ANÁLISIS BIOINFORMÁTICO DE SECUENCIAS L1, E6 Y E7 DE VIRUS PAPILOMA HUMANO DE ALTO Y BAJO RIESGO MÁS FRECUENTES EN LATINOAMÉRICA

ANDRÉS FELIPE ANGULO COLIMBA

UNIVERSIDAD DE NARIÑO
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE BIOLOGÍA
PROGRAMA DE BIOLOGÍA
SAN JUAN DE PASTO
2014

ANÁLISIS BIOINFORMÁTICO DE SECUENCIAS L1, E6 Y E7 DE VIRUS PAPILOMA HUMANO DE ALTO Y BAJO RIESGO MÁS FRECUENTES EN LATINOAMÉRICA

ANDRÉS FELIPE ANGULO COLIMBA

Trabajo de grado presentado como requisito final para optar por el título de Biólogo

Directora
MILENA GUERRERO FLÓREZ
BSc. MSc. Microbiología

Co-directora
CLAUDIA ANDREA SÁNCHEZ
BSc. Bióloga

UNIVERSIDAD DE NARIÑO
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE BIOLOGÍA
PROGRAMA DE BIOLOGÍA
SAN JUAN DE PASTO
2014

Las líneas y conclusiones aportadas son responsabilidad de los autores.

Artículo 1ro Nº 324 de octubre 11 de 1966 emanado por el Honorable Consejo Superior directivo de la Universidad de Nariño.

Nota de aceptación

_____

_____

_____

_____

_____

_____

Milena Guerrero Flórez

Directora

Claudia Andrea Sánchez

Co-directora

Dolly Margot Revelo Romo

Jurado

Edith Mariela Burbano Rosero

Jurado

San Juan de Pasto, Diciembre de 2014.

AGRADECIMIENTOS

A Dios por ser mi apoyo más grande.

A mi Directora de Trabajo de grado Milena Guerrero Flórez, porque sin ella llegar al final de este escalón no hubiera sido posible. Mil y muchas gracias por sus enseñanzas, por sus aportes, por la confianza y sobre todo por la paciencia. Su pasión por la investigación me ayudó a reencontrar el camino. Este trabajo es suyo.

A mi co-directora Claudía Andréa Sánchez, por todos sus aportes, la confianza y todo el apoyo recibido durante la estrucutración del proyecto. Mis mejores deseos para con la nueva etapa que esta desarrollando.

A mis jurados Mariela Burbano y Dolly Revelo por todos sus aportes y ayudarnos a direccionar la estrategia y el desarrollo de esta investigación.

Al profesor Carlos Torres por sus aportes en la elaboración de mapas de distribución.

A Eliana Revelo por el apoyo incondicional antes y durante la ejecución de este trabajo (…), sé que después también lo será.

A mis compañeros por haber compartido de su esencia conmigo, sin ustedes el tiempo en el salón de clases, las prácticas de laboratorio, las salidas al campo o los eternos trasnochos no hubieran sido lo mismo. A mis amigos Ricardo, Raquel, Karen, Camilo, Roger, Eliana, Maira, Lizeth, Miguel, Aura, Ana, Pauline, Carolina y Carlos, gracias.

A todos los docentes del Departamento de Biología por aportar a mi formación profesional.

A la Universidad de Nariño por darme la oportunidad de formarme como biólogo en sus aulas.

… y por último, y no por eso menos importante, a mi familia por todo el amor, el apoyo y la paciencia. Gracias por exigirme a ser mejor persona cada día.

DEDICATORIA

"Como si se tratase de la canción más hermosa, este trabajo lo dedico a mis amados padres y a mi hermano, por hacerme tan feliz. A ellos gracias, no necesito ninguna otra cosa."

(...) Desde más allá de donde siempre mirábamos, te vi.
Te vi más allá del cielo, desde la más profunda oscuridad.
De tan lejos vi tu expresión mojada, deambulante entre las luces y los demonios de la soledad.
También te vi morir, te vi olvidarme y encontrar al errante que ahora mora en tus entrañas.
¿Sientes lo mismo que yo?

Ahora soy solo yo sin ti, de mi capricho, de la gracia del impulso, de la utopía delirante de mi corazón, de mi esencia carnal.
Ahora surjo de más allá de lo profundo, de lo que es y será por siempre más allá de mí.
Eterno como el oscuro infinito del universo y etéreo como el segundo que ya fue.
Ahora soy solo yo sin ti, de mi propia inspiración (…)

B.V.R

RESUMEN

El uso de herramientas de Biología computacional y las secuencias ORF de L1, E6 y E7 reportadas en literatura especializada y en la base de datos de GenBank de los tipos de VPH-AR y VPH-BR más frecuentes en Latinoamérica, permitió analizar las relaciones filogenéticas, variantes de tipos y su distribución geográfica. Se utilizaron rangos de diferente tamaño de fragmento de L1 para construir las filogenias, mientras que el análisis de variabilidad de E6 y E7 permitió identificar mutaciones puntuales para variantes específicas de cada tipo, que posteriormente visualizamos para establecer su distribución en América latina. Se obtuvieron 770 secuencias procedentes de trece países, principalmente Costa Rica, México y Brasil para los tipos de VPH 6, 11, 16, 18, 31, 33, 35, 45, 52, 58 y 97. Se identificaron 24 nuevas mutaciones puntuales para todos los tipos excepto VPH 6, 16, 31 y 35, así como mutaciones que permitieron identificar grupos de variantes de acuerdo a la variabilidad exhibida, los cuales se resumieron en mapas de distribución geográfica según el tipo viral. Esta investigación permitió identificar mutaciones puntuales presentes en los diferentes tipos de VPH, posiblemente relacionadas con la persistencia viral. Los datos sobre patrones de distribución geográfica de tipos de VPH, así como de variabilidad intratipo encontrados en este estudio, son esenciales en estudios epidemiológicos tanto para el desarrollo como para la estimación del impacto de las vacunas contra el VPH en CCU y los programas de cribado del cáncer cervical.

Palabras clave: Bioinformática, Latinoamérica, mutación puntual, ORF, VPH-BR, VPH-AR.

ABSTRACT

The use of computational biology tools and the L1, E6 and E7 ORF sequences reported in specialized literature and Gen Bank data base about the most frequent HR-HPV and LR-VPH types for Latin America, allowed analyze their phylogenetic relationships, variants and geographic distribution. Various L1 lengths were used in order to construct the phylogenies, and the E6 and E7 variability analysis allow the identification of punctual mutations for specific variants of both HR-HPV and LR-VPH types which were used to establish their distribution in Latin America. It was obtained 770 sequences from 13 countries, mainly Costa Rica, Mexico and Brazil for 6, 11, 16, 18, 31, 33, 45, 52, 58 and 97 VPH´s types. It were identified 24 new punctual mutations for all types except 6, 16, 31, and 35 VPH´s types, and it were found some mutations that allow identify variants groups according to the variability exhibited. These variations were summarized in geographical distribution maps according to the viral type. This research was carried out to identify which mutations can define VPH variations and therefore can be a higher risk of cancer development. This investigation identified point mutations present in different types of HPV, possibly related to viral persistence. The distribution patterns data of viral types, and the intratype variability found here, are essentials in epidemiological studies for both development and impact estimation of vaccines against VPH in CCU and screening of cervical cancer.

Key words: Bioinformatics, Latin America, punctual mutation, ORF, HR-HPV, LR-VPH

CONTENIDO

	Pág.
INTRODUCCIÓN	16
1. OBJETIVOS	18
1.1 Objetivo general	18
1.2 Objetivos específicos	18
2. MARCO TEÓRICO	19
3.1 VIRUS DEL PAPILOMA HUMANO (VPH)	19
3.2 Taxonomía y Sistemas de clasificación de VPH	19
3.3 Estructura del genoma viral	21
3.4 Descripción del ciclo infeccioso viral	23
3.4.1 Caracterización de las oncoproteínas E6 y E7 y su relación con el cáncer	24
3.5 Distribución de variantes de VPH	26
3.6 Bioinformática: estructura de datos y algoritmos	26
3.6.3 Alineamiento de secuencias	27
3.6.3.1 Alineamiento de pares de secuencias y algoritmos	27
3.6.3.2 Alineamiento múltiple de secuencias y algoritmos	28
3.6.4 Análisis filogenético y evolución de VPH	29
3.6.4.1 Métodos para la construcción de árboles filogenéticos	30
3. ANTECEDENTES	31
4. METODOLOGÍA	34
4.1 Minería de datos	34
4.2 Análisis de comparación de secuencias mediante alineamientos	36
4.3 Análisis filogenético	37
4.4 Identificación de variantes de tipos virales	37
4.5 Análisis de la distribución geográfica de variantes	37
5. RESULTADOS Y DISCUSIÓN	38
5.1 Revisión de literatura especializada y bases de datos de secuencias	38
5.2 Tipos de VPH seleccionados para ele studio	43
5.3 Relaciones filogenéticas de los VPH-AR y BR más frecuentes en Latinoamérica, mediante análisis de secuencias del gen L1	45
5.4 Análisis de variantes de tipos virales haciendo uso de los oncogenes E6 y E7	53
5.4.1 Variantes moleculares de tipos virales de bajo riesgo	57
5.4.2 Variantes moleculares de tipos virales de alto riesgo	62
CONCLUSIONES	87
RECOMENDACIONES	88
LITERATURA CITADA	89
ANEXOS	108

LISTA DE TABLAS

Pág.

Tabla 1. Funciones de los genes del Papilomavirus Humano 22
Tabla 2. Bases de datos para obtención de secuencias 27
Tabla 3. Genomas y genes de referencia utilizados para el análisis bioinformático 34
Tabla 4. Criterios de inclusión y exclusión bibliográfica 35
Tabla 5. Criterios de inclusión y exclusión de secuencias 36
Tabla 6. Descripción de las características de los estudios incluidos 39
Tabla 7. Número total de secuencias recuperadas según su procedencia 41
Tabla 8. Matriz de distancia pareada entre genotipos de la especie Alfa-9 de secuencias parciales (~1596 - 808 nt) 48
Tabla 9. Matriz de distancia pareada entre genotipos de la especie Alfa-9 de secuencias parciales (667 nt -321 nt.) 50
Tabla 10. Mutaciones más importantes identificadas para las variantes de VPH analizados 54
Tabla 11. Variabilidad genética de los genes E6 y E7 de los VPH-AR y BR analizados 58

LISTA DE FIGURAS

Pág.

Figura 1. Árbol filogenético de VPHs basado en secuencias ORF L1 de 170 tipos 20
Figura 2. Organización circular del genoma de VPH 21
Figura 3. Mecanismo molecular de la infección por VPH 23
Figura 4. Representación esquemática de oncoproteínas E6 y E7 25
Figura 5. Diagrama de flujo para el proceso de selección de artículos 38
Figura 6. Tasas de incidencia de CCU en el mundo estandarizadas por edad 42
Figura 7. Frecuencia (%) de tipos de VPH en Latinoamérica 44
Figura 8. Número de secuencias de genes L1, E6 y E7 por cada genotipo de AR y BR analizados, presentes en la base de datos del GenBank 44
Figura 9. Número total de secuencias para cada gen (L1, E6 y E7) analizado, recuperadas de la base de datos del GenBank. 45
Figura 10. Árbol filogenético construido con el programa MEGA para nueve tipos de VPH usando el rango de secuencias parciales (~1596 - 808 nt) de L1 47
Figura 11. Árbol filogenético construido con el programa MEGA para diez tipos de VPH usando el rango de secuencias parciales (667 - 321 nt) de L1 49
Figura 12. Árbol filogenético construido con el programa MEGA para siete tipos de VPH usando un fragmento de 120 nt de L1 52
Figura 13. Árbol filogenético para variantes de E6 de VPH 6 61
Figura 14. Árbol filogenético para variantes de E6 de VPH 11 62
Figura 15. Árbol filogenético para variantes de E6 de VPH 18 63
Figura 16. Árbol filogenético para variantes de E7 de VPH 18 63
Figura 17. Árbol filogenético para variantes de E7 de VPH 18 64
Figura 18. Árbol filogenético para variantes E6 y E7 VPH 45 65
Figura 19. Distribución geográfica para variantes E6 y E7 de VPH 45 66
Figura 20. Árbol filogenético para variantes de E6 de VPH 16 68
Figura 21. Distribución geográfica para variantes de E6 VPH 16 69
Figura 22. Árbol filogenético para variantes de E7 de VPH 16 70
Figura 23. Distribución geográfica para variantes de E7 VPH 16 71
Figura 24. Árbol filogenético para variantes de E6 de VPH 31 73
Figura 25. Árbol filogenético para variantes de E7 de VPH 31 73
Figura 26. Distribución geográfica para variantes E6 de VPH 31 74
Figura 27. Distribución geográfica para variantes de E7 VPH 31 75
Figura 28. Árbol filogenético de variantes de VPH 33 76
Figura 29. Distribución geográfica y frecuencia (%) de VPH 33 77
Figura 30. Árbol filogenético para variantes de VPH 35 78
Figura 31. Distribución geográfica para variantes de VPH 33 79
Figura 32. Árbol filogenético para variantes de E7 de VPH 52 80
Figura 33. Distribución geográfica de variantes de E7 VPH 52 81
Figura 34. Árbol filogenético para variantes de E6 de VPH 58 82
Figura 35. Distribución geográfica y frecuencia (%) de variantes de E6 VPH 58 83
Figura 36. Árbol filogenético para variantes de E7 de VPH 58 84
Figura 37. Distribución geográfica y frecuencia (%) de variantes de E7 VPH 58 85

LISTA DE ANEXOS

Pág.

Anexo A. Tabla de aminoácidos con abreviaturas empleadas para su identificación y su clasificación según el grupo funcional 108
Anexo B. Listado de artículos seleccionados en la revisión 109
Anexo C. Variabilidad genética de secuencias E6 de VPH 6 111
Anexo D. Variabilidad genética de secuencias E6 de VPH 11 112
Anexo E.Variabilidad genética de secuencias E6 de VPH 18 113
Anexo F. Variabilidad genética de secuencias E7 de VPH 18 114
Anexo G. Variabilidad genética de secuencias E6 de VPH 45 115
Anexo H. Variabilidad genética de secuencias E7 de VPH 45 116
Anexo I. Variabilidad genética de secuencias E6 de VPH 16 117
Anexo J. Variabilidad genética de secuencias E7 de VPH 16 121
Anexo K. Variabilidad genética de secuencias E6 de VPH 31 123
Anexo L. Variabilidad genética de secuencias E7 de VPH 31 124
Anexo M. Variabilidad genética de secuencias E6 y E7 de VPH 33 125
Anexo N. Variabilidad genética de secuencias E6 y E7 de VPH 35 126
Anexo Ñ. Variabilidad genética de secuencias E6 de VPH 52 127
Anexo O. Variabilidad genética de secuencias E7 de VPH 52 128
Anexo P. Variabilidad genética de secuencias E6 de VPH 58 129
Anexo Q. Variabilidad genética de secuencias E7 de VPH 58 130

GLOSARIO

ALGORITMO: secuencia de pasos que conforman una tarea específica, usualmente en el ámbito computacional.

ALINEAMIENTO DE SECUENCIAS: Arreglo mutuo de dos o más secuencias, que muestra donde estas son similares y donde difieren. Un alineamiento óptimo es aquel que muestra la mayor cantidad de correspondencias y la menor cantidad de diferencias.

AMINOÁCIDO: Compuestos cuya estructura incluye un grupo amino, un grupo carboxilo y una cadena lateral de composición variable. Veinte aminoácidos conocidos como "estándar" son los principales monómeros de las cadenas polipeptídicas que forman las proteínas.

ASCUS: Células escamosas atípicas de significancia no determinada, se utiliza para designar aquellas alteraciones citológicas y cambios atípicos que no pueden ser específicamente clasificadas en la citología como lesiones de alto o bajo riesgo.

BASE DE DATOS: Conjunto de datos consistente y usualmente persistente, organizado en un modo específico que permita acceder a la información de forma fácil y rápida.

BIOINFORMÁTICA: De manera simplificada, la aplicación de las tecnologías de la computación y la información al manejo y análisis de información de origen biológico.

CÁNCER DE CUELLO UTERINO (CCU): Es una enfermedad por la que se forman células malignas (cancerosas) en los tejidos del cuello uterino.

DOMINIO: Segmento de cadena polipeptídica que se pliega de forma independiente del resto de la cadena y que usualmente tiene una función específica. También usado para designar al segmento de gen que codifica para éste.

FASTA: (i) Una herramienta desarrollada por Pearson y Lipman para el alineamiento de secuencias de ácidos nucleicos y proteínas. (ii) Uno de los formatos más simples utilizados para almacenar secuencias nucleotídicas o aminoacídicas.

FILOGENÉTICA: La rama de la biología que se ocupa de descubrir las líneas de origen o filogenias de los organismos, a fin de construir las relaciones antepasado-descendientes (relaciones evolutivas) entre los grupos de organismos vivos y extinguidos.

GEN: Es una secuencia de DNA genómico o de RNA que es esencial para especificar uan dterminada función, que no necesita ser traducido a proteína.

GENBANK: Es una base de datos de libre acceso, que contiene secuencias genéticas.

GRUPO EXTERNO (*outgroup*): Es cualquier grupo usado en el análisis que no es incluido en el taxón bajo estudio. Se utiliza para fines comparativos y debe ser lo más cercano posible al grupo interno, preferentemente su grupo hermano.

INCIDENCIA: Número de casos nuevos de una enfermedad que se presentan durante un período determinado, en relación con la población donde ocurren. Generalmente se expresa en forma de tasa.

MOTIVO: Conocidos como bloques, representan usualmente elementos importantes desde el punto de vista estructural o funcional y pueden utilizarse como descriptores de la familia en cuestión.

MUTACIÓN: Proceso mediante el cual el material genético sufre un cambio detectable y heredable, generalmente en la forma de una variación en la secuencia del ADN. Estas variaciones pueden ser puntuales (cambio de un nucleótido por otro) o deberse a la inserción (ganancia) o deleción (pérdida) de un segmento de nucleótidos.

NEOPLASIA: Masa anormal de tejido provocada porque las células que lo constituyen se multiplican a un ritmo superior al normal, estas pueden ser benignas o malignas.

NIC I: Es considerada una lesión escamosa intraepitelial de bajo grado. Se caracteriza por estar confinada a un tercio basal del epitelio cervical.

NIC II: Es considerada una lesión escamosa intraepitelial de alto grado. Confinada a los dos tercios basales del epitelio cervical.

NIC III: En este tipo de lesión, considerada también de alto grado, la displasia es severa y cubre más de los dos tercios del epitelio cervical. En algunos casos incluyendo todo el grosor del revestimiento cervical.

PREVALENCIA: Es la medida del número de casos existentes, llamados casos prevalentes, de una enferemdad en un punto o periodo de tiempo y en una población determinados, sin distinguir si son o no casos nuevos. La prevalencia es un indicador de la magnitud de la presenica de una enfermedad u otro evento de salud en la población.

SECUENCIA: La disposición ordinal de los monómeros que forman parte de las biomoléculas de tipo “polimérico” como los ácidos nucleicos y las proteínas.

VIRUS: Organismos dotados de extraordinaria simplicidad, pertenecen a un nivel de organización subcelular, y marcan la barrera entre lo vivo y lo inerte. No se nutren, no se relacionan, carecen de metabolismo propio y para reproducirse utilizan la maquinaria metabólica de la célula a la que parasitan; su simplicidad estructural y funcional los convierte en parásitos intracelulares obligados, tanto de bacterias (bacteriófagos o fagos), como de las células animales y vegetales.

INTRODUCCIÓN

El Virus del Papiloma Humano (VPH) es uno de los patógenos de transmisión sexual más comunes en todo el mundo, se presume que el 80 % de la población general se infectará en algún momento de su vida con algún tipo de VPH (Muñoz *et al.,* 2003; Cai *et al.,* 2013). El VPH es el principal precursor biológico de los cambios celulares asociados al cáncer cervico-uterino (CCU), pues ha sido identificado en el 99,7 % de las pacientes con cáncer (Kurvinen *et al.,* 2000; Cai *et al.,* 2013). De acuerdo con la Organización mundial de la salud (sigla WHO del inglés *World Health Organization*), con aproximadamente 530.000 nuevos casos por año y 300.000 muertes por año (WHO, 2014), este cáncer constituye la segunda causa de muerte más frecuente en mujeres en el mundo (Burd 2003; WHO, 2014).

La Organización panamericana de la salud (sigla PAHO del inglés *Panamerican Health Organization*, 2013) y la WHO, (2014), estiman que más del 85 % de los casos de muerte por CCU ocurren en países en vía de desarrollo, donde los programas de prevención de cáncer a menudo no cuentan con adecuados programas de cribado (Cogliano *et al*., 2005). Según la PAHO, (2013), en Latinoamérica la tasa de incidencia promedio para esta enfermedad (29,2 / 100.000 mujeres) es casi el doble del promedio mundial (16,2 / 100.000 mujeres). En Colombia este cáncer tiene una alta incidencia, con una tasa de mortalidad de 36,4 / 100.000 mujeres, lo cual lo ubica dentro de los países con el mayor número de muertes por esta causa (Murillo, 2006; Muñoz *et al.,* 2013). Las estadísticas muestran a departamentos como Nariño (26,4 / 100.000) y Amazonas (49,1 / 100.000) con tasas de mortalidad que igualan e incluso superan el promedio nacional y mundial (Molano *et al.,* 2002; Murillo *et al.,* 2008; Cerón, 2013), un patrón que ha sido observado en la mayoría de países latinoamericanos.

Actualmente, se conocen otros cánceres asociados al VPH como el carcinoma de células escamosas de ano, del carcinoma de células escamosas de vulva, vagina, pene, boca y oro-faringe (Squiquera, 2006; Mijares *et al.,* 2007). Por esta razón, el estudio de la infección por este virus tiene una gran importancia y preocupa a la comunidad científica desde el punto de vista sanitario, por sus aspectos epidemiológicos todavía no enteramente dilucidados así como por el amplio espectro clínico, el potencial oncogénico de algunos genotipos, la complejidad de su terapéutica y las frecuentes recidivas que se producen en los pacientes afectados (Castellsagué, 2008; Venegas *et al*., 2011).

Más de 170 tipos de VPH han sido identificados y aproximadamente 40 de estos pueden infectar el tracto anogenital humano (Brebi *et al.,* 2009; Sabol *et al.,* 2012; Cai *et al.,* 2013). Al menos 15 de estos tipos (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 y 82) son considerados de alto riesgo oncogénico (VPH-AR) por su rol en el desarrollo de neoplasia y cáncer (Muñoz *et al.,* 2003; Sabol *et al.,* 2012). Los VPH 6 y 11 son considerados de bajo riesgo (VPH-BR) y causantes de verrugas genitales (*condyloma acuminata*) y una rara papilomatosis respiratoria (PR) (Lacey *et al.,* 2006). Desde el punto de vista evolutivo, los tipos cancerígenos están relacionados entre sí, pero lejanamente relacionados con los virus causantes de verrugas comunes de la piel.

Los VPH-AR más frecuentes y mayormente distribuidos en Latinoamérica y el mundo son los VPH 16 y VPH 18 (Kurvinen *et al.,* 2000; Sabol *et al.,* 2012), los cuales se subdividen en variantes virales (Sabol *et al.,* 2012), producto de la selección y/o fijación de mutaciones puntuales,

inserciones, supresiones y reordenamientos genéticos en su genoma (Bernard *et al.,* 2006; Burk *et al.,* 2013). Tales eventos evolutivos son los responsables (en gran parte) de la gran proporción de las variantes virales, caracterizadas por diferenciarse biológica y etiológicamente en el potencial oncogénico (Sabol *et al.,* 2012; Burk *et al.,* 2013), la persistencia viral y el desarrollo de CCU (Xi *et al.,* 2007; Cornet *et al.,* 2012).

La genotipificación de VPHs se basa en el análisis de secuencias del DNA total o parcial del marco abierto de lectura (sigla ORF del inglés *Open Reading Frame*) del gen L1, su secuencia altamente conservada ha permitido evidenciar relaciones filogenéticas de importancia a nivel evolutivo (de Villers *et al*., 2004; de Villers, 2013). La detección de variantes de tipos se ha fortalecido en los últimos años analizando las secuencias genómicas de las oncoproteínas E6 y E7, por la capacidad de integración al genoma hospedero y en consecuencia su alta variabilidad (Burk *et al*., 2013).

Los estudios sobre la diversidad intratipo de VPH han revelado que las variantes moleculares del virus pueden producir alteraciones que conducen a cambios en los residuos de los dominios funcionales y/o antigénicos de aminoácidos (aa), algunos de estos cambios han indicado que pueden influir en la persistencia de la infección, la morbilidad de la carcinogénesis y la progresión de lesiones precursoras del cáncer; por lo tanto, las variaciones en las secuencias de nucleótidos del virus pueden ser utilizadas como una herramienta en los estudios epidemiológicos, constituyendo biomarcadores para el seguimiento de la infección con relación entre el genotipo y el fenotipo, es decir, pueden ayudar a clarificar la virulencia, la expresión y transformación de células las hospedadoras, así como también pueden ser utilizadas en estudios evolutivos y taxonómicos.

Existen varias técnicas moleculares para la detección de VPH, sin embargo, los eventos evolutivos propios de su genoma suponen un emergente problema para realizar una adecuada y temprana detección; sin embargo, la aparición de nuevos genotipos a nivel mundial, motiva un nuevo enfoque de las investigaciones que permitan detectar, recopilar y detallar el origen y la distribución de dichos genotipos como modelos para entender los mecanismos evolutivos de VPH y la posible aparición de nuevas infecciones asociadas.

La biología computacional ha concedido alcanzar este objetivo gracias al uso de algoritmos y modelos estadísticos que permiten detectar, inferir y predecir procesos biológicos con gran precisión, hasta ahora, la detección de variantes moleculares de VPH se ha limitado a los genotipos 16 y 18 (Yamada *et al.,* 1995; Yamada *et al.,* 1997; Huertas-Salgado *et al*., 2013), dejando de lado centenares de tipos con y sin potencial oncogénico de gran importancia por la frecuencia con que se detectan y la emergente aparición en determinadas poblaciones.

La presente investigación fue realizada utilizando herramientas de bioinformática tales como bases de datos de secuencias, programas para análisis y algoritmos *online de libre acceso* con el objetivo de identificar relaciones filogenéticas, variantes de tipos y distribución geográfica de los VPH-AR y VPH-BR más frecuentes en Latinoamérica. Se utilizaron como blancos de análisis las secuencias de DNA de los genes L1, E6 y E7 del virus que permitieron analizar variaciones en el genotipado mediante la detección de mutaciones puntuales y su efecto sobre la epidemiología y la distribución geográfica de los VPH en las lesiones en CCU.

1 OBJETIVOS

1.1 Objetivo general

Realizar un análisis bioinformático de secuencias L1, E6 y E7 disponibles en bases de datos e identificar relaciones filogenéticas, variantes de tipos y distribución geográfica de los VPH-AR y VPH-BR más frecuentes en Latinoamérica.

1.2 Objetivos específicos

1.2.1 Establecer las posibles relaciones filogenéticas de los tipos de VPH de AR y BR más frecuentes en Latinoamérica.

1.2.2 Identificar las variantes de tipos de los VPH de AR (16, 18, 31, 33, 35, 45, 52 y 58) y BR (6 y 11) analizados.

1.2.3 Determinar la distribución geográfica de las variantes de tipos de VPH analizados.

2 MARCO TEÓRICO

2.1 Virus del Papiloma Humano (VPH)

Los Virus del Papiloma Humano (VPH) son miembros de la familia Papillomaviridae, asociados a la infección de transmisión sexual (ITS) más común, con una prevalencia de alrededor de 500 millones de infecciones en todo el mundo (Villa, 2012). El VPH es considerado como el principal agente causal de las lesiones intraepiteliales que conllevan al desarrollo de CCU, identificado en el 99,7 % de las pacientes con cáncer (WHO, 2014).

Morfológicamente son virus sin envoltura, con DNA de doble cadena y, aunque el genoma viral puede variar ligeramente entre los diferentes genotipos, en general poseen un genoma circular de ~8.000 pares de bases (pb) con un contenido G-C del 41 %, unido a histonas celulares, que codifica entre siete y nueve ORFs, los cuales son transcritos como RNAm policistrónicos (Burd, 2003; Boulet *et al.*, 2007). El VPH presenta una cápside icosaédrica formada por dos proteínas L (*late*) de las cuales la proteína L1 representa el 80 % de las proteínas del virus (proteína mayor de la cápside) y L2 (proteína menor de la cápside) (Villa, 2012). Tiene una cápside de ~55 nm de diámetro compuesta de 72 capsómeros pentaméricos de la proteína L1 que están arreglados en la superficie en forma de un icosaedro (Burd, 2003).

2.2 Taxonomía y Sistemas de clasificación de VPH

La taxonomía de papilomavirus (PV) está basada en filogenias. Los VPH se clasifican en géneros y especies de acuerdo con el Comité Internacional de Taxonomía de Virus (sigla ICTV del inglés International Committee on Taxonomy of Viruses) (de Villiers *et al.*, 2004; de Villiers, 2013). La designación de tipos, subtipos y variantes, comprenden los niveles inferiores en la clasificación (Prado *et al.*, 2005; de Villiers, 2013).

Un “tipo” de VPH se establece cuando la secuencia de DNA del ORF de L1 difiere de cualquier otro tipo que se caracteriza por al menos 10 %, los “subtipos” se caracterizan si las diferencias en homología varían entre un 2 % y 10 % y, si es menos del 2 % constituyen una “variante” (de Villiers *et al.*, 2004; Bernard *et al.*, 2010). Hasta la fecha más de 170 tipos de VPH que infectan el epitelio cutáneo y las mucosas anogenitales han sido completamente descritos (http://www.hpvcenter.se/html/refclones.html) (de Villers, 2013).

De Villiers *et al.* (2004) describen la topología de los árboles filogenéticos, a través de umbrales cuantitativos bajo comparaciones de secuencias de nucleótidos del ORF L1 y características biológicamente distintivas como especie hospedera, tejidos diana, patogenicidad y la organización del genoma, que determinan la clasificación de PV a nivel de género. La nomenclatura de género se basada en el alfabeto griego; donde, cinco géneros abarcan a los PV humanos (Alpha, Beta, Gamma, My y Ny), mientras que los once géneros restantes agrupan uno o varios PV aislados de diferentes mamíferos y aves (Bravo *et al.*, 2010).

De Villiers *et al.* (2004) establecieron los siguientes criterios de clasificación:

(i) Los géneros deben compartir menos del 60% de identidad del gen L1, mientras que deben tener más del 23 % pero menos del 43 % de identidad en la secuencia nucleotídica del genoma.
(ii) Las especies deben compartir entre el 60 % - 70 % de identidad nucleotídica.
(iii) Dentro de las especies los distintos tipos de papilomavirus deben compartir del 71-89 % de identidad nucleotídica en el gen L1.

Los criterios anteriores permitieron la construcción del árbol filogenético de la Figura 1 donde se resume la actual clasificación de los VPHs.

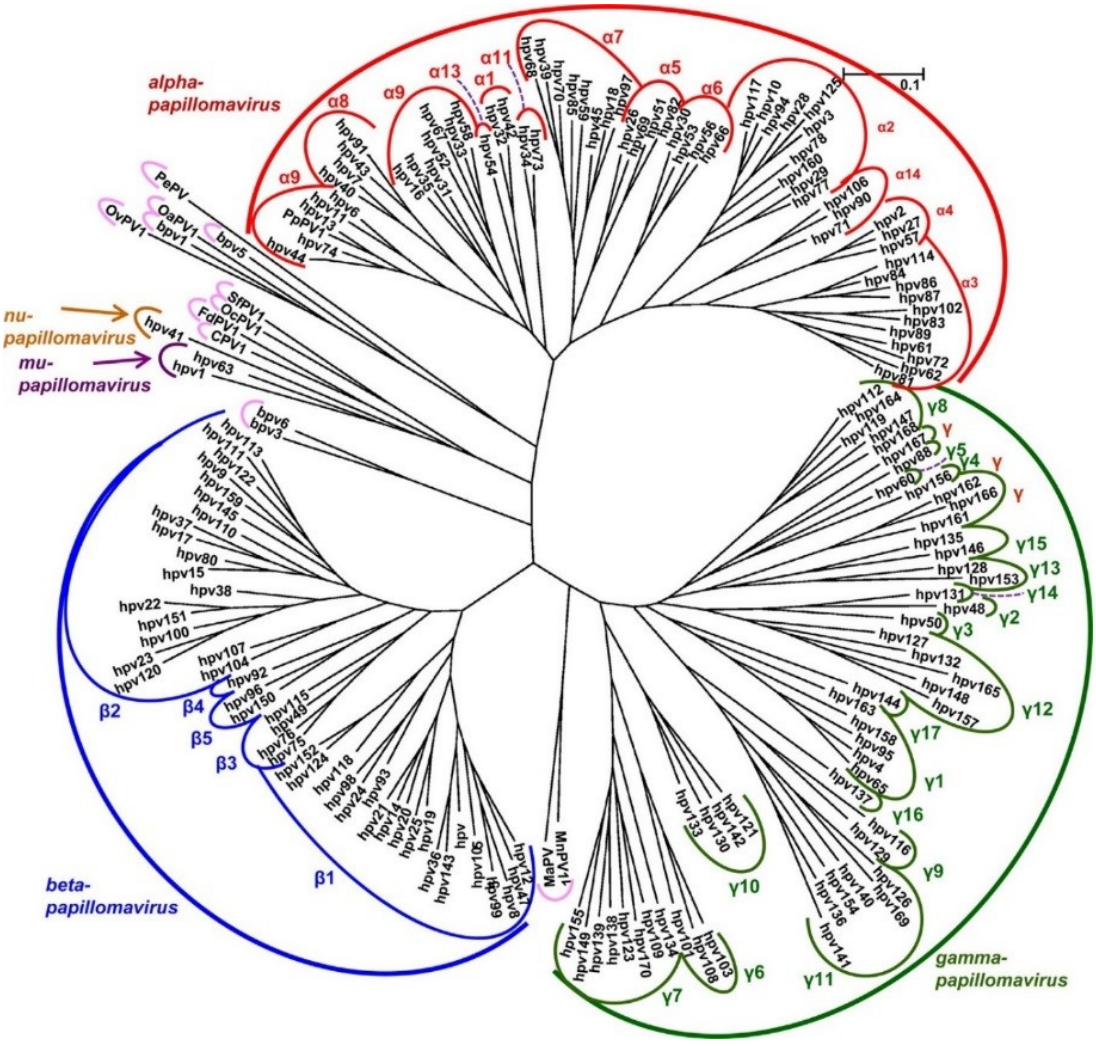


Figura 1. Árbol filogenético de VPHs basado en secuencias ORF L1 de 170 tipos, así como los virus del papiloma animales individuales, clasificados en géneros (alpha-, beta-, gamma-, nu- y mu- papillomavirus), especies (α, β, γ seguido de número) y tipos (ramas internas de cada especie). *Fuente*: De Villers, 2013.

De acuerdo al tipo de tejido que infectan, los PV se han clasificado en cutáneos y mucosos. Los VPH mucosos infectan el tracto anogenital y se clasifican en tipos de VPH-BR y AR, de acuerdo a su potencial para inducir la transformación maligna. Los VPH-BR (tipos 6, 11, 40, 42, 53, 54 y 57) pueden causar cambios leves en el cuello del útero y provocar verrugas genitales, pero no cáncer. Los VPH-AR incluyen los tipos 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 y 82, y se relacionan mayormente con los casos de CCU, de los cuales VPH 16 y 18 son los más peligrosos y frecuentes, detectados en ~50 % y ~15 % del CCU en todo el mundo respectivamente (Muñoz *et al.*, 2003; Castellsagué, 2008; Li *et al.*, 2010).

2.3 Estructura del genoma viral

El genoma del VPH, lo conforman los genes de etapas tempranas de la infección conocidos como genes E (del inglés Early = temprano) y aquellos que se expresan durante las etapas tardías del ciclo replicativo del mismo conocidos como genes L (del inglés Late = tardío) (Rivera *et al.*, 2006). Se conocen seis genes tempranos: E1, E2, E4, E5, E6 y E7 (aunque E4 en realidad es tardío) y dos tardíos: L1 y L2 (Paavonen, 2007; Tena & López, 2007). Los genes E codifican proteínas de la replicación, regulación e integración viral (ésta última define su capacidad carcinogénica) y los genes L codifican proteínas estructurales de la cápside (Burd, 2003) (Figura 2).

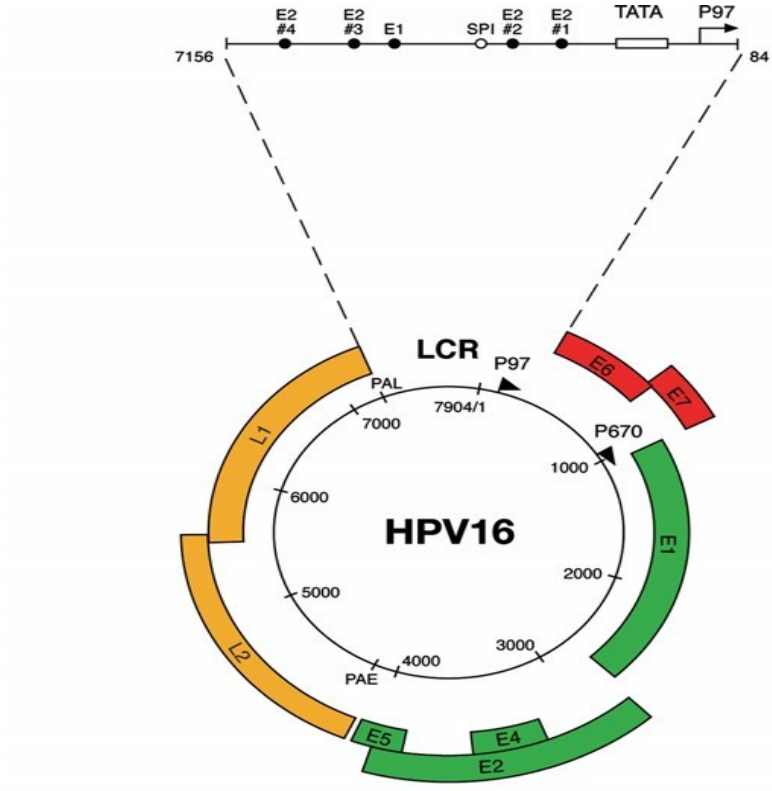


Figura 2. Organización circular del genoma de VPH. Se indican los ORF de genes E (barras rojo y verde) y los genes L (barras amarillo), corriente arriba en la región LCR se encuentra el sitio de unión de E1, los cuatro sitios de unión de E2 (#1-4), el sitio de unión SP1, la caja TATA y P97. P97

y P670 constituyen los promotores del virus y PAE es el sitio de poliadenilación temprana. *Fuente*: Doorbar, 2006.

El pequeño genoma del virus también presenta una región regulatoria no codificante o de control (LCR), la cual contiene sitios de unión a factores proteicos y hormonales del hospedador, necesarios para completar el ciclo de replicación (Burd, 2003). La región LCR consta de tres elementos: una secuencia rica en A y T que sirve como sitio inicial de separación de la doble hélice, una secuencia palindrómica de 18 pb que constituye el sitio de unión de E1 y una secuencia de 12 pb que corresponde al sitio de enlazamiento de E2 (Doorbar, 2006) (Figura 2). En la Tabla 1 se resume la relación de cada una de las regiones codificantes en el proceso de infección viral.

Tabla 1. Funciones de los genes del Papilomavirus Humano*

Genes	Funciones
E1	Reconoce y une al Ori como complejo hexamérico necesario para la replicación del DNA viral. Tiene actividad de ATPasa (Trifosfatasa de adenosina) y DNA helicasa.
E2	Principal regulador de la transcripción de genes virales; se une al promotor de la transcripción viral como un dímero; implicado en la replicación del DNA viral.
E4	Interacciona con la queratina del citoesqueleto y los filamentos intermedios; induce la detención en G2; se cree que facilita el ensamblaje del virus y su liberación.
E5	Induce la proliferación celular no programada; puede activar los receptores del factor de crecimiento y otras proteínas quinasas; inhibe la apoptosis; inhibe el tráfico de los principales complejos de histocompatibilidad en la superficie celular.
E6	Induce la síntesis de DNA, induce la telomerasa; impide la diferenciación celular; interactúa con cuatro clases de proteínas celulares: co-activadores transcripcionales, proteínas implicadas en la polaridad celular y la motilidad, supresores tumorales e inductores de apoptosis, principalmente p53, y la replicación del DNA y factores de reparación.
E7	Induce la proliferación celular no programada; interactúa con la histona acetil transferasa; interactúa con los reguladores negativos del ciclo celular y supresores tumorales, principalmente p105Rb.
L1	Mayor proteína estructural viral; regula el ensamble de capsómeros y la cápside; interactúa con L2, interactúa con los receptores de las células; codifica epítopos neutralizantes.
L2	Menor proteína estructural viral; interactúa con el DNA; interactúa con algunos dominios nucleares; se cree que facilita el ensamblaje del virión, puede interactuar con receptores de células; codifica epítopos neutralizantes del virus lineal.

*Algunas de estas funciones se han descrito en células cultivadas u otros sistemas experimentales. *Fuente*: Fernandes *et al.,* 2013.

2.4 ***Descripción del ciclo infeccioso viral***

El ciclo de replicación del VPH se inicia con la entrada del virus al tejido de revestimiento, facilitado por abrasión o microtrauma en el introito vaginal, donde la partícula viral se une a receptores de superficie específicos como integrinas; los diferentes tipos de VPH entran a la célula por distintas vías, incluyendo la endocitosis y el caveolae (Squiquera, 2006; Letian & Tianyu, 2010). El virus se aloja en el núcleo de células del estrato germinativo, en el epitelio cervical escamoso y/o metaplásico de la zona de transformación, para aprovechar la maduración y diferenciación epitelial en la producción de proteínas para el ensamblaje de nuevos virus (Frattini *et al.,* 1996). En el estrato basal, el virus mantiene un bajo número de copias no integradas (Burd, 2003).

En una primera etapa del ciclo se expresan mayoritariamente las proteínas E6 y E7, implicadas en la transformación celular al interactuar con reguladores celulares antigénicos p53 y pRB (Stubenrauch & Laimins, 1999) (Figura 3). Las células infectadas mantienen niveles bajos de p53 debido a la interacción con E6. E7 interactúa con varias proteínas reguladoras del ciclo celular, como pRb, p107 y p130 (Howie *et al.,* 2009) y facilita la liberación del factor de transcripción E2F, que activa constitutivamente la síntesis del DNA y la proliferación celular (Doorbar, 2005). También se conoce que las proteínas tempranas E1, E2, E4 y E5 se expresan antes de la amplificación del genoma con el fin de mantener el episoma viral (Thomison III *et al.,* 2008).

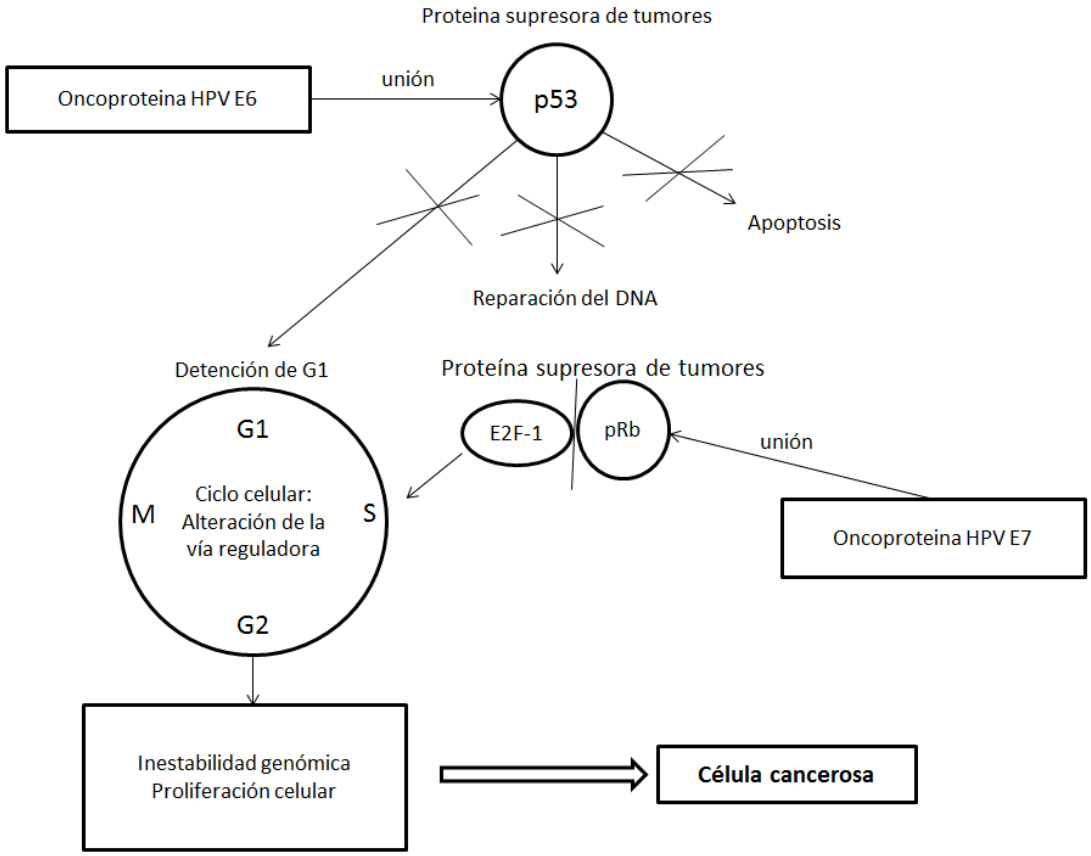


Figura 3. Mecanismo molecular de la infección por VPH, papel de las oncoproteínas E6/E7. *Fuente*: Tena & López, 2007.

