CHAPTER THREE:
PHYLOGENY OF PAPILLOMAVIRUSES OBTAINED IN SKIN LESIONS OF CAPE MOUNTAIN ZEBRA (*Equus zebra zebra*, Linnaeus, 1758)

ABSTRACT

After clinical, histological and molecular diagnosis of sarcoid in free roaming Cape mountain zebra (*Equus zebra zebra*), phylogenetic analysis of the BPV DNA in the E2E5 region demonstrated that sequences corresponded to variants of either BPV-1 or BPV-2. Sequences obtained from lesions in Equidae cluster together as sister taxa to the rest of BPV-1 sequences obtained from Bovidae. The phylogenetic position of BPV-1 and BPV-2 is well nested within a supertaxon containing papillomaviruses infecting Artiodactyla and Equidae. The evolutionary analysis estimate that the most recent common ancestor for all BPV-1 dates back to 1.4 million years (Mya) and in BPV-2 0.55 Mya. While the the last common ancestor for BPV-1 and BPV-2 dates back to 5.34 Mya.

Keywords: Papillomavirus, phylogeny, E2E5, Cape mountain zebra (*Equus zebra*), South Africa
3.1 INTRODUCTION

Papillomaviruses are classified in the family *Papillomaviridae* (Bernard, 2006) and cause disease in humans and numerous animal species by stimulating epithelial proliferation. Papillomaviruses are quantitatively the most important group of viruses associated with benign and malignant neoplasia in humans (Zur Hausen 1991). In animals papillomavirus association has been documented in various species: equids (Angelos *et al*. 1991; Bloch *et al*. 1994; Reid *et al*. 1994; Chambers *et al*. 2003; Löhr *et al*. 2005), felidae (Sundberg *et al*. 2000; Munday *et al*. 2008) camelids (Schulman *et al*. 2003); possums (Perrott *et al*. 2000); giraffes and impalas (Karstad & Kaminjolo 1978), and have been best studied in horses (Chambers *et al*. 2003) and cattle (Borzacchiello & Roperto 2008; Nasir & Campo, 2008). Animal papillomavirus have been studied as agents of disease in animals and as models of human papilloma infection (Campo 2002; Breitburd *et al*. 2006).

The papillomavirus genome is a single molecule of double-stranded, circular DNA and contains approximately 7 900 bp. Eight well defined open reading frames are encoded: L1 and L2 that encode capsid proteins and E1, E2, E4, E5, E6 and E7 that encode proteins involved in replication, transcription and transformation, and a non-coding, regulatory, long control region (LCR) (Bernard *et al*. 2006; Bravo & Alonso 2007). In humans, more than 100 different papillomavirus types, associated with both benign and malignant conditions, have been fully sequenced (De Villiers *et al*. 2004).

Papillomaviruses are thought to be highly species-specific. This hypothesis is, however, frequently rejected as papillomaviruses infecting the same host are not monophyletic, and many distantly related papillomaviruses infect the same host species (Bravo *et al*. 2010). The only known exception being naturally occurring cross-species infections of horses, donkeys and mules with BPV-1 and BPV-2 (Nasir *et al*. 2007) in which they are associated with sarcoids (Lancaster & Olson 1980; Nasir & Reid 1999; Campo 2002). Recently BPV was described in other examples of cross species infection when it was described in the water buffalo (*Bulbalus bulbalus*) (Silvestre *et al*. 2009) and in a European bison (*Bison bonasus*) (Literák *et al*. 2006). Kidney & Berrocal (2008) described sarcoids in tapirs (*Tapirus bairdii*) which histologically resembled equine sarcoids and BPV-1 was demonstrated. Munday *et al*. (2007) described a growth in a domestic cat (*Felis catus*) of which sequencing of the amplicon revealed 98% similarity to human papillomavirus type 9. In a study by Zaugg *et al*. (2005) papillomaviruses found in dogs (*Canis lupis familiaris*) were closely related to human papillomavirus type 5. Papillomaviruses are highly diverse and are likely occur in most mammals and birds (de Villiers *et al*. 2004). At least 50 mammalian species have been
confirmed as being infected by species-specific papillomaviruses (Sundberg et al. 2001). Papillomaviruses appear to be widespread and have been found in a large number of vertebrate species and have been historically assumed to have evolved alongside their hosts (Bernard 1994, Antonsun & McMillan 2006). Phylogenetic studies also suggest that they do not change host species, do not recombine, and have maintained their basic genomic organization for more than 100 million years (Bernard 2006). It has, however, been shown that phylogenetic trees of the viruses and their hosts are not congruent and that this hypothesis is questionable (Bravo et al. 2010).

