Data Article

Human papillomavirus type 13: Genome amplification and characterization data

Nuvia Kantún-Moreno\textsuperscript{a}, Gemaly Elisama Ek-Hernández\textsuperscript{b}, José Reyes Canché-Pech\textsuperscript{a}, Jesús Gilberto Gómez-Carballo\textsuperscript{a}, María del Refugio González-Losa\textsuperscript{a}, Laura Conde-Ferráez\textsuperscript{a,∗}

\textsuperscript{a} Universidad Autónoma de Yucatán, Centro de Investigaciones Regionales, Dr. Hideyo Noguchi, Laboratorio de Virología, Merida, Yucatan, Mexico
\textsuperscript{b} Instituto Tecnológico Superior de Calkiní en el Estado de Campeche, Calkiní, Campeche, Mexico

As for 2020 only two complete genomes of Human papillomavirus type 13 (HPV13) are publicly available in GenBank database. In addition, reports of partial sequences of genetic regions are very limited. Therefore, genomic research that contributes to knowledge of viral components involved in HPV13 pathogenesis, and molecular mechanisms associated to multifocal epithelial hyperplasia (MEH) disease are urged. In the accompanying paper \cite{1}, we aimed to obtain the complete genome sequence of HPV13 associated to MEH disease, obtained from a Mayan boy living in Yucatan, Mexico. Coding sequences were annotated, and viral proteins traduced and deposited in GenBank with accession number MT068446. In this data report, we present the oligonucleotide list used to amplify the complete genome, a graphical abstract of process employed for the amplification of circular HPV13 genome, a representative figure of PCR products obtained for sequencing and multiple sequence alignments with the translated coding sequences of the existing genomes: X62843 is the first HPV13 genome reported \cite{2}; it was generated from a clone obtained from a Turkish patient; DQ344807 was originally obtained from a patient in the Amazonian region \cite{3}. The multiple sequence alignments show the main
viral proteins (predicted). This provides relevant information for future molecular analysis and epidemiological studies because HPV13 is an understudied genotype associated to a neglected disease that appears more commonly in children. Additionally, the description of the methods can help in future sequencing of HPV genomes. We hope that our solutions will help researchers who do not have next-generation sequencing (NGS) platforms. A more comprehensive analysis of this data may be obtained from “Genomic characterization of Human papillomavirus type 13, associated to Multifocal Epithelial Hyperplasia, in a Mayan community” [1].

© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

### Specifications Table

| Subject | Virology |
|---------|----------|
| Specific subject area | Human papillomavirus genomics |
| Type of data | Table, Figure, Alignments, GenBank metadata, FASTA files with amino acid sequences |
| How data were acquired | DNA isolation, PCR amplifications, Gel imaging and Sanger sequencing. Instruments: Mastercyrclyer Ep Gradient thermocycler (Eppendorf, mod. 5341-02537), ABI PRISM 310 Genetic Analyzer (Applied Biosystems, mod. 310–3), Owl EasyCast mini gel electrophoresis system (Thermo Scientific, mod. 7309 B1) and Gel Doc XR System (Bio-Rad, mod. Universal Hood II). Software packages: Geneious R6 software v.6.1.8, Image lab v.2.0.1, Phred, Phrap and Consed v. 29.0, BoxShade v. 3.21, CLC sequence viewer 8.0 |
| Data format | Raw data (primer sequences), Analysed sequence alignments, FASTA formatted sequence files and Image (TIF) |
| Parameters for data collection | A sample from oral cells from a Mayan 11 year old male from Yucatan, with clinical signs of MEH disease, associated to HPV13, was processed, amplified, and sequenced. From genome assembly, coding sequences (CDS) were annotated and translated for predicted viral proteins. Silent mutations and amino acidic changes were identified from alignments with predicted proteins from two previous HPV13 genomes |
| Description of data collection | A rolling cycle amplification (RCA) of the sample was used as template. PCR amplification of overlapping 500 bp fragments were obtained with primers that where designed in-house. Amplicons were Sanger sequenced using BigDyeTM terminator reaction, read using ABI PRISM™ 310 Genetic Analyzer. The reads were trimmed and assembled using Phred, Phrap and Consed (v. 29.0) software. Final sequence of 7831 bp was annotated and formatted for GenBank. Predicted proteins for our HPV13_YUC were pairwise compared with E6, E7, E1, E2, E4, E5, L1 and L2 proteins reported for DQ344807 (from Amazonian [3]) and X62843 (from Turkey [2]), using CLC sequence viewer (v. 8.0) software |
| Data source location | Institution: Universidad Autónoma de Yucatan |
| | City/Town/Region: San Francisco, Tinum, Yucatan |
| | Country: Mexico |
| | Latitude and longitude (and GPS coordinates): 20.766 N and 88.383 W |
| Data accessibility | Repository name: NCBI GenBank |
| | Data identification number: Genome and annotation: MT068446. The direct URL to the data is https://www.ncbi.nlm.nih.gov/nuccore/MT068446 HPV13 proteins data available with this article |
| Related research article | Laura Conde-Ferráez; Gemaly Elisama Ek-Hernández; José Reyes Canché-Pech; Jesús Gilberto Gómez-Carballo; Nuvia Eugenia Kantún-Moreno; María del Refugio González-Losa. Genomic characterization of Human papillomavirus type 13, associated to Multifocal Epithelial Hyperplasia, in a Mayan community. Infect Genet Evol. https://doi.org/10.1016/j.meegid.2020.104595 |
Value of the Data

