Infection, Genetics and Evolution 34 (2015) 44–51

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

High-level of viral genomic diversity in cervical cancers: A Brazilian study on human papillomavirus type 16

Cristina Mendes de Oliveira [a,b,*], Ignacio G. Bravo [b], Nathália Caroline Santiago e Souza [a], Maria Luiza Nogueira Dias Genta [c], José Humberto Tavares Guerreiro Fregnani [d], Maricy Tacla [e], Jesus Paula Carvalho [c], Adhemar Longatto-Filho [f,g,h,i], José Eduardo Levi [a]

[a] Laboratório de Virologia, Instituto de Medicina Tropical, Universidade de São Paulo, São Paulo, Brazil
[b] Infections and Cancer Laboratory, Catalan Institute of Oncology, Barcelona, Spain
[c] Instituto do Câncer do Estado de São Paulo (ICESP), Faculdade de Medicina, São Paulo, Brazil
[d] Teaching and Research Institute, Barretos Cancer Hospital, Barretos, São Paulo, Brazil
[e] Gynecology Department, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil
[f] Molecular Oncology Research Center, Hospital de Câncer de Barretos, Barretos, São Paulo, Brazil
[g] Laboratory of Medical Investigation (LIM) 14, Faculty of Medicine, University of São Paulo, São Paulo, Brazil
[h] Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal
[i] ICVS/3B's – PT Government Associate Laboratory, Braga/Guimarães, Portugal

ARTICLE INFO

Article history:
Received 25 March 2015
Received in revised form 2 July 2015
Accepted 4 July 2015
Available online 6 July 2015

Keywords:
HPV16 diversity
Intralesion
Cervical cancer

ABSTRACT

Invasive cervical cancer (ICC) is the third most frequent cancer among women worldwide and is associated with persistent infection by carcinogenic human papillomaviruses (HPVs). The combination of large populations of viral progeny and decades of sustained infection may allow for the generation of intra-patient diversity, in spite of the assumedly low mutation rates of PVs. While the natural history of chronic HPVs infections has been comprehensively described, within-host viral diversity remains largely unexplored.

In this study we have applied next generation sequencing to the analysis of intra-host genetic diversity in ten ICC and one condyloma cases associated to single HPV16 infection. We retrieved from all cases near full-length genomic sequences.

All samples analyzed contained polymorphic sites, ranging from 3 to 125 polymorphic positions per genome, and the median probability of a viral genome picked at random to be identical to the consensus sequence in the lesion was only 40%. We have also identified two independent putative duplication events in two samples, spanning the L2 and the L1 gene, respectively. Finally, we have identified with good support a chimera of human and viral DNA.

We propose that viral diversity generated during HPVs chronic infection may be fueled by innate and adaptive immune pressures. Further research will be needed to understand the dynamics of viral DNA variability, differentially in benign and malignant lesions, as well as in tissues with differential intensity of immune surveillance. Finally, the impact of intralesion viral diversity on the long-term oncogenic potential may deserve closer attention.

1. Introduction

Papillomaviruses (PVs) are a family of small non-enveloped viruses, with a circular double-stranded DNA genome, that infect epithelia in a wide range of vertebrates and probably all amniotes (Bravo et al., 2010). Human PVs (HPVs) are the most studied members in the family, because of their medical importance. They are related to benign proliferative disorders, such as skin warts, epidermodysplasias or *condylomata acuminata*, and also to malignant neoplasms of the cervix, anus, vagina, vulva, penis and oropharynx (Cubie, 2013). To date, more than 180 HPV types have been fully sequenced (http://pave.niaid.nih.gov/). HPVs are not monophyletic and belong instead into five genera, with different types showing differential association with different diseases (Bernard et al., 2010; Bravo et al., 2010; de Villiers, 2013).

* Corresponding author at: Laboratório de Virologia, LIM 52, Instituto de Medicina Tropical da Universidade de São Paulo, Rua Dr Enéas de Carvalho Aguiar 470, 2o andar, São Paulo, SP CEP 05403-000, Brazil.
E-mail address: cristina.oliveira1@gmail.com (C.M.d. Oliveira).

http://dx.doi.org/10.1016/j.meegid.2015.07.002
1567-1348/

Infection by most HPVs is asymptomatic (Antonsson et al., 2000), but 13 HPV genotypes are classified by the International Agency for Research on Cancer as oncogenic: HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66 (Bouvard et al., 2009). Carcinogenic potential is not evenly distributed among HPVs, and HPV16 alone is responsible for more than 50% of the invasive cervical cancer (ICC) cases worldwide (de Sanjose et al., 2010; Li et al., 2011), and is the most prevalent HPV in other squamous carcinomas associated to HPVs infections (De Vuyst et al., 2009; de Sanjose et al., 2014), as well as in women with normal cytology (Bruni et al., 2010). In South-East Brazil, 2.3% of women on the screening population were positive for HPV16 (Levi et al., 2014), whereas 75% of the ICC cases from the same region were associated with this HPV (de Oliveira et al., 2013). Genetic diversity among HPV16 isolates can reach 2%, and the trend is towards full-genome sequencing of the isolates (Chen et al., 2005, 2009, 2013; Smith et al., 2011).

For rapidly evolving viruses such as HCV or HIV, a wealth of data has been generated during follow-up studies to address the question of the amount and generation pace of intra-host genetic diversity. For PVs infections, however, the common view about these viruses has precluded experimental research about mutation and evolutionary rates. Most studies have addressed changes in the most prevalent, "consensus", sequence during chronic PVs infections based on Sanger sequencing. This technology may however not be appropriate to capture the dynamics of slowly evolving viruses, such as PVs. Very few reports have applied deep sequencing technologies to gain insight into the accumulation of genetic diversity during the course of a PV infection. To help bridging this gap, the aim of the present study is to analyze by means of next-generation sequencing a number of ICC cases associated to HPV16 single infections.

