



Attenuated live vaccine usage affects accurate measures of virus diversity and mutation rates in avian coronavirus infectious bronchitis virus

Enid T. McKinley^a, Mark W. Jackwood^{a,*}, Deborah A. Hilt^a, Jessica C. Kissinger^b, Jon S. Robertson^c, Cornelia Lemke^c, Andrew H. Paterson^{b,c}

^a Department of Population Health, College of Veterinary Medicine, 953 College Station Road, University of Georgia, Athens, GA 30602, USA

^b Department of Genetics, Center for Tropical & Emerging Global Diseases, University of Georgia, 500 D. W. Brooks Drive, Athens, GA 30602, USA

^c Plant Genome Mapping Laboratory, Departments of Crop and Soil Sciences, Plant Biology, and Genetics, University of Georgia, 111 Riverbend Road, Athens, GA 30602, USA

ARTICLE INFO

Article history:

Received 11 January 2011

Received in revised form 14 April 2011

Accepted 15 April 2011

Available online 22 April 2011

Keywords:

Coronavirus

Infectious bronchitis virus

Molecular evolution

Mutation rate

Recombination

Comparative genomics

Sequencing

Polymorphism

ABSTRACT

The full-length genomes of 11 infectious bronchitis virus (IBV) field isolates from three different types of the virus; Massachusetts (Mass), Connecticut (Conn) and California (CAL) isolated over a 41, 25 and 8 year period respectively, were sequenced and analyzed to determine the mutation rates and level of polymorphisms across the genome. Positive selection was not detected and mutation rates ranged from 10^{-4} to 10^{-6} substitutions/site/year for Mass and Conn IBV types where attenuated live vaccines are routinely used to control the disease. In contrast, for CAL type viruses, for which no vaccine exists, positive selection was detected and mutation rates were 10 fold higher ranging from 10^{-2} to 10^{-3} substitutions/site/year. Lower levels of genetic diversity among the Mass and Conn viruses as well as sequence similarities with vaccine virus genomes suggest that the origin of the Mass and all but one of the Conn viruses was likely vaccine virus that had been circulating in the field for an unknown but apparently short period of time. The genetic data also identified a recombinant IBV isolate with 7 breakpoints distributed across the entire genome suggesting that viruses within the same serotype can have a high degree of genetic variability outside of the spike gene. These data are important because inaccurate measures of genetic diversity and mutation rates could lead to underestimates of the ability of IBV to change and potentially emerge to cause disease.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Avian infectious bronchitis virus (IBV) a gamma-coronavirus in the *Coronaviridae* family causes a highly contagious upper-respiratory tract disease of domestic chickens characterized by coughing, sneezing and tracheal rales. Serotype-specific attenuated live vaccines are routinely used to control the disease, yet IBV remains one of the most widely reported respiratory diseases of chickens worldwide resulting in significant economic loss due to reduction in weight gains and feed efficiency, and condemnations at the processing plant. The virus can also cause permanent oviduct damage in layers and some strains of the virus cause nephritis, which when severe can result in significant mortality.

Abbreviations: IBV, infectious bronchitis virus; nsps, nonstructural proteins; ORF, open reading frame; PBS, phosphate buffered saline; RdRp, RNA-dependent RNA-polymerase; RT-PCR, reverse transcriptase-polymerase chain reaction; UTR, untranslated region.

* Corresponding author. Tel.: +1 706 542 5475; fax: +1 706 542 5630.

E-mail address: mjackwoo@uga.edu (M.W. Jackwood).

Infectious bronchitis virus is an enveloped, positive-sense single-stranded RNA virus with a genome size of approximately 27 kb (Holmes, 1991). The 5' two-thirds of the genome, approximately 21 kb, encodes two large overlapping open reading frames (ORFs), 1a and 1b, which encode 15 non-structural proteins (nsps), nsps 2–16 (IBV does not have nsp 1) including a papain-like protease 2 (the PL1 protease is truncated and non-functional in IBV), a main protease, the RNA-dependent RNA-polymerase (RdRp) and several other non-structural proteins involved in making up the replication transcription complex. The remaining one-third of the genome encodes four structural proteins; spike (S), envelope (E), membrane (M) and nucleocapsid (N), and a number of nonstructural proteins. The spike glycoprotein of IBV forms projections on the surface of the virion. Spike is post-translationally cleaved into S1 and S2 subunits with the S1 subunit forming the outermost portion and S2 forming a stalk like structure that is embed in the viral membrane. The S1 subunit contains hypervariable regions that play a role in attachment to host cell receptors, membrane fusion and entry into the host cell, and it contains conformationally-dependent virus-neutralizing and serotype-specific epitopes (Cavanagh et al., 1998; Niesters, 1987).

