



Generation of Entire Human Papillomavirus Genomes by Long PCR: Frequency of Errors Produced During Amplification

Ann-Charlotte M. Stewart,¹ Patti E. Gravitt,² Suzanne Cheng,² and Cosette M. Wheeler^{1,3}

¹University of New Mexico Cancer Research and Treatment Center, Department of Cell Biology, Albuquerque, New Mexico 87131-5226; ²Department of Infectious Diseases (PEG) and Department of Human Genetics (SC), Roche Molecular Systems, Foster City, California

Recently, several improvements of traditional PCR techniques have facilitated the amplification of significantly longer DNA target sequences. Here we report an improved method for amplification of entire human papillomavirus (HPV) genomes. Using *rTth* DNA polymerase, XL (Perkin-Elmer, Foster City, CA), and the accompanying XL PCR buffer system, we have successfully amplified 8-kb genomes from ~10 copies of input reference strain HPV16 DNA. This long PCR (LPCR) method was subsequently used to amplify the entire HPV16 genome from clinical specimens. The fidelity with which the *rTth* DNA polymerase XL amplifies target sequences under our chosen amplification conditions was estimated by partial sequencing of cloned LPCR products generated from cloned reference strain HPV16 genomes. A region spanning the HPV16 E6, E7, and part of the E1 open reading frames (ORFs) was sequenced in 29 clones. A total of 33 nucleotide substitutions were observed in the 23.5 kb sequenced. This corresponds to an error frequency of ~one error per 700 bases. Finally, LPCR methods were used to amplify entire, novel HPV genomes from clinical specimens. LPCR primer pairs were designed for amplification of seven potentially novel HPV types. Amplicons of ~8 kb were generated

from five of the seven HPV types targeted. One of the LPCR-generated novel genomes, CP141, was subsequently cloned and a partial sequence was determined.

The ability of PCR to produce large quantities of DNA from very small amounts of genetic material has revolutionized molecular biology.^(1,2) The impact of PCR technology is probably best exemplified in genome mapping and sequencing.⁽³⁾ Although PCR can be performed with relative ease, the size of the amplification products that can reliably be generated have been generally limited to ~3–4 kb.⁽⁴⁾ Recently, important alterations in amplification conditions have been identified that allow effective amplification of longer DNA targets.^(5,6) The modified conditions for long PCR (LPCR) involve changes of the standard PCR buffer and cycling parameters together with a two-polymerase system that includes a polymerase activity and a 3' → 5' exonuclease activity.^(5,6) Using LPCR protocols, ≤42 kb of the bacteriophage λ genome, 22 kb of human genomic DNA, and 16.3 kb of the human mitochondrial genome have been successfully amplified.^(5–7)

The use of a two-polymerase system including proofreading activity facilitates the amplification of much longer targets but may additionally influence

the fidelity of the amplification.⁽⁵⁾ The reaction conditions used and the resultant fidelity during the amplification reaction are critical to determining the value of LPCR applications. As reviewed by Eckert and Kunkel, the observed error frequencies for *Thermus aquaticus* (*Taq*)-based PCR, in the absence of proofreading activity, can range from approximately one error per 363 nucleotides (3×10^{-3}) to one error per 5411 nucleotides (2×10^{-4}), depending on the reaction conditions used.⁽⁸⁾

Using LPCR technology we have recently developed a model system to allow effective amplification of the entire 7.9-kb human papillomavirus type 16 (HPV16) genome.⁽⁹⁾ HPVs are implicated as one of the causative agents in the etiology of cervical dysplasia and cervical cancer.^(10–12) The study of the natural history and the classification of HPVs has been hampered owing to the lack of efficient in vitro culture systems for these viruses. Utilization of LPCR technology to study other papillomaviruses could greatly facilitate classification and characterization of known and potentially novel HPV genomes.

Here we report improved amplification of HPV target genomes present in cloned and clinical material and demonstrate a rapid protocol for the cloning of LPCR products. Moreover, we present an assessment of the fidelity with which the *rTth* DNA polymerase XL system amplifies its target sequences and, finally,

³Corresponding author.

E-MAIL cwheeler@medusa.unm.edu; FAX (505) 277-8572.

how LPCR technology can be used for the isolation of entire, novel HPV genomes.

MATERIALS AND METHODS

Samples and Preparation of DNA

Cervicovaginal lavages^(13,14) were screened previously for specific HPV DNA types.⁽¹⁵⁾ Ten samples that were strongly positive by HPV16 oligonucleotide hybridization, and two HPV-negative samples were chosen for LPCR analysis. Crude digestion was performed as described previously.⁽¹⁵⁾ DNA was purified further using a standard phenol/chloroform/isoamyl alcohol method, with mixing by inversion rather than vortexing to minimize DNA shearing.

Clinical specimens (crude proteinase digests) screened previously for the presence of potentially novel HPV sequences were obtained from investigations performed at the University of New Mexico, Albuquerque.⁽¹⁶⁾ DNA from these specimens was prepared as described previously.^(16,17) Specimens designated ISXX were all obtained from the recently reported International Biological Study of Cervical Cancer (IBSCC).^(18,19) Gen-

Bank accession numbers are as follows: IS039, U12481; CP6108, U12478; CP141, U12476; CP8061, U12479; MM9, U12491; MM7, U12489; MM4, U12488.

Clone ppH16, the entire HPV16 genome cloned into a single *Bam*HI site in a pBR322 vector,^(20,21) was generously provided by Dr. E.-M. de Villiers (DKFZ Papillomavirus Reference Center, Heidelberg, Germany).

PCR Primers and Amplification

Abutting primers were designed juxtaposing the unique *Bam*HI site for the HPV16 model system.⁽⁹⁾ For the potential novel HPVs^(16,18,22), abutting primers were designed juxtaposing restriction endonuclease (RE) sites present in the MY09/MY11 L1 fragments^(23,24) that would potentially represent unique RE sites within the complete genomes once isolated. The selection of RE sites and primer locations was facilitated by sequence comparisons to other closely related HPV genomes. All primers contained 5'-terminal sequences to permit regeneration of the juxtaposed RE site and a "clamp" of two to five guanines at

the extreme 5' end to facilitate subsequent RE digestion and cloning. Unique RE sites were included at the 5' end of the HPV16 model system primers for directional cloning. The lengths of the primers varied between 22 and 35 bases. Individual primer pairs were selected to have balanced melting temperatures (T_m) that ranged between 65°C and 79°C as determined with Oligo 4.0 (National Biosciences, Plymouth, MN). Sequences of all oligonucleotides used for LPCR amplification are shown in Table 1.

Amplification reactions (50 or 100 μ l) contained 1–1.1 \times *rTth* DNA polymerase, XL buffer (Perkin-Elmer, Foster City, CA), 200 μ M of each dNTP, 0.15–0.4 μ M of primers, 1.0–1.2 mM Mg(OAc)₂, and 1–2 units of *rTth* DNA polymerase, XL per 100 μ l reaction (Perkin-Elmer, Foster City, CA).

