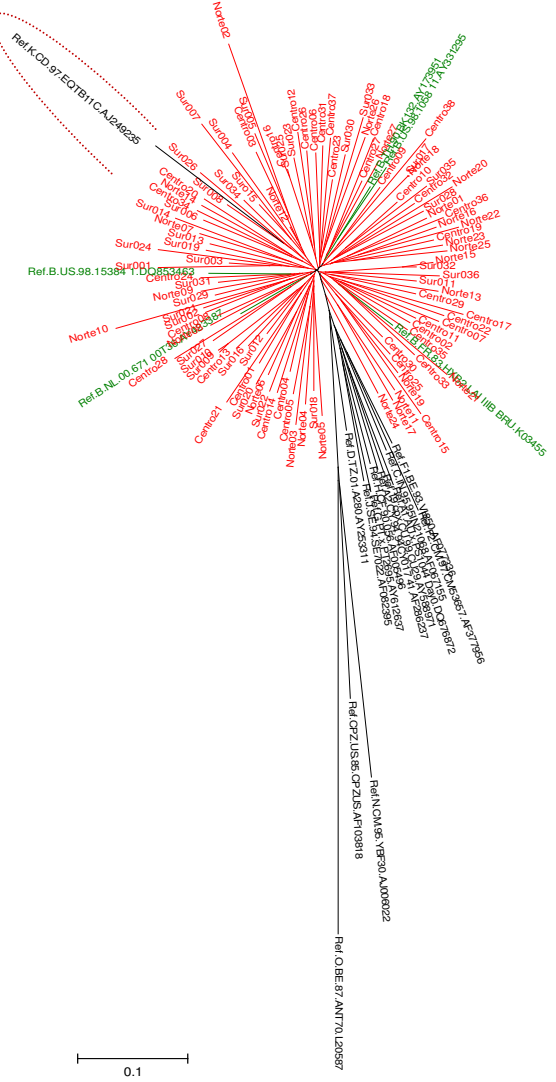
7.3.2 Pruebas para determinar el modelo de nucleótidos más óptimo

Estudios previos involucrando análisis filogenéticos de VIH en diferentes poblaciones y con diferentes objetivos han utilizado consistentemente métodos de Neighbor-Joining para inferir las filogenias y frecuentemente se ha preferido el modelo de Kimura 2 parámetros para estimar las distancias evolutivas [46,47,48,49,50]. Sin embargo, la simplicidad del modelo de Kimura podría hacer que se perdiera parte de la información evolutiva de las secuencias analizadas. Para determinar cuál era el mejor modelo de nucleótidos para representar la información del grupo de secuencias del gen *env* de VIH incluidas en el presente estudio se construyeron diferentes árboles con el método Neighbor-Joining utilizando los diferentes modelos de nucleótidos incluidos en el programa MEGA 4.1 β para estimar las distancias evolutivas (**ver figura 7.11**). Los árboles se construyeron a partir de las secuencias del gen *env* de los virus de pacientes infectados por VIH-1 que participaron en el estudio. Se incluyeron en el alineamiento al menos una secuencias referencia de cada subtipo del grupo M y secuencias referencia del grupo O y N. Todas las secuencias referencia se obtuvieron de la base de datos de los Álamos [42].

Comparación De los diferentes modelos de nucleótidos

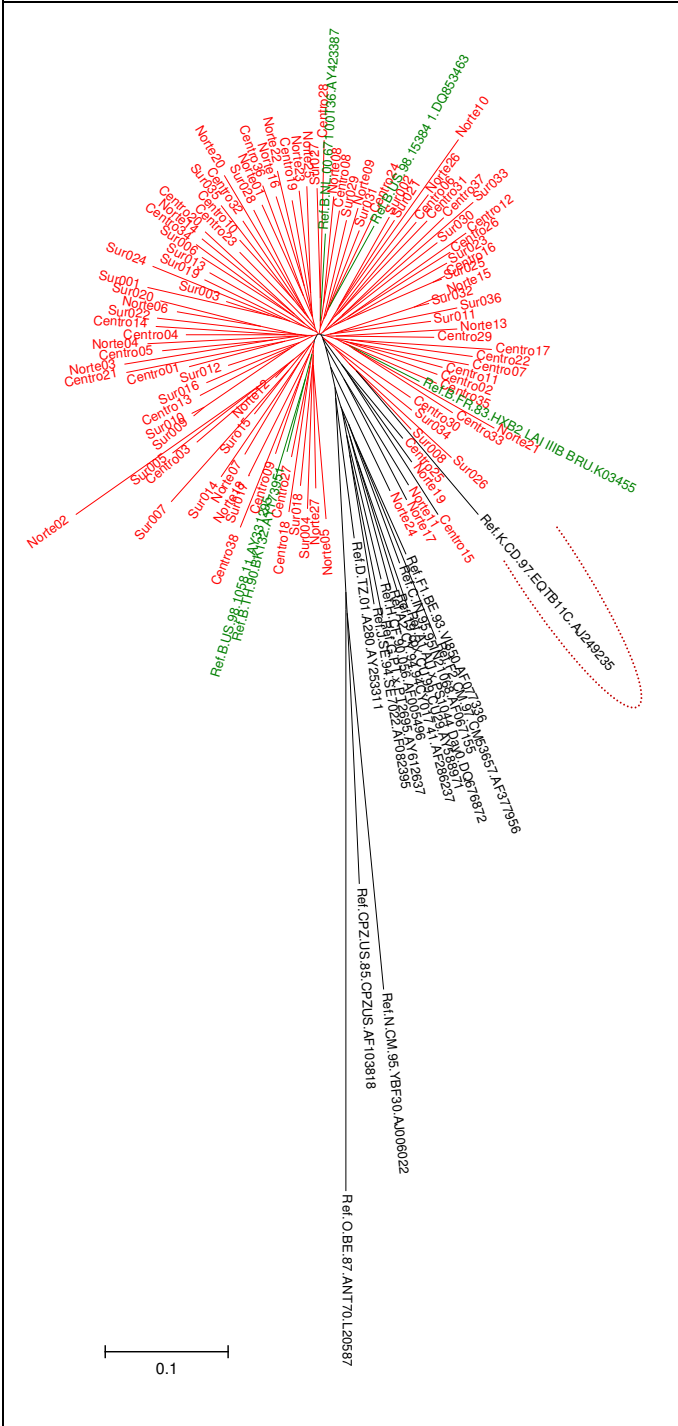


a) Jukes –Cantor (1969)

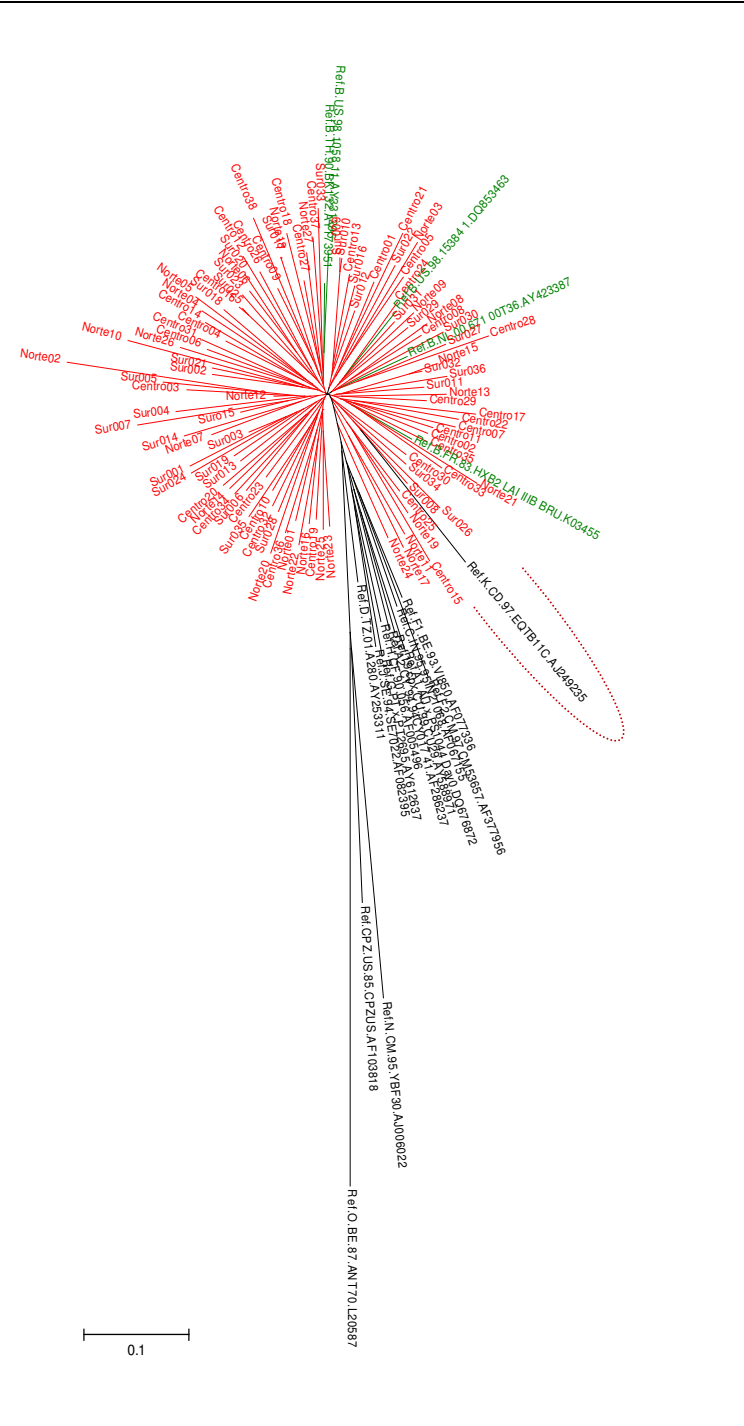


b) Kimura 2
parametros(1981)

Comparación De los diferentes modelos de nucleótidos



b)Tamura 3 Parametros (1992)

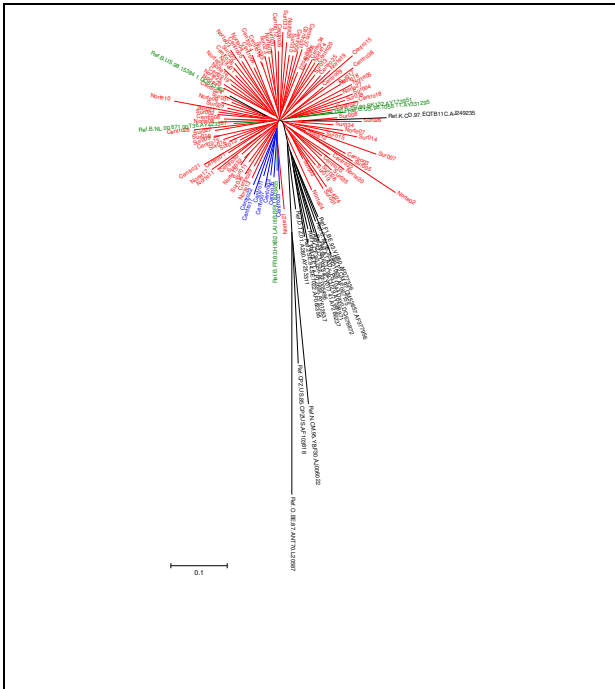


c)Tajima-Nei (1984)