- HPV13 is an understudied genotype, associated to a neglected disease that affects predominantly ethnic groups of the Americas, such as the Maya. Genomic information is important but scarce. Research that contributes to the knowledge of the molecular determinants underlying this pathology are urged. The data presented includes the list of primers successfully used to amplify and characterize the first HPV13 genome from Mexico, using simple Sanger sequencing.
- The genomic data can be of useful of researchers interested on the evolutionary biology of HPV, because multiple proteins alignments of the viral proteins are presented. Researchers from developing countries (with ethnic groups affected by MEH) rarely can afford Next Generation Sequencing platforms, and therefore they employ more traditional methods for molecular studies; therefore the primers list provided can be useful to amplify any region of HPV13 genome.
- The data of the list of primers and protein alignments are a valuable tool that can be used for future experiments, such as studies on molecular epidemiology, HPV13 variants, evolution and phylogenetics. These primers are a valuable tool that can also be evaluated for diagnosis of the associated disease, or for tracking asymptomatic carriers. Other HPV13 genomes can be isolated for study their genetic diversity and gather information about viral evolution.
- The description of the methods can help in sequencing HPV genomes. More research that contributes to the knowledge of HPV13 and the associated pathology are urged. In the Maya of Yucatan, Mexico, it is considered an endemic pathology, and is very frequent mainly in rural areas. However, as it is rare in developing countries and in urban areas, the viral molecular determinants underlying this disease remain unknown.
- The data provide an updated genetic information on silent mutations and amino acidic changes (substitutions, deletions, and additions). The availability of these protein alignments in FASTA format will provide users a starting scaffold for assessing newly obtained HPV13 sequences and screening of new mutations among isolates.

