Does human papillomavirus-negative condylomata exist?

Laila Sara Arroyo Mühr a, Davit Bzhalava a, Camilla Lahgeden a, Carina Eklund a, Hanna Johansson b, Ola Forslund b, Joakim Dillner a, c, Emilie Hultin a

a Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden
b Department of Laboratory Medicine, Lund University, Malmö, Sweden
c Department of Medical Epidemiology & Biostatistics, Karolinska Institutet, Stockholm, Sweden

A R T I C L E  I N F O

Article history:
Received 29 May 2015
Returned to author for revisions 3 July 2015
Accepted 30 July 2015
Available online 28 August 2015

Keywords:
Human papillomavirus
Metagenomic sequencing
Condylomata

A B S T R A C T

Condylomata acuminata is caused by human papillomavirus (HPV). PCR with consensus primers will typically detect HPV in >96% of condylomata. Metagenomic sequencing has found that some “HPV-negative” condylomata do indeed contain HPV. We wished to perform a renewed evaluation of the “HPV-negative” condylomata using deeper metagenomics sequencing. Sequencing of whole genome amplified DNA from 40 apparently “HPV-negative” condylomata detected HPV in 37/40 specimens. We found 75 different HPV types, out of which 43 represented novel putative HPV types. Three types were cloned and established as HPV types 200, 201 and 202. Molluscum contagiosum virus was detected in 24 of the 40 samples.

In summary, deep sequencing enables detection of HPV in almost all condylomata. “HPV-negative” condylomata might largely be explained by clinical misdiagnosis or the presence of viral variants, distantly related HPV types and/or low viral loads.

© 2015 Elsevier Inc. All rights reserved.

Introduction

Human papillomaviruses (HPVs) are a large group of dsDNA-viruses that infect keratinocytes of the skin and mucosa. About 200 different human papillomavirus (HPVs) types have been completely cloned, sequenced and given an official number by the International HPV Reference Center (www.hpvcenter.se, accessed on 2015-05-28), which moved to Karolinska Institutet from Heidelberg in 2013 (Bzhalava et al., 2015). In addition to the established types, a large number of subgenomic HPV sequences representing putatively novel HPV types have been discovered using the broad general primer PCR system FAP (FA-isolates) (Forslund, 2007), CUT PCR (GC-isolates) (Chouhy et al., 2010) and/or using deep sequencing (SE-isolates) (Ekstrom et al., 2011).

Condylomata acuminata, anogenital warts, is a very common sexually transmitted disease caused by HPV infection, most commonly by HPV6 of the Alphapapillomavirus genus (Bernard et al., 2010; Gissmann et al., 1982). A small proportion of condylomata are reported to be “HPV-negative” and, besides true negativity, 3 possible explanations have been been suggested (Walboomers and Meijer, 1997): (i) specimen inadequacy, (ii) detection method insensitivity, and (iii) the existence of novel HPV types that are not detectable by the method.

We have previously pointed out that extended analysis of HPV-associated clinical lesions, such as cervical cancer or condyloma, that are seemingly “HPV-negative” is a promising strategy to discover new HPVs (Ekstrom et al., 2010). In southern Sweden, a systematic condyloma reporting system has performed HPV genotyping of the largest series of condylomata reported to date in the literature (Sturegard et al., 2013). Overall, 96.3% of condylomata were found to be HPV-positive, but a few were seemingly HPV-negative in analyses even with broad HPV-primer PCR systems (Soderlund-Strand et al., 2009).

Metagenomic sequencing can be used for an increased sensitivity in HPV detection (Arroyo et al., 2013; Ekstrom et al., 2011) as well as to obtain a comprehensive and more unbiased identification of the DNA present in samples, as no prior PCR or any prior knowledge of the sequence is required (Bzhalava et al., 2014).

The aim of this study was to apply modern deep sequencing technology, targeting individual “HPV-negative” condylomata, in order to obtain a sensitive virus detection, also of putatively novel HPVs.

Results

The DNA from the 40 condylomata specimens together with 2 negative controls (lab-grade water) were amplified using WGA,
labeled with index tags and sequenced using Illumina deep sequencing. Sequencing results from Illumina MiSeq were used to evaluate DNA library quality, before sequencing them on Illumina NextSeq500. MiSeq sequencing provided about 365 Mb sequencing depth per sample with a read length of 300 bp paired-end, whereas NextSeq sequencing provided about 3.33 Gb sequencing depth per sample with a read length of 150 bp paired-end.