2. Materials and methods

2.1. Clinical specimens and total DNA extraction

Ten invasive cervical cancers containing HPV16 DNA as single infection, determined after Linear Array HPV Genotyping (Roche Molecular Diagnostics, Pleasanton, USA) or PapilloCheck (Greiner Bio-One GmbH, Frickenhausen, Germany) tests were selected for the present study. The ICC cases originated from women participating in a previous study (de Oliveira et al., 2013), and the condyloma originated from a women attended at the Gynecological Department from Hospital das Clínicas – Universidade de São Paulo. The study was approved by the Brazilian National Research Ethical Committee (CONEP n°158.490).

Total DNA was extracted from fresh tissue using QIAamp DNA Mini Kit (Qiagen, Gaithersburg, USA) or NucleoSpin Tissue kit (Macherey–Nagel GmbH&Co, Germany), according to manufacturer's instructions.

2.2. HPV16 E6 and E2 real time PCR

To determine the HPV16 physical status a multiplex real-time PCR for simultaneous detection of the E6 and E2 regions were used. Primers and probes were modified from (Canadas et al., 2010), as follows: primer E6F 5′ACCGGTTAGTATAAAAGCAGACATTTTMT3′, primer E6R 5′GCTCCTGTGGGTCCTGRAAC3′, probe E6 CAL Fluor Orange 560 BHQplus – CACCAAAAGAGAACTGCAA – BHQ1, primer E2F 5′GCRACGAAGTATCCTCTSCTGAA3′, primer E2R 5′AAGGCGAC GGCTTTGGKAT3′ and probe E2 FAM – TAYTAGGCAGCACTTGGCCA ACCA – BHQ1. Real-time PCR was performed in a final volume of 25 μL, containing 5 μL of DNA, 1x qPCR Low Mg^{2+} (Next Generation ECO HotStart qPCR kit), 200 nM of each primer and probe. PCR was carried out according to the protocol: for 95 °C for 2 min, 45 cycles of 95 °C for 15 s, 43 °C for 30 s 60 °C for 45 s, in a 7500 FAST Real Time Systems (Applied Biosystems). In each run, water was included as negative control, an HPV16 complete genome plasmid as episomal control, and SiHa DNA as HPV16 DNA integrated control. Physical state was determined by comparing the CtE6/CtE2 ratio of each sample with the ratio observed in the plasmid control. Samples negative for E2 region were considered as integrated, samples with the same CtE6/CtE2 ratio of the plasmid as pure episomal form and samples with a CtE6/CtE2 ratio lower than the plasmid as mixed (both forms episomal and integrated).

2.3. HPV16-specific long PCR

PCR was performed in 50 μL consisting of 1x long PCR buffer (Fermentas GmbH&Co, Germany), 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 μM of each primer, 1.25 U of long PCR enzyme mix (Fermentas GmbH&Co, Germany) and 100–300 ng of template DNA, generating an amplicon of 7935pb. The primers used (forward-21 5′AAACTAAGGGCGTAACCGA3′ and reverse-7862 5′C GGTTTGCACACACCCATGT3′) were previously described (Sun et al., 2012). Positions corresponding to 5′-end nucleotide of primers in the HPV16 reference K02718 were annotated following primer orientation.

PCR cycling conditions were: 94 °C for 2 min, 10 cycles of (95 °C for 15 s, 60 °C for 30 s and 68 °C for 7 min), followed by 25 cycles of (95 °C for 15 s, 60 °C for 30 s and 68 °C for 7 min with an increment of 5 s by cycle), plus a final extension of 68 °C for 10 min.

Long PCR products were separated by electrophoresis and the approximately 8Kbp band was purified using Invisorb® Fragment Clean Up (Invitek, Berlin, Germany), according to manufacturer's instructions. The purified PCR product was then submitted to deep sequencing on the Ion Torrent platform (Life technologies, CA, USA).

2.4. Deep sequencing

Libraries were generated using 100 ng of the long PCR product and the Ion Shear™ fragmentation method (Life technologies, CA, USA). Adapters with barcode allowing for multiplexing were ligated and purified using the E-Gel® SizeSelected™ system (Life technologies, CA, USA). Samples were quantified using Ion Library Quantification Kit (Life technologies, CA, USA) and 1–5 libraries from different samples were multiplexed. Samples were submitted to emulsion PCR using the Ion PGM™ Template 200 kit (Life technologies, CA, USA) and finally to sequencing reaction using Ion PGM™ 200 Sequencing kit (Life technologies, CA, USA). Sequencing was performed in the Ion 314™ chip (Life technologies, CA, USA) according to the manufacturer's instructions.

2.5. Deep sequence data analyses

Fastq files were extracted from the Ion Torrent *SFF* files, clipped and submitted to quality control analyses using FastQC High Throughput Sequence QC Report software v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), for trimming according to phred scores. Reads were mapped against the HPV16 reference sequence K02718 using the TMAP v0.2.3 software. The mapping results were visualized with IGV viewer (Thorvaldsdottir et al., 2013). Consensus fasta sequences were generated using SAMtools v0.1.19 (Li et al., 2009).

2.6. Read depth analysis

A slide-window analysis was performed to verify differences in the read depth coverage between and/or within the samples, using