Published in final edited form as:

Infect Genet Evol. 2015 June ; 32: 313–321. doi:10.1016/j.meegid.2015.03.029.

Genomic characterization of a turkey reovirus field strain by Next-Generation Sequencing

Yi Tang¹, Huaguang Lu^{1,*}, Aswathy Sebastian², Yin-Ting Yeh³, Craig A. Praul⁴, Istvan U. Albert², and Si-Yang Zheng³

¹Wiley Lab / Avian Virology, Department of Veterinary and Biomedical Sciences, The Pennsylvania State University, University Park, PA 16802

²Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802

³Department of Biomedical Engineering, The Pennsylvania State University, University Park, PA 16802

⁴Genomics Core Facility, The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802

Abstract

The genome of a turkey arthritis reovirus (TARV) field strain (Reo/PA/Turkey/22342/13), isolated from a turkey flock in Pennsylvania (PA) in 2013, has been sequenced using Next-Generation Sequencing (NGS) on the Illumina MiSeq platform. The genome of the PA TARV field strain was 23,496bp in length with 10 dsRNA segments encoding 12 viral proteins. The lengths of the genomic segments ranged from 1,192bp (S4) to 3,959bp (L1). The 5' and 3' conserved terminal sequences of the PA TARV field strain were similar to the two Minnesota (MN) TARVs (MN9 and MN10) published recently and avian orthoreovirus (ARV) reference strains. Phylogenetic analysis of the nucleotide sequences of all 10 genome segments revealed that there was a low to significant nucleotide sequence divergence between the PA TARV field strain and reference TARV and ARV strains. Analysis of the PA TARV sequence indicates that this PA TARV field strain is a unique strain and is different from the TARV MN9 or MN10 in M2 segment genes and ARV S1133 vaccine strain.

Keywords

Genome segments; genotyping; Next-Generation Sequencing; turkey arthritis reovirus

*corresponding author's telephone: 814-863-4369; fax: 814-865-4717; hx115@psu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Avian orthoreoviruses (ARV) belong to the genus *Orthoreovirus* in the family *Reoviridae*. They are non-enveloped viruses and contain a double-stranded RNA (dsRNA) genome with 10 segments. The viral genome is enclosed within a double protein capsid shell with a diameter of 70-80nm (Spandidos and Graham, 1976; van der Heide, 2000). Polyacrylamide gel electrophoresis indicates that the whole genome can be separated into 10 genomic segments including 3 large (L1, L2, L3), 3 medium (M1, M2 and M3) and 4 small (S1, S2, S3 and S4) segments (Nick et al., 1975; Varela and Benavente, 1994). There are 8 structural proteins (λ A, λ B, λ C, μ A, μ B, σ A, σ B and σ C) and 4 non-structural proteins (μ NS, P10, P17 and σ NS) encoded by the segmented genome (Varela et al., 1996).

ARV is one of the most important avian viruses causing clinical diseases in poultry worldwide. ARV affected poultry flocks commonly suffer from viral arthritis/tenosynovitis, enteric disease, immunosuppression, runting-stunting syndrome, and malabsorption syndrome (Sterner et al., 1989; van der Heide, 2000). Most domestic avian species are susceptible to ARV infections, such as broiler breeders (Ide and Dewitt, 1979), layer breeders (De Gussem et al., 2010), broilers (Howell and Walker, 1972), goose (Bezerra et al., 2012; Palya et al., 2003), turkey (Giangaspero et al., 1997; Taber et al., 1976), duck (Baroni et al., 1980; Petek et al., 1973; Rey et al., 1999), pigeon (Vindevogel et al., 1982), and quail (Guy et al., 1987; Magee et al., 1993; Ritter et al., 1986). However, the meat-type chickens are well documented species that are more susceptible to the viral arthritis/tenosynovitis caused by ARV than other avian species (De Gussem et al., 2010).

The outbreaks of turkey arthritis reovirus (TARV) infections initially occurred in the United States in late 1970's and 1980's (Afaleq and Jones, 1989; Levisohn et al., 1980). The recent TARV outbreaks started in Minnesota (MN) in 2009. Research findings by Mor's group suggested that the MN TARV strains obtained from these outbreaks were likely the tropism of ARV variants re-involved with arthritis in turkeys (Mor et al., 2013). A similar TARV infection problem occurred in a commercial turkey flock at 9 weeks of age in Pennsylvania (PA) in summer 2011, and another flock (17 weeks old) of the same company was infected a week later. The affected birds suffered from severe lameness and tenosynovitis, tendon rupture or hemorrhage. Tendon and synovial tissues collected from sick birds were virus isolation positive for TARV. The 2011 TARV cases diagnosed in PA were the first index cases representing TARV infections in the east region of United States, and similar TARV cases were continuously diagnosed in 2013, 2014 and 2015.

Until recently, research findings in TARV sequencing studies, mostly based on the L-class and S-class genomic segments, showed a great genetic divergence between TARVs and ARV vaccine strains (Mor et al., 2014a; Mor et al., 2014b). Other research studies indicated that highly pathogenic ARV variants were capable of infecting not only unvaccinated broiler chickens, but also vaccinated broiler breeders (Dandar et al., 2013; Rosenberger et al., 2013; Troxler et al., 2013). In this paper, we report the complete genome sequence of a PA field strain (Reo/PA/Turkey/22342/13), isolated from TARV infected turkeys showing severe arthritis and lameness in a commercial turkey flock at 14 weeks of age, by using Next-Generation Sequencing (NGS) with the Illumina MiSeq.

2. Materials and methods

2.1 The TARV strain and RNA sample preparation

The PA TARV field strain (Reo/PA/Turkey/22342/13) used in this study was isolated from tendons and synovial tissues of sick birds showing arthritis/tenosynovitis in a PA commercial turkey flock at 14 weeks of age in 2013. Virus isolation and propagation were conducted in LMH (Leghorn Male-chicken Hepatocellular-carcinoma, ATCC CRL-2113) cell cultures as routine cell culture procedures (Heggen-Peay et al., 2002). Viral RNA extraction from the TARV infected cell culture fluid was carried out with a total RNA extraction Kit (QIAGEN, Valencia, CA, USA) following manufacturer's instructions.

