Bioinformatics of Recent Aqua- and Orthoreovirus Isolates from Fish: Evolutionary Gain or Loss of FAST and Fiber Proteins and Taxonomic Implications

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version
doi:10.1371/journal.pone.0068607

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:11717674

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Bioinformatics of Recent Aqua- and Orthoreovirus Isolates from Fish: Evolutionary Gain or Loss of FAST and Fiber Proteins and Taxonomic Implications

Max L. Nibert1**, Roy Duncan2**

1 Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts, United States of America, 2 Department of Microbiology and Immunology, Department of Biochemistry and Molecular Biology, and Department of Pediatrics, Dalhousie University, Halifax, Nova Scotia, Canada

Abstract

Family Reoviridae, subfamily Spinareovirinae, includes nine current genera. Two of these genera, Aquareovirus and Orthoreovirus, comprise members that are closely related and consistently share nine homologous proteins. Orthoreoviruses have 10 dsRNA genome segments and infect reptiles, birds, and mammals, whereas aquareoviruses have 11 dsRNA genome segments and infect fish. Recently, the first 10-segmented fish reovirus, piscine reovirus (PRV), has been identified and shown to be phylogenetically divergent from the 11-segmented viruses constituting genus Aquareovirus. We have recently extended results for PRV by showing that it does not encode a fusion-associated small transmembrane (FAST) protein, but does encode an outer-fiber protein containing a long N-terminal region of predicted α-helical coiled coil. Three recently characterized 11-segmented fish reoviruses, obtained from grass carp in China and sequenced in full, are also divergent from the viruses now constituting genus Aquareovirus, though not to the same extent as PRV. In the current study, we reexamined the sequences of these three recent isolates of grass carp reovirus (GCRV)–HZ08, GD108, and 104–for further clues to their evolution relative to other aqua- and orthoreoviruses. Structure-based fiber motifs in their encoded outer-fiber proteins were characterized, and other bioinformatics analyses provided evidence against the presence of a FAST protein among their encoded nonstructural proteins. Phylogenetic comparisons showed the combination of more distally branching, approved Aquareovirus and Orthoreovirus members, plus more basally branching isolates GCRV104, GCRV-HZ08/ GD108, and PRV, constituting a larger, monophyletic taxon not suitably recognized by the current taxonomic hierarchy. Phylogenetics also suggested that the last common ancestor of all these viruses was a fiber-encoding, nonfusogenic virus and that the FAST protein family arose from at least two separate gain-of-function events. In addition, an apparent evolutionary correlation was found between the gain or loss of NS-FAST and outer-fiber proteins among more distally branching members of this taxon.

Citation: Nibert ML, Duncan R (2013) Bioinformatics of Recent Aqua- and Orthoreovirus Isolates from Fish: Evolutionary Gain or Loss of FAST and Fiber Proteins and Taxonomic Implications. PLoS ONE 8(7): e68607. doi:10.1371/journal.pone.0068607

Editor: Earl G Brown, University of Ottawa, Canada

Received May 3, 2013; Accepted May 30, 2013; Published July 4, 2013

Competing Interests: The authors have declared that no competing interests exist.

** E-mail: E-mails: m nibert@hms.harvard.edu (MLN); roy.duncan@dal.ca (RD)

† These authors contributed equally to this work.

Introduction

Family Reoviridae, subfamily Spinareovirinae (turreted reoviruses) includes nine approved genera, two of which—Aquareovirus and Orthoreovirus—comprise members that are closely related and consistently share nine homologous proteins. Members of the five approved species in Orthoreovirus have 10 dsRNA genome segments and infect reptiles, birds, and mammals; members of the seven approved species in Aquareovirus have 11 dsRNA genome segments and infect fish and putatively shellfish [1]. Despite these differences in segment number and host range, ortho- and aquareoviruses share homologous proteins encoded by nine of their 10 or 11 genome segments [2–5] as well as highly similar particle structures [6–9]. Seven of their nine homologous proteins are structural, i.e., assembled into virions (core RNA-dependent RNA polymerase [RdRp]), core nucleoside triphosphate phosphohydrolase [NTPase], core shell, core turret, core clamp, outer shell, and outer clamp), and the other two are non-structural (NS) proteins required for replication and assembly inside cells (NS factory and NS RNA-binding [RNAb]) (Tables 1, 2, and S1). Ortho- and aquareoviruses are thus likely to have shared a common viral ancestor from which these nine genome segments and their encoded proteins were inherited [2].

Ortho- and aquareovirus proteins that are not consistently homologous across the two genera include two proteins of clear biological significance. One is the outer-fiber protein present in most orthoreoviruses, which anchors atop the core-turret protein at the icosahedral fivefold axes of virions [6,10,11] and mediates attachment to cell-surface receptors [12–16]. The other is the NS fusion-associated small transmembrane (FAST) protein of aquareoviruses and most orthoreoviruses [17,18], which promotes cell-to-cell spread by fostering syncytium formation and release of progeny virions via syncytium-induced cytopathic effects [19,20] (Tables 1, 2, and S1). In members of approved Orthoreovirus species,
the single genome segment not shared by the aquareoviruses is the one that encodes either the outer-fiber protein (in Mammalian orthoreovirus isolates [MRVs]) or the NS-FAST protein (in the Baboon orthoreovirus isolate [BRV]), or both (in Avian orthoreovirus, Nelson Bay orthoreovirus, and Reptilian orthoreovirus isolates [ARVs, NBVs, and RRVs, respectively]) [21–23]. Another NS protein is encoded on this segment in members of approved Orthoreovirus species except RRVs. In members of approved and fully sequenced Aquareovirus species (Aquareovirus A, Aquareovirus C, and Aquareovirus G isolates [AqRVs-A, -C, and -G, respectively]), the two genome segments not shared by aquareoviruses encode three different NS proteins including the FAST protein [2,4,5,24,25] (Tables 1, 2, and S1). The roles of these additional NS proteins encoded on the same genome segments as the fiber and/or FAST proteins in ortho- and aquareoviruses remain poorly understood in most cases, but may involve “luxury/accessory” functions [26] affecting virus–cell interactions in host animals but not essential for virus growth in cultured cells [27].