The thermal cycling conditions included a manual hot start⁽²⁵⁾, which entailed withholding the Mg(OAc)₂ until the reaction mixture had reached 80°C; an initial denaturation step at 94°C for 10 sec; 35–40 cycles in which each cycle consisted of 15 sec at 94°C, 1 min at 62°C, and 4–6 min at 68°C; and a final extension at 72°C for 10 min. All amplifications were performed in a Perkin-

TABLE 1 Sequences of Oligonucleotides used for LPCR Amplification of HPV Genomes

Target sequence	Primer	Sequence ^a	Length ^b	T_d^c	T_m^c
HPV 16	PEG07	5' - GGGGTCTAGAGGATCC CCCATGTACCAATGTTGCAGTAAATCCAGGTGA	38	nd	69
	PEG12	5' - GGGGCGGGCCGGGATCCT TGCCCCAGTGTCCCTATAGGTGGTTTG	36	nd	70
CP141	141BN	5' - GGGGGATCCT TTTTTTTCAGGTGTAGGAGC	26	69	76
	141BP	5' - GGGGGATCC CTATGACGATTTAAATTTTGG	28	66	76
CP141	141HN	5' - GGGGAA GCCTTGCAGATGGTGAGGGG	22	72	81
	141HP	5' - GGGGAA GCCTTGGTGACACGTATAGGTA	24	69	75
CP6108	6108BN	5' - GGGGGATCCT TTTTTAGGTTCTGGGGCAGCAG	28	72	82
	6108BP	5' - GGGGGATCC ATATGATAAGTTATCCTTTTGGGATG	32	68	77
CP8061	8061BN	5' - GGAGATCT GCATAGGGGTCCCTTT	21	65	68
	8061BP	5' - GGAGATCT TACATTTTGGGAGGTG	22	63	67
	8061BgIN	5' - GGAGATCT GCATAGGGGTCCCTTTTTT	25	66	74
	8061BgIP	5' - GGAGATCT TACATTTTGGGAGGTGGATTA	28	66	74
IS039	IS39AN	5' - GGCCTTA AGGTCCACGGTCCAAAAC	22	67	75
	IS39AP	5' - GGCCTTA AGGAACGCTTTTCTTTGGA	23	64	73
	IS39NN	5' - GGCCATGG TGTGCAGGTAAGCCA	21	68	77
	IS39NP	5' - GGCCATGG ATTCTACAATACTAGAGC	24	64	67
	IS39SN	5' - GGGAGTACT TCTGTTAGTGTCAACACA	24	66	68
	IS39SP	5' - GGGAGTACT AATTTAACCATTAGCACACTGC	26	64	68
PAP238 (MM9)	238AEN	5' - GGGGAATTC ATAGAATGTATATATGTCATTAC	27	62	66
	238AEP	5' - GGGGAATTC TACTATATTGGAAGAGTG	24	63	65
PAP291 (MM7)	291BN	5' - GGGGGATCCT TTTTTAGGGGAGGGG	22	71	79
	291BP	5' - GGGGGATCCT TATGATGGCCTTGTATTTTGG	28	68	78
W13B (MM4)	W13BNN	5' - GGCCCATGG TGTGCAGGTAAGC	19	69	74
	W13BNP	5' - GGCCCATGG ATTCTACAATTTTAGAACAGTG	28	66	74

^aOligonucleotide sequences shown in bold indicate 5' clamps added to facilitate RE digestion and cloning. Underlined sequences indicate RE sites.

^bLength of primer binding region, i.e., stretches directly corresponding to HPV sequences.

^cMelting temperatures, T_d and T_m , were determined using Oligo 4.0 (National Biosciences, Plymouth, MN). (nd) Not determined.

Elmer GeneAmp PCR System 9600, using MicroAmp tubes.

Cloning and Sequencing of LPCR Products

The LPCR products were subjected to electrophoresis in 0.8%–1% low melt or regular agarose gels containing ethidium bromide (EtBr). DNA fragments were visualized under long-wavelength (300 nm) UV light, and the bands of correct sizes (~8 kb) were excised from the gel. The DNA was isolated from the agarose using Gelase (Epicentre Technologies, Madison, WI), as recommended by the manufacturer, and ligated to the pGEM-T vector (Promega, Madison, WI). Bacterial colonies that contained recombinant plasmids were identified by α -complementation screening.^(26,27)

Selected colonies were grown in 2-ml cultures, and plasmids were isolated using a crude alkaline denaturation method essentially as described by Birnboim et al.^(28,29) Bacterial cells, grown overnight in LB medium, were pelleted by low speed centrifugation. Cell pellets were resuspended in 300 μ l of 50 mM Tris-HCl, 10 mM EDTA, containing 100 μ g/ml of RNase A. Three hundred microliters of 200 mM NaOH containing 1% SDS was added to the resuspended cells, whereupon they were gently mixed and incubated at room temperature for 5 min to lyse the bacterial cells. Finally, 3.0 M KOAc at pH 5.5 (300 μ l) was added, and the lysed bacterial cells were incubated on ice to precipitate chromosomal DNA. Cell debris and precipitated DNA were pelleted by centrifugation (14,000g) at room temperature for 15 min, and the supernatant was removed promptly. Plasmid DNA was precipitated by addition of 0.7 volumes of isopropanol to the supernatant, mixed by inversion, and pelleted by centrifugation (14,000g) at room temperature for 30 min. The DNA pellet was washed briefly in 70% EtOH, air-dried, and resuspended in 50 μ l of deionized water. RE digestion with *Bam*HI (HPV16) and *Hind*III (CP141) of selected clones was used to confirm the presence of an intact 8-kb insert.

Cloned LPCR products were sequenced by the dideoxy termination method⁽³⁰⁾ using ³⁵S-labeled dATP and Sequenase 2.0 (U.S. Biochemical, Cleveland, OH). Sequencing primers were

selected from the E6 and E7 open reading frames (ORFs) of the HPV16 genome (16-60, 5'-GTTAGTATAAAAAG-CAGACAT-3'; 16-269, 5'-GATGG-GAATCCATATGCT-3'; 16-430, 5'-TGTC-CTGAAGAAAAG-3' and 16-549, 5'-CCCAGCTGTAATCATGCAT-3') or from the pGEM-T vector (*pGEM-F*, 5'-ACGGC-CAGTGAATTGTA-3' and *pGEM-R*, 5'-ACGCGTTGGGAGCTCTC-3'). *pGEM-F* and *pGEM-R* primers were designed to facilitate sequencing efforts and are closer to the pGEM-T insertion site than standard M13 sequencing primers.

Restriction Endonuclease Analysis of CP141

Digests with different REs were carried out as described by the manufacturer (New England Biolabs, Inc., Beverly, MA). The enzymes used are described in Figure 5 (see below). The resulting DNA fragments were separated by electrophoresis in 0.8–2% agarose gels. Physical mapping of cleavage sites followed standard protocols.

RESULTS

Sensitivity in a HPV16 Model System

In a previous report we described a model system for amplification of the entire HPV16 reference genome cloned into a pBR322 vector.⁽⁹⁾ The detection limit in this system was ~1000 copies of linearized input DNA. This LPCR system was used to amplify the entire HPV16 genome directly from cervicovaginal lavages that were previously positive for HPV16 DNA based upon PCR amplification of the 450-bp MY09/MY11 L1 fragment.⁽²³⁾ The sensitivity of this model system, when applied directly to these clinical specimens, was low with only 2 of 21 cervicovaginal lavages producing the expected 7.9-kb HPV16 LPCR product.