It is extremely difficult to control IBV because little or no cross protection occurs between numerous different types of the virus. Coronaviruses evolve by a process of natural selection working on mutations and recombination events (Domingo and Holland, 1997; Moya et al., 2004; Vijaykrishna et al., 2007). These factors contribute to diverse subpopulations of the virus that continually emerge to form new and variant IBV types. In the absence of a specific vaccine, the mutation rate for the hypervariable region in the S1 gene of the 793/B type of IBV was reported as 3×10^{-3} substitutions/site/year (Cavanagh et al., 1998) while the mutation rate for the GA98 virus, which emerged as a result of selection pressure driven by vaccine usage, was reported to be 1.5×10^{-2} substitutions/site/year in the hypervariable region of S1 (Lee and Jackwood, 2001).

Sequence alignments and phylogenetic reconstructions indicate that recombination events play a significant role in coronavirus evolution. Recombination in IBV was demonstrated with co-infection of different IBV types in embryonated eggs (Kottier et al., 1995a). Sequence analysis of the S1 gene from field isolates of IBV also revealed evidence of multiple recombination events (Jia et al., 1995; Wang et al., 1993) and that recombination occurred between vaccine viruses (Mondal and Cardona, 2007). In a recent study, recombination in IBV, which replaced the spike gene with an unknown sequence, likely from another coronavirus, resulted in the emergence of turkey coronavirus (TCov) a new enteric disease of young turkeys (Jackwood et al., 2010).

Characterization of IBV is based on analysis of the spike glycoprotein by either serotyping or genotyping methods (Cavanagh and Gelb, 2008). However, pathogenicity in IBV is associated with spike as well as genes outside of spike (Ammayappan et al., 2009; Cavanagh, 2007; Hodgson et al., 2004). Because typing is focused on the analysis of spike, there is no clear understanding of whether viruses with similar spike proteins also share high sequence similarities in other regions of the genome especially in regions associated with pathogenicity. This gap in knowledge results from the lack of complete genome sequences from viruses of the same serotype.

The objective of this study was to determine the levels of polymorphism across the entire genome of IBV isolates with similar spike genes and to examine the mutation rates for viruses with and without vaccine selection pressure. We sequenced the full-length genome from 11 IBV field isolates of three different serotypes, Massachusetts (Mass), Connecticut (Conn) and California (CAL) isolated over a 41, 25 and 8-year period respectively. The genome sequences were aligned and phylogenetic analysis was conducted to examine genetic relationships between the viruses. Estimates of the rates of mutation were calculated and evidence of recombination was investigated for each of the full-length genome sequences. To our knowledge, this is the first study to identify polymorphisms and determine mutation rates using full-length genome sequences of IB viruses isolated over a multi-year time span.

2. Materials and methods

2.1. Viruses

In this study, we sequenced Mass viruses isolated in 1965, 1972, 1979 and 1985 obtained from Dr. Jack King (Southeast Poultry Research Laboratory, Athens, GA) and one isolated in 2006 obtained from Dr. Pedro Villegas (Poultry Diagnostic and Research Laboratory, Athens, GA). In addition, Conn viruses isolated in 1966, 1972, 1983 and 1991 obtained from Dr. King, and CAL viruses isolated in 1995 and 2003 obtained from Dr. Peter Woolcock (California Animal Health and Food Safety Laboratory System, University of California, Fresno, CA) were sequenced. The specific viruses were all isolated from broiler chickens with clinical signs of disease. The

viruses were propagated in embryonated eggs as described (Gelb and Jackwood, 2008). The strains and embryo passage number are: IBV/Mass/65 (passage 1), IBV/Mass/72 (passage 7), IBV/Mass/79 (passage 2), IBV/Mass/85 (passage 15), IBV/Mass/06 (passage 3), IBV/Conn/66 (passage 3), IBV/Conn/72 (passage 7), IBV/Conn/83 (passage 15), IBV/Conn/91 (passage 1), Cal99/NE15172/95 (passage 3), CAL/CA557/03 (passage 5).

