AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Concurrence of Iridovirus, Polyomavirus, and a Unique Member of a New Group of Fish Papillomaviruses in Lymphocystis Disease-Affected Gilthead Sea Bream

Alberto López-Bueno,[a] Carla Mavian,[a]* Alejandro M. Labella,[b] Dolores Castro,[b] Juan J. Borrego,[b] Antonio Alcami,[a] Alí Alejo[c]

Centro de Biología Molecular Severo Ochoa (Consejo Superior de Investigaciones Científicas and Universidad Autónoma de Madrid), Cantoblanco, Madrid, Spain[a]; Universidad de Málaga, Departamento de Microbiología, Campus Universitario Teatinos, Málaga, Spain[b]; Centro de Investigación en Sanidad Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Valdeolmos, Madrid, Spain[c]

ABSTRACT

Lymphocystis disease is a geographically widespread disease affecting more than 150 different species of marine and freshwater fish. The disease, provoked by the iridovirus lymphocystis disease virus (LCDV), is characterized by the appearance of papilloma-like lesions on the skin of affected animals that usually self-resolve over time. Development of the disease is usually associated with several environmental factors and, more frequently, with stress conditions provoked by the intensive culture conditions present in fish farms. In gilthead sea bream (*Sparus aurata*), an economically important cultured fish species in the Mediterranean area, a distinct LCDV has been identified but not yet completely characterized. We have used direct sequencing of the virome of lymphocystis lesions from affected *S. aurata* fish to obtain the complete genome of a new LCDV-Sa species that is the largest vertebrate iridovirus sequenced to date. Importantly, this approach allowed us to assemble the full-length circular genome sequence of two previously unknown viruses belonging to the papillomaviruses and polyomaviruses, termed Sparus aurata papillomavirus 1 (SaPV1) and Sparus aurata polyomavirus 1 (SaPyV1), respectively. Epidemiological surveys showed that lymphocystis disease was frequently associated with the concurrent appearance of one or both of the new viruses. SaPV1 has unique characteristics, such as an intron within the L1 gene, and as the first member of the *Papillomaviridae* family described in fish, provides evidence for a more ancient origin of this family than previously thought.

IMPORTANCE

Lymphocystis disease affects marine and freshwater fish species worldwide. It is characterized by the appearance of papilloma-like lesions on the skin that contain heavily enlarged cells (lymphocysts). The causative agent is the lymphocystis disease virus (LCDV), a large icosahedral virus of the family *Iridoviridae*. In the Mediterranean area, the gilthead sea bream (*Sparus aurata*), an important farmed fish, is frequently affected. Using next-generation sequencing, we have identified within *S. aurata* lymphocystis lesions the concurrent presence of an additional LCDV species (LCDV-Sa) as well as two novel viruses. These are members of polyomavirus and papillomavirus families, and here we report them to be frequently associated with the presence of lymphocysts in affected fish. Because papillomaviruses have not been described in fish before, these findings support a more ancient origin of this virus family than previously thought and evolutionary implications are discussed.

Lymphocystis is a widely distributed disease affecting over 150 different marine and freshwater fish species (1). It was initially described during the 19th century and is characterized by the appearance of papillomalike lesions on the skin and fins, which develop over prolonged periods of time ranging from weeks to months. Occasionally, lesions on internal organs have been described, but the condition is rarely life-threatening and the outgrowths usually self-resolve, leaving apparently healthy tissue in the affected animal. At the histopathological level, the disease is characterized by the formation of heavily enlarged dermal fibroblasts termed lymphocysts that can increase their size by up to 100 times depending on the affected fish species. Thus, lymphocysts of up to 2,000 μm in diameter have been measured in affected European flounder (*Platichthys flesus*) (2). The lymphocysts are individually surrounded by a thick hyaline capsule and contain characteristic cytoplasmic basophilic inclusion bodies, granular cytoplasm, and a prominent nucleolus. Single cells may coalesce, forming larger nodules corresponding to the typical lesions observed on the fish skin, which may become vascularized or pigmented. The infectious nature of the disease was soon established, and the nature of the infectious agent as a virus was proposed in the first decade of the 1900s (3) but not confirmed until 1962, when the disease was demonstrated to be transmitted by a filterable agent (4) and the viral particle was observed by electron microscopy (5). Lym-

Received 13 July 2016 Accepted 15 July 2016

Accepted manuscript posted online 20 July 2016

Citation López-Bueno A, Mavian C, Labella AM, Castro D, Borrego JJ, Alcami A, and Alí Alejo. 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. doi:10.1128/JVI.01369-16.

Editor: S. R. Ross, University of Illinois at Chicago

Address correspondence to Alí Alejo, alejo@inia.es.

* Present address: Carla Mavian, Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, Florida, USA.

A.L.-B. and C.M. contributed equally to this work.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JVI.01369-16.

phocystis disease virus (LCDV) has since been described as a double-stranded DNA virus of nucleocytoplasmic replication with complex icosahedral particles ranging from 130 to 300 nm in diameter, depending mostly on the species of isolation, being classified within the genus *Lymphocystivirus* (family *Iridoviridae*) (6).

Due to the difficulties in propagating lymphocystiviruses in cell culture, complete genome sequences are at present available only for two distinct isolates, both obtained from fish species belonging to the Pleuronectid order. LCDV-1, the only currently recognized species within the genus (7), was isolated from diseased European flounder in northern Europe (8), and its complete genome sequence was obtained (9). A second virus, termed LCDV-C, was originally isolated from diseased, cultured Japanese flounder (*Paralichthys olivaceus*) in China. Its complete genome sequence was found to be considerably larger than that of LCDV-1 (186,250 nucleotides [nt] compared to 102,653 nt) and showed both a lack of gene order conservation and a relatively low degree of sequence identity in conserved genes, suggesting it should be considered a novel species (10). Several LCDV isolates from different hosts and locations have been described, with partial or complete sequences of the conserved major capsid protein being used to determine their phylogenetic relationships. In general, the existence of different genotypes within the lymphocystiviruses is accepted, with clustering being related to host species rather than geographic location (11).

The gilthead sea bream (*Sparus aurata*) of the order Perciformes is one of the most important cultured fish species in the Mediterranean area and East Atlantic coast of Europe. Lymphocystis disease (LCD) was first described in *S. aurata* in Israel in 1982 (12) and since then has been reported in several countries from this area (13–16). The disease follows the typical course observed in other species, with the appearance of external nodular lesions ascribed to environmental stress factors broadly associated with high-density culture conditions. Variable mortality rates associated with LCD in *S. aurata* may be attributed to secondary bacterial infections in most cases. Lymphocysts show diameters of approximately 250 μm and contain large arrays of icosahedral viral particles ranging from 210 to 235 nm (12, 17). Lesions on skin, fins, and tails usually resolve within 1 month, while viral DNA can be detected beyond this time point as well as in internal organs, including gills, liver, spleen, and kidney (18, 19), suggesting the systemic nature of infection and the establishment of an asymptomatic carrier state in recovered individuals. Genetic analyses showed that all LCDV isolates obtained from *S. aurata* (LCDV-Sa) are very closely related to each other and form a single novel genotype distinct from previously reported isolates from other hosts and locations (20).

We applied next-generation sequencing techniques to determine the complete genomic sequence of the LCDV-Sa isolate obtained from a single LCD-affected individual. Surprisingly, the viral DNA extracted from lesion-derived semipurified material unveiled LCDV-Sa as well as the complete sequence of two other abundant viruses that are novel species in the families *Papillomaviridae* and *Polyomaviridae*. Importantly, the novel papillomavirus is the first member of this family to be isolated from a fish host, supporting a more ancient evolutionary origin than previously suspected.

MATERIALS AND METHODS

Cell culture. HEK293 and Vero cell lines were obtained from the ATCC and cultured in complete Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal bovine serum at 37°C under a constant CO_2 atmosphere. Transfections were performed using Turbofect (ThermoFisher Scientific, Waltham, MA, USA) according to the manufacturer's recommendations.

Fish samples and virus purification. Juvenile gilthead sea bream specimens (5 to 10 g in weight) from different Mediterranean fish farms were sampled between 2005 and 2015. A total of 10 diseased individuals, showing typical external signs of LCD on the whole body surface, were collected during outbreaks of LCD, whereas 12 asymptomatic fish (i.e., without signs of LCD) were obtained from fish populations not previously affected by the disease. Fish were euthanized by anesthetic overdose (MS-222) (Sigma-Aldrich, St. Louis, MO, USA) before sampling. Samples of caudal fin (50 to 100 mg) were aseptically cut off, frozen immediately in dry ice, and preserved at −20°C until further use.

To obtain the sample for next-generation sequencing, one juvenile gilthead sea bream specimen collected from a lymphocystis disease outbreak in a fish farm in southeastern Spain in 2001 was used. Lymphocystis lesions were scraped under aseptic conditions and homogenized (1:10, wt/vol) in TNE buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4). After three freeze-thaw cycles, the sample was sonicated at 40 W for 2 min and centrifuged (10,000 × *g* for 10 min, 4°C) to remove cell debris, and viral particles in the supernatant were concentrated by ultracentrifugation (30,000 × *g*, 1 h, 4°C). The pellet was suspended in TNE buffer and layered over a 20% to 60% (wt/vol) continuous sucrose gradient. After centrifugation at 70,000 × *g* for 2 h at 4°C, the fraction containing the virus particles was withdrawn, ultracentrifuged (40,000 × *g*, 1 h, 4°C), and finally suspended in Tris-saline buffer (10 mM Tris-HCl, 100 mM NaCl, pH 7.4).

The nonexperimental animal research was performed according to the guidelines on animal welfare and experimentation from the European Union and Spanish Government by following the Code for Methods and Welfare Considerations in Behavioral Research with Animals (Directive 86/609EC; RD1201/2005).

Fluorescence and transmission electron microscopy. For electron microscopy analyses, lymphocystis lesions were fixed in a 2% glutaraldehyde solution in phosphate buffer (PBS) for 90 min. Postfixation was carried out with 1% OsO_4 and 1.5% $K_3Fe(CN)_6$ in water at 4°C for 1 h. Samples were dehydrated with acetone and embedded in Epoxi (TAAB 812 resin; TAAB Laboratories Ltd., Reading, United Kingdom) before 80-nm ultrathin sections were obtained. For negative staining, purified virus samples prepared as described above were applied to glow discharge carbon-coated grids and stained with a 2% uranyl acetate aqueous solution. The samples were observed under a JEM 1010 (JEOL Ltd., Tokyo, Japan) microscope and recorded using a TemCam F416 (TVIPS; Gauting, Germany) digital camera. For fluorescence microscopy, cells were washed with PBS, fixed with a neutral buffered 4% formaldehyde solution, and observed under a Zeiss Axio Vert.A1 (Zeiss, Oberkochen, Germany) microscope.

