Conserved structure/function of the orthoreovirus major core proteins

Wanhong Xu^{a,b} and Kevin M. Coombs^{a,b,*}

^aDepartment of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, Manitoba, Canada R3E 0J9

^bManitoba Centre for Proteomics and Systems Biology, 715 McDermot Avenue, Winnipeg, Manitoba, Canada R3E 3P4

Abstract

Orthoreoviruses are infectious agents with genomes of 10 segments of double-stranded RNA. Detailed molecular information is available for all 10 segments of several mammalian orthoreoviruses, and for most segments of several avian orthoreoviruses (ARV). We, and others, have reported sequences of the L2, all S-class, and all M-class genome segments of two different avian reoviruses, strains ARV138 and ARV176. We here determined L1 and L3 genome segment nucleotide sequences for both strains to complete full genome characterization of this orthoreovirus subgroup. ARV L1 segments were 3958 nucleotides long and encode λ A major core shell proteins of 1293 residues. L3 segments were 3907 nucleotides long and encode λ C core turret proteins of 1285 residues. These newly determined ARV segments were aligned with all currently available homologous mammalian reovirus (MRV) and aquareovirus (AqRV) genome segments. Identical and conserved amino acid residues amongst these diverse groups were mapped into known mammalian reovirus λ 1 core shell and λ 2 core turret proteins to predict conserved structure/function domains. Most identical and conserved residues were located near predicted catalytic domains in the λ -class guanylyltransferase, and forming patches that traverse the λ -class core shell, which may contribute to the unusual RNA transcription processes in this group of viruses.

Keywords

Double-stranded RNA virus; Nucleotide sequencing; Crystallographic mapping

1. Introduction

An infectious form of arthritis/tenosynovitis was first recognized in chickens in 1959 (Olson, 1959). Subsequent work established that avian reoviruses (ARV) were responsible for infectious enteritis in turkeys (Gershowitz and Wooley, 1973), viral arthritis/tenosynovitis

*Corresponding author at: Manitoba Centre for Proteomics and Systems Biology, 715 McDermot Avenue, Winnipeg, Manitoba, Canada R3E 3P4. Tel.: +1 204 789 3976; fax: +1 204 789 3926. kcoombs@cc.umanitoba.ca (K.M. Coombs).

Conflict of interest

The authors declare no competing interests.

(Olson, 1978), “pale bird” and runting–stunting syndromes (Kouwenhoven et al., 1978), and gastroenteritis, hepatitis, myocarditis, and respiratory illness in chickens (Olson, 1978; Rosenberger and Olson, 1991; Schiff et al., 2007). ARVs are ubiquitous in commercial poultry and are frequently isolated from acutely infected chicken gastrointestinal and respiratory tracts (Rosenberger and Olson, 1991). Contaminated food and water lead to ARV infections being spread through the fecal-oral route (Jordan and Pattison, 1996). In contrast to mammalian reoviruses (MRV) which are rarely associated with human pathology (Krainer and Aronson, 1959; Joske et al., 1964; Johansson et al., 1996; Hermann et al., 2004; Tyler et al., 2004; Schiff et al., 2007), ARV-induced poultry diseases usually result in low mortality but often produce high morbidity rates that lead to significant economic losses (Olson and Solomon, 1968; Glass et al., 1973; Herenda and Franco, 1996; Calnek et al., 1997). Infectious viral arthritis/tenosynovitis continues to be a major focus for researchers in the poultry industry, where attempts are being made to produce vaccines to combat these illnesses. Despite extensive reports on ARV pathogenesis and its apparent significance on the economics of commercial poultry, the basic aspects of its biology, such as viral factors that influence ARV–host cell interactions and pathogenesis, remain poorly understood.

The ARV are members of the family *Reoviridae*, the only group of dsRNA viruses (out of seven dsRNA virus families) that infect mammals (Mertens, 2004; Schiff et al., 2007). The ARVs are the prototypic members of syncytia-inducing, non-enveloped viruses within the *Orthoreovirus* genus. This genus is divided into three subgroups: non-syncytia-inducing mammalian reovirus (subgroup 1; the prototype of the whole genus), avian reovirus and Nelson Bay virus (subgroup 2), and baboon reovirus (subgroup 3) (Chappell et al., 2005).

Both MRV and ARV are non-enveloped viruses with 10 linear double-stranded RNA gene segments surrounded by a double concentric icosahedral capsid shell (inner shell [also called core] and outer shell) of 70–80 nm diameter (Spandidos and Graham, 1976; Benavente and Martinez-Costas, 2007). The genomic segments of avian reovirus can be resolved into three size classes based on their electrophoretic mobilities, designated L (large), M (medium), and S (small) (Spandidos and Graham, 1976; Benavente and Martinez-Costas, 2007). The orthoreovirus genome consists of three large segments (L1, L2 and L3), three medium sized segments (M1, M2 and M3), and four small segments (S1, S2, S3 and S4). Nine of the segments are monocistronic and encode a single different protein (Spandidos and Graham, 1976; Gouvea and Schnitzer, 1982; Benavente and Martinez-Costas, 2007) while S1 is tricistronic with partially overlapping open reading frames (ORFs) that encode for three proteins (Bodelon et al., 2001; Shmulevitz et al., 2002). Although ARVs share many features with the prototypic MRVs, several notable differences exist including hemagglutination properties, host range, pathogenicity, and syncytium formation (Spandidos and Graham, 1976; Schnitzer, 1985; Ni and Ramig, 1993; Theophilos et al., 1995; Martinez-Costas et al., 1997; Jones, 2000; Zhang et al., 2005; Benavente and Martinez-Costas, 2007).

Genomic coding differences also exist between MRV and ARV. For example, the ARV L1 segment encodes the core shell protein λA (Spandidos and Graham, 1976; Benavente and Martinez-Costas, 2007), whereas the homologous MRV protein ($\lambda 1$) is encoded by the MRV L3 segment (McCrae and Joklik, 1978), and the viral RNA-dependent RNA polymerase protein is encoded by either the L1 segment (MRV: $\lambda 3$; (McCrae and Joklik, 1978)) or L2

segment (ARV: λ B; (Xu and Coombs, 2008)). Differences in the functional properties of homologous ARV and MRV proteins have also been reported. For example, two non-homologous dsRNA-binding proteins (the ARV σ A core protein and the MRV σ 3 major outer capsid protein) are predicted to regulate PKR activation (Schiff et al., 1988; Gonzalez-Lopez et al., 2003) while the ARV σ A core protein displays nucleoside triphosphate phosphohydrolase (NTPase) activity (Yin et al., 2002), ascribed to the non-homologous MRV μ 2 (Noble and Nibert, 1997) and λ 1 (Bisaillon et al., 1997) core proteins. Based on these early comparative studies, it seems likely that additional analysis of ARV will continue to broaden our understanding of the *Reoviridae* family, possibly leading to the identification of novel features that impact on the distinct biological and pathogenic properties of ARV.

Recent advances have allowed sequence determinations of a growing number of virus isolates. Many ARV and MRV genome segment sequences have been reported. In addition, the complete genomic sequences of three prototype strains of MRV have been completed (Wiener et al., 1989; Breun et al., 2001; Yin et al., 2004). In contrast, sequence information from ARV isolates is more limited. The entire complement of S-class genome segments (for example, Chiu and Lee, 1997; Duncan, 1999; Liu and Huang, 2001; Bodelon et al., 2001; Shmulevitz et al., 2002; Kapczynski et al., 2002; Sellers et al., 2004) and M-class genome segments (for example, Touris-Otero et al., 2004; Noad et al., 2006) have been determined for some ARV clones. These sequence determinations include two strains (ARV138 and ARV176) which have served as useful genetic reagents; panels of intertypic reassortants (Duncan and Sullivan, 1998; O'Hara et al., 2001) and several temperature-sensitive mutants (Patrick et al., 2001) have been derived from these strains. Sequence information is also available for some ARV L1 and L3 genome segments (Hsiao et al., 2002; Shen et al., 2007) (Table 1), but not for strains ARV138 and ARV176. Finally, we have also recently determined the sequences of the ARV138 and ARV176 L2 genome segments (Xu and Coombs, 2008). Thus, to complete the entire genomic sequence determinations of two genetically distinct ARV clones that exhibit characteristic differences in genome electrophoretic mobilities, syncytium-induction, and pathogenic capabilities (Duncan and Sullivan, 1998; O'Hara et al., 2001) and which, therefore, are useful genetic reagents to further delineate ARV structure, function and pathogenesis, and to expand the available ARV sequence database, we determined the genomic sequences of the ARV138 and ARV176 L1 and L3 genome segments. These comparative sequence studies, in combination with ongoing genomic and proteomic studies, are intended to better delineate the role(s) of individual ARV proteins in the viral replication cycle.

2. Materials and methods

2.1. Cells and viruses

Avian reovirus strain 138 (ARV138) and strain 176 (ARV176) are laboratory stocks. Virus clones were amplified in the continuous quail cell line QM5 in Medium 199 (Gibco) supplemented to contain 7.5% fetal calf serum (Hyclone), 2 mM glutamine, 100 U/ml penicillin, 100 μ g/ml streptomycin, and 1 μ g/ml amphotericin B, essentially as previously described (Patrick et al., 2001).