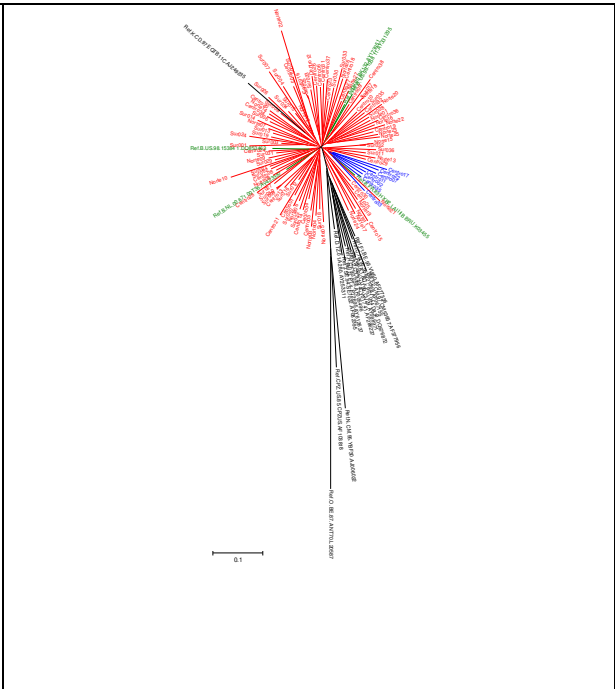
Figura 7.11 Comparación de los diferentes modelos de nucleótidos en la construcción de árboles filogenéticos a partir de secuencias del gen *env*, Relación filogenética de 120 taxas. La historia evolutiva se obtuvo usando el método Neighbor-Joining en todos los casos. Los árboles que se muestran son aquellos que mostraron una distancia genética menor al realizar la suma de todas sus ramas. En el caso del árbol de Jukes Cantor la longitud total fue de 13.14541865 , en Kimura 2 parametros fue de 13.27077729, en el caso de Tamura 3 parámetros fue de 13.37398744, con el modelo Tajima Nei fue de 13.55357458, con el modelo Tamura Kumar fue de 13.72937281 y finalmente con el modelo Maximum Composite Likelihood fue de 13.44171794. Se midió el porcentaje de réplica ó por bootstrap (1000 repeticiones), mostrándose el árbol consenso en cada caso. Todas las inserciones o deleciones que se encontraban en el alineamiento fueron eliminadas exclusivamente después de comparar con las demás secuencias del alineamiento (opción: pairwise deletion). El análisis filogenético se realizó en el programa Mega 4.1β. En rojo se muestran las secuencias de los pacientes, en negro se muestran las secuencias referencia de los subtipos No B y en verde se muestran las referencias del subtipo B. Con una línea punteada roja se indica la posición de la referencia del subtipo K en los diferentes árboles, mientras que con un asterisco se indica la posición de la referencia de subtipo K en el árbol que se construyó con el modelo Maximum Composite Likelihood.

El modelo que demuestro mayor apego a la realidad fue Maximum Composite Likelihood, debido a en el árbol filogenético ninguna OTU se relaciona cercanamente con la referencia del subtipo K (**ver figura 7.11**). Ninguna OTU puede relacionarse de forma cercana al subtipo K ya que el 100% de las secuencias pertenecen al subtipo B. Cabe aclarar que el subtipo B no guarda una relación cercana con el subtipo K.

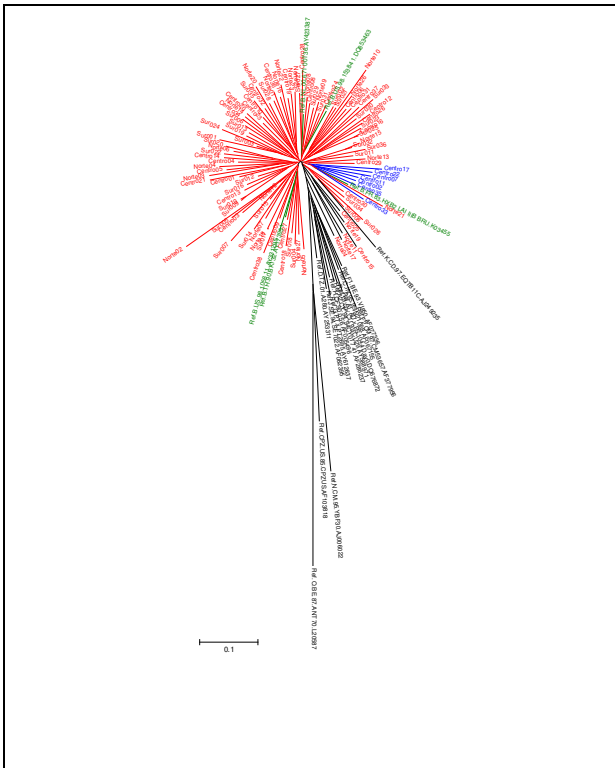
Al construir los diferentes árboles variando el modelo de nucleótidos encontramos que la topología de los árboles variaba en cada caso, a pesar de ello encontramos un agrupamiento entre virus de pacientes provenientes del centro que se repite en todos los casos. En azul se muestran las OTUs provenientes de la misma región geográfica que se agruparon, en rojo se muestran las OTUs no agrupadas, en verde se muestran las referencias del subtipo B, en negro las referencias de los subtipos no B del grupo M; las referencias de los grupos O , N ;y una referencia de SIV (**ver figura 7.12**).



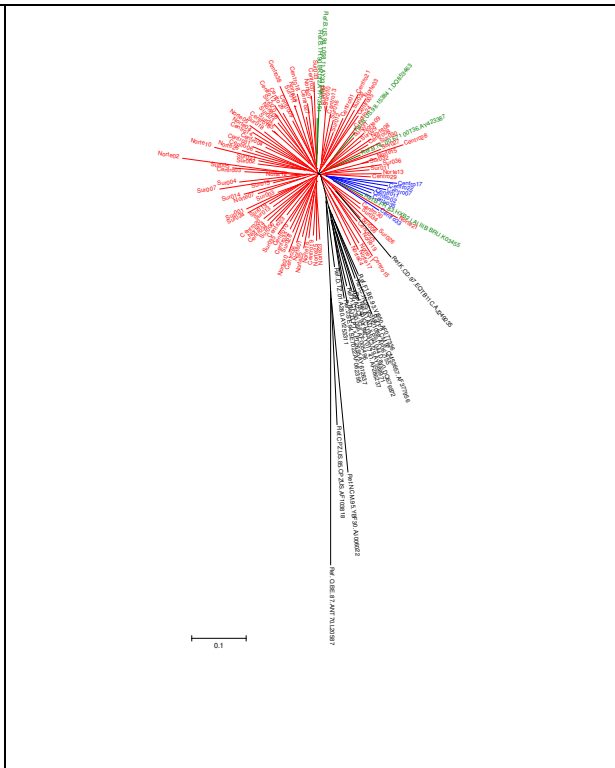
a)Jukes- Cantor



b)Kimura-2 parámetros



c)Tamura-3 parametros



d)Tajima-Nei

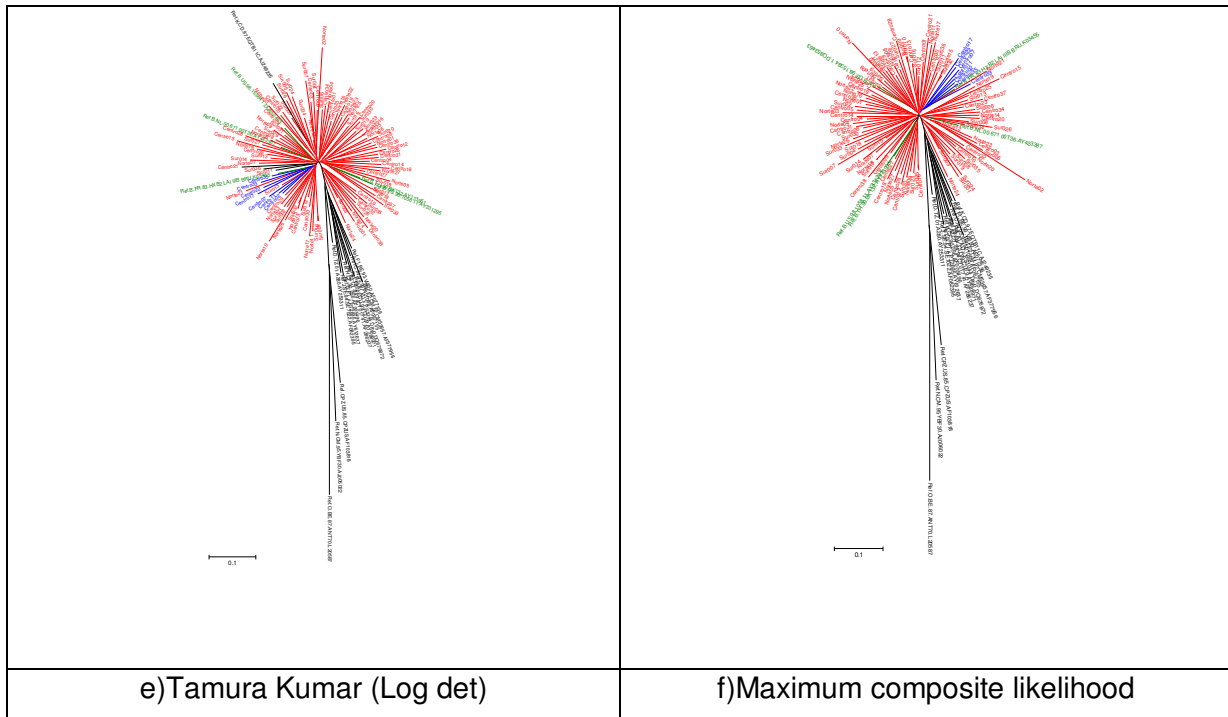


Figura.7.12 Repetición del agrupamiento de los pacientes provenientes del centro al realizar el análisis filogenético del gen *env* con diversos modelos evolutivos, Relación filogenética de 120 taxas . Relación filogenética de 120 taxas. La historia evolutiva se obtuvo usando el método Neighbor-Joining en todos los casos. Los árboles que se muestran son aquellos que mostraron una distancia genética menor al realizar la suma de todas sus ramas. En el caso del árbol de Jukes Cantor la longitud total fue de 13.14541865 , en Kimura 2 parametros fue de 13.27077729, en el caso de Tamura 3 parámetros fue de 13.37398744, con el modelo Tajima Nei fue de 13.55357458, con el modelo Tamura Kumar fue de 13.72937281 y finalmente con el modelo Maximum Composite Likelihood fue de 13.44171794. Se midió el porcentaje de réplica ó por bootstrap (1000 repeticiones), mostrándose el árbol consenso en cada caso. Todas las inserciones o deleciones que se encontraban en el alineamiento fueron eliminadas exclusivamente después de comparar con las demás secuencias del alineamiento (opción: pairwise deletion). El análisis filogenético se realizó en el programa Mega 4.1β. En azul se encuentran las OTUs agrupadas que provienen del Centro de la República Mexicana ,en rojo se encuentran las OTUs no agrupadas de los pacientes, en negro las referencias de subtipos no B del grupo las referencias de VIH-1 Ny O : así como la referencia de SIV..

7.3.3 Análisis filogenético de las secuencias del gen *env* y *pol* de VIH circulante en México mediante un modelo de distancias optimizado.

Debido a que el modelo Maximum Composite Likelihood fue el modelo que representó con mayor exactitud las relaciones filogenéticas entre los virus de pacientes mexicanos provenientes de las diversas regiones de la República

Mexicana, se construyeron los árboles filogenéticos con los genes *env* (**ver figura 7.13**) y *pol* (**ver figura 7.14**) utilizando este modelo.

En los árboles de *env* y *pol* es posible observar que todas las OTUs de los virus provenientes de pacientes mexicanos se agruparon en una misma rama. En el caso del árbol de *env* las OTUs de virus provenientes de pacientes mexicanos se intercalaron con las referencias de las formas recombinantes BF y AB (**ver figura 7.13**). En el árbol de *env* es posible observar que existe una OTU (Sur036) que se aleja de todas las demás, lo cual nos indica que existe una mayor distancia filogenética entre esta secuencia y las demás.

En el caso del árbol de *pol* las OTUs provenientes de virus de pacientes mexicanos únicamente se intercalaron con referencias de subtipo B, las referencias de recombinantes B en ningún caso se agruparon con las OTUs de virus de pacientes mexicanos (**ver figura 7.14**).

