Vol. 130: 95–108, 2018
https://doi.org/10.3354/dao03261

DISEASES OF AQUATIC ORGANISMS
Dis Aquat Org

Published September 10

Isolation, characterization and molecular identification of a novel aquareovirus that infects the endangered fountain darter *Etheostoma fonticola*

L. R. Iwanowicz[1,*], D. D. Iwanowicz[1], C. R. Adams[1], T. D. Lewis[2], T. M. Brandt[3], L. R. Sanders[1], R. S. Cornman[4]

[1]US Geological Survey, Leetown Science Center, 11649 Leetown Road, Kearneysville, WV 25430, USA
[2]US Fish and Wildlife Service, Southwestern Native Aquatic Resources and Recovery Center, PO Box 219, Dexter, NM 88230, USA
[3]US Fish and Wildlife Service, San Marcos Aquatic Resources Center, 500 East McCarty Lane, San Marcos, TX 78666, USA
[4]US Geological Survey, Fort Collins Science Center, 2150 Centre Avenue Building C, Fort Collins, CO 80526, USA

ABSTRACT: The fountain darter *Etheostoma fonticola* (FOD) is a federally endangered fish listed under the US Endangered Species Act. Here, we identified and characterized a novel aquareovirus isolated from wild fountain darters inhabiting the San Marcos River. This virus was propagated in Chinook salmon embryo (CHSE)-214, rainbow trout gonad-2 and fathead minnow cells at 15°C. The epithelioma papulosum cyprini cell line was refractory at all temperatures evaluated. High throughput sequencing technologies facilitated the complete genome sequencing of this virus utilizing ribosomal RNA-depleted RNA extracted from infected CHSE-214 cells. Conventional PCR primer sets were developed for the detection and confirmation of this virus to assist diagnostic screening methods. Phylogenetic analysis suggests this virus belongs to the *Aquareovirus A* genus. This research provides requisite initial data critical to support hatchery and refugia biosecurity measures for this endangered species.

KEY WORDS: Virus · Aquareovirus · *Etheostoma fonticola* · Fountain darter · Endangered

INTRODUCTION

The fountain darter (FOD) *Etheostoma fonticola* is a federally endangered fish listed under the US Endangered Species Act. Like many threatened and endangered fishes of the southwestern USA, FOD populations are threatened due to habitat degradation. These fish exclusively inhabit the headwaters of the San Marcos and Comal Rivers of central Texas, USA (Hubbs et al. 2008). A recovery plan has been developed to support long-term survival (Brandt et al. 1993, Bonner & McDonald 2005). During the 1950s, drought conditions led to the cessation of flow in critical headwaters of Comal Springs that consequently extirpated fountain darters from this system (Schenck & Whiteside 1976). Fountain darters from the San Marcos River were later used to reintroduce this species to the Comal River. Due to the risk of future drought and exotic pathogens, maintenance of captive-reared broodstock is a critical component of the endangered species recovery plan for the fountain darter.

Informed, conservative biosecurity measures are paramount for the sustainability of successful captive rearing programs for endangered species revitalization programs. Given that wild fishes are used as broodstock for such programs, common and exotic pathogens pose a perpetual risk of disease. Fountain darters are currently maintained at the US Fish and Wildlife Service (USFWS) Uvalde National Fish Hatchery and the USFWS San Marcos Aquatic Resources Center. During 2003, an unknown virus was isolated

*Corresponding author: liwanowicz@usgs.gov

from fountain darters originating from the San Marcos River. Subsequently, proactive biosecurity measures were implemented such that darters collected from the San Marcos River would only be held in refugia at the San Marcos Aquatic Resources Center to minimize the risk of unintentional transfer of virus to other stocks, wild populations, or ecosystems. These biosecurity measures were implemented until this virus could be identified and its pathogenicity risk evaluated.

Here, we identified and characterized a novel aquareovirus isolated from wild fountain darters using classical and contemporary methods, including next generation sequencing (NGS) technologies. We identified cell lines (selected from those commonly used by the USFWS for diagnostic screening) and incubation temperatures for practical diagnostic isolation. Diagnostic end-point PCR for the detection and confirmation of this virus were also developed to assist screening methodology. These requisite initial characterization data are critical to support hatchery and refugia biosecurity measures for this endangered species.

MATERIALS AND METHODS

Virus isolation and preparation of virus suspension

During March 2003, fountain darters were collected from the San Marcos River and evaluated for the presence of infectious agents at the USFWS Pinetop Fish Health Center (PTFHC), Pinetop, Arizona (now part of the Southwestern Native Aquatic Resources and Recovery Center). Tissues were processed for virology, and a cytopathic effect (CPE) characterized by foci of syncytia was observed in Chinook salmon embryo (CHSE)-214 cells. A filterable agent that induces this CPE has been periodically isolated from fountain darters collected from the San Marcos River, but not the Comal River (Table 1). While the putative virus was unidentified, it was presumed to be an aquareovirus of unknown significance based on CPE alone.

Table 1. US Fish and Wildlife Service (USFWS) history of cell line isolations of the *Etheostoma fonticola* reovirus (EFReV) based on the observation of syncytial cytopathic effect (CPE). Sample size was 60 or 10 (denoted with an asterisk) individuals unless indicated otherwise

Date	San Marcos River	Comal River
24 March 2003	Yes	–
1 April 2003	–	No
3 June 2003	Yes	–
9 December 2003	No	–
7 January 2004	No	No
3 May 2004	Yes	No
15 June 2006	No	No
4 June 2007	Yes	No
17 September 2008	No	No
9 July 2009	Yes	No
27 January 2010	No*	No
4 April 2010	No*	–
6 June 2010	No*	–
17 August 2010	No*	No*
18 July 2011	No*	No*
21 September 2011	No*	No*
14 November 2011	–	No*
1 January 2012	No*	No*
3 March 2012	No*	No*
14 May 2012	Maybe	No
23 July 2012	No*	No*
25 September 2012	No (49 fish)	No
28 January 2013	No*	No*
25 March 2013	No	No
17 June 2013	No*	No*
23 September 2013	No (59 fish)	No (59 fish)
9 December 2013	No*	No*
14 February 2014	No*	No*
14 May 2014	Yes	No (53 fish)
18 November 2014	No	No
12 May 2015	No	No
12 November 2015	No	No

Cell line susceptibility

We evaluated 5 piscine cell lines commonly used for diagnostic screening of fish viruses to determine their permissibility to viral infection. Cell lines included the CHSE-214, fathead minnow (FHM), rainbow trout gonad (RTG-2), epithelioma papulosum cyprini (EPC), and bluegill fry (BF-2) cultured in Eagle's minimal essential medium with Hanks' balanced salt solution (Sigma Chemical) supplemented with 10% fetal bovine serum (Sigma Chemical) as described in the USFWS and American Fisheries Society – Fish Health Section Blue Book screening method for viral isolation (USFWS and AFS-FHS 2014). Media from freshly split monolayers were removed and cells were inoculated in 10-fold serial dilutions with virus inoculum. Cell lines were incubated at 15 and 25°C for 28 d. Cells were observed daily for evidence of CPE.

Molecular characterization

The CHSE-214 cell line was inoculated with clarified supernatant from cells from the original PTFHC isolation that exhibited CPE. Cells were incubated at

20°C for 120 h. A syncytial CPE was evident 48 h post-inoculation. The supernatant was harvested and centrifuged at 2000 × *g* for 20 min. The pellet and remaining adherent cells in the tissue culture flask were dislodged with a cell scraper and lysed in TRK lysis buffer (700 µl TRK lysis buffer / 1 × 10^7 cells). Total RNA was extracted using an EZNA RNA kit (Omega Bio-tek) as per manufacturer's instructions using the on-column DNase treatment protocol. The RNA was eluted and quantified using a Nanodrop (ThermoFisher Scientific).