2.2 Sequencing

Libraries were constructed from total RNA samples using the TruSeq Stranded Total RNA Sample Prep Kit (Illumina, San Diego, CA, USA) according the manufacturer's protocol with the exception that the initial poly A enrichment step was skipped. Library size and quality was assessed using the Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Library concentration was assessed by qPCR using the KAPA Library Quantification Kit Illumina Platforms (Kapa Biosystems, Wilmington, MA, USA). Libraries were sequenced on an Illumina MiSeq using 150 nt single read sequencing according to the manufacturer's protocol (Illumina, San Diego, CA, USA).

2.3 Viral genome assembly

The overall pipeline for sequence data analysis started with raw NGS reads (Fig. 1). Briefly, the first step was to trim off the adaptors added for sequencing and remove reads mapping to contaminants (e.g., mRNA, rRNA, chicken and human sequences) using both sortMeRNA (Kopylova et al., 2012) and BWA-MEM (Li and Durbin, 2010) methods. Unmatched sequence reads were assembled using *de novo* SPAdes assembly software (ver 3.5.0) (Bankevich et al., 2012). All the assembled contiguous sequences (contigs) were aligned to the reference genome using LASTZ (Harris, 2007) to identify and extract maximally aligned viral contigs. To further improve the contigs, all raw reads of each segment were mapped back to the assembled contigs. Finally, the consensus sequences from the re-mapping reads and LASTZ contig alignment were obtained using SAMtools commands (Li et al., 2009).

2.4 Obtaining 5' and 3' termini

The rapid amplification cDNA ends (RACE) methods were used to obtain the 5' and 3' termini for each of the 10 genome segments. A short oligonucleotide PC3, which was phosphorylated at the 5' end and blocked at the 3' end with dideoxy cytosine, was ligated to the 3' ends of extracted the genomic RNA (Watson et al., 1992). The ligation reaction was performed by T4 RNA ligase (New England Bio Labs, Ipswich, MA, USA). Following the incubation, the ligated dsRNA was purified using agarose gel extraction columns following the manufacturer's instructions (Lot No. 04113KE1, Axygen, Tewksbury, MA, USA). Subsequently, the PC2 complementary primer to the ligated oligonucleotide was combined with gene specific primers in different reactions for 5' and 3' ends of each genomic segment amplification and sequencing, respectively, using the conditions as described above. The

DNA concentration of the purified PCR product was measured using a NanoDrop™1000 (Thermo Scientific, Waltham, MA, USA) spectrophotometer and then submitted to Penn State Genomics Core Facility for Sanger sequencing.

2.5 Sequence analyses

Lasergene 12 Core Suite (DNASTAR, Inc. Madison, WI, USA) was used for Sanger sequencing results assembly, viral ORFs prediction and nucleotide (nt) sequences translation. Sequence similarity was evaluated using BLASTN search in GenBank (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). The alignments of sequences were carried out using the ClustalW 1.83 program (<http://align.genome.jp/>). Neighbor-joining and maximum-likelihood (ML) phylogenetic trees were generated and tree topologies were validated by bootstrap analysis as implemented in MEGA program (Version 5.0) with absolute distances following 1,000 bootstrap replicates (Tamura et al., 2011). Visualizing of NGS and viral genomic data plot were generated by the Circos method (Krzywinski et al., 2009). The S1 gene ORF organization mapping was performed using CLC Genomic Workbench V7.5 software (QIAGEN, Boston, MA, USA). Analysis of whole genome alignments was performed using the mVISTA online platform (<http://genome.lbl.gov/vista/mvista/submit.shtml>). In order to conduct genome comparison of this PA TARV field strain with other reference strains, full genomic sequences of two MN turkey TARV strains (MN9, MN10) and 6 ARV reference strains (S1133, 1733, 138, 176, AVS-B and J18) retrieved from GenBank (Table S1) were used for comparison analysis.

3. Results

3.1 Viral RNA extraction and RT-PCR confirmation of the PA TARV field strain

The PA TARV field strain was freshly propagated in LMH cell cultures for viral RNA extraction in this study, and the extracted RNA was confirmed positive by the S1-based RT-PCR using P1/P4 primers to amplify 1088bp of the S1 gene sequence (Kant et al., 2003).

3.2 Summary of NGS data of PA TARV field strain genome

From the total RNA sample of the PA TRAV filed strain, a total of 1,686,331 reads of length 35-151 nt- following trimming were obtained resulting in 524 Mb of fastq format sequence data. Sequence reads that passed default quality control (QC) filters on the MiSeq platform were aligned to reference sequences of chicken genomic DNA and rRNA databases followed by quality trimming to remove contamination reads and excluding reads with similarities to chicken mRNA or rRNA sequences. As a result, 310,284 reads (18.4%) were identified to be the chicken mRNA source, 1,305,052 reads (77.4%) to be the chicken rRNA source, and 8432 reads (0.5%) to be sequencing adapter (Fig. 1). The remaining 63,935 reads (3.7%) were considered as clean reads and further analyzed using a BLASTN procedure, which revealed 47,217 reads (2.8%) as no hits and 16,718 reads (0.9%) as orthoreovirus (Fig. 1).