Una de las diferencias más notorias entre el árbol de *pol* y *env* fue que en el caso de las ramas del árbol de *env*, las ramas fueron más largas. Mientras que en el árbol de *pol* las ramas presentaron una menor longitud evidenciando la mayor homogeneidad de las secuencias del gen *pol* en comparación con las secuencias del gen *env* (**ver figura 7.13 y 7.14**).

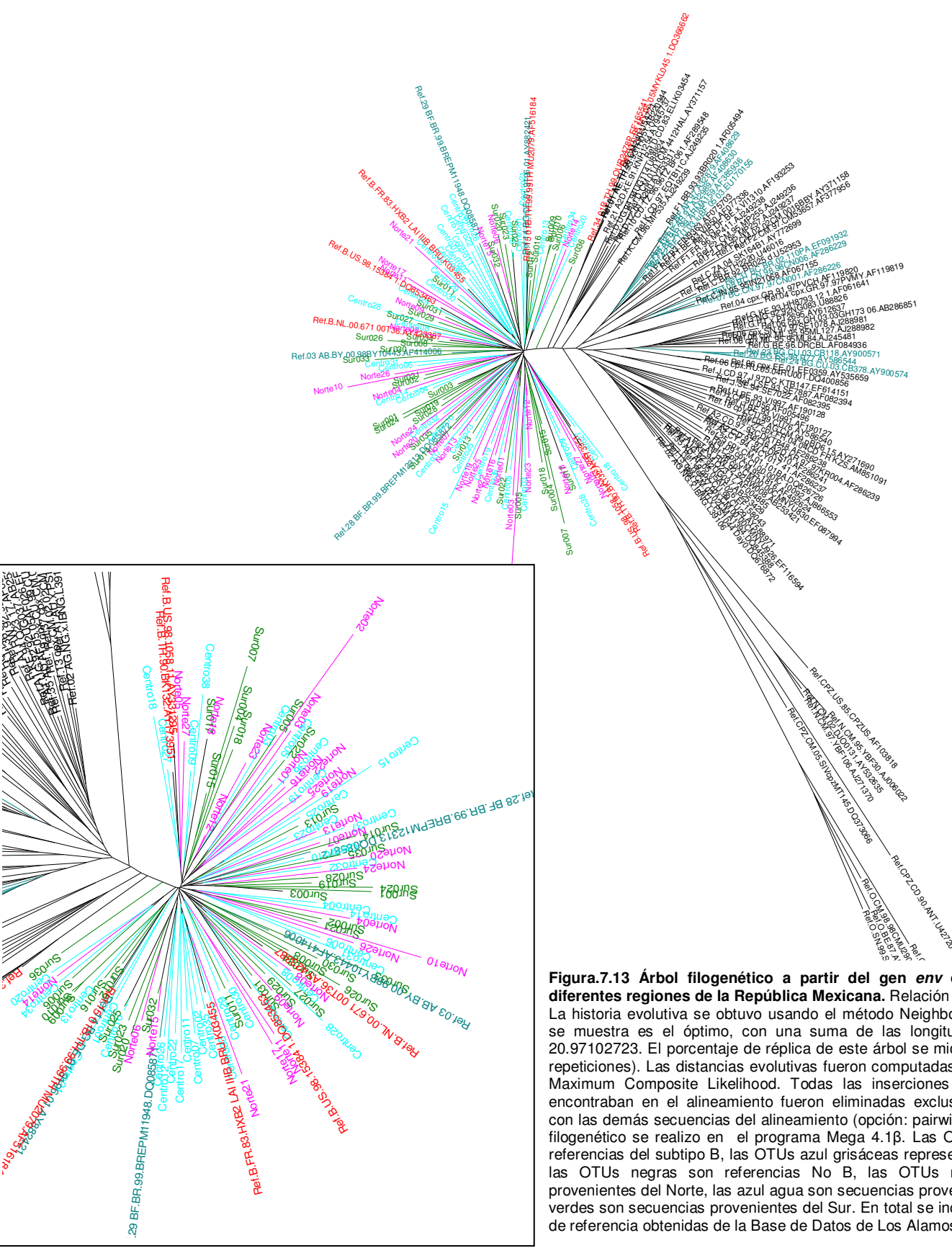


Figura.7.13 Árbol filogenético a partir del gen *env* de VIH circulante en diferentes regiones de la República Mexicana. Relación evolutiva de 194 taxas. La historia evolutiva se obtuvo usando el método Neighbor-Joining. El árbol que se muestra es el óptimo, con una suma de las longitudes de las ramas de 20.97102723. El porcentaje de réplica de este árbol se midió por bootstrap (1000 repeticiones). Las distancias evolutivas fueron computadas por medio del modelo Maximum Composite Likelihood. Todas las inserciones o deleciones que se encontraban en el alineamiento fueron eliminadas exclusivamente al comparar con las demás secuencias del alineamiento (opción: pairwise deletion). El análisis filogenético se realizó en el programa Mega 4.1β. Las OTUs rojas representan referencias del subtipo B, las OTUs azul grisáceas representan recombinantes B, las OTUs negras son referencias No B, las OTUs rosas son secuencias provenientes del Norte, las azul agua son secuencias provenientes del Centro y las verdes son secuencias provenientes del Sur. En total se incluyeron 93 secuencias de referencia obtenidas de la Base de Datos de Los Alamos [43]

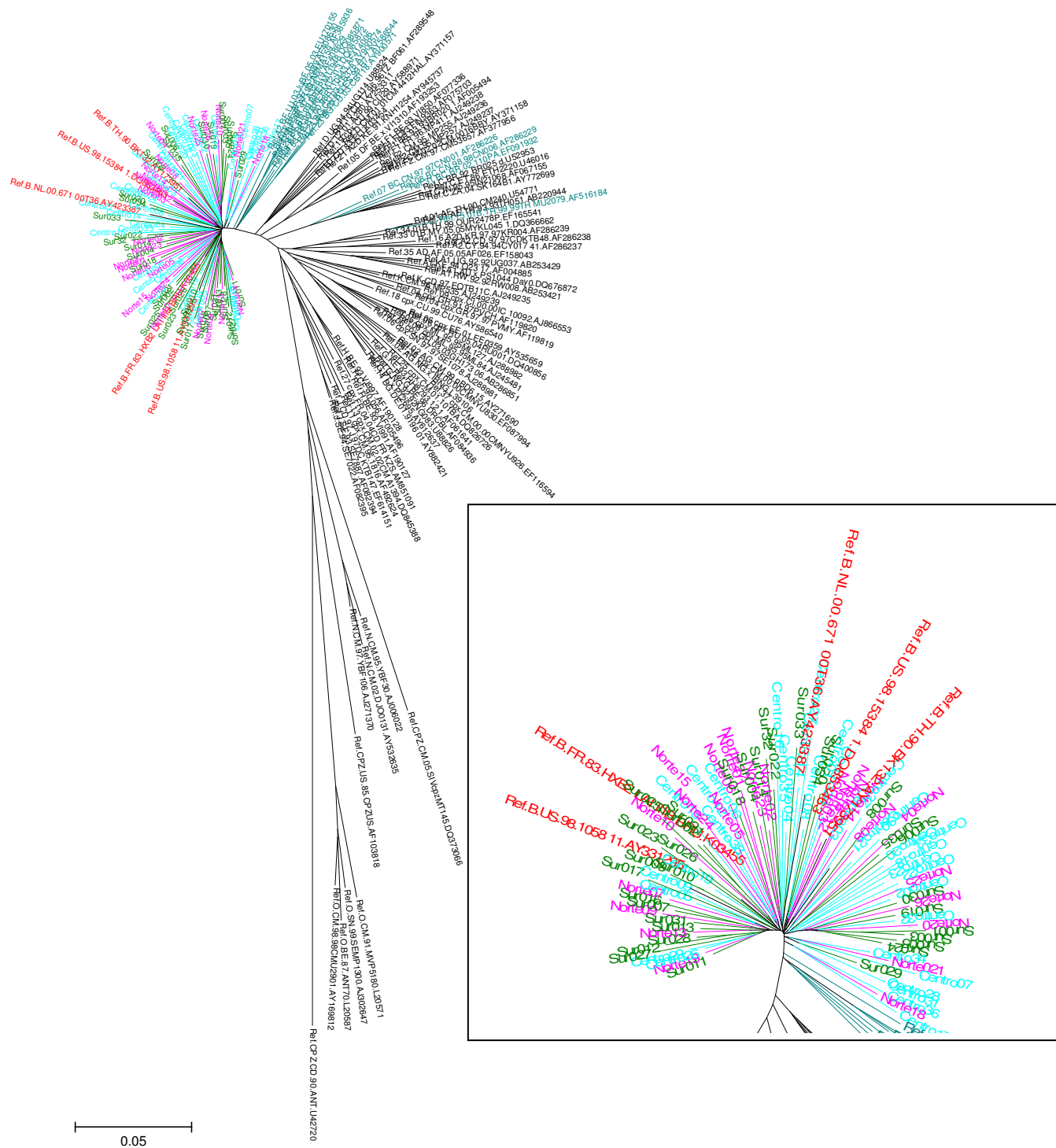


Figura.7.14 Árbol filogenético a partir del gen *pol* de VIH circulante en diferentes regiones de la República Mexicana. Relación evolutiva de 194 taxas. La historia evolutiva se obtuvo usando el método Neighbor-Joining. El árbol que se muestra es el óptimo, con una suma de las longitudes de las ramas de 6.22409539. El porcentaje de réplica de este árbol se midió por bootstrap (1000 repeticiones). Las distancias evolutivas fueron computadas por medio del modelo Maximum Composite Likelihood. Todas las inserciones o deleciones que se encontraban en el alineamiento fueron eliminadas exclusivamente al comparar con las demás secuencias del alineamiento (opción: pairwise deletion). El análisis filogenético se realizó en el programa Mega 4.1β. Las OTUs rojas representan referencias del subtipo B, las OTUs azul grisáceas representan recombinantes B, las OTUs negras son referencias No B, las OTUs rosas son secuencias provenientes del Norte, las azul agua son secuencias provenientes del Centro y la verdes son secuencias provenientes del Sur. En total se incluyeron 93 secuencias de referencia obtenidas de la Base de Datos de Los Alamos [43]

Como ya se había mencionado anteriormente, en el árbol de *env* se observó que algunas OTUs de recombinantes B se agruparon con OTUs de virus mexicanos. La existencia de recombinantes B se descartó analizando con el programa RIP [45] las secuencias de virus que mostraban menor distancia filogenética con las referencias de recombinantes. En las figuras 7.15, 7.16 y 7.17 se muestran las gráficas de similitud de las secuencias que mostraban menor distancia filogenética con algún recombinante B. Como es posible observar ninguna se subtípificó como recombinante. Todas las secuencias pertenecieron al subtipo B.

