Virology 565 (2022) 65–72

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

Discovery of novel fish papillomaviruses: From the Antarctic to the commercial fish market

Simona Kraberger [a], Charlotte Austin [b], Kata Farkas [c], Thomas Desvignes [d], John H. Postlethwait [d], Rafaela S. Fontenele [a], Kara Schmidlin [a], Russell W. Bradley [e], Pete Warzybok [f], Koenraad Van Doorslaer [g], William Davison [c], Christopher B. Buck [h,**], Arvind Varsani [a,i,*]

[a] The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine and School of Life Sciences, Arizona State University, Tempe, AZ, 85287, USA
[b] School of Biological Sciences, University of Canterbury, Christchurch, 8140, New Zealand
[c] School of Natural Sciences, Bangor University, Bangor, LL57 2UW, UK
[d] Institute of Neuroscience, University of Oregon, Eugene, OR, 97403, USA
[e] Santa Rosa Island Research Station, California State University Channel Islands, Camarillo, CA, 93012, USA
[f] Point Blue Conservation Science, Petaluma, California, CA, 94954, USA
[g] School of Animal and Comparative Biomedical Sciences, The BIO5 Institute; Department of Immunobiology; Cancer Biology Graduate Interdisciplinary Program; UA Cancer Center, University of Arizona, Tucson, AZ, 85724, USA
[h] Lab of Cellular Oncology, National Cancer Institute, National Institutes of Health, Bethesda, MD, 20892, USA
[i] Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, 7925, Cape Town, South Africa

ARTICLE INFO

Keywords:
Papillomavirus
Trematomus bernacchii
Melanogrammus aeglefinus
Sparus aurata
Centropristis striata
Larus occidentalis

ABSTRACT

Fish papillomaviruses form a newly discovered group broadly recognized as the *Secondpapillomavirinae* subfamily. This study expands the documented genomes of the fish papillomaviruses from six to 16, including one from the Antarctic emerald notothen, seven from commercial market fishes, one from data mining of sea bream sequence data, and one from a western gull cloacal swab that is likely diet derived. The genomes of secondpapillomaviruses are ~6 kilobasepairs (kb), which is substantially smaller than the ~8 kb of terrestrial vertebrate papillomaviruses. Each genome encodes a clear homolog of the four canonical papillomavirus genes, E1, E2, L1, and L2. In addition, we identified open reading frames (ORFs) with short linear peptide motifs reminiscent of E6/E7 oncoproteins. Fish papillomaviruses are extremely diverse and phylogenetically distant from other papillomaviruses suggesting a model in which terrestrial vertebrate-infecting papillomaviruses arose after an evolutionary bottleneck event, possibly during the water-to-land transition.

1. Introduction

Papillomaviruses have circular double-stranded DNA genomes. The *Papillomaviridae* family is comprises of two subfamilies (*Firstpapillomavirinae* and *Secondpapillomavirinae*) (Van Doorslaer et al., 2018a). The *Firstpapillomavirinae* encompass >50 genera that infect mammals and other terrestrial vertebrates, including various birds (Canuti et al., 2019; Prosperi et al., 2016; Truchado et al., 2018; Van Doorslaer et al., 2017b; Varsani et al., 2014), lizards (Agius et al., 2019), and snakes (Gull et al., 2012; Kubacki et al., 2018). Some are oncogenic and can have disease outcomes that are lethal to their host.

The discovery of divergent papillomaviruses in fish prompted the establishment of a novel subfamily, *Secondpapillomavirinae*, composed of one genus, *Alefpapillomavirus*. Since the discovery of the first fish papillomavirus in a gilthead sea bream (*Sparus aurata*) (Lopez-Bueno et al., 2016), other fish papillomaviruses have been recovered from commercial market samples of haddock (*Melanogrammus aeglefinus*), red snapper (*Lutjanus campechanus*), and rainbow trout (*Oncorhynchus mykiss*) (Tisza et al., 2020), as well as farmed wels catfish (*Silurus glanis*) (Surjan et al., 2021). In the cases of the sea bream and of the wels catfish, the papillomaviruses were identified in individuals exhibiting papillomatous lesions (Lopez-Bueno et al., 2016; Surjan et al., 2021).

* Corresponding author. The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine and School of Life Sciences, Arizona State University, Tempe, AZ, 85287, USA.
** Corresponding author. Lab of Cellular Oncology, National Cancer Institute, National Institutes of Health, Bethesda, MD, 20892, USA.
E-mail addresses: buckc@mail.nih.gov (C.B. Buck), arvind.varsani@asu.edu (A. Varsani).

https://doi.org/10.1016/j.virol.2021.10.007
Received 6 September 2021; Received in revised form 23 October 2021; Accepted 25 October 2021
Available online 2 November 2021
0042-6822/

Unlike mammalian papillomavirus genomes, which are generally about 8 kilobase pairs (kb) in length (Frias-De-Diego et al., 2019), papillomaviruses associated with fish are significantly smaller, with sizes of 5.6–6 kb (Lopez-Bueno et al., 2016). Previously reported fish papillomavirus genomes contain the core four open reading frames (ORFs) with predicted protein products that include two early genes, E1 and E2, which encode protein products involved in replication), and two late genes, L1 and L2, that encode viral capsid proteins (Van Doorslaer et al., 2018a).

Here we report the identification of ten novel fish-associated papillomaviruses in an endeavor to gain further insight into the *Secondpapillomavirinae*. Genomic characterization, including ORFs with features that resemble the papillomavirus E6/E7 proteins (which we call Oncoid), and phylogenetic analyses shed light on this highly divergent group of papillomaviruses.

2. Material and methods

2.1. Sampling and processing for viral DNA

2.1.1. Antarctic fish samples

As part of a study to identify papillomaviruses circulating in Antarctic fishes, we sub-sampled four species of fish that were collected for various ongoing studies on the biology of these animals. All samples were stored at −20 °C till processed.

