Sequencing Errors in Reference HPV Clones

John Meissner

Duke University Medical Center
e-mail: jdmeissner@hotmail.com

"We may fling ourselves into a hammock in a fit of divine carelessness, but we are glad that the netmaker did not make the hammock in a fit of divine carelessness."

—G. K. Chesterton

The first papillomavirus (PV) nucleotide sequences were published fifteen years ago [7,10]. Since then, more than one million bases of papillomavirus sequence have been deposited in GenBank [Dennis Benson, personal communication]. With the number of novel papillomavirus isolates and genomic sequences starting to plateau [29], this is an ideal time for the papillomavirus community to carefully examine problematic clones and redact lingering sequencing errors, thus giving future clinicians and researchers the chance to throw themselves into the challenges of papillomavirus typing or vaccine development without having to worry about the strength of the underlying sequence data.

As indicated by the title, this review has a limited focus. Only reference human papillomavirus (HPV) clones will be examined, and of these, only a few—HPV-1a, HPV-5a, HPV-6b, HPV-16, HPV-18, HPV-35, and HPV-59—will be examined in any detail. The fundamental importance of trustworthy reference sequences does not require justification, nor should the primacy of reference HPV sequence corrections over (and of these reference HPV sequences in particular over) correction of errors in animal PV or nonreference HPV sequences. All three groups of PV sequences have errors. The emphasis on correcting errors in reference HPV clones *first* relates to practical considerations—immediate applicability of corrections, template availability for resequencing—as well as absolute numbers. Furthermore, nonreference sequences can be influenced by sequencing errors in reference clones, and this error propagation is unidirectional. The nucleotide (nt) at position 2926 (numbering based on HPV-16R sequence [28]) of the reference HPV-16 clone, for example, is G [44, *vide infra*], not A, and the nucleotide at the corresponding position of the HPV-16 variant present in SiHa cells is also G, not A, despite what two labs examining HPV integration in SiHa cells independently reported [1, 17, GenBank Accession number AF001599].

Fifteen years ago, sequencing a papillomavirus genome was both a significant undertaking and a remarkable achievement. Reagents and enzymes did not come as numbered tubes in quality-controlled kits. Equipment was basic. Techniques were still being worked out. That these initial sequence determinations contained as few errors as they did—produced under the added pressure of finishing before competing laboratories—is a testament to the scientific abilities of the researchers. Without question, these pioneering efforts should be judged by a different standard. Without resequencing, though, these uncorrected sequences will continue to generate unneccesary confusion.

Today, given the latest automated sequencing apparatus and the appropriate oligonucleotide primers, sequencing a complete HPV genome is a one-week task. Ease of resequencing is obviously not a justification for resequencing, and even if one reference clone a day could be resequenced starting from HPV-1a, no lab would devote the resources to such an undertaking. Most in the papillomavirus community would resent even the suggestion that reference clones up to some numerical or chronological cut-off, or beyond those already corrected, should be completely resequenced. Error-free reference sequences would be at risk of having artifactual corrections forced upon them, and even if true errors were found, the overall importance of these errors might not justify the cost. These concerns are not unwarranted, but, at least so far, they are not supported by available data or defined criteria.

While it is hoped that improved sequencing methods and reagents, accumulated knowledge of invariant properties of HPV genomes, and availability of sequence information from related types guarantee the accuracy of reference "HPV-X" sequence (where X represents a cloned HPV genome whose sequence was determined within the last decade, for example, or one whose sequence is supported by reisolation in an epidemiologic survey, at least over the region sequenced), the unfortunate truth is that no reference HPV clone, regardless of when it was initially sequenced, has proven to be 100% error-free on resequencing, and the first error reported in a reference sequence is generally not the final one.

Admittedly, a selection bias exists, i.e., those clones for which sequence errors are *a priori* suggested by contemporaneous sequencing of the same clone or by disagreements with experimentally determined restriction sites or correctly sequenced variants and closely related types are chosen for resequencing, thus every reference clone examined has sequencing errors. It could be argued, however, that the choice of every reference clone examined thus far has been based solely on whatever genotype a particular group looks for in epidemiologic surveys or whatever reference clone two groups happen to sequence simultaneously—that is, the choice has been somewhat arbitrary. Thus, up to this point, the reference clones that have sequencing errors are those reference clones subject to closer scrutiny.

The difficulty with calling unreported sequencing errors in reference clones "unimportant" is that it assumes that egregious errors are self-correcting, either in the process of assembling the sequence or at the behest of a thorough review. Yet today's reported errors are yesterday's unreported ones, and if the HPV major capsid (L1) protein is any example, this faith in self-correction is not well-founded. Since the L1 protein is involved in receptor interaction, virion assembly, and elicitation of antibody response [28], and because portions of the L1 open reading frame (ORF) are amplified in PCR-based HPV detection and typing protocols [49 and references therein], it is obviously advantageous for this region of the genome be sequenced correctly in reference clones. Nothing about this region of the HPV genome makes it inherently difficult to sequence: It does not have high G+C content, it does not adopt stable secondary structure(s), and guidance for self-correction is provided by amino acid alignment with previously sequenced reference clones, with strict conservation of up to 20% of total amino acid residues for all HPV types [44]. Despite this, the literature and the databanks contain reference clone L1 ORF sequences with artifactual premature terminations, artifactual amino acid insertions and deletions, and artifactual nonconservative amino acid changes [19, 47, 61].

Furthermore, this assessment of "unimportance" implies the existence of an absolute scale of importance for each nucleotide position of papillomavirus genomes. Such a scale could be proposed—indeed, all site-directed mutagenesis at the bench or sequence alignment at the keyboard is directed at some level towards determining a ranking of inviolate or invariant amino acids or nucleotides at specific positions—these studies as well as the practical issue of HPV diagnostics could serve as a foundation for assigning precedence to resequencing proposals. A case could be made for resequencing and correcting errors only in regions of recognized phenotypic and functional significance in reference clones, for example. Under such a proposal, correction of phenotypically silent errors would be given lower priority unless, returning to the L1 ORF, they artifactually altered a recognized splice site or restriction site or potential methylation site or influenced design of degenerate primer sets used in HPV typing. Problems would arise in resequencing of noncoding regions or ORFs whose functions were still being delineated. Restoration of a dropped CG dinucleotide after nt 7717 of the reference HPV-11 sequence [11, 24], for example, has a different impact on different researchers (see Matsukura and Sugase [41]—it is this change, not the one they propose, which creates the "missing" *Ban*I site in the reference HPV-11 clone), and an unpredictable impact on the HPV-11 genome. A string of 11 nucleotides (TAAAACGAAAG) strictly conserved at the 5′ end of the E1 ORF of most if not all HPVs [6, 13] in a region of otherwise low nucleotide conservation across HPV types [29] is likely preserved for reasons beyond those of coding potential, since any of 95 other combinations of nucleotides could accomplish the same task.

