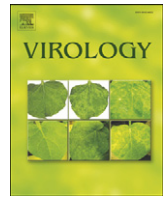




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Determination and analysis of the full-length chicken parvovirus genome

J. Michael Day, Laszlo Zsak\*

Southeast Poultry Research Laboratory, Agricultural Research Service, United States Department of Agriculture, 934 College Station Road, Athens, GA 30605, USA

### ARTICLE INFO

#### Article history:

Received 29 September 2009  
 Returned to author for revision  
 17 November 2009  
 Accepted 17 December 2009  
 Available online 25 January 2010

#### Keywords:

Parvovirus  
 Chicken  
 Turkey  
 Genome  
 Enteric  
 Phylogenetics

### ABSTRACT

Viral enteric disease in poultry is an ongoing problem in many parts of the world. Many enteric viruses have been identified in turkeys and chickens, including avian astroviruses, rotaviruses, reoviruses, and coronaviruses. Through the application of a molecular screening method targeting particle-associated nucleic acid (PAN), we recently described the detection and partial characterization of a novel enteric parvovirus in chickens. Subsequent surveys of intestinal homogenates from turkeys and chickens in the United States revealed widespread occurrence of parvovirus in poultry. Here we report the first full genome sequence of a novel chicken parvovirus, ChPV ABU-P1. ChPV ABU-P1 genome organization, predicted amino acid sequence, and phylogenetic relationships with other described parvoviruses are discussed.

Published by Elsevier Inc.

### Introduction

Enteric disease syndromes are a continual economic burden for the poultry industry throughout the world. The major recognized enteric syndromes in poultry are poult enteritis complex (PEC) and poult enteritis mortality syndrome (PEMS) in turkeys and runting–stunting syndrome (RSS) in broiler chickens (Barnes and Guy, 2003; Barnes et al., 2000; Goodwin et al., 1993). Poultry enteric disease is marked by diarrhea, stunting, and increased time to market; the more severe forms of enteric disease are characterized by immune dysfunction and increased mortality.

Numerous viruses have been detected in or isolated from the intestinal tract of poultry (both diseased and healthy), and have subsequently been implicated in poultry enteric disease. These viruses include the avian reoviruses, rotaviruses, astroviruses and coronaviruses; despite these diagnostic efforts and viral descriptions, the role these viruses play in the enteric syndromes are still poorly understood (Guy, 1998; Pantin-Jackwood et al., 2008a; Pantin-Jackwood et al., 2007a; Pantin-Jackwood et al., 2008b; Pantin-Jackwood et al., 2007b; Reynolds et al., 1987; Spackman et al., 2005).

We recently reported the use of a molecular screening protocol to amplify and identify novel enteric viruses in the intestinal tract of poultry showing characteristic enteric disease signs (Zsak et al., 2008). This random amplification of viral particle-associated nucleic acid identified numerous parvovirus DNA sequences directly from the intestinal homogenates of chickens with signs of enteric disease.

Initial analysis of this chicken-origin parvovirus (ChPV) revealed that it is a novel member of the *Parvovirinae* subfamily within the *Parvoviridae*, and led to the development of a specific molecular diagnostic test targeting the ChPV non-structural (NS) gene (Zsak et al., 2009).

Parvoviruses have been described that infect a wide range of invertebrates and vertebrates, and a thorough knowledge of the parvovirus genome and encoded proteins is essential to an understanding of the evolution of parvoviruses in various hosts, including tissue tropism during viral pathogenesis (Hueffer and Parrish, 2003). Parvoviruses are well known for causing enteric disease in mammals, with the canine parvoviruses recently emerging as important pathogens in the 1970s (Decaro et al., 2009; Truyen, 1999). Parvoviruses have been known for years as the causative agent of Derzsy's disease in young geese, and are also found in Muscovy ducks, where they cause multiple clinical signs ranging from enteritis to hepatitis and infectious myocarditis (Gough, 2008). Parvoviruses have been previously described in chickens based upon their morphology via electron microscopy and upon their genome size (Kisary et al., 1985; Kisary et al., 1984), and enteric disease signs have been attributed to parvovirus-like particles detected in turkey intestinal tracts (Trampel et al., 1983). A recent diagnostic survey of enteric samples collected from United States turkey and chicken flocks revealed widespread presence of parvovirus in U.S. poultry (Zsak et al., 2009). This is the first in-depth molecular characterization and analysis of the full-length genome sequence of ChPV. The analysis includes comparisons to other members of the *Parvovirinae* that infect mammals and birds, including two novel turkey-origin parvoviruses (TuPV) recently sequenced using a similar molecular approach.

\* Corresponding author. Fax: +1 706 546 3161.  
 E-mail address: [Laszlo.Zsak@ars.usda.gov](mailto:Laszlo.Zsak@ars.usda.gov) (L. Zsak).

## Results

### The ChPV ABU-P1 genome

The fully assembled ChPV ABU-P1 genome is 5257 nucleotides (nt) in length and has a basic organization similar to previously described members of the *Parvoviridae* (Fig. 1). The genome is flanked on the 5' and 3' ends by 206 nt direct repeat sequences, each of which contains 39 nt inverted repeats presumed to form a hairpin structure similar to the structures common in the parvoviruses (Fig. 2) (Farkas et al., 2004; Muzyczka and Berns, 2001; Sukhumsirichart et al., 2006). The 206 nt direct repeats are identical and are found in the same orientation 5' to 3' in the genome. The genome contains 32.36% A, 21.95% G, 24.08% T, and 21.61% C, with an A + T content of 56.44% and a C + G content of 43.56%. The overall genomic organization of ChPV ABU-P1 is similar to other parvoviruses, with two major predicted open reading frames (ORFs). The 5' ORF is 2085 nt long, and a blastx search using the putative amino acid sequence revealed a protein similar to the *Parvoviridae* non-structural (NS) proteins NS1, with the greatest similarity to human bocavirus NS1. The major 3' ORF is 2028 nt long, and encodes a protein similar to the *Parvoviridae* capsid protein VP1, with the greatest similarity to avian adeno-associated virus, a member of the *Dependovirus* genus. The ChPV genome also contains a small (306 nt) ORF located between the major 5' and 3' ORFs. This ORF shows no homology to known protein coding regions in the databases.

