

Short Communication

Correspondence

Darren P. Martin
darren@science.uct.ac.za

Received 7 February 2006
Accepted 2 May 2006

Evidence of ancient papillomavirus recombination

Arvind Varsani,^{1,2} Eric van der Walt,² Livio Heath,² Edward P. Rybicki,^{2,3}
Anna Lise Williamson³ and Darren P. Martin³

Electron Microscope Unit¹, Department of Molecular and Cell Biology² and Institute of Infectious Disease and Molecular Medicine³, University of Cape Town, Observatory, Cape Town 7925, South Africa

An open question amongst papillomavirus taxonomists is whether recombination has featured in the evolutionary history of these viruses. Since the onset of the global AIDS epidemic, the question is somewhat less academic, because immune-compromised human immunodeficiency virus patients are often co-infected with extraordinarily diverse mixtures of human papillomavirus (HPV) types. It is expected that these conditions may facilitate the emergence of HPV recombinants, some of which might have novel pathogenic properties. Here, a range of rigorous analyses is applied to full-genome sequences of papillomaviruses to provide convincing statistical and phylogenetic evidence that evolutionarily relevant papillomavirus recombination can occur.

Papillomaviruses (PVs; family *Papillomaviridae*) have been found exclusively in vertebrates, where they display high degrees of species specificity and are only capable of infecting squamous epithelial cells productively (Howley, 1995). Although most human papillomaviruses (HPVs) are benign, a 'high-risk' oncogenic subset contains causal agents of various oral and anogenital cancers (Clifford *et al.*, 2003; Muñoz, 2000; Steller, 2002).

The circular, double-stranded DNA genomes of nearly all PVs are approximately 8 kb in size and may, depending on the species, encode three oncogenes (E5, E6 and E7), two replication proteins (E1 and E2), a protein involved in productive infection (E4) and two capsid proteins (L1 and L2). Linkage of genome replication with host cell cycles, coupled with a low replication-error rate, means that the most variable parts of PV genomes are diverging at a rate of approximately 1% every 40 000–80 000 years (Bernard *et al.*, 1994) – an evolutionary rate around one million times lower than that of most RNA viruses (Halpern, 2000). It has been suggested convincingly that HPVs have been infecting and co-evolving with humans for at least a few hundred thousand years (Bernard, 2005; Calleja-Macias *et al.*, 2005; Chan *et al.*, 1992).

Although various sequence analyses have indicated that PV evolution has occurred in the absence of recent recombination (Chan *et al.*, 1992, 1997; Halpern, 2000), slight differences between internal branches of phylogenetic trees constructed by using different parts of the same PV genomes (Bravo & Alonso, 2004; Schiffman *et al.*, 2005) have raised suspicions that recombination may have occurred, albeit infrequently, between ancestral PV lineages. A rigorous

search for evidence of PV recombination using currently available recombination-detection methods (Posada & Crandall, 2001) has, however, been seriously hindered by the technical difficulty of aligning extremely diverse PV sequences accurately. Nevertheless, a recent statistical analysis of topology differences in trees constructed from inferred amino acid sequences encoded by different genes from α -PVs has provided the most convincing evidence yet of PV recombination (Narechania *et al.*, 2005).

However, an important question remains: do these reported differences in tree topology represent evidence of ancient recombination events or are they artefacts of imperfect sequence alignment and phylogenetic-reconstruction methodologies? Here, we report the combined use of rigorous alignment-accuracy tests and some of the most powerful recombination-detection methods currently available to more conclusively identify and characterize statistically probable recombination events in PV full-genome sequences.

One hundred and five complete PV genome sequences, representing the full scope of currently sequenced PVs (de Villiers *et al.*, 2004) (see Supplementary Table S1, available in JGV Online), were obtained from GenBank. All genomes were 'linearized' at the first nucleotide of their L1 genes and aligned by using POA (Grasso & Lee, 2004) (gap open penalty, 12; gap extension penalty, 4). We constructed five multiple sequence alignments: GP0, containing all 105 sequences; GP1, containing α -PV sequences; GP2, containing δ -, ε -, ζ -, η -, τ -, ι -, κ -, λ -, μ - and ν -PV sequences; GP3, containing ξ -, σ -, π - and γ -PV sequences; and GP4, containing β -PV sequences. Analysis of these alignments with the RDP (Martin & Rybicki, 2000), GENECONV (Padidam *et al.*, 1999), BOOTSCAN (Martin *et al.*, 2005a), MAXIMUM χ^2 (Maynard Smith, 1992), CHIMAERA (Martin