2.2. Sequencing the L genome segments

Genomic dsRNA was extracted from amplified virus P2 stocks with phenol/chloroform (Sambrook et al., 1989). The extracted dsRNA were resolved in 10% SDS-PAGE and L1, L2, and L3 genome segments were separately excised. Individual segment gel bands were collected into microcentrifuge tubes, macerated, and incubated in 1–2 vol. of diffusion buffer (0.5 M ammonium acetate; 10 mM magnesium acetate; 1 mM EDTA, pH 8.0; 0.1% SDS) at 50 °C for 30 min. The macerated gel pieces were pelleted by centrifugation at $10,000 \times g$ for 1 min, supernatants were collected and dsRNA precipitated by ethanol. Each pellet was dried and resuspended in ddH₂O for 3'-ligation-based RT-PCR. All primers used for ligation, RT-PCR, and sequencing were synthesized by Invitrogen. An anchor primer, P-5'-GAAGCCTATCCCTAACCCCTCTC-CTCGGTCTCGATTCTACG 3'-Bio (5'-end phosphorylated and 3'-end biotin-blocked) was ligated to the 3'-end of each genome segment, using T4 RNA ligase according to the manufacturer's instructions (Promega Inc., Madison, USA). Ligated products were precipitated by mixing with 1/2 vol. of (30% PEG 8000 in 30 mM MgCl₂), and centrifuged immediately at $10,000 \times g$ for 30 min. The supernatants were removed and pellets were dried and dissolved in ddH₂O for cDNA synthesis. Full-length cDNA copies of each L1 or L3 genome segment were synthesized using a primer (21-mer) complementary to the anchor primer by SuperScript™ II reverse transcriptase according to the manufacturer's instructions (Invitrogen). Forward and reverse primers corresponding to the 5'-most conserved nucleotides, and the 3'-most conserved nucleotides of the plus-strand, respectively, were designed (Table 2) based upon available corresponding segment sequences of ARV strains from GeneBank. PCR amplification was performed using cDNA and these forward and reverse primers with the Expand Long Template PCR System (Roche). PCR products used for DNA sequencing were gel purified using QIAquick® gel extraction kit according to the manufacturer's instructions (Qiagen).

DNA sequencing was performed in both directions by use of an ABI Prism BigDye Terminator v3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and an Applied Biosystems Genetic Analyzer DNA Model 3100. The first two sequencing reactions were performed with the primers used for PCR amplification. Primers for subsequent reactions were designed from newly obtained sequences to completely sequence each full-length PCR product in both directions. Sequences nearer the ends of each segment were determined from PCR products that were amplified with a primer complementary to the anchor primer and an internal gene-specific primer. Sequences obtained from both directions were assembled and checked for accuracy with SeqMan® (Lasergene®, Version 7.1.0; DNASTAR, Inc.).

2.3. Sequence analyses

Sequences were compiled and analyzed using the Lasergene® software suite (Version 7.1.0; DNASTAR, Inc.) Pairwise sequence alignments were performed using the Wilbur–Lipman method (Wilbur and Lipman, 1983) for highly divergent nucleotide sequences, the Martinez–NW method (Martinez, 1983) for closely related nucleotide sequences, and the Lipman–Pearson method (Lipman and Pearson, 1985) for protein alignments in MegAlign® (Lasergene®). Multiple sequence alignments were performed using Clustal-W (Thompson et al., 1994) and T-Coffee (Notredame et al., 2000), and alignment adjustments were manually

performed as needed in MegAlign[®]. Amino acid alignment images were adjusted in Adobe Photoshop 7.0 (Adobe[®]). Nucleotide compositions and protein molecular weights were calculated by DNA statistics and protein statistics, respectively, in EditSeq[®] (Lasergene[®]). Phylogenetic trees were constructed using Neighbor-Joining and tested with 1000 bootstrap replicates in MEGA version 4 (Tamura et al., 2007).

2.4. 3D structural analyses

Molecular graphics coordinates of the mammalian reovirus (MRV) core $\lambda 1/\lambda 2/\sigma 2$ complex (PDB #1EJ6; (Reinisch et al., 2000)) were assembled into a pentameric vertex with Viper[®] (Reddy et al., 2001). The resulting structures were manipulated with the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco ((Pettersen et al., 2004); supported by NIH P41 RR-01081). Resulting images were imported into Adobe Photoshop and assembled with Adobe Illustrator (Adobe).

3. Results

3.1. Sequence analysis of L1 genome segments and deduced ARV138 and ARV176 λ A proteins

The ARV138 (GeneBank accession number EU707933) and ARV176 (GeneBank accession number EU707934) L1 genome segments were determined to each contain 3958 nucleotides (Table 3). There are no gaps relative to one another when aligned. The 5' and 3' predicted non-coding regions of both L1 segments are 20 and 56 nucleotides long, respectively (Table 3). Nucleotide sequence identity between ARV138 and ARV176 L1 genome segments is 91.4% (Table 4). BLAST searches indicated the ARV L1 genome segments were most similar to the mammalian reovirus (MRV) L3 and aquareovirus (AqRV) L3 genome segments, both of which encode major core shell proteins. Pairwise sequence comparisons between both of these newly determined ARV genome segments and all currently available homologous MRV and AqRV L3 genome segments (see Table 1) showed a range of nucleotide and protein identity values. Preliminary comparative studies of all currently available ARV L-class genome segments indicated that for some L segments the ARV138 and ARV176 sequences were the most distantly related, whereas the ARV strain 601G L1 segment was amongst the most distant from all other ARV L1 segments (data not shown). Similarly, preliminary comparative analyses indicated that the grass carp reovirus (GCRV) and chum salmon reovirus (CSRV) L genes were the most distantly related amongst the AqRV (data not shown). Thus, although all currently available ARV, MRV, and AqRV L-class genome segments were aligned and compared in subsequent analyses, we limited presentation in subsequent tables and figures to these few most-distant clones for clarity. As predicted from taxonomic grouping, MRV L3 genome segments showed greater identity to each other [from a low of 77% between strains Lang (T1L) and Jones (T2J), up to 98% between T1L and strain Dearing (T3D) (Table 4)] than they did to homologous ARV and AqRV genes. Similarly, ARV L1 genome segments showed greater identity to each other than to homologous MRV and AqRV genome segments. Pairwise sequence comparisons between the ARV L1 and the MRV L3 genome segments showed ~51% identity and pairwise comparisons between the ARV L1 and AqRV L3 genome segments revealed ~43%

identity (Table 4). Phylogenetic analysis of the L1 genes clustered the ARV, MRV, and aquareoviruses into three distinct clades (Fig. 2A), with ARV and MRV more closely related to each other than either are to aquareoviruses, reflecting that ARV and MRV belong to different species in *Orthoreovirus* (Duncan, 1999) whereas aquareoviruses are members of the different *Aquareovirus* genus in the *Reoviridae* family.

The ARV L1 genome segment is predicted to encode the core capsid shell protein λ A (Varela and Benavente, 1994). The ARV138 and ARV176 λ A proteins are each deduced to be 1293 amino acids long with no insertions or deletions relative to each other (Fig. 1), amino acid identity of 98.2% (Table 4), and a calculated molecular weight of ~142 kDa (Table 3). Pairwise comparisons of ARV138 and ARV176 λ A protein sequences with homologous MRV λ 1 proteins showed sequence identity of ~44%, and an identity of ~28–32% when compared to homologous AqRV VP3 proteins (Table 4). Amino acid alignments between ARV138 λ A and ARV176 λ A showed a total of 24 mismatches (Fig. 1). Of these, 15 amino acid differences were found within the first 228 residues of the N-terminus. Examining alignments of 16 available ARV λ A amino acid sequences revealed similar results, and in all cases, the greatest sequence divergence was found within the N-terminal 19–91 amino acids (data not shown). The greatest variable region is restricted to the N-terminal 19–51 amino acids. Similar features were also reported in the homologous MRV core shell proteins (Harrison et al., 1999). Alignments of ARV λ A proteins with homologous MRV and AqRV proteins (Fig. 1), and similarly weighted window-averaged analyses of ARV:MRV and ARV:MRV:AqRV protein identities (Fig. 3A) also showed this region near the N-terminus is more variable than other regions. This distinct character is further highlighted by secondary structure, hydrophilicity, and flexibility predictions (similar to those observed in the homologous MRV proteins (Harrison et al., 1999); (data not shown)). Additionally, the first 240 amino acids of this region of (1 are largely disordered and lie beneath the tightly folded, plate-like structure of the main part of the MRV λ 1 shell crystal structure (Reinisch et al., 2000). Window-averaged analyses revealed several regions within the ARV and MRV core shell proteins that were relatively highly conserved when compared to each other. These were found immediately C-terminal to the previously discussed disordered region (ARV and MRV residues 254–272) and several other such regions were located in the C-terminal half of the protein (ARV residues 715–748, 822–847, 892–909, 1043–1056, 1208–1223, and 1273–1279, that correspond with MRV residues 709–742, 816–841, 884–901, 1033–1046, 1198–1213, and 1263–1269, respectively). Addition of the AqRV VP3 protein to the above analyses indicated several notable features. Clustal-W alignments identified 195 amino acid residues that were identical in the 6 aligned sequences (Fig. 1) whereas T-Coffee alignments identified 198 amino acid residues that were identical in the 6 aligned sequences (data not shown; overall average identity = 15.5% [Fig. 3A, horizontal solid line]). Window-averaged regions of very low conservation included the N-terminal ~110 residues in the N-terminal disordered region (due largely to significant gaps in these alignment regions, as also previously reported (Kim et al., 2004)). There also were three other regions (residues ~330–400, ~750–800, and ~850–880) that appeared to have very low conservation; however, these regions were fairly well conserved in ARV:MRV comparisons; corresponding regions of the AqRV were poorly conserved. Three regions showed higher-than-average conservation in the ARV:MRV:AqRV alignments, suggesting

these regions (ARV residues L₂₅₅ITWDSGLCTSFELVPI₂₇₁, G₇₃₄SANLFTP₇₄₁, and P₈₃₆FQVPFARL₈₄₄, that correspond to MRV residues L₂₅₅VTWDAGLCTSFKIVPI₂₇₁, G₇₂₈SANLFTP₇₃₅, and P₈₃₀FQVPYVRL₈₃₈, respectively) contain important structural/functional domains. In addition to the 198 identical residues found in all 6 sequences discussed earlier, alignments with blossom50 weighting indicated that an additional 217 alignment positions contained either identical amino acid residues or conservative substitutions in at least 4 of the 6 aligned sequences.

