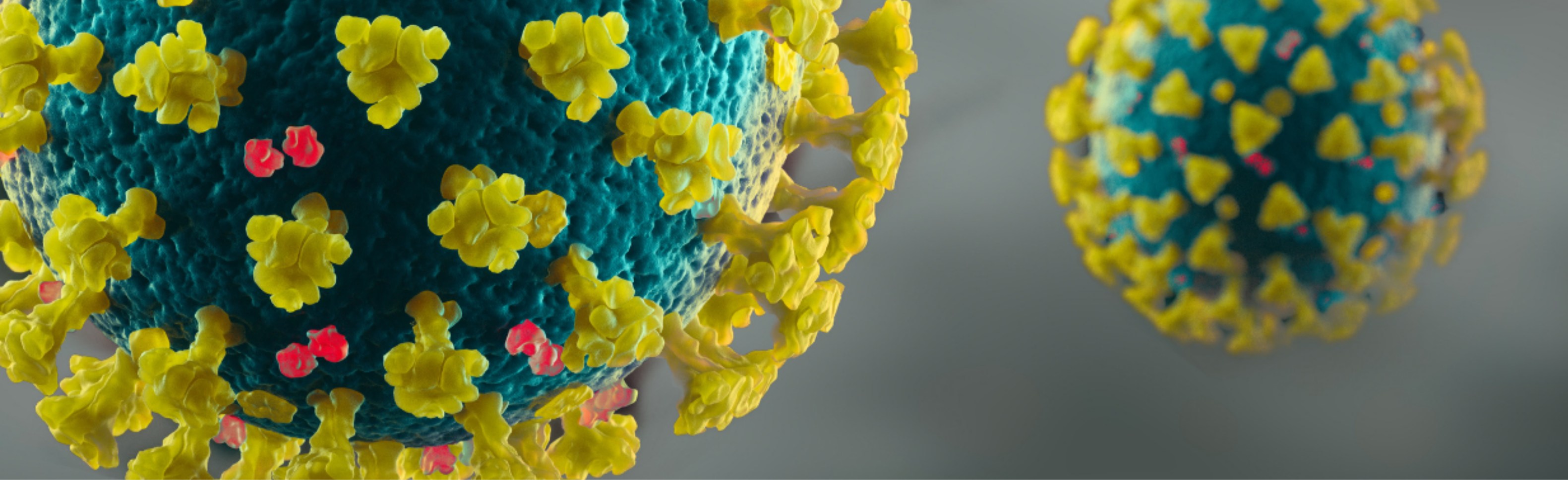

Vaccine Discovery and Development: Lessons from COVID-19

Free eBook

Emerging infectious diseases (EIDs) can evolve into a global healthcare crisis or pandemic. Scientists have previously required years to develop vaccines or therapeutics. The use of high throughput technology can greatly broaden the insights collected during discovery, augment efficiency and safety of handling EIDs, and shorten timelines.

Download this publication for an overview of many lessons learned in virology, immunology, and vaccine research during COVID-19 vaccine development.

Download here

Received: 28 July 2021 | Accepted: 22 September 2021

DOI: 10.1002/jmv.27359

RESEARCH ARTICLE

Genetic variation in the E6 and E7 genes of human papillomavirus type 16 in northeastern Argentina

M. Elina Totaro[1] | Juan A. Gili[2,3] | D. Javier Liotta[1,4] | Theodore G. Schurr[5] | Maria A. Picconi[6] | Inés Badano[1,3]

[1]Laboratorio de Biología Molecular Aplicada, Facultad de Ciencias Exactas, Quimicas y Naturales, Universidad Nacional de Misiones, Posadas, Misiones, Argentina

[2]Laboratorio de Epidemiología Genética, Dirección de Investigación CEMIC-CONICET, Buenos Aires, Argentina

[3]Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

[4]Instituto Nacional de Medicina Tropical-ANLIS "Dr. Malbrán", Puerto Iguazú, Misiones, Argentina

[5]Laboratory of Molecular Anthropology, Department of Anthropology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[6]Servicio de Virus Oncogénicos, Laboratorio Nacional y Regional de Referencia de HPV, Instituto Nacional de Enfermedades Infecciosas-ANLIS "Dr. Malbrán", Buenos Aires, Buenos Aires, Argentina

Correspondence

Badano Inés, Laboratorio de Biología Molecular Aplicada, Facultad de Ciencias Exactas, Químicas y Naturales, Universidad Nacional de Misiones, Av Mariano Moreno No 1375, 1er Piso, Posadas (cp. 3300), Misiones, Argentina.
Email: inesbadano@gmail.com

