Virus Evolution, 2019, 5(2): vez051

doi: 10.1093/ve/vez051
Research article

New insights into Sauropsid *Papillomaviridae* evolution and epizootiology: discovery of two novel papillomaviruses in native and invasive Island geckos

Jessica E. Agius,[1,*,†] David N. Phalen,[1] Karrie Rose,[2,3] and John-Sebastian Eden[4,5,‡]

[1]Faculty of Science, Sydney School of Veterinary Science, University of Sydney, Werombi Road, Camden, New South Wales 2570, Australia, [2]Australian Registry of Wildlife Health, Taronga Conservation Society Australia, Bradleys Head Road, Mosman, New South Wales 2088, Australia, [3]College of Public Health, Medical and Veterinary Sciences, James Cook University, James Cook Drive, Townsville, Queensland 4814, Australia, [4]Marie Bashir Institute for Infectious Diseases and Biosecurity, Faculty of Medicine and Health, Sydney School of Medicine, University of Sydney, Missenden Road, Camperdown, New South Wales 2006, Australia and [5] The Westmead Institute for Medical Research, Centre for Virus Research, Hawkesbury Rd, Westmead, New South Wales 2145, Australia

*Corresponding author: E-mail: jessica.agius@sydney.edu.au

†https://orcid.org/0000-0001-8735-1222

‡https://orcid.org/0000-0003-1374-3551

Abstract

Papillomaviruses cause persistent infections in skin and mucosal membranes and, in at least one species, are also be able to infect a tissue of mesenchymal origin. Infections may either be subclinical or induce proliferative lesions. Of the known papillomaviruses, the majority that have been characterized are from humans and other mammals. Currently, only fifteen complete bird and reptile papillomavirus genomes have been described, and they have been found in birds ($n = 11$), turtles ($n = 2$), and snakes ($n = 2$). Using next-generation sequencing technologies and virus-specific PCR, we have identified two novel papillomavirus genomes, *Hemidactylus frenatus* Papillomavirus 1 and 2 (HfrePV1, HfrePV2), in the widely distributed and highly invasive Asian house gecko (*H.frenatus*) and mute gecko (*Gehyra mutilata*) on Christmas Island and Cocos (Keeling) Islands. HfrePV1 was also detected in critically endangered Lister's geckos (*Lepidodactylus listeri*) in their captive breeding colony on Christmas Island. Tissue-containing virus included epidermis, oral mucosa, and liver (HfrePV1) and epidermis, liver, and colon (HfrePV2). Concurrent infections were found in both *H.frenatus* and *G.mutilata*. Invasive mourning geckos (*Lepidodactylus lugubris*) ($n = 4$), Sri Lankan house geckos (*Hemidactylus parvimaculatus*) ($n = 3$), flat-tailed house geckos (*Hemidactylus platyurus*) ($n = 4$) from the Cocos Islands, and blue-tailed skinks (*Cryptoblepharus egeriae*) ($n = 10$) from Christmas Island were also screened but were not found to be infected. The novel HfrePV1 and HfrePV2 genomes were 7,378 bp and 7,380 bp in length, respectively, and each contained the early (E1, E2, and E7), and late (L1 and L2) open-reading frames. Phylogenetic analysis of the concatenated E1, E2, and L1 proteins from both papillomaviruses revealed that they clustered with, but were basal to, the Sauropsida clade containing bird and reptile viruses. This study sheds light on the evolution of

papillomaviruses and the distribution of pathogens in a highly invasive species impacting endangered populations of geckos.

Key words: evolution; *Hemidactylus frenatus*; lizard; papillomavirus; reptile; Sauropsid.

1. Introduction

Papillomaviruses have been detected in all vertebrate taxon (excluding amphibians), but the majority of known papillomaviruses are of human and other mammalian origin (Van Doorslaer et al. 2017b). Only fifteen complete papillomavirus genomes have been detected and sequenced in birds and reptiles (Van Doorslaer et al. 2017a). Two of these are from snakes and they appear to be most closely related to mammalian papillomaviruses (Lange et al. 2011; Kubacki et al. 2018). The remainder are from a range of often unrelated birds including Adelie penguin (*Pygoscelis adeliae*) (Varsani et al. 2014, 2015), common chaffinch (*Fringella coelebs*) (Prosperi et al. 2016), Atlantic canary (*Serinus canaria*) (Truchado et al. 2018b), yellow-necked Francolin (*Francolinus leucoscepus*) (Van Doorslaer et al. 2009), African grey parrot (*Psittacus erithacus*) (Tachezy et al. 2002), Northern fulmar (*Fulmarus glacialis*) (Gaynor et al. 2015), black-legged kittiwake (*Rissa timorensis*), American herring gull (*Larus agrentatus smithsonianus*), Atlantic puffin (*Fratercula arctica*) and mallard (*Anas platyrhynchos*) (Canuti et al. 2019), and two turtles; green sea turtle (*Chelonia mydas*) and loggerhead sea turtle (*Caretta caretta*) (Herbst et al. 2009). These viruses form the monophyletic Sauropsid (bird and reptile) clade (Van Doorslaer et al. 2017b). To date no papillomaviruses have been characterized from lizards.

The genomes of papillomaviruses are relatively small (~8 kbp) with high genotypic diversity. The coding regions of *Papillomaviridae* typically contain between six and eight open-reading frames (ORFs) yielding two classes of proteins (Fields, Knipe, and Howley 2007). These protein classes include the early genes that encode proteins with regulatory function during virus replication, and the late structural genes involved in virion capsid formation. The early E1 and E2, and late L1 and L2 proteins form the foundation of the viral genome, occurring in all sequenced papillomaviruses (de Villiers et al. 2004). Four early ORFS coding for the E4, E5, E6, and E7 proteins are commonly found within the mammalian clade of papillomaviruses. The E3 and E8 ORFs have additionally been identified in mammalian papillomaviruses, though typically do not encode individual proteins (Zheng and Baker 2006). However, the E8 ^E2 viral transcript generated from splicing the E8 ORF is highly conserved across mammalian papillomaviruses (Dreer, van de Poel, and Stubenrauch 2017). Within the Sauropsid clade all viruses have the E7 ORF, and all but one, the African grey papillomavirus (PePV1) contain the E6 ORF. A unique E9 or X ORF of unknown function is found in the clade containing the majority of the avian papillomaviruses, but is absent in the two turtle papillomaviruses, and the unique papillomavirus from the Northern fulmar (FgPV1) (Van Doorslaer et al. 2017b). The E4 ORF has similarly not been identified in all Sauropsid papillomaviruses, and while predicted to occur in most avian papillomaviruses (Varsani et al. 2014; Truchado, Williams, and Benítez 2018a), has only been genomically annotated in two chelonid (CcPV1 and CmPV1), and two avian (ScPV1 and FlPV1) papillomaviruses (Benson et al. 2012).

Papillomaviruses are epitheliotrophic, infecting the skin and mucosal epithelia of their hosts. These viruses exhibit high host specificity, and with the exception of bovine papillomaviruses 1 and 2 (BPV1, BPV2) which can infect cattle and horses *in vivo*, and fibroblasts from cattle, cats and horses *in vitro* (Gil da Costa et al. 2017), these viruses rarely infect more than one species (Wang and Hildesheim 2003). The majority of papillomaviruses are non-pathogenic and cause subclinical infections in infected individuals or cause benign, often self-limiting papillomatous lesions. A limited number of papillomavirus strains are oncogenic and are associated with the development of malignant neoplasms (Munday and Kiupel 2010). Viruses infecting sauropsids were thought to behave similarly to mammals, infecting and causing lesions on epithelial surfaces such as the oral mucosa, nasolacrimal epithelium, beak, skin of the face, eyelids, feet, and legs (Truchado, Williams, and Benítez 2018a). Recently, however, it has been shown that a papillomavirus identified in tissues from the foot of a Northern fulmar—*Fulmaris glacialis* papillomavirus 1 (FgPV1) can also infect and cause a productive infection in chondrocytes, which are cells of mesenchymal origin (Gaynor et al. 2015). Conversely, in some avian papillomaviruses (American herrings, mallards, black-legged kittiwakes, Adelie penguins, and Atlantic puffins) where infection has been detected, the disease is subclinical, and no papillomatosis are observed (Varsani et al. 2014; Canuti et al. 2019).

Our knowledge of the virosphere is ever expanding. Given the large diversity of papillomaviruses described in mammalian hosts and what has been observed for RNA viruses (Shi et al. 2018), it is reasonable to predict that many non-mammalian, particularly Sauropsid origin papillomaviruses remain undiscovered that should shed light on the origins and evolution of this ubiquitous pathogen. In this study, we use whole genome sequencing to identify and characterize the genomes of two novel Sauropsid papillomaviruses. We show that they have an uncharacteristically wide host range, that their tissue distribution includes liver, a tissue of endodermic origin, and that they cluster basal to the clade-containing turtle and bird papillomaviruses, with speciation likely predating the avian-chelonid divergence.

