Virology 445 (2013) 232–243

Contents lists available at ScienceDirect

Virology

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

Human papillomavirus genome variants

Robert D. Burk [a,b,c,d,*], Ariana Harari [b], Zigui Chen [a]

[a] Department of Pediatrics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx 10461, NY, USA
[b] Department of Microbiology and Immunology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx 10461, NY, USA
[c] Department of Epidemiology and Population Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx 10461, NY, USA
[d] Department of Obstetrics, Gynecology and Women's Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx 10461, NY, USA

ARTICLE INFO

Available online 31 August 2013

Keywords:
HPV
Human papillomavirus variants
Alphapapillomaviruses
HPV variant lineages
HPV evolution

ABSTRACT

Amongst the human papillomaviruses (HPVs), the genus *Alphapapillomavirus* contains HPV types that are uniquely pathogenic. They can be classified into species and types based on genetic distances between viral genomes. Current circulating infectious HPVs constitute a set of viral genomes that have evolved with the rapid expansion of the human population. Viral variants were initially identified through restriction enzyme polymorphisms and more recently through sequence determination of viral fragments. Using partial sequence information, the history of variants, and the association of HPV variants with disease will be discussed with the main focus on the recent utilization of full genome sequence information for variant analyses. The use of multiple sequence alignments of complete viral genomes and phylogenetic analyses have begun to define variant lineages and sublineages using empirically defined differences of 1.0–10.0% and 0.5–1.0%, respectively. These studies provide the basis to define the genetics of HPV pathogenesis.

Introduction and background

Papillomaviruses are small circular double-stranded DNA viruses. They are highly species specific and preferentially infect cutaneous or mucocutaneous epithelium. Papillomavirus genomes have been isolated and characterized from reptiles (Herbst et al., 2009), birds (Terai et al., 2002), marsupials (Bennett et al., 2010) and multiple other mammalian species (for recent review see Bernard et al. (2010)) suggesting an evolutionary history spanning more than 300 million years (Herbst et al., 2009). Papillomaviruses replicate their genomes using the host enzymatic machinery, ensuring a high degree of proof reading with low mutation rates (Rector et al., 2007). As their genomes have evolved, mutations, insertions/deletions (indels) and rearrangements have been selected and/or fixed through host isolation into discrete entities (called types), with the intermediary genomes generally lost (Bernard et al., 2006; Gottschling et al., 2007). More recent evolutionary events are responsible for the large heterogeneity of related viral variants detected in the modern era.

Papillomavirus evolution has been predominantly asexual, although extremely rare recombination events cannot be excluded. This implies that multiple mutations/variations occurring in papillomavirus genomes are not related to genetic distance as in recombining genomes, i.e., linkage disequilibrium, but to sequential accumulation of genetic changes. This process of speciation has been termed lineage fixation (Chen et al., 2005). That is, groups of single nucleotide polymorphisms and/or insertions/deletions (indels) tend to become fixed within viral lineages. Over time, the quantity of these lineage-defining variations increases eventually leading to speciation (entities referred to as types). For papillomaviruses this has been defined when approximately 10% or more of the nucleotide positions differ between 2 genomes that share a most recent common ancestor; this takes millions of years (Chen et al., 2009; Rector et al., 2007).

A distinct human papillomavirus (HPV) "type" is established and curated when the DNA sequence of the L1 open reading frame (ORF) of the cloned viral genome differs from that of any other characterized type by at least 10% (Bernard et al., 2010; de Villiers et al., 2004). Within the PV research community, isolates of the same HPV type are referred to as variants or subtypes when the nucleotide sequences of the L1 ORF differ by less than 10%. The criterion for HPV types has proven extremely stable and useful for basic researchers, clinicians, epidemiologists and immunologists.