Conventional PCR

Total RNA was heat denatured at 65°C for 10 min, immediately cooled in an ice bath and then synthesized into cDNA using a High-Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific). This approach selected for viral mRNA. Heat denaturation >95°C is necessary to melt genomic dsRNA (the estimated melting temperature range for genomic segments here is 91.2 to 93.2°C; Maan et al. 2007). We first attempted to confirm the identity of the viral agent using a PCR assay developed for a broad range of aquareoviruses (Seng et al. 2004). Conditions for the PCR reaction were as described previously; however, GoTaq Green master mix (Promega) was used instead of individual master mix components.

Preparation of RNA for NGS

Ribosomal RNA was depleted from the total RNA sample using a Ribo-Zero rRNA Removal Kit (H/M/R) (Illumina) according to manufacturer's instructions. This approach was selected as not to bias for polyadenylated RNA molecules. The RNA was then prepared for NGS using the Ion Torrent platform. An additional pool of RNA was prepared from CHSE-214 cells infected for 96 h as above for sequencing using the MiSeq platform.

Ion Torrent PGM library preparation and sequencing

An Ion Torrent Total RNA-Seq kit v2 (Life Technologies) was used to prepare a library for sequencing. Briefly, 100 ng of rRNA-depleted RNA was fragmented with RNAse III at 37°C for 10 min, diluted with 1 volume of nuclease-free water and placed on ice. Fragmented RNA was purified with magnetic beads and solutions provided with the kit.

The resulting library was purified using AMPure XPbeads (Beckman Coulter) and the concentration and size determined using a BioAnalyzer DNA High-Sensitivity LabChip (Agilent Technologies). Sample emulsion PCR, emulsion breaking, and enrichment were performed using the Ion Xpress Template Kit, according to manufacturer's instructions. Briefly, an input concentration of 1 DNA template copy/Ion Sphere Particles (ISPs) was added to the emulsion PCR master mix and the emulsion generated using an IKA DT-20 mixer (Life Technologies). Next, ISPs were recovered and template-positive ISPs enriched for using Dynabeads MyOne Streptavidin C1 beads (Life Technologies). ISP enrichment was confirmed using the Qubit 2.0 fluorometer (Life Technologies), and the sample was prepared for sequencing using the Ion Sequencing Kit protocol. The complete sample was loaded on an Ion 314 chip and sequenced on the PGM for 65 cycles.

MiSeq library preparation and sequencing

Total RNA from CHSE-214 cells infected for 96 h was enriched for non-ribosomal RNA using a Ribo Zero kit as described above. Ribosomal RNA depleted RNA was then quantified using a Qubit and a sequencing library was constructed using the NEBNext® Ultra™ RNA library kit for Illumina according to manufacturer protocols (NewEnglandBio Labs). The library was indexed to accommodate other samples included in the sequencing run and quantified using the KAPA Library Quantification Kit for Illumina (Roche) and 10 pM of library was sequenced on a MiSeq (2 × 101 PE; Illumina).

Draft genome assembly

PGM reads were processed with CLC Genomics Workbench v.8.1.1 (Qiagen) to remove adapter sequences and low-quality reads. Bases were trimmed at an error probability of 0.01 and reads shorter than 50 bp were discarded. Over-representation analysis of processed reads identified an additional CCG GAA GGC G motif primarily at the 5′ end of sequences; perfect matches to this sequence at 5′ or 3′ ends of reads were removed and those shorter than 40 bp after this step were discarded.

We used 2 approaches for assembly of the raw reads. First, we used CLC Genomics Workbench to assemble all trimmed reads using automatic kmer and bubble-size selection. Second, we used the same

method to assemble the subset of reads that had TBLASTX matches to the related turbot reovirus (*Scophthalmus maximus* reovirus; SMReV). The latter approach proved slightly better in terms of the lengths of best-matching contigs to SMReV. A single long contig was identified as the putative ortholog of 10 of the 11 SMReV segments in GenBank (accession nos. HM989930–HM989940), whereas 2 non-overlapping contigs were identified as homologous to the 5′ and 3′ ends of SMReV segment 4 (M4). Pairwise nucleotide alignments of these segments were performed with MUSCLE (Edgar 2004) using default parameters.

As most contigs failed to extend to the ends of their SMReV counterparts, we performed an additional round of assembly to aggressively extend the *Etheostoma fonticola* reovirus (EFReV) contigs further. We used 60 bp regions at the 5′ and 3′ ends of contigs as seeds for both BLASTN and TBLASTX searches (e-value set to 1e–2) against the raw sequence reads, as well as for searches with the hash-based program SMALT (https//www.sanger.ac.uk/science/tools/smalt-0) using a hash length of 15. All matching reads were aggressively assembled along with the original contigs using CAP3 (Huang & Madan 1999), with a minimum overlap of 16, a 90% percent identity, and an overhang allowance of 98% in order to allow assembly of small overlaps. Manual alignment of potentially informative reads was also performed. The longest contig matching each SMReV accession was retained for further evaluation, with the exception that we again recovered 2 non-overlapping contigs homologous to the M4 segment.

We evaluated the assembly quality by re-mapping sequence reads with Bowtie2 (Langmead & Salzberg 2012), using the 'local' and 'very-sensitive-local' search modes. Tablet (Milne et al. 2010) was used to view the contigs in order to identify regions of very high or very low coverage that might represent assembly errors. We also evaluated read depth and read-pair concordance in genomic regions that diverged substantially from SMReV. Even so, the sequence identity was generally good between the 2 genomes and all SMReV accession began with the conserved motifs GTTTT and ended with CATC, such that the combination of pairwise alignment, read mapping, and presence of the conserved motifs could be used to guide manual editing and confirm completed segments. Manual editing consisted of trimming unsupported end sequence from contigs and determining whether a few frameshifting indels could be recovered by adding a base that was supported in the read pileup.

Genome finishing

Due to the persistence of a gap in the presumed M4 segment, as well as low-coverage regions, incomplete segment ends, and unresolved frameshifts in the manually edited draft assembly, we performed genome finishing with targeted Sanger sequencing and the data set from the MiSeq run. Primers for targeted sequencing (Table S1 in the Supplement at www.int-res.com/articles/suppl/d130p095_supp.pdf) were designed in Primer3 bundled within Geneious 7.1.5 (Biomatters) using default parameters, and ordered from Integrated DNA Technologies. All PCR amplifications used GoTaq Green master mix (Promega). Thermal cycling conditions for end-point PCR primer sets were 95°C for 5 min; cycle 30×; 95°C for 30 s, 56°C for 30 s, 72°C for 40 s; a final extension at 72°C for 5 min, followed by a 10°C hold. The 5′ and 3′ segment termini were Sanger sequenced using the anchor spacer-ligation method (Attoui et al. 2000, Maan et al. 2007). PCR products were resolved in 2% agarose matrices by electrophoresis at 90 V for 90 min and stained with GelRed™ (Biotium). All PCR products were purified using DNeasy PCR purification kits (Qiagen) and prepared for direct sequencing. Direct sequencing was conducted using BigDye v.3.1 chemistry. Sequences were used to complete segment M4 and manually edit contigs from the PGM sequencing assembly.

