INTRODUCTION

Papillomaviruses (PVs) are small, non-enveloped viruses containing double-stranded, circular genomic DNA. To date, 120 types of human PV (HPV) and 69 non-human mammalian, avian and reptilian PVs have been isolated and characterized (de Villiers et al., 2004; Bernard et al., 2010). PVs infect cutaneous and mucosal epithelial cells and cause a variety of benign and malignant lesions (Chang et al., 1990; Jenson et al., 2001; Howley & Lowy, 2007). PVs seem to have existed for millions of years during which time they appear to have co-evolved with their natural hosts (Rector et al., 2007; Herbst et al., 2009).

Human PVs are known to induce almost all cervical cancers and at least 25% of head and neck cancers (Desai et al., 2009). Since 2006, efficacious virus-like-particle-based cervical-cancer vaccines have been available, but these are only effective in preventing initial infection by the specific HPV types included in the vaccine (Suzich et al., 1995; Ghim et al., 2002; Scarinci et al., 2010). The present vaccines contain virus-like particles (VLPs) of two to four different HPV types and can only prevent 70–75% of cervical cancer cases. The fact that PVs are species specific and cannot be replicated in vitro hindered not only initial development of HPV prophylactic approaches but also the investigations for therapeutic regimens. The lag time from infection with HPVs to the appearance of precancerous lesions can range from a few years to decades, which may result in there being new cervical-cancer patients for many decades to come.

There are numerous naturally infected large domestic animal species that have historically been used as models for HPV research, including cattle, horses, dogs and rabbits. Data obtained from the experiments using genetically engineered transgenic-mouse models and in vivo models using injected cell lines expressing HPV oncpro-
Mouse papillomavirus genomic analysis

Prior to the recently described clinical aspects of a laboratory mouse papillomavirus infection (MusPV; Ingle et al., 2010), seven rodent PVs were identified. Two, MnPV1 and McPV2, are from the African multimammate rats (Mastomys natalensis and Mastomys coucha, respectively) (Tan et al., 1994; Nafz et al., 2008; Bernard et al., 2010). Their prevalence among exotic rats may be high (Schäfer et al., 2010). Another rat PV (RnPV1) was isolated from the oral cavity of a healthy female free-ranging (field sample from Germany) Norway rat (Rattus norvegicus) (Schulz et al., 2009). Rodent-associated PVs have also been identified in Syrian golden hamsters (Mesocricetus auratus; HaOPV, recently renamed to MaPV1) (Iwasaki et al., 1997; Bernard et al., 2010), North American porcupines (Erethizon dorsatum; EdPV1) (Rector et al., 2005) and beavers (Castor fiber) (Schulz et al., 2009). Until recently, the only PV associated with any mouse species was isolated from a zoo colony of European harvest mice (Micromys minutus; MmPV, recently renamed MmiPV) (O’Banion et al., 1988; Sundberg et al., 1988; van Ranst et al., 1992; Van Doorslaer et al., 2007; Bernard et al., 2010). Although there are a number of known rodent PVs, none has been entirely satisfactory as they were isolated from wild or exotic rodent species.

We recently described the clinical features of a novel PV that naturally infected a colony of immunodeficient NMRI-Foxn1nu/Foxn1nu (nude) mice and was transmissible to immunocompetent SRV/Cri-ba/ba mice (Ingle et al., 2010). We report here the detailed genomic DNA characterization of this novel laboratory mouse PV, MusPV.

**RESULTS**

MusPV was isolated from laboratory mice. MusPV infection was very productive in nude mice as evidenced by the large number of the nuclei in the stratum granulosum that were highly positive for the presence of PV genus-specific antigens by immunohistochemistry (Ingle et al., 2010). Transmission electron microscopy of papillomas removed from the initial cases in nude mice revealed abundant intranuclear viral particles (Supplementary Fig. S1, available in JGV Online).