To increase the sensitivity of the LPCR system, we have altered these previously reported conditions. By reducing the reaction volume (50 vs. 100 μ l), prolonging the extension time (6 vs. 4 min), and increasing the number of amplification cycles (40 vs. 35), we obtained higher yields of amplified products. We also found that increasing the concentration of *rTth* polymerase XL from 0.5 to 1 unit per 50- μ l reaction increased specific amplification product

yields significantly. Using this modified protocol, we detected ~10 copies of cloned HPV16 input DNA (Fig. 1A). Application of this protocol and the inclusion of a very gentle phenol extraction step for sample preparation facilitated the successful amplification of full-length circular HPV16 genomes from 10 of 10 cervicovaginal lavages (Fig. 1B). LPCR product yields from the lavage amplifications were also improved by linearizing the circular DNA genome via *Bam*HI restriction (Fig. 1C). Although we achieved RE-verified 8-kb products from 10 of 10 clinical samples, the reaction specificity was quite variable between samples and among replicates (see Fig. 1C). Nonspecific amplicon species range in size from ~100 to 6500 bp. Similar background was seen from amplification of a circular HPV dimer that was regenerated by ligation of HPV plasmid inserts. Linear targets at concentrations >1000 copies were observed to yield two products (3.2 and 5.5 kb) that are produced from internal mispriming sites predicted from the HPV16 reference sequence. No other nonspecific amplification was observed with <50,000 copies of linearized HPV-containing plasmid target or with HPV-negative lavages (data not shown). In all cases, the 7.9-kb target remained the predominant product. Preamplification specimen dilution, *Bam*HI linearization, *rTth* DNA polymerase XL concentration reduction, and amplification using fewer cycles decreased but did not eliminate the nonspecific amplification. The extent to which any combination of the above improved specificity was sample dependent.

Fidelity During LPCR

Knowing that DNA polymerases can introduce errors during PCR-based amplification of target sequences, we wanted to assess the error frequency of our LPCR system. A commercial kit for T/A cloning was used to isolate LPCR products amplified from the HPV16 model system (cloned reference strain HPV16 DNA). Figure 2 presents a schematic description of the entire procedure. Recombinant clones were selected by α -complementation screening and RE verification of inserts (~8 kb). A total of 29 clones were chosen for further analysis. The E6, E7, and part of the E1 ORFs

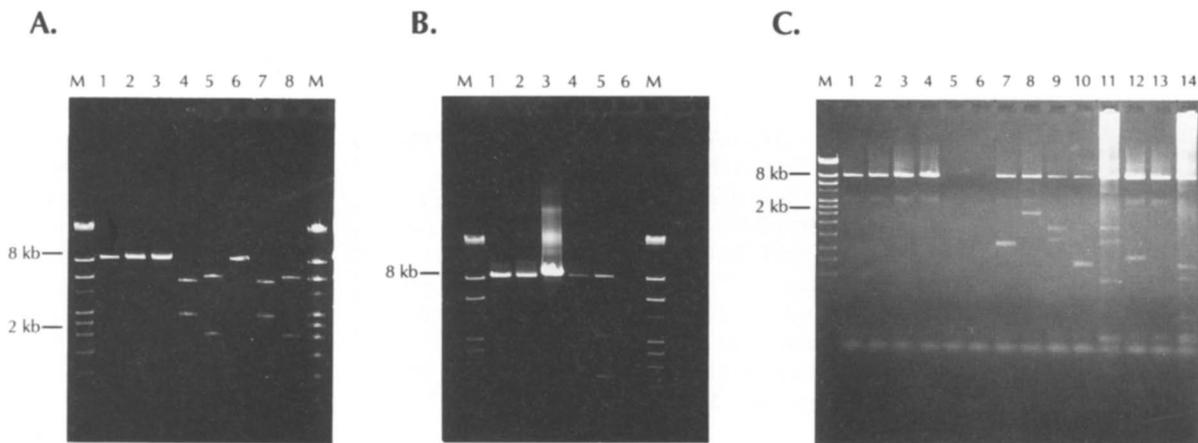


FIGURE 1 (A) Sensitivity of HPV16 model system amplification and RE verification of PCR products. Sensitivity was determined by amplification of linear HPV plasmid. (Lane 1) 10, (lane 2) 50, and (lane 3) 100 input copies. Products from linear plasmid (lanes 1–3) and cervicovaginal lavage specimen (lane 6) amplifications were verified by RE digestion (*HincII*–plasmid, lane 4 and lavage, lane 7; *PvuII*–lanes 5 and 8, plasmid and lavage, respectively). Products were also verified by RE digestion with *BsgI*, *HhaI*, *NcoI*, *NdeI*, *PstI*, and *XmnI* (data not shown). (B) HPV16 amplification from representative cervicovaginal lavages. (Lane 1) 100 copies of HPV16 linear plasmid; (lanes 2–5) full-length HPV genomes were successfully amplified from 10 of 10 HPV16-positive phenol-extracted lavages with variable yields; (lane 6) HPV16-negative lavage. (C) Variable specificity of HPV16 amplification from lavages. Phenol-extracted lavages yielded nonspecific amplification (refer to text for discussion). Amplifications were done in duplicate. (Lanes 1,2) 50 input copies of HPV16 linear plasmid; (lanes 3,4) 100 input copies; (lanes 5,6) HPV-negative lavage; (lane M) molecular weight marker IV (Boehringer Mannheim, Indianapolis, IN). Lavage B2726 was amplified in duplicate at a 1:3 (lanes 7,8) and a 1:15 (lanes 9,10) dilution in TE. *Bam*HI linearization increased product yield from this sample, and reduced nonspecific amplification. (Lanes 11,14) 1:3 dilution; (lanes 12,13) 1:15 dilution.

from these clones were subjected to sequence analysis. Figure 3 shows that 11 of 29 clones contained no errors in the sequenced region (~1 kb). Eight clones contained one error, and 10 clones contained two errors or more. None of the clones contained more than five errors in the region sequenced. In the clones with many errors (e.g., clones F and J in Fig. 3), base substitutions were clustered. All errors detected were simple base substitutions. No frameshift mutations were detected. The total number of errors detected in all clones combined was 33, which corresponds to an error frequency per nucleotide of approximately 1 per 700 bases (33 errors in 23,510 bp sequenced). The most common type of error was a T → C (or A → G) substitution (Figure 3). For semiconservative replication, either strand can serve as the template for producing an error. Only the mutational outcome can be determined and not the intermediate. The T → C (or A → G) transition was found in 23 of the 33 errors detected (70%). These types of substitutions were often found in A/T-rich stretches. The second most common type of error was a C → T (or G → A) transition (Fig. 3). This type of substitution was detected in 5 of the 33 errors found (15%). Only one base

substitution was found at the same position (A → G, position 446) in two different clones (F and Z in Fig. 3).

