



**UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES
CENTRO DE CIENCIAS DE LA SALUD**

**VARIABILIDAD GENÉTICA DE LOS GENES L1 Y L2 DEL VIRUS DEL
PAPILOMA HUMANO (VPH) GENOTIPO 51 EN MUJERES CON
DISPLASIA CERVICAL Y CÁNCER CERVICOUTERINO EN EL CENTRO
DE MÉXICO**

**TESIS QUE PRESENTA
ELÍ DANIEL GARCÍA MARTÍNEZ
PARA OPTAR POR EL GRADO DE
MAESTRÍA EN INVESTIGACIÓN BIOMÉDICA**

**TUTOR PRINCIPAL: DR. ALEJANDRO ROSAS CABRAL
COTUTOR: DR. RAFAEL GUTIÉRREZ CAMPOS
ASESOR: MSP. RAÚL ARIAS ULLOA**

AGUASCALIENTES, AGS., A 31 DE OCTUBRE DE 2022



UNIVERSIDAD AUTÓNOMA
DE AGUASCALIENTES

CENTRO DE CIENCIAS BÁSICAS
Departamento de Química

CARTA DE VOTO APROBATORIO
COMITÉ TUTORAL

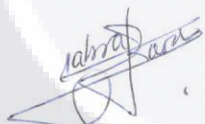
DRA. PAULINA ANDRADE LOZANO
DECANO (A) DEL CENTRO DE CIENCIAS DE LA SALUD
P R E S E N T E


Por medio del presente como **Miembros del Comité Tutorial** designado del estudiante **ELÍ DANIEL GARCÍA MARTÍNEZ** con ID **97309** quien realizó la tesis titulada: **VARIABILIDAD GENÉTICA DE LOS GENES L1 Y L2 DEL VIRUS DEL PAPILOMA HUMANO (VPH) GENOTIPO 51 EN PACIENTES CON DISPLASIA CERVICAL Y CÁNCER CERVICOUTERINO EN EL CENTRO DE MÉXICO**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia damos nuestro consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que nos permitimos emitir el **VOTO APROBATORIO**, para que él pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Ponemos lo anterior a su digna consideración y sin otro particular por el momento, le enviamos un cordial saludo.

ATENTAMENTE
"Se Lumen Proferre"

Aguascalientes, Ags., a 31 de octubre de 2022.


Dr. Alejandro Rosas Cabral
Tutor de tesis


Dr. Rafael Gutiérrez Campos
Co-Tutor de tesis


MSP. Raúl Arias Ulloa
Asesor de tesis

c.c.p.- Interesado
c.c.p.- Secretaría Técnica del Programa de Posgrado

Elaborado por: Depto. Apoyo al Posgrado.
Revisado por: Depto. Control Escolar/Depto. Gestión de Calidad.
Aprobado por: Depto. Control Escolar/ Depto. Apoyo al Posgrado.

Código: DO-SEE-FO-16
Actualización: 00
Emisión: 17/05/19



DICTAMEN DE LIBERACIÓN ACADÉMICA PARA INICIAR LOS TRÁMITES DEL EXAMEN DE GRADO



Fecha de dictaminación dd/mm/aaaa: 31/10/2022

NOMBRE: Elí Daniel García Martínez ID 97309

PROGRAMA: Maestría en Investigación Biomédica LGAC (del posgrado): Biomedicina Molecular y Celular

TIPO DE TRABAJO: (X) Tesis () Trabajo Práctico

TITULO: Variabilidad genética de los genes L1 y L2 del virus del papiloma humano (VPH) genotipo 51 en mujeres con displasia cervical y cáncer cervicouterino en el centro de México

IMPACTO SOCIAL (señalar el impacto logrado): La tesis proyecta la posibilidad de identificar la presencia de variabilidad genética del genotipo 51 de VPH en mujeres con displacia CU

INDICAR SI NO N.A. (NO APLICA) SEGÚN CORRESPONDA:

Elementos para la revisión académica del trabajo de tesis o trabajo práctico:	
Sí	El trabajo es congruente con las LGAC del programa de posgrado
Sí	La problemática fue abordada desde un enfoque multidisciplinario
Sí	Existe coherencia, continuidad y orden lógico del tema central con cada apartado
Sí	Los resultados del trabajo dan respuesta a las preguntas de investigación o a la problemática que aborda
Sí	Los resultados presentados en el trabajo son de gran relevancia científica, tecnológica o profesional según el área
Sí	El trabajo demuestra más de una aportación original al conocimiento de su área
Sí	Las aportaciones responden a los problemas prioritarios del país
no	Generó transferencia del conocimiento o tecnológica
Sí	Cumple con la ética para la investigación (reporte de la herramienta antiplagio)
El egresado cumple con lo siguiente:	
Sí	Cumple con lo señalado por el Reglamento General de Docencia
Sí	Cumple con los requisitos señalados en el plan de estudios (créditos curriculares, optativos, actividades complementarias, estancia, predoctoral, etc)
Sí	Cuenta con los votos aprobatorios del comité tutorial, en caso de los posgrados profesionales si tiene solo tutor podrá liberar solo el tutor
no	Cuenta con la carta de satisfacción del Usuario
Sí	Coincide con el título y objetivo registrado
Sí	Tiene congruencia con cuerpos académicos
Sí	Tiene el CVU del Conacyt actualizado
Sí	Tiene el artículo aceptado o publicado y cumple con los requisitos institucionales (en caso que proceda)
En caso de Tesis por artículos científicos publicados	
N.A.	Aceptación o Publicación de los artículos según el nivel del programa
N.A.	El estudiante es el primer autor
N.A.	El autor de correspondencia es el Tutor del Núcleo Académico Básico
N.A.	En los artículos se ven reflejados los objetivos de la tesis, ya que son producto de este trabajo de investigación.
N.A.	Los artículos integran los capítulos de la tesis y se presentan en el idioma en que fueron publicados
N.A.	La aceptación o publicación de los artículos en revistas indexadas de alto impacto

Con base a estos criterios, se autoriza se continúen con los trámites de titulación y programación del examen de grado:

Sí X
No

FIRMAS

Elaboró:

* NOMBRE Y FIRMA DEL CONSEJERO SEGÚN LA LGAC DE ADSCRIPCIÓN:

DR. RICARDO ERNESTO RAMÍREZ OROZCO

NOMBRE Y FIRMA DEL SECRETARIO TÉCNICO:

DR. LUIS FERNANDO BARBA GALLARDO

* En caso de conflicto de intereses, firmará un revisor miembro del NAB de la LGAC correspondiente distinto al tutor o miembro del comité tutorial asignado por el Decano

Revisó:

NOMBRE Y FIRMA DEL SECRETARIO DE INVESTIGACIÓN Y POSGRADO:

DR. RICARDO ERNESTO RAMÍREZ OROZCO

Autorizó:

NOMBRE Y FIRMA DEL DECANO:

DRA. PAULINA ANDRADE LOZANO

Nota: procede el trámite para el Depto. de Apoyo al Posgrado

En cumplimiento con el Art. 105C del Reglamento General de Docencia que a la letra señala entre las funciones del Consejo Académico: Cuidar la eficiencia terminal del programa de posgrado y el Art. 105F las funciones del Secretario Técnico, llevar el seguimiento de los alumnos.

Elaborado por: D. Apoyo al Posg.
Revisado por: D. Control Escolar/D. Gestión de Calidad.
Aprobado por: D. Control Escolar/ D. Apoyo al Posg.

Código: DO-SEE-FO-15
Actualización: 01
Emisión: 28/04/20

From: Online submission manuscript
Sent: Monday, October 31, 2022 1:30 PM
To: ELI DANIEL GARCIA MARTINEZ
Subject: GAMO - Article successfully registered

SOCIEDAD MEXICANA DE ONCOLOGÍA, A.C.
**MEXICAN JOURNAL
OF ONCOLOGY**
GACETA MEXICANA DE ONCOLOGÍA


PERMANYER MÉXICO
www.permanyer.com

Dear Doctor,

Thank you for your interest in our publication. We are writing to confirm that the article entitled '[Variabilidad genética de los genes L1 y L2 del virus del papiloma humano \(VPH\) genotipo 51 en pacientes con displasia cervical y cáncer cervicouterino](#)' (GAMO/0133/22) has been registered correctly.

The Editorial Committee of Gaceta Mexicana de Oncología will read your manuscript and we will shortly be in contact with you.

Many thanks,

Kind regards,
Saludos cordiales,
Oficina editorial



Temístocles 315, Dept. 404. - Col Polanco, Del. Miguel Hidalgo
México D.F., 11560 | dulce.bustamante@permanyer.com

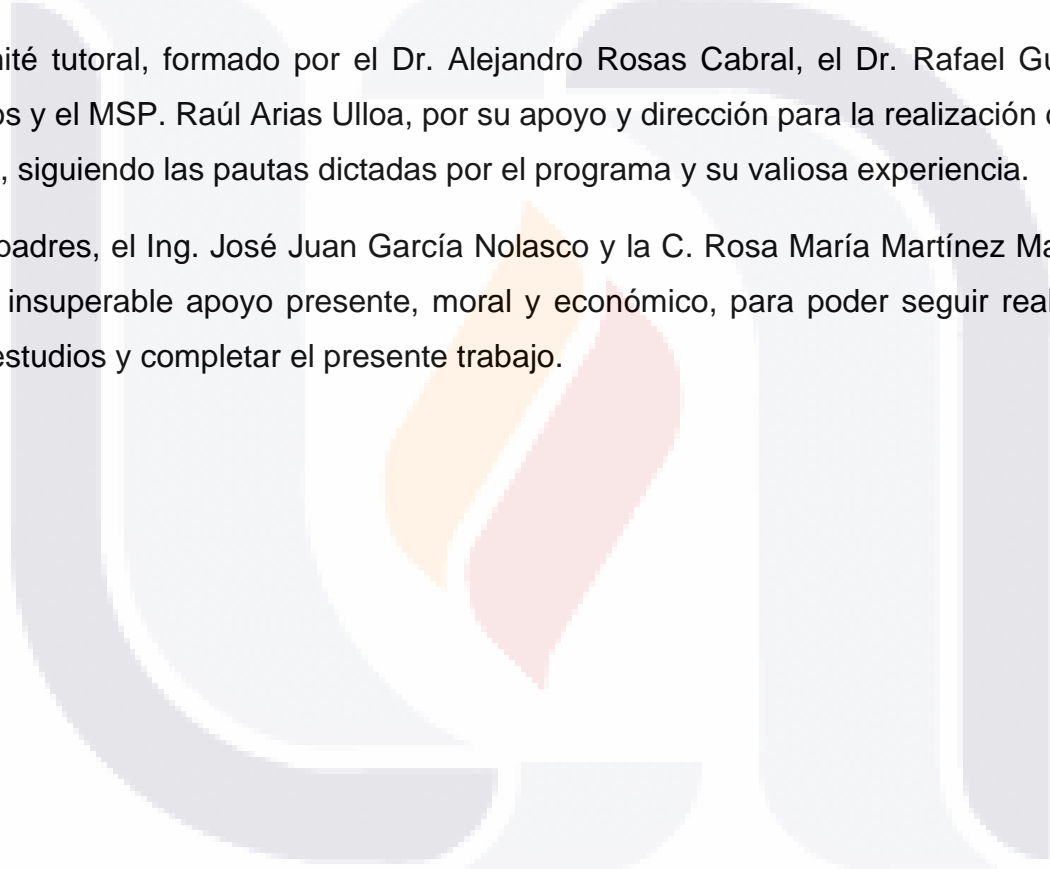
AGRADECIMIENTOS

A la Comisión Nacional de Ciencias y Tecnología (CONACYT) por el apoyo económico prestado en forma de beca para la realización de este trabajo.

A la Universidad Autónoma de Aguascalientes y al Centro de Ciencias de la Salud, por la oferta académica del programa de Maestría en Investigación Biomédica, así como el apoyo institucional, técnico y docente prestado para la realización de los estudios de Maestría que derivaron en la realización de este trabajo.

Al comité tutorial, formado por el Dr. Alejandro Rosas Cabral, el Dr. Rafael Gutiérrez Campos y el MSP. Raúl Arias Ulloa, por su apoyo y dirección para la realización de este trabajo, siguiendo las pautas dictadas por el programa y su valiosa experiencia.

A mis padres, el Ing. José Juan García Nolasco y la C. Rosa María Martínez Martínez, por su insuperable apoyo presente, moral y económico, para poder seguir realizando estos estudios y completar el presente trabajo.



DEDICATORIAS

A mis padres, por su presencia insuperable en mi vida y por siempre alentarme a la superación y realización personal; sin su apoyo nada hubiera sido posible, y pese a las dificultades que se presentaron, siempre han dado lo mejor para que fuesen superadas para toda nuestra familia.

A mi jefe, co-tutor y colaborador, el Dr. Rafael Gutiérrez Campos, que siempre ha mostrado disposición, comprensión y paciencia para orientarme en mi desarrollo académico y profesional, por darme las oportunidades necesarias para desarrollar mis habilidades como no lo hubiera podido hacer en otro lado.

Al resto de mi comité tutorial, el Dr. Alejandro Rosas Cabral y el MSP. Raúl Arias Ulloa, por su apoyo y colaboración en mi proyecto de tesis, con su experiencia y retroalimentación fortalecieron este proyecto.

A mi mejor amigo y siempre leal, el Lic. Jorge Luis Cisneros Ruiz, por su presencia, aliento y presencia durante mis etapas buenas y malas de mis estudios de maestría, sin su aportación y amistad nada en este proyecto hubiera sido posible y hubiera renunciado hace mucho tiempo.

A mi amigo, el Lic. Enrique Gómez Salinas, por ser la voz de la razón en muchos tiempos que necesité escuchar un consejo objetivo pero comprensivo al mismo tiempo. Gracias por tu continuada presencia en mi vida y apoyo para ingresar al programa.

A mis compañeros y amigos de la Maestría, que incluso ante los tiempos inciertos en los que cursamos nuestro programa nunca se dejaron vencer y ante todo pudimos mantener el contacto.

Y a todos los que, sin ser mencionados anteriormente, estuvieron involucrados directa o indirectamente en estos dos años de mi vida.

Este trabajo es para ustedes.

ÍNDICE GENERAL

RESUMEN	11
ABSTRACT	12
IDEA DE INVESTIGACIÓN.....	13
PLANTEAMIENTO DEL PROBLEMA	15
PREGUNTA DE INVESTIGACIÓN	17
JUSTIFICACIÓN	18
MARCO TEÓRICO	20
Fisiopatología del cáncer cervicouterino.	20
Epidemiología del cáncer cervicouterino.	22
Virus del papiloma humano (VPH): genotipos y su papel en el desarrollo del cáncer cervicouterino.	25
OBJETIVOS	30
Objetivo general.	30
Objetivos específicos.....	30
HIPÓTESIS.....	31
METODOLOGÍA	32
Diseño del estudio	32
Tipo de estudio.	32
Características de la muestra.	32
Criterios de selección.....	32
Variables por analizar.	34
Consideraciones éticas.....	38

Técnicas de recolección de datos.....	38
Análisis de laboratorio.	39
Análisis estadístico y bioinformático.....	42
Recursos disponibles para el estudio.....	44
Recursos humanos y materiales.....	44
Recursos financieros.	45
Difusión e impacto de la investigación.....	46
RESULTADOS.....	48
Verificación de la amplificación de los genes L1 y L2 de VPH 51 mediante alineamiento y PCR <i>in silico</i>	48
Estandarización de la técnica de amplificación de los genes L1 y L2 de VPH 51 mediante PCR punto final.....	50
Características de las pacientes analizadas.....	53
Análisis de polimorfismos y árboles filogenéticos en las secuencias génicas	54
Traducción <i>in silico</i> y análisis de mutaciones en secuencias proteicas.....	65
Ensamblaje de consensos y predicción de variaciones estructurales en L1 y L2.	72
DISCUSIÓN.....	76
CONCLUSIONES.....	83
REFERENCIAS.....	84
ANEXO 1: CARTA DE CONSENTIMIENTO INFORMADO.....	91
ANEXO 2: CUESTIONARIO ADJUNTO A LA TOMA DE MUESTRA.....	93
ANEXO 3: PROTOCOLO DE EXTRACCIÓN DE ADN DE MUESTRAS DE HISOPADO CERVICAL.....	95
ANEXO 4: SECUENCIAS DE OLIGONUCLEÓTIDOS UTILIZADAS EN LAS REACCIONES DE PCR.....	98

- A. Oligonucleótidos consenso E6/E798
- B. Oligonucleótidos para la genotipificación de VPH, reacciones multiplex ..98
- C. Oligonucleótidos para la amplificación específica de L1 y L2 de VPH 51
100



ÍNDICE DE TABLAS

Tabla 1: Características sociodemográficas y clínicas de las pacientes analizadas, distribuidas por grupo de lesión. Las variables continuas se muestran como media \pm SD, y las categóricas como n (%), separadas por categorías según corresponda a las politómicas, y a los resultados positivos en el caso de las variables dicotómicas.

.....53

Tabla 2: Polimorfismos detectados en el alineamiento múltiple de secuencias para el fragmento L1_1 y su prevalencia, expresada como porcentaje a la derecha de cada polimorfismo. Cada prevalencia superior al umbral de 20% se resalta en negrita, y cada polimorfismo encontrado en más de un grupo de lesión se señala sombreado.56

Tabla 3: Polimorfismos detectados en el alineamiento múltiple de secuencias para el fragmento L1_2 y su prevalencia, expresada como porcentaje a la derecha de cada polimorfismo. Cada prevalencia superior al umbral de 20% se resalta en negrita, y cada polimorfismo encontrado en más de un grupo de lesión se señala sombreado.56

Tabla 4: Polimorfismos detectados en el alineamiento múltiple de secuencias para el fragmento L2_1 y su prevalencia, expresada como porcentaje a la derecha de cada polimorfismo. Cada prevalencia superior al umbral de 20% se resalta en negrita, y cada polimorfismo encontrado en más de un grupo de lesión se señala sombreado.57

Tabla 5: Polimorfismos detectados en el alineamiento múltiple de secuencias para el fragmento L2_2 y su prevalencia, expresada como porcentaje a la derecha de cada polimorfismo. Cada prevalencia superior al umbral de 20% se resalta en negrita, y cada polimorfismo encontrado en más de un grupo de lesión se señala sombreado.57

Tabla 6: Análisis de neutralidad génica para las secuencias de cada fragmento, empleando D de Tajima. Todas las secuencias alineadas con éxito para cada fragmento son consideradas en el análisis, que emplea 200 iteraciones para realizar una predicción de selección o aneutralidad en el tiempo. m: número de secuencias

empleadas en el análisis. S: número total de sitios variables encontrados en el alineamiento. Ps: proporción de sitios variables respecto al número total de sitios probables de variación. θ : estadístico de significancia para la selección positiva o negativa. π : diversidad nucleotídica relativa. D: estadístico de Tajima.64

Tabla 7: Análisis de neutralidad génica para las secuencias de cada fragmento separadas por grupo de lesión, empleando D de Tajima. Todas las secuencias alineadas con éxito para cada fragmento, y para cada grupo de lesión, son consideradas en el análisis, que emplea 200 iteraciones para realizar una predicción de selección o aneutralidad en el tiempo. m: número de secuencias empleadas en el análisis. S: número total de sitios variables encontrados en el alineamiento. Ps: proporción de sitios variables respecto al número total de sitios probables de variación. θ : estadístico de significancia para la selección positiva o negativa. π : diversidad nucleotídica relativa. D: estadístico de Tajima.65

Tabla 8: Mutaciones no silenciosas detectadas en los fragmentos ensamblados de L1 para cada grupo de lesión, respecto a la secuencia de referencia. Las mutaciones detectadas en más de un grupo de lesión se muestran en sombreado.72

Tabla 9: Mutaciones no silenciosas detectadas en los fragmentos ensamblados de L1 para cada grupo de lesión, respecto a la secuencia de referencia. Las mutaciones detectadas en más de un grupo de lesión se muestran en sombreado.72

Tabla 10: Análisis de neutralidad génica para las secuencias de cada proteína consenso, empleando D de Tajima. Todas las secuencias alineadas con éxito para cada fragmento, y para cada grupo de lesión, son consideradas en el análisis, que emplea 200 iteraciones para realizar una predicción de selección o aneutralidad en el tiempo. m: número de secuencias empleadas en el análisis. S: número total de sitios variables encontrados en el alineamiento. Ps: proporción de sitios variables respecto al número total de sitios probables de variación. θ : estadístico de significancia para la selección positiva o negativa. π : diversidad nucleotídica relativa. D: estadístico de Tajima.74

ÍNDICE DE FIGURAS

Fig. 1: Esquema general de la evolución de las lesiones cervicales asociadas a la infección por VPH. De izquierda a derecha: NILM, LSIL, HSIL e IC. Tomada de (2).
.....21

Fig. 2: Incidencia y mortalidad de cáncer cervicouterino a nivel mundial. Las zonas más oscuras indican una mayor incidencia o mortalidad según corresponda. Tomado de (1).....23

Fig. 3: tasas ajustadas por edad de mortalidad por cáncer cervicouterino a nivel nacional, medidas en casos por cada 100,000 mujeres. La media nacional (5.40 por cada 100,000 se indica en color claro). Tomado de (2).24

Fig. 4: Cladograma circular de los diferentes grupos de papilomavirus que infectan a humanos (y algunos murinos como secuencias externas). Los genotipos de alto riesgo para cáncer cervicouterino (HR-VPH) se indican en rojo. Tomado de (40).26

Fig. 5: Esquema general del arreglo genómico del VPH 16. Incluye los grupos de genes de expresión temprana (E) y tardía (L). El tamaño del genoma de los papilomavirus varía según el genotipo, pero se encuentra normalmente entre 7000 y 8500 pb. Tomado de (40).....27

Fig. 6: Expresión de los genes virales del VPH de alto riesgo durante la infección del epitelio cervical. Los genes E6 y E7 se mantienen activos hasta la etapa de replicación viral, momento en cuya expresión es interrumpida por la fractura del genoma. E4 modula el ensamblaje de las partículas virales, y no se conocen puntos de fractura en este gen en todos los genotipos de alto riesgo. Tomado de (40)...28

Fig. 7: Diferencia en la expresión de E6 y E7 entre genotipos de bajo y alto riesgo, que determinan si la formación de lesiones neoplásicas puede ocurrir o no. Tomado de (40).....28

Fig. 8: Distintos escenarios de la infección por HR-VPH. Una infección activa conduce a la formación de neoplasias que pueden permanecer o no estáticas hasta su reactivación (latencia viral). El sistema inmune contribuye a resolver un gran número de infecciones, lo que explica los números bajos de lesiones de alto grado y cáncer respecto al total de infecciones por VPH. Los fenómenos específicos que

subyacen a la reactivación de la infección aún se encuentran en estudio. Tomado de (40).....29

Fig. 9: Mapa de amplificación de los oligonucleótidos dirigidos contra los genes L1y L2 del VPH 51. El genoma de referencia (GenBank: M62877.1) es circular debido a la forma episomal previa a la ruptura e integración al genoma hospedero. Cada fragmento amplificado incluye sus oligonucleótidos flanqueantes del mismo código de color. Se incluye además el fragmento de 223 pb amplificado en la genotipificación mediante PCR multiplex propuesta por Sotlar et al. (46) que permite la detección de VPH 51 (HPV51_gtype_sotlar). Elaborado con ayuda del software SnapGene (Insightful Science, disponible en www.snapgene.com).48

Fig. 10: Simulación de un gel de agarosa al 2.0% de las secuencias a amplificar de la fig. 9. MW; marcador de peso molecular 100 bp (Promega). Carril 1: HPV_gtype_sotlar. Carril 2: HPV51L1_1. Carril 3: HPV51L1_2. Carril 4: HPV51L2_1. Carril 5: HPV51L2_2. Elaborado con ayuda del software SnapGene (Insightful Science, disponible en www.snapgene.com).49

Fig. 11: Resumen de PCR in silico de los fragmentos de L1 y L2 estudiados. La longitud de cada fragmento corresponde con los estimados teóricos reportados previamente. El valor aceptable tanto de auto complementariedad como complementariedad 3' es menos a 3.00.50

Fig. 12: Genotipificación de muestras modelo, observadas tras una electroforesis en gel de agarosa al 2.0%. Una banda de 223 pb en la segunda reacción indica la presencia de VPH 51. Carriles 1-4: muestras genotipificadas. MW: marcador de peso molecular 100 pb (Jena Bioscience). Carril 5: primera amplificación, E7AS/LCR. Carril 6: muestra negativa para VPH.50

Fig. 13: Estandarización de técnicas de amplificación de L1 y L2 de VPH 51. Cada gel corresponde a un fragmento donde se probaron las variables de $[MgCl_2]$ y temperatura de alineamiento. La observación se realizó en un gel de agarosa al 1.5%. Los resultados positivos se muestran en el anexo 4. Para todos los gels, carril 1: 2.0 mM, 54 °C; carril 2: 2.5 mM, 54 °C; carril 3: 2.0 mM, 56 °C; carril 4: 2.5 mM, 56 °C; carril 5: 2.0 mM, 58 °C; carril 6: 2.5 mM, 58 °C; carril 7: 2.0 mM, 60 °C;

carril 8: 2.5 mM, 60 °C. MW: marcador de peso molecular 100 pb (Jena Bioscience).

.....52

Fig. 14: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias del fragmento L1_1. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -4691.87, B (NILM) = -2188.00, C (LSIL) = -1680.51, D (HSIL) = -1809.20, E (IC) = -2990.45.....60

Fig. 15: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias del fragmento L1_2. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -10333.74, B (NILM) = -3687.07, C (LSIL) = -2724.25, D (HSIL) = -3580.05, E (IC) = -5505.22.....61

Fig. 16: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias del fragmento L2_1. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -3716.46, B (NILM) = -2362.07, C (LSIL) = -1529.09, D (HSIL) = -1968.31, E (IC) = -1768.45.....62

Fig. 17: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias del fragmento L2_2. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -9161.55, B (NILM) = -2700.40, C (LSIL) = -3382.50, D (HSIL) = -3341.66, E (IC) = -3312.79.....63

Fig. 18: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias de aminoácidos resultantes de la traducción in silico del fragmento L1_1. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol:

A (global) = -2393.63, B (NILM) = -1159.52, C (LSIL) = -1326.81, D (HSIL) = -1096.13, E (IC) = -1134.72.....68

Fig. 19: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias de aminoácidos resultantes de la traducción in silico del fragmento L1_2. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol:

A (global) = -10108.37, B (NILM) = -3620.82, C (LSIL) = -2015.33, D (HSIL) = -2999.46, E (IC) = -4326.76.....69

Fig. 20: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias de aminoácidos resultantes de la traducción in silico del fragmento L2_1. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol:

A (global) = -2639.51, B (NILM) = -1866.42, C (LSIL) = -963.60, D (HSIL) = -1287.33, E (IC) = -1127.55.70

Fig. 21: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias de aminoácidos resultantes de la traducción in silico del fragmento L2_2. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol:

A (global) = -4182.39, B (NILM) = -1278.45, C (LSIL) = -1779.56, D (HSIL) = -1466.15, E (IC) = -1844.52.....71

Fig. 22: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias consenso calculadas para cada grupo de lesión. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: L1 (izquierda) = -1654.72, L2 (derecha) = -1584.73.....73

Fig. 23: Modelos 3D de las proteínas consenso L1 calculadas para cada grupo de lesión. En todos los casos, la predicción se realizó mediante configuración estándar a partir del modelo más cercano obtenido en .pb.75



RESUMEN

La infección por VPH es un factor necesario, mas no suficiente, para la aparición de lesiones cervicales neoplásicas que pueden progresar a cáncer cervicouterino. Entre los más de 200 genotipos de VPH caracterizados a la fecha, un grupo de 14 se consideran de alto riesgo para la progresión de las lesiones neoplásicas cervicales a cáncer; entre ellos se encuentra VPH 51, un genotipo con una prevalencia mundial estimada entre 0.5 y 4%. Debido a antecedentes de una prevalencia inusualmente alta de VPH 51 en pacientes con lesiones neoplásicas en evolución y pacientes con cáncer en la región centro de México, es de interés conocer si existen variantes regionales que pudieran tener un impacto adicional en la persistencia de dichas lesiones o en potenciales programas de vacunación si VPH 51 se incorpora al esquema de vacunación. Para ello, se realizó un estudio ambispectivo sobre 44 pacientes positivas a VPH 51 y que presentaran grados variables de evolución de lesiones cervicales o carcinoma, con el fin de determinar la variabilidad genética de los genes L1 y L2 de VPH 51 y determinar si existen variantes de interés. Mediante amplificación por PCR y secuenciación, se obtuvieron las secuencias de L1 y L2 de las muestras de pacientes, y se realizó el análisis bioinformático para determinar la presencia de polimorfismos, filogenia y traducción *in silico* para determinar posibles mutaciones, así como la posibilidad de selección presente sobre las mismas mediante análisis de neutralidad de D de Tajima. Se determinó la presencia de varios polimorfismos de interés, tanto únicos como comunes entre diferentes grupos de lesión, de los cuales sólo unos pocos producían alteraciones en la secuencia de aminoácidos traducida *in silico* de las secuencias consenso para cada grupo. En particular, se encontró en L1 la presencia común de los genotipos V277G y G278S, mientras que en L2 es de especial interés la secuencia variante EPIAP en los aminoácidos 95 a 99, que podría tener potencial inmunogénico para ser utilizado como vacuna. Se determinó que, pese a la presencia de variabilidad natural en las secuencias, no existe evidencia suficiente para afirmar la existencia de variantes regionales de VPH 51 en estos genes, y se recomiendan estudios más numerosos para confirmar o refutar esta evidencia.

ABSTRACT

HPV infection is a necessary, yet not enough, factor for the appearance of neoplastic cervical lesions that can progress to cervical cancer. Among the over 200 HPV genotypes characterized so far, a group of 14 is considered of high risk for the progression of neoplastic lesions to cancer; there HPV 51 can be found, a genotype with an estimated global prevalence between 0.5 and 4%. Due to a background of an unusually high HPV 51 prevalence in patients with neoplastic lesions currently in evolution and cancer patients in Central Mexico, it is of interest to determine if there exist any regional variants that could generate an additional impact in the persistence of such lesions or in potential vaccination programs if HPV 51 is ever added to vaccination schemes. For that, an ambispective study was performed with 44 HPV 51-positive patients in Central Mexico that presented variable degrees of evolution of neoplastic lesions or carcinoma, with the aim to determine the genetic variability of L1 and L2 genes of HPV 51 and determine if variants of interest exist. Using PCR amplification and sequencing, L1 and L2 genetic sequences were obtained from patient samples, and bioinformatic analysis was carried on determining the presence of polymorphisms, phylogeny and in silico translation to determine potential mutations, as well as the possibility of present selection using Tajima's D test of neutrality. The presence of several polymorphisms of interest was determined, both unique and shared among lesion groups, of which only a handful produced alterations in the amino acid sequence translated in silico from consensus sequences for each group. In particular, the ubiquitous presence of V277G and G278S polymorphisms in L1 was detected, as well as the variant sequence EPIAP in amino acid positions 95 o 99, the latter of which could have immunogenic potential to be used as a vaccine. It was concluded that, despite the presence of natural variability among the analyzed sequence, there is not enough evidence to confirm the existence of regional variants of HPV 51 in these genes, and more numerous studies are recommended to confirm or refute this evidence.

IDEA DE INVESTIGACIÓN

El cáncer cervicouterino es el cuarto tipo de cáncer más prevalente en mujeres a nivel mundial, tanto en incidencia como en mortalidad.