3.2. Sequence analysis of L3 genome segments and deduced ARV138 and ARV176 λ C proteins

Both ARV138 and ARV176 L3 genome segments (accession nos. EU707937 and EU707938, respectively) were found to be 3907 nucleotides long, and to align with no insertions and deletions (Fig. 4). Thus, the whole genome sizes of ARV138 and ARV176 are almost identical, with only one base pair difference, attributed solely to the L2 genome segment; 23,492 bp for ARV138 versus 23,493 bp for ARV176 (Table 3). The L3 segment 5' non-translated region has the shortest length (12 bases) among the 10 genome segments (Table 3). Thus, the ARV have short non-translated regions at the 5' and 3' ends of the genome, similar to what has been reported for MRV (Dermody et al., 1991). The nucleotides in the ARV non-translated regions are highly conserved, with a maximum of one mismatch in the 5' non-translated region in 5 of 10 genome segments between ARV138 and ARV176, and a maximum of two mismatches in the 3' non-translated region in the S3 genome segment (data not shown). The first six nucleotides (GCTTTT) of the 5' end and the last five nucleotides (TCATC) of the 3' end of the plus-strand are conserved in all 10 genome segments (data not shown).

Percent nucleotide identity between the ARV138 and ARV176 L3 segments is 73% (Table 4), making it the least conserved of the ARV genome segments. Pairwise comparisons of ARV138 and ARV176 L3 genome segments with homologous MRV or AqRV genome segments revealed nucleotide identity ranging from 42 to 44%, or 38 to 39%, respectively (Table 4). Phylogenetic analysis of the L3 genes clustered the ARV, MRV, and aquareoviruses into three distinct clades (Fig. 2B), and, as seen with other genome segment comparisons, with ARV and MRV more closely related to each other than either are to aquareoviruses. Translation of the ARV L3 genome segment, like the other nine ARV genome segments, is predicted to start from the first AUG in the 5' end of the plus-strand, yielding predicted λ C proteins 1285 amino acids in length and with calculated molecular weights of ~142 kDa (Table 3). The sequence identity between ARV138 and ARV176 λ C proteins is 83.9% with no gaps in the alignment. Protein BLAST search results showed that significant similarities exist between ARV λ C, MRV λ 2, and AqRV VP1 proteins. ARV λ C has been identified as a core turret protein (Hsiao et al., 2002; Zhang et al., 2005), which mediates the guanylyl-transferase reaction in cap formation (Martinez-Costas et al., 1995; Hsiao et al., 2002). The guanylyltransferase domain of ARV λ C has been demonstrated to locate in a 42-kDa N-terminal region, corresponding to the first 384 residues (Hsiao et al., 2002), similar to the location of this domain in MRV (Luongo et al., 2000; Luongo, 2002). Comparisons of ARV138 and ARV176 λ C proteins with homologous MRV or AqRV proteins showed sequence identity of ~28 or ~24%, respectively (Table 3).

Amino acid alignments (Fig. 4) and window-averaged analysis (Fig. 3B) of ARV λ C and MRV λ 2 protein identities revealed generally low overall identity, with only a few regions where averaged identity was $\geq 60\%$. These regions were located within MTase domains (MRV residues 472–490, 565–592, 606–620, and 826–841). Three of these regions remained relatively highly conserved (averaged identity values $>40\%$) after addition of the AqRV VP1 protein to the analysis. These were ARV residues V₅₇₀YSDVDQV₅₇₇, T₆₀₂Y/FTGGSV/LVA/VKCNFPT₆₁₈, and L₈₂₂DLGTGPERPL₈₃₃, the latter of which encompasses putative methyltransferase domain I (Bisaillon and Lemay, 1999). Clustal-W (Fig. 4) and T-Coffee (data not shown) alignments identified 136 and 138 amino acid residues, respectively, that were identical in the 6 aligned sequences (overall average identity = 10.7% [Fig. 3B, horizontal solid line]). In addition, blossom50 weighting alignments indicated that an additional 171 positions contained either identical amino acid residues or conservative substitutions in at least 4 of the 6 aligned sequences.

4. Discussion

4.1. Whole-genome comparisons of avian reovirus genes and proteins

ARV138 was originally isolated from the hock joint of an infected chicken in New Brunswick, Canada (Drastini et al., 1992), whereas ARV176 was isolated from the hock of infected chickens in Georgia, USA (Hieronymus et al., 1983). These two strains have exhibited characteristic differences with respect to their genome electrophoretic mobilities, syncytium-inducing, and pathogenic capabilities (Duncan and Sullivan, 1998; O'Hara et al., 2001). Thus, these two strains may represent some degree of diversity that exists among ARV isolates. In addition, these strains have been used to generate a variety of genetic reagents, including intertypic reassortants and temperature-sensitive mutants. The sequences for S-class and M-class genome segments of ARV138 and ARV176 had been reported (Duncan, 1999; Shmulevitz et al., 2002; Noad et al., 2006). With completion of ARV138 and ARV176 L-class sequences in this, and our previous (Xu and Coombs, 2008) report, we can now make whole-genome comparisons of ARV138 and ARV176.

The whole genome sizes of ARV138 and ARV176 are almost identical, with only one base pair difference, attributed solely to the L2 gene; 23,492 bp for ARV138 versus 23,493 bp for ARV176 (Table 1). The ARV have short non-translated regions at 5' and 3' ends of the genome, similar to what has been reported for MRV (Dermody et al., 1991). The 5' non-translated regions of plus-strands of all 10 genome segments of ARV138 and ARV176 have lengths between 12 and 30 nucleotides, with the longest two being those of S3 (30) and M2 (29), and the shortest two being those of M1 (12) and L3 (12) (Table 4). The 3' non-translated regions of plus-strands of all 10 genome segments of ARV138 and ARV176 have lengths between 33 and 98 nucleotides, with the longest two being those of M2 (98) and M1 (72) and the shortest two being those of S1 (33) and L2 (36) (Table 4). The nucleotides of ARV non-translated regions are highly conserved, with a maximum of one mismatch in the 5' non-translated region in 5 of 10 genome segments between ARV138 and ARV176, and a maximum of two mismatches in the 3' non-translated region in the S3 gene (data not shown). This high degree of conservation implies important functional consequences, and may reflect presence of important signal sequences required for genome segment packaging

or replication. The first 6 nucleotides (GCTTTT) of the 5' end and the last 5 nucleotides (TCATC) of the 3' end of the plus-strand are conserved in all 10 genome segments.

Features of overall nucleotide identity between ARV138 and ARV176 for all 10 genome segments are shown in Fig. 5A. The S2 gene is the most conserved, with overall sequence identity of 94.0% and the L3 gene is the least conserved, with overall sequence identity of 73.0%. The features of the 12 proteins deduced from corresponding nucleotide sequences of ARV138 and ARV176 are summarized in Table 4 and Fig. 5B. The lengths of all deduced proteins are invariant between the two strains. The calculated molecular weights and pI values of the 12 proteins among homologues from the two strains vary by less than 289 Da and 0.4 pH units, respectively. The major core shell protein λA is the most conserved, with overall sequence identity of 98.2%. The core turret protein λC is the least conserved, with overall sequence identity of 83.9%. Comparisons of protein identities between ARV and MRV show that the RNA-dependent RNA polymerase proteins ARV λB , and MRV homolog $\lambda 3$, are the most conserved, with ARV:MRV identity of ~55% (Table 3; see also (Xu and Coombs, 2008)). The relatively high conservation of the RdRp proteins of these different *Orthoreovirus* species implies significant structure/function constraints are placed upon this protein. This is further implied by the observation that the most conserved regions are those that are believed to be involved directly in polymerization and that interact directly with RNA during transcription. The next-most conserved proteins (as determined by comparing percent identity between ARV and MRV proteins) are the major outer capsid protein μB (Noad et al., 2006) and the major core shell protein λA (this study), both with ~43% sequence identities. These relatively high identities also imply significant structure/function constraints are placed upon these proteins, possibly because of the importance of these proteins in maintaining overall structure of the outer shell and core particle, respectively, combined with their relatively buried positions within whole virions, which would make them less susceptible to antibody-induced variation. In contrast, most other virion proteins have ARV:MRV protein identities ~30%, implying either less structure/function constraint and/or more antibody-induced variation (i.e. the outer sigma-class proteins σB and σC).

