# UC Irvine

## UC Irvine Previously Published Works


### Title

Evidence of Diversifying Selection in Human Papillomavirus Type 16 E6 But Not E7 Oncogenes


### Permalink

https://escholarship.org/uc/item/95g3n807



### Journal

Journal of Molecular Evolution, 55(4)

### ISSN

0022-2844



### Authors

DeFilippis, Victor R
Ayala, Francisco J
Villarreal, Luis P



### Publication Date

2002-10-01

### DOI

10.1007/s00239-002-2344-y



### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

J Mol Evol (2002) 55:491–499
DOI: 10.1007/s002369-002-2344-y

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 2002

# Evidence of Diversifying Selection in Human Papillomavirus Type 16 E6 But Not E7 Oncogenes

**Victor R. DeFilippis,[1,*] Francisco J. Ayala,[1] Luis P. Villarreal[2]**

[1] Department of Ecology and Evolutionary Biology, University of California—Irvine, Irvine, CA 92697, USA
[2] Department of Molecular Biology and Biochemistry, University of California—Irvine, Irvine, CA 92697, USA

Received: 30 November 2001 / Accepted: 1 May 2002

**Abstract.** Human papillomavirus type 16 is a common sexually transmitted pathogen capable of giving rise to cervical intraepithelial neoplasia and invasive carcinoma through the expression and activity of two adjacent oncogenes: E6 and E7. Naturally occurring amino acid variation is commonly observed in the E6 protein but to a much lesser extent in E7. In order to investigate the evolutionary mechanisms involved in the generation and maintenance of this variation, we examine 42 distinct E6-E7 haplotypes using codon-based genealogical techniques. These techniques involve estimation of the ratio of nonsynonymous to synonymous substitutions ($d_n/d_s$) and allow testing for directional (positive) natural selection. Positive selection was detected for four codon sites within the E6 oncogene but not in any E7 codons. The amino acid compositions and locations of selected sites are described. Possible sources of natural selection including antiviral immune pressure and polymorphism of host cellular proteins are discussed.

**Key words:** Adaptive evolution — Diversifying selection — DNA virus — Codon substitution model — Phylogeny

*Present address: Oregon Health and Science University, Vaccine and Gene Therapy Institute, 505 NW 185th Ave., Beaverton, OR 97006, USA.
*Correspondence to:* V. R. DeFilippis; *email:* defilipp@ohsu.edu

## Introduction and Background

Human papillomavirus type 16 (HPV16) is a double stranded DNA virus (genus *Papillomavirus*, family *Papovaviridae*) that codes for eight ORFs including two oncogenes: E6 and E7. The virus is the main etiologic agent of cervical cancer, being found in approximately 60% of such tumors (Bosch et al. 1995). Cervical cancer is the second most common malignancy among women, responsible for 500,000 new cases and 200,000 deaths annually (Syrjänen and Syrjänen 2000).

HPV16 replication primarily takes place in the stratified squamous epithelium and virus maturation is intimately linked to the differentiation path of developing keratinocytes. Replication begins with cellular induced transcription of E6 and E7 in basal cells. The proteins are responsible for alteration of the normal cell cycle and inhibition of differentiation making viral DNA replication possible. E6 facilitates degradation of the cellular tumor suppressor protein p53 via the ubiquitin pathway. This degradation requires complex of the E6 protein with p53 and cellular E6-associated protein (E6-AP) (Huibregtse et al. 1991). Direct p53 binding by E6 inhibits p53-mediated transcriptional repression (Lechner et al. 1992) and activation (Pim et al. 1994). E7 is responsible for binding to and inactivating retinoblastoma protein p105RB. This releases cellular transcription factor E2F, thought to be involved in cell proliferation. During typical HPV16 replication, transcription of E6 and E7 is downregulated by the viral E2 protein.

This leads to the resumption of p53 and p105RB function and subsequent normal cellular differentiation and replication.

Overexpression of E6 and/or E7 may lead to permanent inactivation of tumor suppressor proteins that ultimately results in cellular transformation and neoplasia (Hawley-Nelson et al. 1989, Münger et al. 1989, Lambert et al. 1993). This is most often brought about by integration of viral DNA into the host cell genome followed by unregulated, constitutive expression (Schwarz et al. 1985, Smotkin et al. 1986, Baker et al. 1987).

The E6 protein is composed of 158 amino acids, with a molecular mass of 18 kDa. It contains four Cys-X-X-Cys motifs that give rise to two well-conserved zinc fingers. The protein is known to perform additional functions and to interact with other cellular factors including paxillin (Tong et al. 1997) and hD1g/SAP97 (Lee et al. 1997), as well as participate in the upregulation of telomerase (Klingelhutz et al. 1996). The E7 protein is an acidic 21 kDa phosphorylated protein composed of 98 amino acids. It contains two Cys-X-X-Cys motifs forming one zinc finger. The E7 protein interacts with other pRB-related proteins (e.g. p107, p130) and with cellular factors involved in transcriptional control (e.g. AP-1, TBP) (see Zwerschke and Jansen-Dürr 2000).

Papillomaviruses rely on cellular DNA replication machinery and this results in slow synonymous substitution rates near $10^{-8}$/site/year. Yet, naturally-occurring HPV16 polymorphism is demonstrable and has been the focus of numerous studies (Chan et al. 1992, Eriksson et al. 1999, Ho et al. 1991, Wheeler et al. 1997, Yamada et al. 1997). Amino acid variation in E6 and E7 has been examined in the context of clinical manifestations of infection, epidemiology, and host genetic makeup (Tornesello et al. 1997, Bontkes et al. 1998, Zehbe et al. 1998, Brady et al. 1999, van Duin et al. 2000, Berumen et al. 2001, Hu et al. 2001). At both the nucleotide and amino acid levels, E6 genes are considerably more variable than haplotypic E7 genes, although they are genomically adjacent and both encode nonstructural proteins required for viral replication (Tornesello et al. 1997, van Duin et al. 2000). HPV16 genetic diversity often entails variation in the biological properties of the virus (Stoppler et al. 1996), which is likely to result in differences in pathogenicity, carcinomic risk, and perhaps immunogenicity, especially with respect to the E6 and E7 genes (Xi et al. 1997, Yamada et al. 1997). The evolutionary basis of this variation has never been examined, however, and its underlying biological significance remains unknown.