**Cloning MusPV genomic DNA**

Digestion of rolling circle amplification (RCA) product with XbaI, HindIII, SacI or Ncol resulted in distinctive multiple DNA bands totalling approximately 7.5 kb (not shown). Fig. 1(a) shows SacI digests of the rolling circle product. The full-size DNA fragment generated by XbaI digestion was cloned into pUC19 vector. Since the XbaI restriction site was within the L1 gene, resulting in the fragmentation of the L1 gene in the original clone, a set of primers annealing specifically upstream and downstream of the L1 gene was designed, and the L1 gene was cloned into a pCR-XL TOPO vector by using PCR cloning technology (Fig. 1b).

**Sequence analysis of MusPV**

The complete MusPV genomic sequence was 7510 nt with a 46.6 mol% G+C content. This DNA sequence was deposited in GenBank (accession no. GU808564). BLASTN analysis showed no close matches with any other PVs, indicating that MusPV is a novel PV. According to MEGALIGN (DNASTAR) analysis, the MusPV genome shares less than 70% DNA sequence homology with other PVs. The full-length genomic DNA of MusPV shared between 56.2 and 66.3% similarity with those of π-genus rodent PVs (MaPV1, MmiPV, McPV2 and RnPVI) and only 42.9 and 45.3% similarity with MnPV1 (i genus) and EdPV1 (σ genus), respectively. de Villiers et al. (2004) reported that ‘Different genera share less than 60% nucleotide sequence identity in the L1 ORF. Full-length sequences of complete genomes have more than 23%, but less than 43% nucleotide sequence identity when comparing genera of the family *Papillomaviridae*).

As indicated in Table 1, BLASTX comparison of L1 proteins revealed 80, 79 and 78% similarity to the L1 proteins of RnPV1, McPV2 and MmiPV, respectively, but less than 63% with the L1 proteins of MnPV1 and EdPV1. From the
pairwise sequence analyses, higher similarities were also found in the L1 and E1 ORFs with lowest homology in E6 and E7 ORFs.

Analysis of the phylogenetic tree of rodent PVs constructed via alignment of the nucleotide sequences of concatenated E1/E2/L1/L2 sequences revealed that MusPV is most closely related to McPV2 and is genetically distant from EdPV1 and MnPV1 (Fig. 2). These support classification of MusPV within the p-PV genus according to the currently accepted PV classification system (de Villiers et al., 2004).

Genomic organization of MusPV

There are seven ORFs on the sense strand of MusPV genomic DNA. Like other PVs, MusPV genomic DNA can be divided into early (E), late (L) and non-coding (NCR) regions (Fig. 3). The E region comprises 49% of the genomic DNA with five ORFs encoding E6, E7, E1, E2 and E4 proteins. An E5 ORF was not identified. The L region comprises 42% of the genomic DNA with 2 ORFs encoding structural proteins L1 and L2. A possible ORF without a start codon that might encode a proline-rich E4-like protein was also identified within E2. The NCR located between the L1 and E6 genes contained 609 bp and multiple binding sites for transcription factors (TFs) (Supplementary Fig. S2, available in JGV Online). The nucleotide position and size of the ORFs and NCR as well as the predicted molecular mass of putative proteins are detailed in Table 2. The start of the E6 ORF was assigned position 1.

Putative binding-site analysis

Both E6 and E7 proteins of most PVs bind to cellular proteins and are important in regulating the cell cycle (Howley & Lowy, 2007). Similar to many other PVs, the E6 protein of MusPV possesses two zinc-binding domains (CX2C–X29–CX2C) starting at aa positions 26 and 99. The E7 protein has one zinc-binding domain starting at aa 62. Unlike most PVs, which have a pRb-binding domain on their E7 proteins, a putative domain for pRb [retinoblastoma protein (RB) in case of murine genetics] binding was identified on the E6 protein (aa 66–70) of MusPV. E1 and E2 proteins of PVs are mainly involved in replication and transcription. In MusPV, E1 contains one ATP-binding site for the ATP-dependent helicase (GPPDTGKS) and E2 contains a DNA-binding α-helix (GPTNSLKCWRNR). Both L1 and L2 proteins of MusPV have putative