1. Data Description

We sequenced the first HPV13 genome associated to MEH, obtained from a patient from a Maya community in a rural area of Yucatan state, Mexico. After trimmed and final assembly, we reported a length of 7831 bp for HPV13 genome from Yucatan (referred to as HPV13_YUC). Coding sequences were annotated, and viral proteins traduced [1]. The information was deposited in GenBank with accession number MT068446. Data presented in the text includes a list of primers with their optimal annealing temperatures for the amplification of DNA segments that coverage the HPV13 genome by regions, included the long control region (Table 1). A graphic representation of amplification strategy by overlapping PCR products, based on rolling cycle amplification (RCA) method for sequencing of full-length HPV13_YUC genome is shown in Fig. 1. Overlapping PCR products of approximately 500 bp were amplified with each primer set (Fig. 2), according to the strategy mentioned above, all amplicons were purified and sequenced with Sanger method. CDSs were annotated for E6, E7, E1, E2, E4, E5, L2 and L1, the genes were translated and compared with other predicted proteins reported for HPV13 from Amazonian (DQ344807) and Turkey (X62843). Details about amino acid changes were highlighted on protein alignments shown in Fig. 3. Alignments of E6, E7, E1, E2, E4, E5, L2 and L1 proteins are available in FASTA format (Supplemental files, HPV13 protein alignments in FASTA: “E6_alignment.fa”; “E7_alignment.fa”; “E1_alignment.fa”; “E2_alignment.fa”; “E4_alignment.fa”; “E5_alignment.fa”; “L2_alignment.fa”; “L1_alignment.fa”).
| Primer       | Sequence 5′–3′                        | Melting temperature (Tm) °C | Gene region |
|-------------|--------------------------------------|----------------------------|-------------|
| FRAG 1 HPV 13 - 35 F | GAC CGA AAA CGG TTT                     | 52.1                       | URR         |
| FRAG 1 HPV 13 - 534 R | TAA CC TAT CAT GAC CAG CAA TGA AA                 | 51.0                       | E6          |
| FRAG 2 HPV 13 - 399 F | ATG TGC TAA TTC GCT AT                  | 52.0                       | E6          |
| FRAG 2 HPV 13 - 898 R | AAA AAC CAT CCT GAG CAT CC              | 53.1                       | E1          |
| FRAG 3 HPV 13 - 819 F | GTG TGC ACC AAA AAG                     | 52.7                       | E7          |
| FRAG 3 HPV 13 - 1318 R | CTA AC CCA CAA TCA TTT TCC GGT TC       | 52.0                       | E1          |
| FRAG 4 HPV 13 - 1232 F | GGA AAT AAC GGA CAG                     | 52.2                       | E1          |
| FRAG 4 HPV 13 - 1732 R | TGG AT TTA AGA AGT GCC AGT GT           | 52.0                       | E1          |
| FRAG 5 HPV 13 - 1528 F | CAA CAT GTG GGG ACT                     | 56.0                       | E1          |
| FRAG 5 HPV 13 - 1993 R | GGG TCA AAA CCT CTG TGT GC              | 53.1                       | E1          |
| FRAG 6 HPV 13 - 1768 F | AAA TAC AAA GCA GTG                    | 52.9                       | E1          |
| FRAG 6 HPV 13 - 2267 R | TGG CA CCC CTC TAT TGC AAT ACA GT       | 52.3                       | E1          |
| FRAG 7 HPV 13 - 2136 F | GAG GAA GCA GGA AAT                     | 52.7                       | E1          |
| FRAG 7 HPV 13 - 2635 R | TGG AA GCA TTC CCA TTT CTG TCA AA       | 51.2                       | E1          |
| FRAG 8 HPV 13 - 2460 F | GGC AAT CCA ATG AGC ATT G              | 52.3                       | E1          |
| FRAG 8 HPV 13 - 2964 R | CAT TTG CAT TTC AAT TGC CTC            | 51.2                       | E2          |
| FRAG 9 HPV 13 - 2887 F | AGC CAC ATT GGA TTA CAA GT              | 51.9                       | E2          |
| FRAG 9 HPV 13 - 3386 R | GTG GAG TAT GAA G                     | 52.5                       | E2          |
| FRAG 10 HPV 13 - 3269 F | GGG AAA CGT TAC AAT                    | 52.5                       | E2          |
| FRAG 10 HPV 13 - 3768 R | GGG A GGT TAA GGT TAC CAG TGC AT       | 52.5                       | E2          |
| FRAG 11 HPV 13 - 3492 F | CCA GAA CAC ACA AAG                    | 52.5                       | E2          |
| FRAG 11 HPV 13 - 3991 R | CAT TG TGC AAG TGC AAT TAC AAG TGG     | 50.9                       | E5 GAMMA    |
| FRAG 12 HPV 13 - 3740 F | CAC AAA AAC ATG CAC                    | 51.4                       | E2          |
| FRAG 12 HPV 13 - 4239 R | TGG TA AAC CAT GTG TCA CCA TCA TC      | 52.8                       | E5 DELTA    |