2. Materials and methods

2.1 Animal ethics

Animal collection for Christmas Island and Cocos Islands specimens were approved by the University of Sydney Animal Ethics Committee (AEC) (2017/1211). Animal collection for specimens collected from Brisbane, Australia were approved by the Taronga Conservation Society Australia Animal Ethics Committee (3b/12/18). All animals used in this research were in compliance with the NSW Animal Research Act 1985, and the Australian code for the care and use of animals for scientific purposes eighth edition.

2.2 Sample preparation

Samples were collected from an Asian house gecko (*Hemidactylus frenatus*) free-ranging on Christmas Island

(10°29′30.50″S 105°38′49.60″E), that was displaying signs consistent with infection with a novel *Enterococcus* sp. bacterium (Rose et al. 2017). The affected gecko was euthanized with an overdose of Alfaxan (Jurox Animal Health). Diseased epidermis, mucosa, and submucosa of the maxilla were isolated and stored in 95 per cent ethanol at 4°C until processing.

2.3 DNA isolation

Genomic DNA was isolated from alcohol-fixed tissues using a modified animal tissue protocol from the DNeasy Blood and Tissue Kit (Qiagen, Victoria, Australia). Briefly, tissues were rehydrated with four phosphate-buffered saline washes to remove residual fixative, mechanically ground and pre-treated with a lysozyme digestion step (25 mM Tris–HCl pH 8, 2.5 mM EDTA, 1.2% triton X-100, 20 mg/µl lysozyme) recommended for lysis of Gram-positive bacteria. Following enzymatic pre-treatment, samples were digested with proteinase K overnight and the DNA purified following the manufacturer's instructions.

2.4 DNA sequencing

Purified DNA was prepared as shotgun libraries using the Truseq DNA PCR-free library (average insert size of 350 bp), and sequenced on the Illumina HiSeq X Ten platform generating 50 Gb of 150 bp paired-end reads (Macrogen Inc., Seoul, Rep. of Korea).

2.5 Sequence analysis

Raw sequence reads were quality trimmed using Trimmomatic (Bolger, Lohse, and Usadel 2014) with a Phred score of less than 25 removed. Trimmed reads were mapped to the reference assembly *Gekko japonicus* V1.1 (NCBI genome assembly GCF_001447785.1) and then the unmapped reads *de novo* assembled using MEGAHIT 1.0.6 (default parameters) (Li et al. 2015). Contigs were annotated using Basic local Alignment Search Tool (BLAST) (Altschul et al. 1990). A search of the full non-Gecko contig list was performed against the NCBI non-redundant nucleotide and protein databases with an *E*-value cut-off of $1E^{-10}$ using BLAST and DIAMOND (Buchfink, Xie, and Huson 2015) and annotated by class (Eukaryota, Bacteria, Archaea, and Viruses). Here, a single viral contig was identified which corresponded to the complete genome of a novel papillomavirus, *H.frenatus* papillomavirus 1 (HfrePV1). The only other notable organism present was a bacteria, *Enteroccoccus*, which was known and the focus of the original sequencing efforts.

2.6 Genome annotation

The HfrePV1 papillomavirus contig was imported into Geneious Prime (version 2019.0.04) (https://www.geneious.com) and annotated. First, the contig was found to have homologues overlapping terminal regions, suggesting a closed circular genome. Next, ORFs with a minimum nucleotide number of 250 bp were identified, and then predicted based on sequence and structural similarity to known papillomavirus genomes. Predicted ORFs identified in Geneious Prime were translated and queried using the homology search tool HHblits (Max Planck Institute Bioinformatics Toolkit) (Remmert et al. 2012) with default search parameters. Conserved domains qualified as putative predictions if they were of sufficient homology; containing a minimum *E*-value threshold of $1E^{-2}$, and minimum probability and query cover of 70 per cent and 50 per cent respectively. The top five targets for each ORF were listed. When targets didn't qualify using the above parameters, the hit with the highest probability was included. The top hit for each ORF annotated in HHblits were comparatively interrogated and aligned against Geneious prime annotations.

2.7 Motif identification

The non-coding or long-control region (LCR) of HfrePV1 was screened to identify conserved motifs and regions using the Multiple EM Motif Elicitation tool (MEME, version 5.0.2) (Bailey et al. 2015). A total of thirteen LCR nucleotide sequences from available complete Sauropsid papillomavirus genomes (FgPV1, FlPV1, PaPV1, PaPV2, PePV1, FcPV1, ScPV1, CmPV1, CcPV1, AplaPV1, FarcPV1, LsmiPV1, and RtriPV2) were used as input. A maximum number of twenty motifs and their associated *E*-values, lengths, and sites were generated. The HfrePV1 LCR nucleotide sequence was further screened manually to identify additional pattern notations conserved within *Papillomaviridae* and not identified using MEME software.

To resolve the highly divergent putative E7 protein in HfrePV1 identified during ORF annotation, the translated sequence was aligned in Geneious Prime to the E7 proteins (or putative E8 for PePV1) of the aforementioned Sauropsid papillomaviruses using MAFFT version 7.388 (Katoh and Standley 2013) with the E-INS-i algorithm. The alignment was screened at the amino acid level to locate unique synamorphic features and identify conserved domains and residues.

2.8 Pairwise sequence analysis

Pairwise comparisons of sequences at the nucleotide and protein level with HfrePV1 and thirteen Sauropsid papillomaviruses for five ORFs (E1, E2, E7, L1, and L2) were performed using Geneious Prime version 2019.0.4. Specifically, the nucleotide and protein alignments were made using MAFFT with the FFS-NS-i x 1000 and L-INS-i algorithms, respectively.

2.9 Sample screening for HfrePV1 DNA in tissues

DNA from alcohol-fixed liver ($n = 93$), colon ($n = 3$), oral mucosa ($n = 7$), and epidermis ($n = 19$) of blue-tailed skinks (*Cryptoblepharus egeriae*) ($n = 10$), Lister's (*Lepidodactylus listeri*) ($n = 11$), Asian house ($n = 26$), and mute geckos (*Gehyra mutilata*) ($n = 8$) collected on Christmas Island, Asian house ($n = 16$), Sri Lankan house (*Hemidactylus parvimaculatus*) ($n = 3$), mute ($n = 7$), flat-tailed (*Hemidactylus platyurus*) ($n = 4$), and mourning geckos (*Lepidodactylus listeri*) ($n = 4$) on Cocos (Keeling) Islands, and Asian house geckos from Brisbane, Australia ($n = 9$) were extracted using the Animal Tissue Protocol from the DNeasy Blood and Tissue Kit (Qiagen). Tissues were rehydrated and mechanically ground as described above. Following the grinding step, samples were digested with proteinase K for 3 h and the DNA purified following the manufacturer's instructions. DNA was amplified using a conventional PCR assay specific for the HfrePV1 conserved L1 gene (Forward: 5′-TCCACGGTTGCAGC TGTATT-3′ and reverse: 5′-GTGAGCCCGCGTATTATCCA-3′). Amplicons were visualed by agarose gel electrophoresis and then bi-directionally sequenced using amplification primers at the Australian Genome Research Facility (AGRF). Sequence comparisons revealed the presence of a second papillomavirus type (HfrePV2) with approximately 9 per cent nucleotide divergence to HfrePV1.

2.10 Hemidactylus papillomavirus 2 (HfrePV2) genome sequencing

To complete the genome of HfrePV2, the DNA from the PCR positive mandibular mucosal lesion of an Asian house gecko was submitted for whole genome sequencing on the Illumina NovaSeq 6000 platform (AGRF) generating 150 Gb of 150 bp paired-end reads. The raw sequences were processed to identify viral sequences as previous, except the trim reads were also mapped onto the complete HfrePV1 genome with BBMap (Bushnell 2014). This generated a partial genome that was then filled by PCR and a primer walking strategy. Sequence annotation, motif identification, and pairwise sequence analysis of the HfrePV2 genome were identical to HfrePV1.

2.11 Phylogenetic analysis

To cover known papillomavirus diversity, a single representative genome was selected from each of the forty seven classified genera for animal papillomaviruses. When more than one papillomavirus sequence of non-mammalian origin was identified for a genus, all sequences were included in the phylogenetic analysis. All unclassified mammalian papillomaviruses not assigned into a genus were excluded. Full-length sequences for each included viral genome ($n = 57$) were downloaded from the Papillomavirus Episteme (PaVE) (Van Doorslaer et al. 2017a) and translated into proteins. Individual early (E1, E2) and late (L1, L2) core genes for all downloaded sequences, combined with HfrePV1 and HfrePV2 sequences and aligned using MAFFT version 7.388 with the E-INS-i algorithm in Geneious Prime (version 2019.0.04). Each core gene was initially examined individually to assess the quality of the alignments and phylogenetic signal. All core genes with the exclusion of L2 were considered suitable to resolve phylogenetic incongruence. Alignments were trimmed of ambiguously aligned positions using trimAL (version 1.2), conserving at least 50 per cent, 30 per cent, and 60 per cent of positions for E1, E2, and L1 respectively. The trimmed alignments were then concatenated, having a total sequence length of 1,191 amino acids. A maximum likelihood phylogeny of the concatenated alignment was estimated using PhyML (version: 20150402) with the best-fit substitution model (LG+G) determined using ProtTest 3 and 1,000 bootstrap replicates. The phylogeny generated using this approach represented the best topology with nearest-neighbour interchange (NNI) and subtree pruning and re-grafting (SPR) searches. The phylogeny was visualized using Figtree version 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) with midpoint rooting.