Two of these species were collected in the East Antarctic and two in the West Antarctic. An emerald notothen (*Trematomus bernacchii*) and a sharp-spined notothen (*Trematomus pennellii*) were captured from the Ross Sea (East Antarctica) during the austral summer of 2012–2013, and under the 2011/08R animal ethics permit (University of Canterbury, New Zealand). From these two notothens we subsampled the stomach, including its contents, liver and gills. The fish did not show any obvious external pathology. Approximately 0.5 cm^3 of each subsampled tissue was homogenized in 20 ml of SM buffer (0.1 M NaCl, 50 mM Tris/HCl – pH 7.4, 10 mM MgSO4) using a mortar and pestle. The resulting homogenate for each tissue sample was then centrifuged at 6000 g for 10 min to pellet cell debris, and the supernatant was sequentially filtered through 0.45 and then 0.22 μm syringe filters. Viral particles in the filtrate were precipitated with 15% w/v PEG 8000 overnight at 4 °C. The resulting solution was centrifuged at 6000×*g* for 20 min, the supernatant was discarded and the pellet was resuspended in 1 ml of SM buffer. 200 μl of this was used for viral DNA extraction with a High Pure Viral Nucleic Acid Kit (Roche Diagnostics, USA).

Skin samples of 3 mm x 3 mm were necropsied from eight crowned notothen (*Trematomus scotti*) and two painted notothen (*Notothenios larseni*) collected in the West Antarctic Peninsula during the austral fall of 2018 according to protocols approved by the Institutional Animal Care and Use Committees (IACUC) of the University of Oregon, USA (#13-27RRAA). Due to the small sample size of these samples, each was individually homogenized in 200 μl of SM buffer, which was then directly used to extract viral DNA using the High Pure Viral Nucleic Acid Kit (Roche Diagnostics, USA).

2.1.2. Commercial market fishes

Whole uncleaned black sea bass (*Centropristis striata*) (n = 3) and haddock (*Melanogrammus aeglefinus*) (n = 4) specimens were purchased from Maryland (USA) fish vendors between 2018 and 2019. The fishing location of these fishes is unknown. Approximately 0.5 g of scales, muscle, liver, and small samples of other tissues were combined, macerated and resuspended in 15 ml Dulbecco's PBS with Triton™ X-100 (MilliporeSigma, USA) detergent (1% w/v). To this solution and 0.02% Benzonase® Nuclease (MilliporeSigma, USA). The mixture was vortexed followed by incubation in a 37 °C water bath for 30 min. NaCl was added to a final concentration of 0.85 M and the lysate was clarified by centrifugation for 5 min at 5000×*g*. The supernatant was then transferred to a fresh siliconized tube and the centrifugation step was repeated. The resulting supernatant was used for iodixanol/OptiPrep™ (MilliporeSigma, USA) ultracentrifugal step gradients using the protocol outline in Tisza et al. (2020). From the gradient fractions, viral nucleic acids were extracted following a standard DNA extraction protocol (Tisza et al., 2020). Sequencing reads for these samples were deposited in NCBI under the BioProject PRJNA393166.

2.1.3. Western gull cloacal swab samples

Western gulls (*Larus occidentalis*) are generalist predators that have diverse diets including fish, crustaceans and other birds (Pierotti and Annett, 1995). Cloacal swabs were collected in 2011 from 42 Western gull chicks on the South Farallon Islands (part of the Farallon Islands National Wildlife Refuge) located 42 km west of San Francisco. The samples were collected as part of an avian gut microbial study. The swabs were stored in RNAlater and guanidinium isothiocyanate buffer. Viral DNA was extracted from 200 μl of the lysate using the High Pure Viral Nucleic Acid Kit (Roche Diagnostics, USA).

2.2. Enrichment of circular nucleic acids and high throughput sequencing

For each sample, 1–5 μl of the viral DNA extract were used to preferentially amplify circular DNA by rolling circle amplification (RCA) using a TempliPhi 100 kit (GE Healthcare, USA).

The RCA product was then used to generate Illumina sequencing libraries (2 × 150bp) and sequenced on either Illumina 4000 or NextSeq500 sequencers.

2.3. De novo assemblies from high-throughput sequencing and identification of fish papillomaviruses

Sequence reads were filtered for quality, trimmed using Trimmomatic v0.39 (Bolger et al., 2014) or FastP (Chen et al., 2018), and then *de novo* assembled using metaSPades v3.12 (Bankevich et al., 2012) or Megahit (Li et al., 2015). Contigs >1000 nts were analyzed and processed through the Cenote-Taker, virus discovery and annotation pipeline (Tisza et al., 2020). An un-annotated papillomavirus sequence fragment (GenBank accession number FLSL01000248) was detected in a sea bream metagenomic survey. A complete circular map for the virus was curated back to parent read sets (BioProject PRJEB7439) using CLC Genomics Workbench v.21 (http://www.clcbio.com/products/clc-genomics-workbench/).

2.4. Genome amplification and verification

2.4.1. Antarctic fish samples

A 5752 nt contig assembled from the viral DNA of the emerald notothen stomach sample displayed similarities to papillomavirus sequences. This contig was determined to be circular based on terminal redundancy resulting in a putative 5675 nt circular molecule sequence. Based on the sequence of this contig, back-to-back primers in the L1 gene region were designed (BiS_F: 5-'AAC GAC ATG CTA CTG GTA TCA GAC ATC TGG-3'; BiS_R: 5′-CAA TGA TCA TGA AGT TGG AGT CTC CAG CAT C-3′) and used in a polymerase chain reaction (PCR) to recover and verify the papillomavirus genome sequence. The PCR reaction was performed using Kapa HiFi polymerase (Kapa Biosystems, USA) as per manufacturer's recommendations. The amplicon was run on a 0.7% agarose gel, excised, purified and cloned using pJET1.2 plasmid (Thermo Fisher, USA). The recombinant plasmid was Sanger sequenced at Macrogen Inc. (South Korea) by primer walking.

No papillomavirus-like sequences were identified in sharp-spined notothen, crowned notothen or painted notothen tissue samples.

2.4.2. Western gull cloacal swab samples

A 6072 nt contig with similarities to fish papillomaviruses was identified in a Western gull cloacal swab. This contig was determined to be circular based on terminal redundancy resulting in a putative 5994 nt

circular molecule sequence. Based on the sequence of this contig, back-to-back primers in the L1 gene region were designed (W11C2_F: 5-'GTC TTC CCG TAA GAC GTG TGG CTG C-3'; W11C2_R: 5-'GCA GGC TGA CTG TGG TGA TTC TTA GGT G-3′) and used to screen the 42 western gull chick swab DNA extracts. Of the 42, only one (sample ID 3_WEGU 2011_C) was found to be positive by PCR using Kapa HiFi polymerase (Kapa Biosystems, USA) as per manufacturer's recommendations. The amplicon was run on a 0.7% agarose gel, excised, purified and cloned using pJET1.2 plasmid (Thermo Fisher, USA). The recombinant plasmid was Sanger sequenced at Macrogen Inc. (South Korea) by primer walking.