En una segunda etapa se sintetizan las proteínas E1 y E2, se reprime la expresión de E6 y E7 y se inicia la replicación del PV; la proteína E1 es la única proteína con actividad ATPasa y helicasa ATP dependiente, siendo un factor esencial para la replicación del genoma viral, la proteína E2 además de facilitar la síntesis del DNA viral, es la única proteína responsable de regular la expresión de genes virales (particularmente la represión de E6 y E7), por lo que la disrupción de E2 de VPH-AR o la pérdida de su expresión, deviene en células que son más propensas a proliferar y, como consecuencia, más propensas a la generación de lesiones tumorales (Doorbar, 2005; Parish *et al.,* 2006; Thomison III *et al.,* 2008). El ciclo infeccioso finaliza con el ensamblado del virión y la síntesis de las proteínas que lo componen, L1 y L2 encierran al DNA nuevo, conformando una partícula de PV capaz de infectar nuevos hospederos (Burd, 2003; Doorbar, 2006).

2.4.1 Oncoproteínas E6 y E7 y su relación con el cáncer

El ORF E6 codifica una proteína pequeña de ~150 aminoácidos con un peso molecular de 18 kDa, presenta cuatro motivos conservados Cys-XX-Cys que se unen a zinc a través de la

coordinación con los residuos de cisteína (Figura 4A), importante para la activación transcripcional, transformación, inmortalización y asociación con proteínas celulares como p53 (Cole & Danos, 1987; Thomas *et al*., 1999; Howie *et al*., 2009). Estos motivos se conservan en toda la familia del VPH (excepto BPV tipo 3, 4, 6 y VPH tipo 101 y 103 donde no hay gen E6) y la ablación de cualquiera de ellos afecta su actividad (Boulet *et al*., 2007). En los VPH-AR la región COOH-terminal posee un domino PDZ involucrado en la interacción con varias proteínas responsables de la regulación de la proliferación celular, las cuales también presentan dominios PDZ (Boulet *et al*., 2007; Vande Pol & Klingelhutz, 2013).

La proteína E6 tiene actividades enzimáticas y pleiotropismo notable por proteínas de la célula hospedadora, una vez el agente infeccioso entra a la célula hospedera, E6 de VPH-AR se distribuye en el núcleo y citoplasma, donde interactua con hasta doce proteínas diferentes entre las que se destacan p53, E6AP, p300/CBP, p73 y el factor de regulación IRF-3; sin embargo, la relevancia fisiológica de la interacción con estas proteínas no está claramente identificada (Cai *et al*., 2013). La función más importante de E6 es vincularse al supresor tumoral p53, un factor de transcripción sitio específico del DNA hospedador y principal coordinador de la señalización celular expresado en respuesta a daño del DNA, dando como resultado la detención del ciclo celular o la apoptosis; el VPH depende de la maquinaria de síntesis del DNA celular y debe estimular la progresión de la fase S para la replicación de su genoma, por lo que la sobreexpresión de p53 inhibe la replicación viral (Boulet *et al*., 2007; Vande Pol & Klingelhutz, 2013).

E6 se une a p53 a través de una ligasa de ubiquitina celular (E6AP), ubiquitinando lisinas en p53 e iniciando su proteólisis (Van, 2014). La degradación de p53 no pasa por las señales normales de control de la detención del crecimiento en las fases G1/S y G2/M, causando inestabilidad cromosómica y una mejor integración del DNA exógeno (Cai *et al*., 2013). Diferentes ensayos *in vitro* señalan que E6 también se estabiliza por la asociación con proteínas de dominio PDZ que no están implicados en la unión y degradación de p53; sin embargo, contribuyen a la actividad transformante de E6, (Van, 2014). Este dominio está representado por la secuencia de X(S/T/Y)X(V/ L/M) en los VPH-AR (Vande Pol & Klingelhutz, 2013).

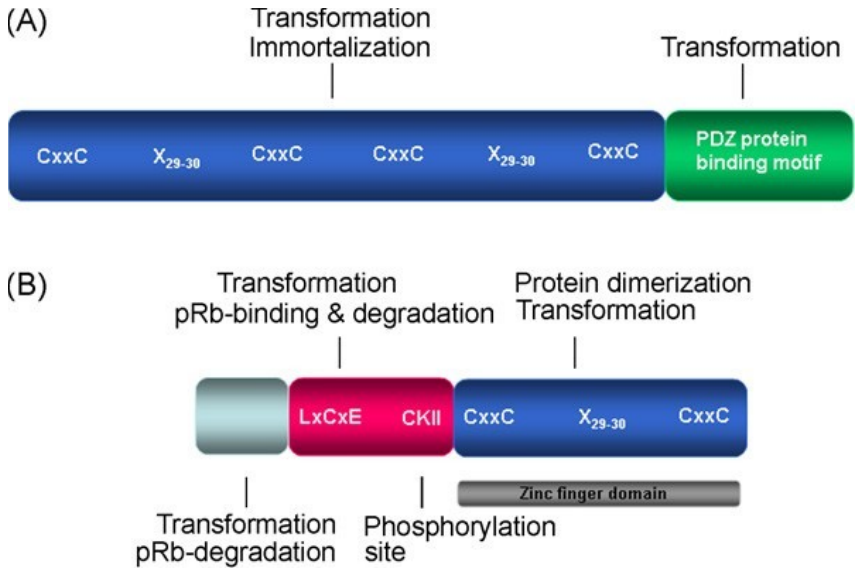


Figura 4. Representación esquemática de oncoproteínas E6 y E7. (A) E6 de ~150 aa, (azul) cuatro motivos CxxC para activación transcripcional, transformación, inmortalización y asociación con proteínas celulares, extremo COOH (verde) dominio PDZ de interacción. (B) E7 con tres regiones conservadas: (i) dominio N-terminal (gris) para transformación celular y degradación de pRb, (ii) dominio de enlazamiento y degradación de pRb (LxCxE - sitio de fosforilación de Kinasa II) (fucsia) y (iii) dominio C-terminal (con dos copias CxxC) de asociación con pRb y unión a zinc (azul). *Fuente*: Boulet *et al*., 2007.

El gen E7 codifica una proteína de ~100 aminoácidos, con tres regiones conservadas (CR) (Figura 4B), el dominio CR1 necesario para la transformación y degradación de pRb, CR2 contiene una secuencia conservada LxCxE de unión a pRb y un sitio de fosforilación de caseína quinasa II (CKII) y COOH-terminal de CR3 posee dos copias del motivo Cys-XX-Cys relacionados con la unión a zinc (Boulet *et al*., 2007).

La unión de E7 a pRb causa la liberación del factor de transcripción E2F, en las células normales pRb es hipofosforilizada a principios de G1, obligando a E2F a formar complejos que funcionan como represores transcripcionales, tras la fosforilación, los complejos se disocian permitiendo a E2F actuar como activador transcripcional; al asociarse la pRb hipofosforilizada, E7 previene su unión con E2F promoviendo la progresión del ciclo celular (Boulet *et al*., 2007; Jiang & Yue, 2014). Por otra parte, pRb es ubiquitinada por E7, esencial para superar eficazmente la detención del ciclo celular; además, E7 interactúa con otros dos miembros de la familia pRb- p107 y p130, que también regulan negativamente la transcripción E2F; sin embargo, se ha determinado que de la estabilidad del motivo LxCxE en el dominio CR2 depende la asociación E7-pRb (McLaughlin-Drubin & Munger, 2009; Jiang & Yue, 2014).

E7 puede inhibir la apoptosis y muerte celular estimulando la expresión de las ciclinas A y E, que abrogan las actividades inhibidoras de crecimiento de la quinasa dependientes de ciclinas (CKIs) para antagonizar la activación de p53 (Cai *et al*., 2013; Jiang & Yue, 2014). La expresión de E7 también induce la aneuploidía (alteraciones cromosómicas numéricas) de la célula hospedera, lo que contribuye a la tumorogénesis (McLaughlin-Drubin & Munger, 2009). E6 y E7 son independientemente capaces de inmortalizar células humanas, pero su expresión combinada conduce a un efecto complementario y sinérgico, que a la vez es responsable de la mayor eficiencia de la transformación (Boulet *et al*., 2007).

2.5 Distribución de variantes de VPH

Los trabajos en variabilidad genética iniciaron con el VPH-AR 16 dada la prevalencia en CCU a nivel mundial, comparando las secuencias nucleotídicas de distintos aislamientos con el virus clon referencial o prototipo- el primer VPH 16 descrito (Picconi *et al.*, 2000). Ho *et al.* (1993) construyeron la primera filogenia basada en fragmentos LCR, de aislamientos procedentes de diferentes continentes. El árbol reveló cinco grupos filogenéticos o ramas, distribuidas de manera diferente dentro de los cinco continentes: la rama Asiática-Americana (AA) con variantes presentes en Latinoamérica y España, dos grupos principalmente africanos (Af-1 y Af-2), la rama de sudeste asiático (As) y la Europea (E) en todas las regiones distintas de África, quedando el clon referencial dentro de la rama E (Tornesello *et al.,* 2004). Posteriormente estudios similares fueron

realizados para VPH 18, encontrando tres ramas: la europea (E), la africana (Af) y la Asiática (As), la cual incluía al clon de referencia (Picconi *et al.*, 2000).

La detección de variantes presenta gran importancia epidemiológica debido a que algunos datos sugieren que mujeres infectadas con ciertas variantes tienen mayor probabilidad de persistencia viral y desarrollar una lesión cervical de alto grado (Tornesello *et al.*, 2011).

2.6 Bioinformática: estructura de datos y algoritmos

La bioinformática es el análisis computarizado de los datos biológicos, en el que la biología, la ciencia computacional y la tecnología de la computación se funden en una sola (Valencia, 2002). Consta de dos objetivos y subcampos complementarios: desarrollar herramientas computacionales y bases de datos y, generar nuevo conocimiento biológico que ayude a entender mejor los sistemas vivos (predicción de genes, estructuras proteicas y expresión génica, modelos evolutivos y análisis de genómica funcional) (Kanehisa & Bork, 2003).

Una secuencia de DNA es una secuencia escrita en cadena sobre el conjunto S de 4 símbolos de nucleótidos: adenina (A), citosina (C), guanina (G) y timina (T): S = {A, C, G, T}, escrito como s = ATTTCTGTAA (Thu-Hang, 2008).

Una secuencia de proteína s' es una cadena sobre el conjunto de 20 símbolos de aminoácidos S = {A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y}, por ejemplo, s' pueden ser: s '= GKGDPKKPRGKMSSYAFFVQTSREEHKKKHPDASVNF (Thu-Hang, 2008); los aminoácidos que corresponden a estas letras se muestran en el Anexo A.

Por otra parte, las bases de datos (BD) biológicas contienen todo tipo de información relacionada con las biomacromoléculas (secuencias, estructura, clasificación, localización, secuencias de motivos, patrones de expresión) (Xiong, 2006). De acuerdo a su contenido, las BD pueden ser catalogadas en BD primarias, BD secundarias y BD especializadas; alguno de las diferentes BD de acuerdo a su contenido se presenta en la Tabla 2.

Tabla 2. Bases de datos de diferentes categorías.

Categoría	Base de dato	Descripción	Referencia
DNA	GenBank	Secuencias de nucleótidos	Benson *et al*., 2000
	EMBL	Secuencias de nucleótidos	Stoesser *et al*., 2002
Proteínas	SWISSROT	Secuencias de aminoácidos	Bairoch & Apweiler, 2002
Genoma	Wormpep	Genoma de *C. elegans*,	Wilson, 1999,
		A. thaliana,	Huala *et al*., 2001,
		E. coli	Blattner *et al*., 1997
Estructuras	PDB	Estructuras	Westbrook *et al*., 2002
	SCOP	Clasificación de estructuras	Murzin *et al*., 1995
Motivos	Pfam	Alineamientos de dominios	Bateman *et al*., 2002

2.6.1 Alineamiento de secuencias

La comparación de secuencias constituyen la base del análisis bioinformático, en donde, diferentes secuencias son confrontadas buscando patrones comunes y estableciendo

correspondencias residuo a residuo (Barton & Sternberg, 1987). Se dice que dos secuencias son homólogas, si ambas descienden de un ancestro común; mide el porcentaje de residuos que son similares en cuanto a propiedades fisicoquímicas; por otro lado, la similaridad es el resultado directo de la observación de un alineamiento de secuencias cuantificado mediante porcentajes (Brown, 2002; Xiong, 2006). Existen dos tipos de alineamientos, el alineamiento de pares de secuencias y el alineamiento múltiple de secuencias.

2.6.1.1 *Alineamiento de pares de secuencias y algoritmos*

El objetivo consiste en encontrar el mejor emparejamiento de dos secuencias, de tal forma que se obtenga una máxima correspondencia entre los residuos (Waterhouse *et al*., 2009). Para un alineamiento pareado existen dos estrategias: alineamiento local y el alineamiento global. El Algoritmo Needleman-Wunsch, utilizado para alinear secuencias de nucleótidos o proteínas a través del alineamiento global, es más útil cuando las secuencias son similares y de más o menos la misma longitud (Needleman &Wunsch, 1970). Para encontrar el óptimo alineamiento global entre dos secuencias x´, y´:

Se construye una matriz D, de tamaño (m+1)x(n+1), donde m y n son las longitudes de las secuencias x' y y´ respectivamente, a partir del *score* de similaridad:

El algoritmo comienza inicializando D0, (0,0), la primera fila y la columna se llenan de acuerdo con el esquema de puntuación seleccionado que representa las penalizaciones por gap en el inicio de cualquier secuencia, el mejor *score* entre las secuencias x´ y y´ va a ser el valor de D(m,n) el último valor de la tabla, cada valor D(i,j) va a ser la puntuación máxima entre cada residuo x´(1... i) y el residuo y´(1... j); es decir, en cada posición de la matriz D, vamos a tener el máximo *score* de los posibles alineamientos locales hasta las posiciones i,j. Si se tiene un alineamiento óptimo hasta D(i-1,j-1), sólo hay tres opciones para lo que va a pasar en la siguiente posición: (i) los residuos para x´(i) y y´(j) coinciden (lo que no significa que sean iguales), (ii) se introduce un gap en x´ o (iii) se introduce un gap en y´.

Una vez obtenida la matriz, sólo queda reconstruir el camino desde la posición D(m,n) hasta la D(0,0). La entrada D(m,n) da la puntuación de alineamiento global. Para ello, habrá que recalcular los valores de D(i-1,j-1), D(i-1,j) y D(i,j-1), para rastrear de qué casilla venía el máximo correspondiente a D(i,j).

El algoritmo Smith-Waterman, **s**e utiliza con el objetivo de encontrar el *score* local más alto entre dos secuencias, las subsecuencias pueden ser parte o la totalidad de dos secuencias; por tanto, la alineación local también es global (Smith & Waterman, 1981). Dadas dos secuencias x´, y´, para hallar segmentos de secuencia muy similares, el primer paso consiste en construir la matriz D:

El algoritmo se inicializa en la primera fila y columna con cero en lugar de penalizaciones por gap, el cero es la condición que se incluye seguidamente para evitar obtener valores de similaridad negativos y la razón es para iniciar una nueva alineación local en lugar de extenderla, el Di, j valor depende de Di-1, J-1 (*match* o *mismatch*), Di, j-1 (inserción), o Di-1, j (supresión); el óptimo alineamiento local puede estar en cualquier lugar en la matriz en lugar de D(m,n), si se quisiera encontrar el par de residuos con la siguiente mejor similitud, el único requisito necesario sería comenzar el proceso desde el siguiente valor máximo de D, no asociado con la anterior traza.

La principal diferencia entre el algoritmo de Needleman-Wunsch y Smith-Waterman es que a los valores negativos en la matriz de puntuación de similitud se ponen a cero. Otra característica es que con Smith-Waterman el *score* se incrementa en una región de alta similaridad y disminuye fuera de tales regiones.

2.6.1.2 Alineamiento múltiple de secuencias y algoritmos

Consiste en alinear múltiples secuencias de tal forma que el máximo número de residuos de cada secuencia coincida con una función de puntuación; dentro de las aplicaciones encontramos, la detección de regiones o posiciones conservadas, la construcción de árboles filogenéticos y la predicción de estructura secundaria y terciaria de proteínas (Cline, 2000). Para realizar alineamientos múltiples existen tres clases de algoritmos que se pueden clasificar en alineamiento progresivo, alineamiento iterativo y alineamiento basado en bloques (Duc, 2012). La herramienta más utilizada es ClustalW, basada en el método progresivo, donde se parte de una serie de alineamientos por pares, para alinear grupos cada vez mayores (ensamblaje progresivo) siguiendo el orden de ramificación de un árbol guía (Thomson *et al*., 1994).

El algoritmo ClustalW, consiste en tres etapas principales, la primera etapa calcula las distancias entre cada par de secuencias por alineación de secuencias pareadas, debido a la simetría, sólo se requiere la parte superior o inferior de la matriz triangular de distancia; la alineación de secuencias utilizando este algoritmo, es calculada por un método de programación dinámica más lento (con relación a otras herramientas) pero preciso con base en un método heurístico, este método calcula una puntuación utilizando fragmentos que coinciden exactamente (k-tuplas), por la mejor alineación entre dos secuencias, menor es la penalidad fija por gap. Estas puntuaciones se calculan como el número de identidades en el mejor alineamiento dividido por el número de residuos comparados (excluyendo las posiciones de gap) (Thomson *et al*., 1994).

La segunda etapa utiliza la matriz de distancia calculada en la primera etapa para construir un árbol de guía, que se utiliza para guiar el proceso final del alineamiento múltiple, este árbol puede ser construido por dos métodos: Neighbour-Joining (NJ), que produce árboles sin raíces con rama longitudes proporcionales a la divergencia estimada o UPGMA (Un-weighted Pair Group Method with Arithmetic mean), que genera árboles ultramétricos, dado que se basa en la identificación de las parejas de secuencias más similares y en el cálculo de la media de las distancias entre ellas y el resto de las secuencias (Thomson *et al*., 1994; Xiong, 2006). En la tercera etapa se utiliza el árbol de guía para alinear progresivamente las secuencias. El término “progresivo” se usa porque ClustalW primero determina las secuencias más relacionadas y luego añade progresivamente las secuencias o grupos de secuencias menos afines, el resultado de la

alineación progresiva, constituye la secuencia final de alineamiento múltiple (Thomson *et al*., 1994).

2.6.2 Análisis filogenético y evolución de VPH

Una filogenia es la descripción de las relaciones entre las especies, poblaciones, individuos o genes que tienen relación ancestro-descendiente (historia evolutiva), La mayoría de las técnicas producen árboles filogenéticos (una estructura matemática utilizada para modelar la historia de la evolución real de las especies o de sus partes) (Delsuc *et al.*, 2005; Choi & Gomez, 2009). La reconstrucción de la relación filogenética utilizando secuencias de DNA, RNA o de aminoácidos es un proceso jerárquico que consta de cuatro pasos: (i) la alineación de secuencias, (ii) la selección del modelo apropiado, (iii) el método de construcción del árbol y (iv) la evaluación de la filogenia resultante (Xiong, 2006; Talianová, 2007).

La evolución de los PV ha sido objeto de intensivos debates (de Villiers *et al.*, 2004; de Villiers, 2013) debido a que los papilomavirólogos al principio no concertaban en que gen y número de genes debían utilizar o si la asignación de nuevos tipos podía ser realizada con amplificados parciales generados por la Reacción en Cadena de la Polimerasa (sigla PCR del inglés *Polymerase Chain Reaction*) o con el genoma completo, por lo que estudios comparativos que utilizaron ORF de E6, L1 o combinados (van Ranst *et al.*, 1992; Chan *et al.*, 1995), generaron árboles filogenéticos con relaciones similares o incluso idénticas; sin embargo, un amplicón altamente conservado de 291 pb demostró que un pequeño segmento del ORF del gen L1 es suficiente para generar comparaciones filogenéticas altamente informativas (de Villiers, 2013).

Por otra parte, entablar conclusiones sobre la evolución de PV puede ser en gran parte especulativo teniendo en cuenta que el conocimiento sobre la diversidad de PV en hospedadores no humanos es desequilibrada y limitada (Gottschling *et al.*, 2011), por lo que han aumentado los esfuerzos en la toma de muestras de PV (en su mayoría no humano) para comprender sus mecanismos evolutivos, que en la actualidad no solo involucran eventos de diversificación por coespeciación (hipótesis rechazada por incongruencias topológicas), sino que también incluye cambios debido a la recombinación en el huéspedero y eventos de duplicación en el mismo, durante la realización de nuevos nichos ecológicos por la radiación adaptativa (Batista *et al.*, 2011).

2.6.2.1 Métodos para la construcción de árboles filogenéticos

La construcción de árboles filogenéticos se basa en la cladística (la clasificación sistemática de los organismos sobre la base de características comunes que se derivan de un ancestro común) y generalmente los métodos de construcción pueden ser de tres tipos: (i) basados en distancia, (ii) basados en caracteres y (iii) probabilísticos o de inferencia bayesiana (Grievink, 2009; Potiny, 2010). Para los análisis filogenéticos de esta investigación, se implementó el método basado en caracteres de Máxima verosimilitud (ML), que de forma general considera las diferencias entre las secuencias de entrada como caracteres discretos y por cada carácter considera que está en un estado en particular (por ejemplo, presente vs. ausente) (Rizzo & Rouchka, 2007).

En el algortimo ML, la verosimilitud se la considera como la probabilidad de observar los datos alineados X dado un modelo θ, $P(X|\theta)$, donde los parámetros de θ incluyen un árbol, longitudes de ramas y un modelo de evolución, el principio de la verosimilitud busca parámetros

que maximicen la probabilidad P (Cavalli-Sforza & Edwards, 1967). La función de verosimilitud se define como:

La máxima verosimilitud es el valor de θ que maximiza L(θ), el procedimiento básico es calcular la probabilidad de cada columna del alineamiento, por tanto la probabilidad del alineamiento entero es simplemente el producto de las probabilidades de cada columna, la verosimilitud del alineamiento está dada por:

A nivel filogenético, el método de ML, intenta estimar la cantidad de cambio real de acuerdo con un modelo establecido, este método evalúa la hipótesis (el árbol) mediante una función (verosimilitud o similitud), que maximiza la probabilidad de obtener datos (matriz de secuencias de DNA) dado un modelo evolutivo, de esta forma, conforme al criterio de ML, el mejor árbol de cuatro secuencias conectadas dos a dos es aquel que presenta el mayor valor de verosimilitud, independiente del número de transformaciones de estado carácter que necesite; una de las grandes ventajas de ML es la capacidad de hacer comparaciones estadísticas entre las topologías y el conjunto de datos, por ejemplo, se puede decir que un árbol es exactamente un por ciento más probable que otro, la principal fortaleza del método ML es que es posible formular hipótesis sobre las relaciones evolutivas (Rizzo & Rouchka, 2007).

3 ANTECEDENTES

Los inicios de la tipificación viral

Danos *et al.* (1982) identificaron por primera vez la secuencia completa de nucleótidos de un papilomavirus humano, el cual fue designado como tipo 1a (7811 pb), concluyendo que la organización general del genoma viral era diferente de otros papovavirus en su momento relacionados (SV40, BKV, polioma). Schwarz y colaboradores (1983) establecieron la secuencia de nucleótidos del VPH 6b (7902 pb) que se comparó con las secuencias de DNA del VPH 1a y el virus del papiloma bovino tipo 1 (VPB 1), concluyendo que las secuencias de aminoácidos de L1 son altamente conservadas.

Seedorf *et al.* (1985) determinaron la secuencia de nucleótidos del VPH 16 (7904 pb); la comparación de homología entre los PVs previamente secuenciados (VPB 1, VPH 1a, VPH 6b), reveló una interrupción en el marco de lectura que codificaba para la proteína E1.

Dartmann *et al.* (1986) a partir del DNA de papilomas laríngeos y de verrugas genitales identificaron la secuencia de nucleótidos de VPH 11 (7931 pb), la cual mostró un alto grado de homología con HPV 6b (82 %). Para el mismo año, Cole y Streeck determinaron la secuencia de nucleótidos de VPH 33 (7909 pb) asociado previamente a CC, estableciendo posteriormente una alta correspondencia entre los ORFs y otras características oncogénicas de VPH 16.

Cole & Danos (1987) determinaron la secuencia de nucleótidos de VPH 18 asociado con el CC, identificando una serie de características específicas para el PV genital y construyeron un árbol filogenético, donde diferenciaron PV humanos de otros PV animales, debido que E1 tiene un ORF mucho más largo y una región LCR característica; a nivel filogenético observaron que VPH 18 está situado entre los virus benignos genitales VPH 6 y VPH 11, y los aislamientos malignos VPH 16 y VPH 33, concluyendo que podría representar una evolución intermedia entre PV oncogénicos. Por su parte, Naghashfar y colaboradores (1987) a partir de una lesión cervical recurrente que se creía relacionada con VPH 18, clonaron a VPH 45, cuya principal distinción consistió en los patrones de digestión generados con enzimas de restricción.

Goldsborough *et al.* (1989) identificaron la secuencia de nucleótidos de VPH 31 (7912 pb), que posteriormente utilizaron para deducir su organización genómica y su relación con el CC; a partir de los ORF de E6 y E7, identificaron los motivos Cys-X-X-Cys, conservados en todos los VPH y distinguibles entre los tipos 16, 18, 31 y 33 asociados a CC. Por su parte, Takami *et al.* (1991) clonaron y caracterizaron a VPH 52, el análisis de secuencias reveló que los ORF E6-E7 contenían varios motivos de unión al DNA (Cys-X-X-Cys), tal como los VPH previamente secuenciados, identificaron que el ORF de E6 contiene un empalme de donantes y aceptores de señales, lo que puede permitir la expresión de la proteína E6; por su parte, el ORF de E7 codifica una secuencia de aminoácidos que se conserva en algunos virus tumorales de DNA y está implicado en la unión de la proteína pRb y en la transformación celular.

Kirii *et al.* (1991) tras realizar la secuenciación VPH 58 (7824 pb), determinaron que a nivel de la secuencia de nucleótidos exhibía una extensa homología con VPH 33 (77 %). De acuerdo a

las secuencias de aminoácidos concluyeron que VPH 58 se encontraba más estrechamente relacionado con los tipo 16, 31 y 33, y relativamente más alejado de VPH 6, 11, 18 y 45.

Marich *et al.* (1992) posterior a realizar la secuenciación de VPH 35 (7851 pb), analizaron la región LCR, revelando un elemento de 20 pb adicionales presentes en todos los tipos de VPH asociados con la proliferación maligna; el análisis filogenético con E6 y E7 de los tipos anogenitales conocidos hasta el momento: 6, 11, 16, 18, 31, 33, 35, 39, 43, 44, 45, y 51, indicó tres subgrupos distintos de VPH; los asociados con lesiones benignas y dos ramas de los tipos de VPH asociados más frecuentemente con la proliferación maligna.

Durante la siguiente década, los VPH aumentaron exponencialmente la atención de la comunidad de papilomavirólogos, debido a que cuantitativamente constituyeron el grupo más importante de los virus asociados con neoplasias benignas y malignas en humanos; sin embargo, con el tiempo otros tipos de PV también fueron identificados en mamíferos, aves y reptiles. Múltiples estudios han determinado que los PV han evolucionado durante millones de años a partir de ancestros comunes dentro de ecosistemas de hospederos específicos, por lo que especialmente para este grupo, la variabilidad en ciertas regiones del genoma, ha jugado un papel fundamental a la hora de describir la nomenclatura y la biología evolutiva de VPH, tal como se entiende a través de análisis de árboles filogenéticos.

Hacia un sistema de clasificación basado en filogenias, variantes y distribución de tipos.

Chan *et al.* (1992) secuenciaron segmentos genómicos de 118 VPH 16, procedentes de Singapur, Brasil, Tanzania y Alemania; la región LCR reveló 38 variantes, la mayoría difirieron por una o varias mutaciones puntuales, los árboles filogenéticos apoyado con el ORF de E5 indicó dos ramas principales: variantes de Singapur pertenecientes a una rama denominada Eurasia; Alemania y Tanzania a una rama de África, mientras que las variantes brasileñas estaban claramente divididas entre las dos ramas, los datos sugirieron dos hipótesis: i) que el VPH 16 se desarrolló por separado durante un largo período en África y Eurasia y ii) los representantes de ambas ramas pudieron haber sido trasladados a Brasil a través de la pasada inmigración colonial.

Van Ranst *et al.* (1992) proporcionaron una clasificación de los VPH relacionados con el tropismo del tejido infectado y el potencial oncogénico de cada tipo de VPH, el análisis fue basado en el alineamiento de los ORFs de E1, E2, E6, L1 L2 y la construcción de los árboles filogenéticos para 18 VPH fueron validados, y constituyeron la base de la taxonomía del virus.

De Villiers *et al.* (2004) describieron la taxonomía de 118 PV de orden superior siguiendo los criterios generales establecidos por la ICTV, los PV genitales humanos fueron considerados como un género el cual denominaron “Alfa”; determinaron que los ensamblajes de orden inferior dentro de cada género deben ser tratados como especies porque están estrechamente relacionados filogenéticamente.

Arias-Pulido *et al. (2005)* teniendo en cuenta el limitado conocimiento de las variantes del VPH 18, seleccionaron muestras al azar de EE.UU y analizaron las variaciones de nucleótidos en las regiones E2, E6, L1 y LCR, el estudio agrupó los genes analizados en tres ramas principales: la Asiática-americana (AA), la europea (E) y la africana (Af), donde la rama Af mostró la mayor variabilidad de nucleótidos en comparación con la secuencia de referencia del VPH 18.

Huertas-Salgado *et al. (2011)* propusieron un nuevo criterio para el agrupamiento y la nomenclatura de variantes moleculares de E6 del VPH 16 basados en análisis descriptivos y comparativos de secuencias de nucleótidos.

De Villers *(2013)* realizó una actualización de la clasificación de los 170 tipos de VPH conocidos en la actualidad haciendo uso de la región conservada en el ORF del gen L1; a su vez, presentaron argumentos para la aplicación de metagenómica, para ahondar en el conocimiento de la familia Papillomaviridae y aportar en la identificación nuevos tipos virales.

Con relación al departamento de Nariño, se destacan dos trabajos meramente descriptivos en salud pública, donde poco se hace relación con el cribado a nivel molecular; sin embargo, dichas investigaciones han llevado al departamento a incursionar en la identificación de los tipos virales circundantes de la zona, información de la cual no se disponía hasta hace tres años.

Sánchez *et al*. (2012) realizaron la detección y tipificación de VPH mediante la técnica *Reverse line blot*- GP5+/GP6+ en 160 mujeres procedentes del municipio de Pasto, los genotipos encontrados fueron VPH 11, 45, 53, 54, 55, 56, 70 y 84, a la vez que no se encontró relación entre los genotipos detectados y factores como la edad, el número de compañeros sexuales, el uso de anticonceptivos, la coinfección con otros microorganismos.

Nicola *(2014)* realizó la tipificación del VPH a partir de lisado celular sin extracción previa de DNA, a través de PCR-L1 + secuenciación y analizó su relación con características poblacionales y lesiones en CCU en una población de mujeres del municipio de Pasto, se logaron detectar los genotipos de VPH 6, 16, 33, 45, 58, 73, 97 y uno no determinado, a la vez no se encontró relación entre los genotipos detectados y la edad, el número de compañeros sexuales, el uso de anticonceptivos o la coinfección con otros microorganismos.

4 METODOLOGÍA

4.1 Minería de datos

La selección de los tipos de VPH para el análisis bioinformático, se realizó teniendo en cuenta la literatura científica y su frecuencia en los reportes a nivel mundial (Clifford *et al*., 2005; Villa, 2012); además, se tuvo en cuenta el estudio de tipificación molecular de VPH realizado por Nicola (2014) para incluir en el análisis filogenético, debido a que constituyen las primeras secuencias de VPH obtenidas en el Departamento de Nariño y dan una idea sobre los genotipos circundantes de la zona. Las secuencias de referencia (control para el análisis de agrupamiento y la identificación de variantes) de los ORF L1, E6 y E7 fueron obtenidas del GenBank (Benson *et al*., 2014) y se confirmaron mediante *screening* sobre genomas completos consignados en la base de datos PAVE (http://pave.niaid.nih.gov), especializada en Papilomavirus, los números de acceso, la longitud, el tipo de VPH y la fuente de publicación se presentan en la Tabla 3.

Tabla 3. Genomas y genes de referencia utilizados para el análisis bioinformático.

VPH	Longitud genoma (pb)	L1 (pb)	E6 (pb)	E7 (pb)	Código de acceso GenBank	Referencias
6A	7902	1502	452	296	X00203	Schwarz *et al*., 1983
6B1	8012	1502	452	296	AF092932	Kovelman *et al*., 1999
6B3	8010	1502	452	296	L41216	Hofmann *et al*., 1995
11	7931	1505	452	296	M14119	Dartmann *et al*., 1986
16	7904	1595	476	296	K02718	Seedorf *et al*., 1985
18	7857	1706	476	317	X05015	Cole & Danos, 1987
31	7912	1514	449	296	J04353	Goldsborough *et al*., 989
33	7909	1499	449	293	M12732	Cole & Streeck, 1986
35	7879	1508	449	299	X74477	Marich *et al*., 1992
45	7858	1619	476	320	X74879	Naghashfar *et al*., 1987
52	7942	1589	446	299	X74481	Takami et al., 1991
58	7854	1574	449	296	D90400	Kirii *et al*., 1991
73	7700	1511	446	293	X94165	Volter *et al*., 1996
97	7843	160	446	320	EF202168	Chen *et al*., 2007

BPV 1*	7945	1 148 7	413	383	X02346	Chen *et al*., 1982

* Secuencia utilizada como *outgroup* en los análisis filogenéticos.

Las secuencias obtenidas del estudio de Nícola (2014) se analizaron a partir de los cromatogramas y archivos .txt, las secuencias en formato FASTA se procesaron para confirmación bioinformática del tipo de VPH, para lo cual se realizó un alineamiento de primers GP5-GP6 (Sánchez *et al*., 2013) sobre cada secuencia utilizando el algoritmo BLASTn (Johnson *et al*., 2008) disponible en NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi); para depurar y editar estas secuencias se utilizó el programa BioEdit V. 7.2.5 (Hall, 2011). Para la confirmación del tipo viral, con cada secuencia se realizó un alineamiento múltiple (comparación de secuencia "*query*" con las disponibles en GenBank) mediante la herramienta Blastn/NCBI y se analizó el e-*value* (probabilidad de coincidencia debido al azar), el *score* del alineamiento, el porcentaje de similaridad entre pares de secuencias, el porcentaje de identidad (superior al 95 %) y la anotación.

Para seleccionar y obtener las secuencias de los genes L1, E6 y E7 de los diferentes genotipos de VPH-AR y BR más frecuentes en Latinoamérica, se realizó una búsqueda exhaustiva a partir de literatura científica y bases de datos biológicas, empleando los términos de búsqueda o palabras clave: "*Human Papillomavirus* type", "*Human Papillomavirus gene* L1, E6 o E7".

Se revisaron artículos científicos de las bases de datos bibliográficos: ScienceDirect (http://www.sciencedirect.com/), Springer (http://www.springer.com/?SGWID=10-102-0-0-0) y Pubmed (http://www.ncbi.nlm.nih.gov/pubmed); para complementar la búsqueda, se revisaron publicaciones científicas en la editorial ELSEVIER (http://www.elsevier.com/journals/title/a), SCielo (http://www.scielo.org/php/index.php?lang=es) y Redalyc (http://www.redalyc.org/homeBasic.oa), de los artículos seleccionados, se realizaron búsquedas manuales de los estudios relacionados para Latinoamérica y, para los datos que fueron publicados más de una vez o de una combinación de datos antiguos y recientes, la referencia se registró con respecto a la fuente primaria. Una vez obtenida la literatura científica, se seleccionaron y clasificaron los artículos más relevantes en el tema con base en: Tipo de publicación, población de estudio, tipo de estudio (Tabla 4).

Tabla 4. Criterios de inclusión y exclusión bibliográfica.

Parámetro	Inclusión	Exclusión
Según el tipo de publicación	Revistas y notas científicas, revisiones sistemáticas, ensayos clínicos, estudios de casos y controles, estudios transversales y series de casos.	Revisiones narrativas, cartas al director, editoriales, comentarios, y comunicaciones o publicaciones de congresos.
Según la población de estudio	Hombres y mujeres diagnosticados con VPH-AR y -BR, en poblaciones de países latinoamericanos, con resultados de citología normal o con alguna anormalidad citológica e infección por	Resultados de países diferentes a Latinoamérica.

	VPH.	
Según el tipo de estudio	Estudios de detección o tipificación molecular de VPH donde se hayan reportado secuencias del virus.	Estudios citológicos y descriptivos.

Los artículos seleccionados fueron codificados y almacenados en una carpeta denominada "VPH Latinoamérica", con el fin de organizar los resultados y acceder a información en análisis posteriores. La codificación partió del número uno (Ej. 1, 2, 3… 24), seguido de la fuente de publicación (Autor, año). Se construyó una base de datos con el programa Microsoft Office Excel 2013®, donde se registraron los artículos de acuerdo con: el código asignado en este estudio, la base de datos de donde se obtuvo, la revista donde se publicó, la fuente de publicación, país de publicación, título del artículo, genes y tipos analizados y el DOI (Digital Object Identifier) (Anexo B).

Se realizó la búsqueda de secuencias genómicas de VPH en la base de datos de secuencias de nucleótidos "Nucleotide" del NCBI (http://www.ncbi.nlm.nih.gov/nuccore), de donde se obtuvo además parte de su anotación. Todas las secuencias se descargaron del GenBank en formato FASTA, se almacenaron en bloc de notas (.txt) y en archivos independientes; se construyó una base de datos con el programa Microsoft Office Excel 2013®, donde se consignó la información correspondiente a: código de acceso a GenBank, país, estado u/o departamento de origen, tipo viral, nombre del gen, tamaño del ORF de la secuencia y la fuente de publicación. Como se describe en la Tabla 5 las secuencias de VPH fueron seleccionadas con base en: el tipo de secuencia, la calidad de la secuencia y su anotación.

Tabla 5. Criterios de inclusión y exclusión de secuencias.

Parámetro	Inclusión	Exclusión
Según el tipo de secuencia	Secuencias de DNA del ORF L1 de VPH (de Villiers, 2013) y ORFs E6 y E7 (García-Vallvé *et al*., 2005).	Secuencias proteicas, secuencias promotoras y regiones no codificantes del genoma de VPH.
Según la calidad de la secuencia	ORFs completos (aprox. 1525 nt) o parciales (fragmentos de 120-808 nt) de las secuencias de genes virales L1, E6 y E7.	ORFs de los genes E1, E2, E4, L2 y región LCR.
Según la anotación de la secuencia	Información sobre: procedencia, año, método molecular con el cual se obtuvieron los ORFs de secuencias de DNA de los genes virales L1, E6 y E7	Secuencias sin información relevante, cuyo análisis genere información dudosa.

4.2 Análisis de comparación de secuencias mediante alineamientos

Alineamientos de secuencias del gen L1. Las secuencias L1 obtenidas del GenBank (http://www.ncbi.nlm.nih.gov/genbank/) se analizaron en dos grupos teniendo en cuenta los

diferentes tamaños de fragmento (i) para secuencias de longitud promedio aproximada de 808 a 1575 nt, (ii) para las secuencias de longitud promedio 269 a 667 nt, de aquí se organizaron y consignaron en un archivo todos los tipos de VPH y se procedió a realizar alineamientos múltiples con la herramienta ClustalW2 (Thompson *et al*., 1994) disponible en EMBL-EBI (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Alineamientos de secuencias E6 y E7. Por cada gen de un tipo viral, se realizaron alineamientos múltiples mediante la herramienta ClustalW2 (Thompson *et al*., 1994), incluyendo las correspondientes secuencias de referencia reportadas para los tipos de VPH descritos en la Tabla 3. Con los resultados de los alineamientos múltiples obtenidos mediante ClustalW2 se construyeron árboles guía con el algoritmo Neighbour Joining disponible en EMBL-EBI (http://www.ebi.ac.uk/Tools/webservices/services/msa/clustalw2_soap), estos árboles se tuvieron en cuenta para orientar el posterior análisis de agrupamiento.

4.3 Análisis filogenético

Construcción y análisis de filogenias con L1. Se construyeron árboles filogenéticos partiendo de los alineamientos múltiples descritos anteriormente, el análisis se realizó con base en la diferencia del tamaño de fragmento y para los agrupamientos se tuvieron en cuenta los criterios reportados por De Villiers *et al*. (2004), para los análisis taxonómicos, epidemiológicos y filogenéticos, según lo cual, para la asignación de un "tipo viral" de VPH la secuencia de DNA del ORF del gen L1 difiere de cualquier otro tipo en al menos un 10 %.

Las aproximaciones filogenéticas fueron realizadas bajo el criterio de máxima verosimilitud (ML), con el programa MEGA v. 6.0.5 (Tamura *et al*., 2013), haciendo uso del modelo de sustitución nucleotídica Kimura 2-p (Chan *et al*., 2011), para el cual se implementó el algoritmo *bootstrap*, que permitió ejecutar búsquedas de árboles óptimos y la obtención de soportes de ramas en un único análisis (Felsenstein, 1985), en las ramas del árbol se representaron los valores de *bootstrap* (1000 réplicas) que fueron significativos (por encima de 50 %). Como grupo externo (*outgroup*) se empleó la secuencia de referencia de Virus Papiloma Bovino tipo 1 (BPV 1), dada la relación filogenética cercana con el Virus Papiloma Humano (Ong *et al*., 1993).

Árboles filogenéticos con E6 y E7. Se construyeron árboles filogenéticos por cada tipo y para cada gen, con el fin de analizar las agrupaciones de variantes y establecer un patrón de acuerdo a la frecuencia de mutaciones específicas de cada gen en el agrupamiento. Para verificar las relaciones intratipo de los árboles E6 y E7 únicamente se emplearon las secuencias de referencia de cada tipo (Tabla 3) (Chen *et al*., 2009; Chen *et al*., 2011).

4.4 ***Identificación de variantes de tipos virales***

A partir de los alineamientos de E6 y E7, la predicción de variantes para cada gen fue realizada haciendo una comparación residuo a residuo, entre cada secuencia de nucleótidos del alineamiento y la secuencia de referencia del tipo analizado, para identificar las variantes se construyeron tablas que resumieron las posiciones donde se presentaron mutaciones puntuales, la asignación de los números de posición para cada nucleótido se realizó a partir de la longitud total del genoma de

referencia de cada tipo. Con el fin de establecer las mutaciones no sinónimas, el cambios de aminoácidos en cada variante se realizó con la traducción *in silico* de los codones, haciendo uso de la herramienta *Translate or Reverse-Translate* del programa BioEdit V. 7.2.5, la asignación de los números de posición para cada aminoácido se realizó a partir de la longitud del gen de referencia de cada tipo (Huertas-Salgado *et al*., 2011). De acuerdo al criterio establecido por De Villiers *et al*. (2004), se consideró una variante si dicha variación entre una secuencia y la secuencia de referencia era menor al 2 %.

4.5 Análisis de la distribución geográfica de variantes

Con el fin de visualizar la distribución geográfica respecto a la frecuencia de las variantes de E6 y E7 de cada tipo viral en Latinoamérica, se construyeron mapas con el sistema de proyección cartográfica Gauss Datum Magna Sirgas mediante el programa ArcGis 10.2.2 (http://www.esri.com/software/arcgis/arcgis-for-desktop/free-trial), la proyección de puntos se realizó a partir de la coordenada plana de cada país en el globo terráqueo.