Amplification and Cloning of Novel HPV Genomes

LPCR methods established in the HPV16 model system were subsequently applied to the amplification of entire novel HPV genomes. The expected error frequency determined in our HPV16 model system (one error per 700 bases, or <0.2%) was predicted to be adequate for characterization of novel HPV genomes. We selected seven potential novel HPV genomes for further study: CP141, CP6108, CP8061, IS039, PAP238 (MM9), PAP291 (MM7), and W13B (MM4).^(16,18,22) One to three abutting primer pairs, with balanced melting temperatures, were designed for each of these potential novel HPVs. All primer pairs are located in the MY09/MY11 fragment of the L1 ORF. The sequences of the LPCR primers are shown in Table 1.

Three cervical specimens from different patients containing CP141 sequences were tested using LPCR conditions as described in Materials and Methods. Primer pair 141HP/HN produced an 8-kb LPCR product from one

of these specimens, 60224C (Fig. 4). We did not detect any products of this size in the other two specimens examined. In all three specimens a similar pattern of smaller nonspecific amplicons, ranging in size from 1 to 4 kb, was detected. These nonspecific products were detected only in amplifications from clinical specimens. Cloned, supercoiled CP141 genome (see below) produced a single, 8-kb band after amplification under similar conditions (data not shown). No amplification products were observed using primer pair 141BP/BN with the three specimens tested.

The primer pair 6108BP/BN was tested for amplification of a potential CP6108 genome in three samples obtained from one patient, 61085V (vulvar), 61085C (cervical), and a cervicovaginal lavage specimen, 61085L. Only the vulvar specimen produced an 8-kb LPCR product that was barely visible following EtBr staining. By decreasing the concentration of Mg(OAc)₂ (1.0 vs 1.1 mM), the yields of the amplification product increased slightly, although the band intensity remained low (data not shown). Also, by increasing the concentration of the *rTth* XL buffer (1.1 vs 1.0×), the yields of the CP6108 product were slightly increased (Fig. 4). The non-

LONG PCR OF HPV GENOMES AND AMPLIFICATION FIDELITY

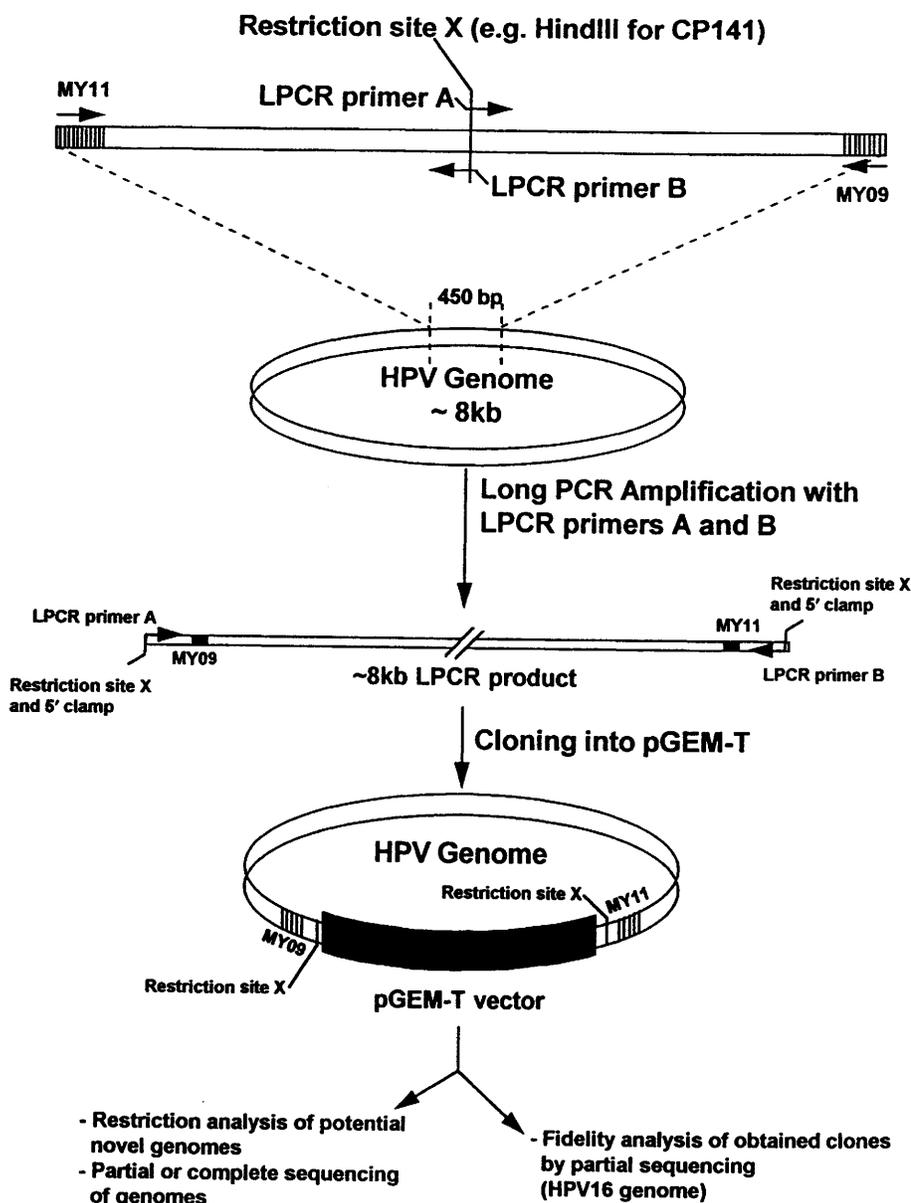


FIGURE 2 Generation, cloning, and analysis of LPCR products from HPV specimens. Abutting LPCR primers A and B were designed in the MY11-MY09 region of the novel HPV genomes. Amplification was performed using these primers as described in Materials and Methods. Complete 8-kb LPCR products were gel purified, cloned into a pGEM-T vector, and analyzed further using RE-based methods and sequencing analysis.

specific amplification of the CP6108BP/BN primer pair was generally very low and appeared as a smear (Fig. 4).

Two specimens, IS039 and IS215 containing IS039 sequences, were tested for LPCR amplification of entire IS039 genomes utilizing three different primer pairs, IS39AN/AP, IS39NN/NP, and IS39SN/SP. One of these specimens, IS039, produced an 8-kb product with primer pairs IS39AN/AP and IS39NN/NP. When the expected 8-kb genome was generated, we also observed abun-

dant background bands that varied in size from 0.5 to 7 kb (Fig. 4). No 8-kb amplification products were observed using primer pair IS39SN/SP with these specimens.

Seven specimens, IS601, IS780, IS461, C009, S214, S271, and S391, were subjected to LPCR amplification of entire PAP238 (MM9) genomes using the primer pair 238AEN/EP. Only one of these specimens, IS780, produced the targeted 8-kb amplification product. This amplification product was barely

visible following EtBr staining (data not shown).