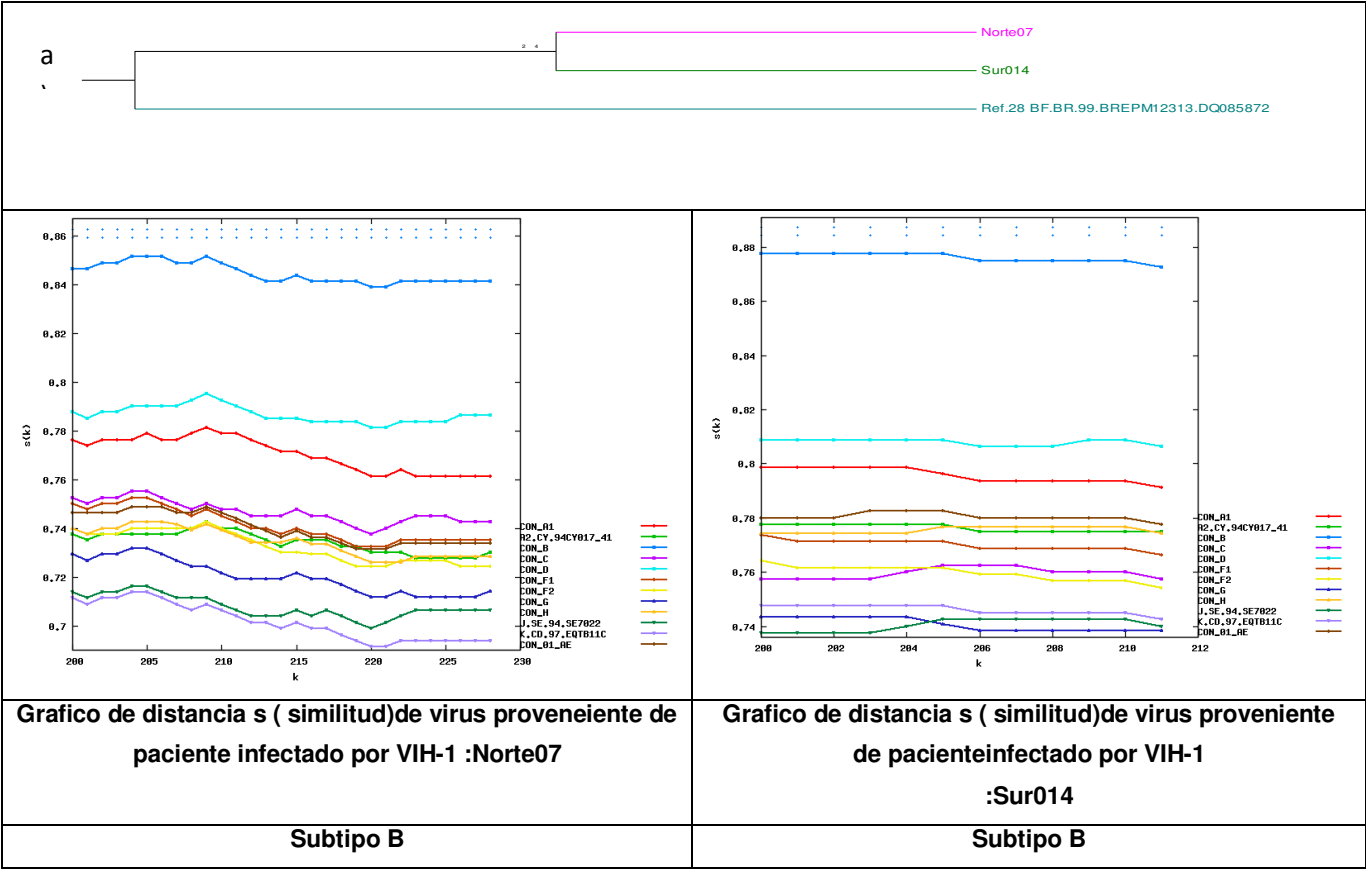


Figura.7.15 Análisis de secuencias cercanas a referencia de recombinante BFBR.99BREPM.2313.DQ85872. Las secuencias cercanas a la referencia BFBR.99BREPM.2313.DQ85872 se analizaron con el programa RIP[45]a) Se muestra la rama que contiene una referencia de recombinante y virus mexicanos (Norte07 y Sur014). Debajo de la rama se muestran los gráficos de similitud que demuestran que los virus Norte07 y Sur014 pertenecen al subtipo B.

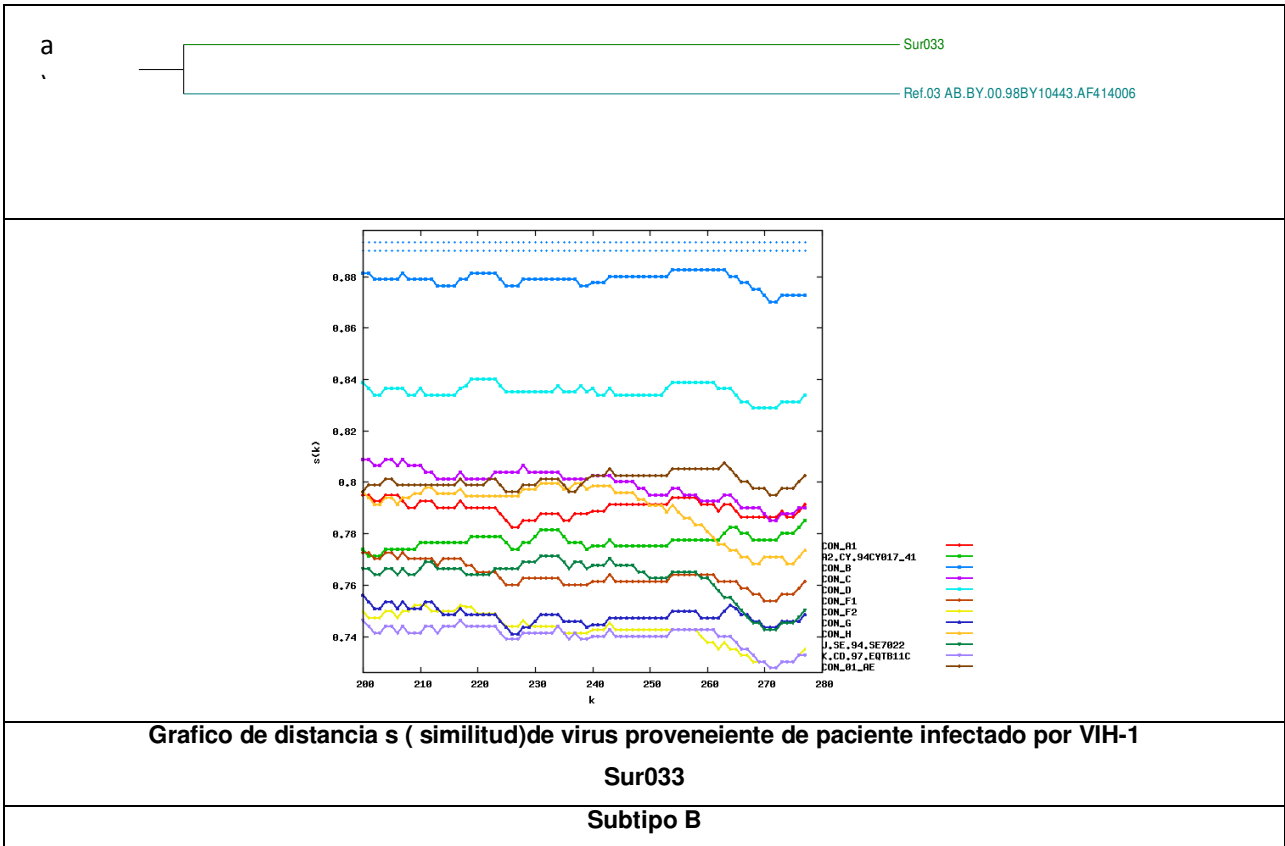


Figura.7.16. Analisis de secuencia cercana a referencia de recombinante AB.BY.00.98BY10443.AF414006. Las secuencia cercana a la referencia **AB.BY.00.98BY10443.AF414006** se analizo con el programa RIP[45]a) Se muestra la rama que contiene una referencia de recombinante y el virus mexicano (Sur033). Debajo de la rama se muestra el gráfico de similitud que demuestra que el virus Sur033 pertenece al subtipo B.

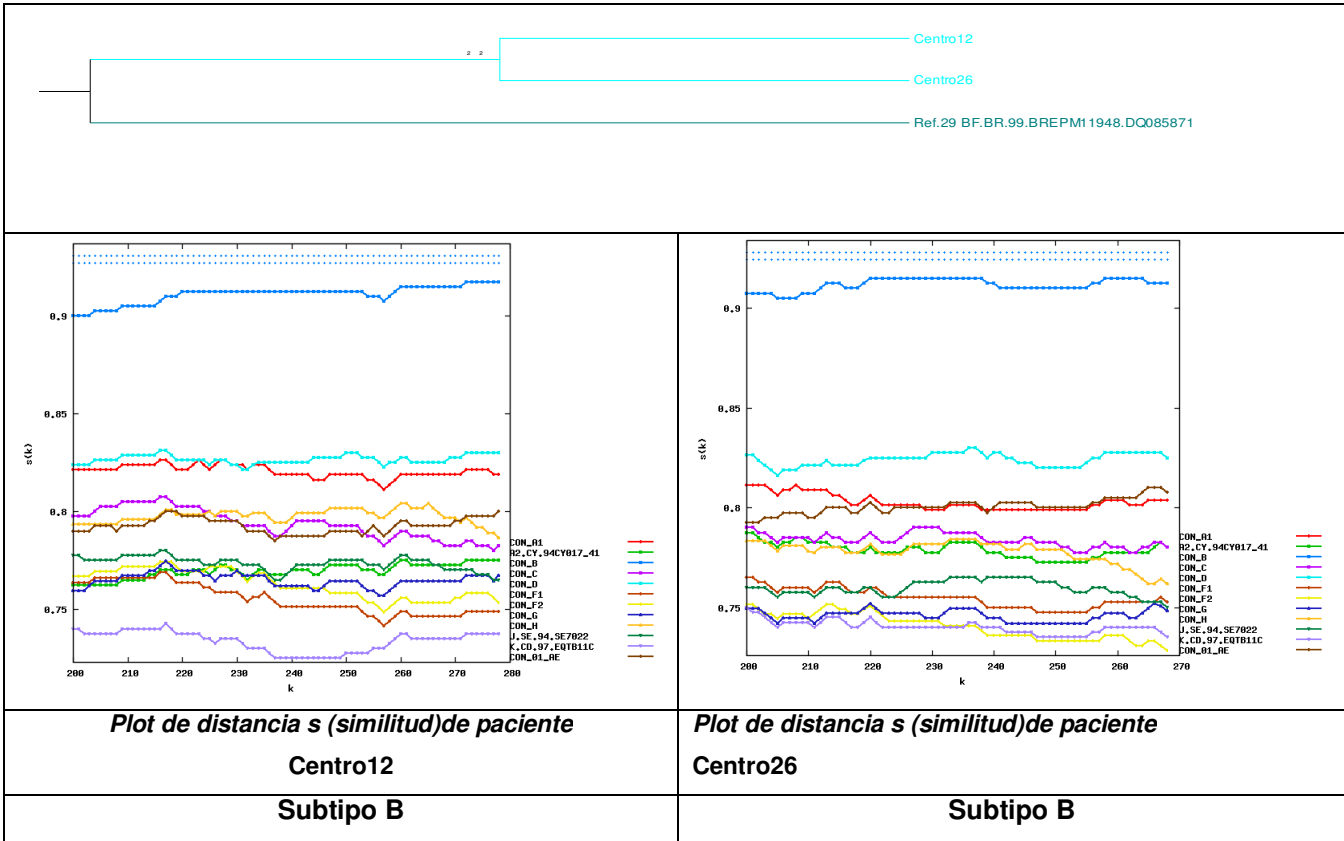


Figura.7.17 Análisis de secuencias cercanas a referencia de recombinante BFBR.99.BREPM11948.DQ85871. Las secuencias cercanas a la referencia BFBR.99.BREPM11948.DQ85871 se analizaron con el programa RIP [45]a) Se muestra la rama que contiene una referencia de recombinante y virus mexicanos (Centro12 y Centro 26). Debajo de la rama se muestran los gráficos de similitud que demuestran que los virus Centro 12 y Centro 26 pertenecen al subtipo B.

En el árbol de env es posible observar una agrupación entre 7 OTUs de virus de pacientes provenientes del Centro de la República Mexicana (**ver figura 7.18**). Los pacientes infectados por estos virus provienen de Jalisco:43%(3) , Morelos 29% (2), D.F. 14% (1) y del Estado de México 14% (1) (**ver figura 7.19**).