2.2. Viral RNA extraction, amplification and genome sequence determination

Viral RNA extraction, reverse transcription and polymerase chain reaction (RT-PCR), library construction and sequencing were conducted as previously described (Jackwood et al., 2010). Briefly, the viruses were filtered through a 0.8 μm filter then through a 0.22 μm filter (Millipore, Billerica, MA) and viral RNA was purified using the High Pure RNA Isolation Kit according to the manufacturer's recommendation (Roche Diagnostic Corporation, Foster City, CA). The RT-PCR amplifications were performed with the Takara RNA LA PCR kit (Takara Bio Inc., Otsu, Shiga, Japan) using a random primer and an amplification primer in a strand displacement amplification reaction following the manufacturer's protocol. The sequence of the random reverse transcription primer was 5'-AGCGGGGTTGTCGAATGTTTGANNNNN-3', and the amplification primer sequence was 5'-AGCGGGGTTGTCGAATGTTTGA-3'. Both primers were obtained from Integrated DNA Technologies, Inc. (IDT, Coralville, IA). A master mix for the RT reaction was prepared, which included MgCl (5 mM), 10X RNA PCR buffer (1 \times), dNTP mixture (1 mM), RNase inhibitor (1 units/ μl), reverse transcriptase (.25 units/ μl), 5' degenerate primer (2.5 μM), and RNA (5.75 μl /reaction) then 10 μl per sample was aliquoted in a thermocycler tube. The RT reaction conditions were 10 min at 30 °C for the primer annealing then 1 h at 50 °C for extension followed by a five-minute incubation at 99 °C for inactivation of the enzyme and a 5-min period at 5 °C. A PCR master mix, which included at the final concentrations: MgCl (2.5 mM), 10X LA PCR Buffer (1 \times), sterilized distilled water (32.25 μl), Takara LA Taq (1.25 U/50 μl), and 5' primer (0.2 μM) was prepared and 10 μl of the RT reaction was added to 40 μl of the mix. The amplification reaction consisted of a 94 °C step for 2 min followed by 30 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 3 min.

Ten PCR reactions were combined for each virus and purified using the QIAquick PCR Purification Kit (QIAGEN, Foster City, CA) then run on a 1% agarose gel. The PCR products were size selected by removing amplicons between 500 and 1500 bp from the gel, followed by purification using the QIAquick (QIAGEN) gel purification kit.

The PCR products were cloned into the pCR-XL-TOPO vector (Invitrogen, Life Technologies, Carlsbad, CA) according to the manufacturer's recommendations and transformed into One Shot TOPO Electrocompetent *E. coli* cells (Invitrogen) using 30 μl of competent cells mixed with 2 μl of the ligation reaction. Electroporation settings were 20 kV and 200 Ω using a BioRad Gene Pulser (BioRad, Hercules, CA), and the electroporated cells were incubated at 37 °C in 480 μl of Super Optimal broth medium for 1 h on a rotary shaker. The cultures were mixed with 70% glycerol and frozen at -80 °C until plated on Q-trays (Genetix, Boston, MA) containing liquid broth agar CAT#3002-032 (MP Biomedicals, LLC, Solon, Oh) with 50 $\mu\text{g}/\text{ml}$ of kanamycin. The Q-trays were pre-warmed at 37 °C before the entire culture (approximately 500 μl) was spread on the plates and incubated overnight at 37 °C, then robotically picked with a Q-BOT (Genetix, Boston, MA).

Plasmid DNA from the libraries of cloned cDNA fragments for each virus were isolated using an alkaline lysis method modified for the 96-well format, and incorporating both Hydra and Tomtek robots (<http://www.intl-pag.org/11/abstracts/P2c.P116.XI.html>).

