Human papillomavirus 33 worldwide genetic variation and associated risk of cervical cancer

Alyce A. Chen [a], Daniëlle A.M. Heideman [b], Debby Boon [b], Zigui Chen [c], Robert D. Burk [c], Hugo De Vuyst [a], Tarik Gheit [a], Peter J.F. Snijders [b], Massimo Tommasino [a], Silvia Franceschi [a], Gary M. Clifford [a,*]

[a] International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon cedex 08, France
[b] Department of Pathology, VU University Medical Center, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands
[c] Albert Einstein College of Medicine, 1300 Morris Park Ave., 10461 Bronx, NY, USA

ARTICLE INFO

ABSTRACT

Human papillomavirus (HPV) 33, a member of the HPV16-related alpha-9 species group, is found in approximately 5% of cervical cancers worldwide. The current study aimed to characterize the genetic diversity of HPV33 and to explore the association of HPV33 variants with the risk for cervical cancer. Taking advantage of the International Agency for Research on Cancer biobank, we sequenced the entire E6 and E7 open reading frames of 213 HPV33-positive cervical samples from 30 countries. We identified 28 HPV33 variants that formed 5 phylogenetic groups: the previously identified A1, A2, and B (sub) lineages and the novel A3 and C (sub)lineages. The A1 sublineage was strongly over-represented in cervical cases compared to controls in both Africa and Europe. In conclusion, we provide a classification system for HPV33 variants based on the sequence of E6 and E7 and suggest that the association of HPV33 with cervical cancer may differ by variant (sub)lineage.

© 2013 Elsevier Inc. All rights reserved.

Background

A subset of human papillomaviruses (HPV) is considered the necessary cause of cervical cancer and a proportion of other anogenital and head and neck carcinomas (Bouvard et al., 2009). While infection with a high-risk HPV is a common event among sexually active women, the majority of infections are cleared within 2 years (Goodman et al., 2008; Insinga et al., 2009; Plummer et al., 2007; Rosa et al., 2008). There is a need to identify the factors that influence a patient's progression from HPV infection to viral persistence, cellular transformation, and cervical cancer. Current evidence suggests that sequence variations within HPV16 and HPV18 may influence viral persistence and clinical outcome (Gheit et al., 2011; Schiffman et al., 2010; Zuna et al., 2009; Sathish et al., 2005; Villa et al., 2000; Berumen et al., 2001).

HPV33 is a high-risk HPV within the same phylogenetic species (alpha-9) as HPV16 (Bernard et al., 2010; de Villiers et al., 2004) and accounts for approximately 5% of cervical cancer cases worldwide, with some variation in this proportion by geographical region (Guan et al., 2012; Li et al., 2011). For example, while HPV33 is found in 5.4% of cervical cancer cases in Eastern Asia, it is only found in 1.7% of cases in Oceania (Li et al., 2011). HPV33 was originally cloned from an invasive cervical carcinoma (Beaudenon et al., 1986), and the entire viral sequence was described shortly after (Cole and Streeck, 1986). Since then, a few studies have described genetic variation in HPV33 worldwide (Chen et al., 2011; Stewart et al., 1996; Godínez et al., 2013) of which the most comprehensive was based upon the whole genome sequencing of 20 HPV33-positive samples (Chen et al., 2011). HPV33 variants were classified into two major lineages, A and B. The A lineage was further divided into two sublineages, A1, which includes the prototype sequence [M12732.1 (Cole and Streeck, 1986)], and A2. This classification is based upon the definition that the full genome sequence of a major variant lineage differs by approximately 1.0% from another variant lineage of the same HPV type, with differences of 0.5–0.9% defining sublineages (Chen et al., 2011).

Previous studies of HPV33 variants and clinical outcome have tended to be small and focused on a particular geographical region or ethnic group (Khouadri et al., 2006; Bokal et al., 2010; Gagnon et al., 2004; Garbuglia et al., 2007; Ntova et al., 2012; Raiol et al., 2009; Wu et al., 2009; Xin et al., 2001). Hence, the aims of the current study were to further characterize the genetic diversity of HPV33 worldwide and to explore the association of well-defined HPV33 variant lineages with the risk for cervical cancer. To do this, we sequenced and analyzed the entire E6 and E7 open reading

frames of HPV33-positive cervical samples stored at the International Agency for Research on Cancer (IARC), including cervical cancer cases and controls collected during more than 20 years of studies on HPV.