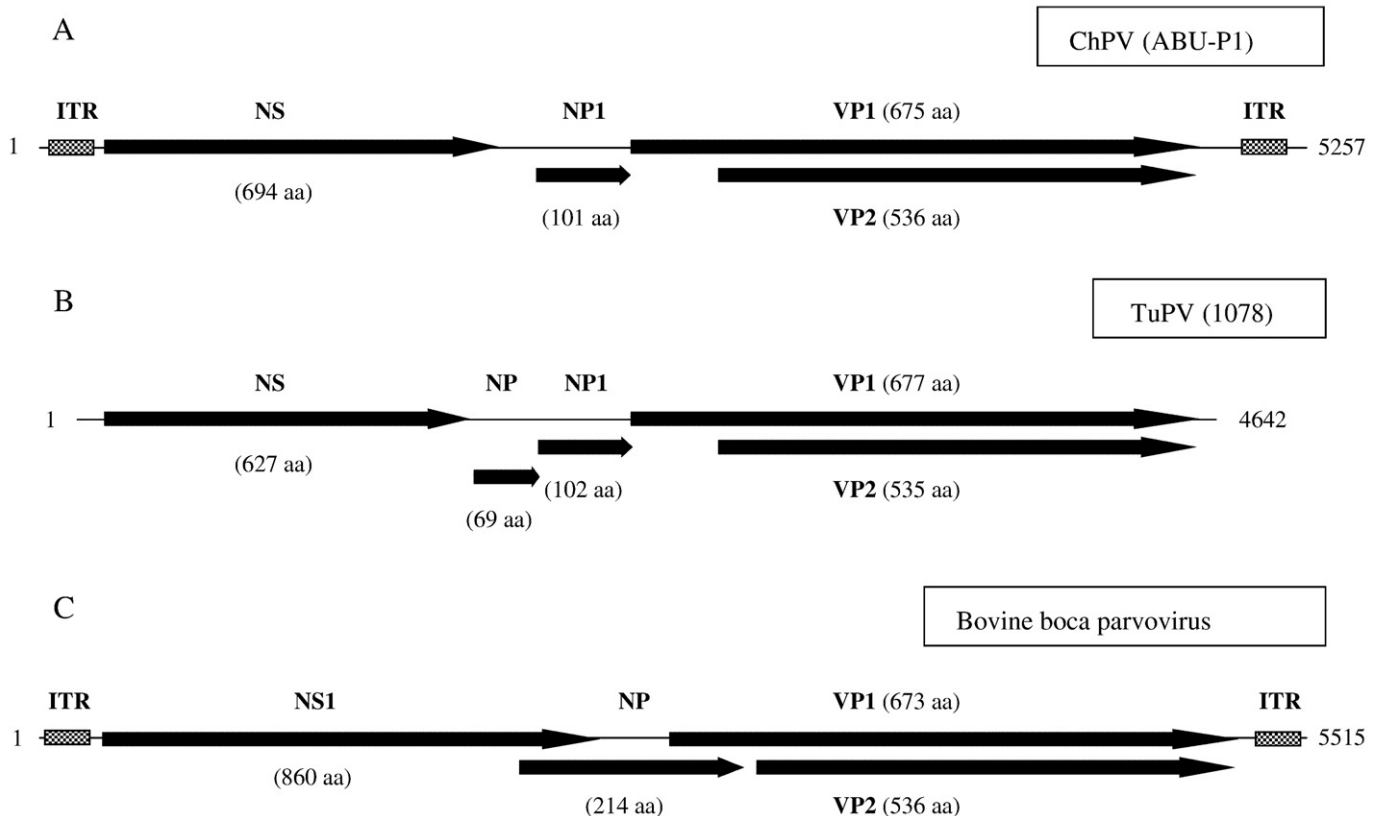
### NS1

The ChPV NS1 ORF is preceded upstream by the 39 bp inverted repeat sequence. The NS1 start codon is in a strong Kozak context (AAGATGG) and is preceded directly upstream by a TATA box, an

extended CAAT box, and a putative Inr-box (Hernandez, 1993; Rupp et al., 1990; Smale and Kadonaga, 2003). NS1 also contains a well-conserved phosphate-binding loop (P-loop) motif (GXXXXGKT...EE), distinguishing it as a member of the ATP and GTP-binding superfamily of proteins (Saraste et al., 1990). At the amino acid level, the ChPV NS1 protein exhibited from 45.2 to 53.9% amino acid similarity (17.1 to 19.0% identity) with NS1 from other parvovirus isolates. The NS1 similarity increased to 89.3 and 100% (99.3 to 83.7% identity) when ChPV was compared with two turkey-origin parvoviruses (TuPV), TuPV 260 and TuPV 1078 respectively. The lower identity noted when comparing the ChPV and TuPV 1078 NS1 coding sequences can be partially explained by the presence of a stop codon beginning at position 1882 in the TuPV 1078 NS1 coding region.