2.5. Fish papillomavirus annotation, and sequence analysis

A dataset of all previously reported fish-associated papillomaviruses (n = 6) was compiled from GenBank (downloaded on 28 June 2021). Open reading frames in the ten papillomaviruses identified in this study were first determined using ORFfinder (https://www.ncbi.nlm.nih.go

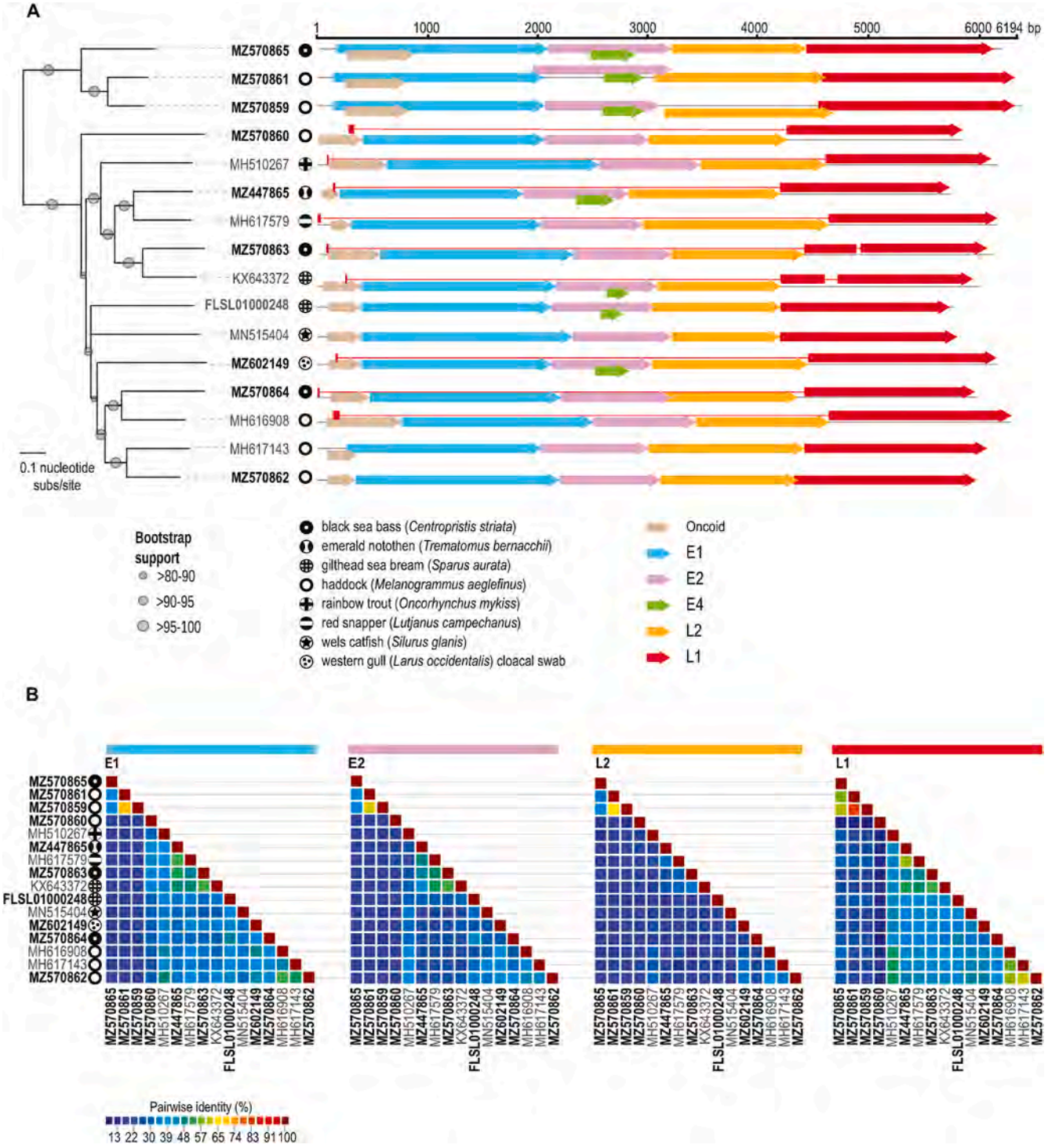

Fig. 1. Phylogenetic and genome analyses of all known fish papillomaviruses. Those recovered in this study are shown in bold. **A.** Neighbor-joining tree of all fish papillomavirus genomes and their genomic organization, with colored arrows representing inferred protein coding sequences. **B.** Pairwise comparisons of the E1, E2, L2 and L1 amino acid sequences.

v/orffinder/) with manual input for putative introns and splice acceptor/donor sites coupled with similarity searches using HHpred (Gabler et al., 2020; Zimmermann et al., 2018). MacVector v. 18.1.5 was used to scan candidate "Oncoid" proteins for the following short linear motifs: Rb interaction: (LI)XCX(ED) or (DEN)(LIMV)XX(LM)(FY)D; casein kinase 2 phosphorylation: (ST)XX(DE); cell division sequence motif (CDSM): DXXCX(TES)X_{1-8}(DE)(DETS)(DE); zinc coordination: CXXC(X_{4-40}C)XXC or CXCXXC; leucine interaction motif: LXXLLX (where X ≠ proline); PDZ interaction motifs: class 1 (STC)X(ACVILF)*, (VLIFY)X(ACVILF)*, class 3 (DE)X(ACVILF) (Kumar et al., 2020). RNA hairpins were predicted using MXfold2 http://www.dna.bio.keio.ac.jp/mxfold2/(Sato et al., 2021).

The pairwise identities of the full papillomavirus genomes and the E1, E2, L1 and L2 amino acid sequences were determined using SDT v1.2 (Muhire et al., 2014). A neighbor-joining phylogenetic tree of the complete papillomavirus genomes was generated. Genomes were aligned with MUSCLE (Edgar, 2004) inferred with Jukes-Cantor substitution model in MEGA 5.2 (Tamura et al., 2013) and midpoint rooted.