<table>
<thead>
<tr>
<th>Host</th>
<th>PVs</th>
<th>Genus</th>
<th>Genome</th>
<th>E6</th>
<th>E7</th>
<th>E1</th>
<th>E2</th>
<th>L2</th>
<th>L1</th>
<th>NCR</th>
<th>Lesion type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M. auratus</strong> (Syrian golden hamster)</td>
<td>MaPV1 (E15111)</td>
<td>π</td>
<td>56.2</td>
<td>50.0</td>
<td>33.3</td>
<td>65.1</td>
<td>52.4</td>
<td>56.2</td>
<td>67.0</td>
<td>45.5</td>
<td>Oral papilloma</td>
</tr>
<tr>
<td><strong>M. minutus</strong> (European harvest mouse)</td>
<td>MmPV (NC_008582)</td>
<td>π</td>
<td>62.7</td>
<td>57.1</td>
<td>63.5</td>
<td>64.9</td>
<td>55.6</td>
<td>60.8</td>
<td>70.2</td>
<td>52.8</td>
<td>Cutaneous papilloma</td>
</tr>
<tr>
<td><strong>M. coucha</strong> (Multimammate rat)</td>
<td>McPV2 (DQ664501)</td>
<td>σ</td>
<td>66.3</td>
<td>61.2</td>
<td>62.7</td>
<td>71.8</td>
<td>61.3</td>
<td>64.3</td>
<td>72.1</td>
<td>51.8</td>
<td>Anogenital warts and others</td>
</tr>
<tr>
<td><strong>R. norvegicus</strong> (Norway rat)</td>
<td>RnPV1 (NC_013196)</td>
<td>τ</td>
<td>58.5</td>
<td>56.3</td>
<td>54.7</td>
<td>66.4</td>
<td>58.2</td>
<td>62.3</td>
<td>68.7</td>
<td>50.1</td>
<td>Healthy oral cells</td>
</tr>
<tr>
<td><strong>E. dorsatum</strong> (North American porcupine)</td>
<td>EdPV1 (NC_006951)</td>
<td>σ</td>
<td>45.3</td>
<td>32.1</td>
<td>29.5</td>
<td>48.7</td>
<td>44.8</td>
<td>40.4</td>
<td>54.1</td>
<td>38.2</td>
<td>Cutaneous papilloma</td>
</tr>
<tr>
<td><strong>M. natalensis</strong> (Multimammate rat)</td>
<td>MnPV1 (NC_001605)</td>
<td>τ</td>
<td>42.9</td>
<td>37.5</td>
<td>40.2</td>
<td>51.9</td>
<td>30.1</td>
<td>42.8</td>
<td>55.9</td>
<td>39.7</td>
<td>Cutaneous papilloma</td>
</tr>
</tbody>
</table>

**Table 1.** MusPV DNA sequence similarities to other rodent PVs

The genomic DNA and ORF sequences of MusPV were compared with those of six rodent PVs collected from GenBank. The results of amino acid sequence comparisons are indicated in parentheses (MEGALIGN program; DNASTAR).

**Fig. 2.** Neighbour-joining phylogenetic tree of rodent PVs based on concatenated E1/E2/L1/L2 ORFs. The PV genera are indicated with their Greek character according to the criteria set forth by de Villiers et al. (2004).
nuclear-localization signals at their C termini (RKKRRK and KRK, respectively). A putative polyadenylation signal for early viral-mRNA transcription is located on the L2 ORF at nt 3844. A putative polyadenylation signal for late genes was located within the NCR at nt 7046.

Potential binding-site analysis of MusPV NCR revealed the relative locations of several important motifs, including binding sites for the cellular TFs (NF1, AP1, OCT1, YY1, GATA1 and SP1), typical palindromic E2 binding sites (E2BSs) and p53 (TRP53 in murine genetics). Polyadenylation and TATA sequences were also identified in NCR of MusPV (Supplementary Fig. S2). Six TF binding sites were common to all rodent PV. At least 13 TF binding sites were shared among those belonging to the π genus, but they were variable among the λ and π genera. The NCR of MusPV had one putative binding site for TRP53.

