

# A novel papillomavirus in Adélie penguin (*Pygoscelis adeliae*) faeces sampled at the Cape Crozier colony, Antarctica

Arvind Varsani,<sup>1,2,3</sup> Simona Kraberger,<sup>1</sup> Scott Jennings,<sup>4</sup> Elizabeth L. Porzig,<sup>5</sup> Laurel Julian,<sup>1</sup> Melanie Massaro,<sup>6</sup> Annie Pollard,<sup>5</sup> Grant Ballard<sup>7</sup> and David G. Ainley<sup>5</sup>

## Correspondence

Arvind Varsani

arvind.varsani@canterbury.ac.nz

<sup>1</sup>School of Biological Sciences and Biomolecular Interaction Centre, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

<sup>2</sup>Department of Plant Pathology and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA

<sup>3</sup>Electron Microscope Unit, Division of Medical Biochemistry, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory, 7700, South Africa

<sup>4</sup>Department of Fisheries and Wildlife, Oregon Cooperative Fish and Wildlife Research Unit, US Geological Survey, Oregon State University, Corvallis, OR, USA

<sup>5</sup>HT Harvey and Associates, Los Gatos, CA 95032, USA

<sup>6</sup>School of Environmental Sciences, Charles Sturt University, Albury, NSW 2640, Australia

<sup>7</sup>Point Blue Conservation Science, Petaluma, CA 94954, USA

Papillomaviruses are epitheliotropic viruses that have circular dsDNA genomes encapsidated in non-enveloped virions. They have been found to infect a variety of mammals, reptiles and birds, but so far they have not been found in amphibians. Using a next-generation sequencing *de novo* assembly contig-informed recovery, we cloned and Sanger sequenced the complete genome of a novel papillomavirus from the faecal matter of Adélie penguins (*Pygoscelis adeliae*) nesting on Ross Island, Antarctica. The genome had all the usual features of a papillomavirus and an E9 ORF encoding a protein of unknown function that is found in all avian papillomaviruses to date. This novel papillomavirus genome shared ~60% pairwise identity with the genomes of the other three known avian papillomaviruses: *Fringilla coelebs papillomavirus* 1 (FcPV1), *Francolinus leucoscepus papillomavirus* 1 (FIPV1) and *Psittacus erithacus papillomavirus* 1. Pairwise identity analysis and phylogenetic analysis of the major capsid protein gene clearly indicated that it represents a novel species, which we named *Pygoscelis adeliae papillomavirus* 1 (PaCV1). No evidence of recombination was detected in the genome of PaCV1, but we did detect a recombinant region (119 nt) in the E6 gene of FIPV1 with the recombinant region being derived from ancestral FcPV1-like sequences. Previously only paramyxoviruses, orthomyxoviruses and avian pox viruses have been genetically identified in penguins; however, the majority of penguin viral identifications have been based on serology or histology. This is the first report, to our knowledge, of a papillomavirus associated with a penguin species.

Received 6 February 2014

Accepted 23 March 2014

## INTRODUCTION

The family *Papillomaviridae* is a large family with over 270 distinct types of papillomaviruses classified into 38 genera (Bernard, 2013; Bernard *et al.*, 2010). A large proportion (~160) of the established papillomavirus types are human

The GenBank/EMBL/DBJ accession number for the sequence determined in this study is KJ173785.

Two supplementary figures and supplementary data are available with the online version of this paper.

isolates, but a significant number of non-human papillomaviruses (113) have been identified in 54 different host species spanning 16 taxonomic orders (Bernard, 2013; Bernard *et al.*, 2010; Rector & Van Ranst, 2013). Over the last 5 years, there has been a rapid expansion in the discovery of novel papillomaviruses, from both human and animal sources. This has been driven by the advent of new molecular tools such as  $\phi$ 29 DNA polymerase used in rolling-circle amplification and next-generation sequencing (NGS) technology.

Papillomaviruses have circular dsDNA genomes of ~8 kb that encode at least six conserved ORFs. The E1 and E2 genes encode regulators of replication and transcription, E6 and E7 encode the transforming proteins, whilst L1 and L2 encode the major and minor capsid proteins, respectively. L1 and L2 assemble into ~60 nm *T*=7 icosahedral virions (Baker *et al.*, 1991; Trus *et al.*, 2005). Papillomaviruses were identified initially as the aetiological agent of skin warts in various animals (Shope & Hurst, 1933; zur Hausen, 2009a, b). As such, papillomaviruses exclusively infect skin and squamous mucosa, although not all types induce neoplastic growth. Despite the identification of a large number of non-human papillomaviruses in various animals, only three papillomaviruses have been recovered from reptiles: diamond python (*Morelia spilota spilota*), loggerhead sea turtle (*Caretta caretta*) and green sea turtle (*Chelonia mydas*) (Herbst *et al.*, 2009; Lange *et al.*, 2011), and three from birds: African grey parrot (*Psittacus erithacus*), common chaffinch (*Fringilla coelebs*) and yellow-necked francolin, (*Francolinus leucoscepus*) (Terai & Burk, 2002; Terai *et al.*, 2002; Van Doorslaer *et al.*, 2009).

Genetic elements of a few viruses have been identified in some penguin species (see Table 1 for details and summary); however, their genomes have not been identified. These include avian poxvirus in an African penguin (*Spheniscus demersus*) (Carulei *et al.*, 2009) from South Africa and Magellanic penguins (*Spheniscus magellanicus*) (Kane *et al.*, 2012) from Argentina, and Newcastle disease virus (NDV) in Adélie penguins (*Pygoscelis adeliae*) from King George Island (Thomazelli *et al.*, 2010).

The majority of work on viruses in penguins has used serological detection of paramyxoviruses in Adélie penguins on Ross, Windmill, Peterson, Midgley, Shirley, Cameron and Chappell islands and near Davis and Dumont d'Urville bases in the Antarctic (Alexander *et al.*, 1989; Austin & Webster, 1993; Morgan & Westbury, 1981, 1988), in rockhopper penguins (*Eudyptes chrysocome*) on the Falkland Islands (Miller *et al.*, 2010), in blue penguins (*Eudyptula minor*) at Port Campbell and on Phillip Island, Australia (Morgan *et al.*, 1985), and in royal (*Eudyptes schlegeli*), king (*Aptenodytes patagonicus*) and rockhopper penguins on Macquarie Island, Australia (Alexander *et al.*, 1989; Morgan *et al.*, 1981). In contrast, some serology work on orthomyxoviruses has shown the presence of reactive antibodies in Adélie penguins on Ross and Peterson Islands in the Antarctic and from chinstrap (*Pygoscelis antarctica*) and gentoo (*Pygoscelis papua*) penguins on King George Island (Baumeister *et al.*, 2004). Some serology work has indicated that there may be flavivirus-like infections in blue penguins at Port Campbell and Phillip Islands and in royal, king and rockhopper penguins on Macquarie Island (Morgan *et al.*, 1981, 1985). Gardner *et al.* (1997) provided some serological evidence of infectious bursal disease virus in Adélie and Emperor (*Aptenodytes forsteri*) penguins around Mawson Station in the Antarctic. Finally a herpesvirus-like infection has been documented in captive African penguins at the Baltimore Zoo, USA (Kincaid *et al.*, 1988) and an Eastern equine

**Table 1.** Summary of viruses with sequence information identified in penguins

Virus classification	Sequencing approach	Genetic region of interest	Penguin species	Location	Reference
Family: Poxviridae, genus: Avipoxvirus	454 GS-FLX platform	VLTF-1 gene P4b region H3L region	African penguin ( <i>Spheniscus demersus</i> )	South Africa	Carulei <i>et al.</i> (2009)
Family: Poxviridae, genus: Avipoxvirus	PCR/Sanger sequencing	P4b core region	Magellanic penguin ( <i>Spheniscus magellanicus</i> )	Argentina	Kane <i>et al.</i> (2012)
Family: Paramyxoviridae, genus: Avulavirus, species: Newcastle disease virus	RT-PCR/Sanger sequencing	Matrix gene region	Adélie penguin ( <i>Pygoscelis adeliae</i> )	King George Island, Antarctica	Tornesello <i>et al.</i> (1997)
	RT-PCR/Sanger sequencing	AI virus matrix gene 15,226 bp product – novel avian paramyxovirus	Rockhopper penguin ( <i>Eudyptes chrysocome</i> )	Falkland Islands	Miller <i>et al.</i> (2010)
Family: Papillomaviridae, genus: unassigned, species: <i>Pygoscelis adeliae</i> papillomavirus 1	Illumina NGS and Sanger sequencing of cloned complete genome-recovered back-to-back-primers	Full genome	Adélie penguin ( <i>Pygoscelis adeliae</i> )	Cape Crozier, Antarctica	This study

**Table 2.** Summary of viruses identified in penguins using serology, viral culturing or histology methods

APMV, Avian paramyxovirus; EEEV, Eastern equine encephalitis virus; MVEV, Murray Valley encephalitis virus; AdeP, Adélie penguin; KinP, King penguin; RocP, Rockhopper penguin; RoyP, Royal penguin. Note: numbers in superscript in the location column identify the antigen used in serology detection in the corresponding row of the positive viral detection column.