Funding information

Agencia Nacional de Promoción Científica y Tecnológica; Fundación Florencio Fiorini; Sistema Nacional de Datos Genómicos; Consejo Interuniversitario Nacional; Codon Code Aligner Grants; Consejo Nacional de Investigaciones Científicas y Técnicas

Abstract

The province of Misiones is considered a region with a high mortality rate due to cervical cancer (CC). To gain insight into this problem, we explored the association between genetic variation in the E6 and E7 oncogenes of HPV16 and the risk of CC. We studied 160 women with cytological diagnoses of negative for intraepithelial lesion or malignity, low-grade squamous intraepithelial lesion, and high-grade squamous intraepithelial lesion/CC and a positive test for HPV16 infection. The genetic characterization of E6 and E7 genes was undertaken through PCR amplification and direct Sanger sequencing. Phylogenetic classification was conducted using Bayesian methods. To estimate the odds ratio (OR) for an association between genetic variants in the E6 and E7 genes and the risk of CC, we used ordinal logistic regression adjusted by age. The final data set comprised 112 samples. Diagnostic single-nucleotide polymorphisms (SNPs) and phylogenetic trees confirmed the presence of Lineage A (95.5%) and D (4.5%) in the samples. For the E6 gene, we identified eleven different sequences, with the most common ones being Lineage A E6 350G (58.9%) and E6 350T (37.5%). The E6 350G was associated with progression to HSIL/CC, with an OR of 19.41 (4.95–76.10). The E7 gene was more conserved than E6, probably due to the functional constraints of this small protein. Our results confirmed the association of the E6 350G SNP with a higher risk of developing CC. These data will contribute to understanding the biological bases of CC incidence in this region.

KEYWORDS

carcinogenesis, cervical cancer, HPV, risk factor, SNP

1 | INTRODUCTION

Cervical cancer (CC) is the third most common cancer among women worldwide, and ranks as the fourth most frequent cancer among women in Argentina, causing 2231 deaths every year.[1] Misiones Province is considered a region with a high mortality rate of CC (15/100 000 individuals) compared to the national average (7/100 000).[2] The primary risk factor identified in the development of this disease is persistent infection with high-risk human papillomavirus (HPV), particularly HPV type 16 (Family *Papillomaviridae*, genus *Alphapapillomavirus*, species A9).[3] During the last 10 years, we have studied risk factors for CC in this population. These early studies

 © 2021 Wiley Periodicals LLC

showed a high prevalence of HPV infection (>30%) in different populations across the province[2,4] with the most common type being HPV16 (9.8%).[2] Since 2011, the introduction of vaccination programs is contributing to a significant reduction in HPV16/18 infection (nearly 95%). However, the prevalence in young unvaccinated women (between 15 and 17 years old) showed 62% rate of HPV infection, with 21% being infected with HPV16/18.[5] Thus, the epidemiology of HPV16 infection and risk factors involved in the development of CC remains as an important issue to address in this population.

The association of HPV16 infection and the development of CC has been confirmed in several studies with an odds ratio (OR) ranging from 5.3 (1.8–15.8) to 24.2 (9.3–62.7).[2,6] Furthermore, the analysis of genetic variation in HPV16 at the noncoding LCR region indicated that different viral lineages were present in the region (93% A and 7% D), probably due to the multiethnic origin of this population.[3] Lineage D was also associated with high-grade lesions and cancer at an OR of 13.8 (1.6-117.0).[3] Aside this initial work, however, no other viral genes were investigated in relation to the development of CC in Misiones.

The HPV E6 gene is an oncogene of 450 base pairs (bp) that encodes a 151 amino acid protein involved in the human cell transformation.[7,8] Several single-nucleotide polymorphisms (SNPs) have been reported across its sequence,[9–11] suggesting that the gene is under diversifying selection.[12] Among them, the 350T→G (E6 350G) mutation is a non-synonymous substitution that results in a Valine to Leucine change in the 83rd amino acid of the oncoprotein 83L→V (L83V). *In vitro* studies have shown that this mutation can affect antigenic presentation and immunogenicity, increasing the capacity of the virus to induce cell transformation compared to the E6 350T allele.[13,14]

The E7 gene is an oncogene of 297 bp that encodes a 98 amino acid protein.[15] Studies in transgenic mice have shown that E7 is a more potent driver for CC than E6.[15] On the other hand, recent large epidemiological studies showed that genetic variation at this locus is very low in CC cases compared to controls.[16,17] This finding suggests that E7 sequence conservation is crucial for CC development.[16,17] Moreover, the limited repertoire of E7 proteins is consistent in different geographic locations and ethnic populations, and has features of purifying selection.[17]

Given this background information for these oncogenes, this study aimed to explore the potential association between HPV16 E6 and E7 variants with the risk of CC among urban women living in northeastern Argentina. The results of this analysis will enhance our understanding of the incidence of CC in this region of South America, and help clarify the role of HPV as a viral genetic risk factor in Argentinian populations.

