Members of the *Astroviridae* are frequently associated with clinical diarrhea in the young of both mammalian and avian hosts (2, 9, 15). In spite of their worldwide distribution and endemic nature, our understanding of their evolution is limited. The majority of previous studies of astrovirus phylogeny have examined relationships among clinical isolates based on diagnostic reverse transcription-PCR amplicons (5, 19). While sequences from short diagnostic amplicons have been successful in assigning isolates to groups, they lack the power to accurately resolve phylogenetic relationships (14). Studies which have attempted to reconstruct the astrovirus phylogeny using genome alignments have done so across astrovirus genera (26). The sequence divergence and differences in codon usage across the *Astroviridae* may confound conclusions about phylogenetic relationships and selection pressure. The current study describes the phylogenetic analysis of multiple genomic sequences of closely related turkey astrovirus (TAstV) clinical isolates collected from commercial turkey flocks across the United States. Representative isolates from each location were randomly selected, and the full-length genomic sequences were determined as previously described (8). Chromatogram data were analyzed using phred, phrap, and consed software (3, 4); sequences were aligned using ClustalW (25) and were edited using GeneDoc (18). Amino acid- and nucleotide-based estimates of phylogeny were generated by using both MrBayes (6) and the hypothesis testing using phylogenies (HYPPH) package (10).

**Evidence of distinct subtypes.** The relationships of the novel clinical TAstV isolates (GenBank accession numbers EU143843 to EU143851) within the *Astroviridae* were first assessed by using predicted capsid amino acid sequences. The topology of this tree (Fig. 1A) was consistent with previous studies demonstrating two major clades containing the genera *Mamastrovirus* and *Avastrovirus* and minor clades corresponding to their host species (29). All of the clinical isolates clustered with TAstV-2/NC/99, TAstV1987, and TAstV2001 (TAstV-2-like) (Fig. 1A). The TAstV-1 capsid sequence was found in a clade with avian nephritis virus, with the distance between the reference TAstV-1 (accession no. CAB95007) and TAstV-2 (TAstV-2/NC/99; accession no. AAF18464) sequences comparable to the distance between human astroviruses (HAstVs) and other mamastroviruses. The sequence analysis of TAstV-1 and TAstV-2 diagnostic amplicons, previously described by Pantin-Jackwood et al. (19) and Cattoli et al. (1), demonstrated that the levels of variation among TAstV-1-like isolates and among TAstV-2-like isolates are comparable to the level of diversity among HAstVs. The phylogenetic analysis of the full-length capsid genes of all TAstV viruses (Fig. 1A) suggests that TAstV-1-like and TAstV-2-like viruses may have originated from separate introductions into the turkey species and that there are at least two TAstV lineages which should be regarded as distinct subtypes instead of serotypes. Within each subtype, there appears the potential for distinct serotypes to exist, as TAstV2001 and TAstV1987 have been reported to represent distinct serotypes (24) and share only 73% nucleotide sequence identity (23). This level of sequence conservation is similar to that of HAstV capsid genes from different serotypes (<80% nucleotide similarity; unpublished observation). These sequence differences suggest that MN/01 may represent a serotype that is distinct from that of TAstV-2/NC/99; however, experimental examination of the serological cross-reactivity of MN/01 with other viruses is needed. Collectively, these findings suggest that the ecology of *Avastrovirus* species may be more complicated than currently appreciated. Interestingly, Lukashov et al. (14) described the phylogenetic evidence of at least two cross-species transmissions within the genus *Mamastrovirus*. This leads one to question if other, as-yet- unidentified astrovirus subtypes exist within mammalian populations.

**Genomic analysis is required to understand phylogeny.** To develop a more-accurate reconstruction of the relationships among the TAstV-2-like viruses, phylogenies were constructed using genomic, open reading frame 1a (ORF1a), ORF1b, and ORF2 sequences from the TAstV-2-like clinical isolates and
the TAstV-2/NC/99 reference sequence (Fig. 1B to E). The analysis of the TAstV-2-like clinical isolates demonstrates variation in phylogenetic relationships across the different ORFs in comparison to the full genome (Fig. 1B to E). While generating this level of data on a routine basis is impractical for diagnostic purposes, it is important to recognize that the region of viral genome analyzed can affect the interpretation of phylogenetic relationships. The initial characterization of a virus based on its capsid sequence is effective for establishing its genus and species; however, to understand the evolutionary history of an isolate during an outbreak, sufficient sequence coverage should be included to ensure the most-accurate relationship possible. This is highlighted by the observation that MI/00 clustered with PA/01 and VA/99 in ORF1a and ORF1b trees (Fig. 1C to D) but was found with CO/01 in the capsid phylogeny tree (Fig. 1E).