Isolation of viral DNA for next-generation sequencing. Semipurified virus particles were treated with DNase I (500 U ml^{-1}) and S7 nuclease (500 U ml^{-1}) to digest free DNA (Roche Diagnostics GmbH, Mannheim, Germany). After proteinase K treatment (0.2 mg ml^{-1}) with 0.5% SDS, viral DNA was extracted with phenol-chloroform and precipitated with sodium acetate and ethanol in the presence of 10 μg of mussel glycogen (Roche Diagnostics GmbH) as the carrier. Viral genomes were separated from low-molecular-weight DNA by electrophoresis in a 0.7% agarose gel and extracted with a QIAEX II gel extraction kit (Qiagen GmbH, Hilden, Germany). Extracted DNA (10 ng) was randomly amplified by multiple-displacement amplification for 2 h 30 min at 30°C in accordance with provider instructions (Ilustra GenomiPhi V2 DNA amplification kit; GE Healthcare, Buckinghamshire, United Kingdom).

Two micrograms of amplified viral DNA was used to construct a li-

TABLE 1 Properties of the assembled genomes from the *S. aurata* lymphocystis lesion-associated virome

System	No. of reads (mean length [bp])	Genome[a]	Length (bp)	No. of mapped reads	GC (%)	Coverage (×)	No. of reads/kbp	Relative abundance[b]
454-Roche	60,872 (348)	LCDV-Sa	208,501	30,567	33.0	51	147	1.0
		SaPyV1	7,299	3,714	52.1	177	509	3.5
		SaPV1	5,748	15,729	39.5	952	2,736	18.6
Illumina	3,512,129 (76)	LCDV-Sa	208,501	1,493,449	33.0	544	7,163	1.0
		SaPyV1	7,299	387,048	52.1	4,030	53,028	7.4
		SaPV1	5,748	1,273,919	39.5	16,844	221,628	30.9

[a] LCDV-Sa, lymphocystis disease virus-Sparus aurata; SaPyV1, Sparus aurata polyomavirus 1; SaPV1, Sparus aurata papillomavirus 1.
[b] Relative abundance compared to the number of LCVD-Sa reads/kbp.

brary using a GS-FLX Titanium system (454 Life Sciences, Roche, Branford, CT, USA) and pyrosequenced with a FLX genome sequencer at the Genomics Unit of the Scientific Park of Madrid (Spain). For Illumina sequencing, 5 μg of amplified viral DNA was used to construct a TruSeq library and sequenced with a Genome Analyzer IIx, which is also hosted in the Scientific Park of Madrid. The output consisted of 60,872 454-Roche reads (average size, 348 bp) and 3,512,129 single reads (average size, 75 bp) from Illumina (Table 1).

Metagenomic analysis, *de novo* assembly, and genome annotation. A subset of 10,000 454-Roche reads longer than 400 bp were compared by tBLASTx with a database containing all viral genomes available in GenBank (downloaded on 18 September 2015). 454-Roche and Illumina reads were *de novo* assembled under stringent parameters (97% identity over 90% overlapping) with Newbler 2.5.3 (Roche, USA) and the *de novo* assembler of CLC Genomics Workbench (trial version), respectively. A large scaffold of 208,501 bp was generated, combining contigs from both technologies. The final LCDV-Sa sequence was obtained by mapping Illumina reads against this scaffold by using Bowtie-2 with default parameters (http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml).

The genome was annotated with the Genome Annotation Transfer Utility (21) and Artemis (22) software using a curated genome of LCDV-C (23) as a template and further refined manually by using the similarity search algorithm BLASTp (http://blast.ncbi.nlm.nih.gov/) on all unassigned open reading frames (ORFs) longer than 100 bp. A circular genome map was drawn with DNA Plotter, and a circular coverage profile was created with the graphics package Matplotlib implemented in Python (24). Nucleotide-to-nucleotide comparisons between LCDV-Sa, LCDV-C (NC_005902.1) and LCDV-1 (NC_001824.1) genomes were calculated using Gepard-1.4 (25) software under default settings. The BLAST-based alignment algorithm PAirwise Sequence Comparison (PASC) was used to estimate the overall degree of nucleotide identity (26).

Two additional contigs were assembled with high coverage from Roche and Illumina reads. Their annotation and genome drawing were carried out with SnapGene (trial version) and BLASTp searches. Splicing sites in S. aurata papillomavirus 1 (SaPV1) were predicted using the NetGene2 Server (http://www.cbs.dtu.dk/services/NetGene2/).

Phylogenetic analyses. To assess the phylogenetic affiliation of LCDV-Sa, a concatenated amino acid sequence from its 26 core genes totaling 12,498 amino acids (23) was aligned with those from a set of representative iridoviruses: FV3 (frog virus 3; NC_005946.1), CMTV (common midwife toad ranavirus; JQ231222.1), EHNV (epizootic haematopoietic necrosis virus; NC_028461.1), SGIV (Singapore grouper iridovirus; NC_006549.1), LCDV-C (lymphocystis disease virus - isolate China; NC_005902.1), LCDV-1 (lymphocystis disease virus 1; NC_001824.1), RSIV (red sea bream iridovirus; AB104413.1), ISKNV (infectious spleen and kidney necrosis virus; NC_003494.1), IIV3 (invertebrate iridescent virus 3; NC_008187.1).

Phylogenetic analysis of SaPV1 was based on the amino acid alignment of L1 proteins from SaPV1 and a set from representative papillomaviruses: OaPV1 (Ovis aries papillomavirus 1; NP_044438.1), AaPV1 (Alces alces papillomavirus 1; NP_041313.1), BPV1 (bovine papillomavirus-1; NP_056744.1), BPV5 (bovine papillomavirus-5; NP_694435.1), EcPV2 (equine papillomavirus 2; YP_002635574.1), EcPV3 (Equus caballus papillomavirus 3; YP_006299864.1), EcPV1 (Equus caballus papillomavirus 1; NP_620513.1), CPV3 (canine papillomavirus 3; YP_717906.1), CPV4 (canine papillomavirus 4; YP_001648805.1), ZcPV1 (Zalophus californianus papillomavirus 1; YP_004346968.1), EdPV1 (Erethizon dorsatum papillomavirus 1; YP_224227.1), HPV41 (human papillomavirus 41; CAA39619.1), CPV1 (canine oral papillomavirus 1; NP_056819.1), PlPV1 (Procyon lotor papillomavirus 1; NC_007150.1), SfPV1 (Sylvilagus floridanus papillomavirus 1; NP_077113.1), CcanPV1 (Castor canadensis papillomavirus 1; YP_008992244.1), HPV1 (human papillomavirus 1; NP_040309.1), HPV63 (human papillomavirus 63; NC_001458.1), MnPV1 (Mastomys natalensis papillomavirus 1; NP_042019.1), BPV7 (Bos taurus papillomavirus 7; YP_406565.1), BpPV1 (Bettongia penicillata papillomavirus 1; YP_003622569.1), HPV9 (human papillomavirus 9; NP_041865.1), HPV49 (human papillomavirus 49; NP_041837.1), HPV92 (human papillomavirus 92; NP_775311.1), HPV5 (human papillomavirus 5; NP_041372.1), EePV1 (Erinaceus europaeus papillomavirus 1; YP_002427696.1), HPV50 (human papillomavirus 50; NP_043429.1), HPV88 (human papillomavirus 88; YP_001672014.1), HPV4 (human papillomavirus 4; NP_040895.1), MaPV1 (Mesocricetus auratus papillomavirus 1; YP_008720077.1), MmiPV1 (Micromys minutus papillomavirus 1; YP_873945.1), CPV2 (Canis familiaris papillomavirus 2; YP_164635.1), RaPV1 (Rousettus aegyptiacus papillomavirus 1; ABC95030.1), ChPV1 (Capra hircus papillomavirus 1; YP_610959.1), BPV3 (Bos taurus papillomavirus 3; AF486184.1), TtPV1 (Tursiops truncatus papillomavirus 1; YP_002117846.1), TtPV2 (Tursiops truncatus papillomavirus 2; YP_656480.1), PsPV1 (Phocoena spinipinnis papillomavirus; NP_542623.1), SscPV1 (Saimiri sciureus papillomavirus 1; YP_009002602.1), HPV16 (Human papillomavirus 16; AIQ82825.1), MmPV1 (Macaca mulata papillomavirus 1; NP_043338.1), HPV6 (human papillomavirus 6b; NP_040304.1), HPV18 (human papillomavirus 18; AAP20601.1), UmPV1 (Ursus maritimus papillomavirus 1; YP_001931973.1), PphPV4 (Phocoena phocoena papillomavirus 4; YP_006470639.1), SsPV1 (Sus scrofa papillomavirus 1; ABS12702.1), TmPV1 (Trichechus manatus latirostris papillomavirus 1; YP_164627.1), MsPV1 (Morelia spilota papillomavirus 1; AEO16190.1), CcPV1 (Caretta caretta papillomavirus 1; YP_002308363.1), CmPV1 (Chelonia mydas papillomavirus 1; ACD39811.1), FgPV1 (Fulmarus glacialis papillomavirus 1; YP_009041476.1), FlPV1 (Francolinus leucoscepus papillomavirus 1; ABX61091.1), PePV1 (Psittacus erithacus papillomavirus 1; NP_647590.1), PaPV1 (Pygoscelis adeliae papillomavirus; YP_009022077.1), FPV1 (Fringilla coelebs papillomavirus; NP_663767.1).