4.2. 3D mapping identical and conserved residues in the orthoreovirus core shell protein

Comparisons of identical amino acid residues, conservative substitutions, and non-conservative substitutions in selected ARV λA , MRV $\lambda 1$, and AqRV VP3 core shell proteins, and mapping these residues in the available MRV $\lambda 1$ atomic structure (Reinisch et al., 2000) indicated that identical residues and conservative substitutions were generally found in patches within the molecule and in close proximity to suspected catalytic domains. Similar to what was found from MRV:AqRV comparisons (Kim et al., 2004), identical and conserved residues surround the 5'-axis (the suspected site of nascent mRNA egress (Zhang et al., 2003; Kim et al., 2004; Mendez et al., 2008)). Previous MRV:AqRV comparisons (Kim et al., 2004) found significantly more identical and conserved residues on the inner shell surface than on the outer shell surface (see Fig. 6 in (Kim et al., 2004)). Addition of ARV to these comparisons greatly reduced the numbers of completely identical and conserved residues on the inner shell surface (Fig. 6E); but examination of these surfaces and of sections (Fig. 6D–I) indicate, similar to what was previously described (Kim et al.,

2004), that most conservation is located on the inner core shell surface, inside the core shell plate-like structure itself, and surrounding highly conserved residues thought to be involved in helicase or 5'-RNA triphosphatase activity (Fig. 6). Previous alignment analyses (Bisaillon and Lemay, 1999; Kim et al., 2004) identified six potential helicase domains. The first three are located within the disordered N-terminal region and are either poorly conserved (or, in the case of putative domain II, absent in some of the AqRV), in ARV, MRV, AqRV alignments (Fig. 1). The remaining three (domains IV, V, and VI) collectively contain 6 highly conserved residues found in all 26 currently available ARV, MRV and AqRV core shell proteins (ARV residues Trp₂₂₆ in domain IV; Pro₃₅₄ in domain V; and Arg₄₃₆, Gly₄₃₈, Arg₄₄₂, and Ala₄₄₃ in domain VI; predicted to be equivalent to MRV residues Trp₂₂₇, Pro₃₅₁, Arg₄₃₀, Gly₄₃₂, Arg₄₃₆, and Ala₄₃₇—see Fig. 1). These data suggest helicase domain VI, which is the most highly conserved, and which is located near the edge of the decameric unit (Fig. 6D and H; green arrow) may have primary enzymatic function. Alternatively, other domains may be catalytically important but there are significant differences in each virus' enzymatic characteristics, which have not yet been fully characterized. Previous alignment analyses (Bisaillon and Lemay, 1999; Kim et al., 2004) also identified two potential 5'-RNA triphosphatase domains in the C-terminal ~200 residues. Several amino acid residues were identified as highly conserved between MRV and AqRV (Kim et al., 2004). Many of these (MRV residues Asp₁₀₉₆, Gly₁₁₀₆, Gly₁₁₃₃, and Glu₁₁₄₀) have completely conserved ARV equivalents (predicted as Asp₁₁₀₆, Gly₁₁₁₆, Gly₁₁₄₃, and Glu₁₁₅₀). However, our newly determined ARV sequences appear to allow us to narrow down the conserved elements of these enzymatic domains; a proline residue conserved in MRV (position 1092) and in the AqRV GCRV (1029) (Kim et al., 2004); see also Fig. 1) is not found in the ARV sequence in this region. Similarly, MRV Arg₁₁₃₆, which aligns with GCRV Arg₁₀₇₃, aligns with Tyr₁₁₄₆ in ARV. The four completely conserved residues (MRV Asp₁₀₉₆, Gly₁₁₀₆, Gly₁₁₃₃, and Glu₁₁₄₀) lie close to each other inside and on the upper core shell surface within generally conserved patches (Fig. 6D, F, and I; arrows and hexagons), consisting also of Gly₁₀₄₃, Pro₁₀₆₁, Arg₁₀₆₉, Val₁₀₇₄, Gly₁₀₇₈, and Pro₁₁₁₁ (which are completely conserved in all 26 available ARV, MRV, and AqRV core shell proteins (Fig. 1)). Three other residues in these patches also are very highly conserved; Arg₁₀₄₂, Gly₁₁₄₂, and Ile₁₁₇₅ are found in 25 of the 26 available ARV, MRV, and AqRV core shell protein sequences (including the six chosen ARV, MRV, and AqRV sequences (Fig. 1); these positions are replaced, respectively, by alanine in AqRV AGCRV, by proline in AqRV AGCRV, and by valine in MRV SC-AL. Finally, these patches also contain numerous conservatively substituted residues, suggesting these regions of the shell have important functions. Kim et al. (2004) had previously noted when mapping conserved AqRV VP3 residues into the MRV λ 1 structure that several highly conserved patches on the upper and lower surfaces of the shell corresponded (Fig. 6 in (Kim et al., 2004)) and postulated that these conserved patches traversed the core shell and might mediate small molecule entry needed to drive RNA transcription. We similarly found patches of identical and conservatively substituted residues on the upper and lower core shell surfaces in ARV:MRV:AqRV structural mapping (Fig. 6). Many of these conserved patches clearly traverse the entire core shell as seen in various cross-sections (see, for example, Fig. 6G, H and I). For example, when viewed from the outside, the largest patches closest to the 5-fold axes in the λ 1.5 molecules (circled in Fig. 6D and H) consist of identical amino acid residues (Asp₅₈₇, Pro₆₁₄, Ser₆₁₉, Thr₆₂₁, Leu₆₃₄, Ala₆₃₅, Pro₆₃₇, Asp₆₄₁, Pro₆₄₂, Phe₆₄₉,

Met₆₅₀, Ala₆₇₅, Asn₆₇₆, Phe₆₈₈, Ala₇₀₆, Ile₇₀₉, Trp₇₁₃, Pro₇₁₄, Pro₇₁₆, Pro₇₄₂, Trp₇₆₄, Arg₈₃₇ and Leu₈₃₈), as well as nearly equivalent numbers of conservatively substituted residues. These same residues, which are not continuous within the primary sequence, also form shell-spanning patches in the λ 1.3 molecules (Fig. 6D and H, boxes), although, because of the different folding of these two alternate λ 1 molecules, these patches in λ 1.3 are located in different regions on the shell than are the patches in the λ 1.5 molecules. The conserved patches containing the putative 5'-RNA-triphosphatase residues described above (Fig. 6D and I) also appear to traverse the core shell. As also previously noted (Kim et al., 2004), some of the conserved patches do not completely traverse the core shell. In addition, a predicted C₂H₂ zinc finger motif for ARV λ A is located between residues 182–198. This zinc finger motif has been previously noted, and described, in MRV (Bartlett and Joklik, 1988) and in MRV:AqRV alignments (Kim et al., 2004) and is completely conserved in the alignment with all other ARV, MRV and AqRV core shell proteins (Fig. 1). The zinc finger motif is located near several predicted helicase domains (Fig. 6, green), including Trp₂₂₆ in domain IV, which is consistent with observations about roles of zinc finger motifs in Nidovirus helicase activity (Seybert et al., 2005) and in herpes simplex virus UL52 protein (Chen et al., 2005).

4.3. 3D mapping identical and conserved residues in the orthoreovirus core turret protein

ARV λ C, MRV λ 2, and AqRV VP1 core turret proteins, and mapping these residues in the available MRV λ 2 atomic structure (Reinisch et al., 2000) indicated that, similar to the other λ proteins, identical and conservative substitutions were not evenly distributed within the core turret. The flaps domain at the top of the molecule (Fig. 7C and D) and the spacer/extended domain near the middle of the molecule (Fig. 7H and I) appear the least conserved, with the lower surfaces of each (Fig. 7D and I) having less conservation. Interestingly, Kim and colleagues had noted several patches as well as the 5-fold channel in the flaps domain were highly conserved in MRV:AqRV comparisons (Kim et al., 2004). Addition of ARV to these comparisons indicated a previously noted patch near the 5-fold axis (see Fig. 4B in (Kim et al., 2004)) was preserved in all aligned sequences, but other patches located on the top of the flaps domain and further from the 5-fold axis were not conserved in ARV. In addition, there appeared to be little conservation at the 5-fold axis (the suspected site of cell attachment protein tethering; σ C in ARV and σ 1 in MRV). These observations suggest the ARV differ somewhat more from the MRV and AqRV within this region of the turret protein. By contrast, the enzymatic domains appear more highly conserved. This is particularly true for the internal and lower regions of the MTase domains (Fig. 7F and G) and the internal and upper regions of the GTase domain (Fig. 7J and L). Most striking, as previously reported from MRV:AqRV comparisons (Kim et al., 2004), are the patches of conservation that include and surround suspected MTase-1 residues, suspected MTase-2 residues, and highly conserved lysine residues in the GTase domain. Previous MRV:AqRV comparisons (Kim et al., 2004) identified nine conserved residues within the MTase-1 domain that might be enzymatically important. Inclusion of the newly determined ARV sequences reduces the number of conserved residues in this suspected catalytic region to 4; MRV T1L residues His₅₂₁, Asp₅₆₁, Asp₅₇₇, and Asp₅₇₉, which correspond to ARV residues His₅₁₈, Asp₅₅₇, Asp₅₇₃, and Asp₅₇₅ (Fig. 7, magenta). Of these four residues, only the histidine is not completely conserved in all 31 currently available homologous ARV λ C,