| FRAG 13 HPV 13 - 4096 F | ACTAACAATCCCTTGCACAT                  | 51.9                       | E5 GAMMA    |
| FRAG 13 HPV 13 - 4595 R | CTACTGGTACATAGCCAGT                   | 51.7                       | L2          |
| FRAG 14 HPV 13 - 4428 F | AAACCTTGAAGGCCTTTGGA                  | 51.5                       | L2          |
| FRAG 14 HPV 13 - 4927 R | AGATGGGCGATATAGAACCAGT               | 51.4                       | L2          |
| FRAG 15 HPV 13 - 4797 F | TGGATGTTCTGTTACACAA                  | 51.4                       | L2          |
| FRAG 15 HPV 13 - 5296 R | ATACATAGACCCCCCTGTGAC                | 52.8                       | L2          |
| FRAG 16 HPV 13 - 5222 F | CTACATAGGCCAGCCATAAC                 | 52.6                       | L2          |
| FRAG 16 HPV 13 - 5721 R | CTTTGTGTGTTAAATAATCAA              | 51.4                       | L2          |
| FRAG 17 HPV 13 - 5591 F | GACATAACATCCCCCAACTGC               | 51.5                       | L2          |
| FRAG 17 HPV 13 - 6090 R | TACCAACACTTAAAGGGTGA               | 53.1                       | L1          |
| FRAG 18 HPV 13 - 6021 F | ACTAGTCAACCTAGTTG                  | 52.0                       | L1          |
| FRAG 18 HPV 13 - 6519 R | AATGCCCTGCAAAATTTGT                 | 51.7                       | L1          |

(continued on next page)
Table 1 (continued)

| Primer       | Sequence 5′ – 3′       | Melting temperature (Tm) °C | Gene region |
|--------------|------------------------|----------------------------|-------------|
| FRAG 19 HPV13 - 6289 F | GAGATGGCCCTCCTTAGAA       | 51.7                       | L1          |
| FRAG 19 HPV13 - 6788 R | AAGAGATGATGTAGTGGCTG       | 51.7                       | L1          |
| FRAG 20 HPV13 - 6517 F | ATTTCTTAAAGGGGCAAGG       | 53.3                       | L1          |
| FRAG 20 HPV13 - 7016 R | CTTATGGCCTGAGATTGTA       | 52.0                       | L1          |
| FRAG 21 HPV13 - 6938 F | CTGGAACITTTGGGCTATCTC      | 52.8                       | L1          |
| FRAG 21 HPV13 - 7437 R | CGAGACAACATACACCATT       | 52.3                       | URR (LCR)   |
| FRAG 22 HPV13 - 7265 F | ATTTCTTTAACAGGGCAGGC       | 51.1                       | URR (LCR)   |
| FRAG 22 HPV13 - 7764 R | CGTTATGGCCTGAGATTGTA       | 52.2                       | URR (LCR)   |
| FRAG 23 HPV13 – 7643 F | CTAGGCCGCCCTTATAT           | 51.1                       | URR (LCR)   |
| FRAG 23 HPV13 – 7847 R | GTTTCAGGTGTGGAGCC           | 54.3                       | URR (LCR)   |

Fig. 1. Graph representation of a two-step simple procedure for the HPV13_YUC sequencing. 1) Rolling Circle Amplification (RCA) method as first amplification step (see methods section). 2) End-point PCR amplification using RCA product as template, and specific primer pairs. HPV13 genome obtained from Amazonian (DQ344807) was used as reference sequence for primers design (see Table 1). A key point of this strategy was optimal primer pairs design to cover the entire genome in 500 pb overlapping fragments for the later assembly of contigs. The template for this amplification was the RCA product, for convenience, it is presented the linear and linear representation of the HPV13 genome. Each color bar is representative of a coding sequence (CDS). All expected amplicons are shown as lines (fragment 1 to fragment 22), with their primer pairs. The last amplicon (fragment 23) representative of promoter region or LCR sequence is show in dotted lines. Primer pair 534 R and 7643 F was used to complete the last fragment of the genome. Amplicons were subjected to Sanger sequencing and the HPV13_YUC genome assembled.
2. Experimental Design, Materials and Methods