Cycle sequencing reactions were performed using the BigDye™ Terminator® Cycle Sequencing Kit Version 3.1 (Applied Biosystems, Foster City, CA) and MJ Research (Watertown, MA) thermocyclers. Finished reactions were filtered through Sephadex filter plates into Perkin-Elmer MicroAmp Optical 96-well plates. A 1/12-strength sequencing reaction on an ABI 3730 was used to sequence each clone from both the 5' and 3' ends. Each viral genome was sequenced to approximately 10× coverage. The accuracy of the sequence is insured by generating sequence reads for both strands.

Gaps and areas with <2× coverage were identified and specific primers were synthesized (IDT) for RT-PCR amplification and sequencing of the ambiguous areas. The RT-PCR was conducted as described above, and the reaction conditions were 42 °C for 60 min, 95 °C for 5 min, then ten 10 cycles of 94 °C for 30 s, 50 °C for 30 s, 68 °C for 90 s, followed by 25 cycles of 94 °C for 30 s, 50 °C for 30 s, 68 °C for 90 s + 5 s/cycle added. The final elongation step was 68 °C for 7 min then the reaction was cooled to 4 °C. The PCR products were directly sequenced in both directions using the amplification primers at a concentration of 15 ng/reaction and the ABI Prism BigDye Terminator v3.0 (Applied Biosystems, Foster City, CA). The amount of cDNA added to the reaction ranged from 20 to 30 ng and the sequencing reactions were analyzed on an ABI 3730 (Applied Biosystems).

The extreme 5' end of each genome was obtained using the 5' RACE system for rapid amplification of cDNA ends, according to manufacturer's protocol (Invitrogen, Life Technologies, Carlsbad, CA). The primer design was based upon an alignment of IBV sequences. The RACE primers are designated; IBV 5'RACE SP1 (5'-CGTATAGAAAAACAAGCGTCAC-3'), IBV 5'RACE SP2 (5'-GTCACGTCTATTTGTATGTCTGCTC-3') and IBV 5'RACE SP3 (5'-TAGCCGACCTTATGCGAGAACC-3'). The extreme 3' end of each genome was obtained using primers designated; M41 3' end reverse (5'-GCTCTAACTCTATACTAGCCTA-3') and Genome 3'F 900 bp (5'-TGACAAGATGAATGAGGAAGGTAT-3').

Sequences for some viruses were verified and sequence gaps were filled by amplification using overlapping primer pairs. Briefly, the primer design tool located at <http://www2.eur.nl/fgg/kgen/primer/Overlapping.Primers.html> was used with a template genome sequence of a virus with the same serotype as the virus to be sequenced. Thirty-two overlapping primer pairs (available upon request) were synthesized (IDT) for each virus and used to amplify the genome. The amplicons were sequenced directly using the amplification primers and the ABI Prism BigDye Terminator v3.0 (Applied Biosystems) as described above.

2.3. Genome sequence analysis

The sequence reads for each virus were assembled using the SeqMan program (DNASTAR, Inc. Madison, WI) and the ORF finder at The National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) was used to predict open reading frames. Genome annotation was entered using the annotation editor SeqBuilder (DNASTAR, Inc.). The BLASTn program (<http://www.ncbi.nlm.nih.gov/BLAST/>) was used to search GenBank (National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>) for similar IBV sequences. The MegAlign program implemented in DNASTAR was used to align sequences and to calculate percent identity and number of substitutions between selected sequences.

2.4. Multiple sequence alignments and phylogenetic analysis

Whole genome alignments were generated using ClustalW and phylogenetic trees were constructed with the Neighbor-Joining

method, Minimum Evolution method, Maximum Parsimony method, and UPGMA with 1000 bootstrap replicates (MEGA4, <http://www.megasoftware.net/index.html> (Tamura et al., 2007)). To reconstruct evolutionary trees, Maximum Likelihood (Tree-Puzzle, <http://www.tree-puzzle.de/>) and Bayesian analysis (BEAST 1.4, <http://beast.bio.ed.ac.uk/>) were conducted.

2.5. Molecular evolutionary rates and recombination detection

A codon-based test of positive selection (Z-test, MEGA4) was used to analyze the numbers of non-synonymous and synonymous substitutions per site (dN/dS ratio) in the spike glycoprotein. Because only 2 CAL type viruses were sequenced herein, additional reference sequences for the CAL type viruses were included in the codon-based test (Z-test) of positive selection; CAV/CAV56b/91 (AF027509), CAV/CAV9437/95 (AF027510), CAL99/CAL99/99 (AY514485) and CA/CA1737/04 (EU925393).