Virus family	Positive viral detection	Penguin species	Location	Reference
<i>Paramyxoviridae</i>	Serology: haemagglutination-inhibition test for the presence of NDV-V4 <sup>1</sup> and APMV-1M <sup>2</sup>	Royal penguin ( <i>Eudyptes schlegeli</i> ), king penguin ( <i>Aptenodytes patagonicus</i> ), rockhopper penguin ( <i>Eudyptes chrysocome</i> )	Nuggets beach (RoyP), <sup>1,2</sup> Sandy Bay (RoyP), <sup>1,2</sup> Hurd point (RoyP), <sup>1,2</sup> Flat creek (RoyP), <sup>1,2</sup> Red River (RoyP), <sup>1</sup> Lusitania Bay (KinP, RocP), <sup>2</sup> Macquarie Island	Morgan <i>et al.</i> (1981)
	Serology: haemagglutination-inhibition test for the presence of NDV-V4, <sup>3</sup> APMV-3M <sup>4</sup> and APMV-8A <sup>5</sup>	Adélie penguin ( <i>Pygoscelis adeliae</i> )	Wilkes base, <sup>3,4,5</sup> Peterson Island, <sup>3,4,5</sup> Midgley Island, <sup>4,5</sup> Shirley Island, <sup>4,5</sup> Cameron Island, <sup>4,5</sup> Chappel Island, <sup>4,5</sup> and near Dumont d'Urville base, <sup>5</sup> Antarctica	Morgan & Westbury (1981)
	Viral isolation: cultured and tested; isolates APMV-7A <sup>6</sup> and APMV-8A <sup>7</sup>	Adélie penguin ( <i>Pygoscelis adeliae</i> )	Wilkes base, <sup>6</sup> Peterson Island, <sup>7</sup> Antarctica	Morgan & Westbury (1981)
	Serology: indirect enzyme immunoassay against APMV-179/78	Adélie penguin ( <i>Pygoscelis adeliae</i> )	Cape Bird, Ross Island, Antarctica	Austin & Webster (1993)
	Serology: Haemagglutination-inhibition test for the presence of AMPV-1M	Blue penguin ( <i>Eudyptula minor</i> )	Port Campbell, Australia	Morgan <i>et al.</i> (1985)
	Viral isolation and electron microscopy	Adélie penguin ( <i>Pygoscelis adeliae</i> )	Vestfol hills, Antarctica	Morgan & Westbury (1988)
	Serology to AMPV1 and AMPV3 Virus isolation: cultured and tested; isolates PV1, <sup>8</sup> PV2, <sup>8</sup> PV3, <sup>8</sup> PV, <sup>8</sup> PV5, <sup>9</sup> PV6,* PV7, <sup>9</sup> PV8 <sup>10</sup> and 78/179 <sup>11</sup>	Humboldt penguins ( <i>Spheniscus humboldti</i> ) King penguin ( <i>Eudyptes chrysolophus</i> ), royal penguin ( <i>Eudyptes schlegeli</i> ), Adélie penguin ( <i>Pygoscelis adeliae</i> )	Peru Nuggets Beach (KinP), <sup>8</sup> Hurd Point (RoyP), <sup>8</sup> Macquarie Island; Wilkes base (AdeP), <sup>9</sup> Peterson Island (AdeP), <sup>10</sup> Cape Bird (AdeP; Ross Island), <sup>11</sup> Antarctica	Smith <i>et al.</i> (2008) Alexander <i>et al.</i> (1989)
	Serology: haemagglutination-inhibition test for the presence of APMV-8 <sup>12</sup> and APMV-10 <sup>13</sup>	Rockhopper penguin ( <i>Eudyptes chrysocome</i> )	Rugges Hill, <sup>12,13</sup> Eagle Hill, <sup>13</sup> Diamond Cove, <sup>13</sup> Falkland Islands	Miller <i>et al.</i> (2010)
	Viral isolation and electron microscopy; isolate AMPV-10/Penguin/Falkland Islands/324/2007	Rockhopper penguin ( <i>Eudyptes chrysocome</i> )	Rugges Hill, Eagle Hill and Diamond Cove, Falkland Islands	Miller <i>et al.</i> (2010)
	Serology: haemagglutination-inhibition test for the presence of NDV; antiserum against NDV-B1	Adélie penguin ( <i>Pygoscelis adeliae</i> ), chinstrap penguin ( <i>Pygoscelis antarctica</i> ), gentoo penguin ( <i>Pygoscelis papua</i> )	King George Island, Antarctica	Thomazelli <i>et al.</i> (2010)
<i>Orthomyxoviridae</i>	Serology: haemagglutination-inhibition test for the presence of AIV (A/fowl/Victoria/1975; H7N6)	Adélie penguin ( <i>Pygoscelis adeliae</i> )	Peterson Island, Antarctica	Morgan and Westbury (1981)
	Serology: screening for antibodies recognizing influenza A virus subtypes A(H1), A(H3), A(H7) and A(H9)	Adélie penguin ( <i>Pygoscelis adeliae</i> ), masked penguin ( <i>Pygoscelis antarctica</i> ), gentoo penguin ( <i>Pygoscelis papua</i> )	Hope Bay, Antarctica Potter Peninsula, Antarctica Potter Peninsula, Antarctica	Baumeister <i>et al.</i> (2004)

Table 2. cont.

Virus family	Positive viral detection	Penguin species	Location	Reference
<i>Flaviviridae</i>	Serology: haemagglutination-inhibition test for the presence of MVEV	Royal penguin ( <i>Eudyptes schlegeli</i> ), king penguin ( <i>Aptenodytes patagonicus</i> ), rockhopper penguin ( <i>Eudyptes chrysocome</i> )	Nuggets (RoyP, KinP), Sandy bay (RoyP, RocP), Flat Creek (RoyP), Lusitania bay (RocP), Caroline cove (RocP), Macquarie Island	Morgan <i>et al.</i> (1981)
<i>Birmaviridae</i>	Serology: haemagglutination-inhibition test for the presence of MVEV	Blue penguin ( <i>Eudyptula minor</i> )	Port Campbell and Phillip Islands, Australia	Morgan <i>et al.</i> (1985)
<i>Herpesviridae</i>	Serology: neutralization assay bursal disease virus	Adélie penguin ( <i>Pygoscelis adeliae</i> ), emperor penguins ( <i>Aptenodytes forsteri</i> )	Auster Rookery, Mawson, Antarctica	Gardner <i>et al.</i> (1997)
<i>Togaviridae</i>	Visual symptoms and histology	Black-footed penguin ( <i>Spheniscus demersus</i> ), Baltimore Zoo, USA	Baltimore, USA	Kincaid <i>et al.</i> (1988)
	Serology: antibody detection to EEEV, PCR confirmation	African penguins ( <i>Spheniscus demersus</i> ); captive birds in an aquarium in Connecticut, USA	Connecticut, USA	Tuttle <i>et al.</i> (2005)

\*No information on host or source provided in the reference.

encephalitis viral-like infection has been noted in African penguins at an aquarium in Connecticut, USA (Tuttle *et al.*, 2005). The serological methods, antigens used in detection, penguin hosts and locations of these studies are summarized in Table 2.

Adélie penguins are sea-ice obligates, occurring only in areas where Antarctic sea ice is present for most of the year and feeding on small fish, krill and squid (Ainley, 2002). Adélie penguins spend their non-breeding period, February to October, typically in small flocks in the pack ice that surrounds the Antarctic continent. Adélie penguins that breed on Ross Island migrate north after completing the nesting season and ride the Ross Gyre in a large clockwise circulation north of the Ross Sea but south of the southern boundary of the Antarctic Circumpolar Current (Ballard *et al.*, 2010). During the austral spring and summer, from late October to early February, all but yearling penguins can be found at colonies located at ice- and snow-free areas having easy access to the ocean and to wind-driven openings in the pack ice, called polynyas (reviewed by Ainley, 2002).

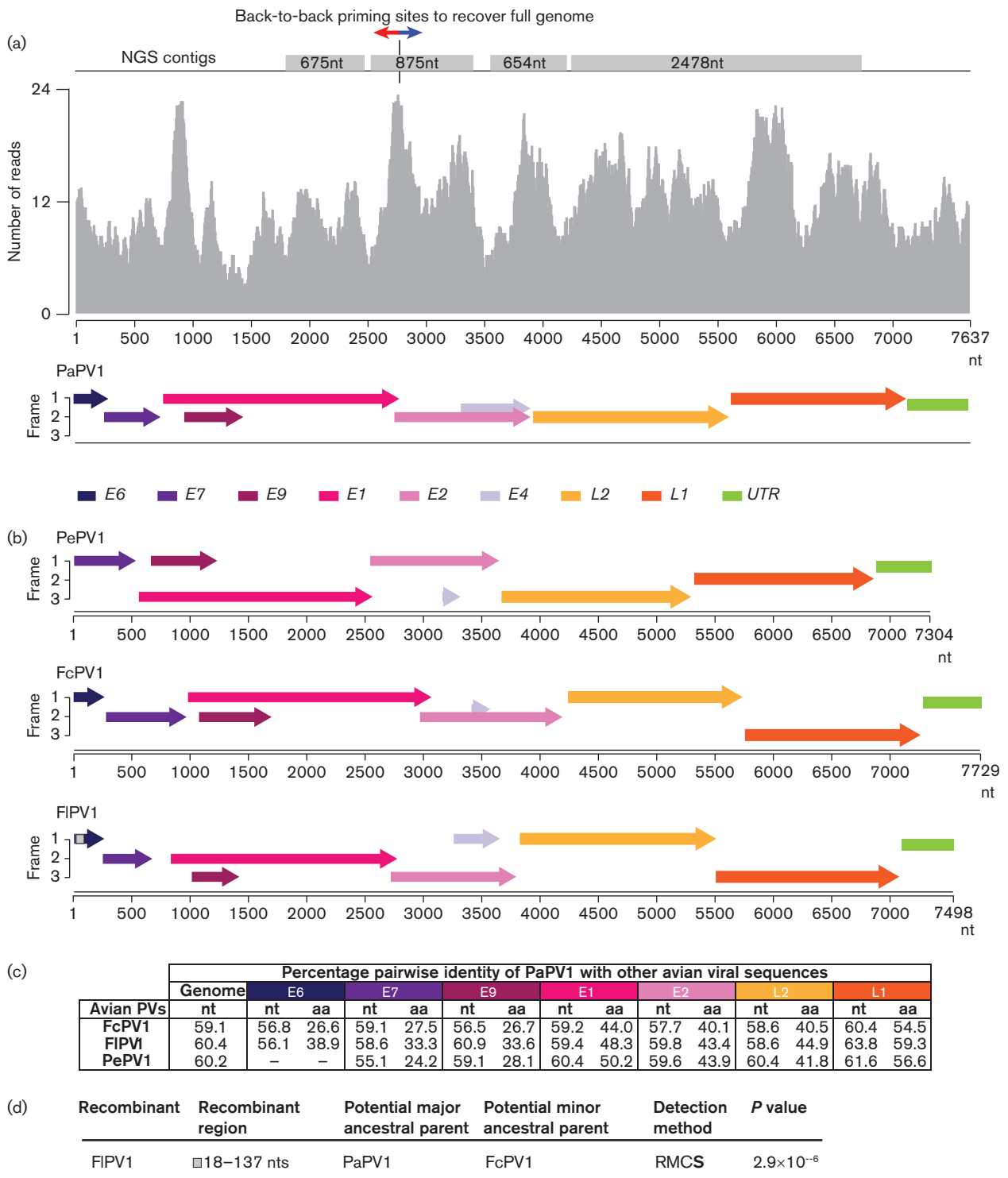
Given that there is so little known about animal viruses in the Antarctic, especially DNA viruses among penguins, and the fact that Adélie penguins exist in a unique Antarctic environment, we undertook a pilot study to identify novel DNA viruses associated with them. We opted for sampling faecal matter as it is non-invasive and we have had significant success in identifying novel circular DNA viruses in the past (Sikorski *et al.*, 2013a, b, c, d). In our pilot study we sampled faecal matter of Adélie penguins in the 2012/2013 breeding season that were nesting at Cape Crozier on Ross Island, Antarctica, one of the southernmost nesting locations for any penguin species.