One final argument against blanket resequencing of reference clones is that the some reference sequences, even if corrected, may prove to be irrelevant. Certain reference clones, for example the reference HPV-14 clone [13, 36], may not be genuine representations of genomes present in the patient material from which they were isolated. Others may be representative of (a minority or majority of)

genomes present in the sample but nonrepresentative of circulating, infectious genomes [34]. Except for the reference HPV-6b clone [14], the authenticity or representativeness of particular reference clones will not be addressed in this review, in part because these distinctions are based on conjecture and limited prevalence studies, in part because an incorrectly sequenced reference clone is, by definition, representative of nothing. The polymorphism (what deVilliers refers to as the "genomic constellation" [15]) of certain HPV genotypes across patients or across continents may render any single reference clone nonrepresentative. A similar constellation that receives little attention is the marked heterogeneity of HPV type-specific genomes visible within patient material from which reference clones were isolated and continue to be isolated [3, 16, 38]. (This heterogeneity should not be confused with the heterogeneity of HPV genomes present in dually-infected or pooled clinical material, from which isolation of reference clones should be strongly discouraged.) Such heterogeneity consists largely, if not completely, of physically or epigenetically [9] altered genomic subpopulations, but the genetic homogeneity of multiple HPV clones from the same sample (or multiple copies integrated at different chromosomal sites) has never been rigorously tested beyond comparison of restriction digest patterns [1, 9] or partial sequences [44, 46]. No evidence for creation of chimeric reference sequences assembled from overlapping PCR products [20] or subgenomic reference clones [38] exists, and *in vitro* examinations of polymerase fidelity in normal cells versus transformed cells (or normal cells versus tumor cells) yield conflicting results [5, 37]. Still, isolation of non-unit-length reference HPV clones from a poorly-differentiated carcinoma would require a deeply held conviction that error-free replication and mismatch repair of HPV genomes was taking place within a rapidly dividing, chromosomally unstable milieu.

HPV-1a

Although the reference HPV-1a clone [10] has not been resequenced in its entirety, this clone has been resequenced in regions where the reference HPV-1a sequence and the sequence of pHPV-1 (3-3), a closely-related North American HPV-1a isolate [23], disagreed [43]. This directed resequencing revealed seven previously unreported sequencing errors in the reference HPV-1a sequence. Unless identical sequencing errors were made at the same nucleotide position(s) in the independently determined pHPV-1 (3-3) and reference HPV-1a clone sequence, it is unlikely that additional errors in the reference HPV-1a sequence remain. The reference HPV-1a sequence in GenBank (accession number V01116) has been routinely updated as sequencing errors were reported. These final seven corrections, however, have yet to be incorporated, the most significant of which reunites the previously split putative E5 ORF (Figure 1). This change also requires renumbering of the reference sequence beyond nt 3885.

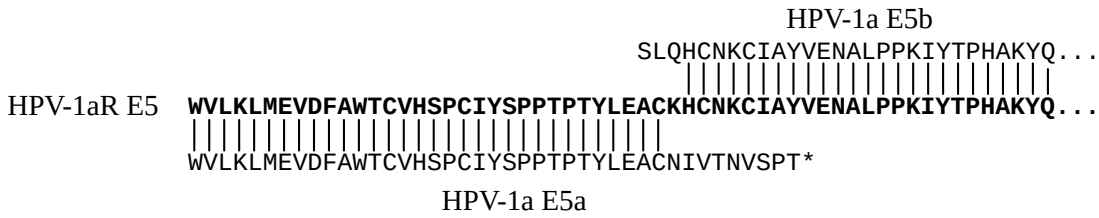

Figure 1. Amino acid alignment of revised HPV-1a E5 ORF (bold center) with previously published E5a (lower) and E5b (upper) ORFs.

It is important to note that pHPV-1 (3-3), the HPV-1a clone deposited in the American Type Culture Collection (ATCC) [23], is not the reference clone. While the reference clone and pHPV-1 (3-3) share greater than 99% nucleotide identity, there are a number of differences in the HPV-1a long control regions (LCRs) of these clones, including a 15-bp deletion in the 5′ end of the pHPV-1 (3-3) LCR relative to the reference clone which is not an artifact of cloning, as it is seen in uncloned HPV-1a genomes present in other North American plantar wart shavings [44]. These differences may prove to have negligible effect on HPV-1a LCR functions. Nevertheless, to avoid confusion, studies involving the reference HPV-1a LCR should specifically mention the source of the "reference" clone. The starting

clone for recent work on E1 binding at the HPV-1a minimal origin of replication was pHPV-1 (3-3) [S. Khan's laboratory, personal communication], which explains the C instead of G at nt 7801 in the pori60–1 subclone [22].

HPV-5a

The published reference HPV-5a [61] and HPV-18 [8] sequences are examples of sequence determination thwarted by bad printing. Font selection and quality of reproduction conspire against confident distinction between C's and G's at certain nucleotide positions. Most of the residues at these positions appear as "G remnants" in reprints—the distinguishing features of these G's are lost on photocopying. This difficulty is compounded by GenBank database entries (M17463 and X05015, respectively), which interpret all of these ambiguous positions as C's, and republished sequences [27] derived from these database entries. In the case of HPV-5a, the situation is further confused by GenBank entries (accession numbers M22961 et al.) for cloned HPV-5a deletion mutants which utilize the reference HPV-5a sequence backbone but enter G's at these ambiguous positions (Table 1). A compact illustration of the problem is provided by proposed corrections to a portion of the reference HPV-5a L1 ORF determined by resequencing of the reference clone [33]. As indicated (Table 2), one to three corrections are necessary depending on which reference sequence is used for comparison.