Reads were trimmed of adapters and poor-quality sequence as above using MiSeq-specific search models. Processed reads were then mapped to each contig corresponding to the 11 segments of the EFReV to resolve frameshift ambiguities, and to provide sufficient coverage to purge low-frequency variants and sequencing errors from the final assembly. A final round of mapping against the finished reference genome was conducted using all reads from the PGM and MiSeq runs to determine coverage metrics on a per segment basis. The final assembly was deposited in GenBank as GCA_001678455.2.

Phylogeny

Sequence identity-based species demarcation criteria for aquareoviruses are based on amino acid identity of the RNA-dependent RNA polymerase (RdRp), and nucleotide or amino acid identities of the outer capsid protein (VP7; segment 10). We performed phylogenetic analyses for both loci. Available RdRp sequences were acquired from the NCBI database and aligned using default parameters of MUS-

CLE, bundled in Geneious (v.7.1.5). They included American grass carp reovirus (YP_001837095), Atlantic salmon reovirus (ABO32573), Atlantic halibut reovirus (AIY69147), avian orthoreovirus (AED99918), chum salmon reovirus (NP_398630), *Etheostoma fonticola* aquareovirus, fall Chinook reovirus (YP_009259508), golden shiner reovirus (NP_938061), grass carp hemorrhagic reovirus (AF284502), grass carp reovirus (AAG10436), grass carp reovirus 109 (AHD25636), grass carp reovirus HeNan988 (AGR-34045), grass carp reovirus HuNan794 (AGG38806), grass carp reovirus Jx02 (AGQ21748), grass carp reovirus HZ08 (ADJ75336), mammalian orthoreovirus 3 (NP_003199418), *Micropterus salmoides* reovirus (AJD09447), *Scopthalmus maximus* reovirus (ADZ-31977), striped bass reovirus (AF450318) and white bream reovirus (AEC53507). Alignments were imported into MEGA v.7.0.14 and were used to determine the best amino acid substitution model using maximum likelihood. Bayesian inference of phylogeny was determined using the LG=G substitution model in MrBayes v.3.2.6 (Ronquist & Huelsenbeck 2003). Mammalian orthoreovirus 3 was set as the outgroup.

Phylogeny based on the segment 10 open reading frame (ORF) (VP7 or homologous segment) nucleotide sequence was determined. When available, we utilized the nucleotide sequences of segment 10 from the same viruses used in the RdRp analysis. We included additional aquareoviruses for this analysis for which RdRp sequence was not available. Viruses and accession numbers are included (see Fig. 3). The HKY+G nucleic substitution model was identified as the best fit. Phylogenetic analysis was also performed using amino acid sequence, but that analysis yielded less resolution.

Investigating sequence-based species demarcation criteria

We applied the Sequence Demarcation Tool (v.1.2) to compare demarcations based on percent identity of RdRp and VP7 (Muhire et al. 2014) with current species classifications. Sequence alignments included in these analyses were identical to those used for the phylogeny. Matrices were produced for each protein using the International Committee on Taxonomy of Viruses (ICTV) species thresholds. Specifically, the matrix color cut-offs for the RdRp were set to 95 and 70%. The upper and lower cutoff values for VP7 were 55 and 35% respectively. Alternative matrix thresholds were then manually selected based on the peak and trough boundaries of the pairwise identity frequency distribution plot.

Homology and conservation of aquareovirus genome segments

Evolutionary conservation of EFReV and its predicted ORFs was assessed by comparison to 8 representative draft or complete genomes representing aquareovirus phylogenetic clusters A, B, C, G as well as an unclassified group of grass carp reoviruses. Note that SMReV is not classified as an aquareovirus in Genbank (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=10979), but can be assigned to *Aquareovirus A* based on recent analyses (Makhsous et al. 2017).

To identify homologous protein coding regions in each of these genomes, translated EFReV ORFs were searched with TBLASTN (default parameters) against the following accessions: *Micropterus salmoides* reovirus (KJ740724.1–KJ740734.1), SMReV (GCA_002829525.1), fall Chinook aquareovirus (FCReV, GCA_002288715.1), chum salmon reovirus (GCA_000866805.1), Aquareovirus C (GCA_000853585.1), American grass carp reovirus (GCF_000879275.1), grass carp reovirus isolate GCR918 (KC201177.1–KC201187.1), and Green River Chinook virus (KC588376.1–KC588385.1). ORFs for each of these genomes were identified with getorf of the EMBOSS package and those encompassing the TBLASTN matches to each EFReV protein were codon-aligned with CLUSTALW (Guo et al. 2014). Alignments were manually edited to remove ORFs of questionable homology, or where possible to extend ORFs in genomes with frame-shift mutations. Note the failure to identify a complete homolog of a given EFReV ORF in another aquareovirus genome may indicate that the homologous segment is genuinely absent, the homologous segment is too divergent to be recognized, or a gap or error exists in either genome. Thus, this analysis emphasizes general patterns across genome segments and recognizes that specific pairwise comparisons may be improved with additional data.

The divergence rate parameter (ω) of each ORF, i.e. the rate that nonsynonymous substitutions accrue relative to synonymous ones, was estimated with the codeml function of the PAML package (Yang 2007). As a tree topology specific to the genomes analyzed must be specified for this analysis, we estimated a maximum likelihood topology from 4 ORFs deemed

to be complete in all of these genomes (ORF1, ORF2, ORF5, and ORF6). Each set of ORFs were aligned at the protein level using CLUSTALW and then reverted to nucleotide sequence. Nucleotide model selection (GTR+G) and tree topology estimation were performed in MEGA7 (Kumar et al. 2016). A neighbor-joining tree was created using the maximum composite likelihood (MCL) substitution model and a gamma value of 0.55 for the distribution of rate variation among sites. A second tree created using maximum likelihood and 5 estimated categories of rate variation had the same topology.

For ORFs 7A and 7B, alignments were manually trimmed to the alignable parts of the coding sequence, due to ambiguities in the boundaries of coding sequence for these overlapping ORFs. For ORF4, a central region that could not be aligned across all species was removed. PAML was configured to specify ω under the 'M0' model (i.e. a single value for the entire alignment for each ORF). When no ORF sequence was available for a given isolate, the corresponding branch was pruned from the guide tree for that run. Two iterations were performed for each ORF to confirm the stability of the parameter estimates. The relative rate of transitions, specified by the parameter κ and co-estimated with ω, was inspected as an indicator of potential misalignment that might have impacted the ω estimate (transition ratios outside of 'typical' values for an organism may denote extreme saturation or alignment error; Rosenberg et al. 2003).

EFReV segment lengths, their median coverage in consecutive 50 bp windows, EFReV ORF boundaries, length and percent identity of their best TBLASTN match, and the estimated ω for each ORF alignment were integrated into a circular comparative plot using Circos (Krzywinski et al. 2009). Note that high scoring pairs (HSPs) with the 2 ORFs on segment 7 are represented as overlapping tracks. Inferred ORF homology and divergence rates across genomes are indicated in Table S2 in the Supplement.

RESULTS

Cell line susceptibility

All cell lines except EPCs were permissive to viral infection. The characteristic CPE consisted of syncytia, and was observed at both temperatures evaluated (Fig. 1). The FHM cultures at 15°C developed CPE with cleared plaques, uncharacteristic of CPE in CHSE-214 cultures at 15°C, while FHM at 25°C developed only mild CPE at equivalent titer inoculation.