Supplementary material is available in JGV Online.

et al., 2005b) and SISCAN (Gibbs *et al.*, 2000) recombination-detection methods as implemented in RDP2 (Martin *et al.*, 2005b) indicated that GP0, GP1, GP2, GP3 and GP4

respectively contained evidence of 224, 168, 45, 27 and 65 unique potential recombination signals (see Supplementary Methods for information on how these methods work and

Event	Gene map	Recombinant sequence(s) (genus-sp.)	Estimated breakpoint positions		Parental sequences		Evidence
			In alignment (alignment no.)	In GenBank sequence (acc. no.)	Minor (genus-sp.)	Major (genus-sp.)	
1		HPV85 (α-7) HPV18 (α -7) HPV45 (α -7) HPV59 (α -7) (GP1) HPV70 (α -7) HPV39 (α -7) HPV30 (α -6) HPV56 (α -6) HPV66 (α -6) HPV53 (α -6) HPV51 (α -5) HPV69 (α -5) HPV82 (α -5) HPV26 (α -5)	8916, 9784 (GP0) 8064, 8677 (GP1)	4451, 4999 (AF131950)	HPV3 (α -10)	HPV54 (α -13)	B, C, G, M, R, S, P
2		PsPV (α-1)	8683, 8834 (GP0)	4133, 4275 (AJ238373)	Unknown	HPV14 (β -1)	c, m, R, P
3		HPV42 (α-1)	9789, 10071 (GP0)	5158, 5382 (M73236)	HPV69 (α -5)	Unknown	g, R, S
4		HPV65 (γ-1) HPV4 (γ -1) HPV50 (γ -2) HPV60 (γ -2)	8702, 8947 (GP0)	3789, 4007 (X70829)	Unknown	HPV5 (β -1)	m, R, S
5		HPV71 (α-15)	10594, 69 (GP0) 9326, 26 (GP1)	5805, 5863 (AB040456)	HPV45 (α -7)	Unknown	G, M, r
6		HPV41 (ν-2)	5480, 6219 (GP0) 5175, 5536 (GP2)	1999, 2347 (X56147)	Unknown	COPV (λ -1)	B, C, M, R, P
7		HPV75 (β-3)	36, 178 (GP4)	5798, 5894 (Y15173)	HPV12 (β -1)	HPV15 (β -2)	g, m, r, S, P

Fig. 1. Characterization of seven recombination events detected amongst papillomavirus (PV) full-genome sequences. Sequences bounded by recombination breakpoints are shaded on the graphical representation of PV genomes. In cases where multiple recombinants (descended from the same recombinant ancestor) are identified for a particular event (events 1 and 4), graphical genome representations, breakpoint positions and accession numbers refer to the sequence with its name in bold type. Breakpoint positions represent the bounds of the strongest recombination signal, but are not necessarily at, or even close to, the breakage sites that occurred during the original recombination event. 'Minor' and 'major' parents refer to the parental sequences respectively contributing the smaller and larger fractions of the recombinant's sequence. Although parental sequences are identified in some cases, the viruses named are all distant relatives of the actual parental sequences. This indicates either that recombination events are ancient or that sparse sampling has left genomes resembling those of the actual parents more closely unsequenced. It is possible to detect recombination even if only sequences distantly resembling one of the recombinant's parents have been sampled – in such cases, one of the parents is listed as 'unknown'. In the evidence column, letters represent methods indicating the presence of recombination with >95% (lower case) and >99% (upper case) confidence: R/r, RDP; G/g, GENECONV; B/b, BOOTSCAN; M/m, MAXIMUM χ^2 ; C/c, CHIMAERA; S/s, SISCAN. The method indicating the clearest evidence of recombination for a particular event is represented by a letter in bold type. Strong phylogenetic evidence in favour of a recombination hypothesis is indicated by P.

RDP project files for detailed analysis results and program settings; all available in JGV Online).