Table 2. Positions of the ORFs and NCRs of MusPV and the predicted molecular masses of the translated proteins

<table>
<thead>
<tr>
<th>ORF</th>
<th>Genome position (nt)</th>
<th>Length nt</th>
<th>Molecular mass (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start</td>
<td>Stop</td>
<td></td>
</tr>
<tr>
<td>E6</td>
<td>1</td>
<td>423</td>
<td>423</td>
</tr>
<tr>
<td>E7</td>
<td>423</td>
<td>755</td>
<td>333</td>
</tr>
<tr>
<td>E1</td>
<td>742</td>
<td>2604</td>
<td>1863</td>
</tr>
<tr>
<td>E2</td>
<td>2534</td>
<td>3685</td>
<td>1152</td>
</tr>
<tr>
<td>E4</td>
<td>3101*</td>
<td>3419</td>
<td>318</td>
</tr>
<tr>
<td>L2</td>
<td>3735</td>
<td>5361</td>
<td>1617</td>
</tr>
<tr>
<td>L1</td>
<td>5291</td>
<td>6901</td>
<td>1611</td>
</tr>
<tr>
<td>NCR</td>
<td>6902</td>
<td>7510</td>
<td>609</td>
</tr>
</tbody>
</table>

*E4 does not have an identifiable start codon in the ORF.

DISCUSSION

MusPV, the first PV found to naturally infect laboratory mice (NMRI-Foxn1nu/ Foxn1 nu, nude), induced florid papillomatosis at the mucocutaneous junctions of the nose and mouth. MusPV could be transferred to immunocompetent mice using a cell-free extract from papillomas (Ingle et al., 2010). Viral particles were identified in the nucleus of keratinocytes in the stratum granulosum of the papillomas by electron microscopy (Supplementary Fig. S1). This novel mouse PV was closely related to several other rodent PVs within the π genus, and should be classified as a member of the π genus according to the rules applied to the classification of PVs (de Villiers et al., 2004). MusPV shares similar genomic organization to all other PVs. It shares the locations of binding sites for the critical cell-regulatory factors, such as RB and TFs, with other PVs. The fact that homologies among the 5 PVs in the π genus (MaPV1, MmiPV, McPV2, RnPV1 and MusPV) were higher than those between PVs in π genus and rodent PVs in other genera, supports recent suggestions of unique co-divergency of rodent PVs in the π genus (Schulz et al., 2009).

The NCR that is located between the E and L regions of PVs contains essential sequences necessary for replication and transcription, including the origin of replication, binding sites for E1 and E2 proteins and various TFs (Zheng & Baker, 2006; Howley & Lowy, 2007). The NCR of MusPV maintained the sites for most of these elements and also shared their locations with the other PVs in the π genus (Supplementary Fig. S2). For instance, four of five PVs in the π genus (MusPV, RnPV1, McPV2 and MmiPV) shared at least 18 common TF sites, while MaPV1 shared only 13 common TF-binding sites with other π-genus PVs. Two other rodent PVs, MnPV1 (λ genus) and EdPV1 (σ genus), displayed dissimilar patterns for TF binding sites compared with those of MusPV. Only six TF binding sites (two E2BSs, two NF1, one SP1 and one GATA1) were common to all rodent PVs.