El virus del papiloma humano (VPH) es el agente etiológico principal involucrado en el desarrollo de neoplasias intraepiteliales cervicales y cáncer cervicouterino a nivel mundial. Existen más de 100 genotipos diferentes de VPH identificados a la fecha, que se clasifican de acuerdo con su riesgo oncogénico en alto riesgo (HR) y bajo riesgo (LR1 y LR2). El grupo de los genotipos de alto riesgo comprende los genotipos 16, 18, 31, 33, 39, 45, 51, 52, 56, 58, 59, 62, 66 y 68. El genotipo 51 (VPH 51) es el séptimo genotipo de alto riesgo más prevalente a nivel mundial (aprox. 4%) y no es tan estudiado globalmente como otros genotipos más prevalentes, tanto en pacientes sin displasia y con ella. Sin embargo, se ha reportado recientemente una prevalencia inusualmente elevada de este genotipo en pacientes con lesiones intraepiteliales de alto grado (HSIL) y carcinoma invasor de células escamosas (IC) en la región centro de México, siendo incluso más prevalente que VPH 16 – considerado globalmente como el principal genotipo involucrado en la progresión de las displasias cervicales a cáncer.

Se cuenta actualmente con esquemas de vacunación contra VPH que cubren genotipos altamente prevalentes como 16, 18, 31 y 33; sin embargo, incluso la vacuna nonavalente (aún disponible de manera limitada), VPH 51 no se encuentra cubierto bajo ningún esquema actual. Sumado a ello, hay evidencia de la existencia de variantes regionales de cada genotipo (menor variabilidad genética que la intergenotípica) que pudieran suponer un riesgo diferente para la infección y desarrollo de lesiones debido a la incapacidad de la inmunización para cubrir las variantes.

Debido a la escasa información respecto a la variabilidad genética de los genes L1 y L2 del VPH 51, particularmente en México, la idea de este proyecto de investigación es analizar dicha variabilidad en pacientes con distintos grados de lesión neoplásica cervical y cáncer cervicouterino. Dicho análisis será realizado de

TESIS TESIS TESIS TESIS TESIS

forma intra e intergrupala para determinar no sólo la variabilidad intratípica global, sino también sitios específicos de mutación que persisten en estadios de lesión de severidad creciente. El objetivo inmediato de esta investigación se perfila en obtener información sobre la posible existencia de variantes y subtipos regionales del VPH 51 en la zona centro de México, cubriendo en adición distintos grados de lesiones cervicales neoplásicas – lo que lo distingue de otros estudios en la materia que se enfocan sólo en la comparación de pacientes con lesiones generalizadas contra las que no presentan desarrollo neoplásico. Se espera obtener una información más precisa sobre esta variabilidad genética que pueda dar pie a futuras investigaciones en otras regiones del genoma del VPH 51, y así obtener un modelo genómico estándar para la región que a su vez contribuya a mejorar los métodos de diagnóstico, prevención y tratamiento de las displasias cervicales y cáncer cervicouterino atribuibles a VPH 51 y su interacción con otros genotipos.

El modelo experimental contemplado comprende el uso de un genoma estándar de VPH 51, con el cual se ha reportado el uso de dos reacciones de PCR punto final para cubrir la extensión total de los ORFs L1 y L2 del VPH 51, previamente confirmado por genotipificación en pacientes negativas para lesión intraepitelial o malignidad (NILM), lesiones intraepiteliales de bajo (LSIL) y alto grado (HSIL) y carcinoma invasor (IC). Cada ORF será cubierto por dos productos cuya secuenciación confirmará su solapamiento y permitirá además su análisis bioinformático para la búsqueda general de polimorfismos y su persistencia en estadios crecientes de severidad.

PLANTEAMIENTO DEL PROBLEMA

El cáncer cervicouterino es el cuarto tipo de cáncer más prevalente en mujeres a nivel mundial, tanto en incidencia como en mortalidad (1), afectando a más de 500 000 mujeres a nivel mundial, y más de 300 000 casos nuevos cada año. Las mayores tasas de incidencia se concentran de manera consistente en regiones de desarrollo bajo o medio-bajo, siendo África el continente con la prevalencia e incidencia más elevadas, seguidas del sureste asiático y América. En México, ocupa el segundo lugar en mortalidad por cáncer en mujeres, sólo después del cáncer de mama (2).

El virus del papiloma humano (VPH) es el agente etiológico principal involucrado en el desarrollo de neoplasias intraepiteliales cervicales y cáncer cervicouterino a nivel mundial. Existen más de 150 genotipos diferentes de VPH identificados a la fecha, que se clasifican de acuerdo con su riesgo oncogénico en alto riesgo (HR) y bajo riesgo (LR1 y LR2). El grupo de los genotipos de alto riesgo comprende los genotipos 16, 18, 31, 33, 45, 51, 52, 56, 58, 59, 62, 66 y 68 (3,4). En general, los genotipos 16 y 18 se encuentran en más del 85% de las muestras de pacientes cuyas lesiones neoplásicas progresan a cáncer, y son los más utilizados en el diagnóstico y pronóstico de progresión en severidad. El genotipo 51 (VPH 51) es el séptimo genotipo de alto riesgo más prevalente a nivel mundial (aprox. 4%) (5) y no es tan estudiado globalmente como otros genotipos más prevalentes, tanto en pacientes sin displasia y con ella. Sin embargo, se ha reportado recientemente una prevalencia inusualmente elevada de este genotipo en pacientes con lesiones intraepiteliales de alto grado (HSIL) y carcinoma invasor de células escamosas (IC) en la región centro de México, siendo incluso más prevalente que VPH 16 (6).

VPH 51 actualmente no está comprendido en ningún esquema de vacunación ofrecido a la población general. Estas vacunas se diseñan con base en secuencias antigénicas presentes en las proteínas de la cápside viral, que usualmente están codificadas por los ORF L1 y L2 del genoma viral (7). Incluso para los genotipos ya cubiertos, hay evidencia de la existencia de variantes regionales (3,8,9) de los

TESIS TESIS TESIS TESIS TESIS

mismos que pudieran suponer un riesgo para la infección y desarrollo de lesiones debido a la incapacidad de la inmunización para cubrir las variantes (10).

La información respecto a la variabilidad genética del VPH 51 aún es escasa, aunque se han detectado algunas variantes regionales. En particular, la información respecto a la variabilidad en los genes L1 y L2 del VPH 51 (7), por lo que son necesarios estudios enfocados a dicho análisis. Al conocimiento, a la fecha sólo hay un estudio reportado en la variabilidad de VPH 51 en cuatro genes: E6, E7, L1 y L2, realizado en China y que ha demostrado que, mientras en E6 y E7 no suele haber una gran variabilidad, sí existe la presencia de sustituciones en las secuencias de L1 y L2 que, si bien no alcanzan para distinguir variantes, sí ejemplifican que incluso en una región específica puede haber polimorfismos de cambio de sentido con potencial implicación en la virulencia, persistencia y potencial terapéutico de VPH 51 (11). Al existir evidencia de una alta prevalencia de VPH 51 en regiones específicas de México (6,12), puede enfocarse la investigación en el estudio preliminar de las diversas variantes regionales del genotipo que coexisten en la región (en caso de haberlas) y si tienen efecto sobre la progresión de lesiones neoplásicas a cáncer.

A la fecha, no hay estudios enfocados a la prevalencia de polimorfismos de VPH 51 para diferentes grados de lesión, por lo que los posibles mecanismos que ocasionarían una diferencia en la respuesta y evolución de la infección todavía se desconocen; no contribuye además que la mayor parte de la información depositada en bases de datos como Gene o Nucleotide (NCBI) provenga de fracciones aisladas de ORFs característicos para VPH 51, mientras que sólo hay evidencia de unos pocos genomas completos secuenciados (13) – los cuales proveen sólo de información básica sobre variantes genotípicas o técnicas de genotipificación.

El panorama entonces permite plantear un nuevo paradigma de investigación alrededor de VPH 51 a nivel regional. No sólo puede establecerse un estudio pionero sobre la variabilidad genética y genómica de VPH 51 para diferentes grados de lesión en pacientes atendidas en la región – para lo cual hay escasa o nula

información disponible y extrapolable para el diagnóstico –, sino que además se contribuye al acervo genético que permita refinar los estudios filogenéticos existentes sobre VPH 51, quizá con el descubrimiento de variantes no identificadas a la fecha. Toda esta información puede contribuir al desarrollo de técnicas de diagnóstico e incluso desarrollo de vacunas contra VPH más específicas para la población mexicana, afectada de diferente manera por los genotipos ya cubiertos por las vacunas disponibles. La información recolectada de los estudios de variabilidad sobre VPH 51, en principio de ORFs individuales y posteriormente a nivel genómico, es esencial para comprender el contexto general de la distribución regional y severidad de la infección por VPH 51, tanto en mono-infección como en interacción con otros genotipos de alto riesgo oncológico.

PREGUNTA DE INVESTIGACIÓN

¿Cuál es el grado de variabilidad en la secuencia de los genes L1 y L2 del virus del papiloma humano (VPH) genotipo 51 en mujeres del centro de México que presentan distintos grados de displasia cervical o cáncer cervicouterino?

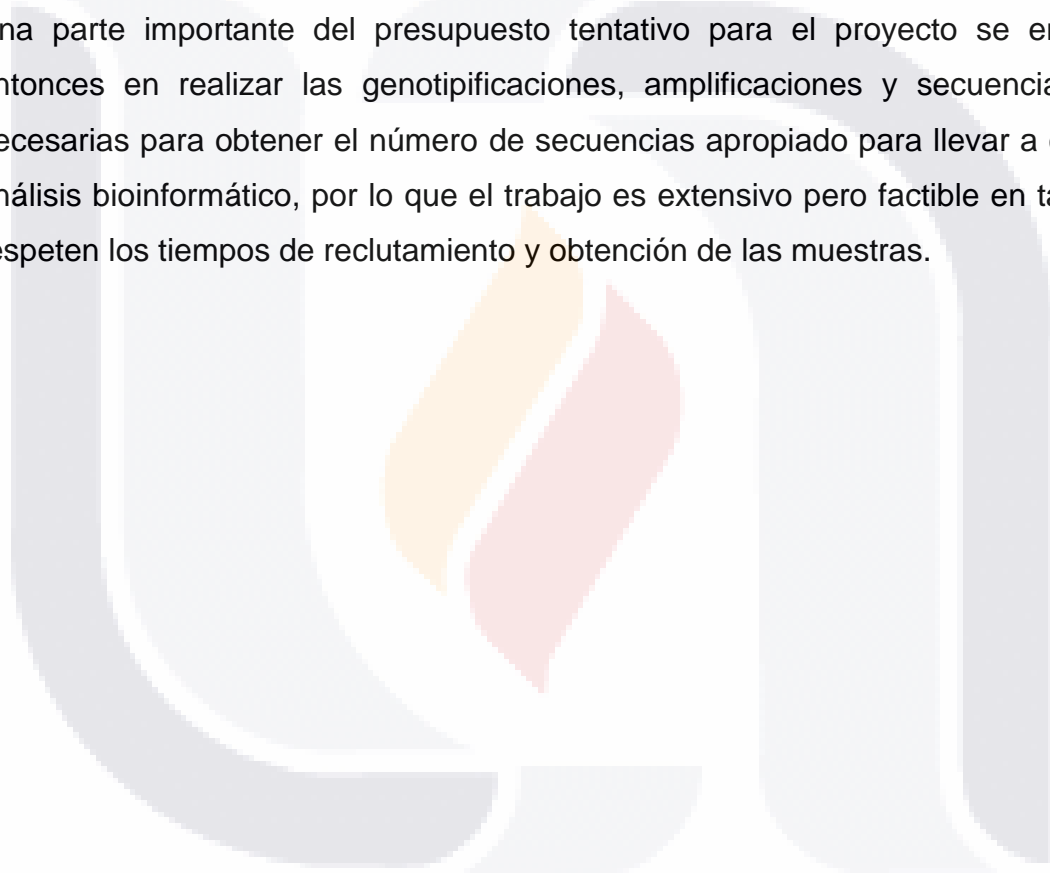
JUSTIFICACIÓN

La nueva información disponible respecto a la prevalencia de genotipos de VPH de alto riesgo que no están cubiertos dentro de los esquemas de vacunación se ha incrementado de manera importante en los últimos años, y dado que dichos genotipos y sus variantes no siempre se encuentran caracterizados al mismo nivel que aquellos genotipos más prevalentes según la literatura, es de importancia realizar estudios para detallar la variabilidad genética regional de los mismos, y así definir si existen diferencias potenciales en virulencia, persistencia y progresión de lesiones neoplásicas causadas por la infección de estos virus tanto en mono como en co-infección. Respecto a VPH genotipo 51, la información respecto a su variabilidad genética es escasa y apenas ha comenzado a caracterizarse, por lo que los bancos de información sobre secuencias de genes y genomas completos aún comprenden unas cuantas variantes aisladas y pocos genomas completos secuenciados. Un estudio de variabilidad a nivel genético, que dé hincapié a estudios de WGS de aislados de VPH 51 obtenidos de la región centro de México – reportada como una zona donde la prevalencia de este genotipo es inusualmente elevada – puede contribuir de forma importante al acervo genético y a una caracterización más profunda de variantes regionales al realizar la comparación con otras secuencias ya reportadas en la literatura.

Este estudio se enfocará en el análisis de variabilidad de dos de los genes conocidos de VPH 51, L1 (que codifica para la proteína de la cápside mayor viral) y L2 (cápside menor). Estos dos genes se cuentan entre los principales estudiados a nivel inter e intragenotípico para la caracterización de variantes, y cuentan con la ventaja adicional de que ya hay reportes sobre reacciones de amplificación específica para estos genes en VPH 51, las cuales se basan en PCR punto final y secuenciación Sanger automatizada – ambas técnicas ya comunes y cuyo costo por reacción ha disminuido drásticamente en las dos últimas décadas, por lo que ya son más accesibles desde el punto de vista económico-operacional. Más aún, es con base en L1 y L2 como se diseñan las vacunas contra VPH de primera y segunda generación, por lo que el conocimiento de las variantes regionales características

podría dar lugar, además, a un posible desarrollo de vacunas dirigidas de modo más específico a la población mexicana donde este genotipo es más prevalente.

Debido a la longitud de ambos genes (cada uno supera 1.5 Kbp), no es confiable realizar una amplificación única para cada gen, por lo que la técnica sugerida emplea dos reacciones para cada secuencia, que se solapan y de esta forma cubren toda la extensión del gen. La región de solapamiento puede servir como una verificación adicional de que la secuenciación se ha llevado a cabo exitosamente. Una parte importante del presupuesto tentativo para el proyecto se enfocará entonces en realizar las genotipificaciones, amplificaciones y secuenciaciones necesarias para obtener el número de secuencias apropiado para llevar a cabo el análisis bioinformático, por lo que el trabajo es extensivo pero factible en tanto se respeten los tiempos de reclutamiento y obtención de las muestras.



MARCO TEÓRICO

Fisiopatología del cáncer cervicouterino.

El cáncer cervicouterino es un tipo de cáncer heterogéneo en morfología y progresión que se origina en la región cervical (14). De acuerdo con su morfología, se puede distinguir entre cáncer de células escamosas – el más común -, adenocarcinoma, y un tercer tipo de cáncer menos común denominado carcinoma adenoescamoso (15). Sin embargo, se sabe que esta clasificación por sí misma puede resultar ambigua, debido al diverso grupo de linajes celulares involucrados en el desarrollo de las lesiones neoplásicas, en particular aquéllos asociados de manera menos específica con uno o un grupo de factores de riesgo como las infecciones virales (16).

Histológicamente, la región cervical está compuesta por tres zonas, de las cuales la zona de transformación (transición entre el ectocérvix y el endocérvix) es la más importante en el desarrollo de las lesiones cervicales que progresan a cáncer (15,17). En esta zona son comunes los fenómenos metaplásicos, que modifican el epitelio columnar y las células de reserva a un fenotipo estratificado, de forma natural desde la pubertad cuando el cérvix se ve expuesto al ambiente ácido de la vagina, pero también como resultado de una respuesta local a la irritación (15). La mayor parte de las lesiones intraepiteliales escamosas inician en esta región o en la región ectocervical, mientras que las encontradas en el endocérvix suelen derivan a procesos patológicos glandulares que derivan en adenocarcinoma (18). Sin embargo, estudios recientes han demostrado que los fenómenos que subyacen al control estricto del proceso metaplásico y su desregulación no son tan sencillos, con poblaciones celulares localizadas de manera adyacente al epitelio de transformación que pudieran jugar un papel regulatorio de la homeostasis de todo el epitelio cervical (19,20).

Las lesiones precancerosas evolucionan de forma progresiva, y su clasificación obedece a varios sistemas ideados de acuerdo a las características histopatológicas observadas en las biopsias (21) (fig. 1). En la actualidad, el sistema Bethesda –

creado en 1988 y refinado en 2001 y 2014 – es el recomendado para clasificar las lesiones cervicales precancerosas tanto escamosas como glandulares (22,23), y de forma general divide las lesiones escamosas en: negativas para lesión intraepitelial o malignidad (NILM), lesiones escamosas intraepiteliales de bajo (LSIL) y alto grado (HSIL) y carcinomas invasores de células escamosas (IC) que corresponden propiamente a las distintas etapas de evolución de cáncer cervicouterino (24). Las lesiones glandulares se clasifican en NILM, células atípicas (AC) y endocarcinomas (EC). Este sistema ha sido adecuado además para incluir interpretaciones también de biopsias líquidas, lo que permite su compatibilidad con métodos menos invasivos de muestreo (23,25).

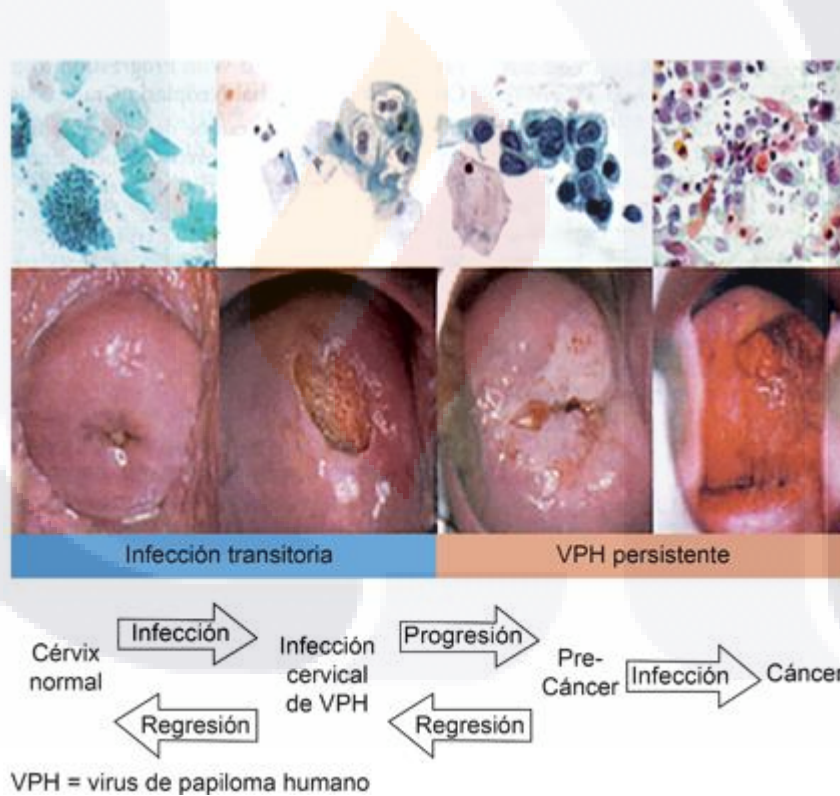


Fig. 1: Esquema general de la evolución de las lesiones cervicales asociadas a la infección por VPH. De izquierda a derecha: NILM, LSIL, HSIL e IC. Tomada de (2).

Entre los factores de riesgo más comunes para cáncer cervicouterino se encuentran:

- Infección por virus del papiloma humano (abordada más adelante).
- Conducta sexual activa: inicio temprano de la actividad sexual, múltiples parejas sexuales, conductas sexuales de alto riesgo (26).
- Fumar, que incrementa hasta dos veces el riesgo de padecer cáncer cervicouterino que una mujer que no fuma, además de otros tipos de cáncer (27,28).
- Sistema inmune comprometido (incluye la infección por VIH) (29).
- Otras infecciones propias del aparato reproductor femenino, tales como *Chlamydia trachomatis*, HSV 2, etc. (30,31)
- Uso prolongado de anticonceptivos orales (carcinogénesis inducida por estrógenos) (32).
- Estatus socioeconómico bajo, asociado con una pobre educación sexual y reproductiva (33).
- Factores endógenos: genéticos, hormonales e inmunológicos (34,35).

Epidemiología del cáncer cervicouterino.

De acuerdo a las estadísticas presentadas en el informe GLOBOCAN 2018, el cáncer cervicouterino es el cuarto tipo de cáncer más prevalente en mujeres a nivel mundial, con aproximadamente 570 000 mujeres afectadas al año y una incidencia de más de 300 000 casos nuevos cada año (1). Estandarizando por edad, el este y sureste africano son las zonas con mayores proporciones de prevalencia e incidencia (en algunos países, superiores a 40 por cada 100 000 mujeres), seguidos por el sureste asiático y América Central y el Caribe (fig. 1), aunque en números brutos China e India contribuyen en conjunto a aproximadamente 35% de todos los casos de cáncer cervicouterino a nivel mundial (36).

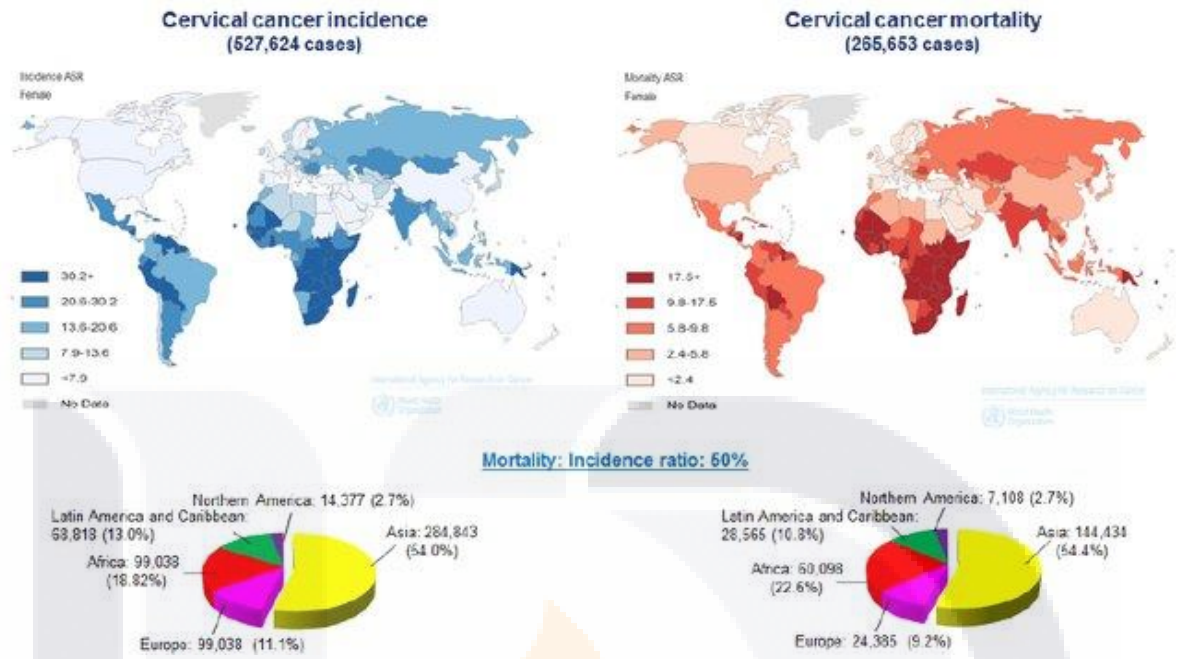


Fig. 2: Incidencia y mortalidad de cáncer cervicouterino a nivel mundial. Las zonas más oscuras indican una mayor incidencia o mortalidad según corresponda. Tomado de (1).

En lo que respecta a México, el cáncer cervicouterino ocupa el segundo lugar en prevalencia y mortalidad (sólo después del cáncer de mama) (2), con una tasa de mortalidad de aprox. 6.0 casos por cada 100 000 mujeres y una tasa de incidencia ligeramente superior a 10 casos por cada 100 000 años-mujer, equiparable con países de un nivel de desarrollo similar como Brasil o Colombia (36). La tasa de mortalidad en México por cáncer cervicouterino ha disminuido considerablemente desde un 2.5% en los años noventa, y un 1.5% adicional en la primera década del siglo XXI (37), pero de manera no homogénea a lo largo del territorio nacional: estados como Morelos, Tamaulipas o Quintana Roo reportan tasas de mortalidad superiores a 9.0 por cada 100 000 mujeres, mientras que Durango y Baja California Sur se ubican por debajo de 2.0 muertes por cada 100 000 mujeres (2). La región centro de México se ubica alrededor de la media nacional (fig. 3).

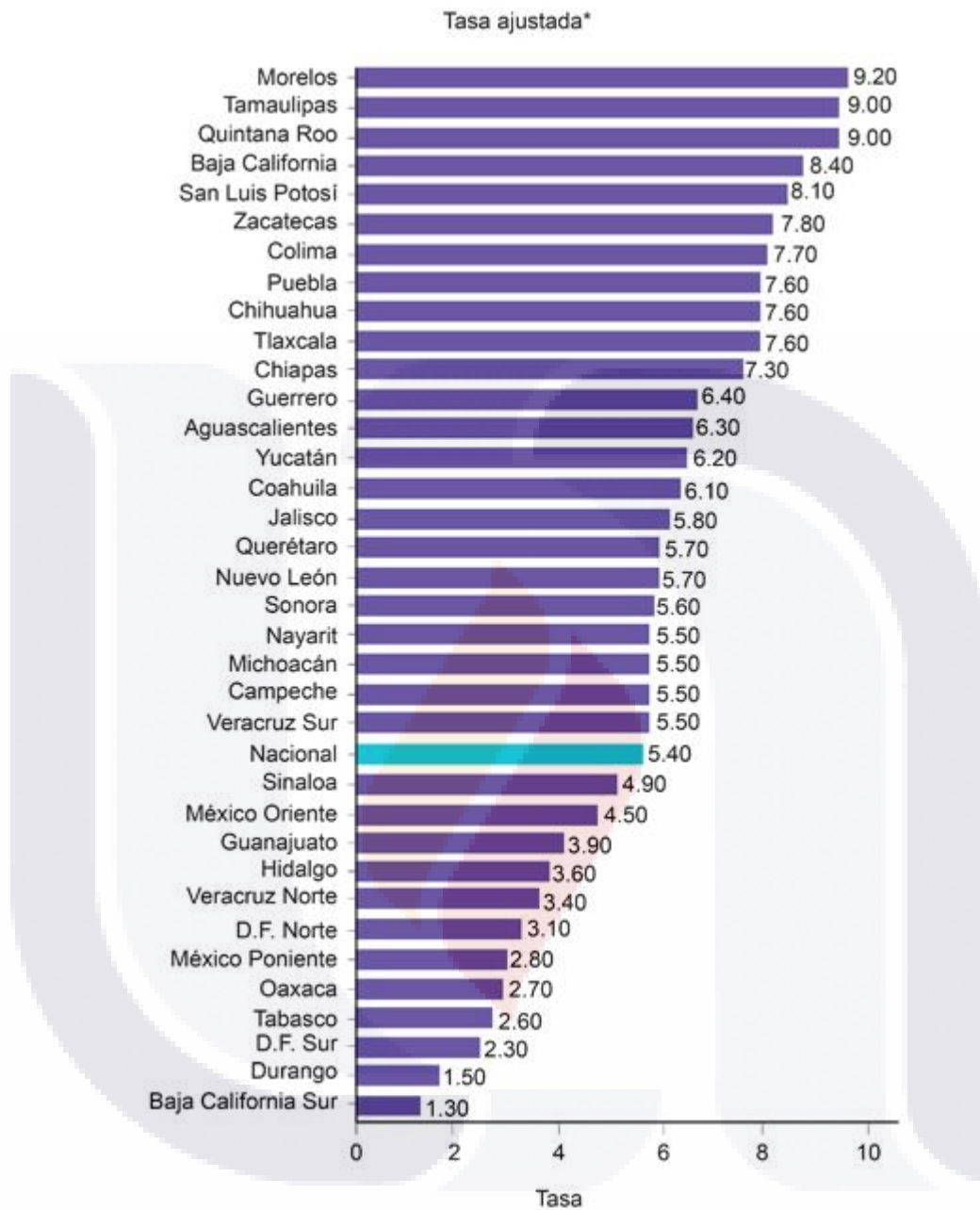


Fig. 3: tasas ajustadas por edad de mortalidad por cáncer cervicouterino a nivel nacional, medidas en casos por cada 100,000 mujeres. La media nacional (5.40 por cada 100,000 se indica en color claro). Tomado de (2).

Virus del papiloma humano (VPH): genotipos y su papel en el desarrollo del cáncer cervicouterino.

El virus del papiloma humano (VPH) es el agente etiológico necesario, más no esencial, para el desarrollo de lesiones neoplásicas cervicales y cáncer cervicouterino (38). Se estima que en más del 90% de todos los casos de este cáncer, uno o más genotipos de VPH están involucrados (39).

A la fecha, se han secuenciado los genomas de más de 150 genotipos de VPH que infectan a humanos, y más de 60 a animales (3,8,40) (fig. 4).. Entre los papilomavirus que infectan al ser humano se pueden encontrar los grupos alfa, beta, gama, mu y nu. La gran mayoría de los virus selectivos que infectan mucosas producen infecciones asintomáticas o pueden producir sólo la aparición de verrugas (genotipos de bajo riesgo, LR-VPH), pero existe un grupo selecto de genotipos (todos alfa) que son los principales asociados a la aparición de lesiones neoplásicas: 16, 18, 31, 33, 39, 45, 51, 52, 56, 58, 59, 62, 66 y 68. En particular, los genotipos 16 y 18 están reportados como los más prevalentes – en casi el 80% de todos los casos de cáncer cervical a nivel mundial (38,41).

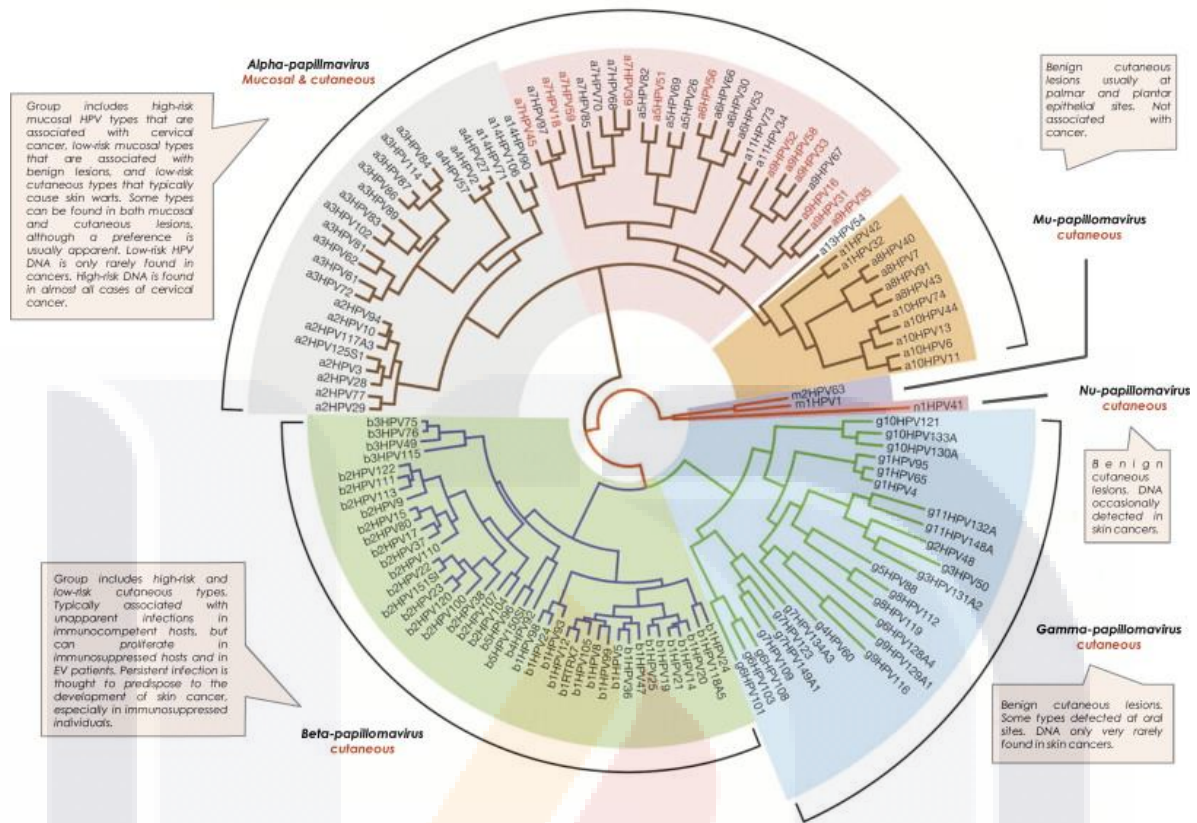


Fig. 4: Cladograma circular de los diferentes grupos de papilomavirus que infectan a humanos (y algunos murinos como secuencias externas). Los genotipos de alto riesgo para cáncer cervicouterino (HR-VPH) se indican en rojo. Tomado de (40).