Since the E1 protein is the most conserved protein across all papillomaviruses with high similarity in the helicase domain, representative papillomavirus E1 amino acid sequences were downloaded from the Papillomavirus Episteme (PaVE) (Van Doorslaer et al., 2017a) and aligned together with all available fish papillomavirus E1 sequences using MAFFT (Katoh and Standley, 2013). The resulting alignment was trimmed using trimAL (Capella-Gutierrez et al., 2009) with gappyout function and this was then used to infer a maximum-likelihood phylogenetic tree using best substitution model LG + G4 with IQ-Tree v2.1.3 (Minh et al., 2020).

To gain better insights into the evolutionary relationships among fish papillomaviruses, the amino acid sequences of the most conserved proteins, i.e., E1, E2 and L1 of the fish and avian papillomaviruses were concatenated and "block" aligned using MUSCLE (Edgar, 2004). This alignment was then used to infer a maximum-likelihood phylogenetic tree using partition models (Chernomor et al., 2016) Q. yeast + F + I + G4 for the E1 protein, Q. pfam + F + I + G4 for the E2 protein and rtREV + F + G4 for the L1 protein in IQ-Tree v2.1.3 (Minh et al., 2020).

All phylogenetic trees were visualized in iTOL v6 (Letunic and Bork, 2019).

3. Results and discussion

3.1. Discovery of novel fish papillomaviruses

Here we present the genomes of ten new papillomaviruses from emerald notothen (n = 1), black sea bass (n = 3), haddock (n = 4), sea bream (n = 1), and from a cloacal swab sample of a western gull (n = 1).

The emerald notothen papillomavirus genome (5.6 kb; GenBank accession MZ447865) (Fig. 1A and B), was identified and recovered from the stomach tissue of an individual in whose liver we had previously identified a novel polyomavirus (Van Doorslaer et al., 2018). Simultaneous co-infection with both a papillomavirus and a polyomavirus has been noted previously in a gilt-head sea bream (Lopez--Bueno et al., 2016). No papillomaviruses were identified in the other Antarctic fish (sharp-spined, crowned and painted notothens).

Seven other papillomavirus genomes were identified in sequence data generated from commercial market fish. Three genomes were identified in black sea bass (5.7-6 kb; GenBank accession MZ570863-MZ570865) and four genomes from haddock (5.7-6.1 kb; GenBank accession MZ570859 and MZ570862) (Fig. 1A and B). An additional genome was detected in a dataset for a red snapper, but it was nearly identical to a sequence we previously reported from a haddock (MH616908). The red snapper and haddock specimens were purchased from different fish markets and were subjected to sequencing in different runs six months apart. The observation illustrates the problem that surface cross-contamination between different fish species is highly likely in the context of a fish market. It thus not possible to confidently assign host tropism for the market-derived fish papillomaviruses.

One papillomavirus more similar to those from fish compared to those from birds and other terrestrial vertebrates was identified in a cloacal swab from a chick of a seabird, the western gull (6 kb; GenBank accession MZ602149) sampled on the Farallon Islands, USA in 2011. This virus likely originated from a fish that was ingested as part of the bird's diet (Fig. 1).

A TBLASTN survey of the GenBank nr database revealed an unannotated papillomavirus-like sequence sea bream (*Sparus aurata*) dataset (5.6 kb; GenBank accession FLSL01000248; BioProject PRJEB7439; BioSample: SAMEA2826833) (Fig. 1).

To gain a better understanding of the genome organization of all the fish papillomaviruses we re-analyzed previously published fish papillomavirus sequences together with the ones we report in this study. Several papillomaviruses identified in commercial market rainbow trout, red snapper, and haddock had previously been annotated using an automated computational method, Cenote-Taker v1 (Tisza et al., 2020). Manual re-annotation confirmed the presence of the major ORFs encoding the canonical E1, E2, L2, and L1 proteins in these samples (Fig. 1).

In well-studied mammalian papillomaviruses, L1 mRNAs typically initiate within the E7 ORF and a large intron encompassing the entire early region and L2 is spliced out. This positions a highly conserved ATG codon found near the 5′ end of the L1 ORF near the 5′ end of the L1 mRNA. The L1 ORFs of many fish papillomaviruses lack a suitable ATG initiator codon at the 5' end of the L1 ORF. It has been proposed that some bird papillomaviruses, such as puffin papillomavirus 1, initiate translation of their L1 protein from a non-canonical GTG initiation codon (Canuti et al., 2019). This is a surprising proposition, in the sense that the major capsid protein must be expressed at high levels in the late phase of the papillomavirus life cycle and GTG initiation codons are not used efficiently (Kearse and Wilusz, 2017). Inspection of the puffin papillomavirus 1 map suggests the alternative explanation that a traditional ATG initiation codon might be encoded on the first exon of a hypothetical L1 mRNA, creating a novel L1 leader peptide at map positions 7600–7662. We invoke a similar solution to the L1 initiation codon puzzle for several fish papillomaviruses (Fig. 1). In some cases, the hypothetical L1 initiator ATG codon is generated by splicing.

The L2 minor capsid proteins of mammalian papillomaviruses have an N-terminal polybasic motif, which is cleaved by host furin proteases during the infectious entry process, and a conserved Cys-X_5-Cys motif (Richards et al., 2006). In all available fish papillomavirus L2 protein sequences, the familiar di-cysteine motif is found near the C-terminus and a potential polybasic furin cleavage site is located between the two cysteines (Duckert et al., 2004; Gabler et al., 2020; Zimmermann et al., 2018).

Cancer-causing HPVs encode two oncogenes, E6 and E7. A hallmark feature of E6 is a C-terminal PDZ-interaction motif and a hallmark feature of E7 oncogenes is an LXCXE motif followed by a potential casein kinase 2 (CK2) phosphorylation site that drives interactions with Rb and related tumor suppressor proteins (Suarez and Trave, 2018). The Rb-interaction motif often overlaps a cell-division sequence motif (Figge and Smith, 1988). The core fold of both E6 and E7 is anchored by sets of paired CXXC motifs that coordinate a zinc ion, supporting speculation that the two genes might have arisen through duplication of a single ancestral zinc-binding protein (Van Doorslaer, 2013). This hypothesis has been difficult to explore because E6 and E7 share no discernible primary sequence similarity with one another beyond their shared CXXC motifs. There is also staggering sequence diversity within each gene class. In an arbitrary example, the E6 proteins of *Alphapapillomavirus* HPV16 (EU118173) and *Gammapapillomavirus* HPV201 (KP692115) share only 26% amino acid identity. Moreover, the HPV201 E6 protein lacks the hallmark PDZ-interaction motif and instead encodes a hallmark pRb interaction motif that is missing from HPV201 E7 (Fig. 2).