**Phylogenetic evidence of recombination across the astrovirus genome.** The Sawyer test for recombination (21) was performed to further analyze the potential recombination event between MI/00 and CO/01, and a breakpoint was identified (P < 0.0001) at nucleotide position 4861. To determine if this region was the only region associated with recombination, the analysis was expanded using GENECONV (21) to test all pairwise comparisons of the entire isolate genomes. Forty-six total recombination events were identified, with at least one recombination event identified in each of the 10 TAstV-2-like isolates (accession numbers EU143843 to EU143851 and AF206663) (Fig. 2A). The distribution of the putative recombination events corresponded with the level of divergence across the three reading frames. ORF1b is the least divergent and had only two putative recombination events. ORF2 is the most divergent and contained the majority of putative recombination events, together with reports by Walter et al. (28) and Pantin-Jackwood et al. (20), suggest that the region around the ORF1b-ORF2 junction is a potential recombination hot spot.

**FIG. 1.** Phylogenetic relationship of TAstV-2-like viruses. (A) A Bayesian phylogenetic tree describing evolutionary relationships among predicted amino acid sequences for human (H AstV), pig (P AstV), sheep (O AstV), mink (M AstV), avian (ANV), and turkey (T AstV) astrovirus capsids. (B to E) Bayesian phylogenetic trees describing the nucleotide relationships among the newly reported TAstV-2-like isolates and the TAstV-2/NC/99 reference sequence. Alignments were constructed using whole genomes (B), ORF1a (C), ORF1b (D), and ORF2 (E). Branch lengths represent the expected number of amino acid substitutions (A) and the expected number of nucleotide substitutions per site (B to E). Posterior support values are shown only for nodes with values less than 0.95. Estimates of phylogeny were made by using the MrBayes program (6).
combination events (Fig. 2A). The evidence of recombination was also assessed using TOPALi (16) to analyze the TAstV-2-like multiple alignment. TOPALi analysis demonstrated similar evidence for recombination, with the strongest support near the junctions between ORFs (Fig. 2B), suggesting an association between recombination and transcriptional signal sequences (7, 13). Furthermore, the finding that at least one putative recombination event was detected in every isolate suggests that recombination may play a key role in astrovirus sequence diversity.

Role of selection in astrovirus capsid evolution. Sequence diversity in astroviruses may also involve host selection pressures. These pressures would presumably explain the existence of distinct serotypes. To address this, TAstV-2-like astroviruses were analyzed for selection using two alignments. The first alignment contained TAstV-2-like capsid sequences (accession numbers EU143843 to EU143851 and AF206663). The second alignment contained TAstV-2-like sequences that were found to have >80% nucleotide similarity to TAstV-2/NC/99, excluding sequences which may belong to different serotypes (MN/01) based on the sequence distances observed (Fig. 1A). For comparison, two HAstV alignments were also analyzed. The first included at least one sequence from each serotype (HAstV), and the second alignment included eight HAstV-4 capsid sequences (HAstV-4). The phylogenies for the four alignments were constructed using HYPHY and analyzed by genetic algorithm for recombination detection (GARD) (11). Breakpoints were recorded in the data input file, and tests for selection were performed using the fixed-effects likelihood (FEL), internal fixed-effects likelihood (IFEL), and random-effects likelihood (REL) models (10, 11) and the partitioning approach for robust inference of selection (PARRIS) (22) methods implemented at http://www.datamonkey.org. Each method uses likelihood-based analysis to identify sites where the rate of nonsynonymous substitution is greater than the rate of synonymous substitution.

The FEL and IFEL methods identified a small number of positively selected sites in the TAstV-2 alignments, while REL and PARRIS found no evidence of positive selection (Tables 1 and 2). More sites were identified as positively selected with IFEL than FEL, suggesting that selective pressure is occurring primarily at the population level (internal branches). Results from the analysis of the HAstV were similar to those for TAstV-2, except that REL identified positively selected sites in the multiple-serotype HAstV data set (Table 1). Sites with
The results from these analyses do not point to one mechanism as the primary means of achieving sequence diversity in astroviruses; instead, they suggest that astroviruses employ all sequence-changing mechanisms available to positive-sense single-stranded RNA viruses and underscore the need for models which allow for all of these factors to be analyzed together.