5 RESULTADOS Y DISCUSIÓN

5.1 Revisión de literatura especializada y bases de datos de secuencias

Durante la búsqueda de artículos se identificaron un total de 2408 resultados, de los cuales el 52,2 % de los artículos estuvo representado por la BD Springer, seguido de PubMed (25,9 %), ScienceDirect (21,6 %) y SCielo (0,3 %), los resultados de búsqueda obtenidos para Redalyc fueron excluidos debido al aporte poco significativo al problema de interés. Las secuencias de nucleótidos incluidas en la revisión, se obtuvieron a partir de 24 artículos publicados entre 1996 y 2014, identificados durante una búsqueda sistemática de literatura en el periodo comprendido entre abril - junio de 2014, el proceso de selección de artículos se resume en la Figura 5.

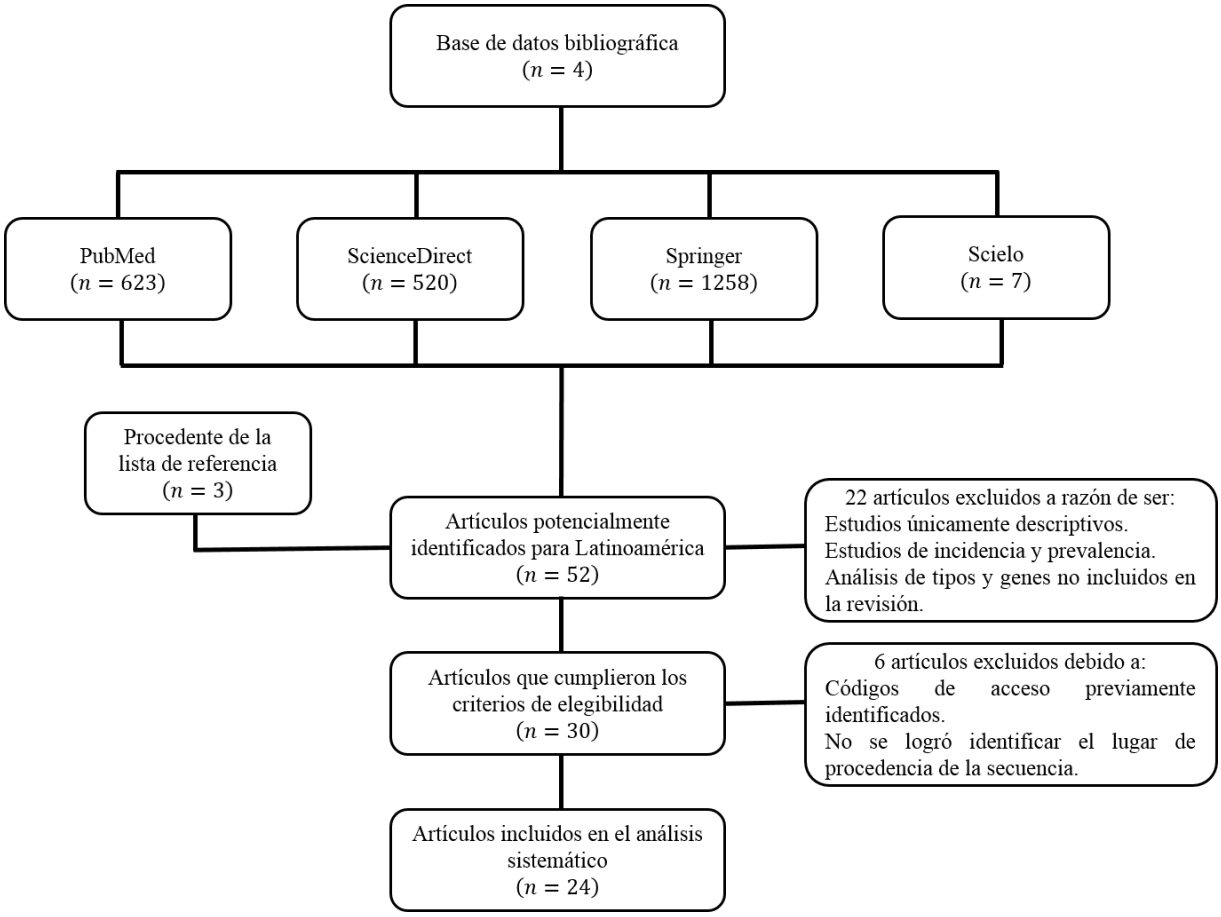


Figura 5. Diagrama de flujo para el proceso de selección de artículos. *Fuente:* Esta investigación.

Se identificaron 52 estudios que cumplieron los criterios de elegibilidad para ser incluidos en la revisión; sin embargo, el análisis a texto completo permitió seleccionar únicamente 30 documentos que cumplían con todos los criterios de inclusión previamente establecidos, de estos fueron excluidos tres trabajos, puesto que las secuencias allí analizadas originalmente pertenecían a trabajos reportados previamente (Li *et al.*, 2009; Ntova *et al.*, 2012; Burk *et al.*, 2013) y tres más, debido a que no fue posible discriminar el país latinoamericano de origen de la secuencia (Ong *et al.*, 1993; Lizano *et al.*, 2006; Jelen *et al.*, 2014) (Figura 5), una breve descripción de los estudios incluidos en la revisión se presentan en la Tabla 6.

La exclusión de las 22 publicaciones restantes se debió a diferentes razones: diez artículos relacionados únicamente con la descripción de variantes (Yamada *et al.*, 1997; Heinzel *et al.*, 1995; Berumen *et al.*, 2001; Cruz *et al*., 2004; González-Losa *et al.*, 2004; Veras *et al.*, 2005; Tonon *et al.*, 2007; Junes-Gill *et al.*, 2008; Huertas-Salgado *et al.*, 2011; Buitrago *et al.*, 2012;), cinco artículos relacionados con la incidencia y prevalencia de los tipos virales (Camara *et al.*, 2003; Oliveira *et al.*, 2006; Baldez da Silva *et al.*, 2009; Canul *et al.*, 2010; Tamegão-Lopes *et al.*, 2014), tres artículos mayormente descriptivo a nivel metodológico (Moreno-Acosta *et al.*, 2008; Schiffman *et al.*, 2010; Cruz *et al.*, 2013), dos artículos relacionaban genes diferentes a los de interés (Calleja-Macias *et al.*, 2004; Raiol *et al.*, 2007) y dos últimos que involucraban tipos virales no relacionados al estudio (Gottschling *et al.*, 2007; Almeida *et al*., 2013) (Figura 5).

Tabla 6. Descripción de las características de los estudios incluidos en la revisión sistemática.

#	Autor	País*	Gen (s)	Método de detección	Tipo viral (s)	Tipo de lesión			
						N**	ASCUS	NIC	CC
1	Stewart *et al.*, 1996	E.U	L1	MY09/11	18, 33, 35, 45, 52, 58				✔
2	DeFilippis *et al.*, 2002	E.U	E6, E7	PCR anidada, LCRS/E7AS, pU1M/pU2R	16				
3	Cerqueira *et al.*, 2003	Brasil	L1	MY09/11	58			✔	
4	Calleja-Macias *et al.*, 2005	E.U	E6, L1	MY09/11	31, 35, 52				
5	Chen *et al.*, 2005[a]	E.U	E6, E7, L1	MY09/11, DB	16	✔	✔	✔	✔
6	Cerqueira *et al.*, 2007	Brasil	L1	MY09/11 + RFLP	18, 33	✔		✔	
7	Cerqueira *et al.*, 2008	Brasil	E6	MY09/11 + RFLP	18	✔		✔	
8	Chen *et al.*, 2009	E.U	E6, E7, L1	MY09/11, DB	18, 45, 97			✔	✔
9	López-Revilla *et al.*, 2009	México	E6	PCR anidada, LCRS/E7AS, Pu1M/Pu2R	16			✔	✔
10	Raiol *et al.*, 2009	Brasil	E6, L1	MY09/11 + RFLP	31, 33, 52	✔		✔	
11	Castro *et al.*, 2011	Brasil	L1	PCR MY09/11	16, 33, 58			✔	✔
12	Chagas *et al.*, 2011	Brasil	E6, E7	Primers t-s E6/E7	31	✔		✔	
13	Chan *et al.*, 2011	Hong Kong	L1	PCR anidada	58	✔	✔	✔	✔
14	Chen *et al.*, 2011[a]	E.U	E6, E7, L1	MY09/11, DB	16, 31, 33, 35, 52, 58			✔	✔
15	Smith *et al.*, 2011	E.U	E6, E7, L1	PCR anidada MY09/11, GP5+/GP6+	16			✔	✔
16	Cornet *et al.*, 2012	Francia	E6	GP5+/GP6+ EISA	16	✔	✔	✔	✔

#	Autor	País*	Gen (s)	Método de detección	Tipo viral (s)	Tipo de lesión			
						N**	ASCUS	NIC	CC
17	Chagas *et al.*, 2013	Brasil	E6, E7	MY09/11, primers t-s E6/E7	31			✔	
18	Chen *et al.*, 2013[a]	E.U	E6, E7, L1	MY09/11, DB	18, 45			✔	✔
19	Chan *et al.*, 2013	Hong Kong	E6, E7	PCR anidada	58	✔	✔	✔	✔
20	de Matos *et al.*, 2013	Brasil	L1	MY09/11 + RFLP	6, 11 [H]				
21	Godínez *et al.*, 2013	España	E6, E7	PCR anidada, GP5+/GP6+	33, 45, 58				✔
22	Chen *et al.*, 2014[a]	Francia	E6, E7	PCR anidada	33	✔	✔	✔	✔
23	Chen *et al.*, 2014b	Francia	E6, E7	PCR anidada	45	✔	✔	✔	✔
24	Zhang *et al.*, 2014	Hong Kong	E6, L1	PCR anidada	52	✔		✔	✔

*País donde se analizaron originalmente las secuencias. ** N= sin lesión.
(1… 23) los números de la columna indican el código del estudio incluido en esta investigación. La técnica MY09/11 utiliza cebadores generales para L1 + PCR; Primers t-s: cebadores tipo específico; DB: *Dot Blot*. H: muestras procedentes de pacientes hombres. *Fuente*: esta investigación.

De los 24 trabajos referenciados (Tabla 6), el 37,5 % fueron realizados en América latina (México y Brasil) , principalmente Brasil, de donde se lograron extraer datos de ocho estudios, el 33,3 % fueron desarrollados en Norteamérica (E.U.), mientras que el continente europeo y asiático con cuatro y tres trabajos respectivamente, sumaron el 29,2 % restante. Para Latinoamérica existe una basta información sobre tipificación viral, que constituye un aporte importante en la epidemiología de VPH; sin embargo, las cifras de publicaciones actuales realizadas en Latinoamérica para secuencias de DNA del VPH no son significativamente comparables con el número de estudios realizados en otras regiones geográficas del planeta donde profundizan sobre patrones de especiación (Chan *et al*., 1992), coevolución (Ong *et al*., 1993) o el desarrollo de vacunas terapéuticas (Ahmed *et al*., 2013).

La técnica de detección del DNA viral ampliamente utilizada en las 24 investigaciones fue la PCR (MY09/MY11 y GP5+/GP6+) y cuatro de ellas previo a la secuenciación (16,7 %), hicieron uso del sistema *Dot Blot* (DB) (Tabla 6), un sistema de hibridación que consiste en una membrana de nitrocelulosa con sondas específicas para VPH-AR y BR (Melo *et al*., 2003). La selección del método de detección del VPH suscita el problema con la variación en las tasas de detección de genotipos, diferentes entre un método y otro, ya sea por factores como (i) la procedencia y estado de la muestra (congelada, fijada con parafina o raspados), (ii) la sensibilidad del método de prueba y (iii) la cobertura de genotipos de VPH en el panel de ensayo.

En la actualidad, no hay consenso sobre el método más adecuado para detectar el VPH; sin embargo, los métodos basados en la PCR (34,8 %) reportan una mayor prevalencia que los métodos de hibridación (32,9 %), a la vez que la PCR anidada (MY09/MY11 y GP5+/GP6+) aumenta la sensibilidad en comparación que una PCR simple, porque el conjunto de cebadores

detectan una amplia gama de tipos, así como los VPH con un bajo número de copias y realizan una mejor caracterización de la muestra infectadas con múltiples cepas de VPH (Syrjänen, 2010).

Tan solo para esta revisión sistemática, el 45,8 % de los estudios reportó la detección de algún tipo de VPH procedente de citología normal (Tabla 6); un porcentajes similar (48,5 %) fue calculado por Quintero-Vega *et al*., cuando compararon casos catalogados con citología normal después de realizar la tipificación del virus por PCR, dicha falencia en el ensayo clínico tradicional puede traducirse en un problema de salud pública generalizado, sobre todo en las zonas de países en desarrollo, donde se presenta mayor prevalencia del virus y no cuentan con adecuados programas de cribado (Quintero-Vega *et al*., 2008). Variaciones en la frecuencia y la prevalencia del virus pueden ser atribuibles a las diferencias en los métodos de detección, la sensibilidad y tipos de muestra analizadas (Castellsagué, 2008; Syrjänen, 2010).

A partir de la revisión sistemática de literatura se lograron obtener un total de 655 códigos de acceso de secuencias a los diferentes tipos de VPH y la búsqueda en NCBI permitió identificar doce nuevos autores con trabajos en categoría "*in press*" (sin publicar), que oscilaban entre los años 2001 a 2014, los cuales aportaron 115 diferentes códigos de acceso, el resumen de la tabla 7 relaciona el número total de las secuencias y procedencia de la misma.

Tabla 7. Número total de secuencias recuperadas según su procedencia.

País	Tipo viral (s)	Gen (s)	Número de secuencias
Costa Rica	16, 18, 31, 33, 35, 45, 52, 58, 97	E6, E7, L1	405
Brasil	6, 11, 13, 16, 18, 31, 33, 35, 45, 52, 58	E6, E7, L1	150
México	16, 52, 58	E6, E7, L1	85
Argentina	16, 18, 45, 52, 58	E6, E7, L1	48
Paraguay	16, 33, 45, 58	E6, E7, L1	25
Cuba	16	E6, L1	19
Perú	16, 33, 45	E6, E7	13
Chile	16, 18, 33, 45	E6, E7, L1	6
Honduras	58	E6, L1	6
Panamá	45, 52	E6, E7, L1	5
Colombia	16, 45, 58	E6, E7, L1	4
Bolivia	52	L1	2
Guatemala	58	E7	2

Fuente: esta investigación.

De los 22 países que conforman la región de América Latina, El Salvador, Nicaragua, Haití, República Dominicana, Puerto Rico, Venezuela, Ecuador, Uruguay y la Guyana francesa no aportaron secuencias a la revisión. Además, como se puede apreciar en la Tabla 7, la mayor parte de las secuencias incluidas en esta revisión provinieron de sólo tres países (Costa Rica, Brasil y México) que en conjunto sumaron el 85 % del total (640 secuencias); por su parte, Chile,

Honduras, Panamá, Colombia, Bolivia y Guatemala con menos de diez secuencias en el registro, aportaron menos del 4 % de secuencias (Tabla 7). Si bien existe evidencia sobre una considerable heterogeneidad de estudios que involucran la tipificación del virus para casi todos los países de la zona, la falta de representatividad de secuencias provenientes de algunas naciones de la región constituye una limitante para el conocimiento de la variabilidad genética exhibida por aislamientos de un mismo tipo en diferentes países.

La infección por VPH es considerada la enfermedad de transmisión sexual más frecuente en todo el mundo (Volkow *et al*., 2001; Rivera *et al*., 2006); sin embargo, según un análisis sobre distintos estudios científicos acerca de la frecuencia de la infección y las enfermedades causadas por VPH en Latinoamérica, realizado por la Organización Panamericana de la Salud (PAHO, 2014), se ha estimado que la tasa de incidencia promedio de CCU de la región (29,2 / 100.000 mujeres) es casi el doble del promedio mundial (16,2) (Luciani & Andrus, 2008; PAHO, 2014), las tasas de incidencia estimadas más altas (ASR / 100.000 mujeres) se observan en Haití (87,3), Bolivia (55,0), Paraguay (53,2) y Perú (48,2) y la más baja en Puerto Rico (8,8), Uruguay (18,8) Cuba (20,2) y Costa Rica (21,5); por su parte, Brasil aporta el mayor número de casos (27,3 % del total regional), seguido de México (17,4 %) y Colombia (9,5 %) (Luciani & Andrus, 2008; Parkin *et al*., 2008; PAHO, 2014) (Figura 6).

Sin embargo, estos resultados podrían estar acompañados del subregistro y falta de estudios moleculares que confirmen el tipo viral, tal como lo demuestra el insuficiente o ausente número de publicaciones en muchos países de la región Centro América y el Caribe.

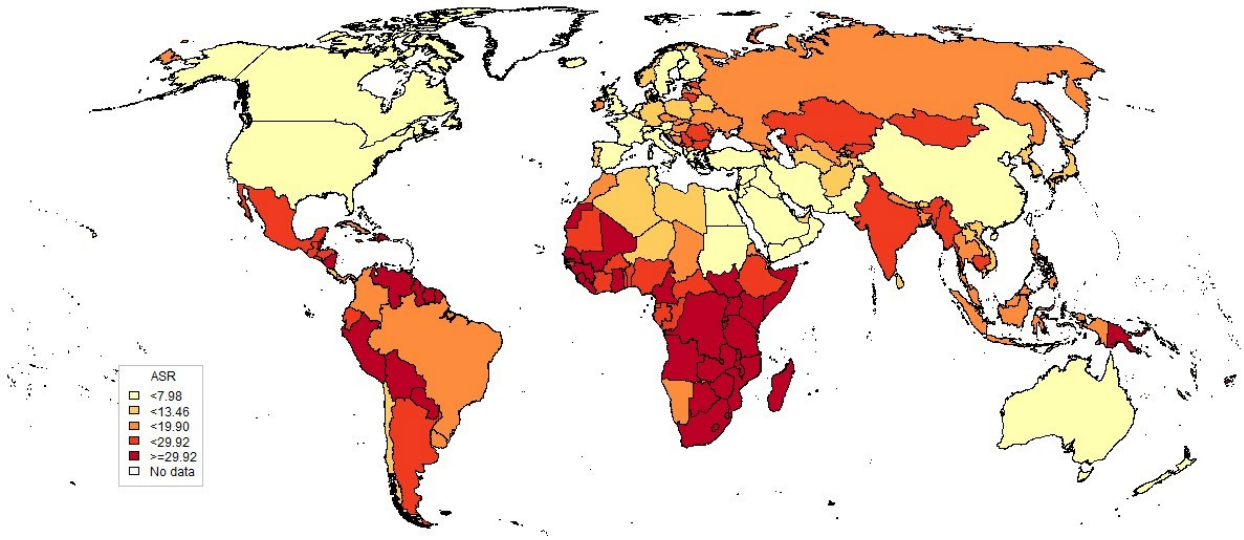


Figura 6. Tasas de incidencia de CCU en el mundo estandarizadas por edad (15-44 años). Tasas de 100.000 mujeres por año. ASR, sigla del inglés, *Age-standardized incidence rates*. Fuente: HPV information centre, 2014 (http://www.hpvcentre.net/link_media/asr_cervical_cancer_world.pdf).

Aun así, la fuerte asociación que se ha establecido entre las infecciones persistentes por ciertos tipos virales y el desarrollo de las lesiones intraepiteliales escamosas (LIE) y cánceres anogenitales (Clifford *et al*., 2003, ACS, 2014; Sehnal *et al*., 2014), la infección por VPH constituye un problema mundial en salud pública, en donde, el costó más alto y significativo

(riesgo de morir) lo pagan las regiones de menor desarrollo social y económico. En Latinoamérica se ha observado un aumento de las tasas de incidencia y mortalidad por CCU en los últimos 30 años; sin embargo, la excepción se presenta en aquellos países que han logrado una transición demográfica con el aumento en la riqueza a raíz de la industrialización, donde los factores demográficos, sociales, económicos y ambientales, así como cambios en el estilo de vida - como los patrones reproductivos –contribuyen en gran medida a este cambio en el comportamiento epidemiológico (Luciani *et al*., 2013, WHO, 2014).

Algunos ejemplos donde se ha visto disminución de la incidencia por CCU, son algunas ciudades de Colombia (Cali, Manizales y Bucaramanga), Cuba, Puerto Rico y Uruguay (Parkin *et al*., 2008; Luciani *et al*., 2013), cuya transición demográfica ha percibido un efecto contrario con relación al aumento de la incidencia del cáncer de mama, similar a los cambios que se produjeron en América del Norte y Europa Occidental en la primera parte del siglo pasado (Luciani *et al*., 2013). En Colombia lo anterior se corrobora con las estadísticas del registro poblacional de cáncer (RCP) y las tasas de incidencia promedio ajustadas por edad de CCU Vs. cáncer de mama, por cada 100.000 mujeres, para las ciudades de Pasto (27,4 Vs. 26,8), Cali (20,5 Vs. 49,2), Manizales (19,5 Vs. 32,2), y Bucaramanga (17,3 Vs. 41,9) (CI5plus, 2014; IACR, 2014).

5.2 Tipos de VPH seleccionados para el estudio

La selección de los diez tipos de VPH involucrados en la revisión, fue concebida con base a la frecuencia en Latinoamérica y la clasificación epidemiológica en cuanto al riesgo de inducir CCU (Figura 7); los genotipos AR (16, 18, 31, 33, 35, 45, 52 y 58) se encuentran asociados al desarrollo de cáncer y los de BR (6 y 11) están asociados con pequeños cambios celulares del epitelio genital, pero no comúnmente a cáncer (Sehnal *et al*., 2014). VPH 16 y 18 se han reportado con factor de alto riesgo a CCU en todas las regiones del mundo (2000; de Sanjosé *et al*., 2007), asociados con aproximadamente el 70 % de todos los casos (Mazarico & Gonzalez-Bosquet, 2012); sin embargo, asociaciones de esta clase para tipos de VPH de menor frecuencia generalmente son limitadas dada la investigación insuficiente en la prevalencia, se necesita conocimiento sobre la incidencia de los diferentes genotipos de VPH para un enfoque adecuado en las medidas preventivas, diagnósticas y terapéuticas.

Como se muestra en la Figura 8, el VPH-AR que aportó el mayor número de secuencias a la revisión fue el genotipo l6 (39,2 %), mientras que los tipos restantes de este grupo en particular no superaron el 12 %; por su parte, VPH 6 y 11 aportaron con el 2,6 y 1,2 % de secuencias para el análisis, respectivamente (Figura 8). A partir del número de secuencias de VPH reportadas para Latinoamérica en esta revisión (Figura 8), se pudo establecer una relación entre la frecuencia de un tipo de VPH con la prevalencia de la infección por un genotipo en particular, de acuerdo con Clifford *et al*., VPH 16 se encuentra dos veces más frecuente que cualquier otro tipo de AR en todas las regiones del mundo, excepto en África subsahariana, donde VPH 35 es igualmente común (Clifford *et al*., 2005).

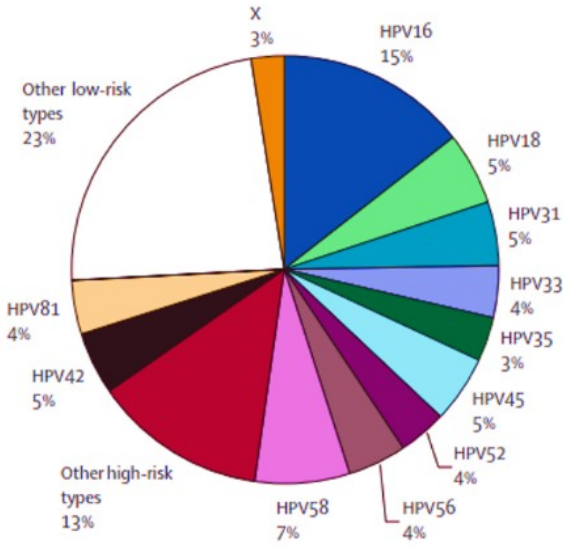


Figura 7. Frecuencia (%) de tipos de VPH en Latinoamérica.
Fuente: Editado de Clifford *et al*., 2005

En las diferentes regiones del mundo, después de VPH 16 y VPH 18, los tipos de AR más frecuentes son: VPH 33 y VPH 56 en Asia, VPH 58, VPH 45 y VPH 31 en América latina, siendo VPH 31 el segundo más frecuente en Europa y África (Figura 7 solo para Latinoamérica) (Clifford *et al*., 2005). Como se observa en la Figura 8, VPH 45 está teniendo un rol epidemiológico importante en la región, lo que lo está haciendo objeto de estudio debido a su origen filogenético en VPH 18 (M. Guerrero, com. pers.). Tiews *et al*., realizaron una correlación entre diferentes VPH-AR junto con diferentes fenotipos de NIC, concluyendo que VPH 45, 18 y 16 pueden constituir un biomarcador que sea útil en la identificación de mujeres con una lesión cervical subyacente con un riesgo elevado de desarrollar lesiones cervicales severas (Tiews *et al*., 2009).

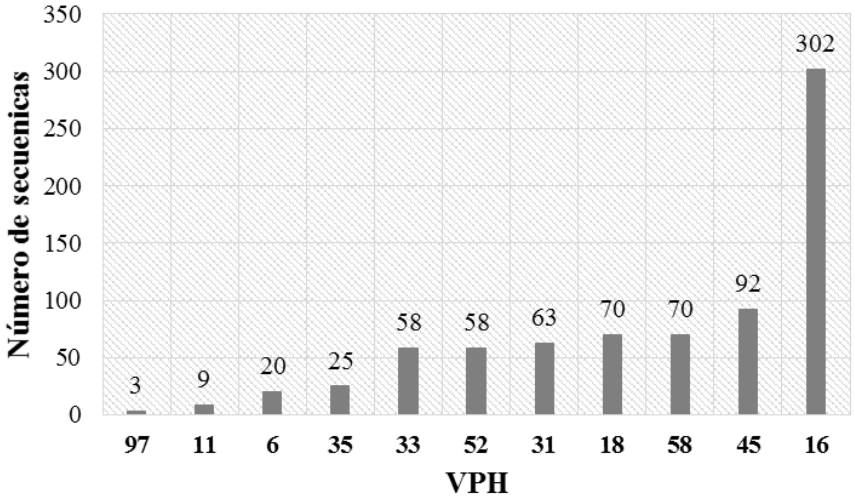


Figura 8. Número de secuencias de genes L1, E6 y E7 por cada genotipo de AR y BR analizados, presentes en la base de datos del GenBank. *Fuente*: esta investigación. *Fuente*: esta investigación.

Según los resultados de este estudio VPH 97 tuvo poca representatividad dentro del grupo (Figura 8), posiblemente debido a su reciente diversificación, lo que lo convierte en un genotipo relativamente nuevo, patogénicamente desconocido, pero que podría estar emergiendo debido a su origen filogenético y la alta distribución y frecuencia mostrada por su antecesor más cercano, el VPH 18 (de Villers, 2013). Este genotipo se caracterizó hasta el 2009 y a la fecha tan solo se conocen tres genomas completos (Chen *et al*., 2009; Landry *et al*., 2012), en 2014, Nicola reportó la presencia de VPH 97 en Colombia, para el departamento de Nariño a partir de la secuenciación de un fragmento de 120 nt, del gen L1 (Nicola, 2014).

Se contabilizaron 770 secuencias ORFs para los genes L1, E6 y E7 procedentes de Latinoamérica, la longitud de los ORFs varió dependiendo del estudio, así por ejemplo, el gen L1 aportó 155 (55,1 %) secuencias completas y 126 (44,9 %) secuencias parciales cuyos fragmentos variaron en una longitud de entre 321 nt a 1596 nt; según lo descrito por De Villers *et al*. (2004) y De Villers, el análisis a partir de pequeños segmentos altamente informativos del gen L1 (291 nt) permite establecer relaciones filogenéticas comparables a las obtenidas mediante análisis de secuencias completas del gen e incluso al análisis de genomas completos (De Villes, 2013). Respecto a los oncogenes virales, el número de secuencias parciales fue menor, en E6 las secuencias parciales representaron el 4,9 % (15) y para E7 el 1,7 % (3), el número total de secuencias obtenidas del GenBank para cada gen se representa en la Figura 9.

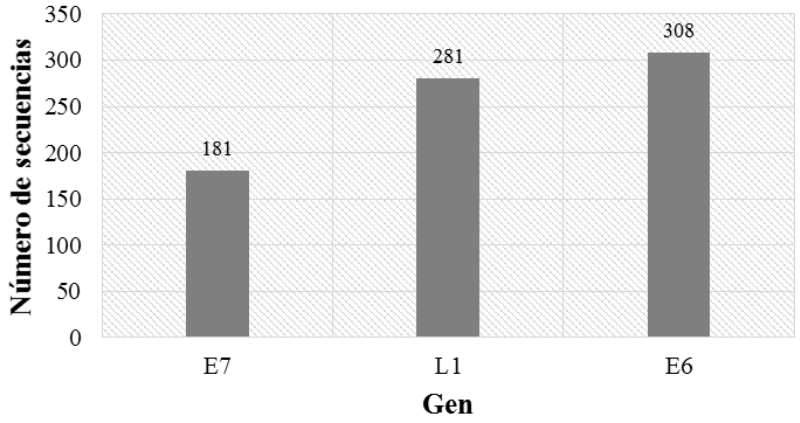


Figura 9. Número total de secuencias para cada gen (L1, E6 y E7) analizado, recuperadas de la base de datos del GenBank. *Fuente*: esta investigación. *Fuente*: esta investigación.

5.3 Relaciones filogenéticas de los VPH-AR y BR más frecuentes en Latinoamérica, mediante análisis de secuencias del gen L1

Se obtuvieron secuencias del gen L1 con longitudes variables dependiendo del tipo, entre 321 nt y 1596 nt; para la construcción de los árboles filogenéticos las secuencias se analizaron en dos grupos: (i) Secuencias de ~1596 - 808 nt, y (ii) secuencias de 667 - 321 nt, los árboles filogenéticos se construyeron utilizando el programa MEGA v. 6.0.5, con el algoritmo ML y el modelo de sustitución nucleotídica de Kimura 2-p, que tiene por objetivo encontrar la estimación

más probable (ML) del número real de eventos mutacionales dado el número de transiciones y transversiones entre cada par de secuencias (Xiong, 2006).

El árbol filogenético construido con las secuencias del grupo (i) se presenta en la Figura 10, construido mediante el alineamiento de 206 ORFs del gen L1 de nueve genotipos distintos de VPH; el alineamiento mostró 1077 sitios variables y 865 sitios potencialmente informativos de parsimonia (sitios conservados que permiten hacer inferencia de las relaciones evolutivas dado el algoritmo), a partir de la topología del árbol se lograron diferenciar dos grandes grupos de especies de PV: Alfa-papilomavirus 7 (Alfa-7) y Alfa-papilomavirus 9 (Alfa-9) con un 59 % de similaridad entre ellos.

El 18,5 % (38/206) de las secuencias pertenecieron a la especie Alfa- 7, de la cual se desprendieron tres ramas, cada uno representando un tipo de VPH diferente: 18, 45 y 97, similar a lo propuesto por De Villers, (2013), donde filogenéticamente VPH 45 y VPH 97 son los tipos virales más estrechamente relacionados con VPH 18 y en conjunto forman un *cluster*, por lo que se cree que comparten un ancestro común más reciente dentro de la especie Alfa-7 (Chen *et al*., 2009).

Dentro de la especie Alfa-7, el genotipo 18 representó la mayor proporción, con un 55,3 % (21/38), secuencias que fueron 13,5 % distantes entre sí, para VPH 45 con el 42,1 % (16/38) de las secuencias, la distancia entre las mismas fue del 10,6 %; de VPH 97 se reportó una única secuencia para Latinoamérica (2,6 %), dada su caracterización reciente, por la cual solo se conecen tres genomas completos procedentes de Costa Rica y Canadá, incluyendo el genotipo de referencia (Chen *et al*., 2009; Landry *et al*., 2012).

La distancia entre los tipos VPH 18 y VPH 45 fue de aproximadamente 19,8 %, la distancia entre el genotipo 97 y los tipos 18 y 45 fue de aproximadamente 18,3 %, dichos porcentajes entre los tipos de la especie Alfa-7, concuerdan con los criterios de clasificación taxonómica propuestos por De Villers *et al*. (2004), en el que un tipo de VPH se diferencia de cualquier otro solo si la variación en el ORF de L1 es mayor al 10 %.

Los genotipos 16, 31, 33, 35, 52 y 58 conformaron el *cluster* de la especie Alfa-9, agrupando el 81,5 % (168/206) de las secuencias analizadas, dentro de Alfa-9 se observó la separación de dos grupos, el primero conformado por los VPH 16, 31 y 35 y el segundo los VPH 52, 33 y 58; el 37,8 % de las secuencias fueron agrupadas dentro del genotipo 16 y mostraron un 90,1 % de similaridad entre sí, el 9,7 % de secuencias correspondientes al VPH 31, fueron 89,8 % similares entre sí, mientras que el 3,9 % de secuencias agrupadas en VPH 35, fueron 95,9 % similares, menos heterogéneas con respecto a las secuencias del resto de tipos.

El 9,3 % de las secuencias agrupadas por VPH 52, fueron 83,6 % similares entre sí, VPH 58 agrupó un total de 35 secuencias (88,5 % similares entre sí), correspondiente al 27 % del agrupamiento, mientras que las secuencias de VPH 33 fueron 89,3 % similares entre ellas. Al analizar el agrupamiento junto con la matriz pareada de similitud (Tabla 8), se pudo establecer una relación más estrecha entre los VPH 58 y VPH 33, dado que la distancia entre estos dos genotipos fue del 15,7 %, en la Tabla 8 se puede observar la distancia pareada entre tipos de VPH de la especie Alfa-9.

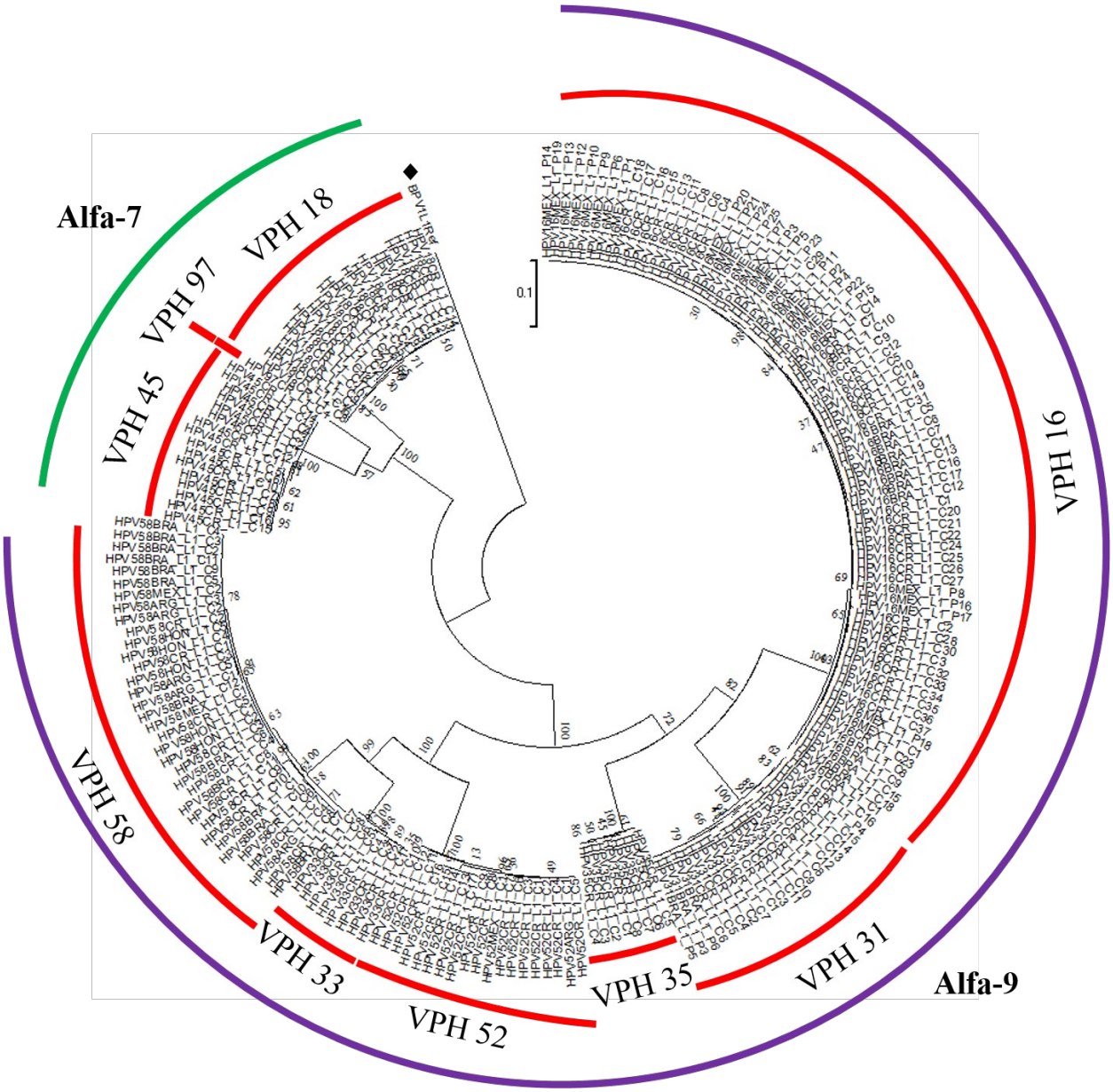


Figura 10. Árbol filogenético construido para nueve tipos de VPH usando el rango de secuencias parciales ~1596 - 808 nt de la región L1 de virus (ORFs) aislados en Latinoamérica, el árbol fue construido con el programa MEGA v. 6.0.5, bajo el algoritmo de Máxima Verosimilitud-ML utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas son representados los valores de *bootstrap* (1000 réplicas) que fueron significativos (por encima de 50 %), la barra de escala indica sustituciones de nucleótidos por sitio y los códigos de las secuencias están indicados a la derecha de los árboles, el símbolo ♦ representa la posición del *outgroup*. *Fuente*: esta investigación.

Tabla 8. Matriz de distancia pareada entre genotipos de la especie Alfa-9 de secuencias parciales (~1596 - 808 nt), obtenida con la herramienta *Compute Between Group Mean Distance* utilizando el algoritmo Kimura 2-p del programa MEGA v. 6.0.5.

	VPH 58	**VPH 16**	**VPH 31**	**VPH 33**	**VPH 35**	**VPH 52**
VPH 58	0					
VPH 16	0,279	0				
VPH 31	0,309	0,248	0			
VPH 33	0,157	0,276	0,291	0		
VPH 35	0,306	0,251	0,255	0,299	0	
VPH 52	0,239	0,304	0,330	0,253	0,288	0

Fuente: esta investigación.

Con el segundo grupo (ii) de secuencias parciales se construyó el árbol filogenético de la Figura 11, para lo cual previamente se realizó un alineamiento múltiple con 75 ORFs del gen L1 de diez genotipos distintos de VPH, la secuencia de BPV 1 constituyó el grupo externo, el alineamiento mostró 294 sitios variables y 226 sitios potencialmente informativos de parsimonia; a partir de la topología del árbol, se identificaron tres grandes *clusters*: Alfa-papilomavirus 6 (Alfa-6), Alfa-papilomavirus 7 (Alfa-7) y Alfa-papilomavirus 9 (Alfa-9), la matriz de distancia pareada indicó que el menor porcentaje de similaridad se presentó entre las especies Alfa-6 y 7 (54,5 %).

La especie Alfa-6 representada por los tipos 6 y 11, aportó el 5,3 % (4/75) de las secuencias al alineamiento, dentro de VPH 6, las secuencias mostraron un porcentaje de similaridad del 86,3 %, mientras que para las VPH 11 fue del 88,4 %. El 9,3 % (7/75) de las secuencias pertenecieron a la especie Alfa- 7, que agruparon a VPH 18 y 45, tal como lo reportado por De Villers (2013), el cálculo de la matriz de distancia pareada arrojó un 81,7 % de similaridad entre estos dos genotipos, similar al porcentaje encontrado para el análisis de las secuencias del grupo (i) (rango 808 -1596 nt).

Los genotipos 16, 31, 33, 35, 52 y 58 conformaron el clado de la especie Alfa-9 con el 85,3 % (64/75) del total de las secuencias alineadas; el 56 % (42/75) de las secuencias se agruparon dentro del genotipo 16, cuyas secuencias fueron 99,1 % similares entre sí, el 13,3 % (10/75) de las secuencias se agruparon en VPH 33, similares entre sí un 98,9 %. El único grupo de secuencias cuyo porcentaje de similaridad fue del 100 %, fue VPH 31, en la tabla 9 se puede visualizar la distancia pareada entre tipos de VPH de la especie Alfa-9, donde se observa que los tipos más relacionados fueron los VPH 52 y 58.

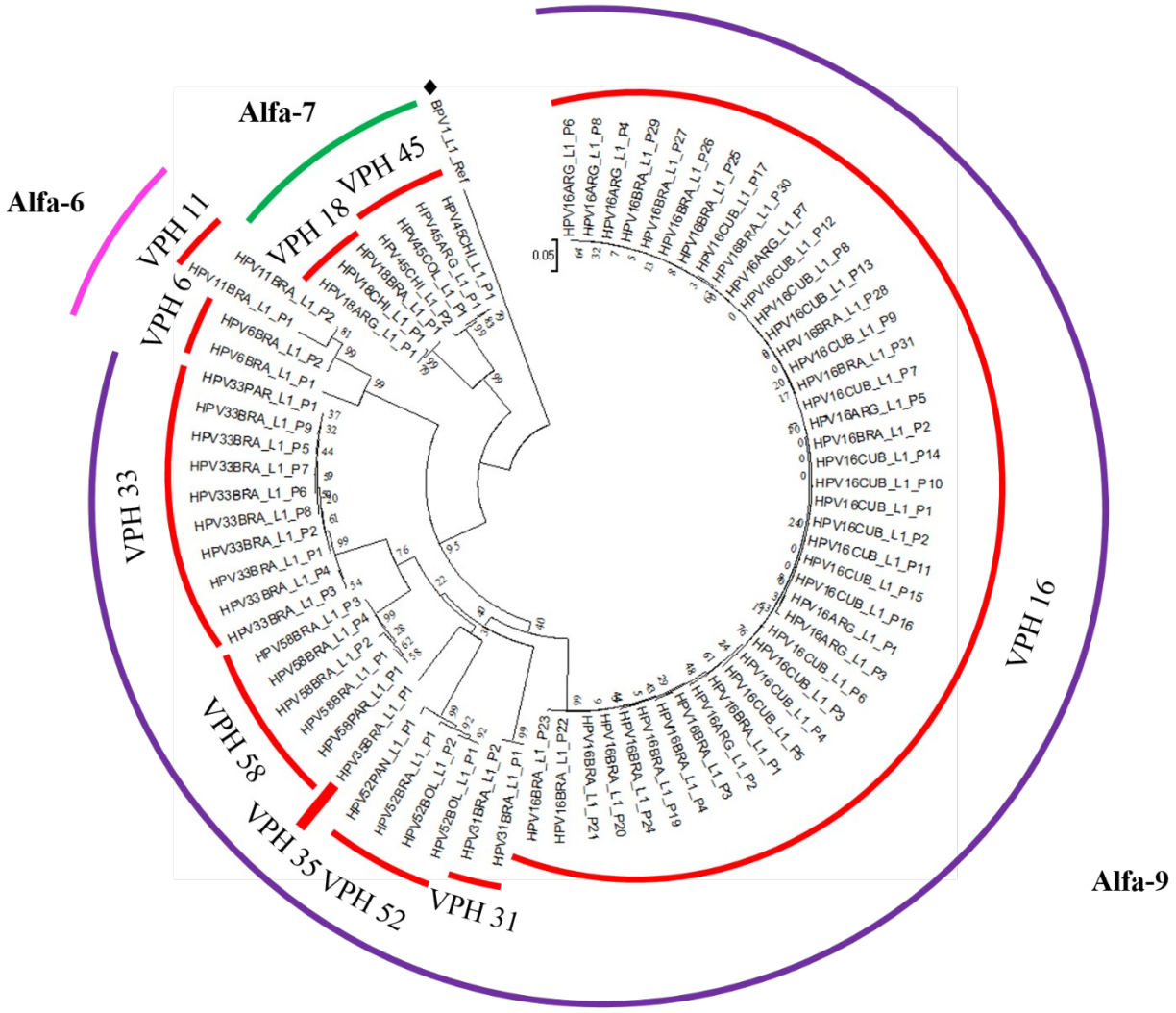


Figura 11. Árbol filogenético construido para diez tipos de VPH usando el rango de secuencias parciales 667 - 321 nt de la región L1 de virus (ORFs) aislados en Latinoamérica, el árbol fue construido con el programa MEGA v. 6.0.5, bajo el algoritmo de Máxima Verosimilitud-ML utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas son representados los valores de *bootstrap* (1000 réplicas) que fueron significativos (por encima de 50 %), la barra de escala indica sustituciones de nucleótidos por sitio y los códigos de las secuencias están indicados a la derecha de los árboles, el símbolo ♦ representa la posición del *outgroup*. *Fuente*: esta investigación.