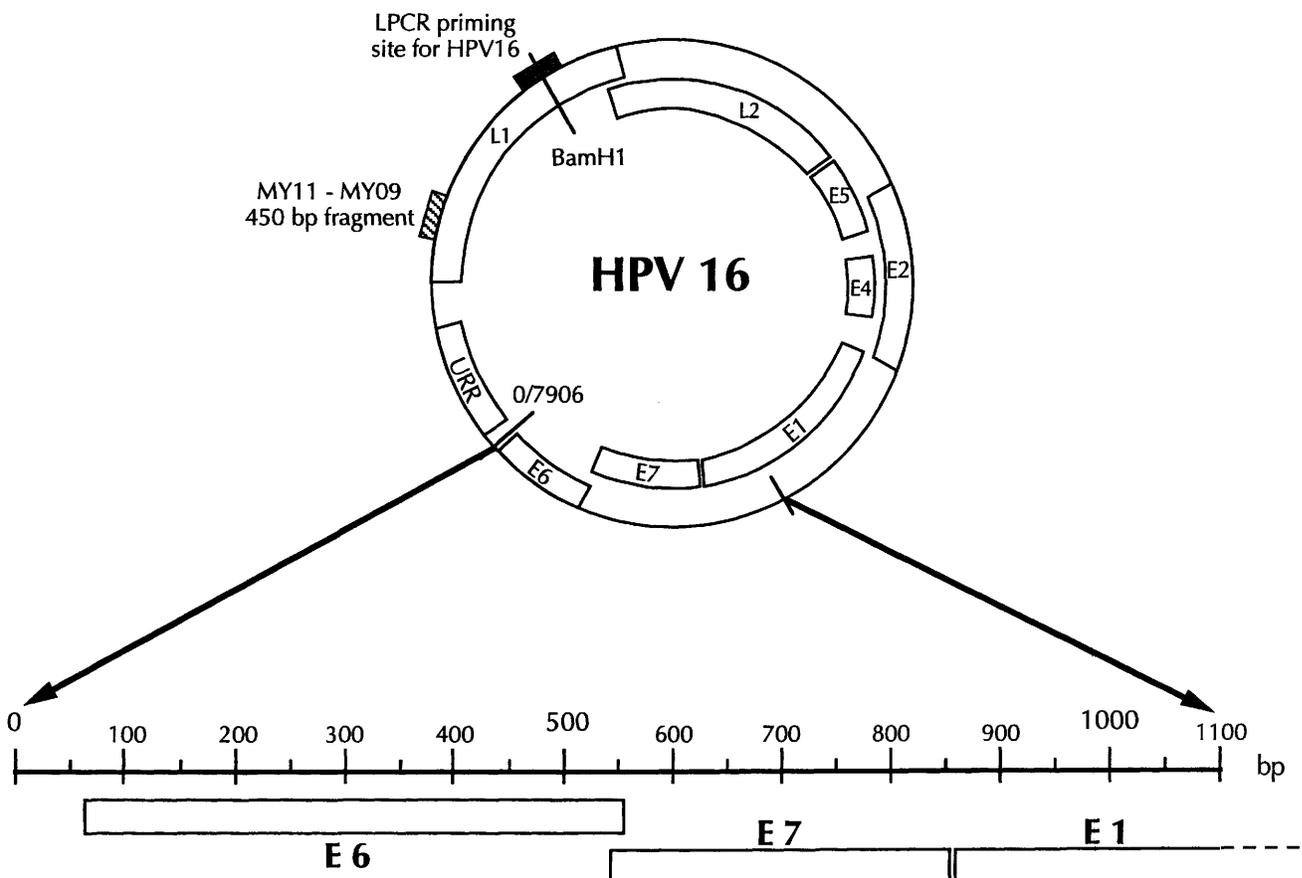
Using primer pairs W13BNN/NP, we tested six clinical specimens, W13B142, W13B385/389, W13B26018, IS887, IS766, and IS1016, for amplification of 8-kb HPV genomes. No amplification products of the expected 8-kb size were detected. An additional primer pair, IS39SN/SP, was applied to the same group of specimens and an 8-kb product was generated from W13B385/398 (Fig. 4).

We did not detect 8-kb LPCR products from any specimens using the primer pairs 8061BP/BN, 8061BgI/BgIP, and 291BP/BN, although optimization experiments were not attempted with these primer pairs.

The five potentially novel HPV genomes that produced an 8-kb amplification product using our LPCR system were subsequently subjected to T/A cloning using the pGEM-T vector (see Fig. 2 for an outline of the entire procedure). We succeeded in cloning only the CP141 LPCR product using this technique. A total of 10 recombinant bacterial colonies from the CP141 pGEM-T ligation were analyzed. Five of these contained an insert of the anticipated 8-kb size when analyzed using RE digestion with *Hind*III. Three of the CP141 clones, CP141-3, CP141-6, and CP141-7, were analyzed further by RE digestion. All three clones exhibited identical RE digestion patterns with the exception of clone CP141-7 that had a restriction fragment length polymorphism (RFLP) when digested with *Bst*NI. A physical map of the complete CP141 genome was constructed and is shown in Figure 5.

Several CP141 clones were subjected to partial sequence analysis using primers derived from the pGEM-T vector, *pGEM-F* and *pGEM-R*. Sequence analysis revealed that an additional adenine had been added to the 3' end of the cloned LPCR products; this addition was necessary for successful T/A cloning. In the sequenced L1 region around the *Hind*III cloning site, CP141 shares 81% of its base pairs with HPV39 DNA. The resulting amino acid sequence located within the 3' part of the L1 ORF spanned from amino acids 310–501 (numerically oriented from the first amino acid of the ORF) in the sequenced CP141 clones. Alignment to the L1 ORF of HPV39 identified 28 of 191 (14.7%) amino acids that were different in CP141 compared

STEWART ET AL.



clone	errors
A	0
B	2
D	2
E	2
F	4
G	1
H	0
I	0
J	5
K	1
L	2
N	0
O	1
P	0
Q	2
R	1
S	0
T	2
U	0
V	1
X	0
Y	1
Z	2
Å	2
Æ	0
AA	1
AB	0
AD	1
AE	0

X = T→C, A→G; Δ = G→A, C→T; O = T→A, A→T; ¶ = G→T

FIGURE 3 (See facing page for legend.)

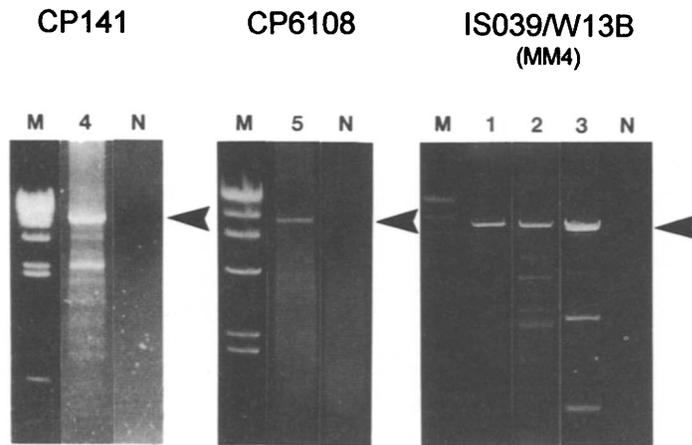


FIGURE 4 Amplification of entire novel genomes using LPCR. The reactions were performed essentially as described in Materials and Methods. (Lane 1) Specimen IS039, primer pair IS39AN/AP; (lane 2) specimen IS039, primer pair IS39NN/NP; (lane 3) specimen 385/(W13B), primer pair IS39SN/SP; (lane 4) specimen 60224C, primer pair 141HN/HP; (lane 5) specimen 61085V, primer pair 6108BN/BP; (lane M) λ /HindIII DNA (23, 130, 9416, 6557, 4361, 2322, 2027, 564, and 125 bp); (lane N) negative control (no input DNA).

to HPV39 (Fig. 6). Conserved residues common to all genital papillomaviruses⁽²⁴⁾ were also conserved in our isolate of CP141 (Fig. 6). One amino acid difference, an alanine substituted for a valine at position 444 was observed in our isolate compared to the published CP141 sequence.⁽¹⁶⁾ The alanine at this position has been observed previously in another CP141 isolate in our laboratory⁽³²⁾ and in the LVX160 MY09/MY11 L1 fragment sequence.⁽³³⁾ Subsequent to these determinations this HPV genome was designated as HPV70 (E.-M. de Villiers, pers. comm.).