MRV λ 2, and AqRV VP1 protein sequences (Fig. 4). Furthermore, previous MRV:AqRV structural mapping identified 14 additional residues making up a patch surrounding the MTase-1 site (Kim et al., 2004). Most of these (MRV residues Asp₄₈₀, Arg₄₈₁, Lys₄₈₅, Asp₄₈₆, Pro₅₂₇, Val₅₅₄, Tyr₅₇₅, Val₅₈₁, Lys₆₁₅, Phe₆₁₈, and Glu₆₅₀) are also not only conserved in both ARV sequences we determined, but also found in all 29 other currently available homologous ARV λ C, MRV λ 2, and AqRV VP1 protein sequences (Fig. 4). In addition, MRV T1L Asn₆₁₇ is also conserved in both our ARV sequences (as Asn₆₁₃) (Fig. 4). An asparagine residue is found at this position in 30 of the available 31 ARV λ C, MRV λ 2, and AqRV VP1 sequences; this position is occupied by serine in the ARV918 sequence (data not shown). Two other previously identified potentially important conserved residues (MRV T1L Val₄₈₉ and Phe₆₅₂ (Kim et al., 2004)) are replaced with Asn₄₈₇ and Tyr₆₄₈, respectively, in both our ARV sequences. Similarly, previous MRV:AqRV comparisons (Kim et al., 2004) identified six conserved residues within the MTase-2 domain that might be catalytically important; addition of the ARV sequences reduces the number of conserved residues in this suspected catalytic region to 4; MRV residues Asp₈₂₇ and Gly₈₂₉, which bind S-adenosyl-L-methionine (Luongo et al., 1998), and which correspond to ARV residues Asp₈₂₃ and Gly₈₂₅; and MRV residues Asp₈₅₀, and Tyr₈₇₂, which surround the S-adenosyl-L-homocysteine binding site (orange-red), and which correspond to ARV residues Asp₈₄₆, and Tyr₈₆₈. Of these four residues, only Gly₈₂₅ and Asp₈₄₆ are completely conserved in all 31 currently available homologous ARV λ C, MRV λ 2, and AqRV VP1 protein sequences (Fig. 4). Previous MRV:AqRV structural mapping also identified 15 additional residues making up a patch surrounding the MTase-2 site, plus 2 MTase-1 residues projecting into this region from adjacent λ 2 molecules (Kim et al., 2004). Most of these (MRV residues Gly₅₆₉ (from an adjacent MTase-1 domain), Thr₈₃₀, Pro₈₃₂, Glu₈₃₃, Arg₈₅₂, Gly₈₉₃, Ala₈₉₅, Asn₉₂₇, Phe₉₅₁, Arg₉₅₆, and Glu₉₅₈) are conserved in all 31 currently available homologous ARV λ C, MRV λ 2, and AqRV VP1 protein sequences (Fig. 4). In addition, MRV T1L Ala₈₉₄ is also conserved in both our ARV sequences (as Ala₈₉₀) (Fig. 4). Alanine is found at this position in 30 of the available 31 ARV λ C, MRV λ 2, and AqRV VP1 sequences; this position is occupied by threonine in the MRV T3 clone 93 sequence (data not shown). Five other previously identified potentially important conserved residues (MRV T1L Pro₅₆₇ (from an adjacent MTase-1 domain), Leu₈₂₈, Ile₈₃₆, Leu₈₃₉, and Thr₈₈₇ (Kim et al., 2004)) are replaced with Ala₅₆₃ (although a proline residue is present in the ARV sequences at position 561), Cys₈₃₄ in the CSRV sequence; and Pro₈₃₂, Phe₈₃₅, and Ser₈₈₃ in both our newly determined ARV sequences, respectively. Highly conserved GTase residues MRV Lys₁₇₁ and Lys₁₉₀ (black; corresponding to ARV Lys₁₆₉ and Lys₁₈₈) are also conserved in all currently known orthoreovirus turret and homologous proteins. Previous MRV:AqRV structural mapping identified 11 additional residues surrounding the GTase site (Kim et al., 2004). Many of these (MRV residues Tyr₁₉₅ and Glu₂₅₈, which interact with Lys₁₉₀; His₂₂₃ and His₂₃₂, shown to be essential for GTase activity (Qiu and Luongo, 2003); and Ala₁₆₈ and Arg₂₇₈) are also not only conserved in both ARV sequences we determined, but also found in all 29 other currently available homologous ARV λ C, MRV λ 2, and AqRV VP1 protein sequences (Fig. 4). Five other previously identified potentially important conserved residues (MRV T1L Gln₅₈, Tyr₁₇₂, Ala₁₉₆, Thr₂₂₈, and Tyr₂₈₃ (Kim et al., 2004)) are replaced, respectively, with Arg₅₈ in MRV T2J and Arg₅₆ in ARV176; and Phe₁₇₀, Gly₁₉₄, Ser₂₂₆, and Leu₂₈₁ in both newly determined ARV sequences. In addition, a number of

additional amino acids in putative guanylyltransferase domain I (Bisaillon and Lemay, 1999) (MRV residues 219–240), most notably MRV residues Asp₂₂₅, Pro₂₂₇, and Gly₂₃₀, are completely conserved in all 31 currently available homologous ARV λ C, MRV λ 2, and AqRV VP1 protein sequences (Fig. 4). Thus, the majority of the ~60 residues found in or near the MTase-1, MTase-2 and GTase sites are identical in all 31 currently available homologous aligned ARV λ C, MRV λ 2, and AqRV VP1 protein sequences, and many of the non-identical residues found in some clones near these suspected catalytic sites are predicted, by blossom50 matrix weighting, to be conservatively substituted. These observations support, and extend, earlier observations that the orthoreovirus lambda-class core proteins are most conserved within predicted catalytic domains, whereas non-identical amino acids are generally located on the surfaces of the various proteins (Kim et al., 2004). Inclusion of these new ARV data to previous MRV:AqRV comparisons refines the analyses and potentially identifies core residues important for the enzymatic functions of these dsRNA viruses.

In conclusion, we report the first sequence determinations of all 10 genome segments of an avian reovirus, which allows whole-genome comparisons of all 10 segments amongst two different avian reovirus isolates as well as comparisons to the 10 cognate mammalian reovirus genes. As seen with most other ARV genome segments, all ARV L genes are more closely related to homologous MRV L genes than they are to homologous AqRV L genome segments. Comparisons of the ARV and MRV sequences showed nucleotide identity values of 42–55%. Comparisons of encoded ARV and MRV λ protein sequences showed amino acid identity values of about 28% for the core turret protein (comparable to amino acid identity values for most other proteins) and 43% for the major core shell protein, amongst the highest ARV:MRV identity values. Structural mapping of identical and conserved amino acid residues between the closely related ARV and MRV, and amongst these two and the more distantly related AqRV, identified several highly conserved regions. These regions represent probable catalytic domains in the λ -class core turret guanylyltransferase, and probable catalytic domains and patches that traverse the λ -class core shell, and which may contribute to the unusual RNA transcription processes in this group of viruses.

Acknowledgments

We thank members of our laboratory for critical reviews of this manuscript and Kolawole Opanubi for expert technical assistance. This research was supported by grant FRN-11630 from the Canadian Institutes of Health Research.

References

- Bartlett JA, Joklik WK. The sequence of the reovirus serotype 3 L3 genome segment which encodes the major core protein lambda 1. *Virology*. 1988; 167:31–37. [PubMed: 3267236]
- Benavente J, Martinez-Costas J. Avian reovirus: structure and biology. *Virus Res*. 2007; 123:105–119. [PubMed: 17018239]
- Bisaillon M, Bergeron J, Lemay G. Characterization of the nucleoside triphosphate phosphohydrolase and helicase activities of the reovirus lambda1 protein. *J Biol Chem*. 1997; 272:18298–18303. [PubMed: 9218469]
- Bisaillon M, Lemay G. Computational sequence analysis of mammalian reovirus proteins. *Virus Genes*. 1999; 18:13–37. [PubMed: 10334035]

- Bodelon G, Labrada L, Martinez-Costas J, Benavente J. The avian reovirus genome segment S1 is a functionally tricistronic gene that expresses one structural and two nonstructural proteins in infected cells. *Virology*. 2001; 290:181–191. [PubMed: 11883183]
- Breun LA, Broering TJ, McCutcheon AM, Harrison SJ, Luongo CL, Nibert ML. Mammalian reovirus L2 gene and lambda2 core spike protein sequences and whole-genome comparisons of reoviruses type 1 Lang, type 2 Jones, and type 3 Dearing. *Virology*. 2001; 287:333–348. [PubMed: 11531411]
- Calnek, BW., Barnes, JH., Beard, CW., McDougald, LR., Saif, YM. *Diseases of Poultry*. Iowa State University Press; Ames, Iowa: 1997.
- Chappell, JD., Duncan, R., Mertens, PPC., Dermody, TS. Genus orthoreovirus. In: Fauquet, CM.Mayo, MA.Maniloff, J.Desselberger, U., Ball, LA., editors. *Virus Taxonomy Eighth Report of the International Committee on Taxonomy of Viruses*. Elsevier Inc; San Diego, CA: 2005.
- Chen Y, Carrington-Lawrence SD, Bai P, Weller SK. Mutations in the putative zinc-binding motif of UL52 demonstrate a complex interdependence between the UL5 and UL52 subunits of the human herpes simplex virus type 1 helicase/primase complex. *J Virol*. 2005; 79:9088–9096. [PubMed: 15994803]
- Chiu CJ, Lee LH. Cloning and nucleotide sequencing of the S4 genome segment of avian reovirus S1133. *Arch Virol*. 1997; 142:2515–2520. [PubMed: 9672610]
- Dermody TS, Schiff LA, Nibert ML, Coombs KM, Fields BN. The S2 gene nucleotide sequences of prototype strains of the three reovirus serotypes: characterization of reovirus core protein sigma 2. *J Virol*. 1991; 65:5721–5731. [PubMed: 1920614]
- Drastini Y, Kibenge FS, McKenna PK, Lopez A. Comparison of eight different procedures for harvesting avian reoviruses grown in Vero cells. *J Virol Methods*. 1992; 39:269–278. [PubMed: 1430071]
- Dryden, KA., Coombs, KM., Yeager, M. The structure of orthoreoviruses. In: Patton, JT., editor. *Segmented Double-stranded RNA Viruses: Structure and Molecular Biology*. Horizon Press; 2008. p. 3-25.
- Duncan R. Extensive sequence divergence and phylogenetic relationships between the fusogenic and nonfusogenic orthoreoviruses: a species proposal. *Virology*. 1999; 260:316–328. [PubMed: 10417266]
- Duncan R, Sullivan K. Characterization of two avian reoviruses that exhibit strain-specific quantitative differences in their syncytium-inducing and pathogenic capabilities. *Virology*. 1998; 250:263–272. [PubMed: 9792837]
- Gershowitz A, Wooley RW. Characterization of two reoviruses isolated from turkeys with infectious enteritis. *Avian Dis*. 1973; 17:406–414. [PubMed: 4351419]
- Glass SE, Nagi SA, Hall CF, Kerr KM. Isolation and characterization of a virus associated with arthritis of chickens. *Avian Dis*. 1973; 17:415–424. [PubMed: 4351420]
- Gonzalez-Lopez C, Martinez-Costas J, Esteban M, Benavente J. Evidence that avian reovirus sigmaA protein is an inhibitor of the double-stranded RNA-dependent protein kinase. *J Gen Virol*. 2003; 84:1629–1639. [PubMed: 12771434]
- Gouvea VS, Schnitzer TJ. Polymorphism of the genomic RNAs among the avian reoviruses. *J Gen Virol*. 1982; 61(Pt 1):87–91. [PubMed: 7119751]
- Harrison SJ, Farsetta DL, Kim J, Noble S, Broering TJ, Nibert ML. Mammalian reovirus L3 gene sequences and evidence for a distinct amino-terminal region of the lambda1 protein. *Virology*. 1999; 258:54–64. [PubMed: 10329567]
- Herenda, DC., Franco, DA. *Poultry Diseases and Meat Hygiene*. Iowa State University Press; Ames: 1996.
- Hermann L, Embree J, Hazelton P, Wells B, Coombs K. Reovirus type 2 isolated from cerebrospinal fluid. *Pediatr Infect Dis*. 2004; 23:373–375.
- Hieronymus DR, Villegas P, Kleven SH. Identification and serological differentiation of several reovirus strains isolated from chickens with suspected malabsorption syndrome. *Avian Dis*. 1983; 27:246–254. [PubMed: 6189477]
- Hsiao J, Martinez-Costas J, Benavente J, Vakharia VN. Cloning, expression, and characterization of avian reovirus guanylyltransferase. *Virology*. 2002; 296:288–299. [PubMed: 12069527]