3.3 De novo assembly of the PA TARV field strain genome

All clean reads from the QC step, including no hits reads, were assembled using SPAdes *de novo* assembler. After BLASTN searching, a total of 10 contigs were generated by software

according to the 10 ARV genome segments. Among all the 10 contigs, the whole lengths of each segment were correctly assembled as the single contig (Table 1 and Fig. 2). Similarity comparison in nt sequences of the 10 contigs to published ARV reference strains from Genbank indicated that all 10 contigs had different homologies with other published reference ARV strains. The highest similarities (sharing 95%-98% nt identities) for general contigs identification were seen between the PA TARV field strain and various classic reference strains (Table 1). To obtain the sequencing depth data and identify single nucleotide variations (SNVs) in the 10 segments, the sequencing reads were mapped back to the assembled contigs. The coverage at every base of the segments is shown on the track 6 of the circos plot (Fig. 2). Read coverage was calculated at 117× on average for all segments. To identify SNVs in each genomic segment of the PA TARV field strain, alignment results were screened using CLC Genomic Workbench software. As a result, a total of 13 SNVs were identified in five genomic segments in the L or M class, and 0 in the S class. Each of the 13 SNVs had a read depth of 100× or above (Table 1).

3.4 The complete genome of the PA TARV field strain

To determine the 5' and 3' termini of each segment not included in NGS assembly contigs, we used the RACE technique combined with Sanger sequencing, thus we successfully obtained the full length sequence of the PA TARV field strain. The complete sequences of all 10 segments of the PA TARV field strain were finally obtained when the 5' and 3' termini Sanger sequencing results from each of the 10 segments were all combined. The full-length sequences of all 10 segments of the PA TARV field strain were deposited into GeneBank in November 2014 (Genbank accession numbers: KP173683 to KP173692, Reo/PA/Turkey/22342/13, Table 1). The genome of the PA TARV field strain is 23,496bp in length with an approximately 50% G+C content and 10 dsRNA segments encoding 12 viral proteins. The lengths of the genomic segments range from 1192bp (S4) to 3959bp (L1) and are similar to those TAVR and ARV reference strains. ORF analysis of nucleotide sequences indicated that 9 of the 10 genome segments (except S1) encoded one single ORF. The 1645bp S1 segment was shown to be tricistronic with partially overlapping ORFs encoded p10 at 25–324bp, p17 at 296–748bp and σC at 633–1613bp, respectively. The length of untranslated regions (UTRs) ranges from 12–30bp at the 5' ends, and from 33–98bp at the 3' ends (Table 2). Other segments coding sequences were predicted to encode 9 putative proteins (λA on L1, λB on L2, λC on L3, μA on M1, μB on M2, μNS on M3, σA on S2, σB on S2, and σNS on S3) with sizes ranging from 368 amino acid (aa) to 1294 aa. All of the 10 fragments share a GCUUUU motif in their 5' UTR and a UCAUC motif in their 3' UTR. The 5' UTRs of the PA TARV field strain are the same as the corresponding sequences in ARV reference strains, but different from other species within the genus (Table 3). The last 5 bases of 3' UTRs sequences of all orthoreoviruses were found to be highly conserved with a UCAUC motif. Moreover, the first and last nucleotides of each segment in TARV were complementary (G-C).

Genomic analysis of the PA TARV field strain revealed that the lengths of all 10 genomic segments were almost identical to the cognate segments of previously reported TARVs and standard ARV reference strains. Each genome segment started and ended with short and highly conserved sequences, however, of which the sequence at the 5' end (GCUUUU) was

different from the sequences determined for other orthoreovirus species. Whereas, the sequence at the 3' end (UCAUC) was a typical segment termination motif and highly conserved within Orthoreovirus genus, thus this PA TARV field strain genome segments belonged to a particular viral species with no segment reassortment between ARV and mammalian reoviruses (MRV). Furthermore, similarity searches using nt and aa sequences of the PA TARV field strain indicated that all 10 segments were highly conserved in the homologous proteins of reference strains, and some of the genomic features were also found in the J18 duck strain. All these sequence data indicate that the PA TARV field strain belongs to the ARV species.

3.5 Sequence comparison

The PA TARV field strain nt and aa sequences were compared to their homolog with two TARVs (MN9 and MN10) and three ARV reference strains (S1133, 138 and AVS-B). The nt sequence variations are visible on the track 7, 8, 9, 10, 11 of the circos plot (Fig. 2); and different identity values (Table S2) are seen between these strains and their genome segments. The PA TARV strain was found to be different from the MN9 and MN10 strains at the M2 segment μ B encoding gene because they shared only 75.5% nt identities and 89.7% aa identities; while they shared high sequence identities at the other 9 segments (nt: 91.4–98.7%; aa: 97.1–99.0%) (Table S2). When the PA TARV field strain was compared with the 3 ARV strains (S1133, 138, AVS-B): their aa and nt sequence identities were remarkably low at S1 segment σ C encoding genes (48.3%–59.3%), low at S3 segment σ B encoding genes (70.1%–79.2%) and S4 segment σ NS encoding genes (74.2%–77.8%); and moderate to high at all other segments (>80%) except L3 segment (nt: 72.5%) of the S1133 strain (Table S2). The σ C ORF and the encoded proteins of the one PA and two MN TARV field strains were seen to be highly conserved in length in comparison with ARV strains (Fig. 3), but its length was 5aa longer than the J18 ARV. The predicted sizes of p10 and p17 ORFs of the PA TARV field strain are 99aa and 150aa, 1–3aa and 4–10aa, respectively, which are longer than the corresponding ORF regions of ARV strains.

3.6 Phylogenetic analyses

The evolutionary relationships of the PA TARV field strain with the *Orthoreovirus* genus members, particularly other ARV strains and TARV strains, were determined by phylogenetic analysis. The phylogenetic trees (Fig. 4) were constructed based on nt sequences of 3 L and 3 M genome segments and 4 σ -class genes using the maximum likelihood method with bootstrapping. All the phylogenetic trees illustrated a great divergence between ARV, TARV and MAR strains. For most genomic segments or genes, the PA TARV field strain and the MN TARV strains formed a turkey-origin TARV group that was different from the chicken-origin ARV strains and duck-origin ARV strains. The TARV group showed a closer evolutionary relationship with chicken ARV strains than duck ARV in all segments or genes except M2. In the M2 phylogenetic tree, four TRAV strains including PA TARV field strain divided into group 1 and group 2. The PA TARV field strain and MN12 form TARV group 1, which is closely related to chicken and duck ARV strains but is distinct from the TARV group 2 (formed by MN9 and MN10) with more than 14% nt divergence. Phylogenetic analyses also revealed that the segments or genes encoding

for outer capsid proteins (μ B, σ B and σ C) exhibit noticeably higher divergence than other segments for ARVs as indicated in their sequence comparisons.