The three LPCR products successfully amplified from IS039, W13B, and CP6108 were additionally subjected to RE digestion-based cloning techniques. The LPCR products from IS039, W13B, and CP6108 were digested with *Afl*III, *Sca*I, and *Bam*HI, respectively. Analysis of the RE digests in 1% agarose gels revealed 8-kb products suggesting that a single RE site for each selected enzyme was present in the targeted HPV genomes. The digested products were subsequently ligated to a similarly restricted vector. Competent cells were transformed, and several (>200) colonies were screened. No clones with the ex-

pected 8-kb insert were detected for any of the LPCR products.

DISCUSSION

We have demonstrated that LPCR methods can be applied to the successful amplification of previously characterized as well as potentially novel HPV genomes directly from crude preparations of clinical specimens. It is likely that these methods will subsequently facilitate the isolation of other complete viral genomes.

Although linearization via *Bam*HI restriction of circular HPV genomes present in lavage specimens improved LPCR product yields, our results suggest that the nicking that occurs during preparation of the nucleic acid and during thermal cycling creates a sufficient quantity of the relaxed form of the HPV DNA molecules to allow amplification of the full-length genomes. This observation is consistent with previous reports that supercoiled DNA cannot be heat denatured⁽³⁴⁾ and that the amplification rate with linearized DNA templates is significantly better than that with supercoiled templates.⁽³⁵⁾ Because PCR requires strand separation using

high heat denaturation (94°C), only linear or relaxed circle forms of the HPV DNA can serve as the templates for amplification. RE-mediated "nicking" of the DNA circle relaxes the supercoil, thus increasing the effective concentration of HPV DNA by causing a rise in the proportion of HPV templates suitable for amplification.

Our results also suggest that the non-specific amplification products obtained with the HPV16 genome from either lavages or purified cloned DNA are generated from the circular and/or supercoiled form of the HPV genome, as linear HPV and non-HPV targets have failed to generate background products with the exception of the 3.2- and the 5.5-kb products described above. *Bam*HI linearization would thus have been expected to eliminate the nonspecific amplification. Although this was not observed, linearization of the lavage-derived HPV DNA was not performed under conditions that are optimal for *Bam*HI restriction of supercoiled DNA templates. Thus, it may be possible to improve specificity further by optimizing the linearization protocol.

Our fidelity studies demonstrate that this LPCR method is comparable to previously reported PCR amplification systems.⁽⁸⁾ The total number of base substitutions detected in all clones was 33 in 23,510 bp sequenced (~1/700). Most errors found were transitions, and most of these are likely to be polymerase errors, although some of these errors may have arisen from DNA damage. The most common type of DNA damage that occurs during incubation at high temperatures is deamination of cytosine to produce uracil.⁽⁸⁾ Faithful replication of a template uracil by DNA polymerase will result in a C → T (or G → A) transition mutation. A second common type of DNA damage that may explain the nature of these errors is spontaneous base release resulting from hydrolysis of the N-glycosidic bond.⁽⁸⁾ This type of damage is also seen often during incubation at high temperatures and low pH. The resulting apurinic/apyrimidinic sites in the DNA can inhibit DNA synthesis or alternatively be bypassed by the DNA

FIGURE 3 Map of errors introduced during LPCR in ORFs E6, E7, and E1 (partial) of the HPV16 genome. The complete HPV16 genome was amplified, and HPV16 LPCR products were cloned as described in Materials and Methods. A total number of 29 individual clones were subjected to dideoxy sequencing of the region depicted and an assessment of LPCR fidelity performed. The number of errors in each individual clone is shown at right.

STEWART ET AL.

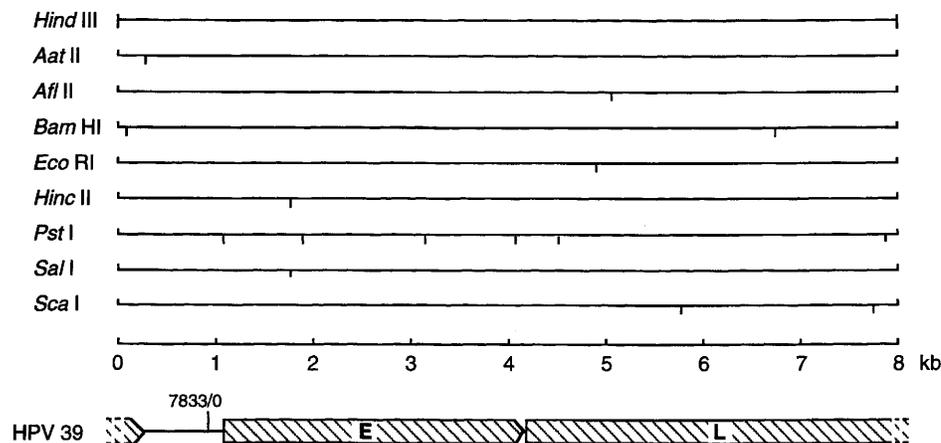


FIGURE 5 Physical map of the CP141 (HPV70) genome. The cleavage sites for REs were determined and are plotted on a linear map of the virus genome. The *Hind*III site is arbitrarily assigned as coordinate 0. The CP141 map has been aligned with the map of the HPV39 genome.⁽³¹⁾ The areas of the HPV39 genome corresponding to the early (E) and the late (L) regions are indicated at the *bottom*. The following RE sites are not present in the CP141 genome: *Ava*I, *Bgl*II, *Bgl*III, *Eco*RV, *Hpa*I, *Mlu*I, *Not*I, *Sac*II, and *Sma*I.

polymerase to create base-substitution mutations, most frequently transversions.

It is possible that optimizations,⁽³⁶⁾ including decreasing the number of cycles; lowering concentrations of dNTPs, divalent cations and primers; and altering thermal cycling parameters, can further improve the results presented in this report. Our analysis of fidelity was conducted, most probably, using LPCR products obtained from the plateau of amplification (as determined from the number of cycles used). The fidelity of this system may be different during the logarithmic phase. Additionally, because these studies were limited to amplification of 8-kb targets under specific

conditions, it is not known whether differences in fidelity will be observed for longer targets or when alternative amplification conditions are employed. It is reasonable to conclude, however, that LPCR using the *rTth* XL system described will be generally useful for applications that previously would have tolerated the fidelity of PCR-based methods.