Finally, amino acid sequences of a set of representative VP1/large T antigen proteins were used to perform phylogenetic analyses of S. aurata polyomavirus 1 (SaPyV1) (accession numbers for VP1/large T antigen protein sequences are given in parentheses, with a minus indicating no sequence): Marbled_EelPyV1 (marbled eel polyomavirus; AKL71464.1/−), GiantGuitarFish_PyV1 (giant guitarfish polyomavirus; YP_009116786.1/YP_009116784.1), Trematomus_PyV1 (Trematomus polyomavirus; YP_009134738.1/YP_009134737.1), BassPyV1 (black sea bass polyomavirus; AIX88121.1/YP_009110676.1), JEECV (Japanese eel endothelial cell-infecting virus; YP_009103994.1/−), ButcherBird_PyV1 (butcher

bird polyomavirus; YP_008873519.1/YP_008873518.1), CPyV (crow polyomavirus; YP_529828.1/YP_529827.1), GHPyV (goose hemorrhagic polyomavirus; NP_849170.1/NP_849169.1), FPyV (finch polyomavirus; YP_529834.1/YP_529833.1), APyV (budgerigar fledgling disease virus 1; YP_004061429.1/YP_004061428.1), CaPyV (canary polyomavirus; YP_005454255.1/YP_005454254.1), BPyV (bovine polyomavirus; NP_040788.2/NP_040787.1), MPtV (murine pneumotropic virus; NP_041232.1/NP_041234.1), BatPyV (Myotis polyomavirus VM-2008; YP_002261489.1/YP_002261488.1), SqPyV (squirrel monkey polyomavirus; YP_001531349.1/YP_001531348.1), WUPyV (WU polyomavirus; YP_001285488.1/YP_001285487.1), KIPyV (KI polyomavirus; YP_001111259.1/YP_001111258.1), BKPyV (BK polyomavirus; YP_717940.1/YP_717939.1), SA12 (simian virus 12; YP_406555.1/YP_406554.1), JCPyV (JC polyomavirus; NP_043512.1/NP_043511.1), SV40 (simian virus 40; YP_003708382.1/YP_003708381.1), SLPyV (California sea lion polyomavirus 1; YP_003429323.1/YP_003429322.1), HPyV6 (human polyomavirus 6; YP_003848919.1/YP_003848918.1), HPyV7 (human polyomavirus 7; YP_003848924.1/YP_003848923.1), ChPyV (chimpanzee polyomavirus; YP_004046683.1/AAW12702.1), HaPyV (hamster polyomavirus; AAA67118.1/YP_009111410.1), MPyV (murine polyomavirus strain A2; AAB59901.1/AAB59902.1), TSPyV (Trichodysplasia spinulosa-associated polyomavirus; YP_003800007.1/YP_003800006.1), OraPyV1 (Bornean orangutan polyomavirus; YP_003264534.1/YP_003264533.1), LPyV (B-lymphotropic polyomavirus; YP_003429323.1/NP_848007.2), OraPyV2 (Sumatran orangutan polyomavirus; YP_009175031.1/YP_009175030.1), MCPyV (Merkel cell polyomavirus; YP_009111421.1/ABY65888.1).

Clustal Omega alignments were manually edited with Jalview (http://www.jalview.org/) to trim low conserved terminal regions from L1, large T, and VP1 proteins. The resulting alignments preserve the full-length amino acid sequence of HPV16 and most of the sequence of SV40 large T (amino acids 120 to 579) and VP1 (amino acids 28 to 344) proteins. Maximum-likelihood phylogenetic trees were constructed with bootstrap 1,000, under the LG substitution model in R (package Phangorn). The best-fit model of amino acid substitution was selected with the ProtTest 2.4 server according to the lowest Akaike information criterion (AIC) (27). Trees were drawn with Dendroscope (28).

Plasmid construction. The region containing the complete L1 gene from SaPV1 and including 143 nt before and 231 nt after the predicted initiator and stop codons of the L1 protein was PCR amplified from purified viral DNA samples using oligonucleotides L1F-143 and L1R-231. The PCR fragment was cloned using InFusion recombination technology (Clontech Laboratories, Inc., Mountain View, CA, USA) into a BamHI-digested phrGFPII-1 (Stratagene, ThermoFisher Scientific) plasmid. The resulting plasmid (pCMV-SaPV1-L1) contains the SaPV1 L1 gene under the control of a strong late cytomegalovirus (CMV) promoter. To insert the SaPV1 predicted intron sequence into the coding sequence of the fluorescent reporter hrGFP, the Q5 site-directed mutagenesis kit (New England BioLabs Inc., Ipswich, MA, USA) was used according to the provider's recommendations. Plasmid phrGFP11-1 (Stratagene) was used as a template to generate three plasmids with insertions situated after the second coding triplet of hrGFP. These plasmids contained, respectively, the L1 predicted intron sequence (phrGFP-saPV1-L1-Intron), the sequence corresponding to the processed nucleotides predicted to be retained after intron splicing (phrGFP-Processed), and a modified intron with mutations in predicted splicing donor and acceptor sites (phrGFP-saPV1-L1-Mutated-Intron). The sequences of all plasmids were verified by Sanger sequencing.

Nucleic acid extraction, PCR, and reverse transcription (RT)-PCR analyses. Total DNA from caudal fin samples was extracted using an E.Z.N.A. tissue DNA kit (VWR, Omega Bio-tek, Norcross, GA, USA) in accordance with the manufacturer's instructions. Total RNA was obtained using the RNeasy minikit (Qiagen) with on-column digestion of genomic DNA using RNase-free DNase I (Qiagen) according to the manufacturers' instructions. Conventional PCR and nested-PCR protocols were used for viral detection. All PCRs were carried out in 50-μl volumes using the GoTaq G2 Flexi DNA polymerase kit (Promega Corporation, Madison, WI, USA). For LCDV detection, a 609-bp fragment of the viral MCP gene was amplified by PCR using the primers LCDVs-F and LCDVs-R as previously described (29). For nested PCR, specific primers were designed (RT-LCDV-F and RT-LCDV-R2) that generate a 238-bp amplicon within the 609-bp fragment of the MCP gene (nucleotide positions 173 to 410 of the LCDV-Sa MCP gene; GenBank accession no. GU320728).

For the detection of SaPyV1 and SaPV1, PCR and nested-PCR protocols using specific primers were designed to amplify a fragment of 602 bp of large T antigen gene (Polyo602-F and Polyo602-R) and a 420-bp amplicon (Polyo420-F and Polyo420-R) within this or a fragment of 650 bp of the L1 gene (L1-3′-F and L1-3′-R) and an included 152-bp amplicon (NL1-3′-F and NL1-3′-R).

For RT-PCR analyses, first-strand synthesis was carried out using Superscript III reverse transcriptase (Invitrogen, Carlsband, CA, USA) and the gene-specific reverse oligonucleotide L1R + 679. The products of the reaction were subsequently used for conventional PCR using OneTaq polymerase (New England BioLabs, Inc.) and oligonucleotides L1F + 336 and L1R + 679 producing fragments of 421 and 349 bp for the predicted precursor and processed sequences, respectively. All oligonucleotide sequences and PCR amplification procedures used in this work are available upon request.

Accession number(s). Metagenomic reads were deposited into the Sequence Read Archive (SRA) under accession number SRP067009. The assembled and annotated complete viral genome sequences were deposited in Genbank under accession numbers KX643370 (LCDV-Sa), KX643371 (SaPyV1), and KX643372 (SaPV1).

RESULTS

Presence of novel fish viruses in the virome of lymphocystis disease-affected fish. Lymphocystis disease-affected juvenile gilthead sea bream (0.5 to 5 g) develop lesions that typically cover the whole body, including the fins (Fig. 1A). Thin sections of these lesions observed by electron microscopy showed the typical large number of viral particles in ordered arrays (Fig. 1B). To obtain the complete genome sequence of LCDV-Sa, we purified viral particles by gradient centrifugation from pooled lesions obtained from a single LCD-affected gilthead sea bream specimen. Viral DNA was further purified and subjected to both Roche-454 and Illumina next-generation sequencing techniques. A first analysis of the sequences obtained showed an abundance of viral reads, as expected. While most of the viral reads were assigned to the *Iridoviridae* family, a significant proportion was unexpectedly assigned to the *Papillomaviridae* and *Polyomaviridae* families, indicating the presence of additional viruses in the sample (Fig. 1C). Using reads from both deep-sequencing technologies, we were able to assemble three independent complete viral genomes corresponding to the LCDV-Sa isolate and to two putative novel fish viruses, tentatively named Sparus aurata polyomavirus 1 (SaPyV1) and Sparus aurata papillomavirus 1 (SaPV1). The independence of the three assembled contigs is supported by the different coverage as well as different GC% value for each genome (Table 1), and the circular nature of the SaPyV1 and SaPV1 genomes was confirmed by PCR and Sanger sequencing. Even though a bias toward small circular DNA molecules has been estimated in the range of 50 to 100× (30), the relative abundance of SaPV1 and SaPyV1 genome copies compared to that of LCDV-Sa rules out a contaminating origin for these two small viruses. When the viral sample was negatively stained and examined by transmission electron microscopy, LCDV-Sa virions were frequently observed alongside smaller icosahedral particles with measured diameters ranging