### VP1

The predicted VP1 ORF begins at position 2998 in the ChPV genome. The predicted VP1 start codon (ATG) overlaps with the stop codon (AAT) of the predicted 306 nt ORF located between the ChPV NS1 and VP1 ORFs. VP1 is preceded upstream by a TATA box and a polyadenylation signal (AATAAA) follows the ORF; this poly-A signal immediately precedes the 39 nt inverted repeat at the 5' end of the genome. One of two downstream ATG codons that begin 417 and 420 nt downstream to the start of the putative VP1 ORF may serve as the start codon for a VP2 capsid protein that shares a stop codon (TAA) with VP1. The second of these ATG codons has the stronger Kozak consensus sequence (ATGATGG) compared to the translation initiation context of the first codon (GAAATGA) and remains in frame (Kozak, 1987). A third start codon with a favorable translation initiation context (GACATGG) is present at position 3067 in the ChPV genome, and may represent the beginning of a putative VP3 ORF as has been described in the goose and Muscovy duck parvoviruses



**Fig. 1.** A graphical representation of the genome organization of (A) ChPV ABU-P1, (B) the TuPV 1078, and (C) bovine bocavirus, a parvovirus with an overall genome organization similar to ChPV ABU-P1. The full genome length in nucleotides and the number of amino acids in each encoded protein are indicated. ITR = inverted terminal repeat.

ChPV (ABU-P1)

1 GAGTCTTGCTTTCTCTAGTCAACCTACTTCCATCCTCCATGGAAGTATTATTAAGTGTAGTTTTTATACCTTATCGCTAG  
 81 GGGAATCCCTGGCGCTTGGCAACAGCGGGCTTCGCCCGCGCCCGAAGGGCGTCTTAGAAGGCGTGACTATGCATATGAGG  
 161 GATAACATGGAAGTTGGTGGTTGGTTGTAGAGAGAAAGCAAGACTCATAACGGGTGGTATGACCTATCAATGATTGACTA  
 241 GCTGAGTCATGCTGACTTCGGATTTCGCTGGGTATATAAGGCGGTGTCTGGAGCGGCAATCACCCATCCACCTTAAGTGG  
 321 TTCCGGGTCACTCTCAGAGCGCTCATTGCCCCTTTGACAGGCGAACATCCGCACGTTGATTTCGGCTGCCAGTCTGCTGCTG  
 401 AGCTGGTAAAGATCGCATTCCCAACACGCGGTGGGTTTTCCGCTATTCTAAATCTTCCAGCGATTTAATCGTGGGACAGA  
 481 AAAATATGCGTGTATTCGAGTGTGTAAGTCCGGTTACACTTCTCCGGATTTCAGGCGAATTCGTGGCTTCTAGATTTCCA  
 561 CCCAACCACGAATTGAGACAAGAAATAGAAATTTTATCGGGAACCCAGAATGGGACGATCCAGATTTGACATGGATAGA  
 641 TTTCCACATCAATTGGCGAAACACGCCATAGACGTGAAACTAACTCAATCCAAATTTACAAATTTTCTGGCAAATGG  
 721 AAAAATCTAGTCAGTCGGGAAGATATCATCTCCATTACTGTTCGTAGAGGATGACTTCCAGGCAAAATACATGGATA  
 801 GTCAAAGAATGAGAAGGGAATTTGCAACAAMGTGCAGAAATGTGCACATGTTTTGTCTGATTTCCCAAAAGTCA  
 881 AATTTGGGGATCGCTGGTGAATATCCAAAACGCTATGCTTTCACTGAATAGAGGATACTCACCGAGAATTTGGGAAACCAA  
 961 TCCCAACACAGTGAATCCATACTCATTTCTAAAATACTACATGTTCAATTTCTGCTAAAATGAGCTTTCTGAGAGGGGCT  
 1041 ACTCAAAATTAGCTGAAAAATTACCAATTTGGTACAAGATATGCTAGATTTGAAGGGTCTCTGACCTGGACGAGGGTCC  
 1121 CGAAATCCATCGCAGGAATTAACCTCCAGGGGACACTTACTCTGTATGCTATGTAAGTACTGAAACTGGTTGGGGGAAACG  
 1201 GTGGGGGAGACCGTGTATCAAAACAGTATCATTGAAAAATATGCATGGAAGCATTAAAGATATGCAGAGAAAAATCAC  
 1281 ATTTTACCATGAAAGCATTTAAGCTAAAATTTCCAGACAAATTCATGCAGTTCAGTTCGCGAAATCAGGGGTTAGTGAA  
 1361 ATTAGAAGAAAATAAACCTGTATTTGTAGACTTATCAACGAGCAGATGCGTGGGAAATAGCTAAAATCAATACATG  
 1441 GAGATGTTGACACAGCTGATCTAGATGATAATTTAGCCATCCGGTTATGCAGCTACCAAGGATATAGTCCAAAGTACGTT  
 1521 A R L I L C W L S G Q A G K K N A L Y F H **G P A N T G**  
 1601 GCTAGACTGATATTGTGCTGGCTATCCGGACAAGCAGGAAAAAACGCATTATATTTCCACGGACCAGCCAATACTGG  
 1681 **K T** M M A E S I C K M V Q I Y G N V N H N N K N F P  
 1761 TAAAACAATGATGGCAGAAAGCATATGCAAAATGGTACAATATACGGTAACGTAATCATAACAACAAAAATCCCAC  
 1841 L N D C H N K A V L W W **E E** C S M T E E H E S A K C  
 1921 TCAACGACTGTCACAACAAAGCAGTACTTTGGTGGGAAGAATGTAGCATGACGGAAGAACAGTAGAATCTGCAAAATGT  
 2001 I M G G S S V R I D K K N Q D S V L L C K T P I V I T  
 2081 ATCATGGGGGATCCTCAGTGAAGATAGATAAAAAAATCAAGATTCAGTTCATGTAATAATCCAAATCGTTATTAC  
 2161 S N N D I T Q V S S R N A I S T V H A A A I R A R C  
 2241 TTCTAATAACGATATCACTCAAGTTTCAAGCAGGAATGCCATCTCAACAGTTCATGCAGCTGCAATAAGAGCAAGATGCT  
 2321 L K F T F N N W L T S N W G L I T V D Q M Y Q F L C W  
 2401 TAAAATTTACATTTAACAATTTGGCTAACTTCAAATTTGGGGACTGATAACCGTTGATCAAATGTATCAATTTTATGCTGG  
 2481 G E L E G I P T V D S S L L K D H P E F N G L F E K A N  
 2561 GGAGAATTAGAGGGCATTCCGACCGTGGATAGTCTCTGAAAGATCATCCGAATTTAACGGGACTTTGCCTTATAACCA  
P K G K F C A D C V N Q F S T A T N L T V C P A C S  
ACCCAAAGTAAATTTGTGCCGATTGTGTTAACAGTTTCAAGTTCAGCACTGCAACTAACCTGACCGTGTGCCAGCTGCTCCG  
G W T R K P F S E E E P P Y T G S E S G L F E K A N  
GATGGACAAGAAAGCCATTCTCCGAAGAAGAGGAACCCCCATACTGGTTCTGAATCCGGGCTCTTTGAGAAAGCCAAC  
E G K N E V P G I Y P V N S G S R L I P E A P I L Q K  
GAAGTAAAGATGAGTCCAGGGATTTATCCCGTAAATAGTGGATCCAGGTTAATCCCTGAGGCCCAATCTTACAGAA  
P S T Y W E T E I P S E S G F P D E P D E P S A  
ACCATCGACCTATTGGGAGACGGAGATCCCCAGCGAATCGAGTGGATTTCCAGACGAGCCAGACTTCCCGCAAGCGCAA  
K L L L K V D D L H L E S D L K V E Q D P L E C K K C  
AACTTCTCTGAAAGTGGACGATCTCCACCTAGAAAAGCGATCTGAAAAGTGGAGCAGGATCCTCTGGAGTGCAAAAAGTGC  
C L V Q \*  
TGTCTGGTTCAAAGTCCGAAGTAAAAGCATAACATTTACTGATGATCTACCTTCCCTTACTGAGTTATTTTGC  
GACGAAACGCTGCCCCATCCCCGACATCCCCAATCAATCTCCCTGAACCTCTCGAGTGGTACAATGGGGGACTTCAA  
M C W Y N A N I