- Johansson PJ, Sveger T, Ahlfors K, Ekstrand J, Svensson L. Reovirus type 1 associated with meningitis. *Scand J Infect Dis*. 1996; 28:117–120. [PubMed: 8792475]
- Jones RC. Avian reovirus infections. *Rev Sci Technol*. 2000; 19:614–625.
- Jordan, FTW., Pattison, M. *Poultry Diseases*. W.B. Saunders Company; London: 1996. Reoviridae.
- Joske RA, Keal DD, Leak PJ, Stanley NF, Walters MN. Hepatitis–encephalitis in humans with reovirus infection. *Arch Inter Med*. 1964; 113:811–816.
- Kapczynski DR, Sellers HS, Simmons V, Schultz-Cherry S. Sequence analysis of the S3 gene from a turkey reovirus. *Virus Genes*. 2002; 25:95–100. [PubMed: 12206313]
- Kim J, Tao Y, Reinisch KM, Harrison SC, Nibert ML. Orthoreovirus and aquareovirus core proteins: conserved enzymatic surfaces, but not protein–protein interfaces. *Virus Res*. 2004; 101:15–28. [PubMed: 15010214]
- Kouwenhoven B, Vertommen M, Eck JHV. Runting and leg weakness in broilers: involvement of infectious factors. *Vet Sci Commun*. 1978; 2:253–259.
- Krainer L, Aronson BE. Disseminated encephalomyelitis in humans with recovery of hepatoencephalitis virus. *J Neuropathol Exp Neurol*. 1959; 18:339–342.
- Lipman DJ, Pearson WR. Rapid and sensitive protein similarity searches. *Science*. 1985; 227:1435–1441. [PubMed: 2983426]
- Liu HJ, Huang PH. Sequence and phylogenetic analysis of the sigma A-encoding gene of avian reovirus. *J Virol Methods*. 2001; 98:99–107. [PubMed: 11576636]
- Luongo CL. Mutational analysis of a mammalian reovirus mRNA capping enzyme. *Biochem Biophys Res Commun*. 2002; 291:932–938. [PubMed: 11866455]
- Luongo CL, Contreras CM, Farsetta DL, Nibert ML. Binding site for S-adenosyl-L-methionine in a central region of mammalian reovirus lambda2 protein. Evidence for activities in mRNA cap methylation. *J Biol Chem*. 1998; 273:23773–23780. [PubMed: 9726986]
- Luongo CL, Reinisch KM, Harrison SC, Nibert ML. Identification of the guanylyltransferase region and active site in reovirus mRNA capping protein lambda2. *J Biol Chem*. 2000; 275:2804–2810. [PubMed: 10644745]
- Martinez HM. An efficient method for finding repeats in molecular sequences. *Nucleic Acids Res*. 1983; 11:4629–4634. [PubMed: 6866775]
- Martinez-Costas J, Grande A, Varela R, Garcia-Martinez C, Benavente J. Protein architecture of avian reovirus S1133 and identification of the cell attachment protein. *J Virol*. 1997; 71:59–64. [PubMed: 8985323]
- Martinez-Costas J, Varela R, Benavente J. Endogenous enzymatic activities of the avian reovirus S1133: identification of the viral capping enzyme. *Virology*. 1995; 206:1017–1026. [PubMed: 7856076]
- McCrae MA, Joklik WK. The nature of the polypeptide encoded by each of the 10 double-stranded RNA segments of reovirus type 3. *Virology*. 1978; 89:578–593. [PubMed: 716218]
- Mendez II, Weiner SG, She YM, Yeager M, Coombs KM. Conformational changes accompany activation of reovirus RNA-dependent RNA transcription. *J Struct Biol*. 2008; 162:277–289. [PubMed: 18321727]
- Mertens P. The dsRNA viruses. *Virus Res*. 2004; 101:3–13. [PubMed: 15010213]
- Ni Y, Ramig RF. Characterization of avian reovirus-induced cell fusion: the role of viral structural proteins. *Virology*. 1993; 194:705–714. [PubMed: 8503184]
- Noad L, Shou JY, Coombs KM, Duncan R. Sequences of avian reovirus M1, M2 and M3 genes and predicted structure/function of the encoded mu proteins. *Virus Res*. 2006; 116:45–57. [PubMed: 16297481]
- Noble S, Nibert ML. Core protein mu2 is a second determinant of nucleoside triphosphatase activities by reovirus cores. *J Virol*. 1997; 71:7728–7735. [PubMed: 9311857]
- Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. 2000; 302:205–217. [PubMed: 10964570]
- O'Hara D, Patrick M, Cepica D, Coombs KM, Duncan R. Avian reovirus major mu-class outer capsid protein influences efficiency of productive macrophage infection in a virus strain-specific manner. *J Virol*. 2001; 75:5027–5035. [PubMed: 11333882]

- Olson NO. Transmissible synovitis of poultry. *Lab Invest.* 1959; 8:1384–1393. [PubMed: 14428822]
- Olson, NO. Reovirus infections. In: Hofstad, MS., editor. *Diseases of Poultry*. Iowa State University Press; Ames, Iowa: 1978.
- Olson NO, Solomon DP. A natural outbreak of synovitis caused by the viral arthritis agent. *Avian Dis.* 1968; 12:311–316. [PubMed: 4872845]
- Patrick M, Duncan R, Coombs KM. Generation and genetic characterization of avian reovirus temperature-sensitive mutants. *Virology.* 2001; 284:113–122. [PubMed: 11352672]
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera—a visualization system for exploratory research and analysis. *J Comp Chem.* 2004; 25:1605–1612. [PubMed: 15264254]
- Qiu T, Luongo CL. Identification of two histidines necessary for reovirus mRNA guanylyltransferase activity. *Virology.* 2003; 316:313–324. [PubMed: 14644613]
- Reddy VS, Natarajan P, Okerberg B, Li K, Damodaran KV, Morton RT, Brooks C, Johnson JE. Virus particle explorer (VIPER), a website for virus capsid structures and their computational analyses. *J Virol.* 2001; 75:11943–11947. [PubMed: 11711584]
- Reinisch KM, Nibert ML, Harrison SC. Structure of the reovirus core at 3.6 Å resolution. *Nature.* 2000; 404:960–967. [PubMed: 10801118]
- Rosenberger, JK., Olson, NO. Reovirus infections. In: Calnek, BW, Burnes, MJ, Beard, CW, Reid, WM., Yoder, HW., editors. *Diseases of Poultry*. Iowa State University Press; Ames, Iowa: 1991.
- Sambrook, J., Fritsch, EF., Maniatis, T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory; Cold Spring Harbor: 1989.
- Schiff LA, Nibert ML, Co MS, Brown EG, Fields BN. Distinct binding sites for zinc and double-stranded RNA in the reovirus outer capsid protein sigma 3. *Mol Cell Biol.* 1988; 8:273–283. [PubMed: 3275869]
- Schiff, LA., Nibert, ML., Tyler, KL. Orthoreoviruses and their replication. In: Knipe, DM., Howley, PM., editors. *Fields Virology*. Lippincott Williams & Wilkins; Philadelphia: 2007. p. 1853–1915.
- Schnitzer TJ. Protein coding assignment of the S genes of the avian reovirus S1133. *Virology.* 1985; 141:167–170. [PubMed: 3976177]
- Sellers HS, Linnemann EG, Pereira L, Kapczynski DR. Phylogenetic analysis of the sigma 2 protein gene of turkey reoviruses. *Avian Dis.* 2004; 48:651–657. [PubMed: 15529990]
- Seybert A, Posthuma CC, van Dinten LC, Snijder EJ, Gorbalenya AE, Ziebuhr J. A complex zinc finger controls the enzymatic activities of nidovirus helicases. *J Virol.* 2005; 79:696–704. [PubMed: 15613297]
- Shen PC, Chiou YF, Liu HJ, Song CH, Su YP, Lee LH. Genetic variation of the lambda A and lambda C protein encoding genes of avian reoviruses. *Res Vet Sci.* 2007; 83:394–402. [PubMed: 17336355]
- Shmulevitz M, Yameen Z, Dawe S, Shou J, O'Hara D, Holmes I, Duncan R. Sequential partially overlapping gene arrangement in the tricistronic S1 genome segments of avian reovirus and Nelson Bay reovirus: implications for translation initiation. *J Virol.* 2002; 76:609–618. [PubMed: 11752152]
- Spandidos DA, Graham AF. Physical and chemical characterization of an avian reovirus. *J Virol.* 1976; 19:968–976. [PubMed: 987252]
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol.* 2007; 24:1596–1599. [PubMed: 17488738]
- Theophilos MB, Huang JA, Holmes IH. Avian reovirus sigma C protein contains a putative fusion sequence and induces fusion when expressed in mammalian cells. *Virology.* 1995; 208:678–684. [PubMed: 7747439]
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; 22:4673–4680. [PubMed: 7984417]
- Touris-Otero F, Martinez-Costas MCSM, Benavente J. Avian reovirus morphogenesis occurs within viral factories and begins with the selective recruitment of sigma NS and lambda A to mu NS inclusions. *J Mol Biol.* 2004; 341:361–374. [PubMed: 15276829]