Tabla 9. Matriz de distancia pareada entre genotipos de la especie Alfa-9 de secuencias parciales (667 nt -321 nt.), obtenida con la herramienta *Compute Between Group Mean Distance* utilizando el algoritmo Kimura 2-p del programa MEGA v. 6.0.5

	VPH 52	VPH 16	VPH 33	VPH 58	VPH 35	VPH 31
VPH 52	0					
VPH 16	0,329	0				
VPH 33	0,296	0,291	0			
VPH 58	0,242	0,268	0,166	0		
VPH 35	0,266	0,286	0,266	0,253	0	
VPH 31	0,344	0,308	0,275	0,302	0,265	0

Fuente: esta investigación.

Al comparar las tablas de matrices de distancia pareada (Tabla 8 y 9), no se encontraron diferencias respecto a los resultados de agrupamiento, esto concuerda con el planteamiento de De Villers, (2013) en el sentido de que el análisis filogenético con secuencias de L1 de VPH, fragmentos o completas, no difiere significativamente y por lo tanto, en el análisis filogenético se pueden utilizar secuencias de longitudes de hasta menos del 20 % del gen L1, estos resultados son comparables a los obtenidos por Nicola (2014), donde se analizaron fragmentos de 120-130 nt del gen L1 y se obtuvieron conclusiones similares.

Una característica notable del genoma de PV, radica en la estabilidad global de los ORFs debido a la alta fidelidad en la replicación del genoma viral por la usurpación de las polimerasas anfitrionas de corrección celular, aunado a la evolución clonal a través de las mutaciones, la selección o deriva génica (Narechania *et al*., 2005) dicha estabilidad, ha sugerido un modelo de evolución monofilético que ha mantenido rasgos como la oncogenicidad y el tropismo celular (Van Ranst *et al*., 1992; Narechania *et al*., 2005) y puede ser observada en las figuras 10 y 11, donde los árboles filogenéticos resultantes clasificaron con precisión los diferentes tipos de VPH de acuerdo con el potencial oncogénico (AR y BR).

Sin embargo, se ha demostrado que dicha monofilia se rompe cuando las filogenias de genes tempranos (L1, L2) y tardíos (E5, E6, E7) son comparadas cualitativamente (Bravo & Alonso, 2004; de Villers *et al*., 2004); lo cual se debe a que la tasa de divergencia de proteínas es más elevada en las proteínas tempranas que en las proteínas tardías, y aumenta en el orden de: L1 <L2 <E6 E7 ≈ <E5 (Bravo & Alonso, 2004), por ende, la razón por la cual el enfoque de la mayoría de los estudios de clasificación de PV, ha hecho uso únicamente del gen L1 como el estándar para el análisis de evolución de: géneros, especies, tipos y subtipos, se debe a que constituye una región altamente conservada en estructura y función (Narechania *et al*., 2005).

Sobre la base de similitud filogenética, los Alfa-virus se pueden clasificar en 15 especies o clados (de Villers *et al*., 2004), los filogramas construidos anteriormente a partir de los tipos de AR (16, 18, 31, 33, 35, 45, 52, 58) y BR (6 y11) más frecuentes en Latinoamérica, delimitaron únicamente tres de estas especies (Alfa-6, Alfa-7 y Alfa-9), el grado en que estas agrupación

filogenéticas (especies) comparten una función biológica, no está del todo claro, por ejemplo, la forma en que las especies difieren por el tropismo de un tejido específico, como Alfa-6 que puede encontraste con mayor frecuencia en el tejido del tracto genital, pero también en el de las vías respiratorias o Alfa-7 y 9 por su afinidad con el epitelio del cuello uterino (Clifford & Franceschi, 2008).

De acuerdo a los manifiestos epidemiológicos, la prevalencia de los tipos individuales en la población se correlaciona con la persistencia viral, a su vez, la persistencia y la progresión a cáncer están altamente asociados; los tipos de VPH más cancerígenos se encuentran concentrados en las especies Alfa-9 y Alfa-7, y se distinguen de otras especies, por el riesgo elevado de progresión a cáncer debido a la persistencia (Schiffman *et al*., 2005; Clifford & Franceschi, 2008).

Respecto a los agrupamientos anteriores, con las secuencias previamente reportadas por Nícola (2014), provenientes de mujeres con lesiones en CCU y ASCUS, se construyó un árbol filogenético (Figura 12) utilizando secuencias de DNA de aproximadamente 120 nt del gen L1 de VPH, los resultados de este análisis mostraron mayor frecuencia del genotipo viral 16 (5/13), seguido de VPH 6 y 58, los genotipos restantes correspondieron a VPH 33, 73, 97 y 45 para un total de siete genotipos diferenciados.

La frecuencia de distribución de tipos reportada por Nicola (2014), se relaciona con los informes publicados para la población general de Colombia, donde la distribución de tipos específicos señala que VPH 16 es el genotipo más frecuente, siendo esto similar a los datos reportados en todo el mundo, a VPH 16 se le suman los genotipos 31, 58, 18, 45 y 33 de acuerdo a la frecuencia exhibida; sin embargo, esta frecuencia puede cambiar dependido de la región, el tipo de estudios y el método de detección (Castellsagué *et al*., 1997; Muñoz *et al*., 2004; Muñoz *et al*., 2013).

La frecuencia de los tipos VPH 73 y 97 no había sido reportada antes para Colombia. Dada la procedencia de los aislamientos asociados positivamente con el cáncer (NIC) a nivel mundial han sido clasificados dentro de los tipos de AR; sin embargo, como los datos proceden de un pequeño número de casos, carecen de evidencia concluyente, por ende, ponen de antemano la necesidad de aumentar las investigaciones y establecer sistemas de vigilancia que permitan conocer la prevalencia de estas infecciones emergentes a lo largo del tiempo en una población.

La identificación de nuevos genotipos (genotipos emergentes) y el remplazamiento de los tipos oncogénicos más frecuentes (VPH 16, VPH 18) por genotipos de menor frecuencia (VPH 45, VPH 58, VPH 73, VPH 97) es un evento bastante común en muchas partes del mundo, que podría estar relacionado con los esquemas de vacunación actuales, por ejemplo, en Latinoamérica las vacunas están diseñadas para prevenir los dos tipos oncogénicos más frecuentes, VPH 16 y VPH 18 (Murillo *et al*., 2008), estas vacunas podrían modular positivamente la prevalencia de infecciones de diferentes genotipos a través de la sustitución del genotipo viral vacunal, lo que podría conducir a un aumento en la prevalencia de VPH-AR no vacunales, dado que los VPH son virus de DNA altamente estables, las presiones selectivas de la vacunación pueden desocupar nichos ecológicos ocupados actualmente por los VPH 16 y 18, este fenómeno es conocido como el reemplazamiento cepa patógena inducida por vacuna y fue descrita previamente en un programa de vacunación

contra *Streptococcus pneumoniae*, dirigido contra siete serotipos, que resultó en el aumento de otros serotipos (Peralta *et al*., 2014).

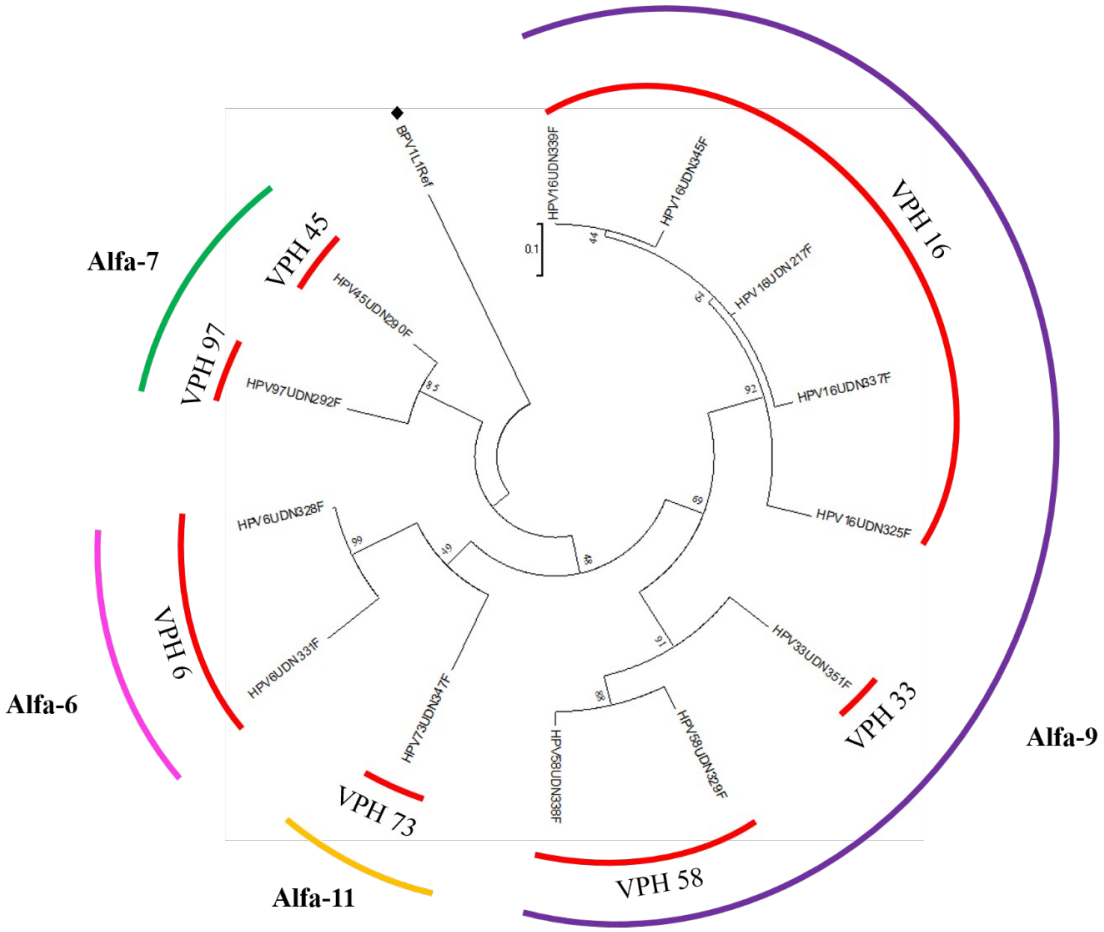


Figura 12. Árbol filogenético construido para siete tipos de VPH usando un fragmento de 120 nt del gen L1 (ORFs), el árbol fue construido con el programa MEGA v. 6.0.5, bajo el algoritmo de ML utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas son representados los valores de *bootstrap* (1000 réplicas) que fueron significativos (por encima de 50 %), la barra de escala indica sustituciones de nucleótidos por sitio, y los códigos de las secuencias están indicados a la derecha de los árboles, el símbolo ♦ representa la posición del *outgroup*. *Fuente*: esta investigación.

En conclusión, dado que el gen L1 de los VPH analizados se encuentra altamente conservado, el análisis de grupos de secuencias de diferente tamaño de fragmento para este gen no mostró diferencias significativas para definir los genotipos virales y establecer criterios taxonómicos confiables, sin embargo, para precisar en las relaciones histórico-evolutivas y para construir filogenias mucho más informativas, se deben utilizar fragmentos con el mayor número de sitios conservados o de una longitud similar, debido a que el análisis de agrupamiento generado

para todo los tamaños de fragmento combinados, generó clados no resueltos para los fragmento inferiores a 808 nt (datos no mostrados).

5.4 Análisis de variantes de tipos virales haciendo uso de los oncogenes E6 y E7

El análisis de la variabilidad genética ha demostrado ser esencial para una mejor comprensión de la historia natural y la evolución del VPH (Castellsagué, 2008); las variantes moleculares de VPH-AR y BR han sido estudiadas de amplia forma en Asia y Europa, pero muy poco en Latinoamérica, por lo que la revisión que se presenta a continuación involucra el análisis de variantes moleculares para los genotipos de AR y BR más frecuentes en esta región, con el fin de proporcionar información acerca del significado biológico de la variabilidad intratipo de los genes E6 y E7, que en última instancia podría ser utilizada en futuros estudios para el desarrollo de ensayos epidemiológicos de vacunación, funcionales y moleculares en el manejo y control de estas infecciones.

Teniendo en cuenta la extensa información relacionada por cada tipo y sus variantes moleculares, a continuación se resumen las principales mutaciones puntuales identificadas para los genes E6 y E7 y su relevancia para este estudio (Tabla 10), dado que constituyen los principales resultados de esta investigación, de gran interés para definir las alteraciones que pueden interferir con las propiedades funcionales o antigénicas en estas oncoproteínas, cualquier cambio a nivel de secuencia de nucleótidos en estas proteínas puede alterar su expresión y potencial oncogénico de VPH, aumentando la probabilidad de aparición, desarrollo y mantenimiento de malignidad en las células epiteliales del cuello uterino (Liu *et al*., 2006; Liu & Baleja, 2008).

De forma general para ambos genes, se identificaron 25 nuevas mutaciones puntuales a lo largo de región latinoamericana (Tabla 10), el 56 % de las mutaciones puntuales (14/25) constituyeron sustituciones aminoacídicas sobre la cadena polipeptídica, mientras que el 44 % (11/25) restante conservó el aminoácido original. Si bien el número de nuevas mutaciones puntuales, en su mayoría, ocurren no más de una, son de gran interés para definir las alteraciones que pueden interferir con las propiedades funcionales o antigénicas específicas de las proteínas, que con el tiempo, pueden establecerse en una población si hay algún mecanismo que selecciona positivamente para la variación o, si son funcionalmente neutrales, por la expansión selectiva de la población hospedera infectada con la variante que presenta dicha mutación (deriva genética) (Villa *et al*., 2000; Bernard *et al*., 2006).

Las mutaciones puntuales de la Tabla 10, descritas previamente por autores, fueron seleccionadas para el resumen debido a la función ejercida sobre el episoma viral; las mutaciones en E6 de AR sobre los codones que configuran los polipéptidos CXXC, son importantes porque de su estabilidad depende la modulación (positiva o negativa) de la activación transcripcional, la transformación, la inmortalización y la asociación con las proteínas celulares hospedadoras (p53, p73); las mutaciones sobre el dominio PDZ (~ aa 149-152) pueden favorecer la activación de la telomerasa y la síntesis desmedida de secuencias repetidas de telómeros, actividad característica de la mayoría de las líneas celulares y tumores, debido a la degradación de las proteínas de señalización celular y supresoras tumorales: hDlg1, hDlg4, hScrib, MUPP1, PTPN3, SAP97, MAGI 1, 2 y 3, (Boulet *et al*., 2007; Vande Pol & Klingelhutz, 2013).

Tabla 10. Mutaciones más importantes identificadas para las variantes de VPH analizados.

Tipo VPH	Gen	Mutación	Importancia	Referencias
VPH 6	*E6*	A397G	Descrita por primera vez para Latinoamérica. Mutación no sinónima (D99G).	Esta investigación.
VPH 11	*E6*	A538G	Descrita por primera vez para Latinoamérica. Mutación no sinónima (E146G).	Esta investigación.
VPH 18	*E6*	C287G	Descrita por primera vez para Brasil.	Esta investigación.
			Biomarcador de transmisión viral.	Chang *et al*., 2005
	E7	A864G	Mutación no sinónima (N92/S/K).	
			Puede modular positivamente la interacción con proteínas supresoras de tumores.	Todorovic *et al.*, 2012
VPH 45	*E6*	A124C, 259T/A, A284C, T482C, G487A y A497G	Descritas por primera vez para Brasil.	Esta investigación.
		A124C	Mutación no sinónima (K8T).	Esta investigación.
		G259T/A	Mutación no sinónima (C53F/Y).	Esta investigación.
		G487A	Mutación no sinónima (S129N).	Esta investigación.
		T418G	Mutación no sinónima (L106R) sobre el tercer motivo conservado CXXC, puede implicar perdida de la integridad y no degradación de p53.	Crook *et al*., 1991; Foster *et al*., 1994; Howie *et al*., 2009
	E7	G600A, A603C, A718C, G808T y G822T	Descritas por primera vez para Brasil.	Esta investigación.
		G600A	Mutación no sinónima (R5Q/K) sobre la región RC1, implicada en la estabilidad del dominio.	Brokaw *et al*., 1994
		A603C	Mutación no sinónima (E6A) sobre la región RC1, que puede inhibir la asociación estable con pRb.	Brokaw *et al*., 1994
		G808T	Mutación no sinónima (E74D).	
		G822T	Mutación no sinónima (S79I).	

Tipo VPH	Gen	Mutación	Importancia	Referencias
VPH 16	*E6*	T350G	Mutación no sinónima (L83V) observada en todo el mundo, asociada a la persistencia del virus y mayor riesgo de neoplasia cervical.	Villa *et al*., 2000
		G176A, T310G, C335T	Mutaciones no sinónimas (D25E, F69L, y H78Y respectivamente) que modulan el potencial oncogénico transformador de VPH en lesiones escamosas glandulares.	Villa *et al*., 2000
		T290A	Mutación no sinónima (C63S), modifica el segundo motivo conservado CXXC manteniendo el potencial interactivo con p53.	Park & Androphy, 2002
	E7	A645G, A647C/G	Mutaciones no sinónimas (L28S, N29P/S respectivamente), mantienen asociación estable con pRb, sin embargo, limitan su posterior degradación.	Giarrè *et al*., 2001
		G649T	Mutación no sinónima (D30F), sobre el dominio CR2 dentro del sitio de fosforilación de pRb.	
VPH 31	*E6*	C285T, A475G	Mutaciones no sinónimas (H60Y, K123R respectivamente) relacionadas con variaciones en sitios de unión o degradación de p53, afecta directamente la persistencia de la enfermedad.	Foster *et al*., 1994
		C520T, G523A	Mutaciones no sinónimas (A138V, C139Y respectivamente), modifican el cuarto motivo conservado CXXC, expresión de E6 defectuosas para la unión y degradación de p53.	Foster *et al*., 1994
		C537G	Mutación no sinónima (R144g), puede abolir la degradación de MAGI 1 y SAP97 (proteínas supresoras tumorales).	Zhang *et al*., 2007
	E7	C626T, G695A	Mutaciones no sinónimas (H23Y, E46K respectivamente) modifican el sitio de unión de pRb al dominio CR2, afectando la función inmortalizadora de E7.	Barbosa *et al*., 1990
		A743G	Mutación no sinónima (K62E/N) que altera la adaptabilidad y la patogenicidad del genotipo.	Liu *et al*., 2014

Tipo VPH	Gen	Mutación	Importancia	Referencias
VPH 33	*E6*	A449T	Descritas por primera vez para Latinoamérica.	Esta investigación.
			Mutación no sinónima (E114V).	
		C521T	Mutación no sinónima (A138V), modifica el cuarto motivo conservado CXXC.	Foster *et al*., 1994
VPH 35	*E6*	G543T	Mutación no sinónima (R145I), puede abolir la degradación de MAGI 1 y SAP97.	Zhang *et al*., 2007
	E7	T628C	Mutación no sinónima (Y23H), modifica el motivo funcional LXCXE, sitio de alta afinidad a pRb.	Liu *et al*., 2006
VPH 52	*E6*	A143C, C348G	Mutaciones no sinónimas (E14D, L83V respectivamente) correlacionada estadísticamente con la persistencia de NIC II y III.	Aho *et al*., 2003; Liu *et al*., 2014
	E7	A681G	Descrita por primera vez para Latinoamérica.	Esta investigación.
		C672T	Mutación no sinónima (T37I) que modifica el sitio de fosforilación de pRb.	
		C733T	Mutación no sinónima (H61Y) que modifica el segundo motivo conservado CXXC.	Park & Androphy, 2002
VPH 58	*E6*	A136G, T143C, G193A y A260C	Descritas por primera vez para Latinoamérica.	Esta investigación.
	E7	C632T, G760C/A	Mutaciones no sinónimas (T20I, G63H/D/S respectivamente), correlacionada estadísticamente con la persistencia de NIC II y III.	Chan *et al*., 2002
		A587G, A603T A614G y T778C	Descritas por primera vez para Latinoamérica.	Esta investigación.
		A587G, A614G	Mutaciones no sinónimas (N5S, D15G respectivamente), modifican el dominio CR1 en el sitio de unión de pRB.	
		G694A	Mutación no sinónima (G41R) dentro del sitio de fosforilación de pRb.	

Fuente: esta investigación.

Mutaciones en E7 de AR sobre los dominios homólogos: CR1, CR2 y CR3, pueden modular (positiva o negativamente) la interacción con una amplia gama de proteínas celulares hospedadoras relacionadas con el crecimiento y la transformación celular (pRb, p107 y p130), la apoptosis, la transcripción de genes y la síntesis del DNA viral, en especial mutaciones en CR2 que alojan una región (LXCXE) requerida para asociación con el supresor tumoral pRb (Boulet *et al*., 2007; Todorovic *et al*., 2012). En contraste, las funciones fisiológicas de las proteínas E6 y E7 de BR siguen sin estar claros, debido a los escasos estudios relacionados (Liu & Baleja, 2008).

En el contexto completo del ciclo infeccioso viral, la función de la mayoría de las mutaciones en los oncogenes E6 y E7 sigue siendo desconocida, debido a que muchas de las mutaciones que se indujeron para asignar funciones en estos genes, lo más probables es que interrumpieran la estructura general de las oncoproteínas, más que los sitios de interacciones específicos proteína-proteína, lo que fácilmente pudiera modificar sus efectos sobre la transformación e inmortalización celular; sin embargo, éstos resultados evidencian la dinámica evolutiva del VPH y sugieren un estudio más intenso para identificar los procesos moleculares que involucran al hospedador, con el fin de entender la oncogénesis inducida por los VPH, lo que podría proporcionar una base para el desarrollo de novedosos compuestos antivirales y anticancerosos para los tipos presentes en Latinoamérica.

El resumen de la variabilidad genética exhibida por los genes E6 y E7 de cada tipo de VPH de AR y BR más frecuentes en Latinoamérica se presentan en la Tabla 11.

5.4.1 Variantes moleculares de tipos virales de bajo riesgo

Todas las secuencias de VPH 6 y VPH 11, provinieron de Brasil; las variantes detectadas de VPH 6, se agruparon en tres sublinajes de acuerdo al subtipo de mayor similaridad: A, B1 y B3 (Figura 13), cuyo patrón de agrupamiento y frecuencia de variantes por linaje, ha sido reportado de forma similar por Kocjan *et al*. (2009) y Kocjan *et al*. (2011). La distancia entre grupos de linajes fue de aproximadamente 0,87 %; a su vez, la variante del grupo A compartió con la secuencia de referencia de VPH 6A (No. Acceso X00203) 98 % de similaridad, las variantes de B3 fueron 99,6 % similares y las variantes de B1 fueron 99,7 % similares entre sí. Las mutaciones puntuales A221T y G473A permitieron diferenciar la variante del grupo A, la mutación C479T únicamente se presentó en B3 y las sustituciones que permitieron la agrupación de variantes en B1 fueron C251G, A365T, C392C y la mutación múltiple A323C/T (Anexo C), reportada previamente en África y Europa por Kocjan *et al*. (2009) y Kocjan *et al*. (2011).

Tabla 11. Variabilidad genética de los genes E6 y E7 de los VPH-AR y BR analizados.

Tipo VPH	Gen	Número de secuencias analizadas(ORFs)[I]	Número de variantes identificadas[II]	Número de mutaciones puntuales[III]		Número de mutaciones NO sinónimas[IV]	Grupos filogenéticos establecidos[V]
				Trn*	Trv**		
VPH 6	E6[a]	18	18/18 (100%)	4/9 (44,4%)	5/9 (55,6%)	2/9 (22,2%)	A (1/18), B1 (12/18), B3 (5/18)[p]
VPH 11	E6[b]	7	4/7 (57,1%)	3/3 (100%)	-	1/3 (33,3%)	A1 (4/7), A2 (3/7)[q]
VPH 18	E6[c]	25	25/25 (100%)	12/15 (80%)	3/15 (20%)	4/15 (26,7%)	A1 (5/25), A2 (12/25), B (8/25)[r]
	E7[d]	21	14/21 (66,7%)	10/11 (90,9%)	1/11 (9,1%)	4/11 (36,4%)	A1 (8/21), A2 (6/21), B (7/21)[s]
VPH 45	E6[e]	36	34/36 (94,4%)	7/16 (43,75%)	9/16 (56,25%)	8/16 (50%)	A1 (2/36), A2 (2/36), A3 (5/36), A4 (10/36), B (15/36)
	E7[f]	36	34/36 (94,4%)	2/9 (22,2%)	7/9 (77,8%)	6/9 (66,7%)	
VPH 16	E6[g]	140	120/140 (85,7%)	24/44 (54,5%)	20/44 (45,5%)	28/44 (63,6%)	E (44/140), AA (32/140), NA1 (10/140), Af-2 (6/140), As (1/140),
Tipo VPH	**Gen**	**Número de**	**Número de**	**Número de mutaciones**		**Número de**	**Grupos**

		secuencias analizadas(ORFs)[I]	variantes identificadas[II]	puntuales[III]		mutaciones NO sinónimas[IV]	filogenéticos establecidos[V]
				Trn*	Trv**		
VPH 16	E7[h]	43	24/43 (55,9%)	8/13 (61,5%)	5/13 (38,5%)	3/13 (23,1%)	A (21/43), B (21/43), C (1/43)[p]
VPH 31	E6[i]	22	17/22 (77,3%)	15/20 (75%)	5/20 (25%)	10/20 (50%)	A (9/22), B (5/22), C (8/22)[q]
	E7[j]	19	19/19 (100%)	15/16 (93,75%)	1/16 (6,25%)	9/16 (56,25%)	A (9/19), B (5/19), C (5/19)[r]
VPH 33	E6[k]	20	10/20 (50%)	4/12 (33,3%)	8/12 (66,7%)	9/12 (75%)	A (12/20), B (6/20), C (2/20)[s]
	E7[k]	20	13/20 (65%)	3/6 (50%)	3/6 (50%)	3/6 (50%)	
VPH 35	E6[l]	8	8/8 (100%)	7/10 (70%)	3/10 (30%)	4/10 (40%)	A (4/8), B (2/8), C (2/8)[t]
	E7[l]	8	2/8 (25%)	2/2 (100%)	-	1/2 (50%)	
VPH 52	E6[m]	18	14/18 (77,8%)	9/17 (53%)	8/17 (47%)	5/17 (27,8%)	A1 (6/18), A2 (2/18), B (2/18), C (5/18), D (3/18)[v]
	E7[n]	17	10/17 (58,8%)	9/13 (69,2%)	4/13 (30,8%)	7/13 (53,8%)	A (7/17), B (2/17), C (5/17), D (3/17)[w]
Tipo VPH	**Gen**	**Número de secuencias**	**Número de variantes**	**Número de mutaciones puntuales[III]**		**Número de mutaciones**	**Grupos filogenéticos**

		analizadas(ORFs)[I]	identificadas[II]	Trn*	Trv**	NO sinónimas[IV]	establecidos[V]
VPH 58	E6[ñ]	14	14/14 (100%)	8/10 (80%)	2/10 (20%)	3/10 (20%)	A1 (5/14), A2 (2/14), A3 (3/14), B (4/14)[x]
	E7[o]	16	16/16 (100%)	12/16 (75%)	4/16 (25%)	10/16 (62,5%)	A1 (2/16), A2 (3/16), A3 (4/16), B1 (3/16), B2 (4/16)[y]

Los superíndices **a** (Anexo C), **b** (Anexo D), **c** (Anexo E), **d** (Anexo F), **e** (Anexo G), **f** (Anexo H), **g** (Anexo I), **h** (Anexo J), **i** (Anexo K), **j** (Anexo L), **k** (Anexo M), **l** (Anexo N), **m** (Anexo Ñ), **n** (Anexo O), **ñ** (Anexo P) y **o** (Anexo Q), simbolizan la variabilidad genética identificada para cada uno de los genes y tipo de VPH analizados, la cual se encuentra representada en cada uno de los anexos señalados.
I Número de secuencias descargadas del GenBank.
II Las variantes fueron identificadas al comparar residuo a residuo, cada una de las secuencias del gen E6 o E7, con la secuencia E6 o E7 del genoma de referencia de cada VPH analizado.
III Las mutaciones puntuales fueron identificadas a partir del alineamiento múltiple de secuencias. ***Trn:** mutación puntual de tipo transición, ****Trv:** mutación puntual de tipo transversión.
IV Las mutaciones No sinónimas se identificaron a partir de la traducción *in silico* de los alineamientos múltiples.
V Los grupos filogenéticos fueron establecidos a partir de la variabilidad nucleotídica exhibida por cada grupo y la topología de los árboles filogenéticos representados en las figuras que indican los superíndices: **p** (Anexo R), **q** (Anexo S), **r** (Anexo T), **s** (Anexo V), **t** (Anexo W), **v** (Anexo X), **w** (Anexo Y), **x** (Anexo Z), **y** (Anexo AA).
Fuente: esta investigación.

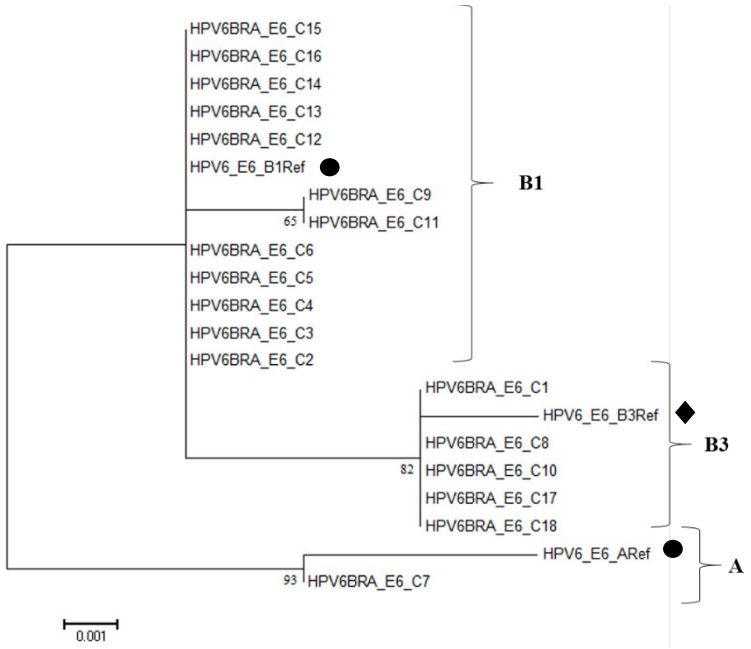


Figura 13. Árbol filogenético que representa las relaciones entre las variantes de E6 de VPH 6, construido con el programa MEGA v. 6.0.5, bajo el criterio de Máxima Verosimilitud-ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p; en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %). El símbolo ♦ designa la secuencia del VPH 6A (No. Acceso X00203). El símbolo ● indica secuencias de los sublinajes procedentes de lesiones malignas: VPH 6B1 (No. Acceso AF092932) y VPH 6B3 (No. Acceso L41216), utilizados en el agrupamiento debido a que exhiben la variabilidad genética necesaria para establecer la clasificación de variantes genómicas del VPH 6. *Fuente*: esta investigación.

A partir de las variaciones en las secuencias de VPH 11 (Anexo D), el análisis de agrupamiento indicó dos grupos filogenéticos designados como A1 y A2 (Figura 14), similar a la nomenclatura propuesta por Danielewski *et al*. (2011) y Burk *et al*. (2011) comparando genomas completos del genotipo 11 procedentes de diferentes partes del mundo. La distancia genética entre grupos fue de aproximadamente 0,5 %; las variantes del grupo A1 compartieron 99,9 % de similaridad, las variantes del grupo A2 fueron 100 % similares; partiendo del análisis de similaridad, así como del limitado número de posiciones nucleotídicas que mutaron a lo largo de las secuencias E6, concluimos que para este genotipo dichas secuencias se encuentran altamente conservadas, similar a lo establecido por Burk *et al*., para genomas completos descritos hasta 2011. El 25 % de las secuencias agrupadas en A1 constituyeron una variante, diferenciadas por la mutación puntual A538G, mientras que las sustituciones nucleotídicas T137C y C380T permitieron la agrupación de variantes en A2 (Anexo D).

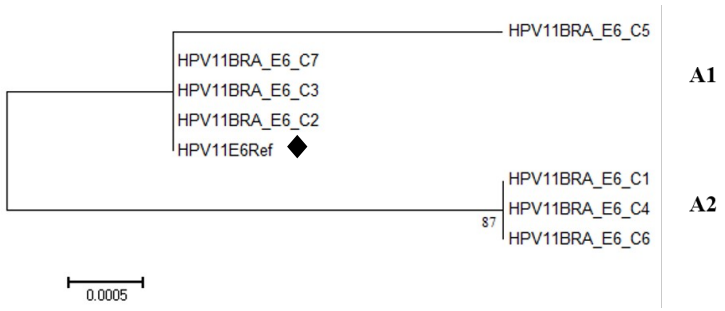


Figura 14. Árbol filogenético que representa las relaciones entre las variantes de E6 de VPH 11, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p; en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), A1 y A2 señalan los grupos de variantes identificadas. El símbolo ♦ designa la secuencia de referencia del VPH 11 (No. Acceso M14119.1). *Fuente*: esta investigación.

5.4.2 Variantes moleculares de tipos virales de alto riesgo

VPH 18 y VPH 45 comparten la misma rama filogenética en el árbol de PVs, conformando el segundo grupo de tipos más importantes en cuanto a carcinogenicidad (de Villers *et al*., 2013). Las secuencias E6 de VPH 18 provinieron de Brasil (4) y Costa Rica (21), las secuencias de E7 únicamente de Costa Rica (21); a partir del análisis de agrupamiento realizado para las variantes de cada gen, se establecieron tres grupos filogenéticos: A1, A2 y B (Figura 15 y 16), distantes entre sí 1,5 y 1,15 % respectivamente, similar al agrupamiento obtenido por Chan *et al*. (2009) y Chan *et al*. (2013); las variantes agrupadas en A1 compartieron 99,8 % de similaridad, las variantes de A2 fueron 100 % para E6 y 99,6 % similares para E7, por su parte, las secuencias del grupo B, constituyeron el grupo más heterogéneo de variantes, siendo 0,5 % diferentes entre sí.

La distribución y frecuencia de grupos de variantes del gen E6 determinado para Brasil y Costa Rica se representan en la Figura 17 con las principales mutaciones puntuales características de cada grupo; diferencialmente las variantes del grupo A1, sin distribución en Brasil, presentaron una mutación única en la posición nt.287, para A2 la mutación A232G está presente únicamente en Costa Rica, con una frecuentica inferior al 10 % del total de variantes, mientras que las mutaciones T317C y C342T en promedio, no superaron el 55 % de total de variantes del grupo B en Costa Rica (Anexo E).

La distribución y frecuencia de grupos de variantes del gen E7 determinado únicamente para Costa Rica se representan en la Figura 17 con las principales mutaciones puntuales características de cada grupo; las secuencias agrupadas en A1 fueron 100 % similares a la secuencia de referencia (No. Acceso X05015), las variantes de A2 presentaron la mutación C751T, que permitió su diferenciación con respecto al resto de grupos, dicha agrupación presentó un 100 % de similaridad entre sus variantes, en las secuencias del grupo B, 99,5 % similares entre sí, se pudo observar que la mutación C593T estuvo presente en el 100 % de las variantes, mientras que C640T y A864G aproximadamente para el 85,7 y 66,7 % del grupo (Anexo F).

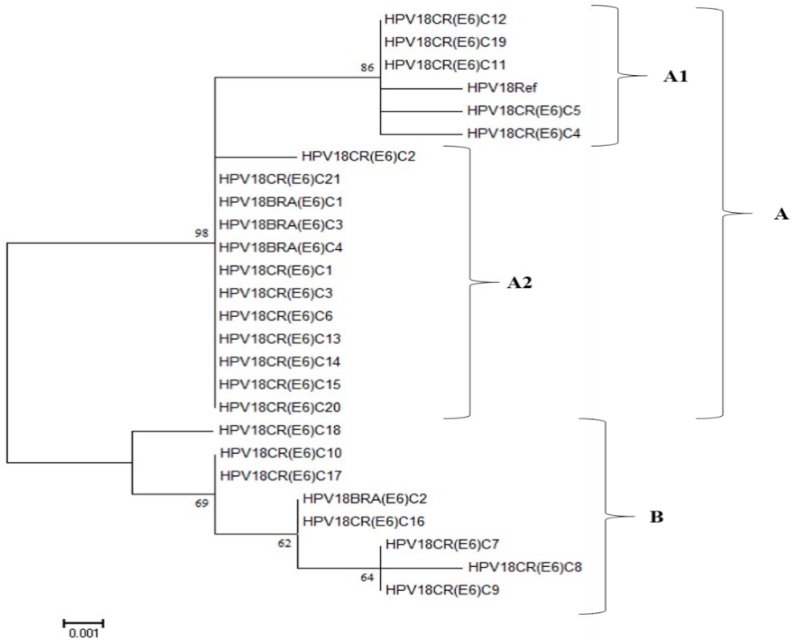


Figura 15. Árbol filogenético que representa las relaciones entre las variantes de E6 de VPH 18, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %); A1, A2 y B señalan los grupos de variantes. *Fuente*: esta investigación.

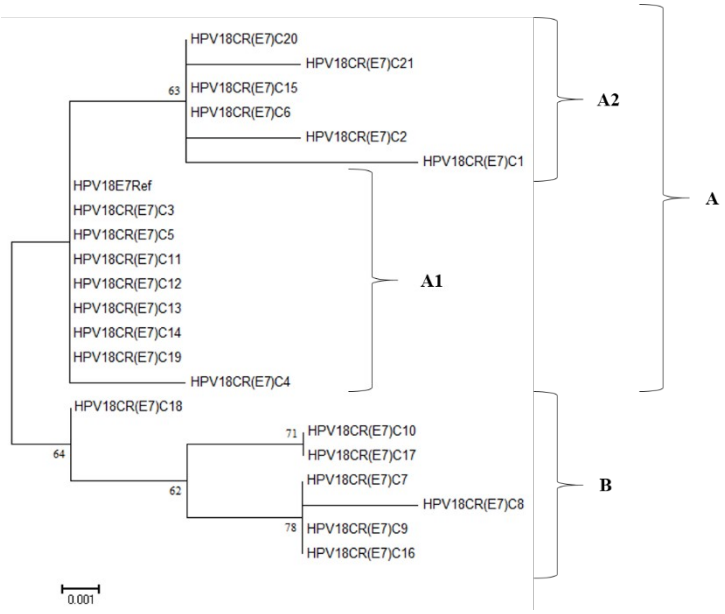


Figura 16. Árbol filogenético que representa las relaciones entre las variantes de E7 de VPH 18, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %); A1, A2 y B señalan los grupos de variantes.

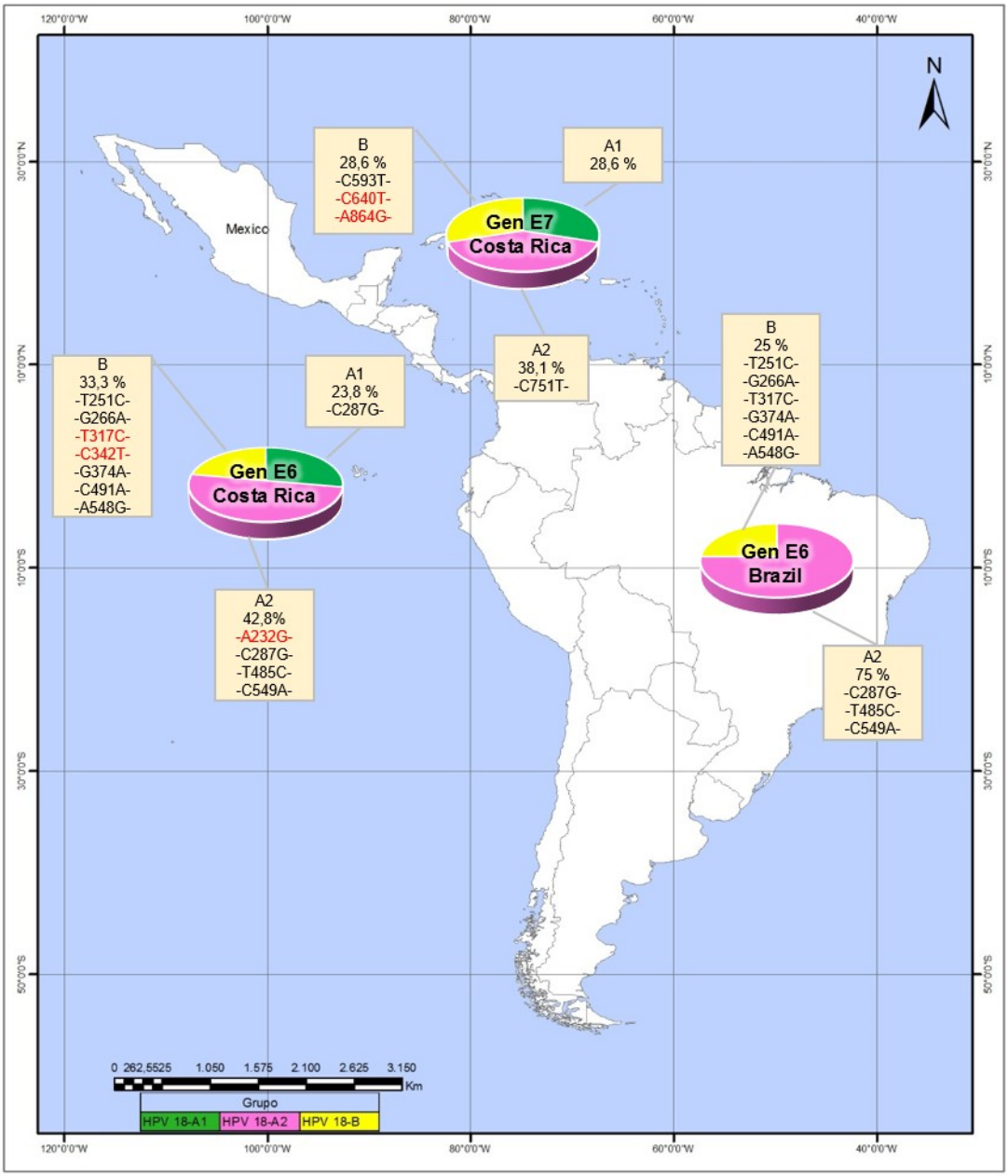


Figura 17. Mapa de distribución geográfica y frecuencias (%) de grupos de variantes de E6 (Brasil y Costa Rica) y E7 (Costa Rica) de VPH 18, construido con el programa ArcGis 10.2; donde A1, A2 y B representan los grupos de variantes y en rojo se distinguen las mutaciones puntuales cuya frecuencia no fue igual al 100 %. *Fuente*: esta investigación.

Para VPH 45 se analizaron 36 ORFs de genes E6 y E7 respectivamente, procedentes de Costa Rica (16), Brasil (10), Paraguay (5), Perú (3) y Panamá (2); teniendo en cuenta que los ORFs de E6 y E7 procedían del mismo aislamiento, el análisis de agrupamiento se realizó concatenando los oncogenes, la topología del árbol filogenético (Figura 18) y las variaciones nucleotídicas (Anexo G y H) ayudaron a delimitar el número de grupos y subgrupos; sin embargo, se evidenció una gran variabilidad intratipo, por lo cual pocas mutaciones constituyeron patrones para delimitar *clusters*.

El *cluster* A fue dividido en cuatro subgrupos debido a la alta variabilidad nucleotídica observada, aun así, dos variantes no pudieron ser subagrupadas debido a la ambigüedad en el patrón mutacional (Figura 18), A1 agrupó las secuencias que fueron 100 % similares a la de referencia (No. Acceso X74879), las variantes de A2 presentaron un 99,9 % de similaridad entre sí, A3 y sus secuencias fueron 99,7 % similares y A4 constituyó la subagrupación de variantes más heterogéneos, reduciendo la similaridad al 99,6 %; por su parte, las variantes del grupo B fueron 99,5 % similares entre sí.