Materials and methods

Origin of clinical specimens

The IARC has coordinated cervical cancer case series, cervical cancer case-control studies, and population-based HPV prevalence surveys in a large number of countries around the world (Bosch et al., 1995; Muñoz et al., 2003; Franceschi et al., 2003; Clifford et al., 2005; Li et al., 2006; Bardin et al., 2008; Dondog et al., 2008; Keita et al., 2009; Sherpa et al., 2010; Alibegashvili et al., 2011; De Vuyst et al., 2012; Aruhuri et al., 2012; Sideri et al., 2009) and as yet unpublished studies from Fiji, Bhutan, and Senegal. The collection of samples has spanned a period of over 20 years, but predates the introduction of HPV vaccines. Informed consent was obtained from all participants, and the studies were approved by the IARC Ethical Review Committee. Cervical samples (exfoliated cells or tissue biopsy specimens) derived from these studies have been comprehensively genotyped for 37 HPV types by using a standardized and well-validated protocol (General Primer GP5+/6+PCR–EIA followed by reverse line blot assay) (van den Brule et al., 2002) in one centralized laboratory (Molecular Pathology Unit, Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands). All HPV33-positive cervical samples in the IARC biobank were selected for the current analysis, without exclusion, and were categorized into the following regions: Africa, Asia and Oceania, Europe, and South America. Country specific details are noted in Table 1.

PCR and DNA sequencing

DNA extraction from stored samples was performed using the High Pure PCR Template Preparation Kit (Roche, Mannheim, Germany), and DNA isolates were subjected to β-globin PCR to assess sample quality, as described previously (Hesselink et al., 2009). Sequencing of the entire HPV33 E6 and E7 region (nucleotides 109–866) was performed as described previously (Godínez et al., 2013) using a series of HPV33 specific primer pairs that were designed to amplify overlapping regions of the HPV33 E6 and E7 open reading frames in order to cover the entire E6 and E7 region.

To reveal single nucleotide polymorphisms (SNPs), sequences of the specimens were aligned to the prototype HPV33 sequence (NCBI accession number M12732) using multalin software (http://multalin.toulouse.inra.fr/ multalin/). SNPs that were observed in only one sample were confirmed by re-examination of the sequence traces. Isolates that did not classify into existing lineage categories were confirmed by manual re-examination of the sequencing traces and with additional sequencing, where necessary. Multiple sequence traces for each sample were compiled to provide one sequence encompassing the entire HPV33 E6 and E7 region. All sequences were submitted to GenBank (accessions KC862070–KC862080, KC881011–KC881020, and KF536962–KF536968).

Phylogenetic analysis

Unrooted consensus trees were built using the Phylogeny Inference Package (PHYLIP), version 3.69 (Felsenstein, 1989). This included generating 10,000 bootstraps using the F84 model of DNA distances, clustering with the unweighted pair group method with arithmetic mean (UPGMA), and applying the majority rule extended, or greedy, method of consensus. Trees created with a

Table 1
Geographic distribution of 213 HPV33-positive cervical samples

Region/country	No.
AFRICA	**56**
Algeria	2
Guinea	17
Kenya	3
Mali	1
Morocco	3
Nigeria	4
Senegal	13
South Africa	10
Tanzania	1
Uganda	2
ASIA and OCEANIA	**87**
Bhutan	21
China	6
Fiji	3
India	13
Korea	5
Mongolia	20
Nepal	3
Thailand	8
Vanuatu	1
Vietnam	7
EUROPE	**36**
Georgia	7
Italy	5
Poland	10
Spain	14
SOUTH AMERICA	**34**
Argentina	3
Brazil	6
Chile	7
Colombia	5
Paraguay	11
Peru	2
TOTAL	**213**

maximum-likelihood method showed similar results and are not described further. For the principal tree analysis, all unique sequence variants found in IARC samples were included. In a subsequent analysis, IARC variants were supplemented by other unique E6/E7 sequences reported in the literature, including 2 from Garbuglia et al. (2007), 6 from Khouadri et al. (2006), 1 from Wu et al. (2009), and 4 from Chen et al. (2011) and personal communication. The IARC samples that contained a variant that appeared to belong to a newly described (sub)lineage were full-genome sequenced as described previously (Chen et al., 2011) to establish the phylogenetic classification and nomenclature.

Case-control analysis

Samples were classified as either controls [including normal, atypical squamous cells of unknown significance (ASCUS), low-grade intraepithelial lesion (LSIL), or cervical intraepithelial neoplasia (CIN) 1] or cases [squamous cell carcinoma, adenocarcinoma, adenosquamous cell carcinoma, or unspecified invasive cervical cancer]. Samples from population-based HPV prevalence studies for which histology and cytology were unavailable were also classified as controls ($n=15$). Samples reported as CIN2, CIN3, or high-grade squamous intraepithelial lesion (HSIL) were excluded from the case-control analysis ($n=15$), but were included in the previously described phylogenetic analysis. Region-specific associations between variant (sub)lineage and case-control status were assessed by 2-sided p-values arising from Fisher's exact test