3.7 Overall genomic constellation of the TARV

Figure 5 illustrates the sequence similarity values and divergence of individual genes and their encoded proteins in comparison of the PA TARV field strain with 2 TARV and 6 ARV reference strains. The greatest sequence similarity values were seen in 9 of the 10 genome segments of TARV MN9 or MN10 strain, with the exception of M2 segment which showed high diversity. However, the highest identity ($>90\%$ nt) of M2 segment was found between the PA TARV field strain and chicken ARV-1733 strain. Considerable genetic relatedness of the PA TARV field strain and ARV reference strains was found along 9 of the 10 genome segments, with exception for the 3' end of the S1 segment corresponding to the σ C coding region. The concatenated genes revealed that only L1, L2, M1, M3 and S2 genome segments were found to be closely related between PA TARV field strain and the chicken-origin ARV reference strains, while the remaining 5 genome segments appeared to be more divergent, particularly for the L3, S1 and S3 segments. In addition, L3, S1, S3 and S4 segments of the PA TARV field strain and S1133 strain were only low to moderately related to each other, which indicated that the PA TARV field strain was very diverse from the ARV vaccine strain. In comparison with the duck-origin ARV J18, the PA TARV field strain shared very low sequence identities with it through the whole genome, particularly the S1 segments ($<50\%$).

4. Discussion

ARV was first identified as the causative agent to a highly pathogenic poultry disease characterized as arthritis and tenosynovitis (Olson and Kerr, 1966; Olson et al., 1957; Walker et al., 1972). Thereafter, many ARV isolations from various locations in the United States were reported, and ARV infection in poultry were well described and documented (Deshmukh and Pomeroy, 1969; Johnson, 1972; Van der Heide et al., 1974). Extensive ARV infections in PA poultry have been continuously diagnosed in our laboratory from 2011 to the present time. We have made over 300 ARV isolates so far, mostly from broilers and turkeys, some from layers and a small number from other avian species. We will publish our research findings on these ARV isolates soon. Nonetheless, a complete genomic characterization of a PA broiler ARV field strain, by using traditional genome sequencing procedures (Banyai et al., 2011), was published recently (Tang and Lu, 2015).

Until recently, research findings in TARV sequencing studies, mostly based on the L-class and S-class genomic segments, showed a great genetic divergence between TARVs and ARV vaccine strains (Mor et al., 2014a; Mor et al., 2014b). Other research studies indicated that highly pathogenic ARV variants were capable of infecting not only unvaccinated broiler chickens, but also vaccinated broiler breeders (Dandar et al., 2013; Rosenberger et al., 2013; Troxler et al., 2013). In this study, we employed Illumina NGS technology to characterize the complete genome of the PA TARV field strain. By *de novo* assembling the viral genome, the 10 contigs corresponding to the 10 genomic segments of the PA TARV field strain were generated. The assembled contigs covered the most areas of the 10 genomic segments, only

missed the genomic termini. In NGS, the genomic termini are often under-represented in the obtained viral reads, which is likely due to a variety of reasons including the RNA second structure and protein binding site at genomic termini, viral RNA population, reads trimming before assembly, and *de novo* assembly software error. Specifically designed primers in genomic termini for cDNA synthesis have improved the number of termini sequences reads, but have not completely resolved the problem (Marston et al., 2013). When mapping reads back to the consensus sequence, the coverage was observed at 117× on average for all segments, indicating a sufficient redundancy to identify the SNVs in viral sequences (Kuroda et al., 2010; Kwok et al., 2012). In this study, a total of 13 SNV sites of the PA TARV field strain were observed in consensus sequence presenting in the low level population. This approach is also useful to determine high proportion SNV changes within the same virus strain using consensus sequencing, which can be targeted directly by deep sequencing the relevant genomic regions (Marston et al., 2013).

Sequence comparison of all S-class encoding genes between the PA TARV field strain and MN9 or MN10 TARVs revealed that they shared very high nt (>93.4%) and aa (>97.2%) sequence identities. Whereas, when compared with ARV reference strains of S-class encoding genes, the PA TARV field strain shared very low sequence identities with them in σ C gene (nt: <54.2%; aa: <59.3%) and low in σ B and σ NS genes (nt: <70.1%; aa: <78.4%). In σ A gene, the shared sequence identities of the PA TARV field strain with the 138 strain (nt: 92%; aa: 97.6%) were higher than with other ARV strains (nt: 88.9%; aa: 97.3%), which indicated that ARV 138 strain was more closely related to the PA TARV field strain than other ARVs. The sequence divergences of the σ B and σ C encoding genes were significantly higher than those of other genes between TARVs and ARVs, which should be reasonable true because the σ B and σ C proteins are components of the ARV virion outer capsid (Benavente and Martinez-Costas, 2007). Particularly, σ C gene displayed the feature of highest level of sequence divergence and rapid evolution; therefore, the σ C gene could be used as a genetic marker for rapid differentiation between ARV and TARV.