El genoma de la mayor parte de los alfa papilomavirus (y encontrándose análogos también en los grupos beta y gama) sigue la estructura básica bien definida para VPH 16 (3,4,40) (fig. 5). Se trata de un genoma de ADN circular de cadena sencilla que está conformado por dos grandes grupos de marcos de lectura: genes de expresión temprana (genes E) que incluyen los genes para la replicación y oncogenes necesarios para la estabilidad y supervivencia viral, y genes de expresión tardía (L) que codifican para las proteínas mayores (L1) y menores (L2) de la cápside. Además, se incluye una región larga de control (LCR) que sirve como región promotora y reguladora de la expresión génica viral, y uno o más puntos de fractura donde se rompe la estructura circular y se pasa de una conformación genómica episomal a la integrativa, que se inserta en el genoma hospedero (3).

La expresión adecuada de los genes, en particular de los oncogenes E6 y E7, es clave para el desarrollo de la neoplasia intraepitelial y es una de las características

definitivas que distinguen a los genotipos de alto riesgo del resto (fig. 6). La interacción de E6 y E7 con reguladores clave del ciclo celular y la homeostasis, como p53 y pRB, se encuentra elevada en los genotipos HR, mientras que en los LR la expresión de E6 y E7 es nula o mínima y el efecto oncogénico no existe (42,43). Sin embargo, durante la infección suele ocurrir un periodo de latencia donde el virus se inactiva o la lesión recede debido al fracaso de la infección (fig. 7); los casos que desarrollan lesiones de alto grado y carcinomas son el resultado de una reactivación de la infección, la integración exitosa del genoma viral al hospedero y la desregulación producto de la actividad oncogénica (8,44,45).

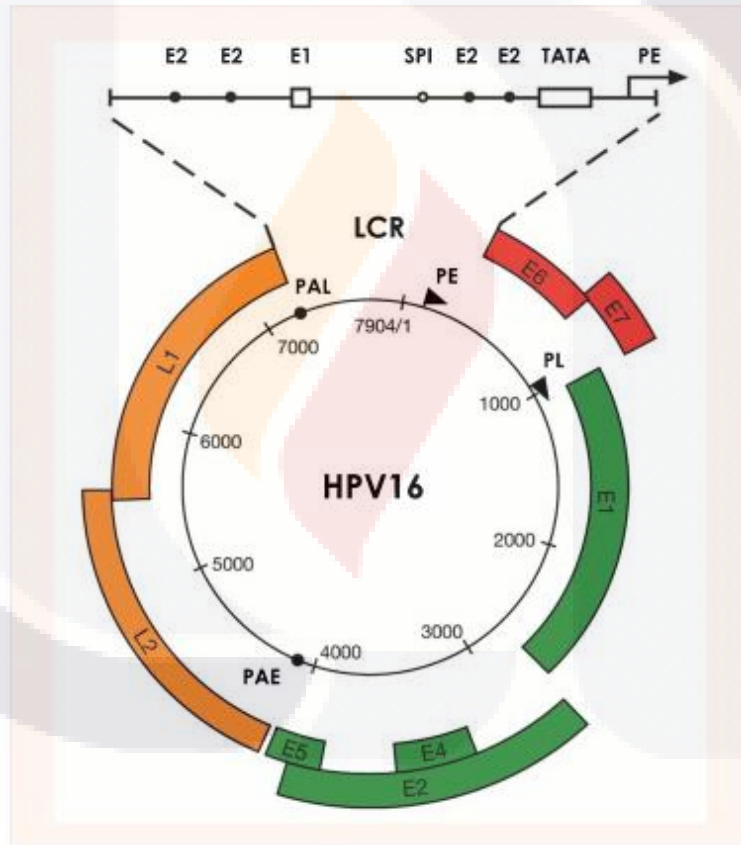


Fig. 5: Esquema general del arreglo genómico del VPH 16. Incluye los grupos de genes de expresión temprana (E) y tardía (L). El tamaño del genoma de los papilomavirus varía según el genotipo, pero se encuentra normalmente entre 7000 y 8500 pb. Tomado de (40).

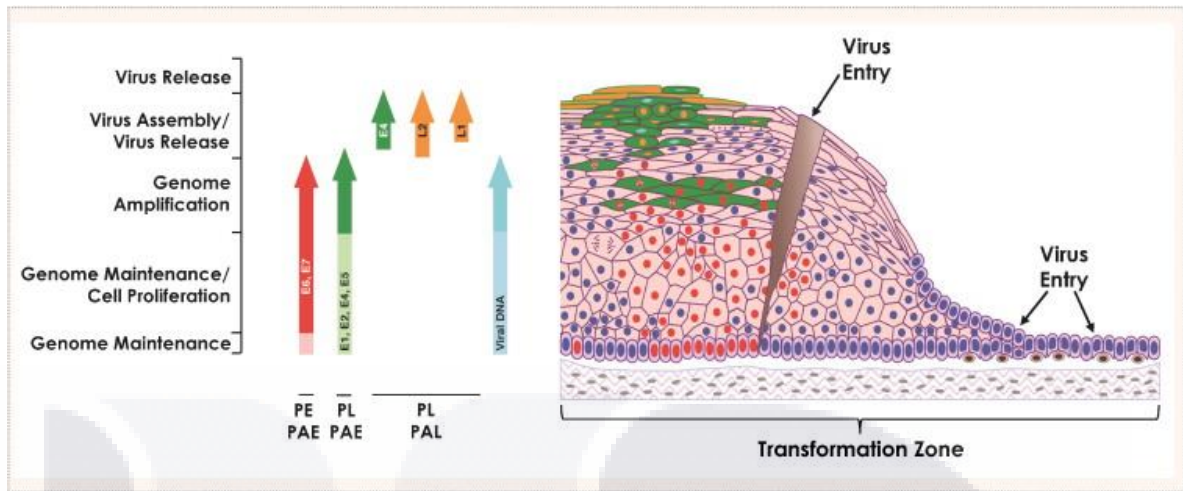


Fig. 6: Expresión de los genes virales del VPH de alto riesgo durante la infección del epitelio cervical. Los genes E6 y E7 se mantienen activos hasta la etapa de replicación viral, momento en cuya expresión es interrumpida por la fractura del genoma. E4 modula el ensamblaje de las partículas virales, y no se conocen puntos de fractura en este gen en todos los genotipos de alto riesgo. Tomado de (40).

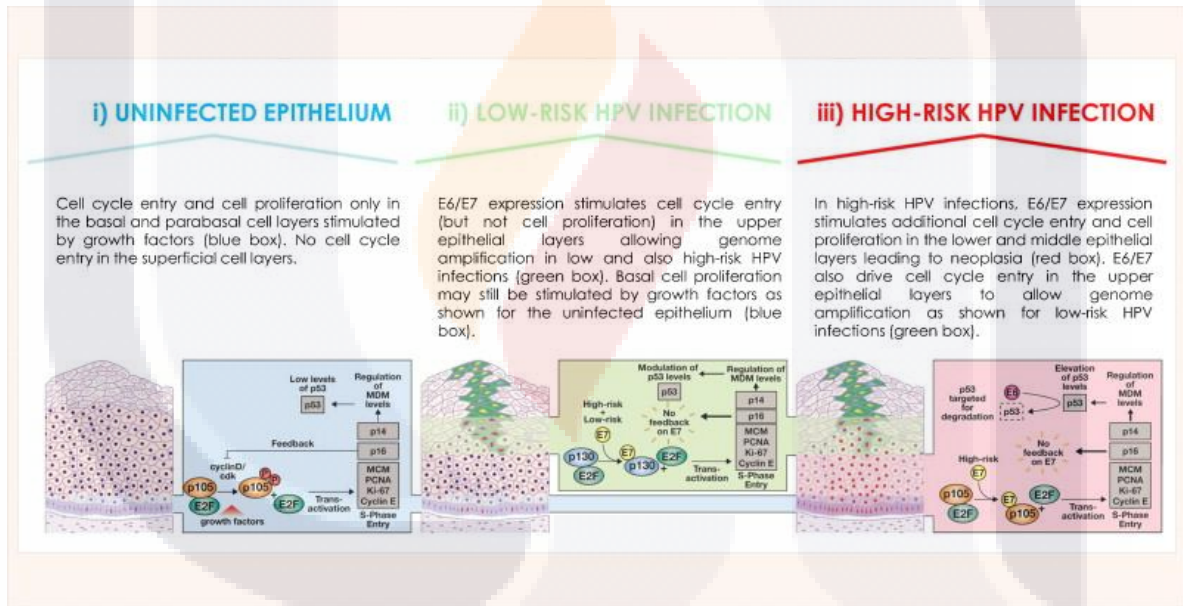


Fig. 7: Diferencia en la expresión de E6 y E7 entre genotipos de bajo y alto riesgo, que determinan si la formación de lesiones neoplásicas puede ocurrir o no. Tomado de (40).

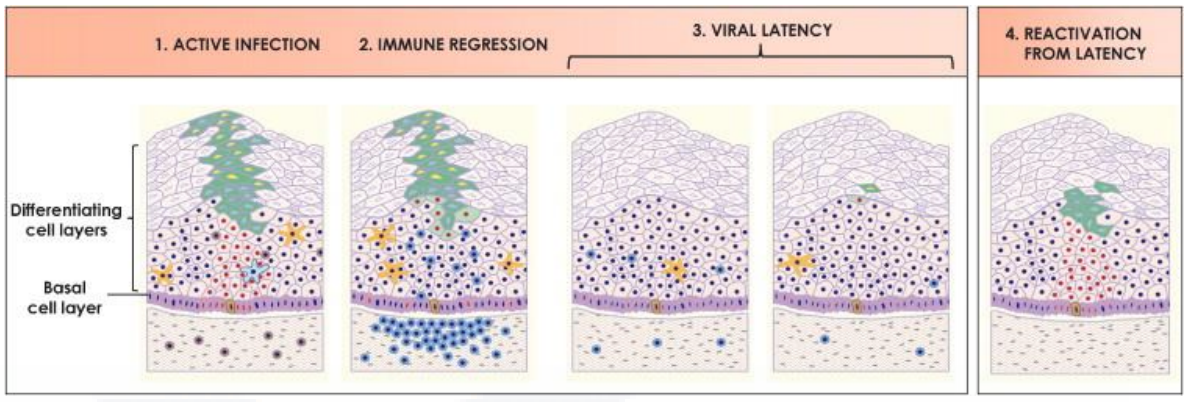
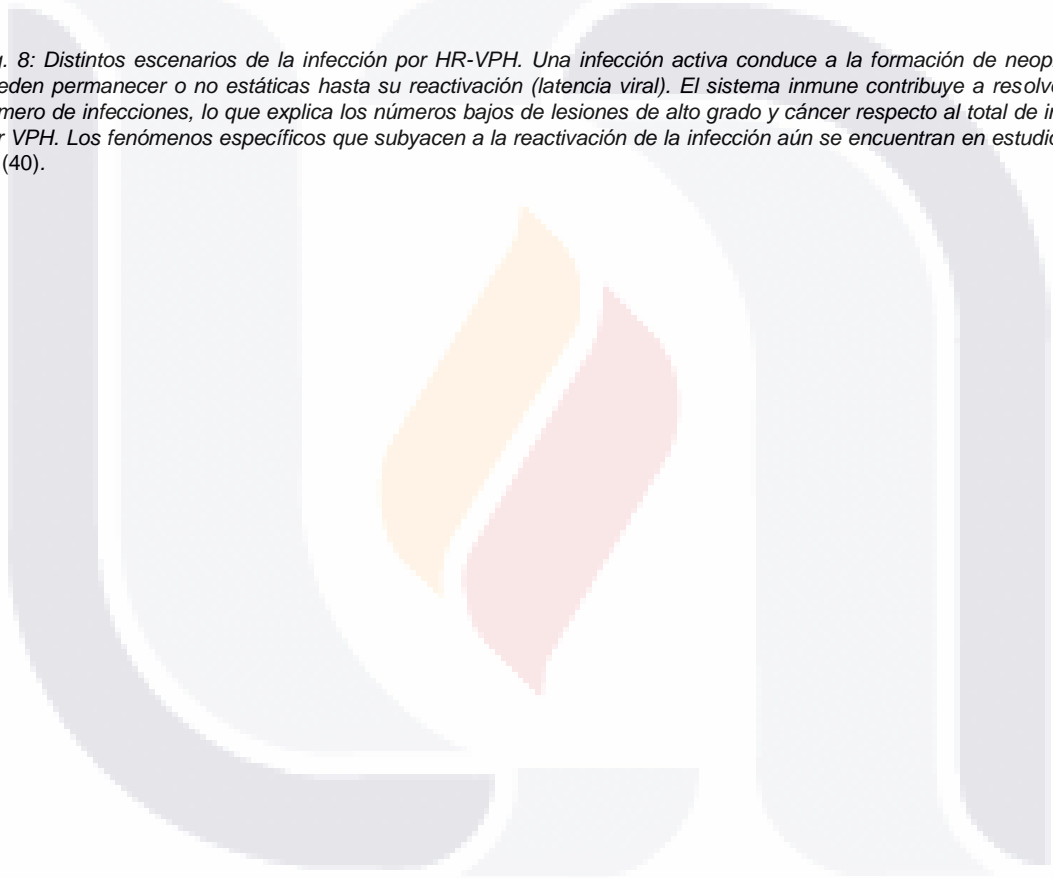


Fig. 8: Distintos escenarios de la infección por HR-VPH. Una infección activa conduce a la formación de neoplasias que pueden permanecer o no estáticas hasta su reactivación (latencia viral). El sistema inmune contribuye a resolver un gran número de infecciones, lo que explica los números bajos de lesiones de alto grado y cáncer respecto al total de infecciones por VPH. Los fenómenos específicos que subyacen a la reactivación de la infección aún se encuentran en estudio. Tomado de (40).



OBJETIVOS

Objetivo general.

Determinar el grado de variabilidad genética en los genes L1 y L2 del virus del papiloma humano (VPH) genotipo 51 en pacientes con displasias cervicales de gravedad creciente y cáncer cervicouterino en la región centro de México.

Objetivos específicos.

- Estandarizar las técnicas de amplificación de los genes L1 y L2 específicas para el VPH genotipo 51 a partir de ADN de muestras de pacientes con distintos grados de lesiones cervicales y cáncer cervicouterino.
- Analizar las secuencias obtenidas de la amplificación de los genes L1 y L2 del VPH 51 mediante alineamientos múltiples y filogenia.
- Determinar la ubicación y prevalencia de polimorfismos en las secuencias obtenidas de lesiones cervicales de distinto grado de los genes L1 y L2 de VPH 51.
- Agrupar las secuencias obtenidas mediante cladograma y comparar la distancia génica de los clusters calculados.
- Analizar el posible efecto sobre la estructura de las proteínas L1 y L2 de los polimorfismos más prevalentes sobre la secuencia consenso de los principales clusters calculados mediante análisis filogénico de las secuencias analizadas.
- Comparar los clusters de secuencias calculados con la variabilidad genética reportada para VPH 51 en otros estudios.

HIPÓTESIS

- La variabilidad genética de los genes L1 y L2 de VPH 51 en la región centro de México es diferente respecto a la reportada en otras partes del mundo.
- El grado de variabilidad genética de los genes L1 y L2 de VPH 51 en la región centro de México disminuye conforme la gravedad de la lesión neoplásica cervical aumenta.



METODOLOGÍA

Diseño del estudio

Tipo de estudio.

El estudio fue de tipo cuantitativo, observacional, descriptivo y retrospectivo. Se dividirá a los sujetos de estudio de acuerdo con su diagnóstico histopatológico confirmado, correspondiente a uno de los cuatro principales grados de lesión intraepitelial cervical del sistema Bethesda: negativo para lesión intraepitelial o malignidad (NILM), lesión intraepitelial escamosa de bajo grado (LSIL), lesión intraepitelial escamosa de alto grado (HSIL) y carcinoma invasor (IC).

Características de la muestra.

El universo de estudio comprendió pacientes del sexo femenino que hubiesen sido atendidas en servicios de Medicina Preventiva en el Hospital General de Zona (HGZ) 1 y el Hospital de la Mujer de Aguascalientes. Debido a que la disponibilidad de muestras para cada tipo de lesión a considerar en el estudio es variable, se tomó un muestreo a conveniencia (no aleatorio) de acuerdo con la disponibilidad de pacientes, sin estratificación por edad o estatus socioeconómico. Tomando en consideración la experiencia previa en la recolección y resultados reportados de genotipificación para VPH 51 en las mujeres del universo de estudio, se estimó inicialmente un tamaño muestral de 20 pacientes por cada grado de lesión (NILM, LSIL, HSIL y IC), para un N = 80.

Criterios de selección.

Criterios de inclusión	Criterios de exclusión
<ul style="list-style-type: none">• Pacientes del sexo femenino.• Pacientes en edad reproductiva.• Paciente con evaluación histopatológica y diagnóstico de lesión cervical intraepitelial de	<ul style="list-style-type: none">• Paciente cuya información sociodemográfica, clínica e historial gineco-obstétrico no haya sido recabado al momento del enrolamiento.

<p>acuerdo con el sistema de clasificación Bethesda.</p> <ul style="list-style-type: none"> • Paciente con diagnóstico confirmado de infección por VPH. • Paciente que haya consentido su participación en el estudio y haya firmado el consentimiento informado (en el caso de menores de edad, consentimiento adicional del padre o tutor legal). 	<ul style="list-style-type: none"> • Paciente sin evaluación y diagnóstico, o con diagnóstico no definitivo de lesión cervical intraepitelial de acuerdo al sistema de clasificación Bethesda. • Paciente sin infección por VPH 51, aun presentando infección por otros genotipos de VPH. • Paciente que no haya consentido su participación en el estudio, no haya firmado el consentimiento informado o éste no se encuentre disponible al momento de la recepción de la muestra (en el caso de menores de edad, consentimiento adicional del padre o tutor legal).
<p>Criterios de eliminación</p>	
<ul style="list-style-type: none"> • Muestra insuficiente, de mala calidad o indebidamente manejada al momento de su recepción para el análisis molecular. • Paciente que, aun firmado el consentimiento informado, decida voluntariamente no seguir participando en cualquier etapa del estudio. 	

Variables por analizar.

Variable	Tipo de variable	Operacionalización	Dimensiones	Indicadores	Unidad de medida	Índice	Valor
Grado de lesión	Categórica Politómica Ordinal	Diagnóstico histopatológico de lesión cervical intraepitelial	<ul style="list-style-type: none"> Sin lesión Lesión escamosa intraepitelial Carcinoma 	Sistema Bethesda	N/A	<ul style="list-style-type: none"> Negativo para lesión cervical o malignidad Lesión escamosa intraepitelial de bajo grado Lesión escamosa intraepitelial de alto grado Carcinoma invasor 	<ul style="list-style-type: none"> NILM LSIL HSIL IC
Edad	Continua De intervalo	Tiempo de vida extrauterina cumplido de la paciente, expresado como número entero	N/A	Años cumplidos	Años	N/A	<ul style="list-style-type: none"> < 18 años. 18-27 años. 28-37 años 38-47 años. 48-57 años > 58 años.
Nivel socioeconómico	Categórica Politómica Nominal	Categoría asignada de acuerdo con la capacidad económica y social percibida de la paciente	<ul style="list-style-type: none"> Clase alta Clase media Clase baja 	N/A	N/A	N/A	<ul style="list-style-type: none"> Alta Media Baja
Escolaridad	Categórica Politómica Nominal	Último grado escolar completado	<ul style="list-style-type: none"> Nivel básico 	Grado escolar completado	Grado escolar	N/A	<ul style="list-style-type: none"> Primaria Secundaria

			<ul style="list-style-type: none"> Nivel medio superior Nivel superior 				<ul style="list-style-type: none"> Bachillerato Licenciatura Posgrado
Número de hijos	Continua De intervalo	Número de nacimientos vivos que ha tenido la paciente a lo largo de su vida reproductiva	<ul style="list-style-type: none"> Nulípara Unípara Múltipara 	Número de hijos nacidos vivos	Número de hijos	N/A	<ul style="list-style-type: none"> 0 hijos 1 hijo > 1 hijo
Edad del primer parto	Continua De intervalo	Edad en la que la paciente tuvo su primer nacimiento vivo	N/A	Años cumplidos al momento del primer nacimiento vivo	Años	N/A	<ul style="list-style-type: none"> 12-17 años 18-23 años 24-29 años 30-35 años > 35 años
Edad de inicio de la vida sexual	Continua De intervalo	Edad a la que la paciente tuvo su primera relación sexual consensuada	N/A	Años cumplidos al momento de la primera relación sexual consensuada	Años	N/A	<ul style="list-style-type: none"> 11-16 años 17-22 años 23-28 años >29 años
Uso de condón durante las relaciones sexuales	Categoría Dicotómica	La paciente utiliza condón como método de protección al sostener relaciones sexuales	N/A	Uso afirmativo o negativo	N/A	<ul style="list-style-type: none"> Usa condón No usa condón 	<ul style="list-style-type: none"> SÍ NO
Infecciones vaginales frecuentes	Categoría Dicotómica	La paciente sufre constantemente de infecciones vaginales, diagnosticadas como tales por un médico	N/A	Presenta o no presenta infecciones con frecuencia	N/A	<ul style="list-style-type: none"> Presenta infecciones con frecuencia No presenta infecciones con frecuencia 	<ul style="list-style-type: none"> SÍ NO
Uso de anticonceptivos	Categoría Dicotómica	La paciente admite utilizar algún método	N/A	Admite o no el uso de anticonceptivos	N/A	<ul style="list-style-type: none"> Usa anticonceptivos 	<ul style="list-style-type: none"> SÍ

		anticonceptivo actualmente				<ul style="list-style-type: none"> No usa anticonceptivos 	<ul style="list-style-type: none"> NO 	
Uso de tabaco	Catagórica Dicotómica	La paciente consume tabaco	N/A	Admite o no el uso de tabaco	N/A	<ul style="list-style-type: none"> Usa tabaco No usa tabaco 	<ul style="list-style-type: none"> SÍ NO 	
Hábitos de consumo de tabaco	Continua De intervalo	Frecuencia y cantidad de tabaco utilizado por la paciente que lo consume	<ul style="list-style-type: none"> Frecuencia de consumo de tabaco Cantidad de consumo de tabaco 	<ul style="list-style-type: none"> Número de veces al día que se fuma Número de cigarrillos que se consumen por vez 	<ul style="list-style-type: none"> Número de veces por día Número de cigarrillos 	N/A	<ul style="list-style-type: none"> 1 vez/día 2 veces/día > 2 veces/día 1-3 cigarrillos 4-6 cigarrillos > 7 cigarrillos 	
AHF de cáncer cervicouterino	Catagórica Dicotómica	La paciente tiene antecedentes familiares cercanos de cáncer cervicouterino	N/A	Admite o no antecedentes	N/A	<ul style="list-style-type: none"> Sin antecedentes Al menos un antecedente 	<ul style="list-style-type: none"> NO SÍ 	
AHF de cáncer de otro tipo	Catagórica Dicotómica	La paciente tiene antecedentes familiares cercanos de cáncer distinto al cervicouterino	N/A	Admite o no antecedentes	N/A	<ul style="list-style-type: none"> Sin antecedentes Al menos un antecedente 	<ul style="list-style-type: none"> NO SÍ 	
Número de polimorfismos en la secuencia	Continua Absoluta De Proporción	<p>Conteo absoluto de polimorfismos en una secuencia respecto al consenso</p> <p>Conteo relativo de polimorfismos en una secuencia respecto al total de sitios probables</p>	<p>N/A</p> <p>N/A</p>	<p>Número de polimorfismos en una secuencia</p> <ul style="list-style-type: none"> Sublinaje Variante Genotipo 	<p>Número de polimorfismos</p> <p>% de sitios polimórficos</p>	N/A	<ul style="list-style-type: none"> Sublinaje Variante Genotipo 	<p>Número de polimorfismos</p> <ul style="list-style-type: none"> < 1% 1 – 10% > 10%

Prevalencia de polimorfismos	Continua De Proporción	Proporción de polimorfismos presentes en una secuencia o cluster y que persisten en dos o más grados de lesión contiguos	N/A	<ul style="list-style-type: none"> • Mutación • Polimorfismo raro • Polimorfismo común 	% de sitios polimórficos respecto al consenso	<ul style="list-style-type: none"> • Mutación • Polimorfismo raro • Polimorfismo común 	<ul style="list-style-type: none"> • < 1% • 1 – 10 % • > 10 %
Posición de polimorfismos prevalentes	Catórgica Politómica Nominal	Localización puntual de los polimorfismos persistentes cuya prevalencia ha sido determinada	N/A	Posición después del sitio de inicio del consenso	Posición (base)	N/A	Número de posición
Tipo de polimorfismos prevalentes	Catórgica Politómica Nominal	Efecto del polimorfismo sobre la expresión génica y la estructura primaria de la proteína codificada de acuerdo a la traducción <i>in silico</i>	<p>Sin efecto</p> <p>Reemplazo de un aa</p> <p>Proteína trunca</p> <p>Cambio de marco de lectura</p>	Efecto producido en la proteína resultante en la traducción	N/A	<p>Sin cambio de aa</p> <p>Cambio de aa</p> <p>Interrupción de la traducción</p> <p>Desplazamiento</p>	<p>Silenciosa</p> <p>Cambio de sentido</p> <p>Sin sentido</p> <p>Cambio de ORF</p>
Neutralidad y selección génica	Continua De Proporción	Medida de divergencia que analiza la selección o neutralidad (no selección) de un polimorfismo en un conjunto de secuencias alineadas	N/A	D de Tajima	Adimensional	<ul style="list-style-type: none"> • Selección negativa • Selección positiva • No selección 	<ul style="list-style-type: none"> • $D < 0$ • $D > 0$ • 0

Consideraciones éticas.

Acorde con la normatividad vigente y los convenios internacionales aplicables a la investigación con sujetos humanos, y atendiendo a los requerimientos comprendidos en la Norma Oficial Mexicana NOM-012-SSA3-2012, Que establece los criterios para la ejecución de proyectos de investigación para la salud en seres humanos, se sometió este protocolo de investigación a consideración de los Comités de Ética de la Universidad Autónoma de Aguascalientes, HGZ 1 Aguascalientes y Hospital de la Mujer de Aguascalientes para su revisión y deliberación. El protocolo debió ser aprobado por todos los comités pertinentes antes de comenzar el reclutamiento de las pacientes. Para solicitar la participación a las pacientes de estudio, en todos los casos se les informó debidamente del contenido, protocolos y alcances de la investigación, y se solicitó su consentimiento a través de la Carta de Consentimiento Informado (Anexo 1). Ninguna muestra enviada al laboratorio fue procesada si el Consentimiento Informado no hubo sido enviado y debidamente firmado junto con el espécimen.

Atendiendo además a las recomendaciones oportunas de los Comités de Ética, se hicieron las modificaciones adecuadas al protocolo antes, durante y después de su ejecución para garantizar en todo momento la integridad de la información almacenada de las pacientes, resultados de las pruebas de laboratorio y el análisis de datos ejecutados. Para ello, se guardó de manera segura y confidencial toda evidencia de dichos datos como bitácoras, bases electrónicas y fotografías, y sólo fueron divulgadas atendiendo los lineamientos éticos ya mencionados.

Técnicas de recolección de datos.

Las pacientes fueron reclutadas al momento de su atención en los Módulos de Medicina Preventiva de las unidades receptoras, para su revisión de rutina, o de aquéllas que fueron atendidas como parte de seguimiento como parte de un diagnóstico previo de lesiones cervicales. Se informó a las pacientes del contenido, procedimiento y alcances del estudio, y aquéllas que desearon participar voluntariamente firmaron el consentimiento informado (Anexo 1). El médico o pasante llenó el cuestionario adjunto (Anexo 2) con información sociodemográfica y

clínica de la paciente de forma anónima, y enviado debidamente relleno junto con la muestra destinada al análisis molecular.

Se tomaron dos tipos de muestras: un raspado cervical estándar para la prueba de citología cervical de acuerdo al Manual de Procedimientos para la Toma de la Muestra de Citología Cervical, y un raspado empleando citobrush almacenado en un tubo eppendorf con 1 mL de medio de transporte (de 10 mM de Tris pH 8.0, 0.4 M de NaCl, 2 mM de EDTA disódico pH 8.0, 50% de etanol gbm) a temperatura ambiente para su análisis posterior. De acuerdo con los resultados de la citología, en caso de encontrarse un resultado anormal, se tomó a la paciente una biopsia cervical que fue analizada en la Clínica de Displasias para determinar el diagnóstico histopatológico preciso. El diagnóstico definitivo de la lesión fue vaciado en el formato del cuestionario, y dicho cuestionario fue enviado junto con la muestra almacenada en medio de transporte, debidamente identificada con el número de folio asignado a la paciente, al Laboratorio de Virología e Ingeniería Genética de la Universidad Autónoma de Aguascalientes para llevar a cabo el análisis molecular.

Análisis de laboratorio.

Las muestras de citobrush debidamente recibidas fueron registradas en una bitácora de acuerdo con su número de folio asignado, donde además fue vaciada la información recabada en el cuestionario anexo para su análisis posterior. Las muestras fueron preservadas selladas a temperatura ambiente alejadas de la luz hasta su procesamiento.

El medio de transporte fue recuperado en un tubo limpio, y la secreción mucosa contenida en el citobrush fue recuperada con 1 mL de medio de colecta (misma composición del medio de transporte, pero sustituyendo el etanol por H₂O gbm) en un segundo tubo. Las muestras se sometieron a centrifugación a 13 000 rpm durante 7 min para recuperar la fracción celular, se eliminó el sobrenadante, se evaporó el exceso de etanol por incubación a 60 °C durante 5-10 min, y el ADN se extrajo del paquete celular utilizando el Blood-Animal-Plant DNA Preparation Kit Kit (Jena Bioscience, Jena, Alemania) siguiendo las instrucciones del fabricante para

el procesamiento de muestras de células animales (Anexo 3) y eluyendo las dos partes de la muestra en un mismo vial. La integridad del ADN fue verificada mediante electroforesis en gel de agarosa al 1% en buffer SB, y revelando mediante tinción con bromuro de etidio al 0.005% y observación utilizando un transiluminador de radiación UV.