2.1. DNA isolation

A sample from oral cells from a Mayan 11 year old male from San Francisco, Tinum, Yucatan State, Mexico (20.766 N and 88.383 W), with clinical signs of MEH disease and positive to HPV13 was processed. This sample is part of a repository of oral swabs of virology laboratory of Centro de Investigaciones Regionales, Dr. Hideyo Noguchi.

Total genomic DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) following the manufacturer's instructions for cells protocol. Briefly, buccal cells were washed twice with 500 µL 1X PBS (phosphate buffered saline) and then, the cell pellet was resuspended in 200 µL 1X PBS and 20 µL proteinase K. The sample was mixed by vortex and incubated for 10 min at 56 °C in a thermomixer. All subsequent steps were carried out at room temperature and following the spin column method for DNA isolation. The quantity and quality (purity and integrity) of the DNA were evaluated with a NanoDrop 2000 (Thermo Scientific) and by PCR using β-globin primers GH20 and PCO4 [4]. Finally, 260 bp amplicon was confirmed on 1.2% agarose gel with 1X TAE buffer and ethidium bromide staining.

2.2. Primers design

Specific primers were designed on the previously reported genome from the Amazonian [3], using Geneious R6 software v.9.1.6 (Biomatters Ldt) [5], for amplification of HPV13_YUC genome. The following criteria were applied to design primers: GC content from 45 to 60%, length of 18 to 25 bases, Tm between 53 and 58 °C with a Tm optimal of 55 °C and short amplicons of approximately 500 pb named as fragment 1 to 23. Primers were designed to amplify overlap

Fig. 2. Agarose gel electrophoresis of fragments 1 to 12 from HPV13 genome amplified by PCR using primers on Table 1. Lane 2 –11, 13- 14 show 500 pb amplicons obtained from RCA template.
Fig. 3. Multiple sequence alignments of translated CDS from HPV13 from Turkey, Amazonian and Yucatan (HPV13_YUC). Each panel correspond to predicted proteins: A) E6, B) E7, C) E1, D) E2, E) E4, F) E5, G) L2 and H) L1. ID proteins from Turkey: CAA44647.1, CAA44648.1, CAA44649.1, CAA44650.1, CAA44651.1, CAA44652.1, CAA44653.1, CAA44654.1. ID proteins from Amazonian: ABC79057.1, ABC79058.1, ABC79059.1, ABC79060.1, ABC79061.1, ABC79062.1, ABC79063.1, ABC79064.1. Identical amino acids are shaded in black. Clustal consensus key: “*”, fully conserved residue, “:” conservation of strong groups,”.” conservation of weak groups and blank space, no consensus. Alignments of the amino acid sequences represented in this figure is included in supplemental material as FASTA alignment files.
N. Kantún-Moreno, G.E. Ek-Hernández and J.R. Canché-Pech et al. / Data in Brief 35 (2021) 106955

Fig. 3. Continued
**Panel E) E4**

| E4_ABC79064.1                  | MKRNYGKYNL1AASQYLHHLVLYKLKYPPLGLLH1TPPPPHFPPQCPAARPRKVCR 60 |
| E4_CAA44651.1                  | MKRNYGKYNL1AASQYLHHLVLYKLKYPPLGLLH1TPPPPHFPPQCPAARPRKVCR 60 |
| E4_HPv13_YUC consensus         | .......................................................................................... |
| E4_ABC79064.1                  | SLVNDHELHVLETPTHKALCYSOTTFQTVTTTSTTLL1ITTITGTTTVQCRLH 118 |
| E4_CAA44651.1                  | SLVNDHELHVLETPTHKALCYSOTTFQTVTTTSTTLL1ITTITGTTTVQCRLH 118 |
| E4_HPv13_YUC consensus         | .......................................................................................... |