The highest identity rates were found in L2 segment coding λ B protein between the PA TARV field strain and ARV reference strains, which is likely because the λ B is the highly conserved viral RNA-dependent RNA polymerase (RdRp) protein (Xu and Coombs, 2008). Sequence comparisons among the M-class genes demonstrated that the sequence divergences of the M2 segment and the μ B protein were higher than those of the other M-class genes and their encoded proteins. Sequence identities shared between the PA TARV field strain and MN TARVs were low in μ B encoding gene (nt: 75.5%; aa: 89.7%) which is not unexpected because the position of the μ B, like σ B and σ C, in the outer capsid probably results in more immune selection on the μ B than μ A and μ NS. Selection pressures might restrict variability within the μ B, σ B and σ C genes to be included in the requirement to form a stable interaction with each other. M2 sequences of additional 17 strains (14 turkey-origin and 3 chicken-origin) including MN6 (KJ874301) and MN12 (KJ874294) retrieved from GeneBank were used for nt similarity comparison. Our blast results showed that they all shared high similarities (nt>91%) at M2 segment with the PA TARV field strain in this study. Sequence comparisons of other segments with these reference strains may reveal differences between them; especially full genome sequence comparisons between TARVs

should be conducted. However, such comparisons are not available currently due to the lack of all 10 segments or full genome sequences of other TARVs.

Phylogenetic analysis of individual segments and genes revealed various patterns of clustering with reference strains, and such relationship was supported by high bootstrap values. In general, the nt and aa sequence identities should be greater than 75% and 85%, respectively, between homologous orthoreovirus genes or within a same species group; whereas, if the nt and aa identities are less than 60% and 65%, respectively, between distinct species groups (Urbano and Urbano, 1994). For most genomic segments, the PA TARV field strain clustered together with MN TARV strains. However, in M2 segment phylogenetic analysis, 4 TARVs were evolved into two genetically heterogeneous groups, revealing the existence of at least two predominant TARV genotypes. The phylogeny of sigma-class genes showed that the TARV strains formed a different cluster from ARV reference strains, whereas the TARV strains were only distantly related to the 091, J18 and D20-99 waterfowl strains. In σ C gene, this PA TARV and 3 MN TARVs formed a separated cluster, which was different from most chicken-origin and duck-origin ARVs. Although the σ C gene phylogenetic relationships between TARVs and the ARV GEL13a98M strain (Kant et al., 2003) were close, the host species were different, the TARV was a turkey-origin, the GEL13a98M strain was a chicken origin, cross infections each other of the two species remained unclear. These findings suggested that reassortment between the ARVs from malabsorption syndrome and ARVs from viral arthritis could have played a role in the origin of the TARVs emerged previously in the USA and re-emerging recently, which agreed with a number of published research studies on ARV genes (Day et al., 2007; Hsu et al., 2005; Su et al., 2006). Collectively, all 10 segment-based phylogenies permitted insight into the origin of the unique genetic configuration of the PA TARV field strain in this study. This PA TARV field strain implies a greater genetic diversity than documented ARV reference strains, thus it is a novel or newly emerging TARV strain, similar as MN9 or MN10 TARVs.

Our research findings revealed that the complete genome of the PA TARV field strain exhibited distinct molecular features when compared with chicken-origin ARV and duck-origin ARV strains, and its M2 segment was remarkably different from the TARV MN9 or MN10 strains although they shared high similarities in other segments. These interesting results provide scientific insights into the evolutionary relationship between species origins and provide important genetic information to better understand stabilities or variations of TARV field strains. In conclusion, we determined the whole virus genome sequence and completed full genomic sequence characterization on this PA TARV field strain, which exhibits a number of different properties and pathogenicity compared with the members in *Orthoreovirus* genus. This is the first report that the TARV full genomic features were determined from a single virus strain.

Our research findings have demonstrated that this new PA TARV field strain shares high nt and aa sequences similarity and genomic features with MN9 or MN10 TARV strains. The topological heterogeneity observed among the phylogenetic trees and the sequence pairwise comparison results have suggested that genetic reassortment of the L, M and S segments might have occurred between re-emerging TARV strain and ARV strains. Therefore, full genomic analysis of the PA TARV field strain revealed that there were distinct molecular

sequence characteristics in comparison with standard ARV vaccine strains. The research findings provide full genomic data for better understanding the evolutionary genomic changes of a newly emerging TARV field strains or variants.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This research study was funded by the Pennsylvania Department of Agriculture, Animal Health and Diagnostic Commission's 2014 research project, PDA Contract ME No. 44134401; The National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, Grant No. UL1 TR000127.

Reference

- Afaleq AA, Jones RC. Pathogenicity of three turkey and three chicken reoviruses for poults and chicks with particular reference to arthritis/tenosynovitis. *Avian pathology : journal of the W.V.P.A.* 1989; 18:433–440. [PubMed: 18679874]
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology.* 2012; 19:455–477. [PubMed: 22506599]
- Banyai K, Dandar E, Dorsey KM, Mato T, Palya V. The genomic constellation of a novel avian orthoreovirus strain associated with runting-stunting syndrome in broilers. *Virus genes.* 2011; 42:82–89. [PubMed: 21116842]
- Baroni A, Bertoncin P, D'Aprile PN, Felluga B. Ultrastructural histopathology of chick embryo chorioallantoic membrane infected with an avian reovirus. *Avian Pathol.* 1980; 9:341–354. [PubMed: 18770272]
- Benavente J, Martinez-Costas J. Avian reovirus: structure and biology. *Virus research.* 2007; 123:105–119. [PubMed: 17018239]
- Bezerra DA, da Silva RR, Kaiano JH, Silvestre RV, de Souza Oliveira D, Linhares AC, Gabbay YB, Mascarenhas JD. Detection of avian group D rotavirus using the polymerase chain reaction for the VP6 gene. *J Virol Methods.* 2012; 185:189–192. [PubMed: 22820073]
- Dandar E, Balint A, Kecskemeti S, Szentpali-Gavaller K, Kisfali P, Meleg B, Farkas SL, Banyai K. Detection and characterization of a divergent avian reovirus strain from a broiler chicken with central nervous system disease. *Archives of virology.* 2013; 158:2583–2588. [PubMed: 23771766]
- Day JM, Pantin-Jackwood MJ, Spackman E. Sequence and phylogenetic analysis of the S1 genome segment of turkey-origin reoviruses. *Virus genes.* 2007; 35:235–242. [PubMed: 17265142]
- De Gussem J, Swam H, Lievens K, De Herdt P. Reovirus tenosynovitis in a flock of layer breeders. *Avian pathology : journal of the W.V.P.A.* 2010; 39:169–170. [PubMed: 20544421]
- Deshmukh DR, Pomeroy BS. Avian reoviruses II. Physicochemical characterization and classification. *Avian diseases.* 1969; 13:243–251. [PubMed: 4307738]
- Giangaspero M, Vanopdenbosch E, Nishikawa H, Tabbaa D. Prevalence of antibodies against respiratory viruses (parainfluenza virus type 3, respiratory syncytial virus, reovirus and adenovirus) in relation to productivity in Syrian Awassi sheep. *Trop Anim Health Prod.* 1997; 29:83–91. [PubMed: 9203308]
- Guy JS, Levy MG, Ley DH, Barnes HJ, Gerig TM. Experimental reproduction of enteritis in bobwhite quail (*Colinus virginianus*) with *Cryptosporidium* and reovirus. *Avian Dis.* 1987; 31:713–722. [PubMed: 2831866]
- Harris RS. Improved pairwise alignment of genomic DNA. ProQuest. 2007