In the past few years, there have been reports of several fish reoviruses whose full-length sequences reveal their divergence from the viruses currently constituting genus Aquareovirus. These recent isolates include piscine reovirus (PRV) from Atlantic salmon
Salmo salar L. [28] and three isolates from grass carp Chennopharyngodon idella: grass carp reovirus (GCRV; HZ08, GCRV-GD108, and GCRV104 [29–31], the first two of which share nearly identical sequences (>97% aa identity overall) and thus appear to be strains of the same tentative new species. GCRV-HZ08/GD108 and GCRV104 possess the 11-segment profile of approved Aquareovirus members, whereas PRV has the 10-segment profile of approved Orthoreovirus members. We have recently reported that PRV is also like some orthoreoviruses, and unlike approved aquareoviruses, in encoding an outer-fiber protein, p53, and also in not encoding an NS-FAST protein [32] (Tables 1, 2, and S1). The additional NS protein encoded by PRV, p13, is instead a cytotoxic integral-membrane protein that localizes to cytosolic compartments, not to the plasma membrane as do FAST proteins. Interestingly, this novel p13 membrane protein is encoded on the same genome segment as the outer-clamp protein of PRV, the first example among ortho- and aquareoviruses of an outer-clamp protein encoded on a bicistronic segment. PRV is also unique in that its outer-fiber protein is encoded on a monocistronic segment, rather than on a bi- or tricistronic, segment. Nonetheless, based on these and other supportive findings, we have suggested that PRV would be best classified at present as a new species in genus Orthoreovirus [32], where it would represent the only fish-virus species identified to date (tentative species “Piscine orthoreovirus”).

In our recent report on PRV, we noted the also-recent discoveries of GCRV-HZ08/GD108 and GCRV104, but neglected to examine or discuss these additional new fish reoviruses in much detail, other than recognizing the previously overlooked sequences of their outer-clamp proteins [29–32]. In this report, we focus on these viruses, their encoded proteins, and their relationships to other ortho- and aquareoviruses. The results provide new insights into the evolution of this monophyletic taxon, identify an apparent evolutionary correlation between the gain or loss of NS-FAST and outer-fiber proteins among its more distally branching members, and prompt a reconsideration of the taxonomic hierarchy in current family Reoviridae.

Results and Discussion

GCRV-HZ08/GD108 and GCRV104 Encode Outer-Fiber Proteins

Both Wang et al. and Ye et al. [29,30] have recently reported that genome segment 7 of GCRV-HZ08/GD108 encodes a 512-aa protein (calculated mass 56 kDa; hence p56) with sequence similarities to MRV outer-fiber protein σ1 (Tables 1, 2, and S1). Ye et al. additionally noted that GCRV-GD108 p56 shares sequence similarities with adeno virus fiber protein, including portions of its shaft region [30]. Based on these findings, it seems reasonable to expect that GCRV-HZ08/GD108 p56 is a structural component that anchors atop the core turret at the fivefold axes of virions and mediates attachment to cell-surface receptors, as in the case of MRV σ1 [6,10,12–16] and similarly to the case of adeno virus fiber protein [33–35]. As noted by Ye et al., the presence of an outer-fiber protein encoded by GCRV-GD108 raises interesting questions about the evolution of this virus and its relationships to other ortho- and aquareoviruses [30].

Additional details about the location and nature of the sequence similarities between GCRV-HZ08/GD108 p56 and orthoreovirus or adeno virus fiber proteins are shown here in Fig. 1. These similarities appear to be based primarily in structural motifs associated with known fiber proteins: both α-helical coiled-coil and β-spiral motifs in the case of GCRV-HZ08/GD108 p56 and MRV σ1 [36,37], and β-spiral motifs alone in the case of adenovirus fiber protein [38]. MRV σ1, like closely related fiber proteins from ARV, NBV, and RRV isolates [32], has a long region of strongly predicted coiled-coil structure in the N-terminal half of its sequence [36,37]. Within the region of predicted coiled coil, heptad repeats of hydrophobic residues consistent with this structure are regularly evident [36,37], and the presence of this structure has been confirmed by X-ray crystallography of the ARV σC protein [39]. We have recently reported that this motif is also present in the N-terminal half of PRV p35, encoded by monocistronic genome segment S4 [32], and we newly observed for the current report that it is additionally present in the N-terminal half of GCRV-HZ08/GD108 p56, encoded by monocistronic segment 7 (Fig. 1A, red underline at left, red lettering for hydrophobic residues in heptad-repeat pattern at right; Fig. 1B, red bars at top). Upon analyzing the sequences of the other recent isolate, GCRV104, as deposited in GenBank by Fan et al. [31], we newly identified a long region of coiled-coil motif in the N-terminal half of the 511-aa protein (calculated mass 55 kDa; p55) encoded by its monocistronic segment 7 as well (Fig. 1A, B).