2 | METHODS

2.1 | Study design and bioethics

This is a retrospective case-control study investigating genetic risk factors, viral infection, and CC. The study samples were selected from the LaBiMAp database (2004–17), Misiones (Argentina). The inclusion criteria were cytological diagnosis of negative for intraepithelial lesion or malignity (NILM), low-grade squamous intraepithelial lesion (L-SIL), high-grade squamous intraepithelial lesion (H-SIL), CC *in situ/invasive*, and a positive test for HPV16 infection. Age information was also available for each participant. Samples were selected without personal information to preserve patient anonymity. The Ethics Committee of the "Departamento de Docencia e Investigación, Comité de Bioética, Hospital Dr. Ramón Madariaga, Posadas, Misiones" approved the study design.

2.2 | Study population and biological samples

A total of 160 women were positive for HPV 16, including 56 controls (NILM) and 104 cases (48 L-SIL, and 56 H-SIL/CC). Biological samples consisted of total DNA extracted from cervical scrapings ($n = 134$) or biopsies ($n = 26$). These DNA extracts were obtained during 2004–2017 by using commercial kits (ADN PuriPrep-S Kit, K1205-250, Inbio HighWay®).

2.3 | E6 and E7 genetic characterization

The genetic characterization of the E6 and E7 genes in HPV16 samples was conducted through PCR amplification and direct sequencing of a single 749 bp fragment with primers E6/E7-Fw (ATGCACCAAAAGAGAACTGC) and E6/E7-Rv (TGGTTTCTGRGA ACAGATGG). We performed template amplification in a 50 µl reaction mix containing 100 ng of DNA in PCR buffer (Tris-HCL 20 mM pH 8.3, KCl 20 mM, $(NH_4)_2SO_4$ 10 mM, and $MgSO_4$ 2 mM), 1 µM of each primer, 200 µM of each dNTP, and 2.5U of EasyTaq DNA Polymerase (AP111-01, Transgenbiotech), using the following conditions: denaturation at 95°C for 5 min followed by 40 cycles of 94°C for 1 min, 61°C for 1 min and 72°C for 1 min 30 s, and then a final elongation at 72°C for 3 min. Positive amplicons were purified with the ADN PuriPrep-GP Kit (K1206-100 Inbio HighWay®) and then Sanger sequenced using the original primers through sequencing services (Macrogen Inc).

2.4 | Sequence analysis

The E6/E7 HPV16 sequences were read and analyzed using CodonCode aligner software v.3.0.1 (CodonCode Corporation). The readable sequences were unequivocally aligned and the positions of the SNPs were identified according to the Reference genome K02718 available at Papillomavirus episteme.[18] To identify a novel variant, BLAST programs were used.[19] The sequence data were used to estimate gene diversities with MEGA 6.[20] In addition, the coding regions were translated into amino acid sequences. To designate each variant, the following nomenclature was employed: the name of the gene and the nucleotide position (number) followed by the

current change, for example, E6 350G. For an amino acid change, the letter preceding the number refers to the reference amino acid and the letter after it refers to its substitution.

2.5 | Phylogenetic analysis

The phylogenetic analysis of E6 and E7 gene sequences was conducted using the Bayesian method implemented in BEAST v1.7.2.[21] The priors were as follows: HKY; a relaxed (uncorrelated lognormal) molecular clock; a constant demographic growth[22]; and a substitution rate of 2.39×10^{-8} s/s/y (substitutions per site per year).[23] The Monte Carlo Markov Chain was run for 10 million generations, sampling every 1000th generation to achieve an Effective Sample Size (ESS) > 200. All BEAST run logs were analyzed with the TRACER program v1.5.[24] A maximum clade credibility tree was constructed with the Tree Annotator tool[21] after discarding 2% of the sampling. The tree was then visualized with FigTree v1.4.0 software.[25]

2.6 | GenBank accession number

The sequences described in this study were deposited in GenBank under the following accession numbers: MN053339–MN053388; MN160917–MN061890.