- Heggen-Peay CL, Cheema MA, Ali RA, Schat KA, Qureshi MA. Interactions of poult enteritis and mortality syndrome-associated reovirus with various cell types in vitro. *Poultry science*. 2002; 81:1661–1667.
- Howell SH, Walker LL. Synthesis of DNA in toluene-treated *Chlamydomonas reinhardtii* (DNA replication-chloroplast DNA-cell cycle-electron microscopy). *Proceedings of the National Academy of Sciences of the United States of America*. 1972; 69:490–494. [PubMed: 4621835]
- Hsu HW, Su HY, Huang PH, Lee BL, Liu HJ. Sequence and phylogenetic analysis of P10- and P17-encoding genes of avian reovirus. *Avian diseases*. 2005; 49:36–42. [PubMed: 15839410]
- Ide PR, Dewitt W. Serological incidence of avian reovirus infection in broiler-breeders and progeny in Nova Scotia. *The Canadian veterinary journal. La revue veterinaire canadienne*. 1979; 20:348–353. [PubMed: 231481]
- Johnson DC. Diagnosis, pathology, and etiology of tenosynovitis in broilers and broiler breeders. *Avian diseases*. 1972; 16:1067–1072. [PubMed: 4344861]
- Kant A, Balk F, Born L, van Roozelaar D, Heijmans J, Gielkens A, ter Huurne A. Classification of Dutch and German avian reoviruses by sequencing the sigma C protein. *Veterinary research*. 2003; 34:203–212. [PubMed: 12657212]
- Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*. 2012; 28:3211–3217. [PubMed: 23071270]
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome research*. 2009; 19:1639–1645. [PubMed: 19541911]
- Kuroda M, Katano H, Nakajima N, Tobiume M, Ainai A, Sekizuka T, Hasegawa H, Tashiro M, Sasaki Y, Arakawa Y, Hata S, Watanabe M, Sata T. Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. *PloS one*. 2010; 5:e10256. [PubMed: 20428231]
- Kwok H, Tong AH, Lin CH, Lok S, Farrell PJ, Kwong DL, Chiang AK. Genomic sequencing and comparative analysis of Epstein-Barr virus genome isolated from primary nasopharyngeal carcinoma biopsy. *PloS one*. 2012; 7:e36939. [PubMed: 22590638]
- Levisohn S, Gur-Lavie A, Weisman J. Infectious synovitis in turkeys: Isolation of tenosynovitis virus-like agent. *Avian pathology : journal of the W.V.P.A.* 1980; 9:1–4. [PubMed: 18770233]
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010; 26:589–595. [PubMed: 20080505]
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. [PubMed: 19505943]
- Magee DL, Montgomery RD, Maslin WR, Wu CC, Jack SW. Reovirus associated with excessive mortality in young bobwhite quail. *Avian Dis*. 1993; 37:1130–1135. [PubMed: 8141744]
- Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, David D, de Lamballerie X, Fooks AR. Next generation sequencing of viral RNA genomes. *BMC genomics*. 2013; 14:444. [PubMed: 23822119]
- Mor SK, Sharafeldin TA, Porter RE, Goyal SM. Molecular characterization of L class genome segments of a newly isolated turkey arthritis reovirus. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2014a; 27:193–201.
- Mor SK, Sharafeldin TA, Porter RE, Ziegler A, Patnayak DP, Goyal SM. Isolation and characterization of a turkey arthritis reovirus. *Avian diseases*. 2013; 57:97–103. [PubMed: 23678736]
- Mor SK, Verma H, Sharafeldin TA, Porter RE, Jindal N, Ziegler A, Goyal SM. Characterization of S class gene segments of a newly isolated turkey arthritis reovirus. *Virology*. 2014b; 464-465:33–44. [PubMed: 25043587]
- Nick H, Cursiefen D, Becht H. Structural and growth characteristics of two avian reoviruses. *Arch Virol*. 1975; 48:261–269. [PubMed: 170888]
- Olson NO, Kerr KM. Some characteristics of an avian arthritis viral agent. *Avian diseases*. 1966; 10:470–476. [PubMed: 4961651]