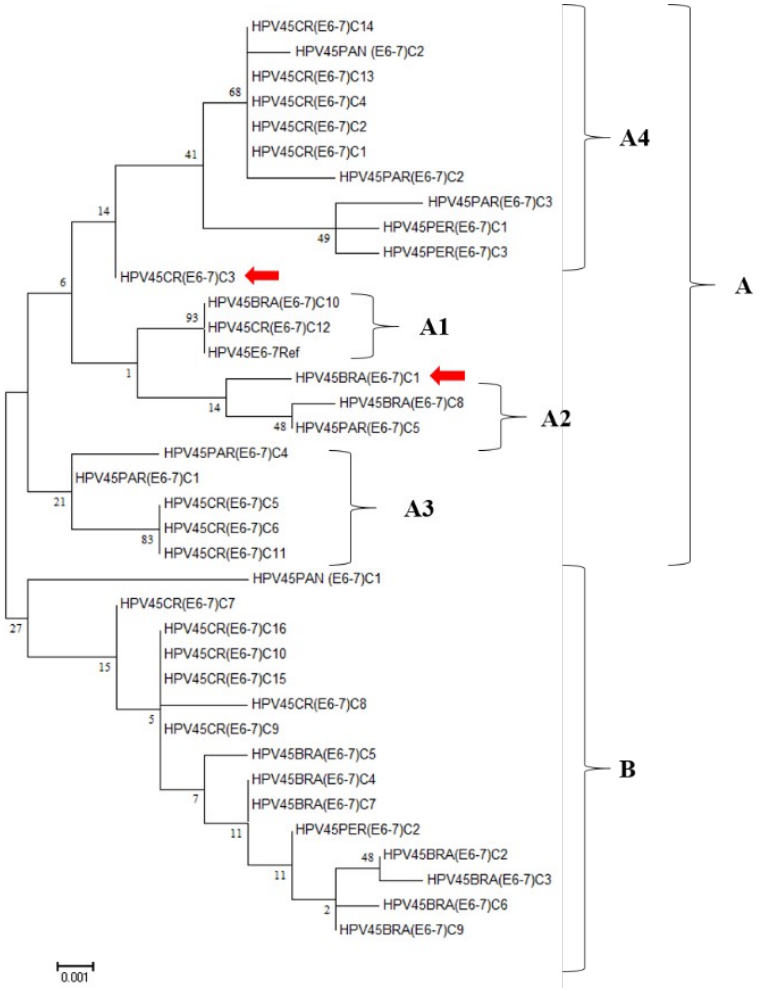


Figura 18. Árbol filogenético que representa las relaciones entre las variantes de VPH 45 con base en las secuencias de los genes E6 y E7, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), las flechas rojas indican que la secuencia no pudo ser asignada a ningún *cluster*, A1, A2, A3, A4 y B señalan los grupos de variantes. *Fuente*: esta investigación.

La distribución y frecuencia de los grupos de variantes para E6 y E7 se representan en la Figura 19, con las principales mutaciones puntuales características de cada agrupación; las secuencias agrupadas en A1 no presentaron mutaciones puntuales, en A2 la frecuencia conjunta de las mutaciones T482C, A603C, A718C, G808T y G822T representó el patrón diferencial para este agrupamiento, las presencia conjunta de las mutaciones puntuales A284C, T482C, A497G y

A603C, diferenciaron las variantes agrupadas en A3, mientras que para A4 la mayoría de mutaciones puntuales difirieron en frecuencia a excepción de T162A, T163C y T482C, observadas en todas las variantes del grupo, algo similar para el grupo B de secuencias, donde únicamente A124C y T482C estuvieron presentes para todas las variantes (Anexo G y H).

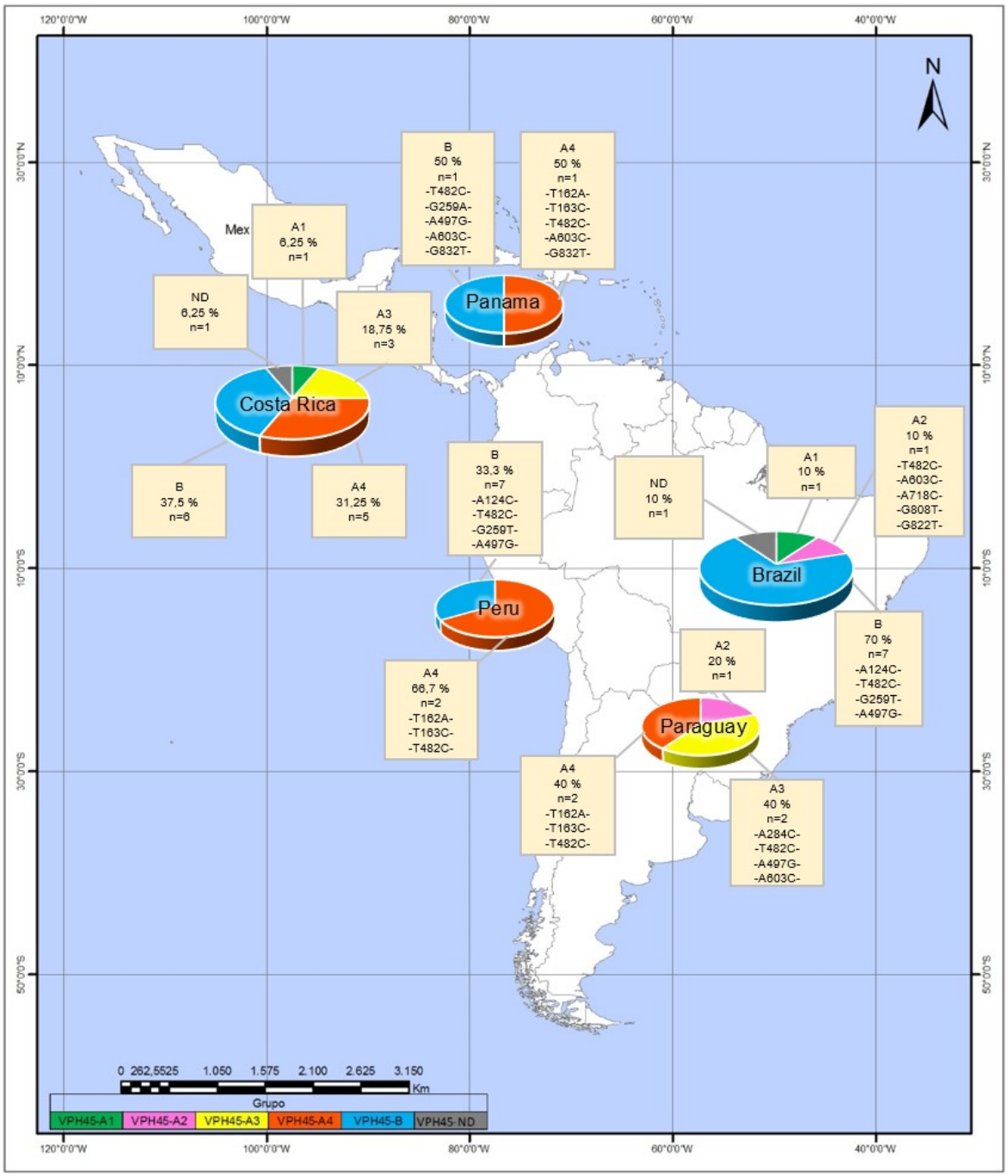


Figura 19. Distribución geográfica y frecuencia (%) de variantes E6 y E7 de VPH 45 en Latinoamérica, construido con el programa ArcGis 10.2, donde A1, A2, A3, A4 y B representan los grupos de variantes, ND representa las secuencias que no pudieron ser agrupadas. *Fuente*: esta investigación. *Fuente*: esta investigación.

El tipo 16 es el VPH-AR más estudiado a nivel molecular, su importancia radica en la frecuencia con la que se ha encontrado en CCU a nivel mundial (Muñoz *et al*., 2003); para el gen E6 se analizaron 140 ORFs procedentes de nueve países: México (51), Costa Rica (37), Argentina (29), Brasil (10), Perú (5), Paraguay (3), Cuba (2), Colombia (2) y Chile (1), a partir del análisis de agrupamiento se identificaron cinco de las seis ramas filogenéticas mundialmente distribuidas asociadas con diferentes regiones geográficas y grupos étnicos (Yamada *et al*., 1997; Wu *et al*., 2006) y se clasificaron bajo el sistema de nomenclatura para variantes moleculares de E6 del VPH 16 propuesto por Huertas-Salgado *et al*. (2011).

El 31,4 % (44/140) de las variantes pertenecieron a las rama E, el 22,85 % (32/140) a la rama AA, la rama AN 1 y Af-1 agruparon el 7,1 % (10/140) y 4,3 % (6/140) de las variantes, respectivamente, mientras que la rama menos representativa dentro del análisis fue la As con apenas una variante; por otra parte, el 33,5 % (47/140) restante estuvo representado por variantes que no pudieron ser clasificadas bajo el sistema de Huertas-Salgado *et al*. (2011) debido al patrón mutacional de sus secuencias (Figura 20).

La distribución y frecuencia de las ramas filogenéticas para variantes de E6 se representan en la Figura 21, las principales mutaciones puntuales características de cada agrupación se resumen a continuación: las variantes de la rama E se asocian con la presencia o ausencia de un cambio de T a G o T a A en el nt 350, todas las variantes de la de la rama As pueden ser identificadas de acuerdo a mutaciones únicas en la posición T178, las variantes de la rama AA presentan las mutaciones comunes G145T, T286A, A289G, T350G y A532G (este último cambio permite diferenciarla de la rama NA 1) y puede presentar mutaciones adicionales menos frecuentes como C335T y T183G, las variantes de la rama NA 1 tienen en común las mutaciones G145T, T286A, A289G y pueden presentar mutaciones adicionales como C335T, T350G y T183G y, las mutaciones C143G, G145T, T286A, A289G y C335T describen las variantes de Af-2 (Anexo I).

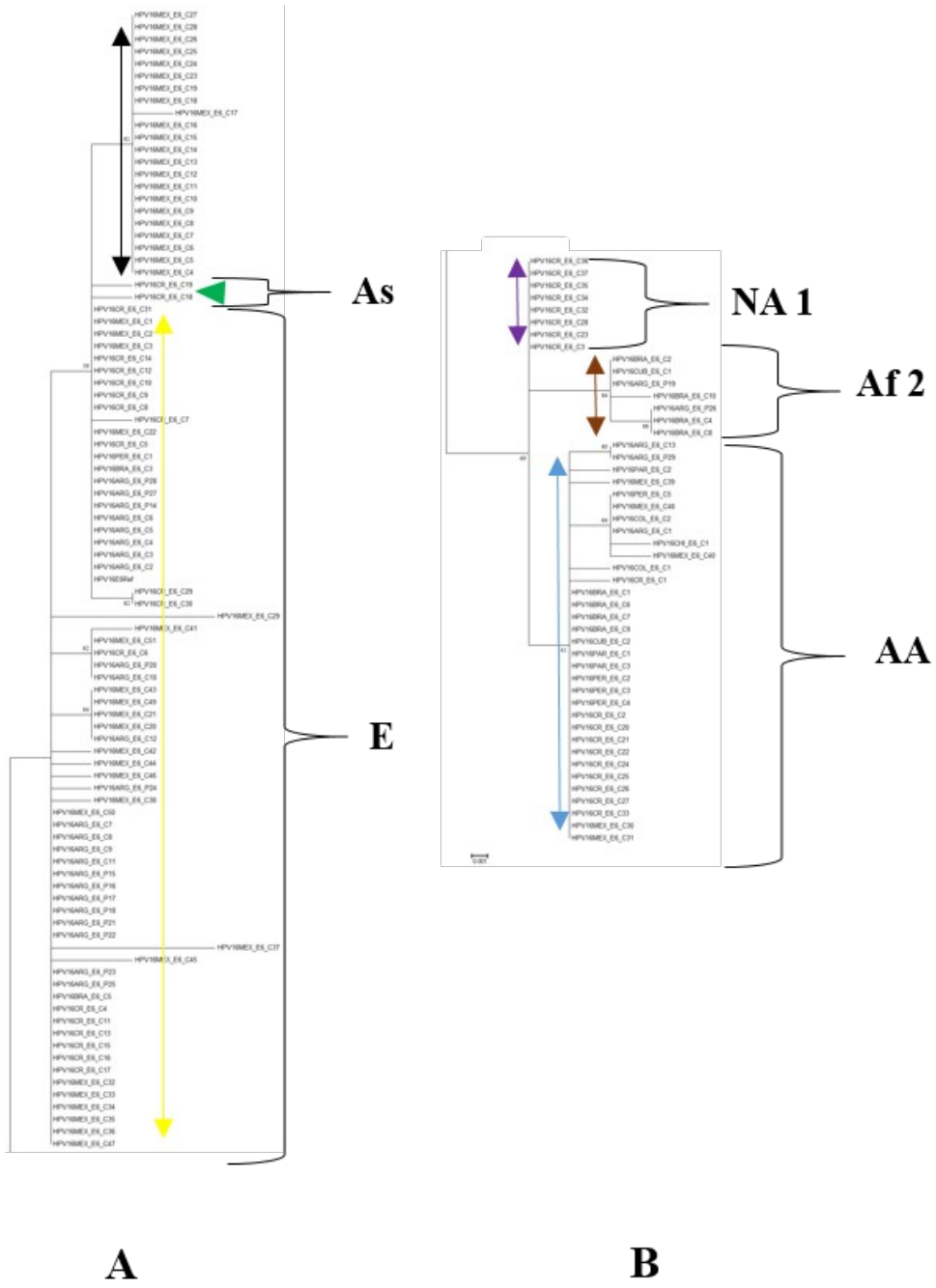


Figura 20. Árbol filogenético que representa las relaciones entre las variantes de E6 de VPH 16, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %); A y B son utilizados para diferenciar dos grandes *clusters*, el *cluster* A agrupa las ramas E, As y las variantes sin rama definida y B es el segundo agrupa a las ramas NA-1, Af-2 y AA, el color de las flechas es según la rama filogenética y se relaciona con el anexo I: Rama Europea-E (amarillo), rama Asia-As (verde), rama Asiático-americana-AA (azul celeste), rama América del norte 1-NA1 (purpura) y rama africana 2- Af-2 (marrón). *Fuente*: esta investigación.

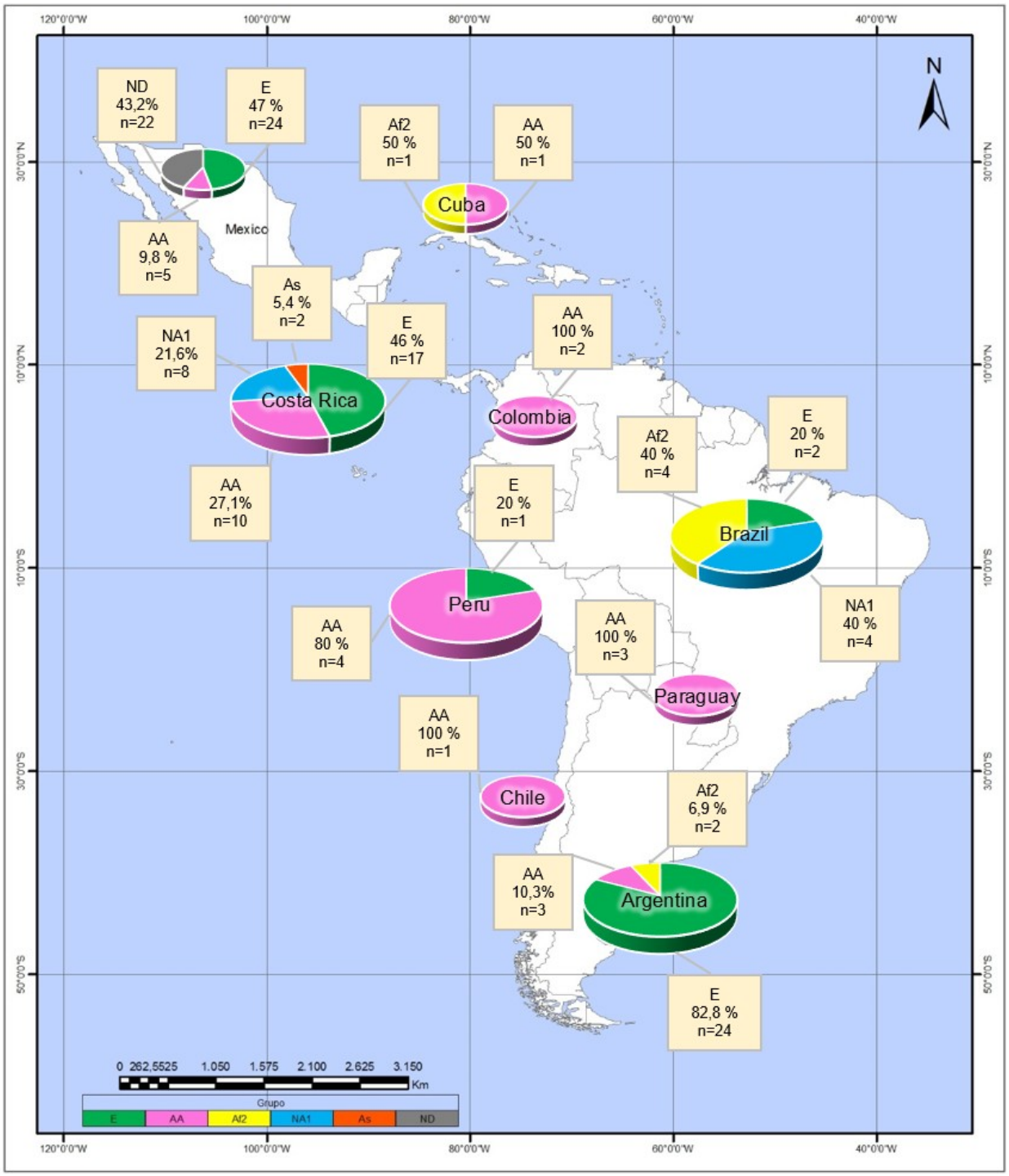


Figura 21. Distribución geográfica y frecuencia (%) de ramas filogenéticas de E6 VPH 16 en Latinoamérica, construido con el programa ArcGis 10.2. Rama Europea-E, rama Asia-As, rama Asiático-americana-AA, rama América del norte 1-NA1 y rama africana 2- Af-2. *Fuente*: esta investigación.

Las secuencias E7 provinieron de Costa Rica (37), México (5) y Brasil (1), las variantes detectadas par este gen, a partir del análisis filogenético fueron agrupadas en tres *clusters*: A, B y C; el grupo A incluyó variantes que podrían tener un origen europeo, dado que la secuencia de referencia (No. Acceso J04353) que origina la rama europea para E6 se ubicó en este grupo, B incluyó variantes de la rama AA, dadas las mutaciones T732C, T789C y T795G, las cuales diferencian esta agrupación y concuerda con lo reportado por De Boer *et al*. (2004) y Tornesello *et al*. (2004), por otra parte, para C se agrupó una única variante 58 % distante del resto de grupos.

En comparación con las secuencias de E6, las secuencias de E7 presentaron menor variabilidad con relación al número de mutaciones puntuales presentes en ellas. Esto evidencia que las secuencias de E6 de VPH 16 son mucho más variables y que las de E7, que tienden a ser más conservadas, según lo dicho por Huertas-Salgado *et al*. (2011).

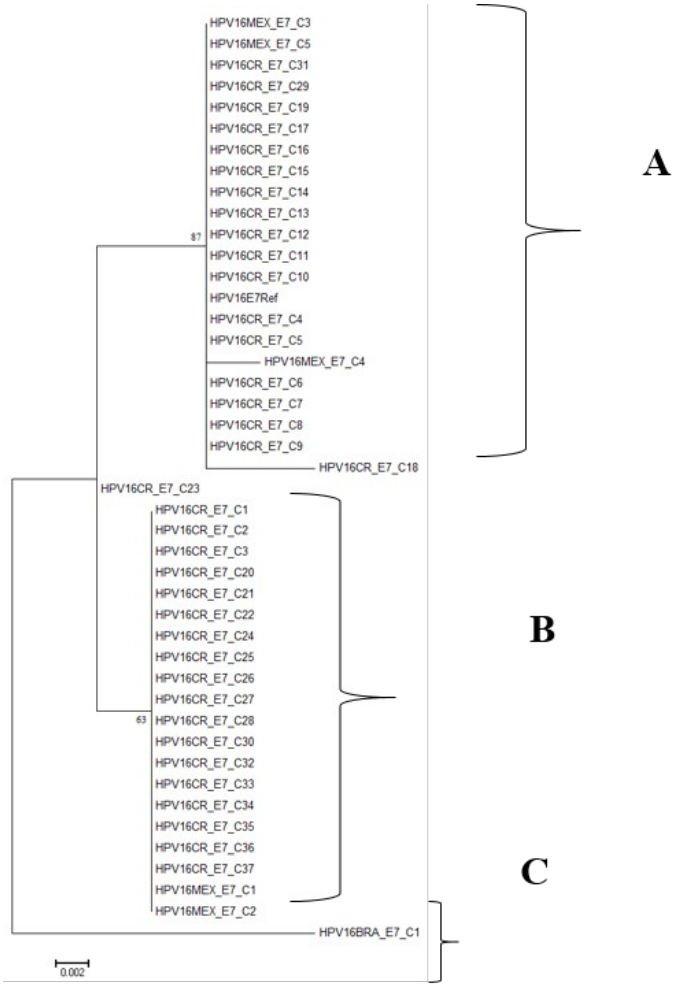


Figura 22. Árbol filogenético que representa las relaciones entre las variantes de E7 de VPH 16, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, En las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), A, B y C representan los grupos de variantes. *Fuente*: esta investigación.

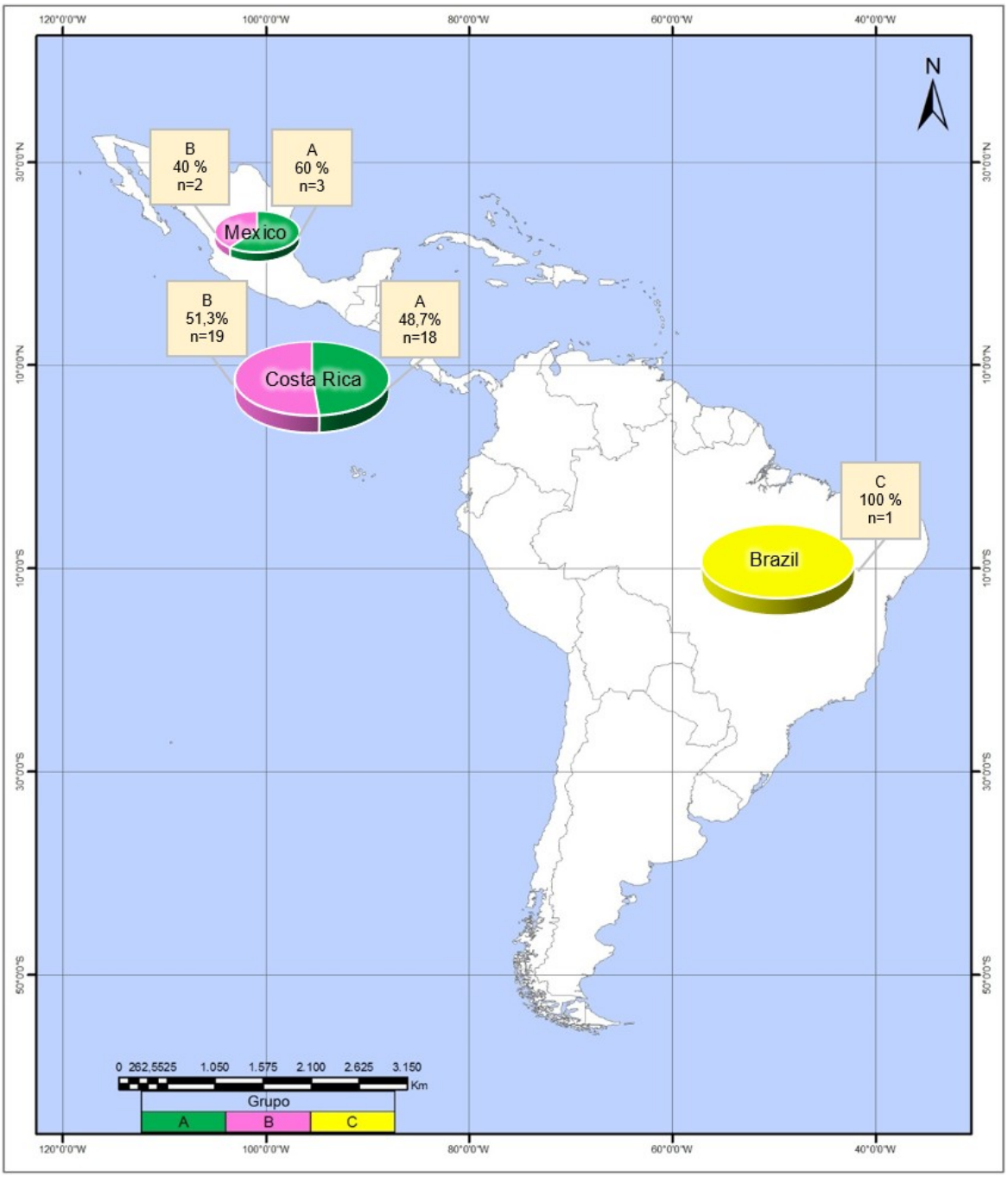


Figura 23. Distribución geográfica y frecuencia (%) de variantes de E7 VPH 16 en Latinoamérica. *Fuente*: esta investigación, construido con el programa ArcGis 10.2, A, B y C representan los grupos de variantes. *Fuente*: esta investigación.

Para VPH 31 fueron analizados 22 ORFs de E6 provenientes de Costa Rica (16) y Brasil (6) y 19 ORFs de E7 de Costa Rica (16) y Brasil (3); para establecer la relación entre grupos de variantes, para cada gen generamos un árbol filogenético mediante el cual se evidenciaron las mismas tres agrupaciones: A, B y C (Figura 24 y 25), similar a la topología propuesta por Chen *et al*. (2011). Para ambos genes, las secuencias agrupadas en A, se encontraron más relacionadas con la secuencia de referencia (No. Acceso J04353), similares entre sí 99,2 % para E6 y 99,3 % para E7, el porcentaje de similaridad de las variantes agrupadas en B para E6 fue del 98,4 %, mientras que para E7 fue del 98,6 % y para las secuencias agrupadas en C, los porcentajes de similaridad indicaron que las variantes de este grupo fueron las más heterogéneas, con 98,25 % (E6) y 98,3 % (E7).

La distribución y frecuencia de los grupos de variantes para E6 se representan en la Figura 26, con las principales mutaciones puntuales características de cada agrupación; las variantes en A presentaron mutaciones puntuales dispersas, que no constituyeron un patrón diferencial para el agrupamiento, mientras que las variantes de los grupos B y C compartieron mutaciones en las posiciones nt 320 y 520, B se diferencia de C en las transiciones A297G y G475A, entretanto, las mutaciones A261T y G404A, A428G observadas en las variantes agrupadas en C se presentaron con una frecuencia inferior al 87,5 % (Anexo K).

La distribución y frecuencia de los grupos de variantes para E7 se representan en la Figura 27, con las principales mutaciones puntuales características de cada agrupación; las variantes en A presentaron mutaciones puntuales dispersas, que no constituyeron un patrón diferencial para el agrupamiento, a excepción de la sustitución A473G, presente en todas las variantes del árbol, las variantes de los grupos B y C compartieron mutaciones en las posiciones nt 670, 695 y 743, sin embargo, C se diferencia de B en la transición G580A; la proporción de grupo de variantes entre países indicó que el grupo B únicamente estuvo presente en Costa Rica (Anexo L).

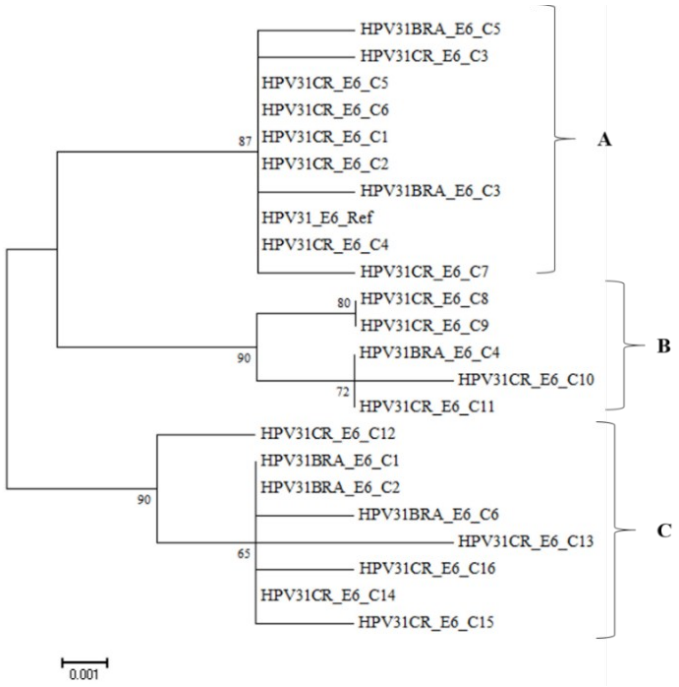


Figura 24. Árbol filogenético que representa las relaciones entre las variantes de E6 de VPH 31, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), A, B y C representan los grupos de variantes. *Fuente*: esta investigación.

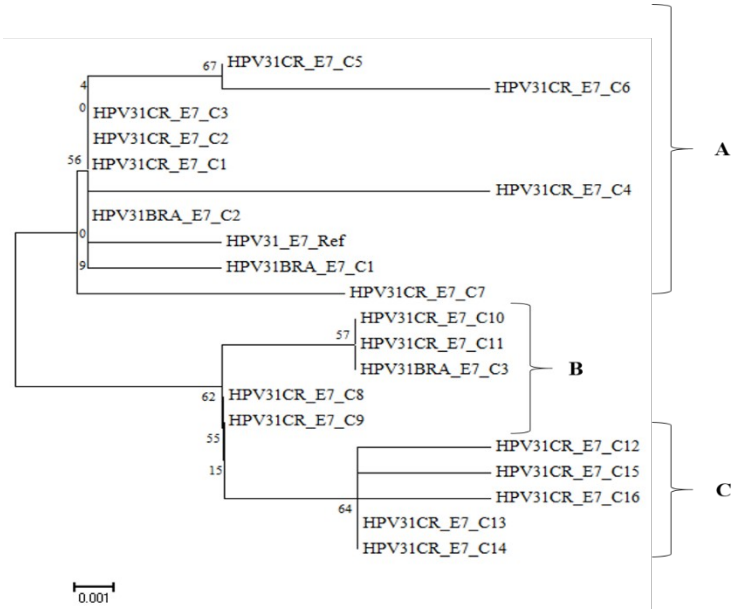


Figura 25. Árbol filogenético que representa las relaciones entre las variantes de E7 de VPH 31, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), A, B y C representan los grupos de variantes. *Fuente*: esta investigación.

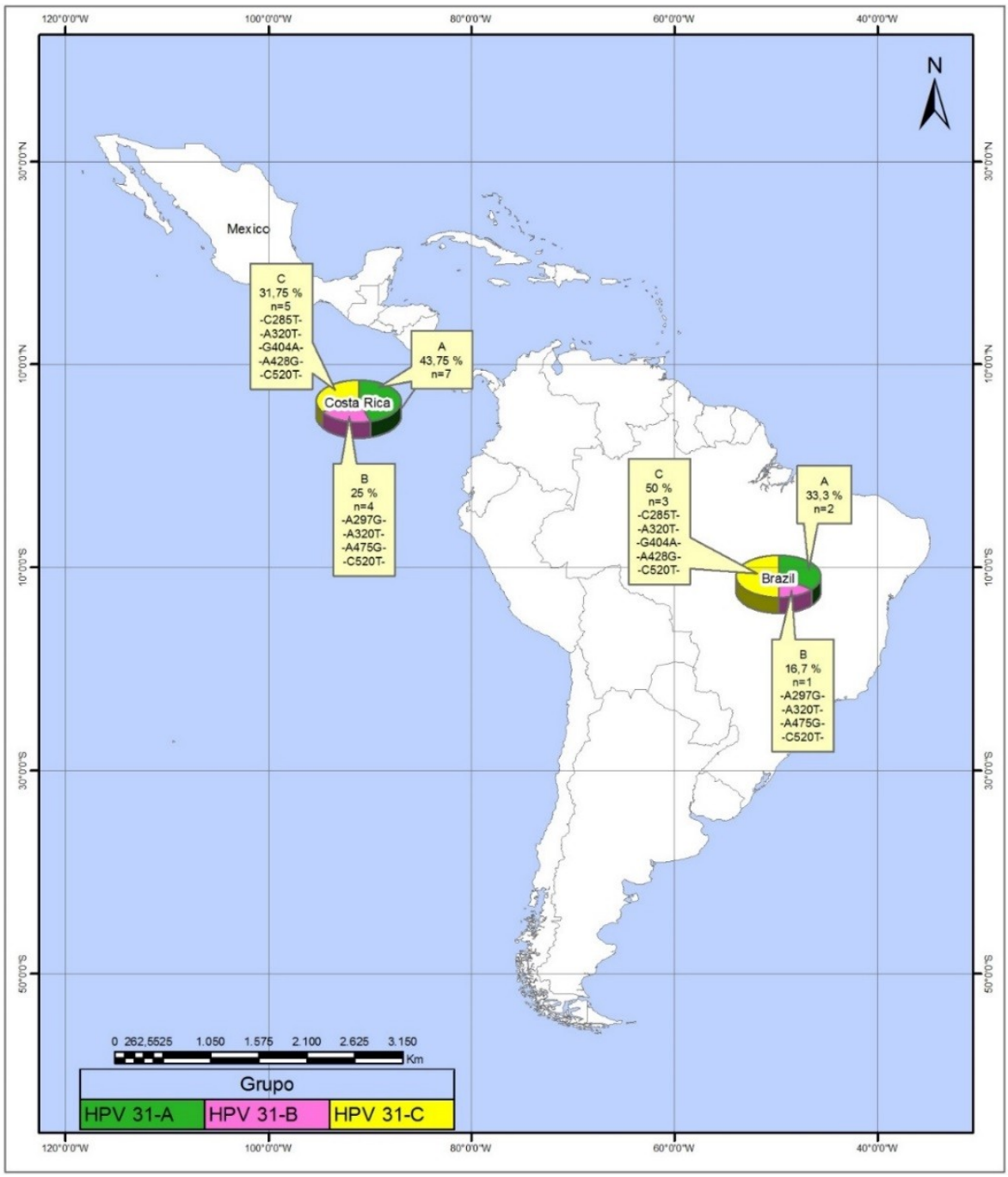


Figura 26. Distribución geográfica y frecuencia (%) de variantes E6 de VPH 31 procedentes de Brasil y Costa Rica, construido con el programa ArcGis 10.2, donde, A, B y C representan los grupos de variantes. *Fuente*: esta investigación.

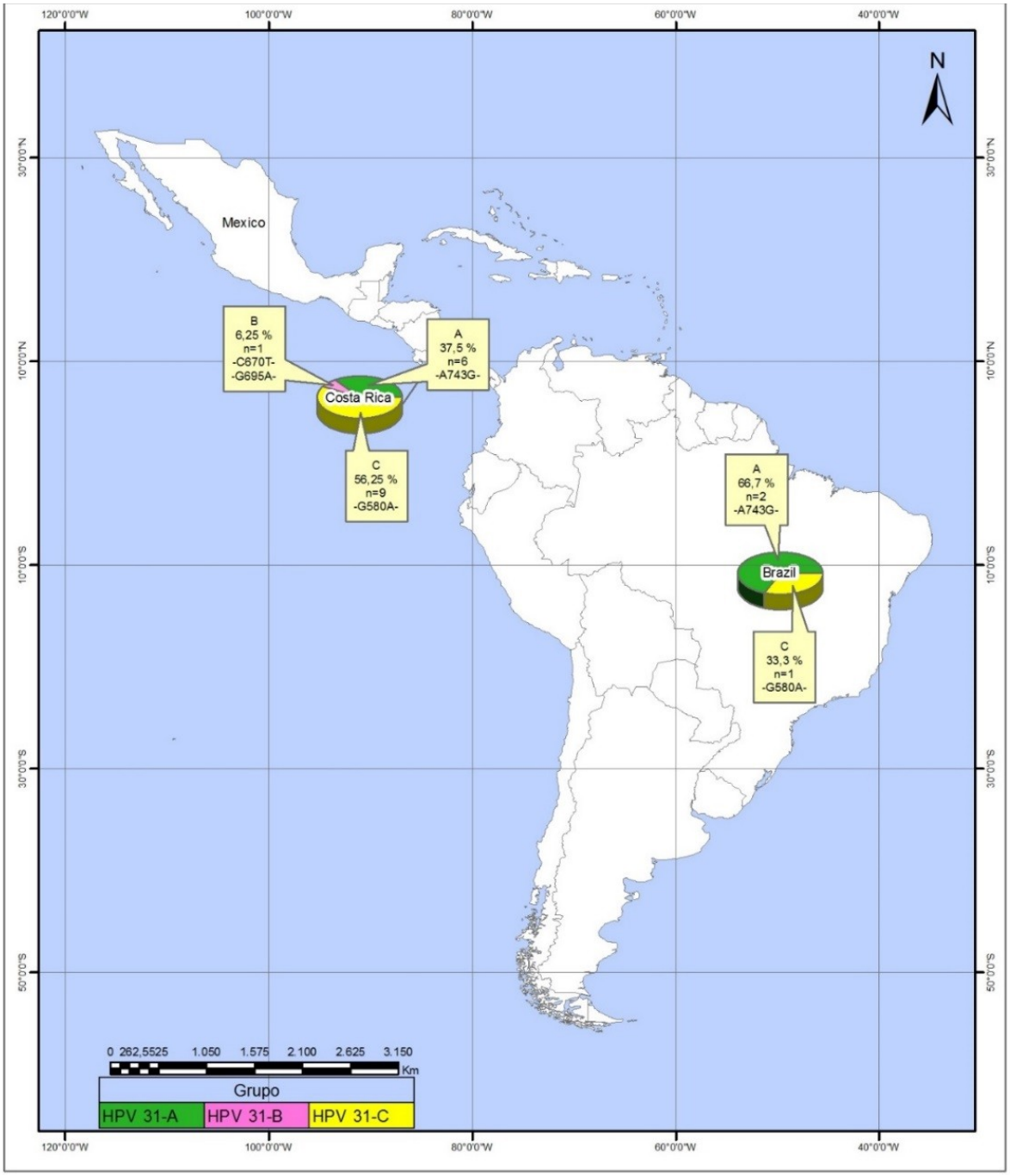


Figura 27. Distribución geográfica y frecuencia (%) de variantes de E7 VPH 31 identificadas para Brasil y Costa Rica, construido con el programa ArcGis 10.2, donde, A, B y C representan los grupos de variantes. *Fuente*: esta investigación.

Para VPH 33 fueron analizados 20 ORFs de E6 y E7 procedentes de los mismo aislamientos y por ende con la misma distribución de países: Costa Rica (8), Brasil (5), Paraguay (5),Chile (1) y Perú (1); para establecer la relación entre grupos de variantes, concatenaos los genes para generar un único árbol filogenético mediante el cual se evidenciaron tres agrupaciones: A, B y C (Figura 28), similar a la topología propuesta por Chen *et al*. (2011). Las secuencias agrupadas en A, se encontraron más relacionadas con la secuencia de referencia (No. Acceso M12732) y fueron 100% similares a esta, el porcentaje de similaridad de las variantes agrupadas en B fue del 99, 14 % y para las secuencias agrupadas en C, el porcentajes de similaridad indicó que las variantes de este grupo fueron las más heterogéneas con un 98,1 %.

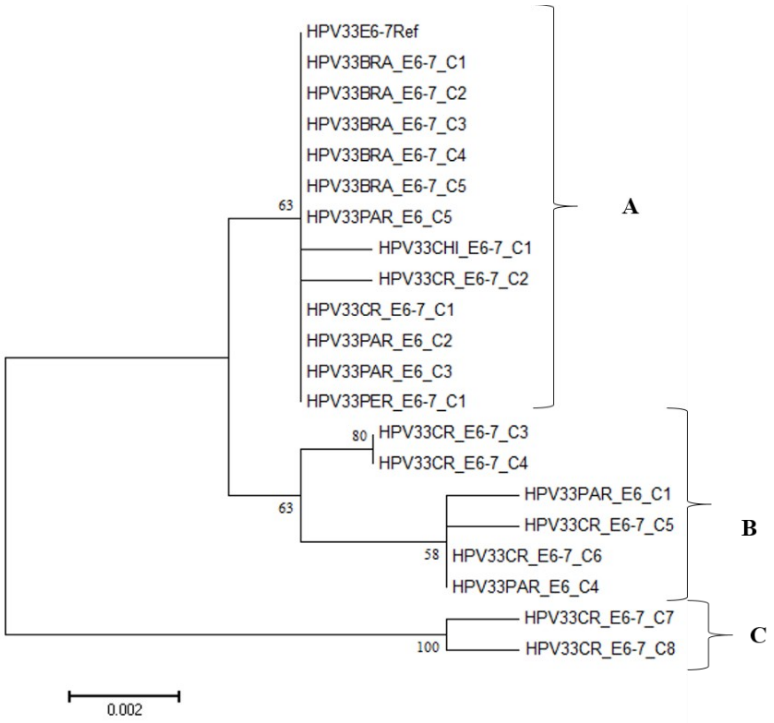


Figura 28. Árbol filogenético que representa las relaciones entre las variantes de VPH 33 con base en las secuencias de los genes E6 y E7, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), A, B y C representan los grupos de variantes. *Fuente*: esta investigación.

La distribución y frecuencia de los grupos de variantes para VPH 33 se representan en la Figura 29, con las principales mutaciones puntuales características de cada agrupación; las variantes en A presentaron mutaciones puntuales dispersas, que no constituyeron un patrón diferencial para el agrupamiento, mientras que en las variantes del grupo B la mutación A213C se encuentra en el 100 % de las variantes, seguido de A862T presente en 83,3 % de las mismas, otras mutaciones puntuales se presentan con una frecuencia entre el 66,7 % (A364C y A480T) y el 50 % (A737G), las variantes del grupo C, presentaron nueve puntos de mutación comunes en las posiciones nt 315, 355, 387, 521, 534, 722, 833, 850 y 862 (Anexo M).

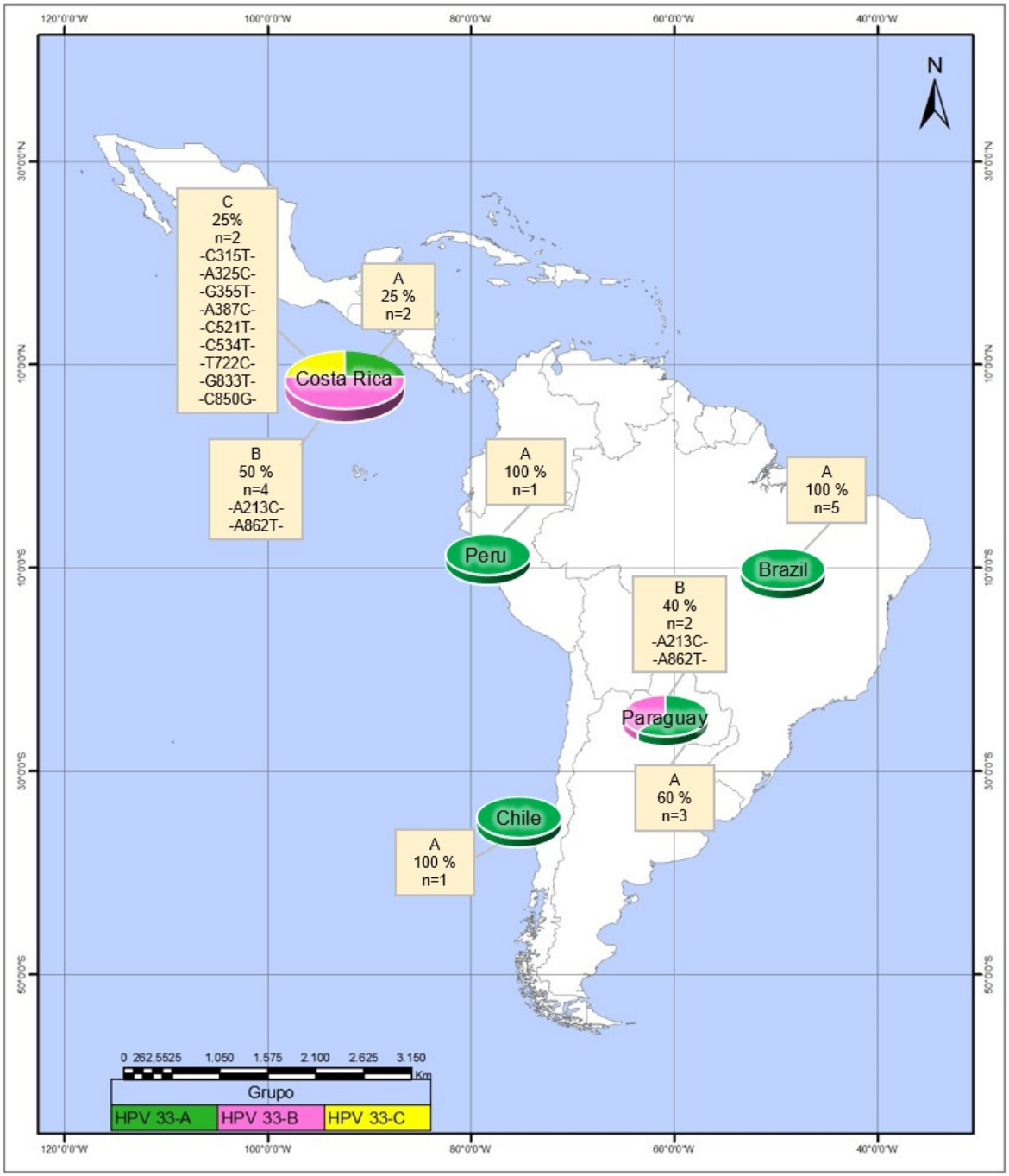


Figura 29. Distribución geográfica y frecuencia (%) de variantes de VPH 33 identificadas para Latinoamérica, construido con el programa ArcGis 10.2, donde, A, B y C representan los grupos de variantes. *Fuente*: esta investigación.