La presencia de la infección por VPH fue determinada mediante genotipificación empleando PCR anidada multiplex de acuerdo al protocolo de Sotlar et al. (46). Los oligos utilizados comprendieron una primera reacción (Anexo 4A) y una reacción multiplex secundaria (Anexo 4B). La primera reacción se llevó a cabo utilizando 1X del buffer de reacción (Invitrogen), 1.5 mM de MgCl₂, 0.2 mM de dNTPs, 0.8 μM de cada oligonucleótido, 2 U de Taq polimerasa y 200-400 ng de ADN extraído, a un volumen final de 50 μL. La primera amplificación consistió en una desnaturalización inicial a 94 °C por 5 minutos, seguido de 40 ciclos de desnaturalización a 94 °C por 1 min, alineamiento a 55 °C por 1 min y extensión a 72 °C por 1 min; terminando con una extensión final a 72 °C durante 10 min. La segunda reacción utilizó la misma mezcla que la primera excepto dNTPs (0.4 mM) y Taq polimerasa (2.5 U), se utilizarán 5-7 μL del producto de PCR obtenido en la primera reacción como ADN molde, y cada coctel se llevó a cabo en un tubo por separado. La segunda amplificación consistió en una desnaturalización inicial a 94 °C durante 5 min, seguido de 30 ciclos de desnaturalización a 94 °C por 30 s, alineamiento a 53 °C por 45 s y extensión a 72 °C durante 45 s; terminando con una extensión final a 72 °C durante 5 min. Los productos de amplificación se observaron en un gel de agarosa al 1.5 % revelado con bromuro de etidio y radiación UV.

Una vez confirmada la presencia de VPH 51, se procedió a la amplificación específica de los genes L1 y L2 de este genotipo mediante las reacciones de PCR punto final reportadas por Xu et al. (11) sujetas a estandarización respecto a la temperatura de alineamiento y la concentración de MgCl₂. Debido al tamaño de los genes (alrededor de 2 Kpb cada uno), a fin de cubrir toda su extensión se llevaron a cabo dos reacciones por separado para cada gen, utilizando los oligonucleótidos del Anexo 4C y las siguientes condiciones de reacción: 1X de buffer de reacción (Invitrogen), 0.4 mM de dNTPs, 0.8 μM de cada oligonucleótido según la reacción

correspondiente, 2.5 U de Taq polimerasa (Invitrogen) y 200-400 ng de ADN total extraído. La concentración de MgCl₂ fue estandarizada utilizando un gradiente de concentración desde 1.5 mM a 2.5 mM, a intervalos de 0.25 mM. El programa de reacción utilizado consistió en una desnaturalización inicial a 94 °C por 5 min, seguido de 35 ciclos de desnaturalización a 94 °C por 45 s, alineamiento a una temperatura por estandarizar durante 60 s, y extensión a 72 °C durante 90 s; se realizará una extensión final a 72 °C durante 7 min, y los productos amplificados fueron conservados a -20 °C hasta su uso. Para estandarizar la temperatura de alineamiento, se hizo uso de la función VeriFlex del termociclador Verity (Applied Biosciences), que permite establecer un gradiente de temperaturas para un paso determinado de la reacción – en este caso, se utilizó un gradiente de 52 a 56 °C a incrementos de 1 °C. El resultado fue determinado mediante electroforesis en gel de agarosa al 1.25% en buffer TBE y revelando con bromuro de etidio y radiación UV. La mejor concentración de MgCl₂ y temperatura de alineamiento óptima fue determinada en conjunto, de acuerdo con los siguientes criterios:

- 1) Amplificación exitosa de los fragmentos esperados.
- 2) Ausencia de otras bandas, indicadoras de una amplificación inespecífica.
- 3) Buena eficiencia de amplificación.
- 4) Producto de PCR no degradado.

El programa final seleccionado fue utilizado para amplificar los fragmentos de los genes L1 y L2 de todas las muestras seleccionadas y confirmadas como VPH 51 positivas. Se determinó además la concentración y pureza de los productos de PCR mediante cuantificación espectrofotométrica a 260, 280 y 230 nm utilizando el equipo NanoDrop 2000 (ThermoFisher Scientific, Reino Unido), contando como aceptable si cumplía con todos los requisitos siguientes (de lo contrario, se repite la amplificación):

- Presenta una concentración mínima de 200 ng/μL (necesaria para la secuenciación).
- Presenta un cociente A260/A280 entre 1.8 y 2.0 (sin contaminación por ARN).

- Presenta un cociente A260/A230 mayor a 2.0 (sin contaminación por proteínas o compuestos fenólicos).
- Presenta un espectro UV/visible no alterado significativamente, consistente con un máximo de absorción a 260 nm, sin otros picos prominentes (en particular a 230 nm o inferior).

La secuenciación fue llevada a cabo utilizando los servicios del Laboratorio Nacional de Biotecnología Agrícola, Médica y Ambiental (LANBAMA) (IPICYT, San Luis Potosí, MX), el cual cuenta con servicio de secuenciación Sanger automatizada. Para cumplir con los requerimientos establecidos para un correcto procesamiento de las muestras enviadas, los productos de PCR cuantificados fueron diluidos con H₂O gbm estéril a una concentración no menor a 50 ng/μL en un volumen mínimo de 50 μL, de los cuáles se tomaron 9 μL para realizar una electroforesis de corroboración y 1 μL para verificar la concentración. Las muestras fueron enviadas en refrigeración a LANBAMA junto con la hoja de información anexa solicitada, que requiere que cada muestra sea debidamente identificada junto con su concentración, método de purificación y evidencia fotográfica de la electroforesis de corroboración. Se enviaron además viales de oligonucleótidos propios utilizados para la secuenciación, que fueron los mismos utilizados para las reacciones de PCR, a una concentración de 1.5 μM y a un volumen mínimo de 10 μL/muestra.

Los resultados de la secuenciación fueron recibidos por medio electrónico en formato .txt, .doc y .ab1, y guardados para su análisis posterior.

Análisis estadístico y bioinformático.

Con la información proporcionada por el cuestionario adjunto, se elaborará una base de datos en Microsoft Excel donde se vaciaron las respuestas que corresponden a las variables sociodemográficas y clínicas. La base además contiene celdas para vaciar los resultados de la genotipificación de VPH (incluyendo la presencia de otros genotipos aparte de 51). Dicha información fue cotejada con los registros de la bitácora para la concordancia de resultados en caso necesario.

Las variables sociodemográficas y clínicas fueron analizadas mediante estadística descriptiva, de manera global y grupal, reportándose en forma de media y desviación estándar (variables continuas) y porcentajes (variables categóricas). Para la comparación intergrupala se utilizaron pruebas de contraste como ANOVA o Kruskal-Wallis (continuas) y chi cuadrada (categóricas), manejando como hipótesis nula la no diferencia intergrupala y verificando previamente normalidad y homocedasticidad (prueba de Kolmogorov-Smirnov y Levene respectivamente), interpretando un valor $p < 0.05$ como estadísticamente significativo. Estos análisis se llevaron a cabo utilizando los softwares SPSS 24.0 (IBM) y GraphPad Prism 8.0.

Las secuencias recibidas fueron convertidas a formato FASTA para llevar a cabo alineamientos simples y múltiples:

- Alineamiento simple: cada secuencia fue comparada con otras ya existentes en bases de datos utilizando la herramienta BLASTn (NBCI) (47), para verificar que se tratase de una correspondiente a VPH 51 y no a otro genotipo, gen u organismo.
- Alineamiento múltiple: aquellas secuencias confirmadas como pertenecientes a VPH 51 fueron sometidas a un alineamiento múltiple utilizando la herramienta MUSCLE (48).

Como resultado del alineamiento múltiple se realizó el conteo absoluto y relativo de polimorfismos por secuencia, obteniendo un consenso preliminar por grupos y de forma global. Las secuencias y los consensos fueron traducidos *in silico* empleando la herramienta Translate ExPASy) (49) para identificar los efectos de los polimorfismos sobre la estructura primaria de la proteína más probable de acuerdo con los ORF hipotéticos, y un modelaje tridimensional utilizando el software UCSF Chimera (50).

El análisis filogenético se llevó a cabo utilizando el programa MEGA X (51), para calcular clusters en un cladograma mediante el método de máxima verosimilitud y, de ser posible, usando matrices Bayesianas para optimizar el método. Una vez calculadas las distancias entre cada cluster, se procedió a realizar la prueba de

neutralidad de los polimorfismos más prevalentes y persistentes según la prueba de D de Tajima (52).

Recursos disponibles para el estudio

Recursos humanos y materiales.

La parte experimental del estudio se llevó a cabo en el Laboratorio de Virología e Ingeniería Genética (edificio 23) de la Universidad Autónoma de Aguascalientes. En dichas instalaciones se dispone del material y reactivos necesarios para procesar y analizar las muestras: cabinas de bioseguridad, termociclador Verity (Applied Biosystems), refrigerador de ultracongelación (-20 °C), cámaras de electroforesis y transiluminador. Los insumos consumibles y reactivos también pueden almacenarse en el laboratorio.

Respecto al análisis de datos, la Universidad Autónoma de Aguascalientes posee acceso a los softwares estadísticos MiniTab 18.0, SPSS 24.0 y GraphPad 8.0 – este último será el prioritario para el procesamiento y presentación de datos por su conveniencia. No se dispone actualmente en la Universidad de acceso a software bioinformático especializado como MEGA X, el cual fue descargado aparte para su uso.

Para la recolección de las muestras, fue fundamental la participación de los Módulos de Medicina Preventiva y Salud Reproductiva de las unidades receptoras de pacientes, por lo cual este proyecto fue llevado a cabo en colaboración con el HGZ 1 y el Hospital de la Mujer de Aguascalientes, atendiendo la necesidad de la práctica privada en caso necesario también como fuente de muestras. La toma de muestras fue llevada a cabo por el personal adscrito o pasante del servicio que deseara participar en el proyecto y que estuviese capacitado para la correcta toma de muestras. El medio de transporte fue proporcionado listo para su uso por el Laboratorio de Virología e Ingeniería Genética. El diagnóstico histopatológico de las lesiones cervicales fue llevado a cabo en la Clínica de Displasias respectiva por personal capacitado.

Recursos financieros.

La mayor parte del recurso económico fue obtenido sometiendo este proyecto a consideración en la Universidad Autónoma de Aguascalientes como un proyecto de investigación por parte del Laboratorio de Virología e Ingeniería Genética – ya que se cuenta con el antecedente del proyecto PIBB18-4, un estudio similar al propuesto (pero sobre el gen E6 y un grupo limitado de lesiones) que ya recibió financiamiento suficiente por parte de la Universidad.

- La mayor parte del recurso financiero se destinó al servicio externo de secuenciación, el cual tiene un costo (a la fecha de mayo 2020) de 120 pesos/secuencia sin importar su tamaño, costo que pudo disminuir si se envía un número grande de secuencias a la vez.
- Una segunda parte del presupuesto la abarcó la genotipificación de las muestras obtenidas. Como sólo las muestras positivas para VPH 51 serán consideradas posteriormente, se calculó que se requerirán al menos 300 muestras (entre todos los grupos de lesión) para genotipificación, a un costo estimado sólo en reactivos de 50 pesos/genotipificación como mínimo.
- Una parte menor se destinó a la amplificación de los fragmentos a secuenciar. Esta reacción se llevó a cabo sólo en las muestras positivas a VPH 51, con lo que se debieron cubrir 320 reacciones (sin contar las necesarias para estandarización) al mismo costo estimado de 50 pesos/cuatro reacciones (dos para L1 y dos para L2).

La tabla de balance entre el presupuesto probable y los costos de operación se muestra a continuación:

CONCEPTO	Presupuesto	Costos
INGRESOS		
Presupuesto ordinario (2 periodos anuales)	\$35,000.00	-
Presupuesto extraordinario (1 periodo)	\$40,000.00	-

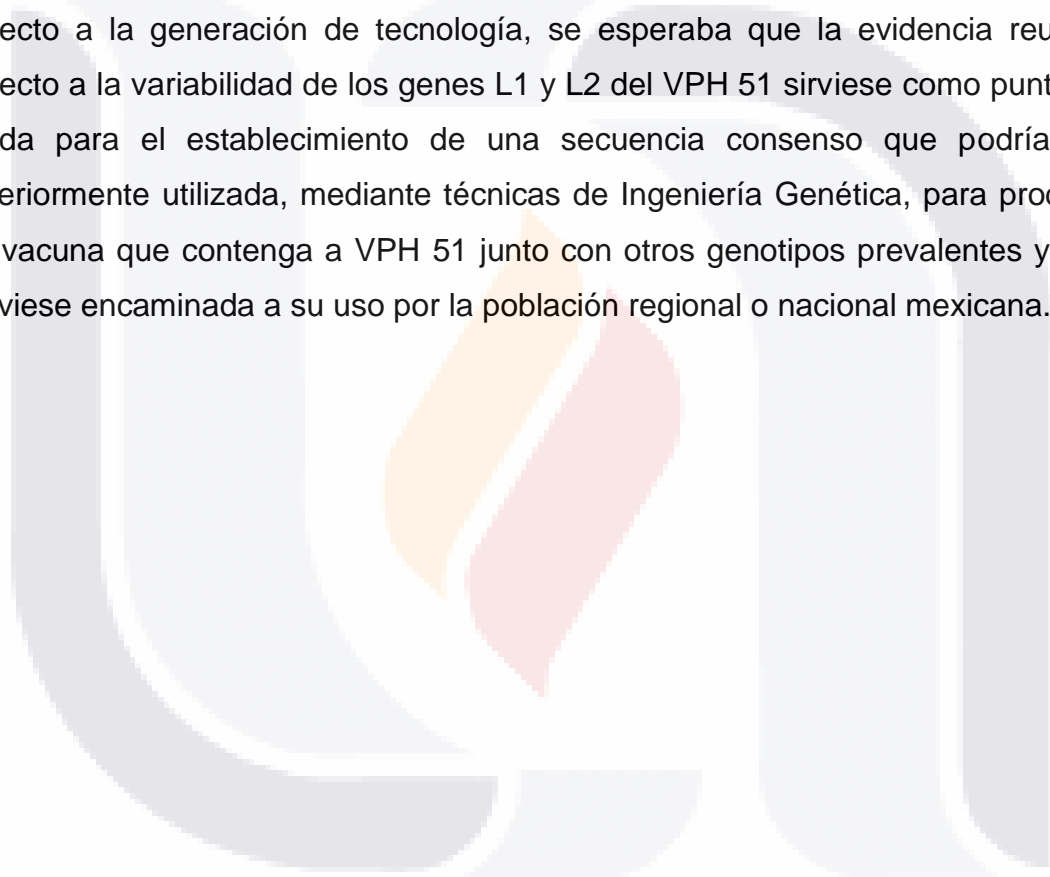
EGRESOS		
Servicios de secuenciación (320 secuencias, máximo \$120.00/secuencia)	-	\$38,400.00
Genotipificación de VPH (300 muestras)	-	\$15,000.00
Amplificación específica de L1 y L2 de VPH 51 (100 muestras, ya contando estandarización)	-	\$10,000.00
Otros consumibles (incl. gastos de envío)	-	\$10,000.00
TOTALES		
Balances estimados	\$75,000.00	\$73,400.00
Balance final estimado		\$1,600.00

El balance positivo era candidato de ser incrementado aún más si se reduce el costo por secuencia obtenida (\$100/secuencia a partir de 50 secuencias enviadas simultáneamente, además de ahorrarse en envíos) y haciendo uso de insumos y reactivo ya disponible en el laboratorio; el ahorro no se estimó al momento, pero podría haberse elevado hasta \$10,000.00 de acuerdo con la disponibilidad de recursos.

Difusión e impacto de la investigación

Esta investigación forma parte de una línea de reciente creación, encaminada al análisis genómico y aplicaciones clínicas y tecnológicas del conocimiento generado en torno al VPH genotipo 51. Partiendo de la evidencia que sugiere una inusual prevalencia de este genotipo en la región, los resultados de esta investigación y otras serían presentados, primero por separado y después en conjunto, aprovechando los espacios de difusión científica disponibles como Seminarios de Investigación Institucionales, Congresos Nacionales en las áreas de Genética Humana, Virología y/o Ginecología y otros que se consideren necesarios, ya sea en forma de ponencias o carteles.

Para la publicación de los resultados, al final de esta investigación se elaboró un artículo en forma para su envío a una revista indexada especializada en el ramo, a fin de dar a conocer los hallazgos a la comunidad científica. Se esperaba que, al agrupar la evidencia resultante del análisis de diversos genes del VPH 51, los datos generados sirviesen para establecer una línea formal de análisis a nivel genómico, mucho más especializado, que contribuyesen a incrementar el acervo genético existente del VPH 51 en bases de datos bioinformáticas, y cuyas secuencias serían de libre acceso para otros grupos de investigación con líneas similares. Finalmente, respecto a la generación de tecnología, se esperaba que la evidencia reunida respecto a la variabilidad de los genes L1 y L2 del VPH 51 sirviese como punto de partida para el establecimiento de una secuencia consenso que podría ser posteriormente utilizada, mediante técnicas de Ingeniería Genética, para producir una vacuna que contenga a VPH 51 junto con otros genotipos prevalentes y que estuviese encaminada a su uso por la población regional o nacional mexicana.



RESULTADOS

Verificación de la amplificación de los genes L1 y L2 de VPH 51 mediante alineamiento y PCR *in silico*

La primera verificación por realizar fue la concordancia de las secuencias reportadas por Xu et al. (11), que de acuerdo con la secuencia de referencia M62877.1 (GenBank) permitirían la amplificación de las regiones genómicas ilustradas en las figs. 9 y 10. La amplificación *in silico* fue verificada empleando Primer-BLAST (53), y un resumen de los resultados se muestra en la fig. 11. Para ninguno de los casos se mostraron amplificaciones inespecíficas potenciales, excepto para otros genomas aislados correspondientes a VPH 51.

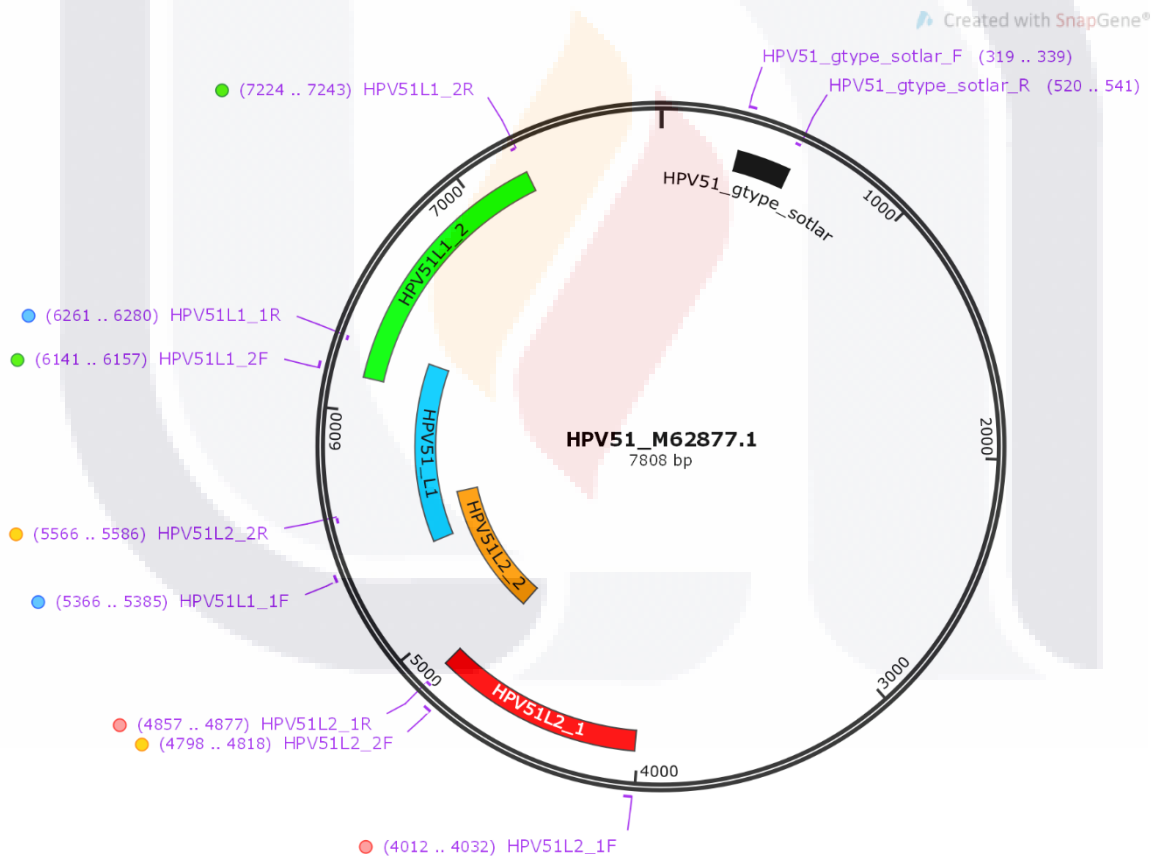


Fig. 9: Mapa de amplificación de los oligonucleótidos dirigidos contra los genes L1 y L2 del VPH 51. El genoma de referencia (GenBank: M62877.1) es circular debido a la forma episomal previa a la ruptura e integración al genoma hospedero. Cada fragmento amplificado incluye sus oligonucleótidos flanqueantes del mismo código de color. Se incluye además el fragmento de 223 pb amplificado en la genotipificación mediante PCR multiplex propuesta por Sotlar et al. (46) que permite la detección de VPH 51 (HPV51_gtype_sotlar). Elaborado con ayuda del software SnapGene (Insightful Science, disponible en www.snapgene.com).

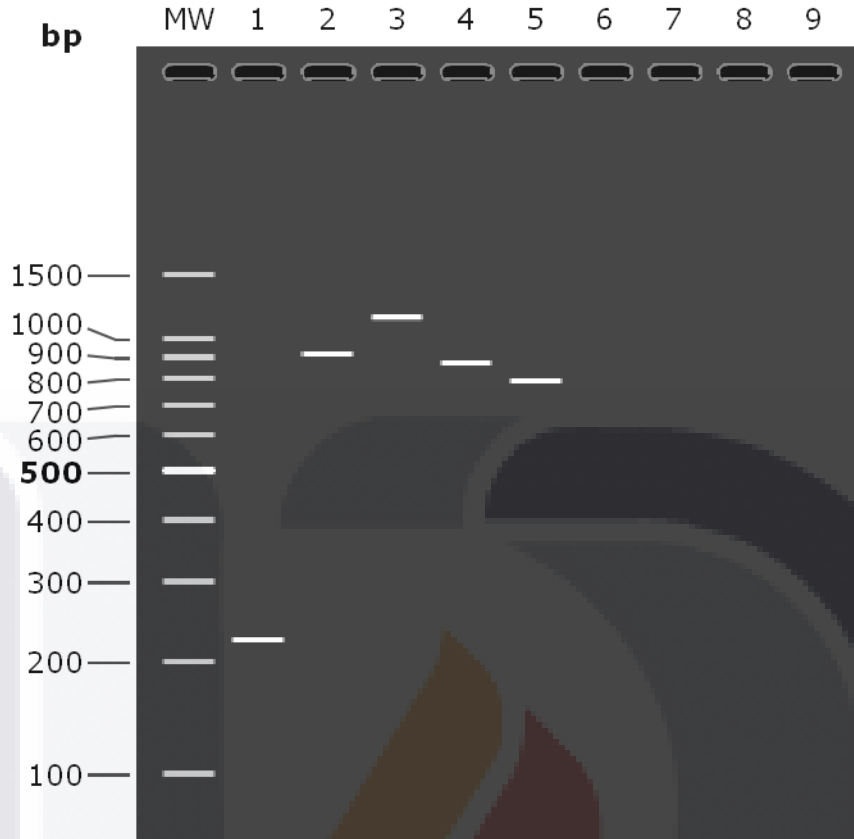


Fig. 10: Simulación de un gel de agarosa al 2.0% de las secuencias a amplificar de la fig. 9. MW; marcador de peso molecular 100 bp (Promega). Carril 1: HPV_gtype_sotlar. Carril 2: HPV51L1_1. Carril 3: HPV51L1_2. Carril 4: HPV51L2_1. Carril 5: HPV51L2_2. Elaborado con ayuda del software SnapGene (Insignifal Science, disponible en www.snapgene.com).

Sin embargo, al observar los parámetros determinados por la PCR *in silico*, se observaron dos situaciones diferentes. Mientras que los oligonucleótidos destinados a la amplificación del gen L1 (VPH51L1_1 y VPH51L1_2) mostraron relativamente poca auto complementariedad 3' (> 3.00), los oligonucleótidos para L2 sí mostraron valores más elevados de auto complementariedad; más aún, los oligonucleótidos del primer fragmento de L2 (HPV51L2_1) poseen un %GC más bajo del recomendado, lo que podría ocasionar problemas para lograr una amplificación exitosa. Sin embargo, la alta especificidad de los oligonucleótidos para su amplificación, en especial debido a la ausencia de amplificación en otros genotipos relacionados – en especial VPH 82 – o secuencias endógenas de *Homo sapiens*, es suficiente para probar la estandarización, aunque con posibles modificaciones respecto al protocolo original.

Primer pair 1										
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity	
HPV51L1_1	Forward primer	GGTATTGCCACATCTCCTA	Plus	20	5366	5385	56.38	50.00	3.00	2.00
	Reverse primer	TTTGCTCCCTGCGTAAATGA	Minus	20	6280	6261	57.81	45.00	3.00	1.00
	Product length	915								
Primer pair 1										
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity	
HPV51L1_2	Forward primer	GGATTTCGCTGCCCTAC	Plus	17	6141	6157	54.92	58.82	3.00	0.00
	Reverse primer	TACCCACATACACCCGTACA	Minus	20	7243	7224	57.77	50.00	4.00	2.00
	Product length	1103								
Primer pair 1										
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity	
HPV51L2_1	Forward primer	TGCAACATCCCAATTAACACTAC	Plus	21	4012	4032	54.16	38.10	4.00	0.00
	Reverse primer	TGTGGATGGCTTACTAATAAAA	Minus	21	4877	4857	52.22	33.33	6.00	4.00
	Product length	866								
Primer pair 1										
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity	
HPV51L2_2	Forward primer	TAGCTGCTCCCCGCTTGATA	Plus	21	4798	4818	60.76	52.38	5.00	4.00
	Reverse primer	CACAATTCGAGACACAGTGTC	Minus	21	5586	5566	59.80	52.38	4.00	2.00
	Product length	789								

Fig. 11: Resumen de PCR in silico de los fragmentos de L1 y L2 estudiados. La longitud de cada fragmento corresponde con los estimados teóricos reportados previamente. El valor aceptable tanto de auto complementariedad como complementariedad 3' es menos a 3.00.

Estandarización de la técnica de amplificación de los genes L1 y L2 de VPH 51 mediante PCR punto final

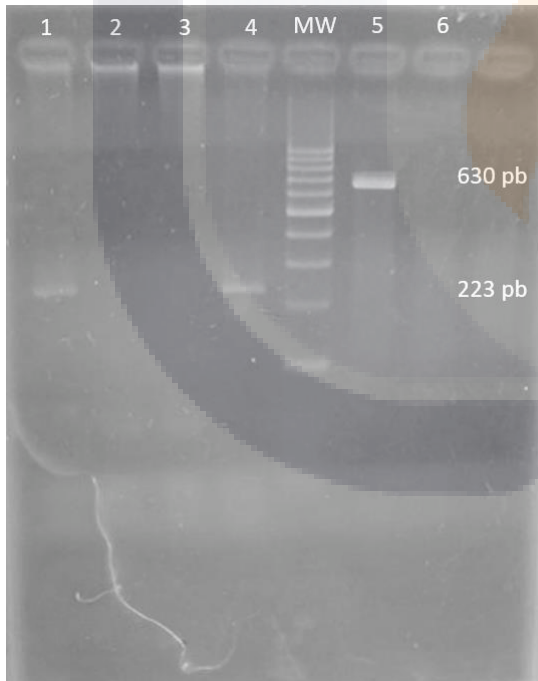


Fig. 12: Genotipificación de muestras modelo, observadas tras una electroforesis en gel de agarosa al 2.0%. Una banda de 223 pb en la segunda reacción indica la presencia de VPH 51. Carriles 1-4: muestras genotipificadas. MW: marcador de peso molecular 100 pb (Jena Bioscience). Carril 5: primera amplificación, E7AS/LCR. Carril 6: muestra negativa para VPH.

Con las muestras de ADN previamente destinadas para estandarización (no incluidas dentro del conjunto final de muestras de estudio), se realizó la verificación de integridad y genotipificación, con el fin de comprobar que las muestras siguieran siendo viables para su uso en la estandarización de la amplificación de L1 y L2. De las 10 muestras almacenadas, se verificó que 7 contenían ADN de buena calidad y con una amplificación exitosa para el fragmento de VPH 51 (fig. 12). Con un volumen aproximado de 30 µL por muestra, se procedió a la estandarización con estas muestras.

Con base en experiencia previa con la amplificación del gen E6 siguiendo el mismo protocolo (11), se optó por incrementar la concentración de dNTPs en la reacción de 0.2 a 0.3 mM para la amplificación de los fragmentos del gen L1 y a 0.4 mM para los del gen L1, con el fin de garantizar una proporción aproximada de 300 pb/0.1 mM de dNTPs en la mezcla y evitar un agotamiento prematuro que pudiera resultar en la síntesis de amplicones incompletos. La concentración de oligonucleótidos no mostró interferencia en la reacción exitosa, así que su concentración (0.8 μ M) no fue modificada inicialmente (ver más abajo). Las variables por estandarizar fueron, entonces:

- 1) La concentración de $MgCl_2$, que fue probada para cada reacción en concentraciones de 1.5 a 2.5 mM, a un gradiente de 0.5 mM.
- 2) La temperatura de alineamiento, que debido a la auto complementariedad de algunos oligonucleótidos fue elevada de un punto central de 54 a 56 $^{\circ}C$, en un gradiente de 2 $^{\circ}C$ desde 54 hasta 60 $^{\circ}C$.

Los resultados se muestran en la fig. 13. De acuerdo con las amplificaciones, se observó un requerimiento mínimo de 2.0 mM de $MgCl_2$ para lograr una amplificación mínima (a 1.5 mM no se observó amplificación en ningún caso), en conjunto con una temperatura de 56 y 58 $^{\circ}C$ – de ahí que 57 $^{\circ}C$ sea la temperatura candidata por ser el punto medio entre ambas. Sin embargo, el rendimiento no es igual para todas las reacciones: en particular, para el caso de la amplificación de los fragmentos del gen L2 el rendimiento es marcadamente inferior al encontrado para el gen L1. Se desea que la temperatura de alineamiento sea igual o similar para ambas reacciones, por lo que los resultados de estandarización obtenidos hasta ahora para L2 no fueron definitivos hasta esta etapa.

Con el fin de optimizar el rendimiento sin comprometer la especificidad de las reacciones, se llevó a cabo un ajuste de la concentración de $MgCl_2$ y oligonucleótidos de manera simultánea a diferentes temperaturas, partiendo del hallazgo central de 57 $^{\circ}C$, considerando las siguientes condiciones: 1.5 mM de $MgCl_2$, 0.6 μ M de cada oligonucleótido en cada reacción, y temperaturas de alineamiento de 57, 59 y 61 $^{\circ}C$; los resultados se muestran en la fig. 13. Se logró un

buen rendimiento sin presencia de amplificaciones inespecíficas amplificando los fragmentos L1_1 y L2_2 a 57 °C, y los fragmentos L1_2 y L2_1 a 59 °C.