2.12 Screening short-read archive database

The NCBI short-read archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra) was accessed to screen published high-throughput raw DNA sequencing data of reptiles (NCBI taxa ID: 8504) for papillomavirus discovery. SRA data generated from this search were filtered to eight relevant reptilian genera (*Gehyra* [NCBI taxa ID: 95113], *Crytodactylus* [NCBI taxa ID: 96732], *Hemidactylus* [NCBI taxa ID: 47727], *Agamura* [NCBI taxa ID: 401555], *Gekko* [NCBI taxa ID: 8565], *Alsophylax* [NCBI taxa ID: 1208010], *Cnemaspis* [NCBI taxa ID: 221557], and *Pseudoceramodactylus* [NCBI taxa ID: 1355907]), which were represented by 200 sequence libraries (as of 4 December 2018). The SRA data generated from the select genera search were further filtered to include only genomic DNA libraries. Complete datasets available post-filtering were downloaded from the NCBI database using NCBI SRA toolkit (version 2.8.0) and converted into FASTQ format using fastq-dump (version 2.5x). The available FASTQ reads were queried against a reference papillomavirus database containing the core gene protein sequences of HfrePV1 and HfrePV2 using Diamond (version: 0.9.22), with parameters (matrix BLOSUM50, max target seqs 1, max hsps 1) and an *E*-value threshold of $1E^{-5}$. Sequence hits in FASTQ format were extracted using seqtk (version 1.3).

2.13 Histological screening

Formalin-fixed specimens from Christmas Island ($n = 103$) and Cocos Islands ($n = 202$) invasive Asian house geckos were placed in dorsal recumbency and a sagittal midline incision made through the body wall from rostrum to pectoral girdle. Tissues were paraffin-embedded using routine histological processing, and sectioned at 5 μm. Tissue sections were stained with haematoxylin and eosin (H&E). Skin and mucosal surfaces (oral cavity, conjunctiva, and cloaca) were examined microscopically for evidence of papillomatous lesions. Papillomatous lesions were defined as branching fronds of hyperplastic epidermis supported by a fibrovascular stalk (Schmidt, Reavill, and Phalen 2015).

3. Results

3.1 Papillomavirus discovery

During epidemiological and genomic investigations of a novel enterococcus disease outbreak in Christmas Island and Cocos Islands geckos (Rose et al. 2017), DNA from an enterococcus infected animal was subjected to whole genome sequencing on the Illumina HiSeq platform. While examining the microbial sequences by *de novo* assembly and annotation of reads that did not map to the *Gekko japonicus* V1.1 genome, we considered that viral sequences might be present. From a total of 2,847,697 putative non-Gecko sequence contigs, 385 were annotated as viral; however, once excluding non-specific and likely endongenous elements, only a single contig remained. The sequence (length = 7473 bp) partially aligned to the L1 gene of human papillomavirus type 39 with 42.7 per cent amino acid identity. The trimmed sequence reads where then remapped onto the papillomavirus sequence, which demonstrated a mean coverage depth of 45×. The approximate size and coverage suggested a novel papillomavirus genome was present in the sample, which was then further suggested by the identification that the sequence could be cirularized with an overlapping identical region (95 nt) present at the termini, which collapsed to produce a final genome length of 7,378 nt for the virus, and here after named HfrePV1.

A second related virus, HfrePV2, was then identified in the liver of *H.frenatus* by PCR screening of the conserved L1 gene using primers designed from HfrePV1. The HfrePV2-positive sample was then subjected to whole genome sequencing on an Illumina NovaSeq platform. The trimmed sequences were mapped onto the HfrePV1 genome to determine the completeness and coverage depth. The HfrePV2 mapped reads spanned only 58 per cent of the genome with a mean coverage depth of 4×. To complete the HfrePV2 genome sequence, overlapping PCRs were designed and sequenced by Sanger reaction.

3.2 Features of the HfrePV genome

The complete genomes of the novel HfrePV1 (Fig. 1) and HfrePV2 were 7,378 and 7,380 bp in length respectively, with a G+C content of 51.6 per cent for both viruses, and 89 per cent

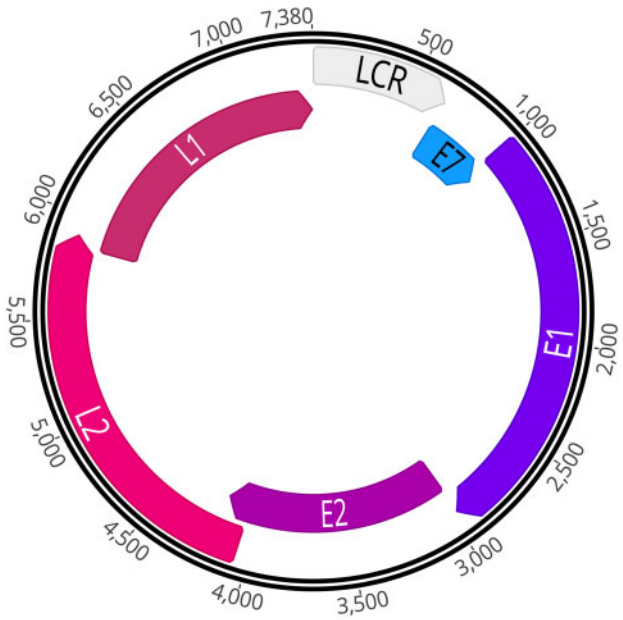

Fig. 1. Genomic organization of the novel *Hemidactylus frenatus* papillomavirus 1 (HfrePV1) identified in the Asian house gecko showing the LCR, and early and late genes. The genome of the novel *H.frenatus* papillomavirus 2 (HfrePV2) is not shown as the organization is essentially identical to HfrePV1.

nucleotide similarity to each other. Annotation of the HfrePV1 and HfrePV2 genomes using Geneious Prime ORF prediction software identified eight ORFs on the same coding strand (greater than 250 nt), and in positions in the genome orthologous to known Sauropsid papillomavirus genes (Fig. 2a). Comparisons to reference protein sequences by profile hidden Markov model confirmed five of these eight canonical papillomavirus ORFs based on homology scores (*E*-value $<1E^{-2}$, probability >70%, and query cover >50%) (Fig. 2a). These included the early E1, E2, and E7 genes (see Supplementary Tables S1, S2, and S5), and late L1 and L2 genes (see Supplementary Tables S7 and S8), with no significant homology scores found to E4, E6, and E9 ORFs (see Supplementary Tables S3, S4, and S6). The LCR of HfrePV1 (Fig. 3) defined as the sequence between the late (stop codon of L1) and early genes (start codon of E7) was 654 bp long (8.9% of genome), contained a polyadenylation (PolyA) site (276–281 bp) and TATA box (285–289 bp) in a region rich in AT content. The Nf1 (TTGGC) (303–307 bp) and Sp1 (GGCGGG) (361–366 bp) binding sites were also located. The LCR region for HfrePV2 (Fig. 3), 659 bp long (8.9% of genome) similarly contained a PolyA site (278–283 bp) and TATA box (294–297 bp) in an AT rich region. No Nf1 or Sp1 binding sites matching the motifs identified in HfrePV1 were evident, though similar motifs TTGGG (305–308 bp) and GGTGGG (331–336 bp) with a single nucleotide difference occurred in corresponding regions of the LCR. Three canonical E2-binding sites ($ACCN_{2\text{-}11}GGT$) were identified at analogous regions of the LCR for both viruses. These binding sites were found at (122–133 bp|129–137 bp), (538–554 bp|542–558 bp), and (630–645 bp|634–650 bp) for HfrePV1 and HfrePV2, respectively.

3.3 Genomic organization of HfrePV1

The conserved early E1 (665aa) (see Supplementary Table S1) and E2 (408aa) (see Supplementary Table S2), and late L1 (511aa) (see Supplementary Table S7) and L2 (619aa) (see Supplementary Table S8) proteins each met the homology criteria when queried with HHblits (see Supplementary Tables S1, S2, S7, and S8). Three potential early ORFs, including E4 (167aa), E6 (121aa), and E9 (164aa) were identified on the basis of genomic location, but when queried using HHblits profile alignments produced low homology hits with no significant identity to papillomavirus proteins. Instead the highest scoring hits were bacterial in origin with no characteristic motifs, and lacking reliable evidence for any distinguishable E4 (see Supplementary Table S3), E6 (see Supplementary Table S4), and E9 (see Supplementary Table S6) ORFs within the genome of HfrePV1. The putative E7 ORF (122aa) met the homology criteria (see Supplementary Table S5) and contained unique features typical of E7 proteins including the conserved LxCxE-binding domain at position 22-26aa, and zinc-binding motif (CxxC) in the C-terminal region at 79–109aa (Fig. 2b and c). HfrePV1 additionally contained the non-folded CR1 and CR2 motifs associated with E7 (Fig. 2b).