Table 1 Reference HPV-5a sequence ambiguous residues

Nt position	Published[1]	GenBankA[2]	GenBankB[3]	HPV-5b[4]
1927	g	c	g	g
2573	c	c	g	g
2576	g	c	g	g
2840	g	c	g	g
3065	g	c	g	g
4364	g	c	g	g
5529	?g	c	g	g
6175	?g	c	g	g
6265	g	c	g	g
6733	g	c	g	g
7120	g	c	g	g
7680	g	c	g	t
7683	g	c	g	g

[1] The distinguishing features of these already light g's are lost on photocopying. Zachow, K.R. et al. [61].

[2] Accession number M17463. This is also what appears in *Human Papillomaviruses 1994* [27].

[3] Accession numbers M22961, M18452, M18453, and M18454. Ostrow, R.S. et al. [46].

[4] GenBank accession number D90252.

Table 2 Reported reference HPV-5a L1 ORF nucleotide residues

Nt position	Published[1]	GenBankA[2]	GenBankB[3]	Corrected[4]	HPV-5b[5]
6175	?g	c	g	g	g
6265	g	c	g	g	g
6502	c	c	c	g	g

[1] Zachow, K.R. et al. [61].

[2] Accession number M17463. This is also what appears in *Human Papillomaviruses 1994* [27].

[3] Accession numbers M22961, M18452, M18453, and M18454. Ostrow, R.S. et al. [46]

[4] Kawase, M. et al. [33]

[5] GenBank accession number D90252

Where the reference HPV-5a sequence is legible, other sequencing errors have been reported [12]. Because this clone has not been completely resequenced, additional errors may be discovered. The changes proposed by Deau et al. [12] at nt 4055–4058 bring the sequence more in line with the reference HPV-5b sequence. The proposed T to C correction at nt 140 is unarguably present in the accompanying autoradiograph [12]. However, the method of resequencing the reference clone is not clear. If, as with the clinical samples described in this report, cloned PCR amplicons were used, and these proposed changes were not confirmed on a second clone, it raises the possibility of PCR artifact. Prior to correcting the reference sequence at these locations, these changes should be confirmed by direct sequencing of the reference clone, if this was not performed.

HPV-6b

No errors have been reported in the reference HPV-6b sequence [54], although attention to and resequencing of the reference clone outside of the LCR has been limited [30]. Three groups and at least six papers have been devoted completely or in part to characterizing the 5′ end of HPV-6a and HPV-6b LCRs, in order to address the authenticity of a 94-bp deletion in this region of the reference HPV-6b clone [24, 52 and references therein]. An apparent consensus on the inauthenticity of the reference HPV-6b clone selected for sequencing has been reached, yet the reference HPV-6b sequence remains unmodified in the database [GenBank accession number X00203]. It may be helpful to take a second look at some of the data (Figure 2) and offer an alternative interpretation, especially in light of Heinzel et al.'s [24] sequencing of a "reference" HPV-6b clone that does not contain the 94-bp deletion:

1. The reference HPV-6b clone [14] actually consists of two clones, p21 and pAmp2—p21 is the "short" HPV-6b subgenomic *Eco*RI-*Bam*HI double-digestion fragment (from nt 2188–4722) cloned into pBR322, while pAmp2 is the "long" subgenomic fragment (from nt 4722–2188, harboring the LCR) cloned into pBR322. pAmp2 was reportedly difficult to clone [14], and the 5′ LCR of pAmp2 is ∼120 bp shorter than the 5′ LCR of reference HPV-6a or HPV-6d clones as well as the corresponding region of HPV genomic DNA isolated from the HPV-6a, HPV-6b, and HPV-6d positive clinical material from which these clones were obtained [4].
2. Initial templates for reference HPV-6b sequence determination [54] included p21 and pAmp2 *as well as* [emphasis added] full-length *Bam*HI or *Eco*RI (single-digest) genomes cloned into λL47 and subcloned into pBR322, provided by Lutz Gissmann.

3. The initial full-length *Eco*RI reference HPV-6b/pBR322 clone provided to Kasher and Roman (sequence 1 in Figure 2, this clone was reportedly a reconstruction of the p21 and pAmp2 clones [32]) did not contain the reference HPV-6b sequence (HPV-6b in Figure 2) or the "corrected" sequence (HPV-6bR in Figure 2) reported by Heinzel et al. [24] in this region of the LCR. No information was given by Kasher and Roman on the full-length *Bam*HI reference HPV-6b/pBR322 clone provided as purified DNA, beyond an indication that its *Ava*II restriction pattern differed from all other clones analyzed [32]. They did not report the nucleotide sequence of this clone.
4. The Materials and Methods section of the Heinzel et al. paper [24] includes the following: "The original isolates of HPV-6b and HPV-11 were obtained by one of us (H.-U. B.) in the form of pBR322 clones in 1983 from Lutz Gissmann at the German Cancer Research Center . . . "

In order to rectify the Heinzel et al. [24] sequencing results with statements 2 and 3 above, two assumptions are required:

A. The plural "clones" of Heinzel et al. [24] in statement 4 refer to HPV-6b and HPV-11 full length clones, rather than the HPV-6b p21 and pAmp2 clones.
B. The reference HPV-6b clone obtained by H.-U. Bernard as described in Heinzel et al. [24] was not the full-length *Eco*RI clone provided to Kasher and Roman [32], but rather was identical to the full-length *Bam*HI clone provided to Kasher and Roman. Kasher and Roman's flow diagram (Figure 2 in [32]) may be correct, but it seems implausible that clones reconstructed from the pAmp2 clone could *de novo* regenerate the exact sequence of the 94-bp deletion. The simpler interpretation is that the clones obtained were the full-length genomes cloned into λL47 and subcloned into pBR322, with Heinzel et al. [24] sequencing the full-length *Bam*HI clone and Kasher and Roman [32] sequencing everything but this *Bam*HI clone, i.e., those clones and subclones which suffered deletions/duplications during manipulation and propagation in *E. coli* (sequences 1, 2, and 3 in Figure 2).