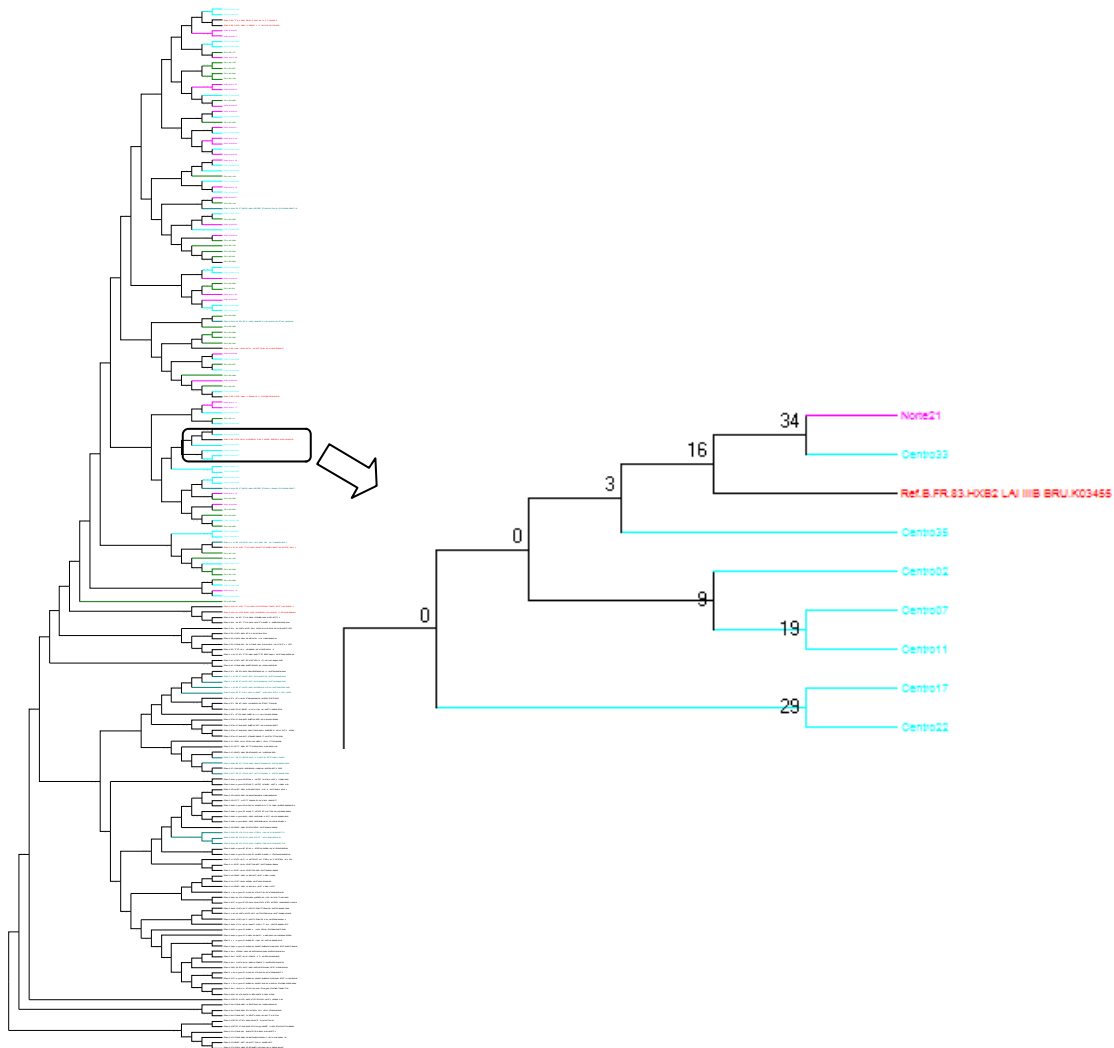


Figura.7.18 Agrupación de 8 virus provenientes del Centro de la República Mexicana analizando el gen *env* Relación evolutiva de 194 taxas. La historia evolutiva se obtuvo usando el método Neighbor-Joining. a) El árbol que se muestra es el óptimo, con una suma de las longitudes de las ramas de 20.97102723. El porcentaje de réplica de este árbol se midió por bootstrap (1000 repeticiones). Las distancias evolutivas fueron computadas por medio del modelo Maximum Composite Likelihood. Todas las inserciones o deleciones que se encontraban en el alineamiento fueron eliminadas exclusivamente al comparar con las demás secuencias del alineamiento (opción: pairwise deletion). El análisis filogenético se realizó en el programa Mega 4.1β. Las OTUs rojas representan referencias del subtipo B, las OTUs azul grisáceas representan recombinantes B, las OTUs negras son referencias No B, las OTUs rosas son secuencias provenientes del Norte, las azul agua son secuencias provenientes del Centro y las verdes son secuencias provenientes del Sur. En total se incluyeron 93 secuencias de referencia obtenidas de la Base de Datos de Los Alamos [43]b) A la izquierda se muestra una agrupación de OTUs provenientes del Centro de la República Mexicana al lado de cada rama se observan los valores de bootstrap, estos son menores al 70%.

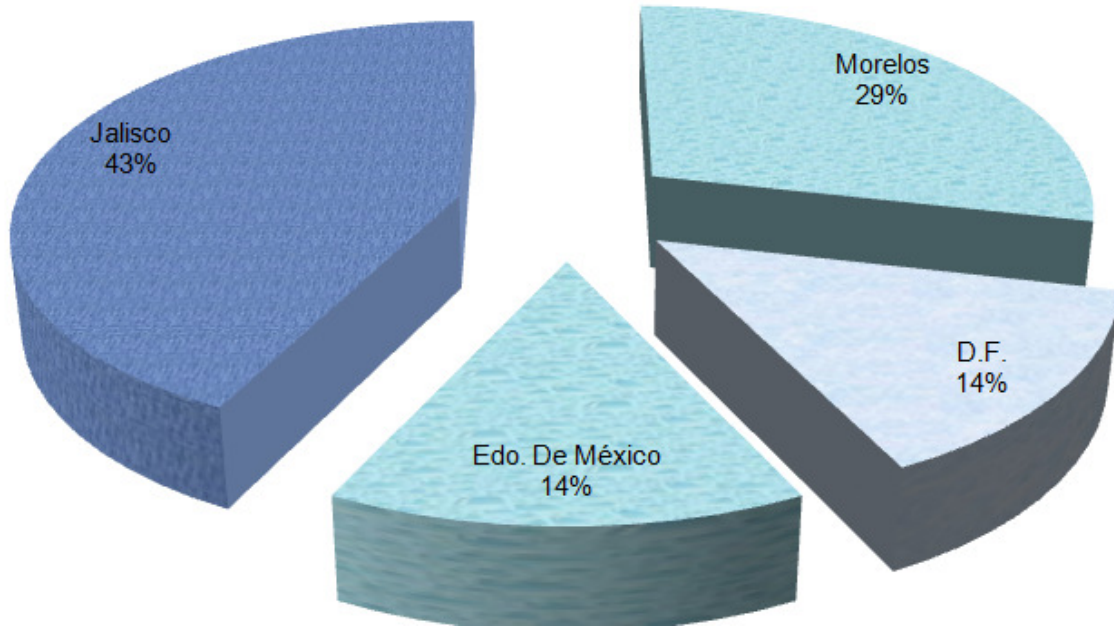


Figura.7.19 Procedencia de pacientes agrupados del Centro del País . Estos provenían de: Jalisco:43%(3) de , Morelos 29%(2), D.F. 14%(1) y del Estado de México14% (1).

El 28.57% de los pacientes que formaron el agrupamiento en el árbol eran mujeres, mientras que el 71.43% eran hombres. La mediana de células CD4+ fue de 190 cel/ μ l, esta fue muy similar a la mediana de los pacientes no agrupados del Centro del Norte y del Sur (**ver figura 7.20**)

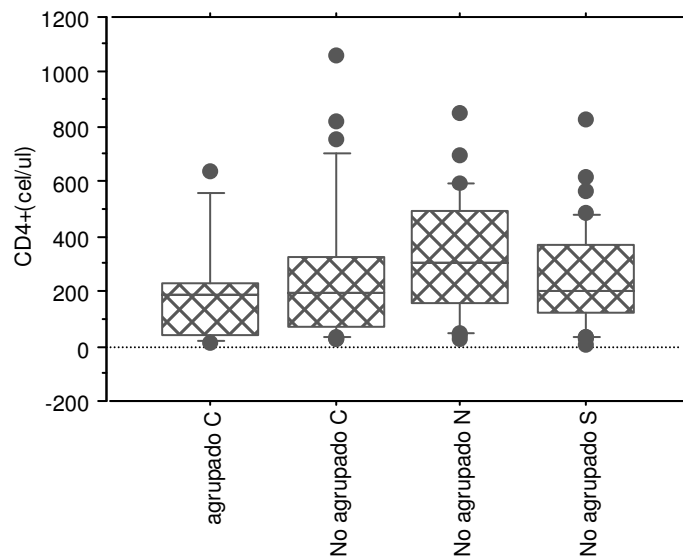


Figura.7.20 Comparación de la mediana de células CD4+ de pacientes agrupados del Centro. Se realizó dicha comparación por medio de diagramas de cajas y bigotes. El rectángulo caja representa los cuartiles Q1 y Q3. La línea horizontal que divide al rectángulo representa el cuartil

Q2 el cual es la mediana. Los bigotes representan los límites inferiores y superiores. Los puntos que se salen de los límites superiores e inferiores son datos atípicos.

Por medio de la prueba Kruskal Wallis se demostró que no existía diferencia significativa entre los pacientes agrupados y los no agrupados de las regiones Centro, Norte y Sur de la República Mexicana; ya que el valor corregido de P para los rangos es mayor a $\alpha=0.05$ por tanto se acepta la hipótesis nula (H_0), la cual nos indica que no existe diferencia significativa entre las medias de los pacientes agrupados del centro y los no agrupados del Centro, Norte y Sur de la República Mexicana (ver tabla 7.7 y tabla 7.8).

Tabla 7.7 Valores de los parámetros obtenidos por la prueba Kruskal Wallis aplicada a los valores de células CD4⁺ de pacientes agrupados del Centro y no agrupados del Centro, Norte y Sur de la República Mexicana.

Prueba Kruskal-Wallis CD4 ⁺ (cel/ml)	
Grupo variable: Agrupación	
Grados de libertad	3
# Grupos	4
#Rangos	1
H	3.075
Valor de P	0.3803
H corregida para los rangos	3.075
Corrección del valor de P de los rangos	0.3803

Tabla 7.8 Información de los rangos obtenidos por la prueba Kruskal Wallis aplicada a los valores de células CD4⁺ de pacientes agrupados del Centro y no agrupados del Centro, Norte y Sur de la República Mexicana.

	#	Σ Rango	Rango promedio
Agrupado C	7	273.000	39.000
No Agrupado C	31	1486.000	47.935
No Agrupado N	27	1567.000	58.037
No Agrupado S	36	1825.000	50.694

Al comparar la mediana del $\text{Log}_{10}(\text{carga viral})$ de los pacientes agrupados del centro, encontramos que era muy similar a la mediana de los pacientes no agrupados del Centro, del Norte y del Sur de la República Mexicana (**ver figura 7.21**)

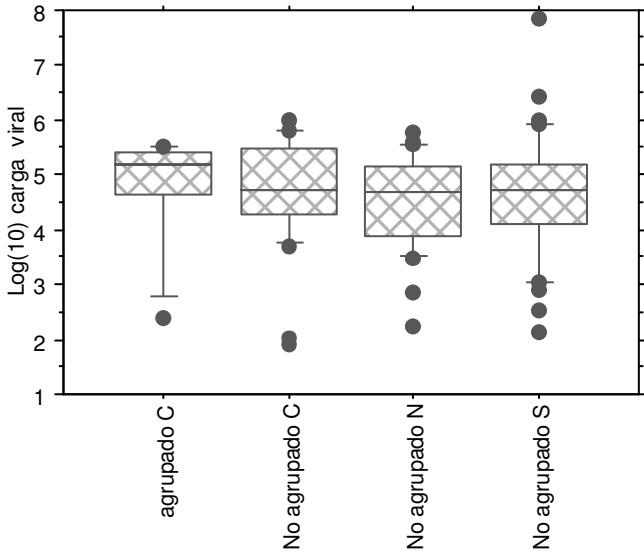


Figura.7.21 Comparación de la mediana de $\text{Log}_{10}(\text{carga viral})$ de pacientes agrupados del Centro. Se realizó dicha comparación por medio de diagramas de cajas y bigotes. El rectángulo caja representa los cuartiles Q1 y Q3. La línea horizontal que divide al rectángulo representa el cuartil Q2 el cual es la mediana. Los bigotes representan los límites inferiores y superiores. Los puntos que se salen de los límites superiores e inferiores son datos atípicos

Por medio de la prueba Kruskal Wallis se demostró que no existía diferencia significativa entre los pacientes agrupados y los no agrupados de las regiones Centro, Norte y Sur de la República Mexicana; ya que el valor corregido de P para los rangos es mayor a $\alpha=0.05$ por tanto se acepta la hipótesis nula (H_0), la cual nos indica que no existe diferencia significativa entre las medias de los pacientes agrupados del centro y los no agrupados del Centro, Norte y Sur de la República Mexicana (**ver tablas 7.9 y 7.10**).