## RESULTS AND DISCUSSION

### Identification and recovery of a novel papillomavirus

Two 1 m<sup>2</sup> trays were placed onto the ground in a subcolony before breeding commenced to passively collect faecal material from Adélie penguins (as part of a diet study) at Cape Crozier on Ross Island, Antarctica (Fig. S1, available in the online Supplementary Material). After processing the faecal samples for DNA viruses, NGS of the viral DNA using an Illumina sequencing platform and *de novo* sequence assembly, we identified four contigs (654, 675, 875 and 2478 nt) that had significant BLASTX (Altschul *et al.*, 1990) hits to papillomavirus proteins (Fig. 1). All four contigs had hits to avian papillomaviruses spanning the E1, E2, L1 and L2 domains. This was not surprising given that these four proteins are relatively well conserved in all papillomaviruses.

As the 875 nt contig had a significant BLASTX hit to E2 of *Francolinus leucoscepus papillomavirus* 1 (FIPV1; 75 % coverage, 73 % identity and E-value  $8 \times 10^{-52}$ ) and given that E2 is an essential papillomavirus regulatory protein (reviewed by McBride, 2013), we designed back-to-back



**Fig. 1.** (a) Genome organization of *Pygoscelis adeliae papillomavirus 1* (PaPV1) with mapping of *de novo* assembled NGS contigs and posterior mapping of paired-end reads from Illumina sequencing to the genome of PaPV1 recovered by back-to-back primers, cloned and sequenced using Sanger sequencing. The position of the primer pair is shown (interface between E1 and E2 ORFs). (b) Genome organizations of other avian papillomaviruses: *Psittacus erithacus timneh papillomavirus 1* (PePV1), *Fringilla coelebs papillomavirus 1* (FcPV1) and *Francolinus leucoscepus papillomavirus 1* (FIPV1). (c) Pairwise identities of PaPV1 with the genome, ORFs and proteins of other avian papillomaviruses. (d) Summary of recombination analysis results of avian papillomaviruses using detection methods RDP (R), GENECONV (G), BOOTSCAN (B), MAXCHI (M), CHIMAERA (C), SISCAN (S) and 3SEQ (T) implemented in the recombination detection software, RDP4. The *P* value is for the detection method highlighted in bold (S).

(abutting) primers to recover the full genome of this novel papillomavirus. Using the back-to-back primers, we recovered a complete genome, which was cloned and Sanger sequenced. The Sanger-sequenced papillomavirus genome from Adélie penguin faecal matter was 7637 nt (Fig. 1). BLASTN analysis of this genome indicated that it was most similar, albeit with only 18% coverage and sharing 67% identity, with an E-value of  $4 \times 10^{-45}$ , to the genome of *Psittacus erithacus timneh papillomavirus 1* (PePV1). However, analysis using SDT v.1.0 (Muhire *et al.*, 2013) showed that the novel Adélie penguin papillomavirus genome shared only ~60% genome-wide pairwise identity with *Fringilla coelebs papillomavirus 1* (FcPV1), PePV1 and FIPV1 (see Fig. 1 for the genome organizations of FcPV1, PePV1 and FIPV1).

Posterior mapping (using Geneious v.6.1.4; Biomatters, <http://www.geneious.com/>) of the paired-end Illumina sequencing reads to the complete genome sequence of the papillomavirus derived from PCR amplification, cloning and Sanger sequencing showed complete but variable coverage (mean of 12-fold coverage; Fig. 1). It is worth noting that NGS is a valuable tool for viral diagnostics when working with environmental samples and especially with viral sequences that are novel, but *de novo* assembly of the reads may not yield complete genomes. Hence, it is essential to use the data from the *de novo* assembly to devise strategies to recover and sequence complete genomes. We

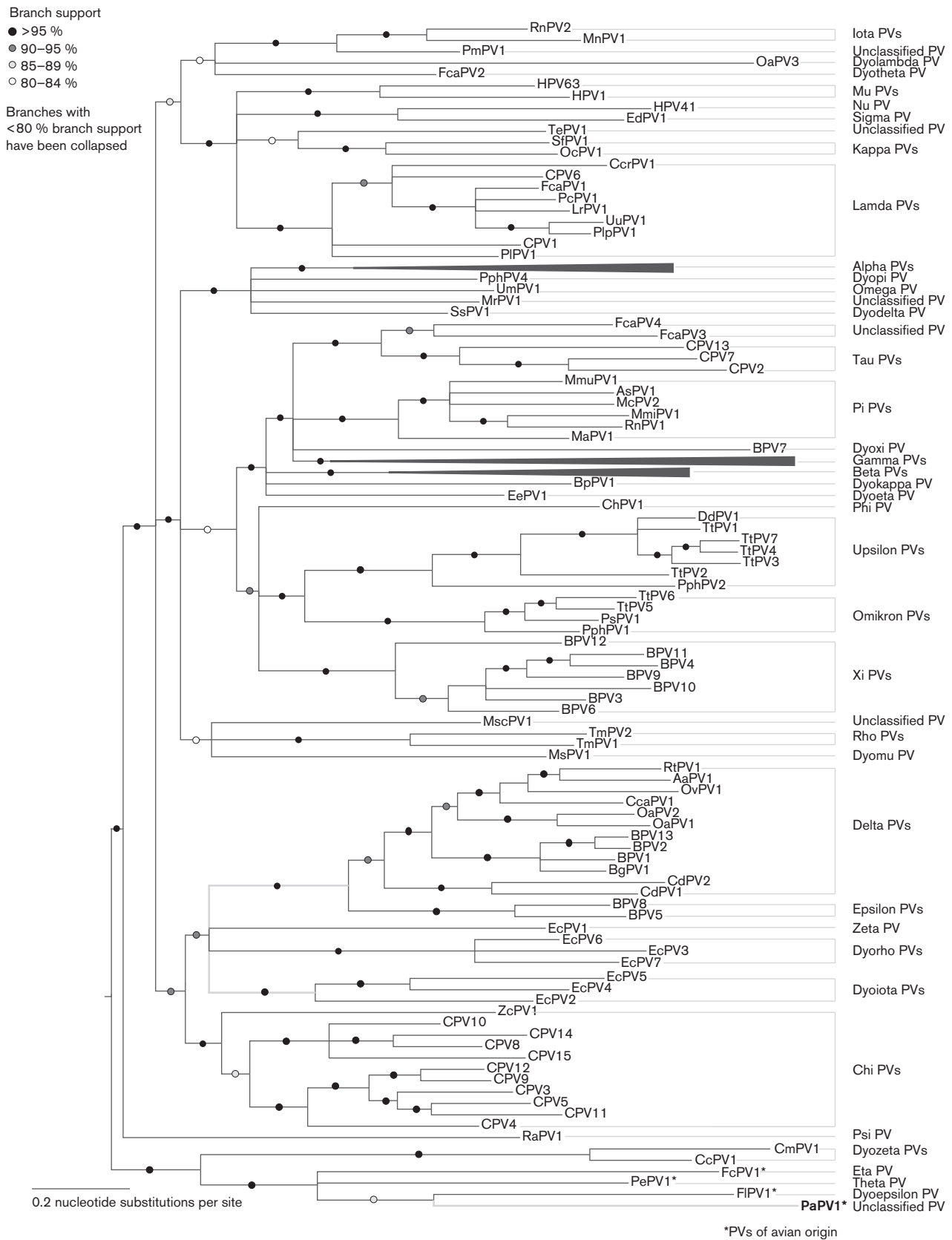
have previously successfully employed a strategy of using NGS *de novo* assembly contig-informed recovery (PCR with proofreading DNA polymerase) of complete ssDNA viruses (Dayaram *et al.*, 2013, 2014; Kraberger *et al.*, 2013; Sikorski *et al.*, 2013b, d). We employed the same strategy in this study to recover the novel papillomavirus. Furthermore, the advantage of this strategy is that, by archiving the cloned viral genome, it can be used downstream for future research.

