JOURNAL OF VIROLOGY, May 1983, p. 557–566
0022-538X/83/050557-10$02.00/0

Vol. 46, No. 2

Comparative Analysis of the Human Type 1a and Bovine Type 1 Papillomavirus Genomes

OLIVIER DANOS,[1]† LINDA W. ENGEL,[2] ELLSON Y. CHEN,[3] MOSHE YANIV,[1] AND PETER M. HOWLEY[2]*

Unité des Virus Oncogènes, Departement de Biologie Moléculaire, Institut Pasteur, 75724 Paris Cedex 15, France[1]; Laboratory of Pathology, National Cancer Institute, Bethesda, Maryland 20205[2]; and Department of Molecular Biology, Genentech Inc., South San Francisco, California 94080[3]

Received 10 January 1983/Accepted 23 February 1983

The DNA sequences of the genomes of the bovine type 1 and human type 1a papillomaviruses were compared. The overall organization of both genomes is very similar. Three areas of maximal homology were found in the L1 and E1/E2 genes, and at the beginning of L2. The conservation of homologous amino acid sequences encoded in the open reading frames argues that these segments represent real genes or exons. Within these segments, however, only certain domains of the putative proteins are preferentially conserved. Two polypeptide chains show homologous arrangement of the cysteine residue clusters Cys-X-X-Cys, despite a lack of conservation of the rest of the amino acid sequence. A significant sequence divergence in a region where the three reading frames are open suggests that papillomavirus genomes have evolved not solely by accumulation of point mutations. Conserved sequences were also found in the noncoding region, and their possible involvement in regulation of viral gene expression is discussed.

Papillomaviruses are members of the Papovaviridae family and induce benign squamous epithelial tumors in a wide variety of higher vertebrates (for a review, see reference 20). The proliferative lesions caused by most papillomaviruses (such as human papillomavirus type 1a [HPV-1a]) are limited to the squamous epithelium. However, there is a subgroup of papillomaviruses (including bovine papillomavirus type 1 [BPV-1]) which induce fibropapillomas consisting of both a proliferative dermal fibroblastic component and the proliferative squamous epithelial component. The full vegetative cycle of the papillomaviruses is expressed only in the terminally differentiating squamous epithelium of the papillomas. Although these self-limiting tumors usually regress, some papillomas may undergo malignant progression (18, 28). The molecular genetics of papillomaviruses have not been well understood, mainly due to the lack of a permissive in vitro tissue culture system. In the past few years, the molecular cloning of a number of papillomavirus genomes has allowed their biochemical characterization and the study of their genetic diversity (6, 16, 17). By using molecularly cloned DNA, the complete genomes of HPV-1a (7) and BPV-1 (4) have been sequenced.

† To whom reprint requests should be addressed.

Although they share a number of common features, BPV-1 and HPV-1a differ in many of their biological properties. Whereas the plantar and palmar warts caused by HPV-1a are entirely epithelial, the paragenital warts (3) and cutaneous fibropapillomas in cattle caused by BPV-1 (20) consist of both epithelial and underlying mesenchymal components. Furthermore, BPV-1 is one of a subgroup of papillomaviruses which is oncogenic in hamsters (12) and capable of transforming mouse fibroblasts (9). In contrast, HPV-1a is nontumorigenic when inoculated into hamsters and does not induce foci on mouse fibroblasts. Both BPV-1 and HPV-1a DNAs are able to replicate as extrachromosomal plasmids in cells. The BPV-1 genome remains exclusively extrachromosomal as a stable multicopy (10 to 200 copies per cell) plasmid in transformed mouse cells (23), and recently La Porta and Taichman have shown that HPV-1a DNA can persist as a stable plasmid in cultured human epidermal cells (21). Sequence analysis indicated that the double-stranded circular genomes are organized in a similar fashion, each containing a single coding DNA strand on which open reading frames are similarly distributed. Yet under standard stringent hybridization conditions, little if any homology can be detected between BPV-1 and HPV-1a DNAs (22).

In this report we present a detailed analysis of

the sequence comparison between the BPV-1 and HPV-1a genomes with the following questions in mind. (i) What is the nature of the conserved DNA sequences which have been demonstrated by nonstringent hybridization experiments (17; L. T. Chow and T. R. Broker *in* M. L. Pearson and N. L. Sternberg, ed., *Gene Transfer and Cancer*, in press), and can they be related to the common features of the two viruses, such as the viral capsid structure, the tropism for squamous epithelial cells, or the ability of the DNA to persist and replicate as a stable plasmid in transformed and latently infected cells? (ii) Do the differences between the two genomes correlate with the particular biological activity and host specificity of each virus? (iii) How has the evolution of papillomavirus genomes progressed, and in what way did it produce the observed diversity of the viral family?

MATERIALS AND METHODS

Sequence data. The modifications introduced in the previously reported HPV-1a DNA sequence (7) are the following: T added between positions 538 and 539; C deleted at positions 700 and 707; G added between positions 834 and 835; G added between positions 5106 and 5107; C deleted in 5494; at position 6513, change CG to GGCTG. The numbering system of the HPV-1a DNA sequence has been changed to start with the G of the *Hpa*I site (5'GTTAAC 3') at position 4275. The modifications introduced in the published BPV-1 DNA sequence (4) are the following: G changed to C at position 7305; G deleted at position 7585.

Computer analysis. The nucleic acid sequences of each genome were analyzed by using the program of Queen and Korn (29). Distribution of termination codons was determined with a program in which groups of 10 codons were observed and a vertical bar drawn when at least one terminator codon occurred within the set (7). Comparison of the BPV-1 and HPV-1a sequences employed the enhanced matrix procedure of Maizel and Lenk (25). A filtration factor of 5 required the identity of 5 successive nucleotides for a dot to be denoted on the graph.