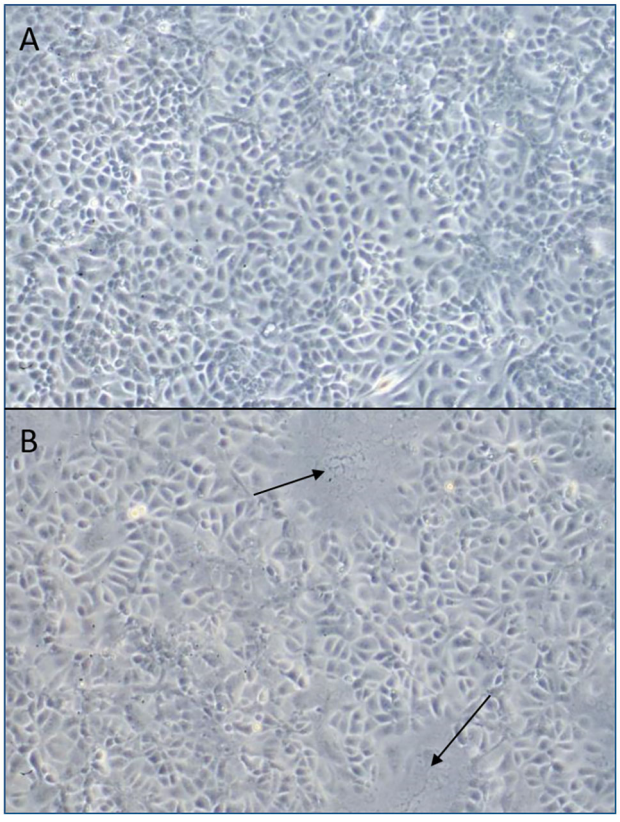

Fig. 1. (A) Control and (B) infected Chinook salmon embryo 214 cells. Cells were infected with *Etheostoma fonticola* reovirus and incubated for 72 h. Cytopathic effect of syncytia formation (arrows) was observed

Conventional PCR

Attempts to amplify product from infected CHSE-214 cells with the degenerate aquareovirus primer set were unsuccessful. *In silico* evaluation of the universal degenerate primer set confirmed that this primer set was not adequate to amplify this virus or several others for which nucleotide sequence was not available at the time of development of the universal method (see Fig. S1 in the Supplement). The application of NGS was therefore utilized to confirm the identity of the virus.

Genome sequencing

The combination of Ion Torrent and Illumina RNA sequencing of ribosomal RNA depleted RNA from

CHSE-214 infected cells led to the almost complete sequencing of all EFReV segments (Fig. S2). Dideoxy sequencing was required to close a gap in segment 4 (nt 449 to 1066). The initial PGM sequencing led to a total of 1 720 373 reads (average = 102 nt) of 4.8% of reads mapped to the final EFReV genome. A large minority of reads (43.4%) mapped to contigs with homology to a mycoplasma, indicating contamination of the cell culture but irrelevant to the reconstruction of EFReV. The PGM run was deposited under BioProject PRJNA431729. The MiSeq sequencing run yielded 9 218 424 reads (average length 100 nt). A slightly smaller proportion of reads mapped to the final EFReV genome (3.5%).

The final EFReV reference genome was 23 958 bp of dsRNA, consisting of 11 segments. The average composite mapping coverage of each segment ranged from 241 to 9133× coverage (Fig. 2, Table S3 in the Supplement). Length-normalized sequencing depth was greatest for segment 10 and lowest for segment 1. Given that these are dsRNA viruses, the difference in coverage is likely the result of viral transcription in addition to genome replication with in the CHSE-214 cell line at the time of harvest. BLASTX queries of all segments in the NCBI database best matched aquareoviruses, and both genome organization and protein identity was consistent with other aquareoviruses (Table 2). The terminal sequences of the segments were GUUUUA(U/G/A) and (A/U)UCAUC. All conserved terminal sequences were consistent with that observed for *Aquareovirus A*, with the exception that the 5′ terminal motif for S4 was typical of *Aquareovirus G*. The GC content was 56.5% and consistent with that from other aquareovirus (Attoui et al. 2002). Sequences for the 11 segments were deposited into GenBank and a RefSeq genome (NC_030405–NC_030416; GCA_001678455.2) has been compiled (Iwanowicz et al. 2016).

Phylogeny

Based on pairwise analysis of the RdRp protein, VP2, the EFReV was most similar (≥90%) to the aquareoviruses previously classified as *Aquareovirus A* (Fig. 3A). Amino acid sequence identity ranged from 90 to 96%. The phylogenetic tree based on this gene depicts a well-supported (posterior probability = 1.0) clade consisting of striped bass reovirus, EFReV, chum salmon reovirus, Atlantic salmon reovirus, SMReV, and *Micropterus salmoides* reovirus (MSReV). The Atlantic halibut reovirus (unassigned) was the sole representative of a distinct branch between *Aquareovirus A* and *Aquareovirus B* representatives. Two other well-defined clades of aquareoviruses that infect cyprinids were resolved. One included *Aquareovirus C* and *Aquareovirus G*, while the other clade included a distinct set of unclassified reoviruses from Chinese grass carp. Based on analyses with the species demarcation tool, pairwise identity cut-offs for the RdRp of these aquareovirus species was 85 to 90% (Fig. 4).

Similar relationships were identified in the phylogenetic tree topology for these viruses based on nucleotide sequences for VP7 (Fig. 3B). This higher resolution (nucleotide) analysis resolved the *Aquareovirus A* representatives that infect salmonids as a distinct well-supported clade (posterior probability = 1.0); albeit this may be an artefact of limited representation of salmonid aquareovirus sequence availability for comparison. Nucleotide % identity ranged from 71 to 76% within the clade containing the new EFReV. Pairwise identity of the EFReV and aquareoviruses that resolved in other clades ranged from 33 to 46%.

We identified and graphically depicted the most highly conserved protein coding regions across aquareovirus species relative to the EFReV (Fig. 2). While the %ID was high between the EFReV, SMReV, MSReV and chum salmon reovirus (*Aquareovirus A*), the length of orthologous regions in the latter were much shorter relative to EFReV. Across these presumed *Aquareovirus A* genomes, the putative FAST protein within segment 7 was identified in only 5 genomes and had the highest rate of amino acid divergence ($\omega = 0.62$). Note that this ORF is predicted to have a non-standard start codon (CTG) in EFReV, MSReV, and SMReV, but a standard ATG start was present in fall Chinook aquareovirus (genotype group B1) and *Aquareovirus C*. In general, the larger genomic segments were identified as being under greater purifying selection (indicated by lower omega values) than the smaller genomic segments.