We have investigated E6 and E7 with an emphasis on the role of natural selection using 42 different haplotypes. Diversifying selection acting on a protein-coding region can be estimated by examining the ratio of nonsynonymous ($d_n$) to synonymous substitutions ($d_s$) per site (Hughes and Nei 1988, 1989). We employed genealogical, codon-based models developed by Nielsen and Yang (1998) and Yang et al. (2000) that incorporate maximum likelihood estimates of $d_n$ and $d_s$.

## Materials and Methods

### *E6 and E7 Haplotype Sequences*

In the summer of 1998 cervical swabs were collected as part of an epidemiological study in Oaxaca, Mexico. Dacron-tipped swabs containing exfoliated cervical cells were stored in 800 μL of specimen transport medium (Digene, Beltsville, MD). DNA was then extracted using QIAmp tissue extraction kits (Qiagen, Valencia, CA) according to the manufacturer's protocol. The PCR amplification of E6 was as described by Alvarez-Salas et al. (1995); the amplification of E7 was as described by Fujinaga et al. (1994). E6 and E7 PCR products were sequenced directly using the dye dideoxy-termination method on an ABI 377 DNA sequencer (PE Applied Biosystems, Foster City, CA). The primers for sequencing reactions were the same used for the PCR amplifications. Both strands were sequenced and the five newly obtained sequences have been deposited in GenBank under accession numbers AY089951–AY089955.

Thirty-seven of the E6-E7 DNA sequences analyzed here are from Seedorf et al. (1985), Tornesello et al. (1997), Zehbe et al. (1998), and van Duin et al. (2000). We only use nonidentical E6–E7 haplotype sequences. The 42 total E6-E7 haplotypes were aligned by eye using GeneDoc version 2.5 (Nicholas and Nicholas 1997). There are no insertions/deletions. We used E6 codons 9–158 (corresponding to nucleotide numbers 110–556) and E7 codons 2–97 (corresponding to nucleotide numbers 565–855). Stop codons are not included in the analysis. Alignments are available from VRD upon request.

### *Phylogenies*

Phylogenies were reconstructed separately for E6 and E7. We used both the neighbor-joining algorithm (Saitou and Nei 1986) as implemented by the program PAUP version 4.0b3 (Swofford 1999), and a maximum-likelihood algorithm (Felsenstein 1981) as implemented by the program FastDNAml with transition/transversion ratio set to 2. All available pairwise distance estimation methods were used for implementing the neighbor-joining algorithm. Likelihood scores for all trees were compared following Hasegawa et al. (1991) and the gene tree with the highest likelihood score was used in the analysis. For the E6 gene, this tree is generated using FastDNAml; for E7, this tree is generated using the neighbor-joining method with distance estimation performed under the general reversible model. Tree topologies are available from VRD upon request.

### *Analysis of Positive Selection*

In order to determine whether positive selection impacts the evolution of E6 or E7, seven codon-substitution models were used. These models view the codon as the fundamental unit of evolutionary change and take into account genealogic history when calculating parameters. Log likelihood scores evaluate the quality of the fit of the input data to the conditions of the model. In these models, $d_n/d_s$ was estimated for separate classes of codons that are assumed to evolve independently of one another.

**Table 1.** Codon-based models used (after Yang et al. 2000)

| Model | Parameters | $d_n/d_s > 1$? |
|---|---|---|
| M0 (Invariant) | $d_n/d_s$ | Yes |
| M1 (Neutral) | $p_0$ | No |
| M2 (Selection) | $p_0$, $p_1$, $d_n/d_s$ | Yes |
| M3 (Discrete) | $p_0$, $p_1$, … $p_{K-2}$; $d_n/d_s(0)$ $d_n/d_{s(1)}$,…$d_n/d_{s(K-1)}$ | Yes |
| M5 (Gamma) | $\alpha$, $\beta$ | Yes |
| M7 (Beta) | $p$, $q$ | No |
| MS (Beta&$\omega$) | $p_{beta}$, $p$, $q$, $d_n/d_s$ | Yes |

$d_n/d_{s(i)}$ = Per site ratio of nonsynonymous substitutions to synonymous substitutions for site class $i$.
$p_n$ = Frequency of codons falling into site class n.
$\alpha$, $\beta$ = Gamma distribution shape parameters.
$p$, $q$ = Beta distribution shape parameters.
$p_{beta}$ = Frequency of codons falling into beta 0, 1 site class

When $d_n/d_s = 0$ for a particular codon site, it is assumed that nonsynonymous (i.e. amino acid) substitutions are eliminated by natural selection. When $d_n/d_s = 1$ it is assumed that amino acid changes have approximately the same probability of fixation as synonymous substitutions, which are selectively neutral with respect to protein evolution, so that the site is not subject to natural selection. When $d_n/d_s > 1$ it is assumed that nonsynonymous substitutions have a greater probability of fixation than synonymous substitutions and the site is evolving under the influence of diversifying selection, so that amino acid changes are adaptive. When $0 < d_n/d_s < 1$, it is assumed that amino acid changes are slightly deleterious (nearly neutral) and their fate is determined by a mixture of selection, mutation, and drift.