All observed fish papillomaviruses encode at least one short ATG-initiated ORF near the 5' end of the E1 ORF (Fig. 1). In light of the

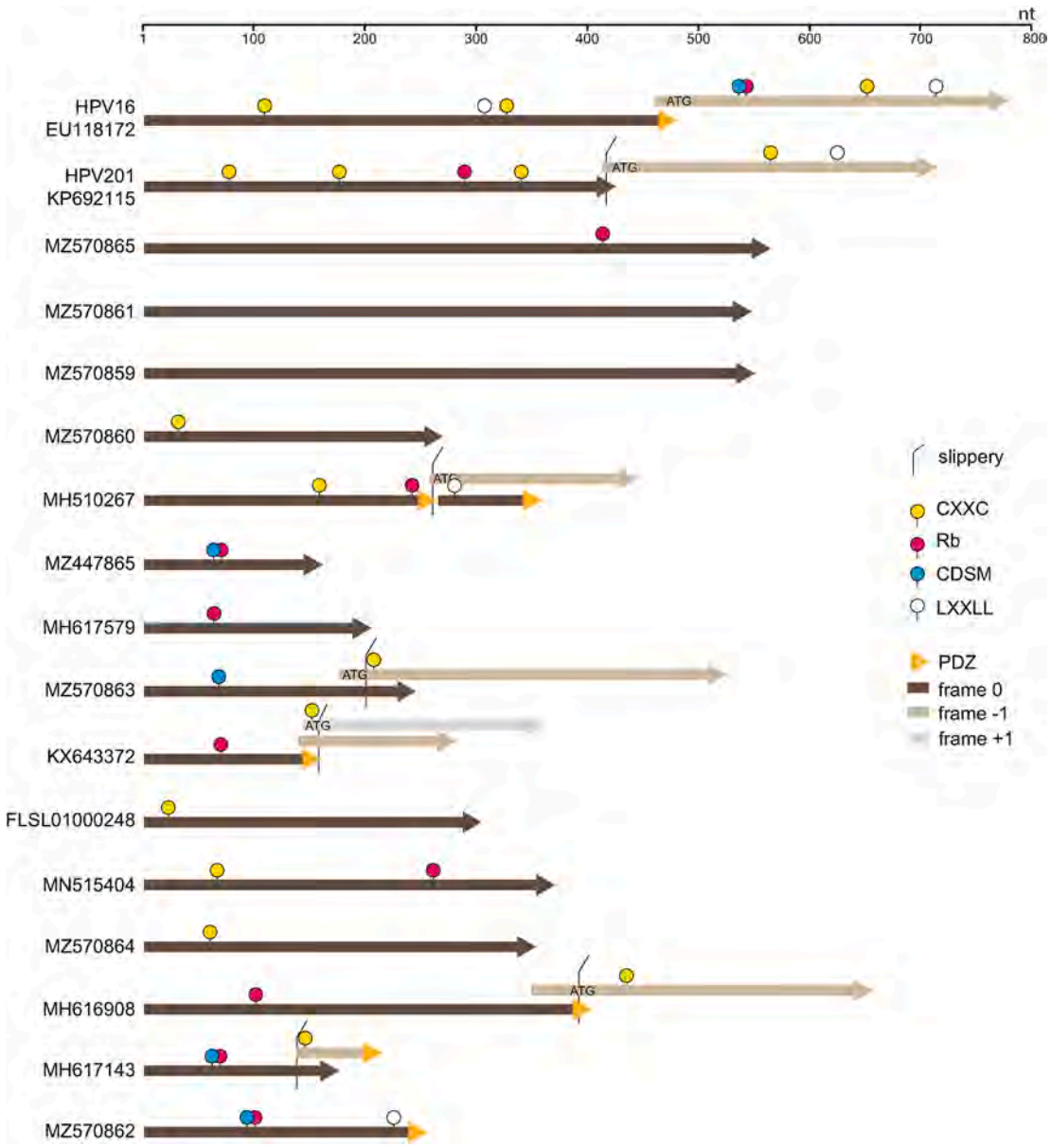

Fig. 2. Short linear motifs observed in candidate "Oncoid" proteins. Potential protein sequences encoded by open reading frames found near the 5′ end of the E1 ORF were scanned for the presence of various short linear peptide motifs of interest (see Materials and Methods for search syntax). The E6 and E7 oncoproteins of HPVs 16 and 201 are shown for reference.

poor conservation of HPV E6/E7 proteins, it was unsurprising that attempts to align protein sequences encoded by the E6/E7-syntenic fish papillomavirus ORFs against known terrestrial papillomavirus oncogenes did not reveal discernible similarities. We therefore resorted to scanning the genes for short linear peptide motifs. About half of the candidate fish papillomavirus "Oncoid" genes encode either paired CXXC motifs or a Rb-interaction motif or both (Fig. 2). In several instances (e.g., MH617143 MH617579, and MZ447865) the LXCXE-CK2-CXXC segment gives high-probability hits for the solved structure of HPV16 E7 in HHPred searches (Gabler et al., 2020; Zimmermann et al., 2018).

A haddock-associated papillomavirus, MH617143, encodes a downstream Oncoid ORF with a potential zinc-coordinating motif and a C-terminal PDZ interaction motif. There are no suitable splice acceptor signals that would potentiate expression of the downstream ORF as a second exon. In the specific cases of sea bass and sea bream papillomaviruses (MZ570863 and KX643372, respectively) the overlap between the upstream ATG-initiated Oncoid ORF and an overlapping downstream ORF have a −1 frame relationship with a TTTAAAC "slippery" motif in the overlap as well as a predicted RNA pseudoknot just downstream of the first ORF stop codon (Fig. 2). These features closely resemble the organization of the programmed −1 ribosomal frameshifting machinery that promotes expression of the fused ORF1ab polyprotein in coronaviruses (Bhatt et al., 2021; Brierley et al., 1992). We hypothesize that fish papillomavirus species with apparently split Oncoid ORFs might express a single fused Oncoid protein via translational frameshifting. Interestingly, nearly all HPVs exhibit a −1 overlap between the E6 and E7 ORFs and, in some cases, the overlap includes a canonical TTTAAAC slippery motif and a predicted RNA pseudoknot just downstream of the E6 stop codon. It would be interesting to experimentally test the hypothesis that some HPVs express a fused E6:E7 protein via programmed −1 ribosomal frameshift mechanism (Harger et al., 2002).