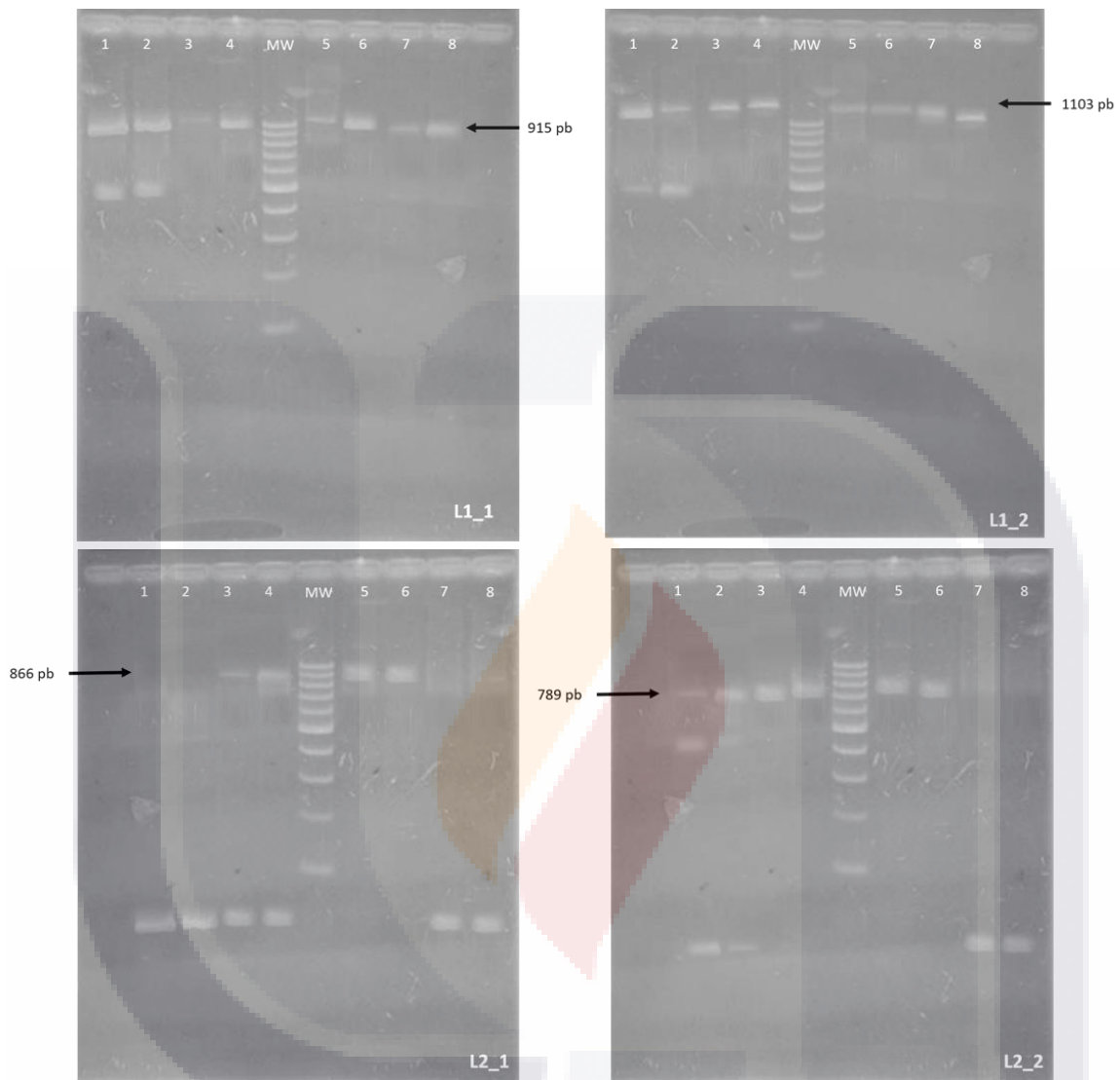


Fig. 13: Estandarización de técnicas de amplificación de L1 y L2 de VPH 51. Cada gel corresponde a un fragmento donde se probaron las variables de $[MgCl_2]$ y temperatura de alineamiento. La observación se realizó en un gel de agarosa al 1.5%. Los resultados positivos se muestran en el anexo 4. Para todos los gels, carril 1: 2.0 mM, 54 °C; carril 2: 2.5 mM, 54 °C; carril 3: 2.0 mM, 56 °C; carril 4: 2.5 mM, 56 °C; carril 5: 2.0 mM, 58 °C; carril 6: 2.5 mM, 58 °C; carril 7: 2.0 mM, 60 °C; carril 8: 2.5 mM, 60 °C. MW: marcador de peso molecular 100 pb (Jena Bioscience).

Características de las pacientes analizadas

Fueron secuenciadas con éxito 44 muestras de pacientes distribuidas en los cuatro grupos de lesión: NILM (n = 7), LSIL (n = 17), HSIL (n = 13) e IC (n = 7). Las características sociodemográficas y clínicas, recogidas mediante el cuestionario anexo, se resumen en la tabla 1.

Tabla 1: Características sociodemográficas y clínicas de las pacientes analizadas, distribuidas por grupo de lesión. Las variables continuas se muestran como media \pm SD, y las categóricas como n (%), separadas por categorías según corresponda a las politómicas, y a los resultados positivos en el caso de las variables dicotómicas.

Variable	NILM (n = 7)	LSIL (n = 17)	HSIL (n = 13)	IC (n = 7)	P
Edad (años)	26.85 \pm 4.91	36.12 \pm 13.20	34.38 \pm 9.02	39.14 \pm 5.43	0.134
Nivel socioeconómico	Bajo: 1 (14.3) Medio: 6 (85.7)	Bajo: 1 (5.9) Medio: 16 (94.6)	Bajo: 1 (7.7) Medio: 12 (92.3)	Medio: 7 (100.0)	0.768
Escolaridad	Primaria: 1 (14.3) Secundaria: 3 (42.9) Bachillerato: 2 (28.6) Superior: 1 (14.3)	Primaria: 2 (11.8) Secundaria: 11 (64.7) Bachillerato: 2 (11.8) Superior: 2 (11.8)	Primaria: 5 (38.5) Secundaria: 4 (30.8) Bachillerato: 4 (30.8)	Secundaria: 2 (28.6) Bachillerato: 3 (42.9) Superior: 2 (28.6)	0.094
Número de hijos	2.00 \pm 0.82	2.82 \pm 2.07	2.92 \pm 2.78	2.14 \pm 0.90	0.695
Edad primer parto (años)	21.00 \pm 3.11	19.50 \pm 3.11	19.64 \pm 2.84	23.14 \pm 3.58	0.076
Edad inicio vida sexual (años)	17.29 \pm 1.98	18.06 \pm 3.70	17.69 \pm 1.49	21.14 \pm 2.04	0.035
Uso de condón	2 (28.6)	4 (23.5)	4 (30.8)	1 (14.3)	0.871
Infecciones vaginales	5 (71.4)	6 (35.3)	9 (69.2)	4 (57.1)	0.222
Anticonceptivos hormonales	2 (28.6)	1 (5.7)	4 (30.8)	3 (42.9)	0.182
Uso de tabaco	4 (57.1)	11 (64.7)	7 (53.8)	4 (57.1)	0.944
No. De cigarrillos consumidos (día)	1.75 \pm 0.50	2.82 \pm 1.40	4.00 \pm 2.16	3.50 \pm 1.92	0.179

Tiempo de consumo de tabaco (años)	8.25 ± 5.38	4.36 ± 4.01	11.57 ± 9.43	12.50 ± 5.00	0.070
AHF cáncer cervicouterino	1 (14.3)	3 (17.6)	2 (15.4)	4 (42.9)	0.466
AHF otros tipos de cáncer	0 (0.0)	2 (11.8)	2 (15.4)	2 (28.6)	0.483

De acuerdo con los resultados obtenidos, no se encontraron diferencias significativas en la mayoría de las variables del cuestionario, salvo en el caso de la edad de inicio de la actividad sexual ($p = 0.035$), donde el análisis *post hoc* de Tukey determinó que las pacientes del grupo IC tienen una media de edad de inicio de la actividad sexual mayor que la del resto de los grupos (21.14 años, $p = 0.043$).

Análisis de polimorfismos y árboles filogenéticos en las secuencias génicas

Las secuencias de los fragmentos L1_1, L1_2 y L2_2 fueron recibidas del servicio de secuenciación en formatos .docx (secuencia en texto) y .ab1 (electroferograma). Para analizar las secuencias, las secuencias en formato .docx fueron transformadas a formato .txt y comparadas mediante BLASTn (NCBI) con las secuencias presentes en la literatura y verificar su identidad como secuencias pertenecientes a los genes L1 y L2 de VPH 51. Posteriormente, las secuencias confirmadas como pertenecientes al genoma de VPH 51 fueron procesadas manualmente para eliminar los primeros 25-50 nt de cada secuencia según su alineamiento con la secuencia más cercana a la reportada previamente, y cada secuencia perteneciente a cada fragmento fue convertida a formato FASTA. En total, 44 secuencias concordantes con alineamientos de secuencias ya reportadas de L1 y L2 de VPH 51 fueron consideradas, mientras que 10 (22.72%) fueron eliminadas debido a que reportaban una mayor similitud con VPH 82 (filogenéticamente cercano con VPH 51 y considerado sólo de potencial riesgo oncológico).

Las secuencias reunidas bajo un mismo archivo FASTA (un archivo por cada segmento analizado) fueron sometidas a un alineamiento múltiple de secuencias (MSA) empleando el programa MUSCLE (48), y fueron visualizadas empleando los

programas Jalview (54) y MEGA (51). El número de polimorfismos por segmento fueron calculados manualmente por conteo simple e identificación de posición a partir de la posición 0 definida según la posición del codón de inicio correspondiente al marco de lectura adecuado para los fragmentos L1_1 y L2_1, y del primer punto donde se presentara una conservación superior al 80% hasta la identificación del codón de paro en las secuencias L1_2 y L2_2, en ambos casos tomando en cuenta la secuencia de referencia de los genes L1 o L2 según correspondiera. Los polimorfismos identificados, así como su prevalencia, se muestran en las tablas 2-5. Se eligió un umbral de relevancia de 20% de prevalencia para un polimorfismo en su grupo de lesión (acorde con el número de muestras en cada uno). Los principales hallazgos al momento de realizar la comparación e identificación de polimorfismos comunes fueron los siguientes:

- L1_1: en este segmento se encontraron cuatro polimorfismos en el grupo NILM, todos con una prevalencia de 28.57%. Dos de estos polimorfismos, A126C y G495T, también se identificaron en el grupo HSIL, con una prevalencia de 40%. En el grupo LSIL, sólo dos (A129T y A266C) de los cinco identificados tienen una prevalencia superior al 20%, y en IC se identificó una inserción, Ins[794C], no presente en ninguno de los otros grupos.
- L1_2: un mayor número de polimorfismos fueron encontrados en este fragmento: 16 en NILM, 12 en LSIL, 17 en HSIL y 13 en IC, dentro de los cuales HSIL reúne la mayor proporción de polimorfismos con una prevalencia superior a 20% (hasta 50% respecto a la secuencia de referencia en algunos de ellos). Para polimorfismos comunes, sólo T270C se encontró en dos grupos – NILM y HSIL – mientras que el resto fueron únicos para su grupo.
- L2_1: para este fragmento, el hallazgo más notorio es la presencia universal de tres polimorfismos en todos los grupos de lesión – Ins[285A], G292A y Del[298]. Asimismo, el nucleótido G96 también se encuentra modificado en todos los grupos de lesión, pero la base de sustitución es distinta (G96A en LSIL, HSIL e IC, y G96T también se encontró en HSIL; en todos los casos la prevalencia de este polimorfismo fue superior al 20%).

- L2_2: en este fragmento no se encontraron polimorfismos compartidos entre los grupos, pero se identificó C149A en todas las muestras del grupo NILM, C142A en el 60% de las HSIL, y los cinco polimorfismos de sustitución identificados en IC tuvieron una prevalencia superior al umbral.

Tabla 2: Polimorfismos detectados en el alineamiento múltiple de secuencias para el fragmento L1_1 y su prevalencia, expresada como porcentaje a la derecha de cada polimorfismo. Cada prevalencia superior al umbral de 20% se resalta en negrita, y cada polimorfismo encontrado en más de un grupo de lesión se señala sombreado.

NILM		LSIL		HSIL		IC	
A126C	28.57	A129T	23.08	A126C	40.00	Ins612C	14.29
C247T	28.57	A158G	7.69	G171A	10.00		
G495T	28.57	A266C	23.08	C251T	40.00		
C527G	28.57	T615G	15.38	G255A	20.00		
		T616G	15.38	A429G	20.00		
				G495T	40.00		
				C532G	40.00		
				C541T	20.00		
				A615T	20.00		

Tabla 3: Polimorfismos detectados en el alineamiento múltiple de secuencias para el fragmento L1_2 y su prevalencia, expresada como porcentaje a la derecha de cada polimorfismo. Cada prevalencia superior al umbral de 20% se resalta en negrita, y cada polimorfismo encontrado en más de un grupo de lesión se señala sombreado.

NILM		LSIL		HSIL		IC	
G132A	28.57	T2G	7.69	C119T	20.00	Ins14T	28.57
A142C	28.57	T5C	7.69	A134C	50.00	Ins80T	14.29
A213G	28.57	A31C	7.69	A145C	20.00	Ins119	14.29
T234C	14.29	G131A	38.46	A216G	50.00	G147A	42.86
A243G	28.57	A141C	15.38	A246G	20.00	A158C	28.57
T270C	28.57	A212G	38.46	T273C	20.00	A158T	14.29
A346C	28.57	A345C	23.08	T270C	20.00	A234G	42.86
T360C	28.57	A443G	23.08	T279C	20.00	A264G	42.86
G430A	28.57	A484C	23.08	A317T	10.00	T291C	28.57
A444G	28.57	A545G	23.08	A344C	40.00	T291A	14.29
G492A	14.29	A221C	7.69	T363C	40.00	A368C	42.86
A531C	14.29	Del[546:550]	7.69	T350C	20.00	A473G	42.86
G532C	14.29			A447G	40.00	G521A	28.57
C533G	14.29			T453G	20.00		
C536G	14.29			G525A	20.00		
C539G	14.29			C551T	20.00		
				A560T	20.00		

Tabla 4: Polimorfismos detectados en el alineamiento múltiple de secuencias para el fragmento L2_1 y su prevalencia, expresada como porcentaje a la derecha de cada polimorfismo. Cada prevalencia superior al umbral de 20% se resalta en negrita, y cada polimorfismo encontrado en más de un grupo de lesión se señala sombreado.

NILM		LSIL		HSIL		IC	
Ins7A	42.86	A28G	15.38	G96T	40.00	G3C	28.57
T46G	28.57	G94A	23.08	G96A	20.00	G4C	28.57
T137G	28.57	G96A	23.08	G97A	10.00	G7A	28.57
C244T	28.57	G150A	15.38	A192C	20.00	C8A	28.57
Ins285A	100.00	A274G	23.08	C238T	40.00	G96A	14.29
G292A	100.00	Ins285A	100.00	G247A	10.00	Ins284A	100.00
Del[298]	100.00	G292A	100.00	Ins285A	100.00	G291A	100.00
G324A	28.57	Del[298]	100.00	G292A	100.00	Del[297]	100.00
C381T	14.29	C305T	7.69	Del[298]	100.00	G328C	14.29
C435A	14.29	G329A	15.38	G329A	40.00	G421T	14.29
C435T	14.29	G329C	23.08	G329C	10.00	A433C	14.29
Ins553G	100.00	C433A	38.46	G362A	10.00	InsG536	100.00
A561G	14.29	T436C	15.38	C381T	40.00	T539C	100.00
A561T	14.29	Ins536G	100.00	Ins536G	100.00	Ins[TA]559	100.00
Ins575T	100.00	T540C	100.00	C433A	50.00		
Ins582T	100.00	Del[559:560]	100.00	C436T	20.00		
T588A	14.29	A561C	100.00	A557C	100.00		
T588C	14.29	C562T	100.00	C558T	100.00		
C717A	28.57	T610C	7.69	C629A	30.00		
A719G	14.29	G695T	7.69	C686T	20.00		
C722A	14.29	T696G	7.69	C696T	50.00		

Tabla 5: Polimorfismos detectados en el alineamiento múltiple de secuencias para el fragmento L2_2 y su prevalencia, expresada como porcentaje a la derecha de cada polimorfismo. Cada prevalencia superior al umbral de 20% se resalta en negrita, y cada polimorfismo encontrado en más de un grupo de lesión se señala sombreado.

NILM		LSIL		HSIL		IC	
T14G	14.29	T110C	7.69	T104C	30.00	T176G	28.57
C149A	100.00	C126T	7.69	C142A	60.00	T176C	28.57
T501A	28.57	G193A	23.08	G210A	40.00	G334T	42.86
C517G	28.57	C233A	38.46	A325C	10.00	G336T	28.57
C518A	28.57	C412A	23.08	T478G	10.00	C371G	28.57
C533G	28.57	A435C	30.77	T516G	10.00		
A578C	28.57	C440A	7.69	A523G	20.00		
G614A	28.57	A443C	7.69	G640A	40.00		
A615G	28.57	T533G	15.38	G745A	20.00		
G620T	28.57	T600G	23.08	T835A	20.00		
C634G	28.57	C605G	7.69				
T645A	28.57	A647C	23.08				
T647A	28.57						
T692A	28.57						
T834G	28.57						

NILM		LSIL	HSIL	IC
T785G	14.29			

Con los alineamientos respectivos, también se construyeron árboles filogenéticos utilizando el programa MEGA X (51), empleando el método de Tamura-Nei a 200 iteraciones y demás configuración estándar. Los árboles finales fueron calculados, para cada fragmento, primero para todas las secuencias en conjunto y posteriormente para cada grupo de lesión. Los árboles de cada fragmento se observan en las figuras 14-17, junto con sus log likelihoods (LLs). Tras realizar una comparación del cálculo de probabilidad de las ramificaciones en cada árbol y del árbol global, se observó lo siguiente:

- L1_1: en el árbol global, las muestras de los grupos LSIL y HSIL tienden a agruparse en los mismos clústeres, mientras que para NILM e IC tienden a dispersarse de acuerdo con sus respectivas similitudes entre los clústeres formados por LSIL y HSIL. Las muestras IC80 e IC126 son las más lejanas al resto, de manera consistente con la observación de un alto número de polimorfismos únicos para estas dos secuencias. Para los árboles individuales, se observa que tanto en NILM como en IC las muestras tienden a agruparse por pares en una alta proporción de los árboles calculados, en LSIL cada muestra se separa en una baja proporción de los árboles (indicativo de una divergencia dispersa entre la mayor parte de las muestras) y para HSIL casi todos los árboles calculados (valores de ramas cercanos al 100) generan los mismos clústeres.
- L1_2: para el árbol global, se observa que IC60 e IC80 presentan la mayor divergencia, mientras que para el resto de las muestras se observa una baja probabilidad de formación de subgrupos que combinen NILM con los otros grupos de lesión. Se identifica aquí también que LSIL y HSIL tienden a formar un mayor número de clústeres comunes a todos los árboles calculados, y el resto de las muestras se dispersan entre dichos grupos. Al momento de separar por grupo de lesión, en el grupo NILM las muestras se organizan por pares de casi la misma distancia génica, LSIL tiende a conservar el

comportamiento de divergencia por pares de muestras en una alta proporción de árboles calculados, para HSIL cuatro de las muestras (HSIL64, HSIL83, HSIL94 y HSIL113 no forman clústeres sino que se van separando sucesivamente del clúster principal formado por el resto de las muestras, y en IC se identifica que, a diferencia del resto de las ramas del árbol, el clúster de IC60 e IC80 siempre se observa independientemente de la iteración actual.

- L2_1: NILM78 y NILM101, tanto en el árbol global como en el individual de NILM, se observan como muestras con una divergencia marcada respecto a otros clústeres formados. Las muestras LSIL y HSIL, una vez más, tienden a agruparse entre ellas mismas en el árbol global para casi todos los calculados, con una alta probabilidad de verosimilitud para los árboles individuales de ambos grupos. Respecto a IC, la muestra que presenta una mayor distancia de las demás es IC9, que es la única en no formar subgrupo con otra excepto con la secuencia de referencia.
- L2_2: aquí, NILM95 y NILM124 son las dos muestras de su grupo que divergen más temprano tanto en el árbol global como en el intragrupal, con el resto de las muestras distribuyéndose entre diferentes clústeres en el árbol global. LSIL58 y LSIL69 muestran un comportamiento similar en su grupo, lo mismo que HSIL11 e IC9. Para el resto de los casos, la distribución obedece a una divergencia por pares con probabilidad alta, pero en este caso las muestras NILM también muestran un comportamiento a agruparse entre ellas mismas, similar a LSIL y HSIL.

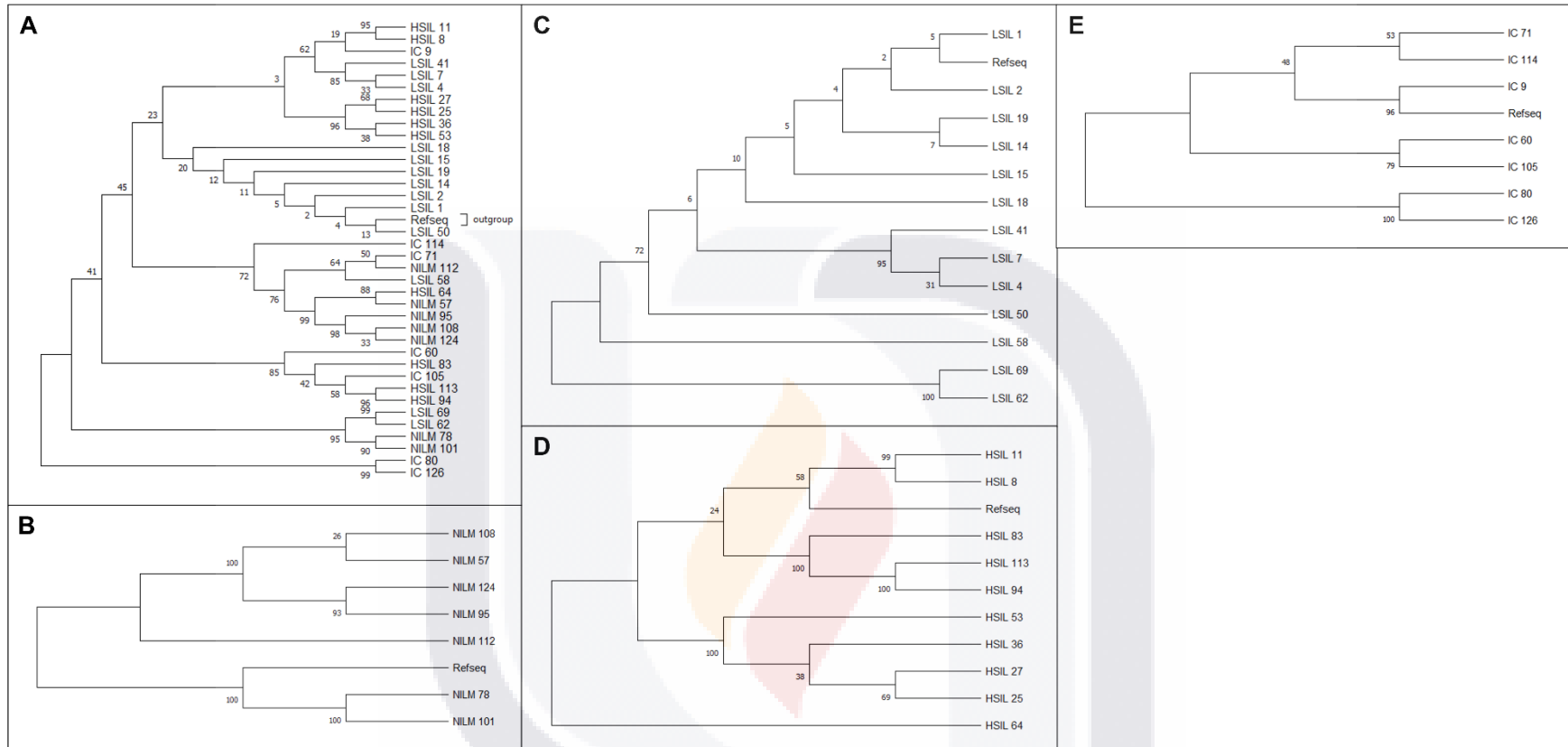


Fig. 14: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias del fragmento L1_1. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -4691.87, B (NILM) = -2188.00, C (LSIL) = -1680.51, D (HSIL) = -1809.20, E (IC) = -2990.45.

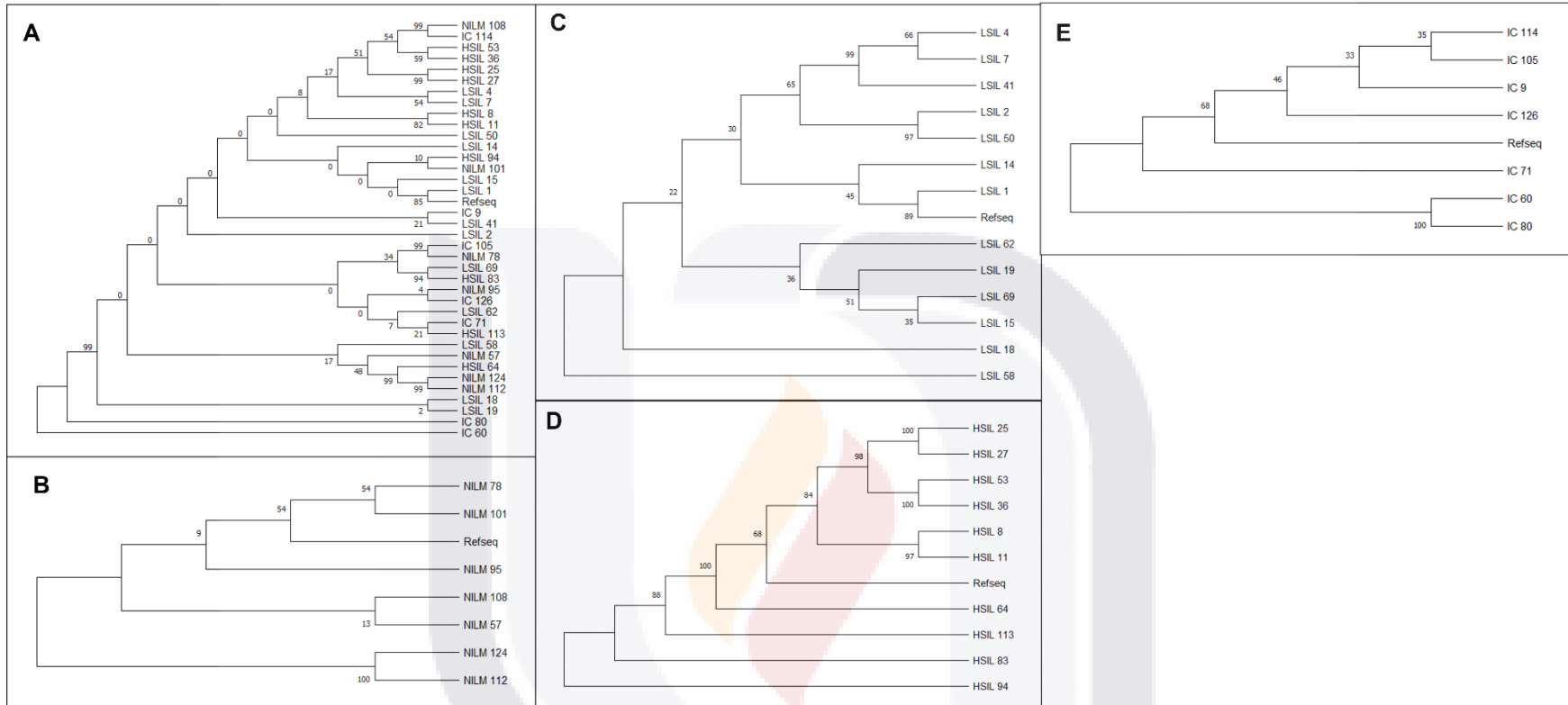


Fig. 15: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias del fragmento L1_2. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -10333.74, B (NILM) = -3687.07, C (LSiL) = -2724.25, D (HSiL) = -3580.05, E (IC) = -5505.22.

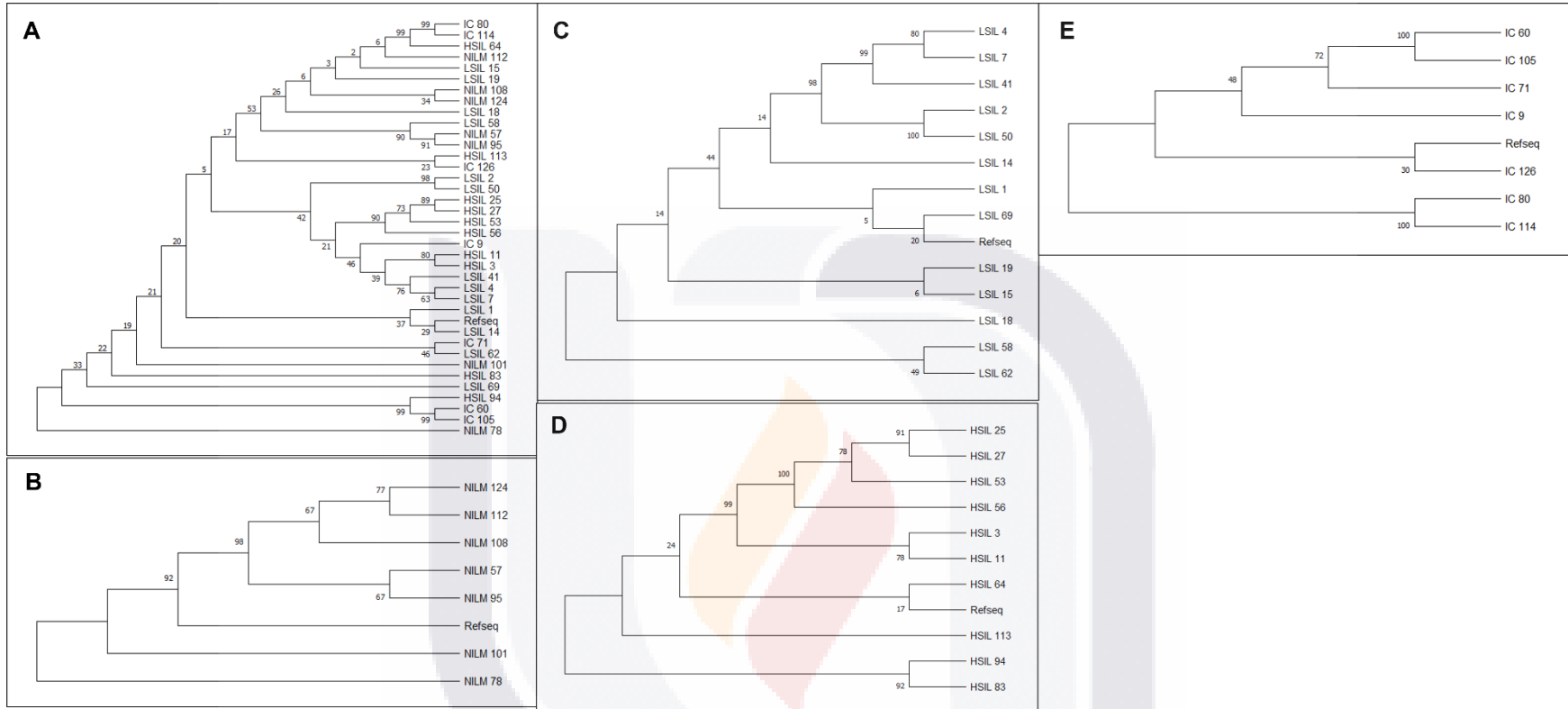


Fig. 16: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias del fragmento L2_1. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -3716.46, B (NILM) = -2362.07, C (LSIL) = -1529.09, D (HSIL) = -1968.31, E (IC) = -1768.45.