3.4 Genomic organization of HfrePV2

The annotated early E1 (656aa) (see Supplementary Table S1) and E2 (412aa) (see Supplementary Table S2), and late L1 (508aa) (see Supplementary Table S7) and L2 (616aa) (see Supplementary Table S8) proteins met the homology criteria when queried on HHblits software. Three potential early ORFs, including E4 (186aa), E6 (104aa), and E9 (164aa) were identified but returned similar results to HfrePV1, where the sequences did not significantly match against known papillomavirus sequences. The putative E7 ORF (123aa) (see Supplementary Table S5) met the homology criteria, and also contained the LxCxE-binding domain (22-26aa), and zinc-binding motif (CxxC) at the C-terminal region (80-110aa), and similarly contained the non-folded CR1 and CR2 motifs (Fig. 2b and c).

3.5 Comparative genomics

Both the HfrePV1 and HfrePV2 genomes contain the fewest early genes compared to known papillomaviruses of Sauropsid origin, and lack distinguishable E4, E6, and E9 ORFs. It remains unclear if these genes are indeed missing or too divergent that they cannot be aligned against the currently described diversity of papillomaviruses. Interestingly, a sister species to HfrePV1 and HfrePV2 isolated from the Northern fulmar, similarly did not contain the E9 ORF. The remaining avian papillomavirus genomes all have E9 genes (Fig. 4).

Pairwise nucleotide sequence identities between HfrePV1, HfrePV2, and other papillomaviruses of Sauropsid origin ($n=13$) ranged from a minimum 12.42 per cent and 12.93 per cent (E7) to a maximum 48.31 per cent and 48.88 per cent (L1) for HfrePV1 and HfrePV2 to *Serinus canaria* papillomavirus 1 (ScPV1) (see Supplementary Table S9). A minimum and maximum pairwise protein-level identity of 6.47 per cent and 6.43 per cent (E7) and 40.07 per cent and 39.85 per cent (L1) was found between HfrePV1 and HfrePV2 to other papillomaviruses of Sauropsid origin (see Supplementary Table S10).

3.6 Phylogenetic analysis of HfrePV1 and HfrePV22

In order to determine the origins and evolutionary history of HfrePV1 and HfrePV2, sequences were compared to known, publically available animal papillomaviruses. Initially, publically available SRAs available at NCBI were mined to identify further novel papillomaviruses in existing sequencing data. Specifically, the novel HfrePV1 and HfrePV2 protein sequences were compared against SRAs of the following reptilian taxanomic groups: *Gehyra*, *Crytodactylus*, *Hemidactylus*, *Agamura*,

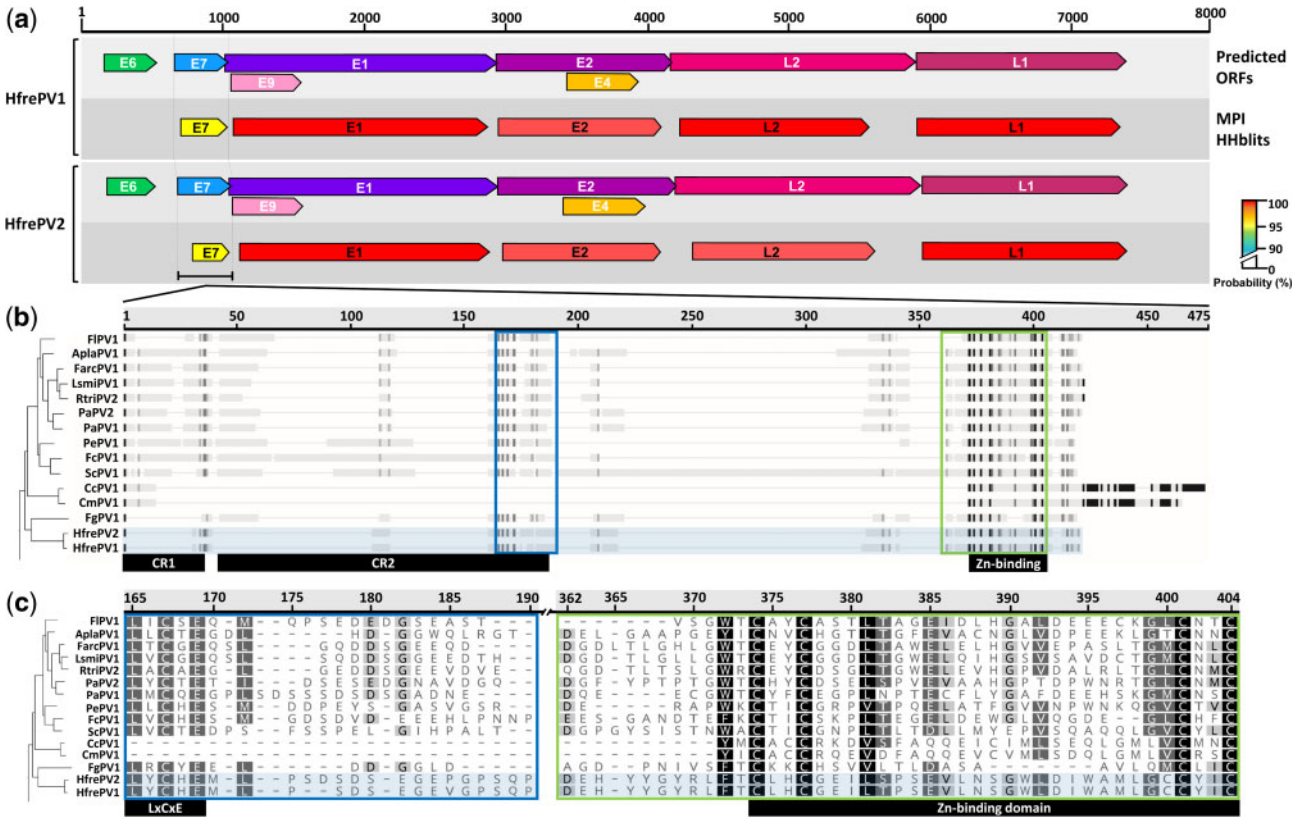

Fig. 2. Sequence and motif annotation of HfrePV1 and HfrePV2. (a) Putative and confirmed early and late ORFs identified in HfrePV1 and HfrePV2 genomes using ORF predictor—Geneious Prime (light grey) and MPI HHblits (dark grey), respectively. Predictions of each conserved domain based on homology probability scores using HHblits are illustrated. (b) The complete alignment of the E7-conserved protein domains of HfrePV1 and HfrePV2 (light blue shading) with other Sauropsid papillomaviruses. The alignment is shaded according to the level of conservation among residues (black to grey). The unfolded CR1 and CR2 regions are illustrated. (c) Amino acid level view of part of the alignment shown in (b). The alignment is shaded by conservation. The LxCxE (pRb binding) and C-terminal zinc-binding domains [N-terminal for *Chelonia mydas* papillomavirus 1 (CmPV1) and *Caretta caretta* papillomavirus 1 (CcPV1)] for each Sauropsid papillomaviruses are indicated.

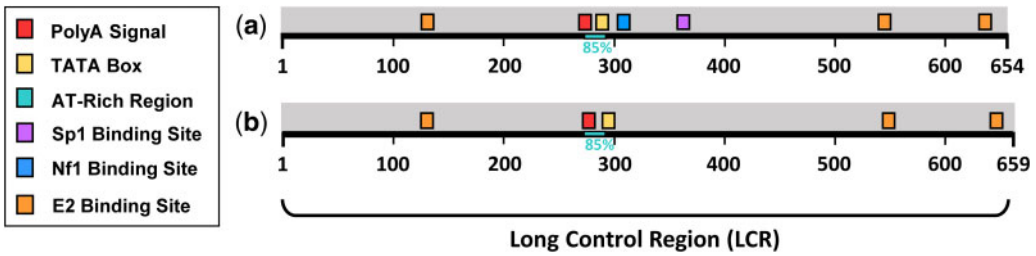

Fig. 3. Representation of the LCR motifs identified using multiple EM motif elicitation (MEME) and sequence screening in (a) HfrePV1 and (b) HfrePV2. Coloured boxes represent different regulatory elements as per key provided. Percentages refer to the frequency of A and T nucleotides.

Gekko, *Alsophylax*, *Cnemaspis*, and *Pseudoceramodactylus*, however, no further papillomaviruses were identified. Therefore, phylogenetic relationships of the novel viruses to other papillomaviruses were inferred using multi-locus alignments of concatenated amino acid sequences of three proteins (E1, E2, and L1) from fifty five available complete papillomavirus genomes spanning the known diversity of the genus (see Supplementary Table S11). The L2 protein did not qualify as a suitable candidate to infer evolutionary history due to poor sequence alignment, and was therefore not included within the concatenation. The multi-gene phylogenetic tree (Fig. 5) illustrates that the novel HfrePV1 and HfrePV2 forms a distinct genetic lineage, clusters within the Sauropsid papillomavirus clade, and it along with FgPV1 form a cluster that is basal to the remaining known Sauropsid *Papillomaviridae*. This phylogenetic tree is in agreement with the evolutionary relationships proposed in Van Doorslaer et al. (2017b), which equally illustrates the two distinct monophyletic clades (mammalian and sauropsid papillomaviruses), the divergence of FgPV1 from the avian and chelonid papillomavirus clade, and the clustering of both snake papillomaviruses (MsPV1 and BcosPV1) within the mammalian monophyletic clade.