These seven codon models estimate different sets of parameters using different assumptions (see Table 1). M0 assumes that all codons have the same $d_n/d_s$, which is averaged over all sites (see Goldman and Yang 1994). M1 assumes that all sites fall into one of two classes (see Nielsen and Yang 1998). The first class is at frequency $p_0$ and has $d_n/d_{s(0)} = 0$, (no amino acid variation is tolerated), and the second class is at frequency $p_1 = 1 - p_0$ and has $d_n/d_{s(1)} = 1$. Model M2 allows for an additional class of codons at frequency $p_2$ for which $d_n/d_s$ may take any value between 0 and infinity. M3 separates all codons into $K$ different site classes, so that for each class $d_n/d_s$ can be between 0 and infinity and codons may fall into each class at any frequency (Yang et al. 2000). We started with $K = 3$ and increased the number of classes by one until the log likelihood did not improve. M5 assumes a discrete gamma distribution for $d_n/d_s$ among sites in the interval $(0, \infty)$ with shape parameter $\alpha$ (Yang et al. 2000). In this model, $d_n/d_s$ can also vary from 0 to infinity. M7 assumes a discrete beta distribution for $d_n/d_s$ among sites in the interval (0, 1) with shape parameters $p$ and $q$ (Yang et al. 2000). This model only allows $0 \leq d_n/d_s \leq 1$. M8 is a more general version of M7 that allows for an additional class of sites with $d_n/d_s > 1$ and thus uses two site classes (Yang et al. 2000). Codons in site class 1 (at frequency $p_0$) are assumed to have beta-distributed $d_n/d_s$ values between 0 and 1, and codons in site class 2 (at frequency $1 - p_0$) are assumed to have the same $d_n/d_s$ that can vary between 0 and infinity. All models were implemented using the program CODEML in the PAML software package version 3.0c (Yang 2001).

In order to assess the influence of positive selection on a particular coding region, a likelihood ratio test (LRT) is performed to compare nested models (Nielsen and Yang 1998, Yang et al. 2000). All models involve estimates of genealogy branch lengths, transition/transversion rate ratio ($\kappa$), and nucleotide frequencies. Specific models estimate additional parameters such as $d_n/d_s$, site class frequencies, and distribution shape parameters. Some models are less general equivalents of others, since they estimate subsets of parameters (e.g. those that do not allow for positively selected sites are less general than those that do). In this case, twice the log likelihood difference between the two models follows a $\chi^2$ distribution with degrees of freedom equal to the difference in the number of parameters estimated between the models. For instance, M2 and M3 are more general versions of M1. When $d_n/d_s$ estimates are greater than 1 in the more general models and these have significantly better likelihood scores, the existence of positively selected sites is probable (i.e. a null hypothesis of no positively selected sites is rejected). Likewise, M7 is nested within M8. When $d_n/d_s > 1$ in M8, the LRT tests the alternative hypothesis of the presence of positively selected sites. Anisimova and colleagues (2001) have analyzed the effectiveness of this approach and found that the LRT is powerful when used under similar conditions including sequence length, number of taxa, and level of divergence.

## Results

Table 2 shows the results of the likelihood analysis, including parameter estimates for different models. Table 3 gives the results of the positive selection LRTs comparing nested models within each gene. All tests for positively selected sites within E6 rejected a null hypothesis of no such sites. All tests for positively selected sites within E7 fail to reject a null hypothesis of no such sites. Results from M3 with $K = 3$ site classes are given since K > 3 did not improve the likelihood.

For E6, M3 with $K = 3$ gave the highest likelihood. This model indicates that approximately 3.22% of the E6 codon sites fall into a site class with $d_n/d_s \approx 9.8994$. Furthermore, both M5 and M8 indicate that positively selected sites are present in E6. Four codon sites were consistently shown to have $d_n/d_s > 1$. These are E6 codons 17, 21, 34, and 90 (see Figure 1). Table 4 displays the amino acids observed in those positions, their frequencies, and their biochemical characteristics. For E7, M8 gave the highest likelihood, but no sites with $d_n/d_s > 1$ were detected under any model. Additionally, the M8 likelihood score is not significantly better than for any other model.

In order to examine qualitative amino acid diversification at codon sites identified by PAML as likely to be under positive selection in E6, we employed the diversification method of Hughes et al. (1990). This method characterizes nonsynonymous substitutions

**Table 2.** Results of PAML estimation for E6 and E7 genes

| Gene | Model | ln Likelihood | Parameter Estimates |
|---|---|---|---|
| E6 | M0 | −974.780024 | $d_n/d_s = 0.5622$ |
| | M1 | −961.838218 | $p_0 = 0.7312$ |
| | M2 | −950.996257 | $p_0 = 0.6422$; $p_1 = 0.3279$; $d_n/d_s = 11.6461$ |
| | M3 (K = 3) | −950.405998 | $p_0 = 0.5989$; $p_1 = 0.3689$; $d_n/d_{s(0)} = 0.4295$; $d_n/d_{s(1)} = 0.001$; $d_n/d_{s(2)} = 9.8994$ |
| | M5 | −960.435008 | $\alpha = 0.0574$; $\beta = 0.0133$ |
| | M7 | −962.219003 | $p = 0.0178$; $q = 0.0376$ |
| | M8 | −950.425030 | $p_{beta} = 0.9678$; $p = 0.6583$; $q = 1.6973$; $d_n/d_s = 9.8858$ |
| E7 | M0 | −457.651984 | $d_n/d_s = 0.2755$ |
| | M1 | −457.866279 | $p_0 = 0.6788$ |
| | M2 | −457.648022 | $p_0 = 0.0000$; $p_1 = 0.0000$; $\omega = 0.4011$ |
| | M3 (K = 3) | −457.648066 | $p_0 = 0.3296$; $p_1 = 0.3329$;. $d_n/d_{s(0)} = 0.2755$; $d_n/d_{s(1)} = 0.2755$; $d_n/d_{s(2)} = 0.2755$ |
| | M5 | −457.651936 | $\alpha = 115.3644$; $\beta = 416.1659$ |
| | M7 | −457.653412 | $p = 121.3345$; $q = 289.0995$ |
| | M8 | −457.649207 | $p_{beta} = 0.9999$; $p = 47.9240$; $q = 125.6978$; $d_n/d_s = 0.2900$ |