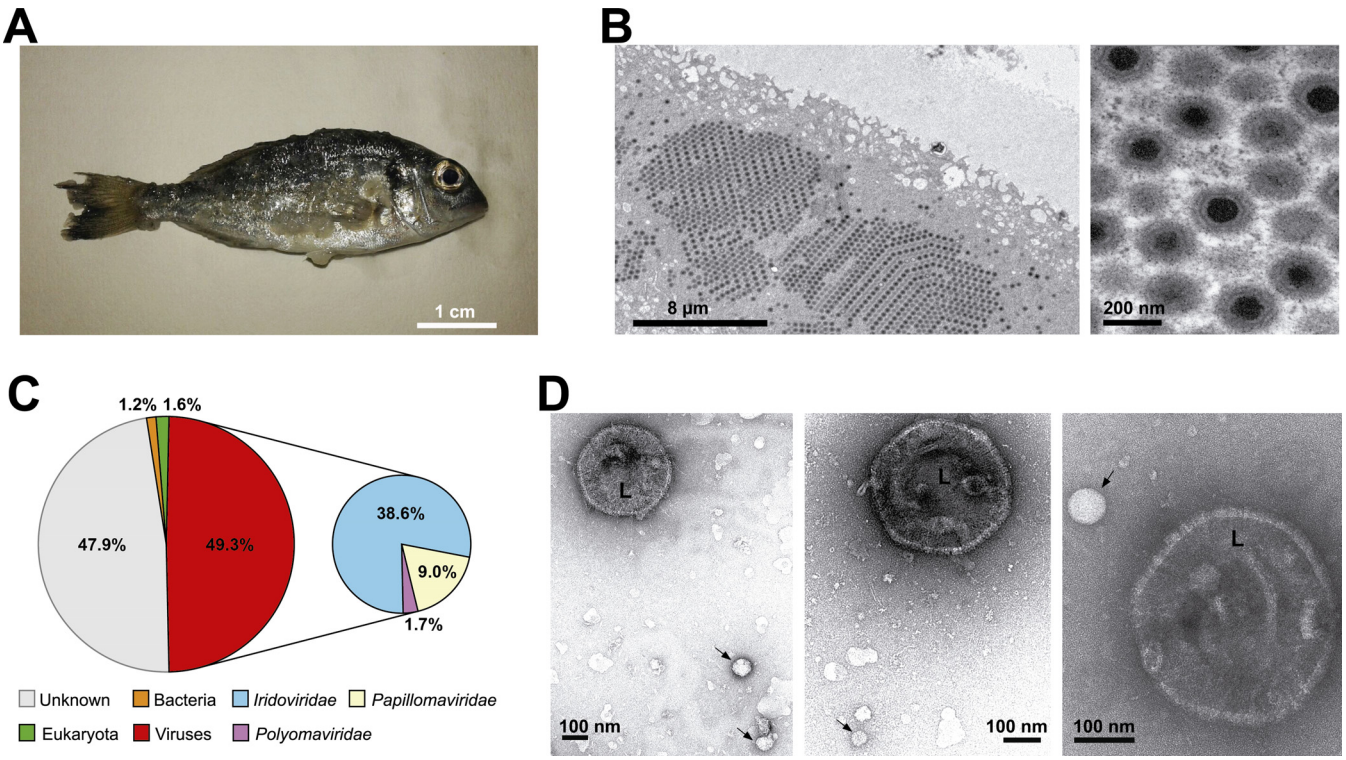

FIG 1 Virome of lymphocystis disease-affected gilthead sea bream. (A) Picture of a sample juvenile gilthead sea bream with lymphocystis collected from a Mediterranean Sea farm (southern Spain) in 2001. (B) Transmission electron micrographs of an ultrathin section of skin with lymphocystis lesions showing arrays of LCDV-Sa particles. (C) Taxonomic profile based on a BLASTx comparison of 10,000 metagenomic reads obtained from lesions of a lymphocystis-affected juvenile gilthead sea bream against GenBank-nr database. (D) Transmission electron micrographs of a negatively stained semipurified viral sample derived from gilthead sea bream lymphocysts. LCDV-Sa particles (L) and putative additional icosahedral particles (arrows) are indicated. Scale bars are shown.

from 51 to 82 nm (Fig. 1D), which may correspond to any of the additional viruses described. In summary, our data revealed the coexistence of three abundant viruses from the families *Iridoviridae*, *Papillomaviridae*, and *Polyomaviridae* within a lymphocystis-diseased gilthead sea bream specimen.

Concurrent detection of three distinct viruses is frequent in lymphocystis lesions. The surprising detection of a polyomavirus and a papillomavirus besides LCDV in lymphocystis lesions led us to assess their prevalence in additional samples. For this, PCR procedures previously established for LCDV or newly devised based on the obtained sequences of SaPyV1 and SaPV1 were used to screen a collection of caudal fin samples from asymptomatic and lymphocystis-diseased juvenile gilthead sea bream specimens collected at different Mediterranean farms (Table 2). In asymptomatic fish, we were unable to detect the presence of either SaPyV1 or SaPV1, neither by direct nor by nested-PCR procedures. Interestingly, in most of the examined individuals (10 of 12), the presence of small amounts of LCDV-Sa DNA was detected by nested PCR. As expected, all of the lymphocystis-diseased fish analyzed contained LCDV-Sa DNA, with 100% of the samples being positive by direct PCR. In stark contrast to the samples from asymptomatic fish, we were able to detect either SaPV1 or SaPyV1 in all the samples from diseased fish analyzed, with 6 out of 10 samples containing all three viruses concurrently.

LCDV-Sa is a novel species of the genus *Lymphocystivirus*. The full-length genome sequence of LCDV-Sa (Fig. 2A) is 208,501 bp in length and has a GC content (33.0%) that is higher than that of LCDV-C (27.2%) or LCDV-1 (29.1%). To calculate the overall degree of nucleotide identity, we used a BLAST-based alignment algorithm (PASC) that does not take into account positional information on the genomes. Using this approach, we calculated that LCDV-Sa shared only 54.69% identity with LCDV-C or 38.95% identity with LCDV-1 while LCDV-1 and LCDV-C shared 41.53% identity with each other. Accepted viral species within other genera of the *Iridoviridae* family show identity values above 85%. Using the same criterion, LCDV-Sa represents a new species within the *Lymphocystivirus* genus. Further, we performed dot plot analysis to assess the degree of colinearity among the fully sequenced LCDV isolates. As shown in Fig. 2B, LCDV-Sa showed evidence of heavy genomic rearrangements, compared to LCDV-C, and an almost complete absence of colinearity stretches with LCDV-1.

The LCDV-Sa genome contains 183 putative ORFs, including orthologues of all 26 conserved iridovirus core genes (23). To assess the phylogenetic position of LCDV-Sa, we used the concatenated sequence of proteins encoded by the 26 core genes to calculate maximum-likelihood trees, including representatives of all groups of vertebrate iridoviruses. As shown in Fig. 2C, LCDV-Sa clusters confidently within the genus *Lymphocystivirus*, where it appears to be more closely related to the LCDV-C virus than to the LCDV-1 virus.

In terms of gene content, BLASTp analyses were performed on all annotated ORFs (see Table SA1 in the supplemental material).

TABLE 2 Detection of LCDV-Sa, SaPyV1, and SaPV1 in *Sparus aurata* fins

Fish group and sample	Origin/yr	Detection of virus[a]					
		LCDV-Sa		SaPyV1		SaPV1	
		PCR	nPCR	PCR	nPCR	PCR	nPCR
Asymptomatic fish							
1	Spain/2010	−	+	−	−	−	−
2	Spain/2010	−	+	−	−	−	−
3	Spain/2010	−	−	−	−	−	−
4	France/2010	−	+	−	−	−	−
5	France/2010	−	+	−	−	−	−
6	Spain/2012	−	+	−	−	−	−
7	Spain/2014	−	+	−	−	−	−
8	Spain/2014	−	+	−	−	−	−
9	Turkey/2014	−	+	−	−	−	−
10	Turkey/2014	−	+	−	−	−	−
11	Turkey/2014	−	−	−	−	−	−
12	Spain/2015	−	+	−	−	−	−
Diseased fish							
1	Spain/2005	+	+	+	+	+	+
2	Spain/2005	+	+	+	+	+	+
3	Spain/2005	+	+	+	+	+	+
4	Spain/2005	+	+	+	+	+	+
5	Spain/2011	+	+	−	−	−	+
6	Spain/2011	+	+	−	+	−	−
7	Spain/2012	+	+	−	−	−	+
8	Spain/2012	+	+	−	+	+	+
9	Spain/2012	+	+	+	+	+	+
10	Italy/2012	+	+	−	+	−	−

[a] nPCR, nested PCR.

Out of 145 ORFs showing significant BLASTp hits, 129 presented best hits to LCDV-C annotated ORFs, supporting the closer relationship of LCDV-Sa to this virus, while 10 were more similar to ORFs from LCDV-1. Two other genes (*LCDVSa001L* and *LCDVSa050R*) encoded proteins with similarity to those identified in the recently described scale drop disease virus (SDDV) (31). Importantly, four new genes encoding proteins with significant BLASTp hits not previously annotated in lymphocystiviruses were identified. Three of the encoded proteins were most similar to proteins from fish species: LCDVSa168L was most similar to an uncharacterized protein from *Cynoglossus semilaevis*, LCDVSa179R showed a best BLASTp hit with the insulin-like growth factor 2a from *Danio rerio*, and LCDVSa017R showed highest similarity to a predicted induced myeloid leukemia cell differentiation protein Mcl-1 homolog from the fish species *Notothenia coriiceps*. One protein was most similar to a protein from a spider species (LCDVSa033R, best hit to the zinc finger protein 271 from *Stegodyphus mimosarum*).

The identification of a polyomavirus in *S. aurata* supports the existence of a novel, phylogenetically ancient group of polyomaviruses. The second complete viral genome identified in the gilthead sea bream lymphocystis sample corresponded to a circular, 7,299-nt-long DNA with a 52.1% GC content. This genome contains five nonoverlapping ORFs carried on opposite DNA strands in an organization reminiscent of that found in polyomaviruses, which are bidirectionally transcribed from a regulatory region that includes the origin of replication (Fig. 3A). Moreover, a BLASTp similarity search (see Table SA1 in the supplemental material) readily identified both a potential large T antigen as well as a VP1 protein, which are characteristic of the *Polyomaviridae* family. The best hits were obtained with the recently described black sea bass polyomavirus 1 (BassPyV1) (32), a virus with a genome very similar in length (7.37 kbp) that represents the largest polyomavirus genome reported. Consistently, phylogenetic analyses based on the amino acid sequence of the large T antigen and the VP1 proteins revealed that this novel polyomavirus clustered with the only other currently available polyomaviruslike full-length viral genomes isolated from fish: BassPyV1, Trematomus polyomavirus 1, and giant guitarfish polyomavirus 1 (Fig. 3C and D). Interestingly, while the large T antigen of this fish clade of polyomavirus was clearly related to those from genus *Avipolyomavirus*, their VP1 proteins were phylogenetically more related to polyomaviruses of the genus *Wukipolyomavirus*. A third ORF product potentially corresponding to a late viral transcript was annotated as VP2, as it shared significant homology to the proteins identified as VP2 in BassPyV1 and Trematomus polyomavirus 1.

The overall percent nucleotide identity (calculated using PASC), compared to other currently known polyomaviruses, was highest to BassPyV1 (33.59%), followed by Trematomus polyomavirus 1 (31.57%), and dropping to below 18% for the next hits, which included polyomaviruses from mammalian and avian species. A relatively low identity (17.47%) to the giant guitarfish polyomavirus 1 may be due to the atypical genome length of this virus (3,963 bp).