A previously described endogenous viral element of the papillomavirus L1 protein was identified in the platypus genome (Cui and Holmes 2012). This sequence was highly divergent and outside the diversity here describing exogenous papillomaviruses, the sequence was therefore omitted from further analysis (data not shown).

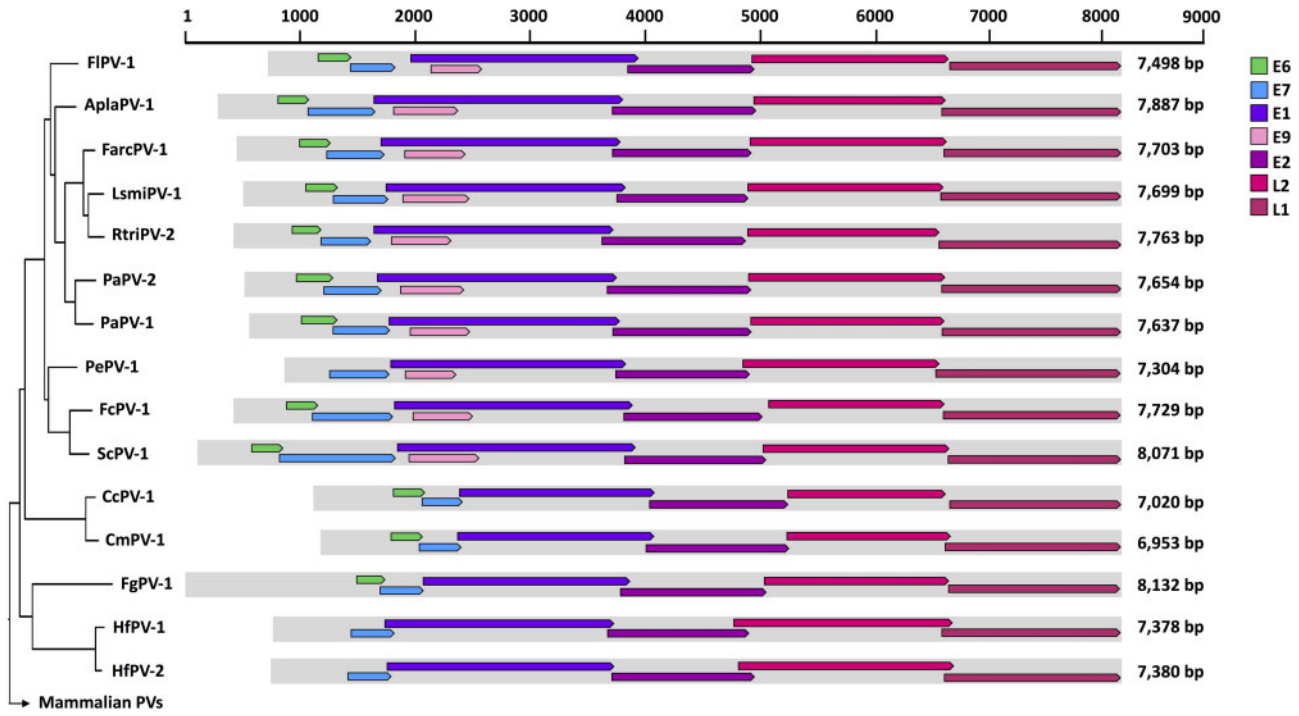

Fig. 4. Illustration of the genomic structure of *Hemidactylus frenatus* papillomavirus 1 (HfrePV1) and *Hemidactylus* papillomavirus 2 (HfrePV2) as compared to known papillomaviruses from the Sauropsid clade; *Francolinus leucoscepus* papillomavirus 1 (FlPV1), *Anas platyrhnchos* papillomavirus 1 (AplaPV1), *Fratercula arctica* papillomavirus 1 (FarcPV1), *Larus smithsonianus* papillomavirus 1 (LsmiPV1), *Rissa tridactyla* papillomavirus 2 (RtriPV2), *Pygoscelis adeliae* papillomavirus 2 (PaPV2), *Pygoscelis adeliae* papillomavirus 1 (PaPV1), *Pstttacus erithacus timneh* papillomavirus 1 (PePV1), *Fringeilla coelebs* papillomavirus 1 (FcPV1), *Serinus canaria* papillomavirus 1 (ScPV1), *Caretta caretta* papillomavirus 1 (CcPV1), *Chelonia mydas* papillomavirus 1 (CmPV1), and *Fulmar glacialis* papillomavirus 1 (FgPV1). ORFs for each papillomavirus are colour coded and refer to a specific early (E1, E2, E6, E7, and E9) and late proteins (L1 and L2). Numbers indicate nucleotide positions.

3.7 Prevalence of papillomaviruses

Papillomavirus DNA was amplified from DNA extracted in the liver, epidermis, oral mucosa, and colon from invasive Asian house geckos and mute geckos on Christmas Island and Cocos Islands, and critically endangered and endemic Lister's geckos on Christmas Island. Both variants of papillomavirus were found in specimens from both islands (Table 1). Of the papillomavirus variants detected, HfrePV2 dominated ($n = 19$) compared to HfrePV1 ($n = 12$). In two specimens on Christmas Island, an Asian house gecko and mute gecko, both variants of the virus were detected in a single tissue. The prevalence of papillomavirus infection in Christmas Island and Cocos Islands reptiles is shown in Table 1. The livers from Asian house geckos ($n = 9$) collected on the Australian mainland (Brisbane, Queensland) were screened for the presence of papillomavirus DNA, but returned negative results.

3.8 Histological screening

Papillomatous lesions were not seen in any of the lizard specimens collected and examined from Christmas Island and Cocos Islands.

4. Discussion

The *Hemidactylus frenatus* papillomavirus sequences identified in this study represent the first report of papillomaviruses found in lizards, and only the third and fourth papillomaviruses identified in reptiles in the Sauropsid clade of papillomavirues. The new viruses infect at least three species of gecko (Asian house, mute and Lister's gecko) and high prevalences of infection were found on both Cocos Islands and Christmas Islands. They appear to cryptically circulate through these species as disease has not been associated with them, a feature that is shared with recently discovered papillomaviruses of gulls, puffins, and ducks (Canuti et al. 2019). While clearly within the Sauropsid clade, these new reptiles viruses have unique genomic features that distinguish them from previously studied papillomaviruses.

Based on the findings of this study and other recent studies, all Sauropsid papillomaviruses retain the two regulatory proteins, E1 and E2, and structural proteins, L1 and L2 found in all papillomaviruses, as well as the accessory E7 protein. Members of the clade Sauropsida differ from mammalian papillomaviruses in that they do not contain an E5 protein (Rector and Van Ranst 2013). The E5 ORF occurs only in the genomes of a few polyphyletic mammalian papillomavirus lineages, therefore it is likely that this viral gene was not lost, but never emerged in bird and reptile papillomaviruses (Willemsen, Félez-Sánchez, and Bravo 2019). They also differ in that the E4 protein is not conserved, and is only annotated as a functional E4 protein in GenBank in four of the Sauropsid papillomaviruses (CcPV1, CmPV1, FlPV1, and ScPV1). Based on current evidence, the E9 protein appears to have been acquired by the Sauropsid papillomaviruses (Van Doorslaer et al. 2017b) after the clade containing the bulk of the avian papillomaviruses separated from clades containing the turtles, avian FgPV1 and both lizard (HfrePV1 and HfrePV2) papillomaviruses. The HfrePV1 and HfrePV2 also appear to have also lost the E6 protein, meaning that they only code for five proteins, which is the fewest number of proteins found in any papillomavirus, and are the only Sauropsid papillomavirus not to have an E6 protein. Although we cannot exclude the possibility that the