Besides containing all reported reference HPV-6b sequences and sequences resulting from passage or subcloning of reference HPV-6b clones [24,32,54] in this region of the LCR, the alignment in Figure 2 also includes the sequence for HPV6-T70 [31] and reference or nonreference HPV-6a sequences where subtype was confirmed by *Pst*I restriction-digest pattern [18, 24, 50, 52]. Bracketing this region by the L1 ORF stop codon and the late polyadenylation signal and regrouping the delineated sequence into triplets (this recasting is intended merely to break up the monotony of the purine-thymidine repeats—it does not imply coding potential) allows for a clear division into three consensus subregions: the upstream conserved 24-bp repeat elements (noted by Roman numerals **I** and **IIa** in Figure 2), the central alternating purine-thymidine tract (with a 66-bp sequence implicated in down-regulation of transcription [48] noted as "negative regulatory element"), and a downstream polyglot tail, less repetitive but with some suggestion of iterated 18-bp blocks. It can also be appreciated from this alignment that, ignoring 1 and 2-bp deletions, alterations of this region consist entirely of duplication of 24-bp repeat element **I** and deletion/excision of fragments flanked by short direct repeats, with elimination of flanking bases equivalent to either the "left" or "right" repeat. Identical deletions have been described for HPV-5a mutants [46].

The current reference HPV-6b sequence is something of a pariah, and it would be foolish to defend its maintenance in the database. Whether the HPV-6b clone characterized by Heinzel et al. [24] should be distributed as the reference HPV-6b clone, or whether HPV-6b should be replaced by the more prevalent HPV-6a [25] as the reference HPV-6, are matters for the papillomavirus community to decide. Before writing off the reference HPV-6b sequence to "cloning artifact", however, it is important to recognize the distinction between an inauthentic clone and a nonrepresentative one. Kasher and Roman [32] convincingly demonstrated that insertions and deletions in the 5′ end of the LCR of HPV-6b clones can arise as a result of *in vitro* propagation, and Boshart and zur Hausen [4] did not find genomes with a deletion in the 5′ end of the LCR upon reexamination of the clinical material which yielded the reference HPV-6b clone. To deny the possibility that a subpopulation of HPV-6b genomes with a 94-bp deletion represented by pAmp2 ever existed *in vivo*, though, is somewhat extreme, since legitimate clones displaying similar (albeit smaller) deletions in this region exist—HPV6-T70 [31], for example, or HPV-6vc [50] (sequences 2 and 4, respectively, in Figure 2). As well, the sensitivity of the reexamination of the clinical material was undoubtedly less than that possible with current techniques.

It is also important to note that the complete HPV-6a sequence reported recently [25] is not the sequence of the reference HPV-6a clone [24], and that, according to current definitions [60], HPV-6a and HPV-6b would not qualify for designation as separate subtypes.

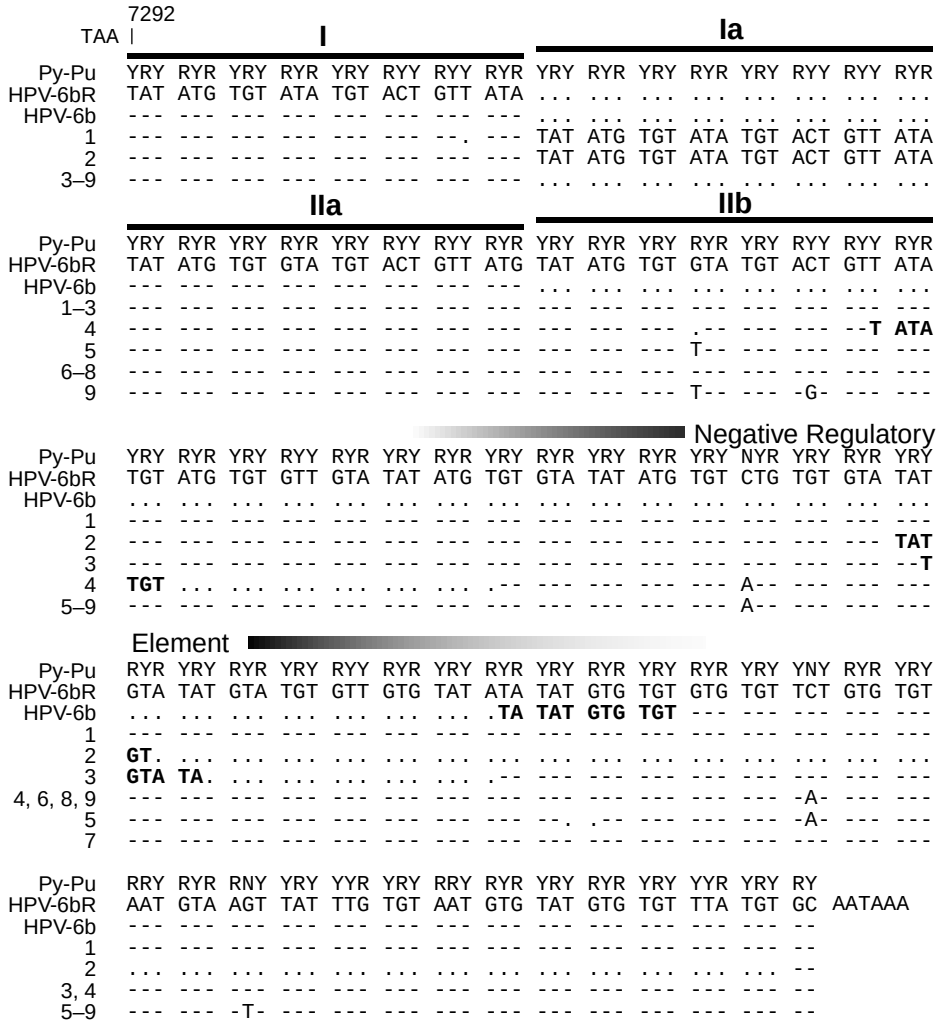

Figure 2 Legend

Sequence	"Subtype"	Description	Reference	GenBank Acc. No.
Py-Pu	—	pyrimidine-purine consensus	—	—
HPV-6bR	HPV-6b	revised reference HPV-6b	[24]	S80200
HPV-6b	HPV-6b	published reference HPV-6b	[54]	X00203
1	HPV-6b	reference clone sequence #1	[32]	?M22106*
2	HPV-6b	reference clone sequence #2	[32]	—
		HPV6-T70	[31]	L22694
3	HPV-6b	reference clone sequence #3	[32]	—
4	HPV-6a	HPV-6vc	[50]	—
5	HPV-6a	reference HPV-6a sequence†	[24]	S80200
6	HPV-6a	full length HPV-6a sequence	[25]	L41216
		isolate 1094	[52]	L36842
7	HPV-6a	HPV6-W50	[18]	L22693
		isolate 1086	[52]	L36841
8	HPV-6a	isolate 1083	[52]	L36839
9	HPV-6a	isolate 1084	[52]	L36840

*The GenBank entry indicates this is the sequence for the pHPV6b-ML2 clone, yet pHPV6b-ML2 is described as having a "#2" sequence, not a "#1" sequence [32].