RESULTS AND DISCUSSION

Overall comparison of genome structure. The BPV-1 and HPV-1a genomes consist of 7,944 and 7,814 base pairs, respectively. The distribution of the open reading frames on the coding strand of each viral DNA is shown in Fig. 1. The two genomes have been aligned by using the *Hpa*I site whose relative position was identical in a homologous segment on the physical maps. The numbering system of HPV-1a nucleotide sequence has been changed to correspond to that of BPV-1 to start with the first of the six nucleotides of the *Hpa*I site (previously at position 4275). Corrections included in the HPV-1a and BPV-1 sequences from those previously published are listed above.

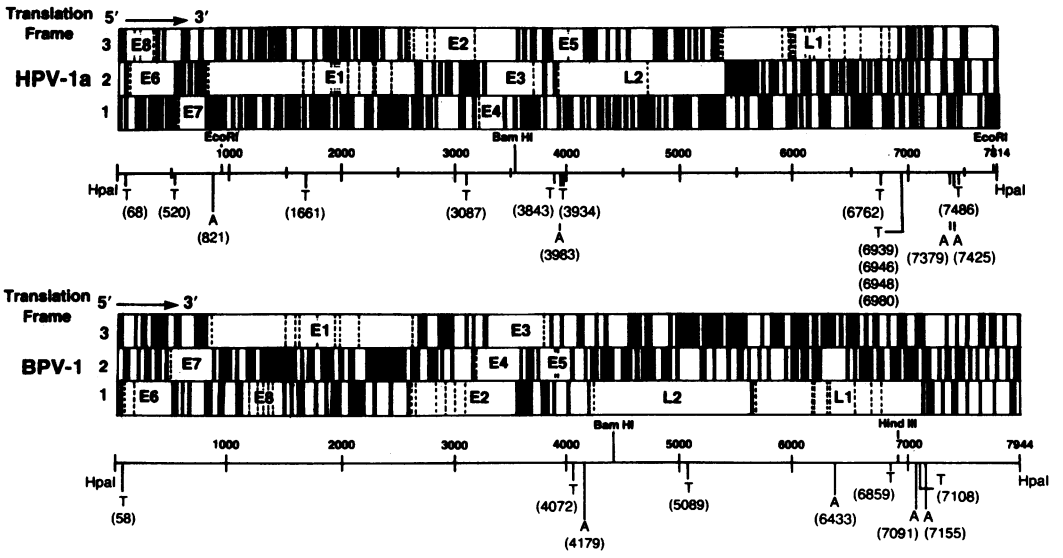

FIG. 1. Distribution of termination codons of HPV-1a and BPV-1 genomes on their respective sense strands. Groups of 10 codons are observed in each of the three reading frames, and a vertical bar is drawn when at least one termination codon occurs among them (7). The width of the heavy bars reflects successive groups of 10 codons, each containing at least one termination codon. Open reading frames potentially encoding a peptide greater than 90 amino acids in size have been designated E1 through E8 and L1 and L2. Dashed vertical lines superimposed on these regions indicate positions of the translational initiation codon ATG. The letter A refers to potential polyadenylation recognition signals (AATAAA). The letter T indicates the presence of a potential TATA sequence, either TATAAA or TATATA.

The overall organization of the genomes is the same for these two papillomaviruses and clearly contrasts with the genomic organization of the polyomaviruses simian virus 40, BK virus, or polyoma (30), in which early and late functions are encoded on different DNA strands. We have not observed any significant blocks of homology between the DNA sequences of the polyomaviruses and the papillomaviruses. Thus, the papillomaviruses should be considered as a separate viral family from the polyomaviruses, distinguishable by their characteristic genomic organization and by their specific tropism for squamous epithelial cells.

As depicted in Fig. 1, all of the coding sequences in the papillomavirus DNAs are presumed to be on one strand. The existence of coding sequences on the other strand of papillomavirus DNA cannot be absolutely ruled out on the basis of the sequence analysis, but two observations make it unlikely: (i) there is no conservation of open reading frames between the antisense strands of BPV-1 and HPV-1a (data not shown), and (ii) all the polyadenylated RNA species detected and mapped from the BPV-1 genome are derived from a single strand (1, 15; L. Engel, C. A. Heilman, and P. M. Howley, submitted for publication).

It is possible to functionally divide the papillomavirus genomes into regions. The *Bam*HI-*Hin*dIII 69% subgenomic fragment of BPV-1 DNA is sufficient for neoplastic transformation of mouse fibroblasts (17, 24), and all of the transcripts present in transformed mouse cells (15) or in BPV-1-induced hamster tumors map within this fragment (1, 11). We have designated this part of the BPV-1 genome and the corresponding HPV-1a segment as the E region. The open reading frames located within this segment have been designated E1 through E8. The designation has been made based on size of the open reading frame, homology, and location within the genome. The remaining 31% fragment in each genome contains two large open reading frames designated L1 and L2. Transcripts from this region are detected in productively infected bovine fibropapillomas (1; Engel et al., submitted for publication), and transcripts from this region can direct the synthesis of the 55,000-molecular-weight major structural protein VP1 in vitro (L. Engel, C. Heilman, and P. Howley, manuscript in preparation). The E and L regions are separated at the 3′ end of the L region by a noncoding segment.

The enhanced graphic matrix procedure was used to analyze the nucleic acid sequences for features of possible biological interest and to reveal the spatial patterns of such features. Comparison of the two different sequences presented in Fig. 2A shows domains of similarity, regions of divergence, and insertions and deletions. In this analysis, the identity of five successive nucleotides is denoted as a dot on the graph. The schematic redrawing of this comparison in Fig. 2b highlights the three regions of maximal homology revealed by this analysis. The two largest homologous regions mapping in the L1 and E1/E2 open reading frames have been previously detected by electron microscopy heteroduplex observation (5) and blotting experiments with hybridization conditions of reduced stringency (16, 22).