Para VPH 35 fueron analizados 8 ORFs de E6 y E7 procedentes de los mismo aislamientos y de un único país (Costa Rica); para establecer la relación entre grupos de variantes, concatenamos los genes para generar un único árbol filogenético mediante el cual se evidenciaron tres agrupaciones: A, B y C (Figura 30), similar a la topología propuesta por Chen *et al*. (2011). Las secuencias agrupadas en A, se encontraron más relacionadas con la secuencia de referencia (No. Acceso X74477) y fueron 99,7 % similares entre sí, las mutaciones C131A y A295T en E6 permitieron la diferenciación de este grupo, el porcentaje de similaridad de las variantes agrupadas en B fue del 100 % y se encontró diferenciado por la transversión G543T de E6 y las mutaciones T628C y T675C de E7, mientras que para las secuencias agrupadas en C, el porcentajes de similaridad fue del 99,9 % y tres mutaciones únicas en las posiciones nt 127, 136 y 341 de E6 permitieron la diferenciación del grupo (Anexo N).

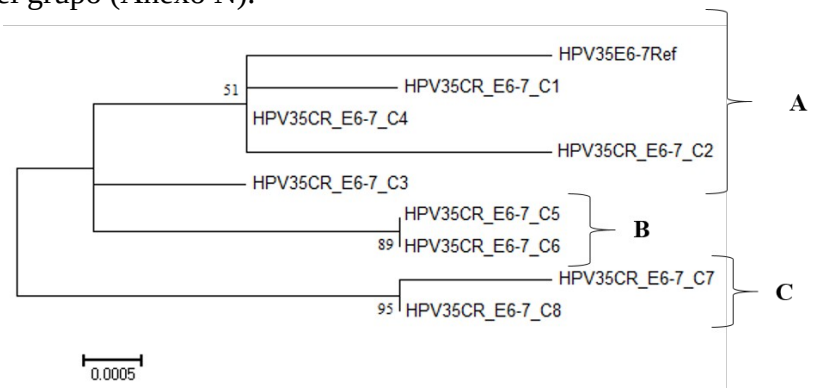


Figura 30. Árbol filogenético que representa las relaciones entre las variantes de VPH 35 con base en las secuencias de los genes E6 y E7, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), A, B y C representan los grupos de variantes. *Fuente*: esta investigación.

Para el gen E6 de VPH 52 fueron analizados 18 ORFs provenientes de Costa Rica (17) y Argentina (1); para establecer la relación entre grupos de variantes, generamos un árbol filogenético mediante el cual se evidenciaron cinco agrupaciones: A1, A2, B, C y D (Figura 31), similar a la topología propuesta por Chen *et al*. (2011). Las secuencias agrupadas en A1, se encontraron más relacionadas con la secuencia de referencia (No. Acceso X74481), similares entre sí 99,7 %, las variantes en A2 fueron 100 % similares entre sí, al igual que para las variantes agrupadas en B, el porcentaje de similaridad de las variantes agrupadas en C fue del 99,7 %, mientras que el porcentaje de similaridad indicó las variantes de este grupo D fueron las más heterogéneas con un 99,2 % entre sí.

La distribución y frecuencia de los grupos de variantes de E6 para VPH 52se representan en la Figura 32, con las principales mutaciones puntuales características de cada agrupación; las variantes en A1 presentaron mutaciones puntuales dispersas, que no constituyeron un patrón diferencial para el agrupamiento, mientras que en A2 las mutaciones A251G y A379G definieron el grupo, en las variantes del grupo B las mutaciones G350C y A379G diferenciaron este *cluster* del resto de agrupaciones, las variantes del grupo C, presentaron conjuntamente las mutaciones G350C y A530G , mientras que para D cuatro posiciones definieron la agrupación: C200T, C237T, G350C y G425T (Anexo Ñ).

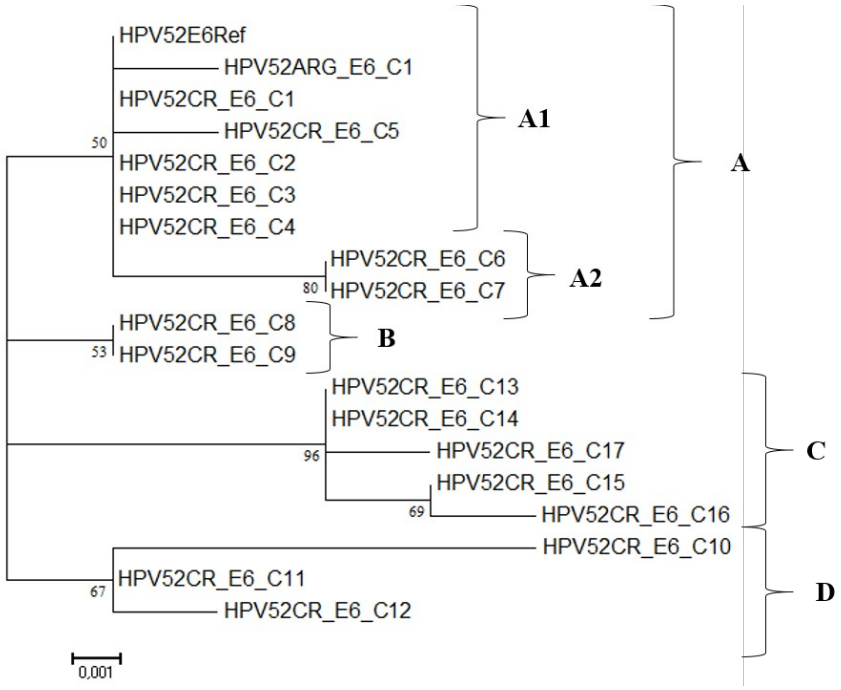


Figura 31. Árbol filogenético que representa las relaciones entre las variantes de E6 de VPH 52, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p. En las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), A1, A2, B, C y D representan los grupos de variantes. *Fuente*: esta investigación.

Para el gen E7 de VPH 52 fueron analizados 17 ORFs provenientes de Costa Rica; para establecer la relación entre grupos de variantes, generamos un árbol filogenético mediante el cual se evidenciaron cuatro agrupaciones: A, B, C y D (Figura 33), similar a la topología propuesta por Chen *et al*. (2011). Las secuencias agrupadas en A, se encontraron más relacionadas con la secuencia de referencia (No. Acceso X74481) y 100 % similares entre sí, igual que las variantes agrupadas en B, donde las mutaciones C751T y A801G diferencian el grupo, las variantes agrupadas en C fueron 98,1 % similares entre sí, constituyendo el grupo de variantes más heterogéneas, diferente del resto de grupos por la frecuencia conjunta de las mutaciones T573C, C766A y A801G, mientras que las variantes agrupadas en D fueron 99,3 % similares entre sí, el 83,3 % (10/12) de las mutaciones puntuales de E7 se presentaron en las variantes de este grupo, sin embargo, las mutaciones comunes para todas las variantes fueron T573C, A706G, G707A, T727G, G742A y A801G (Anexo O).

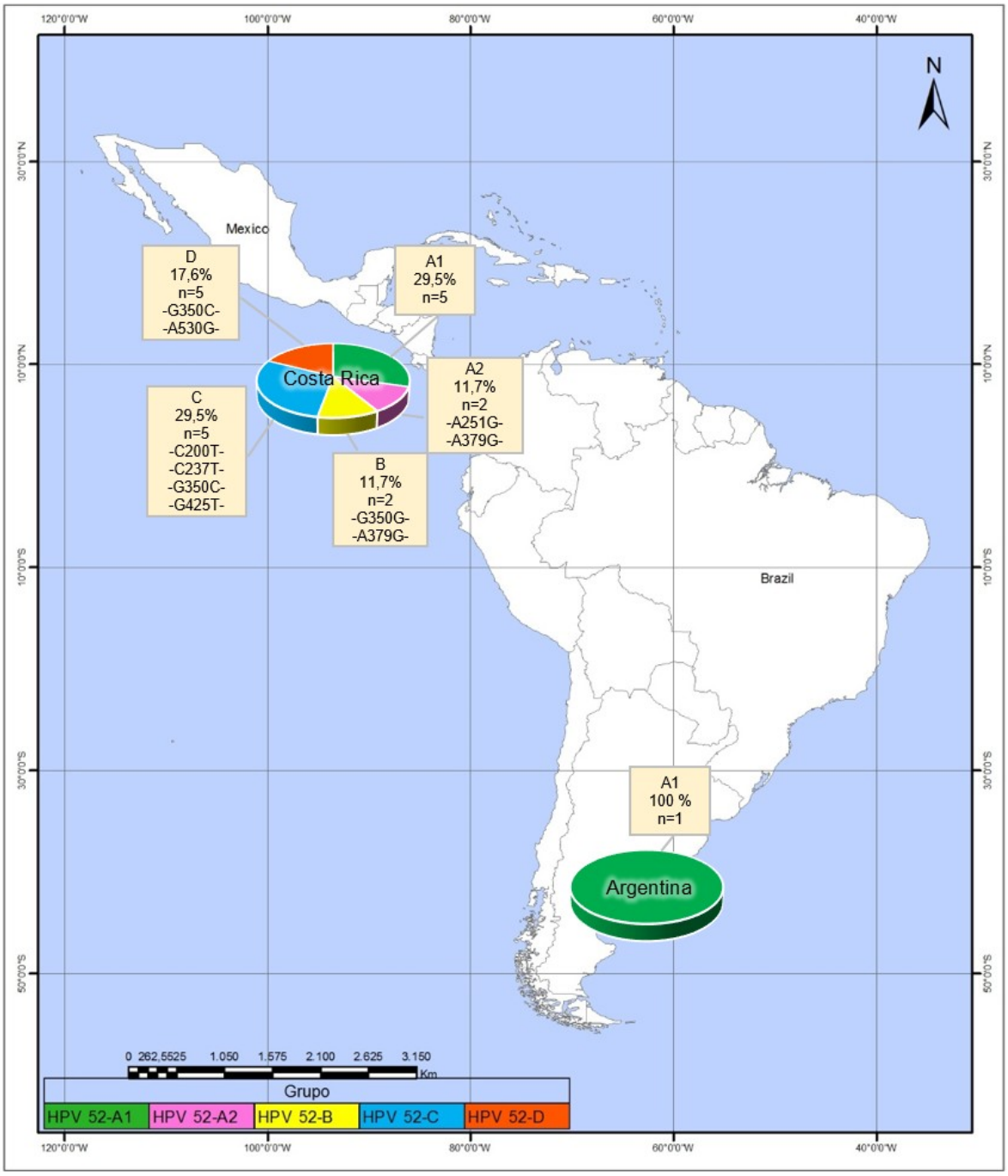


Figura 32. Distribución geográfica y frecuencia (%) de variantes de E6 VPH 52 identificadas para Argentina y Costa Rica, construido con el programa ArcGis 10.2, donde, A1, A2, B, C y D representan los grupos de variantes. *Fuente*: esta investigación.

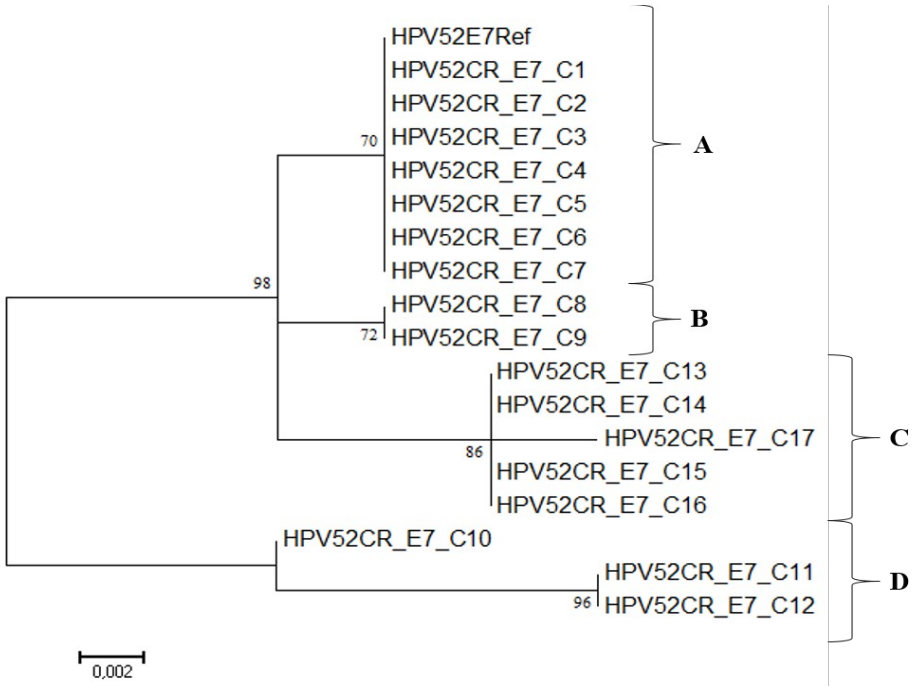


Figura 33. Árbol filogenético que representa las relaciones entre las variantes de E7 de VPH 52, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), A, B, C y D representan los grupos de variantes. *Fuente*: esta investigación.

Para el gen E6 de VPH 58 fueron analizados 14 ORFs provenientes de Costa Rica (11), Argentina (3) y Honduras (1); para establecer la relación entre grupos de variantes, generamos un árbol filogenético mediante el cual se evidenciaron cuatro agrupaciones: A1, A2, A3 y B (Figura 34). Las secuencias agrupadas en A1, se encontraron más relacionadas con la secuencia de referencia (No. Acceso D90400), 99,8 % similares entre sí, las variantes en A2 fueron 100 % similares entre sí, al igual que las variantes del grupo A3, el porcentaje de similaridad de las variantes agrupadas en B fue del 99,7 %.

La distribución y frecuencia de los grupos de variantes de E6 para VPH 58 se representan en la Figura 35, con las principales mutaciones puntuales características de cada agrupación; la mutación C307T está presente en todos los grupos de variantes, sin embargo, para la subagrupación A1 esta variación constituye un patrón de diferenciación, debido a su altísima frecuencia en comparación con otras mutaciones dispersas en diferentes posiciones, en A2 se presentaron conjuntamente las mutaciones C187T y C307T y en A3 las mutaciones C187T, C307T y C373A constituyeron es el patrón diferencial; las variantes del grupo B se diferenciaron del resto de agrupaciones por la frecuencia conjunta de las mutaciones G193A, G245A, C307T y A398G (Anexo P).

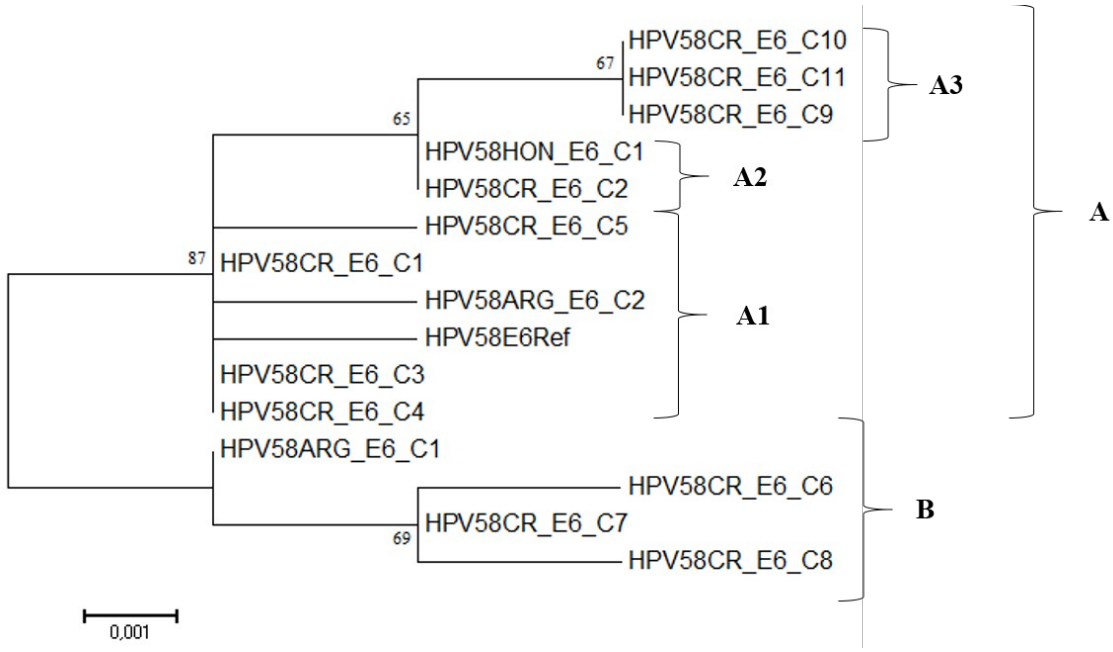


Figura 34. Árbol filogenético que representa las relaciones entre las variantes de E6 de VPH 58, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), A1, A2, A3 y B representan los grupos de variantes. *Fuente*: esta investigación.

Para el gen E7 de VPH 58 fueron analizados 16 ORFs provenientes de Costa Rica (11), Guatemala (2), Argentina (1) Brasil (1) y Colombia (1); para establecer la relación entre grupos de variantes, generamos un árbol filogenético mediante el cual se evidenciaron cinco agrupaciones: A1, A2, A3, B1 y B2 (Figura 36). Las secuencias agrupadas en A1, se encontraron más relacionadas con la secuencia de referencia (No. Acceso D90400), 99,8 % similares entre sí, las variantes en A2 fueron 99,9 % similares entre sí, mientras que la que las variantes del grupo A3 fueron 100 % similares, al igual que las variantes de B1 y B2.

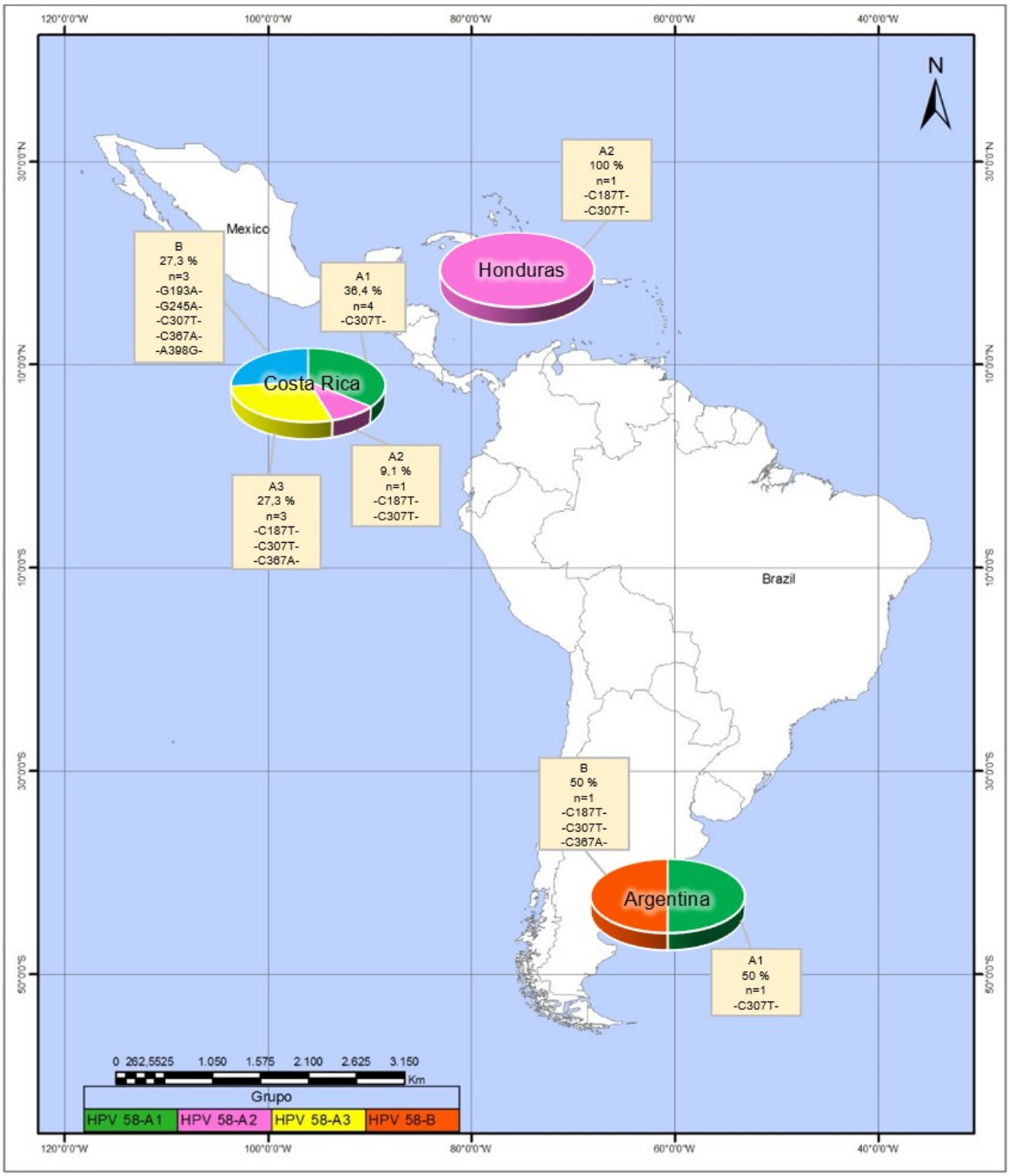


Figura 35. Distribución geográfica y frecuencia (%) de variantes de E6 VPH 58 identificadas para Latinoamérica, construido con el programa ArcGis 10.2, donde, A1, A2, A3 y B representan los grupos de variantes. *Fuente*: esta investigación.

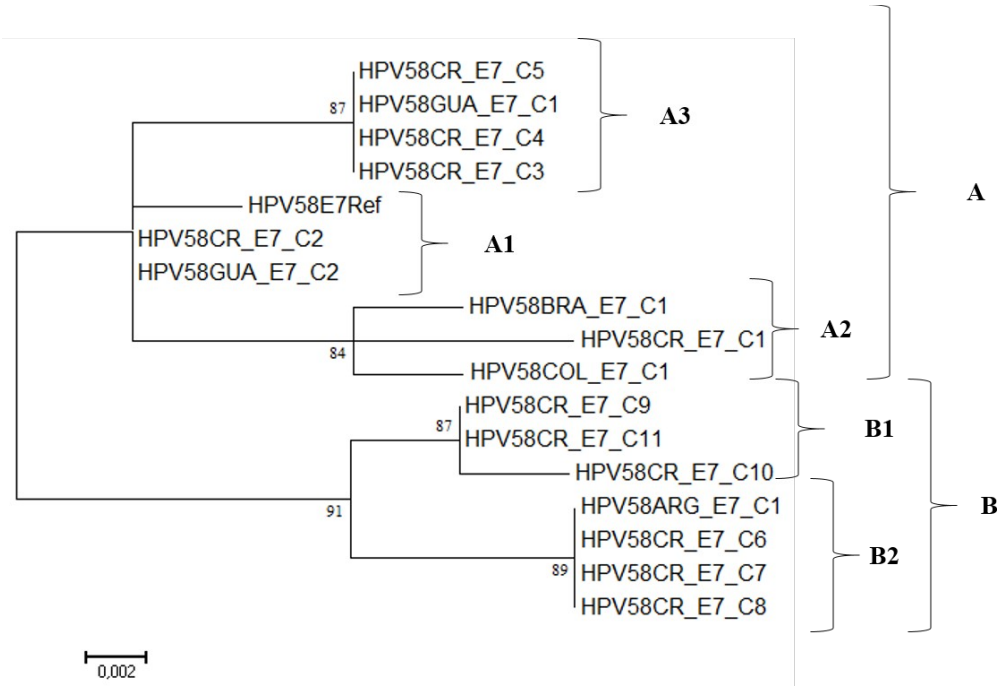


Figura 36. Árbol filogenético que representa las relaciones entre las variantes de E7 de VPH 58, construido con el programa MEGA v. 6.0.5, bajo el criterio de ML, utilizando el modelo de sustitución de nucleótidos Kimura 2-p, en las ramas se representan los valores de *bootstrap* (1000 réplicas) que fueron significativos (> 50 %), A1, A2, A3, B1 y B2 representan los grupos de variantes. *Fuente*: esta investigación.

La distribución y frecuencia de los grupos de variantes de E7 para VPH 58 se representan en la Figura 37, con las principales mutaciones puntuales características de cada agrupación; la mutación T744G está presente en todos los grupos de variantes, sin embargo, para la subagrupación A1 esta variación constituye un patrón de diferenciación, debido a su altísima frecuencia en comparación con otras mutaciones dispersas en diferentes posiciones, en A2 se presentaron conjuntamente las mutaciones G694A, T744G y G761A y, en A3 las mutaciones C632T, T744G y G760C constituyeron es el patrón diferencial; las variantes del grupo B1 se diferenciaron del resto de agrupaciones por la frecuencia conjunta de las mutaciones T744G, A793G, C798T, C801A, C840T y T852C, para B2, las mutaciones características del grupo fueron T744G, G760C, G761A, A793G, C798T, C801A y C840T (Anexo Q).

Los datos y los conocimientos sobre la distribución geográfica intratípica de variantes para los tipos de VPH más frecuentes en Latinoamérica, podría ayudar a establecer una base de datos sobre la diversidad y la patogenicidad de las diferentes variantes de tipos, con lo que se puede diseñar y optimizar los protocolos de diagnóstico a fin de reducir la enfermedad.

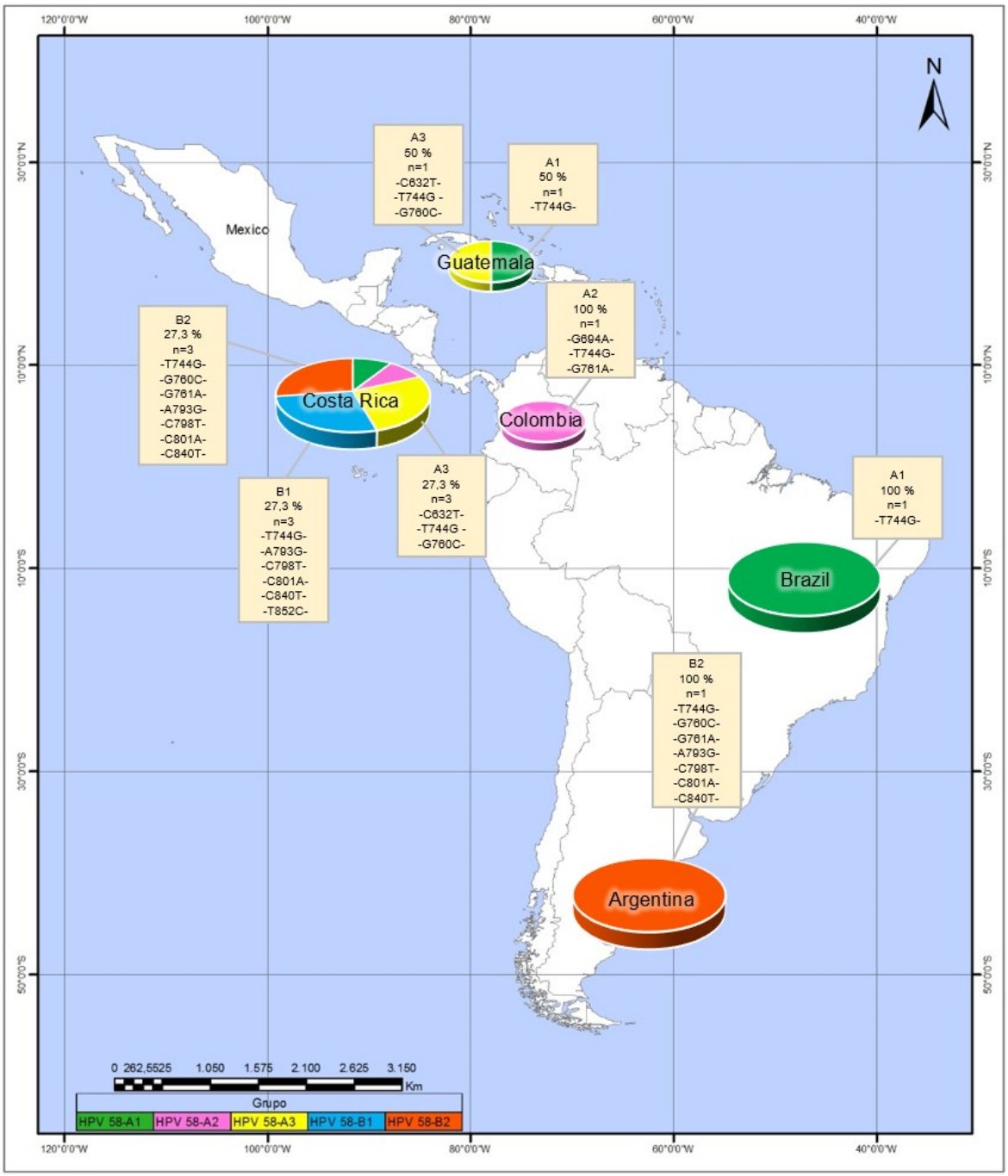


Figura 37. Distribución geográfica y frecuencia (%) de variantes de E7 VPH 58 identificadas para Latinoamérica, construido con el programa ArcGis 10.2, donde, A1, A2, A3, B1 y B2 representan los grupos de variantes. *Fuente*: esta investigación.

En síntesis, cada una de las variantes descritas en este trabajo representa el esfuerzo que se ha realizado para caracterizar la diversidad genética de VPH en Latinoamérica, que si bien no se comparan en número con los avances que proceden de otras regiones geográficas, como Asía y Europa, durante el proceso de esta investigación se identificaron nuevas mutaciones puntuales en tipos específicos del virus, importante para desarrollar un mejor pronóstico de la enfermedad, debido a que algunas de estas alteraciones ocurren en residuos de dominios funcionales, que pueden modular negativamente o positivamente en la mayoría de los casos la persistencia de la infección, la morbilidad de la carcinogénesis y la progresión de lesiones precursoras del cáncer.

Por otra parte, se hace hincapié en que se realicen estudios funcionales a nivel *in vitro,* para validar la función de cada mutación y la combinación de estas sobre la proteína mutada y el contexto del ciclo viral, sobre todo en genotipos diferentes a VPH 16 y VPH 18, donde los estudios son aún más limitados. Por ejemplo, la documentación de VPH 58 y VPH 45 es escasa, a pesar de que se conoce muy bien que en algunas zonas la región de América latina la prevalencia de estos tipos tiende a igualar la frecuencia de los genotipos VPH 16 y 18 designados en su momento como de mayor frecuencia, lo que hace pensar en si se está dando en Latinoamérica un remplazamiento de genotipos, como alguna vez se observó en algunas zonas del continente africano, donde VPH 31 iguala y hasta puede superar en frecuencia a VPH 16.

Es fundamentalmente importante establecer bases de datos sobre las variantes del VPH y la diversidad genómica intratipo de sus secuencias para (i) desarrollar y herramientas de diagnóstico y vacunas eficaces, (ii) estudios epidemiológicos debido a que variaciones específicas de nucleótidos pueden ser utilizadas como marcadores para analizar el seguimiento de la transmisión en una población, (iii) analizar la relación entre en genotipo y el fenotipo (ciertas variantes pueden estar asociadas positiva o negativamente con la enfermedad cervical, dependiendo de sus diferentes propiedades biológicas y funcionales) y (iv) los estudios evolutivos y taxonómicos, como el que se pretendió aquí al identificar grupos de variantes de acuerdo a mutaciones puntuales recurrentes.

CONCLUSIONES

El 85 % de las secuencias analizadas fueron provenientes de Costa Rica, Brasil y México donde se pudo determinar la mayor diversidad de genotipos.

De los 22 países que conforman América latina, Brasil aportó el mayor número de secuencias de VPH -BR 6 y 11, y permitió identificar la gran variabilidad de estos genotipos distribuidos a nivel mundial, lo que evidenció la necesidad de ampliar los estudios de secuenciación de VPH en la región.

Se construyeron filogenias altamente informativas, que relacionaron la historia evolutiva de los tipos analizados y el potencial oncogénico de los mismos, usando diferentes tamaños de fragmento del gen L1; sin embargo, las comparaciones deben realizarse con fragmentos de tamaños equivalentes y provenir de regiones altamente informativas de parsimonia.

Si bien la variabilidad observada de los genotipos incluidos en el análisis es menor o igual a la que se ha presentado en otros continentes, la presencia de ciertas mutaciones puntuales y su relación con la severidad de la lesión de la cual provenían no fue concluyente.

A partir del análisis de variantes del gen E6, para Brasil se describieron las siguientes nuevas mutaciones puntuales respecto al tipo viral: A538G en VPH 11, C287G en VPH 18, A124C, 259T/A, A284C, T482C, G487A y A497G en VPH 45. Para el gen E7 fueron: G600A, A603C, A718C, G808T y G822T en VPH 45, estas podrían servir para el diseño de nuevas metodologías diagnósticas.

Para el gen E6 en Latinoamérica se describieron las siguientes nuevas mutaciones puntuales: A449T en VPH 33, A136G, T143C, G193A y A260C en VPH 58. Para el gen E7 fueron: A681G en VPH 52, A587G, A603T, A614G y T778C en VPH 58.

Se identificó la presencia en Latinoamérica de las sustituciones de aminoácidos C287G, N92S/K (VPH 18), L83V (VPH 16), K62E/D (VPH 31), E14D (VPH 52) T20I y G63H/S (VPH 58), correlacionadas positivamente con la persistencia del virus y la NIC grado II y III, que podrían conferir a estos genotipos ventajas evolutivas que el permitan escapar de la neutralización efectiva generada por las vacunas. La identificación de variantes donde se presenten dichas sustituciones significaría discriminar entre pacientes con un mayor riesgo de desarrollar CCU.

RECOMENDACIONES

Enfatizar en la secuencias de genomas completos de VPH en Latinoamérica, para así realizar un análisis que permita identificar en su totalidad la variabilidad genética, identificar las variantes para todos los ORFs en el genoma, más la región no codificante- LCR y clasificar por sistemas de variantes, para a su vez correlacionar con los factores demográficos y epidemiológicos que caracterizan a la población de donde proceden los diferentes aislamiento, con el fin de identificar el mayor número de factores que conlleven a CCU.

Es recomendable ampliar esta investigación, con un mayor número de secuencias, sobre todo de países de la región caribe, cuyas tasas de prevalencia del virus y mortalidad por cáncer igualan el promedio mundial y de los cuales poco o nada se conoce de la variabilidad y genotipos exhibidos para esta zona.

Se debería establecer el monitoreo permanente de diversidad genómica de los VPH como una estrategia de seguimiento a nuevas variantes, tipos o subtipos virales con el fin de actualizar permanentemente las técnicas de detección molecular.

Se recomienda ampliar las investigaciones a nivel local, con el fin de conocer la variabilidad genética de los tipos que convergen en la zona y una vez identificadas las variantes, enfatizar y precisar el efecto de las mutaciones en las lesiones del CCU.

LITERATURA CITADA

ACS. (2014). American Cancer Society. Disponible en: http://www.cancer.org/

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., … & Venter, J. C. (2000). The genome sequence of Drosophila melanogaster. Science, 287(5461), 2185-2195.

Aho, J., Hankins, C., Tremblay, C., Lang, F., Forest, P., Pourreaux, K., … & Canadian Women's HIV Study Group. (2003). Molecular analysis of human papillomavirus type 52 isolates detected in the genital tract of human immunodeficiency virus-seropositive and -seronegative women. Journal of Infectious Diseases, 188(10), 1517-1527.

Almeida, A., Medeiros do Amaral, C., Soares, I., Cartaxo, M., Chagas, B., Bezerra, M., … & de Freitas, A. (2013). Prevalence and Genetic Variability in Capsid L1 Gene of Rare Human Papillomaviruses (HPV) Found in Cervical Lesions of Women from North-East Brazil. BioMed Research International, 2013 (546354), 1-7.

Ahmed, A., Bissett, S., & Beddows, S. (2013). Amino acid sequence diversity of the major human papillomavirus capsid protein: Implications for current and next generation vaccines. Infection, Genetics and Evolution, 18, 151-159.

Arias-Pulido, H., Peyton, C. L., Torrez-Martínez, N., Anderson, D. N., & Wheeler, C. M. (2005). Human papillomavirus type 18 variant lineages in United States populations characterized by sequence analysis of LCR-E6, E2, and L1 regions. Virology, 338(1), 22-34.

Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Research, 28(1), 45-48.

Baldez Da Silva, M. F., Chagas, B. S., Guimar, V., Kats, L., Feliz, P., Miranda, P. M., … & Beçak, W. (2009). HPV31 and HPV33 incidence in cervical samples fromwomen in Recife, Brazil". Genetics and Molecular Research, 8(4), 1437-1443.

Barbosa, M. S., Edmonds, C., Fisher, C., Schiller, J. T., Lowy, D. R., & Vousden, K. H. (1990). The region of the HPV E7 oncoprotein homologous to adenovirus E1a and Sv40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. EMBO Journal, 9(1), 153-160.

Barton, G. J., & Sternberg, M. J. (1987). A strategy for the rapid multiple alignment of protein sequences: Confidence levels from tertiary structure comparisons. Journal of Molecular Biology, 198(2), 327-337.

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., & Sonnhammer, E. L. (2002). The Pfam protein families database. Nucleic Acids Research, 30(1), 276-280.

Batista, M. V., Ferreira, T. A., Freitas, A. C., & Balbino, V. Q. (2011). An entropy-based approach for the identification of phylogenetically informative genomic regions of Papillomavirus. Infection, Genetics and Evolution, 11(8), 2026-2033.

Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2014). GenBank. Nucleic Acids Research, 42 (1), D32-D37.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A., & Wheeler, D. L. (2002). GenBank. Nucleic Acids Research, 30(1), 17-20.

Bernard, H. U., Calleja-Macias, I., & Terence, S. (2006). Genome variation of human papillomavirus types: Phylogenetic and medical implications. International Journal of Cancer, 118(5), 1071-1076.

Bernard, H. U., Burk, R., Chen, Z., van Doorslaer, K., Hausen, H., & de Villiers, E. (2010). Classification of papillomaviruses (HPVs) based on 189 PV types and proposal of taxonomic amendments. Virology, 401(1), 70-79.

Bernard, H. U., Chan, S. Y., Manos, M. M., Ong, C. K., Villa, L. L., Delius, H., Peyton, C. L., … & Wheeler, C. M. (1994). Identification and Assessment Of Known And Novel Human Papillomaviruses by Polymerase Chain Reaction Amplification, Restriction Fragment Length Polymorphisms, Nucleotide Sequence, and Phylogenetic Algorithms. The Journal of Infectious Diseases, 170(5), 1077-1085.

Berumen, J., Ordoñez, R. M., Lazcano, E., Salmeron, J., Galvan, S. C., Estrada, R. A., … & Madrigal-de la Campa, A. (2001). Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study. Journal of the National Cancer Institute, 93(17), 1325-1330.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., & Riley, M. (1997). The complete genome sequence of Escherichia coli K-12. Science, 277(5331), 1453-1462.

Boulet, G., Horvath, C., Vanden, D., Sahebali, S., & Bogers, J. (2007). Human papillomavirus: E6 and E7 oncogenes. The International Journal of Biochemistry & Cell Biology, 39(11), 2006-2011.

Bravo, I., & Alonso, A. (2004). Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malig nant or benign growth. Journal of Virology, 78(24), 13613-13626.

Bravo, I., de Sanjose, S., & Gottschling, M. (2010). The clinical importance of understanding the evolution of papillomaviruses. Trends in Microbiology, 18(10), 432-438.

Brebi, P., Ili, C. G., López, J., García, P., Melo, A., Montenegro, S., … & Roa, J. C. (2009). Detección y tipificación de virus papiloma humano en adenocarcinoma de cuello uterino

mediante reverse line blot, Región de La Araucanía, Chile. Revista Médica de Chile, 137(3), 377-382.

Brokaw, J. L., Yee, C. L., & Münger, K. (1994). A mutational analysis of the amino terminal domain of the human papillomavirus type 16 E7 oncoprotein. Virology, 205(2), 603-607.

Brown, T. A. (2002). Genomes. 2nd edition. Oxford: Wiley-Liss; Disponible en: http://www.ncbi.nlm.nih.gov/books/NBK21128/

Buitrago, O., Morales, N., Martín, C., Huertas, A., Moreno, P., Martínez, T., … & Molano, M. (2012). Presencia y persistencia de variantes moleculares del ORF E6 y ORF E7 del VPH 58 en mujeres con citología normal que pertenecen a la cohorte de Bogotá, Colombia. Revista Colombiana de Cancerología, 16(1), 40-48.

Burd, E. (2003). Human papillomavirus and cervical cancer. Clinical Microbiology Reviews, 16(1), 1-17.

IACR. (2014). International Association of Cancer Registries. Disponible en: http://ci5.iarc.fr/

Burk, R., Chen, Z., Harari, A., Smith, B., Kocjan, B., Maver, P., … & Poljak, M. (2011). Classification and nomenclature system. Classification and nomenclature system for Human Alphapapillomavirus variants: general features, nucleotide landmarks and assignment of HPV6 and HPV11 isolates to variant lineages. Acta Dermatovenerol Alp Panonica Adriat, 20(3), 113-123.

Burk, R., Harari, A., & Chen, Z. (2013). Human papillomavirus genome variants. Virology, 445(1-2), 232-243.

Cai, Y. P., Yang, Y., Zhu, B. L., Li, Y., Xia, X. Y., Zhang, R. F., … & Xiang, Y. (2013). Comparison of human papillomavirus detection and genotyping with four different prime sets by PCR-sequencing. Biomedical and Environmental Sciences, 26(1), 40-47.

Calleja-Macias, I. E., Kalantari, M., Huh, J., Ortiz-Lopez, R., Rojas-Martinez, A., Gonzalez-Guerrero., J … & Barrera-Saldaña, M. (2004). Genomic diversity of human papillomavirus-16, 18, 31, and 35 isolates in a Mexican population and relationship to European, African, and Native American variants. Virology, 319(2), 315-323.

Calleja-Macias, I. E., Villa, L. L., Prado, J. C., Kalantari, M., Allan, B., Williamson, A. L., … & Bernard, H. U. (2005).Worldwide genomic diversity of the high-risk human papillomavirus types 31, 35, 52, and 58, four close relatives of human papillomavirus type 16. Journal of Virology, 79(21), 13630-13640.

Camara, G., Cerqueira, M., Oliveira, A., O Silva, E., Carvalho, L., & Martins, C. (2003). Prevalence of Human Papillomavirus Types in Women with Pre-neoplastic and Neoplastic Cervical Lesions in the Federal District of Brazil. Mem Inst Oswaldo Cruz, Rio de Janeiro, 98(7), 879-883.

Canul, J., Rosado, I., Suárez, N., Colli, G., Conde-Ferráez, L., Canto de Cetina, T. … & González, R. (2010). High prevalence and low E6 genetic variabilityof human papillomavirus 58 in women with cervical cancer and precursor lesions in Southeast Mexico. Mem Inst Oswaldo Cruz, Rio de Janeiro, 105(2), 144-148.

Castellsagué, X., Ghaffari, A., Daniel, R. W., Bosch, F. X., Muñoz, N., & Shah, K. V. (1997). Prevalence of penile human papillomavirus DNA in husbands of women with and without cervical neoplasia: a study in Spain and Colombia. The Journal of Infectious Diseases, 176(2), 353-361.

Castellsagué, X. (2008). Natural history and epidemiology of HPV infection and cervical cancer. Gynecologiconcology, 110(3-2), 4-7.

Castro, M. M., Farias, I. P., Borborema-Santos, C. M., Correia, G., & Astolfi-Filho, S. (2011). Prevalence of human papillomavirus (HPV) type 16 variants and rare HPV types in the central Amazon region. Genet Mol Res, 10(1), 186-196.

Cavalli-Sforza, L. L., & Edwards, A. W. (1967). Phylogenetic analysis. Models and estimation procedures. The American Journal of Human Genetics, 19(3-1), 233-257.

Cerón, C. (2013). Estrategia de vacunación contra el VPH. Revista Universidad y Salud, 15(1), 5.