The purpose of this investigation was to study the phylogenetic relationships of the papillomaviruses DNA obtained from tumours in the Cape mountain zebra (Equus zebra zebra), in South Africa.

3.2 MATERIALS AND METHODS

3.2.1 Study population and sample collection

Specimens of sarcoid tumours included in this study were obtained from Cape mountain zebras (Equus zebra zebra) from the Gariep Dam Nature Reserve (Free State Province) (n=9), Bontebok National Park (Western Cape Province) (n=2) and Mountain Zebra National Park (Eastern Cape Province) (n=1). Additionally, blood samples were collected from healthy, sarcoid-unaffected zebras (n=51) in these parks as well as from parks where sarcoïds had never been observed.

The sarcoid tumour samples were collected by means of a biopsy punch. Each comprised a 10 mm x 3 mm x 3 mm piece of tissue which was stored in a glass tube at 4 °C until further analysis. Blood samples were collected in EDTA-buffered tubes by venipuncture of the jugular vein and were also stored at 4 °C.

Histopathologically confirmed equine sarcoid (n=1) and bovine papilloma (n=1) specimens were obtained from private veterinarians and were included in the study to serve as positive control material.

3.2.2 DNA extraction

DNA was extracted from 200 µl of blood or 25 mg of tissue (sarcoid) using the QIAamp®DNA extraction kit (Southern Cross Biotechnologies) according to the manufacturer’s instructions. Extracted DNA was eluted in 100 µl elution buffer and stored at 4 °C until further analysis.
3.2.3 Conventional PCR amplification of a region of the E5 open reading frame

Conventional PCR was used for the amplification of 637 bp DNA fragment of a region of the E5 ORF of both BPV-1 and BPV-2 as described in Chapter 2.

3.2.3.1 Real-time PCR

The hybridization probe real-time PCR assay (as described in Chapter 2) developed for the detection and differentiation of BPV-1 and BPV-2 DNA in specimens was used to analyse the samples.

3.2.3.1.1 Primer and hybridization probe design

The conventional PCR primer set, the hybridization probes, PCR amplification mixture and reaction condition used were as described in Chapter 2. Specimens from histopathologically confirmed equine sarcoid from a horse (n=1) and a bovine papilloma (n=1) were included as positive controls.

3.2.3.2 Molecular cloning and sequencing

Conventional PCR was used for the amplification of 637 bp DNA fragment of a region of the E5 ORF of both BPV-1 and BPV-2. The primers (F_3610: 5’-GCT AAC CAG GTA AAG TGC TAT C-3’; R_4247: 5’-TGC TTG CAT GTC CTG TAC AGG T-3’), PCR amplification mixture and reaction conditions were as described in Chapter 2. The obtained amplicons were purified, cloned into the pGEM®-T vector (Promega pGEM–T Easy Vector System, Promega, Madison, USA), and transformed into competent JM 109 *E. coli* cells (Promega, Madison, USA) as described in Chapter 2. Recombinant plasmid DNA was isolated, directly sequenced using the ABI BigDye™ Terminator Cycle Sequencing Ready Reaction kit (PE Applied Biosystems) and analyzed on an ABI 3100 sequencer at Inqaba Biotec (Pretoria, South Africa). Sequences were assembled and edited using the GAP4 program of the Staden package (version 1.6.0 for Windows) and aligned with published sequences of related genera using ClustalX (version 1.81 for Windows) (Staden *et al.* 2000).