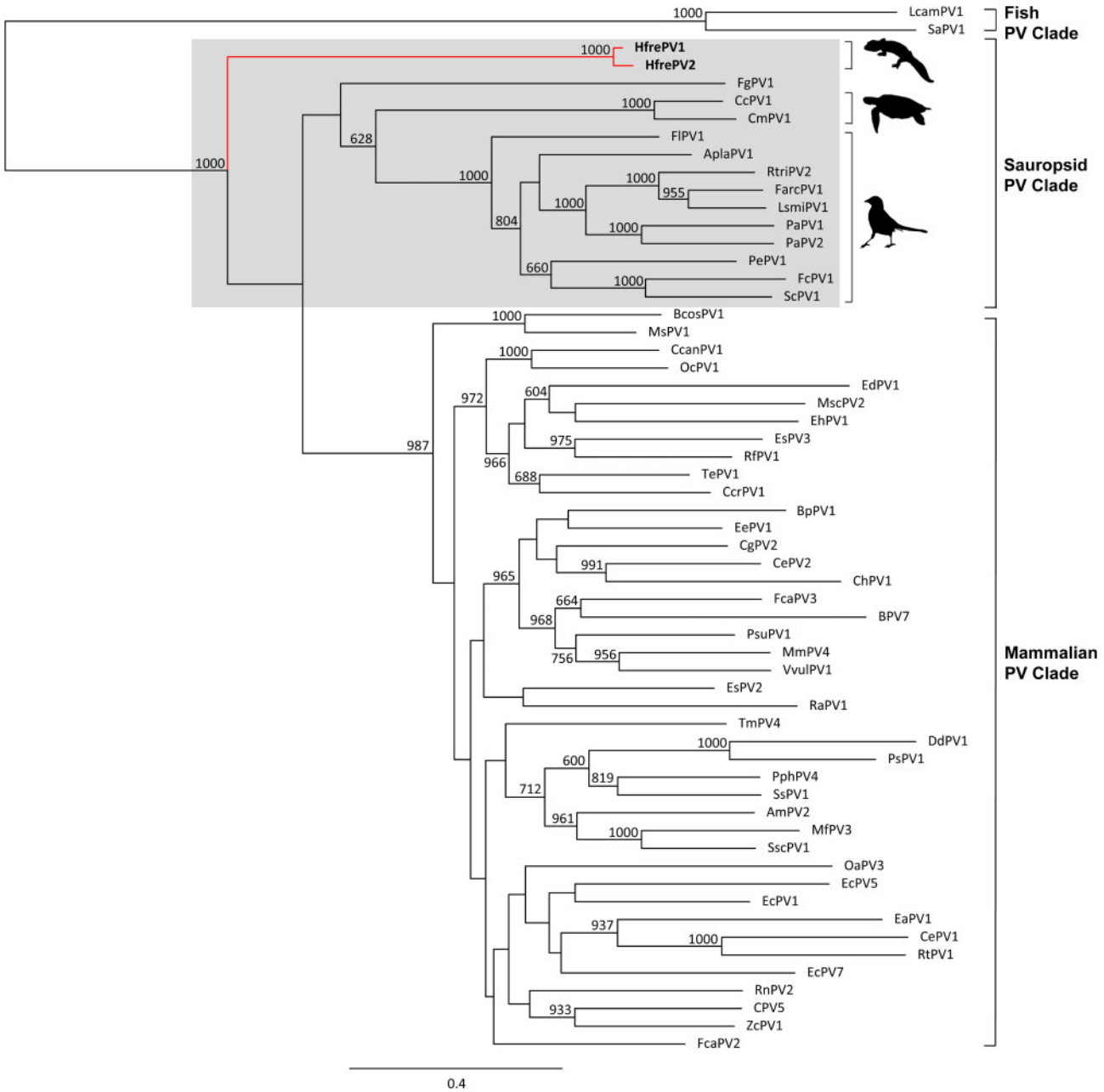

Fig. 5. Phylogenetic tree of concatenated early (E1, E2) and late (L1) papillomavirus proteins. The evolutionary history was inferred from fifty nine amino acid sequences including the novel HfrePV1 and HfrePV2 using a maximum likelihood approach. The percentage of trees in which the associated taxa clustered together is illustrated next to the branches, and is derived from 1000 bootstrap replicates with values hidden when less than 60 per cent. The Sauropsid clade is highlighted in grey, with HfrePV1 and HfrePV2 denoted in bold (red branch outline).

E6 proteins (and also other missing proteins—E4 and E9) are so divergent that they cannot be readily identified.

The genomes of the two novel viruses places them in a monophyletic cluster within the Sauropsid papillomavirus clade, and basal to all currently described Sauropsid papillomaviruses (excluding both snake papillomaviruses), indicating that the ancestor of HfrePV1 and HfrePV2 speciated prior to the split between avian and reptile papillomaviruses, possibly coinciding with the palaeozoic period (Van Doorslaer et al. 2017b). This is consistent with the host species phylogenies and suggestive of host-virus co-divergence model. HfrePV1 and HfrePV2 demonstrated highest homology and form a divergent group with the avian FgPV1, which together formed a sister clade containing the remaining Sauropsid papillomaviruses. HfrePV1, HfrePV2, and FgPV1 all lacked the E4 and E9 proteins. Finding an avian papillomavirus that is more closely related to the gecko viruses than other avian viruses is difficult to explain assuming a virus-host codivergence model, and therefore as previously suggested, the presence of FgPV1 in a bird may represent an ancient host switching event from a primitive reptile to a primitive bird. However, the long branches between HfrePV1, HfrePV2, FgPV1, and the remaining Sauropsid papillomaviruses would also imply some uncertainty as to the ordering and events that underpin the emergence of these viruses. Furthermore, recent work

Table 1. Prevalence of HfrePV1 and HfrePV2 in lizard species from Christmas Island, Cocos (Keeling) Islands and Brisbane, Australia.

Country	Species	Tissue	Positive	Negative	PV species
Christmas Island	*Hemidactylus frenatus*	Liver[a]	8[b]	20	($n = 3$ HfrePV1), ($n = 6$ HfrePV2)
		Colon	1	2	($n = 1$ HfrePV2)
		Oral mucosa	1	1	($n = 1$ HfrePV1)
		Epidermis[a]	3[b]	0	($n = 1$ HfrePV1), ($n = 3$ HfrePV2)
	Gehyra mutilata	Liver[a]	1	7	($n = 1$ HfrePV1), ($n = 1$ HfrePV2)
	Lepidodactylus listeri	Oral mucosa	2	3	($n = 2$ HfrePV1)
		Liver	0	6	
		Epidermis	1	3	($n = 1$ HfrePV1)
	Cryptoblepharus egeriae	Liver	0	10	
		Epidermis	0	10	
Total tissue samples			17	62	
Total lizard specimens			15	40	
Cocos (Keeling) Islands	*Hemidactylus frenatus*	Liver	10	6	($n = 3$ HfrePV1), ($n = 7$ HfrePV2)
	Gehyra mutilata	Liver	1	6	($n = 1$ HfrePV2)
	Lepidodactylus lugubris	Liver	0	3	
		Epidermis	0	1	
	Hemidactylus parvimaculatus	Liver	0	3	
	Hemidactylus platyurus	Epidermis	0	1	
		Liver	0	3	
Total tissue samples			11	23	
Total lizard specimens			11	23	
Brisbane, Australia	*Hemidactylus frenatus*	Liver	0	9	
Total tissue samples			0	9	
Total lizard specimens			0	9	

[a]Single tissue sample with HfrePV1 and HfrePV2 DNA detected.
[b]Cases with positive samples for both liver and epidermal tissues.

highlights that the origins and evolution of most viral families are complex with co-divergence and host-switching events both evidently contributing (Shi et al. 2018). As described previously, both snake papillomavirus genomes characterized (MsPV1 and BcosPV1) clustered with papillomaviruses of mammalian origin rather than within the Sauropsid papillomavirus clade containing all other reptile papillomavirus genomes (Lange et al. 2011; Kubacki et al. 2018). The phylogenetic position of both snake viruses confounds the hypothesis of co-speciation, and may be explained by the occurrence of a cross-species infection event, or sample contamination with mammalian viral DNA.

The E7 protein, the only accessory early protein identified within the genome of HfrePV1 and HfrePV2, like the E6 protein is hypothesized to upregulate DNA replication, inhibit cell checkpoints, and block apoptosis in mammalian hosts (Ruttkay-Nedecky et al. 2013; Truchado, Williams, and Benítez 2018a). The predicted E7 protein identified within the genome of HfrePV1 and HfrePV2 contains motifs typical of the E7 ORF found in other species, including the LxCxE region and modified zinc-binding domain ($CxxC_{n23}CxxC$) (Fig. 2c). HfrePV1 and HfrePV2 also contain the non-folded CR1 motif which is required for cellular transformation and pRB degradation (McLaughlin-Drubin and Munger 2009). The non-folded CR2 motif is also present in HfrePV1 and HfrePV2 E7 protein (Fig. 2b), which is responsible for acting as a high-affinity binding site for pRb and pRb-like proteins (Todorovic et al. 2012). The CR1 and CR2 domains are present within all Sauropsid E7 ORFs, with the exception of both chelonian papillomaviruses (CcPV1 and CmPV1) (Fig. 2b) (Herbst et al. 2009). The absence of the CR1 and CR2 motifs in both turtle papillomaviruses and the presence of an N-terminal zinc-binding domain in place of a C-terminal zinc-binding domain (Fig. 2b and c) suggests that these modified E7 proteins are remnants of the classical E7 protein observed in avian and mammalian papillomaviruses (Herbst et al. 2009).