Genomic diversification of aquareoviruses

Patterns of similarity across the genome accord well with the phylogenetic pattern at the marker loci. We identified a core genome conserved in the aquareovirus and grass carp reovirus clade that consists of segments 1 to 6. Absence of segment 4 in Green River Chinook virus is likely an artefact, as the genome reference for that species consists of only 10 segments. Protein sequence in this ‘core genome’

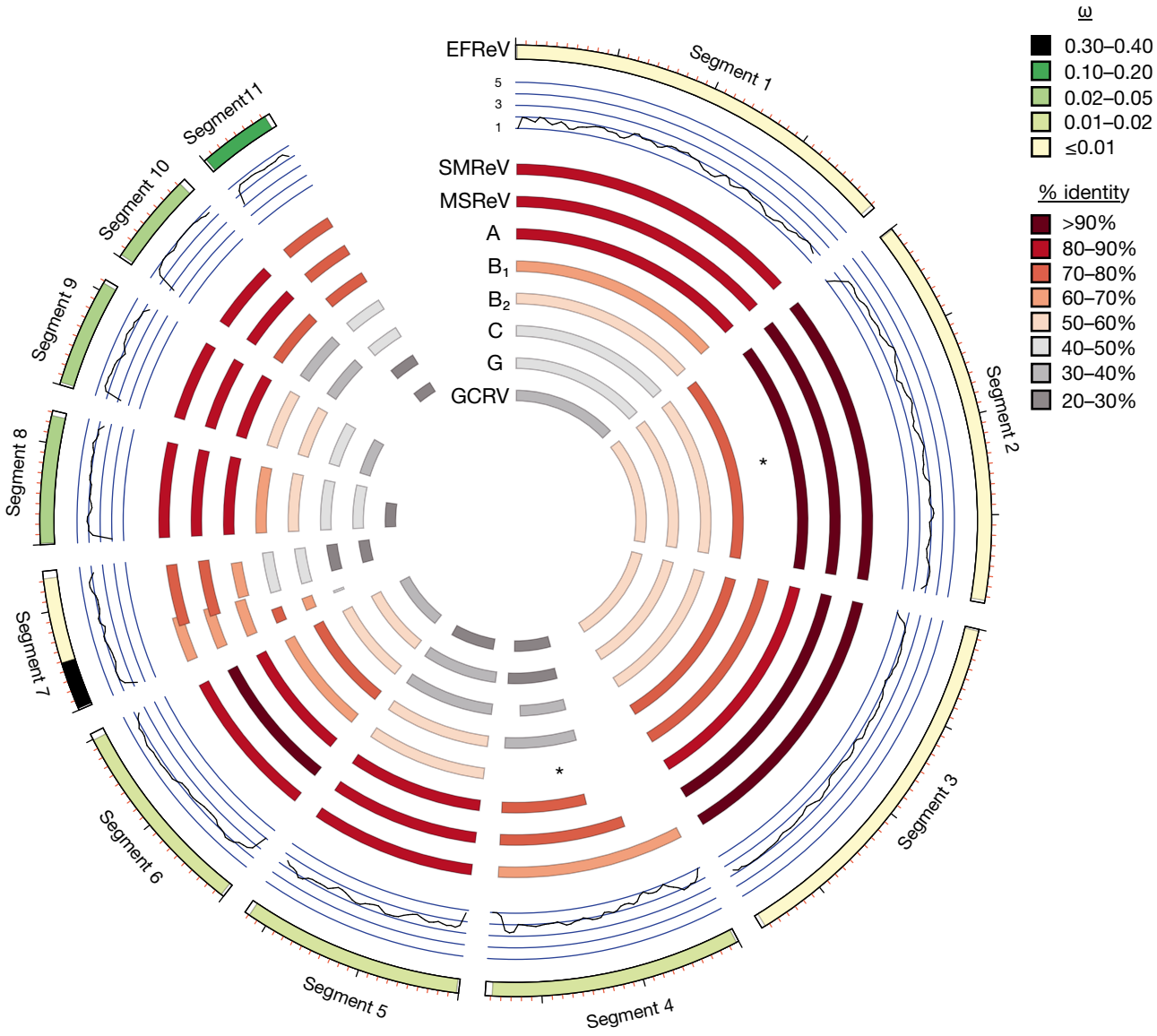

Fig. 2. Circular representation of divergence among aquareovirus genomes, including representative *Aquareovirus A*, *B*, *C*, *G*, and unclassified cyprinid reoviruses (GCRV). Depth of sequencing coverage (mean coverage in 50 bp windows, $\log_{10}$ scale) across the length of the dsRNA genomic segment is indicated below each *Etheostoma fonticola* reovirus (EFReV) segment. Coordinates of the best high scoring segment pair (HSPs; ⋆) in each genome for each EFReV open reading frame (ORF) identified by TBLASTN are depicted in each track, with the % amino acid identity denoted by the indicated color gradient. Estimates of the parameterized ratio of nonsynonomous to synonymous substitution (ω) are indicated by shading the respective EFReV coding region according to the indicated color scale. HSPs identify pairwise blocks of amino acid similarity, whereas omega represents the evolutionary rate of the ORF as a whole integrated across all included genomes

evolves at least several fold less rapidly than in the remaining segments (mean ω of 0.011 and 0.137, respectively; see Table S2). Phylogenetic analyses at the standard loci revealed some of the grass carp reoviruses to be divergent from the aquareovirus clade (results given above). Other genomic features are concordant with this hypothesis. In addition to lacking obvious homologs for EFReV ORFs 7A, 7B, 9, 10, and 11, several of the proteins encoded by the conserved segments (including ORFs 1, 2, 3, and 5) have 5′ regions in grass carp reovirus that are uniquely divergent in structure or composition. HSPs between grass carp reovirus and EFReV were generally the shortest and had the lowest percent identity. Of note, we identified that the 4 shortest segments (8 to 11) had high levels of expression and higher rates of evolutionary divergence relative to segments 1 to 6. The estimated transition rate bias was relatively low, approaching equality (kappa = 1) in rates of transition and transversion substitutions. Transi-

Table 2. Characteristics of genome segments, conserved terminal untranslated regions (UTRs), predicted protein identity biochemistry of *Etheostoma fonticola* reovirus. ORF: open reading frame. Protein pI and MW were calculated using the Compute pI/MW tool (http://web.expasy.org/compute_pi/pi_tool-doc.html)

Segment	Nucleotide				Deduced protein		
	Accession no.	Segment length (bp)	Conserved terminal UTRs	Nucleotide coordinates of ORF	pI	MW (kDa)	Predicted protein identity
S1	KU194213	3944	5′-GUU UUA U 3′-AUU CAU C	14-3907	6.04	140.64	VP1, putative guanyl/ methyl transferase
S2	KU194214	3866	5′-GUU UUA U 3′-AUU CAU C	13-3837	8.69	140.87	VP2, RNA-dependent RNA polymerase
S3	KU194215	3687	5′-GUU UUA U 3′-UUU CAU C	19-3648	5.91	131.05	VP3, putative helicase, NTPase
S4	KU194216	2554	5′-GUU UUA A 3′-AUU CAU C	25-2484	5.67	89.09	NS89
S5	KU194217	2236	5′-GUU UUA U 3′-AUU CAU C	22-2190	7.08	79.93	VP4, putative NTPase
S6	KU194218	2057	5′-GUU UUA U 3′-AUU CAU C	29-1990	4.47	69.31	VP5, outer capsid protein
S7	KU194219	1400	5′-GUU UUA G 3′-UUU CAU C	17-643 489-1325	6.29 9.27	31.98 23.41	NS32 NS23, putative FAST protein
S8	KU194220	1317	5′-GUU UUA U 3′-UUU CAU C	13-1266	8.87	45.42	VP6
S9	KU194221	1118	5′-GUU UUA G 3′-AUU CAU C	26-1078	6.52	38.04	NS38
S10	KU194222	987	5′-GUU UUA G 3′-AUU CAU C	28-924	7.54	32.31	VP7, outer capsid protein
S11	KU194223	783	5′-GUU UUA G 3′-UUU CAU C	25-732	7.77	25.37	NS25
Consensus			5′-GUU UUA $^{A}/_{G}/_{U}$ 3′-$^{A}/_{U}$UU CAU C				

tion rate biases reported for other viruses are generally higher, although estimates can be impacted by long branches (Duchêne et al. 2015). The consistency of kappa across segments suggests no major alignment errors affected estimates of evolutionary rate.