†The sequence in this region is almost identical to the published sequence of isolate 1082. L36837, the GenBank entry for isolate 1082, is incorrect at nt 68 according to its published sequence.

Figure 2. (Previous page) Alignment of all published reference and nonreference partial or complete HPV-6a and HPV-6b nucleotide sequences which include sequence information from the 5′ end of the LCR between the L1 ORF stop codon and the late polyadenylation signal. Nucleotide position based on published reference HPV-6b sequence [54]. Alignment uses reference HPV-6b sequence published by Heinzel et al. [24] as the "true" reference sequence. Agreement of any other sequence with this sequence at a particular nucleotide position is indicated by a dash, with a gap indicated by a dot, and disagreement indicated by the specific residue at that nucleotide position. Nucleotides in bold flanking a deletion are in agreement with the Heinzel et al.-determined reference HPV-6b sequence, but placement 5′ or 3′ to the deletion is arbitrary. See text and accompanying legend for further details.

HPV-16

Proposed revisions to the sequence of the reference HPV-16 clone [16, 56] reported piecemeal were summarized in *Human Papillomaviruses 1995* and incorporated in a sequence called HPV-16R [28]. The only additional correction (Table 3) found on complete resequencing of the reference HPV-16 clone is a G instead of A at nt 2926 [44]. This G was confirmed by *Bst*NI digestion of the reference clone (the correction restores a *Bst*NI site) [44], and has been reported as well in a nonreference HPV-16 clone [35, R. Kovelman, personal communication]. It is also found at the corresponding position of the integrated HPV-16 variants present in the CaSki and SiHa cervical carcinoma cell lines [44, GenBank accession number AF001599].

It is important to note that the HPV-16R sequence is a composite sequence, with two consensus nucleotides at positions 1139 and 6242 (present in all or almost all HPV-16 variants) grafted onto the sequence of the idiosyncratic reference HPV-16 clone [16, 28 and references therein, 56]. As yet, no clone or completely characterized HPV-16 isolate has been identified with a sequence identical to HPV-16R.

Table 3 Reference HPV-16 and HPV-18 sequence corrections

Nt position	Sequence	Corrected
HPV-16[1]		
2926	a	g
HPV-18[2]		
287	c	g
2856–2859	tgcg	gcgt
3084–3085	cg	gc
3275	c	g
5701	c	g
6460	c	g
6625	c	g
6842	c	g

[1] Nucleotide positon and comparison based on HPV-16R sequence [28].

[2] Nucleotide positon and comparison based on HPV-18 sequence as entered in GenBank, accession number X05015.

HPV-18

Two corrections to the sequence of the reference HPV-18 clone [3, 8] have been previously reported [2, 42]. Complete resequencing of the reference HPV-18 clone revealed no additional sequencing errors [44] when compared to the original published sequence. As already noted in the discussion of HPV-5a, C's and G's at some nucleotide positions are difficult to distinguish in the published reference HPV-18 sequence. This ambiguity is the likely cause of numerous data entry errors (Table 3) for the reference HPV-18 sequence [GenBank accession number X05015] which also appear in republished sequences obtained from the database [27]. The constant rediscovery of some of these "sequencing errors" [26, 58] could have been prevented by a closer examination of the published sequence, specifically the amino acid translations above the ambiguous nucleotides. Only at nt 287, 3275, and 6842, where these G's occupy wobble positions of their respective codons, can the intended nucleotides not be discerned.

Of greater concern (Table 4) are the partial reference HPV-18 sequences, published and/or in the database [21, 40, 53, 57, 59], which predate the complete reference HPV-18 sequence publication. With the exception of an unpublished 117-bp sequence from the L1 ORF [55], none of these partial sequences are identical either to each other in regions of overlap or to the complete (corrected) reference sequence. Most of these differences are likely attributable to sequencing errors. However, the partial sequence reported by Thierry et al. [59] shares too much in common with the sequences of HPV-18 variants found in HeLa or C4–1 [45] cells to dismiss these differences as coming from a poorly sequenced reference clone. Since there is no evidence to indicate that the *Bam*HI subclone used to determine this sequence came from something other than the reference HPV-18 clone (F. Thierry and E. Schwarz, personal communications), either the sample from which the reference HPV-18 clone was isolated [3] was dually infected with reference HPV-18 genomes and C4–1-like genomes, or contamination of the reference clone took place at some point between isolation and determination of this partial sequence.

Table 4 Previous partial or complete sequence determinations of reference HPV-18 clone

		Nucleotides			
Reference	GenBank Acc. No.	Published	In Database	Correspond to[1]	In conflict[2]
[8]	X05015	7857	7857	1–7857	9
[57]	X04773/1750	1750	7857	7844–1737	3
[40]	X04354	1188	1188	7648–979	20
[21, 59]	M14710[3]	1052	1053[4]	6929–124	10
[53]	A06328	1000[5]	1000	6–93	1
[55]	A06329	–[6]	117	5730–5846	0

[1] Nucleotide positions relative to reference HPV-18 sequence, GenBank accession number X05015.

[2] Number of nucleotides conflicting with corrected reference HPV-18 sequence; see Table 3.

[3] M14710 sequence and A06328 sequence (brackets) determined from the same *Bam*HI subclone of the reference HPV-18 clone (F. Thierry and E. Schwarz, personal communications).

[4] Additional nucleotide in database represents data entry error.

[5] The A06328 sequence is a composite sequence of reference HPV-18 clone sequence (initial 88 nt) and sequence determined from HeLa HPV-18 variant and C4-1 HPV-18 variant cDNA clones.

[6] Sequence unpublished.

HPV-35

The reference HPV-35 clone [38] is actually two subgenomic *Bam*HI fragments cloned into pBR322, deposited in the ATCC under the designations HPV35 clone 2A and HPV35 clone 2B. Two laboratories independently determined the reference HPV-35 sequence [13, 39], and the differences between the two reported sequences are profound. Both groups specifically state that the reference HPV 35 clones deposited in the ATCC were used as starting templates for sequencing, ruling out the possibility that one of the groups used a "different" reference HPV-35 clone or clones.