- Tyler KL, Barton ES, Ibach ML, Robinson C, Campbell JA, O'Donnell SM, Valyi-Nagy T, Clarke P, Wetzel JD, Dermody TS. Isolation and molecular characterization of a novel type 3 reovirus from a child with meningitis. *J Infect Dis*. 2004; 189:1664–1675. [PubMed: 15116303]
- Varela R, Benavente J. Protein coding assignment of avian reovirus strain S1133. *J Virol*. 1994; 68:6775–6777. [PubMed: 8084013]
- Wiener JR, Bartlett JA, Joklik WK. The sequences of reovirus serotype 3 genome segments M1 and M3 encoding the minor protein mu 2 and the major nonstructural protein mu NS, respectively. *Virology*. 1989; 169:293–304. [PubMed: 2523177]
- Wilbur WJ, Lipman DJ. Rapid similarity searches of nucleic-acid and protein data banks. *Proc Natl Acad Sci USA*. 1983; 80:726–730. [PubMed: 6572363]
- Xu W, Coombs KM. Avian reovirus L2 genome segment sequences and predicted structure/function of the encoded RNA-dependent RNA polymerase protein. *Virol J*. 2008; 5:153. [PubMed: 19091125]
- Yin HS, Su YP, Lee LH. Evidence of nucleotidyl phosphatase activity associated with core protein sigma A of avian reovirus S1133. *Virology*. 2002; 293:379–385. [PubMed: 11886258]
- Yin P, Keirstead ND, Broering TJ, Arnold MM, Parker JSL, Nibert ML, Coombs KM. Comparisons of the M1 genome segments and encoded μ 2 proteins of different reovirus isolates. *Virol J*. 2004; 1(6)
- Zhang X, Tang J, Walker SB, O'Hara D, Nibert ML, Duncan R, Baker TS. Structure of avian orthoreovirus virion by electron cryomicroscopy and image reconstruction. *Virology*. 2005; 343:25–35. [PubMed: 16153672]
- Zhang X, Walker SB, Chipman PR, Nibert ML, Baker TS. Reovirus polymerase lambda 3 localized by cryoelectron microscopy of virions at a resolution of 7.6 angstrom. *Nat Struct Biol*. 2003; 10:1011–1018. [PubMed: 14608373]

Fig. 1.

Alignment of the deduced ARV138 and ARV176 λA amino acid sequences. All 26 currently available homologous ARV λA, MRV λ1, and AqRV VP3 proteins (determined for each clone shown in Table 1) were aligned, both by T-Coffee (Notredame et al., 2000) (data not shown) and by Clustal-W (Thompson et al., 1994), with only minor differences in the alignments created by different gap penalties (data not shown). Only the two most-distant ARV, MRV, and AqRV sequences (see text for details) are shown for clarity. Clones are: MRV–T1L (GenBank no. NC_004255) and T2J (GenBank no. NC_004256); ARV–ARV138 (GenBank no. EU707933) and ARV176 (GenBank no. EU707934); AqRV–Grass Carp reovirus (GCRV) (GenBank no. AF260513) and Chum Salmon reovirus (CSRV) (GenBank No. NC_007584). Amino acid residues that are identical in at least four of the sequences are indicated by black background shading. The single letter amino acid code is used. Six previously identified putative helicase domains (labeled I–VI) (Bisaillon and Lemay, 1999) are indicated with solid horizontal lines above the sequence, and previously identified putative 5'-RNA triphosphatase domains (labeled I–II) (Bisaillon and Lemay, 1999) are indicated with dashed horizontal lines above the sequences. Amino acid residues that are completely conserved in all 26 sequences are indicated by (●) above the sequences.

Fig. 2. Phylogenetic tree analyses of the prototype ARV L1 (A) and L3 (B) genes and homologous genes in other reoviruses. Abbreviations are as defined in the legend to Fig. 1 and Table 1. Lines are proportional in length to nucleotide substitution. Alignments were performed by Neighbor-Joining and tested with 1000 bootstrap replicates in MEGA version 4 (Tamura et al., 2007). The ARV138 and ARV176 clones are indicated with arrows.

Fig. 3.

Window-averaged scores for sequence identity among the ARV, AqRV, and MRV λ -class major core shell proteins (A) and λ -class core turret proteins (B). To provide consistent weighting to the averaged scores, only the two most-distant clones from each of the three groups (ARV: ARV138 and ARV176; AqRV: GCRV and CSRV; MRV: T1L and T2J—see text for details) were used. Identity scores averaged over running windows of 15 amino acids and centered at consecutive amino acid positions are shown for ARV:MRV comparisons (dashed lines) and ARV:MRV:AqRV comparisons (solid line). The global identity scores for each of the compared sequence sets are indicated by the horizontal lines. Previously identified enzymatic motifs are indicated above the plots in (B).

Fig. 4.

Alignment of the deduced ARV138 and ARV176 λC amino acid sequences. All 31 currently available homologous ARV λC, MRV λ2, and AqRV VP1 proteins (determined for each clone shown in Table 1) were aligned, both by T-Coffee (Notredame et al., 2000) (data not shown) and by Clustal-W (Thompson et al., 1994), with only minor differences in the alignments created by different gap penalties (data not shown). Only the two most-distant ARV, MRV, and AqRV sequences (see text for details) are shown for clarity. Clones are: MRV–T1L (GenBank no. NC_004259) and T2J (GenBank no. NC_004260); ARV–ARV138 (GenBank no. EU707937) and ARV176 (GenBank no. EU707938); AqRV–Grass Carp reovirus (GCRV) (GenBank no. AF260511) and Chum Salmon reovirus (CSRV) (GenBank no. NC_007582). Amino acid residues that are identical in at least four of the sequences are indicated by black background shading. The single letter amino acid code is used.

Previously identified putative guanylyltransferase domains (labeled I–II) (Bisaillon and Lemay, 1999) are indicated with solid horizontal lines above the sequences, and previously identified putative methyltransferase domains (labeled I–IV) (Bisaillon and Lemay, 1999) are indicated with dashed horizontal lines above the sequences. Amino acid residues that are completely conserved in all 31 sequences are indicated by (●) above the sequences.

Fig. 5. Pairwise sequence identities of ARV138 and ARV176 genes (A) and deduced proteins (B). The genome segments and proteins are organized in decreasing rank (left to right) with the sequence identity value shown within each vertical bar.

Fig. 6. Localization of conserved, non-conserved, and identical amino acids in ARV, MRV, and AqRV core shell protein. The MRV core crystal structure (Reinisch et al., 2000) asymmetric unit (PDB #1EJ6) was assembled into a pentameric vertex aggregate with Viper[®] (Reddy et al., 2001) and manipulated with Chimera[®] (Pettersen et al., 2004). (A) Low-resolution, cutaway model of the reovirus core structure (modified from (Dryden et al., 2008) with permission). (B–C) Blow-up of indicated $\lambda 1$ molecules in “A”; with “C” rotated 90° towards the viewer to better illustrate the decameric organization of 10 $\lambda 1$ molecules. The $\lambda 1$ molecules that approach the 5′-axis (designated $\lambda 1.5$) are shown in salmon and the $\lambda 1$ molecules that approach the 3′-axis (designated $\lambda 1.3$) are shown in magenta. (D) Same top view of decameric structure as “C”, but in “D”–“G”, amino acids that are identical in all 6 ARV, MRV, and AqRV sequences (see Fig. 1) are shown in darker versions of each motif

color (firebrick and deep pink, respectively), amino acids that represent conservative substitutions (as determined by blossom50 matrix) are shown in lighter versions of each motif color (salmon and hotpink, respectively), and non-conserved amino acids are shown in white and grey, respectively, for λ 1.5 and λ 1.3. (E) Bottom view of decamer with coloration same as in “D”; in addition, the N-terminal 240 residues, which are visualized only in λ 1.3 molecules, are depicted in pale yellow (for non-conservative substitutions), cyan for conservative substitutions, and blue for identical amino acids, respectively. (F) Slabbed section of “D”, approximately as indicated by labelled horizontal line in “B”; note that apparent “hole” in structure is caused by linear transverse section of convex shell. (G) Slabbed section of “E”, approximately as indicated by labelled horizontal line in “B”. (H) Slabbed cross-section through middle of “D”, as indicated by labelled line “H”. (I) Slabbed cross-section through edge of “D”, as indicated by labelled line “I”. The highly conserved C₂H₂ zinc finger motif is indicated in black (and with large arrow heads) in one λ 1.3 molecule, highly conserved residues in putative helicase domains are indicated in green, and highly conserved residues in putative 5' RNA triphosphatase domains are indicated in gold (and with narrow arrows) in only some molecules. Some highly conserved patches (see text) are indicated in some λ 1.5 (circles and hexagons) and λ 1.3 molecules (squares and hexagons) in “D”, “H”, and “I”.