Sequencing data from NextSeq consisted of 753,923,432 reads whereof 363,182 were viral reads (0.05% of total sequences). The majority of the viral reads, 329,636 reads, were related to HPV sequences (91%). HPV sequences (with a cutoff of at least 5 reads per HPV type) were found in 37/40 specimens (92.5%), representing a total of 75 different HPV types or putative types within the genera Parvoviridae, Iridoviridae, Herpesviridae, Baculoviridae, Mimiviridae, Molluscum contagiosum virus is shown in Table 1. HPV6 and HPV180 were the most frequent genotypes, detected in 30 and 18 samples respectively. 13 established HPV types were detected in between 2 to 10 samples, whereas 12 established HPV types were only present in single samples (Table 1). Only 3 samples were negative for any known or putative HPV type. Molluscum contagiosum virus was detected with a total of 27,538 reads, corresponding to 7.6% of all viral reads. Molluscum positivity were detected in 24/40 samples (Table 1). 22 out of those 24 were also positive for HPV sequences. The rest of the virus related positivity were detected in 24/40 samples (Table 1). 22 out of those 24 samples had more than 10,000 HPV reads. 7 samples had more than 10,000 HPV reads (Table 2). Based only on established HPV types, 31/40 samples were positive for multiple HPV types, being positive for up to 8 different established types.

DNA sequences of novel putative HPV types

Overall, 46 novel HPV DNA sequences belonging to Beta and Gamma genera were detected from the deep sequencing of the 40 condyloma specimens in this study. Six of the novel putative HPV types represented complete HPV genomes (GenBank accession number KP692114 to KP692119). Three of these were cloned (isolates “SE370”, “SE371” and “SE372”). Clones and sequences were sent to the International HPV Reference Center where they were re-sequenced and established as new HPV types with the designation HPV200, HPV201 and HPV202 (for “SE370”, “SE371” and “SE372”, respectively) and reported in Table 1 as established HPV types. They belonged to the genus Gamma-papillomavirus, but with only 79%, 68% and 82% similarity to the most closely related type (Table 3). HPV200 did not have any ORF corresponding to the E4 gene. The sequences from the remaining 43 novel putative types are deposited in GenBank, accession numbers KP692120 to KP692178. Raw unassembled HPV related reads were deposited in the SRA database with BioProject ID: 278583.

Comparison with previous 454 sequencing of the same specimens

A previous study sequenced these 40 apparently HPV-negative condyloma specimen in pools of 4 samples with 454 pyrosequencing on a GS Junior instrument (Roche) (12). HPV sequences were detected in 5/10 pools from a total of 35 different HPV types: 13 now established HPV types (HPV 5, 6, 50, 57, 58, 66, 105, 124, 130, 150, 165 (former isolate K7), 175 (former isolate SE87) and 180 (former isolate FA69)), 1 known putative HPV sequence (SE8) as well as 21 novel putative HPV sequences all belonging to the Gamma genus. Read lengths of singletons and contigs ranged from 89 bp to 7356 bp and samples were considered positive as long as 1 read was detected. Among the 13 established HPV types detected, only 5 (HPV6, 50, 165, 175 and 180) generated contigs > 700 bp.

The deeper sequencing performed in this study confirmed positivity (with a cutoff at > 5 reads per sample) for 11 of the 13 established
Nu papillomaviruses are presented in red, green, blue, orange and purple colors, respectively.

Discussion

Overall, we found 75 different HPV types in the 40 “HPV-negative” condyloma samples (27 established HPV types, 5 known putative HPV types and 43 novel putative types). Several samples were found to contain established HPV types known to cause condylomatata, in particular HPV6, although the PCR methods used should have detected these viruses (Aubin et al., 2008; de Villiers et al., 2004; Sturegard et al., 2013). Possible explanations include the presence of viral variants with genomic substitutions in the sequences targeted by PCR primers or probes. Deep sequencing overcomes this problem by detecting HPV types without any prior sequence information, independent of PCR.