HfrePV1 and HfrePV2, like PePV1; another Sauropsid papillomavirus from an African grey parrot, and multiple bovine papillomavirus, appear to have lost the E6 protein (Rector and Van Ranst, 2013). The E6 protein plays an important role in establishing infection and permitting virus replication to occur in differentiated cells, by impacting concentrations and the activity of the p53 protein, interfering with innate immunity, blocking apoptosis, and maintaining chromosomal stability (Rector and Van Ranst 2013). Mechanistically, however, the E6 protein both duplicates and compliments the activity of the E7 protein. Therefore, it appears that the E6 protein is not necessarily required for establishment of virus infection or for virus replication in HfrePV1, HfrePV2, or PePV1, and the E7 protein alone may be sufficient for these functions.

The E4 and E9 ORFs also appeared to be absent in the genome of HfrePV1 and HfrePV2. The E4 protein encoded by chelonid and some, but not all avian species (Terai, DeSalle, and Burk 2002), is considered a late protein due to its co-expression with L1 and L2 (Doorbar 2013). The precise role of this protein is not clear, though it likely plays a productive part in the papillomavirus lifecycle, contributing to amplification efficiency and virus synthesis, accumulating in high concentrations in cells during the lytic phase, and interacting with cytokeratin and facilitating virion release (Doorbar 2013). Although some key functions have been proposed, viral replication can still occur in papillomaviruses where the E4 ORF has been made inactive (Raj et al. 2004). The putative E9 ORF identified in the avian clade of the Sauropsid papillomaviruses (which excludes FgPV1 and both gecko and chelonid viruses) is completely embedded within the N-terminal regulatory domain of the core E1 protein and appears to be acquired by an ancestor of this clade (Van

Doorslaer et al. 2017b). The role of the E9 is unknown and its absence of robust motifs and transmembrane domains has lead others to speculate that this ORF may not encode a functional protein (Van Doorslaer et al. 2009).

The non-coding LCR region of HfrePV1 and HfrePV2 located between the late and early genes contained a single PolyA signal, TATA box, and three putative canonical E2 binding sites. HfrePV1 additionally contained Sp1 and Nf1 binding sites within the LCR. The conserved polyA site and TATA box from HfrePV1 and HfrePV2 were separated by a total of 3 and 10 nucleotide bases respectively, and clustered around the AT rich region, unlike other papillomaviruses within the Sauropsid clade, where the polyA signal and TATA box were separated by at least 250 nucleotide bases. The three E2-binding sites identified in HfrePV1 and HfrePV2, like other Sauropsid papillomaviruses, do not follow the 'traditional' four E2-binding site pattern found in most mammalian papillomaviruses (Rogers Waltke, and Angeletti 2011). Similarly, the nucleotide spacer identified within the E2-binding sites of previously discovered Sauropsids and both novel papillomaviruses also differed from mammals, ranging from 2 to 11 instead of six nucleotides in length. The E2-binding sites in papillomaviruses are required for viral replication, with E2 proteins functioning in the activation or repression of virion transcription (contigent on the nature of the binding site) (McBride 2013). Differences in the number of E2 binding sites within the LCR of different papillomaviruses may not correlate with the ability to establish infection, with further studies needed to determine the significance of the variation in the number of binding sites (McBride 2013).

These viruses can be classified as new papillomaviral types. Pairwise alignment of the L1 ORF revealed that HfrePV1 and HfrePV2 shared a maximum nucleotide sequence homology of 48.31 per cent and 48.88 per cent to the next most similar protein sequence from the avian Atlantic canary papillomavirus (ScPV1). As HfrePV1 and HfrePV2 share less than 60 per cent nucleotide identity to the most highly conserved coding region within the papillomavirus genome, these novel papillomaviruses would be considered as a potentially new taxonomic genus (de Villiers et al. 2004). Pairwise nucleotide identity of the L1 gene differed between HfrePV1 and HfrePV2 by more than 10 per cent (10.63%), therefore this homology score would classify both viruses as new papillomaviral types (de Villiers et al. 2004; Van Doorslaer et al. 2018). The discovery and addition of both novel virus types within a new genus adds an important milestone in the understanding of papillomavirus evolution, particularly within the Sauropsida clade, but also highlights that there is likely a large diversity of papillomaviruses in lizards and other basal vertebrates awaiting discovery. For example, there are estimated to be more than 10,000 extant species of reptiles (Pincheira-Donoso et al. 2013; Uetz and Stylianou 2018), so it is highly likely that further new reptile papillomaviruses will be discovered. In silico attempts to detect novel papillomaviruses by screening publicly available DNA sequences amplified from selected reptile species were unsuccessful in this study. However, given that the tissue source of these sequences is not known, and that limited reptilian genera were examined, additional screening of existing reptile sequences available in public databases and liver, skin, and oral mucosal samples from other reptile species is still indicated.

In most instances papillomaviruses are host specific. Our observations show that HfrePV1 and HfrePV2 have a much wider host range than most papillomavirus as both were found to infect a gecko species in the genera *Hemidactylus*, and *Gehyra*, and HfrePV1 was also detected in a species in the genus *Lepidodactylus*. These cross-species transmission events within species affected by papillomaviruses have also been observed to occur in gulls and mallards (Canuti et al. 2019), indicating host-specificity constraints can relax when genetically closely related hosts are involved. Whether HfrePV1 and HfrePV2 can infect other species within these genera or other genera is not known. Attempts to detect these papillomaviruses in the geckos, *L.lugubris*, *H.parvimaculatus*, and *H.platyurus*, and a skink, *C.egeriae*, were unsuccessful. Whether failure to detect these viruses in these additional species is the result of a low sample size, spatial differences in habitat use (e.g. terrestrial versus arboreal) and thus exposure between infected and non-infected species, or resistance to infection (Weitzman, Gibb, and Christian 2018), will require further investigation.

The detection of papillomaviruses in two of the most widely distributed and highly invasive reptiles; the Asian house gecko and mute gecko and the possible wide host range of HfrePV1 and HfrePV2 is of concern. The highly adaptable nature of these cryptic geckos and their sympatry with diverse faunal assemblages of different ecological niches may facilitate papillomavirus transmission, speciation, and evolution thus posing an infectious disease risk to other reptiles (Shah, Doorbar, and Goldstein 2010; Gottschling et al. 2011; Van Doorslaer 2013). The detection of HfrePV1 in both invasive gecko species and the native Lister's geckos already suggests that this virus is moving between them. Whether the origin of one or both of these viruses is the Lister's gecko, the Asian house gecko or the mute gecko is not known, but the cross-species interactions uncovered by this discovery demonstrate the risk of transmission of other unknown pathogens and the potential impact of invasive species. Although a small sample, the absence of HfrePV in Asian house geckos from Brisbane, Queensland which is where the Asian house geckos first colonized suggests that virus might not present on the Australian mainland. This may then support the idea that these viruses evolved in a different species other than the Asian house gecko. Further work is therefore needed to examine the broader distribution of these viruses across the Australian mainland and where the Asian house gecko has invaded.

Papillomavirus infections have historically been thought to be restricted to skin and mucous membranes. However, the recent detection of productive infection in cartilage by the FgPV1 in a bird indicates that papillomaviruses may be able to infect more tissues than previously thought. In this current study, as expected, both HfrePV1 and HfrePV2 were detected in skin and HfrePV1 was detected in oral mucosa. However, they were both also consistently detected in the liver, suggesting that they infected one or more cell types in this organ. While concentrations in the liver samples were sufficient to suggest that replication was occurring in it, we cannot completely rule out the possibility that these lizards were viremic, and the virus was actually in blood and not in the liver cells. If this was the case, it would be very unusual as viremia does not occur in other papillomavirus infections, and it would still imply that virus replication was occurring at a location other than an epithelial surface. HfrePV2 was also detected in a colon sample. Whether the colon represents a location for HfrePV2 infection or if virions passing through it were detected is not known. Additional studies, such as in situ hybridization, will be required to determine the true tissue distribution and cell targets of these viruses.

In conclusion, two novel papillomavirus types within a novel genus across three species of gecko were identified in this study. In depth sequence analysis of both genomes has provided insights into the evolution of bird and reptile

papillomaviruses, adding to the current and very limited diversity of Sauropsid genomes characterized to date. Further work will be required to elucidate the pathogenesis and origins of these viruses.

Data availability

The annotated genome sequence of HfrePV1 and HfrePV2 are available on GenBank (accession numbers: MK207055 and MN194600). Raw reads will be made available on GenBank for the final article.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Acknowledgements

The authors thank Christmas Island National Park, Parks Australia, the Westmead Institute for Medical Research, the Sydney School of Veterinary Science, University of Sydney, and Taronga Conservation Society Australia for their logistical support. We also wish to acknowledge the Sydney Informatics Hub and University of Sydney Core Research Facilities for providing research computing services, and Wei-Shan Chang for her support and sharing of Queensland samples.

Funding

This work was supported by the Australian Government's National Environmental Science Program [grant number NESP 2.3.5], and Australia and Pacific Science Foundation [grant number APSF 17/6].

Conflict of interest: None declared.

References

Altschul, S. F. et al. (1990) 'Basic Local Alignment Search Tool', *Journal of Molecular Biology*, 215: 403–10.