**Table 3.** Results of likelihood ratio tests of positive selection

| Gene | Model 1 | Model 2 | 2 (Lik. Diff.) | *p* value | $d_n/d_s > 1$ Detected? | Positively Selected Codons |
|---|---|---|---|---|---|---|
| E6 (150 codons) | M1 | M2 | 21.683922 | 0.00001956 | Yes | 17 [$P(d_n/d_s > 1) = 0.9997$]<br>21 [$P(d_n/d_s > 1) = 0.9969$]<br>34 [$P(d_n/d_s > 1) = 0.9341$]<br>90 [$P(d_n/d_s > 1) = 0.9380$] |
| | M1 | M3 | 22.86444 | 0.00013477 | Yes | 17 [$P(d_n/d_s > 1) = 1.0000$]<br>21 [$P(d_n/d_s > 1) = 0.9998$]<br>34 [$P(d_n/d_s > 1) = 0.9921$]<br>90 [$P(d_n/d_s > 1) = 0.9854$] |
| | M7 | M8 | 23.587946 | 0.00003045 | Yes | 17 [$P(d_n/d_s > 1) = 1.0000$]<br>21 [$P(d_n/d_s > 1) = 0.9997$]<br>34 [$P(d_n/d_s > 1) = 0.9890$]<br>90 [$P(d_n/d_s > 1) = 0.9831$] |
| E7 (97 codons) | M1 | M2 | 0.436514 | 0.80391881 | No | None |
| | M1 | M3 | 0.436426 | 0.97938744 | No | None |
| | M7 | M8 | 0.008410 | 0.99979539 | No | None |

with respect to charge, polarity, or other user-specified property. It allows numerical comparison of frequency of substitutions that result in ''radical'' versus ''conservative'' changes in amino acids with respect to the property of interest. Diversification analysis of E6 codons revealed no significant excesses of radical amino acid substitutions (data not shown). In other words, radical amino acid substitutions were not more likely to occur than conservative substitutions at these sites. This applies to changes in amino acid charge, polarity, and hydrophobicity. Our analysis was performed on all positively selected sites using all taxa, unique positively selected site haplotypes, site 17 only (see below), and site 90 only (see below).

## Discussion

It is likely that E6 contains codon sites that are evolving under the influence of diversifying selection. E7 does not appear to contain any such sites, despite the presence of amino acid variation. It is of interest to note that while genealogical techniques have uncovered numerous proteins from ribo- and retroviruses that experience diversifying selection this is the first such demonstration in a DNA virus. The processes that account for directional selection operating on E6 are difficult to identify experimentally due mostly to the complex reproductive strategy of the virus and the unknown characteristics of the individual host environments from which these haplotypes were obtained. In addition, complications associated with culturing the virus *in vitro*, as well as its low mutation rate, effectively prevent experimental examination of its evolutionary dynamics. Nevertheless, a significant amount of information is known about the functions of E6 and E7 with respect to viral replication and therefore we will focus on two aspects that may be related to external selective forces: Immunogenicity and functional interaction of viral and cellular proteins.