The region of predicted coiled coil in the MRV σ1 sequence is followed by a long region with sequence similarity to the β-spiral motif region of adeno virus fiber protein [37,38,40], and the presence of this structure has been confirmed by X-ray crystallography of both MRV σ1 [41] and ARV σC [42]. A long region with similarity to the β-spiral motif region of adeno virus fiber protein is also present following the coiled-coil motif in both GCRV-HZ08/GD108 p56 and GCRV104 p55 (Fig. 1A, cyan underline at left, cyan shading at right; Fig. 1B, cyan bars at top). Although the hydrophobic-repeat pattern is not as regular in the β-spiral motif as in the coiled-coil motif, hydrophobic residues tend to occur at every other position within regions expected to form β-strands (Fig. 1B, cyan and blue lettering at bottom), interspersed by regions of more polar residues expected to be β-turns or loops (Fig. 1B, green lettering at bottom). The presence of this second type of fiber-protein motif in GCRV-HZ08/GD108 p56 and GCRV104 p55 supports the interpretation that these proteins probably share both structural and functional similarities with the outer-fiber proteins of orthoreoviruses.

Also of note for GCRV-HZ08/GD108 p56 is that a smaller region of strongly predicted α-helical coiled coil and associated heptad repeats follows the predicted β-spiral region (Fig. 1A). Indeed, though not strongly predicted by coiled-coil algorithms, a short region of MRV σ1 within the overall β-spiral region was predicted to assume a coiled-coil structure based on the presence of heptad repeats [37], and that structure has been recently confirmed by X-ray crystallography [16]. Thus, it seems reasonable to interpret the current findings for GCRV-HZ08/GD108 p56 to indicate that it is likely to contain a similar, second region of coiled coil following the β-spiral region.

At the sequence termini of MRV σ1 [36,37] and the other orthoreovirus outer-fiber proteins including PRV p35 [32] are regions (short at the N-terminus, longer at the C-terminus) that do not exhibit clear structure-based fiber motifs. The same is notably true for GCRV-HZ08/GD108 p36 and GCRV104 p55 (Fig. 1A). Based on analogies with MRV σ1 [6,10,43–45], the short N-terminal region of these fish-reovirus proteins appears likely to form a base domain involved in anchoring the fiber atop the core-turret protein at the virion surface, and the longer C-terminal region appears likely to form a head domain, at the distal end of the projecting fiber, involved in binding to cell-surface receptors.
Fiber and FAST Proteins in Reoviruses

The two new tentative species of aquareoviruses represented by GCRV-HZ08/GD108 and GCRV104 remain biologically less well characterized than several other aquareoviruses to date. Notably, unlike other previously described aquareoviruses, GCRV-GD108 seems not to induce syncytium formation in cell culture [30]. There are no reports regarding the fusogenic potential of GCRV104. We therefore used bioinformatics approaches to search for FAST-protein homologs encoded by these viruses. FAST proteins share several common features, including their small size (<200 aa), a single transmembrane domain (TMD) located <40 aa from the N terminus, a cluster of basic residues on the C-terminal side of the TMD, sites for modification by fatty acids (N-terminal myristoylation, or palmitoylation on membrane-proximal Cys residues), a short amphipathic or hydrophobic motif that can be located on either side of the TMD, and C-terminal cytosolic endodomains with predicted propensity for intrinsic disorder [18].

Using four different algorithms (HMMTOP, SOSUI, TMHMM, and TMpred; see Materials and Methods), we identified potential TMDs in both of the functionally undefined (“other”) NS proteins of GCRV-HZ08/GD108, NS41 (361 aa) and NS11/9 (95 or 83 aa; see below) (Tables 2 and 3; Fig. 2 A, B). GCRV-HZ08/GD108 NS41 is the sole predicted translation product of genome segment 8. All four algorithms predict this protein contains one or more TMDs; one near the N terminus is also predicted to be a signal peptide, and the other near the C terminus has an adjacent cluster of basic residues (Fig. 2A). NS41 lacks an N-terminal myristoylation consensus sequence, but both predicted TMDs contain nearby Cys residues that might be palmitoylated. The C-terminal region of NS41 is also enriched in Arg and Pro residues, similarly to the AqRV-A FAST proteins [5,24], and this region has predicted propensity for intrinsic disorder. Importantly, however, NS41 is 2–3 times larger than any other known FAST protein, and the locations of its predicted TMDs are inconsistent with FAST-protein membrane topology [18].

We newly identified protein NS11/9 as a predicted translation product of GCRV-HZ08/GD108, from a previously unrecognized, N-proximal ORF of genome segment 11 in both of these viruses (Tables 2 and 3; Fig. 3B). This small ORF is fully embedded within the larger p35 ORF, which we recently determined to encode a homolog of the ortho- and aquareovirus outer-clamp proteins, a similar arrangement as found in PRV genome segment S1 for encoding outer-clamp protein p37 and cytotoxic integral membrane protein p13 [32]. The ORF for NS11 contains 95 codons; however, a second in-frame Met codon in a better context for translation initiation (purine at the −3 position) might instead translate an 83-aa product, NS9 (Fig. 2B). All four of the indicated algorithms predict GCRV-HZ08 and/or GCRV-GD108 NS11/9 has one or more TMDs, the first of which would be absent if the second in-frame Met codon functions as the start codon (Fig. 2B). NS11/9 is the right size for a FAST protein but lacks several other defining features, namely a cluster of basic residues near the predicted TMD and a single, N-proximal TMD. We therefore predict that GCRV-HZ08/GD108 NS11/9, as well as GCRV-HZ08/GD108 NS41, may be additional examples of nonfusogenic, integral membrane proteins encoded by ortho- or aquareoviruses, similar to PRV p13 [32].