2.7 | Statistical analysis

Ordinal logistic regression adjusted by age was used to estimate OR and 95% confidence intervals (CIs) of an association between genetic variants in the E6 and E7 genes and the risk for CC. This estimate has the advantage of producing a unified OR value, taking under consideration the three progressive stages of Pap cytology (namely, progressive model: NILM→LSIL→ HSIL+), but also generates separate ORs for each individual state (cut H-SIL+/NILM and cut LSIL/NILM). The denotation "H-SIL+" was used to refer to H-SIL+CC cases clustered together. H-SIL was used as a threshold state because it is a landmark for treatment in clinical practice. All calculations were made using Stata 14.0 (Stata Corp LLC). Genetic diversity was estimated by using DnaSP 5.1.[26]

3 | RESULTS

3.1 | Study population

A total of 48 samples (30%) could not be successfully amplified or sequenced and, thus, were excluded from the study. These failures were most likely attributable to low-quality DNA obtained from paraffin-embedded biopsies ($n = 11$) or for being old cytological samples collected more than a decade ago (2005–2006; $n = 37$). Subsequently, 112 samples that represented 29 NILM, 48 L-SIL, and 35 H-SIL+ individuals, respectively, were analyzed. The age distribution for each group was as follows: NILM = mean age 28.6 years ($SD = 7.8$; range 19–52; median 27); L-SIL = mean age 31 years ($SD = 9.9$; range 18–54; median 29); H-SIL+ = mean age 36.6 years ($SD = 11.7$; range 18–61; median 33); and total population = mean age 31.5 years ($SD = 10.7$; range 18-61; median 29). The age distribution was statistically significant across study groups ($p = 0.005$), and therefore the association analysis was adjusted by age.

3.2 | Genetic characterization of E6 and E7 oncogenes

The sequence data for the E6 and E7 genes are shown in Table 1. Eleven different sequences for E6 were identified, with only six of them having non-synonymous changes, which occurred at amino acids L83V, R10G, D25N, D25H, and F69L. The most common variant was E6 350G or L83V (58.9%, 66/112), which was found alone (54.4%) or in combination with the amino acid signature for Lineage D (Q14H, F78Y, and L83V) (4.5%). The E7 gene was less variable, exhibiting five different sequences but only two non-synonymous variants, H51N (4.2%) and G88R (1.0%), both being previously reported. Using these data, the genetic diversity for E7 was estimated to be lower than that for E6.

3.3 | Phylogenetic analysis

The phylogeny for E6 and E7 concatenated DNA for 82 sequences is shown in Figure 1. The tree topology revealed that the Misiones samples belonged to three main branches. These included lineage A E6 350G (54.4%) and E6 350T (41.1%), and lineage D (4.5%). E6 350G SNP as a recurrent mutation, having occurred in parallel in these different lineages.

3.4 | Distribution of HPV variants according to the lesion grade

Table 2 shows the frequency of E6 350G variants according to the Pap cytology examination and the association analysis. The OR value for the progressive model showed a positive trend with an OR of 1.74 (0.84–3.64) and a significant association with the H-SIL+ stage with an OR of 19.41 (4.95–76.10). Due to their low frequencies (<5%), other SNPs were not analyzed.

4 | DISCUSSION

This is the first study to address the genetic variation of E6 and E7 in Misiones, Argentina. Both genes are involved in the transformation process through the interaction with p53 and pRb human proteins, respectively.[27] Thus, their genetic variation has been of upmost

TABLE 1 Genetic characterization of E6 and E7 genes in cervical samples from Misiones

	E6													Pap cytology			
Ref	109 T	131 A	145 G	176 G	217 G	286 T	289 A	310 T	335 C	350 T	403 A	532 A	Protein	NILM	LSIL	HSIL+	Total
Seq_01	.	.	.	.	.	.	.	.	.	.	.	.	Prototype	15	16	11	42
Seq_02	.	.	.	.	.	.	.	.	.	G	.	.	L83V	13	28	16	57
Seq_03	.	.	.	.	.	.	.	.	.	.	G	.	Prototype	0	0	1	1
Seq_04	C	.	.	.	.	.	.	.	.	G	.	.	L83V	0	1	1	2
Seq_05	.	.	.	.	A	.	.	.	.	G	.	.	L83V	0	1	0	1
Seq_06	.	.	.	A	.	.	.	.	.	.	.	.	D25N	0	1	0	1
Seq_07	.	G	.	.	.	.	.	.	.	.	.	.	R10G	0	1	0	1
Seq_08	.	.	.	.	.	.	.	G	.	.	.	.	F69L	1	0	0	1
Seq_09	.	.	.	C	.	.	.	.	.	G	.	.	D25H L83V	0	1	0	1
Seq_10	.	.	T	.	.	A	G	.	T	G	.	.	Q14H F78Y L83V	0	0	1	1
Seq_11	.	.	T	.	.	A	G	.	T	G	.	G	Q14H F78Y L83V	0	0	4	4
Total														29	49	34	112

	E7										Pap cytology			
Ref	603 T	627 C	642 A	658 G	712 C	732 T	789 T	795 T	823 G	Protein	NILM	LSIL	HSIL+	Total
Seq_1	.	.	.	.	.	.	.	.	.	Prototype	22	34	28	84
Seq_2	.	.	.	C	.	.	.	.	.	Prototype	1	0	0	1
Seq_3	.	.	.	.	.	.	.	.	C	H51N	2	0	2	4
Seq_4	.	.	.	.	A	.	?	?	?	G88R	0	1	0	1
Seq_5	.	.	.	.	.	C	C	G	.	Prototype	0	0	3	3
Total											25	35	33	93