3.2.3.3 Phylogenetic analysis

The phylogenetic position of BPV-1 and -2 and the relationships with its close relatives were reconstructed based on a concatenated alignment of the E1E2L1 genes, as described by Gottschling
et al. (2007) and Bravo et al. (2010). Briefly, the sequences were aligned at the amino acid level with MUSCLE (http://www.drive5.com/muscle/) (Edgar 2004), filtered for ambiguously aligned positions with GBLOCKS (http://molevol.cmima.csic.es/castresana/Gblocks.html) (Castresana 2000) visualized for manual correction and back-translated to the nucleotide level using PAL2NAL (http://www.bork.embl.de/pal2nal) (Suyama et al. 2006). Maximum Likelihood (ML) phylogenetic analysis was performed with RAxML v7.2.5 (http://wwwkramer.in.tum.de/exelixis/software.html) (Stamatakis 2006a) using the GTR+ model of evolution and the CAT approximation of rate heterogeneity (Stamatakis 2006b), which introduced nine partitions that corresponded to each of the codon positions of each of the three genes, and with 10 000 non-parametric bootstraps.

Fine relationships within this superclade, to which BPV belongs, were further reconstructed with the same procedure, while rerunning the GBLOCKS-filtering step using only members of the superclade and including RaPV1 as an outgroup. A dated Bayesian phylogenetic analysis was performed with BEAST v1.4.8 (http://beast.bio.ed.ac.uk) (Drummond & Rambaut 2007) with the GTR+ model of evolution, using an uncorrelated log normal relaxed clock, which introduced three partitions that corresponded to each of the codon positions, and unlinking parameters across codon positions. A uniform prior for the age of the most recent common ancestor of the superclade was fixed between 95.3 and 113 millions of years (Mya), based on the estimated dating for divergence time for the crown of Laurasiatheria (Bininda-Edmonds et al. 2007). Two independent chains of 50 million steps were calculated and analyzed with a burn-in of ten million steps and combined after a Bayes factor analysis on likelihood values for consistency (Suchard et al. 2001). The mutation rate was estimated to have an average of 0.0125 mutations per site per Mya (95% posterior density between 0.0110 and 0.0141). Values for average relative contributions to mutation rate were 0.49, 0.31 and 2.21, for each codon position respectively. The age for the most recent common ancestor for BPV-1 and BPV-2 was estimated to be 5.34 Mya (95% posterior density between 3.17 and 7.62), and this estimation was used in subsequent analyses.

A final phylogenetic analysis focusing on the genomic region analysed here and amplification was performed, including the C termini of the E2 gene, the E5a gene and the E5b gene (51, 56 and 69 parsimony informative positions, respectively). A dated Bayesian phylogenetic analysis was performed with BEAST v1.4.8 (http://beast.bio.ed.ac.uk) (Drummond & Rambaut 2007) with the GTR+ model of evolution, for both strict clock and uncorrelated log normal relaxed clock, introducing three partitions that corresponded to each of the codon positions, and unlinking parameters across codon positions. A normal prior for the age of the most recent common ancestor of BPV-1 and BPV-2 superclade was fixed with average 5.3 Mya and standard deviation 1.0 Mya.
There were no differences in the likelihood values for the trees obtained under strict clock and under relaxed clock assumptions, and the simpler, strict clock model was preferred. Mutation rate was estimated to have an average of 0.017 mutations per site per Mya (95% posterior density between 0.009 and 0.026). Values for average relative contributions to mutation rate were 0.996, 0.613 and 1.391, for each codon position respectively. Evolutionary relationships of the papillomaviruses were run and the hierachial taxonomonic relationships were constructed at Genomics and Health, Centre for Public Health Research, Valencia, Spain.

3.3 RESULTS

3.3.1 Real-time PCR

The real-time PCR results are shown in Table 2 (Chapter 2). DNA obtained from all sarcoid tissue specimens tested positive for the presence of BPV DNA. For the blood samples collected from zebra from sarcoid-affected parks, 52% was positive for the presence of BPV DNA. Surprisingly, of the blood samples collected from zebras from sarcoid-unaffected parks, 41% were positive.