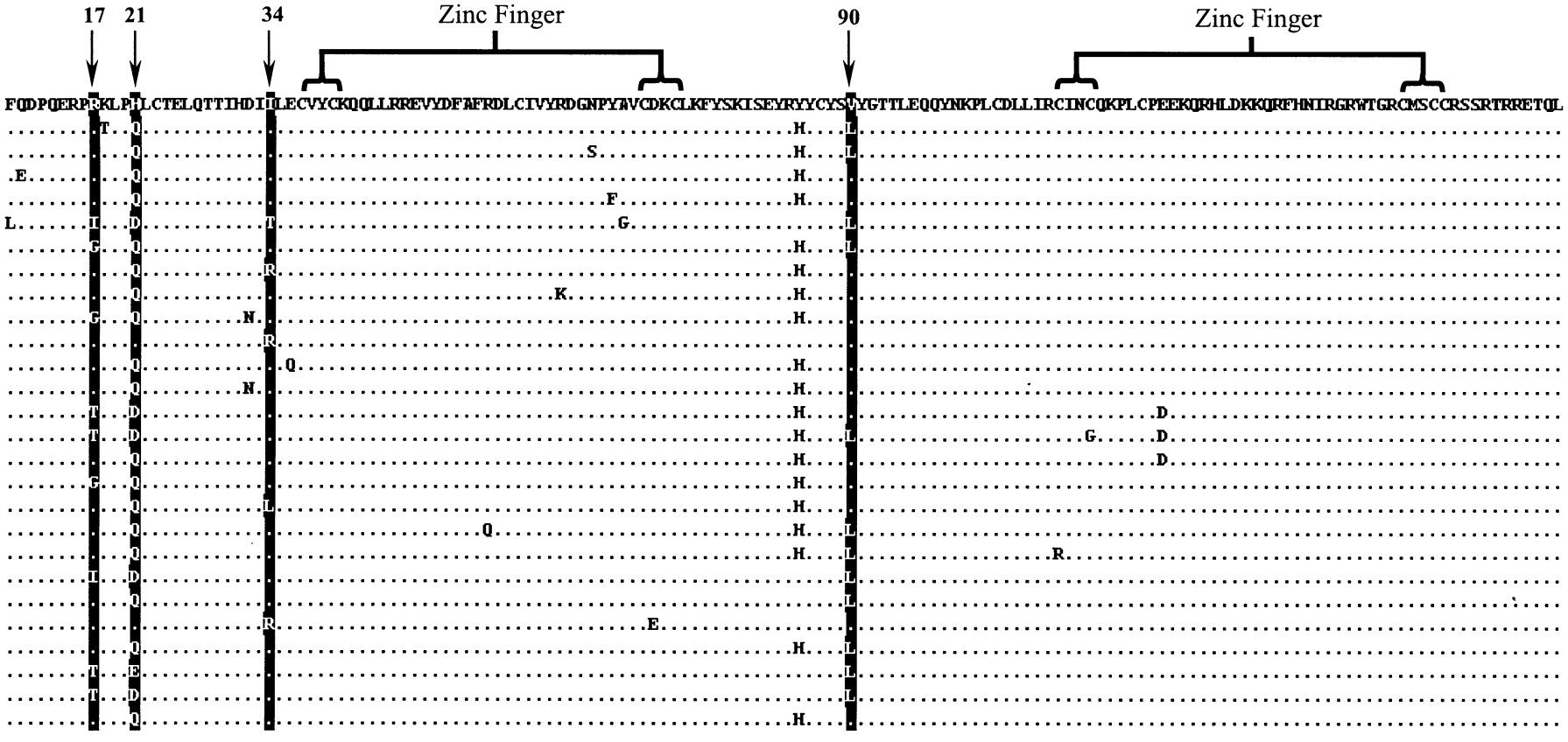


**Fig. 1.** Alignment of unique E6 amino acid sequences used in this study (sites 9–158). Sites identified as having $d_n/d_s > 1$ are indicated.

**Table 4.** Amino acids observed in positively selected E6 sites and their corresponding biochemical traits

| Site | Amino Acid | Frequency | Charge | Polarity | Hydrophobicity |
|---|---|---|---|---|---|
| 17 | R | 28 | Positive | Polar | Hydrophilic |
| | G | 7 | Neutral | Polar | Hydrophilic |
| | T | 6 | Neutral | Polar | Hydrophilic |
| | I | 2 | Neutral | Non-polar | Hydrophobic |
| 21 | Q | 32 | Neutral | Polar | Hydrophilic |
| | D | 7 | Negative | Polar | Hydrophilic/Acidic |
| | H | 3 | Positive | Polar | Hydrophilic |
| | E | 1 | Negative | Polar | Hydrophilic/Acidic |
| 34 | I | 38 | Neutral | Non-polar | Hydrophobic |
| | R | 3 | Positive | Polar | Hydrophilic |
| | L | 1 | Neutral | Non-polar | Hydrophobic |
| | T | 1 | Neutral | Polar | Hydrophilic |
| 90 | V | 25 | Neutral | Non-polar | Hydrophobic |
| | L | 18 | Neutral | Non-polar | Hydrophobic |

*Immune Selection*

Immunogenic peptides [specifically cytotoxic T-lymphocye (CTL) epitopes] that exist in proteins of intracellular parasites may be subjected to diversifying selection imposed by cell-mediated immune responses. Protein variants containing epitopes against which a host can react will experience a smaller probability of surviving T-cell attack than a coexisting protein variant that does not (see Gould and Bangham 1998 for review). Selective pressure can thus direct amino acid evolution in viral proteins away from sequences capable of being recognized by a particular host or host population. This has been demonstrated mostly in viruses with RNA genomes that are capable of rapid, intrahost evolutionary change (see Phillips et al. 1991, Nowak et al. 1995, Wolinsky et al. 1996, McMichael and Phillips 1997, da Silva and Hughes 1999, Kuiken et al. 1999, Haydon et al. 2001).

Intrahost variation is essentially nonexistent in DNA viruses such as HPVs that use host enzymes for genome replication. Yet, while HPVs evolve slowly compared to ribo- and retroviruses, they evolve much more rapidly than humans. An "equilibrium" distribution of amino acid sequences that maximizes the survival of a virus population in the context of a particular host MHC gene pool could potentially arise. Evidence suggests that Epstein-Barr virus (EBV), a DNA virus, experiences local allele frequency alteration in response to host population MHC makeup (de Campos-Lima et al. 1993). The long-term effect of MHC makeup on $d_n/d_s$ ratios within EBV CTL epitopes may not be driving amino acid change, however (Khanna et al. 1997).

If the observed diversification of the E6 gene is driven by immune selection, it might be expected that the E6 protein would be highly immunogenic relative to E7. While a number of studies have identified CTL epitopes within both E6 and E7 proteins (in fact, few amino acids in either protein fall outside known CTL epitopes), it has not been demonstrated that either protein is more or less immunogenic than the other (see Dillner 1990, Strang et al. 1990, Comerford et al. 1991, Altmann et al. 1992, Kast et al. 1994, Sarkar et al. 1995, Bauer et al. 1995, Ressing et al. 1995, Dunn et al. 1997).

One of the E6 amino acid sites identified as evolving under positive selection (site 17) is potentially very informative. We have found four amino acids in this position (R = 64.3%, G = 16.7%, T = 14.3%, and I = 4.8%). Under M3, the maximum likelihood $d_n/d_s$ estimate for this site is ~9.9. Ellis and colleagues (1995) found a small but significant association between HLA-B7 positive individuals and an E6 mutant at this amino acid position (R → G). This site is located within a HLA-B7 CTL epitope, and while the mutation does not affect HLA binding it drastically alters the epitope's exposure to $CD8^+$ T-cell receptors (Ellis et al. 1995). This may result in different survival probabilities of strains related to (1) host HLA-B makeup and (2) amino acid composition of HLA-B viral epitopes. In other words, certain E6 site 17 variants may be better adapted for replication within hosts of a specific HLA makeup. Bontkes and colleagues (1998) failed to detect a significant association of any E6 or E7 variants with specific HLA types in a sample of 40 cancer patients. Thus it is possible that immune selection is only a localized phenomenon, if it exists at all.