Note: The prototype sequence is indicated as reference (NC-001526). The nucleotide positions for SNPs are at the top of the table. For amino acids, we use the one-letter code (D = Aspartic acid; F = Phenylalanine; G = Glycine; H = Histidine; L = Leucine; N = Asparagine; Q = Glutamine; R = Arginine; V = Valine; Y = Tyrosine). The position of amino acid change is stated numerically. The letter preceding this number refers to the reference amino acid and the letter after refers to its substitution.

interest as a molecular marker of cancer progression.[8,14,18,28] E6 350G variant was found in 44% of controls and 65.7% of cervical lesion cases. The latter value falls into the range described in other local studies from the urban population of the city of Córdoba (67%), indigenous Guarani populations from Misiones (50%), and Quechua descendants from Jujuy (>60%).[29–31] Yet, none of these studies addressed the strength of the association between this viral marker and CC in the Argentinean population.

In this study, ordinal logistic regression was used to make this assessment. The results showed an OR value of 1.74 (0.84–3.64) for the general progressive model, thus establishing a positive trend in the association between the E6 350G variant and cervical lesions. Interestingly, the association for H-SIL cases was significant at an OR of 19.41 (4.95–76.10). This OR is higher than that described by Cornet et al.[11] for Argentina (OR = 5.81; 95% CI = 1.64–21.12) and Central and South America (OR = 4.69; 95% CI = 2.07–10.66),[11] adding supporting evidence for the role of 350G in the development of CC in Latin America.

However, this association is not supported in populations from Europe and Central Asia (OR = 0.42; 95% CI = 0.27–0.64).[11] The reason that there is no linkage between the E6 350G SNP and disease risk in European and Asian populations remains unclear, but has been discussed elsewhere.[11] For example, early studies showed that this SNP was not frequent in populations from Africa and Southeast Asia.[9,14] Therefore, the described associations are likely population dependent,[11] meaning any extrapolation of our results to other populations or locations should be taken with caution.