There were additional potential, putative TF sites in NCRs of each rodent PV (Supplementary Fig. S2). NCRs of most PVs have the TRP53 (formally p53) binding site for its transcriptional functions. The NCR of MusPV has one. Surprisingly, the NCR of McPV2 has no binding sites for TRP53 while MaPV1 has four. MusPV also differs from most of the other rodent PVs in regard to the presence of a putative E4 ORF and its putative RB binding site. With the exception of RnPV1, other rodent PVs do not have typical E4 ORFs. MusPV has an E4 ORF, without a start codon, within the E2, which may encode a proline-rich protein. While PV E7 proteins may be oncogenic in some PV types via binding with tumour-suppressor proteins such as the RB (formally pRb) family of proteins (Pim & Banks, 2010), the E6 and E7 proteins might also be involved in several other functions such as modulating innate immune responses and downregulation of major histocompatibility loci (Ghittoni et al., 2010). In the case of MusPV, one putative RB-binding site was found on the E6 protein but not on the E7 protein. It is interesting to mention that of all currently known rodent PVs only MnPV1 has an RB-binding site on the E7 protein (Tan et al., 1994). RB-
binding sites on the E6 protein were also observed in *Phocoena spinipinnis papillomavirus* type 1 (PsPV1) and bottlenose dolphin PV types 1, 2 and 3 (TrPV1–3), which do not have identifiable E7 ORFs. These PVs have longer E6 regions than any other PVs and lack E7 ORFs, but have their putative RB-binding sites on E6 and E1, respectively (Rehtanz *et al.*, 2006; Van Bressem *et al.*, 2007; Rector *et al.*, 2008). The full effect of the differences in the number of TF-binding sites, as well as TRP53-binding sites on NCRs, the presence and absence of potential E4 ORFs and the different locations of the RB-binding sites of these rodent PVs, is not yet known.

Rodent PVs, other than MusPV, are of limited value as models for HPV infection of animals because they are, for the most part, from exotic or wild species, not standard laboratory animals. We demonstrated that immunocompetent laboratory mice inoculated with a cell-free homogenate of papillomas from the nude mice developed clinically evident papillomas at the sites of inoculation after 5 weeks (Ingle *et al.*, 2010). Some rodent hosts, such as European harvest mice (MmiPV), North American porcupine (EdPV1) or beavers are not tractable laboratory animals for experimental purposes. The two potential laboratory–rodent PVs, MaPV1 and RnPV1, failed to produce transmissible, clinically apparent lesions in presumably naïve recipients. For instance, MaPV1 was isolated from a papillomatous lesion on the tongue of Syrian hamsters, which were co-treated with abrasion and a chemical carcinogen, 7, 12- dimethyl benzanthracene, known to cause papillomas without a viral association. MaPV1 could also be easily identified in DNA extracted from normal skin and mucocutaneous tissues of healthy hamsters by PCR (Iwasaki *et al.*, 1997). The laboratory–rat PV (RnPv1) was identified only from clinically normal tissue (Schulz *et al.*, 2009). McPV2 was isolated from anogenital wart-like lesions, which appeared on rats at approximately 8 months of age. McPV2 was detected by immunohistochemistry in clinically normal tissues (Nafz *et al.*, 2008).

Although MusPV could potentially become a very useful model to study PV pathogenesis, from which data could be translated to help us to better understand HPV diseases, it would be prudent to bear in mind that the genomic homogeneity between MusPV and HPV is low. When compared to HPV16 in the α genus, which comprises most of the oncogenic mucosotrophic PVs, the genome homology between MusPV and HPV16 was 49.8%. For E1 and L1 proteins, which usually have the highest degree of homogeneity among all PVs, the degrees of similarity were 53.7 and 58.1%, respectively. For E6 and E7 oncoproteins, they were 44.7 and 39.9%, respectively. For the NCR, their homology was 45.6%. While this appears to be a deterrent, these differences might reflect differences in host specificity such that biological activity, currently under investigation, might be quite relevant.

In this study, the genomic DNA of MusPV was cloned, analysed and compared with PVs isolated from other rodent species. This information will be helpful not only in understanding papillomavirus infection and pathogenesis of this novel mouse PV in general, but also in developing a more readily available model for the study of HPV.

**METHODS**

Transmission electron microscopy (EM). Previously formalin-fixed and embedded samples were obtained from the Advanced Centre for Treatment Research and Education in Cancer (ACTREC) in India. EM blocks were incubated in propylene oxide (Electron Microscopy Science) for 1 h to eliminate excess plastic. Samples were collected, reinfiltated in a mixture of LX1112 (Ladd Research Industries) and propylene oxide (1:1) for 1 h, twice in LX112 for 1 h each, and subsequently embedded in LX1112 overnight. Sections were cut on an LKB ultramicrotome at 80 Å, mounted on 200-mesh copper grids, stained and viewed under a Phillips CM-12 transmission electron microscope operating at 60 kV.

