Virology 440 (2013) 1–7

Contents lists available at SciVerse ScienceDirect

Virology

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

Metagenomic sequencing of "HPV-negative" condylomas detects novel putative HPV types

Hanna Johansson [a], Davit Bzhalava [b], Johanna Ekström [a], Emilie Hultin [b], Joakim Dillner [a,b], Ola Forslund [a,*]

[a] Department of Laboratory Medicine, Division of Medical Microbiology, Lund University, Malmö, Sweden
[b] Departments of Laboratory Medicine, Medical Epidemiology & Biostatistics, Karolinska Institute and Departments of Clinical Microbiology and Pathology, Karolinska Hospital, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 13 December 2012
Accepted 31 January 2013
Available online 22 March 2013

Keywords:
Human papillomavirus
High-throughput sequencing
Condyloma

ABSTRACT

Condylomas are caused by human papillomavirus (HPV), but may in rare cases be "negative for HPV" by PCR. Metagenomic sequencing can be used for an unbiased assessment of the presence of virus. Ten swab sample pools, each containing four cases of "HPV-negative" condylomas, were subjected to metagenomic sequencing. One pool contained Molluscum contagiosum. Five pools contained HPV, of which three pools contained novel putative HPV-types. The 12 samples in these three pools were sequenced individually. Six of these contained HPV and two contained Molluscum contagiosum. Altogether, 1337 HPV-related reads were detected, representing 23 novel putative Gammapapillomaviruses, 10 established HPV types (genital HPV types 6, 57, 58 and 66, Betapapillomavirus types 5, 105, 124, and Gammapapillomavirus types 50, 130, 150) and two described HPV sequences (KC7 and FA69). Complete genomes of Gammapillomavirus FA69 and SE87 were compiled. Metagenomic sequencing reveals that seemingly "HPV-negative" condylomas contain known and previously unknown HPV types.

Introduction

Condylomas are a very common sexually-transmitted disease caused by infection by human papillomavirus (HPV), most commonly by HPV 6 of the Alphapapillomavirus genus (Gissmann et al., 1982; Bernard et al., 2010). The HPVs are a large group of small, non-enveloped dsDNA-viruses that infect keratinocytes of the skin and mucosa. There are 156 established HPV types (Bernard et al., 2010; Bottalico et al., 2011; Chouhy et al., 2010; de Villiers et al., 2004; Kohler et al., 2011; Kovanda et al., 2011) and several complete genomes representing putative HPV types (Foulongne et al., 2012; Li et al., 2012). In addition to the established types, a large number of subgenomic HPV sequences representing novel putative 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 sequencing (SE-isolates) (Ekstrom et al., 2011).

The oncogenic HPV types are the main cause of cervical cancer, being found in close to 100% of cervical tumors (Walboomers et al., 1999). These HPV types also cause other types of malign mucosal tumors, such as anal, vulvar, and oral cancers (IARC, 2012). A few HPV negative cervical cancers exist, and four possible explanations (in addition to true negativity) have been suggested (Walboomers and Meijer, 1997): (i) specimen inadequacy, (ii) loss of the L1 gene due to integration, (iii) detection method insensitivity, and (iiii) the existence of still unidentified 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 condylomas, that are seemingly "HPV-negative" is a promising strategy to discover new, pathogenic HPVs (Ekstrom et al., 2010). In southern Sweden, a systematic condyloma reporting system has performed HPV genotyping of the largest series of condylomas reported to date in the literature (Sturegård et al., 2013). This has enabled collection of a sizable number of condyloma samples that are seemingly negative by broad HPV-primer PCR system (Soderlund-Strand et al., 2009). High-throughput sequencing has, for many years, been instrumental in the characterization of bacterial metagenomes and the discovery of new viruses (Didelot et al., 2012; Human Microbiome Project, 2012; Mokili et al., 2012; Willner et al., 2012), and has, in recent years, also been used for virus detection in skin samples (Ekstrom et al., 2011; Foulongne et al., 2012).

In the present study, we used metagenomic sequencing of "HPV-negative" condylomas in order to obtain a comprehensive map of which known or previously unknown viruses are present in these lesions.

Results

Forty samples of apparently HPV-negative condylomas were pooled in ten pools with four samples in each. The samples were

* Corresponding author. Fax: +46 40 337312.
E-mail address: ola.forslund@med.lu.se (O. Forslund).

0042-6822/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2013.01.023

subjected to whole genome amplification (WGA) and sequenced using 454 technology. In three of the pools, putative novel HPV-types were detected and the twelve samples that had been included in these pools were also sequenced individually.