### Genomic elements of the novel papillomavirus from Adélie penguin faecal matter

The Adélie penguin faeces-derived papillomavirus genome had eight ORFs, which encoded proteins with significant homology to the E6 (288 nt), E7 (489 nt), E9 (513 nt), E1 (2028 nt), E2 (1179 nt), E4 (639 nt), L2 (1674 nt) and L1 (1521 nt) proteins of avian papillomaviruses (Fig. 1). Among the avian viruses, it is important to note that PePV1 does not have an ORF encoding E6 (Fig. 1). Furthermore, the E9 gene has only been identified in avian papillomaviruses (Tachezy *et al.*, 2002; Terai *et al.*, 2002; Van Doorslaer *et al.*, 2009). Within the genome, we identified seven E2-binding sites (ACCN<sub>5-7</sub>GGT), four Sp1-binding sites (GGGCGG), four Nf1-binding sites (TTGGC) and two AP1-binding sites (TGANTCA) (Fig. S2). In the E1 and E2 protein sequences, we identified the putative leucine zipper domain (LDNHQSLLENEQQLQDANHL and LERLITDLQ-SKETDLEKEHQTL, respectively; leucines shown in bold)

PaPV1-E6	M--P-FPVTLGQLLYNTRRDLDDICIL	CRMCHTPLTAEIILQRELEDRESNQ-----WHRAPFCGVHTDRDQVLWLSL	[80]
FIPV1-E6	MSIP-IPYTLQVCKTAGKNLSEITIR	CRYCSGVLSVLEIILQVEEAAETRDRLYPGFWARQKGGIQ-----WVSTC	[80]
FcPV1-E6	MDIKRVPVTFSTILRWGKPKKVKVT	CKYCRVRLSLELQLEEDSRASQKLFLEIARS-GEKILTR-----YRSGC	[80]
PaPV1-E6		<b>NSC</b> ITRQRRRNRCNRYAYPRFRP--	[108]
FIPV1-E6		<b>AKC</b> -----RLFLTNA	[108]
FcPV1-E6		<b>CRC</b> -----RLFVD--	[108]
<b>CX<sub>2</sub>CX<sub>41-43</sub>CX<sub>2</sub>C metal-binding motif</b>			
PaPV1-E7	MLTR-----	-----DPSGLRPRPFKDCDDGELSTTEELEVDLY	[80]
FIPV1-E7	MRDR-----	-----DPTALRQAEPMWL-----PDIEVVIALD	[80]
PePV1-E7	MRTM-----	-----RYHPTETDDSSPDEGE-----TGNHAVTHLL	[80]
FcPV1-E7	MRNRLPIGAQPPGGQPSNNPSENNFNPEDDWILLDTSSSGSETETAETSPSSSESAEFDWEVKLIVTGNQDPITL		[80]
PaPV1-E7	VDIIE-----	-----DSSSDDDS-----DFEGYYDCTLE---T <b>LMCQE</b> GPLSDSSSDSDSGADNEAPPPPPGA	[160]
FIPV1-E7	SDISS-----	-----DSDDDSNSTTSTDLDDAFYGGWT---FBCETGRIL <b>CSEQM</b> QPSDEDEDGSEASTVPA	[160]
PePV1-E7	MQLQE-----	-----QLHALNYPTDSDSDSTD---GEELVFRLTVEESASEDDNDRSRLSVDDVGAEVDDVEVEGAYGGVA	[160]
FcPV1-E7	QELEDLNQVLVAELGHPATVYSIQDQADGASYPPGSPTSTSTFP---FETPQGNQFTSMAVAGSSAAADPPSPFSPAS		[160]
PaPV1-E7	PQPLPHVEDPTIEELLQE--	FGNNDNDQEECG-----WTCYFCEGLNPTECFLYGAFDEEHKGMCSNCYVS-TLSDF	[240]
FIPV1-E7	EDALLHDIIV-----	-----SG-----WTCAYCASTLTAGEIDLHGALDEEECKGLCNTCYVGGRLLEHL	[240]
PePV1-E7	SDN <b>LCHH</b> SMDDPEYSGAS-VGSRPDGYDERAP-----	WKCTICGRPVTPQELATFGVVNPWNKQGVCTVCFHG-QQERF	[240]
FcPV1-E7	VVD <b>LVCHH</b> SMGDSVDVEEHLNPNNAVTPPEESGANDTEFK <b>TCISKPLTEGELDEWGLVQGE</b> ---GLCHFCGFGAGVVDF		[240]
PaPV1-E7	EE---	[245]	
FIPV1-E7	FDIDI	[245]	
PePV1-E7	NSIWG	[245]	
FcPV1-E7	FP---	[245]	
<b>LXCXE motif</b>			
<b>CX<sub>2</sub>CX<sub>21-23</sub>CX<sub>2</sub>C metal-binding motif</b>			

**Fig. 2.** Conserved zinc-binding motifs in the E6 and E7 proteins and the conserved pRB LxCxE binding motif in the E7 protein.



**Fig. 3.** ML phylogenetic tree of the nucleotide sequences (substitution model GTR+G+I4) of the major capsid gene L1 using all papillomavirus (PV) types available from PAvE. AaPV, *Alces alces papillomavirus*; AsPV, *Apodemus sylvaticus papillomavirus*; BgPV, *Bos grunniens papillomavirus*; BpPV, *Bettongia penicillata papillomavirus*; BPV, *Bos taurus*

*papillomavirus*; CcaPV, *Capreolus capreolus papillomavirus*; CcPV, *Caretta caretta papillomavirus*; CcrPV, *Crocota crocuta papillomavirus*; CdPV, *Camelus dromedarius papillomavirus*; CgPV, *Colobus guereza papillomavirus*; ChPV, *Capra hircus papillomavirus*; CmPV, *Chelonia mydas papillomavirus*; CPV, *Canis familiaris oral papillomavirus*; CPV, *Canis familiaris papillomavirus*; DdPV, *Delphinus delphis papillomavirus*; EcPV, *Equus caballus papillomavirus*; EdPV, *Erethizon dorsatum papillomavirus*; EePV, *Erinaceus europaeus papillomavirus*; FcaPV, *Felis catus papillomavirus*; FcPV, *Fringilla coelebs papillomavirus*; FIPV, *Francolinus leucoscepus papillomavirus*; HPV, *Human papillomavirus*; LrPV, *Lynx rufus papillomavirus*; MaPV, *Mesocricetus auratus papillomavirus*; McPV, *Mastomys coucha papillomavirus*; MfPV, *Macaca fascicularis papillomavirus*; MmiPV, *Micromys minutus papillomavirus*; MmPV, *Macaca mulatta papillomavirus*; MmuPV, *Mus musculus papillomavirus*; MnPV, *Mastomys natalensis papillomavirus*; MrPV, *Myotis ricketti papillomavirus*; MscPV, *Miniopterus schreibersii papillomavirus*; MsPV, *Morelia spilota papillomavirus*; OaPV, *Ovis aries papillomavirus*; OaPV, *Ovis aries papillomavirus*; OcPV, *Oryctolagus cuniculus papillomavirus*; OvPV, *Odocoileus virginianus papillomavirus*; PaPV, *Pygoscelis adeliae papillomavirus*; PcPV, *Puma concolor papillomavirus*; PePV, *Psittacus erithacus papillomavirus*; PhPV, *Papio hamadryas papillomavirus*; PlpPV, *Panthera leo persica papillomavirus*; PIPV, *Procyon lotor papillomavirus*; PmPV, *Peromyscus maniculatus papillomavirus*; PphPV, *Phocoena phocoena papillomavirus*; PpPV, *Pan paniscus papillomavirus*; PsPV, *Phocoena spinipinnis papillomavirus*; PtPV, *Pan troglodytes papillomavirus*; RaPV, *Rousettus aegyptiacus papillomavirus*; RnPV, *Rattus norvegicus papillomavirus*; RnPV, *Rattus norvegicus papillomavirus*; RtPV, *Rangifer tarandus papillomavirus*; SfPV, *Sylvilagus floridanus papillomavirus*; SsPV, *Sus scrofa domesticus papillomavirus*; TePV, *Talpa europaea papillomavirus*; TmPV, *Trichechus manatus latirostris papillomavirus*; TtPV, *Tursiops truncatus papillomavirus*; UmPV, *Ursus maritimus papillomavirus*; UuPV, *Uncia uncia papillomavirus*; ZcPV, *Zalophus californianus papillomavirus*.

and in the E2 protein, a putative DNA recognition helix sequence (GQTGQLKTIRYRLQTGPYT) was found. Within the E6 and E7 proteins, we identified the conserved zinc-binding motifs (E6: CX<sub>2</sub>CX<sub>42</sub>CX<sub>2</sub>C; E7, CX<sub>2</sub>CX<sub>21</sub>CX<sub>2</sub>C; Fig. 2). For all the avian papillomaviruses, we identified a conserved pRB LxCxE binding motif in the E7 protein (Fig. 2). It is important to note that, similar to other avian papillomaviruses, only a single zinc-binding motif was identified in the E6 of the novel papillomavirus in contrast to E6 of mammalian papillomaviruses, which have two domains (Van Doorslaer *et al.*, 2009).