3.2. Phylogenetic and genetic similarity analyses

The genome-wide phylogeny showed two major fish papillomavirus clades, one containing emerald notothen papillomavirus and another comprised of two haddock-associated papillomaviruses (MZ570861 and MZ570859) and one sea bass papillomavirus (MZ570865) (Fig. 1A). A distinctive feature of the second clade is that the candidate Oncoid ORF is overprinted in the +1 frame of the E1 ORF (Fig. 1). The smaller clade comprising three viruses, with overprinted Oncoid genes, share 59–75% genome-wide similarity (Supplementary data 1). Members of the larger clade, which houses all other known fish papillomavirus genomes, share 58–66% genome-wide similarity within the group and 57–60% between members of the two groups. The E1 and L1 genes are the most conserved genes among all the terrestrial vertebrate papillomaviruses and this is also the case for the fish group (Fig. 1). The four core proteins share the following pairwise identities among the fish papillomavirus group: E1: 23–73%, E2:21–65%, L2: 17–83% and L1: 21–69% (Fig. 1B).

Analysis of E1 proteins from papillomaviruses across all vertebrate host groups revealed that papillomaviruses from fish form a distantly related group of viruses sister to avian and reptile papillomaviruses

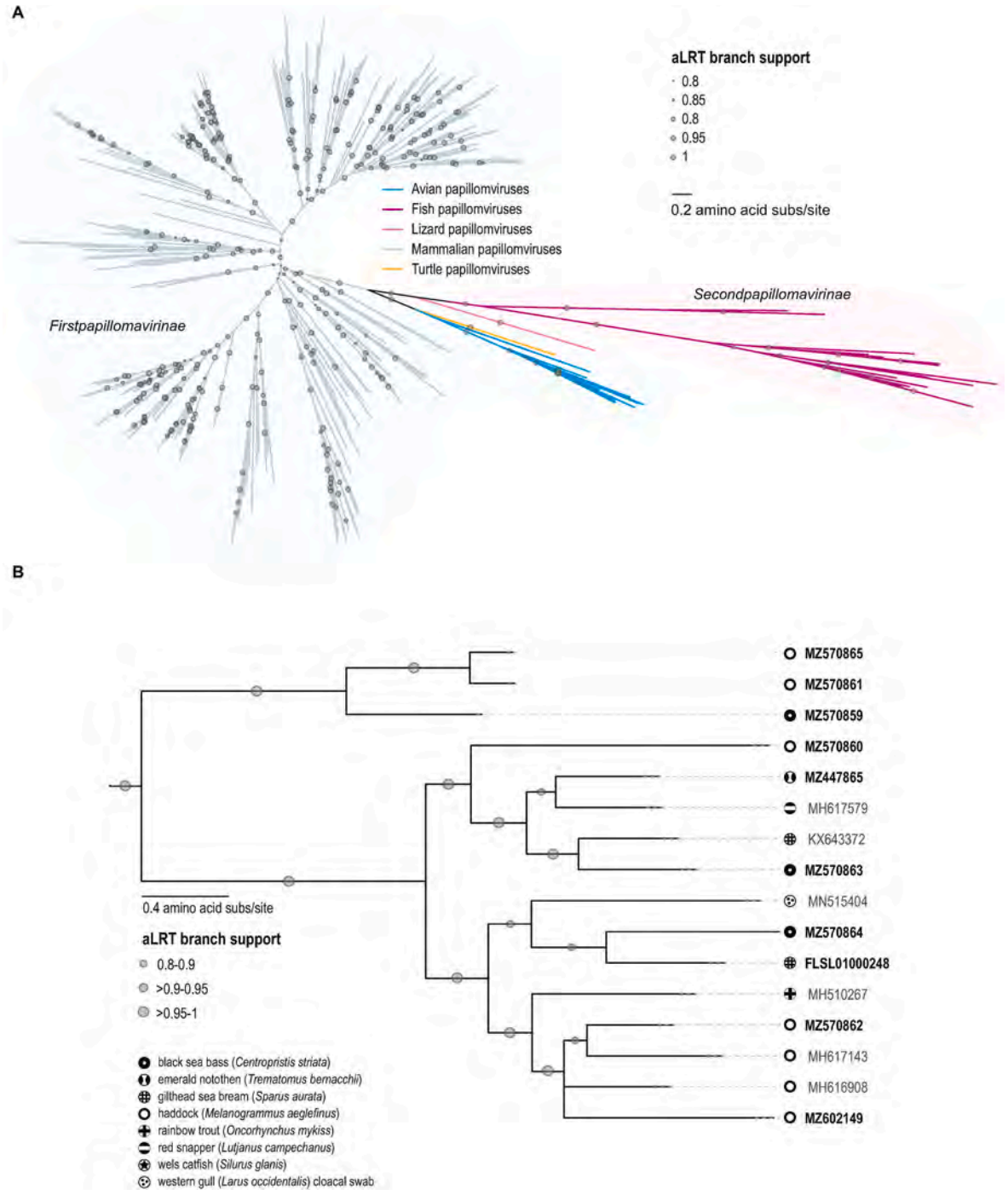

Fig. 3. Phylogenetic analyses of core fish papillomavirus proteins. **A.** Unrooted Maximum likelihood phylogeny of E1 proteins from all known papillomaviruses highlighting the highly divergent fish papillomavirus sub-family (*Secondpapillomavirinae*) compared to the terrestrial vertebrate papillomavirus sub-family (*Firstpapillomavirinae*). **B.** Maximum-likelihood phylogeny of concatenated E1, E2 and L1 amino acid sequences of the fish papillomaviruses and rooted with the avian papillomavirus sequences.