In this study, we compared data from modern deep sequencing using Illumina with previous data using 454 sequencing. As expected, more HPV types were detected with the deeper sequencing. It also generated longer contigs of HPV sequences. A total of 46 putative novel types were detected, belonging to either Beta or Gamma genera. Beta- and gamma-HPVs are frequently detected on healthy skin (Antonsson et al., 2003; Ekstrom et al., 2013; Forslund et al., 2004; Fouloungne et al., 2012) and therefore, presence of these types might be due to biological contamination from adjacent healthy genital skin. Interestingly, the Merkel cell polyomavirus which is known to be commonly present in healthy skin was not detected.

Clinical distinction between condyloma and mollusca is not an easy task. Presence of molluscum contagiosum virus is also known to exist concomitantly with presence of HPV (Castronovo et al., 2012). In our study, molluscum contagiosum was detected in 24/40 condyloma samples and 22 out of those 24 also contained HPV.
sequences. This leaves only a single condyloma specimen where we did not find neither HPV, nor molluscum contagious virus. The detection of multiple viruses in the diseased tissue does not imply causality, as this study did not analyze matched controls.

In summary, deep sequencing of apparently “HPV-negative” condylomata specimen greatly reduces the number of “HPV-negative” specimens. Sequencing of apparently HPV-negative specimens from diseases known to be HPV-associated continues to identify a large number of previously unknown HPV sequences.

**Importance**

Conventional methods fail to detect papillomaviruses (HPV) in a small proportion of genital warts (condylomata), a disease established as caused by HPV. Modern deep sequencing is not based on prior knowledge of sequence information, can increase the analytic sensitivity of HPV detection as well as detect viral variants and distantly related viruses. We sequenced 40 apparently HPV-negative genital warts and observed that 37/40 samples did contain HPV, including 43 previously unknown apparently new HPV types. This suggests that seemingly HPV-negative condylomata might be explained by e.g., clinical misdiagnosis or presence of viral variants, low viral loads or presence of previously unknown HPV types.

**Material and methods**

**Patients**

40 swab samples of apparently HPV-negative condylomata from 21 women and 19 men (Johansson et al., 2013) were analyzed with Illumina sequencing technology. Swab samples were collected with a pre-wetted (0.9% NaCl) cotton-tipped swab rolled over the condyloma and stored in 1 mL saline. Samples were then centrifuged (5 min, 3500 × g) and 500 μL of the supernatant were left for pellet re-suspension. 200 μL were used for DNA extraction with MagNA Pure LC using the Total Nucleic Acid Kit (Roche) and DNA was eluted in 100 μL elution buffer.

Informed consent was obtained from all participants, the study adhered to the declaration of Helsinki and was approved by the Ethical Review Committee of Lund University (Sweden).

**Sample preparation**

Sample adequacy was assayed using beta-globin PCR, visualized by gel electrophoresis in 11 cases and by real-time PCR in 29 cases (Sturegard et al., 2013). In order to amplify the DNA in the sample, extracted DNA from 40 individual samples and 2 negative controls (water) were subjected to whole genome amplification (WGA) using GenomiPhi High Yield (GE Health Care, United Kingdom). WGA was performed according to manufacturer’s instructions with some modifications. 5 μL of sample were diluted with 20 μL sample buffer. The mix was incubated at 95 °C for 3 min and then cooled on ice. 22.5 μL of reaction buffer was mixed with 2.5 μL of enzyme mix and added to the diluted sample. Samples were incubated at 30 °C for 7 h and inactivated at 65 °C for 10 min. Products were dissolved by diluting samples 1:2 in PCR-Grade water. All amplification products were quantified with QuantiFluor-ST (Promega, US) and a quality analysis (Bioanalyzer, Agilent) was performed to check the DNA fragments size distribution prior library preparation.

**Deep sequencing**

For each sample, 50 ng of WGA DNA were tagmented and amplified using the Nextera DNA Sample Preparation kit according to the user guide revision B (Illumina). Two sample tags, a unique combination for each sample, were used in the short PCR amplification step to facilitate multiplex sequencing. All individual libraries were validated, normalized to 4 nM and pooled.