Tabla 7.9 Valores de los parámetros obtenidos por la prueba Kruskal Wallis aplicada a los valores de \log_{10} (carga viral) de pacientes agrupados del Centro y no agrupados del Centro, Norte y Sur de la República Mexicana.

Prueba Kruskal-Wallis \log_{10}(Carga viral)	
Grupo variable: Agrupación	
Grados de libertad	3
# Grupos	4
#Rangos	16
H	1.391
Valor de P	0.7076
H corregida para los rangos	1.391
Valor de P corregido de los rangos	0.7075

Tabla 7.10 Información de los rangos obtenidos por la prueba Kruskal Wallis aplicada a los valores de \log_{10} (carga viral) de pacientes agrupados del Centro y no agrupados del Centro, Norte y Sur de la República Mexicana.

	#	Σ Rango	Rango promedio
Agrupado C	7	273.000	39.000
No Agrupado C	31	1486.000	47.935
No Agrupado N	27	1567.000	58.037
No Agrupado S	36	1825.000	50.694

8.0 Discusión de Resultados

Al inicio del proyecto se planteó como objetivo analizar por lo menos 100 secuencias del gen *env*, al término de éste se lograron analizar 101. Para la realización de este análisis se tomó una muestra de pacientes de la cohorte de Epidemiología Molecular del Centro de Investigación en Enfermedades Infecciosas.

La elección de los pacientes participantes se realizó aleatoriamente, únicamente se trató de mantener una proporción similar entre el número de pacientes del Centro (37%), Norte (27%) y Sur (36%) de la República Mexicana. La edad promedio de los pacientes fue de 34.7 años lo cual concuerda con el rango de edad que se encuentra mayormente afectada por VIH, es decir personas entre 30-40 años [51]. Con relación al porcentaje de hombres y mujeres participantes en el estudio, se tiene un 67.33% de hombres y un 32.67% de mujeres. El porcentaje de mujeres es ligeramente mayor al reportado (29%) [3], lo cual puede ser un reflejo de la feminización de la epidemia en México o un simple sesgo de muestreo para la cohorte del presente estudio.

Los pacientes participantes en el estudio en general se encontraban en una fase avanzada de la enfermedad, ya que la mediana de células CD4⁺ fue de 212 células / μ l y la mediana de la carga viral fue de 52,846 partículas virales/ml. No se encontró diferencia significativa entre el conteo de células CD4⁺ (**ver tabla 7.4**) y log₁₀carga viral (**ver tabla 7.2**) entre los pacientes provenientes del Centro, Norte y Sur de la República Mexicana lo cual nos indica que nuestra muestra es homogénea y no existen sesgos de muestreo entre los pacientes provenientes de las diferentes regiones analizadas.

Uno de los objetivos principales fue determinar el subtipo de VIH predominante en el país utilizando el gen *env*. Como antecedente se tenía el estudio de Epidemiología Molecular de VIH más grande realizado en el país (1500

secuencias del gen *pol*), el cual se realizó en el laboratorio CIENI, lugar donde se realizó el presente trabajo. En este estudio se analizaron 101 secuencias del gen *env*, las cuales se subtipificaron por 3 diferentes métodos: El primer método para subtipificar las secuencias fue utilizando el programa REGA [44], con este programa únicamente fue posible subtipificar el 12% de las secuencias, el programa se ve limitado ya que cuenta con pocas secuencias referencia. Este hecho limita la subtipificación ya que el programa construye un alineamiento con las secuencias referencia y la secuencia problema. El alineamiento resultante se utiliza para inferir la relación filogenética entre la secuencia problema y las referencias por medio de la construcción de un árbol filogenético. El segundo método que se utilizó para subtipificar las secuencias fue por el programa RIP [45], con este programa fue posible subtipificar un 87% de las secuencias. La limitante que presentó fue que requiere secuencias mayores a 400 pb, por esta razón un 13 % de las secuencias no se pudieron subtipificar por este método. Las secuencias no subtipificadas por REGA y RIP se subtipificaron construyendo un árbol filogenético con al menos una secuencia referencia de cada subtipo del grupo M y una referencia del grupo M y O, así como una referencia de SIV (**ver figura 7.6**). En este árbol fue posible observar que todas las OTUs de virus mexicanos se agrupaban en una misma rama de la cual se desprendían las referencias del subtipo B. A pesar de ello se observaron dos OTUs que se alejaban del resto (Sur003y Sur005). Por tal razón se decidió construir un árbol filogenético únicamente compuesto por secuencias de los virus mexicanos y una secuencia referencia de SIV, para determinar la relación filogenética entre los virus que se encuentran alejados (Sur003 y Sur005) y los virus mexicanos ya subtipificados. Esto para correlacionar únicamente los virus mexicanos entre sí. Mediante este análisis es claro que el virus Sur003 se relacionó cercanamente con el virus Sur001, el cual pertenece al subtipo B. El virus Sur001, a su vez se relacionó cercanamente con el virus Norte02 el cual pertenece también al subtipo B (**ver**

figura 7.7).Estos análisis sugieren fuertemente que ambos virus pertenecen al subtipo B.

Una de las razones por las cuales se encontraron alejadas estas dos secuencias de el resto de las secuencias referencias en la **figura 7.6** podría deberse a que ninguna referencia del subtipo B es mexicana , pudiendo representar variantes locales ligeramente divergentes. Otro factor por lo cual se muestran alejadas las secuencias Sur003 y Sur005 podría ser que éstas secuencias representen un cluster de diferenciación distinto en los virus mexicanos. A pesar de esta posibilidad las secuencias pertenecieron al subtipo B ya que el rango de variación intra-clado es amplio de 15-20%[16].

El 100% de las secuencias del gen *env* se subtipificaron como B, lo cual concuerda con los resultados obtenidos en un estudio previo realizado en el CIENI, en el cual se analizaron más de 1500 secuencias del gen *pol*. Se cree que la epidemia en nuestro país está fuertemente influenciada por la epidemia de Estados Unidos debido a los fenómenos migratorios. Sin embargo en Estados Unidos se ha reportado un aumento en la prevalencia de virus pertenecientes a los subtipos no B [22]. Una de las razones por las cuales no se puede observar este fenómeno en los estudios realizados en México, se puede deber a que el diagnóstico se realiza de forma tardía y a que el aumento en la prevalencia de subtipos no B sea reciente. Así los individuos infectados por virus pertenecientes a subtipos no B aun no habrían sido diagnosticados. Otra posibilidad es que las redes de transmisión en la población mexicana sean estrechas, en este sentido, los migrantes que adquieren la infección en Estados Unidos tenderían a infectarse por otros mexicanos. Fue además la alta homogeneidad de los virus circulantes en México reflejaría una baja introducción de subtipos no B por migrantes de otras regiones del mundo.

Se realizó una comparación de los árboles filogenéticos de los genes *env* y *pol* de los virus participantes en el estudio (**ver figura 7.8**). El método que se utilizó fue

Neighbor Joining y el modelo Kimura 2 parámetros debido a que es el modelo evolutivo más utilizado en los análisis filogenéticos del VIH en la literatura. En general encontramos que tanto en el árbol de *env* como en el de *pol* las OTUs del Centro, Norte y Sur se distribuyen homogéneamente. A pesar de ello en el árbol de *env* fue posible observar ciertas agrupaciones entre algunas OTUs. En este árbol fue posible observar un agrupamiento entre 7 virus provenientes del Centro y algunos pequeños agrupamientos entre virus provenientes del Sur. Algo que resultó inesperado fue que en el árbol de *env* se observó una agrupación entre el virus mexicano Sur008 con una referencia del subtipo K. Esto no corresponde con los resultados obtenidos en la subtipificación ya que el virus Sur008 pertenece al subtipo B y éste no mantiene una relación filogenética cercana con el subtipo K (**ver figura 8.1**). Es sabido que el subtipo K se encuentra fuertemente relacionado con el subtipo F [52] y no así con el subtipo B (**ver figura 8.2**), por tanto esta relación podría reflejar un artefacto del modelo.

Sur008	181bp	Report	HIV-1 Subtype B	98.0	<p>Approximate recombination pattern with > 75% bootstrap confidence © REGA HIV-1 Subtyping Tool</p>
--------	-------	------------------------	-----------------	------	---

Figura 8.1 Reporte de subtipificación de virus Sur008 por medio del programa REGA. En dicho reporte se muestra el nombre de la secuencia analizada, el tamaño de la secuencia, el subtipo al que pertenece, el porcentaje de similitud con la secuencia referencia del subtipo y un diagrama que ilustra la región que se subtipificó.

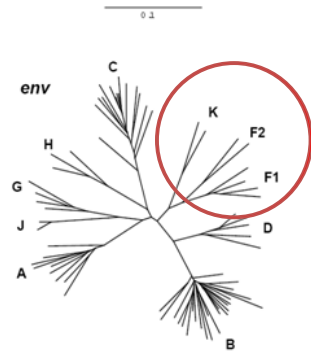


Figura8.2 Relación filogenética a partir de secuencias del gen *env* entre los diferentes subtipos del grupo M de VIH-1. Como se puede observar el subtipo K tiene mayor relación con el subtipo F [52].

La razón por la cual se obtuvo la relación entre una referencia K y el virus mexicano perteneciente al subtipo B (Sur008) (**ver figura 7.10**) podría deberse a que el modelo Kimura 2 parámetros considera únicamente la tasa de transversión (cambio de una purina a una pirimidina y viceversa) y la transición (cambio de una purina a purina), sin considerar que el VIH tiene una proporción desequilibrada entre purinas y pirimidinas, ya que las secuencias de VIH son ricas en adenina (**ver figura 8.3**).

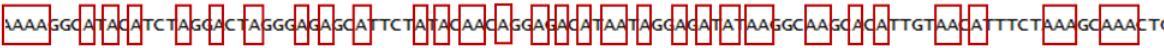


Figura8.3 Fragmento de secuencia del gen *env* del VIH. Las secuencias de VIH tienen una proporción desequilibrada entre purinas y pirimidinas ya que las secuencias son ricas en adenina. En rojo se indican las adeninas.

Debido a que el modelo Kimura 2 parámetros presenta una relación errónea entre la referencia de subtipo K y el virus mexicano Sur008, se decidió construir diferentes árboles variando el modelo de nucleótidos. Esto se realizó para determinar cuál era el modelo óptimo para el análisis filogenético de las secuencias del gen *env*.

Al comparar la topología de cada árbol nos dimos cuenta que variaba en cada caso. A pesar de ello fue posible distinguir un agrupamiento entre siete pacientes

del Centro en todos los arboles (**ver figura 7.12**). Esto sugiere fuertemente que este agrupamiento es real.