### Classification of the novel papillomavirus and maximum-likelihood (ML) phylogenetic analysis

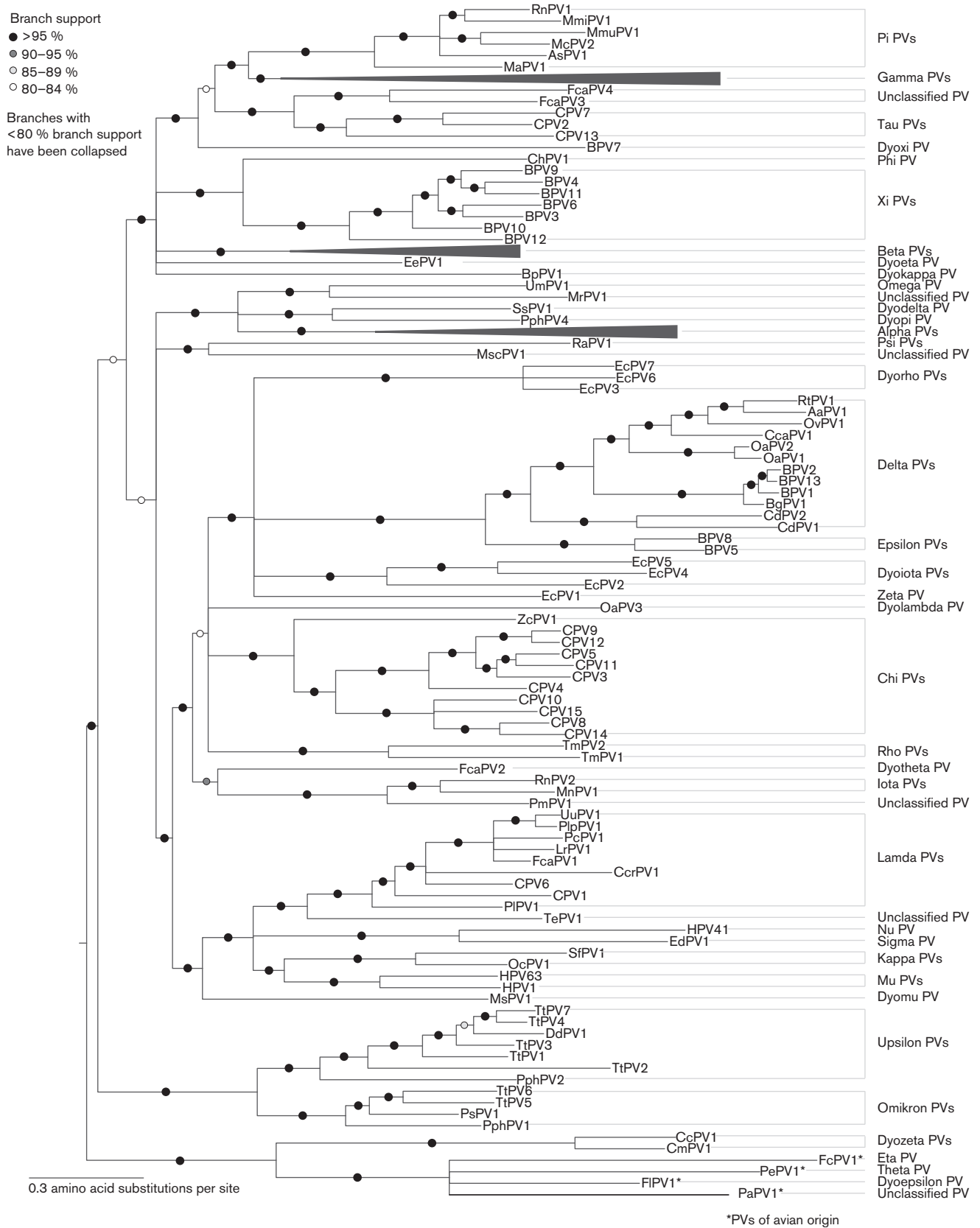
The classification of papillomaviruses is based on the sequence analysis of the major capsid protein L1, i.e. pairwise identity coupled with phylogenetic support (Bernard, 2013; Bernard *et al.*, 2010; de Villiers *et al.*, 2004). In general, L1 sequences of papillomaviruses with <60% pairwise nucleotide identity would be assigned a new genus and those with >60% but <70% would be a new species, whereas those with >70% but <90% identity would be novel types. To date, most papillomavirus pairwise identities have been determined by deriving pairwise comparisons of sequences from global alignments of L1 rather than pairwise alignments of two individual sequences, and hence the pairwise identity values will be slightly inflated for more divergent sequences (see discussions in Muhire *et al.*, 2013 and Varsani *et al.*, 2014). SDT v.1.0 aligns two sequences at a time using MUSCLE (CLUSTAL W and MAFFT options available) and determines pairwise identity (with pairwise deletion of gaps) for all possible sequence pairs within a dataset.

We calculated the pairwise identities of L1 nucleotide sequences using SDT v.1.0 (Muhire *et al.*, 2013) with MUSCLE (Edgar, 2004) pairwise alignments. The Adélie penguin papillomavirus L1 shared 63.8% pairwise identity

with the L1 of FIPV (Fig. 1) and between 55–62.5% with all other papillomavirus L1 sequences (Supplementary Data S1). However, if we calculated the pairwise identities from a global alignment (with pairwise deletion of gaps), it shared 57.6% pairwise identity with FIPV and <56.3% identity with all other papillomaviruses. Interestingly, the L1 of FIPV shared 58.9% (54.3% from global alignment) and 64.7% (58.1% from global alignment) pairwise identity to FcPV and PeCV, respectively. FIPV, FcPV and PeCV are each members of unique genera, namely *Dyoepsilon*, *Theta* and *Eta*. Therefore, based on our analysis, it is clear that the papillomavirus found in this study is a novel species, which we named *Pygoscelis adeliae papillomavirus* 1 (PaPV1), and is probably a member of a new genus. Our species proposal for PaPV1 was supported by the ML phylogenetic analysis of the L1 sequences (Fig. 3). The mid-point rooted L1 ML phylogenetic tree clearly showed that PaPV1 is part of an avian papillomavirus clade and is phylogenetically most closely related to FIPV from *Francolinus leucoscepus*.

In order to determine the evolutionary relationship of PaPV1 with other papillomaviruses, we generated amino acid concatenated alignments of E1–E2–L1–L2. ML phylogenetic (mid-point rooted) analysis of the concatenated E1–E2–L1–L2 protein sequences (Fig. 4) showed that the avian papillomaviruses are basal to all the other papillomavirus types and are monophyletic. It is interesting to note that these avian papillomaviruses share a last common ancestor with *Chelonia mydas papillomavirus* 1 and *Caretta caretta papillomavirus* 1 isolated from turtles (Herbst *et al.*, 2009). All the sauropsid papillomaviruses, with the exception of *Morelia spilota spilota papillomavirus* 1 (MsPV1), are basal to all other papillomaviruses. In general, this fits well with the co-evolution theory of papillomaviruses that has received considerable attention for decades with other support being provided (Bernard *et al.*, 2006; Rector *et al.*, 2005). However, the discovery of sequences such as MsPV1 raises some serious questions on this theory. We cannot discount the





**Fig. 4.** ML phylogenetic tree of the concatenated amino acid alignment of E1, E2, L1 and L2 proteins (substitution model LG + F) using all PV types available from PAVe. AaPV, *Alces alces papillomavirus*; AsPV, *Apodemus sylvaticus papillomavirus*; BgPV, *Bos grunniens papillomavirus*; BpPV, *Bettongia penicillata papillomavirus*; BPV, *Bos taurus papillomavirus*; CcaPV,

*Capreolus capreolus papillomavirus*; CcPV, *Caretta caretta papillomavirus*; CcrPV, *Crocota crocuta papillomavirus*; CdPV, *Camelus dromedarius papillomavirus*; CgPV, *Colobus guereza papillomavirus*; ChPV, *Capra hircus papillomavirus*; CmPV, *Chelonia mydas papillomavirus*; CPV, *Canis familiaris oral papillomavirus*; CPV, *Canis familiaris papillomavirus*; DdPV, *Delphinus delphis papillomavirus*; EcPV, *Equus caballus papillomavirus*; EdPV, *Erethizon dorsatum papillomavirus*; EePV, *Erinaceus europaeus papillomavirus*; FcaPV, *Felis catus papillomavirus*; FcPV, *Fringilla coelebs papillomavirus*; FIPV, *Francolinus leucoscepus papillomavirus*; HPV, *Human papillomavirus*; LrPV, *Lynx rufus papillomavirus*; MaPV, *Mesocricetus auratus papillomavirus*; McPV, *Mastomys coucha papillomavirus*; MfPV, *Macaca fascicularis papillomavirus*; MmiPV, *Micromys minutus papillomavirus*; MmPV, *Macaca mulatta papillomavirus*; MmuPV, *Mus musculus papillomavirus*; MnPV, *Mastomys natalensis papillomavirus*; MrPV, *Myotis ricketti papillomavirus*; MscPV, *Miniopterus schreibersii papillomavirus*; MsPV, *Morelia spilota papillomavirus*; OaPV, *Ovies aries papillomavirus*; OaPV, *Ovis aries papillomavirus*; OcPV, *Oryctolagus cuniculus papillomavirus*; OvPV, *Odocoileus virginianus papillomavirus*; PaPV, *Pygoscelis adeliae papillomavirus*; PcPV, *Puma concolor papillomavirus*; PePV, *Psittacus erithacus papillomavirus*; PhPV, *Papio hamadryas papillomavirus*; PlpPV, *Panthera leo persica papillomavirus*; PIPV, *Procyon lotor papillomavirus*; PmPV, *Peromyscus maniculatus papillomavirus*; PphPV, *Phocoena phocoena papillomavirus*; PpPV, *Pan paniscus papillomavirus*; PsPV, *Phocoena spinipinnis papillomavirus*; PtPV, *Pan troglodytes papillomavirus*; RaPV, *Rousettus aegyptiacus papillomavirus*; RnPV, *Rattus norvegicus papillomavirus*; RnPV, *Rattus norvegicus papillomavirus*; RtPV, *Rangifer tarandus papillomavirus*; SfPV, *Sylvilagus floridanus papillomavirus*; SsPV, *Sus scrofa domesticus papillomavirus*; TePV, *Talpa europaea papillomavirus*; TmPV, *Trichechus manatus latirostris papillomavirus*; TtPV, *Tursiops truncatus papillomavirus*; UmPV, *Ursus maritimus papillomavirus*; UuPV, *Uncia uncia papillomavirus*; ZcPV, *Zalophus californianus papillomavirus*.

fact that other factors such as inter- and intra-generic recombination (Angulo & Carvajal-Rodríguez, 2007; Bravo & Alonso, 2007; Carvajal-Rodríguez, 2008; García-Vallvé *et al.*, 2005; Narechania *et al.*, 2005b; Van Doorslaer, 2013; Varsani *et al.*, 2006), inter-family recombination (Woolford *et al.*, 2007), adaptive radiation and inter-species transmission (Gottschling *et al.*, 2011b; Van Doorslaer, 2013) have played a key role in papillomavirus evolution.

## Recombination

Evidence for both inter- and intra-species recombination has been provided for papillomaviruses by various research groups (Angulo & Carvajal-Rodríguez, 2007; Bravo & Alonso, 2007; Carvajal-Rodríguez, 2008; Garcia-Vallvé *et al.*, 2005; Gottschling *et al.*, 2011a; Narechania *et al.*, 2005a; Rector *et al.*, 2008; Shah *et al.*, 2010; Varsani *et al.*, 2006) and has been reviewed by Van Doorslaer (2013). Detection of inter-species and moreover inter-genera recombination is difficult due to divergent sequences and hence poor alignments.