(Fig. 3A). A maximum likelihood phylogeny of the E1, E2 and L1 proteins from all fish papillomaviruses (Fig. 3B), rooted with the avian papillomaviruses, confirmed the two major fish papillomavirus clades observed in the genome-wide phylogeny (Fig. 1A). It should be noted that the Western gull cloacal swab-derived papillomavirus is most closely related to papillomaviruses from haddock, gilthead sea bream and black seabass (Figs. 1A and 3B), neither of which are found in the Farallon Island foraging area. Fish that are part of the primary diet of Western gulls are northern anchovy (*Engraulis mordax*), juvenile rockfish (*Sebastes* spp.), and other observed fish prey include Pacific whiting (*Merluccius productus*), Jack mackerel (*Trachurus symmetricus*), Pacific saury (*Cololabis saira*), midshipman (*Poricthys* spp.), white croaker (*Genyonemus lineatus*), spotted cusk eel (*Chilara taylori*), and jacksmelt (*Atherinopsis californiensis*) (Ainley, 1990). Papillomaviruses from the same host species do not always cluster together, for example the three isolates recovered from sea bass resolved in separate clades. A similar situation was also observed for the haddock papillomaviruses. However, an important caveat is that the sea bass and haddock samples were collected from a commercial market and the fish samples might have been cross contaminated by serial handling of different fish species. This caveat is less likely to apply to the two highly divergent papillomaviruses identified in single-species surveys of sea bream (KX643372 and FLSL01000248). Based on current data the fish papillomavirus phylogeny appears to not follow the host phylogeny given the multiple placements of viruses from a single host across the papillomavirus phylogeny, and therefore infers that host switching, recombination and/or multiple within host divergence events have occurred. A phenomenon that has been observed in other papillomavirus groups such as those from Weddell seals (Smeele et al., 2018).

4. Concluding remarks

Here we report nine novel fish papillomaviruses derived directly from fish tissues, one of which is the first identification of a fish papillomavirus in an Antarctic fish, and an additional papillomavirus from a western gull cloacal sample that groups with the fish papillomaviruses and was likely diet derived. These papillomaviruses belong to the *Secondpapillomavirinae* subfamily. Within an Antarctic context, this is the third animal species in which papillomaviruses have so far been identified; the others were firstpapillomaviruses found in Adélie penguin (*Pygoscelis adeliae*) (Van Doorslaer et al., 2017b; Varsani et al., 2014) and Weddell seal (*Leptonychotes weddellii*) (Smeele et al., 2018). It is apparent from the sample set described here that fishpapillomaviruses are highly diverse and likely present in diverse forms in many more, yet unsampled, fish species. Current genome sequence data suggest that all fish papillomaviruses derive from a common ancestor and are highly divergent from other papillomaviruses, however, more investigation on fish, and early diverging vertebrates such as sharks and rays as well as early branching terrestrial vertebrates such as amphibians, reptiles, and birds, is needed to gain insight into the origins, evolutionary relationships, and disease outcomes of the fish papillomaviruses in the context of the entire papillomavirus family.

CRediT authorship contribution statement

Simona Kraberger: Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Charlotte Austin:** Methodology, Investigation, Writing – review & editing. **Kata Farkas:** Methodology, Validation, Formal analysis, Investigation, Writing – review & editing. **Thomas Desvignes:** Methodology, Formal analysis, Investigation, Writing – review & editing. **John H. Postlethwait:** Methodology, Formal analysis, Investigation, Writing – review & editing. **Rafaela S. Fontenele:** Methodology, Formal analysis, Investigation, Writing – review & editing. **Kara Schmidlin:** Methodology, Investigation, Writing – review & editing. **Russell W. Bradley:** Methodology, Investigation, Writing – review & editing. **Pete Warzybok:** Methodology, Investigation, Writing – review & editing. **Koenraad Van Doorslaer:** Methodology, Investigation, Writing – review & editing. **William Davison:** Methodology, Formal analysis, Investigation, Writing – review & editing. **Christopher B. Buck:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Arvind Varsani:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to Alison McBride, and Karl Munger for their advice on papillomavirus oncogenes. The field work in the Ross Sea was supported by a grant (K057) awarded to WD from Antarctica New Zealand. The field work on the West Antarctic Peninsula was supported by the National Science Foundation (NSF) polar program grant OPP-1543383 (JHP, TD). The molecular work described in this study for the fish from Ross Sea was supported by the Center of Evolution and Medicine Venture Fund (Center of Evolution and Medicine, Arizona State University, USA) grant awarded to AV. The molecular data acquisition on fish from the West Antarctic Peninsula was supported by the NSF polar grant OPP-1947040 (awarded to TD, AV and JHP). In addition, the NSF polar grant OPP-1947040 also partially supported SK, TD, JHP, RSF and AV. The commercial market fish work was funded in part by the NIH Intramural Research Program, with support from the NCI Center for Cancer Research awarded to CB. The collection of cloacal swabs from Western gulls was conducted by Point Blue Conservation Science with the support of the California Academy of Sciences and the U.S. Fish and Wildlife Service under cooperative agreement number 81640-5-J046.