Phylogenetic analyses based on VP1 and large T antigen proteins suggest that the polyomavirus identified within the gilthead sea bream lymphocystis lesions shared a common ancestor with

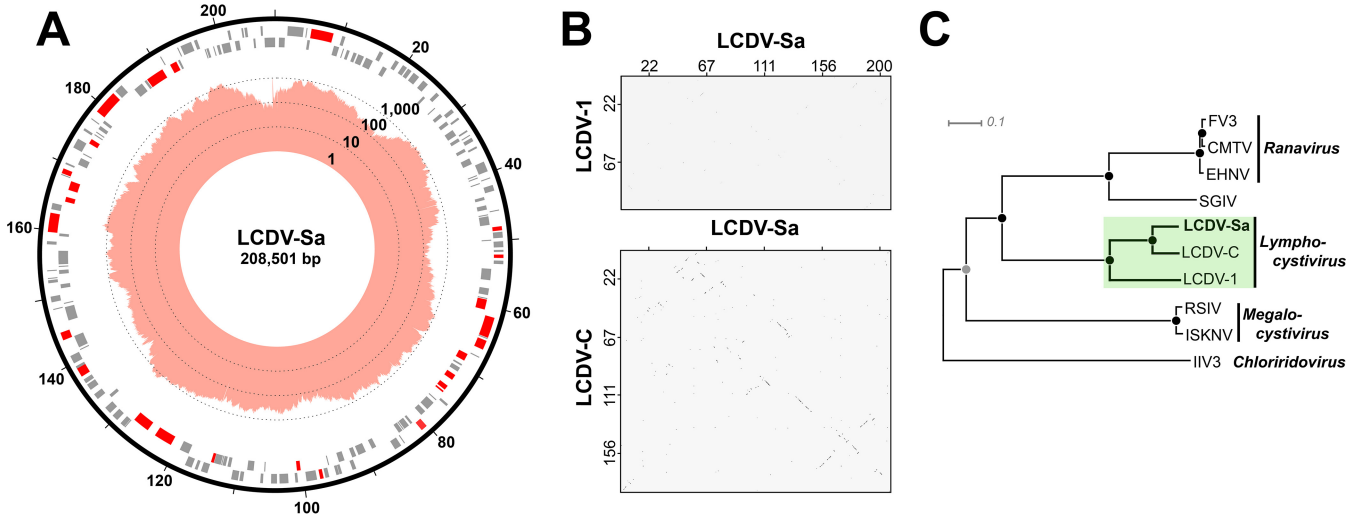

FIG 2 Genome analysis and phylogeny of LCDV-Sa. (A) Graphical circular map of the LCDV-Sa genome. The outer scale is numbered clockwise in kilobase pairs. Predicted putative genes are denoted by rectangles, and those in red correspond to iridovirus core genes. The internal histogram shows the coverage in logarithmic scale along the LCDV-Sa genome achieved by Illumina metagenomic sequencing. (B) Dot plot comparison of LCDV-Sa versus LCDV-1 and LCDV-C. Scale is numbered in kilobase pairs. (C) Phylogenetic analysis showing the relationship of LCDV-Sa to representative iridoviruses. The unrooted phylogenetic tree is based on aligned amino acid sequences encoded by 26 concatenated core genes. The evolutionary relationships were inferred through the maximum-likelihood method using R with 1,000 bootstrap replicates. Gray and black circles represent branch nodes with >75% and >90% bootstrap values, respectively. The scale bar indicates the number of amino acid substitutions per residue. The green rectangle indicates members of the genus *Lymphocystivirus*.

the other fish polyomaviruses. Therefore, we propose to name this novel polyomavirus Sparus aurata polyomavirus 1 (SaPyV1) and suggest that SaPyV1 may be a member of a putative new genus of fish-infecting polyomaviruses.

First description of a papillomavirus with unique characteristics in a fish host. The third complete viral genome identified in the virome of lymphocystis disease-affected sea bream, proposed to be named Sparus aurata papillomavirus 1 (SaPV1), is 5,748 bp long (Fig. 3B) and has a GC content of 39.5%. While this genome is significantly smaller than that of the previously described papillomaviruses (about 8 kbp), the assembled genome nevertheless presented a typical papillomavirus organization, with seven distinct ORFs carried on the same strand. Similarity searches identified distant orthologues of the early E1 and E2 proteins involved in replication and transcription and late structural proteins L1 and L2 (see Table SA1 in the supplemental material). The region situated between the end of the L1 ORF and the beginning of the E1 ORF was only 408 bp long and may correspond to the long control region that usually contains transcriptional and replicative regulatory signals in other papillomaviruses. This region included two small ORFs encoding proteins of 52 (sORF1) and 70 (sORF2) amino acids, which showed no similarities to other proteins in the databases. Interestingly, the sORF1 protein contains both a pRB binding motif (LXCXE) and a C-terminal PDZ class 2 binding motif, which are elements typically present in the longer E7 and E6 proteins found in most known papillomaviruses.

Interestingly, the potential major structural protein (L1)-encoding sequence of SaPV1 is interrupted by an intron (see below). This original gene structure within the *Papillomaviridae* family was confirmed by deep sequencing (Illumina) and PCR-Sanger sequencing. Demarcation criteria and taxonomy of the family *Papillomaviridae* are based on the nucleotide sequence of the L1 gene (http://pave.niaid.nih.gov). The low nucleotide similarity of the SaPV1 L1 gene to other L1 genes in the *Papillomaviridae* family indicates that it is a distant member of this group. A phylogenetic analysis based on the amino acid sequence of the complete L1 protein (Fig. 3B) showed that SaPV1 clustered with a confidently supported group, including avian and reptile papillomaviruses. However, the inferred evolutionary distance is large, as reflected in the length of the branch separating SaPV1 from the other viruses in this group.

The conserved genomic organization, the similarity of the main proteins, and the phylogenetic analysis support the discovery of SaPV1 as the first member of the *Papillomaviridae* family in fish.

The SaPV1 L1 major structural protein is expressed from a spliced transcript. The unusual structure for the SaPV1 L1 gene suggested that a functional full-length protein might be expressed from a spliced mRNA. Indeed, putative splicing acceptor and donor sites were consistently predicted as shown in Fig. 4A. To analyze whether the predicted sequence can act as an intron, we first inserted the corresponding nucleotides into a plasmid containing the coding sequence of an hrGFP reporter gene. When this plasmid was transfected into either Vero (data not shown) or HEK293 cells, a strong fluorescence signal was detected in an elevated fraction of cells (Fig. 4B). Both the number of hrGFP-expressing cells and the intensity per cell were similar to those obtained when the control plasmids (bearing no insertion or containing an in-frame insertion at the same position of the amino acids expected to be retained after splicing) were transfected. Additionally, a control plasmid in which the splicing donor and acceptor sites were mutated did not produce any fluorescent signal, suggesting that hrGFP expression from the intron-bearing plasmid is due to a correct splicing of the precursor mRNA. Next, we analyzed by RT-PCR the size of SaPV1 L1-specific mRNA transcripts after

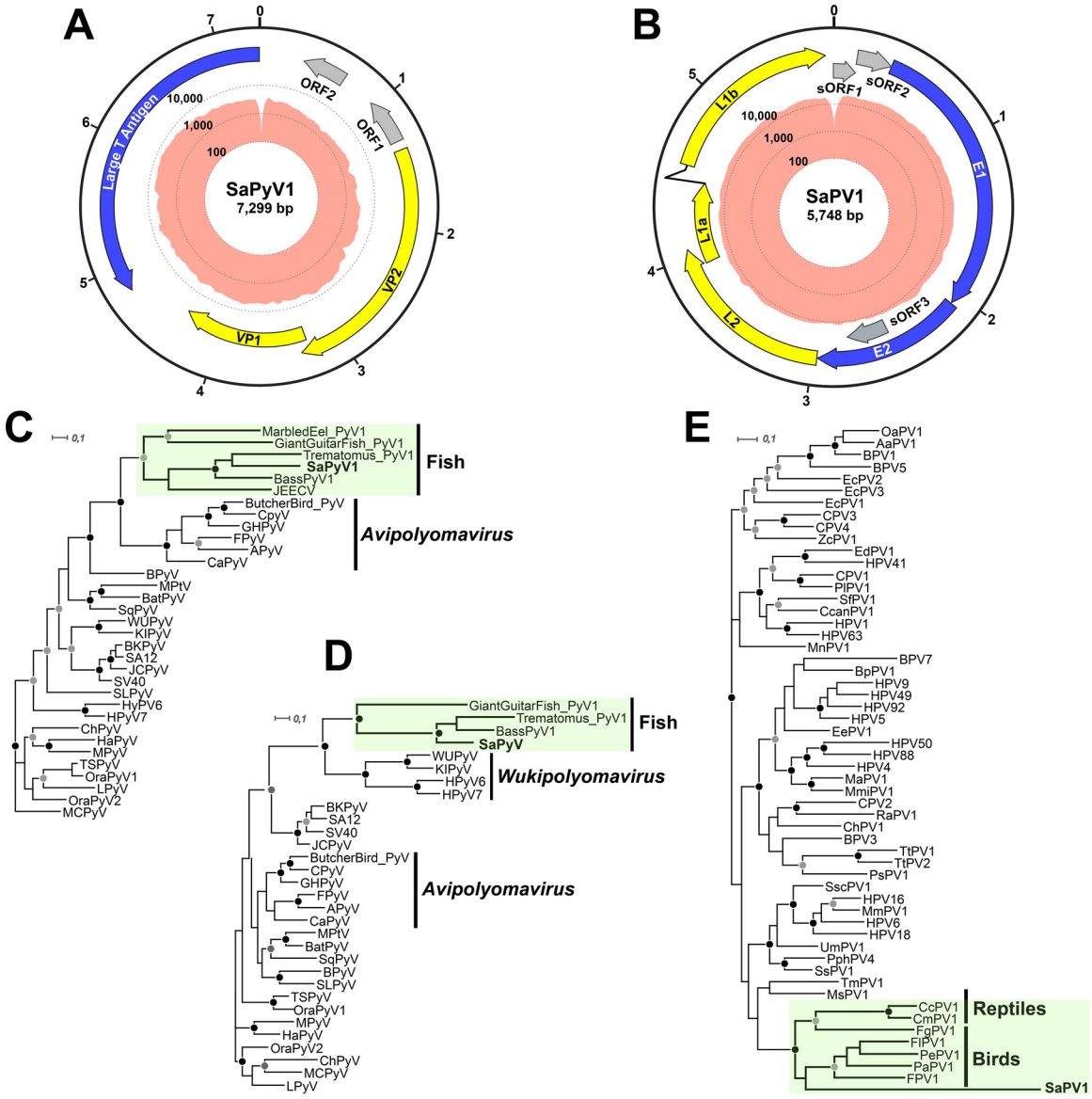

FIG 3 Genome analyses and phylogeny of SaPyV1 and SaPV1. (A and B) Circular graphs of SaPyV1 (A) and SaPV1 (B) genomes are shown. The outer scales are numbered clockwise in kilobase pairs. Predicted genes are indicated by arrows. Yellow, blue, and gray colors indicate structural, nonstructural, and unknown-function proteins, respectively. The internal histograms show the coverage in logarithmic scale along the genomes achieved by Illumina metagenomic sequencing. (C and D) Phylogenetic trees of polyomaviruses based on the amino acid sequences of the large T antigen (C) and VP1 (D) proteins. (E) Phylogenetic analysis of papillomaviruses based on aligned amino acid sequences of the SaPV1 L1 protein and the L1 protein of other representative papillomaviruses. The evolutionary relationships were inferred through the maximum-likelihood method using R with 1,000 bootstrap replicates. The trees were arbitrarily rooted. Gray and black circles represent branch nodes with >75% and >90% bootstrap values, respectively. The scale bar indicates the number of amino acid substitutions per residue. Green rectangles indicate fish polyomaviruses and bird-reptile papillomaviruses.

transfection of a plasmid carrying a 2,131-bp-long DNA fragment that includes the complete L1 gene sequence under the control of a strong CMV promoter. As shown, a PCR band corresponding to the expected size of the spliced product was detected, and this band was smaller than that obtained when the unprocessed plasmid DNA was used as a template. (Fig. 4C). The nucleotide sequence of the PCR product confirmed that the splicing site used corresponded to the predicted positions. Finally, we performed RT-PCR on samples from lymphocystis-affected sea bream (data not shown). This showed the presence of a spliced SaPV1 L1 mRNA within lesions, indicating that the viral gene is actively transcribed as well as spliced out during infection *in vivo*.