**Panel F) E5**

| E5_CAA44652.1                  | MERIPVDYSTCQTSKSLLPLVIALTVCSVISITIILGSEIFVYVNLVIIIYLLW 60 |
| E5_ABC79061.1                  | MERIPVDYSTCQTSKSLLPLVIALTVCSVISITIILGSEIFVYVNLVIIIYLLW 60 |
| E5_HPv13_YUC consensus         | .......................................................................................... |
| E5_CAA44652.1                  | LTTPLQFYLTITSLCFLPACVHYQILQTCE 91 |
| E5_ABC79061.1                  | LTTPLQFYLTITSLCFLPACVHYQILQTCE 91 |
| E5_HPv13_YUC consensus         | .......................................................................................... |

**Panel G) L2**

| L2_CAA44653.1                  | MAHGRRKRKASATQYLQTCAKSGTCDPPV1IPKVQGNTLADKILKWSLGFVGGGLIGC 60 |
| L2_ABC79062.1                  | MAHGRRKRKASATQYLQTCAKSGTCDPPV1IPKVQGNTLADKILKWSLGFVGGGLIGC 60 |
| L2_HPv13_YUC consensus         | .......................................................................................... |
| L2_CAA44653.1                  | TGSCTRGTGYVPGVTSPRAPIAGPTARPPTVPTTVGTDPSPSVLWSEAIINSVPED 120 |
| L2_ABC79062.1                  | TGSCTRGTGYVPGVTSPRAPIAGPTARPPTVPTTVGTDPSPSVLWSEAIINSVPED 120 |
| L2_HPv13_YUC consensus         | .......................................................................................... |
| L2_CAA44653.1                  | PLPPVHGGEITTSQASATPAILDVSSTTTQNTTSSTISFPNVFSEFSITQQSISSAHAV 180 |
| L2_ABC79062.1                  | PLPPVHGGEITTSQASATPAILDVSSTTTQNTTSSTISFPNVFSEFSITQQSISSAHAV 180 |
| L2_HPv13_YUC consensus         | .......................................................................................... |
| L2_CAA44653.1                  | FISPSIISPHSTEDIPLOPTIVSSSDNPASTQVPTVATVARFLGLYSRAHLQVVTDEA 240 |
| L2_ABC79062.1                  | FISPSIISPHSTEDIPLOPTIVSSSDNPASTQVPTVATVARFLGLYSRAHLQVVTDEA 240 |
| L2_HPv13_YUC consensus         | .......................................................................................... |
| L2_CAA44653.1                  | FISSPQRLTIFDNTYEGIDISLQFAHINTHEFDEAMDIIRHLFAITSRGGLVFRS 300 |
| L2_ABC79062.1                  | FISSPQRLTIFDNTYEGIDISLQFAHINTHEFDEAMDIIRHLFAITSRGGLVFRS 300 |
| L2_HPv13_YUC consensus         | .......................................................................................... |
| L2_CAA44653.1                  | IQRGSGMTSCHRIGGRVHFPDKDPSISAASEFELHLLVAAAQCHSLPDLYEADDPI 360 |
| L2_ABC79062.1                  | IQRGSGMTSCHRIGGRVHFPDKDPSISAASEFELHLLVAAAQCHSLPDLYEADDPI 360 |
| L2_HPv13_YUC consensus         | .......................................................................................... |
| L2_CAA44653.1                  | PVAVTSGSLSLASSTPEAQSSLSLSSAPWNTTVPLSPGDPFIPQPPGDPITFTAPTVPYN 420 |
| L2_ABC79062.1                  | PVAVTSGSLSLASSTPEAQSSLSLSSAPWNTTVPLSPGDPFIPQPPGDPITFTAPTVPYN 420 |
| L2_HPv13_YUC consensus         | .......................................................................................... |

**Fig. 3. Continued**
DNA fragments (Table 1). Each expected fragment with at least 20 nt overlapping with the fragment further one. The coverage was 98%, from the non-coding LCR to L1 gene. To prediction of oligonucleotide secondary structures (hairpins, self-dimers and heterodimers) were checked using OligoAnalyzer program (https://www.idtdna.com/calc/analyzer) [6], avoid their formation or minimal secondary structures. Further, non-specific hybridizations with the human genome were analysed using primer-blast tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast) [7].