Fig. 2. Full-length nucleotide sequence of the ChPV genome. The terminal repeat sequences are shaded gray, and the palindromic (hairpin-forming) regions are overlined with an arrow. TATA boxes, Inr's, and CAAT boxes are underlined and indicated with a label. Start and stop codons for each of the major and minor ORFs are boxed; a putative alternative start codon for the VP2 ORF is boxed and shaded. The P-loop motif and downstream residues involved in NTP binding are shaded and in bold type. Putative polyadenylation signals are underlined.

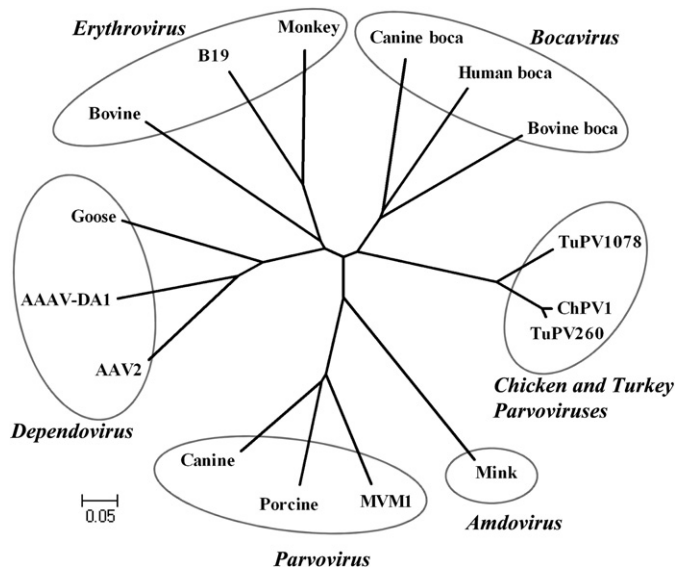
2641 AGCTCAATGGGTGGAAGGATACATTGGAAGTGGAGGGTGCATTTTCTAATTCATG TGTGGTATAATGCTAATATTA  
 T Q I Q T V E G D P C T N P D V F H E L Y F P I K Q H  
 2721 CACAGTTACAGTAGAAGGTGATCCTGTACTAATCCTGATGTTTTTTCACGAGTTATATTTCCCTATCAAGCAACAC  
 A M D A R M L K M K I E D K I Q K G D F T Q D I H L D  
 2801 GCCATGGACGCACGCATGCTGAAAATGAAAATCGAAGACAAAATACAAAAGGGCGACTTCACACAAGACATACACCTAGA  
 M T L T Q C I Q T A L S D D P A G I W P C V D R L A  
 2881 CATGACATTAACAATGCATTCAAACAGCACTTAGCGACCCAGCAGGGATATGGCCATGTGTAGACAGGTTAGCTA  
 N Y C N V I L A S I G T Q \*  
**TATA Box** M K Y K A P K G Y V P S L P  
 2961 ATTACTGCAATGTATATATGGCTTCTATAGGCACACAATGAAATACAAAGCTCCGAAAGCTACGTCCTTCACTGCCCA  
 T T D E E A Y H H T R W W L D H V E N Y G K N K A I G  
 3041 CCACTGATGAAGAAGCATATCACACACTAGATGGTGGTGGATCATGTTGAAAATATGGAAAAATAAAGCTATTGGT  
 R G K G K A R K E L T P Q Q K A E R K R F F I T Q A Q  
 3121 AGAGGCAAAGGAAAAGCTAGGAAAGAGCTCACACCACAACAAAAGCTGAACGCAAACGCTTTTTTATACAGCAGGCACA  
 K N K K P N Y G I P S S S Q H K E F F K N H Q G A Q  
 3201 AAAAAATAAAAAACCTAATTACGGCATCCCATCTTCATCGCAGCACAAAGAGTTTTCAAAATCATCAGGGGGCGCAAG  
 G K K A K Q K T K H I E N Q A P Q R Q E D P E E G P A  
 3281 GGGGAAAAGCAAAGCAAACCAACATATAGAAAATCAAGCACCACAAGGCAAGAGGACCCAGAGGAAGGGCCCGCA  
 P K Q P R V E E N P Y N Q E E L E E M M A D E N E P D  
 3361 CCTAAACAACCTAGAGTAGAAGAAAACCATATAATCAAGAAGATTAGAAGAAATGATGCGCAGATGAAAATGAACCTGA  
 Q S G I P M E L S D G P G T V G G G G G G G I G H Y  
 3441 TCAATCTGGGATCCCTATGGAATTAAGCGACGGCCTGGGACTGTGGGAGGAGGTGGTGGTGGGGGAATCGGGCATTCTA  
 T G N W N C D T I W A E N T C K T Y A S R H C V C L M  
 3521 CCGGTAATTGGAATTGTGATACTATATGGGTGAAAATACATGTAACCACTATGCAAGCCGCCATTGTGTCTGTCTTATG  
 R D M D Q Y K A I G N A N G Q H G L D T E N Q Y F  
 3601 CGTGACATGATCAATATAAAGCAATTGGCAACGCTAACGGGCAACACGGGTTAGACACTGAAAACCAAACAATACTT  
 G F S T P W N Y I D F N Q Y S V H F S P R D W Q Q L  
 3681 TGGTTTCTTACACCATGGAATATATGACTTAAATCAATACAGCGTGCCTTTAGTCCACGAGATGGCAACAACTAG  