**Accession numbers.** Accession numbers for capsid amino acid sequences are as follows: CAB95007 (TAstV-1), AAAS31787 (TAstV2001), AAVS31786 (TAstV2017), AAF18464 (TAstV-2/NC99), BAA92849 (ANV1), BAB21617 (ANV2), NP_059946 (ovine AstV), NP_795336 (mink AstV), CAB95000 (porcine AstV), and AAC13556 (feline AstV). Accession numbers for nucleotide sequences are as follows: AF206663 (TAstV2017/NC99), AY769615 (TAstV1987), AY769616 (TAstV2001), AY720892 (HAstV-1), AY720892 (HAstV-1 ORF2), L23513 (HAstV1 ORF2), L13745 (HAstV2), L06802 (HAstV2 ORF2), AF290724 (HAstV-3 ORF1b), AF117209 (HAstV-3 ORF2), AY720891 (HAstV-4), AB025801 (HAstV-4 ORF2), AB025804 (HAstV-4 ORF2), EU244027 (HAstV-4 ORF2), EU078582 (HAstV-4 ORF2), Z33883 (HAstV-4 ORF2), DQ208633 (HAstV-5), AB037273 (HAstV-5 ORF2), AB037274 (HAstV-5 ORF2), U15136 (HAstV-5 ORF2), AB292077 (HAstV-6 ORF1b), Z46658 (HAstV-6 ORF2), AF248738 (HAstV-7 ORF1b), Y08632 (HAstV-7 ORF2), AF248738 (HAstV-7 ORF2), AF260508 (HAstV-8), Z66541 (HAstV-8 ORF2), ABO131031 (HAstV Katano23-6), ABO131030 (HAstV Katano24), AFO131031 (HAstV ORF2, unclassified serotype), and ABO131031 (HAstV ORF2, unclassified serotype). Accession numbers for novel isolates are as follows: EU143843 (TAstV/AK/98), EU143844 (TAstV/CA/00), EU143845 (TAstV/CO/01), EU143846 (TAstV/MI/01), EU143847 (TAstV/MN/01), EU143848 (TAstV/VA/01), EU143850 (TAstV/TX/00), and EU143851 (TAstV/VA/99).

We thank Chris Ashwell, Shelly Nolin, and Audrey O’Nan of the Domestic Animal Genomics Laboratory at North Carolina State University for their help with sequencing and Sergei L. Kosakovsky Pond, Barb Sherry, and Liz Turpin for their many helpful discussions and critical reviews of the manuscript.

This work was supported in part by P30 DK34987 and the North Carolina Agricultural Foundation.

**REFERENCES**


---

**TABLE 1. Sites (codons) selected by different methods**

<table>
<thead>
<tr>
<th>Isolate</th>
<th>No. of aa aligned</th>
<th>Positive selection</th>
<th>Purifying selection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FEL</td>
<td>IFEL</td>
</tr>
<tr>
<td>HAstV</td>
<td>727</td>
<td>5 (1)</td>
<td>6 (3)</td>
</tr>
<tr>
<td>TAstV-4</td>
<td>727</td>
<td>1 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>TAstV-2-like</td>
<td>771</td>
<td>1 (0)</td>
<td>9 (5)</td>
</tr>
<tr>
<td>TAstV-2</td>
<td>804</td>
<td>4 (1)</td>
<td>6 (3)</td>
</tr>
</tbody>
</table>

1. Alignments were done for capsid sequences from each serotype for HAstV, capsid sequences for HAstV-4, capsid sequences in the TAstV-2-like clade (TAstV-2-like), and TAstV-2-like capsid sequences with 80% similarity to TAstV-2/NC99 (TAstV-2).

2. aa, amino acids.

3. Number of significant sites at a P of ≤0.1 (P = 0.05), except where indicated otherwise.

4. ND, not determined.

---

**TABLE 2. Positively selected sites in TAstV-2 isolates**

<table>
<thead>
<tr>
<th>Codon position in TAstV-2 sequence</th>
<th>TAstV-2-like</th>
<th>TAstV-2 (&gt;80%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FEL</td>
<td>IFEL</td>
</tr>
<tr>
<td>12</td>
<td>0.080</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>365</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>387</td>
<td>0.010</td>
<td>0.098</td>
</tr>
<tr>
<td>391</td>
<td>0.076</td>
<td>0.082</td>
</tr>
<tr>
<td>473</td>
<td>0.083</td>
<td>0.081</td>
</tr>
<tr>
<td>477</td>
<td></td>
<td>0.091</td>
</tr>
<tr>
<td>485</td>
<td></td>
<td>0.048</td>
</tr>
<tr>
<td>493</td>
<td></td>
<td></td>
</tr>
<tr>
<td>598</td>
<td>0.060</td>
<td>0.015</td>
</tr>
<tr>
<td>602</td>
<td>0.043</td>
<td>0.072</td>
</tr>
<tr>
<td>636</td>
<td>0.033</td>
<td>0.042</td>
</tr>
</tbody>
</table>

1. Only P values of ≤0.1 are shown.

2. Sequences were found to have a >80% similarity to TAstV-2/NC99.


17. Reference deleted.