Prior to sequencing, the 42 individual indexed libraries (40 specimens and 2 negative controls) were denatured and diluted, resulting in a 15 pM DNA solution. PhiX control was spiked at 1% and the pool was sequenced by paired-end 301 + 301 cycles using MiSeq version 3 reagent kit (Illumina, US). The sequencing preparations were made according to the user guides Preparing Libraries for Sequencing on the MiSeq revision C, Reagent Preparation Guide revision A and MiSeq System User Guide revision K. For an even deeper sequencing, the 42 individual indexed libraries were sequenced using the NextSeq500 instrument. The pool was denatured and diluted, resulting in a 2.6 pM DNA solution. PhiX control was spiked at 1% and the pool was sequenced by paired-end 151 + 151 cycles using NextSeq 500 High Output reagent kit (Illumina, US). The sequencing preparations were made according to the user guides Denaturing and Diluting

**Table 3**

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>E6</th>
<th>E7</th>
<th>E1</th>
<th>E2</th>
<th>E4</th>
<th>L2</th>
<th>L1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HPV200 (SE370)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of nucleotides (nt)a</td>
<td>7137</td>
<td>420</td>
<td>282</td>
<td>1797</td>
<td>1206</td>
<td>x</td>
<td>1509</td>
<td>1545</td>
</tr>
<tr>
<td>Number of amino acids (aa)</td>
<td>-</td>
<td>140</td>
<td>94</td>
<td>599</td>
<td>402</td>
<td>x</td>
<td>503</td>
<td>515</td>
</tr>
<tr>
<td>HPV200 vs HPV48 Gamma-2% identity on nt level (aa level)</td>
<td>-</td>
<td>78 (72)</td>
<td>81 (84)</td>
<td>79 (80)</td>
<td>82 (79)</td>
<td>x</td>
<td>77 (73)</td>
<td>79 (86)</td>
</tr>
<tr>
<td><strong>HPV201 (SE371)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of nucleotides (nt)a</td>
<td>7299</td>
<td>423</td>
<td>297</td>
<td>1827</td>
<td>1257</td>
<td>549</td>
<td>1554</td>
<td>1527</td>
</tr>
<tr>
<td>Number of amino acids (aa)</td>
<td>-</td>
<td>141</td>
<td>99</td>
<td>609</td>
<td>419</td>
<td>183</td>
<td>518</td>
<td>509</td>
</tr>
<tr>
<td>HPV201 vs HPV163 Gamma-20% identity on nt level (aa level)</td>
<td>-</td>
<td>62 (47)</td>
<td>64 (50)</td>
<td>66 (51)</td>
<td>67 (54)</td>
<td>-</td>
<td>65 (55)</td>
<td>68 (68)</td>
</tr>
<tr>
<td><strong>HPV202 (SE372)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of nucleotides (nt)a</td>
<td>7344</td>
<td>432</td>
<td>300</td>
<td>1818</td>
<td>1179</td>
<td>369</td>
<td>1497</td>
<td>1587</td>
</tr>
<tr>
<td>Number of amino acids (aa)</td>
<td>-</td>
<td>144</td>
<td>100</td>
<td>606</td>
<td>393</td>
<td>123</td>
<td>499</td>
<td>529</td>
</tr>
<tr>
<td>HPV202 vs HPV140 Gamma-11% identity on nt level (aa level)</td>
<td>-</td>
<td>87 (89)</td>
<td>86 (90)</td>
<td>87 (90)</td>
<td>88 (86)</td>
<td>90 (86)</td>
<td>77 (79)</td>
<td>82 (92)</td>
</tr>
</tbody>
</table>

*a Including STOP codon.*
Bioinformatic analysis

Sequences obtained from MiSeq and NextSeq platforms and the indexing tags, included in the Illumina primers, were used to assign the sequence reads to the originating samples. The bioinformatic analysis started with quality checking where sequences were trimmed according to their Phred quality scores (Ewing and Green, 1998). Quality checked reads were then screened against the human reference genome hg19, as well as bacterial, phage and vector sequences downloaded from GenBank using BWA-MEM (Li the human reference genome hg19, as well as bacterial, phage and vector sequences were trimmed according to their Phred quality scores (Ewing and Green, 1998). Quality checked reads were then screened against the human reference genome hg19, as well as bacterial, phage and vector sequences downloaded from GenBank using BWA-MEM (Li and Durbin, 2010) and reads with > 95% identity over 75% of their length to human DNA were removed from further analysis. The rest of the sequences were normalized (http://ged.msu.edu/papers/2012-diginorm) to discard redundant data and reduce sampling variation and sequencing errors. The normalized dataset was then processed for assembly using the Trinity (Haas et al., 2013), SOAPdenovo-Trans (http://soap.genomics.org.cn/) and IDBA-UD (Peng et al., 2012) assemblers into contiguous sequences (contigs). Reads before assembly were re-mapped to assembled contigs. The use of several assembly algorithms and re-mapping of all singleton reads to assembled contigs was used to validate assembly results (Bzhalava et al., 2013; Meiring et al., 2012). Assembled contigs were then subjected to taxonomic classification using GenBank and paracel blast (www.paracel.com) to classify them as i) previously known sequences, ii) related to previously known sequences, or iii) unrelated to any previously known sequences.