Diagnostic PCR

The PCR primers designed here consistently amplified product of the appropriate size for the target region (Fig. S3). They did not amplify spurious products in negative controls under the conditions used. These primer sets effectively amplified a single product using an annealing temperature of 56°C except the primer set for segment 4. At this annealing temperature a double band was observed (not shown); however, increasing the annealing temperature to 60°C resolved a single amplicon. While we did not evaluate if these primers amplified sequence of other aquareoviruses, *in silico* analyses of the publicly available aquareovirus sequences identify these primer pairs as specific to the EFReV. It should be emphasized that there are limited aquareovirus reference sequences in GenBank. Thus, if these primers are used in practice for diagnostic screening, amplicons should be sequenced to confirm virus identity. Additionally, total RNA needs to be melted at 100°C for 3 min prior to cDNA synthesis in order to include genomic RNA as a target for diagnostic screening (Maan et al. 2007).

DISCUSSION

The family *Reoviridae* is comprised of 9 genera that infect a range of host organisms including plants, fungi, insects, birds, molluscs, mammals, and fishes. Reoviruses that infect fishes belong to the *Aquareovirus* and a new, yet to be erected genus (Kibenge et al. 2013, Sibley et al. 2016); aquareoviruses have also been identified in molluscs. They are non-enveloped, icosahedral virions that contain a genome comprised of 11 segments of double-stranded RNA, and are morphologically and physicochemically similar to orthoreoviruses that infect mammals. For decades,

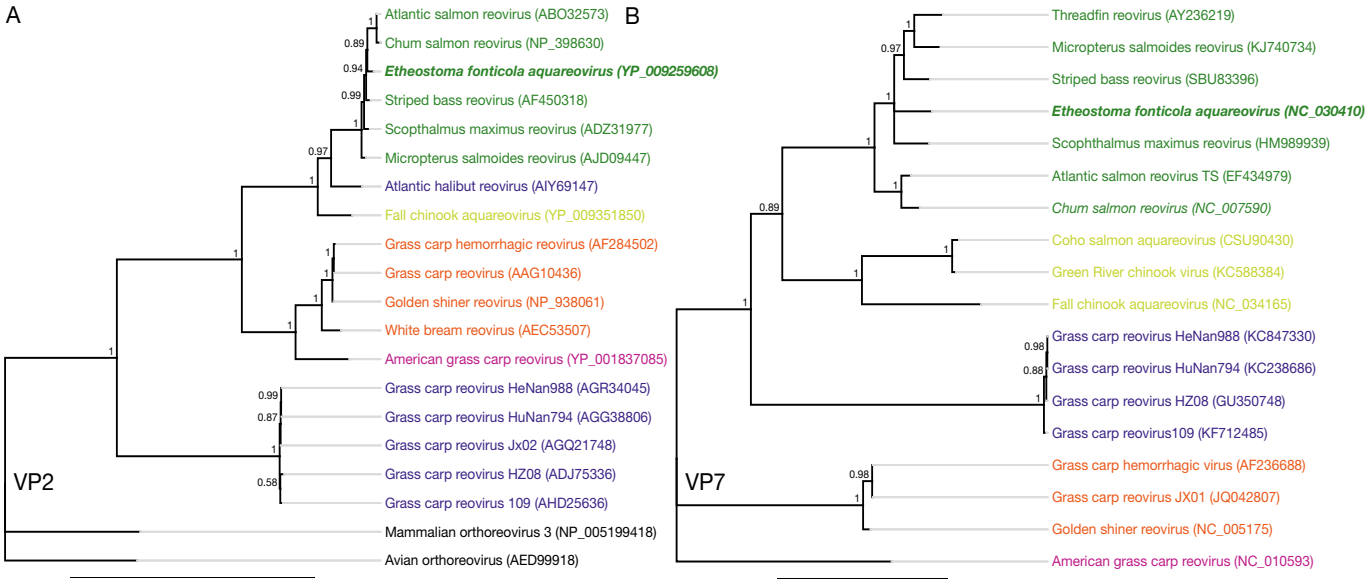

Fig. 3. Phylogenetic analysis of reoviruses based the (A) RNA-dependent RNA polymerase (RdRp) protein and (B) VP7 nucleotide sequences. Analyses were conducted via Bayesian analysis. Representatives of *Aquareovirus A*, *B*, *C*, and *G* are indicated in green, gold, orange, and magenta respectively. Unclassified reoviruses and indicated in blue and those that infect ectotherms are black. The *Etheostoma fonticola* reovirus is highlighted with *italicized*, **boldface** font

aquareoviruses have been isolated from moribund and clinically normal fish and shellfish (Lupiani et al. 1995). Given the regularity that these orphan viruses are isolated from apparently healthy individuals, the significance of these viruses as causative agents of disease is often ambiguous. The pathogenicity of representatives from this genus, however, ranges from apathogenic to highly virulent (Subramanian et al. 1994, Crane & Carlile 2009). Thus, it would be prudent to consider the potential significance of such viruses in fish species of high management interest and social value. Of note, virulent aquareoviruses are typically observed in fishes maintained in culture conditions, which either reflects a surveillance bias or perhaps biological relevance of stress (Cusack et al. 2001, Qiu et al. 2001, Blindheim et al. 2015).

There are 7 species, formerly genogroups (A to G), of the genus *Aquareovirus* recognized by the ICTV (King et al. 2011). A majority of the aquareoviruses isolated to date are classified in groups A and B, and several are unclassified. Species demarcation criteria in the genus have been based on RNA–RNA hybridization and sequence analysis of segments 2 and 10 (Rangel et al. 1999, Attoui et al. 2002). Here, we did not perform RNA–RNA dot blot hybridization to assign this virus to an *Aquareovirus* species; however, a combination of NGS and dideoxy sequencing was utilized to sequence the complete genome. We compared nucleotide and deduced amino acid sequences from this virus to those available in public databases to ascribe phylogeny. This includes sequence for segments 2 and 10 for a number of the type species in the genus *Aquareovirus*. At the time of this analysis, complete genomes were available for 4 of the 7 type species in addition to a number of unclassified reoviruses that infect grass carp (Wang et al. 2012). While not the focus of this investigation, our results suggest that this divergent clade of grass carp reoviruses is more different from the established genera of aquareoviruses sensu stricto than genogroups are from each other, and are not aquareoviruses by a genome content definition. Despite advances and reduced costs associated with viral genome sequencing, there is a paucity of aquareovirus sequences in publicly available data sets, and most represent viruses isolated from grass carp. Previous research has suggested that members of an aquareovirus species share >95% amino acid sequence identity within the RdRp (Attoui et al. 2002). This is also one of the species demarcation benchmarks accepted by the ICTV. Based on this criterion alone, the EFReV is not a definitive member of species *Aquareovirus A*. This novel virus, however, shared 70% nucleotide identity within VP7 to the chum salmon reovirus, which is much greater than the <55% identity benchmark for species demarcation. Our analyses using the species demarcation tool suggest that the EFReV is an *Aquareovirus A*, and

