Complete genomes of two *Human papillomavirus 11* isolates recovered from inverted sinonasal papillomas in humans

Sarah Bouzidi[a], Julien Puech[b,c], Marta Fulla[d], Xavier González-Compta[d], Hélène Pere[b,c], Laia Alemany[e,f,g], David Veyer[b,c] and Ignacio G. Bravo[a#]

Affiliations
a. Laboratory MIVEGEC (Univ Montpellier, CNRS, IRD), French National Center for Scientific Research (CNRS), Montpellier, France.
b. Laboratoire de Virologie, Service de Microbiologie, hôpital européen Georges Pompidou, Assistance Publique-Hôpitaux de Paris, Paris, France.
c. Unité de Génomique Fonctionnelle des Tumeurs Solides, Centre de Recherche des Cordeliers, INSERM, Université Paris, Paris, France.
d. Otorhinolaryngology Department, Hospital Universitari de Bellvitge, L'Hospitalet de Llobregat, Spain.
e. Unit of Molecular Epidemiology and Genetics (UNIC EMG), Cancer Epidemiology Research Program, Catalan Institute of Oncology, ICO, L'Hospitalet, Barcelona, Spain.
f. Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet, Barcelona, Spain.
g. CIBER en Epidemiología y Salud Pública (CIBERESP), Madrid, Spain.

Running title
HPV11 genomes from inverted sinonasal carcinomas

Corresponding author's email address
Address correspondence to Ignacio G. Bravo, Ignacio.bravo@cnrs.fr

Abstract
We present here two complete *Human papillomavirus 11* (HPV11) genomes isolated from one transitional and from one squamous inverted sinonasal papillomas in humans. Both genomes belong to the HPV11_A2 sublineage. These are the first HPV genomes isolated from this rare proliferative disease.

Announcement

Inverted sinonasal papillomas (ISP) are rare proliferative diseases of the nasal fossae and of the ethmoidal and maxillary sinuses, commonly associated to occupational exposures to organic solvents and to chronic inflammation (1). The role of Human papillomaviruses (HPVs) as etiological agents of ISP and/or the interpretation of HPV detection for the clinical prognosis of the disease are not well characterized yet (2, 3). In a large clinical study (4), only 2 out 46 ISP tested positive for HPVs. Here we have characterised the viral genome component in these two ISP samples. These are the first HPV genomes isolated from ISPs.

DNA extracts were obtained from formalin-fixed, paraffin-embedded samples as described (4). Protocols were approved by the ethics committee of the Catalan Institute of Oncology, which required no informed consent to use archived samples (4). One sample originated from a woman living in Spain, diagnosed in 2009 of a transitional inverted sinonasal carcinoma, and the other sample originated from a man living in Spain, diagnosed in 2011 of a squamous inverted sinonasal carcinoma. The possible HPV genomic material present in the extracted DNA was captured using probes covering all HPV genotypes identified by the International Papillomavirus Reference Center (https://www.hpvcenter.se/human_reference_clones/), using SeqCap EZ library reagents (Roche NimbleGen) (5). An Illumina MiSeq system was used to sequence post-capture libraries (150 bp paired-end reads). Quality trimming and adapter clipping were performed on the raw reads using Trimmomatic v0.38 (6). Full viral genome sequences were assembled and annotated with breseq v0.38.1 (7) by aligning viable reads against the HPV11 GenBank M14119 reference genome, and incorporating polymorphic sites showing a frequency above 0.5, including indels. For phylogenetic analysis, we aligned first all available HPV11 full-genome GenBank entries using MAFFT v7.505 (8), and calculated a HPV11 reference tree at the nucleotide level with RAxML-NG v1.1.0 (9), under the GTR+G+I evolutionary model. We added then the novel sequences to the pre-computed alignment using MAFFT v7.505 (8) and applied the evolutionary placement algorithm of RAxML v8.2.12 (10) to place the novel sequences in the pre-computed HPV11 reference tree. Genetic distance was evaluated with dist.dna, in the ape R package (11), using the TN93 substitution model.

The annotated HPV11 genomes are available at the GenBank under the OR669303 and OR669304 accession numbers. The circular viral genomes are both 7933bp and have a G+C content of 41.1%. The viral genomes were complete and we could close the circular dsDNA molecules in both cases, with an average read depth of 4,216 and 6,987 respectively. They contained the complete *E6*, *E7 E1 E2*, *E5_γ*, *E5_δ*, *L2* and *L1* ORFs, as well as the complete upstream regulatory region. The two viral genomes were identical in the *L1* ORF, and differed in 5 positions across the full genome, corresponding to a divergence of $6.3\ 10^{-4}$ substitutions per site. Both sequences were placed in the extant HPV11 genetic diversity, and unequivocally assigned to lineage A, sub-lineage A2, which is the most common HPV11 lineage (12).

Data availability

Accession number	Bioproject	Biosample	SRA(SRR acc.number)
OR669303	PRJNA1039588	SAMN38222584	SRR26801485
OR669304	PRJNA1039588	SAMN38222583	SRR26801484

References

1. Bishop JA. 2017. OSPs and ESPs and ISPs, Oh My! An Update on Sinonasal (Schneiderian) Papillomas. Head Neck Pathol 11:269–277.
2. Lawson W, Schlecht NF, Brandwein-Gensler M. 2008. The role of the human papillomavirus in the pathogenesis of Schneiderian inverted papillomas: an analytic overview of the evidence. Head Neck Pathol 2:49–59.
3. Syrjänen K, Syrjänen S. 2013. Detection of human papillomavirus in sinonasal papillomas: systematic review and meta-analysis. Laryngoscope 123:181–192.
4. Fulla M, Szafarowski T, Frias-Gomez J, Quiros B, Clavero O, Gomà M, Pavon MA, Jurek-Matusiak O, Lares HR, Mañós M, Alemany L, Mena M, Gonzalez X. 2020. Human Papillomavirus and Factors Associated with Recurrence in Sinonasal Inverted Papillomas from Poland and Spain. Head Neck Pathol 14:758–767.
5. Holmes A, Lameiras S, Jeannot E, Marie Y, Castera L, Sastre-Garau X, Nicolas A. 2016. Mechanistic signatures of HPV insertions in cervical carcinomas. npj Genomic Med 1:16004.
6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.
7. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151:165–188.
8. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30:772–780.
9. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455.
10. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
11. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.
12. Jelen MM, Chen Z, Kocjan BJ, Hošnjak L, Burt FJ, Chan PKS, Chouhy D, Combrinck CE, Estrade C, Fiander A, Garland SM, Giri AA, González JV, Gröning A, Hibbitts S, Luk TNM, Marinic K, Matsukura T, Neumann A, Oštrbenk A, Picconi MA, Sagadin M, Sahli R, Seedat RY, Seme K, Severini A, Sinchi JL, Smahelova J, Tabrizi SN, Tachezy R, Tohme Faybush S, Uloza V, Uloziene I, Wong YW, Židovec Lepej S, Burk RD, Poljak M. 2016. Global Genomic Diversity of Human Papillomavirus 11 Based on 433 Isolates and 78 Complete Genome Sequences. J Virol 90:5503–5513.