The positions of the homologous regions relative to the diagonal are important. The first region localized to the C-terminal half of E1 and the N-terminal half of E2 lies on the major diagonal, indicating that the distance between the *Hpa*I site and the N-terminal half of the E2 gene is conserved in both viral genomes. The second region of homology, located in the N-terminal portion of L2, is shifted up from the diagonal, resulting from an insertion of roughly 280 base pairs (bp) in the BPV-1 sequence or an equivalent deletion from the HPV-1a sequence in the region where E3, E4, and E5 are encoded (roughly, between positions 3000 and 4000). Interestingly, E5 is 297 bp long and has no equivalent on the HPV-1a genome (the HPV-1a E5 reading frame shown in Fig. 1 does not correspond to the one of BPV-1, as it overlaps with L2 and encodes a peptide with a totally unrelated amino acid sequence). This region of BPV-1 also contains the polyadenylation site (position 4179), which is used for transcripts detected in mouse transformed cells (15) and for the E set of transcripts in the bovine fibropapilloma (Engel et al., submitted for publication). The third region of homology in L1 is shifted back toward the major diagonal, indicating an insertion of about 120 bp in the HPV-1a sequence. Interestingly, the L2 gene present in this region is 114 bp longer in HPV-1a. This analysis illustrates the possible involvement of an insertion-deletion process in the diversification of papillomavirus genomes. Consequently, the 131 additional base pairs present in BPV-1 relative to HPV-1a (7,944 bp versus 7,814 bp) are not evenly distributed throughout the genome.

The first region of homology includes the C-terminal half of E1 and the N-terminal half of E2. The two genes overlap for 56 bp in both genomes. The DNA sequence conservation is in the range of 65 to 70%, a degree of homology that was not detectable by standard procedures of DNA-DNA hybridization. In the C-terminal half of E1 (last 300 amino acids, Fig. 3a) there is 52% amino acid homology, which increases to 62% if conservative changes such as lysine to arginine or glutamate to aspartate are permitted.

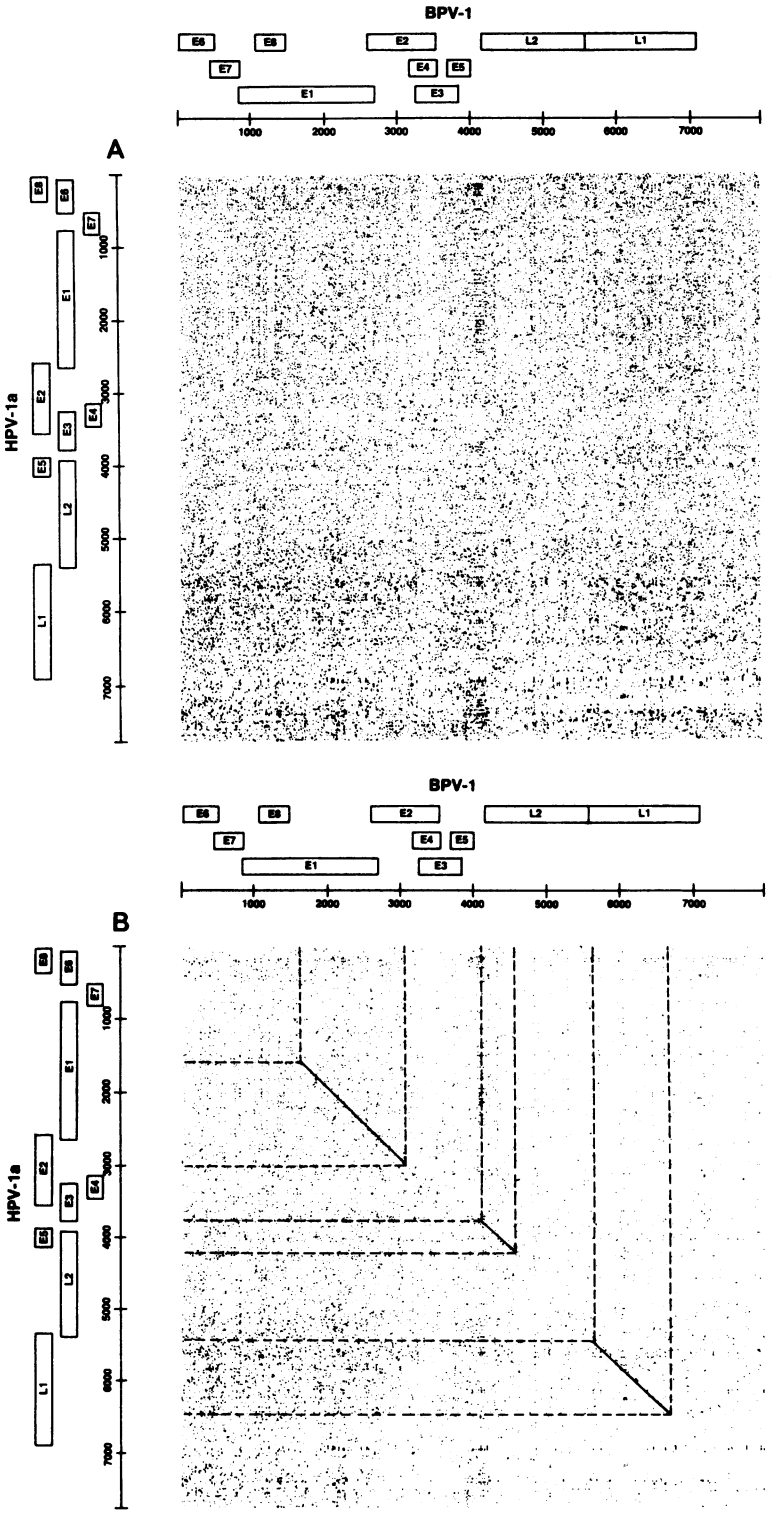

FIG. 2. (A) Graphic matrix of HPV-1a and BPV-1 DNA sequences. HPV-1a is on the vertical axis, and BPV-1 is on the horizontal axis. Sequences were filtered by collating five nucleotides at a step, employing the program of Maizel and Lenk (25). (B) Schematic representation highlighting regions of partial homology.