In an attempt to resolve these discrepancies, these same clones were obtained from the ATCC, and resequencing of a 300-bp region of the E1 ORF [44], selected because nucleotide disagreements and gaps in the alignment were particularly concentrated along this span (Figure 3), revealed 100% identity with the Delius and Hofmann-determined sequence [13], which is deposited in GenBank under accession number X74477 and identified by the authors [13, 27] as HPV-35H. Apparently, autoradiographs were not preserved for the sequence reported by Marich et al. [39, T. Dubensky, personal communication], so it is suggested that HPV-35H be treated as the definitive reference HPV-35 sequence until complete re-resequencing of the reference clones is performed or evidence in support of the Marich et al. [39] HPV-35 sequence is found. Interestingly, both of these sequences have a *Pvu*II restriction site at nt 125 which is not present in the original experimentally-determined HPV-35 restriction map [38]. The use of the HPV-35H sequence instead of the Marich et al.-determined HPV-35 sequence bolsters the hypothesis advanced by Qu et al. [49] that total number of nucleotide mismatches influences MY09/MY11 primer set amplification efficiency of HPV-35 positive samples (see Qu et al.'s Table 4 [49]. The MY09 primer's penultimate base in this table should be T, not G).

```
        1530
        |
HPV-35H ttggtgtgtggccgcatttggaatagccccaagtgtggcggaaagtttaaaaacattaattaaaccatattgtttata
HPV-35  ----------------------------------------.............a-c--t--a-----aaca---g

HPV-35H tatacatatacaatgtttatcgtgttcatgggg.tatggtaattctagcattattacgatttaaatgtgcaaaaaaca
HPV-35  --------------------------------c------------------------------.--cg-----cg-

                                                                                1760
                                                                                 |
HPV-35H gaacaacaattgaaaaactatt.at.caaaattgctatgtatttcagctgcaagtatgctaatacaaccaccaaaatt
HPV-35  ----------------.-----g--g-t-------------------------------------------------
```

Figure 3. Alignment of two nucleotide sequences of reference HPV-35 clone determined independently [13, 39] in a region of the E1 ORF. Numbering based on reference HPV-35 sequence deposited in GenBank under the accession number X74477 and identified as HPV-35H. A dash indicates an identical nucleotide at that position, a dot indicates a gap.

HPV-59

As mentioned in *Human Papillomaviruses 1995* [28], a ∼150-bp *Pst*I fragment in the 5′ half of the E2 ORF from nt 2974–3117 is inverted in the reference HPV-59 sequence, i.e., the sequence as published [51] and in GenBank (accession number X77858) shares less than 40% nucleotide identity with the closely related reference HPV-18, 39, 40, 45, and 70 sequences in this region, while the reverse-complement of this sequence shares greater than 70% nucleotide identity with these sequences (150-bp segments upstream and downstream from nt 2974–3117 share greater than 70% nucleotide identity and 60% nucleotide identity, respectively, with these closely related types [44]). Whether this *Pst*I fragment is actually inverted in the reference clone is unknown, but seems unlikely, unless this inversion represents what is *bona fide* present in the HPV-59 genome. HPV-59 was cloned at its single *Bam*HI site [41], ruling out the possibility of cloning artifact (from religation of subgenomic *Pst*I-digest fragments, for example) during isolation. Rather, this sequence was most likely assembled incorrectly

from sequenced HPV-59 subgenomic *Pst*I fragments. According to Rho et al., HPV-59 *Pst*I, *Hinc*II, *Acc*I, *Ban*I, *Ssp*I, and *Taq*I digests cloned into pUC19 were used as initial templates for sequencing [51]. At this point, it became irrelevant how the small *Pst*I fragment was oriented in the reference clone, because none of these other enzymes used for subcloning have a restriction site between nt 2974 and 3117. Unless a subgenomic fragment generated by a different enzyme which encompassed this *Pst*I fragment was also sequenced, it would be impossible to provide orientation for the *Pst*I fragment except by comparison with closely related sequences. Sequencing of an *Ssp*I fragment from nt 2791–3692 could have reached this region, as could sequencing of a *Ban*I fragment from nt 2921–4420 (how any single-digest HPV-59 *Ban*I fragment could have been cloned into pUC19 without first blunting the ends is not mentioned). If these fragments or nested fragments generated by BAL31 exonuclease [51] in this region were not sequenced, then this inversion is almost certainly artifactual.

Acknowledgments

Portions of the work presented in this article were performed at the University of North Carolina-Chapel Hill and the Centers for Disease Control and Prevention in Atlanta, GA.

References

[1] Baker, C.C., Phelps, W.C., Lindgren, V., Braun, M.J., Gonda, M.A. and Howley, P.M. (1987) Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J. Virol.* **61**, 962–971.

[2] Baker, C.C. (1993) The genomes of the papillomaviruses. In: S.J. O'Brien (Ed.), *Genetic Maps: Locus Maps of Complex Genomes*, Sixth Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1.134–1.146.

[3] Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W. and zur Hausen, H. (1984) A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. *EMBO J.* **3**, 1151–1157.

[4] Boshart, M. and zur Hausen, H. (1986) Human papillomaviruses in Buschke-Löwenstein tumors: Physical state of the DNA and identification of a tandem duplication in the noncoding region of a human papillomavirus 6 subtype. *J. Virol.* **58**, 963–966.

[5] Boyer, J.C., Thomas, D.C., Maher, V.M., McCormick, J.J. and Kunkel, T.A. (1993) Fidelity of DNA replication by extracts of normal and malignantly transformed human cells. *Cancer Res.* **53**, 3270–3275.

[6] Campione-Piccardo, J., Montpetit, M.L., Gregoire, L. and Arella, M. (1991) A highly conserved nucleotide string shared by all genomes of human papillomaviruses. *Virus Genes* **5**, 349–357.

[7] Chen, E.Y., Howley, P.M., Levinson, A.D. and Seeburg, P.H. (1982) The primary structure and genetic organization of the bovine papillomavirus type 1 genome. *Nature* **299**, 529–534.

[8] Cole, S.T. and Danos, O. (1987) Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. *J. Mol. Biol.* **193**, 599–608.

[9] Danos, O., Katinka, M. and Yaniv, M. (1980) Molecular cloning, refined physical map and heterogeneity of methylation sites of papilloma virus type 1a DNA. *Eur. J. Biochem.* **109**, 457–461.