Due to the difficulty of aligning the PV sequences reliably, we thought it likely that many of the apparent recombination signals were in fact alignment artefacts. Artefactual signals were identified by using two tests of alignment

consistency in the genome region surrounding each of the potential recombination events. In both tests, sequences that had been used to identify a specific recombination signal were selected from the alignment and realigned in isolation. In the first test, the region of sequence between estimated recombination breakpoints was realigned by using CLUSTAL W (default settings) (Thompson *et al.*, 1994) and the pattern

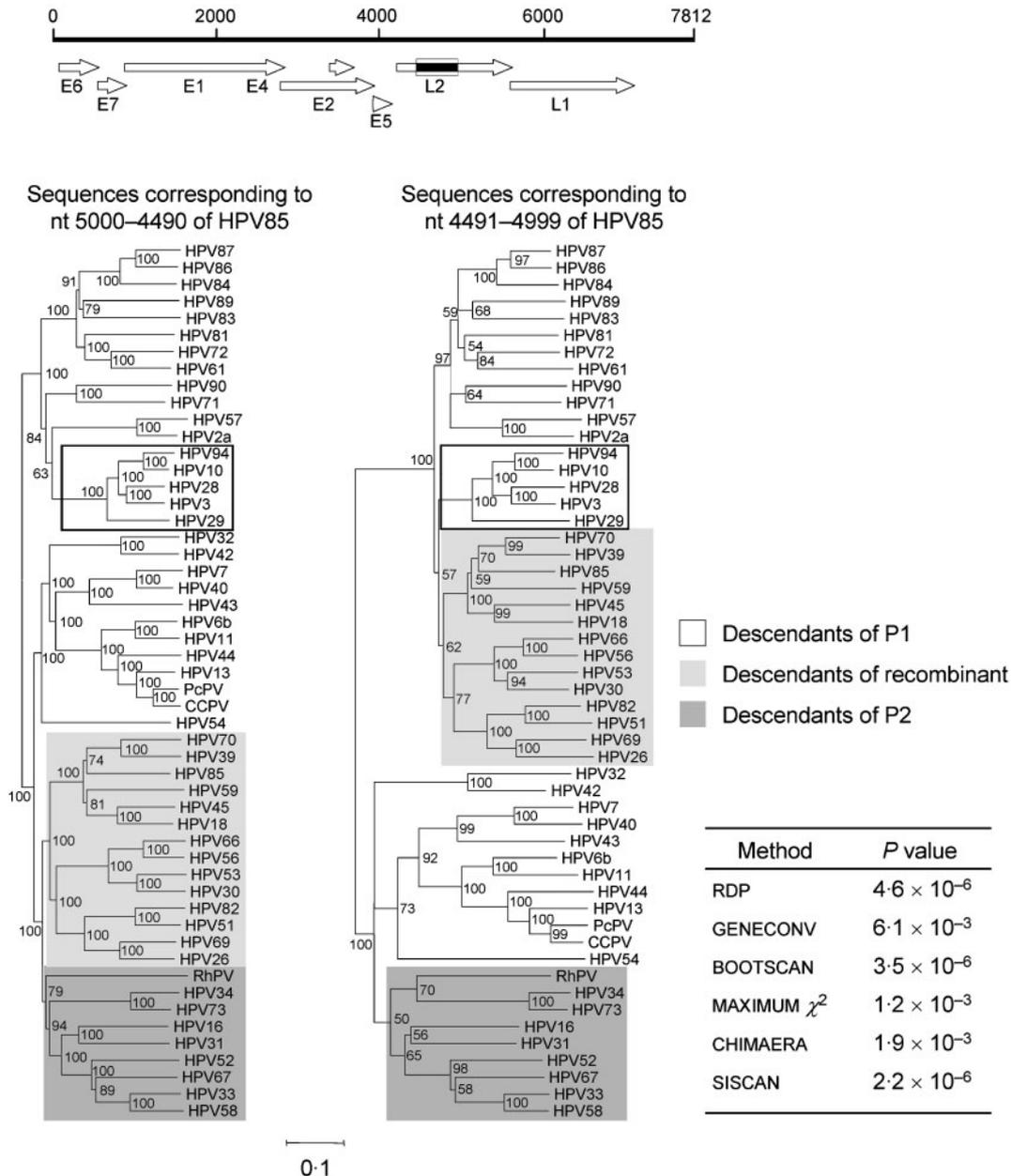


Fig. 2. Phylogenetic evidence of an ancient recombination event amongst ancestral HPV α viruses. These unrooted neighbour-joining trees (1000 bootstrap replicates, Jukes–Cantor distances) were constructed by using sequences with presumed homology to HPV85 (GenBank accession no. AF131950) nucleotide positions 1–4490 and 5000–7812 (the tree on the left) and 4491–4999 (the tree on the right). A shaded box on the gene map of HPV85 (above the trees) indicates the genome fragment used to construct the tree on the right. In both trees, branches with $<50\%$ bootstrap support have been collapsed. Various recombination-detection methods indicate, with a high degree of significance, that the sequences identified in light grey are the recombinant descendants of sequences ancestral to those identified in white and darker grey.

of nucleotide matches and mismatches between the three sequences in the original alignment was compared with that in the new alignment by using a χ^2 test. A χ^2 *P* value of <0.1 constituted a test failure. The second test involved realignment of the three full-length sequences with both POA and CLUSTAL W and reanalysis for recombination with RDP2. Inability to detect evidence for the same recombination signal in either the CLUSTAL W or POA realignments with a Bonferroni *P* value correction identical to that used for the initial signal detection constituted a test failure. CLUSTAL W and POA apply substantially different alignment algorithms, implying that recombination signals detectable in both POA and CLUSTAL W alignments have occurred in consistently alignable genome regions. Of the 529 potential recombination signals, only ten passed both tests. As three very similar signals were detected in the GP0 and either GP1 or GP2 alignments, these signals together represent evidence of at least seven unique interspecies PV recombination events (Fig. 1).