2.3. Rolling circle amplification of HPV DNA

Multiply primed RCA was performed with the TempliPhi 100 amplification kit (Amersham Biosciences) according to the manufacturer’s instructions. This method is employed for the exponentially amplify of dsDNA HPV by rolling circle amplification (RCA), using random hexamers
and phi29 DNA polymerase (Fig. 1, step 1). First, 0.5 μL of total DNA from oral cells, or water (negative control), was transferred into a 0.5-mL tube with 5 μL of TempliPhi sample buffer, containing exonuclease-protected random hexamers. The sample was denatured for 3 min at 95 °C and afterwards immediately were placed on ice. In other tube, a premix was prepared on ice by mixing, 5 μL of TempliPhi reaction buffer (containing dNTPs and salts), 450 mM additional dNTPs and 0.2 μL of TempliPhi Enzyme Mix, containing the phi29 DNA polymerase and exonuclease-protected random hexamers in 50% glycerol. Afterwards, the cooled denatured sample was added to the premix and gently vortexed. The amplification solution was incubated overnight, approximately 16 h at 30 °C, followed by 10 min at 65 °C to inactivate the phi29 DNA polymerase, and stored at −20 °C until further analysis. Plasmid pUC19 was amplified by RCA as positive control.

A successful RCA was verified by a rare cutting restriction enzyme, BsmBI (New England Biolabs) according to described in [1] to linearize the amplified genomes. The digestion were visualized by 0.6% agarose gel electrophoresis in 1X TAE Buffer and ethidium bromide staining.

2.4. End-point PCR amplification

RCA product was used as template for amplifying the fragments from 1 to 23 with specific primer pairs for each region (Fig. 1, step 2). The PCR reactions were performed in final volume of 20 μL using 10x buffer reaction, 10 mM dNTPs, 50 mM de MgCl₂, 10 pM forward and reverse primers, 1 U Taq polymerase (Thermo Scientific), 10 ng RCA template DNA and water. The reactions were performed in a Mastercycler Ep Gradient thermocycler (Eppendorf) with an amplification profile of 94 °C for 9 min followed by 38 cycles at 94 °C for 1 min, annealing temperature for 1 min by each primer set and extension at 72 °C for 1 min; a final extension at 72 °C for 5 min. Additionally, PCR reactions were added as amplification controls: ultrapure water as template (negative control) to discard contamination. Amplification products were electrophoresed in a 1.2% agarose gel with a 100 pb DNA ladder (Invitrogen). Bands were visualized by staining with ethidium bromide on a Gel Doc XR system (Bio-rad) and picture processed with Image lab v.2.0.1 software. To purify the expected PCR products, ExoSAP-IT (USB, Cleveland) was used according to described in [1].

2.5. Genome sequencing, assembly, and gene annotation

All purified PCR products (fragments 1 to 23) were sequenced according to BigDyeTM terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) following the manufacturer’s instructions, and analyzed on a ABI PRISM™ 310 Genetic Analyzer (Applied Biosystems). Chromatograms (.ab1 format files) were analysed and the reads were trimmed and assembled using Phred, Phrap and Consed software package (v. 29.0) [8], using call scores and quality values (QV) defaults. Final editing of the sequences was performed manually, by inspection of the chromatograms, and consensus at each position of the genome. Coding sequences (CDS) were manually annotated considering the previously reported genomes (ID: DQ344807 and X62843). Final assembled sequence was manually checked with for quality, discarding gaps and missing regions; some nucleotides from the ends were low-quality and manually removed, thus obtaining a partial but almost complete sequence of 7831 bp, named for convenience of the annotation as “consed_hpv13_YUC”. The information is available from GeneBank (MT068446).