DISCUSSION

We report the serendipitous finding of two novel fish viruses associated with the development of lymphocystis disease in gilthead sea bream. Since *S. aurata* is extensively farmed in the Mediterranean area, severe economic loss due to this disease is a current concern. However, lymphocystis disease virus isolates from this species and location have been poorly characterized to date. The complete genome sequence of LCDV-Sa obtained here was

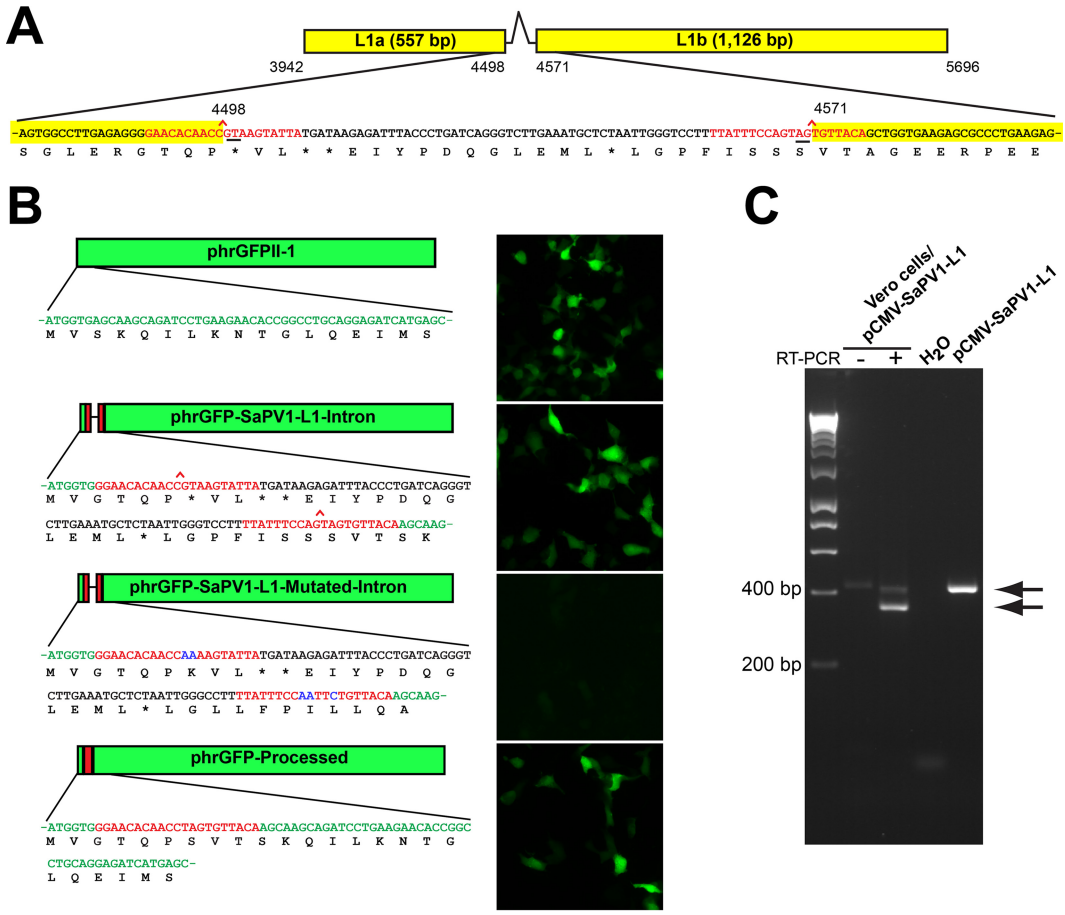

FIG 4 The SaPV1 L1 major structural protein is expressed from a spliced mRNA. (A) Diagram showing the predicted structure of the SaPV1 L1-encoding gene containing two exons (L1a and L1b). The intron sequence is shown, with splice donor and acceptor consensus sequences in red and the splice site indicated. Relevant nucleotide positions and translation are included. (B) HEK293 cells were transfected with plasmids as indicated on the left, and fluorescence was observed at 48 h after transfection. (C) RT-PCR analysis of RNA from Vero cells transfected with the plasmid bearing the SaPV1 L1 region using L1-specific oligonucleotides performed in the absence (−) or presence (+) of reverse transcriptase. For reference, the PCR using the same oligonucleotide pair was performed using the purified plasmid as the template or water as a negative control. Arrows on the right indicate the positions corresponding to the sizes of the unprocessed (top) and spliced (bottom) SaPV1 L1 messenger. Sizes (in base pairs) of selected markers are indicated on the left.

208,501 bp in length, significantly longer than those of the two lymphocystiviruses previously sequenced, LCDV-C (186,250 bp) and LCDV-1 (102,653 bp), and it is, therefore, the largest known vertebrate iridovirus genome. Moreover, it provides genetic evidence that LCDV-Sa should be considered a novel and distinct virus species within the genus *Lymphocystivirus*. Importantly, we have identified a set of at least eight previously undescribed genes, which provide useful markers to analyze the evolutionary relationships within this group. Two of them were found to be most similar to genes from the recently described scale drop disease virus (SDDV), a distant member of the megalocytiviruses infecting tropical fish in Southeast Asia (31). The genomic GC content value of SDDV (37%) is very different from that of other megalocytiviruses (55%) but closer to that of LCDV-Sa (33%), suggesting the existence of a common ancestor for viruses belonging to both genera of fish-infecting iridoviruses.

Acquisition of genes from the host is thought to be a common mechanism driving the evolution of large nucleocytoplasmic DNA viruses, including the iridoviruses (33), and has been proposed as the origin of the dihydrofolate reductase-encoding gene uniquely found in the fish-infecting ranavirus epizootic hematopoietic necrosis virus (EHNV) (34). In this regard, we have found evidence for at least four different LCDV-Sa genes possibly derived from a fish host or from invertebrate species, suggesting that complex interactions among both vertebrate and invertebrate iridoviruses might have shaped the evolution of this viral group.

To date, three different species of LCDV, isolated from different host species and geographic locations, have been fully sequenced. The elevated frequency of genomic rearrangements among them appears to be a common feature of other vertebrate iridoviruses (23, 35) and is different from the extensive conservation of a central genome core in the related double-stranded DNA (dsDNA) *Poxviridae* family, which is probably related to their different replicative mechanisms. This result, together with the elevated branch length observed among the three LCDV viruses in our phylogenetic analyses, might reflect that the lymphocystiviruses have evolved over an extended time period in heterogeneous host species and locations. This is in contrast to the short branch length observed within the group of amphibianlike ranavi-

ruses, which seem to be in the midst of an adaptive radiation process (34, 36).

The *Polyomaviridae* is a family of small nonenveloped icosahedral dsDNA viruses that have been previously described only in mammalian and avian hosts. In recent years, the use of new sequencing technologies has increased the number of known polyomaviruses in humans (37, 38) and other animal species. The polyomaviruses have been classified into three different genera (39), with the mammalian viruses being included in either *Orthopolyomavirus* or *Wukipolyomavirus* and all the avian viruses within the genus *Avipolyomavirus*. Besides their differences in overall nucleotide similarity, biological properties such as broad host range, tissue tropism, and generalized ability to cause disease in the infected host distinguish the avipolyomaviruses from the other two groups. A polyomavirus isolated from a mixed tissue sample of apparently healthy black sea bass (*Centropristis striata*) (BassPyV1) was first described in 2015 (32), and two additional complete polyomavirus genome sequences obtained from fish species have been deposited in data banks since then (NC_026944.1 and NC_026244.1). Therefore, SaPyV1 is one of the first polyomaviruses reported in fish, and its genome reproduces the main hallmarks described for BassPyV1, including genome size and gene content. Moreover, in the phylogenetic analyses based on the amino acid sequence alignments of either large T antigen or VP1, SaPyV1 clustered together with the known polyomaviruses isolated from fish, suggesting a possible monophyletic origin of this group.

All polyomaviruses isolated from fish showed an overall nucleotide sequence identity that was below the accepted species demarcation threshold of 81% (39). Collectively, they showed sequence identities below 25% to other polyomaviruses. In contrast, the closest similarities among viruses from different genera were found to be at least 40%. Therefore, we propose that fish polyomaviruses should tentatively be considered distinct species within a new and distant fourth genus of the *Polyomaviridae* family. While we have detected SaPyV1 in skin samples and found it associated with lymphocystis lesions in affected gilthead sea bream, there is currently little information available about the ability to cause disease, the host range, or the tissue tropism of the fish polyomaviruses, which will constitute important aspects in the study of this novel virus group. Because polyomavirus evolution is thought to occur, at least in part, by codivergent speciation with the host (40), the finding of polyomaviruses in fish suggests that further members of the family might be found in amphibians and reptiles, which split later than fish and before birds and mammals during vertebrate evolution (41).

As noted before for BassPyV1 and described in our phylogenetic analysis of SaPyV1, the large T antigen of fish polyomaviruses is closely related to that found in the Japanese eel endothelial cell-infecting virus (JEECV). Currently JEECV is not assigned to any known viral family due to the lack of sequence similarities beyond that of the large T antigen, which was proposed to have been acquired by recombination from a polyomalike virus coinfecting a fish host (42). Our findings support this hypothesis, providing evidence for the existence of a potential source of the large T antigen gene found in JEECV.

Papillomaviruses are small, nonenveloped dsDNA epitheliotropic viruses that typically cause benign lesions of the skin and mucous membranes. This has been frequently described as the most successful vertebrate virus family, and currently, more than 300 distinct papillomavirus types, classified into 37 different genera, are known, with more than 200 different strains derived from human hosts (43). While most papillomaviruses have been isolated from mammals, isolates from five bird and three reptile species have also been described. This has prompted the ideas that the host range of papillomaviruses might be restricted to amniotes and that this virus family may have emerged at the time of divergence of this group of vertebrates, around 330 million years ago (44). Therefore, SaPV1 represents the first, and to date the only, papillomavirus detected in fish, pushing back the date of appearance of this virus family by up to 120 million years, at the time of appearance of the ray-finned fishes, Actinopterygii (41).