To examine the phylogenetic support for each individual event, two neighbour-joining trees (1000 bootstrap replicates, Jukes–Cantor distances) were constructed with PHYLIP version 3.61 (Felsenstein, 1989) by using two portions of the relevant multiple sequence alignment bounded by the approximated recombination breakpoints for that event (e.g. Fig. 2). We considered ‘movement’ of the identified recombinant sequence(s) between branches of the two trees with an excess of 60% bootstrap support to constitute good phylogenetic evidence in favour of recombination having occurred during the history of the sequence(s).

There was strong phylogenetic support favouring a recombination hypothesis for four of the seven events (events 1, 2, 6 and 7 in Fig. 1; Fig. 2). Importantly, one of these four events (event 1) has been localized to the same approximate region of species 5, 6 and 7 α -PV genomes (Fig. 2) identified previously, using a completely independent set of phylogeny-based statistical tests, as potentially having a recombinant origin (Narechania *et al.*, 2005). However, use of recombination-detection methods that directly identify specific sequences responsible for aberrant phylogenetic signals has enabled us to identify more accurately the portion of the L2 gene exchanged during this recombination event.

Our phylogenetic analyses also revealed that the identified recombination events are probably all ancient, in some cases predating the diversification of major PV groupings (events 1 and 4 in Fig. 1; Fig. 2). This is significant in that it means that viable PV recombinants, once formed, can spread and thrive.

Although we have detected PV recombination and have characterized a number of individual events, we feel that there is limited scope for further analysis. For example, whereas five of the seven detected events have breakpoints within the L2 gene, it is not possible to infer the presence of a L2 recombination hotspot. This is because the absolute

need for reliable nucleotide sequence alignment necessarily biases recombination analyses in favour of breakpoint detection in conserved genomic regions such as L2. The extreme genetic diversity amongst PVs also makes it difficult to envision accurate estimation of recombination rates by using model-based population genetic approaches.

Nevertheless, recombination is a potent evolutionary mechanism and evidence of viable PV recombinants is important. In light of our results and those of Narechania *et al.* (2005), it is of some concern that the global AIDS epidemic has created favourable conditions for the generation and emergence of recombinant HPVs. Whilst the higher incidence of multiple HPV infections in human immunodeficiency virus (HIV) patients (Chaturvedi *et al.*, 2005a, b; Haas *et al.*, 2005; Hameed *et al.*, 2001) is likely to support the generation of recombinants, HIV-induced immune dysfunction may facilitate both short-term survival and eventual adaptation of recombinants to new niches within humans.

Acknowledgements

This work was supported by a grant from the South African Poliomyelitis Research Foundation and the South African National Bioinformatics Network.