GCRV104 initially appeared to lack an NS protein with membrane-interaction potential. However, closer inspection of its genome segment 11, which encodes predicted proteins NS8 (newly identified here; not annotated in GenBank) and NS15 from sequential ORFs, revealed that the NS15 ORF remains open for 64 codons upstream of the predicted NS15 Met start codon. Within this extended reading frame, there are 10 potential non-AUG start codons that reside in a preferred context (i.e., differ by only 1 nt from an AUG start codon and with purines in the −3 and +4 positions) (Fig. 2C). Notably, the FAST proteins of both AqRV-A and AqRV-C isolates are translated from such non-AUG start codons [5,24,25]. All four of the indicated algorithms predict this extended potential N-terminal region of GCRV104 NS15 may contain a TMD with a cluster of basic residues on the C-terminal side (Fig. 2C). The N-terminally extended NS15 protein would also have an N-terminal domain consistent with the size of the FAST protein ectodomains, a C-terminal domain enriched in Arg and Pro residues, and several Cys residues that might be palmitoylated. If GCRV104 is fusogenic (although there is currently no evidence that this is the case), then the N-terminally extended NS15 protein would be the only viable FAST-protein homolog that appears to be encoded by this virus.

Phylogenetic Comparisons

For performing phylogenetic comparisons of ortho- and aquareoviruses more globally than on a protein-by-protein basis, we have previously adopted the approach of aligning concatenated sequences of the nine proteins that are consistently homologous across the two genera [32] (Tables 1, 2, and S1). Comparing these nine-protein sequence concatenations between virus pairs reveals a maximum of 63% aa-sequence identity between representatives of the different species or tentative species (Table 4). We used a similar approach again here, but in order to include outgroup viruses in new phylogenetic comparisons, we limited the concatenated sequence alignments to those of three core proteins with known enzymatic functions (core RdRp, core NTPase, and core turret [guanylyl/methyltransferase]), which are consistently ho-
Figure 2. Putative membrane proteins encoded by GCRV-HZ08/GD108 and GCRV104. In each panel, the genomic plus strand is indicated by the heavy line above and the encoded protein(s) by boxes below. Numbers indicate positions of protein start and stop codons (above) and overall strand length (right). (A) Putative membrane protein NS41 encoded by GCRV-HZ08/GD108 segment 8. Transmembrane regions predicted by the indicated algorithms are indicated by gray bars for GCRV-HZ08 (darker) and GCRV-GD108 (lighter). (B) Putative membrane protein NS11/9 encoded by GCRV-HZ08/GD108 segment 11. Start-codon environment for each of the two ORFs is shown; for the NS11/9 ORF, the two potential in-frame start codons are shown. The potential NS9 product is shaded yellow. Predicted NS11/9 sequences of both GCRV-HZ08 and GCRV-GD108 are shown, with differences in cyan letters and the NS9 product background-shaded in yellow. Transmembrane regions predicted by the indicated algorithms are indicated by gray bars for NS11/9 of GCRV-HZ08 (darker) and GCRV-GD108 (lighter). (C) Putative membrane protein NS15 encoded by GCRV104 segment 11. Start-codon environment for each of the two ORFs is shown; for the NS15 ORF, the extended upstream region without in-frame stop codons preceding the first in-frame Met codon is also shown, and positions of potential, in-frame non-AUG start codons. Transmembrane regions predicted by the indicated algorithms are indicated for N-terminally extended NS15 by gray bars; yellow background shading indicates the non-N-terminally extended NS15 product.

doi:10.1371/journal.pone.0068607.g002
molecular across subfamily Spinareovirinae. Representative members of six of the seven other approved genera in the subfamily [1] were included as outgroup viruses (Table S2). One notable outcome of these new comparisons with three-protein sequence concatenations is that the branch topology of ortho- and aquareoviruses in the resulting phylogram (Fig. 3A) is identical to that obtained with nine-protein sequence concatenations in our recent PRV study [32]. Furthermore, the newly included outgroup viruses adjoin the aqua/orthoreovirus clade on a single, well-defined branch in this phylogram (Fig. 3A), indicating that the combination of more distally branching, approved Aquareovirus and Orthoreovirus members plus more basally branching isolates PRV, GCRV104, and GCRV-HZ08/GD108, constitute a larger, monophyletic taxon that we discuss in more detail below.

The only approved Orthoreovirus species not represented in the preceding phylogram is *Reptilian orthoreovirus*, for which full-length core-protein sequences have not been reported. To obtain tentative placement of an RRV isolate in these analyses, we performed new phylogenetic comparisons using concatenated sequence alignments of the previously reported outer-clamp protein of a python RRV isolate [23] and partial sequences of the outer-shell (T = 13) protein of this virus being newly reported here (GenBank accession no. KF182340), plus concatenated alignments of these two homologous structural proteins from the other ortho- and aquareoviruses. The resulting phylogram (Fig. 3B) exhibits the same branch topology among the viruses as in the preceding analysis with concatenated core-protein alignments (Fig. 3A), and furthermore place python RRV in a subclade with both BRV and Broome virus (BroV), a recent megachiropteran/pipistrelle (megabat/fruit bat, flying fox) isolate that is the prototype strain of a tentative new *Orthoreovirus* species [46], as has been previously reported [23,46]. Other pteropine isolates, as well as their zoonotic relatives obtained from humans with respiratory disease, constitute species *Nelson Bay orthoreovirus* [22,47], whereas microchiropteran/vesper-vitilionid (microbat/insensituous bat, evening bat) isolates to date are members of species *Mammalian orthoreovirus* [48].