A unique aspect of SaPV1 is the expression strategy for its major capsid protein L1, as there are no other examples of splicing events occurring within an L1-encoding ORF (45). It will be of interest to address whether this strategy is conserved in other, yet to be identified papillomaviruses from fish or other hosts. Of note, splicing of the gene encoding structurally related major capsid viral proteins has been described in Acanthamoebae polyphaga mimivirus (46) and the recently described giant virus faustovirus (47), where a complex gene structure with 10 introns has been proposed to allow rapid adaptation of the virion to novel environments.

As described before, the genome structure of SaPV1 contains the elements conserved in all previously described papillomaviruses, including genes encoding L1 and L2 structural proteins, as well as the E1 and E2 early proteins involved in viral transcription and replication (48). This corresponds precisely to the basic papillomavirus genome structure that was proposed to represent that of a putative ancestral proto-papillomavirus (49). Later acquisition of additional proteins E6 and E7, which evolved at a higher rate, is thought to have allowed adaptation to putative new ecological niches, driving papillomavirus diversification (48). Intriguingly, we have identified a small ORF (sORF1) in the region of the SaPV1 genome where E6 and E7 are generally encoded. The protein encoded by sORF1 contains elements typically found in E6 (PDZ-binding motif) (50) or E7 (pRb-binding motif) proteins (51), although it is currently unknown whether it is expressed during infection. An atypical E7 protein from the rhesus papillomavirus type 1 has also been shown to contain a functional PDZ-binding motif (52).

Importantly, the epidermis of adult fish is a stratified squamous epithelium in which all three layers preserve the mitotic capacity. Since the papillomavirus replication cycle is tightly coupled to the differentiation state in infected squamous epithelia, with late protein expression occurring only in differentiated, nonreplicating cells, it will be of importance to study how the permanent mitosing capacity of fish skin epithelia in all strata impinge on the biology of SaPV1. The transition toward terminally differentiated epithelia such as those encountered in mammalian skin occurs during the appearance of Amphibia, in which metamorphic but not larval life forms present this characteristic trait (53). Therefore, it will be of particular interest to study how and when during papillomavirus evolution the acquisition of functional E6 and E7 proteins might reflect the need to adapt to this new environment.

Although fish papillomaviruses have not been not previously reported, the existence of neoplastic epidermal papillomas in several fish species has been described. In several cases, viruslike particles of a different nature have been encountered in affected cells,

although a viral etiology has not been established for any of them (54–57). The finding of the first complete genome of a bona fide fish papillomavirus will provide tools to further study these particular cases.

In our report, LCDV-Sa was detected by nested PCR in 83.3% of the asymptomatic fish tested, with no evidence of accompanying SaPyV1 or SaPV1. This is consistent with previously reported prevalence values of subclinical LCDV infections in gilthead sea bream juveniles, in which the quantitative PCR-estimated viral load in caudal fin ranged between 1 copy and 3.3×10^2 copies of viral DNA per mg of tissue (58). In addition, 100% of a yellow perch population in Alberta (Canada) was found to contain LCDV, with only 2.6% of the fish showing lymphocystis lesions (59). These evidences support an accepted scenario of highly prevalent and frequently inapparent infection with LCDV in several fish species, with lymphocyst formation induced by unknown mechanisms related to physical, environmental, or immunological stress. The frequent detection of SaPyV1 and/or SaPV1 in LCD-affected gilthead sea bream might reflect an opportunistic increased capacity of these viruses to replicate in the presence of an LCDV infection. Alternatively, a direct role of either or both of these viruses in the trigger or development of the condition might be postulated. It is interesting to note that early studies on LCDV described the presence within lymphocysts of two distinct viral particle forms with different diameters (hexagonal particles of 200 to 250 nm and smaller forms ranging from 65 to 100 nm) (60). In addition, purification of LCDV by gradient centrifugation was reported to yield two distinct virus populations, one corresponding to LCDV and a second to an undetermined lymphocystis-associated virus with smaller particles (61). However, further epidemiologic studies together with virus isolation and experimental infection challenges must be performed in order to establish the possible role of the newly described SaPyV1 and SaPV1 in lymphocystis disease development.

We have identified two novel fish viruses with important implications for the origin and evolution of the *Polyomaviridae* and *Papillomaviridae* families. The life cycles of these novel viruses, as well as the molecular mechanisms by which their interaction with an iridovirus might contribute to trigger or sustain lymphocystis development, remain to be explored.

ACKNOWLEDGMENTS

We thank Milagros Guerra of the electron microscopy unit at the CBMSO.

This work was supported by grant AGL2009-08711 from the Spanish Ministerio de Economía y Competitividad and grant P12-RNM-2261 from Junta de Andalucía. A.L-B. was the recipient of a Ramón y Cajal RYC-2010-06300 fellowship from the Spanish Ministerio de Economía y Competitividad.

FUNDING INFORMATION

This work, including the efforts of Juan J. Borrego, was funded by Junta de Andalucia (P12-RNM-2261). This work, including the efforts of Alberto López-Bueno, was funded by Ministerio de Economía y Competitividad (RYC-2010-06300). This work, including the efforts of Alí Alejo, was funded by Ministerio de Economía y Competitividad (AGL2009-08711).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. **Smail DA, Munro ES.** 2012. The virology of teleosts, p 186–291. *In* Roberts RJ (ed), Fish pathology, 4th ed, Wiley-Blackwell, Oxford, United Kingdom.
2. **Weissenberg R.** 1949. Studies on lymphocystis tumor cells of fish; the osmiophilic granules of the cytoplasmic inclusions and their interpretation as elementary bodies of the lymphocystis virus. Cancer Res **9:**537–542.
3. **Weissenberg R.** 1914. Über infectiöse Zellhypertrophie bei Fischen (Lymphocystiserkrankung). Sitzungsber Kgl Preuss Akad Wiss Sitz Physik Mathem **16:**792.
4. **Wolf K.** 1962. Experimental propagation of lymphocystis disease. Virology **18:**249–256. http://dx.doi.org/10.1016/0042-6822(62)90011-9.
5. **Walker R.** 1962. Fine structure of lymphocystis virus of fish. Virology **18:**503–505. http://dx.doi.org/10.1016/0042-6822(62)90047-8.
6. **Chinchar VG, Yu KH, Jancovich JK.** 2011. The molecular biology of frog virus 3 and other iridoviruses infecting cold-blooded vertebrates. Viruses **3:**1959–1985. http://dx.doi.org/10.3390/v3101959.
7. **Jancovich JK, Chinchar VG, Hyatt A, Miyazaki T, Willimans T, Zhang QY.** 2012. Family Iridoviridae, p 193–210. *In* King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego, CA.
8. **Darai G, Delius H, Clarke J, Apfel H, Schnitzler P, Flügel RM.** 1985. Molecular cloning and physical mapping of the genome of fish lymphocystis disease virus. Virology **146:**292–301. http://dx.doi.org/10.1016/0042-6822(85)90012-1.
9. **Tidona CA, Darai G.** 1997. The complete DNA sequence of lymphocystis disease virus. Virology **230:**207–216. http://dx.doi.org/10.1006/viro.1997.8456.
10. **Zhang Q-Y, Xiao F, Xie J, Li Z-Q, Gui J-F.** 2004. Complete genome sequence of lymphocystis disease virus isolated from China. J Virol **78:**6982–6994. http://dx.doi.org/10.1128/JVI.78.13.6982-6994.2004.
11. **Kitamura S-I, Jung S-J, Kim W-S, Nishizawa T, Yoshimizu M, Oh M-J.** 2006. A new genotype of lymphocystivirus, LCDV-RF, from lymphocystis diseased rockfish. Arch Virol **151:**607–615. http://dx.doi.org/10.1007/s00705-005-0661-3.
12. **Paperna I, Sabnai HI.** 1982. An outbreak of lymphocystis in Sparus aurata L. in the Gulf of Aqaba, Red Sea. J Fish Dis **5:**433–437. http://dx.doi.org/10.1111/j.1365-2761.1982.tb00500.x.
13. **Menezes J, Ramos MA, Pereira TG.** 1987. Lymphocystis disease: an outbreak in *Sparus aurata* from Ria Formosa, south coast of Portugal. Aquaculture **67:**222–225. http://dx.doi.org/10.1016/0044-8486(87)90037-8.
14. **Le Deuff RM, Renault T.** 1993. Lymphocystis outbreaks in farmed sea bream, Sparus aurata, first report on French Mediterranean coast. Bull Eur Assoc Fish Pathol **13:**130–133.
15. **Haddad-Boubaker SES, Bouzgarou N, Fakhfakh E, Khayech M, Mohamed SB, Megdich A, Chéhida NB.** 2013. Detection and genetic characterization of lymphocystis disease virus (LCDV) isolated during disease outbreaks in cultured gilt-head sea bream Sparus aurata in Tunisia. Fish Pathol **48:**101–104. http://dx.doi.org/10.3147/jsfp.48.101.
16. **Basurco B, Marcotegui MA, Rueda A, Tiana A, Castellanos A, Tarazona J, Munoz MJ, Coll JM.** 1990. First report of lymphocystis disease in *Sparus aurata* (Linnaeus) in Spain. Bull Eur Assoc Fish Pathol **10:**71–73.
17. **Dezfuli BS, Lui A, Giari L, Castaldelli G, Mulero V, Noga EJ.** 2012. Infiltration and activation of acidophilic granulocytes in skin lesions of gilthead seabream, Sparus aurata, naturally infected with lymphocystis disease virus. Dev Comp Immunol **36:**174–182. http://dx.doi.org/10.1016/j.dci.2011.06.017.
18. **Kvitt H, Heinisch G, Diamant A, Kvitt H.** 2008. Detection and phylogeny of Lymphocystivirus in sea bream Sparus aurata based on the DNA polymerase gene and major capsid protein sequences. Aquaculture **275:**58–63. http://dx.doi.org/10.1016/j.aquaculture.2008.01.007.
19. **Cano I, Ferro P, Alonso MC, Sarasquete C, Garcia-Rosado E, Borrego JJ, Castro D.** 2009. Application of in situ detection techniques to determine the systemic condition of lymphocystis disease virus infection in cultured gilt-head seabream, Sparus aurata L. J Fish Dis **32:**143–150. http://dx.doi.org/10.1111/j.1365-2761.2008.00970.x.
20. **Cano I, Valverde EJ, Lopez-Jimena B, Alonso MC, Garcia-Rosado E, Sarasquete C, Borrego JJ, Castro D.** 2010. A new genotype of Lymphocystivirus isolated from cultured gilthead seabream, Sparus aurata L, and