El primer árbol que se construyó para comparar los modelos fue el árbol en el cual se usó el modelo Jukes-Cantor (**ver figura 7.11 a**). Este modelo asume una única tasa de sustitución de nucleótidos por lo que es el más sencillo. Este árbol mostró una topología similar a la observada en el árbol construido por el modelo Kimura 2 parámetros y también mostró una relación cercana entre el virus Sur008 y la referencia del subtipo K. Por esta razón se consideró que este modelo no era el adecuado para representar las relaciones filogenéticas entre los virus mexicanos. El segundo modelo a comparar fue el modelo Tamura 3 parámetros, este modelo es una modificación del modelo Kimura 2 parámetros que considera la tasa de transversiones, tasa de transiciones y la proporción GC y AT como parámetros [17]. A pesar de considerar la proporción de GC y AT lo cual se adapta más a la realidad de las secuencias de VIH, continuo apareciendo una relación cercana entre la referencia K y el virus Sur008 (**ver figura 7.11 c**). Al compararse el modelo Tajima Nei y Tamura Kumar (Log det) encontramos la misma relación entre el virus Sur008 y la referencia del subtipo K (**ver figura 7.11 d y e**). El único modelo que no mostró esta relación fue el modelo Maximum Composite Likelihood (**ver figura 8.4**). En este modelo se obtiene la sumatoria del Log de verosimilitud (likelihood) para todos los pares de secuencias del alineamiento, la cual se maximiza cambiando los parámetros comunes en el patrón de sustitución de los nucleótidos en cada par de secuencias. El modelo Maximum Composite Likelihood difiere de los modelos que calculan las distancias filogenéticas de forma independiente, lo cual provoca grandes errores a menos que se cuente con secuencias suficientemente grandes. El uso del modelo Maximum Composite Likelihood reduce considerablemente estos errores ya que estima un solo conjunto de parámetros para todos los pares de secuencias para calcular las distancias evolutivas [53]. A pesar de que este modelo resultó ser el más apropiado para el análisis filogenético de las secuencias de virus mexicanos,

no se usa frecuentemente en los análisis filogenéticos del VIH. Una de las razones puede ser que es relativamente nuevo ya que fue publicado por primera vez en el año 2004.

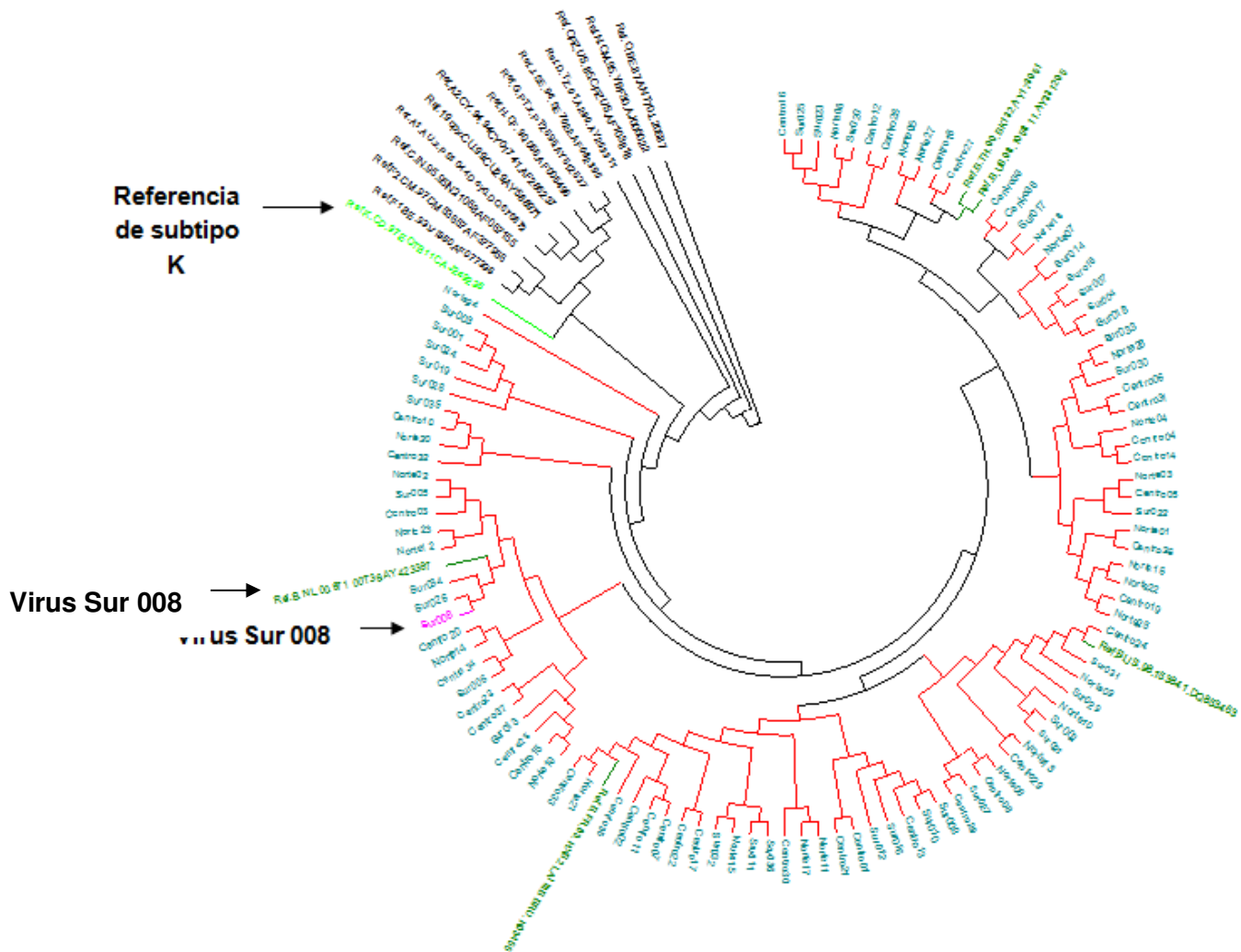


Figura 8.4 Árbol filogenético del gen *env* con el modelo Maximum Composite Likelihood. Este árbol es una representación radial del árbol de la figura 7.11 inciso e. En este árbol se puede observar que la referencia del subtipo K no tiene una relación cercana con ningún virus mexicano. En negro se encuentran las referencias de subtipos no B, en verde hoja se encuentran las referencias de subtipo B, en verde claro se encuentra la referencia K, en gris se encuentran las OTUs de virus mexicanos y en rosa se encuentra el virus Sur008.

Ya determinado el modelo óptimo se construyeron los árboles filogenéticos de los genes *env* y *pol* con el modelo Maximum Composite likelihood. En el árbol de *pol* las OTUs de virus mexicanos se agrupan en una misma rama con las referencias B lo cual nos indica que pertenecen a dicho subtipo (**ver figura 7.14**). Fue posible observar que los virus mexicanos se distribuían de forma homogénea, es decir no existieron agrupamientos por zona, lo cual podría deberse a que los virus circulantes presentaron una alta homogeneidad en el gen *pol*. También se observó que la longitud de las ramas es menor que en el caso del árbol de *env* (**ver figura 7.13**), esto se debe a que las secuencias del gen *pol* guardan menor distancia genética (**ver figura 7.14**).

En el caso del árbol de *env*, fue posible observar que las secuencias mexicanas se agruparon con referencias del subtipo B y recombinantes B (**ver figura 7.13**), lo cual nos sugiere que las secuencias estaban fuertemente relacionadas con virus del subtipo B. Se descartó la existencia de recombinantes B analizando con el programa RIP [45] las secuencias de virus que se encontraban próximos a las referencias de recombinantes (**ver figura 7.15, 7.16 y 7.17**). Esta relación entre virus mexicanos y recombinantes refleja posibles artefactos del análisis. También fue posible observar que las ramas del árbol tenían mayor longitud que en el árbol del *pol* (**ver figura 7.14**), lo cual se atribuye a una mayor distancia genética entre las secuencias. En general las OTUs de los virus mexicanos se distribuyeron de forma homogénea, a pesar de ello se encontró una agrupación entre siete virus de la región Centro. Los pacientes infectados por estos virus agrupados no presentan una diferencia significativa en los valores de células CD4+ (**ver figura 7.20**) o carga viral (**ver figura 7.21**), tampoco presentaron una línea de contagio evidente, es decir no se encontró una relación entre la procedencia geográfica o el grupo de riesgo de los pacientes .

Al revisar los valores de bootstrap encontramos que eran inferiores al 70% lo cual nos indica que no se puede afirmar con certeza que el agrupamiento exista. A

pesar de esto este agrupamiento se repite al utilizar todos los modelos evolutivos (**ver figura 7.12**), por esta razón podemos sugerir que el agrupamiento es real.

En ocasiones los valores de bootstrap obtenidos en el análisis de secuencias poco divergentes resultan bajos al incluir un número elevado de taxás a analizar. Este valor puede mejorar si se disminuye el número de taxás, ya que el número de combinaciones aumenta exponencialmente al aumentar el número de taxás a analizar. A pesar de esto, la razón principal de que se observen valores inferiores al 70% es el hecho de que las secuencias analizadas son altamente homogéneas ya que al disminuir el número de taxás los valores de bootstrap no mostraron una mejoría y continuaron siendo en general menores al 70% (**ver figura 8.5**).

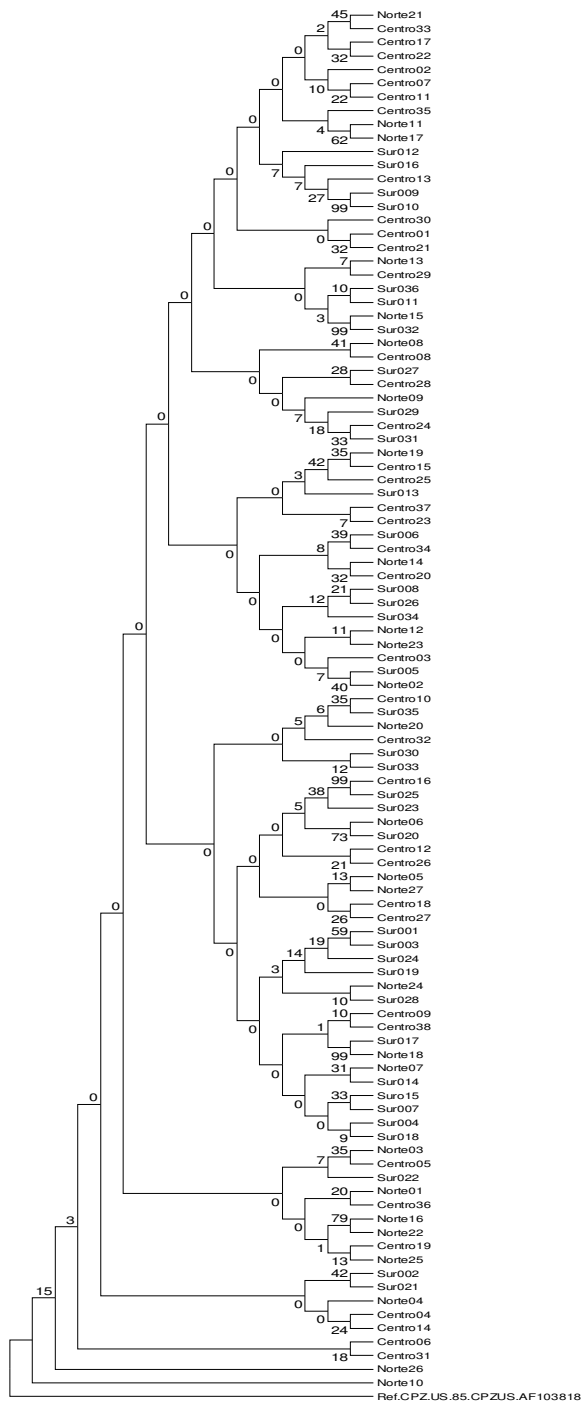


Figura 8.5 Arbol filogenético de secuencias Mexicanas utilizando únicamente un grupo externo. Los valores de bootstrap son muy similares a los obtenidos en el árbol construido con el método Neighbor Joining y el modelo Maximum Composite Likelihood con todas las referencias de los subtipos del grupo M ,referencias de los grupos Oy N,asi como referencias de SIV(figura 7.13).