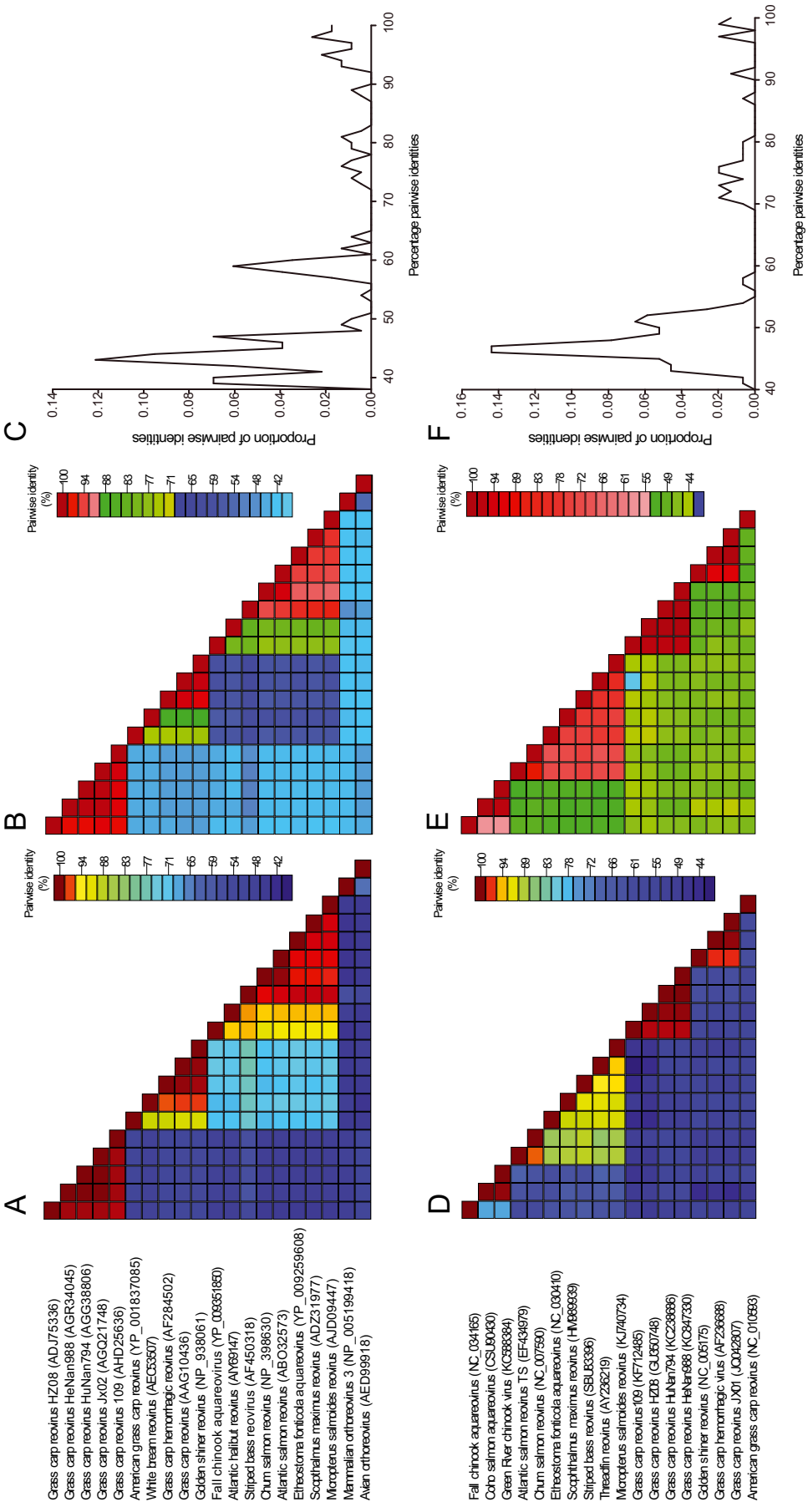

Fig. 4. Species demarcation tool (SDT) analysis of the RNA-dependent RNA polymerase (RdRp) protein and VP7 nucleotides. Analysis of sequences using ICTV demarcation thresholds for the (A) RdRp protein and (D) VP7 nucleotide sequence, with (B,E) less stringent thresholds indicated by (C,F) pairwise-identity thresholds indicated by the SDT tool. Troughs in the pair-wise identity plots identify best supported taxa thresholds

that the amino acid sequence identity criteria for the RdRp should be relaxed to 85–90% for this protein. As more aquareovirus genomes are sequenced there will likely be a need to revisit species demarcation criteria for this genus with perhaps the implementation of a method that utilizes a concatenated, multilocus approach.

We used putative universal aquareovirus primers at the onset of this research for molecular identification. That universal primer set did not amplify product, emphasizing that while these primers may be useful for many aquareoviruses, the assay is not universal. Pan-specific PCR primers have been developed by others, but it has become clear that there is more sequence diversity in aquareoviruses than previously suspected (Makhsous et al. 2017). Advances in sequencing technologies have provided a means to identify novel viruses at an unprecedented pace (Ho et al. 2014). The sequencing approaches applied here utilized ribosomal RNA depleted RNA from infected cells rather than enriched or purified viral genome. The combination of Ion Torrent PGM and Illumina MiSeq sequencing effectively captured complete ORFs from all segments except for segment 4, and captured terminal ends of over half of the segments. It also facilitated the identification of host cell viral response genes (not discussed here). This approach circumvents the need for ultracentrifugation, is amenable to multiplexing and is feasible for a typical diagnostic lab with the means to out-source sequencing to an external genomics core facility. It is critical to note here that a ribosomal depletion method was utilized rather than the conventional poly A enrichment method given that aquareoviruses and many other viruses are not polyadenylated. While some research teams have identified fish viruses in publicly available sequence databases, aquareoviruses are likely uncommon due in part to the frequent use of poly A selection for NGS transcriptome library preparation. Adoption of such an approach would augment the public sequence database, facilitate the development of specific diagnostic methods and clarify of taxonomic relationships within this genus.

Chen et al. (2015) suggested that aquareoviruses can be split into marine and freshwater origin. That work is based on the sequence analysis of the *Micropterus salmoides* reovirus and the assertion by the authors that *M. salmoiodes* (largemouth bass) is a brackish water fish. The fountain darter exclusively inhabits freshwater headwaters of 2 rivers in the USA. The San Marcos River is approximately 75 river miles (~120 rkm) long, and there is no straightforward connection to the marine environment other than migratory birds and invasive invertebrates. Regardless, the distinction of marine and freshwater aquareoviruses is less clear following this observation, and is likely a simple artefact of inadequate surveillance or comprehensive sequence-based identification of this genera.

From a practical diagnostician perspective, we established isolation and molecular identification tools. Similar to typical aquareoviruses, the CPE manifested by the EFReV is that of large syncytia (Racine et al. 2009). This CPE was observed in cell lines commonly used by fish health diagnosticians. Based on the conditions tested here, the preferred cell line for fountain darter aquareovirus isolation was the CHSE-214 cell line at 15°C. A combination of virus isolation on this cell line and specific amplification of the viral genome using the diagnostic primers described here should suffice for ongoing, annual screening and surveillance of this virus in wild fountain darters. Based on relative expression values for each segment, the S10 (VP7) is the best target sequence to enhance detection sensitivity given the greater sequence representation.

Here, we identified an aquareovirus isolated from apparently healthy fountain darters, sequenced the complete genome, identified diagnostic PCR primers, and identified permissive cell lines to facilitate future isolation and identifications of this virus. It is currently unknown whether this virus is a pathogen of concern in the endangered fountain darter. This virus is similar to those associated with disease (Chang et al. 2002, Ke et al. 2011) and does not directly interface with marine or brackish waters. Challenge studies are required to determine the pathogenicity of the fountain darter aquareovirus. Furthermore, it is possible that these viruses are only pathogenic in stressful environmental conditions. Stressors including elevated water temperature and habitat restriction due to drought conditions may lead to elevated cortisol, possibly increasing the risk of viral outbreaks. Stress is commonly associated with outbreaks of persistent viral infections in fish (Gadan et al. 2013). Given that aquareoviruses have been demonstrated to be pathogens in some fish species and circumstantially under stressful condition, establishing pathogenicity of this virus in cultured fountain darters is a critical priority.