 V N N Y R V R P R A V H V K I F N L Q V I Q K T V T  
 3761 TTAATAATTATTTCGAGATGGCGCCACGCGCTACACGTGAAAATCTTTAACCTGCAAGTCATCCAGAAAACCGTAACA  
 D S G T Q Y S N D L T G T I Q I F A D Q E G R Y P R L  
 3841 GATAGCGGTACCCAGTACTCTAACGACTTGACAGGAACATCCAAATATTTGAGATCAGGAAGGAGATATCCTAGATT  
 L Y P N Q T T L M G P F P S N I Y Y L P Q A Y M T  
 3921 GTTATACCTTAACCAAACAACCTAATGGGTCCATTTCCATCCAAATATATACTATCTGCCTCAATACGCATATGACAA  
 S M F N S G G P T V S S W W G F P T R E S A F Y C L D  
 4001 GCATGTTTAAATCGGGTGGACCAACAGTATCATCGTGGTGGGGATTTCCAACACGCGAAAGCGCATTTTACTGTCTAGAC  
 E S P S Q M L R T G N E W E T T F I F P S S T Q W C S  
 4081 GAAAGCCCTCGCAAATGTTAAGGACCGTAAGTGGGAACCCACATTTATCTTTCCATCGAGTACACAATGGTGTAG  
 N K I S T V S I T Q R Q N P L Y D T W N V N G R G D  
 4161 CAATAAATATCAACCGTGTCCATTACACAAAGACAAAACCCATATACGATACATGGAACGTCATGGCAGAGGCGATG  
 D A K R A G N F C T W R S P W Y P G P Y I Y L T D T T S A  
 4241 ATGCTAAAAGAGCAATTTCTGTACTTTGGAGAAGCCCATTTTCCATCCAGGCTTACATTTACCTAACCGTACATCCGCA  
 A S Q R L E N L A N V A V G P D G M P L A P G M P Q Y  
 4321 GCATCACAGCGTTTAGAAAATCTTGCTAACGTTGCCGTAGGCCAGACGGCATGCCTCTAGCACCAGGAATGCCACAATA  
 R S E S D K D E Y L H T T F W V P K K V K M N E G D I  
 4401 CAGATCAGAATCCGATAAAGACGAATATCTACATACATTTTGGGTACCTAAAAAGTTAAGATGAACGAAGGATATAC  
 R N A Q I D V S T A T K V Q V D T S T L Y N T C P R S  
 4481 GTAACGCGCAAATAGATGTATCCACTGCAACCAAAGTACAAGTAGATACATCGACACTATACAATACATGTCCTAGATCA  
 G Y M S T V S Q G G A P A P I V W S G C V P G M I W D  
 4561 GGATACATGTCCACTGTTTACAGGGAGGGGACCCGACCTATGTCTGGTCAAGGTTGTACTGGCTGATGATGGGA  
 N R P A T Y F D P I W Q Q F P E T D E Q F K I I S Q  
 4641 CAATAGGCGGCCACCTACTTTGATCCGATATGGCAACAAATCCAGAAACAGACGAAACAAATTCAAAATATCTCTCAAT  
 L G G G I P V K E S P G H I F V K V T P K P T G G A A N G  
 4721 TGGGGGTATACCTGTTAAAGAAATCACCTGGTCACTTTTGTCAAAGTACCGCTAAGCCTAAGCCTGCAAAATGGT  
 L I D E Y C T F T C T V E I E W E L E P L N T H R W N  
 4801 TTGATCGATGAATATTGTACTTTTACGTGACTGTGCAAAATAGAAATGGGAATTAGAACCCTCAACACTCACAGATGGAA  
 M R S M I S Y E T T D A K A G R Q I V D E N G Q Y Q  
 4881 CATGAGAAGTATGATTTCTGACGAGACCAGGATGCAAAAGTGGAAAGCAAAATGTGCGATGAAAATGGTCAATACCAGG  
 V N V N S A D I T R L Y M T K R A P R T N \*  
 4961 TCAATGTTAACTCGGCTGACATTACTCGTCTGTATATGACCAAGCGCGCCGCGGACCAAC TAA TGTAACTTTCAAAA  
 5041 ATTTCAATAAAGAGTCTTGCTTTCTCTAGTCAACCTACTTCCATCCTCCATGGAAGTATTATTAAGTGTAGTTTTTATAC  
 5121 CTTATCGTAGGGGAATCCCTGGCGCTTGGCAACAGCGGGCTTCGCCCGCCCGAAGGGCGTCTTAGAAGCGGTGACTA  
 5201 TGCATATGAGGGATAACATGGAAGTTGGTGGTTGGTTGTAGAGAGAAAGCAAGACTC