Cerqueira, D. M., Camara, G. N., da Cruz, M. R., Silva, E. O., Brígido, M. Carvalho, L. G., … & Martins, C. R. (2003). Variants of human papillomavirus types 53, 58 and 66 identified in Central Brazil. Virus Genes, 26(1), 83-87.

Cerqueira, D. M., de S Moraes, D., Camara, G. N., Amaral, F. A., Oyama, C. N., dos Santos, M. Q., … & Martins, C. R. (2007). High HPV genetic diversity in women infected with HIV-1 in Brazil. Archives of Virology, 152(1), 75-83.

Cerqueira, D. M., Raiol, T., Véras, N. M., von Gal, N., Amaral, F. A., de Macedo, M., … & Martins, C. R. (2008). New variants of human papillomavirus type 18 identified in central Brazil. Virus Genes, 37(2), 282-287.

Chagas, B. S., Batista, M. V., Guimarães, V., Balbino, V. Q., Crovella, S., & Freitas. A. C. (2011). New variants of E6 and E7 oncogenes of human papillomavirus type 31 identified in Northeastern Brazil. Gynecologic Oncology, 123(2), 284-288.

Chagas, B. S., Batista, M. V., Crovella, S., Gurgel, A. P., Silva, Neto, J., … & Freitas, A. C. (2013). Novel E6 and E7 oncogenes variants of human papillomavirus type 31 in Brazilian women with abnormal cervicalcytology. Infection Genetics and Evolution, 16, 13-8.

Chan, S. Y., Delius, H., Halpern, A. L., & Bernard, H. U. (1995). Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy. Journal of Virology, 69(5), 3074-3083.

Chan, S. Y., Ho, L., Ong, C., Chow, V., Drescher, B., Dürst, M., … & Mgaya, H. N. (1992). Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. Journal of Virology, 66(4): 2057-2066.

Chan, P. K., Luk, A. C., Park, J. S., Smith-McCune, K. K., Palefsky, J. M., Konno, R., … & Banks, L. (2011). Identification of human papillomavirus type 58 lineages and the distribution worldwide. Journal of Infectious Diseases, 203(11), 1565-1573.

Chan, P. K. S., Mak, K. H., Cheung, J. L. K., Tang, N. L. S., Chan, D. P. C., Lo, K. K., & Cheng, A. F. (2002). Genotype spectrum of cervical human papillomavirus infection among sexually transmitted disease clinic patients in Hong Kong. Journal of Medical Virology, 68 (9), 273-277.

Chang, C. H., Chen, T. H., Hsu, R. C., Chou, P. H., Yang, J. J., & Hwang, G. Y. (2005). The prevalence of HPV-18 and variants of E6 gene isolated from cervical cancer patients in Taiwan. Journal of Clinical Virology, 32(1), 33-37.

Chan, P. K., Zhang, C., Park, J. S., Smith-McCune, K. K., Palefsky, J. M., Giovannelli, L., … & Banks, L. (2013). Geographical distribution and oncogenic risk association of human papillomavirus type 58 E6 and E7 sequencevariations. International Journal of Cancer, 132(11), 2528-2536.

Chen, E., Howley, P., Levinson, A., & Seeburq, P. (1982). The primary structure and genetic organization of the bovine papillomavirus type 1 genome. Nature, 299(5883), 529-534.

Chen, Z., DeSalle, R., Schiffman, M., Herrero, R., & Burk, R. D. (2009). Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97. Journal of Virology. 83(3), 1443-1455.

Chen, A. A., Heideman, D. A., Boon, D., Gheit, T., Snijders, P. J., Tommasino, M., … & IARC HPV Variant Study Group.(2014). Human papillomavirus 45 genetic variation and cervical cancer risk worldwide. Journal of Virology, 88(8), 4514-4521.

Chen, A. A., Heideman, D. A., Boon, D., Chen, Z., Burk, R. D., … & Clifford, G. M. (2014). Human papillomavirus 33 worldwide genetic variation and associated risk of cervical cancer. Virology, 448, 356-362.

Chen, Z., Fu, L., Herrero, R., Schiffman, M., & Burk, R. D. (2007). Identification of a novel human papillomavirus (HPV97) related to HPV18 and HPV45. International Journal of Cancer, 121, 2947-2952.

Chen, Z., Schiffman, M., Herrero, R., Desalle, R., Anastos, K., Segondy, M., … & Burk, R. D. (2011). Evolution and taxonomic classification of human papillomavirus 16 (HPV16)-related variant genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. PLoS One, 6(5), e20183.

Chen, Z., Schiffman, M., Herrero, R., DeSalle, R., Anastos, K., Segondy, M., … & Burk, R. D. (2013). Evolution and taxonomic classification of alphapapillomavirus 7 complete genomes: HPV18, HPV39, HPV45, HPV59, HPV68 and HPV70. PLoS One, 8(8), e72565.

Chen, Z., Terai, M., Fu, L., Herrero, R., DeSalle, R., & Burk, R. D. (2005). Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. Journal of Virology, 79(11), 7014-7023.

Chen, Q., Luo, Z. Y., Lin, M., Yang, L., Yang, L., & Ju, G. Z. (2012).Evaluation of the genetic variability of human papillomavirus type 52. International Journal of Molecular Medicine, 30(3), 535-544.

Choi, K., & Gomez, S. (2009). Comparison of phylogenetic trees through alignment of embedded evolutionary distances. BMC Bioinformatics, 10 (423), 1-15.

CI5plus. (2014). CI5: Cancer Incidence in Five Continents. Disponible en: http://ci5.iarc.fr/CI5plus/Default.aspx

Clifford, G., & Franceschi, S. (2008). Members of the human papillomavirus type 18 family (alpha-7 species) share a common association with adenocarcinoma of the cérvix. International Journal of Cancer, 122(7), 1684-1685.

Clifford, G. M., Gallus, S., Herrero, R., Muñoz, N., Snijders, P. J., Vaccarella, S., … & IARC HPV Prevalence Surveys Study Group. (2005). Worldwide distribution of human papillomavirus types in cytologically normal women in the International Agency for Research on Cancer HPV prevalence surveys: a pooled analysis. Lancet, 366(9490), 991-998.

Clifford, G. M., Smith, J.S., Aguado, A., & Franceschi, S. (2003). Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: a meta-analysis. British Journal of Cancer, 89, 101-105.

Cline, M. (2000). Protein sequence alignment reliability: prediction and measurement [Tesis de Doctorado]. California (EE. UU.). University of California.

Cogliano, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F. & WHO International Agency for Research on Cancer. (2005). Carcinogenicity of human papillomaviruses. The Lancet Oncology, 6(4), 204.

Cole, S. & Danos, O. (1987). Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome: Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. Journal of Molecular Biology, 193(4), 599-608.

Cole, S., & Streeck, R. (1986). Genome organization and nucleotide sequence of human papillomavirus type 33, which is associated with cervical cancer. Journal of Virology, 58(3), 991-995.

Cornet, I., Gheita, T., Franceschia, S., Vignata, J., Burkb, R., Syllaa, B., … & Wabinga, H. (2012). Human Papillomavirus Type 16 Genetic Variants: Phylogeny and Classification Based on E6 and LCR. Journal of Virology, 86(12), 6855-6861.

Crook, T., Tidy, J. A., & Vousden, K. H. (1991). Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation. Cell, 67(3), 547-556.

Cruz, M. R., Cerqueira, D. M., Cruz, W. B., Camara, G. N., Brigido, M. M., Silva, E. O., … & Martins, C. R. (2004). Prevalence of human papillomavirus type 16 variants inthe Federal District, Central Brazil. Mem. Inst. Oswaldo Cruz, 9 (3), 281-282.

Cruz, J., Márquez, Quintero, M., Bastidas, M., L., & Puig-Pons, J. (2013). Estudio de variantes intra-tipo del virus del papiloma humano tipo16, por análisis nucleotídico de la región MY09-MY11. Revista de Obstetricia y Ginecología Venezolana, 73(3), 187-194.

Danielewski, J. A., Garland, S. M., McCloskey, J., Hillman, R. J., & Tabrizi, S. N. (2013). Human Papillomavirus Type 6 and 11 Genetic Variants Found in 71 Oral and Anogenital Epithelial Samples from Australia. PLoS ONE, 8(5), e63892.

Danos, O., Katinka, M., & Yaniv, M. (1982). Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. The EMBO Journal, 1(2), 231-236.

Dartmann, K., Schwarz, E., Gissmann, L., & zur Hausen, H. (1986). The nucleotide sequence and genome organization of human papilloma virus type 11. Virology, 151(1), 124-130.

de Boer, M. A., Peters, L. A., Aziz, M. F., Siregar, B., Cornain, S., Vrede, M. A., … & Fleuren, G. J. (2004). Human papillomavirus type 16 E6, E7, and L1 variants in cervical cancer in Indonesia, Suriname, and The Netherlands. Gynecologic Oncology, 94(2), 488-494.

DeFilippis, R.A., Goodwin, E.C., Wu, L., & DiMaio, D. (2003). Endogenous human papillomavirus E6 and E7 proteins differentially regulate proliferation, senescence, and apoptosis in HeLa cervical carcinoma cells. Journal of Virology, 77 (2), 1551-1563.

de Matos, R. P., Sichero, L., Mansur, I. M., do Bonfim, C. M., Bittar, C., Nogueira, R. L., ... & Rahal, P. (2013). Nucleotide and phylogenetic analysis of human papillomavirus types 6 and 11 isolated from recurrent respiratorypapillomatosis in Brazil. Infection, Genetics and Evolution, 16 (1), 282-289.

de Sanjosé, S., Diaz, M., Castellsagué, X., Clifford, G., Bruni, L., Muñoz, N., … & Bosch, F. X. (2007). Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. Lancet Infectious Diseases, 7(7), 453-459.

de Villiers, E. M. (2013). Cross-roads in the classification of papillomaviruses. Virology, 445(1-2), 2-10.

de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U., & Zur Hausen, H. (2004). Classification of papillomaviruses. Virology, 324(1), 17-27.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics, 6(5), 361-375.

Doorbar, J. (2005). The papillomavirus life cycle. Journal of Clinical Virology, 32, 7-15.

Doorbar, J. (2006). Molecular biology of human papillomavirus infection and cervical cancer. Clinical Science (Lond), 110(5), 525-541.

Duc, P. (2012). Event-based Phylogeny Inference and Multiple Sequence Alignment [Tesis de Maestría]. Providence, Rhode Island. Brown University.

Ferenczy, A., & Franco, E. (2002). Persistent human papillomavirus infection and cervical neoplasia. Lancet Oncology, 3 (1), 11-16.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using bootstrap. Evolution, 39(4), 783-791.

Fernandes, J., Galvao, J., & Araújo, T. (2013). Biology and natural history of human papillomavirus infection. Open Access Journal of Clinical Trials, 3(5), 1-12.

Foster, S. A., Demers, G. W., Etscheid, B. G., & Galloway, D. A. (1994). The ability of human papillomavirus E6 proteins to target p53 for degradation in vivo correlates with their ability to abrogate actinomycin D-induced growth arrest. Journal of Virology, 68(9), 5698-5705.

Frattini, M., Lim, H., & Laimins, L. (1996). In vitro synthesis of oncogenic human papillomaviruses requires episomal genomes for differentiation-dependent late expression. PNAS. 93(7), 3062-3067.

García-Vallvé, S., Alonso, A., & Bravo, I. G. (2005). Papillomaviruses: different genes have different histories. Trends Microbiology, 13(11), 514-521.

Giarrè, M., Caldeira, S., Malanchi, I., Ciccolini, F., Leão, M. J., & Tommasino, M. (2001). Induction of pRb degradation by the human papillomavirus type 16 E7 protein is essential to efficiently overcome p16INK4a-imposed G1 cell cycle Arrest. Journal of Virology, 75(10), 4705-4712.

Godínez, J. M., Heideman, D. A. M., Gheit, T., Alemany, L., Snijders, P. J. F., … & Bravo, I. G. (2013). Differential presence of papillomavirus variants in cervical cancer: an analysis for HPV33, HPV45 and HPV58. Infection, Genetics and Evolutión, 13 (4), 96-104.

Goldsborough, M. D., DiSilvestre, D., Temple, G. F., & Lorincz, A. T. (1989). Nucleotide sequence of human papillomavirus type 31: a cervical neoplasia-associated virus. Virology, 171(1), 306-311.

González-Losa, M., Mier, L., Teran, M. A, Puerto-Solís, M., & García-Carrancá, A. (2004). Molecular variants of HPV type 16 E6 among Mexican women with LSIL and invasive cancer. Journal of Clinical Virology, 29(2), 95-98.

Gottschling, M., Göker, M., Stamatakis, A., Bininda-Emonds, O. R., Nindl, I., & Bravo, I. G. (2011). Quantifying the phylodynamic forces driving papillomavirus evolution. Molecular Biology and Evolution, 28(7), 2101-2113.

Gottschling, M., Köhler, A., StockXeth, E., & Nindl, I. (2007). Phylogenetic analysis of beta-papillomaviruses as inferred from nucleotide and amino acid sequence data. Molecular Phylogenetics and Evolution, 42 (3), 213-222.

Grievink, L. (2009). Lineage specific evolution and phylogenetics analysis [Tesis de Doctorado]. Palmerston (New Zealand). Massey university.

Hall, T. (2011). BioEdit: biological sequence alignment editor written for Windows 95/98/NT/2000/XP/7. Version 7.2.5. Disponible en: http://www.mbio.ncsu.edu/bioedit/bioedit.html

Heinzel, P., Chan, S., Ho, L., O'Connor, M., Balaram, P., … & Pfister, H. (1995). Variation of human papillomavirus type 6 (HPV-6) and HPV-11 genomes sampled throughout the world. Journal of Clinical Microbiology, 33(7), 1746-1754.

Ho L, Chan S. Y., Burk R. D., Das, B C., Fujinaga, K., Icenogle, J., … & Mavromara-Nazos, P. (1993). The Genetic Drift of Human Papillomavirus Type 16 Is a Means of Reconstructing Prehistoric Viral Spread and the Movement of Ancient Human Populations. Journal of Virology, 67(11), 6413-6423.

Hofmann, K. J., Cook, J. C., Joyce, J. G., Brown, D. R., Schultz, L. D., George, H. A., … & Jansen, K. U. (1995). Sequence determination of human papillomavirus type 6a and assembly of virus-like particles in Saccharomyces cerevisiae. Virology, 209(2), 506-518.

Howie, H. L., Katzenellenbogen, R. A., & Galloway, D. A. (2009). Papillomavirus E6 proteins. Virology, 384(2), 324-334.

Huala, E., Dickerman, A. W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., … & Rhee, S. Y. (2001). The Arabidopsis information resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Research, 29(1), 102-105.

Huertas-Salgado, A., Martín-Gámez, D. C., Moreno, P., Murillo, R., Bravo, M. M., Villa, L., … & Molano, M. (2011). E6 molecular variants of human papillomavirus (HPV) type 16: an updated and unified criterion for clustering and nomenclature. Virology, 410(1), 201-215.

IACR. (2014). International Association of Cancer Registries. Disponible en: http://www.iacr.com.fr

Jelen, M. M., Chen, Z. Kocjan, B., Burt, F., Chan, P. K. Chouhy, D., … & Poljak, M. (2014). Global genomic diversity of human papillomavirus 6 based on 724 isolates and 190 complete genome sequences. Journal of Virology, 88(13), 7307-7316.

Jiang, P., & Yue, Y. (2013). Human papillomavirus oncoproteins and apoptosis (Review). Experimental and Therapeutic Medicine, 7(1), 3-7.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., … & Madden, T. (2008). NCBI BLAST: a better web interface. Nucleic Acids Research, 36(2), W5-W9.

Junes-Gill, K., Noronha, V., Sichero, L., Villa, L. L., Maciag, P., & Mello, W. (2008). Human Papillomavirus Type 16 Variants in Cervical Cancer From an Admixtured Population in Brazil. Journal of Medical Virology, 80 (14), 1639-1645.

Kanehisa, M., & Bork, P. (2003). Bioinformatics in the post-sequence era. Nature Genetics, 33, 305-310.

Kirii, Y., Iwamoto, S., & Matsukura, T. (1991). Human papillomavirus type 58 DNA sequence. Virology, 185(1), 424-427.

Kocjan, B. J., Jelen, M. M., Maver, P. J., Seme, K., & Poljak, M. (2011). Pre-vaccination genomic diversity of human papillomavirus genotype 6 (HPV 6): a comparative analysis of 21 full-length genome sequences. Infection, Genetics and Evolution, 11(7), 1805-1810.

Kocjan, B. J., Poljak, M., Cimerman, M., Gale, N., Potocnik, M., Bogovac, Z., … & Seme, K. (2009). Prevaccination genomic diversity of human papillomavirus genotype 6 (HPV 6). Virology, 391(2), 274-283.

Kovelman, R., Bilter, G. K., Roman, A., Brown, D. R., & Barbosa, M. S. (1999). Human papillomavirus type 6: classification of clinical isolates and functional analysis of E2 proteins. Journal of General Virology, 80(9), 2445-2451.

Kurvinen, K., Yliskoski, M., Saarikoski, S., Syrjänen, K., & Syrjänen, S. (2000). Variants of the long control region of human papillomavirus type 16. European Journal of Cancer, 36(11), 1402-1410.

Lacey, C. J., Lowndes, C. M., & Shah, K. V. (2006). Chapter 4: Burden and management of non-cancerous HPV-related conditions: HPV-6/11 disease. Vaccine, 3, 35-41.

Landry, M. E., Salit, E. E., Rodrigues-Coutlée, C., Money, D., Rebbapragada, A., Tinmouth, J., … & Coutlée, F. (2012). Ano-genital human papillomavirus type 97 infection is detected in Canadian men but not women at risk or infected with the human immunodeficiency virus. Virology Journal, 9(243) 1-4.

Letian, T., & Tianyu, Z. (2010). Cellular receptor binding and entry of human papillomavirus. Virology Journal, 7(2), 1-7.

Li, Y., Li, Z., He, Y., Kang, Y., Zhang, X., Cheng, M., … & Xu, C. (2009). Phylogeographic analysis of human papillomavirus 58. Science in China Series C, 52(12), 1164-1172.

Li, N., Franceschi, S., Howell-Jones, R., Snijders, P. J., & Clifford, G. M. (2011). Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. International journal of cancer, 128(4), 927-935.

Liu, X., Clements, A., Zhao, K., & Marmorstein, R. (2006). Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. Journal of Biological Chemistry, 281(1), 578-586.

Liu, Y., & Baleja, J. D. (2008). Structure and function of the papillomavirus E6 protein and its interacting proteins. Frontiers in Bioscience, 13, 121-134.

Liu, M., He, Z., Xi, L., Li, J., Liu, F., Liu, Y., … & Ke, Y. (2014) The Distribution and Common Amino Acid Polymorphisms of Human Papillomavirus (HPV)-31 Variants in 2700 Women from Northern China. PLoS ONE, 9(6), e99141.

Liu, M., He, Z., Xi, L., Li, J., Liu, F., Liu, Y., … & Ke, Y. (2014). The Distribution and Common Amino Acid Polymorphisms of Human Papillomavirus (HPV)-31 Variants in 2700 Women from Northern China. PLoS ONE, 9(6), e99141.

Lizano, M., De la Cruz-Hernández, E., Carrillo-García, A., García-Carrancá, A., Ponce de Leon-Rosales, S., Dueñas-González, A., … & Mohar A. (2006). Distribution of HPV16 and 18 intratypic variants in normal cytology, intraepithelial lesions, and cervical cancer in a Mexican population. Gynecologic Oncology, 102(2), 230-235.

Lizano-Soberón, M., Carrillo-García, A., & Contreras-Paredes, A. (2009). Infección por virus del Papiloma Humano: Epidemiología, Historia Natural y Carcinogénesis. Cancerología, 4, 205-216.

López-Revilla, R., Pineda, M. A., Ortiz-Valdez, J., Sánchez-Garza, M., & Riego, L. (2009). Human papillomavirus type 16 variants in cervical intraepithelial neoplasia and invasive carcinoma in San LuisPotosí City, Mexico. Infect Agent Cancer, 4(3), 1-9.

Luciani, S., Cabanes, A., Prieto-Lara, E., & Gawryszewski, V. (2013). Cervical and female breast cancers in the Americas: current situation and opportunities for action. Bulletin of the World Health Organization, 91, 640-649.

Luciani, S., & Andrus, J. K. (2008) A Pan American Health Organization strategy for cervical cancer prevention and control in Latin America and the Caribbean. Reproductive Health Matters, 16(32), 59-66.

Marich, J. E., Pontsler, A. V., Rice, S. M., McGraw, K. A., & Dubensky, T. W. (1992). The phylogenetic relationship and complete nucleotide sequence of human papillomavirus type 35. Virology, 186(2), 770-776.

Massimi, P., Gammoh, N., Thomas, M., & Banks, L. (2004). HPV E6 specifically targets different cellular pools of its PDZ domain containing tumour suppressor substrates for proteosome-mediated degradation. Oncogene, 23, 8033-8039.

Mazarico, E., & Gonzalez-Bosquet, E. (2012). Prevalence of infection by different genotypes of human papillomavirus in women with cervical pathology. Gynecologic Oncology, 125(1), 181-185.

McLaughlin-Drubin, M. E., & Munger, K. (2009). The human papillomavirus E7 oncoprotein. Virology, 384(2), 335-344.

Melo, A., Montenegro, S., Hooper, T., Capurro, I., Roa, J., & Roa, I. (2003). Tipificación del virus papiloma humano (VPH) en lesiones preneoplásicas y carcinoma del cuello uterino en mujeres de la IX Región-Chile. Revista Médica de Chile, 131(12), 1382-1390.

Mijares, A., Suárez, C., Castro, R., Agudo, A., Pérez, C., & Fuentes, C. (2007). Tipificación del virus de papiloma humano relación con el carcinoma de cavidad oral. Revista Venezolana de Oncología, 19(4), 321-331.

Molano, M., Posso, H., Weiderpass, E., Van Den Brule, A. J. C., Ronderos, M., Franceschi, S., … & Munoz, N. (2002). Prevalence and determinants of HPV infection among Colombian women with normal cytology. British Journal of Cancer, 87(3), 324-333.

Moreno-Acosta, P., Molano, M., Huertas, A., de Gomez, M.S., Romero, A., Gonzalez, M., … & Garcia-Carranca, A. (2008). A non-radioactive PCR-SSCP analysis allows to distinguish between HPV 16 European and Asian-American variants in squamous cell carcinomas of the uterine cervix in Colombia. Virus Genes, 37, 22-30.

Muñoz, M., Camargo, M., Soto de León, S. C., Sanchez, R., Parra, D., Pineda, A. C., … & Patarroyo, M. A. (2013). Human Papillomavirus Detection from Human Immunodeficiency Virus-Infected Colombian Women's Paired Urine and Cervical Samples. PLoS ONE, 8(2), e56509.

Muñoz, N., Méndez, F., Posso, H., Molano, M., van den Brule, A. J., Ronderos, M, … & Muñoz, A. (2004). Incidence, duration, and determinants of cervical human papillomavirus infection in a cohort of Colombian women with normal cytological results. The Journal of Infectious Diseases, 190(12), 2077-2087.

Muñoz, M., Mendoza, J., Téllez, L., Noguera, M., Moret, O., López, M., … & Toro, M. (2003). Detección de VPH-16 y 18 en muestras de cérvix de mujeres que acuden a centros asistenciales de la ciudad de Mérida, Venezuela. Revista Biomédica, 14(2), 61-8.

Murillo, R. (2006). Vacunación contra el virus del papiloma humano en Colombia. Revista Colombiana de Cancerología, 10(2), 85-97.

Murillo, R., Almonte, M., Pereira, A., Ferrer, E., Gamboa, O. A., Jeronimo, J, & Lazcano-Ponce, E. (2008).Cervical cancer screening programs in Latin America and the Caribbean. Vaccine, 26(11), 37-48.

Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: a structural classification of protein of databases for the investigation of sequence and structures. Journal of Molecular Biology, 247(4), 536-540.

Naghashfar, Z., Rosenshein, N., Lorincz, A., Buscema, J., Shah, K. (1987). Characterization of human papillomavirus type 45, a new type 18-related virus of the genital tract. Journal of General Virology, 68(12), 3073-3079.

Narechania, A., Chen, Z., DeSalle, R., & Burk, R. D. (2005). Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. Journal of virology, 79(24), 15503-15510.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology, 48(3), 443-453.

Nicola, N. (2014). Tipificación del virus del papiloma humano – VPH y su relación con características poblacionales y lesiones en cáncer de cuello uterino en mujeres del municipio de Pasto [Tesis de pregrado]. San Juan de Pasto (Colombia). Universidad de Nariño.

Ntova, C. K., Kottaridi, C., Chranioti, A., Spathis, A., Kassanos, D., Paraskevaidis, E., ... & Karakitsos, P. (2012). Genetic variability and phylogeny of high risk HPV type 16, 18, 31, 33 and 45 L1 gene in Greek women. International Journal of Molecular Sciences, 13(1), 1-17.

Oliveira, L. H., Rosa, M. L., Pereira, C. R., Vasconcelos, G. A., Silva, R. A., Barrese, T. Z., … Cavalcanti, S. M. (2006). Human papillomavirus status and cervical abnormalities in women from public and private health care in Rio de Janeiro State, Brazil. Revista do Instituto de Medicina Tropical de São Paulo, 48(5), 279-285.

Ong, C. K., Chan, S. Y., Campo, M. S., Fujinaga, K., Mavromara-Nazos, P., Labropoulou, V., ... & Villa, L. (1993). Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. Journal of Virology, 7(11), 6424-6431.

Paavonen, J. (2007). Human papillomavirus infection and the development of cervical cancer and related genital neoplasias. International Journal of Infectious Diseases, 11(2), 3-9.

PAHO. (2014). Pan American Health Organization. Disponible en: http://www.paho.org/hq/

Parish, J. L., Kowalczyk, A., Chen, H. T., Roeder, G. E., Sessions, R., Buckle, M., … Gaston, K. (2006). E2 Proteins from High- and Low-Risk Human Papillomavirus Types Differ in Their Ability To Bind p53 and Induce Apoptotic Cell Death. Journal of Virology, 8(9), 4580-4590.

Park, R. B., & Androphy, E. J. (2002). Genetic analysis of high-risk e6 in episomal maintenance of human papillomavirus genomes in primary human keratinocytes. Journal of Virology, 76 (22), 11359-11364.

Parkin, D. M., Almonte, M., Bruni, L., Clifford, G., Curado, M. P., & Piñeros, M. (2008).Burden and trends of type-specific human papillomavirus infections and related diseases in the latin america and Caribbean region. Vaccine, 26(11), L1-15.

Peralta, R., Vargas-De-León, C., Cabrera, A., & Miramontes, P. (2014). Dynamics of High-Risk Nonvaccine Human Papillomavirus Types after Actual Vaccination Scheme. Computational and Mathematical Methods in Medicine, 542923, 1-8.

Picconi, M., Alonio, L., García, A., Lizano, M., Cervantes, G., Distefano, A., … & Teysie, A. (2000). Variantes moleculares de virus papiloma humano (VPH) tipos 16 y 18 en adenocarcinomas de cérvix. Medicina (Buenos Aires), 60, 89-94.

Potiny, S. (2010). An improved phylogenetic tree comparison method [Tesis de Maestría].Wilmington (EE. UU.). University of North Carolina.

Prado, J. C., Calleja-Macias, I. E., Bernard, H. U., Kalantari, M., Macay, S. A, Allan, B., … & Villa, L. L. (2005). Worldwide genomic diversity of the human papillomaviruses-53, 56, and 66, a group of high-risk HPVs unrelated to HPV-16 and HPV-18. Virology, 340(1), 95-104.

Qu, W., Jiang, G., Cruz, Y., Chang, C. J., Ho, G. Y., Klein, R. S., … & Burk, R. D. (1997). PCR detection of human papillomavirus: comparison between MY09/MY11 and GP5+/GP6+ primer systems. Journal of Clinical Microbiology, 35(6), 1304-1310.

Quintero-Vega, J. F., Cruz-Gómez, M. B., Márquez, L., & Puig-Pons J. (2008). Detección y tipificación de virus del papiloma humano (VPH) mediante PCR- RFLP. Revista de Obstetricia y Ginecología de Venezuela, 68(1), 25-31.

Raiol, T., Marreco, D., Rojas, M., Soares, P., Dias, E., & Renata, C. (2007). New HPV-16 European and non-European variants in Central Brazil. Virus Genes, 35, 1-4.

Raiol, T., Wyant, P. S., de Amorim, R. M., Cerqueira, D. M., Milanezi, N., Brígido, M., … & Martins CR. (2009). Genetic variability and phylogeny of the high-risk HPV-31, -33, -35, -52, and -58 in central Brazil. Journal of Medical Virology, 81(4), 685-692.

Rivera, R., Delgado, J., Painel, V., Barrero, R., & Larraín, R. (2006). Mecanismo de infección y transformación neoplásica producido por virus papiloma humano en el epitelio cervical. Revista chilena de obstetricia y ginecología, 71(2), 135-140.

Rizzo, J., & Rouchka, E. (2007). Review of Phylogenetic Tree Construction. Bioinformatics Review. Obtenido en http://bioinformatics.louisville.edu/lab/localresources/papers/TR-ULBL-2007-01.pdf

Sabol, I., Matovina, M., Si-Mohamed, A., & Grce, M. (2012). Characterization and Whole Genome Analysis of Human Papillomavirus Type 16 E1-13 ˆVariants. PLoS ONE, 7(7), e41045.

Sánchez, C., Suárez, K., Yépez, M., & Guerrero, M. (2013). Infección por VPH en mujeres del municipio de Pasto, Colombia con resultados de citología normal. Universidad y Salud, 15(1), 7-21.

Schiffman, M., Herrero, R., Desalle, R., Hildesheim, A., Wacholder, S., Rodriguez, A. C., … & Burk, R. D. (2005). The carcinogenicity of human papillomavirus types reflects viral evolution. Virology, 337(1), 76-84.

Schwarz, E., Dürst, M., Demankowski, C., Lattermann, O., Zech, R., Wolfsperger, E., … & zur Hausen, H. (1983). DNA sequence and genome organization of genital human papillomavirus type 6b. The EMBO Journal, 2(12), 2341-2348.

Seedorf, K., Krämmer, G., Dürst, M., Suhai, S., & Röwekamp, W. G. (1985). Human papillomavirus type 16 DNA sequence. Virology, 145(1), 181-185.

Sehnal, B., Dusek, L., Cibula, D., Zima,T., Halaska, M., Driak, D., … & Slama, J. (2014). The relationship between the cervical and anal HPV infection in women with cervical intraepithelial neoplasia. Journal of Clinical Virology, 59(1), 18-23.

Sichero, L., Ferreira, S., Trottier, H., Duarte-Franco, E., Ferenczy, A., Franco, E. L., … & Villa, L. L. (2007). High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18. International Journal of Cancer, 120, 1763-1768.

Smith, B., Chen, Z., Reimers, L., van Doorslaer, K., Schiffman, M., Desalle, R., … & Burk, R. D. (2011). Sequence imputation of HPV16 genomes for genetic association studies. PLoS One, 6(6), e21375.

Smith, T. F., & Waterman, M. S. (1981). Waterman. Identication of common molecular subsequences. Journal of Molecular Biology, 147, 195-197.

Squiquera, L. (2006). Human Papillomavirus. Review of the different therapeutic modalities. Revista Argentina de Dermatología, 87, 28-41.

Stoesser, G., Baker, W., van den Broek, A., Camon, E., Garcia-Pastor, M, Kanz., … & Vaughan, R. (2002). The EMBL nucleotide sequence database. Nucleic Acids Research, 30(1), 21-26.

Stubenrauch, F., & Laimins, L. (1999). Human papillomavirus life cycle: active and latent phases. Seminars in Cancer Biology, 9(6), 379-386.

Stewart, A. C., Eriksson, A. M., Manos, M. M., Bosch, F. X., Peto, J., … & Wheeler, C. M. (1996). Intratype variation in 12 human papillomavirus types: a worldwide perspective. Journal of Virology, 70(5): 3127-3136.

Syrjänen, S. (2010).The role of human papillomavirus infection in head and neck cancers. Annals of Oncology, 21(7), vii243-vii245.

Takami, Y., Kondoh, G., Saito, J., Noda, K., Sudiro, T., Sjahrurachman, A., … & Hakura, A. (1991). Cloning and characterization of human papillomavirus type 52 from cervical carcinoma in Indonesia. International Journal of Cancer, 48(4), 516-522.

Talianová, M. (2007). Survey of molecular phylogenetics: A review. Plant Soil Environ, 53(9), 413-416.

Tamegão-Lopes, B. P., Sousa-Júnior, E. C., Passetti, F., Ferreira, C. G., de Mello, W. A., & Duarte-Silvestre, R. V. (2014) Prevalence of human papillomavirus infection and phylogenetic analysis of HPV-16 E6 variants among infected women from Northern Brazil. Infectious Agents and Cance, 9 (25), 1-7.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution, 30(12), 2725-2729.

Tena, D., & López, J. (2007). Human papillomavirus infection and cervical cancer: pathogenesis and epidemiology. Communicating Current Research and Educational Topics and Trends in Applied Microbiology, 680-687. Recuperado de http://www.formatex.org/microbio/pdf/pages680-688.pdf

Thomas, M., Pim, D., & Banks, L. (1999).The role of the E6-p53 interaction in the molecular pathogenesis of HPV. Oncogene, 18(53), 7690-7700.

Thompson, J., Higgins, D., & Gibson, T. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22(22), 4673-4680.

Thomison III, J., Thomas, L., & Shroyer, K. (2008). Human papillomavirus: molecular and cytologic/histologic aspects related to cervical intraepithelial neoplasia and carcinoma. Human Pathology, 39(2), 154-166.

Thu-Hang, N. (2008). Comparison of multiple sequence alignment programs in practise [Tesis de Maestría]. Aarhus (Dinamarca): University of Århus.

Tiews, S., Steinberg, W., Schneider, W., & Hanrath C. (2009). Determination of the diagnostic accuracy of testing for high-risk (HR) human papillomavirus (HPV) types 16, 18 and 45 in precancerous cervical lesions: preliminary data. Journal of Clinical Virology, 46(3), S11-5.

Todorovic, B., Hung, K., Massimi, P., Avvakumov, N., Dick, F., Shaw, G., … & Mymryk, J. (2012). Conserved Region 3 of Human Papillomavirus 16 E7 Contributes to Deregulation of the Retinoblastoma Tumor Suppressor. Journal of Virology, 86(24),13313-13323.

Tonon, S. A, Basiletti, J., Badano, I., Alonio, L. V., Villa, L. L., Teyssie, A. R, … & Picconi M. A. (2007).Human papillomavirus type 16 molecular variants in Guarani Indian women from Misiones, Argentina. International Journal of Infectious Diseases, 11(1), 76-81.

Tornesello, M. L., Duraturo, M. L., Salatiello, I., Buonaguro, L., Losito, S., Botti, G., … & Buonaguro, F. M. (2004). Analysis of Human Papillomavirus Type-16 Variants in Italian Women With Cervical Intraepithelial Neoplasia and Cervical Cancer. Journal of Medical Virology, 74(1), 117-126.

Tornesello, M. L., Losito, S., Benincasa, G., Fulciniti, F., Botti, G., Greggi, S., … & Buonaguro, F. M. (2011). Human papillomavirus (HPV) genotypes and HPV16 variants and risk of adenocarcinoma and squamous cell carcinoma of the cervix. Gynecologic Oncology, 121(1), 32-42.

Valencia, A (2002). Bioinformatics: biology by other means. Bioinformatics, 18(12), 1551-1552.

Van, G. (2014). The Role of Ubiquitin and Ubiquitin-Like Modification Systems in Papillomavirus Biology. Viruses, 6(9), 3584-3611.

Van Ranst, M., Kaplan, J. B., & Burk, R. D. (1992). Phylogenetic classification of human papillomaviruses: correlation with clinical manifestations. Journal of General Virology, 73(10), 2653-2660.

Vande Pol, S. B., & Klingelhutz, A. J. (2013). Papillomavirus E6 oncoproteins. Virology, 445(1-2), 115-137.

Venegas, C., Hernández, D., González, M., & Lorenzo, C. (2011). Infección por virus del papiloma humano: asociación entre infección genital y buccal. Revista de Obstetricia y Ginecología de Venezuela, 71(3), 164-173.

Veras, V. S., Cerqueira, D. M., & Martins, C. R. (2005). L1 sequence of a new human papillomavirus type-58 variant associated with cervical intraepithelial neoplasia. Brazilian Journal of Medical and Biological Research, 38(1), 1-4.

Villa, L. L. (2012). Cervical Cancer in Latin America and the Caribbean: The Problem and the Way to Solutions. Cancer Epidemiology, Biomarkers & Prevention, 21(9), 1409-1413.

Villa, L. L., Sichero, L., Rahal, P., Caballero, O., Ferenczy, A., Rohan, T., … & Franco, E. (2000). Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasma. Journal of General Virology, 81, 2959-2968.

Volkow, P., Rubí, S., Lizano, M., Carrillo, A., Vilar-Compte, D., García-Carrancá, A., … & Mohar A.(2001). High prevalence of oncogenic human papillomavirus in the genital tract of women with human immunodeficiency virus. Gynecologic Oncology, 82(1), 27-31.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics, 25(9), 1189-1191.

Wilson, R. K. (1999). How the worm was won. The C. elegans genome sequencing project. Trend genetics, 15(2), 51-58.

WHO. (2014). World Health Organization: WHO. Disponible en: http://www.who.int/en/

Wu, Y., Chen, Y., Li, L., Yu, G., He, Y., & Zhang, Y. (2006). Analysis of mutations in the E6/E7 oncogenes and L1 gene of human papillomavirus 16 cervical cancer isolates from China. Journal of General Virology, 87(5), 1181-1188.

Xi, L. F., Koutsky, L. A., Hildesheim, A., Galloway, D. A., Wheeler, C. M., Winer, R. L., … & Kiviat, N. B. (2007). Risk for High-Grade Cervical Intraepithelial Neoplasia Associated with Variants of Human Papillomavirus Types 16 and 18. Cancer Epidemiolology, Biomarkers & Prevention, 16(1), 4-10.

Xiong, J. (2006). Essential Bioinformatics. Estados Unidos de América: Cambridge University Press.

Yamada, T., Wheeler, C., Halpern, A., Stewart, A., Hildeshein, A., & Jenison, S. (1995). Human Papillomavirus Type 16 Variant Lineages in United States Populations Characterized by Nucleotide Sequence Analysis of the E6, L2, and L1 Coding Segments. Journal of Virology, 69(12), 7743-7753.

Yamada, T., Manos, M. M., Peto, J., Greer, C. E., Munoz, N., Bosch, F. X., … &Wheeler, C. M. (1997). Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. Journal of Virology, 71(3), 2463-2472.

Zhang, Y., Dasgupta, J., Ma, R. Z., Banks, L., Thomas, M., & Chen, X. (2007). Structures of a Human Papillomavirus (HPV) E6 Polypeptide Bound to MAGUK Proteins: Mechanisms of Targeting Tumor Suppressors by a High-Risk HPV Oncoprotein. Journal of Virology, 81(7), 3618-3626.

Zhang, C., Park, J. S., Grce, M., Hibbitts, S., Palefsky, J. M., Konno, R., … & Chan, P. (2014). Geographical distribution and risk association of human papillomavirus genotype 52-variant lineages. Jounal of Infection Disease, 210(10),1600-1604.

ANEXOS

Anexo A

Tabla de aminoácidos con abreviaturas empleadas para su identificación y su clasificación según el grupo funcional.

Aminoácido	**Abreviatura**		**Clasificación**
	Tres letras	*Una letra*	
Alanina	Ala	A	Alifático
Arginina	Arg	R	Básico
Asparagina	Asn	N	Ácido
Ácido aspártico	Asp	D	Ácido
Cisteína	Cys	C	Azufrado
Glutamina	Gln	Q	Ácido
Ácido glutámico	Glu	E	Ácido
Glicina	Gly	G	Alifático
Histidina	His	H	Básico
Isoleucina	Ile	I	Alifático
Leucina	Leu	L	Alifático
Lisina	Lys	K	Básico
Metionina	Met	M	Azufrado
Fenilalanina	Phe	F	Aromático
Prolina	Pro	P	Alifático
Serina	Ser	S	Hidroxilado
Treonina	Thr	T	Hidroxilado
Triptófano	Trp	W	Aromático
Tirosina	Tyr	Y	Aromático
Valina	Val	V	Alifático

Anexo B

Listado de artículos seleccionados en la revisión.

Código	BD	Revista	Autor (s)	País	Gen (s)	Tipo (s)	DOI
1	PubMet	J Virol	Stewart et al., 1996	E.U	L1	18, 33, 35, 45, 52, 58	
2	Springer	J Mol Evol	DeFilippis et al., 2002	E.U	E6, E7,	16	10.1007/s00239-002-2344-y
3	PubMet	Virus gene	Cerqueira et al., 2003	Brasil	L1	58	
4	PubMet	J Virol	Calleja-Macias et al., 2005	E.U	E6, L1	31, 35, 52	10.1128/JVI.79.21.13630-13640.2005
5	PubMet	J Virol	Chen et al., 2005	E.U	E6, E7, L1	16	10.1128/JVI.79.11.7014-7023.2005
6	Springer	Arch Virol	Cerqueira et al., 2007	Brasil	L1	11, 18, 33	10.1007/s00705-006-0828-6
7	Springer	Virus genes	Cerqueira et al., 2008	Brasil	E6	18	10.1007/s11262-008-0263-8
8	PubMet	J Virol	Chen et al., 2009	E.U	E6, E7, L1	18, 45	10.1128/JVI.02068-08
9	Springer	Infec Agent Cancer	López-Revilla et al., 2009	México	E6	16	10.1186/1750-9378-4-3
10	PubMet	J Med Virol	Raiol et al., 2009	Brasil	E6, L1	31, 33, 52	10.1002/jmv.21432
11	PubMet	Genet Mol Res	Castro et al., 2011	Brasil	L1	13, 16, 33, 58	10.4238/vol10-1gmr992.
12	Science Direct	Gynecol Oncol	Chagas et al., 2011	Brasil	E6, E7	31	10.1016/j.ygyno.2011.07.008
13	PubMet	J Infect Dis	Chan et al., 2011	Hong Kong	L1	58	10.1093/infdis/jir157
14	PubMet	PLoS One	Chen et al., 2011	E.U	E6, E7, L1	16, 31, 33, 35, 52, 58	10.1371/journal.pone.0020183

15	PubMet	PLoS One	Smith et al., 2011	E.U	E6, E7, L1	16	10.1371/journal.pone.0021375
16	PubMet	J Virol	Cornet et al., 2012	Francia	E6	16	10.1128/JVI.00483-12
17	Science Direct	Infect Genet Evol	Chagas et al., 2013	Brasil	E6, E7	31	10.1016/j.meegid.2013.02.001
18	PubMet	PLoS One	Chen et al., 2013	E.U	E6, E7, L1	18, 45	10.1371/journal.pone.0072565
19	PubMet	Int J Cancer	Chan et al., 2013	Hong Kong	E6, E7	58	10.1002/ijc.27932
20	Science Direct	Infect Genet Evol	de Matos et al., 2013	Brasil	L1	6, 11 H	10.1016/j.meegid.2012.12.033
21	Science Direct	Infect Genet Evol	Godínez et al., 2013	España	E6, E7	33, 45, 58	10.1016/j.meegid.2012.09.011
22	PubMet	J Virol	Chen et al., 2014b	Francia	E6, E7	45	10.1128/JVI.03534-13
23	PubMet	J Virol	Chen et al., 2014a	Francia	E6, E7	33	10.1016/j.virol.2013.10.033
24	PubMet	J Infect Dis	Zhang et al., 2014	Hong Kong	E6, L1	52	10.1093/infdis/jiu310

Anexo C

Variabilidad genética de secuencias E6 de VPH 6. Los nucleótidos conservados se marcan con un punto (•) y sin cambio en la secuencia de aminoácidos con guion (-). Secuencia prototipo de VPH 6A: Sec. Ref. (No Acceso X00203). H: histidina, Q: glutamina, D: aspartato, G: glicina.