References

Agius, J.E., Phalen, D.N., Rose, K., Eden, J.S., 2019. New insights into Sauropsid Papillomaviridae evolution and epizootiology: discovery of two novel papillomaviruses in native and invasive Island geckos. Virus Evol 5 vez051.
Ainley, D.G., 1990. The Feeding Ecology of Farallon Seabirds. Seabirds of the Farallon Islands, Ecology, Dynamics and Structure of an Upwelling-System Community.
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.
Bhatt, P.R., Scaiola, A., Loughran, G., Leibundgut, M., Kratzel, A., Meurs, R., Dreos, R., O'Connor, K.M., McMillan, A., Bode, J.W., Thiel, V., Gatfield, D., Atkins, J.F., Ban, N., 2021. Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. Science 372, 1306–1313.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Brierley, I., Jenner, A.J., Inglis, S.C., 1992. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. J. Mol. Biol. 227, 463–479.
Canuti, M., Munro, H.J., Robertson, G.J., Kroyer, A.N.K., Roul, S., Ojkic, D., Whitney, H. G., Lang, A.S., 2019. New insight into avian papillomavirus ecology and evolution from characterization of novel wild bird papillomaviruses. Front. Microbiol. 10, 701.
Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.
Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.
Chernomor, O., von Haeseler, A., Minh, B.Q., 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol. 65, 997–1008.
Duckert, P., Brunak, S., Blom, N., 2004. Prediction of proprotein convertase cleavage sites. Protein Eng. Des. Sel. 17, 107–112.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinf. 5, 113.
Figge, J., Smith, T.F., 1988. Cell-division sequence motif. Nature 334, 109.
Frias-De-Diego, A., Jara, M., Escobar, L.E., 2019. Papillomavirus in Wildlife. Frontiers in Ecology and Evolution 7, 406.
Gabler, F., Nam, S.Z., Till, S., Mirdita, M., Steinegger, M., Soding, J., Lupas, A.N., Alva, V., 2020. Protein sequence analysis using the MPI bioinformatics toolkit. Curr Protoc Bioinformatics 72, e108.
Gull, J.M., Lange, C.E., Favrot, C., Dorrestein, G.M., Hatt, J.M., 2012. Multiple papillomas in a diamond python, Morelia spilota spilota. J. Zoo Wildl. Med. 43, 946–949.
Harger, J.W., Meskauskas, A., Dinman, J.D., 2002. An "integrated model" of programmed ribosomal frameshifting. Trends Biochem. Sci. 27, 448–454.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kearse, M.G., Wilusz, J.E., 2017. Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev. 31, 1717–1731.
Kubacki, J., Ramsauer, A.S., Bachofen, C., Favrot, C., Nicolier, A., Fraefel, C., Tobler, K., 2018. Complete genome sequence of a Boa (Boa constrictor)-specific papillomavirus type 1 isolate. Microbiol Resour Announc 7 e01159-18.
Kumar, M., Gouw, M., Michael, S., Samano-Sanchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J.A., Bukirova, D., Calyseva, J., Palopoli, N., Davey, N.E., Chemes, L.B., Gibson, T.J., 2020. ELM-the eukaryotic linear motif resource in 2020. Nucleic Acids Res. 48, D296–D306.
Letunic, I., Bork, P., 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47, W256–W259.
Li, D., Liu, C.M., Luo, R., Sadakane, K., Lam, T.W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676.
Lopez-Bueno, A., Mavian, C., Labella, A.M., Castro, D., Borrego, J.J., Alcami, A., Alejo, A., 2016. Concurrence of iridovirus, polyomavirus, and a unique member of a new group of fish papillomaviruses in lymphocystis disease-affected gilthead sea bream. J. Virol. 90, 8768–8779.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534.
Muhire, B.M., Varsani, A., Martin, D.P., 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9, e108277.
Pierotti, R.J., Annett, C.A., 1995. Western Gull: Larus Occidentalis. American Ornithologists, Union.
Prosperi, A., Chiari, M., Zanoni, M., Gallina, L., Casa, G., Scagliarini, A., Lavazza, A., 2016. Identification and characterization of Fringilla coelebs papillomavirus 1 (FcPV1) in free-living and captive birds in Italy. J. Wildl. Dis. 52, 756–758.
Richards, R.M., Lowy, D.R., Schiller, J.T., Day, P.M., 2006. Cleavage of the papillomavirus minor capsid protein, L2, at a furin consensus site is necessary for infection. Proc. Natl. Acad. Sci. U. S. A. 103, 1522–1527.
Sato, K., Akiyama, M., Sakakibara, Y., 2021. RNA secondary structure prediction using deep learning with thermodynamic integration. Nat. Commun. 12, 941.
Smeele, Z.E., Burns, J.M., Van Doorsaler, K., Fontenele, R.S., Waits, K., Stainton, D., Shero, M.R., Beltran, R.S., Kirkham, A.L., Berngartt, R., Kraberger, S., Varsani, A., 2018. Diverse papillomaviruses identified in Weddell seals. J. Gen. Virol. 99, 549–557.
Suarez, I., Trave, G., 2018. Structural insights in multifunctional papillomavirus oncoproteins. Viruses 10, 37.
Surjan, A., Fonagy, E., Eszterbauer, E., Harrach, B., Doszpoly, A., 2021. Complete genome sequence of a novel fish papillomavirus detected in farmed wels catfish (Silurus glanis). Arch. Virol. 166, 2603–2606.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
Tisza, M.J., Pastrana, D.V., Welch, N.L., Stewart, B., Peretti, A., Starrett, G.J., Pang, Y.S., Krishnamurthy, S.R., Pesavento, P.A., McDermott, D.H., Murphy, P.M., Whited, J.L., Miller, B., Brenchley, J., Rosshart, S.P., Rehermann, B., Doorbar, J., Ta'ala, B.A., Pletnikova, O., Troncoso, J.C., Resnick, S.M., Bolduc, B., Sullivan, M.B., Varsani, A., Segall, A.M., Buck, C.B., 2020. Discovery of several thousand highly diverse circular DNA viruses. Elife 9, e51971.
Truchado, D.A., Moens, M.A.J., Callejas, S., Perez-Tris, J., Benitez, L., 2018. Genomic characterization of the first oral avian papillomavirus in a colony of breeding canaries (Serinus canaria). Vet. Res. Commun. 42, 111–120.
Van Doorslaer, K., 2013. Evolution of the papillomaviridae. Virology 445, 11–20.
Van Doorslaer, K., Chen, Z., Bernard, H.U., Chan, P.K.S., DeSalle, R., Dillner, J., Forslund, O., Haga, T., McBride, A.A., Villa, L.L., Burk, R.D., Ictv report, C., 2018a. ICTV virus taxonomy profile: papillomaviridae. J. Gen. Virol. 99, 989-990.
Van Doorslaer, K., Kraberger, S., Austin, C., Farkas, K., Bergeman, M., Paunil, E., Davison, W., Varsani, A., 2018. Fish polyomaviruses belong to two distinct evolutionary lineages. J. Gen. Virol. 99, 567–573.
Van Doorslaer, K., Li, Z., Xirasagar, S., Maes, P., Kaminsky, D., Liou, D., Sun, Q., Kaur, R., Huyen, Y., McBride, A.A., 2017a. The Papillomavirus Episteme: a major update to the papillomavirus sequence database. Nucleic Acids Res. 45, D499–D506.
Van Doorslaer, K., Ruoppolo, V., Schmidt, A., Lescroel, A., Jongsomjit, D., Elrod, M., Kraberger, S., Stainton, D., Dugger, K.M., Ballard, G., Ainley, D.G., Varsani, A., 2017b. Unique genome organization of non-mammalian papillomaviruses provides insights into the evolution of viral early proteins. Virus Evol 3, vex027.
Varsani, A., Kraberger, S., Jennings, S., Porzig, E.L., Julian, L., Massaro, M., Pollard, A., Ballard, G., Ainley, D.G., 2014. A novel papillomavirus in Adelie penguin (Pygoscelis adeliae) faeces sampled at the Cape Crozier colony, Antarctica. J. Gen. Virol. 95, 1352–1365.
Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A.N., Alva, V., 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J. Mol. Biol. 430, 2237–2243.