3.3.2 Sequencing results

A 637 bp amplicon of a region of the BPV E5 ORF could be amplified from all the zebra sarcoid tumour samples (n=12) using conventional PCR. Unfortunately, no amplification product could be obtained from blood samples using conventional PCR. The PCR amplicons obtained from the sarcoid tissue were cloned and sequenced. The obtained sequences were deposited in GenBank (accession numbers HQ541333-HQ541354). The number of base differences per sequence from analysis between sequences were conducted in MEGA4 (Tamura et al. 2007). All results are based on the pairwise analysis of 11 sequences. There were a total of 456 positions in the final dataset.

3.3.3 Phylogenetic reconstruction

In a broad selection of papillomaviruses compiled in a phylogenetic tree (Figure 10) the phylogenetic position of BPV-1 and BPV-2 is well nested within the D superfamily (in green) (Gottschling et al. 2007; Bravo & Alonso 2007) which contains different papillomaviruses infecting the dog (CsPV3 and CPV4), horse (EcPV1 and EcPV4), cow (BPV-1, BPV-2, BPV-5 and BPV-8), sheep (OvPV1 and OvPV2), European elk (EEPV), roe deer (CcPV), deer (DPV) and white-tailed deer (RPV). This tree was constructed using the concatenated E1-E2-L1 genes, aligned at the amino acid level with back-translated into nucleotides, and calculated with RAxML. Support after
1 000 bootstrap replicates given on each nodes. Color codes highlight the four papillomavirus supertaxa. The tree was rooted with sequences from birds and turtles in grey. The scale bar is measured in substitutions per site. Bovine papillomavirus-1 and BPV-2 belong to the supertaxon encompassing delta, epsilon and zeta papillomaviruses, infecting different species within Laurasiatheria. This is consistent with previous descriptions; BPV-1 belongs unambiguously to a large clade encompassing members of delta, epsilon, zeta, dyoeta and chi-papillomaviruses (Chan et al. 1995; Bravo & Alonso 2007; Gottschling et al. 2007; Bravo et al. 2010; Vanderstraeten et al. 2010). The zebra sequences used in this study showed the highest similarity with BPV-1 and BPV-2 and their positions are indicated with an asterix in Figure 10.
*Indicates the position of the zebra samples from this study

**Figure 10** Best known maximum likelihood phylogenetic tree for papillomaviruses.
Figure 11  Bayesian phylogenetic reconstruction of the subfamily D (green section in the previous figure) for sequences grouping together with BPV-1 and BPV-2 in the global papillomavirus tree.

A Bayesian phylogenetic reconstruction of subfamily D papillomavirus tree demonstrates the divergence time estimates for this subfamily D (the green section in the previous figure) in Figure 11. This tree has been rooted with the bat papillomavirus (RaPV) sequence (in grey) as an outgroup showing divergence time of 95.3-113 million years based on estimated dating for the divergence time for the crown of the superorder Laurasiatheria (Bininda-Edmonds et al. 2007). The position of the zebra sequences in this study, with BPV-1 and -2, is indicated by an asterix. An explanatory value for the abbreviations used in the tree is given in Table 3. The age for the most recent common ancestor for BPV-1 and BPV-2 was estimated to be 5.34 Mya (95% posterior density between 3.17 and 7.62).
<table>
<thead>
<tr>
<th>Abreviation</th>
<th>Genus papillomavirus</th>
<th>Host species</th>
<th>Family</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>RaPV</td>
<td>Alpha papillomavirus</td>
<td>Bat papillomavirus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CsPV03</td>
<td>Omikron papillomavirus</td>
<td>Dog papillomavirus</td>
<td>Canidae</td>
<td>Carnivora</td>
</tr>
<tr>
<td>CPV04</td>
<td>Alpha papillomavirus</td>
<td>Dog papillomavirus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcPV02</td>
<td>Alpha papillomavirus</td>
<td>Horse papillomavirus</td>
<td>Equidae</td>
<td>Perissodactyla</td>
</tr>
<tr>
<td>Z1EcPV</td>
<td>Zeta papillomavirus</td>
<td>Horse papillomavirus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eBPV05: eBPV08</td>
<td>Epsilon papillomavirus</td>
<td>Cow papillomavirus</td>
<td>Bovidae</td>
<td></td>
</tr>
<tr>
<td>d4BPV02: d4BPV01</td>
<td>Delta papillomavirus</td>
<td>Cow papillomavirus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d3OvPV02: d3BPV01</td>
<td>Delta papillomavirus</td>
<td>Sheep papillomavirus</td>
<td>Ovidae</td>
<td></td>
</tr>
<tr>
<td>CcPV01</td>
<td>Zeta papillomavirus</td>
<td>Roe deer papillomavirus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d1EEPV</td>
<td>Delta papillomavirus</td>
<td>European elk papillomavirus</td>
<td>Cervidae</td>
<td></td>
</tr>
<tr>
<td>d1RPV</td>
<td>Delta papillomavirus</td>
<td>White-tailed deer papillomavirus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d2DPV</td>
<td>Delta papillomavirus</td>
<td>Deer papillomavirus</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reference: Genomics and Health, Centre for Public Health Research, Valencia, Spain.
Figure 12  A Bayesian tree for BPV-1 and BPV-2 obtained from different animal species. Values on branches account for Bayesian posterior probability. Values on nodes account for node age median value. Colour coded is according to the host species (horse, orange; donkey, magenta; zebra, red; cow, black).