### Phylogenetic Distributions of Fiber and FAST Proteins

We next annotated the preceding phylogenograms according to whether each virus possesses an outer-fiber or NS-FAST protein (Fig. 3B), revealing an interesting pattern with at least five seemingly important implications. (i) Based on newly presented sequence analyses in this report, it appears that representatives of the four most basally branching species—GCRV104 and GCRV-HZ08/GD108 on the *Aquareovirus* side and PRV and MRV on the *Orthoreovirus* side of the phylograms—may share both the possession of an outer-fiber protein and the lack of an NS-FAST protein. It therefore seems probable that the last common viral ancestor of all these species was a non-fusogenic virus with an ancestral fiber protein. Fiber-protein sequences from both sides of the phylograms contain β-helical coiled-coil motifs in each, but this may not strongly support common ancestry because this motif is so widespread in nature. On the other hand, fiber-protein sequences from both sides of the phylograms also contain β-spiral motifs, which are much less widespread and hence argue more strongly for common ancestry of these fiber proteins. The consistent relative locations of the coiled-coil and β-spiral motifs within these protein sequences also argue for common ancestry. Thus, ortho- and aquareoviruses seem likely to have shared a last common viral ancestor from which 10, not just nine, genome segments and their encoded proteins, including the outer-fiber but not a functional FAST protein, were inherited.

(ii) More distally branching viruses on both sides of the phylograms have gained an NS-FAST protein. The most parsimonious explanation for the extant fusogenic viruses is two separate gain-of-function events, one after the non-fusogenic PRV and MRV branchpoints leading to the fusogenic orthoreoviruses and the other probably after the GCRV104 and GCRV-HZ08/GD108 branchpoints leading to the fusogenic aquareoviruses (Fig. 3B).

Sequence comparisons support two separate evolutionary trajectories leading to the FAST protein family. The aquareovirus FAST proteins share identity at 20% of the alignment positions over most of their N-terminal 75 aa (Fig. 4A). Over this same interval, the AqRV-C and AqRV-G FAST proteins are more closely related to each other (59% identity) than either is to the AqRV-A FAST protein (≤28% identity), a pattern of sequence conservation that correlates with the topology of the phylograms based on other proteins (Fig. 3). Moreover, all aquareovirus FAST proteins are encoded on bicistronic genome segments that also encode NS proteins of 269–278 aa (Tables 1 and 2), which are also all homologous, suggesting a single evolutionary event led to the gain of fusion activity in these aquareoviruses. Conversely, there is

### Table 3. Predicted transmembrane proteins with no clearly defined functions in aqua- and orthoreoviruses.

<table>
<thead>
<tr>
<th>Features</th>
<th>Functionally unassigned (&quot;other&quot;) NS proteins from representative strains of <em>Aquareovirus</em> and <em>Orthoreovirus</em> species:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome segment</td>
<td>AqRV-A</td>
</tr>
<tr>
<td>Protein (aa)</td>
<td>7</td>
</tr>
<tr>
<td>Additional ORF(s)</td>
<td>FAST</td>
</tr>
<tr>
<td>Genome segment</td>
<td>11</td>
</tr>
<tr>
<td>Protein (aa)</td>
<td>35</td>
</tr>
<tr>
<td>Property</td>
<td>soluble</td>
</tr>
<tr>
<td>Additional ORF(s)</td>
<td>none</td>
</tr>
</tbody>
</table>

*Encoding genome segments are indicated by size rank (largest to smallest).

*Soluble or transmembrane protein predictions were obtained using algorithms HMMTOP, SOSUI, TMHMM, and TMPred. Question mark indicates lack of a consensus prediction with the assignment reflecting the majority prediction.

Identities of the other proteins, if any, encoded by the same genome segment as the indicated NS protein. 

doi:10.1371/journal.pone.0068607.003
Figure 3. Phylogenetic analyses of aqua- and orthoreovirus structural proteins. (A) Maximum-likelihood (PhyML 3.0) unrooted phylogram of concatenated aqua- and orthoreovirus core proteins (core RdRp, core NTPase, and core turret). Homologous proteins from representative members of other genera in subfamily Spinareovirinae were included as outgroups. Program-estimated values for invariant proportion and gamma shape parameter were 0.011 and 1.781, respectively. Branches with support values $\geq 90\%$ are not labeled, and those with support values $<50\%$ are collapsed into polytomies. (B) Maximum-likelihood (PhyML 3.0) phylogram of concatenated aqua- and orthoreovirus outer-capsid proteins (outer...
essential no identifiable sequence conservation between the aqua- and orthoreovirus FAST proteins, and the orthoreovirus FAST proteins alone are more divergent than those encoded by aquareoviruses, with <1% sequence identity shared by all members of the group (Fig. 4B). Different orthoreovirus FAST proteins do, however, share conserved sequences and/or structural motifs. For example, ARV and NBV FAST proteins share 33% overall sequence identity and an identical arrangement of structural motifs [17]; BroV and RRV FAST proteins have an identical N-terminal decapetide sequence [46]; and BRV, BroV, and RRV FAST proteins have N-terminal myristoylation consensus sequences, which are known to be functional in RRV and BRV [49,50]. The orthoreovirus FAST proteins also have different genome-segment coding arrangements. ARV and NBV FAST proteins are encoded on tricistronic genome segments that also encode the fiber protein and a second small NS protein, RRV encodes its FAST protein on a bicistronic genome segment that also encodes the fiber protein, and BRV and BroV FAST proteins are encoded on bicistronic genome segments encoding a second small NS protein (Tables 1 and 2). This diversity among the orthoreovirus FAST proteins could reflect either several different gain-of-function events or a single event followed by extensive divergent evolution accompanied by gene deletions/insertions or lateral gene transfer. Phylogenetic comparisons of the aquareovirus FAST proteins, both groups nonetheless share the defining features of the FAST protein family. The origin of 2–3 distinct clades of FAST proteins with conserved features could arise via convergent evolution from unrelated ancestral precursors or divergent evolution from a common ancestral protein that was nonfusogenic. Regarding the latter option, the three presumed nonfusogenic fish viruses (PRV, GCRV104, and GCRV-HZ08/GD108), which lie close to the inferred bifurcation separating the ortho- and aquareovirus clades (Fig. 3), all potentially encode membrane-interacting NS proteins (Table 3). We have already demonstrated that one of these proteins, p13 of PRV, is a cytotoxic, integral membrane protein [32]. Moreover, as discussed above, four different algorithms predict the NS41 and NS11/9 proteins of GCRV-HZ08/GD108, and the N-terminally extended NS15 protein of GCRV104, may have TMDs, and all three of these GCRV proteins also have one or more additional features of a FAST protein (Fig. 2). It is therefore tempting to speculate that NS41, NS11/9, and N-terminally extended NS15 might all reflect divergent evolution from a membrane-interacting but nonfusogenic FAST protein ancestor.