Fig. 2 (continued).

(Zadori et al., 1995). The ChPV VP1 protein shares from 46.6 to 59.0% amino acid similarity (10.1 to 24.5% identity) with VP1 from other parvoviruses. The VP1 similarity increases to 100 and 95.1% (99.7 and 79.2% identity) when compared to VP1 from TuPV 260 and TuPV 1078, respectively.

#### Phylogenetic analysis

ClustalW was used to align the complete ChPV genome coding region with complete parvovirus sequences from mammals and birds available in the databases, and with two recently sequenced TuPV



**Fig. 3.** A phylogenetic tree prepared using the full coding sequence of the indicated parvoviruses. Representative genera within the *Parvovirinae* are indicated. The evolutionary relationships were inferred using the Neighbor-Joining method. Phylogenetic analyses were conducted using MEGA4. GenBank Accession numbers for ChPV, GU214704; TuPV 1078, GU214705; and TuPV 260, GU214706.

genome coding regions. The Neighbor-Joining tree was produced using MEGA4. The turkey and chicken sequences clustered together, and were clearly separate from the other members of the *Parvovirinae* (Fig. 3). A previous phylogenetic analysis using a conserved region of the ChPV and TuPV NS1 gene produced a similar tree (Zsak et al., 2008), and a phylogenetic analysis of the ChPV and TuPV VP1 gene produced an identical tree when compared to the full-length parvovirus coding region (data not shown).

## Discussion

This is the first analysis of the complete nucleotide sequence from the ChPV ABU-P1 strain. Small viral particles (15 to about 25 nm) resembling parvovirus have been identified for years in the intestinal tracts of poultry (Kisary et al., 1984; Trampel et al., 1983; Woolcock and Shivaprasad, 2008). The use of a particle-associated nucleic acid (PAN) sequencing technique has recently allowed the partial molecular characterization of this novel ChPV (Zsak et al., 2008), and led to the identification and sequencing of two additional avian parvoviruses from turkeys: TuPV 260, originally detected in the intestinal tract of a turkey from California, and TuPV 1078, originally detected in the pooled intestinal tracts from a North Carolina turkey farm. All three of these poultry isolates are very similar to each other, yet each differs significantly from other members of the *Parvovirinae*. The high amino acid identities noted among the turkey and chicken isolates suggests that they diverged at sometime in the recent past from a common ancestor. It is known from studies of another member of the *Parvovirinae*, canine parvovirus (CPV), that the parvoviruses, despite their single-stranded DNA genome, have a mutation rate that approaches that of RNA viruses, which allows rapid evolution and host adaptation (Hoelzer et al., 2008; Shackelton et al., 2005). Interestingly, ChPV and TuPV do not group phylogenetically with the parvoviruses that infect geese (GoPV) and Muscovy ducks (MdPV), parvoviruses that are closely related to adeno-associated virus 2 (AAV-2), a member of the *Dependovirus* genus (Fig. 3) (Zadori et al., 1995; Zsak et al., 2008). Further, the ChPV NS protein shares only 19.1% amino acid identity with the MdPV NS protein.

The ITR sequences located at each end of the ChPV genome (Fig. 2) are each 206 nt long, somewhat shorter than those found in the

prototypical human parvovirus B19 (383 nt) but longer than those found in some other autonomously replicating parvoviruses, namely the murine parvoviruses that have terminal palindromes of 115 and about 200 nt long in the same genome (Astell et al., 1979; Deiss et al., 1990). The terminal palindromes found at the 5' and 3' ends of most autonomously replicating parvoviruses bear no nucleotide identity with one another. Interestingly, this is not the case with ChPV, which has identical palindrome sequences at its 5' and 3' end (Fig. 2), a trait it shares with parvovirus B19 (Deiss et al., 1990; Muzyczka and Berns, 2001). It would be interesting to determine if the 5' end of the ChPV genome exists in an alternative orientation, as is the case with B19, MdPV and GoPV, in which the 5' sequence can be found in one of two "flip" or "flop" orientations, with one orientation being the inverted complement of the other (Deiss et al., 1990; Zadori et al., 1995).

The ChPV NS1 deduced amino acid sequence contains highly conserved motifs important for the initiation of parvovirus replication (Zsak et al., 2008), including a well-conserved phosphate-binding loop ("P-loop") motif specifically involved in the binding of nucleoside triphosphates by this protein (Saraste et al., 1990). This motif is required for pathogenesis in the human parvovirus B19 and is found in many parvovirus isolates (Momoeda et al., 1994). The 5' major ORF of ChPV appears to encode the viral capsid proteins VP1, VP2, and VP3. These three proteins together comprise the parvovirus virion (Muzyczka and Berns, 2001) and are responsible for the production of neutralizing antibodies during an infection (Saikawa et al., 1993).