2.6. Comparisons among HPV13 proteins

To identify mutations in predicted proteins of HPV13_YUC, multiple sequence alignments were carried out among proteins previously reported from HPV13 genomes (DQ344807 and
X62843). For sequence alignments, we used ClustalW with default settings and translated alignment with Blosum 62 cost with Geneious software (v.6.1 Biomatters) [5]. Amino acids changes from alignments were shaded with BoxShade (https://embnet.vital-it.ch/software/BOX_form.html) (v. 3.21, written by K. Hofmann and M. Baron) using the parameters: RTF new as output format, consensus line with symbols, 0.5 as fraction on sequence and ALIN as input format. Alignments were exported in FASTA format with CLC sequence viewer 8 software (www.clcbio.com).

Ethics Statement

This protocol was reviewed and approved by the scientific and bioethical committee of the Universidad Autónoma de Yucatan (CEI-00001–2016). All participants' tutors signed informed consent.

CRediT Author Statement

Nuvia E. Kantún-Moreno, PhD. Writing-original draft preparation. Formal analysis, visualization, data curation. Gemaly E. Ek-Hernández, MSc. Investigation, visualization. José Reyes Canché-Pech, MSc. Methodology, investigation. Jesús G. Gómez-Carballo, MSc. Investigation, supervision, writing-review & editing. María del Refugio González-Losa, PhD. Resources, Writing-review & editing. Laura Conde-Ferráez, PhD. Conceptualization, supervision, project administration, visualization, writing-review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

This study was supported by Universidad Autónoma de Yucatan (internal registry CIRB2016–0006). This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.106955.

References

[1] L. Conde-Ferráez, G.E. Ek-Hernández, J.R. Canché-Pech, J.G. Gómez-Carballo, N. Kantún-Moreno, M.R. González-Losa, Genomic characterization of Human papillomavirus type 13, associated to Multifocal Epithelial Hyperplasia, in a Mayan community, Infect. Genet. Evol. (2020), doi:10.1016/j.meegid.2020.104595.
[2] M. Van Ranst, A. Fuse, P. Fiten, E. Beuken, H. Pfister, R.D. Burk, G. Opdenakker, Human papillomavirus type 13 and pygmy chimpanzee papillomavirus type 1: comparison of the genome organizations, Virology 190 (1992) 587–596, doi:10.1006/viro.1992.1089–w.
[3] C.M. Borborema-Santos, M.M. Castro, P.J. Santos, S. Talhari, S. Astolfi-Filho, Oral focal epithelial hyperplasia: report of five cases, Braz. Dent. J. 17 (2006) 79–82, doi:10.1590/s0103-64402006000100018.
[4] R.K. Saiki, T.L. Bugawan, G.T. Horn, K.B. Mullis, H.A. Erlich, Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes, Nature 324 (1986) 163–166, doi:10.1038/324163a0.

[5] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, et al., Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics 28 (2012) 1647–1649, doi:10.1093/bioinformatics/bts199.

[6] R. Owczarzy, A.V. Tataurov, Y. Wu, J.A. Manthey, K.A. McQuisten, H.G. Almabrazi, et al., IDT SciTools: a suite for analysis and design of nucleic acid oligomers, Nucleic Acids Res. 36 (2008) W163–W169, doi:10.1093/nar/gkn198.

[7] J. Ye, G. Coulouris, I. Zaretskaya, et al., Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinform. 13 (2012) 1–11, doi:10.1186/1471-2105-13-134.

[8] M. Machado, W.C. Magalhães, A. Sene, et al., Phred-Phrap package to analyses tools: a pipeline to facilitate population genetics re-sequencing studies, Investig. Genet. 2 (2011) 2–7, doi:10.1186/2041-2223-2-3.