Protein-protein interactions, such as those between pathogen epitopes and MHC proteins are influenced by amino acid charge (Monos et al. 1984). Hughes and Hughes (1995) present evidence of greater than expected amino acid charge variation in the peptide binding regions of human class I HLA molecules. In order to explore further this possibility we examined the biochemical extremism of specific amino acid

changes that took place in E6 codons identified as diversifying. The results indicate that diversification of these sites is essentially random with respect to charge, polarity, and hydrophobicity. While this in no way proves immune selection is not influencing evolution at these sites, when coupled with the results of Bontkes et al. (1998), it downplays the significance of avoidance of epitope binding in the viral life cycle.

*Viral-Cellular Protein Functional Interaction*

E6 and E7 are responsible for direct interaction with cellular proteins p53 and p105RB respectively (Scheffner et al. 1990, Werness et al. 1990, Huibregtse et al. 1991, Scheffner et al. 1993, Dyson et al. 1989, Jones et al. 1997). The specific mechanisms of E6 and E7 interaction with other cellular proteins are not well researched. Naturally occurring polymorphism in these host proteins may promote polymorphism in viral proteins to allow for successful binding. In contrast with selection to discourage epitope-MHC binding, the function of this variation is to *facilitate* protein–protein binding. As such, we might expect less radical diversification of amino acids with respect to biochemical traits such as charge.

p53 is in fact known to be polymorphic in human populations, and to vary according to latitude and ethnicity (Beckman et al. 1994), but p105Rb is nearly monomorphic (Yandell et al. 1989, Schubert and Hansen 1996). One particular p53 substitution (codon 72 R $\rightarrow$ P) is the subject of most polymorphism analyses. Van Duin and colleagues (2000) uncovered a significant association between cervical cancer patients homozygous for the p53 codon 72 R variant and the HPV16 E6 codon 90 L variant. This E6 amino acid site appears in our study to be evolving under diversifying selection but the two amino acids observed here (V and L) are not radically different in charge, polarity, or hydrophobicity (these amino acids are also similar in weight and surface area). Interestingly, p53 protein variants have been shown to be differentially susceptible to E6-induced degradation based on amino acid site 72 composition (Storey et al. 1998). Therefore, selective pressure driven by ability to interact with p53 is likely to exist.

Recently, E6 and E7 proteins have been the foci of efforts to develop therapeutic vaccines for cervical cancer, since they are overexpressed in tumor cells (see Murakami et al. 1999 for review). In addition, E6 and E7 have been considered potential molecular targets for HPV antiviral treatments (see Phelps et al. 1998 for review). The evolutionary potential of E6 and E7 could have serious implications for the efficacy and development of epitope-based vaccination strategies and of antiviral medications. Therefore, such topics should be of great importance to researchers investigating these genes for therapeutic reasons. Site-directed mutagenesis studies targeting positively selected sites could provide invaluable information regarding virus-host interactions and the usefulness of targeting this protein in anti-HPV therapy.

*Acknowledgments.* We thank Ziheng Yang for his useful comments and advice regarding PAML usage. We also thank Austin Hughes for his statistical assistance and for kindly providing software to analyze amino acid diversity. This research was supported by the UC Irvine Center for Virus Research and NIH Minority International Research Training grant TW-00023

## References

Altmann A, Jochmus-Kudielka I, Frank R, Gausepohl H, Moebius U, Gissmann L, Meuer SC (1992) Definition of immunogenic determinants of the human papillomavirus type 16 nucleoprotein E7. Eur J Cancer 28:326–333

Alvarez-Salas LM, Wilczynski SP, Burger RA, Monk BJ, Dipaolo JA (1995) Polymorphism of the HPV-16 E6 gene of cervical carcinoma. Int J Oncol 7:261–266

Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585–1592

Baker CC, Phelps WC, Lindgren V, Braun MJ, Gonda MA, Howley PM (1987) Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. J Virol 61:962–971

Bauer S, Heeg K, Wagner H, Lipford GB (1995) Identification of H-2Kb binding and immunogenic peptides from human papilloma virus tumour antigens E6 and E7. Scand J Immunol 42: 317–323

Beckman G, Birgander R, Sjalander A, Saha N, Holmberg PA, Kivela A, Beckman L (1994) Is p53 polymorphism maintained by natural selection? Hum Hered 44:266–270

Berumen J, Ordonez RM, Lazcano E, Salmeron J, Galvan SC, Estrada RA, Yunes E, et al. (2001) Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study. J Natl Cancer Inst 93:1325–1330

Bontkes HJ, van Duin M, de Gruijl TD, Duggan-Keen MF, Walboomers JM, Stukart MJ, et al. (1998) HPV 16 infection and progression of cervical intra-epithelial neoplasia: analysis of HLA polymorphism and HPV 16 E6 sequence variants. Int J Cancer 78:166–171

Bosch F, Manos M, Muñoz N, Sherman M, Jansen A, Peto J, et al. (1995) Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. J Natl Cancer Inst 87:796–802

Brady CS, Duggan-Keen MF, Davidson JA, Varley JM, Stern PL (1999) Human papillomavirus type 16 E6 variants in cervical carcinoma: relationship to host genetic factors and clinical parameters. J Gen Virol 80:3233–3240

Chan SY, Ho L, Ong CK, Chow V, Drescher B, Durst M, et al. (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

Comerford SA, McCance DJ, Dougan G, Tite JP (1991) Identification of T- and B-cell epitopes of the E7 protein of human papillomavirus type 16. J Virol 65:4681–4690

da Silva J, Hughes AL (1999) Molecular phylogenetic evidence of cytotoxic T lymphocyte (CTL) selection on human immunodeficiency virus type 1 (HIV-1) [Letter]. Mol Biol Evol 16: 1420–1422