References

- Bernard, H.-U. (2005).** The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses. *J Clin Virol* **32** (Suppl. 1), S1–S6.
- Bernard, H.-U., Chan, S.-Y. & Delius, H. (1994).** Evolution of papillomaviruses. *Curr Top Microbiol Immunol* **186**, 33–54.
- Bravo, I. G. & Alonso, Á. (2004).** Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malignant or benign growth. *J Virol* **78**, 13613–13626.
- Calleja-Macias, I. E., Villa, L. L., Prado, J. C. & 16 other authors (2005).** Worldwide genomic diversity of the high-risk human papillomavirus types 31, 35, 52, and 58, four close relatives of human papillomavirus type 16. *J Virol* **79**, 13630–13640.
- Chan, S.-Y., Ho, L., Ong, C.-K. & 8 other authors (1992).** Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. *J Virol* **66**, 2057–2066.
- Chan, S.-Y., Bernard, H.-U., Ratterree, M., Birkebak, T. A., Faras, A. J. & Ostrow, R. S. (1997).** Genomic diversity and evolution of papillomaviruses in rhesus monkeys. *J Virol* **71**, 4938–4943.
- Chaturvedi, A. K., Dumestre, J., Gaffga, A. M. & 8 other authors (2005a).** Prevalence of human papillomavirus genotypes in women from three clinical settings. *J Med Virol* **75**, 105–113.
- Chaturvedi, A. K., Myers, L., Hammons, A. F., Clark, R. A., Dunlap, K., Kissinger, P. J. & Hagensee, M. E. (2005b).** Prevalence and clustering patterns of human papillomavirus genotypes in multiple infections. *Cancer Epidemiol Biomarkers Prev* **14**, 2439–2445.
- Clifford, G. M., Smith, J. S., Plummer, M., Muñoz, N. & Franceschi, S. (2003).** Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *Br J Cancer* **88**, 63–73.

- de Villiers, E.-M., Fauquet, C., Broker, T. R., Bernard, H.-U. & zur Hausen, H. (2004). Classification of papillomaviruses. *Virology* **324**, 17–27.
- Felsenstein, J. (1989). PHYLIP – Phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166.
- Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**, 573–582.
- Grasso, C. & Lee, C. (2004). Combining partial order alignment and progressive multiple sequence alignment increases alignment speed and scalability to very large alignment problems. *Bioinformatics* **20**, 1546–1556.
- Haas, S., Park, T. W., Voigt, E., Buttner, R. & Merkelbach-Bruse, S. (2005). Detection of HPV 52, 58 and 87 in cervicovaginal intraepithelial lesions of HIV infected women. *Int J Mol Med* **16**, 815–819.
- Halpern, A. L. (2000). Comparison of papillomavirus and immunodeficiency virus evolutionary patterns in the context of a papillomavirus vaccine. *J Clin Virol* **19**, 43–56.
- Hameed, M., Fernandes, H., Skurnick, J., Moore, D., Kloser, P. & Heller, D. (2001). Human papillomavirus typing in HIV-positive women. *Infect Dis Obstet Gynecol* **9**, 89–93.
- Howley, P. M. (1995). *Papillomavirinae: the viruses and their replication*. In *Fields Virology*, 3rd edn, pp. 2045–2076. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia, PA: Lippincott-Raven.
- Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562–563.
- Martin, D. P., Posada, D., Crandall, K. A. & Williamson, C. (2005a). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses* **21**, 98–102.
- Martin, D. P., Williamson, C. & Posada, D. (2005b). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260–262.
- Maynard Smith, J. (1992). Analyzing the mosaic structure of genes. *J Mol Evol* **34**, 126–129.
- Muñoz, N. (2000). Human papillomavirus and cancer: the epidemiological evidence. *J Clin Virol* **19**, 1–5.
- Narechania, A., Chen, Z., DeSalle, R. & Burk, R. D. (2005). Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *J Virol* **79**, 15503–15510.
- Padidam, M., Sawyer, S. & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218–225.
- Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* **98**, 13757–13762.
- Schiffman, M., Herrero, R., DeSalle, R. & 15 other authors (2005). The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology* **337**, 76–84.
- Steller, M. A. (2002). Cervical cancer vaccines: progress and prospects. *J Soc Gynecol Investig* **9**, 254–264.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.