[10] Danos, O., Katinka, M. and Yaniv, M. (1982) Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among Papovaviridae. *EMBO J.* **1**, 231–236.

[11] Dartmann, K., Schwarz, E., Gissmann, L. and zur Hausen, H. (1986) The nucleotide sequence and genome organization of human papilloma virus type 11. *Virology* **151**, 124–130.

[12] Deau, M.-C., Favre, M., Jablonska, S., Rueda, L.-A. and Orth, G. (1993) Genetic heterogeneity of oncogenic human papillomavirus type 5 (HPV5) and phylogeny of HPV5 variants associated with epidermodysplasia verruciformis. *J. Clin. Micro.* **31**, 2918–2926.

[13] Delius, H. and Hofmann, B. (1994) Primer-directed sequencing of human papillomavirus types. *Curr. Top. Microbiol. Immunol.* **186**, 13–31.

[14] de Villiers, E.-M., Gissmann, L. and zur Hausen, H. (1981) Molecular cloning of viral DNA from human genital warts. *J. Virol.* **40**, 932–935.

[15] de Villiers, E.-M. (1994) Human pathogenic papillomavirus types: an update. *Curr. Top. Microbiol. Immunol*. **186**, 1–12.

[16] Dürst, M., Gissmann, L., Ikenberg, H. and zur Hausen, H. (1983) A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proc. Natl. Acad. Sci. USA* **80**, 3812–3815.

[17] El Awady, M.K., Kaplan, J.B., O'Brien, S.J. and Burk, R.D. (1987) Molecular analysis of integrated human papillomavirus 16 sequences in the cervical cancer cell line SiHa. *Virology* **159**, 389–398.

[18] Farr, A., Wang, H., Kasher, M.S. and Roman, A. (1991) Relative enhancer activity and transforming potential of authentic human papillomavirus type 6 genomes from benign and malignant lesions. *J. Gen. Virol*. **72**, 519–526.

[19] Favre, M., Kremsdorf, D., Jablonska, S., Obalek, S., Pehau-Arnaudet, G., Croissant, O. and Orth, G. (1990) Two new human papillomavirus types (HPV54 and 55) characterized from genital tumours illustrate the plurality of genital HPVs. *Int. J. Cancer* **45**, 40–46.

[20] Forslund, O. and Hansson, B.G. (1996) Human papillomavirus type 70 genome cloned from overlapping PCR products: complete nucleotide sequence and genomic organization. *J. Clin. Microbiol*. **34**, 802–809.

[21] Garcia-Carranca, A., Thierry, F. and Yaniv, M. (1988) Interplay of viral and cellular proteins along the long control region of human papillomavirus type 18. *J. Virol.* **62**, 4321–4330.

[22] Gopalakrishnan, V., Walker, S. and Khan, S.A. (1995) Stimulation of human papillomavirus type 1a DNA replication by a multimerized AT-rich palindromic sequence. *Virology* **214**, 301–306.

[23] Heilman, C.A., Law, M.-F., Israel, M.A. and Howley, P.M. (1980) Cloning of human papilloma virus genomic DNAs and analysis of homologous polynucleotide sequences. *J. Virol.* **36**, 395–407.

[24] Heinzel, P.A., Chan, S.-Y., Ho, L., O'Connor, M., Balaram, P., Campo, M.S., Fujinaga, K., Kiviat, N., Kuypers, J., Pfister, H., Steinberg, B.M., Tay, S.-K., Villa, L.L. and Bernard, H.-U. (1995) Variation of human papillomavirus type 6 (HPV-6) and HPV-11 genomes sampled throughout the world. *J. Clin. Microbiol*. **33**, 1746–1754.

[25] Hofmann, K.J., Cook, J.C., Joyce, J.G., Brown, D.R., Schultz, L.D., George, H.A., Rosolowsky, M., Fife, K.H. and Jansen, K.U. (1995) Sequence determination of human papillomavirus type 6a and assembly of virus-like particles in *Saccharomyces cerevisiae*. *Virology* **209**, 506–518.

[26] Hofmann, K.J., Neeper, M.P., Markus, H.Z., Brown, D.R., Müller, M. and Jansen, K.U. (1996) Sequence conservation within the major capsid protein of human papillomavirus (HPV) type 18 and formation of HPV-18 virus-like particles in *Saccharomyces cerevisiae*. *J. Gen. Virol*. **77**, 465–468.

[27] *Human Papillomaviruses 1994: A compilation and analysis of nucleic acid and amino acid sequences*. Myers, G., Delius, H., Icenogle, J., Bernard, H.-U., Favre, M., Van Ranst, M. and Wheeler, C. (Eds.) Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM.

[28] *Human Papillomaviruses 1995: A compilation and analysis of nucleic acid and amino acid sequences*. Myers, G., Delius, H., Icenogle, J., Bernard, H.-U., Baker, C., Halpern, A. and Wheeler, C. (Eds.) Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM.

[29] *Human Papillomaviruses 1996: A compilation and analysis of nucleic acid and amino acid sequences*. Myers, G., Baker, C., Wheeler, C., Halpern, A., McBride, A. and Doorbar, J. (Eds.) Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM.

[30] Icenogle, J.P., Sathya, P., Miller, D.L., Tucker, R.A. and Rawls, W.E. (1991) Nucleotide and amino acid sequence variation in the L1 and E7 open reading frames of human papillomavirus type 6 and type 16. *Virology* **184**, 101–107.

[31] Kasher, M.S. and Roman, A. (1988) Characterization of human papillomavirus type 6b DNA isolated from an invasive squamous carcinoma of the vulva. *Virology* **165**, 225–233.

[32] Kasher, M.S. and Roman, A. (1988) Alterations in the regulatory region of the human papillomavirus type 6 genome are generated during propagation in *Escherichia coli*. *J. Virol.* **62**, 3295–3300.

[33] Kawase, M., Orth, G., Jablonska, S., Blanchet-Bardon, C., Rueda, L.-A. and Favre, M. (1996) Variability and phylogeny of the L1 capsid protein gene of human papillomavirus type 5: contribution of clusters of nonsynonymous mutations and of a 30-nucleotide duplication. *Virology* **221**, 189–198.

[34] Kirnbauer, R., Taub, J., Greenstone, H., Roden, R., Dürst, M., Gissmann, L., Lowy, D.R. and Schiller, J.T. (1993) Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. *J. Virol.* **67**, 6929–6936.