To address this issue, the genomes of all the papillomavirus types were downloaded from the Papillomavirus Episteme Database (PAVE; <http://pave.niaid.nih.gov/>) and linearized such that they all began in roughly the same coding region (E6 or right after the untranslated region). These sequences were aligned using MAFFT (Katoh & Standley, 2013) and analysed for recombination using RDP4 (Martin *et al.*, 2010). We specifically analysed for detectable recombination events involving avian papillomavirus sequences. Given the poor alignments, we used the same approach as described by Varsani *et al.* (2006) to recheck the recombination events by realigning the recombinant and two putative parental sequences before repeating the recombination analysis. We identified a single credible recombinant region (199 nt) in the E6 gene of FIPV1 with the recombinant region being derived from ancestral FcPV1-like sequences (Fig. 1).

Based on the theory that papillomaviruses have co-evolved with their hosts and the hypothesis that they have ‘fine-tuned’ their genomes for successful infection of their hosts, it is thought that recombinant regions in papillomaviruses will not get fixed in their populations (Van Doorslaer, 2013). In contrast, recombination seems to play a significant role in ssDNA viruses in altering host range and pathogenicity, and the adaptive potential has been proven in *in vitro* evolution studies, as reviewed by Martin *et al.* (2011). Perhaps the fact that ssDNA viruses have high substitution rates of  $\sim 10^{-3}$ – $10^{-4}$  substitutions per site per year (Duffy & Holmes, 2008, 2009; Firth *et al.*, 2009; Grigoras *et al.*, 2010; Harkins *et al.*, 2009, 2014; Lefeuvre *et al.*, 2010) compared with papillomaviruses of  $7.1 \times 10^{-9}$  to  $9.6 \times 10^{-9}$  substitutions per site per year (Shah *et al.*, 2010) indicates that ssDNA viruses are recombining more frequently in order to repair the accumulation of deleterious mutations in their genomes.

## CONCLUSION

Random-primed rolling-circle replication and NGS technologies have enabled the rapid exploration of the circular DNA viral landscape. Despite the fact that viruses are perhaps the most abundant biological entity on Earth, with the majority present within the oceans, knowledge of Antarctic virology is limited. The bulk of virology in penguins has focused on serological detection of paramyxoviruses, orthomyxoviruses, flaviviruses and birnaviruses (Table 2), with limited nucleic acid characterization of avian poxviruses and paramyxoviruses (Table 1).

Through the sampling of faecal matter or rectal smears of rodents, two novel papillomaviruses, *Peromyscus maniculatus papillomavirus* 1 (host, deer mouse) and *Rattus norvegicus papillomavirus* 2 (host, Norway rat) have been discovered in the last decade (Phan *et al.*, 2011; Schulz *et al.*, 2012). Faecal sampling is a non-invasive method and hence

extremely useful in sensitive (including areas protected for unique attributes, as was our study area) and, to some extent, harsh environments such as the Antarctic. In this study, we identified a novel papillomavirus, PaPV1, in faecal matter of Adélie penguins at Cape Crozier on Ross Island, Antarctica, with PaPV1 being most closely related to the three other known avian papillomaviruses (PePV1, FIPV1 and FcPV1).

The chances of contamination by the only other common bird in the region, the South Polar skua (*Stercorarius maccormicki*), having defecated over a penguin subcolony, more specifically into a faecal tray are extremely slim. South polar skuas feed on Adélie penguin eggs and young chicks, with subcolony edge nests being by far the most vulnerable. In all observed cases, they take the egg or chick far away from the subcolony before feeding on it. Adult Adélie penguins actively protect their subcolonies and their nests, and hence South Polar skuas are not able to spend any time in the interior of a subcolony. Adélie penguins would seriously injure and even kill a South Polar skua if it landed and spent time in the interior of a subcolony.

As far as we know, we have not seen any cutaneous papillomas or skin carcinomas during 17 years of research at the Cape Crozier colony (Lyver *et al.*, 2014), which has about 270 000 breeding pairs of Adélie penguins. However, it must be noted that papillomaviruses have been recovered from individuals with healthy skin, including an isolate of FIPV1 (Van Doorslaer *et al.*, 2009) from a yellow-necked francolin. We speculate that the virus infects the cloaca of Adélie penguins and hence its presence in faecal material. Nonetheless, the discovery of this novel papillomavirus will enable scientists to develop specific probes to screen for PaPV1 infection in Adélie penguin colonies (and possibly other penguin species) and will provide insights into its epidemiology.

The fact that the genome sequences of the papillomaviruses from Adélie penguin (faeces), African grey parrot (cutaneous wart), common chaffinch (cutaneous papilloma) and yellow-necked francolin (non-symptomatic skin) are basal to all other papillomaviruses does go some way in supporting the co-evolution hypothesis. Furthermore, Van Doorslaer *et al.* (2009) noted that the avian papillomaviruses had unusual E6 and E7 ORFs. The E6 of avian papillomaviruses have only a single zinc-binding domain unlike their mammalian counterparts. Van Doorslaer *et al.* (2009) demonstrated that the mammalian C-terminal domain of E6 clusters with avian E6 providing more support for the hypothesis by Cole & Danos (1987) that a duplication in the E6 (second zinc-binding domain) took place in the 310 million years separating the mammals and birds from their common ancestor.

Antarctic virology is possibly unique, given the setting and ecosystem, and so far the amount of virus research in this region has been limited. At a genomic level, López-Bueno *et al.* (2009) discovered high viral diversity in Lake Limnopolar on Livingston Island, and recently Swanson

*et al.* (2012) characterized the genome of a novel bacteriophage in the McMurdo Dry Valley soils of Victoria Land. To our knowledge, PaPV1 is the second animal-associated viral genome to be determined from the Antarctic, the first being that of an adenovirus from deceased South Polar skuas on King George Island (Park *et al.*, 2012). Finally, there is likely to be an increase in the rate of discovery of novel viruses at the species and genus level and also in possible members of new viral families over the next decade, a trend driven primarily by new molecular techniques and new sequencing platforms.

## METHODS

**Sample collection and viral DNA isolation.** Two faecal collection trays (Fig. S1) were placed within a subcolony of Adélie penguins at Cape Crozier on Ross Island to allow the collection of faecal matter without any fine-grained substrate over the 2012/2013 breeding season. These trays were 1 m<sup>2</sup> and were constructed of a square wooden frame supporting a heavy-duty stainless steel cloth that allowed some moisture but no solid material to pass through. The trays were placed within an area where the penguins nest, and generally one or two nests were built within them; other penguins on the periphery, including their chicks, probably ejected their faeces into them. Hence, the faecal trays collected faecal material of both the adults and young from these nests for the duration of the breeding season (late October to late January).

At the end of the season (late January 2013), 50 ml faecal sample from each tray was collected in a 50 ml sterile tube using a sterile wooden spatula. Samples were frozen within 6 h of collection and transferred to the laboratory in a frozen state. Approximately 25 ml of Adélie penguin faecal matter from each tray was resuspended in 50 ml SM buffer [0.1 M NaCl, 50 mM Tris/HCl (pH 7.4), 10 mM MgSO<sub>4</sub>] and processed and homogenized as described previously (Sikorski *et al.*, 2013a, b, d). Briefly, the homogenized faecal matter was centrifuged at 10 000 g for 20 min and the resulting supernatant was filtered sequentially through 0.45 and 0.2 µm pore-sized syringe filters. In order to concentrate any virions in the solution, 3 g PEG 8000 (Sigma) was added to 20 ml filtrate and mixed gently to dissolve the PEG 8000. The faecal/15% (w/v) PEG 8000 solution was then centrifuged for 20 min at 10 000 g and the resulting pellet was resuspended in 2 ml SM buffer overnight at 4 °C. Viral DNA was purified from the resulting resuspension using a High Pure Viral Nucleic Acid kit (Roche Diagnostics).

**Viral DNA enrichment, Illumina sequencing and *de novo* sequence assembly.** Circular DNA molecules from the viral DNA purification were enriched using rolling-circle amplification with a TempliPhi kit (GE Healthcare). The RCA-enriched material together with the total purified viral DNA was sequenced at Beijing Genomics Institute (Hong Kong) on an Illumina HiSeq 2000 platform.

The paired-end sequence reads were assembled using Abyss v.1.3.5 (Simpson *et al.*, 2009) with a kmer setting of 64. The translation products of the assembled contigs (>500 nt) were analysed for homology to known protein sequences available in public databases using BLASTX (Altschul *et al.*, 1990).

**Recovery of complete papillomavirus genome and Sanger sequencing.** The novel papillomavirus was amplified using 10 µm each of the primers PaPV-F 5'-GACTTGCTAGAAAAGGAGCAT-CAGACG-3' and PaPV-R 5'-AGTCTCCTTCGACTGCAGATCTGT-TATC-3' with KAPA Long Range HotStart DNA polymerase (Kapa Biosystems) using the following protocol: initial denaturation at

95 °C for 3 min, followed by 25 cycles of 95 °C for 15 s, 60 °C for 15 s and 68 °C for 8 min, with a final extension at 68 °C for 8 min. The amplicon was resolved on a 0.7% agarose gel and the ~7.5 kb fragment was gel purified and ligated into pJET1.2 plasmid vector (Thermo Fisher). The recombinant plasmid was used to transform *Escherichia coli* DH5 $\alpha$  cells. Plasmid DNA from transformed *Escherichia coli* colonies was isolated and Sanger sequenced independently by primer walking at Macrogen (Korea). The sequence contigs were assembled using DNAbaser v.4 (Heracle BioSoft S.R.L.).

**Pairwise comparisons and ML phylogenetic analysis.** The L1 sequences of all papillomavirus type sequences available at the PAVe (<http://pave.niaid.nih.gov/>) were downloaded on 14 December 2013. These were aligned using MAFFT v.7.13 (Katoh & Standley, 2013) and the resulting alignment was used to infer a ML phylogenetic tree with PHYML (Guindon *et al.*, 2010) with GTR+I+G4 chosen as the best-fit model using jModelTest (Posada, 2009), with approximate likelihood ratio test (aLRT) (Anisimova & Gascuel, 2006) branch support. Branches with less than 79% aLRT support were collapsed using Mesquite v.2.7 (<http://mesquiteproject.org/>). The tree was mid-point rooted.