The present analysis does not include members of the *Densovirinae* subfamily, which infect insects and other invertebrates (Muzyczka and Berns, 2001). ChPV and the closely related TuPVs were compared across their entire coding region with representative members of the *Parvovirinae* subfamily, which includes the *Dependovirus*, *Bocavirus*, *Erythrovirus*, *Amdovirus* and *Parvovirus* genera (Fig. 3). It is clear from the multiple sequence alignments and phylogenetic analysis that ChPV, along with the closely related TuPVs, represents a distinct member of the *Parvovirinae* subfamily and should be recognized as the prototypical member of a novel genus within the *Parvovirinae*.

## Materials and methods

### Assembly of the prototypical chicken parvovirus (ChPV ABU-P1) genome

The chicken parvovirus strain chosen for this analysis was originally isolated in Hungary from the intestines of chickens showing signs of a stunting syndrome (Kisary et al., 1984), and this original cesium chloride gradient-purified parvovirus (provided by J. Kisary) was propagated in specific pathogen free (SPF) chickens as described previously (Kisary, 1985; Zsak et al., 2009). A sequence-independent polymerase chain reaction (PCR) protocol was employed to amplify particle-associated nucleic acid (PAN) present in ChPV ABU-P1-positive chicken intestinal homogenates, and has been described elsewhere in detail (Zsak et al., 2008). The randomly amplified PAN was ligated into the TOPO-TA cloning vector (Invitrogen) and the ligation mix was used to transform competent TOP-10 *E. coli* cells (Invitrogen). Using this approach, a total of 768 clones were identified and sequenced using the M13 forward and reverse primers on an AB-3730 automated DNA sequencer. The sequenced clones were used as query sequences to search the GenBank non-redundant nucleotide (nr/nt) database using the blastn algorithm and the non-redundant (nr) protein database using the blastx algorithm. Four contigs with an average of eight-fold coverage and lengths of 1417 nt, 933 nt, 1069 nt, and 739 nt were assembled from 27, 24, 19, and 21 clones, respectively. These contigs had no significant nucleotide similarity to database sequences, but the deduced amino acid sequence from each contig

revealed significant similarity to the members of the family *Parvoviridae* in the database. PCR primers were designed using the four assembled contigs and were subsequently used to close gaps of 58, 443, and 205 nt between the contigs and assemble the complete ChPV ABU-P1 genome. The right terminal region of contig 4 contained a portion of the right inverted terminal repeat (ITR). Primers designed using this partial ITR used in combination with forward and reverse primers within contig 1 (left terminal region of the genome) and contig 4 resulted in the cloning of the ITR regions at both the right and left genomic termini.

#### Sequence analysis

The ChPV and TuPV genomes and ORFs were aligned with each other and with selected parvovirus sequences available in the databases using ClustalW (Thompson et al., 1994). Searches for conserved domains and transcription factor binding sites within the ChPV genome were performed with GeneQuest (DNASar/Laser-gene8) and the Conserved Domain Database (CDD) Search Service v2.17 at the National Center for Biotechnology Information (NCBI) website (Marchler-Bauer et al., 2009). Phylogenetic analysis and tree construction were performed with MEGA4 using the Neighbor-Joining method (Tamura et al., 2007).

#### Acknowledgments

Thanks to Fenglan Li for excellent technical assistance and to the South Atlantic Area sequencing facility for outstanding support.