Acknowledgements. The authors thank S. K. Buckholtz for his assistance with PCR and sequencing. The authors also thank staff at the PTFHC and M. Rodarte from the Southwestern Native Aquatic Resources and Recovery Center for

virus isolation and cell line sensitivity experiments. This study was funded by the USFWS and the USGS Quick Response and Fisheries Programs. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the US Government. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the USFWS.

LITERATURE CITED

Attoui H, Billoir F, Cantaloube JF, Biagini P, de Micco P, de Lamballerie X (2000) Strategies for the sequence determination of viral dsRNA genomes. J Virol Methods 89: 147–158

Attoui H, Fang Q, Mohd Jaafar F, Cantaloube JF, Biagini P, de Micco P, de Lamballerie X (2002) Common evolutionary origin of aquareoviruses and orthoreoviruses revealed by genome characterization of Golden shiner reovirus, Grass carp reovirus, Striped bass reovirus and golden ide reovirus (genus *Aquareovirus*, family *Reoviridae*). J Gen Virol 83:1941–1951

Blindheim S, Nylund A, Watanabe K, Plarre H, Erstad B, Nylund S (2015) A new aquareovirus causing high mortality in farmed Atlantic halibut fry in Norway. Arch Virol 160:91–102

Bonner TH, McDonald DL (2005) Threatened fishes of the world: *Etheostoma fonticola* (Jordan & Gilbert 1886) (Percidae). Environ Biol Fishes 73:333–334

Brandt TM, Graves KG, Berkhouse CS, Simon TP, Whiteside BG (1993) Laboratory spawning and rearing of the endangered fountain darter. Prog Fish-Cult 55:149–156

Chang SF, Ngoh-Lim GH, Kueh LFS, Qin QW, Seng EK, Sin YM (2002) Initial investigations into two viruses isolated from marine food fish in Singapore. Vet Rec 150:15–16

Chen ZY, Gao XC, Zhang QY (2015) Whole-genome analysis of a novel fish reovirus (MsReV) discloses aquareovirus genomic structure relationship with host in saline environments. Viruses 7:4282–4302

Crane MSJ, Carlile G (2009) Aquareoviruses. In: Mahy B, Van Regenmortel MHV (eds) Desk encyclopedia of animal and bacterial virology. Elsevier, Oxford, p 429–435

Cusack RR, Groman DB, MacKinnon AM, Kibenge FSB, Wadowska D, Brown N (2001) Pathology associated with an aquareovirus in captive juvenile Atlantic halibut *Hippoglossus hippoglossus* and an experimental treatment strategy for a concurrent bacterial infection. Dis Aquat Org 44:7–16

Duchêne S, Ho SY, Holmes EC (2015) Declining transition/transversion ratios through time reveal limitations to the accuracy of nucleotide substitution models. BMC Evol Biol 15:36

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Gadan K, Sandtrø A, Marjara IS, Santi N, Munang'andu HM, Evensen Ø (2013) Stress-induced reversion to virulence of infectious pancreatic necrosis virus in naïve fry of Atlantic salmon (*Salmo salar* L.). PLOS ONE 8:e54656

Guo Y, Ye F, Sheng Q, Clark T, Samuels DC (2014) Three-stage quality control strategies for DNA re-sequencing data. Brief Bioinform 15:879–889

Ho T, Tzanetakis IE (2014) Development of a virus detection and discovery pipeline using next generation sequencing. Virology 471–473:54–60

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868–877

Hubbs C, Edwards RJ, Garrett GP (2008) An annotated checklist of the freshwater fishes of Texas, with keys to identification of species, 2nd edn. Texas Academy of Science, Edinburg, TX

Iwanowicz LR, Iwanowicz DD, Adams CR, Lewis TD, Brandt TM, Cornman RS, Sanders L (2016) Complete genome sequence of a novel aquareovirus that infects the endangered fountain darter, *Etheostoma fonticola*. Genome Announc 4:e01405-16

Ke F, He LB, Pei C, Zhang QY (2011) Turbot reovirus (SMReV) genome encoding a FAST protein with a non-AUG start site. BMC Genomics 12:323

Kibenge MJT, Iwamoto T, Wang Y, Morton A, Godoy MG, Kibenge FSB (2013) Whole-genome analysis of piscine reovirus (PRV) shows PRV represents a new genus in family *Reoviridae* and its genome segment S1 sequences group it into two separate sub-genotypes. Virol J 10:230

King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) (2011) Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. Elsevier, Amsterdam

Krzywinski M, Schein J, Birol I, Connors J and others (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Lupiani B, Subramanian K, Samal SK (1995) Aquareoviruses. Annu Rev Fish Dis 5:175–208

Maan S, Rao S, Maan NS, Anthony SJ, Attoui H, Samuel AR, Mertens PP (2007) Rapid cDNA synthesis and sequencing techniques for the genetic study of bluetongue and other dsRNA viruses. J Virol Methods 143:132–139

Makhsous N, Jensen NL, Haman KH, Batts WN, Jerome KR, Winton JR, Greninger AL (2017) Isolation and characterization of the fall Chinook aquareovirus. Virol J 14:170

Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D (2010) Tablet—next generation sequence assembly visualization. Bioinformatics 26:401–402

Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLOS ONE 9:e108277

Qiu T, Lu RH, Zhang J, Zhu ZY (2001) Complete nucleotide sequence of the S10 genome segment of grass carp reovirus (GCRV). Dis Aquat Org 44:69–74

Racine T, Hurst T, Barry C, Shou J, Kibenge F, Duncan R (2009) Aquareovirus effects syncytiogenesis by using a novel member of the FAST protein family translated from a noncanonical translation start site. J Virol 83: 5951–5955

Rangel AAC, Rockemann DD, Hetrick FM, Samal SK (1999) Identification of grass carp haemorrhage virus as a new genogroup of aquareovirus. J Gen Virol 80:2399–2402

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Rosenberg MS, Subramanian S, Kumar S (2003) Patterns of transitional mutation biases within and among mammalian genomes. Mol Biol Evol 20:988–993

Schenck JR, Whiteside BG (1976) Distribution, habitat preference and population size estimate of *Etheostoma fonticola*. Copeia 1976:697–703

Seng EK, Fang Q, Lam TJ, Sin YM (2004) Development of a rapid, sensitive and specific diagnostic assay for fish aquareovirus based on RT-PCR. J Virol Methods 118: 111–122

Sibley SD, Finley MA, Baker BB, Puzach C, Armién AG, Giehtbrock D, Goldberg TL (2016) Novel reovirus associated with epidemic mortality in wild largemouth bass (*Micropterus salmoides*). J Gen Virol 97: 2482–2487

Subramanian K, Mcphillips TH, Samal SK (1994) Characterization of the polypeptides and determination of genome coding assignments of an aquareovirus. Virology 205: 75–81

USFWS and AFS-FHS (US Fish and Wildlife Service and American Fisheries Society-Fish Health Section) (2014) Standard procedures for aquatic animal health inspections. In: AFS-FHS blue book: suggested procedures for the detection and identification of certain finfish and shellfish pathogens

Wang Q, Zeng W, Liu C, Zhang C, Wang Y, Shi C, Wu S (2012) Complete genome sequence of a reovirus isolated from grass carp, indicating different genotypes of GCRV in China. J Virol 86:12466

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Editorial responsibility: James Jancovich, San Marcos, California, USA

Submitted: March 6, 2018; Accepted: June 16, 2018
Proofs received from author(s): August 23, 2018