Over 160 HPV types have been fully characterized (see ⟨http://www.hpvcenter.se/html/refclones.html⟩); approximately 60 of these are predominantly detected in mucosal epithelia and sort to the genus *Alphapapillomavirus* (alpha-PV) (Bernard et al., 2010; de Villiers et al., 2004). Human alpha-PV infections are involved in the development of both benign and malignant disease, e.g., condylomata acuminatum/respiratory papillomatosis and cervical/anal/head and

* Corresponding author at: Department of Pediatrics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx 10461, NY, USA. Fax: +1 718 430 8975.
E-mail address: robert.burk@einstein.yu.edu (R.D. Burk).

0042-6822/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2013.07.018

neck cancers, respectively. Cervical cancer is the most common gynecologic malignancy and one of the leading causes of cancer mortality in women worldwide (Jemal et al., 2011). Most oncogenic or high-risk (HR) types associated with invasive cervical cancer (Li et al., 2011; Munoz et al., 2003; Smith et al., 2007) are clustered in one clade of the alpha-HPVs that contains species groups alpha-5, alpha-6, alpha-7 and alpha-9 (Burk et al., 2009; Schiffman et al., 2005) and account for ~90% of all cervical cancers worldwide (Li et al., 2011; Smith et al., 2007). Despite phylogenetic relatedness, HPV variants can differ in pathogenicity (Berumen et al., 2001; Burk et al., 2003; Chan et al., 2013; Cornet et al., 2013b; Giannoudis and Herrington, 2001; Hecht et al., 1995; Sabol et al., 2012; Schiffman et al., 2010; Villa et al., 2000; Xi et al., 2007). Because of their medical importance, the predominant research focus on papillomaviruses is their association with pathologic and oncogenic lesions. Thus, this review will focus exclusively on human papillomaviruses and predominantly on the HPVs within the genus *Alphapapillomavirus*.

In this review, we update knowledge of variant lineages and sublineages using currently available complete genome sequences of human *Alphapapillomavirus* from species groups: alpha-3 (HPV61), alpha-5 (HPV26, 51, 69 and 82), alpha-6 (HPV30, 53, 56, 66), alpha-7 (HPV18, 39, 45, 59, 68, 70, 85 and 97), alpha-9 (HPV16, 31, 33, 35, 52, 58 and 67), alpha-10 (HPV6 and 11), alpha-11 (HPV34 and 73) and alpha-13 (HPV54). We summarize the rationale for the grouping of related HPV types and highlight HPV16 lineages as an example of the increased complexity in a dichotomized classification into European and non-European groups by reviewing the use of an alphanumeric nomenclature. The establishment of a coherent classification and nomenclature system for HPV variant lineages will facilitate comparisons amongst studies that directly determine HPV sequences, especially as next-generation sequencing of PCR products and complete sample DNA (metagenomics) is rapidly being applied in new studies (Barzon et al., 2011; Conway et al., 2012; Ekstrom et al., 2011; Mokili et al., 2013). Moreover, specific variants occur on lineages that are stable and have correlated changes and diagnostic polymorphisms throughout the genome that can be used by all investigators (Alizon et al., 2011; Chen et al., 2005). Without a system for naming variant lineages, investigators have had to rely on referring to specific changes at nucleotide positions.

Historical perspective and previous reviews of HPV variants

This review builds upon the pioneering work of early investigators in the field that appreciated the significance of studying HPV variants. The earliest work in the field predated PCR and relied upon restriction enzyme digestion polymorphisms and/or Southern blot hybridization to recognize related viral variants (de Villiers et al., 1981; Mounts et al., 1982). Previous reviews of viral variants can be found from the era when sequencing DNA was cumbersome, time consuming and technically challenging (Bernard, 1994; Bernard et al., 1994). The lab of Bernard was the first to recognize the association of HPV16 (Ho et al., 1993) and HPV18 (Ong et al., 1993) viral variants with human population migrations and continent of origin (see Bernard (1994)) for review of this topic). The lab of CM Wheeler provided some of the early evidence for the intratype variation of HPV types associated with cervical neoplasia and HPV16 and HPV18 heterogeneity (Arias-Pulido et al., 2005; Stewart et al., 1996; Yamada et al., 1997, 1995). With the demonstration of an association of HPV16 variants and cervical cancer (Berumen et al., 2001; Cornet et al., 2013b; Hildesheim et al., 2001; Sichero et al., 2007; Xi et al., 2006), many of the studies on variants explored this aspect of the biology of variants as described in a number of excellent reviews (Bernard et al., 2006; Giannoudis and Herrington, 2001; Hildesheim and Wang, 2002; Lizano et al., 2009; Sichero and Villa, 2006). This current review focuses predominantly on the study of alpha-HPV variants using complete genome analyses.