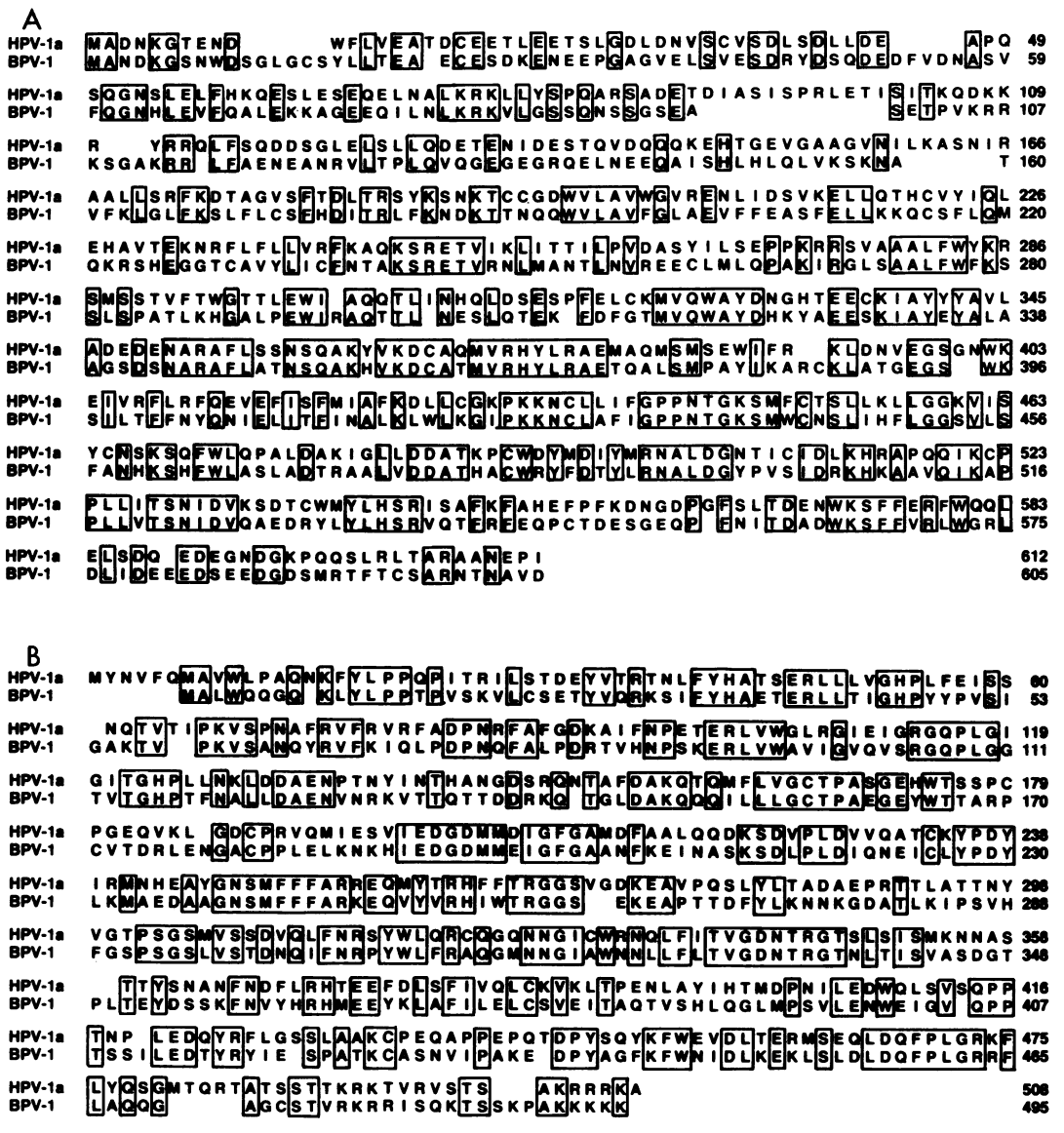

FIG. 3. Amino acid conservation between HPV-1a and BPV-1 genes. (A) E1 open reading frame. (B) L1 open reading frame. The amino acids are denoted in one-letter codes as follows: A, alanine; V, valine; L, leucine; I, isoleucine; P, proline; F, phenylalanine; W, tryptophan; M, methionine; G, glycine; S, serine; T, threonine; C, cysteine; Y, tyrosine; N, asparagine; Q, glutamine; D, aspartic acid; E, glutamic acid; K, lysine; R, arginine; H, histidine.

Many changes in the DNA sequences involve the third base of codons and hence are not reflected as amino acid changes.

Sequence conservation is localized to particular domains of E1 and E2. Several insertions have had to be introduced in the N-terminal part of E1 to match BPV-1 and HPV-1a sequences (Fig. 3a), whereas the C-terminal halves align perfectly. In the N-terminal two-thirds of E2, 37% of the amino acids are identical, whereas only 4.5% are the same for the C-terminal one-third. The conserved E2 domain stops where the overlap with E3 and E4 coding regions begins (see Fig. 1). This variability could reflect a double function of the E2 proteins, one function being common to the viral family and the other particular to the virus type.