- Olson NO, Shelton DC, Munro DA. Infectious synovitis control by medication; effect of strain differences and pleuropneumonia-like organisms. *American journal of veterinary research*. 1957; 18:735–739. [PubMed: 13470226]
- Palya V, Glavits R, Dobos-Kovacs M, Ivanics E, Nagy E, Banyai K, Reuter G, Szucs G, Dan A, Benko M. Reovirus identified as cause of disease in young geese. *Avian pathology : journal of the W.V.P.A.* 2003; 32:129–138. [PubMed: 12745366]
- Petek M, D'Aprile PN, Cancellotti F. Biological and physico-chemical properties of the infectious bursal disease virus (IBDV). *Avian Pathol.* 1973; 2:135–152. [PubMed: 18777392]
- Rey ME, D'Andrea E, Calvert-Evers J, Paximadis M, Boccardo G. Evidence for a phytoreovirus associated with tobacco exhibiting leaf curl symptoms in South Africa. *Phytopathology*. 1999; 89:303–307. [PubMed: 18944775]
- Ritter GD, Ley DH, Levy M, Guy J, Barnes HJ. Intestinal cryptosporidiosis and reovirus isolation from bobwhite quail (*Colinus virginianus*) with enteritis. *Avian Dis.* 1986; 30:603–608. [PubMed: 3021104]
- Rosenberger J, Rosenberger S, Markis M, Rosenberger J. Pathogenicity and control of recent (2011–2012) reovirus isolates from broiler and turkey flocks presenting with viral arthritis and tenosynovitis. *The*. 2013
- Spandidos DA, Graham AF. Physical and chemical characterization of an avian reovirus. *J Virol.* 1976; 19:968–976. [PubMed: 987252]
- Sterner FJ, Rosenberger JK, Margolin A, Ruff MD. In vitro and in vivo characterization of avian reoviruses. II. Clinical evaluation of chickens infected with two avian reovirus pathotypes. *Avian diseases*. 1989; 33:545–554. [PubMed: 2549942]
- Su YP, Su BS, Shien JH, Liu HJ, Lee LH. The sequence and phylogenetic analysis of avian reovirus genome segments M1, M2, and M3 encoding the minor core protein muA, the major outer capsid protein muB, and the nonstructural protein muNS. *Journal of virological methods*. 2006; 133:146–157. [PubMed: 16337282]
- Taber R, Alexander V, Whitford W. Persistent reovirus infection of CHO cells resulting in virus resistance. *J Virol.* 1976; 17:513–524. [PubMed: 176428]
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*. 2011; 28:2731–2739. [PubMed: 21546353]
- Tang Y, Lu H. Genomic characterization of a broiler reovirus field strain detected in Pennsylvania. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2015; 31C:177–182.
- Troxler S, Rigomier P, Bilic I, Liebhart D, Prokofieva I, Robineau B, Hess M. Identification of a new reovirus causing substantial losses in broiler production in France, despite routine vaccination of breeders. *The Veterinary record*. 2013; 172:556. [PubMed: 23636701]
- Urbano P, Urbano FG. The Reoviridae family. *Comparative immunology, microbiology and infectious diseases*. 1994; 17:151–161.
- van der Heide L. The history of avian reovirus. *Avian diseases*. 2000; 44:638–641. [PubMed: 11007012]
- Van der Heide L, Geissler J, Bryant ES. Infectious tenosynovitis: serologic and histopathologic response after experimental infection with a Connecticut isolate. *Avian diseases*. 1974; 18:289–296. [PubMed: 4368815]
- Varela R, Benavente J. Protein coding assignment of avian reovirus strain S1133. *J Virol.* 1994; 68:6775–6777. [PubMed: 8084013]
- Varela R, Martinez-Costas J, Mallo M, Benavente J. Intracellular posttranslational modifications of S1133 avian reovirus proteins. *Journal of virology*. 1996; 70:2974–2981. [PubMed: 8627773]
- Vindevogel H, Meulemans G, Pastoret PP, Schwes A, Calberg-Bacq CM. Reovirus infection in the pigeon. *Ann Rech Vet.* 1982; 13:149–152. [PubMed: 6301348]
- Walker ER, Friedman MH, Olson NO. Electron microscopic study of an avian reovirus that causes arthritis. *Journal of ultrastructure research*. 1972; 41:67–79. [PubMed: 4627607]
- Watson R, Anthony F, Pickett M, Lambden P, Masson GM, Thomas EJ. Reverse transcription with nested polymerase chain reaction shows expression of basic fibroblast growth factor transcripts in

- human granulosa and cumulus cells from in vitro fertilisation patients. Biochemical and biophysical research communications. 1992; 187:1227–1231. [PubMed: 1417798]
- Xu W, Coombs KM. Avian reovirus L2 genome segment sequences and predicted structure/function of the encoded RNA-dependent RNA polymerase protein. Virology journal. 2008; 5:153. [PubMed: 19091125]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

1. The Illumina Next-Generation Sequencing (NGS) was used to conduct full genome deep sequencing for a turkey arthritis reovirus (TARV) field strain.
2. This is the first report that the TARV full genomic features were determined from a single virus strain.
3. A total of 13 single nucleotide variation (SNV) sites were determined in the L and M genome segments of the TARV strain; 0 in the S segment.
4. The predicted sizes of p10 and p17 ORFs of the TARV strain were longer than those of reference strains.

Fig. 1.

A flow chat of genome sequencing steps for the turkey arthritis reovirus (TARV) field strain (Reo/PA/Turkey/22342/13) detected in Pennsylvania (PA) of the USA. The left is data analysis pipeline for viral genome assembly; The right is a pie chart of the homology search results for NGS reads.

Fig. 2.

Circular representation of full genome of the PA TARV field strain (Reo/PA/Turkey/22342/13). Track 1: Consensus sequence; Track 2: Open reading frames (ORFs); Track 3: σ C gene sanger sequencing result; Track 4: Sequence variations between NGS and Sanger sequencing result in σ C gene; Track 5: Assembled contigs by *de novo* assembly (SPAdes); Track 6: Sequencing depth of NGS, the axis of coverage track corresponds to 0,110,220,330,440,550 reads from inside to outside; Track 7-11: Sequence variations of the PA TARV (Reo/PA/Turkey/22342/13) in comparison with 2 MN TRAVs of MN9, MN10, and 3 ARV reference strains of 138, S1133 and AVS-B, respectively.