Cloning of LPCR products was achieved by the T/A method^(37–39) suggesting that this enzyme system can add a single nucleotide to the 3' end of target molecules as has been reported similarly for PCR performed with *Taq* DNA polymerase.⁽⁴⁰⁾ This observation facilitates the rapid cloning of LPCR products and

may in some cases eliminate the need to design amplification primers that contain RE sites for subsequent use in cloning. It should be noted, however, that we observed variable success in applying this cloning method to LPCR products obtained from novel HPV genomes. We were only able to use T/A cloning successfully when a large quantity of LPCR amplicon was present as determined by EtBr staining in agarose gels. This may reflect a relatively low percentage of added 3' adenosines. It is possible that specific primer sequences affect the particular nucleotide added to the 3' termini by the *rTth* XL. Specific 3'-exonuclease activity has been reported for several other DNA polymerases.⁽⁴¹⁾ Alternatively, the addition of nucleotides to the 3' terminus of the PCR product may generally be less frequent⁽⁴¹⁾ under the selected LPCR conditions or the 3' → 5' exonuclease enzymatic activity might remove such nucleotides. Attempts were made to clone IS039 LPCR products by inactivating the *rTth* XL using proteinase K digestion and heat inactivation and, subsequently, adding 3' adenosines by incubating the LPCR products with *Taq* polymerase for 15 min at 72°C to facilitate T/A cloning. A total of 50 white clones were screened, but no clones with the correct insert size were detected using this approach (data not shown).

Our difficulties in cloning LPCR products using RE-based methods could also be an effect of remaining *rTth* DNA polymerase XL enzyme activity that could create blunt-ended products by filling the recessed 3' termini or that could act as an exonuclease to remove the protruding 3' termini created by digestion of the DNA with restriction enzymes. We have not implemented protocols to inactivate the *rTth* XL before RE digestion. We did, however, gel purify all RE-digested LPCR products prior to ligation attempts. Further studies are needed to determine the optimal cloning conditions for LPCR products.

The application of LPCR methods to the isolation of unknown or partially characterized infectious agents represents an important biochemical tool for future investigations. In the study of HPVs, isolation of novel genomes has been limited to the use of labor-intensive cloning in plasmid or phage libraries. These techniques require the presence of relatively high copy numbers of

	311						356
CP141	PSGSMV	SSDSQLFNKP	YWLHKAQGHI	NGICWHNQLF	ITVVDTRST		
Hpv39	PSGSMV	TSDSQLFNKP	YWLHKAQGHN	NGICWHNQLF	LTVVDTRST		
		*		*	^	^	^
	357						406
CP141	NFTLSACTET	AIPAVYSPTK	FKEYTRHVEE	YDLQFIFQLC	TITLTADVMA		
Hpv39	NFTLSTSIES	SIPSTYDPSK	FKEYTRHVEE	YDLQFIFQLC	TVTLTADVMS		
	^	***	*	^	^	*	^
	407						456
CP141	YIHTMNPAIL	DNWNIGVTPP	PSASLVDTYR	YLQSAAIACQ	KDAPTPEKKD		
Hpv39	YIHTMNSSIL	DNWNFAVAPP	PSASLVDTYR	YLQSAAITCQ	KDAPAPEKKD		
	^	^	^	^	^	*	^
	457						501
CP141	PYDDLKFWNV	DLKEKFSDEL	DQFPLGRKFL	LQVQARRRPT	IGPRK		
Hpv39	PYDGLKFWNV	DLREKFSLEL	DQFPLGRKFL	LQARVRRRPT	IGPRK		
	*	^	^	^	^	*	^

FIGURE 6 Amino acid sequence alignment of translated ORF L1, amino acids 311–501, from CP141 and HPV39. The alignment is numerically oriented from the first amino acid of the L1 ORF of the HPV39 genome. Asterisks (*) mark amino acid differences in the two sequences shown. Carats (^) mark conserved amino acid residues common to all genital papillomaviruses.

HPV genomes. The methods described here utilize microliter quantities of clinical specimens to obtain complete HPV genomes. Using a previously determined sequence for the MY09/MY11 region of the HPV L1 ORF, we were able to design a strategy that (1) amplified the entire circular genome of a previously uncharacterized HPV; (2) was amenable to T/A cloning of the ~8-kb genome; and (3) resulted in the ability to produce a linearized, intact HPV genome following RE digestion of the cloned genome.

It should be noted that the use of short PCR sequences to designate novel HPVs is supported by these investigations in which predicted entire novel HPV genomes were isolated using primers derived from short novel sequences. Current criteria for designation of a new HPV type requires <10% nucleotide similarity between the L1, E6, and E7 ORFs and any previously described HPV. Variants of HPV types are designated by $\leq 2\%$ difference in these sequences.⁽⁴²⁾ Fidelity studies presented in this report indicate that the expected error frequency of our LPCR method would be approximately one per 700 bases or <0.2%. On the basis of these data, LPCR products can be used to isolate and appropriately designate novel HPVs. LPCR methods represent valuable tools for future taxonomy studies of HPVs and potentially other uncharacterized infectious agents. The ability to use DNA molecules generated by LPCR methods in functional biologic assays remains to be determined.

ACKNOWLEDGMENTS

We thank the Roche Molecular Systems (RMS) DNA synthesis group for synthesis of oligonucleotides and the members of the IBSCC Study Group for clinical specimens. We are especially grateful to Cheri Peyton and Sheng-Yung Chang for invaluable information, and to Vanessa Padilla, John Herron, and April Martinez for excellent technical assistance. We thank Russ Higuchi, Henry Erlich, John Sninsky, Nick Matwiyoff, and Jim Neidhart for their continuing support and encouragement; Ray Apple for initiating and facilitating these collaborative studies; and Tom Myers and Rick Abramson for manuscript review. A.-C. Stewart is a recipient of a postdoctoral research training fellowship from International Agency for Research on Can-