Little homology is detected in the region of the two genomes where the E2, E3, and E4 open reading frames overlap (approximately between

TABLE 1. Nucleotide positions of the open reading frames of HPV-1a and BPV-1 DNA

Open reading frame	HPV-1a DNA position			BPV-1 DNA position		
	First nucleotide	First methionine	Nucleotide preceding stop codon	First nucleotide	First methionine	Nucleotide preceding stop codon
L1	5346	5412	6935	5596	5608	7092
L2	3893	3896	5416	4171	4186	5592
E1	773	812	2647	813	849	2663
E2	2568	2592	3557	2581	2608	3525
E3	3263		3793	3267		3836
E4	3157	3181	3432	3173	3191	3550
E5	3876		4151	3713		4009
E6	86	104	523	49	91	501
E7	502	529	807	449	479	859
E8	45	165	353	1099	1204	1479

positions 3250 and 3600 on each genome), yet the organization of the open reading frames in this region of the two genomes has been conserved. It seems likely that each of these open reading frames encodes viral peptides in that the 5′ ends of bodies of mRNAs have been mapped near the ends of these open reading frames in BPV-1-transformed cells (15), and RNA species in HPV-1-infected COS cells mapped to this region (Chow and Broker, in press). A region of a genome encoding three different peptides would necessarily be under evolutionary constraints in that changes in the DNA sequence of this region would be strongly counterselected. It seems likely, therefore, that the divergence detected between the genomes in this region may be the result of a combination of genetic events. In addition to the accumulation of point mutations, one can propose the hypothesis of an exchange of genetic material between the virus and its host, between different viruses, or both. The presence of sequences related to the human *Alu* repeat in this region of the HPV-1a genome (7) supports this hypothesis.

The second region of homology shown in Fig. 2 corresponds to the N-terminal 50 amino acids of L2, which are 60% identical. The remaining part of the L2 genes have diverged considerably, except for the sequences encoding the protein C-terminus, which starts the third region of homology between both viruses. This region includes the open reading frame for the L1 protein. The 100-bp A-T-rich region of the HPV-1a genome, located after L1, is displayed on the Fig. 2 computer matrix as a horizontal zone where the mean density of matches is lower. No such large A-T-rich region is noted for the BPV-1 DNA sequence.

Further analysis of the open reading frames. On each genome, we have arbitrarily chosen to consider as open reading frames any region with a coding capacity larger than 90 amino acids. The coordinates of these reading frames are presented in Table 1. The lengths and the calculated molecular weights for putative proteins encoded in these open reading frames are listed in Table 2. These two tables show that coding segments of comparable sizes are present at the same position on the two genomes, except for E5 and E8. Neither the E5 nor the E8 coding regions has any equivalent at the corresponding position on the other genome (see Fig. 1). Whether they are only randomly occurring open reading frames which are not utilized or whether they specifically encode peptides for each virus is not known.

The examination of amino acid sequences also reveals similarities between the predicted peptides of the two genomes in other coding regions. Thus, with the possible exceptions of E5 and E8, each of the open reading frames identified in Tables 1 and 2 is likely to be a genuine gene or exon. As an example, E3 has no ATG in either of the two genomes, but the amino acid homologies in the C-terminal part (not shown) strongly suggest that it is an exon. Chow and Broker have shown, using HPV-1a-simian virus 40 hybrids propagated in COS cells, that an HPV-1a-specific RNA is spliced in the region corresponding to the 5′ end of E3 (Chow and Broker, in press). Also, Heilman et al. have mapped the 5′ end of the body of an mRNA species in BPV-1-transformed mouse cells to the 5′ end of E3 (15).

The case of E6 and E7 is of particular interest. Only sparse amino acid sequence homology is detected between the corresponding open reading frames for each genome; however, for both BPV-1 and HPV-1a, each of these polypeptides is cysteine rich (E6 contains 10.9 and 7.1% cysteine in BPV-1 and HPV-1a, respectively). Fig. 4 shows that the motif Cys-X-X-Cys is repeated four times in the amino acid sequence of E6 on both genomes, with conserved dis-

TABLE 2. Amino acid capacity of the open reading frames of HPV-1a and BPV-1 DNA[a]

Open reading frame	HPV-1a DNA		BPV-1 DNA	
	Amino acid capacity	Amino acid capacity after the first methionine	Amino acid capacity	Amino acid capacity after the first methionine
L1	530 (60,080)	508 (57,638)	499 (56,092)	495 (55,556)
L2	508 (55,463)	507 (55,335)	474 (50,475)	469 (50,020)
E1	625 (71,230)	612 (69,878)	617 (69,407)	605 (68,197)
E2	330 (37,870)	322 (36,926)	315 (35,490)	306 (34,310)
E3	177 (20,172)		190 (21,329)	
E4	92 (10,113)	84 (9,316)	126 (13,296)	120 (12,564)
E5	92 (10,721)		99 (11,473)	
E6	146 (16,962)	140 (16,318)	151 (17,434)	137 (15,851)
E7	102 (11,526)	93 (10,500)	137 (14,597)	127 (13,637)
E8	103 (12,155)	63 (7,535)	127 (14,783)	92 (10,658)

[a] The numbers in parentheses are the sizes of the putative proteins (in daltons).

tances maintained between the repeating units. In E7, the same motif is repeated twice, with a spacing of 29 amino acid residues for each of the two viruses (data not shown). We interpret this to mean that the tertiary structure of each of these pairs of peptides has been maintained with little conservation of the primary amino acid sequence. Such a repeated pattern of cysteine residues is reminiscent of a conserved feature of the C-terminus of the small T antigen of a variety of polyomaviruses, including simian virus 40, BK virus, and polyoma, where the motif Cys-X-Cys-X-X-Cys is repeated twice (14).