The different sequences of BPV-1 and BPV-2 are demonstrated in Figure 12, clustered according to the host species. This tree is rooted and timed according to the position in the supertaxon tree. The zebra sequences are demonstrated in red; horse in orange; donkey in magenta and cattle in black. The zebra sequences form sister clades to both BPV-1 (accession numbers HQ541334, HQ541335, HQ541336, HQ541337, HQ541342, HQ541344, HQ541346, HQ541349, ) and BPV-2 (accession numbers HQ541352, HQ541353) and another clade, to BPV-2 (accession number HQ541354). The age for the most recent common ancestor for BPV-1 variants was estimated to be 1.40 Mya (95% posterior density between 0.64 and 2.26), while for BPV-2 variants were estimated to be 0.55 Mya (95% posterior density between 0.17 and 0.98).
Table 4  Estimates of evolutionary divergence between the BPV-1 and BPV-2 sequences obtained from zebra. The number of base differences per sequence from analysis between sequences were conducted in MEGA4 (Tamura et al. 2007). Codon positions included were 1st + 2nd + 3rd + Noncoding. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 456 positions in the final dataset.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. BPV-1 genome X02346</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. HQ541336</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. HQ541337</td>
<td>17</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. HQ541349</td>
<td>15</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. HQ541344</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. HQ541342</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. HQ541346</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. HQ541335</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. HQ541334</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. BPV-2 genome M20219</td>
<td></td>
<td>57</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>65</td>
<td>65</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. HQ541352</td>
<td>63</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>72</td>
<td>72</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. HQ541353</td>
<td>62</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>71</td>
<td>71</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>13. HQ541354</td>
<td>59</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>68</td>
<td>68</td>
<td>2</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

Estimated evolutionary divergences between the BPV-1 and -2 sequences obtained from zebra sarcoid were compared with published sequences by determining the number of base differences per sequence (Table 4). Some sequences were identical, whereas others differed from each other by between 1-72 bp. The BPV-1 sequences HQ541342, HQ541344 and HQ541346 were identical. The rest of the BPV-1 sequences (HQ541334, HQ541335, HQ541336, HQ541337, HQ541349) differed from the published BPV-1 sequence by 1-17 nucleotides. The obtained BPV-2 sequences (HQ541352, HQ541353, HQ541354) differed from the published BPV-2 sequences by only 2-6 nucleotides.


3.4 DISCUSSION

During the last several years the importance of knowledge about the evolution of pathogens such as papillomaviruses has been stressed (Bernard 2005; Bravo et al. 2010), as information about their phylogeny can contribute to their classification. If the evolutionary history of the papillomaviruses is understood, and how ancestral properties have been influenced, the clinical picture can be evaluated with respect to benign or malignant (van Ranst et al. 1992; Bravo & Alonso 2007).