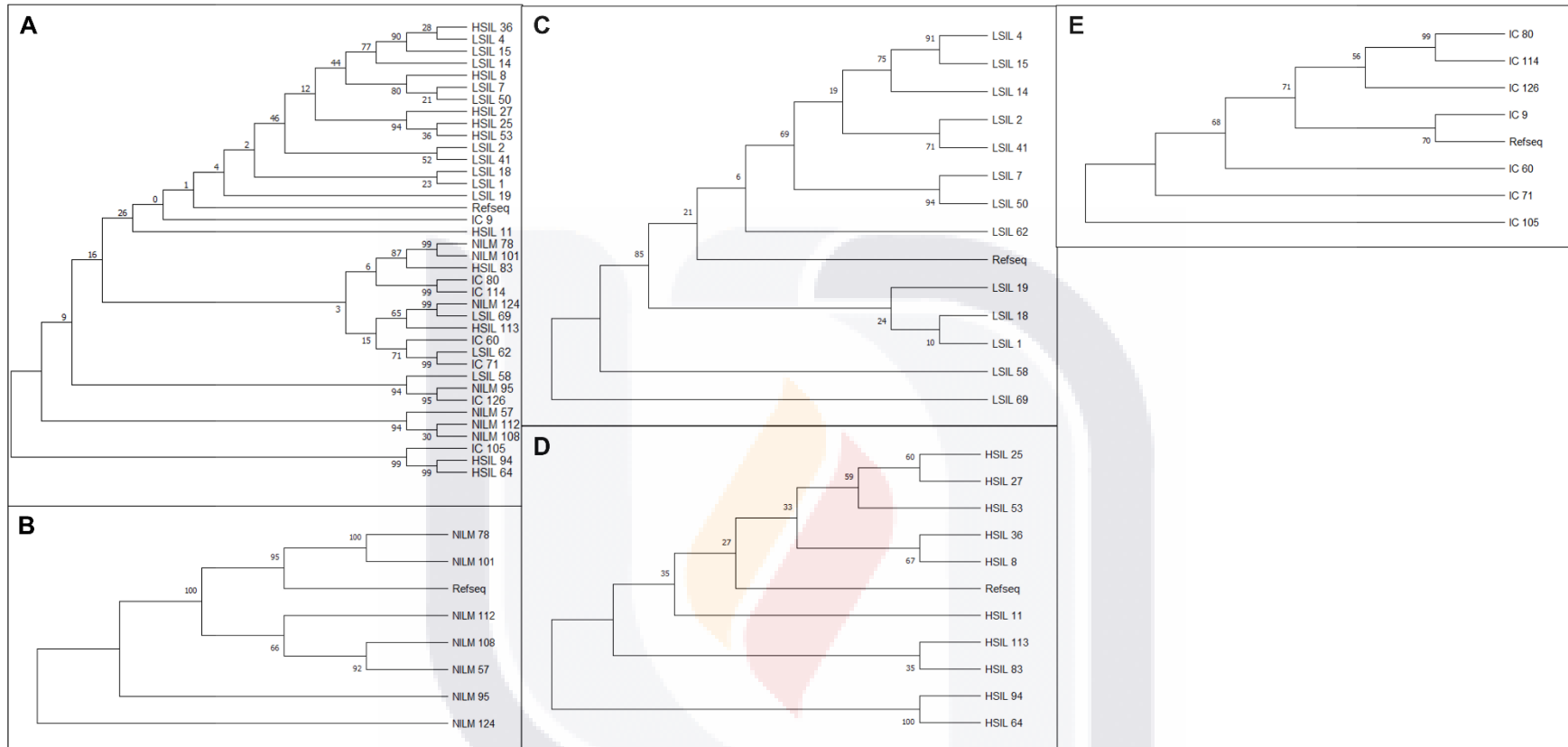


Fig. 17: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias del fragmento L2_2. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -9161.55, B (NILM) = -2700.40, C (LSIL) = -3382.50, D (HSIL) = -3341.66, E (IC) = -3312.79.

A continuación, para verificar cómo se comportaba la selección de los polimorfismos detectados, la distancia génica y determinar si existiría la probabilidad de divergencia con estas secuencias en el futuro, se realizó una prueba de neutralidad génica empleando el método de D de Tajima (52) en el programa MEGA X (51). La prueba se realizó para cada fragmento, primero tomando todas las secuencias independientemente de su grupo de lesión, y posteriormente agrupándolas por grupo de lesión. Un valor de $|D| > 2$ se consideró como indicativo de selección génica positiva o negativa (dependiendo del signo), significativo si $\theta < 0.05$. Los resultados del análisis global para cada fragmento se muestran en la tabla 6. Aquí, el fragmento L1_2 y L2_2 poseen de manera más marcada un mayor número de sitios variables (997 y 911 respectivamente) en comparación con L1_1 y L1_2, que ya se ha demostrado poseen una mayor estabilidad. De acuerdo con el análisis, L1_1 y L1_2 poseen un indicativo de selección negativa (mayor tendencia a la divergencia que el valor crítico); sin embargo, ninguno de los cuatro fragmentos posee una aneutralidad significativa, por lo que se considera que no divergen en el tiempo más allá de los límites estadísticos establecidos por el modelo.

Tabla 6: Análisis de neutralidad génica para las secuencias de cada fragmento, empleando D de Tajima. Todas las secuencias alineadas con éxito para cada fragmento son consideradas en el análisis, que emplea 200 iteraciones para realizar una predicción de selección o aneutralidad en el tiempo. m: número de secuencias empleadas en el análisis. S: número total de sitios variables encontrados en el alineamiento. Ps: proporción de sitios variables respecto al número total de sitios probables de variación. θ : estadístico de significancia para la selección positiva o negativa. π : diversidad nucleotídica relativa. D: estadístico de Tajima.

Secuencias	m	S	Ps	θ	π	D
L1_1	38	537	0.50234	0.11956	0.05963	-1.89132
L1_2	38	997	0.73579	0.17512	0.07499	-2.16295
L2_1	38	318	0.33974	0.08086	0.03233	-2.25655
L2_2	38	911	0.64564	0.15367	0.08228	-1.75679

Por su parte, el análisis por grupos, mostrado en la tabla 7, identificó en L1 para LSIL y HSIL (0.022 y 0.029) y en L2_1 para LSIL, HSIL e IC (0.015, 0.043 y 0.040) una selección significativamente mayor respecto al valor crítico, pero que no es indicativa de una mayor acumulación de polimorfismos respecto al tiempo (ninguno de sus valores D es inferior a -2). Por otro lado, en L1_2 y L2_2, en ambos casos para LSIL, sí hay una mayor acumulación de polimorfismos, pero su selección no

es significativa. Se identificó entonces que, si bien en todos los casos hay secuencias candidatas a ofrecer una mayor selección génica respecto al crítico, la selección aún permanece estable y, a nivel génico, no hay indicativo de divergencias actuales respecto a las secuencias de referencia que indiquen la presencia persistente de variantes únicas, o su probabilidad de aparición constante.

Tabla 7: Análisis de neutralidad génica para las secuencias de cada fragmento separadas por grupo de lesión, empleando D de Tajima. Todas las secuencias alineadas con éxito para cada fragmento, y para cada grupo de lesión, son consideradas en el análisis, que emplea 200 iteraciones para realizar una predicción de selección o aneutralidad en el tiempo. m: número de secuencias empleadas en el análisis. S: número total de sitios variables encontrados en el alineamiento. Ps: proporción de sitios variables respecto al número total de sitios probables de variación. θ : estadístico de significancia para la selección positiva o negativa. π : diversidad nucleotídica relativa. D: estadístico de Tajima.

Secuencias	m	S	Ps	θ	π	D
L1_1_NILM	8	132	0.138510	0.053420	0.061760	0.852047
L1_1_LSIL	14	68	0.072110	0.022675	0.016058	-1.291377
L1_1_HSIL	11	81	0.085624	0.029233	0.023083	-1.004755
L1_1_IC	8	425	0.411822	0.158829	0.164244	0.187204
L1_2_NILM	8	320	0.272109	0.104946	0.083819	-1.104389
L1_2_LSIL	14	225	0.197368	0.062063	0.029005	-2.399592
L1_2_HSIL	11	353	0.300426	0.102570	0.068395	-1.613995
L1_2_IC	8	748	0.598879	0.230973	0.192182	-0.923297
L2_1_NILM	8	208	0.227074	0.087577	0.065073	-1.406906
L2_1_LSIL	14	43	0.048260	0.015176	0.009053	-1.759923
L2_1_HSIL	11	115	0.128348	0.043820	0.028693	-1.657509
L2_1_IC	8	94	0.105145	0.040552	0.038631	-0.257667
L2_2_NILM	8	358	0.391257	0.150898	0.117877	-1.200999
L2_2_LSIL	14	498	0.491124	0.154435	0.082201	-2.116299
L2_2_HSIL	11	477	0.447887	0.152916	0.131609	-0.675720
L2_2_IC	8	473	0.499472	0.192634	0.136370	-1.604253

Traducción *in silico* y análisis de mutaciones en secuencias proteicas.

Considerando los sitios elegidos para el procesamiento de secuencias génicas, ya previamente descrito, se llevó a cabo la traducción *in silico* de todas las secuencias empleando la herramienta ExPASy (49). Los seis posibles marcos de lectura fueron obtenidos, y de ahí para cada secuencia se seleccionó la secuencia de aminoácidos más probable a generar el fragmento de la proteína L1 o L2 llevando a cabo un alineamiento simple BLASTp (47). La secuencia seleccionada fue convertida a formato FASTA, para su posterior alineamiento múltiple de la misma forma que las

TESIS TESIS TESIS TESIS TESIS

secuencias génicas, empleando configuraciones para secuencias de aminoácidos estándares, y con éstos se construyeron árboles filogenéticos globales y por grupo de lesión para cada fragmento, empleando el método de James-Taylor-Thornton (JTT) a 200 iteraciones. Los árboles calculados se ilustran en las figuras 18 a 21. Realizando las observaciones pertinentes, y comparando con los árboles obtenidos con las secuencias génicas con anterioridad, estos fueron los hallazgos más relevantes:

- L1_1: en el árbol global, se observa que las distancias entre secuencias NILM son inferiores respecto al árbol génico, y existe una mayor probabilidad de encontrar las mismas ramas en un árbol inferido del modelo al reducirse la cantidad de sitios variables en el alineamiento múltiple de proteínas. Los LL de los cuatro árboles intragrupal son similares, y menores que los encontrados en los árboles de secuencias génicas.
- L1_2: la distribución de las muestras LSIL y HSIL en el árbol global es similar al caso del árbol génico, donde se agrupan entre clústeres ellas mismas, con las muestras NILM e IC intercaladas entre dichos clústeres. Sin embargo, sí hay una diferencia importante en la distribución de las muestras NILM en su árbol intergrupar, donde si se observaba una divergencia por pares en las secuencias génicas, en este nuevo árbol cada una de las muestras diverge individualmente de la rama primigenia a intervalos regulares. De entre todos los árboles de alineamientos de secuencias de aminoácidos, los de L1_2 son los que poseen un menor valor LL (por ejemplo, $LL \text{ árbol global} = -10108.37$), por lo que la divergencia en este fragmento ocurre de manera más pronunciada que en el resto de los fragmentos.
- L2_1: los árboles globales, tanto de las secuencias génicas como de aminoácidos, poseen una distribución similar, donde la formación de clústeres es poco pronunciada, y predomina la divergencia de muestras individuales con una probabilidad similar entre las ramas. El comportamiento dentro de cada grupo de lesión tampoco varía considerablemente respecto al árbol génico, excepto en el caso de HSIL, donde HSIL56, HSIL64, HSIL83 y HSIL114 divergen individualmente en lugar de hacerlo por pares. Aquí se

tiene una estabilidad relativa, caracterizada por bajos LLs (particularmente en LSIL, donde el valor de LL es superior a -1000.00, indicativo de una baja divergencia).

- L2_2: respecto al árbol génico global, donde se observaba un gran número de pequeños clústeres, en el árbol proteico se distinguen tres principales ramas que divergen de la rama primigenia, y a partir de las cuales comienzan a ramificarse cada una de las muestras. El comportamiento de los árboles en cada grupo presenta un comportamiento diferente sus respectivos árboles génicos – excepto para NILM. En particular, en LSIL se distingue un patrón de divergencia simétrica que no se presentaba en el árbol génico, mientras que en HSIL hay una divergencia heterogénea entre las muestras. Los LLs identificados son similares entre sí para los árboles intragrupal, ligeramente superiores a los de L2_1.

Para estas secuencias también se realizó un análisis de neutralidad mediante D de Tajima (52), cuyos resultados se exhiben en las tablas 8 (globales) y 9 (por grupo de lesión). Comparado con el análisis de neutralidad de las secuencias génicas, se obtiene un número de sitios variables en los análisis globales equivalente con la traducción *in silico*, marcadamente mayor en L1_2 (S = 399) y L2_2 (S = 283), y mostrando en L1_1 una selección negativa indicativa para L1_1, pero no significativa ($D = -2.08$, $\theta = 0.102$). Para el resto de los fragmentos no se encontraron selecciones indicativas o significativas. Separando por grupo de lesión, se determinó que L1_1_HSIL y L2_1_LSIL poseían selecciones no superiores a las críticas pero significativas con el análisis inferencial, mientras que L1_1_NILM y L2_2_LSIL presentaban selecciones mayores a la crítica, pero no significativas.

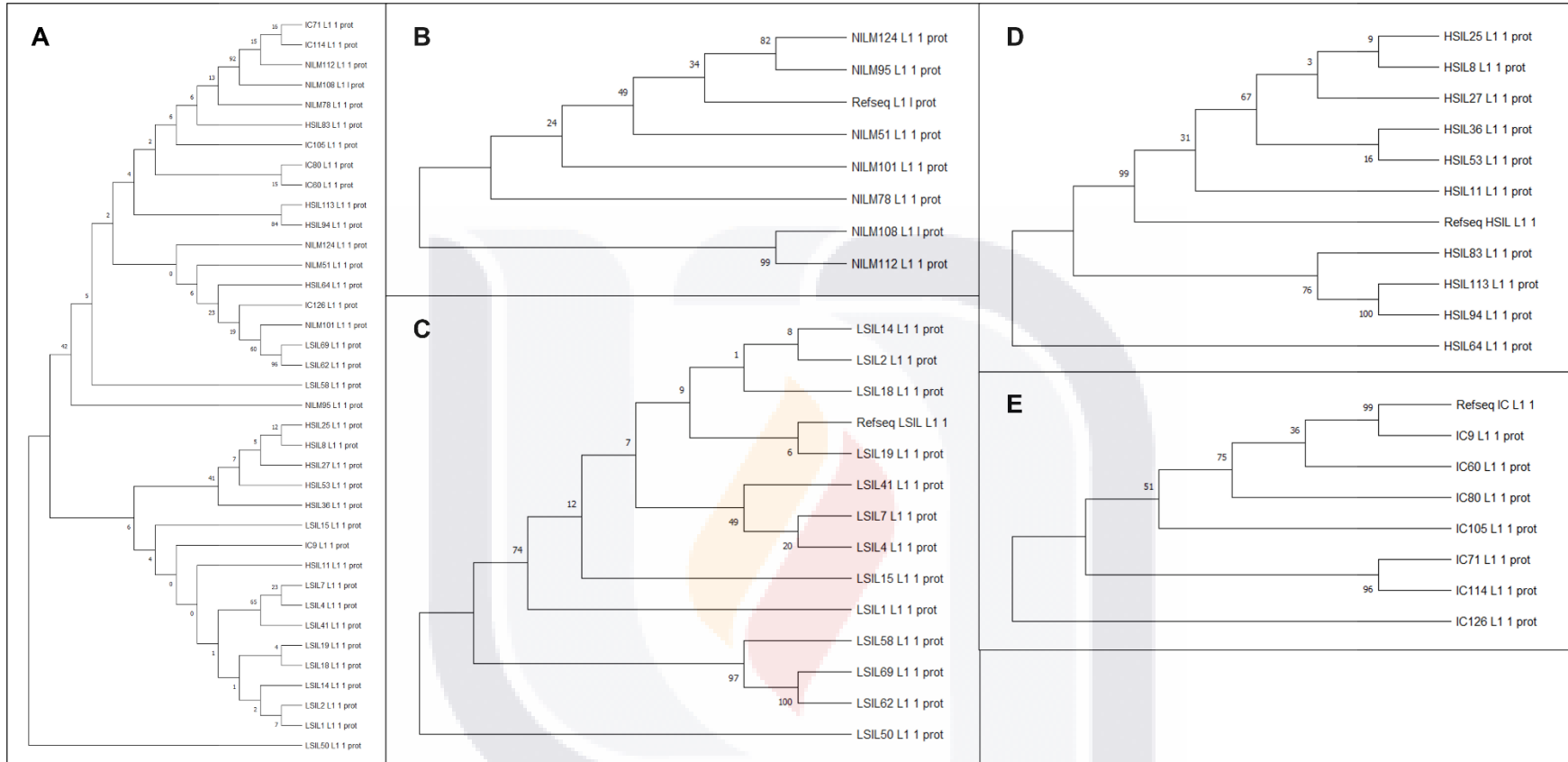


Fig. 18: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias de aminoácidos resultantes de la traducción in silico del fragmento L1_1. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -2393.63, B (NILM) = -1159.52, C (LSIL) = -1326.81, D (HSIL) = -1096.13, E (IC) = -1134.72.

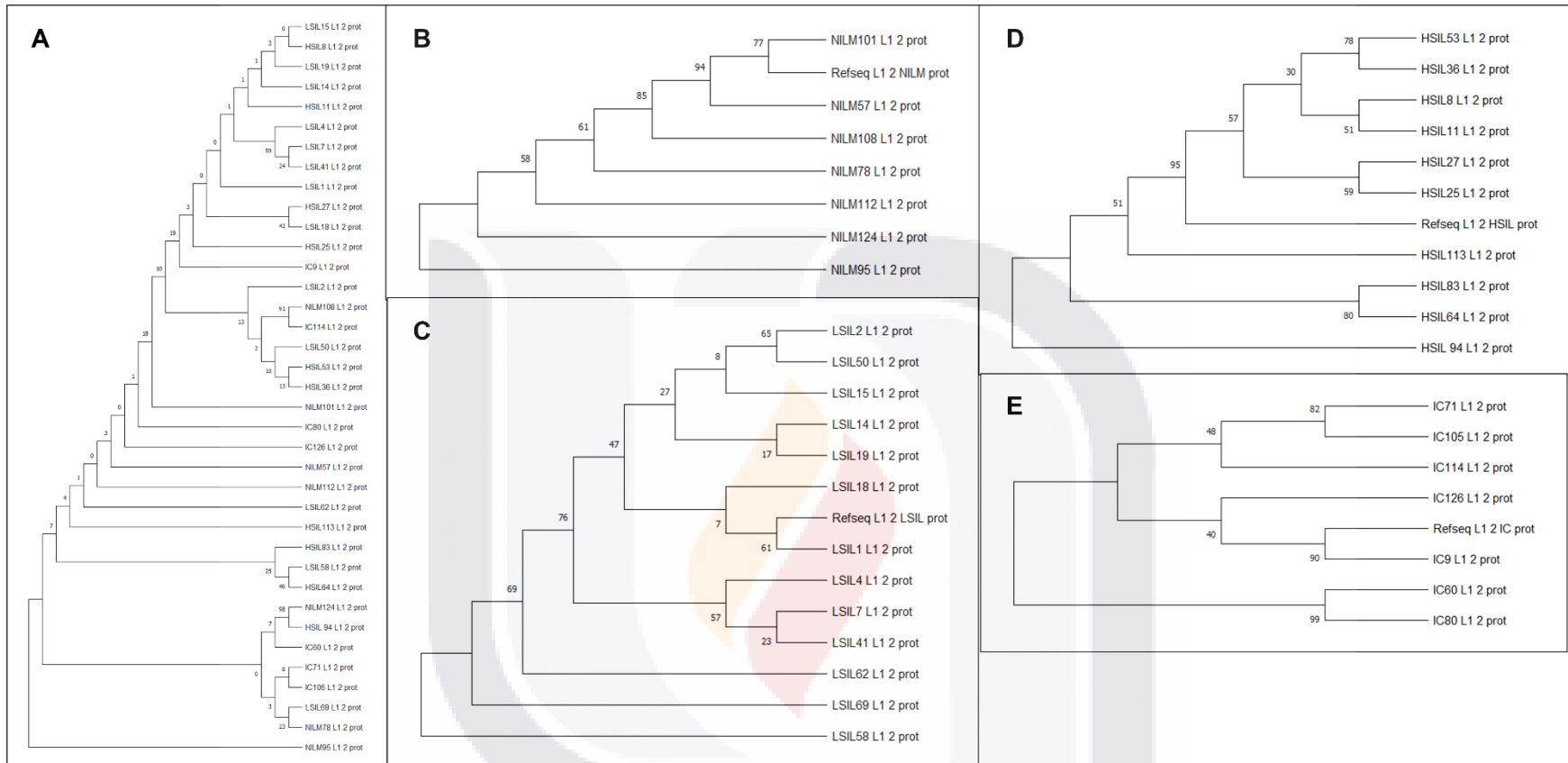


Fig. 19: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias de aminoácidos resultantes de la traducción in silico del fragmento L1_2. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -10108.37, B (NILM) = -3620.82, C (LSIL) = -2015.33, D (HSIL) = -2999.46, E (IC) = -4326.76.

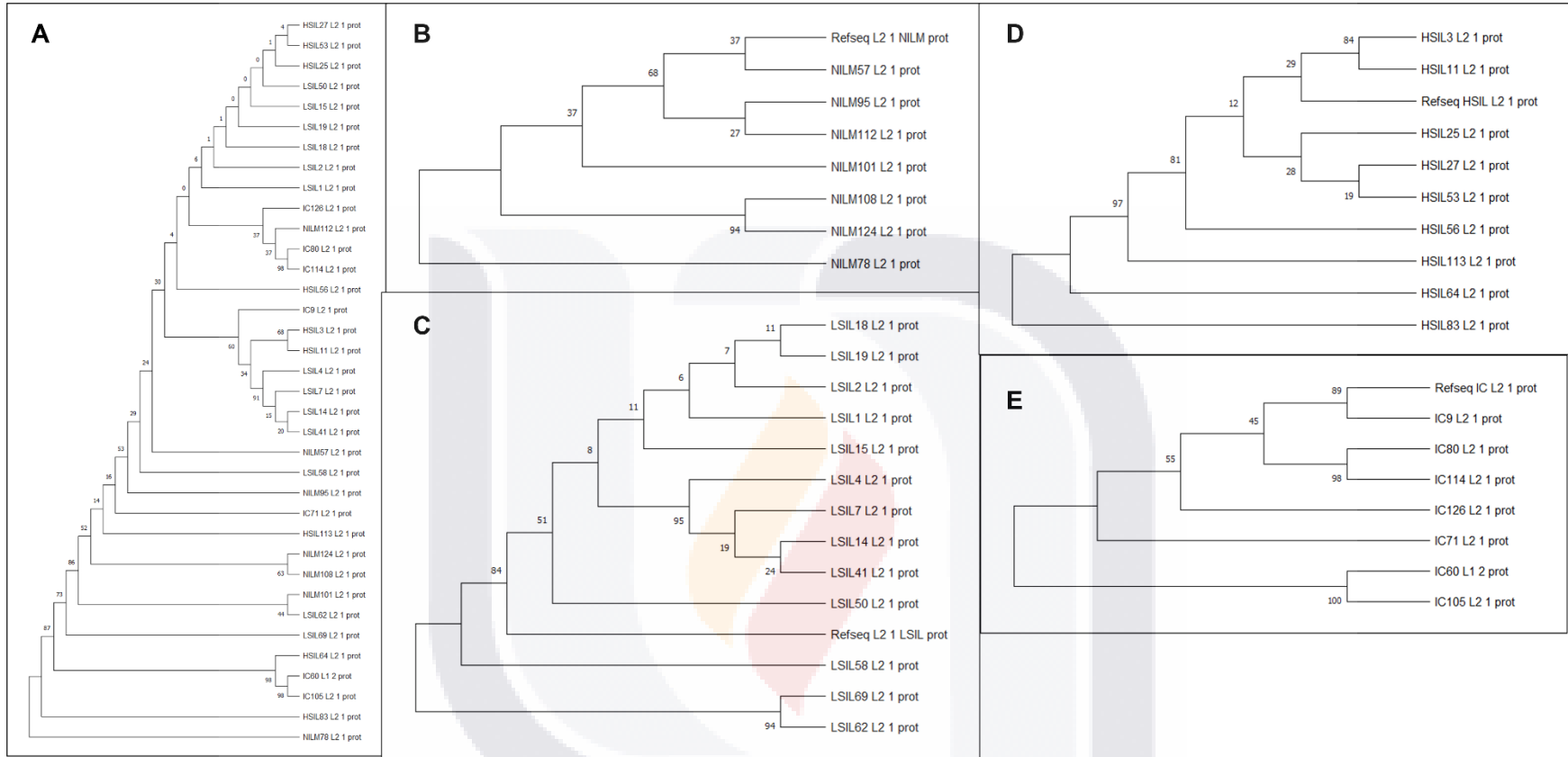


Fig. 20: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias de aminoácidos resultantes de la traducción in silico del fragmento L2_1. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -2639.51, B (NILM) = -1866.42, C (LSIL) = -963.60, D (HSIL) = -1287.33, E (IC) = -1127.55.

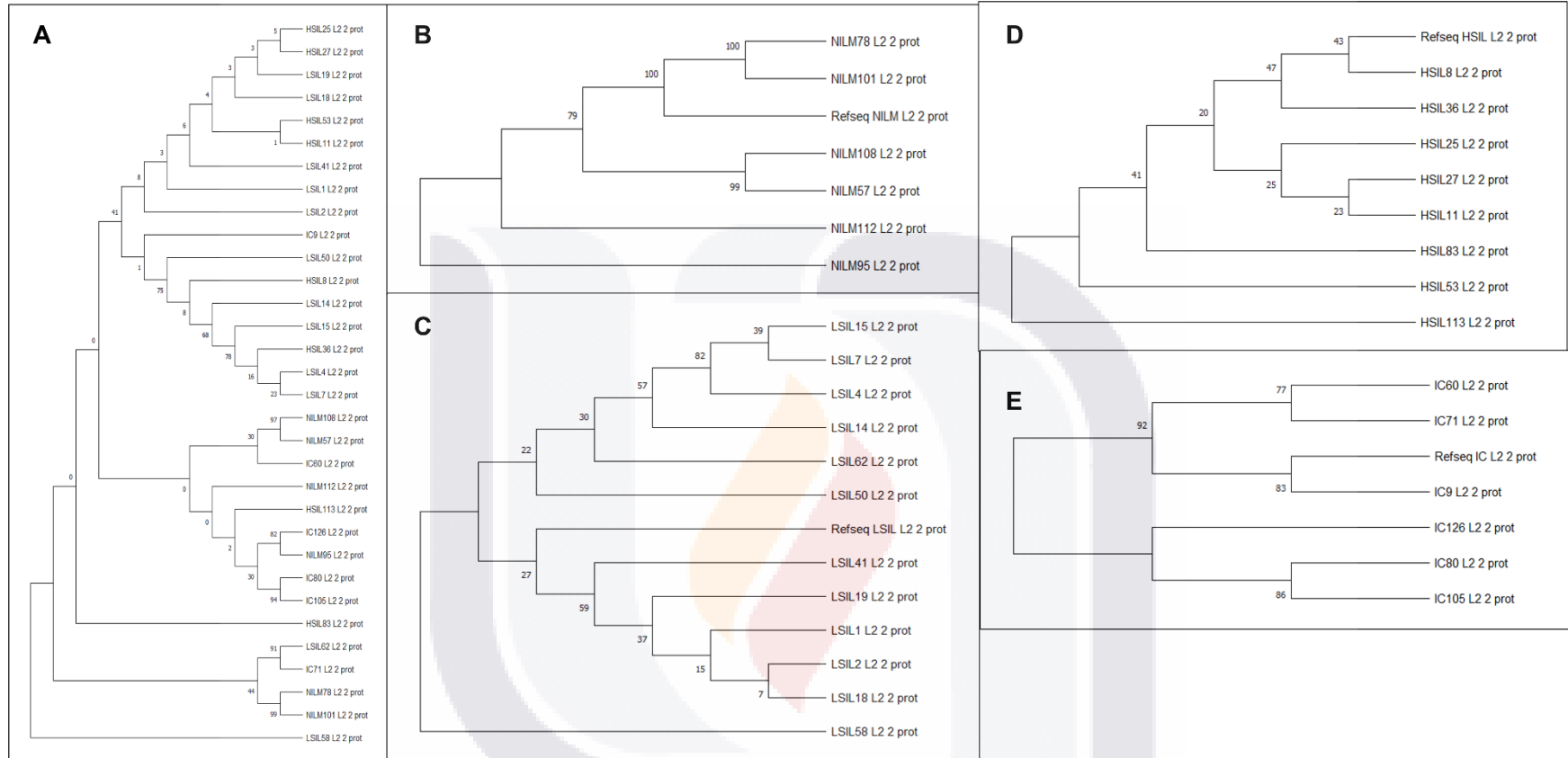


Fig. 21: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias de aminoácidos resultantes de la traducción in silico del fragmento L2_2. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: A (global) = -4182.39, B (NILM) = -1278.45, C (LSIL) = -1779.56, D (HSIL) = -1466.15, E (IC) = -1844.52.

Ensamblaje de consensos y predicción de variaciones estructurales en L1 y L2.

Tomando en cuenta los alineamientos múltiples obtenidos de las secuencias de aminoácidos, se tomaron las secuencias consenso calculadas según el programa Jalwiew (54) y considerando las posiciones relativas a las secuencias de referencia de cada fragmento, se realizó un ensamblaje para obtener una secuencia única de L1 y L2 para cada grupo de lesión. Dichas secuencias fueron sometidas a un alineamiento múltiple para realizar el análisis filogénico y la identificación de variaciones en la secuencia de aminoácidos, cuyos resultados se muestran en las tablas 8 y 9, y la fig. 22. De las mutaciones no silenciosas detectadas, en L1 se concentran todas en la segunda sección (proveniente de L1_2) de la proteína, siendo V277G común a todos los grupos de lesión, y G278S presente en LSIL y HSIL; el resto son variaciones únicas en NILM e IC. En L2, en cambio, la secuencia EPIAP del aminoácido 95 a 99 está presente en todos los grupos de lesión, así como una inserción de glicina en la posición 179 y D180H; también se encuentra presente T300K en NILM y HSIL.

Tabla 8: Mutaciones no silenciosas detectadas en los fragmentos ensamblados de L1 para cada grupo de lesión, respecto a la secuencia de referencia. Las mutaciones detectadas en más de un grupo de lesión se muestran en sombreado.

NILM	LSIL	HSIL	IC
L212A	V277G	V277G	A211C
A214G	G278S	G278S	A214V
A215H			L221Q
Ins[LVCK]221			V277G
Ins[MVIPC]250			G278S
Ins[Y]259			Ins[LTII]287
V277G			Ins[LL]298
G278C			Ins[L]312
			A313T
			T314Q
			Ins[IL]348
			Ins[W]410

Tabla 9: Mutaciones no silenciosas detectadas en los fragmentos ensamblados de L1 para cada grupo de lesión, respecto a la secuencia de referencia. Las mutaciones detectadas en más de un grupo de lesión se muestran en sombreado.