De Campos-Lima P-O, Gavioli R, Zhang Q-J, Wallace LE, Dolcetti R, Rowe M, et al. (1993) HLA-A11 epitope loss iso-

lates of Epstein-Barr virus from a highly A11+ population. Science 260:98–100

Dillner J (1990) Mapping of linear epitopes of human papillomavirus type 16: the E1, E2, E4, E5, E6 and E7 open reading frames. Int J Cancer 46:703–711

Dunn LA, Evander M, Tindle RW, Bulloch AL, de Kluyver RL, Fernando GJ, et al. (1997) Presentation of the HPV16E7 protein by skin grafts is insufficient to allow graft rejection in an E7-primed animal. Virology 235:94–103

Dyson N, Howley PM, Munger K, Harlow E (1989) The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science 243:934–937

Ellis JR, Keating PJ, Baird J, Hounsell EF, Renouf DV, Rowe M, et al. (1995) The association of an HPV16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. Nat Med 1:464–470

Eriksson A, Herron JR, Yamada T, Wheeler CM (1999) Human papillomavirus type 16 variant lineages characterized by nucleotide sequence analysis of the E5 coding segment and the E2 hinge region. J Gen Virol 80:595–600

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Fujinaga Y, Okazawa K, Nishikawa A, Yamakawa Y, Fukushima M, Kato I, Fujinaga K (1994) Sequence variation of human papillomavirus type 16 E7 in preinvasive and invasive cervical neoplasias. Virus Genes 9:85–92

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Gould KG, Bangham CR (1998) Virus variation, escape from cytotoxic T lymphocytes and human retroviral persistence. Semin Cell Dev Biol 9:321–328

Hasegawa M, Kishino H, Saitou N (1991) On the maximum likelihood method in molecular phylogenetics [comment]. J Mol Evol 32:443–445

Hawley-Nelson P, Vousden K, Hubbert N, Lowy D, Schiller J (1989) HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. EMBO J 8:3905–3910

Haydon DT, Bastos AD, Knowles NJ, Samuel AR (2001) Evidence for positive selection in foot-and-mouth diesease virus capsid genes from field isolates. Genetics 157:7–15

Ho L, Chan SY, Chow V, Chong T, Tay SK, Villa LL, Bernard HU (1991) Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. J Clin Microbiol 29:1765–1772

Hu X, Pang T, Guo Z, Mazurenko N, Kisseljov F, Ponten J, Nister M (2001) HPV16 E6 gene variations in invasive cervical squamous cell carcinoma and cancer in situ from Russian patients. Br J Cancer 84:791–795

Hughes AL, Hughes MK (1995) Natural selection on the peptide-binding regions of major histocompatibility complex molecules. Immunogenetics 42:233–243

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962

Hughes AL, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol Biol Evol 7:515–524

Huibregtse JM, Scheffner M, Howley PM (1991) A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. EMBO J 10:4129–4135

Jones DL, Thompson DA, Munger K (1997) Destabilization of the RB tumor suppressor protein and stabilization of p53 contribute to HPV type 16 E7-induced apoptosis. Virology 239:97–107

Kast WM, Brandt RM, Sidney J, Drijfhout JW, Kubo RT, Grey HM, et al. (1994) Role of HLA-A motifs in identification of potential CTL epitopes in human papillomavirus type 16 E6 and E7 proteins. J Immunol 152:3904–3912

Khanna R, Slade RW, Poulsen L, Moss DJ, Burrows SR, Nicholls J, Burrows JM (1997) Evolutionary dynamics of genetic variation in Epstein-Barr virus isolates of diverse geographical origins: Evidence for immune pressure-independent genetic drift. J Virol 71:8340–8346

Klingelhutz AJ, Foster SA, McDougall JK (1996) Telomerase activation by the E6 gene product of human papillomavirus type 16. Nature (London) 380:79–82

Kuiken CL, Foley B, Guzman E, Korber BTM (1999) Determinants of HIV-1 protein evolution. In: Crandall KA (ed) The Evolution of HIV. Johns Hopkins University Press, Baltimore, pp 432–468

Lambert PF, Pan H, Pitot HC, Liem A, Jackson M, Griep AE (1993) Epidermal cancer associated with expression of human papillomavirus type 16 E6 and E7 oncogenes in the skin of-transgenic mice. Proc Natl Acad Sci USA 90:5583–5587

Lechner MS, Mack DH, Finicle AB, Crook T, Vousden KH, Laimins LA (1992) Human papillomavirus E6 proteins bind p53 in vivo and abrogate p53-mediated repression of transcription. EMBO J 11:3045–3052

Lee SS, Weiss RS, Javier RT (1997) Binding of human virus oncoproteins to hD1g/SAP97, a mammalian homolog of the Drosophila discs large tumor suppressor protein. Proc Nat Acad Sci USA 94:6670–6675

McMichael AJ, Phillips RE (1997) Escape of human immunodeficiency virus from immune control. Annu Rev Immunol 15: 271–296

Monos DS, Tekolf WA, Shaw S, Cooper HL (1984) Comparison of structural and functional variation in class I HLA molecules: the role of charged amino acid substitutions. J Immunol 132:1379–1385

Münger K, Phelps W, Bubb V, Howley P, Schlegel R (1989) The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. J Virol 63:4417–4421

Murakami M, Gurski KJ, Steller MA (1999) Human papillomavirus vaccines for cervical cancer. J Immunother 22:212–218

Nicholas KB, Nicholas HB (1997) GeneDoc: A tool for editing and annotating multiple sequence alignments. Distributed by the author: www.psc.edu/biomed/genedoc