9.0 Conclusiones

De acuerdo a los resultados obtenidos el subtipo predominante en la República Mexicana fue el subtipo B con un 100% de los virus circulantes en diversas regiones del país. Esta observación está de acuerdo con los resultados del estudio previo realizado en el CIENI. En contraste estos resultados no concuerdan con el creciente aumento en la prevalencia de subtipos no B en los Estados Unidos. A pesar de ello no podríamos descartar esta tendencia en el país ya que en México el diagnóstico de VIH se realiza de forma tardía, por tanto es posible que no se hayan diagnosticado estos casos.

Encontramos que el modelo Kimura 2 parámetros no es el modelo más adecuado para representar filogenias del VIH, por esta razón probamos los diferentes modelos de nucleótidos y encontramos que el modelo que representó con mayor exactitud las relaciones filogenéticas fue el modelo Maximum Composite Likelihood ya que mostró menos artefactos en los árboles.

Al analizar las secuencias de *pol* y *env* con el modelo Maximum Composite Likelihood encontramos que en los dos casos todos los virus mexicanos se agrupan con las referencias B, lo cual corrobora los resultados de la subtipificación. También fue posible observar que existe una mayor distancia genética entre las secuencias del gen *env* en comparación con las secuencias del gen *pol*. Esto es de esperarse ya que *env* es más variable que *pol*.

Al analizar los árboles de *env* y *pol* encontramos que en general los virus se distribuían de forma homogénea en los árboles filogenéticos, es decir en general no se observaron agrupaciones por regiones geográficas. A pesar de ello fue posible localizar una agrupación de siete virus provenientes del Centro en el árbol de *env*, la cual se repite al utilizar todos los modelos evolutivos.

Los pacientes infectados por los virus agrupados no presentaron una diferencia significativa en sus valores de carga viral o células CD4⁺ en comparación con los

pacientes infectados por virus no agrupados. También se encontró que dichos pacientes no presentaron ningún tipo de relación evidente por procedencia geográfica.

Las secuencias de los genes *env* y *pol* son altamente homogéneas ya que los valores de bootstrap fueron menores al 70%, esto se pudo comprobar al disminuir el número de taxonías en los análisis filogenéticos sin observar un aumento en los valores de bootstrap.

A pesar de haber encontrado un agrupamiento de virus del Centro de la República Mexicana no es posible hablar de que existe una diferenciación entre virus del Centro, Norte y Sur del país ya que : existe una alta homogeneidad entre los virus circulantes en las diferentes regiones geográficas y el número de pacientes agrupados es pequeño y no resulta representativo. Por tanto para verificar estos resultados se tendría que ampliar la muestra.

10.0 Referencias:

1. ONUSIDA, *Informe sobre la epidemia mundial del SIDA 2008*. 2008.
2. ONUSIDA, *Informe sobre la epidemia mundial del SIDA 2007*.2007
3. Operativa, C.D.d.I., *El VIH/SIDA en México. Datos al 14 de Noviembre de 2008*, S.d.S. Pública, Editor. 2008, Centro Nacional para la Prevención y el Control del VIH/SIDA: México.2
4. José Ángel Córdova Villalobos, S.P.d.L.R., José Luis Valdespino, ed. *25 años de SIDA en México. Logros, desaciertos y retos*. ed. P. Edición. 2008, Instituto Nacional de Salud Publica: México. 182.
5. ONUSIDA, *El sida y las relaciones sexuales entre varones*. Colección Practicas Optimas del ONUSIDA. 2000.2000
6. Hoffmann Christian, J.K.R.y.B.S.K., *HIV Medicine 2007*. 15 ed. 2007, Paris: Flying Publiser
7. TermeJ.M.; Calvignac S.;Duc Dondon M.;Gazzolo, Jordan A; *E box motifs as mediators of proviral latency of human retroviruses . Retrovirology 2009.6:81*
8. David M. Knipe, P.P.M.H., MD; Diane E. Griffin, MD, PhD; Robert A. Lamb, PhD, ScD; Malcolm A. Martin, MD; Bernard Roizman, ScD; Stephen E. Straus, MD, *Field' Virology*. 5 ed. 2007.
9. <http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2000/partI/Karn.pdf>
10. Castaño M.E.; Urcuqui S.; Proteínas celulares de las protinas regulatoria y accesorias del VIH-1; Colombia Médica 2004. **35**(2):p112:120.
11. John M. Coffin; Stephen H. Hughes; Harold E. Varmus, *Retroviruses*. 1ed. 1997;USA: Cold Spring Harbor Laboratory Press.
12. David M. Knipe; Peter M. Houley, Diane E. Griffin, Stephen E. Straus, Robert A. Lamb; Malcolm A. Martin, Bernard Roizman, Stephen E. Straus, *Fundamental Virology*.4 ed. 2000
13. Kindt J. Thomas, R.A.G., Barbara A. Osborne, *Inmunología de Kuby*. sexta ed. 2007, México: McGraw Hill Interamericana. 574.
14. Shriner D.;Rodriguez A.G.;Nickle D.C.;Mullins J. I.; *Pervasive Genomic Recombination of HIV-1 in Vivo*. Genetics 2004.167: 1573-1583.
15. Requejo, H.I.Z., *Worldwild molecular epidemiology of HIV* Rev Saúde Pública 2006. **40**: p. 331-345.
16. Taylor B.S.;Sobieszczyk M.E.;McCutchan F.E.;Hammer S.M.; *The Challenge of HIV-1 Subtype Diversity* The New England Journal of Medicine 2009. **358**(15): 1590-1602.
17. A.J. Kandathil, S.; Ramalingam, R. Kannagai; Shoba David ;G. Sridharan , *Molecular Epidemiology of HIV* Indian J Med Res 2005. **121**:333-344.
18. Hemelaar, J., et al., *Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004*. AIDS, 2006. **20**(16): p. W13-23.

19. Rivera- Morales Lydia G.; Novitsky Vladimir A.; Trujillo Roberto; La Valle-Montalvo Carlos; Cano Domínguez Carlos; Ramos-Jiménez Javier; Jiménez-Ríos Eusebio; Flores-Flores Leopoldo; Lopez-Guillen Paulo; Gilbert Peter; Vannberg Fredrik; Tamez-Guerra Reyes; Rodríguez-Padilla Cristina; Essex Max, *The Molecular Epidemiology of HIV Type 1 of Men in Mexico*, Aids Resserch and Human Retroviruses 2001.17(1):pp 87-92.
20. Eyzaguirre L. ; Brouwer K.C.; Nadai Y.; Patterson T.L.; Ramos R.; Firestone Cruz M.; Orozovich P.; Strathdee S.A.; Carr J.K., *First Molecular Report of HIV Type 1 in Injection Drug Users and Female Sex Workers along the U.S.- Mexico Border*, Aids Resserch and Human Retroviruses, 2007, **23**(2):pp331-334.
21. Claudia, G.M. *Jornadas del CIENI. Análisis de resistencia primaria en VIH circulante en México*. 2009. México D.F.
22. Pyne M.T., Holzmayer V., Hackett J., Hillyard D ; *16th Conference on Retrovirus and Opporturtunistic Infection .Analysis of HIV Subtype Prevalence and Geographic Distribution in a Large U.S. Population.*; Canada Montreal 2009.
23. Brawn T.A., *Genome.2* ed.2002; USA :John Wiley & Sons Inc.
24. http://www.ccg.unam.mx/~vinuesa/Courses_bioinfo_mol_phylo_ecol_evolution_PV.html
25. <http://bacteria.fciencias.unam.mx/bioinformatica/Filogenias.htm>
26. Saitou N. and Nei M, *The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees*. Mol. Biol.Evol., 1987. **4**(4): p. 406-425.
27. <http://www.tamps.cinvestav.mx/~ertello/bioinfo/sesion15.pdf>
28. <http://www.tamps.cinvestav.mx/~ertello/bioinfo/sesion16.pdf>
29. Madame Curie bioscienceDatabase:
//www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=eurekah&part=A40239#ch246
30. Masatoshi Nei,, *Molecular evolution and phylogenetics*. 1 ed 2000, New York: Oxford University Press.
31. Nei M.; Rzhetsky A., *Tests of Applicability of Several Substitution Models for DNA Sequence Data*. Mol. Biol. Evol., 1995. **12**(1): p. 131-151
32. Tamura K. and Kumar S.; *Evolutionary Distance Estimation Under Heterogeneous Substitution Pattern Among Lineages*. Mol. Biol. Evol., 2002. **19**(10): p. 1727-1736.
33. Tamura K., Nei M, and Kumar S., *Prospects for inferring very large phylogenies by using the neighbor-joining method*. PNAS, 2004. **101**(30): p. 11030-11035.
34. Bosa D., Posada.D.. *Using models of nucleotide evolution to build phylogenetic trees*. Developmental and Comparative Immunology, 2005. **29**: p. 211-227

35. Quiagen QIAamp Viral RNA Mini Kit:

<http://www1.qiagen.com/Products/RnaStabilizationPurification/CellViralRnaPurificationSystems/QiaampViralRnaMiniKit>.

36. Invitrogen SuperScript™ One-Step RT-PCR with Platinum® Taq :

http://www.invitrogen.com/etc/medialib/en/filelibrary/pdf.Par.35445.File.dat/superScript_onestepRTPCR_man.pdf

37. WHO-UNAIDS. Guidelines for Standard HIV Isolation and Characterization

Procedures. 2nd ed., Geneva:WHO/HTTP/IVR/HVI 02.01; 2002

38. Invitrogen Platinum® Taq DNA Polymerase:

http://tools.invitrogen.com/content/sfs/manuals/platinumtaq_pps.pdf

39. Quiagen QIAquick PCR Purification Kit:

<http://www1.qiagen.com/Products/DnaCleanup/GelPcrSiCleanupSystems/QIAquickPCRPurificationKit>.

40. Roche marcadores de peso molecular:

https://www.roche-applied-science.com/PROD_INF/applic/mwm.pdf

41. Applied Biosystems BigDye® Terminator v1.1 Cycle Sequencing:

Kit:http://awcmee.massey.ac.nz/Data/BDTv1.1_Protocol_04337036.pdf

42. Applied Biosystems ABI-Prism 3100 Avant Genetic Analyzer:

http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_039893.pdf

43. Base de datos de los Alamos:

<http://www.hiv.lanl.gov/content/index>

44. Base de datos de la Universidad de Standford (REGA):

<http://dbpartners.stanford.edu/RegaSubtyping/>.

45. Base de datos de Los Álamos (RIP):

<http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html>

46. Ahumada-Ruiz S.; Flores-Figueroa D.; Toala-Gómez I.; Thamson M.; *Analysis of HIV-1 pol sequences from Panama: Identification of phylogenetic clusters within subtype B and detection of antiretroviral drug resistance mutations* Infection, Genetics and Evolution 2009.

47. Ndembi N., Abraha A.; Pilch H., Ichimura H., Mbanya D.; Kaptue L., Salata R., Arts E.J., *Molecular Characterization of Human Immunodeficiency Virus Type 1 (HIV-1) and HIV-2 in Yaoundé, Cameroon: Evidence of Major Drug Resistance Mutations*

in Newly Diagnosed Patients Infected with Subtypes Other than Subtype B Journal of Clinical Microbiology 2008.26(1): p. 177-184.

48. Allen G. Rodrigo, Gerald H. Learn Jr.; *Computational and Evolutionary Analysis of HIV molecular sequence* 1 ed 2001; USA: Kluwer Academic Publishers.

49. http://www.congresogesida.es/diapositivas/posters_NO_orales/P-75.pdf

50. Fu-xiang W., Hiu Z, Hong L.; Hai-zhou Z, Wei-huia L., Yi-ming S., Jin Z.; *Subtype and sequence analysis of HIV-1 strains in Heilongjiang Province*; Chinese Medical Journal 2007.120(22):2006-2010.

51. http://www.censida.salud.gob.mx/descargas/2009/cifras/05_Edad_y_Sexo.pdf

52. <http://www.hiv.lanl.gov/content/immunology/pdf/1999/4/nomenclature.pdf>

53 Kumar V., Dey A.; Singh A.; *MEGA: A Bio Computational Software for Sequence and Phylogenetic Analysis*; World Congress on Engineering 2009. Inglaterra Londres 2009.