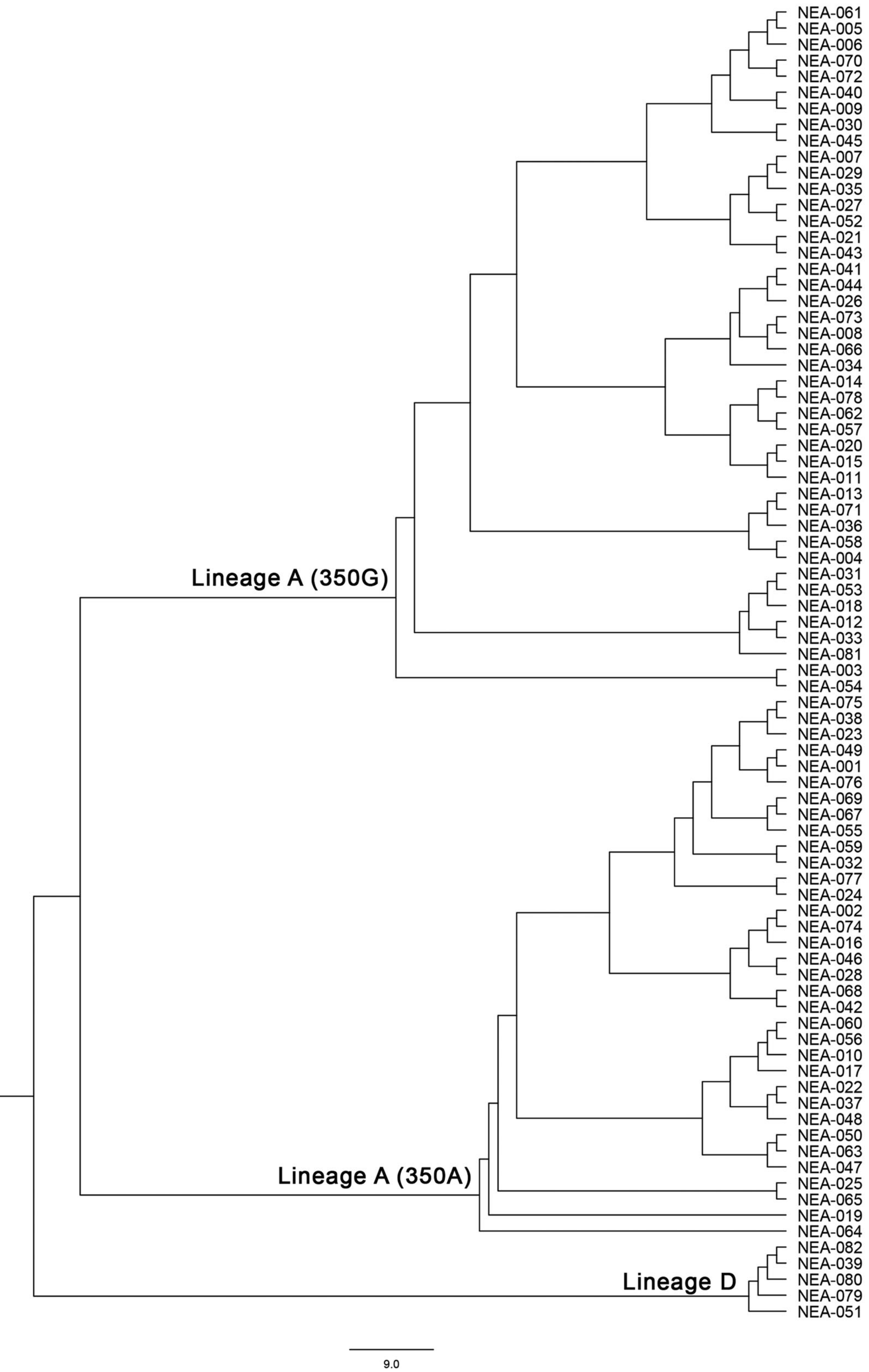

FIGURE 1 Evolutionary relationships of HPV16 isolates from Misiones. The evolutionary history of E6 and E7 concatenated genes. This cladogram was inferred using Bayesian inference. The samples from Misiones have an "NEA" prefix in their label. The recurring nature of the mutation E6 350G along different branches of the HPV16 evolution (homoplasy) is shown at the nodes of the tree

TABLE 2 Association analysis between E6 SNPs and cervical lesions

HPV16	NILM	LSIL	HSIL+	OR[a]	95% CI	OR[b]	95% CI
E6 350T	16	18	12	1	Ref	1	Ref
E6 350G	13	30	18	1.74[c]	(0.84–3.64)	1.47[c]	(0.69–3.10)
				2.76[d]	(0.79–9.66)	2.92[d]	(0.82–10.32)
				19.41[e]	**(4.95–76.10)**	**22.72**[e]	**(5.62–91.75)**

Abbreviations: CI, confidence interval; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; NILM, negative for intraepithelial lesion or malignity; OR, odds ratio; SNP, single-nucleotide polymorphism.

[a]Ordinal logistic regression adjusted by age.

[b]Ordinal logistic regression adjusted by age and lineage.

[c]General for the progression NILM-LSIL-HISL+.

[d]Cut point: LSIL versus NILM.

[e]Cut point: HSIL versus NILM.

From an evolutionary point of view, the E6 350G SNP has arisen independently along different HPV16 lineages. In this study, we found the 350G SNP in both Lineage A and D samples, consistent with previous reports.[8] Interestingly, in vitro studies have shown that keratinocytes infected with E6 350G presented a higher capacity for cell transformation compared to E6 350T, irrespective of their evolutionary origin.[12] In this data set, the risk of E6 350G adjusted by lineage was significant at an OR of 22.72 (5.62–91.75), supporting the clinical importance of this SNP along different branches of HPV16 evolution.

Finally, other non-synonymous SNPs were identified in the E6 (R10G, D25N, D25H, and F69L) and E7 (H51N and G88R) genes, although they occur at low frequencies (<5%). As a result, the epidemiological implications of these mutations remain unclear. In vitro studies have found that D25E variant downregulated E-cadherin to facilitate an epithelial-to-mesenchymal transition.[14] Other studies have reported that the R10G variant is relatively frequent in oral tonsillar squamous cell carcinoma compared to CC (19% and 4%, respectively), raising questions about tissue tumor induction and development by HPV in other types of cancers.[32]

Finally, the low sequence diversity of the E7 compared to E6 gene was confirm. The E7 gene function is critical for earlier phases of the viral life cycle, particularly for genome amplification and replication.[15] Taking into consideration the small size of this protein, the sequence might represent the optimal solution for its function, leading to evolutionary constraints on its sequence diversification.[16] Similarly, previous studies have suggested that the conservation of E7 is important for carcinogenesis.[17]

5 | CONCLUSION

Genetic diversity of HPV16 genes E6 and E7 was assessed in women from northwestern Argentina. Results confirmed the association of E6 350G SNP with a higher risk of developing high-grade cervical lesions. These data will contribute to understanding the biological bases of CC incidence in this region, and will be useful for elucidating the distribution of HPV16 variants and their risk associations around the world.