cer/World Health Organization. Funding for this project was provided in part by National Institutes of Health grant A132917 and the University of New Mexico Cancer Research and Treatment Center (UNM-CRTC). These studies utilized the CRTC Molecular Analysis Facility.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Mullis, K.B. and F.A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155**: 335-350.
- Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.
- Rose, E.A. 1991. Applications of the polymerase chain reaction to genome analysis. *FASEB J.* **5**: 46-54.
- Erlich, H.A., D.H. Gelfand, and J.J. Sninsky. 1991. Recent advances in the polymerase chain reaction. *Science* **252**: 1643-1651.
- Barnes, W.M. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl. Acad. Sci.* **91**: 2216-2220.
- Cheng, S., C. Fockler, W.M. Barnes, and R. Higuchi. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci.* **91**: 5695-5699.
- Cheng, S., R. Higuchi, and M. Stoneking. 1994. Complete mitochondrial genome amplification. *Nature Genet.* **7**: 350-351.
- Eckert, K.A. and T.A. Kunkel. 1991. DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Applic.* **1**: 17-24.
- Cheng, S., S.-Y. Chang, P. Gravitt, and R. Respass. 1994. Long PCR. *Nature* **369**: 684-685.
- McCance, D.J. 1986. Human papillomaviruses and cancer. *Biochim. Biophys. Acta.* **823**: 195-205.
- Pfister, H. 1987. Human papillomaviruses and genital cancer. *Adv. Cancer Res.* **48**: 113-147.
- zur Hausen, H. 1987. Papillomaviruses in human cancer. *Cancer* **59**: 1692-1696.
- Hildesheim, A., M.H. Schiffman, P.E. Gravitt, A.G. Glass, C.E. Greer, T. Zhang, D.R. Scott, B.B. Rush, P. Lawler, M.E. Sherman, R.J. Kurman, and M.M. Manos. 1994. Persistence of type-specific human papillomavirus infection among cytologically normal women. *J. Infect. Dis.* **169**: 235-240.
- Schiffman, M.H., M.M. Manos, and C. Greer (unpubl.).
- Bauer, H.M., C.E. Greer, and M.M. Manos. 1992. Determination of genital HPV infection using consensus PCR. In *Diagnostic molecular pathology: A practical approach.* (C.S. Herrington and J.O.D. McGee, eds.), pp. 131-152. Oxford University Press, Oxford, UK.
- Peyton, C.L. and C.M. Wheeler. 1994. Identification of five novel human papillomaviruses in the New Mexico triethnic population. *J. Infect. Dis.* **170**: 1089-1092.
- Schiffman, M.H., H.M. Bauer, A.T. Lorincz, M.M. Manos, J.C. Byrne, A.G. Glass, D.M. Cadell, and P.M. Howley. 1991. Comparison of Southern blot hybridization and polymerase chain reaction methods for the detection of human papillomavirus DNA. *J. Clin. Microbiol.* **29**: 573-577.
- Peyton, C.L., A.M. Jansen, C.M. Wheeler, A.-C. Stewart, J. Peto, F.X. Bosch, N. Muñoz, A.R. Teyssie, M. Torroella, H.R. Wabinga, Sarjadi, C. Ngelangel, and M.M. Manos. 1994. A novel human papillomavirus sequence from an international cervical cancer study. *J. Infect. Dis.* **170**: 1093-1095.
- Bosch, F.X., M.M. Manos, N. Muñoz, M. Sherman, A.M. Jansen, J. Peto, M.H. Schiffman, V. Moreno, R. Kurman, K.V. Shah, and the IBSCC study group. 1995. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *J. Natl. Cancer Inst.* **87**: 796-802.
- Dürst, M., L. Gissman, H. Ikenberg, and H. zur Hausen. 1983. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proc. Natl. Acad. Sci.* **80**: 3812-3815.
- Seedorf, K., G. Krämmer, M. Dürst, S. Suhai, and W.G. Röwekamp. 1985. Human papillomavirus type 16 DNA sequence. *Virology* **145**: 181-185.
- Manos, M.M., J. Waldman, T.Y. Zhang, C.E. Greer, G. Eichinger, M.H. Schiffman, and C.M. Wheeler. 1994. Epidemiology and partial nucleotide sequence of four novel genital human papillomaviruses. *J. Infect. Dis.* **170**: 1096-1099.
- Manos, M.M., Y. Ting, D.K. Wright, A.J. Lewis, T.R. Broker, and S.M. Wolinsky. 1989. The use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. *Cancer Cells* **7**: 209-214.
- Bernard, H.-U., S.-Y. Chan, M.M. Manos, C.-K. Ong, L.L. Villa, H. Delius, C.L. Peyton, H.M. Bauer, and C.M. Wheeler.

STEWART ET AL.

1994. Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction amplification, restriction fragment length polymorphisms, nucleotide sequence and, phylogenetic algorithms. *J. Infect. Dis.* **170**: 1077-1085.
25. Chou, Q., M. Russell, D.E. Birch, J. Raymond, and W. Bloch. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Res.* **20**: 1717-1723.
26. Horwitz, J.P., J. Chua, R.J. Curby, A.J. Tomson, M.A. Da Rooge, B.E. Fisher, J. Mauricio, and I. Klundt. 1964. Substrates for cytochemical demonstration of enzyme activity. I. Some substituted 3-indolyl- β -D-glycopyranosides. *J. Med. Chem.* **7**: 574-575.
27. Ullman, A., F. Jacob, and J. Monod. 1967. Characterization by *in vitro* complementation of a peptide corresponding to an operator-proximal segment of the β -galactosidase structural gene of *Escherichia coli*. *J. Mol. Biol.* **24**: 339-343.
28. Birnboim, H.C. and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**: 1513-1523.
29. Birnboim, H.C. 1983. A rapid alkaline extraction method for the isolation of plasmid DNA. *Methods Enzymol.* **100**: 243-255.
30. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463-5467.
31. Volpers, C. and R.E. Streech. 1991. Genome organization and nucleotide sequence of human papillomavirus type 39. *Virology* **181**: 419-423.
32. Peyton, C.L. and C.M. Wheeler (unpubl.).
33. Ong, C.-K., H.-U. Bernard, and L.L. Villa. 1994. Identification of genomic sequences of three novel human papillomavirus sequences in cervical smears of Amazonian Indians. *J. Infect. Dis.* **170**: 1086-88.
34. Marguet, E. and P. Forterre. 1994. DNA stability at temperatures typical for hyperthermophiles. *Nucleic Acids Res.* **22**: 1681-1686.
35. Linz, U., U. Delling, and H. RübSamen-Waigmann. 1990. Systematic studies on parameters influencing the performance of the polymerase chain reaction. *J. Clin. Chem. Clin. Biochem.* **28**: 5-13.
36. Gelfand, D. and J. Sninsky. 1993. Fidelity questions and answers. *Amplifications* **11**: 16-19.
37. Holton, T.A. and M.W. Graham. 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucleic Acids Res.* **19**: 1156.
38. Marchuk, D., M. Drumm, A. Saulino, and F.S. Collins. 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* **19**: 1154.
39. Mead, D.A., N.K. Pey, C. Herrstadt, R.A. Marcil, and L.M. Smith. 1991. A universal method for the direct cloning of PCR amplified nucleic acid. *BioTechnology* **9**: 657-663.
40. Clark, J.M. 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* **16**: 9677-9686.
41. Hu, G. 1993. DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3' end of a DNA fragment. *DNA Cell Biol.* **12**: 763-770.
42. Van Ranst, M.A., R. Tachezy, H. Delius, and R.D. Burk. 1993. Taxonomy of the human papillomaviruses. *Papillomavirus Rep.* **4**: 61-65.

Received April 17, 1995; accepted in revised form June 9, 1995.



Generation of entire human papillomavirus genomes by long PCR: frequency of errors produced during amplification.

A C Stewart, P E Gravitt, S Cheng, et al.

Genome Res. 1995 5: 79-88

Access the most recent version at doi:[10.1101/gr.5.1.79](https://doi.org/10.1101/gr.5.1.79)

References This article cites 39 articles, 8 of which can be accessed free at:
<http://genome.cshlp.org/content/5/1/79.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>