The L region contains two open reading frames, L1 and L2. The amino acid sequence of the L1 protein is highly conserved (45% of identical amino acids and 16.5% of conservative changes—see Fig. 3b). The BPV-1 and HPV-1a L1 proteins seem to have diverged by accumulation of point mutations; no insertions or deletions are observed in the sequence. The conservation of many stretches of amino acids between the two proteins is reflected by the immunological cross-reaction observed between BPV-1 and HPV-la capsid proteins detectable with antisera raised against disrupted virions (19, 27). These conserved stretches of amino acids between the two proteins are apparently not exposed on the surface of the virions in that antiserum to intact virus is type specific and generally does not cross-react among different papillomaviruses.

As mentioned above, only the terminal portions of the L2 protein are conserved. L2 may be a minor component of the capsid, and the variability of its internal region suggests that it might play a role in host restriction. In both viruses, L1 and L2 have a highly basic C-terminus. Such a terminal basic stretch is also found in the VP2 and VP3 minor capsid proteins of polyomaviruses (10), and it may constitute a site of interaction with the DNA inside the virion.

In the BPV-1 genome, the L1 and L2 genes are located in the same translational frame, and the putative methionine start codon is located 15 bases after the L2 stop codon (Fig. 1, Table 1). In the HPV-1a genome, the L1 and L2 genes are located in different translational frames, and the first ATG of L1 overlaps the end of L2, preceding the stop codon by five bases (Fig. 1, Table 1). As shown in Fig. 3B, the homology between the L1 proteins begins with the second methionine of the HPV-1a protein. On both genomes, the DNA sequence just upstream of the start codon for the L1 protein contains a conserved potential splice acceptor site. These sequences are, for BPV-1 and HPV-1a, respectively, 5'-

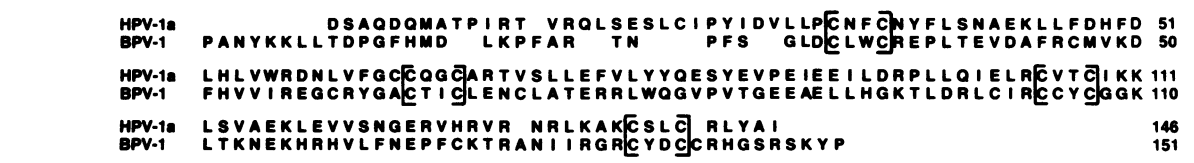

FIG. 4. Amino acid sequences of HPV-1a and BPV-1 E6 open reading frames. Amino acids are denoted in one-letter codes. See the legend to Fig. 3 for abbreviations. The four bracketed sets of amino acids contain the conserved amino acid structural feature Cys-X-X-Cys.

TTTTTGCAG(ATG)-3′ and 5′-GTTTTTCA G(ATG)-3′. Engel et al. (submitted for publication) have mapped the end of the body of a 1,700-base mRNA encoding the L1 protein in bovine papillomas, using S1 analysis to this splice junction. Chow and Broker (in press) have shown that a splice site is used in this region of the HPV-1a genome to produce an RNA body corresponding to the L1 gene(s). Together, these observations suggest that the HPV-1a L1 protein starts with the second ATG of the open reading frame and the BPV-1 L1 protein with the first ATG.

Noncoding region and putative regulatory elements. The region extending between the end of L1 and the start of E6 is noncoding on both genomes, and it contains several conserved features which may constitute regulatory signals for viral gene expression and DNA replication. The size of the noncoding region is 982 bp for HPV-1a and 943 for BPV-1. It is composed, 5′ to 3′, of an A-T-rich region, two successive polyadenylation sites, directly repeated sequences, and, just upstream of the E6 ATG, sequences which could constitute a promotor for polymerase II.

The A-T-rich region is more extensive on the HPV-1a genome and covers about 100 bp (90% A or T) just after the L1 stop codon. The 3′ ends of the "wart-specific" transcripts from the BPV-1 L region map to the region containing the two polyadenylation sites (5′-AATAAA-3′) in the BPV-1 genome (1; Engel et al., submitted for publication), and S1 mapping experiments indicate that it is the second of these sites which is utilized in the biosynthesis of the BPV-1 L region mRNAs at position 7155 (Engel et al., submitted for publication). On the HPV-1a genome, the two polyadenylation sites are found 443 and 489 bp after the end of L1, and their localization fits with the R-loop mapping of 3′ ends of transcripts from HPV-1a-simian virus 40 hybrid molecules transcribed in COS-1 cells (Chow and Broker, in press).

A region of repeated sequences found in both genomes is shown in Fig. 5. Interestingly, the BPV-1 10-bp repeat unit is complementary to the HPV-1a repeated sequence. This implies that if this sequence has any regulatory function, it would be independent of its orientation in the genome with respect to the transcriptional direction, a feature which is characteristic of a transcriptional activator (2, 26). Lusky et al. have localized a sequence in the BPV-1 genome to the 3′ end of the E transcriptional unit at position 4391 to 4454 which does not include these repeat units and which functions as an activator of gene expression (M. Lusky, L. Berg, H. Weiher, and M. Botchan, submitted for publication). In the assays performed, BPV-1 DNA fragments including the noncoding region and these repeat units did not have transcriptional enhancer activity (Lusky et al., submitted for publication). The precise function of 10-bp repeat units in the noncoding region, therefore, remains to be elucidated.