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

Nowak MA, May RM, Phillips RE, Rowland-Jones S, Lalloo DG, McAdam S, et al. (1995) Antigenic oscillations and shifting immunodominance in HIV-1 infections [see comments]. Nature 375:606–611

Phelps WC, Barnes JA, Lobe DC (1998) Molecular targets for human papillomaviruses: prospects for antiviral therapy. Antivir Chem Chemother 9:359–377

Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, et al. (1991) Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition [see comments]. Nature 354:453–459

Pim D, Storey A, Thomas M, Massimi P, Banks L (1994) Mutational analysis of HPV-18 E6 identifies domains required for p53 degradation in vitro, abolition of p53 transactivation in vivo and immortalisation of primary BMK cells. Oncogene 9:1869–1876

Ressing ME, Sette A, Brandt RM, Ruppert J, Wentworth PA, Hartman M, et al. (1995) Human CTL epitopes encoded by human papillomavirus type 16 E6 and E7 identified through in vivo and in vitro immunogenicity studies of HLA-A*0201-binding peptides. J Immunol 154:5934–5943

Saitou N, Nei M (1986) The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. J Mol Evol 24:189–204

Sarkar AK, Tortolero-Luna G, Nehete PN, Arlinghaus RB, Mitchell MF, Sastry KJ (1995) Studies on in vivo induction of cytotoxic T lymphocyte responses by synthetic peptides from E6 and E7 oncoproteins of human papillomavirus type 16. Viral Immunol 8:165–174

Scheffner M, Huibregtse JM, Vierstra RD, Howley PM (1993) The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell 75:495–505

Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM (1990) The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell 63: 1129–1136

Schubert EL, Hansen MF (1996) A previously unknown polymorphism located within the RB1 locus only present in Asian individuals. Hum Hered 46:118–120

Schwarz E, Freese UK, Gissmann L, Mayer W, Roggenbuck B, Stremlau A, zur Hausen H (1985) Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. Nature 314:111–114

Seedorf K, Krammer G, Durst M, Suhai S, Rowekamp WG (1985) Human papillomavirus type 16 DNA sequence. Virology 145:181–185

Smotkin D, Wettstein FO (1986) Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. Proc Natl Acad Sci USA 83:4680–4684

Stoppler MC, Ching K, Stoppler H, Clancy K, Schlegel R, Icenogle J (1996) Natural variants of the human papillomavirus type 16 E6 protein differ in their abilities to alter keratinocyte differentiation and to induce p53 degradation. J Virol 70:6987–6993

Storey A, Thomas M, Kalita A, Harwood C, Gardiol D, Mantovani F, Breuer J, et al. (1998) Role of a p53 polymorphism in the development of human papillomavirus-associated cancer [see comments]. Nature 393:229–234

Strang G, Hickling JK, McIndoe GA, Howland K, Wilkinson D, Ikeda H, Rothbard JB (1990) Human T cell responses to human papillomavirus type 16 L1 and E6 synthetic peptides: identification of T cell determinants, HLA-DR restriction and virus type specificity. J Gen Virol 71:423–431

Swofford DL (1999) PAUR*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland, MA

Syrjanen K, Syrjanen S (2000) Papillomavirus Infections in Human Pathology. John Wiley and Sons, Ltd., West Sussex, UK

Tong X, Howley PM (1997) The bovine papillomavirus E6 oncoprotein interacts with paxillin and disrupts the actin cytoskeleton. Proc Natl Acad Sci USA 94:4412–4417

Tornesello ML, Buonaguro FM, Meglio A, Buonaguro L, Beth-Giraldo E, Giraldo G (1997) Sequence variations and viral genomic state of human papillomavirus type 16 in penile carcinomas from Ugandan patients. J Gen Virol 78:2199–2208

van Duin M, Snijders PJF, Vossen MTM, Klaassen E, Voorhorst F, Verheijen RHM, et al. (2000) Analysis of human papillomavirus type 16 E6 variants in relation to p53 codon 72 polymorphism genotypes in cervical carcinogenesis. J Gen Virol 81:317–325

Werness BA, Levine AJ, Howley PM (1990) Association of human papillomavirus types 16 and 18 E6 proteins with p53. Science 248:76–79

Wheeler CM, Yamada T, Hildesheim A, Jenison SA (1997) Human papillomavirus type 16 sequence variants: identification by E6 and L1 lineage-specific hybridization. J Clin Microbiol 35: 11–19

Wolinsky SM, Korber BT, Neumann AU, Daniels M, Kunstman KJ, Whetsell AJ, et al. (1996) Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection [see comments]. Science 272:537–542

Xi LF, Koutsky LA, Galloway DA, Kuypers J, Hughes JP, Wheeler CM, et al. (1997) Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia [see comments]. J Natl Cancer Inst 89:796–802

Yamada T, Manos MM, Peto J, Greer CE, Munoz N, Bosch FX, Wheeler CM (1997) Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. J Virol 71:2463–2472

Yandell DW, Campbell TA, Dayton SH, Petersen R, Walton D, Little JB, et al. (1989) Oncogenic point mutations in the human retinoblastoma gene: their application to genetic counseling [see comments]. N Engl J Med 321:1689–1695

Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yang Z (2000) Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. J Mol Evol 51:423–432

Yang Z (2001) Phylogenetic analysis by maximum likelihood (PAML), London

Yang Z, Nielsen R, Goldman N, Pedersen A-MK (2000) Codon-Substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Zehbe I, Wilander E, Delius H, Tommasino M (1998) Human papillomavirus 16 E6 variants are more prevalent in invasive cervical carcinoma than the prototype. Cancer Res 58:829–833

Zwerschke W, Jansen-Durr P (2000) Cell transformation by the E7 oncoprotein of human papillomavirus type 16: interactions with nuclear and cytoplasmic target proteins. Adv Cancer Res 78: 1–29