(iii) Two, or possibly three, evolutionary loss-of-function events are required to explain the extant fiber-lacking ortho- and aquareoviruses. A single loss-of-fiber event after the GCRV104 and GCRV-HZ08/GD108 branchpoints is sufficient to explain the extant, fiber-lacking aquareoviruses (Fig. 3B). For the orthoreoviruses, the polytomy at the base of the BRV/BroV/RRV clade (Fig. 3B) complicates interpretation somewhat, but the most parsimonious explanation is that a single loss-of-fiber event occurred after the shared ancestor of BRV and BroV diverged from the ancestor of RRV. Alternatively, two separate loss-of-fiber events may have led to BRV and BroV, respectively. Additional sequencing of RRVs and other, related isolates may help to clarify this point.

Table 4. Pairwise comparisons of concatenated aqua- and orthoreovirus protein sequences.

<table>
<thead>
<tr>
<th>Virus*</th>
<th>Pairwise identity score (%) with:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AqRV-A</td>
</tr>
<tr>
<td>AqRV-A</td>
<td>100</td>
</tr>
<tr>
<td>AqRV-C</td>
<td>100</td>
</tr>
<tr>
<td>AqRV-G</td>
<td>100</td>
</tr>
<tr>
<td>GCRV-HZ08</td>
<td>100</td>
</tr>
<tr>
<td>GCRV104</td>
<td>100</td>
</tr>
<tr>
<td>PRV</td>
<td>100</td>
</tr>
<tr>
<td>MRV</td>
<td>100</td>
</tr>
<tr>
<td>ARV</td>
<td>100</td>
</tr>
<tr>
<td>NBV</td>
<td>100</td>
</tr>
<tr>
<td>BRV</td>
<td>100</td>
</tr>
<tr>
<td>BroV</td>
<td>100</td>
</tr>
</tbody>
</table>

*The abbreviation and representative strain for each virus is defined in Table 1. GenBank accession numbers are listed in Table S1.

*Concatenated sequences of the nine homologous proteins from each virus were compared using EMBoss Stretcher.

doi:10.1371/journal.pone.0068607.t004
Of the 12 approved or tentative species of ortho- and aquareoviruses represented in the phylograms, four have a fiber but lack a FAST protein, five have a FAST but lack a fiber protein, three have both fiber and FAST proteins, and none lack both.

Figure 4. Comparisons of aqua- and orthoreovirus FAST proteins. (A) MAFFT alignment of aquareovirus FAST proteins in Clustal format. Residues conserved among AqRV-A isolates in this figure are colored magenta; residues conserved among AqRV-C and AqRV-G isolates are colored red. (B) MAFFT alignment of orthoreovirus ARV and NBV FAST proteins in Clustal format. Residues conserved among ARV isolates in this figure are colored magenta; residues conserved among ARV isolates are colored red. (C) MAFFT alignment of orthoreovirus BRV, BroV, and RRV FAST proteins in Clustal format. Residues conserved in BroV and RRV are colored red. In A–C, TMDs predicted by TMHMM are background-shaded in gray; polybasic clusters following the TMDs are background-shaded in yellow. (C) Maximum-likelihood (PhyML 3.0) unrooted phylogram of aqua- and orthoreovirus FAST proteins. Program-estimated values for invariant proportion and gamma shape parameter were 0.007 and 3.598, respectively. Branches with support values ≥90% are not labeled, and those with support values <50% are collapsed into polytomies. AqRV-A.1 is Scophthalmus maximus reovirus (see Table S1) and AqRV-A.2 is Atlantic salmon reovirus Canada/2009 (GenBank accession no. ACN38055); AqRV-C.1 is Golden shiner reovirus (see Table S1), AqRV-C.2 is Grass carp reovirus 873 (GenBank accession no. AAM92738), and AqRV-C.3 is Channel catfish reovirus 730 (GenBank accession no. ADP50119); ARV.1 is strain 176 (see Table S1), ARV.2 is turkey reovirus strain NC/98 (GenBank accession no. ABL96273), and ARV.3 is Stellar sea lion reovirus (GenBank accession no. AED99910); NBV.1 is Nelson Bay virus (see Table S1), NBV.2 is Pulau virus (GenBank accession no. ADG50119), and NBV.3 is Xi River reovirus (GenBank accession no. ADE40974). For the other viruses shown in this figure, see Table S1.

doi:10.1371/journal.pone.0068607.g004
both fiber and FAST proteins (Fig. 3B). We conclude that having either of these proteins is essential, probably due to their respective functions in cell-to-cell spread. We conclude in contrast, however, that having both of these proteins may be evolutionarily disfavored. Perhaps a duplication of their respective functions enhances virus replication or cell/host injury in ways that make longer-term maintenance in nature unsustainable in certain hosts. Based on the phylogenams, it appears that the gain of a FAST protein during evolution of these viruses may commonly precede, and portend, the loss of a fiber protein during their subsequent evolution.