NILM	LSIL	HSIL	IC
D95E	D95E	D95E	D95E

L96P	L96P	L96P	L96P
W97I	W97I	W97I	W97I
H98A	H98A	H98A	H98A
H99P	H99P	H99P	H99P
Ins[G]179	Ins[G]179	Ins[G]179	Ins[G]179
D180H	D180H	D180H	D180H
Ins[LRFYD]255	Ins[G]229	T300K	Ins[TS]464
Ins[WWCATWCFCV]274	Ins[EA]303		Ins[RRWHCGACM]486
T300K	Ins[AEV]344		
Ins[R]486			

El análisis de los árboles filogenéticos mostró una mayor probabilidad de agrupar las ramas en L2 que en L1, aunque en ambos casos el valor LL es indicativo de una poca divergencia. En el caso de L1, IC es la secuencia que más diverge, y LSIL y HSIL son las últimas en separarse, acorde con el análisis de variaciones puntuales donde ambos fragmentos tienen la misma secuencia. Para L2, las secuencias tienden a agruparse más entre sí que con la secuencia de referencia, debido al número mayor de variaciones comunes a todos los fragmentos respecto a la referencia. En este caso, LSIL es la que presenta menor número de variaciones, y la que diverge primero; NILM y HSIL se agrupan al final. El análisis de neutralidad con D de Tajima (tabla 10) muestra una baja selección negativa, cercana a la neutralidad y significativa tanto en L1 como en L2; esto indica que las variaciones al hacer el análisis de los fragmentos completos son poco frecuentes y no indican una constante aparición de variantes divergentes entre las muestras analizadas.

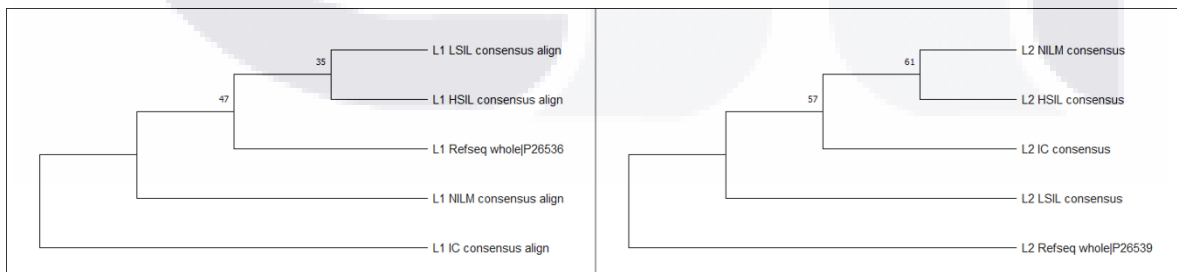


Fig. 22: Árboles filogenéticos calculados a partir de los alineamientos múltiples de las secuencias consenso calculadas para cada grupo de lesión. Los valores en las intersecciones de las ramas indican la probabilidad de 0 a 100 de que dichas ramas sean predichas por el modelo de 200 iteraciones. Refseq se utiliza como outlier para el cálculo de los árboles. Log likelihoods de cada árbol: L1 (izquierda) = -1654.72, L2 (derecha) = -1584.73.

Tabla 10: Análisis de neutralidad génica para las secuencias de cada proteína consenso, empleando D de Tajima. Todas las secuencias alineadas con éxito para cada fragmento, y para cada grupo de lesión, son consideradas en el análisis, que emplea 200 iteraciones para realizar una predicción de selección o aneutralidad en el tiempo. m : número de secuencias empleadas en el análisis. S : número total de sitios variables encontrados en el alineamiento. Ps : proporción de sitios variables respecto al número total de sitios probables de variación. Θ : estadístico de significancia para la selección positiva o negativa. π : diversidad nucleotídica relativa. D : estadístico de Tajima.

Secuencias	m	S	Ps	Θ	π	D
L1_whole	5	11	0.020873	0.010019	0.008918	-0.790615
L2_whole	5	15	0.02994	0.014371	0.012375	-1.015434

Finalmente, para L1 se realizó una verificación de la estructura de las proteínas consenso (fig. 23), modelando mediante el mismo método la secuencia de referencia para comparación. En el caso de LSIL y HSIL, no hay variación en el plegamiento o la estructura de las cadenas, mientras que para NILM sólo se observan las cadenas de la estructura central de mayor tamaño debido a las inserciones detectadas. En IC sí se observa un plegamiento anormal de las cadenas, con deficiencias en el ensamblado hexamérico debido al mayor número de inserciones detectadas. Sin embargo, pese a ello, aún en IC se mantiene la estructura base conocida de L1 en VPH 51. En el caso de L2, este análisis no fue posible debido a que se desconoce la estructura exacta de L2 incluso en genotipos más comunes como VPH 16, por lo que se requiere un modelaje *de novo* que no se puede llevar a cabo debido al número insuficiente de secuencias consenso para obtener un modelo probabilístico confiable (se requieren al menos 5 ensamblados por grupo de lesión para obtener un modelo mínimo).

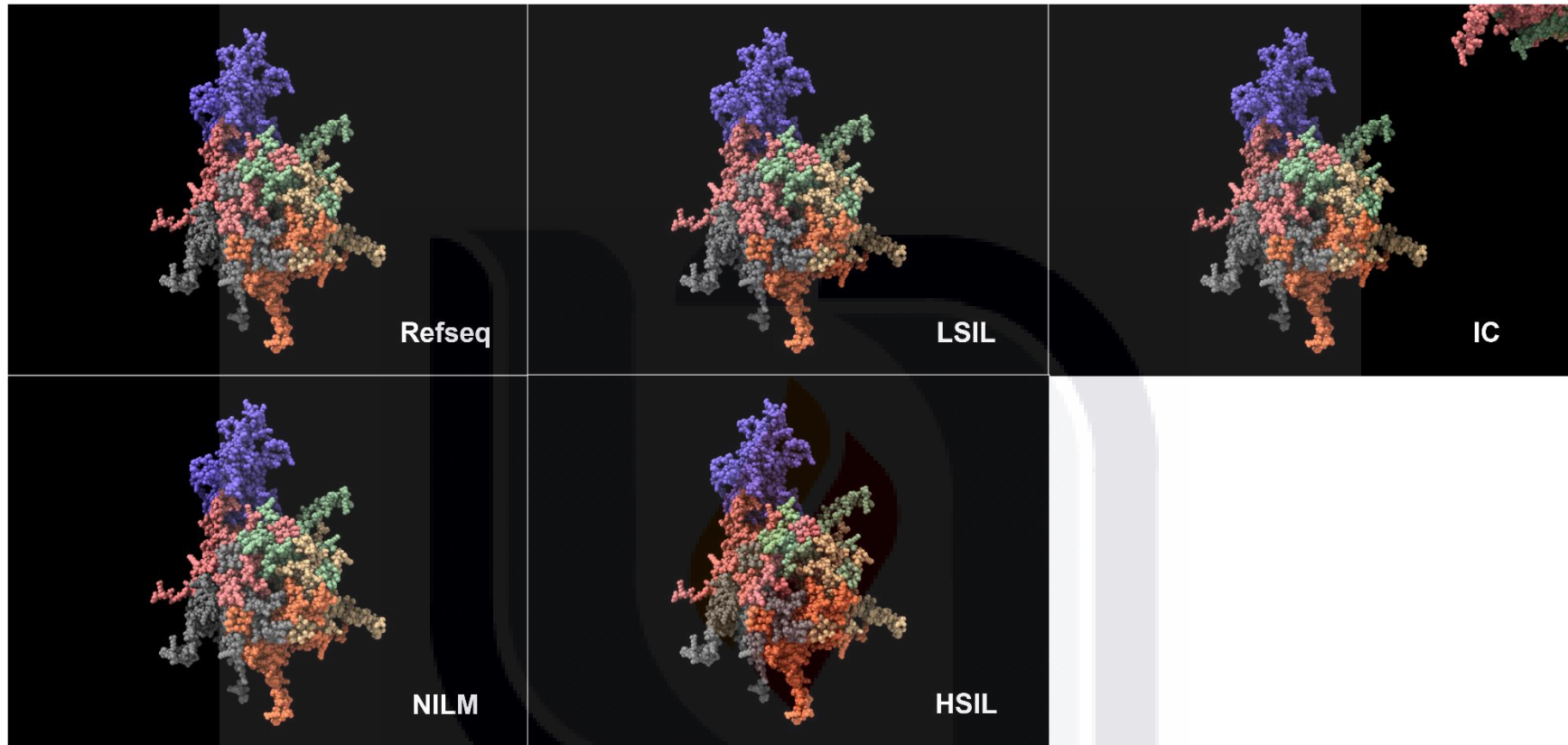


Fig. 23: Modelos 3D de las proteínas consenso L1 calculadas para cada grupo de lesión. En todos los casos, la predicción se realizó mediante configuración estándar a partir del modelo más cercano obtenido en .pb.

DISCUSIÓN

El papel del virus del papiloma humano en el desarrollo del cáncer cervicouterino es ampliamente conocido desde hace casi 40 años; sin embargo, los mecanismos por los cuales sólo un grupo selecto de genotipos de entre los más de 200 existentes (de los cuales no todos infectan al humano) son considerados de alto riesgo para el desarrollo del cáncer cervicouterino no siempre se explican con claridad. Si bien alrededor del 85% de todos los casos de cáncer cervicouterino reportados a nivel mundial incluyen la infección con los genotipos 16 y/o 18 del VPH, de manera regional otros genotipos han sido analizados, y su prevalencia ha resultado inusualmente alta.

El genotipo 51 del VPH fue caracterizado por primera vez en 1991 (55), y está considerado dentro de los genotipos de alto riesgo para el desarrollo de cáncer cervicouterino, así como de otros tipos de cáncer de células escamosas como el cáncer oral, anal y vaginal (40); esto pese a que la evidencia sugiere que se encuentra en el séptimo lugar de los genotipos de alto riesgo en prevalencia de casos de mujeres infectadas con VPH, con una prevalencia mundial estimada entre 0.5 y 4% de todos los casos confirmados (56). A lo largo de las últimas dos décadas, sin embargo, han surgido estudios que indican una prevalencia más elevada que la reportada a nivel mundial, como ya se ha mencionado con anterioridad.

De manera más local, un estudio llevado a cabo por Gutiérrez-Campos et al. en 2019 identificó que, en pacientes con carcinoma invasor invariablemente del grado de avance, la prevalencia de VPH 51 superaba a la de VPH 16, encontrándose en casi la mitad de las pacientes con carcinoma, y con una prevalencia más elevada que el promedio en grados precursores de lesión cervical neoplásica (6). Un estudio posterior en la región occidental de México realizado por Molina-Pineda en 2020 encontró un resultado similar, pero extendiendo el estudio a la comparación de la prevalencia entre mono- y co-infecciones, y determinaron que VPH 51 se encontraba en una prevalencia similar tanto en solitario como en presencia de al menos otro genotipo de alto riesgo, en particular VPH 16 (57). En este estudio no llevó a cabo un análisis de co-infecciones entre VPH 51 y otros genotipos, pero se

TESIS TESIS TESIS TESIS TESIS

buscará optimizar este tipo de análisis considerando la presencia de otros genotipos de importancia y emergentes, en particular debido a la evidencia de que VPH 51 pudiera interactuar con VPH 16, lo que explicaría su alta prevalencia conjunta en casos de carcinoma invasor en la región centro-occidente de México.

Por su parte, Hao et al. encontraron en un estudio con 53 pacientes con cáncer cervicouterino de diferentes edades que el VPH 51 exhibe una tendencia bimodal en la distribución de edad (58), coincidencia parcial encontrada en este trabajo, donde la edad de inicio de la vida sexual, normalmente correlacionada con la edad propia de la paciente que reporta, fue estadísticamente significativa mayor para las pacientes del grupo de carcinoma invasor (IC) respecto a las que no presentaban lesiones malignas o que presentaban lesiones precancerosas. Hao et al. concluyeron en la importancia de realizar el escrutinio no sólo de lesiones malignas, sino también de lesiones benignas que pudieran modificar su comportamiento en el futuro cuando la edad de la paciente se incrementara. Gianella et al. confirmaron parcialmente este resultado en un estudio de cohorte retrospectiva con 1332 pacientes, donde encontraron una tendencia de incremento de riesgo significativa de progresión a HSIL asociada con la edad creciente de la paciente y la infección por VPH 51 (59).

Los reportes de variantes en VPH 51 también han incluido otras regiones del genoma, tales como LCR para determinar sublinajes específicos de baja prevalencia en la población general (60). Si bien los reportes de regiones como E6 o E7 aún son escasos para VPH 51, sí hay evidencia en otros genotipos de mayor prevalencia y baja variabilidad (como VPH 58) que estas regiones son muy estables entre los diferentes genotipos (58), y la caracterización mediante homología hace factible su uso para determinar linajes de manera más precisa, en particular en poblaciones muy heterogéneas como la mexicana. Los cambios observados en variantes de interés también pueden mostrar efecto sobre epítomos virales, incluso en variantes ya plenamente caracterizadas (61).

Respecto a la etnia, la mayor parte de los estudios de prevalencia de VPH 51 se concentran en Asia suroriental (particularmente China), donde VPH 51 aparece

constantemente como uno de los genotipos de mayor prevalencia fuera de los más prevalentes VPH 16 y VPH 18 (62,63). También en otras regiones como África central el genotipo 51 aparece mencionado, pero con una prevalencia menor en comparación con otros genotipos locales de alto riesgo, en particular VPH 33 (64). La mayor parte de estos estudios se realizan sobre población general e indistintamente de la presencia o estado de evolución de una lesión neoplásica o carcinoma; por ejemplo, el estudio de Gallegos-Bolaños et al. donde encontraron una prevalencia elevada de co-infección de VPH 51 y VPH 52 en pacientes con lesiones neoplásicas, por lo que no se descarta la posibilidad de que el riesgo se incremente si existe la persistencia de genotipo 51 con otros genotipos de alto riesgo (12), aunque falta analizar detalladamente la correlación de dichas co-infecciones con el diagnóstico de una lesión neoplásica de determinado grado.

La variabilidad presente en los genes L1 y L2 del VPH ha sido estudiada por años, y es de especial interés al considerar que el desarrollo de las vacunas profilácticas actualmente distribuidas en el mercado se basa en la obtención de partículas similares a virus (VLPs), las cuales se componen de proteínas de la cápside viral pero carecen del genoma de origen, por lo que no tienen potencial replicativo al ser administradas; como coadyuvantes en la construcción de las VLPs se suelen utilizar otros genomas virales base como bacteriófagos, que completan los espacios en la geometría de la cápside e incrementan su capacidad inmunogénica (65). Las vacunas profilácticas basadas en VLPs de L1, como Gardasil-4 o Gardasil-9 (Merck & Co), son las más empleadas a nivel mundial como profilácticos, mientras que las vacunas basadas en L2 tienen un serio inconveniente: L2 no puede auto ensamblarse para formar VLPs por su cuenta, por lo que la construcción se los vectores de expresión que contengan suficientes partículas inmunogénicas de L2 es un reto mayor (66). Aún con esta dificultad, se han logrado progresos en la obtención de algunas vacunas potenciales que han llegado a etapas variables de ensayo clínico, empleando VLPs de diversas fuentes y que son generalmente de tipo multimérico (66–68). La predicción de las secuencias inmunogénicas en el caso de L2 se hace generalmente mediante análisis *in silico*, de forma similar a como se hace de manera rutinaria en genotipos como VPH16 o VPH18 (69), que consiste de

TESIS TESIS TESIS TESIS TESIS

forma general en la identificación de péptidos en la secuencia de L2 candidatos a activar la respuesta inmune, y posteriormente ensamblarlos en vectores quiméricos que incluyan además adyuvantes para potenciar la respuesta inmune. Dado que L2 tiende a presentar una menor variabilidad que L1 entre subtipos y variantes de un genotipo dado de VPH (65), L2 ha ganado interés como candidato a la reformulación y reestructuración del diseño de las vacunas contra VPH, que podrían incluir a múltiples genotipos dado el potencial inmunogénico cruzado de L2 (67). En este estudio se encontró en L2 la persistencia entre grupos de lesión (y pacientes con ausencia de lesiones cervicales) del polipéptido EPIAP en las posiciones 95 a 99, que al encontrarse lejos de las regiones de señalización nuclear tiene potencial de ser utilizado en el diseño de una vacuna multimérica. Esta evidencia debe ser contrastada con la información disponible de la estructura de L2, dado que este polipéptido variante no se encuentra registrado en la literatura o bases de datos de las variantes más conocidas de VPH 51. La filogenia sugiere que, además, la variabilidad de L2 en VPH 51 de manera regional es estable y corresponde a una variabilidad natural, por lo que un diseño empleado según la evidencia encontrada entre estos estudios y las variantes encontradas en otro lugar, si es que son identificadas apoyándose en un modelo cuasiexperimental, reduciendo así la necesidad de diseñar vacunas regionales y abaratando así el precio de los esquemas de vacunación (70). Respecto a L1, se tiene la presencia común de V277G en todos los grupos de lesión, así como de G278S en LSIL y HSIL, que no presentan impacto en el ensamblaje de los monómeros en la proteína completa según las predicciones del modelo *in silico*. En IC se encontró un ensamblaje anormal en dos de las cadenas, que de acuerdo con el modelo pueden ensamblarse entre sí debido a la presencia de múltiples inserciones de aminoácidos en dichas posiciones; sin embargo, para esta etapa la expresión de L1 se encuentra interrumpida por la integración del genoma viral (71), por lo que no existiría impacto potencial para la formación de partículas virales, pero sí tiene potencial para determinar si esta capacidad de dimerización anómala presenta una capacidad aumentada de infectividad o persistencia del virus en la célula epitelial – lo cual,

TESIS TESIS TESIS TESIS TESIS

dado que estas inserciones no se encuentran reportadas, no es posible determinar hasta este punto.

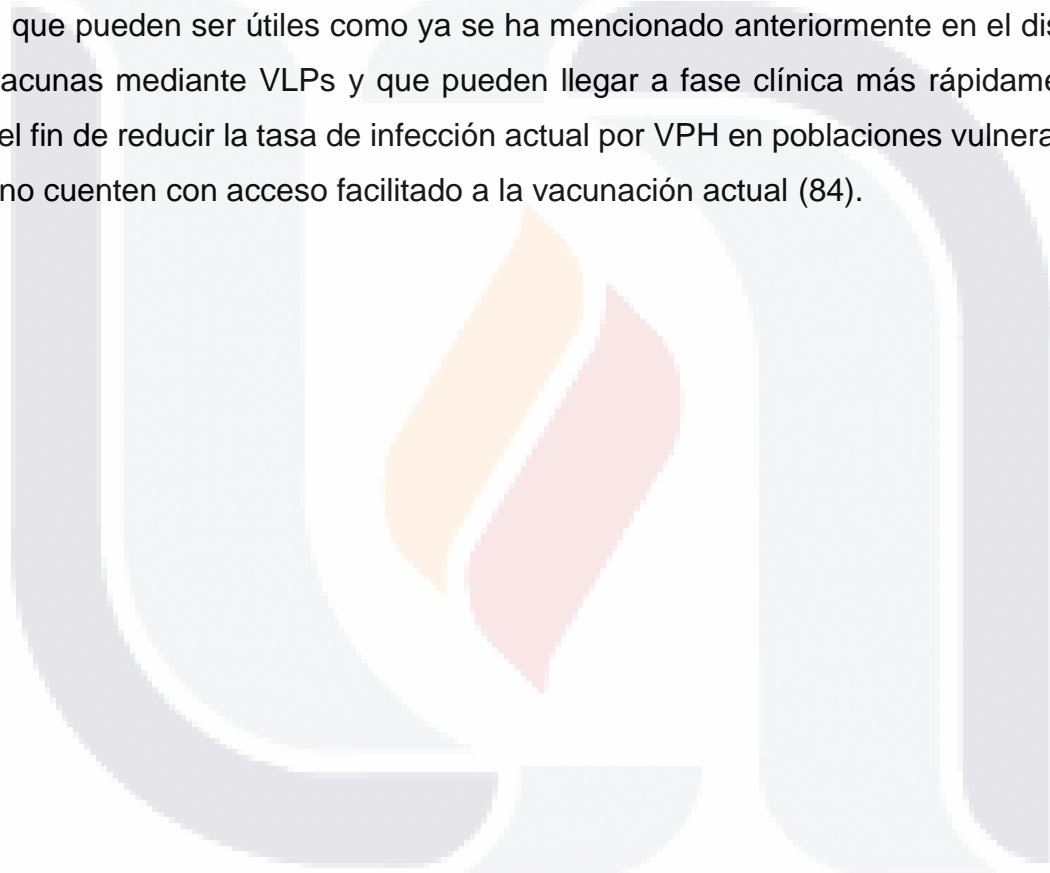
La relación entre los genotipos surgentes de VPH en lesiones cervicales, la variabilidad genética de dichos genotipos y la posibilidad de instalar esquemas de vacunación o reestructurar los existentes requiere, en primer lugar, una caracterización completa de cómo se comportan dichos genotipos en las poblaciones locales (72). En este estudio se observó que, si bien existe una cierta variabilidad natural de VPH 51 circulante en la población femenina afectada (independientemente del grado de lesión), dicha variación no está sujeta a una selección constante que conduzca a la aparición de nuevas variantes en estado activo; la evidencia reunida indica que la divergencia génica en L1 y L2 (particularmente la sección C-terminal de L1) corresponde a la variabilidad esperada según los modelos filogénicos empleados, y que la futura aparición de dichas variantes es poco probable, aunque el incremento de la prevalencia del genotipo 51 podría modificar esto en el futuro.

En cuanto a la filogenia en VPH, particularmente en los genotipos de alto riesgo mejor caracterizados como VPH 16, ha habido avances que permiten elucidar con mayor claridad la evolución del virus con el hospedero. Por ejemplo, Mandishore et al. encontraron que el VPH sí puede co-evolucionar con la respuesta adaptativa del infectado, lo cual tiene efecto significativo sobre el tropismo epitelial y la capacidad de transmisión, particularmente en estadios iniciales o intermedios de la enfermedad (NILM, LSIL e incluso CIN2/3 en HSIL) (73). Por su parte, un metaanálisis publicado por Bowden et al. sugirió que dicha variabilidad asociada a la divergencia y aparición de variantes podría estar relacionada con una mayor persistencia y reducido tiempo de evolución de la lesión cervical a cáncer (74). Empleando técnicas de secuenciación masiva, también ha sido posible determinar, en el caso de HPV 16 en varios de sus genes la presencia de variantes en la secuencia que no solamente determinar un sublinaje, sino que también pueden variar entre hospederos y son reflejo de una constante adaptación del virus a la respuesta inmune (75).

Una vez conocido esto, se puede decidir si el genotipo 51 en cuestión es candidato a la formulación de nuevas vacunas que lo incluyan de manera específica (76), y parece ser que al momento no existe evidencia suficiente de variantes regionales circulantes, al menos no basándose únicamente en las estructuras de L1 y L2. Se recomienda entonces realizar el estudio de secuenciación de genoma completo de las secuencias más variantes en L1 y L2 (por ejemplo, IC60 o IC80, en caso de disponibilidad) para confirmar esta aseveración, o de lo contrario reforzar el banco de secuencias presente con dichos genomas completos secuenciados. Esto debido a evidencia reciente que sugiere que dichas variantes en VPH 51 sí existen, y que dependiendo de las variantes utilizadas en la secuenciación y caracterización del modelo *in silico* pueden determinar si dichas variantes son relevantes o no: un caso a mencionar es el estudio de la región capitalina y metropolitana en México por Ortiz-Gutiérrez et al., que determinó la presencia constante de las mutaciones I21L, E26R, I52L, V71G y F72I en L1 de VPH 51 en pacientes co-infectadas con VPH y VIH (77) – dichas mutaciones no fueron detectadas en este estudio, y una caracterización más profunda podría determinar si estas mutaciones fueron excluidas por el método de secuenciación aquí empleado, o si están presentes en una región específica de la República Mexicana. Ejemplos de cómo podría llevarse este método podrían incluir el uso de técnicas específicas como TAME-seq, que tiene la capacidad de detectar la presencia de variantes de baja prevalencia mediante enriquecimiento post-secuenciación (78), combinado con análisis *in silico* de genoma completo como fue empleado por Liu et al. para determinar alteraciones en la filogenia de varios genotipos de VPH en poblaciones no caucásicas (79), que generalmente están excluidas de estos estudios en el hemisferio occidental. Un ejemplo integrado de este uso se encuentra en Mirabello et al., que realizaron estudios GWAS asociados al diagnóstico histopatológico en VPH 16 sobre 3200 mujeres, y que permitieron encontrar variantes no reportadas de los sublinajes A1/A2 sobre dicha población (80).

Una extensión de los estudios de variabilidad ha extendido el paradigma más allá de la presencia de variantes en el genoma, y lo ha extendido también al epigenoma y los mecanismos de silenciamiento génico del hospedero. Así, por ejemplo, se ha

conocido que ciertos miRNAs también presentan modificaciones entre pacientes y poblaciones que pueden estar asociadas a una progresión más lenta o acelerada de la enfermedad, incluso a cáncer (81). También la metilación del genoma viral ha sido estudiada, y se han encontrado variantes en dichos patrones, en particular sobre L1 y LCR de genotipos como VPH 16, VPH 18 e incluso VPH 51, de significancia clínica por potenciales alteraciones de la replicación del genoma viral y expresión de genes de la cápside (82). Con esta información también ha sido posible generar modelos de proteínas E6, E7 y L1 mediante predicción de dominios (83), que pueden ser útiles como ya se ha mencionado anteriormente en el diseño de vacunas mediante VLPs y que pueden llegar a fase clínica más rápidamente, con el fin de reducir la tasa de infección actual por VPH en poblaciones vulnerables que no cuenten con acceso facilitado a la vacunación actual (84).



CONCLUSIONES

La variabilidad genética encontrada en L1 y L2 de VPH51 en la región centro de México es similar a la encontrada en otras partes del mundo, con una estabilidad relativa tanto en L1 como en L2, salvo en el caso de la variante EIPAP del aminoácido 95 a 99 de L2, la cual es la variante de mayor potencial detectada en ambos genes.

La variabilidad genética encontrada en L1 y L2 de VPH51 es menor a la variabilidad crítica sugerente de la aparición constante de nuevas variantes en la población femenina del centro de México.

La mayor parte de los polimorfismos detectados tanto en L1 como en L2 de VPH51 son silenciosos. Los que llegan a prevalecer en las secuencias consenso calculadas son tanto únicos a cada grupo de lesión, como algunos encontrados en dos o más grupos de manera simultánea.

Las mutaciones V277G de L1, y la variante EPIAP 95-99, Ins[G]179 y D180H de L2 en VPH51 son comunes a todos los grupos de lesión, aunque no generan variaciones estructurales significativas. En el caso de L1 en IC, existe una variación en ensamblaje debido a la presencia de múltiples inserciones en el extremo C-terminal del modelo.

EPIAP 95-99 en L2 no se encuentra reportada en la literatura, por lo que se considera conveniente analizar en el futuro esta secuencia con potencial inmunogénico que pudiera contribuir a la posibilidad de obtener vacunas dirigidas contra VPH51 más efectivas en la población del centro de México.

REFERENCIAS

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018 Nov 1;68(6):394–424.
2. Hernández-Hernández D, Apresa-García T, Patlán-Pérez R. Panorama epidemiológico del cáncer cervicouterino. *Rev Med Inst Mex Seguro Soc*. 2015;53(Supl 1-2):S154-161.
3. Burk RD, Harari A, Chen Z. Human papillomavirus genome variants. *Virology*. 2013 Oct 1;445(1–2):232–43.
4. Chen Z, Schiffman M, Herrero R, DeSalle R, Anastos K, Segondy M, et al. Classification and evolution of human papillomavirus genome variants: Alpha-5 (HPV26, 51, 69, 82), Alpha-6 (HPV30, 53, 56, 66), Alpha-11 (HPV34, 73), Alpha-13 (HPV54) and Alpha-3 (HPV61). *Virology*. 2018 Mar 1;516:86–101.
5. Piana A, Sotgiu G, Cocuzza C, Musumeci R, Marras V, Pischedda S, et al. High HPV-51 Prevalence in Invasive Cervical Cancers: Results of a Pre-Immunization Survey in North Sardinia, Italy. *PLoS One*. 2013 May 22;8(5).
6. Rafael Gutiérrez Campos, Angélica Malacara Rosas, Elvia Gutiérrez Santillán, Mireya Delgado Gutiérrez, Rusland Enrique Torres Orozco, Elí Daniel García Martínez, Luis Fernando Torres Bernal ARC. Unusual Prevalence of High-Risk Genotypes of Human Papillomavirus in a Group of Women With Neoplastic Lesions and Cervical Cancer From Central Mexico. *PLoS One*. 2019;14(4):e0215222.
7. Godi A, Epifano I, Bissett SL, Dell'Anna T, Piana A, Cocuzza C, et al. Amino acid motifs in both the major and minor capsid proteins of HPV51 impact antigenicity and infectivity. *J Gen Virol*. 2015 Jul 1;96(7):1842–9.
8. Bernard HU, Calleja-Macias IE, Dunn ST. Genome variation of human papillomavirus types: Phylogenetic and medical implications. *Int J Cancer*. 2006 Mar 1;118(5):1071–6.
9. Schiller JT, Müller M. Next generation prophylactic human papillomavirus vaccines. *Lancet Oncol*. 2015;16(5):e217–25.
10. de Araujo Souza PS, Sichero L, Maciag PC. HPV variants and HLA polymorphisms: The role of variability on the risk of cervical cancer. *Futur Oncol*. 2009;5(3):359–70.
11. Xu J, Tan L, Wang T, Cui F, Ding X, Wan Q, et al. Genetic variability of human papillomavirus type 51 E6, E7, L1 and L2 genes in Southwest China. *Gene*. 2019 Mar 30;690:99–112.
12. Gallegos-Bolaños J, Rivera-Domínguez JA, Presno-Bernal JM, Cervantes-Villagrana RD. High prevalence of co-infection between human papillomavirus

- (HPV) 51 and 52 in Mexican population. *BMC Cancer*. 2017 Aug 8;17(1):531.
13. Siqueira JD, Alves BM, Prellwitz IM, Furtado C, Meyrelles ÂR, Machado ES, et al. Identification of novel human papillomavirus lineages and sublineages in HIV/HPV-coinfected pregnant women by next-generation sequencing. *Virology*. 2016;493:202–8.
 14. Benedet JL, Bender H, Jones H, Ngan HY, Pecorelli S. FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. *Int J Gynecol Obstet*. 2000 Aug 1;70(2):209–62.
 15. Doorbar J, Griffin H. Refining our understanding of cervical neoplasia and its cellular origins. Vol. 7, *Papillomavirus Research*. Elsevier B.V.; 2019. p. 176–9.
 16. Quddus MR, Manna P, Sung CJ, Kerley S, Steinhoff MM, Lawrence WD. Prevalence, distribution, and viral burden of all 15 high-risk human papillomavirus types in adenosquamous carcinoma of the uterine cervix: A multiplex real-time polymerase chain reaction-based study. *Hum Pathol*. 2014 Feb;45(2):303–9.
 17. Reich O, Regauer S, McCluggage WG, Bergeron C, Redman C. Defining the Cervical Transformation Zone and Squamocolumnar Junction. *Int J Gynecol Pathol*. 2017 Nov;36(6):517–22.
 18. Martens JE, Smedts FMM, Ploeger D, Helmerhorst TJM, Ramaekers FCS, Arends JW, et al. Distribution Pattern and Marker Profile Show Two Subpopulations of Reserve Cells in the Endocervical Canal. *Int J Gynecol Pathol*. 2009 Jul;28(4):381–8.
 19. Herfs M, Yamamoto Y, Laury A, Wang X, Nucci MR, McLaughlin-Drubin ME, et al. A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer. *Proc Natl Acad Sci U S A*. 2012 Jun 26;109(26):10516–21.
 20. Chumduri C, Gurumurthy RK, Berger H, Koster S, Brinkmann V, Klemm U, et al. Transition of Wnt signaling microenvironment delineates the squamocolumnar junction and emergence of squamous metaplasia of the cervix. *bioRxiv*. 2018 Oct 16;443770.
 21. de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int J Cancer*. 2017 Aug 15;141(4):664–70.
 22. Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, et al. The 2001 Bethesda System: Terminology for reporting results of cervical cytology. *J Am Med Assoc*. 2002 Apr 24;287(16):2114–9.
 23. Nayar R, Wilbur DC. The Pap Test and Bethesda 2014. *Acta Cytol*. 2015 May 19;59(2):121–32.
 24. Nayar R, Wilbur DC. The bethesda system for reporting cervical cytology:

Definitions, criteria, and explanatory notes. The Bethesda System for Reporting Cervical Cytology: Definitions, Criteria, and Explanatory Notes. Springer International Publishing; 2015. 1–321 p.