The E1, E2, L1 and L2 papillomavirus type sequences downloaded from PAVe were translated and aligned as individual datasets using MAFFT v.7.13 (Katoh & Standley, 2013). The resulting E1, E2, L1 and L2 alignments were used to generate a concatenated dataset of aligned E1, E2, L1 and L2. These were used to infer a ML phylogenetic tree with PHYML (Guindon *et al.*, 2010) using LG+F, as the amino acid substitution model chosen as the best-fit model using ProtTest v.2.4 (Abascal *et al.*, 2005); branches with less than 79% aLRT support were collapsed and the tree was mid-point rooted.

All analyses of the pairwise identities were carried out using SDT v.1.0 (Muhire *et al.*, 2013).

**Recombination analysis.** Recombination was analysed using RDP4 (Martin *et al.*, 2010), using the RDP (Martin & Rybicki, 2000), GENECONV (Padidam *et al.*, 1999), BOOTSCAN (Martin *et al.*, 2005), MAXCHI (Smith, 1992), CHIMAERA (Posada & Crandall, 2001), SISCAN (Gibbs *et al.*, 2000) and 3SEQ (Boni *et al.*, 2007) methods. Recombination was considered credible in sequences only if they were detected by more than three methods having significant *P* values coupled with strong phylogenetic support of recombination.

## ACKNOWLEDGEMENTS

The field work was supported the US National Science Foundation (NSF) under grant ANT-0944411, with logistics supplied by the US Antarctic Program. Field work was conducted under Animal Care and Use Permit #4130 through Oregon State University, Corvallis, OR, USA, and Antarctic Conservation Act Permit #2006-010 from NSF. S.K. is supported by a School of Biological Sciences (University of Canterbury, New Zealand) scholarship.

## REFERENCES

- Abascal, F., Zardoya, R. & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105.
- Ainley, D. G. (2002). *The Adélie Penguin: Bellwether of Climate Change*. New York: Columbia University Press.
- Alexander, D. J., Manvell, R. J., Collins, M. S., Brockman, S. J., Westbury, H. A., Morgan, I. & Austin, F. J. (1989). Characterization of paramyxoviruses isolated from penguins in Antarctica and sub-Antarctica during 1976–1979. *Arch Virol* **109**, 135–143.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- Angulo, M. & Carvajal-Rodríguez, A. (2007). Evidence of recombination within human alpha-papillomavirus. *Virol J* **4**, 33.
- Anisimova, M. & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* **55**, 539–552.
- Austin, F. J. & Webster, R. G. (1993). Evidence of ortho- and paramyxoviruses in fauna from Antarctica. *J Wildl Dis* **29**, 568–571.
- Baker, T. S., Newcomb, W. W., Olson, N. H., Cowser, L. M., Olson, C. & Brown, J. C. (1991). Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. *Biophys J* **60**, 1445–1456.
- Ballard, G., Toniolo, V., Ainley, D. G., Parkinson, C. L., Arrigo, K. R. & Trathan, P. N. (2010). Responding to climate change: Adélie penguins confront astronomical and ocean boundaries. *Ecology* **91**, 2056–2069.
- Baumeister, E., Leotta, G., Pontoriero, A., Campos, A., Montalti, D., Vigo, G., Pecoraro, M. & Savy, V. (2004). Serological evidences of influenza A virus infection in Antarctica migratory birds. *Int Congr Ser* **1263**, 737–740.
- Bernard, H. U. (2013). Taxonomy and phylogeny of papillomaviruses: an overview and recent developments. *Infect Genet Evol* **18**, 357–361.
- Bernard, H. U., Calleja-Macias, I. E. & Dunn, S. T. (2006). Genome variation of human papillomavirus types: phylogenetic and medical implications. *Int J Cancer* **118**, 1071–1076.
- Bernard, H.-U., Burk, R. D., Chen, Z., van Doorslaer, K., zur Hausen, H. & de Villiers, E.-M. (2010). Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* **401**, 70–79.
- Boni, M. F., Posada, D. & Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* **176**, 1035–1047.
- Bravo, I. G. & Alonso, A. (2007). Phylogeny and evolution of papillomaviruses based on the E1 and E2 proteins. *Virus Genes* **34**, 249–262.
- Carulei, O., Douglass, N. & Williamson, A. L. (2009). Phylogenetic analysis of three genes of Penguinpox virus corresponding to Vaccinia virus G8R (VLTF-1), A3L (P4b) and H3L reveals that it is most closely related to Turkeypox virus, Ostrichpox virus and Pigeonpox virus. *Virol J* **6**, 52.
- Carvajal-Rodríguez, A. (2008). Detecting recombination and diversifying selection in human alpha-papillomavirus. *Infect Genet Evol* **8**, 689–692.
- Cole, S. T. & Danos, O. (1987). Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome: phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. *J Mol Biol* **193**, 599–608.
- Dayaram, A., Goldstien, S., Zawar-Reza, P., Gomez, C., Harding, J. S. & Varsani, A. (2013). Identification of starling circovirus in an estuarine mollusc (*Amphibola crenata*) in New Zealand using metagenomic approaches. *Genome Announc* **1**, e00278-13.
- Dayaram, A., Galatowitsch, M., Harding, J. S., Argüello-Astorga, G. R. & Varsani, A. (2014). Novel circular DNA viruses identified in *Procordulia grayi* and *Xanthocnemis zealandica* larvae using metagenomic approaches. *Infect Genet Evol* **22**, 134–141.
- de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U. & zur Hausen, H. (2004). Classification of papillomaviruses. *Virology* **324**, 17–27.
- Duffy, S. & Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. *J Virol* **82**, 957–965.

- Duffy, S. & Holmes, E. C. (2009). Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *J Gen Virol* **90**, 1539–1547.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.
- Firth, C., Charleston, M. A., Duffy, S., Shapiro, B. & Holmes, E. C. (2009). Insights into the evolutionary history of an emerging livestock pathogen: porcine circovirus 2. *J Virol* **83**, 12813–12821.
- García-Vallvé, S., Alonso, A. & Bravo, I. G. (2005). Papillomaviruses: different genes have different histories. *Trends Microbiol* **13**, 514–521.
- Gardner, H., Kerry, K., Riddle, M., Brouwer, S. & Gleeson, L. (1997). Poultry virus infection in Antarctic penguins. *Nature* **387**, 245.
- Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**, 573–582.
- Gottschling, M., Bravo, I. G., Schulz, E., Bracho, M. A., Deaville, R., Jepson, P. D., Van Bressem, M.-F., Stockfleth, E. & Nindl, I. (2011a). Modular organizations of novel cetacean papillomaviruses. *Mol Phylogenet Evol* **59**, 34–42.
- Gottschling, M., Göker, M., Stamatakis, A., Bininda-Emonds, O. R., Nindl, I. & Bravo, I. G. (2011b). Quantifying the phylodynamic forces driving papillomavirus evolution. *Mol Biol Evol* **28**, 2101–2113.
- Grigoras, I., Timchenko, T., Grande-Pérez, A., Katul, L., Vetten, H. J. & Gronenborn, B. (2010). High variability and rapid evolution of a nanovirus. *J Virol* **84**, 9105–9117.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321.
- Harkins, G. W., Delpont, W., Duffy, S., Wood, N., Monjane, A. L., Owor, B. E., Donaldson, L., Saumtally, S., Triton, G. & other authors (2009). Experimental evidence indicating that mastreviruses probably did not co-diverge with their hosts. *Virol J* **6**, 104.
- Harkins, G. W., Martin, D. P., Christoffels, A. & Varsani, A. (2014). Towards inferring the global movement of beak and feather disease virus. *Virology* **450–451**, 24–33.
- Herbst, L. H., Lenz, J., Van Doorslaer, K., Chen, Z., Stacy, B. A., Wellehan, J. F., Jr, Manire, C. A. & Burk, R. D. (2009). Genomic characterization of two novel reptilian papillomaviruses, *Chelonia mydas* papillomavirus 1 and *Caretta caretta* papillomavirus 1. *Virology* **383**, 131–135.
- Kane, O. J., Uhart, M. M., Rago, V., Pereda, A. J., Smith, J. R., Van Buren, A., Clark, J. A. & Boersma, P. D. (2012). Avian pox in Magellanic penguins (*Spheniscus magellanicus*). *J Wildl Dis* **48**, 790–794.
- Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780.
- Kincaid, A. L., Bunton, T. E. & Cranfield, M. (1988). Herpesvirus-like infection in black-footed penguins (*Spheniscus demersus*). *J Wildl Dis* **24**, 173–175.
- Kraberger, S., Stainton, D., Dayaram, A., Zawar-Reza, P., Gomez, C., Harding, J. S. & Varsani, A. (2013). Discovery of *Sclerotinia sclerotiorum* hypovirulence-associated virus-1 in urban river sediments of Heathcote and Styx Rivers in Christchurch City, New Zealand. *Genome Announc* **1**, e00559-13.
- Lange, C. E., Favrot, C., Ackermann, M., Gull, J., Vetsch, E. & Tobler, K. (2011). Novel snake papillomavirus does not cluster with other non-mammalian papillomaviruses. *Virol J* **8**, 436.
- Lefevre, P., Martin, D. P., Harkins, G., Lemey, P., Gray, A. J. A., Meredith, S., Lakay, F., Monjane, A., Lett, J. M. & other authors (2010). The spread of tomato yellow leaf curl virus from the Middle East to the world. *PLoS Pathog* **6**, e1001164.
- López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A. & Alcamí, A. (2009). High diversity of the viral community from an Antarctic lake. *Science* **326**, 858–861.
- Lyver, P. O. B., Barron, M., Barton, K. J., Ainley, D. G., Pollard, A., Gordon, S., McNeil, G., Ballard, G. & Wilson, P. R. (2014). Trends in the breeding population of Adélie penguins (*Pygoscelis adeliae*) in the Western Ross Sea, 1981–2012: a coincidence of climate and resource extraction effects. *PLoS ONE* **9**, e91188.
- Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562–563.
- Martin, D. P., Posada, D., Crandall, K. A. & Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses* **21**, 98–102.
- Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefevre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462–2463.
- Martin, D. P., Biagini, P., Lefevre, P., Golden, M., Roumagnac, P. & Varsani, A. (2011). Recombination in eukaryotic single stranded DNA viruses. *Viruses* **3**, 1699–1738.
- McBride, A. A. (2013). The papillomavirus E2 proteins. *Virology* **445**, 57–79.
- Miller, P. J., Afonso, C. L., Spackman, E., Scott, M. A., Pedersen, J. C., Senne, D. A., Brown, J. D., Fuller, C. M., Uhart, M. M. & other authors (2010). Evidence for a new avian paramyxovirus serotype 10 detected in rockhopper penguins from the Falkland Islands. *J Virol* **84**, 11496–11504.
- Morgan, I. R. & Westbury, H. A. (1981). Virological studies of Adélie penguins (*Pygoscelis adeliae*) in Antarctica. *Avian Dis* **25**, 1019–1026.
- Morgan, I. R. & Westbury, H. A. (1988). Studies of viruses in penguins in the Vestfold Hills. *Hydrobiologia* **165**, 263–269.
- Morgan, I. R., Westbury, H. A., Caple, I. W. & Campbell, J. (1981). A survey of virus infection in sub-Antarctic penguins on Macquarie Island, Southern Ocean. *Aust Vet J* **57**, 333–335.
- Morgan, I. R., Westbury, H. A. & Campbell, J. (1985). Viral infections of little blue penguins (*Eudyptula minor*) along the southern coast of Australia. *J Wildl Dis* **21**, 193–198.
- Muhire, B., Martin, D. P., Brown, J. K., Navas-Castillo, J., Moriones, E., Zerbini, F. M., Rivera-Bustamante, R., Malathi, V. G., Briddon, R. W. & Varsani, A. (2013). A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus *Mastrevirus* (family *Geminiviridae*). *Arch Virol* **158**, 1411–1424.
- Narechania, A., Chen, Z., DeSalle, R. & Burk, R. D. (2005a). Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *J Virol* **79**, 15503–15510.
- Narechania, A., Terai, M. & Burk, R. D. (2005b). Overlapping reading frames in closely related human papillomaviruses result in modular rates of selection within E2. *J Gen Virol* **86**, 1307–1313.
- Padidam, M., Sawyer, S. & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218–225.
- Park, Y. M., Kim, J. H., Gu, S. H., Lee, S. Y., Lee, M. G., Kang, Y. K., Kang, S. H., Kim, H. J. & Song, J. W. (2012). Full genome analysis of a novel adenovirus from the South Polar skua (*Catharacta macrorhynchos*) in Antarctica. *Virology* **422**, 144–150.
- Phan, T. G., Kapusinszky, B., Wang, C., Rose, R. K., Lipton, H. L. & Delwart, E. L. (2011). The fecal viral flora of wild rodents. *PLoS Pathog* **7**, e1002218.