Fig. 7. Localization of conserved, non-conserved, and identical amino acids in ARV, MRV, and AqRV turret proteins. The MRV core crystal structure (Reinisch et al., 2000) asymmetric unit (PDB #1EJ6) was assembled into a pentameric vertex aggregate with Viper® (Reddy et al., 2001), and manipulated with Chimera® (Pettersen et al., 2004). (A) Low-resolution, cutaway model of the reovirus core structure (modified from (Dryden et al., 2008) with permission). (B) Blow-up of indicated $\lambda 2$ turret in 'A', with front-most $\lambda 2$ molecule in space-filling mode and color-coded according to domain; pink for the N-terminal GTase domain (amino acids 1–380), grey for the extended region (aa. 381–432), yellow for the

MTase-1 domain (residues 433–691), green for the spacer domain (aa. 692–805), blue for the MTase-2 domain (aa. 806–1022), and purple for the flap domain (residues 1023–1289); the four other λ 2 molecules in the turret are depicted in cyan in wire backbone mode. In “C”–“L”, amino acids that are identical in all six ARV, MRV, and AqRV sequences (see Fig. 4) are shown in darker versions of each domain color (deep pink, dim grey, goldenrod, dark green, blue, and purple, respectively), amino acids that represent conservative substitutions (as determined by blossom50 matrix) are shown in lighter versions of each domain color (hotpink, dark grey, yellow, green, cyan, and plum, respectively), and non-conserved amino acids are shown in white. Highly conserved lysine residues (Lys₁₇₁ and Lys₁₉₀ in the MRV sequence, shown to be important for guanylyltransferase activity (Luongo et al., 2000; Luongo, 2002)) are indicated in black, highly conserved MTase-1 residues His₅₂₁, Asp₅₆₁, Asp₅₇₇, and Asp₅₇₉ (see text) are depicted in magenta, and highly conserved MTase-2 residues Asp₈₂₇, Gly₈₂₉, Asp₈₅₀, and Tyr₈₇₂ (see text) are shown in orange-red. (C) (top view) and (D) (bottom view) of flaps domain; (E) (top view) and (F) (bottom view) of MTase domains; (H) (top view) and (I) (bottom view) of spacer and extended region domains; and (J) (top view) and (K) (bottom view) of GTase domain. (G and L) Slabbed sections (cut approximately mid-way and viewed from top) of MTase and GTase domains, respectively, to allow better visualization of catalytic sites and indicated highly conserved residues.

Table 1

Nucleotide sequences used in this study^d.

Strain	GeneBank accession number	
	Core shell	Core spike
<i>ARV^b</i>		
138	EU707933	EU707937
176	EU707934	EU707938
S1133	AY641735	AY652693
1733	AY641741	AF384171
2408	AY641742	AY652694
601G	AY641736	AY652699
916	AY641737	AY652701
918	AY641738	AY652700
919	AY641739	AY652697
1017-1	AY641740	DQ238096
OS161	AY641743	AY652696
R2	AY641744	DQ238095
C-98	EU616735	EU616737
T-98	EU616739	EU616738
750505	DQ238093	AY652695
T6	DQ238094	AY652698
<i>MRV^c</i>		
T1L	NC 004255	NC 004259
T2J	NC 004256	NC 004260
T3D	NC 004274	EF494436
BYD1	DQ664186	DQ664185
SC-A	EF029088	DQ885990
Neth/85	NA	AF378004
SV59	NA	AF378006
9	NA	AF378007
18	NA	AF378008
87	NA	AF378009
93	NA	AF378010
T4N	NA	NA
<i>AqRV^d</i>		
GCRV	AF260513	AF260511
GCHV	AF284503	NA
GSRV	NC 005168	NC 005166
AGCRV	NC 010586	NC 010584
CSRV	NC 007584	NC 007582
ASRV	NA	NA

^aNA, nucleotide sequences not available.

^bARV, avian reovirus.

^cMRV, mammalian reovirus. T1L, type 1 Lang; T2J, type 2 Jones; T3D, type 3 Dearing; T4N, type 4 Ndelle.

^dAqRV, Aquareovirus. GCRV, Grass carp reovirus; GCHV, Grass carp hemorrhagic virus; GSRV, Golden shiner reovirus; AGCRV, American grass carp reovirus; CSRV, Chum salmon reovirus; ASRV, Atlantic salmon reovirus.

Table 2

Primers used for amplification of ARV L1 and L3 genome segments.

Primer ^a	Sequence (5'–3')	Position	Target gene
ARVL1F	GCTTTTCTCCGAACGCCG	1–18	λA
ARVL1R	GATGAATAATCTCCAACGAGAGTCG	3958–3934	
ARVL3F	GCTTTTCCACCCATGGCTCAG	1–21	λC
ARVL3R	GATGAATAACACCCTTCTACTGG	3907–3885	

^aF, forward primer; R, reverse primer.

Table 3
Genome-segment lengths, non-translated regions, and encoded proteins of ARV138 and ARV176.

Genome segment	Base pairs ^d	5' NTR ^b (no. of bases)	3' NTR (no. of bases)	ORF ^c	Codons ^d	Protein	Molecular weight (kDa) ^e	
							ARV138	ARV176
L1	3,958	20	56	21–3899	1293	λA	142.3	142.2
L2	3,829 ^f	13 ^g	36	14–3790 ^h	1259	λB	139.7	139.8
L3	3,907	12	37	13–3867	1285	λC	141.9	142.2
M1	2,283	12	72	13–2208	732	μA	82.0	82.2
M2	2,158	29	98	30–2057	676	μB	73.1	73.3
M3	1,996	24	64	25–1929	635	μC	70.9	70.8
S1	1,643	24	33	25–318	98	p10	10.3	10.3
				293–730	146	p17	16.9	16.9
S2	1,324	15	58	630–1607	326	σC	34.9	34.8
S3	1,202	30	68	16–1263	416	σA	46.1	46.1
S4	1,192	23	65	31–1131	367	σB	40.9	40.9
Total	23,492 ⁱ			24–1124	367	σNS	40.5	40.6

^aTotal nucleotides on each strand.
^bNTR, non-translated region.
^cNucleotide positions indicated for starting and ending codons.
^dTotal number of amino acids in deduced protein.
^eMolecular weight calculated from deduced protein and rounded to closest 0.1 kDa.
^f3830 for ARV176, from (Xu and Coombs, 2008).
^g14 for ARV176, from (Xu and Coombs, 2008).
^h15–3791 for ARV176, from (Xu and Coombs, 2008).
ⁱ23,493 for ARV176.

Table 4

Percent identities of the L-class homologous genome segments and encoded proteins of ARV, MRV and aquareoviruses^a.

Strain	ARV138	ARV176	ARV601G	T1L	T2J	T3D	T4N	GCRV	CSRV
ARV138		98, 98, 84	97, -, 83	44, 55, 28	44, 55, 27	44, 55, 27	-, 55, -	32, 42, 24	28, 41, 25
ARV176	91, 85, 73		97, -, 98	44, 55, 27	44, 55, 28	44, 55, 27	-, 55, -	32, 42, 23	28, 41, 24
ARV601G	86, -, 73	86, -, 95		44, -, 28	44, -, 28	44, -, 28	-, -, -	32, -, 23	28, -, 24
T1L	50, 55, 43	51, 55, 42	51, -, 43		96, 92, 87	99, 99, 93	-, 97, -	32, 42, 27	29, 41, 25
T2J	51, 55, 43	51, 55, 44	50, -, 44	77, 75, 72		96, 92, 87	-, 91, -	32, 42, 27	28, 40, 25
T3D	50, 55, 43	51, 55, 43	51, -, 43	98, 96, 76	77, 76, 74		-, 98, -	32, 42, 27	29, 41, 25
T4N	-, 56, -	-, 56, -	-, -, -	-, 89, -	-, 75, -	-, 90, -		-, 42, -	-, 41, -
GCRV	44, 49, 38	44, 49, 39	43, -, 39	43, 48, 39	43, 47, 40	42, 48, 39	-, 47, -		48, 58, 44
CSRV	42, 47, 38	43, 47, 38	43, -, 38	43, 47, 41	42, 46, 40	42, 47, 39	-, 47, -	58, 59, 52	

^a -, sequences not available. Percent amino acid identities indicated in upper triangle; the first number is for the core shell protein (this study), the second number is for the RNA-dependent RNA polymerase (Xu and Coombs, 2008), and the third number is for the core-spike protein (this study). Percent nucleotide identities are in lower triangle, in bold; the order of the numbers is the same as described for proteins.