Although most authors based phylogenetic studies on the BPV E1, E2 and L1 genes, E5 was used in this study. E5 is expressed in over 75% of sarcoid lesions suggesting that its expression is maintained throughout the disease (Nasir & Reid 1999; Carr et al. 2001a,b; Chambers et al. 2003).

Papillomavirus gene sequences retrieved from zebra sarcoid tumours can be identified as variants of either BPV-1 or BPV-2. Most BPV-1 sequences gained from non-bovine hosts cluster together and define sister taxa to canonical bovine associated BPV-1. The presence of BPV-1 sequences in zebra sarcoids highlights again the importance of this virus in the aetiology and pathogenesis of sarcoid in the Equidae family (Yuan et al. 2007).

Bovine papillomavirus -1 and -2 belonged unequivocally to the delta-epsilon-zeta papillomavirus superclade (Figure 10). Within this superclade, papillomavirus infecting Canidae (CsPV03 and CPV04) and infecting Cetartiodactyla (eBPV05, eBPV04, d4BPV02, d4BPV01, d3OvPV02, d3OvPV01, CcPV01, d1EEPV, d1RPV, d2DPV) were respectively, monophyletic i.e. they share a recent common ancestor (Figure 11). However, papillomavirus that infect horse and papillomavirus that infect cow are not monophyletic respectively. Strict host specificity of papillomaviruses prevents close contact between viruses but a variety of phylogenetically different hosts can be infected by BPV-1 and BPV-2, which display a wider host range, infecting close species such as the waterbuffalo (Silvestre et al. (2009), but also more distant species such as the horse (Bloch et al. 1994), donkey (Reid et al. 1994), Florida manatees (Trichechus manatus latirostis) (Bossart et al. 2002), tapir (Kidney & Berrocal 2008) and the zebra.

In Figure 12 the sequences obtained from the zebras clustered together (red) and the other BPV-1 sequences (yellow) were all obtained from horse sarcoid or skin from all over the world. A different branch with BPV-1 was obtained from a donkey and another branch contained BPV-1 from bovine and horseskin. They could all relate to BPV-1 genome accession number X02346. In BPV-2 the sequences obtained from the zebras clustered in two branches (red) and again the other
BPV-2 sequences were obtained from horse skin or sarcoid samples. They could all relate to BPV-2 genome accession number M020219. The small differences in nucleotides in the zebra sequences, as seen in Table 4, play a role in this separation of zebra sequences.

The sequences analysed in this study were obtained from free-ranging zebra. The sarcoids originated from two game parks which rules human contact out as far as their playing a possible role in the transmission of the neoplasm. This is in contrast to the situation in horses and donkeys which are handled by humans or possibly share facilities with cattle or sheep in the normal daily farming activities. These zebras were, however, brought into the areas which had been previously inhabited by farm animals when the parks concerned were proclaimed in 1931 and 1972 respectively. Other game species share the grazing in the parks. Evolutionary incongruence is seen in BPV, these viruses infecting the same host (*Bos taurus*) are found in other species as well. For BPV-1, the last common ancestor was 1.35 Mya, and for BPV-2, 0.55 Mya, in this host switch, there was no human intervention as, for instance, could be caused by sharing the same facilities with horses and cattle. The other explanation could be that the common ancestor to BPV-1 and BPV-2, (5.34 Mya) already had the ability to infect other hosts species but was unable to gain effective access to them.

More information might have been obtained using the more variable upstream regulatory region (URR) that harbours transcription factor-binding sites and controls gene expression. It consists of a cluster of genes involved in the initial destabilization of the host cell (E4,5,6,7) and genome replication (E1,2) and a cluster of late genes that encode the capsid proteins where more BPV variants have been described. The E5, besides being detectable in a larger percentage of samples, can however, also be used in the phylogenic calculations (personal communication: Bravo).
3.5 REFERENCES