ACKNOWLEDGMENTS

This study was partially supported by research grant from the Agencia Nacional de Promoción Científica y Técnica (ANPCyT) PICTO-UNaM 2011-0106 (to DJL and IB), Subsidio Florencio Fiorini para Investigación en Ciencias Biomédicas, Academia Nacional de Medicina, 2012 (to IB), a Grant of the the Sistema Nacional de Datos Genómicos (SNDG), Ministerio de Ciencia, Tecnología e Innovación (MINCYT), Consejo Interinstitucional de Ciencia y Tecnología (CICyT) (to IB), a Scholarship of Beca CIN del programa estratégico de formación de recursos humanos en investigación y desarrollo (PERHID) Resol. CE 1334-18 (to MET), and a Scholarship of the SNDG-MINCYT-CICyT (to MET). The software used in the study was partially supported by a Codon Code Aligner Grants license program (IB). El IB and JAG are members of Consejo Nacional de Investigaciones Científico y Tecnológicas (Conicet). None of the funding agencies have been involved in the study design, data collection, analysis, or paper writing and submission.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

M. Elina Totaro: Conceptualization, Validation, Investigation, Data curation, Writing – review and editing. **Juan A. Gili:** Data curation, Formal analysis, Writing – review and editing. **D. Javier Liotta:** Supervision, Writing – review and editing. **Theodore G. Schurr:** Writing – review and editing. **Maria A. Picconi:** Writing – review and editing. **InésBadano:** Conceptualization, Validation, Investigation, Supervision, Writing – review and editing.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in GeneBank at (https://www.ncbi.nlm.nih.gov/genbank/), reference numbers MN053339–MN053388; MN160917–MN061890.