Bailey, T. L. et al. (2015) 'The MEME Suite', *Nucleic Acids Research*, 43: W39–49.

Benson, D. A. et al. (2012) 'GenBank', *Nucleic Acids Research*, 41: D36–42.

Bolger, A. M., Lohse, M., and Usadel, B. (2014) 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data', *Bioinformatics*, 30: 2114–20.

Buchfink, B., Xie, C., and Huson, D. H. (2015) 'Fast and Sensitive Protein Alignment Using DIAMOND', *Nature Methods*, 12: 59–60.

Bushnell, B. (2014) *BBMap: A Fast, Accurate, Splice-Aware Aligner* <https://sourceforge.net/projects/bbmap/>, last accessed 18 Sep 2019.

Canuti, M. et al. (2019) 'New Insight into Avian Papillomavirus Ecology and Evolution from Characterization of Novel Wild Bird Papillomaviruses', *Front Microbiol*, 10: 701.

Cui, J., and Holmes, E. C. (2012) 'Evidence for an Endogenous Papillomavirus-Like Element in the Platypus Genome', *Journal of General Virology*, 93: 1362–6.

de Villiers, E.-M. et al. (2004) 'Classification of Papillomaviruses', *Virology*, 324: 17–27.

Doorbar, J. (2013) 'The E4 Protein; Structure, Function and Patterns of Expression', *Virology*, 445: 80–98.

Dreer, M., van de Poel, S., and Stubenrauch, F. (2017) 'Control of Viral Replication and Transcription by the Papillomavirus E8^E2 Protein', *Virus Research*, 231: 96–102.

Fields, B. N., Knipe, D. M., and Howley, P. M. (2007) *Fields Virology* (Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins).

Gaynor, A. M. et al. (2015) 'Identification of a Novel Papillomavirus in a Northern Fulmar (*Fulmarus glacialis*) with Viral Production in Cartilage', *Veterinary Pathology*, 52: 553–61.

Gil da Costa, R. M. et al. (2017) 'An Update on Canine, Feline and Bovine Papillomaviruses', *Transboundary and Emerging Diseases*, 64: 1371–9.

Gottschling, M. et al. (2011) 'Quantifying the Phylodynamic Forces Driving Papillomavirus Evolution', *Molecular Biology and Evolution*, 28: 2101–13.

Herbst, L. H. et al. (2009) 'Genomic Characterization of Two Novel Reptilian Papillomaviruses, *Chelonia mydas* Papillomavirus 1 and *Caretta caretta* Papillomavirus 1', *Virology*, 383: 131–5.

Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.

Kubacki, J. et al. (2018) 'Complete Genome Sequence of a Boa (Boa Constrictor)-Specific Papillomavirus Type 1 Isolate', *Microbiology Resource Announcements*, 7: e01159.e18 .

Lange, C. E. et al. (2011) 'Novel Snake Papillomavirus Does Not Cluster with Other Non-Mammalian Papillomaviruses', *Virology Journal*, 8: 436.

Li, D. et al. (2015) 'MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph', *Bioinformatics*, 31: 1674–6.

McBride, A. A. (2013) 'The Papillomavirus E2 Proteins', *Virology*, 445: 57–79.

McLaughlin-Drubin, M. E., and Munger, K. (2009) 'The Human Papillomavirus E7 Oncoprotein', *Virology*, 384: 335–44.

Munday, J. S., and Kiupel, M. (2010) 'Papillomavirus-Associated Cutaneous Neoplasia in Mammals', *Veterinary Pathology*, 47: 254–64.

Pincheira-Donoso, D. et al. (2013) 'Global Taxonomic Diversity of Living Reptiles', *PLoS One*, 8: e59741.e41.

Prosperi, A. et al. (2016) 'Identification and Characterization of *Fringilla coelebs* Papillomavirus 1 (FcPV1) in Free-Living and Captive Birds in Italy', *Journal of Wildlife Diseases*, 52: 756–8.

Raj, K. et al. (2004) 'E1∧E4 Protein of Human Papillomavirus Type 16 Associates with Mitochondria', *Journal of Virology*, 78: 7199.

Rector, A., and Van Ranst, M. (2013) 'Animal Papillomaviruses', *Virology*, 445: 213–23.

Remmert, M. et al. (2012) 'HHblits: Lightning-Fast Iterative Protein Sequence Searching by HMM-HMM Alignment', *Nature Methods*, 9: 173.

Rogers, A., Waltke, M., and Angeletti, P. C. (2011) 'Evolutionary Variation of Papillomavirus E2 Protein and E2 Binding Sites', *Virology Journal*, 8: 379.

Rose, K. et al. (2017) 'Emergent Multisystemic Enterococcus Infection Threatens Endangered Christmas Island Reptile Populations', *PLoS One*, 12: e0181240.

Ruttkay-Nedecky, B. et al. (2013) 'Relevance of Infection with Human Papillomavirus: The Role of the p53 Tumor Suppressor Protein and E6/E7 Zinc Finger Proteins (Review)', *International Journal of Oncology*, 43: 1754–62.

Schmidt, R. E., Reavill, D. R., and Phalen, D. N. (2015), *Pathology of Pet and Aviary Birds*, 2nd edn. Hoboken: Wiley.

Shah, S. D., Doorbar, J., and Goldstein, R. A. (2010) 'Analysis of Host-Parasite Incongruence in Papillomavirus Evolution Using

Importance Sampling', *Molecular Biology and Evolution*, 27: 1301–14.

Shi, M. et al. (2018) 'The Evolutionary History of Vertebrate RNA Viruses', *Nature*, 556: 197–202.

Tachezy, R. et al. (2002) 'Avian Papillomaviruses: The Parrot *Psittacus erithacus* Papillomavirus (PePV) Genome Has a Unique Organization of the Early Protein Region and Is Phylogenetically Related to the Chaffinch Papillomavirus', *BMC Microbiol*, 2: 19.

Terai, M., DeSalle, R., and Burk, R. D. (2002) 'Lack of Canonical E6 and E7 Open Reading Frames in Bird Papillomaviruses: *Fringilla coelebs* Papillomavirus and *Psittacus erithacus timneh* Papillomavirus', *Journal of Virology*, 76: 10020.

Todorovic, B. et al. (2012) 'Conserved Region 3 of Human Papillomavirus 16 E7 Contributes to Deregulation of the Retinoblastoma Tumor Suppressor', *Journal of Virology*, 86: 13313.

Truchado, D. A. et al. (2018b) 'Genomic Characterization of the First Oral Avian Papillomavirus in a Colony of Breeding Canaries (*Serinus canaria*)', *Veterinary Research Communications*, 42: 111–20.

——, Williams, R. A. J., and Benítez, L. (2018a) 'Natural History of Avian Papillomaviruses', *Virus Research*, 252: 58–67.

Uetz, P., and Stylianou, A. (2018) 'The Original Descriptions of Reptiles and Their Subspecies', *Zootaxa*, 4375: 257–64.

Van Doorslaer, K. (2013) 'Evolution of the Papillomaviridae', *Virology*, 445: 11–20.

—— et al. (2009) 'Identification of Unusual E6 and E7 Proteins within Avian Papillomaviruses: Cellular Localization, Biophysical Characterization, and Phylogenetic Analysis', *Journal of Virology*, 83: 8759.

—— et al. (2017a) 'The Papillomavirus Episteme: A Major Update to the Papillomavirus Sequence Database', *Nucleic Acids Research*, 45: D499–D506.

—— et al. (2017b) 'Unique Genome Organization of Non-Mammalian Papillomaviruses Provides Insights into the Evolution of Viral Early Proteins', *Virus Evolution*, 3: vex027.

—— et al. (2018) 'ICTV Virus Taxonomy Profile: Papillomaviridae', *Journal of General Virology*, 99: 989–90.

Varsani, A. et al. (2014) 'A Novel Papillomavirus in Adelie Penguin (*Pygoscelis adeliae*) Faeces Sampled at the Cape Crozier Colony, Antarctica', *Journal of General Virology*, 95: 1352–65.

—— et al. (2015) 'Identification of an Avian Polyomavirus Associated with Adelie Penguins (*Pygoscelis adeliae*)', *Journal of General Virology*, 96: 851–7.

Wang, S. S., and Hildesheim, A. (2003) 'Chapter 5: Viral and Host Factors in Human Papillomavirus Persistence and Progression', *JNCI Monographs*, 2003: 35–40.

Weitzman, C. L., Gibb, K., and Christian, K. (2018) 'Skin Bacterial Diversity Is Higher on Lizards than Sympatric Frogs in Tropical Australia', *PeerJ*, 6: e5960–e60.

Willemsen, A., Félez-Sánchez, M., and Bravo, I. G. (2019) 'Genome Plasticity in Papillomaviruses and De Novo Emergence of E5 Oncogenes', *Genome Biology and Evolution*, 11: 1602–17.

Zheng, Z. M., and Baker, C. C. (2006) 'Papillomavirus Genome Structure, Expression, and Post-Transcriptional Regulation', *Frontiers in Bioscience*, 11: 2286–302.