(c) A final important point illustrated by the phylogenams is that the division between genus Aquareovirus and genus Orthoreovirus is hardly well demarcated. Possession of a fiber protein and lack of a FAST protein are properties that now seem to extend to particular viruses in both genera and thus can no longer serve as differentiating traits. At present, the dividing line appears to be best represented by the number of genome segments, 10 or 11, in members of the respective genera [32] (Fig. 3). The branchpoint for outgroup viruses in Fig. 3A suggests this same boundary for defining both genera as monophyletic taxa; however, we found that the position of this branchpoint in related phylogenams was sensitive to which proteins were analyzed and which methods were used, raising concerns about one of the genera appearing paraphyletic in certain analyses, depending on where the genus boundary has been decided to be drawn.

The shallow phylogenetic divide between ortho- and aquareoviruses, as well as a preference to avoid paraphyletic taxa, leads us to suggest that consideration should be given by the International Committee on Taxonomy of Viruses (ICTV) to redefining the taxonomic hierarchy in family Reoviridae. We suggest three alternatives for possible restructuring based on current findings. One alternative is to eliminate genera Aquareovirus and Orthoreovirus and to move existing species in these former genera, as well as the tentative species represented by PRV, GCRV-HZ08/GD108, and GCRV104, into the larger new genus “Orthaquareovirus”, encompassing the entire monophyletic taxon of both 10- and 11-segmented viruses. Advantages of this alternative are that it not only recognizes the close evolutionary relationship of aquare- and orthoareoviruses but also directly eliminates the question of where to draw the Aquareovirus/Orthoreovirus genus boundary to avoid one of them appearing paraphyletic in certain analyses. A second alternative is creation of supergenus “Orthaquareoviraceae” to encompass the entire monophyletic taxon that includes all currently approved ortho- and aquareoviruses as well as PRV, GCRV104, and GCRV-HZ08/GD108. The supergenus level of classification is not yet approved by the ICTV, though sometimes used with other organisms and proposed for use with viruses as well [51]. The suffix “-aceae” is suggested here because it is sanctioned for use at the tribe level, which also falls between family and genus, by the international organization that oversees algal, fungal, and plant nomenclature [http://www.iapt-taxon.org/ nomen/main.php]. A third, more complex alternative is to elevate family Reoviridae to order “Reovirales”, allowing current subfamilies Sedoreovirinae and Spinareovirinae to become families “Sedoreoviridae” and “Spinareoviridae” and current genera Aquareovirus and Orthoreovirus to be grouped under new subfamily “Orthaquareovirinae”. In this third scenario as well as the second, the tentative new species represented by PRV, GCRV104, and GCRV-HZ08/GD108 could remain as diverged species in genera Orthoreovirus and Aquareovirus, respectively, or could be assigned to new genera in subfamily “Orthaquareovirinae” or supergenus “Orthaquareoviraceae” if future virus isolates so dictate.

GCRV Nomenclature

Freshwater farming of grass carp is a global industry, and aquareovirus infections of the young of these fish can cause a hemorrhagic disease associated with high mortality [52]. The original GCRV isolate, 873, was obtained in the 1980’s from a fish farm in China and has turned out to be closely related to golden shiner reovirus, the prototype of species Aquareovirus C [2]. Other GCRV isolates closely related to 873, as indicated by partial sequences in GenBank, are 875, 876, and 991 [2], as well as 096 and JX01 (Table 5). A distinctive isolate of GCRV, PB01-155, was obtained in 2001 from a fish farm in Arkansas, USA, and has since been recognized as the prototype of species Aquareovirus G and designated American grass carp reovirus (AGCRV) [4]. Other AGCRV isolates closely related to PB01-155, as indicated by partial sequences in GenBank, are PB04-125 and PB04-151 [4] (Table 5). The more recently reported GCRV isolates from China that we have addressed here–HZ08/GD108 and 104 [29-31] are clearly divergent from those in Aquareovirus C and Aquareovirus G, and thus should be recognized to represent two new species. GCRV104 so far has no closely related isolates found in GenBank, whereas GCRV-HZ08 and GD108, in addition to being closely related to each other, are also closely related to GCRV isolates 106, 918, and HuNan794, for which complete sequences have very recently been added to GenBank, and isolates 097, JX02, HA-2011, ZS11, QC11, YX11, QY12, NC11, JS12, HS11, and HN12, for which partial sequences are present in GenBank (Table 5). Clearly, referring to any of these isolates as simply “GCRV” is now inadequate, since it appears that four different Aquareovirus species are represented among them, encompassing strong potential for important biological differences. Future authors should therefore take care to emphasize for which GCRV isolate they are reporting new results, preferably indicating a species affiliation as well. Wang et al. [29] in particular have reached similar conclusions relating to GCRV diversity by referring to GCRVs 873, HZ08/GD108, and 104 as respective representatives of GCRV “groups” I-III, but ICTV recognition of the HZ08/GD108 and 104 “groups” as distinct, new species is needed to formalize this classification for the benefit of future studies.

Aquareoviruses Infecting Invertebrate Hosts?