25. Nayar R, Wilbur DC. The Bethesda System for Reporting Cervical Cytology: A Historical Perspective. *Acta Cytol.* 2017 Aug 1;61(4–5):359–72.
26. Cooper D, Hoffman M, Carrara H, Rosenberg L, Kelly J, Stander I, et al. Determinants of sexual activity and its relation to cervical cancer risk among South African Women. *BMC Public Health.* 2007;7:341.
27. Kjellberg L, Hallmans G, Ahren AM, Johansson R, Bergman F, Wadell G, et al. Smoking, diet, pregnancy and oral contraceptive use as risk factors for cervical intra-epithelial neoplasia in relation to human papillomavirus infection. *Br J Cancer.* 2000 Apr;82(7):1332–8.
28. Fonseca-Moutinho JA. Smoking and Cervical Cancer. *ISRN Obstet Gynecol.* 2011;2011.
29. Dugué PA, Rebolj M, Garred P, Lynge E. Immunosuppression and risk of cervical cancer. Vol. 13, *Expert Review of Anticancer Therapy.* Taylor & Francis; 2013. p. 29–42.
30. Stone KM, Zaidi A, Rosero-Bixby L, Oberle MW, Reynolds G, Larsen S, et al. Sexual behavior, sexually transmitted diseases, and risk of cervical cancer. *Epidemiology.* 1995;6(4):409–14.
31. Zhu H, Shen Z, Luo H, Zhang W, Zhu X. Chlamydia trachomatis infection-associated risk of cervical cancer: A meta-analysis. *Med (United States).* 2016;95(13):e3077.
32. Gierisch JM, Coeytaux RR, Urrutia RP, Havrilesky LJ, Moorman PG, Lowery WJ, et al. Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: A systematic review. Vol. 22, *Cancer Epidemiology Biomarkers and Prevention.* American Association for Cancer Research; 2013. p. 1931–43.
33. Benard VB, Johnson CJ, Thompson TD, Roland KB, Lai SM, Cokkinides V, et al. Examining the association between socioeconomic status and potential human papillomavirus-associated cancers. *Cancer.* 2008 Nov 3;113(S10):2910–8.
34. Shields TS, Falk RT, Herrero R, Schiffman M, Weiss NS, Bratti C, et al. A case-control study of endogenous hormones and cervical cancer. *Br J Cancer.* 2004 Jan 12;90(1):146–52.
35. Martínez-Nava GA, Fernández-Niño JA, Madrid-Marina V, Torres-Poveda K. Cervical cancer genetic susceptibility: A systematic review and meta-analyses of recent evidence. *PLoS One.* 2016 Jul 1;11(7).
36. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide

analysis. *Lancet Glob Heal*. 2020 Feb 1;8(2):e191–203.

37. Bosetti C, Rodríguez T, Chatenoud L, Bertuccio P, Levi F, Negri E, et al. Trends in cancer mortality in Mexico, 1981–2007. *Eur J Cancer Prev*. 2011 Sep;20(5):355–63.
38. Bosch FX, Lorincz A, Muñoz N, Meijer CJLM, Shah K V. The causal relation between human papillomavirus and cervical cancer. Vol. 55, *Journal of Clinical Pathology*. BMJ Publishing Group; 2002. p. 244–65.
39. Nicolás-Párraga S, Gandini C, Pimenoff VN, Alemany L, de Sanjosé S, Xavier Bosch F, et al. HPV16 variants distribution in invasive cancers of the cervix, vulva, vagina, penis, and anus. *Cancer Med*. 2016 Oct 1;5(10):2909–19.
40. Doorbar J, Quint W, Banks L, Bravo IG, Stoler M, Broker TR, et al. The biology and life-cycle of human papillomaviruses. Vol. 30, *Vaccine*. Elsevier Ltd; 2012.
41. de Sanjose S, Quint WGV, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol*. 2010 Nov;11(11):1048–56.
42. Schiller JT, Day PM, Kines RC. Current understanding of the mechanism of HPV infection. Vol. 118, *Gynecologic Oncology*. Elsevier; 2010. p. S12–7.
43. Saeed M, Alshammari FD, Alam MJ, Sarim KM, Ahmad K, Hussain T, et al. A Synopsis on the Role of Human Papilloma Virus Infection in Cervical Cancer. *Curr Drug Metab*. 2018 Aug 29;19(9):798–805.
44. Narisawa-Saito M, Kiyono T. Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: Roles of E6 and E7 proteins. *Cancer Sci*. 2007 Oct;98(10):1505–11.
45. Cornet I, Gheit T, Iannacone MR, Vignat J, Sylla BS, Del Mistro A, et al. HPV16 genetic variation and the development of cervical cancer worldwide. *Br J Cancer*. 2013 Jan 15;108(1):240–4.
46. Sotlar K, Diemer D, Dethleffs A, Hack Y, Stubner A, Vollmer N, et al. Detection and typing of human papillomavirus by E6 nested multiplex PCR. *J Clin Microbiol*. 2004 Jul 1;42(7):3176–84.
47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990 Oct 5;215(3):403–10.
48. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019 Jul 1;47(W1):W636–41.
49. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003 Jul 1;31(13):3784–8.
50. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC,

et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem*. 2004 Oct 1;25(13):1605–12.

51. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018 Jun 1;35(6):1547–9.
52. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123(3).
53. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012 Dec 18;13(1):134.
54. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;
55. Lungu O, Crum CP, Silverstein S. Biologic properties and nucleotide sequence analysis of human papillomavirus type 51. *J Virol*. 1991 Aug;65(8):4216–25.
56. Sabet F, Mosavat A, Ahmadi Ghezeldasht S, Basharkhah S, Shamsian SAA, Abbasnia S, et al. Prevalence, genotypes and phylogenetic analysis of human papillomaviruses (HPV) in northeast Iran. *Int J Infect Dis*. 2021 Feb 1;103:480–8.
57. Molina-Pineda A, López-Cardona MG, Limón-Toledo LP, Cantón-Romero JC, Martínez-Silva MG, Ramos-Sánchez HV, et al. High frequency of HPV genotypes 59, 66, 52, 51, 39 and 56 in women from Western Mexico. *BMC Infect Dis*. 2020;20(1).
58. Hao S, Wang C, Liu S, He J, Jiang Y. HPV genotypic spectrum in Jilin province, China, where non-vaccine-covered HPV53 and 51 are prevalent, exhibits a bimodal agespecific pattern. *PLoS One*. 2020;15(3).
59. Giannella L, Giorgi Rossi P, Delli Carpini G, Di Giuseppe J, Bogani G, Gardella B, et al. Age-related distribution of uncommon HPV genotypes in cervical intraepithelial neoplasia grade 3. *Gynecol Oncol*. 2021;161(3).
60. Xi J, Chen J, Xu M, Yang H, Luo J, Pan Y, et al. Genetic variability and functional implication of the long control region in HPV-16 variants in Southwest China. *PLoS One*. 2017;12(8).
61. El-Aliani A, Alaoui MA El, Chaoui I, Ennaji MM, Attaleb M, Mzibri M El. Naturally occurring capsid protein variants L1 of human papillomavirus genotype 16 in Morocco. *Bioinformation*. 2017 Aug 31;13(8):241–8.
62. Rahmat F, Kuan JY, Hajiman Z, Shakrin NNSM, Roos NAC, Mustapa M, et al. Human Papillomavirus (HPV) Prevalence and Type Distribution in Urban Areas of Malaysia. *Asian Pac J Cancer Prev*. 2021 Sep 1;22(9):2969–75.
63. Zhu X, Wang Y, Lv Z, Su J. Prevalence and genotype distribution of high-risk

HPV infection among women in Beijing, China. *J Med Virol*. 2021 Aug 1;93(8):5103–9.

64. Baisley KJ, Andreasen A, Irani J, Nnko S, Changalucha J, Crucitti T, et al. HPV prevalence around the time of sexual debut in adolescent girls in Tanzania. *Sex Transm Infect*. 2020 May 1;96(3):211–9.
65. Jagu S, Kwak K, Schiller JT, Lowy DR, Kleanthous H, Kalnin K, et al. Phylogenetic considerations in designing a broadly protective multimeric L2 vaccine. *J Virol*. 2013 Jun;87(11):6127–36.
66. Yadav R, Zhai L, Kunda NK, Muttill P, Tumban E. Mixed Bacteriophage MS2-L2 VLPs Elicit Long-Lasting Protective Antibodies against HPV Pseudovirus 51. *Viruses*. 2021 Jun 1;13(6).
67. Pouyanfard S, Spagnoli G, Bulli L, Balz K, Yang F, Odenwald C, et al. Minor Capsid Protein L2 Polytope Induces Broad Protection against Oncogenic and Mucosal Human Papillomaviruses. *J Virol*. 2018 Feb 15;92(4).
68. Seitz H, Ribeiro-Müller L, Canali E, Bolchi A, Tommasino M, Ottonello S, et al. Robust In Vitro and In Vivo Neutralization against Multiple High-Risk HPV Types Induced by a Thermostable Thioredoxin-L2 Vaccine. *Cancer Prev Res (Phila)*. 2015 Oct 1;8(10):932–41.
69. Gupta B, Kumar A, Sridevi P. A Comprehensive in Silico Analysis for Identification of Immunotherapeutic Epitopes of HPV-18. *Int J Pept Res Ther*. 2021 Dec 1;27(4):2717–26.
70. Cavazzana-Calvo M, Payen E, Negre O, Wang G, Hehir K, Fusil F, et al. Transfusion independence and HMGA2 activation after gene therapy of human β -thalassaemia. *Nature*. 2010;467(7313):318–22.
71. Liu C, Du H, Wang C, Belinson JL, Yang B, Zhang W, et al. HPV L1 and P16 Expression in CIN1 to Predict Future CIN2. *Int J Gynecol Pathol*. 2017;36(3):281–8.
72. Garbuglia AR, Lapa D, Sias C, Capobianchi MR, Del Porto P. The Use of Both Therapeutic and Prophylactic Vaccines in the Therapy of Papillomavirus Disease. *Front Immunol*. 2020 Feb 18;11:188.
73. Dube Mandishora RS, Gjøtterud KS, Lagström S, Stray-Pedersen B, Duri K, Chin'ombe N, et al. Intra-host sequence variability in human papillomavirus. *Papillomavirus Res*. 2018;5.
74. Bowden SJ, Bodinier B, Kalliala I, Zuber V, Vuckovic D, Doulgeraki T, et al. Genetic variation in cervical preinvasive and invasive disease: a genome-wide association study. *Lancet Oncol*. 2021;22(4).
75. Shen-Gunther J, Cai H, Zhang H, Wang Y. Abundance of HPV L1 intra-genotype variants with capsid epitopic modifications found within low- And high-grade Pap smears with potential implications for vaccinology. *Front Genet*. 2019;10(MAY).

76. Gonzalez-Bosquet E, Gibert M, Serra M, Hernandez-Saborit A, Gonzalez-Fernandez A. Candidate HPV genotypes not included in the 9-valent vaccine for prevention of CIN 2-3. *Int J Gynecol Cancer*. 2020 Jul 1;30(7):954–8.
77. Ortiz-Gutiérrez F, Sánchez-Minutti L, Martínez-Herrera JF, Torres-Escobar ID, Pezzat-Said EB, Márquez-Domínguez L, et al. Identification of Genetic Variants of Human Papillomavirus in a Group of Mexican HIV/AIDS Patients and Their Possible Association with Cervical Cancer. *Polish J Microbiol*. 2021;70(4).
78. Lagström S, Umu SU, Lepistö M, Ellonen P, Meisal R, Christiansen IK, et al. TaME-seq: An efficient sequencing approach for characterisation of HPV genomic variability and chromosomal integration. *Sci Rep*. 2019;9(1).
79. González-Hernández LA, Flores-Miramontes MG, Aguilar-Lemarroy A, Quintanilla-Peña KS, Martin-Amaya-Barajas FL, Ramos-Solano M, et al. HPV genotypes detected by linear array and next-generation sequencing in anal samples from HIV positive men who have sex with men in Mexico. *Arch Virol*. 2018;163(4).
80. Mirabello L, Yeager M, Cullen M, Boland JF, Chen Z, Wentzensen N, et al. HPV16 Sublineage Associations with Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women. *J Natl Cancer Inst*. 2016;108(9).
81. Huang SQ, Zhou ZX, Zheng SL, Liu DD, Ye XH, Zeng CL, et al. Association of variants of miRNA processing genes with cervical precancerous lesion risk in a southern Chinese population. *Biosci Rep*. 2018;38(3).
82. Simanaviciene V, Pependikyte V, Gudleviciene Z, Zvirbliene A. Different DNA methylation pattern of HPV16, HPV18 and HPV51 genomes in asymptomatic HPV infection as compared to cervical neoplasia. *Virology*. 2015 Oct 1;484:227–33.
83. Ye M, Li S, Luo P, Tang X, Gong Q, Mei B. Genetic variation of E6, E7, and L1 genes of human papillomavirus 51 from central China. *J Med Virol*. 2022;94(6).
84. Arbyn M, Xu L, Simoens C, Martin-Hirsch PPL. Prophylactic vaccination against human papillomaviruses to prevent cervical cancer and its precursors. *Cochrane database Syst Rev*. 2018 May 9;5(5).

ANEXO 1: CARTA DE CONSENTIMIENTO INFORMADO

CARTA DE CONSENTIMIENTO INFORMADO PARA PARTICIPACIÓN EN PROYECTOS DE INVESTIGACIÓN CLÍNICA.

Lugar: Universidad Autónoma de Aguascalientes, Laboratorio de Ingeniería Genética

Fecha: _____

Yo, _____, por medio de la presente autorizo mi participación en el proyecto de investigación titulado: **Variabilidad genética en los genes L1 y L2 del Virus del Papiloma Humano genotipo 51 en pacientes con displasia cervical y cáncer cervicouterino en el centro de México**, registrado ante el Comité Estatal de Investigación en Salud del ISSEA con el número _____.

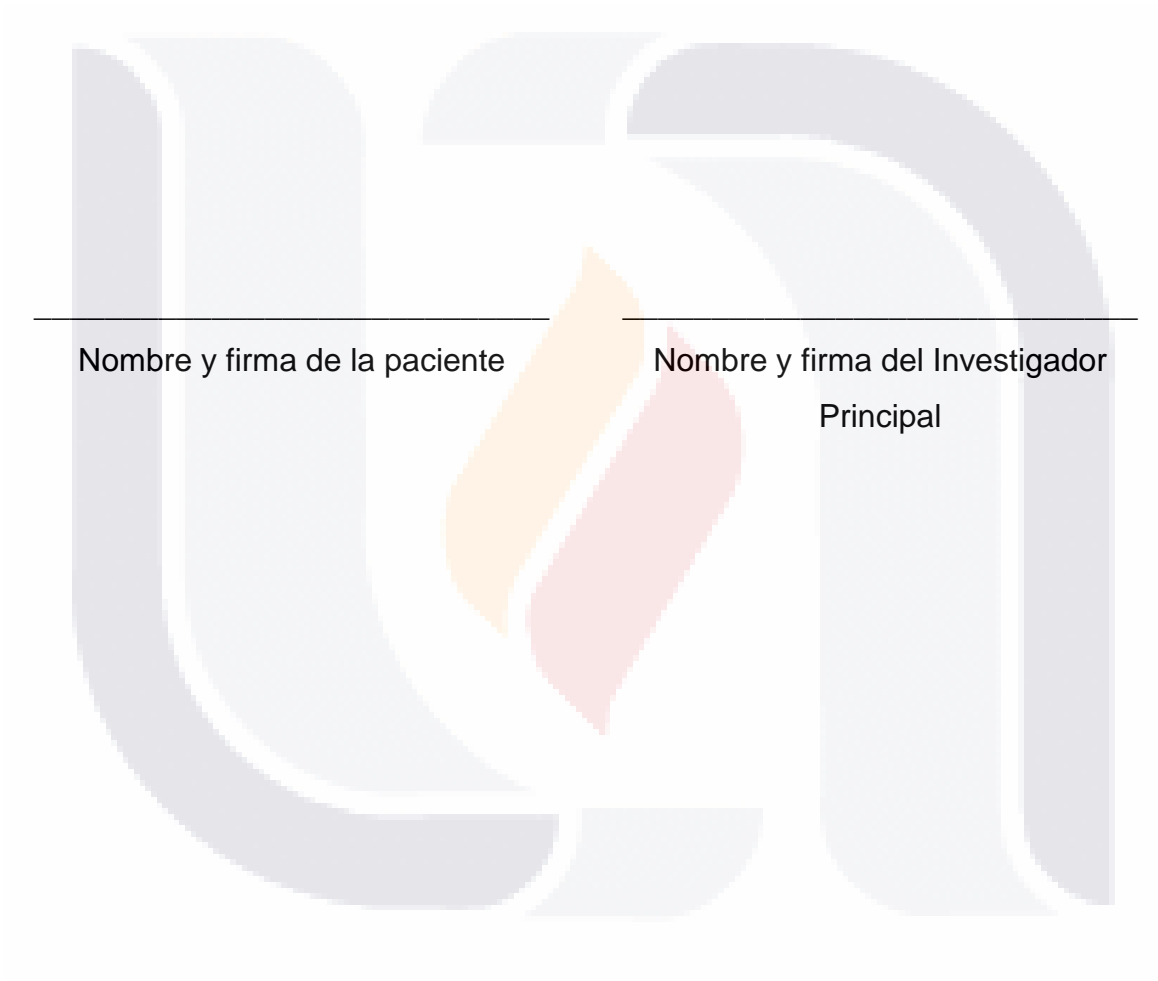
Se me ha explicado que mi participación consistirá en **la toma de muestra de raspado cervical para la determinación de la presencia de VPH 51 y el análisis de la secuencia de sus genes L1 y L2**. Declaro que se me ha informado ampliamente sobre los posibles riesgos, inconvenientes y molestias derivados de mi participación en el estudio, que son los siguientes: **molestias, calambres, retorcijones, dolor o leve sangrado**. El Investigador Principal ha respondido cualquier pregunta y aclarado cualquier duda que le he planteado acerca de los procedimientos que se llevarán a cabo, los resultados que deriven del análisis de la muestra que proporciono, el impacto de la investigación o cualquier otro asunto relacionado con la misma.

Entiendo además que **conservo el derecho de no continuar mi participación en el estudio en cualquier momento en que lo considere conveniente**, sin que esto conlleve penalización alguna por parte del equipo de investigación o la Unidad Médica.

El Investigador Principal me ha dado garantías de que respetará mi identidad en las presentaciones o publicaciones que deriven de este estudio y de que los datos

relacionados con su privacidad serán manejados en forma confidencial. También se ha comprometido a proporcionarme la información actualizada que se obtenga durante el estudio, aunque esta pudiera hacerme cambiar de parecer respecto a mi permanencia en el mismo.

El Investigador me ha informado que mi muestra no podrá ser utilizada para ningún otro fin al que se ha establecido.



ANEXO 2: CUESTIONARIO ADJUNTO A LA TOMA DE MUESTRA

Fecha: _____ No. de folio: _____

EL PRESENTE CUESTIONARIO ES TOTALMENTE CONFIDENCIAL; EN ÉL NO APARECE SU NOMBRE, DOMICILIO, NI NINGÚN OTRO DATO PERSONAL, POR LO QUE LE ROGAMOS QUE LOS DATOS QUE APORTE SEAN ABSOLUTAMENTE VERÍDICOS.

a) INFORMACIÓN A SER PROPORCIONADA POR EL MÉDICO TRATANTE

Diagnóstico e información clínica:

b) INFORMACIÓN A SER PROPORCIONADA POR LA PACIENTE

1. Edad: _____
2. Escolaridad: _____
3. ¿Cuántos hijos tiene? _____
4. ¿A qué edad tuvo su primer hijo? _____
5. ¿A qué edad inició su vida sexual? _____
6. ¿Cuántos compañeros sexuales ha tenido? _____
7. ¿Siempre ha utilizado protección (condón) en sus relaciones sexuales?

SÍ	NO
----	----

De no ser así, ¿con cuántas parejas ha utilizado condón? _____
8. ¿Con qué frecuencia se realiza su Papanicolaou? _____

9. ¿Padece de infecciones vaginales frecuentes (diagnosticadas por su médico)?

SÍ NO

10. ¿Toma usted anticonceptivos?

SÍ NO

Si su respuesta es afirmativa, ¿durante cuántos años los ha tomado o los tomó?

11. ¿Fuma o fumó usted?

SÍ NO

Si su respuesta es afirmativa, ¿durante cuánto tiempo fumó o ha fumado?

¿Cuántos cigarrillos al día fuma o fumó en promedio?

12. ¿Ha habido casos de cáncer cervicouterino en su familia?

SÍ NO

Si su respuesta es afirmativa, ¿quién lo padeció? _____

13. ¿Ha habido casos de otro(s) tipo(s) de cáncer en su familia?

SÍ NO

14. Si su respuesta es afirmativa, ¿quién lo padeció? _____

15. Considera que, de acuerdo a su situación socioeconómica actual, pertenece a clase:

ALTA

MEDIA

BAJA

Nombre y firma del médico tratante: _____

ANEXO 3: PROTOCOLO DE EXTRACCIÓN DE ADN DE MUESTRAS DE HISOPADO CERVICAL

(Adaptado del inserto contenido con el Blood-Animal-Plant DNA Preparation Kit (Jena Bioscience, Jena, Germany)).

Componentes del Kit

(50 preps; todo se preserva a temperatura ambiente excepto donde se indique).

- Blood Lysis Buffer (55 mL) **[no utilizado en esta modalidad del protocolo]**.
- Lysis Buffer (16 mL).
- RNasa A (7.5 mg).
 - Resuspendida a 50 mg/mL con H₂O gbm y preservada a -20 °C.
- Proteinasa K (5 mg).
 - Resuspendida a 10 mg/mL con H₂O gbm y preservada a -20 °C.
- Binding Buffer (16 mL).
- Activation Buffer (6 mL).
- Washing Buffer (12 mL).
 - Se añaden 48 mL de etanol absoluto antes de su uso (volumen final = 60 mL).
- Elution Buffer (5 mL).
- Columnas y tubos colectores de 2 mL (50 cada uno).

Material adicional

- Termoblock o baño de calor a 60 °C.
- Baño de hielo.
- Vórtex.
- Microcentrífuga (min. 10 000 g).
- Tubos eppendorf de 1.5 o 2 mL, estériles.

Preparación de ADN a partir de células animales

A. Lisis celular.

1. A la muestra de tejido homogenizado o paquete celular, añadir 300 μ L de Lysis Buffer y 2 μ L de RNasa A. Agitar vigorosamente con vórtex por 30-60 s.
2. Añadir 8 μ L de Proteinasa K y mezclar por pipeteo.
3. Incubar a 60 °C durante 20 min. Retirar del termoblock y dejar enfriar por 5 min.
4. Añadir 300 μ L de Binding Buffer y agitar brevemente con vórtex.
5. Incubar el tubo en baño de hielo por 5 min.
6. Centrifugar a 10 000 g por 5 min.

B. Activación de la columna.

1. Montar una columna en un tubo colector de 2 mL.
2. Añadir 100 μ L de Activation Buffer en la columna.
3. Centrifugar a 10 000 g por 30 s. Descartar el flujo del tubo y montar.
4. Utilizar inmediatamente la columna para la siguiente etapa.

C. Carga de la columna.

1. Tomar el sobrenadante del tubo centrifugado en la etapa A, paso 6, y colocarlo directamente en la columna.
2. Centrifugar a 10 000 g por 1 min.
3. Descartar el flujo del tubo colector.

D. Lavado primario.

1. Añadir 500 μ L de Washing Buffer en la columna.
2. Centrifugar a 10 000 g por 1 min.
3. Descartar el flujo del tubo colector.

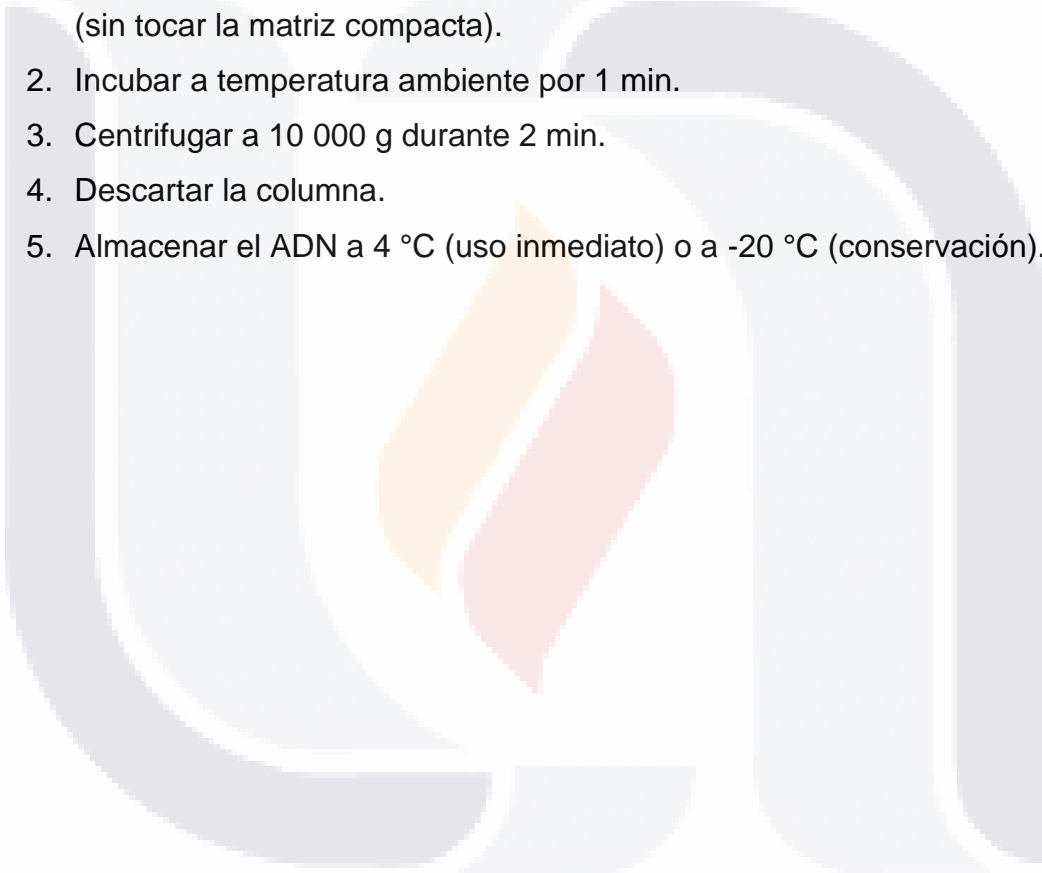
E. Lavado secundario.

1. Añadir 500 μ L de Washing Buffer en la columna.
2. Centrifugar a 10 000 g por 1 min.

3. Descartar el flujo del tubo colector.
4. Centrifugar una vez más a 10 000 g por 1 min para remover residuos de Washing Buffer de la columna.
5. Descartar el flujo y el tubo colector, y montar la columna en un tubo eppendorf limpio y rotulado de 1.5 o 2 mL (tubo de elución).

F. Elución del ADN.

1. Añadir 40-50 μ L de Elution Buffer directamente en el centro de la columna (sin tocar la matriz compacta).
2. Incubar a temperatura ambiente por 1 min.
3. Centrifugar a 10 000 g durante 2 min.
4. Descartar la columna.
5. Almacenar el ADN a 4 °C (uso inmediato) o a -20 °C (conservación).



ANEXO 4: SECUENCIAS DE OLIGONUCLEÓTIDOS UTILIZADAS EN LAS REACCIONES DE PCR

A. Oligonucleótidos consenso E6/E7

Oligo	Amplicón (pb)	Secuencia (5'-3')
E7AS	630	TCATCCTCCTCCTCTGAG
LCRS		AAGGGAGTAACCGAAACGGT

B. Oligonucleótidos para la genotipificación de VPH, reacciones multiplex

Coctel	Genotipo de VPH	Amplicón (pb)	Secuencia (5'-3')	
I	16	457	CAC AGT TAT GCA CAG AGC TGC	
			CAT ATA TTC ATG CAA TGT AGG TGT A	
	18	322	CAC TTC ACT GCA AGA CAT AGA	
			GTT GTG AAA TCG TCG TTT TTC A	
	31	263	GAA ATT GCA TGA ACT AAG CTC G	
			CAC ATA TAC CTT TGT TTG TCA A	
	59	215	CAA AGG GGA ACT GCA AGA AAG	
			TAT AAC AGC GTA TCA GCA GC	
	45	151	GTG GAA AAG TGC ATT ACA GG	
			ACC TCT GTG CGT TCC AAT GT	
	II	33	398	ACT ATA CAC AAC ATT GAA CTA
				GTT TTT ACA CGT CAC AGT GCA
6/11		334	TGC AAG AAT GCA CTG ACC AC	

Coctel	Genotipo de VPH	Amplicón (pb)	Secuencia (5'-3')
	58	274	TGC ATG TTG TCC AGC AGT GT
			GTA AAG TGT GCT TAC GAT TGC
			GTT GTT ACA GGT TAC ACT TGT
	52	229	TAA GGC TGC AGT GTG TGC AG
			CTA ATA GTT ATT TCA CTT AAT GGT
	56	181	GTG TGC AGA GTA TGT TTA TTG
TTT CTG TCA CAA TGC AAT TGC			
III	35	358	CAA CGA GGT AGA AGA AAG CAT C
			CCG ACC TGT CCA CCG TCC ACC G
	42	277	CCC AAA GTA GTG GTC CCA GTT A
			GAT CTT TCG TAG TGT CGC AGT G
	43	219	GCA TAA TGT CTG CAC GTA GCT G
			CAT GAA ACT GTA GAC AGG CCA AG
44	163	TAA ACA GTT ATA TGT AGT GTA CCG	
		TAT CAG CAC GTC CAG AAT TGA C	
IV	68	333	GCA GAA GGC AAC TAC AAC GG
			GTT TAC TGG TCC AGC AGT GG
	39	280	GAC GAC CAC TAC AGC AAA CC
			TTA TGA AAT CTT CGT TTG CT
	51	223	GAG TAT AGA CGT TAT AGC AGG
			TTT CGT TAC GTT GTC GTG TAC G
66	172	TTC AGT GTA TGG GGC AAC AT	

Coctel	Genotipo de VPH	Amplicón (pb)	Secuencia (5'-3')
			AAA CAT GAC CCG GTC CAT GC

C. Oligonucleótidos para la amplificación específica de L1 y L2 de VPH 51

Gen	Oligo	Amplicón (pb)	Secuencia (5'-3')
L1	HPV51L1-1F	915	GGT ATT GCC CAC ATC TCC TA
	HPV51L1-1R		TTT GCT CCC TGC GTA AAT GA
	HPV51L1-2F	1103	GGA TTT CGC TGC CCT AC
	HPV51L1-2R		TAC CCA CAT ACA CCC GTA CA
L2	HPV51L2-1F	866	TGC AAC ATC CCA ATT AAC TAC
	HPV51L2-1R		TGT GGA TGG CTT ACT AAT AAA
	HPV51L2-2F	789	TAG CTG CTC CCC GCT TGT ATA
	HPV51L2-2R		CAC AAT TCG AGA CAC AGG TGC