Cloning MusPV. Naturally occurring papillomas were collected at the time of necropsy from NMRI- Foxn1nu/Foxn1nu (nude) mice housed at ACTREC (Ingle *et al.*, 2010). Samples were snap frozen in liquid nitrogen and stored at −80 °C. Parts of frozen papillomas were minced finely with a scalpel and total DNA was extracted as previously described (Rector *et al.*, 2004b) with the exception that the DNA samples were precipitated with 2-propanol. The extracted DNA, suspended in TE buffer (10 mM Tris and 10 mM EDTA), was shipped to the University of Louisville.

To increase the number of viral-DNA copies, RCA was performed on a tenfold dilution of this DNA extract as previously described (Rector *et al.*, 2004a; Joh *et al.*, 2009). Two-microlitre aliquots of amplified DNA were incubated with restriction enzymes, and the digestion patterns were viewed on a 0.8% agarose gel. A single DNA fragment that resulted from the digestion with *Xba*I was cloned into pUC19 using standard molecular biological techniques. Briefly, 10 μl of the RCA product was digested with *Xba*I. Digested DNA fragments were gel purified (QIAquick Gel Extraction kit), and cloned into pUC19. Ligation of DNA samples were done using a Rapid DNA ligation kit (Roche). The resulting clones were initially sequenced with reverse and forward M13 primers to confirm they contained PV DNA. Additionally, since *Xba*I cut the L1 gene, the full-length L1 gene was amplified by PCR from RCA DNA and cloned into pCR-XL TOPO vector (Invitrogen).

DNA sequencing. To facilitate the sequencing procedure of the MusPV genomic DNA, an EZ::TN5, KAN-2 Insertion kit (Epitcience Biotechnologies) was utilized (Joh *et al.*, 2009). DNA sequencing of the colonies obtained was done at the nucleic acids core facility at the University of Louisville using primers furnished by the manufacturer of the kit (KAN-2 FP1 and RP1). Assembly of the full-length PV genomic sequence and its alignment with other PV DNAs was performed with the SeqMan II program (DNASTAR). Confirmations for uncertain or missing sequences were completed using primers designed to be specific to MusPV.

Sequence similarity and phylogenetic analysis of MusPV. The MusPV genomic-DNA sequence was surveyed using the BLASTN and the BLASTX programs in GenBank and its ORFs identified using the EditSeq program (DNASTAR). The molecular mass (kDa) of each putative viral protein encoded was calculated using the Mapdraw program (DNASTAR). Pairwise sequence analyses of MusPV with other rodent PVs were performed by comparative analysis. The nucleotide and amino acid sequences of the other rodent PVs were downloaded from GenBank, and the similarities of each genome, the ORFs and the
NCR were analysed by using the clustal w setting in the MEGALIGN program (DNASTAR). A neighbour-joining phylogenetic tree was constructed based on the comparison of concatenated E1, E2, L1 and L2 of rodent PVs using the MEGALIGN program (DNASTAR). (Joh et al., 2009; Van Doorslaer et al., 2007; Schulz et al., 2009).

Analyses of binding domains. To find putative binding motifs for TFs, the NCR sequences of MusPV as well as six existing rodent PVs were analysed with the Transcription Element Searching System (TESS; http://www.cbil.upenn.edu/cgi-bin/tess/tess) (Schug & Overton, 1997). Putative binding domains on MusPV-encoded proteins were also searched for by comparing amino acid sequences of MusPV proteins with binding domains known to be present on the proteins of other PVs.

ACKNOWLEDGEMENTS

We thank Irine Chepkoech for her assistance in electron microscopy and manuscript review. This work was supported in part by the University of Louisville Clinical and Translational Science Pilot Grant Program.

REFERENCES