Genus Aquareovirus is currently defined to encompass reovirus-like isolates that are obtained from aquatic, poikilothermic vertebrates (fish) or invertebrates (shellfish) and have 11 genome segments [1]. Shellfish isolates include ones from mollusks (oysters and clams) [53,54] and crustaceans (crabs and shrimp) [55-57]. The prospect of there being such shellfish aquareoviruses is intriguing, but should perhaps be met with some skepticism regarding their natural hosts or taxonomy, since they suggest an unusually broad range of productive infection by viruses from a single genus in the absence of any vector/host relationships among the hosts. Indeed, Meyers et al. [54,58] have shown that the 11-segmented American oyster isolate 13P2 does not productively infect oysters and have argued that putative aquareoviruses obtained from oysters or clams are more likely to be fish viruses that simply accumulated in these shellfish upon filter feeding of virus-contaminated water. Reovirus-like isolates found replicating in a variety of crab species, on the other hand, have been subsequently shown to possess 12 or 10, rather than 11, genome segments and to be phylogenetically divergent from Aquareovirus members [59-63]. Reovirus-like isolates from shrimp have not been genetically characterized. To date, therefore, all sequence-characterized members of genus Aquareovirus are ones that infect vertebrate fish, and all sequence-characterized members of genus
Orthoreovirus, including tentative member PRV, are also ones that infect vertebrates. Thus, until new results may convince us otherwise, we regard the existing genus Aquareovirus, alternatively proposed new larger genus "Orthaquareovirus", alternatively proposed supergenus "Orthaquareovirae", and/or alternatively proposed new subfamily "Orthaquareovirinae" to be constituted solely by vertebrate viruses.

Materials and Methods

Sequences and Basic Analyses

GenBank accession nos. for most of the sequences analyzed in this report are listed in Tables S1 and S2, and a few others are found in figure legends. For some proteins, accession nos. for the protein sequences have not been assigned, and in those cases GenBank accession nos. for the encoding nucleotide sequences are instead listed in Table S1. These nucleotide sequences were analyzed with the Expasy Translate tool as implemented at http://web.expasy.org/translate/ to identify open reading frames and to generate protein sequences for subsequent analysis. Molecular mass and pI values for the proteins were obtained by using the Expasy Compute pI/Mw tool as implemented at http://web.expasy.org/compute_pi/.

-α-helical coiled-coil motifs were detected using Paircoil2 [65] as implemented at http://groups.csail.mit.edu/cb/paircoil2/. Adenovirus-like β-spiral motifs were detected by sequence/structure similarity using HHpred [64] as implemented at http://toolkit.tuebingen.mpg.de/hhpred.

Sequence/Structure Analyses

To concatenate the chosen protein sequences for each virus, they were first joined serially under one FASTA header. The protein order was the same for each virus. The protein sequences of each virus were then separated from one another by a boundary string (WWWWW), which was found to consistently align between the proteins when analyzed by MAFFT 6.85 [70] with default settings (except for maxiterate = 10) as implemented at http://www.ebi.ac.uk/Tools/mafft/. After confirming that these boundary strings had indeed aligned by viewing the output in Clustal format, the alignment was repeated to obtain the output in Pearson/FASTA format. The alignment was then edited to remove the boundary strings and submitted for phylogenetic analyses. For Table 4, the concatenated sequences with boundary strings were compared pairwise using EMBOSS Stretcher with default values as implemented at http://www.ebi.ac.uk/Tools/psa/emboss_stretcher/, and the resulting length and identity values were then corrected to subtract the boundary strings before calculating percent identity.

Phylogenetic Analyses

To concatenate the chosen protein sequences for each virus, they were first joined serially under one FASTA header. The protein order was the same for each virus. The protein sequences of each virus were then separated from one another by a boundary string (WWWWW), which was found to consistently align between the proteins when analyzed by MAFFT 6.85 [70] with default settings (except for maxiterate = 10) as implemented at http://www.ebi.ac.uk/Tools/mafft/. After confirming that these boundary strings had indeed aligned by viewing the output in Clustal format, the alignment was repeated to obtain the output in Pearson/FASTA format. The alignment was then edited to remove the boundary strings and submitted for phylogenetic analyses. For Table 4, the concatenated sequences with boundary strings were compared pairwise using EMBOSS Stretcher with default values as implemented at http://www.ebi.ac.uk/Tools/psa/emboss_stretcher/, and the resulting length and identity values were then corrected to subtract the boundary strings before calculating percent identity.

For phylogenetic analyses, the MAFFT alignment was submitted to PhyML 3.0 [71] as implemented at http://www.lirmm.fr/PhyML/index_en.html, using the LG substitution model, empirical equilibrium frequencies, program-estimated invariant-proportion value and gamma-shape value, and four rate categories. The starting tree was obtained by BioNJ and
optimized by both branch length and tree topology. Tree improvement was performed according to the best of nearest neighbor interchange and subtree pruning and regrafting. Branch support values (%) were estimated by the approximate likelihood ratio test (aLRT) with SH-like criteria. Trees were rendered from the TkTree v1.19b3 as implemented at http://www.phylogeny.fr/to collapse branches with less than 50% support, followed by re-rendering with FigTee v1.4 for cosmetic refinement. The only 

Supporting Information

Table S1 GenBank accession nos. of ortho- and aquareovirus proteins compared in this study.

References


Table S2 GenBank accession nos. of proteins from outgroup viruses compared in this study.

Acknowledgments

We thank Prof. Said A. Gabrial (University of Kentucky) for helpful discussions and comments on the manuscript.

Author Contributions

Conceived and designed the experiments: MLN RD. Performed the experiments: MLN RD. Analyzed the data: MLN RD. Wrote the paper: MLN RD.