Fig 3.

Comparisons of ORF organizations of S1 segments between the PA TARV field strain (Reo/PA/Turkey/22342/13) and 5 ARV reference strains. Yellow color arrow = p10 ORF, blue color arrow = p17 ORF, pink color arrow = σ C ORF. The beginning and ending nucleotide numbers were labeled on each arrow. (Note: S1 sequence comparison was not done with other TARVs due to the lack of complete S1 segment sequences)

Fig. 4. Phylogenetic trees constructed by avian reovirus (ARV) and mammalian reovirus (MRV) based on nucleotide sequences of the 10 (3L, 3M and 4 S) homologous genome segments of orthoreovirus species. Blue color= chicken-origin ARV, red color= turkey-origin ARV or TARV, green color=duck-origin ARV, and purple color=MRV; the PA TARV field strain (Reo/PA/Turkey/22342/13) marked in red dot.

Fig. 5.

The mVISTA method for whole genome nucleotide alignment; this figure illustrate alignment results of the PA TARV field strain (Reo/PA/Turkey/22342/13) in comparisons with 2 MN TARVs (MN9, MN10) and 6 ARV reference strains (138, S1133, AVS-B, 1733, 176 and J18) retrieved from GenBank (Table S1); Areas in pink color represent 90% similarities; and areas in white represent < 90% similarities. The scale bar measures approximate length of the concatenated genome.

Table 1

De novo assembly contigs of the Pennsylvania turkey arthritis reovirus (TARV) field strain (Reo/PA/Turkey/22342/13)

Segment GenBank Accession No.	Contig length (bp)	Highest similarity to TARV reference strain in GenBank	Identities (%)	SNVs	Reads	Contig location
L1	3882	Turkey/USA/MN/2011/TERV-MN6 L1 λA gene (KJ865912)	96	3	2919	22-3903
L2	3805	Turkey/USA/MN/2011/TARV-MN2 L2 λB gene (KJ865893)	92	2	3118	14-3819
L3	2883	Turkey/USA/MI/2013/TARV-MN12 L3 λC gene (KJ865888)	98	1	2755	16-3898
M1	2199	Turkey/USA/MN/2014/TARV-MN13 μA gene (KJ874318)	95	5	1897	13-2211
M2	2158	Turkey/USA/MI/2013/TARV-MN12 μB gene (KJ874294)	99	0	882	12-2150
M3	1909	Turkey/USA/MN/2011/TERV-MN3 μNS gene (KJ874275)	95	2	1943	25-1932
S1	1597	TARV MN3 S1 p10, p17, σC genes (KF872234)	99	0	1003	15-1611
S2	1308	TARV MN10 σA gene, partial (KF872254)	95	0	466	10-1317
S3	1172	TK/MN/USA/D-AB15/2011 S3 σB gene (KF183931)	97	0	1075	18-1189
S4	1169	turkey/MO/SEP-828/05 S4 σNS gene (EU400282.1)	97	0	662	13-1181

Table 2

General genome features of the Pennsylvania turkey arthritis reovirus (TARV) field strain (Reo/PA/Turkey/22342/13), all 10 genome segments and GenBank accession numbers

Genome segment	Size	Length (nt) of the			Sequence at the termini		Protein size(aa)	Encoded protein
		5' end	ORF	3' end	5' end	3' end		
L1 (KP173683)	3959	21	3882	56	GCUUUU	UCAUC	1294	λA(core shell)
L2 (KP173684)	3907	12	3858	37	GCUUUU	UCAUC	1286	λB(core RdRp)
L3 (KP173685)	3829	14	3780	36	GCUUUU	UCAUC	1260	λC(core turret)
M1 (KP173686)	2283	12	2199	72	GCUUUU	UCAUC	733	μA(core NTPase)
M2 (KP173687)	2158	29	2031	98	GCUUUU	UCAUC	677	μB(outer shell)
M3 (KP173688)	1996	24	1908	64	GCUUUU	UCAUC	636	μNS(NS factory)
S1 (KP173689)	1645	24	300	33	GCUUUU	UCAUC	100	p10(NS FAST)
			453				151	p17(NS other)
			981				327	σC(outer fiber)
S2 (KP173690)	1324	15	1251	58	GCUUUU	UCAUC	417	σA(core clamp)
S3 (KP173691)	1203	30	1104	68	GCUUUU	UCAUC	368	σB(outer clamp)
S4 (KP173692)	1192	23	1104	65	GCUUUU	UCAUC	368	σNS(NS RNAb)

Table 3

Comparison of segment 5' and 3' non-coding regions of the Pennsylvania turkey arthritis reovirus (TARV) field strain (Reo/PA/Turkey/22342/13) among members of Orthoreovirus genus

Orthoreovirus species	Host	Terminal region sequences (5' to 3')
Reo/PA/Turkey/22342/13	Turkey	GCUUUUU...UAUUCAUC
Avian orthoreovirus (ARV)-S1133	Broiler Chicken	GCUUUUU...UAUUCAUC
Avian orthoreovirus (ARV)- J18	Muscovy Duck	GCUUUUU...UA ^U / _C UCAUC
Broome virus (BroV)	Little Red Flying Fox	GUCAA...UCAUC
Baboon orthoreovirus (BRV)	Yellow Baboon	GUAAA...UCAUC
Nelson bay virus (NBV)	Grey-headed Flying Fox	GCUUUA...UCAUC
Pulau virus (PuV)	Fruit Bat	GCUUUA...UCAUC
Melaka virus (MELV)	Human	GCUUUA...UCAUC
Mammalian orthoreovirus (MRV)-1	Mink	GCUA...UCAUC
Mammalian orthoreovirus (MRV)-2	Human	GCUA...UCAUC
Mammalian orthoreovirus (MRV)-3	Masked Civet Cats	GCUA...UCAUC