A TATAAA sequence, a characteristic sequence component of a Polymerase II transcriptional promoter in mammalian cells, is located approximately 30 bp before the ATG for E6 in both genomes (see Fig. 5). The dinucleotide CA is also conserved just before the ATG, and either base might serve as a cap site. Sequences about 100 bp upstream from the cap site are known to be required for a functional polymerase II promoter (8). In the potential promoter described here, conserved sequences are found around the *Hpa*I site of BPV-1 and HPV-1a (Fig. 5). The conservations of sequences in the noncoding regions leads us to speculate that this region may function as part of a polymerase II promoter element; however, this hypothesis must be confirmed by more precise mapping of 5′ ends of the viral transcripts and functional studies on these regions.

Conclusion. Initial evaluations of the sequence data for BPV-1 and HPV-1a (4, 7) were limited by the absence of genetic studies on these viruses. A similar situation existed for the human and woodchuck hepatitis B viruses in which DNA sequence comparison (13) has proven to be extremely useful in the absence of extensive genetic and functional studies. We therefore undertook the comparative analysis of the primary genomic structures of BPV-1 and HPV-1a presented in this manuscript to describe in further detail the salient features of the papillomavirus genomes. The genomes have been aligned through homologous conserved regions, and the overall genetic organization of the two viruses is quite similar. Open reading frames localized within the same regions of the genomes are related by various degrees of conserved sequences, indicating the likelihood that these coding regions represent bona fide genes or exons. Homologous noncoding sequences have been localized in corresponding regions of the genome, indicating that they may be involved in the regulation of papillomavirus gene expression.

This comparison suggests two means for papillomavirus genome evolution: a genetic drift by accumulation of point mutations and the acquisition of specific information by exchange of genetic material. This point, however, must be further evaluated by future comparison of DNA sequences of other papillomaviruses. Minor differences in the organization of the genome have been pointed out. The specific biological features of each of these two viruses, including host restriction, tissue tropism, oncogenicity, and

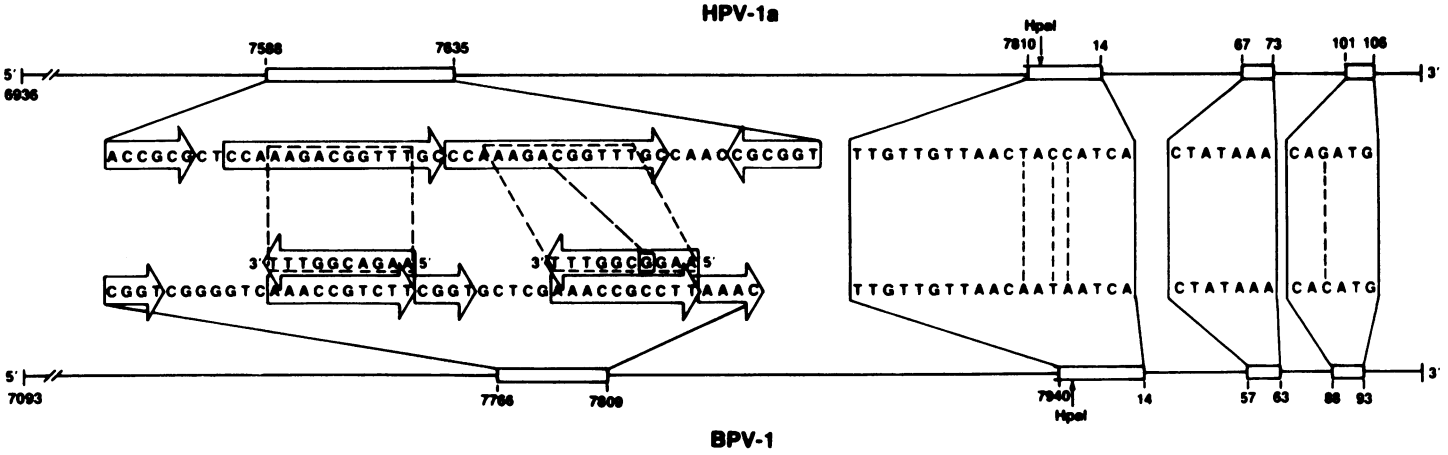

FIG. 5. Conserved sequences of the HPV-1a and BPV-1 noncoding regions. The 10-base repeat unit and partial repeat units in the BPV-1 sequence between position 7766 and 7809 as well as the related 15-base repeat unit and partial repeat unit in the HPV-1a genome between position 7668 and 7695 are enclosed in open arrows indicating the 5′ to 3′ direction of the DNA strand. The homologous regions surround the *Hpa*I site of each genome. The conserved TATAAA sequence and the possible cap site for the putative transcriptional unit are also indicated. Dashed lines joining two nucleotides indicate nonidentity of bases within homologous sequences.

transformation capability, have to be examined with respect to these unique features of their genetic composition.

ACKNOWLEDGMENTS

We thank J. V. Maizel, Jr., and J. Owens for valuable assistance in the enhanced graphic matrix analysis presented in Fig. 2. We thank J. V. Maizel, Jr., and W. T. McAllister for critical reviews of the manuscript. We are also grateful to S. Hostler for assistance in the preparation of this manuscript.