- Posada, D. (2009). Selection of models of DNA evolution with jModelTest. *Methods Mol Biol* 537, 93–112.
- Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* 98, 13757–13762.
- Rector, A. & Van Ranst, M. (2013). Animal papillomaviruses. *Virology* 445, 213–223.
- Rector, A., Van Doorslaer, K., Bertelsen, M., Barker, I. K., Olberg, R. A., Lemey, P., Sundberg, J. P. & Van Ranst, M. (2005). Isolation and cloning of the raccoon (*Procyon lotor*) papillomavirus type 1 by using degenerate papillomavirus-specific primers. *J Gen Virol* 86, 2029–2033.
- Rector, A., Stevens, H., Lacave, G., Lemey, P., Mostmans, S., Salbany, A., Vos, M., Van Doorslaer, K., Ghim, S. J. & other authors (2008). Genomic characterization of novel dolphin papillomaviruses provides indications for recombination within the *Papillomaviridae*. *Virology* 378, 151–161.
- Schulz, E., Gottschling, M., Ulrich, R. G., Richter, D., Stockfleth, E. & Nindl, I. (2012). Isolation of three novel rat and mouse papillomaviruses and their genomic characterization. *PLoS ONE* 7, e47164.
- Shah, S. D., Doorbar, J. & Goldstein, R. A. (2010). Analysis of host-parasite incongruence in papillomavirus evolution using importance sampling. *Mol Biol Evol* 27, 1301–1314.
- Shope, R. E. & Hurst, E. W. (1933). Infectious papillomatosis of rabbits: with a note on the histopathology. *J Exp Med* 58, 607–624.
- Sikorski, A., Argüello-Astorga, G. R., Dayaram, A., Dobson, R. C. & Varsani, A. (2013a). Discovery of a novel circular single-stranded DNA virus from porcine faeces. *Arch Virol* 158, 283–289.
- Sikorski, A., Dayaram, A. & Varsani, A. (2013b). Identification of a novel circular DNA Virus in New Zealand fur seal (*Arctocephalus forsteri*) fecal matter. *Genome Announc* 1, e00558-e13.
- Sikorski, A., Kearvell, J., Elkington, S., Dayaram, A., Argüello-Astorga, G. R. & Varsani, A. (2013c). Novel ssDNA viruses discovered in yellow-crowned parakeet (*Cyanoramphus auriceps*) nesting material. *Arch Virol* 158, 1603–1607.
- Sikorski, A., Massaro, M., Kraberger, S., Young, L. M., Smalley, D., Martin, D. P. & Varsani, A. (2013d). Novel myco-like DNA viruses discovered in the faecal matter of various animals. *Virus Res* 177, 209–216.
- Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Res* 19, 1117–1123.
- Smith, J. M. (1992). Analyzing the mosaic structure of genes. *J Mol Evol* 34, 126–129.
- Smith, K. M., Karesh, W. B., Majluf, P., Paredes, R., Zavalaga, C., Hoogesteijn Reul, A., Stetter, M., Braselton, W. E., Puche, H. & Cook, R. A. (2008). Health evaluation of free-ranging Humboldt penguins (*Spheniscus humboldti*) in Peru. *Avian Dis* 52, 130–135.
- Swanson, M. M., Reavy, B., Makarova, K. S., Cock, P. J., Hopkins, D. W., Torrance, L., Koonin, E. V. & Taliansky, M. (2012). Novel bacteriophages containing a genome of another bacteriophage within their genomes. *PLoS ONE* 7, e40683.
- Tachezy, R., Rector, A., Havelkova, M., Wollants, E., Fiten, P., Opendakker, G., Jenson, B., Sundberg, J. & Van Ranst, M. (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. & Burk, R. D. (2002). *Felis domesticus* papillomavirus, isolated from a skin lesion, is related to canine oral papillomavirus and contains a 1.3 kb non-coding region between the E2 and L2 open reading frames. *J Gen Virol* 83, 2303–2307.
- Terai, M., DeSalle, R. & Burk, R. D. (2002). Lack of canonical E6 and E7 open reading frames in bird papillomaviruses: *Fringilla coelebs* papillomavirus and *Psittacus erithacus timneh* papillomavirus. *J Virol* 76, 10020–10023.
- Thomazelli, L. M., Araujo, J., Oliveira, D. B., Sanfilippo, L., Ferreira, C. S., Brentano, L., Pelizari, V. H., Nakayama, C., Duarte, R. & other authors (2010). Newcastle disease virus in penguins from King George Island on the Antarctic region. *Vet Microbiol* 146, 155–160.
- Tornesello, M. L., Buonaguro, F. M., Meglio, A., Buonaguro, L., Beth-Giraldo, E. & Giraldo, G. (1997). Sequence variations and viral genomic state of human papillomavirus type 16 in penile carcinomas from Ugandan patients. *J Gen Virol* 78, 2199–2208.
- Trus, B. L., Buck, C. B., Cheng, N., Lowy, D. R., Steven, A. C. & Schiller, J. T. (2005). Localization of the HPV-16 minor capsid protein L2 by difference imaging. *Microsc Microanal* 11 (Suppl. 2), 642–643.
- Tuttle, A. D., Andreadis, T. G., Frasca, S., Jr & Dunn, J. L. (2005). Eastern equine encephalitis in a flock of African penguins maintained at an aquarium. *J Am Vet Med Assoc* 226, 2059–2062, 2003.
- Van Doorslaer, K. (2013). Evolution of the *Papillomaviridae*. *Virology* 445, 11–20.
- Van Doorslaer, K., Ould M'hamed Ould Sidi, A., Zanier, K., Rybin, V., Deryckère, F., Rector, A., Burk, R. D., Lienau, E. K., van Ranst, M. & Travé, G. (2009). Identification of unusual E6 and E7 proteins within avian papillomaviruses: cellular localization, biophysical characterization, and phylogenetic analysis. *J Virol* 83, 8759–8770.
- Varsani, A., van der Walt, E., Heath, L., Rybicki, E. P., Williamson, A. L. & Martin, D. P. (2006). Evidence of ancient papillomavirus recombination. *J Gen Virol* 87, 2527–2531.
- Varsani, A., Martin, D. P., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Murilo Zerbini, F. & Brown, J. K. (2014). Revisiting the classification of curtoviruses based on genome-wide pairwise identity. *Arch Virol*.
- Woolford, L., Rector, A., Van Ranst, M., Ducki, A., Bennett, M. D., Nicholls, P. K., Warren, K. S., Swan, R. A., Wilcox, G. E. & O'Hara, A. J. (2007). A novel virus detected in papillomas and carcinomas of the endangered western barred bandicoot (*Perameles bougainville*) exhibits genomic features of both the *Papillomaviridae* and *Polyomaviridae*. *J Virol* 81, 13280–13290.
- zur Hausen, H. (2009a). Papillomaviruses in the causation of human cancers – a brief historical account. *Virology* 384, 260–265.
- zur Hausen, H. (2009b). The search for infectious causes of human cancers: where and why. *Virology* 392, 1–10.