ORCID

Juan A. Gili https://orcid.org/0000-0001-7924-0776

Inés Badano https://orcid.org/0000-0003-3033-3802

REFERENCES

1. Bruni L, Albero G, Serrano B, et al. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in Argentina. Summary Report 17 June 2019. Accessed December 23, 2020.
2. Badano I, Pedrozo RW, Ruíz Díaz LS, et al. Human papillomavirus (HPV) detection and Papanicolaou cytology in low-resource women in Posadas city, Misiones, Argentina. *Rev Argent Microbiol*. 2011;43(4): 263-267. https://doi.org/10.1590/S0325-75412011000400005
3. Badano I, Totaro ME, Culasso AC, et al. Genetic characterization and clinical implications of human papillomavirus type 16 (HPV16) variants from northeastern Argentina. *Infect Genet Evol*. 2015;29: 103-109. https://doi.org/10.1016/j.meegid.2014.11.013
4. Tonon SA, Picconi MA, Zinovich JB, et al. Prevalencia de la infección cervical por virus papiloma humano (HPV) en población caucásica y guaraní, residente en la provincia de Misiones, Argentina [Prevalence of cervical infection by human papilloma virus (HPV) in the Caucasian and Guaraní populations residing in the province of Misiones, Argentina]. *Rev Argent Microbiol*. 2003;35(4):205-213.
5. González JV, Deluca GD, Correa RM, et al. Strong reduction in prevalence of HPV16/18 and closely related HPV types in sexually active adolescent women following the introduction of HPV vaccination in Argentina. *Papillomavirus Res*. 2020;10:100208. https://doi.org/10.1016/j.pvr.2020.100208
6. Badano I, Sanabria DJ, Totaro ME, et al. Mitochondrial DNA ancestry, HPV infection and the risk of cervical cancer in a multiethnic population of northeastern Argentina. *PLOS One*. 2018;13(1): e0190966. https://doi.org/10.1371/journal.pone.0190966
7. Vande Pol SB, Klingelhutz AJ. Papillomavirus E6 oncoproteins. *Virology*. 2013;445(1–2):115-137. https://doi.org/10.1016/j.virol.2013.04.026
8. Martinelli M, Musumeci R, Sechi I, et al. Prevalence of Human Papillomavirus (HPV) and other Sexually Transmitted Infections (STIs) among Italian women referred for a colposcopy. *Int J Environ Res Public Health*. 2019;16(24):5000. https://doi.org/10.3390/ijerph16245000
9. Yamada T, Manos MM, Peto J, et al. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. *J Virol*. 1997;71(3):2463-2472. https://doi.org/10.1128/JVI.71.3.2463-2472.1997
10. Cornet I, Gheit T, Franceschi S, et al. IARC HPV variant study group. Human papillomavirus type 16 genetic variants: phylogeny and classification based on E6 and LCR. *J Virol*. 2012;86(12):6855-6861. https://doi.org/10.1128/JVI.00483-12
11. Cornet I, Gheit T, Iannacone MR, et al. HPV16 genetic variation and the development of cervical cancer worldwide. *Br J Cancer*. 2013; 108(1):240-244. https://doi.org/10.1038/bjc.2012.508
12. Chen Z, Terai M, Fu L, Herrero R, DeSalle R, Burk RD. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. *J Virol*. 2005;79(11):7014-7023. https://doi.org/10.1128/JVI.79.11.7014-7023.2005
13. Sichero L, Sobrinho JS, Villa LL. Oncogenic potential diverge among human papillomavirus type 16 natural variants. *Virology*. 2012; 432(1):127-132. https://doi.org/10.1016/j.virol.2012.06.011
14. Tan G, Duan M, Li Y, et al. Distribution of HPV 16 E6 gene variants in screening women and its associations with cervical lesions progression. *Virus Res*. 2019;273:197740. https://doi.org/10.1016/j.virusres.2019.197740
15. Roman A, Munger K. The papillomavirus E7 proteins. *Virology*. 2013; 445(1–2):138-168. https://doi.org/10.1016/j.virol.2013.04.013
16. Safaeian M, van Doorslaer K, Schiffman M, et al. Lack of heterogeneity of HPV16 E7 sequence compared with HPV31 and HPV73 may be related to its unique carcinogenic properties. *Arch Virol*. 2010;155(3): 367-370. https://doi.org/10.1007/s00705-009-0579-2
17. Mirabello L, Yeager M, Yu K, et al. HPV16 E7 genetic conservation is critical to carcinogenesis. *Cell*. 2017;170(6):1164-1174. https://doi.org/10.1016/j.cell.2017.08.001
18. Van Doorslaer K, Li Z, Xirasagar S, et al. The Papillomavirus Episteme: a major update to the papillomavirus sequence database. *Nucleic Acids Res*. 2017;45(D1):D499-D506. https://doi.org/10.1093/nar/gkw879
19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410. https://doi.org/10.1016/S0022-2836(05)80360-2
20. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725-2729. https://doi.org/10.1093/molbev/mst197
21. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007;7:214. https://doi.org/10.1186/1471-2148-7-214
22. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol*. 2005;22(5):1185-1192. https://doi.org/10.1093/molbev/msi103
23. Chen Z, DeSalle R, Schiffman M, et al. Niche adaptation and viral transmission of human papillomaviruses from archaic hominins to modern humans. *PLOS Pathog*. 2018;14(11):e1007352. https://doi.org/10.1371/journal.ppat.1007352
24. Rambaut A, Suchard M, Xie D, Drummond A. Tracer v1.6. 2014. http://beast.bio.ed.ac.uk/Tracer
25. Rambaut A. FigTree v.1.4.0. 2012. http://tree.bio.ed.ac.uk/software/figtree/
26. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34(12):3299-3302. https://doi.org/10.1093/molbev/msx248
27. Sichero L, Picconi MA, Villa LL. The contribution of Latin American research to HPV epidemiology and natural history knowledge. *Braz J Med Biol Res*. 2020;53(2):e9560. https://doi.org/10.1590/1414-431X20199560
28. Marongiu L, Godi A, Parry JV, Beddows S. Human papillomavirus type 16 long control region and E6 variants stratified by cervical disease stage. *Infect Genet Evol*. 2014;26:8-13. https://doi.org/10.1016/j.meegid.2014.05.009
29. Picconi MA, Alonio LV, Sichero L, et al. Human papillomavirus type-16 variants in Quechua aboriginals from Argentina. *J Med Virol*. 2003;69(4):546-552. https://doi.org/10.1002/jmv.10343
30. Tonon SA, Basiletti J, Badano I, et al. Human papillomavirus type 16 molecular variants in Guarani Indian women from Misiones, Argentina. *Int J Infect Dis*. 2007;11(1):76-81. https://doi.org/10.1016/j.ijid.2006.01.002
31. Mosmann JP, Monetti MS, Frutos MC, Kiguen AX, Venezuela RF, Cuffini CG. Mutation detection of E6 and LCR genes from HPV 16 associated with carcinogenesis. *Asian Pac J Cancer Prev*. 2015;16(3): 1151-1157. https://doi.org/10.7314/apjcp.2015.16.3.1151
32. Du J, Nordfors C, Näsman A, et al. Human papillomavirus (HPV) 16 E6 variants in tonsillar cancer in comparison to those in cervical cancer in Stockholm, Sweden. *PLOS One*. 2012;7(4):e36239. https://doi.org/10.1371/journal.pone.0036239

How to cite this article: Totaro ME, Gili JA, Liotta DJ, Schurr TG, Picconi MA, Badano I. Genetic variation in the E6 and E7 genes of human papillomavirus type 16 in northeastern Argentina. *J Med Virol*. 2021;1-7. https://doi.org/10.1002/jmv.27359