LITERATURE CITED

1. **Amtmann, E., and G. Sauer.** 1982. Bovine papillomavirus transcription: polyadenylated RNA species and assessment of the direction of transcription. J. Virol. **43**:59–66.
2. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
3. **Campo, S. M., M. H. Moar, H. M. Laird, and W. F. H. Jarrett.** 1981. Molecular heterogeneity and lesion specificity of cutaneous bovine papillomaviruses. Virology **113**:323–335.
4. **Chen, E. Y., P. M. Howley, A. D. Levinson, and P. H. Seeburg.** 1982. The primary structure and genetic organization of the bovine papillomavirus type 1 genome. Nature (London) **299**:529–534.
5. **Croissant, O., V. Testanière, and G. Orth.** 1982. Mise en évidence et localisation de régions conservées dans les génomes du papillomavirus humain 1a et du papillomavirus bovin 1 par analyse d'"hétéroduplex" au microscope électronique. C.R. Acad. Sci. **294**:581–586.
6. **Danos, O., M. Katinka, and M. Yaniv.** 1980. Molecular cloning refined physical map and heterogeneity of methylation sites of papillomavirus type 1a DNA. Eur. J. Biochem. **109**:457–461.
7. **Danos, O., M. Katinka, and M. Yaniv.** 1982. Human papillomavirus 1a DNA sequence: a novel type of genome organization among papovaviridae. EMBO (Eur. Mol. Biol. Organ.) J. **1**:231–236.
8. **Darnell, J. E.** 1982. Variety in the level of gene control in eukaryotic cells. Nature (London) **297**:365–371.
9. **Dvoretzky, I., R. Shober, S. K. Chattopadhayay, and D. R. Lowy.** 1980. A quantitative *in vitro* focus assay for bovine papillomavirus. Virology **103**:369–375.
10. **Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. van de Voorde, H. van Heuverswys, J. van Herreveghe, G. Volckaert, and M. Ysebaert.** 1978. Complete nucleotide sequence of SV40 DNA. Nature (London) **273**:113–120.
11. **Freese, U. K., P. Schulte, and H. Pfister.** 1982. Bovine papillomavirus-induced tumor contains a virus specific transcript. Virology **117**:257–261.
12. **Friedman, J. C., J. P. Levy, J. Lasneret, M. Thomas, M. Boiron, and J. Bernard.** 1963. Induction de fibromes souscutanes chez le hamster dore par inoculation d'extraits acellulaires de papillomes bovins. C.R. Acad. Sci. **257**:2328–2331.
13. **Galibert, F., T. S. Chen, and E. Mandart.** 1982. Nucleotide sequence of a cloned woodchuck hepatitis virus genome: comparison with the hepatitis B virus sequence. J. Virol. **41**:51–65.
14. **Ghosh, P. K., M. Piatak, V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1979. Transcription of the SV40 genome in virus transformed cells and early lytic infection. Cold Spring Harbor Symp. Quant. Biol. **44**:31–39.
15. **Heilman, C. A., L. A. Engel, D. R. Lowy, and P. M. Howley.** 1982. Virus specific transcription in bovine papillomavirus-transformed mouse cells. Virology **119**:22–34.
16. **Heilman, C. A., M. F. Law, M. A. Israel, and P. M. Howley.** 1980. Cloning of human papillomavirus genomic DNAs and analysis of homologous polynucleotide sequences. J. Virol. **36**:395–407.
17. **Howley, P. M., M.-F. Law, C. A. Heilman, L. W. Engel, M. C. Alonso, W. D. Lancaster, M. A. Israel, and D. R. Lowy.** 1980. Molecular characterization of papillomavirus genomes, p. 233–247. *In* M. Essex, G. Todaro, and H. zur Hausen (ed.), Viruses in naturally occurring cancers. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Jarrett, W. F. H., P. E. McNeil, H. M. Laird, B. W. O'Neil, J. Murphy, M. S. Campo, and M. H. Moar.** 1980. Papillomaviruses in benign and malignant tumors of cattle, p. 215–222. *In* M. Essex, G. Todaro, and H. zur Hausen (ed.), Viruses in naturally occurring cancers. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Jenson, A. B., J. R. Rosenthal, C. Olson, F. Pass, W. D. Lancaster, and K. Shah.** 1980. Immunological relatedness of papillomaviruses from different species. J. Natl. Cancer Inst. **64**:495–500.
20. **Lancaster, W. D., and C. Olsen.** 1982. Animal papillomaviruses. Microbiol. Rev. **46**:191–207.
21. **La Porta, R. F., and L. B. Taichman.** 1982. Human papilloma viral DNA replicates as a stable episome in cultured epidermal keratinocytes. Proc. Natl. Acad. Sci. U.S.A. **79**:3393–3397.
22. **Law, M. F., W. D. Lancaster, and P. M. Howley.** 1979. Conserved polynucleotide sequences among the genomes of papillomaviruses. J. Virol. **32**:199–207.
23. **Law, M.-F., D. R. Lowy, I. Dvoretzky, and P. M. Howley.** 1981. Mouse cells transformed by bovine papillomavirus contain only extrachromosomal viral DNA sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:2727–2731.
24. **Lowy, D. R., I. Dvoretzky, R. Shober, M.-F. Law, L. Engel, and P. M. Howley.** 1980. In vitro tumorigenic transformation by a defined subgenomic fragment of bovine papillomavirus DNA. Nature (London) **287**:72–74.
25. **Maizel, J. V., Jr., and R. P. Lenk.** 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:7665–7669.
26. **Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, and P. Chambon.** 1981. The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. **9**:6047–6060.
27. **Orth, G., F. Breitburd, and M. Favre.** 1978. Evidence of antigenic determinants shared by the structural polypeptides of (Shope) rabbit papillomavirus and human papillomavirus type 1. Virology **91**:243–255.
28. **Orth, G., M. Favre, F. Breitburd, O. Croissant, S. Jablonska, S. Obalek, M. Jarzabek-Chorzelska, and G. Rzesa.** 1980. Epidermodysplasia verruciformis: a model for the role of papillomavirus in human cancer, p. 259–282. *In* M. Essex, G. Todaro, and H. zur Hausen (ed.), Viruses in naturally occurring cancers. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Queen, C. L., and L. J. Korn.** 1979. Computer analysis of nucleic acids and proteins. Methods Enzymol. **65**:595–609.
30. **Tooze, J.** 1981. DNA tumor viruses, p. 799–936. *In* Molecular biology of tumor viruses, 2nd ed., part 2/ revised. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.