The study of HPV variants from other genera provided evidence for significant intratype variation of HPV5 and HPV8 from individuals with epidermodysplasia verruciformis (Deau et al., 1993, 1991). In fact, the intratype heterogeneity of HPVs from the genera *Betapapillomavirus* and *Gammapapillomavirus* has remained basically unexplored, and there is currently a significant effort to characterize novel types from these genera (Bottalico et al., 2011; Kohler et al., 2011; Li et al., 2012). With implementation of Next-Gen sequencing (Barzon et al., 2011; Ekstrom et al., 2011), it is anticipated that there will be rapid advances in studying these genome variants (Bernard, 2013), particularly if associated with pathologic lesions.

Although the co-evolution of human populations and HPV16 and HPV18 variants is well supported, the geographic associations for variants of other types remains unresolved (Calleja-Macias et al., 2004, 2005; Chan et al., 1997; Heinzel et al., 1995; Matos et al., 2013; Prado et al., 2005). Global studies of HPV variant lineages from worldwide populations are needed to better understand the relationship between HPV and the recent and past evolution and dispersion of their human hosts.

HPV variants and cervical cancer pathogenesis

The natural history of HPV and subsequent development of cervical cancer follows a set of stages from sexual exposure to an oncogenic HPV (Bouvard et al., 2009), persistence of infection, development of a precancerous lesion and progression of the precancerous lesion to invasive disease (for recent review see Schiffman and Wentzensen (2013)). Each stage of a pernicious papillomavirus infection leading to cancer (Burk, 1999) may be influenced by both the host and viral genome variations and environmental factors (Hildesheim and Wang, 2002; Wheeler, 2008). Since the outcome of cervical cancer for any specific HPV exposure is extremely rare, most studies have used case-control designs to investigate the association of specific variants with precancer and cervix cancer. Histologic diagnosis is the end point of choice, although many studies have used cytologic outcomes. The most robust histologic "biomarker" of precancer is cervical intraepithelial neoplasia grade 3 lesions (CIN3) (Luhn et al., 2013), nevertheless studies also utilize CIN2 to enhance power by forming a high-grade class. Thus, if we only consider oncogenic HPV types, the contribution of viral variants to cancer development could be related to (1) acquisition of infection given exposure (differences in infectivity), (2) long-term persistence (e.g., greater than 1 or 2 years and could reflect viral differences in immune clearance or other properties), (3) development of precancer usually given persistence (differences in viral dysregulation of cellular differentiation and/or accumulation of somatic cellular changes in the premalignant or stromal cells) and lastly, (4) invasion of HPV containing cells. It is assumed that most invasive cervix cancer goes through all stages, however there are significant differences between the more common squamous cell carcinoma and the rarer adenocarcinoma of the cervix (Kim et al., 2013; Schiffman et al., 2007; Schiffman and Wentzensen, 2013).

Analysis of the pathogenic affects of HPV variants for this review encompassed evaluation of over 50 manuscripts spanning nearly 20 years. The topic turned out to be extremely complicated because of the complexity of the interaction of the virus with differences in host genetics (de Araujo Souza et al., 2009; Hildesheim and Wang, 2002; Xi et al., 2006) and prevalence of circulating HPV variants in different populations and/or geographic regions. In addition, description or characterization of