[35] Kovelman, R., Bilter, G.K., Glezer, E., Tsou, A.Y. and Barbosa, M.S. (1996) Enhanced transcriptional activation by E2 proteins from the oncogenic human papillomaviruses. *J. Virol.* **70**, 7549–7560.

[36] Kremsdorf, D., Favre, M., Jablonska, S., Obalek, S., Rueda, L.A., Lutzner, M.A., Blanchet-Bardon, C., Van Voorst Vader, P.C. and Orth, G. (1984) Molecular cloning and characterization of the genomes of nine newly recognized human papillomavirus types associated with epidermodysplasia verruciformis. *J. Virol.* **52**, 1013–1018.

[37] Kubota, Y., Murakami-Murofushi, K, Shimada, Y., Ogiu, T. and Oikawa, T. (1995) Reduced fidelity of DNA synthesis in cell extracts from chemically induced primary thymic lymphomas of mice. *Cancer Res*. **55**, 3777–3780.

[38] Lorincz, A.T., Quinn, A.P., Lancaster, W.D. and Temple, G.F. (1987) A new type of papillomavirus associated with cancer of the uterine cervix. *Virology* **159**, 187–190.

[39] Marich, J.E., Pontsler, A.V., Rice, S.M., McGraw, K.A. and Dubensky, T.W. (1992) The phylogenetic relationship and complete nucleotide sequence of human papillomavirus type 35. *Virology* **186**, 770–776.

[40] Matlashewski, G., Banks, L., Wu-Liao, J., Spence, P., Pim, D. and Crawford, L. (1986) The expression of human papillomavirus type 18 E6 protein in bacteria and the production of anti-E6 antibodies. *J. Gen. Virol.* **67**, 1909–1916.

[41] Matsukura, T. and Sugase, M. (1995) Identification of genital human papillomaviruses in cervical biopsy specimens: segregation of specific virus types in specific clinicopathologic lesions. *Int. J. Cancer* **61**, 13–22.

[42] Meissner, J. (1993) *Taq*I is a single cut enzyme for HPV-18. Nucleic Acids Res. 21, 1041.

[43] Meissner, J. (1994) Complete nucleotide sequencing of an HPV-1a variant and determination of extant errors in the prototype HPV-1a sequence. *Virus Genes* **9**, 189–191.

[44] Meissner, J. Unpublished observation.

[45] Ong, C-K., Chan, S.-Y., Campo, M.S., Fujinaga, K., Mavromara-Nazos, P., Labropoulou, V., Pfister, H., Tay, S.-K., ter Meulen, J., Villa, L.L. and Bernard, H.-U. (1993) Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. *J. Virol.* **67**, 6424–6431.

[46] Ostrow, R.S., Zachow, K.R. and Faras, A.J. (1987) Molecular cloning and nucleotide sequence analysis of several naturally occuring HPV-5 deletion mutant genomes. *Virology* **158**, 235–238.

[47] Parton, A. (1990) Nucleotide sequence of the HPV16 L1 open reading frame. *Nucleic Acids Res.* **18**, 3631.

[48] Pattison, S., Skalnik, D.G. and Roman, A. (1997) CCAAT displacement protein, a regulator of differentiation-specific gene expression, binds a negative regulatory element within the 5′ end of the human papillomavirus type 6 long control region. *J. Virol.* **71**, 2013–2022.

[49] Qu, W., Jiang, G., Cruz, Y., Chang, C.J., Ho, G.Y.F., Klein, R.S. and Burk, R.D. (1997) PCR detection of human papillomavirus: comparison between MY09/MY11 and GP5+/GP6+ primer systems. *J. Clin. Micro.* **35**, 1304–1310.

[50] Rando, R.F., Groff, D.E., Chirikjian, J.G. and Lancaster, W.D. (1986) Isolation and characterization of a novel human papillomavirus type 6 DNA from an invasive vulvar carcinoma. *J. Virol.* **57**, 353–356.

[51] Rho, J., Roy-Burman, A., Kim, H., de Villiers, E.-M., Matsukura, T. and Choe, J. (1994) Nucleotide sequence and phylogenetic classification of human papillomavirus type 59. *Virology* **203**, 158–161.

[52] Roman, A. and Brown, D. (1995) Sequence variation in the extreme 5′ end of the human papillomavirus type 6a long control region. *J. Infect. Dis.* **171**, 697–700.

[53] Schneider-Gädicke, A. and Schwarz, E. (1986) Different human cervical carcinoma cell lines show similar transcription patterns of human papillomavirus type 18 early genes. *EMBO J.* **5**, 2285–2292.

[54] Schwarz, E., Dürst, M., Demankowski, C., Lattermann, O., Zech, R., Wolfsperger, E., Suhai, S. and zur Hausen, H. (1983) DNA sequence and genome organization of genital human papillomavirus type 6b. *EMBO J.* **2**, 2341–2348.

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

[56] Seedorf, K., Krämmer, G., Dürst, M., Suhai, S. and Röwekamp, W.G. (1985) Human papillomavirus type 16 DNA sequence. *Virology* **145**, 181–185.

[57] Seedorf, K., Oltersdorf, T., Krämmer, G. and Röwekamp, W. (1987) Identification of early proteins of the human papilloma viruses type 16 (HPV 16) and type 18 (HPV 18) in cervical carcinoma cells. *EMBO J.* **6**, 139–144.

[58] Stewart, A.-C. M., Eriksson, A.M., Manos, M.M., Muñoz, N., Bosch, F.X., Peto, J. and Wheeler, C.M. (1996) Intratype variation in 12 human papillomavirus types: a worldwide perspective. *J. Virol.* **70**, 3127–3136.

[59] Thierry, F., Heard, J.M., Dartmann, K. and Yaniv, M. (1987) Characterization of a transcriptional promoter of human papillomavirus 18 and modulation of its expression by simian virus 40 and adenovirus early antigens. *J. Virol.* **61**, 134–142.

[60] Van Ranst, M.A., Tachezy, R., Delius, H. and Burk, R.D. (1993) Taxonomy of the human papillomaviruses. *Papillomavirus Rep.* **4**, 61–65.

[61] Zachow, K.R., Ostrow, R.S. and Faras, A.J. (1987) Nucleotide sequence and genome organization of human papillomavirus type 5. *Virology* **158**, 251–254.