#### References

- Astell, C.R., Smith, M., Chow, M.B., Ward, D.C., 1979. Structure of the 3' hairpin termini of four rodent parvovirus genomes. *Cell* 17, 691–703.
- Barnes, H.J., Guy, J.S., 2003. Poultry enteritis mortality syndrome. In: Saif, Y.M., Barnes, H.J., Glisson, J.R., Fadly, A.M., McDougald, L.R., Swayne, D.E. (Eds.), *Disease of Poultry*, 11th ed. Iowa State Press, pp. 1171–1180.
- Barnes, H.J., Guy, J.S., Vaillancourt, J.P., 2000. Poultry enteritis complex. *Rev. Sci. Tech.* 19 (2), 565–588.
- Decaro, N., Desario, C., Parisi, A., Martella, V., Lorusso, A., Miccolupo, A., Mari, V., Colaiani, M.L., Cavalli, A., Di Trani, L., Buonavoglia, C., 2009. Genetic analysis of canine parvovirus type 2c. *Virology* 385 (1), 5–10.
- Deiss, V., Tratschin, J.D., Wietz, M., Siegl, G., 1990. Cloning of the human parvovirus B19 genome and structural analysis of its palindromic termini. *Virology* 175, 247–254.
- Farkas, S.L., Zadori, Z., Benko, M., Essbauer, S., Harrach, B., Tijssen, P., 2004. A parvovirus isolated from royal python (*Python regius*) is a member of the genus *Dependovirus*. *J. Gen. Virol.* 85 (Pt 3), 555–561.
- Goodwin, M.A., Davis, J.F., McNulty, M.S., Brown, J., Player, E.C., 1993. Enteritis (so-called runting stunting syndrome) in Georgia broiler chicks. *Avian Dis.* 37, 451–458.
- Gough, R.E., 2008. *Parvovirus* infections. In: Saif, Y.M. (Ed.), *Diseases of Poultry*, 12th ed. Blackwell, pp. 397–404.
- Guy, J.S., 1998. Virus infections of the gastrointestinal tract of poultry. *Poultry Sci.* 77, 1166–1175.
- Hernandez, N., 1993. TBP, a universal eukaryotic transcription factor? *Genes Dev.* 7 (7B), 1291–1308.
- Hoelzer, K., Shackelton, L.A., Parrish, C.R., Holmes, E.C., 2008. Phylogenetic analysis reveals the emergence, evolution and dispersal of carnivore parvoviruses. *J. Gen. Virol.* 89 (Pt 9), 2280–2289.
- Hueffer, K., Parrish, C.R., 2003. *Parvovirus* host range, cell tropism and evolution. *Curr. Opin. Microbiol.* 6 (4), 392–398.
- Kisary, J., 1985. Experimental infection of chicken embryos and day-old chickens with parvovirus of chicken origin. *Avian Pathol.* 14, 1–7.
- Kisary, J., Nagy, B., Bitay, Z., 1984. Presence of parvovirus in the intestine of chickens showing stunting syndrome. *Avian Pathol.* 13, 339–343.
- Kisary, J., Avalosse, B., Miller-Faures, A., Rommelaere, J., 1985. The genome structure of a new chicken virus identifies it as a parvovirus. *J. Gen. Virol.* 66, 2259–2263.
- Kozak, M., 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15, 8125–8148.
- Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.L., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Bryant, S.H., 2009. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.* 37, 205–210.
- Momoeda, M., Wong, S., Kawase, M., Young, N.S., Kajigaya, S., 1994. A putative nucleoside triphosphate binding domain in the nonstructural protein of B19 parvovirus is required for cytotoxicity. *J. Virol.* 68, 8443–8446.
- Muzyczka, N., Berns, K.L., 2001. *Parvoviridae: the viruses and their replication*. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, 4th ed., Vol. 2. Lippincott, Williams and Wilkins, pp. 2327–2359.
- Pantin-Jackwood, M.J., Spackman, E., Day, J.M., 2007a. Pathology and virus tissue distribution of Turkey origin reoviruses in experimentally infected Turkey poults. *Vet. Pathol.* 44 (2), 185–195.
- Pantin-Jackwood, M.J., Spackman, E., Day, J.M., Rives, D., 2007b. Periodic monitoring of commercial turkeys for enteric viruses indicates continuous presence of astrovirus and rotavirus on the farms. *Avian Dis.* 51 (3), 674–680.
- Pantin-Jackwood, M.J., Day, J.M., Jackwood, M.W., Spackman, E., 2008a. Enteric viruses detected by molecular methods in commercial chicken and turkey flocks in the United States between 2005 and 2006. *Avian Dis.* 52 (2), 235–244.
- Pantin-Jackwood, M.J., Spackman, E., Day, J.M., 2008b. Pathogenesis of type 2 turkey astroviruses with variant capsid genes in 2-day-old specific pathogen free poults. *Avian Pathol.: J. W.V.P.A.* 37 (2), 193–201.
- Reynolds, D.L., Saif, Y.M., Theil, K.W., 1987. Enteric viral infections of turkey poults: incidence of infection. *Avian Dis.* 31, 272–276.
- Rupp, R.A., Kruse, U., Multhaup, G., Gobel, U., Beyreuther, K., Sippel, A.E., 1990. Chicken NFI/TGGCA proteins are encoded by at least three independent genes: NFI-A, NFI-B and NFI-C with homologues in mammalian genomes. *Nucleic Acids Res.* 18 (9), 2607–2616.
- Saikawa, T., Anderson, M., Momoeda, M., Kajigaya, S., Young, N.S., 1993. Neutralizing linear epitopes of B19 parvovirus cluster in the VP1 unique and VP1–VP2 junction regions. *J. Virol.* 67, 3004–3009.
- Saraste, M., Sibbald, P.R., Wittinghofer, A., 1990. The P-loop—a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* 15, 430–434.
- Shackelton, L.A., Parrish, C.R., Truyen, U., Holmes, E.C., 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc. Natl. Acad. Sci. U. S. A.* 102 (2), 379–384.
- Smale, S.T., Kadonaga, J.T., 2003. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72, 449–479.
- Spackman, E., Pantin-Jackwood, M., Day, J., Sellers, H., 2005. The pathogenesis of turkey origin reoviruses in turkeys and chickens. *Avian Pathol.* 34 (4), 291–296.
- Sukhumsirichart, W., Attasart, P., Boonsaeng, V., Panyim, S., 2006. Complete nucleotide sequence and genomic organization of hepatopancreatic parvovirus (HPV) of *Penaeus monodon*. *Virology* 346 (2), 266–277.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Trampel, D.W., Kinden, D.A., Solorzano, R.F., Stogsdill, P.L., 1983. *Parvovirus*-like enteropathy in Missouri turkeys. *Avian Dis.* 27 (1), 49–54.
- Truyen, U., 1999. Emergence and recent evolution of canine parvovirus. *Vet. Microbiol.* 69 (1–2), 47–50.
- Woolcock, P.A., Shivaprasad, H.L., 2008. Electron microscopic identification of viruses associated with poulter enteritis in turkeys grown in California 1993–2003. *Avian Dis.* 52, 209–213.
- Zadori, Z., Stefancsik, R., Rauch, T., Kisary, J., 1995. Analysis of the complete nucleotide sequences of goose and muscovy duck parvoviruses indicates common ancestral origin with adeno-associated virus 2. *Virology* 212 (2), 562–573.
- Zsak, L., Strother, K.O., Day, J.M., 2009. Development of a polymerase chain reaction procedure for detection of chicken and turkey parvoviruses. *Avian Dis.* 53, 83–88.
- Zsak, L., Strother, K.O., Kisary, J., 2008. Partial genome sequence analysis of parvoviruses associated with enteric disease in poultry. *Avian Pathol.* 37 (4), 435–441.