Senegalese sole, Solea senegalensis (Kaup). J Fish Dis **33:**695–700. http://dx.doi.org/10.1111/j.1365-2761.2010.01164.x.
21. **Tcherepanov V, Ehlers A, Upton C.** 2006. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. BMC Genomics **7:**150. http://dx.doi.org/10.1186/1471-2164-7-150.
22. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–945. http://dx.doi.org/10.1093/bioinformatics/16.10.944.
23. **Eaton HE, Metcalf J, Penny E, Tcherepanov V, Upton C, Brunetti CR.** 2007. Comparative genomic analysis of the family Iridoviridae: re-annotating and defining the core set of iridovirus genes. Virol J **4:**11. http://dx.doi.org/10.1186/1743-422X-4-11.
24. **Hunter JD.** 2007. Matplotlib: a 2D graphics environment. Comput Sci Eng **9:**90–95. http://dx.doi.org/10.1109/MCSE.2007.55.
25. **Krumsiek J, Arnold R, Rattei T.** 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics **23:**1026–1028. http://dx.doi.org/10.1093/bioinformatics/btm039.
26. **Bao Y, Chetvernin V, Tatusova T.** 2014. Improvements to pairwise sequence comparison (PASC): a genome-based web tool for virus classification. Arch Virol **159:**3293–3304. http://dx.doi.org/10.1007/s00705-014-2197-x.
27. **Abascal F, Zardoya R, Posada D.** 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21:**2104–2105. http://dx.doi.org/10.1093/bioinformatics/bti263.
28. **Huson DH, Scornavacca C.** 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol **61:**1061–1067. http://dx.doi.org/10.1093/sysbio/sys062.
29. **Kitamura S-I, Jung S-J, Oh M-J.** 2006. Differentiation of lymphocystis disease virus genotype by multiplex PCR. J Microbiol **44:**248–253.
30. **Kim K-H, Chang H-W, Nam Y-D, Roh SW, Kim M-S, Sung Y, Jeon CO, Oh H-M, Bae J-W.** 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. Appl Environ Microbiol **74:**5975–5985. http://dx.doi.org/10.1128/AEM.01275-08.
31. **de Groof A, Guelen L, Deijs M, van der Wal Y, Miyata M, Ng KS, van Grinsven L, Simmelink B, Biermann Y, Grisez L, van Lent J, de Ronde A, Chang SF, Schrier C, van der Hoek L.** 2015. A novel virus causes scale drop disease in Lates calcarifer. PLoS Pathog **11:**e1005074–21. http://dx.doi.org/10.1371/journal.ppat.1005074.
32. **Peretti A, FitzGerald PC, Bliskovsky V, Pastrana DV, Buck CB.** 2015. Genome sequence of a fish-associated polyomavirus, black sea bass (Centropristis striata) polyomavirus 1. Genome Announc **3:**e01476–14. http://dx.doi.org/10.1128/genomeA.01476-14.
33. **Filée J, Pouget N, Chandler M.** 2008. Phylogenetic evidence for extensive lateral acquisition of cellular genes by nucleocytoplasmic large DNA viruses. BMC Evol Biol **8:**320–313. http://dx.doi.org/10.1186/1471-2148-8-320.
34. **Jancovich JK, Brémont M, Touchman JW, Jacobs BL.** 2010. Evidence for multiple recent host species shifts among the ranaviruses (family Iridoviridae). J Virol **84:**2636–2647. http://dx.doi.org/10.1128/JVI.01991-09.
35. **Tsai C-T, Ting J-W, Wu M-H, Wu M-F, Guo I-C, Chang C-Y.** 2005. Complete genome sequence of the grouper iridovirus and comparison of genomic organization with those of other iridoviruses. J Virol **79:**2010–2023. http://dx.doi.org/10.1128/JVI.79.4.2010-2023.2005.
36. **Stöhr AC, López-Bueno A, Blahak S, Caeiro MF, Rosa GM, Alves de Matos AP, Martel A, Alejo A, Marschang RE.** 2015. Phylogeny and differentiation of reptilian and amphibian ranaviruses detected in Europe. PLoS One **10:**e0118633. http://dx.doi.org/10.1371/journal.pone.0118633.
37. **DeCaprio JA, Garcea RL.** 2013. A cornucopia of human polyomaviruses. Nat Rev Microbiol **11:**264–276. http://dx.doi.org/10.1038/nrmicro2992.
38. **Feltkamp MCW, Kazem S, van der Meijden E, Lauber C, Gorbalenya AE.** 2013. From Stockholm to Malawi: recent developments in studying human polyomaviruses. J Gen Virol **94:**482–496. http://dx.doi.org/10.1099/vir.0.048462-0.
39. **Johne R, Buck CB, Allander T, Atwood WJ, Garcea RL, Imperiale MJ, Major EO, Ramqvist T, Norkin LC.** 2011. Taxonomical developments in the family Polyomaviridae. Arch Virol **156:**1627–1634. http://dx.doi.org/10.1007/s00705-011-1008-x.
40. **Perez-Losada M, Christensen RG, McClellan DA, Adams BJ, Viscidi RP, Demma JC, Crandall KA.** 2006. Comparing phylogenetic codivergence between polyomaviruses and their hosts. J Virol **80:**5663–5669. http://dx.doi.org/10.1128/JVI.00056-06.
41. **Kumar S, Hedges SB.** 1998. A molecular timescale for vertebrate evolution. Nature **392:**917–920. http://dx.doi.org/10.1038/31927.
42. **Mizutani T, Saijo M, Oba M, Ono S-I, Kurane I, Miura E, Sayama Y, Nakanishi A, Ochiai H, Sakai K, Wakabayashi K, Tanaka N, Morikawa S.** 2011. Novel DNA virus isolated from samples showing endothelial cell necrosis in the Japanese eel, Anguilla japonica. Virology **412:**179–187. http://dx.doi.org/10.1016/j.virol.2010.12.057.
43. **Egawa N, Egawa K, Griffin H, Doorbar J.** 2015. Human papillomaviruses; epithelial tropisms, and the development of neoplasia. Viruses **7:**3863–3890. http://dx.doi.org/10.3390/v7072802.
44. **Rector A, Van Ranst M.** 2013. Animal papillomaviruses. Virology **445:**213–223. http://dx.doi.org/10.1016/j.virol.2013.05.007.
45. **Buck CB, Day PM, Trus BL.** 2013. The papillomavirus major capsid protein L1. Virology **445:**169–174. http://dx.doi.org/10.1016/j.virol.2013.05.038.
46. **Azza S, Cambillau C, Raoult D, Suzan-Monti M.** 2009. Revised Mimivirus major capsid protein sequence reveals intron-containing gene structure and extra domain. BMC Mol Biol **10:**39. http://dx.doi.org/10.1186/1471-2199-10-39.
47. **Klose T, Reteno DG, Benamar S, Hollerbach A, Colson P, La Scola B, Rossmann MG.** 2016. Structure of faustovirus, a large dsDNA virus. Proc Natl Acad Sci U S A **113:**6206–6211. http://dx.doi.org/10.1073/pnas.1523999113.
48. **Van Doorslaer K.** 2013. Evolution of the Papillomaviridae. Virology **445:**11–20. http://dx.doi.org/10.1016/j.virol.2013.05.012.
49. **García-Vallvé S, Alonso A, Bravo IG.** 2005. Papillomaviruses: different genes have different histories. Trends Microbiol **13:**514–521. http://dx.doi.org/10.1016/j.tim.2005.09.003.
50. **Ganti K, Broniarczyk J, Manoubi W, Massimi P, Mittal S, Pim D, Szalmas A, Thatte J, Thomas M, Tomaić V, Banks L.** 2015. The human papillomavirus E6 PDZ binding motif: from life cycle to malignancy. Viruses **7:**3530–3551. http://dx.doi.org/10.3390/v7072785.
51. **Roman A, Munger K.** 2013. The papillomavirus E7 proteins. Virology **445:**138–168. http://dx.doi.org/10.1016/j.virol.2013.04.013.
52. **Tomaić V, Gardiol D, Massimi P, Ozbun M, Myers M, Banks L.** 2009. Human and primate tumour viruses use PDZ binding as an evolutionarily conserved mechanism of targeting cell polarity regulators. Oncogene **28:**1–8. http://dx.doi.org/10.1038/onc.2008.365.
53. **Yoshizato K.** 2007. Molecular mechanism and evolutional significance of epithelial-mesenchymal interactions in the body- and tail-dependent metamorphic transformation of Anuran larval skin. Int Rev Cytol **260:**213–260. http://dx.doi.org/10.1016/S0074-7696(06)60005-3.
54. **Wellings SR, Chuinard RG, Bens M.** 2006. A comparative study of skin neoplasms in four species of pleuronectid fishes. Ann N Y Acad Sci **126:**479–501. http://dx.doi.org/10.1111/j.1749-6632.1965.tb14296.x.
55. **Carlisle JC.** 1977. An epidermal papilloma of the Atlantic salmon II: ultrastructure and etiology. J Wildl Dis **13:**235–239. http://dx.doi.org/10.7589/0090-3558-13.3.235.
56. **Fujimoto Y, Madarame H, Yoshida H, Moriguchi R, Kodama H, Iizawa H.** 1986. Pathomorphological observations on epidermal papilloma of flatfish (Liopsetta obscura). Jpn J Vet Res **34:**81–103.
57. **McAllister PE, Nagabayashi T, Wolf K.** 1978. Viruses of eels with and without stomatopapillomas. Ann N Y Acad Sci **298:**233–244.
58. **Valverde EJ, Cano I, Labella A, Borrego JJ, Castro D.** 2016. Application of a new real-time polymerase chain reaction assay for surveillance studies of lymphocystis disease virus in farmed gilthead seabream. BMC Vet Res **12:**71. http://dx.doi.org/10.1186/s12917-016-0696-6.
59. **Palmer LJ, Hogan NS, van den Heuvel MR.** 2012. Phylogenetic analysis and molecular methods for the detection of lymphocystis disease virus from yellow perch, Perca flavescens (Mitchell). J Fish Dis **35:**661–670. http://dx.doi.org/10.1111/j.1365-2761.2012.01391.x.
60. **Koch EA, Dolowy WC, Spitzer RH, Greenberg S, Brown ER.** 1976. Postulated developmental forms in the life cycle of the lymphocystis virus. Cancer Biochem Biophys **1:**163–166.
61. **Robin J, Berthiaume L.** 1981. Purification of lymphocystis disease virus (LDV) grown in tissue culture. Evidences for the presence of two types of viral particles. Rev Can Biol **40:**323–329.