In total, we obtained 104,740 sequence reads. These were 43% bacterial sequences (45,018 reads), 28% human sequences (29,062 reads), 4% viral sequences (4269 reads), 0.3% phage sequences (292 reads), 9% other sequences (9568 reads), and 16% unknown sequences (16,531 reads). As "other", we have classified reads with similarities to other sequences present in GenBank (e.g., plants, plant viruses, synthetic constructs, or non-human viruses). Sequences which were not similar to any sequence deposited in GenBank were classified as "unknown".

Among the viral reads, 52% (2207/4269) were related to viruses known to infect humans, and were retained for further analysis. Since swab samples do not contain enough DNA for 454 sequencing, we used whole genome amplification (WGA) to increase the amount of DNA prior to the high throughput sequencing. WGA can introduce errors such as chimeras or genome rearrangements (Lasken and Stockwell, 2007). Therefore, the reads were processed with stringent default settings to remove chimeras using both an automatic process and CodonCode Aligner (version 4.0.3, CodonCode Corporation), with the result that only 63% (1385/2207) of the sequences related to human viruses remained. Among these sequences, 96.5% (1337/1385) were related to HPV and 3.5% (48/1385) were Molluscum contagiosum sequences (Table 1). Among the 1 337 HPV-related reads, 11% (145/1337) belonged to subgenomic fragments of novel putative HPV types (Tables 1 and 2). Most of the HPV-related reads, (1064 reads) were detected in the individually sequenced samples (Table 1).

Five out of ten pools were HPV-positive, out of which two contained sequences from HPV6 and three contained sequences from novel putative HPV types (Table 1). In four of the pools, no viral reads were detected. Sequencing of the individual samples

Table 1
Schematic overview of sequenced samples. Pooled samples (four samples in each pool) from the first sequencing round are seen to the left and the corresponding single samples to the right. Only samples from three pools (pool 3, 5, and 10) were sequenced as single samples.

First seq	Virus type	Reads	Fragments/ contigs	Mean length (bp)	Second seq	Sex (age)	Virus type	Reads	Fragments/ contigs	Mean length (bp)
Pool 1	HPV6	24	5	1564 (391–4098)						
	HPV66	1	1	295						
Pool 2	Molluscum contagiosum virus subtype 1	12	12	222 (89–469)						
Pool 3	HPV57	2	2	139 (105–173)	Sample 1	F (41)	Molluscum contagiosum virus subtype 1	3	3	NA
	SE92	2	1	311	Sample 2	F (28)	–	–	–	–
	SE95	1	1	691	Sample 3	M (41)	HPV57	2	2	492.5 (489–496)
	SE96	1	1	109			HPV150	1	1	484
							SE95	1	1	401
					Sample 4	M (40)	Molluscum contagiosum virus subtype 1	33	25	465 (414–520)
Pool 4	–	–	–	–						
Pool 5	FA69	40	13	329 (91–800)	Sample 5	F (38)	–	–	–	–
	SE92	2	1	311	Sample 6	F (33)	SE101	3	1	504
	SE93	1	1	173	Sample 7	M (30)	HPV6	25	6	625 (144–2143)
	SE94	3	1	240			HPV isolate FA69	858	1	Complete
	SE95	8	1	1047			HPV isolate KC7	24	4	1029 (473–1415)
	SE96	3	1	109			SE87	80	1	Complete
	SE97	1	1	189	Sample 8	M (24)	SE8[a]	8	1	629
	SE98	1	1	145			SE104	1	1	506
	SE99	1	1	148			SE105	1	1	501
	SE100	1	1	418			SE109	25	1	2302
							SE110	1	1	531
							SE113	1	1	507
							SE114	2	1	330
Pool 6	–	–	–	–						
Pool 7	HPV6	177	7	819 (391–1347)						
	HPV124	1	1	264						
Pool 8	–	–	–	–						
Pool 9	–	–	–	–						
Pool 10	SE102	1	1	382	Sample 9	F (26)	HPV6	3	1	2143
	SE103	1	1	338			HPV50	3	2	686 (497–876)
	SE106	1	1	375			HPV58	15	9	385 (218–488)
					Sample 10	F (25)	HPV5	1	1	485
							HPV105	5	3	487 (474–496)
							HPV130	2	2	345 (247–443)
							SE107	1	1	476
							SE116	1	1	458
					Sample 11	M (28)	–	–	–	–
					Sample 12	F (25)	–	–	–	–

[a] Previously published (Ekstrom et al., 2011).