VPH 6	Posiciones de sustituciones de nucleótidos en E6								Código GenBank
	221	251	323	365	392	397	473	479	
Sec. Ref	**A**	**C**	**A**	**A**	**C**	**A**	**G**	**C**	**X00203**
BRA1	T	G	C	T	T	·	A	T	KC285843
BRA2	T	G	T	T	T	·	A	·	KC285844
BRA3	T	G	T	T	T	·	A	·	KC285845
BRA4	T	G	T	T	T	·	A	·	KC285846
BRA5	T	G	T	T	T	·	A	·	KC285847
BRA6	T	G	T	T	T	·	A	·	KC285848
BRA7	T	·	·	·	·	·	A	·	KC285849
BRA8	T	G	C	T	T	·	A	T	KC285850
BRA9	T	G	T	T	T	G	A	·	KC285851
BRA10	T	G	C	T	T	·	A	T	KC285852
BRA11	T	G	T	T	T	G	A	·	KC285853
BRA12	T	G	T	T	T	·	A	·	KC285854
BRA13	T	G	T	T	T	·	A	·	KC285855
BRA14	T	G	T	T	T	·	A	·	KC285842
BRA15	T	G	T	T	T	·	A	·	KC285841
BRA16	T	G	T	T	T	·	A	·	KC285840
BRA17	T	G	C	T	T	·	A	T	KC285839
BRA18	T	G	C	T	T	·	A	T	KC285838
Mutación no sinónima	-	H 50 Q	-	-	-	D 99 G	-	-	

Señalada en rojo la mutación descrita por primera vez en este estudio.

Anexo D

Variabilidad genética de secuencias E6 de VPH 11. Los nucleótidos conservados se marcan con un punto (•) y sin cambio en la secuencia de aminoácidos con guion (-). Secuencia de referencia VPH 11: Sec. Ref. E: (No Acceso M14119.1). E: glutamina, G: glicina.

VPH 11	Posiciones de sustituciones de nucleótidos en E6			Código GenBank
	137	380	538	
Sec. Ref.	**T**	**C**	**A**	KC285856
BRA1	C	T	·	KC285857
BRA2	·	·	·	KC285858
BRA3	·	·	·	KC285859
BRA4	C	T	·	KC285860
BRA5	·	·	G	KC285861
BRA6	C	T	·	KC285862
BRA7	·	·	·	KC285856
Mutación no sinónima	-	-	E 146 G	

Señalada en rojo la mutación descrita por primera vez en este estudio.

Anexo E

Variabilidad genética de secuencias E6 de VPH 18. Los nucleótidos conservados se marcan con un punto (•) y sin cambio en la secuencia de aminoácidos con guion (-). Secuencia de referencia VPH 18: Sec. Ref. (No Acceso X05015). E: glutamina, G: glicina, H: histidina, Y: tirosina, N: asparagina, K: lisina, R: arginina.

VPH 18 (E6)	Posiciones de sustituciones de nucleótidos en E6															Código GenBank
	149	232	251	266	287	317	342	374	437	485	491	502	548	549	551	
Sec. Ref.	**T**	**A**	**T**	**G**	**C**	**T**	**C**	**G**	**G**	**T**	**C**	**A**	**A**	**C**	**A**	**X05015**
BRA1	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	EF661657.1
BRA2	•	•	C	A	G	C	•	A	•	C	A	•	G	A	•	EF661656.1
BRA3	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	EF661655.1
BRA4	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	EF661654.1
CR1	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	EF202145
CR2	•	G	•	•	G	•	•	•	•	C	•	•	•	A	•	EF202143
CR3	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	EF202146
CR4	•	•	•	•	G	•	•	•	•	•	•	G	•	•	•	EF202147
CR5	C	•	•	•	G	•	•	•	•	C	•	•	•	•	•	EF202148
CR6	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	EF202149
CR7	•	•	C	A	G	C	T	A	•	C	A	•	G	A	•	EF202154
CR8	•	•	C	A	G	C	T	A	•	C	A	•	G	A	G	EF202155
CR9	•	•	C	A	G	C	T	A	•	C	A	•	G	A	•	EF202153
CR10	•	•	C	A	G	•	•	A	•	C	A	•	G	A	•	EF202152
CR11	•	•	•	•	G	•	•	•	•	•	•	•	•	•	•	kC470208
CR12	•	•	•	•	G	•	•	•	•	•	•	•	•	•	•	kC470209
CR13	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	kC470210
CR14	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	kC470211
CR15	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	kC470212
CR16	•	•	C	A	G	C	•	A	•	C	A	•	G	A	•	kC470221
CR17	•	•	C	A	G	•	•	A	•	C	A	•	G	A	•	kC470227
CR18	•	•	C	A	G	•	•	A	A	C		•	G	A	•	kC470229
CR19	•	•	•	•	G	•	•	•	•	•	•	•	•	•	•	EF202144
CR20	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	EF202150
CR21	•	•	•	•	G	•	•	•	•	C	•	•	•	A	•	EF202151
Mutación no sinónima	-	E 43 G	-	-	-	-	H 80 Y	-	-	-	N 129 K	H 133 R	-	-	-	

Anexo F

Variabilidad genética de secuencias E7 de VPH 18. Los nucleótidos conservados se marcan con un punto (•) y sin cambio en la secuencia de aminoácidos con guion (-). Secuencia de referencia VPH 18: Sec. Ref. (No Acceso X05015). H: histidina, Y: tirosina, E: glutamina, K: lisina, N: asparagina, S: serina.

VPH 18 (E7)	Posiciones de sustituciones de nucleótidos en E7										Código GenBank
	593	640	664	692	736	751	864	865	874	906	
Sec. Ref.	**C**	**C**	**T**	**G**	**A**	**C**	**A**	**C**	**C**	**A**	**X05015**
CR1	·	·	·	A	·	T	·	·	·	G	EF202145
CR2	·	·	·	·	·	T	·	·	T	·	EF202143
CR3	·	·	·	·	·	·	·	·	·	·	EF202146
CR4	·	·	·	·	·	·	·	T	·	·	EF202147
CR5	·	·	·	·	·	·	·		·	·	EF202148
CR6	·	·	·	·	·	T	·	·	·	·	EF202149
CR7	T	T	·	·	·	·	G	·	·	·	EF202154
CR8	T	T	·	·	G	·	G	·	·	·	EF202155
CR9	T	T	·	·	·	·	G	·	·	·	EF202153
CR10	T	T	·	·	·	·	·	A	·	·	EF202152
CR11	·	·	·	·	·	·	·	·	·	·	kC470208
CR12	·	·	·	·	·	·	·	·	·	·	kC470209
CR13	·	·	·	·	·	·	·	·	·	·	kC470210
CR14	·	·	·	·	·	·	·	·	·	·	kC470211
CR15	·	·	·	·	·	T	·	·	·	·	kC470212
CR16	T	T	·	·	·	·	G	·	·	·	kC470221
CR17	T	T	·	·	·	·	·	A	·	·	kC470227
CR18	T	·	·	·	·	·	·	·	·	·	kC470229
CR19	·	·	·	·	·	·	·	·	·	·	EF202144
CR20	·	·	·	·	·	T	·	·	·	·	EF202150
CR21	·	·	C		·	T	·	·	·	·	EF202151
Mutación no sinónima	H2Y	-	-	E35K	-	-	N92S/ N92K	-	-	-	

Anexo G

Variabilidad genética de secuencias E6 de VPH 45. Los nucleótidos conservados se marcan con un punto (•) y sin cambio en la secuencia de aminoácidos con guion (-). Secuencia de referencia VPH 45: Sec. Ref. (No Acceso X74879.1). K: lisina, T: treonina, L: leucina, D: aspartato, E: glutamina, R: arginina, Q: glutamina, C: cisteína, F: fenilalanina, Y: tirosina, S: serina, N: asparagina.

VPH 45 (E6)	Posiciones de sustituciones de nucleótidos en E6														Código GenBank
	124	146	150	162	163	182	226	237	259	284	418	482	487	497	
Sec. Ref.	**A**	**A**	**T**	**T**	**T**	**C**	**C**	**C**	**G**	**A**	**T**	**T**	**G**	**A**	**X74879.1**
BRA1	•	•	•	•	•	•	•	•	•	•	•	•	A	•	JQ976786.1
BRA2	•	•	•	•	•	•	•	•	T	•	•	•	•	•	JQ976787.1
BRA3	C	•	•	•	•	•	•	•	T	•	•	•	•	•	JQ976788.1
BRA4	C	•	•	•	•	•	•	•	T	•	•	C	•	G	JQ976789.1
BRA5	C	•	•	•	•	•	•	•	T	•	•	C	•	G	JQ976790.1
BRA6	•	•	•	•	•	•	•	•	•	•	•	C	•	•	JQ976791.1
BRA7	C	•	•	•	•	•	•	•	T	•	•	C	•	G	JQ976792.1
BRA8	•	•	•	•	•	•	•	•	•	C	•	C	•	•	JQ976793.1
BRA9	•	•	•	•	•	•	•	•	T	•	•	C	•	•	JQ976794.1
BRA10	•	•	•	•	•	•	•	•	•	•	•	•	•	•	JQ976785.1
PAN1	•	G	C	•	•	•	•	•	A	•	•	C	•	G	KF591365
PAN2	•	•	•	A	C	•	•	G	•	•	•	C	A	•	KF591364
PAR1	•	•	•	•	•	•	•	•	•	C	•	C	•	G	KF591367
PAR2	•	•	•	A	C	•	•	G	T	•	•	C	A	•	KF591346
PAR3	•	•	•	A	C	•	G	•	•	•	•	C	A	•	JQ976800
PAR4	•	•	•	•	•	•	•	•	•	C	•	C	•	G	JQ976801
PAR5	•	•	•	•	•	•	•	•	•	•	•	C	•	•	JQ976802
PER1	•	•	•	A	C	•	G	•	•	•	•	C	•	•	JQ976831
PER2	C	•	•	•	•	•	•	•	•	•	•	C	•	•	JQ976832
PER3	•	•	•	A	C	•	G	•	•	•	•	C	A	•	JQ976833
CR1	•	•	•	A	C	•	•	G	•	•	•	C	A	•	EF202157
CR2	•	•	•	A	C	•	•	G	•	•	•	C	A	•	EF202158
CR3	•	•	•	•	•	•	•	•	•	•	•	C	A	•	EF202159
CR4	•	•	•	A	C	•	•	G	•	•	•	C	A	•	EF202160
CR5	•	•	•	•	•	•	•	•	•	C	•	C	•	G	EF202161
CR6	•	•	•	•	•	•	•	•	•	C	•	C	•	G	EF202162
CR7	C	•	•	•	•	•	•	•	•	•	•	C	•	G	EF202166
CR8	C	•	•	•	•	G	•	•	T	•	G	C	•	G	EF202165
CR9	C	•	•	•	•	•	•	•	T	•	•	C	•	G	EF202167
CR10	C	•	•	•	•	•	•	•	T	•	•	C	•	G	EF202164
CR11	•	•	•	•	•	•	•	•	•	C	•	C	•	G	EF202163
CR12	•	•	•	•	•	•	•	•	•	•	•	•	•	•	EF202156
CR13	•	•	•	A	C	•	•	G	•	•	•	C	A	•	KC470253
CR14	•	•	•	A	C	•	•	G	•	•	•	C	A	•	KC470254
CR15	C	•	•	•	•	•	•	•	T	•	•	C	•	G	KC470259
CR16	C	•	•	•	•	•	•	•	T	•	•	C	•	G	KC470260
Mutación no sinónima	K8T	-	-	-	L21T	D27E	T42R	Q46E	C53F / C53Y	-	L106R	-	S129N	-	

Señaladas en rojo las mutaciones descritas por primera vez en este estudio.

Anexo H

Variabilidad genética de secuencias E7 de VPH 45. Los nucleótidos conservados se marcan con un punto (•) y sin cambio en la secuencia de aminoácidos con guion (-). Secuencia de referencia VPH 45: Sec. Ref. (No Acceso X74879.1). R: arginina, Q: glutamina, K: lisina, E: glutamato, A: alanina, I: isoleucina, M: metionina, D: aspartato, S: serina.

VPH 45 (E7)	Posiciones de sustituciones de nucleótidos en E7									Código GenBank
	591	599	600	603	718	772	808	822	832	
Sec. Ref.	**T**	**C**	**G**	**A**	**A**	**T**	**G**	**G**	**G**	**X74879.1**
BRA1	·	·	·	C	·	·	T	T	·	JQ976786.1
BRA2	·	·	A	C	C	·	·	T	·	JQ976787.1
BRA3	·	·	A	C	C	·	·	T	·	JQ976788.1
BRA4	·	·	A	C	C	·	·	T	·	JQ976789.1
BRA5	·	·	A	C	C	·	·	·	·	JQ976790.1
BRA6	·	·	A	C	C	·	·	T	·	JQ976791.1
BRA7	·	·	A	C	C	·	·	T	·	JQ976792.1
BRA8	·	·	·	C	C	·	T	T	·	JQ976793.1
BRA9	·	·	A	C	C	·	·	T	·	JQ976794.1
BRA10	·	·	·	·	·	·	·	·	·	JQ976785.1
PAN1	C	A	A	C	·	·	·	·	T	KF591365
PAN2	·	·	·	C	·	G	·	·	T	KF591364
PAR1	·	·	·	C	·	·	·	·	T	KF591367
PAR2	·	·	·	C	C	·	T	·	T	KF591346
PAR3	·	·	·	C	C	·	T	T	·	JQ976800
PAR4	·	·	·	C	·	·	·	T	·	JQ976801
PAR5	·	·	·	C	C	·	T	T	·	JQ976802
PER1	·	·	·	C	·	·	·	T	·	JQ976831
PER2	·	·	A	C	C	·	·	T	·	JQ976832
PER3	·	·	·	·	·	·	·	T	·	JQ976833
CR1	·	·	·	C	·	·	·	·	T	EF202157
CR2	·	·	·	C	·	·	·	·	T	EF202158
CR3	·	·	·	C	·	·	·	·	T	EF202159
CR4	·	·	·	C	·	·	·	·	T	EF202160
CR5	·	·	·	C	C	·	T	·	T	EF202161
CR6	·	·	·	C	·	·	T	·	T	EF202162
CR7	·	·	A	C	C	·	·	·	T	EF202166
CR8	·	·	A	C	C	·	·	·	T	EF202165
CR9	·	·	A	C	C	·	·	·	T	EF202167
CR10	·	·	A	C	C	·	·	·	T	EF202164
CR11	·	·	·	C	C	·	T	·	T	EF202163
CR12	·	·	·	·	·	·	·	·	·	EF202156
CR13	·	·	·	C	·	·	·	·	T	KC470253
CR14	·	·	·	C	·	·	·	·	T	KC470254
CR15	·	·	A	C	C	·	·	·	T	KC470259
CR16	·	·	A	C	C	·	·	·	T	KC470260
Mutación no Sinónima	-	-	R5Q/ R5K	E 6 A	-	I 62 M	E 74 D	S 79 I	E 82 D	

Señaladas en rojo las mutaciones descritas por primera vez en este estudio.

Anexo I

Variabilidad genética de secuencias E6 de VPH 16. Los nucleótidos conservados se marcan con un punto (•) y sin cambio en la secuencia de aminoácidos con guion (-). Secuencia de referencia VPH 16: Sec. Ref. (No Acceso K02718.1). R: arginina, G: glicina, T: treonina, I: isoleucina, Q: glutamina, H: histidina, D: aspartato, L: leucina, V: valina, N: asparagina, Y: tirosina, E: glutamato, C: cisteína, S: serina, K: lisina, F: fenilalanina.

VPH 16	Posiciones de sustituciones de nucleótidos en E6																																									
	109	127	131	132	143	145	146	173	176	178	182	183	185	188	189	206	234	250	256	286	289	290	295	306	310	334	335	350	351	354	360	361	374	403	404	424	433	447	457	507	532	535	Código Genbank
Sec. Ref.	T	A	A	G	C	G	T	C	G	T	A	T	T	G	A	C	A	T	C	T	A	T	T	A	T	A	C	T	T	A	C	A	C	A	A	A	G	A	T	G	A	A	K02718.1
MEX1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·		·	·	·	·	·	·	·	·	·	T	·	·	·	·	·	·	·	EU880264.1
MEX2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880265.1
MEX3	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880266.1
MEX4	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880267.1
MEX5	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880268.1
MEX6	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880269.1
MEX7	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880270.1
MEX8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880271.1
MEX9	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880272.1
MEX10	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880254.1
MEX11	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880255.1
MEX12	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880256.1
MEX13	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880257.1
MEX14	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880258.1
MEX15	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880259.1
MEX16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880260.1
MEX17	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	A	·	·	·	·	·	·	·	·	·	EU880261.1
MEX18	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880262.1
MEX19	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880263.1
MEX20	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880244.1
MEX21	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880245.1
MEX22	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880246.1
MEX23	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880247.1
MEX24	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880248.1
MEX25	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880249.1
MEX26	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880250.1
MEX27	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880251.1
MEX28	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880252.1
MEX29	·	·	·	·	·	·		A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	G	G	G	·	·	·	·	·	·	·	·	·	·	EU880253.1
MEX30	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	EU880235.1
MEX31	·	·	·	·	·	T		·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	EU880236.1
MEX32	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880237.1
MEX33	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880238.1
MEX34	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880239.1
MEX35	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880240.1
MEX36	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880241.1
MEX37	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	T	C	C	·	·	·	·	·	·	·	·	·	G	A	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880242.1
MEX38	·	·	·	·	·	·		·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	EU880243.1
MEX39	·	·	·	·	·	T	·	·	·	·	·	·	·	C	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KJ465992.1
MEX40	·	·	·	·	·	T	·	·	·	·	·	G	G	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KJ465993.1

VPH 16	Posiciones de sustituciones de nucleótidos en E6																																									Código Genbank
	109	127	131	132	143	145	146	173	176	178	182	183	185	188	189	206	234	250	256	286	289	290	295	306	310	334	335	350	351	354	360	361	374	403	404	424	433	447	457	507	532	535	
MEX41	·	·	·	·	·	·	·	·	A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	G	·	·	·	·	·	·	KJ465994.1
MEX42	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KJ465995.1
MEX43	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	·	G	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KJ465996.1
MEX44	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KJ465997.1
MEX45	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	T	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KJ465998.1
MEX46	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	G	KJ465999.1
MEX47	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY089951.1
MEX48	·	·	·	·	·	T	·	·	·	·	·	G	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	AY089952.1
MEX49	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY089953.1
MEX50	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY089954.1
MEX51	·	·	·	·	·	·	·	·	A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY089955.1
ARGC1	·	·	·	·	·	T	·	·	·	·	·	G	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466853.2
ARGC2	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466839.2
ARGC3	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466834.2
ARGC4	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466833.2
ARGC5	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466831.2
ARGC6	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466830.2
ARGC7	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466763.2
ARGC8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466746.2
ARGC9	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466745.2
ARGC10	·	·	·	·	·	·	·	·	A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466744.2
ARGC11	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466697.2
ARGC12	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466699.2
ARGC13	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466505.2
ARGP14	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291261.1
ARGP15	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291262.1
ARGP16	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291263.1
ARGP17	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291264.1
ARGP18	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291265.1
ARGP19	?	?	?	·	G	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291266.1
ARGP20	?	?	?	·	·	·	·	·	A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291267.1
ARGP21	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291268.1
ARGP22	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291269.1
ARGP23	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291255.1
ARGP24	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291256.1
ARGP25	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291257.1
ARGP26	?	?	?	·	G	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	·	·	·	·	·	·	G	·	·	·	·	·	·	G	·	KC291258.1
ARGP27	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291259.1
ARGP28	?	?	?	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291260.1
ARGP29	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KC291254.1
BRA1	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	AY098923.1
BRA2	·	·	·	C	G	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY098922.1
BRA3	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY098919.1
BRA4	C	·	·	T	G	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	AY098918.1
BRA5	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY112663.1
BRA6	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	AY112662.1
BRA7	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466854.2
BRA8	C	·	·	T	G	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	KF466623.1
BRA9	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466516.2
BRA10	·	·	·	C	G	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	HM057182.1

VPH 16	Posiciones de sustituciones de nucleótidos en E6																																										Código Genbank
	109	127	131	132	143	145	146	173	176	178	182	183	185	188	189	206	234	250	256	286	289	290	295	306	310	334	335	350	351	354	360	361	374	403	404	424	433	447	457	507	532	535	
CHI1	·	·	·	·	·	T	·	·	·	·	·	G	·	·	·	·	·	·	·	A	G	·	G	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466529
COL1	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	A	G	·	KF466522
COL2	·	·	·	·	·	T	·	·	·	·	·	G	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466528
CUB1	·	·	·	C	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466535
CUB2	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466520
PAR1	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466523
PAR2	·	·	·	·	·	T	·	·	A	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466524
PAR3	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	G	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466525
PER1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF466838
PER2	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466518
PER3	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466519
PER4	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF466532
PER5	·	·	·	·	·	T	·	·	·	·	·	G	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	KF464533
CR1	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	A	·	·	·	G	·	AF402678
CR2	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	AY686579
CR3	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY686582
CR4	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY686580
CR5	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY686581
CR6	·	·	·	·	·	·	·	·	A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	AY686583
CR7	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	AY686584
CR8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644274.1
CR9	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644272.1
CR10	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644267.1
CR11	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644268.1
CR12	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644259.1
CR13	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644286.1
CR14	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644284.1
CR15	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644283.1
CR16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644282.1
CR17	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644280.1
CR18	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644261.1
CR19	·	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644287.1
CR20	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	HQ644278.1
CR21	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	HQ644255.1
CR22	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	HQ644253.1
CR23	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644257.1
CR24	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	HQ644285.1
CR25	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	HQ644276.1
CR26	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	HQ644269.1
CR27	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	HQ644265.1
CR28	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644263.1
CR29	·	·	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644275.1
CR30	·	·	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644266.1
CR31	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644264.1
CR32	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644273.1
CR33	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	G	·	HQ644270.1
CR34	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644254.1
CR35	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644281.1
CR36	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644279.1
CR37	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	A	G	·	·	·	·	·	T	G	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ644277.1

VPH 16	Posiciones de sustituciones de nucleótidos en E6																																									
	109	127	131	132	143	145	146	173	176	178	182	183	185	188	189	206	234	250	256	286	289	290	295	306	310	334	335	350	351	354	360	361	374	403	404	424	433	447	457	507	532	535	Código Genbank
Mutación no sinónima	-	-	R10G R10T R10I	-	-	Q14H Q14D	L15V	H24N H24Y	D25N D25E	-	I27T	I27R I27T	L28V	E29Q	E29G	-	-	-	-	-	-	C63S	D64E	K68N	F69L	-	H78Y	L83V	-	Y84C	-	T86R	Q99K	-	I101F	-	-	K115R	-	R135Q	-	-	

***Color según la rama filogenética: Rama E (amarillo), rama As (verde), rama AA (azul celeste), rama NA1 (purpura), y rama Af-2 (marrón).**

Anexo J

Variabilidad genética de secuencias E7 de VPH 16. Los nucleótidos conservados se marcan con un punto (·) y sin cambio en la secuencia de aminoácidos con guion (-). Secuencia de referencia VPH 45: Sec. Ref. (No Acceso K02718.1). L: leucina, S: serina, N: asparagina, P: prolina, D: aspartato, F: fenilalanina.

VPH 16 (E7)	Posiciones de sustituciones de nucleótidos en E7												Código GenBank
	644	645	646	647	648	649	650	666	732	789	795	846	
Sec. Ref.	**T**	**A**	**A**	**A**	**T**	**G**	**A**	**G**	**T**	**T**	**T**	**T**	**K02718.1**
BRA1	C	G	C	C	C	T	T	·	·	C	G	·	HM057182.1
CR1	·	·	·	·	·	·	·	·	C	C	G	·	AF402678
CR2	·	·	·	·	·	·	·	·	C	C	G	·	AY686579
CR3	·	·	·	·	·	·	·	·	C	C	G	·	AY686582
CR4	·	·	·	·	·	·	·	·	·	·	·	·	AY686580
CR5	·	·	·	·	·	·	·	·	·	·	·	·	AY686581
CR6	·	·	·	·	·	·	·	·	·	·	·	·	AY686583
CR7	·	·	·	·	·	·	·	·	·	·	·	·	AY686584
CR8	·	·	·	·	·	·	·	·	·	·	·	·	HQ644274.1
CR9	·	·	·	·	·	·	·	·	·	·	·	·	HQ644272.1
CR10	·	·	·	·	·	·	·	·	·	·	·	·	HQ644267.1
CR11	·	·	·	·	·	·	·	·	·	·	·	·	HQ644268.1
CR12	·	·	·	·	·	·	·	·	·	·	·	·	HQ644259.1
CR13	·	·	·	·	·	·	·	·	·	·	·	·	HQ644286.1
CR14	·	·	·	·	·	·	·	·	·	·	·	·	HQ644284.1
CR15	·	·	·	·	·	·	·	·	·	·	·	·	HQ644283.1
CR16	·	·	·	·	·	·	·	·	·	·	·	·	HQ644282.1
CR17	·	·	·	·	·	·	·	·	·	·	·	·	HQ644280.1
CR18	·	·	·	G	·	·	·	·	·	·	·	C	HQ644261.1
CR19	·	·	·	·	·	·	·	·	·	·	·	·	HQ644287.1
CR20	·	·	·	·	·	·	·	·	C	C	G	·	HQ644278.1
CR21	·	·	·	·	·	·	·	·	C	C	G	·	HQ644255.1
CR22	·	·	·	·	·	·	·	·	C	C	G	·	HQ644253.1
CR23	·	·	·	·	·	·	·	·	·	C	G	·	HQ644257.1
CR24	·	·	·	·	·	·	·	·	C	C	G	·	HQ644285.1
CR25	·	·	·	·	·	·	·	·	C	C	G	·	HQ644276.1
CR26	·	·	·	·	·	·	·	·	C	C	G	·	HQ644269.1
CR27	·	·	·	·	·	·	·	·	C	C	G	·	HQ644265.1

CR28	·	·	·	·	·	·	·	·	C	C	G	·	HQ644263.1
CR29	·	·	·	·	·	·	·	·	·	·	·	·	HQ644275.1
CR30	·	·	·	·	·	·	·	·	C	C	G	·	HQ644266.1
CR31	·	·	·	·	·	·	·	·	·	·	·	·	HQ644264.1
CR32	·	·	·	·	·	·	·	·	C	C	G	·	HQ644273.1
CR33	·	·	·	·	·	·	·	·	C	C	G	·	HQ644270.1
CR34	·	·	·	·	·	·	·	·	C	C	G	·	HQ644254.1
CR35	·	·	·	·	·	·	·	·	C	C	G	·	HQ644281.1
CR36	·	·	·	·	·	·	·	·	C	C	G	·	HQ644279.1
CR37	·	·	·	·	·	·	·	·	C	C	G	·	HQ644277.1
MEX1	·	·	·	·	·	·	·	·	C	C	G	·	AY089951.1
MEX2	·	·	·	·	·	·	·	·	C	C	G	·	AY089952.1
MEX3	·	·	·	·	·	·	·	·	·	·	·	·	AY089953.1
MEX4	·	·	·	·	·	·	·	A	·	·	·	·	AY089954.1
MEX5	·	·	·	·	·	·	·	·	·	·	·	·	AY089955.1
Mutación no sinónima	-	L 28 S	-	N29P/ N29S	-	D 30 F	-	-	-	-	-	-	

Anexo K

Variabilidad genética de secuencias E6 de VPH 31. Los nucleótidos conservados se marcan con un punto (·) y sin cambio de aminoácidos con guion (-). Secuencia de referencia de VPH 31: Sec. Ref. (No Acceso: J04353.1). I: isoleucina, L: leucina, H: histidina, Y: tirosina T: treonina, A: alanina, E: glutamina, K: lisina, R: arginina, V: valina, C: cisteína, G: glicina.

VPH 31 (E6)	Posiciones de sustituciones de nucleótidos en E6																				Código GenBank
	176	248	261	285	297	305	306	309	320	326	372	383	404	428	430	475	520	523	532	537	
Sec. Ref.	**C**	**T**	**A**	**C**	**A**	**T**	**T**	**A**	**A**	**A**	**G**	**A**	**G**	**A**	**G**	**A**	**C**	**G**	**G**	**C**	J04353.1
BRA1	·	·	·	T	·	·	·	·	T	·	·	·	A	G	·	·	T	·	·	·	DQ057308.1
BRA2	·	·	·	T	·	·	·	·	T	·	·	·	A	G	·	·	T	·	·	·	DQ057305.1
BRA3	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	JX185395.1
BRA4	·	C	·	·	G	·	·	·	T	·	·	·	·	·	·	G	T	·	·	·	GU595167.1
BRA5	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	·	·	·	·	FJ202003.1
BRA6	·	·	·	T	·	C	·	·	T	·	·	·	A	G	·	·	T	·	·	·	FJ202002.1
CR1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537667
CR2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537669
CR3	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	·	·	HQ537670
CR4	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537671
CR5	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537672
CR6	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537673
CR7	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537674
CR8	·	·	C	·	G	·	·	·	T	·	·	·	·	·	·	G	T	·	·	·	HQ537676
CR9	·	·	C	·	G	·	·	·	T	·	·	·	·	·	·	G	T	·	·	·	HQ537677
CR10	·	C	·	·	G	·	·	·	T	·	·	·	·	·	A	G	T	·	·	·	HQ537680
CR11	·	C	·	·	G	·	·	·	T	·	·	·	·	·	·	G	T	·	·	·	HQ537681
CR12	·	·	·	T	·	·	·	·	T	·	A	·	A	·	·	·	T	·	·	·	HQ537682
CR13	·	·	·	T	·	·	·	·	T	G	·	G	A	G	·	·	T	·	·	·	HQ537684
CR14	·	·	·	T	·	·	·	·	T	·	·	·	A	G	·	·	T	·	·	·	HQ537685
CR15	·	·	·	T	·	·	·	·	T	·	·	·	A	G	·	·	T	·	A	·	HQ537686
CR16	·	·	·	T	·	·	·	C	T	·	·	·	A	G	·	·	T	·	·	·	HQ537687
Mutación no sinónima	-	-	I52L	H60Y	T64A	-	-	-	-	-	E89K	-	-	-	R108K	K123R	A138V	C139Y	R142K	R144G	

Anexo L

Variabilidad genética de secuencias E7 de VPH 31. Los nucleótidos conservados se marcan con un punto (·) y sin cambio de aminoácidos con guion (-). Secuencia de referencia de VPH 31: Sec. Ref. (No Acceso: J04353.1). H: histidina, Y: tirosina, L: leucina, S: serina, P: prolina, D: aspartato, G: glicina, E: glutamina, K: lisina.

VPH 31 (E7)	Posiciones de sustituciones de nucleótidos en E7																Código GenBank
	574	580	592	626	642	645	648	654	670	695	697	716	743	745	807	822	
Sec. Ref.	**A**	**G**	**T**	**C**	**T**	**C**	**A**	**C**	**C**	**G**	**A**	**A**	**A**	**G**	**T**	**G**	J04353.1
BRA1	·	·	C	·	·	·	·	·	·	·	·	·	G	·	·	·	JX185396.1
BRA2	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	HQ149098.1
BRA3	·	·		T	·	·	·	·	T	A	·	·	G	·	·	·	GU595168.1
CR1	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	HQ537667
CR2	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	HQ537669
CR3	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	HQ537670
CR4	·	·	·	·	C	·	·	·	·	·	G	·	G	·	C	·	HQ537671
CR5	·	·	·	·	·	·	·	·	·	·	·	·	G	C	·	·	HQ537672
CR6	·	·	·	·	·	·	·	·	·	·	·	G	G	C	·	A	HQ537673
CR7	G	·	·	T	·	·	·	·	·	·	·	·	G	·	·	·	HQ537674
CR8	·	·	·	·	·	·	·	·	T	A	·	·	G	·	·	·	HQ537676
CR9	·	·	·	·	·	·	·	·	T	A	·	·	G	·	·	·	HQ537677
CR10	·	·	·	T	·	·	·	·	T	A	·	·	G	·	·	·	HQ537680
CR11	·	·	·	T	·	·	·	·	T	A	·	·	G	·	·	·	HQ537681
CR12	·	A	·	·	·	T	·	·	T	A	·	·	G	·	·	·	HQ537682
CR13	·	A	·	·	·	·	·	·	T	A	·	·	G	·	·	·	HQ537684
CR14	·	A	·	·	·	·	·	·	T	A	·	·	G	·	·	·	HQ537685
CR15	·	A	·	·	·	·	·	T	T	A	·	·	G	·	·	·	HQ537686
CR16	·	A	·	·	·	·	G	·	T	A	·	·	G	·	·	·	HQ537687
Mutación no sinónima	-	-	-	H23Y	L28S	P29L	D30G	S32L	-	E46K	-	-	K62E / K62D	-	L83S	G88E	

Anexo M

Variabilidad genética de secuencias E6 y E7 de VPH 33. Los nucleótidos conservados se marcan con un punto (·) y sin cambio de aminoácidos con guion (-). Secuencia de referencia de VPH 33: Sec. Ref. (No Acceso M12732.1). K: lisina, N: asparagina, P: prolina, T: treonina, S. serina, F: fenilalanina, I: isoleucina, L: leucina, V: valina, H: histidina, E: glutamato, A: alanina, Q: glutamina.

VPH 33	Posiciones de sustituciones de nucleótidos en E6												E7						Código GenBank
	213	214	227	315	325	355	364	387	449	480	521	534	606	722	737	833	850	862	
Sec. Ref.	**A**	**C**	**C**	**C**	**A**	**G**	**A**	**A**	**A**	**A**	**C**	**C**	**G**	**T**	**A**	**G**	**C**	**A**	M12732.1
BRA1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	JQ976760
BRA2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	JQ976761
BRA3	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	JQ976762
BRA4	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	JQ976763
BRA5	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	JQ976764
CR1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537689
CR2	·	·	·	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	HQ537690
CR3	C	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	T	HQ537695
CR4	C	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	T	HQ537696
CR5	C	·	·	·	·	·	C	·	·	T	·	·	·	·	G	·	·	T	HQ537698
CR6	C	·	·	·	·	·	C	·	·	T	·	·	·	·	·	·	·	T	HQ537701
CR7	·	·	·	T	C	T	·	C	·	·	T	T	A	C	·	T	G	T	HQ537705
CR8	·	A	·	T	C	T	·	C	·	·	T	T	·	C	·	T	G	T	HQ537707
CHI1	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KF5369
PAR1	C	·	·	·	·	·	C	·	·	T	·	·	·	·	·	·	·	·	JQ976767
PAR2		·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	JQ976768
PAR3		·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	JQ976769
PAR4	C	·	·	·	·	·	C	·	·	T	·	·	·	·	·	·	·	T	JQ976770
PAR5	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	JQ976771
PER1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	JQ976775
Mutación no sinónima	K 35 N	P 36 T	S 40 F	-	I 73 L	V 83 L	N 86 H	K 93 N	E 114 V	-	A 138 V	-	V 12 I	-	-	-	T 93 S	Q 97 L	

Señalada en rojo mutación descrita por primera vez en este estudio.

Anexo N

Variabilidad genética de secuencias E6 y E7 de VPH 35. Los nucleótidos conservados se marcan con un punto (·) y sin cambio de aminoácidos con guion (-). Secuencia de referencia de VPH 35: Sec. Ref. (No Acceso X74477.1). H: histidina, Y: tirosina, I: isoleucina, V: valina, W: triptófano, R: arginina.

VPH 35	Posiciones de sustituciones de nucleótidos en E6										E7		Código GenBank
	124	127	131	136	179	295	326	341	508	543	628	675	
Sec. Ref.	**A**	**T**	**C**	**T**	**C**	**A**	**A**	**T**	**A**	**G**	**T**	**T**	**X74477.1**
CR1	G	·	A	·	·	T	·	·	·	·	·	·	HQ537708
CR2	·	·	A	·	·	T	G	·	G	·	·	·	HQ537709
CR3	·	·	A	·	·	T	G	·	·	T	·	·	HQ537710
CR4	·	·	A	·	·	T	·	·	·		·	·	HQ537711
CR5	·	·	A	·	·	T	·	·	·	T	C	C	HQ537715
CR6	·	·	A	·	·	T	·	·	·	T	C	C	HQ537716
CR7	·	C	A	C	T	T	·	C	·	T	·	·	HQ537729
CR8	·	C	A	C	·	T	·	C	·	T	·	·	HQ537730
Mutación no sinónima	-	-	-	-	H 24 Y	-	I 73 V	W 78 R	-	R 145 I	Y 23 H	-	

Anexo Ñ

Variabilidad genética de secuencias E6 de VPH 52. Los nucleótidos conservados se marcan con un punto (·) y sin cambio de aminoácidos con guion (-). Secuencia de referencia de VPH 52: Sec. Ref. (No Acceso X74481.1). E: glutamina, D: aspartato, R: arginina, K: lisina, M: metionina, I: isoleucina, L: leucina, V: valina.

VPH 52	Posiciones de sustituciones de nucleótidos en E6																	Código GenBank
	143	200	237	251	252	253	296	308	348	350	379	404	425	471	503	530	548	
Sec. Ref.	**A**	**C**	**C**	**A**	**C**	**G**	**G**	**T**	**C**	**G**	**A**	**T**	**G**	**C**	**G**	**A**	**A**	**X74481.1**
ARG1	C	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	KJ675787.1
CR1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537731
CR2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537732
CR3	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537733
CR4	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	HQ537734
CR5	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	HQ537737
CR6	·	·	·	G	·	·	·	·	·	·	G	·	·	·	·	·	·	HQ537738
CR7	·	·	·	G	·	·	·	·	·	·	G	·	·	·	·	·	·	HQ537739
CR8	·	·	·	·	·	·	·	·	·	C	G	·	·	·	·	·	·	HQ537740
CR9	·	·	·	·	·	·	·	·	·	C	G	·	·	·	·	·	·	HQ537741
CR10	·	·	·	·	A	A	A	C	·	C	·	·	·	·	·	G	·	HQ537743
CR11	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	G	·	HQ537745
CR12	·	·	·	·	·	·	·	·	G	C	·	·	·	·	·	G	·	HQ537746
CR13	·	T	T	·	·	·	·	·	·	C	·	·	T	·	·	·	·	HQ537747
CR14	·	T	T	·	·	·	·	·	·	C	·	·	T	·	·	·	·	HQ537748
CR15	·	T	T	·	·	·	·	·	·	C	·	·	T	A	·	·	·	HQ537749
CR16	·	T	T	·	·	·	·	·	·	C	·	·	T	A	T	·	·	HQ537750
CR17	·	T	T	·	·	·	·	·	·	C	·	·	T	·	·	·	G	HQ537751
Mutación no sinónima	E 14 D	-	-	-	-	R 51 K	M 65 I	-	L 83 V	-	K 93 R	-	-	-	-	-	-	

Anexo O

Variabilidad genética de secuencias E7 de VPH 52. Los nucleótidos conservados se marcan con un punto (·) y sin cambio de aminoácidos con guion (-). Secuencia de referencia de VPH 57: Sec. Ref. (No Acceso X74481.1). T: treonina, I: isoleucina, S: serina, G: glicina, Y: tirosina, D: aspartato, H: histidina, N: asparagina, L: leucina, R: arginina.

VPH 52	Posiciones de sustituciones de nucleótidos en E7												Código GenBank
	573	662	681	706	707	727	733	742	751	766	801	848	
Sec. Ref.	**T**	**C**	**A**	**A**	**G**	**T**	**C**	**G**	**C**	**C**	**A**	**T**	**X74481.1**
CR1	·	·	·	·	·	·	·	·	·	·	·	·	HQ537731
CR2	·	·	·	·	·	·	·	·	·	·	·	·	HQ537732
CR3	·	·	·	·	·	·	·	·	·	·	·	·	HQ537733
CR4	·	·	·	·	·	·	·	·	·	·	·	·	HQ537734
CR5	·	·	·	·	·	·	·	·	·	·	·	·	HQ537737
CR6	·	·	·	·	·	·	·	·	·	·	·	·	HQ537738
CR7	·	·	·	·	·	·	·	·	·	·	·	·	HQ537739
CR8	·	·	·	·	·	·	·	·	T	·	G	·	HQ537740
CR9	·	·	·	·	·	·	·	·	T	·	G	·	HQ537741
CR10	A	·	·	G	A	G	·	A	·	·	G	·	HQ537743
CR11	A	T	·	G	A	G	T	A	·	·	G	G	HQ537745
CR12	A	T	·	G	A	G	T	A	·	·	G	G	HQ537746
CR13	C	·	·	·	·	·	·	·	·	A	G	·	HQ537747
CR14	C	·	·	·	·	·	·	·	·	A	G	·	HQ537748
CR15	C	·	·	·	·	·	·	·	·	A	G	·	HQ537749
CR16	C	·	·	·	·	·	·	·	·	A	G	·	HQ537750
CR17	C	·	G	·	·	·	·	·	·	A	G	·	HQ537751
Mutación no sinónima	-	T 37 I	-	S 52 G	-	Y 59 D	H 61 Y	D 64 N	-	H 72 N	-	L 99 R	

Señalada en rojo mutación descrita por primera vez en este estudio.

Anexo P

Variabilidad genética de secuencias E6 de VPH 58. Los nucleótidos conservados se marcan con un punto (·) y sin cambio de aminoácidos con guion (-). Secuencia de referencia de VPH 58: Sec. Ref. (No Acceso D90400.1). V: valina, I: isoleucina, D: aspartato, E: glutamato, N: asparagina.

VPH 58	Posiciones de sustituciones de nucleótidos en E6										Código GenBank
	136	143	187	193	245	260	307	367	373	398	
Sec. Ref.	**A**	**T**	**C**	**G**	**G**	**A**	**C**	**C**	**A**	**A**	**D90400.1**
ARGC1	·	·	·	·	A	·	T	·	·	G	JX896419.1
ARGC2	·	·	·	·	·	·	T	·	G	·	DQ866823.1
CR1	·	·	·	·	·	·	T	·	·	·	HQ537754
CR2	·	·	T	·	·	·	T	·	·	·	HQ537755
CR3	·	·	·	·	·	·	T	·	·	·	HQ537756
CR4	·	·	·	·	·	·	T	·	·	·	HQ537757
CR5	G	·	·	·	·	·	T	·	·	·	HQ537758
CR6	·	C	·	A	A	·	T	·	·	G	HQ537766
CR7	·	·	T	A	A	·	T	·	·	G	HQ537767
CR8	·	·	T	A	A	C	T	·	·	G	HQ537768
CR9	·	·	T	·	·	·	T	A	·	·	HQ537772
CR10	·	·	T	·	·	·	T	A	·	·	HQ537773
CR11	·	·	T	·	·	·	T	A	·	·	HQ537774
HON1	·	·	T	·	·	·	T	·	·	·	KC190280
Mutación no sinónima	-	-	-	-	V46I	-	-	D86E	-	N97D	

Señaladas en rojo las mutaciones descritas por primera vez en este estudio.

Anexo Q

Variabilidad genética de secuencias E7 de VPH 58. Los nucleótidos conservados se marcan con un punto (·) y sin cambio de aminoácidos con guion (-) y un sitios desconocidos con interrogación (?). Secuencia de referencia de VPH 58: Sec. Ref. (No Acceso D90400.1). N: asparagina, S: serina, T: treonina, I: isoleucina, G: glicina, R: arginina, H: histidina, C: cisteína, A: alanina, E: glutamato.

VPH 58	Posiciones de sustituciones de nucleótidos en E7															Código GenBank
	587	603	614	632	694	744	760	761	775	778	793	798	801	840	852	
Sec. Ref.	**A**	**A**	**A**	**C**	**G**	**T**	**G**	**G**	**T**	**T**	**A**	**C**	**C**	**C**	**T**	**D90400.1**
ARG1	·	·	·	·	·	G	C	A	·	·	G	T	A	T	·	KJ675787.1
BRA1	·	·	·	·	A	G	·	A	C	·	·	·	·	·	·	JX896414.1
CR1	·	·	G	·	A	G	·	A	·	C	·	·	·	·	·	HQ537754
CR2	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	HQ537755
CR3	·	·	·	T	·	G	A	·	·	·	·	·	·	·	·	HQ537756
CR4	·	·	·	T	·	G	A	·	·	·	·	·	·	·	·	HQ537757
CR5	·	·	·	T	·	G	A	·	·	·	·	·	·	·	·	HQ537758
CR6	G	·	·	·	·	G	C	A	·	·	G	T	A	T	·	HQ537766
CR7	G	·	·	·	·	G	C	A	·	·	G	T	A	T	·	HQ537767
CR8	G	·	·	·	·	G	C	A	·	·	G	T	A	T	·	HQ537768
CR9	·	·	·	·	·	G	·	·	·	·	G	T	A	T	C	HQ537772
CR10	·	·	·	·	·	G	·	·	C	·	G	T	A	T	C	HQ537773
CR11	·	·	·	·	·	G	·	·	·	·	G	T	A	T	C	HQ537774
COL1	?	T	·	·	A	G	·	A	·	·	·	·	·	·	·	JQ976849
GUA1	?	·	·	T	·	G	A	·	·	·	·	·	·	·	·	JQ976864
GUA2	?	·	·	·	·	G	·	·	·	·	·	·	·	·	·	JQ976858
Mutación no sinónima	N5S	-	D14G	T20I	G41R	-	G63H/ G63D/ G63S	-	-	C69R	T74A	T76E	-	-	-	

Señaladas en rojo las mutaciones descritas por primera vez en este estudio.