All papillomavirus-related contigs and singletons were checked as described to identify possible artificial “chimeric” sequences (contigs containing sequences originating from different viruses) (Bzhalava et al., 2013). Shortly, the sequence that aligned to its most closely related sequence in GenBank was divided into three equal segments. If at least one of the segments differed in similarity to the corresponding overlapping parts with more than 5% (for example if segment 1 was 88% similar and segment 2 was 94% similar) the sequence was considered as a “possible chimera” and not analyzed further.

Papillomavirus-related contigs having ≥ 90% identity with each other were clustered. The total number of sequencing reads from each cluster was calculated and the longest contig was selected as a representative to remove redundancy. Clustering analysis was conducted separately for contigs longer and shorter than 700 bp length. Based on the percent similarity to the closest related type in GenBank, HPV related contigs were classified as known established, known putative and novel putative HPV types. For each of the HPV types, the total number of reads was calculated and specimens were considered positive if at least 5 reads from the same HPV type were detected in the same specimen.

All analyses were performed using in-house R (www.R-project.org/), python (www.python.org) and bash (http://www.gnu.org/software/bash/) scripts that ran on a high performance (40 core, 2 TB RAM) Linux server.

Characterization of HPV200, 201 and 202

The complete genome sequence was obtained for 6 previously unknown HPV sequences, putatively representing new HPV types. 3 of the 6 complete putative novel HPV genomes, “SE370,” “SE371” and “SE372” isolates, were amplified and cloned from their respective specimens in 2 overlapping fragments each. The fragments were amplified using PrimeSTAR GXL DNA Polymerase from TAKARA, according to manufacturer’s instructions. The first fragment containing the L1 region was amplified using the primers SE370AF 5’-TGCAGCAGACATGATCCTG-3’ and SE370AR 5’-ACCTTCGCCTTCCTTGT-3’ and SE371AF 5’-TCCACAGTGAGGCACTTGT-3’ and SE371AR 5’-GACCATGCACAGTGCTC-3’ and SE372AF 5’-ACATGGCCCTATACCACGG-3’ and SE372AR 5’-TTGAGATTGACACAGCGCC-3’ for “SE370”, “SE371” and “SE372” isolates, respectively. The second fragment was amplified using the primers SE370BF 5’-GACGACAGCAGAAGGGTTGT-3’ and SE370BR 5’-CTAGGACCAGGACACGGC-3’ and SE371BF 5’-GGTTTCGGTTGACAGATTG-3’ and SE371BR 5’-CTAAACGCGCACTAAGG-3’ and SE372BF 5’-TCCAGCTG- CAGGTTTTGTGT-3’ and SE372BR 5’-AAACGCCAAACACCTCTC-3’ for “SE370”, “SE371” and “SE372” isolates respectively. The PCR program for all fragments consisted of 30 cycles at 98 °C for 10 s, 55 °C for 15 s and 68 °C for 8 min. The PCR products were gel-purified and cloned using the Zero Blunt® TOPO® PCR Cloning kit (Invitrogen) and the pCRTM-Blunt II-TOPO® vector, according to manufacturer’s instructions and incubated overnight.

The complete cloned genomes for the 3 novel HPV types were also sequenced by Sanger sequencing to confirm the sequence obtained with the MiSeq instrument from each index sample, using designed primers for conventional primer walking. Clones were sent to the International HPV Reference Center (Karolinska Institutet, Sweden) that confirmed the DNA sequence and assigned the submitted clones the novel HPV type number 200, 201 and 202 for “SE370”, “SE371” and “SE372”, respectively.

Acknowledgments

Supported by the Swedish Cancer Society, grant number 11 0569, and by the Swedish Research Council, grant number K2012-57X-21044-04-6. All authors declare no conflicts of interest.

References