To determine mutation rates, the full length genomes as well as the coding sequences for ORF1ab, S1, E, M and N gene from each virus within the same virus type were aligned using CLUSTALX 2.0.8 (Larkin et al., 2007) and manually edited in Jalview (Waterhouse et al., 2009). The Bayesian Markov chain Monte Carlo (MCMC) implemented in BEAST v1.4.8 (Drummond and Rambaut, 2007) was used to estimate mutation rates of the aligned sequences and the sampling dates for each isolate were used as calibration points. The mutation rates were analyzed using the general time-reversible (GTR) (Rodriguez et al., 1990) substitution model with gamma-distributed among site rate variation with four gamma rate categories (Yang, 1994). The SRD06 model was used to partition codon positions and an uncorrelated log-normal relaxed clock model was selected. Each MCMC chain was run for 10 million states and sampled at every 1000 states. Tracer v1.4.1 (<http://tree.bio.ed.ac.uk/software/tracer/>) was used to confirm convergence of MCMC chains with 10% of each chain discarded as burn-in.

Complete genome nucleotide sequences for Cal99/NE15172/95, IBV/CK/CA99/99, CA/CA557/03, IBV/Conn/66, IBV/Conn/72, IBV/Conn/83, IBV/Conn/91, IBV/Mass/65, IBV/Mass/72, IBV/Mass/79, IBV/Mass/85, IBV/Mass/06 and the IBV/CK/A2 China strain (used as out-group) were aligned using CLUSTALX 2.0.8 (Larkin et al., 2007). To identify recombinants as well as major and minor parents, the data set was scanned using a Recombination Detection Program (RDP) v2 with implemented algorithms GENECONV, BootScan, MaxChi, Chimera and SiScan. Similarity plot and bootscan analyses were performed using the SimPlot program (Lole et al., 1999) to further identify recombination events and recombination breakpoints. The window width and step size were set to 1000 bp and 40 bp, respectively.

3. Results

3.1. Nucleotide sequence accession numbers

Genome sequences generated in this study were submitted to the GenBank database and assigned the following accession numbers; IBV/Mass/65 (FJ904720), IBV/Mass/72 (FJ904721), IBV/Mass/79 (FJ904722), IBV/Mass/85 (FJ904723), IBV/Mass/06 (FJ904713), IBV/Conn/66 (FJ904716), IBV/Conn/72 (FJ904717), IBV/Conn/83 (FJ904718), IBV/Conn/91 (FJ904719), Cal99/NE15172/95 (FJ904714), CA/CA557/03 (FJ904715).

GenBank accession numbers for reference genome sequences used in this study are: TCoV/VA-74/03 (GQ427173); TCoV/TX-GL/01 (GQ427174); TCoV/TX-1038/98 (GQ427176); TCoV/IN-517/94 (GQ427175); TCoV/IN-540/94 (EU022525); TCoV/MN-ATCC (EU022526); TCoV/Canada-MG10

Table 1
Codon-based test (Z-test) of positive selection for S1 gene sequences.

Isolate	P value ^a					
	CAL99/NE15172/95	CAL99/CAL99/99	CAL/CA557/03	CA/CA1737/04	CAV/CAV56b/91	CAV/CAV9437/95
CAL99/NE15172/95						
CAL99/CAL99/99	0.00					
CAL/CA557/03	0.00	1.00				
CA/CA1737/04	0.00	0.01	0.00			
CAV/CAV56b/91	0.00	0.00	0.00	0.00		
CAV/CAV9437/95	0.00	0.00	0.00	0.00	1.00	
		IBV/Conn/66	IBV/Conn/72	IBV/Conn/83		IBV/Conn/91
IBV/Conn/66						
IBV/Conn/72	0.00					
IBV/Conn/83	0.00		0.15			
IBV/Conn/91	0.00		0.15	0.07		
		IBV/Mass/65	IBV/Mass/72	IBV/Mass/79	IBV/Mass/85	IBV/Mass/06
IBV/Mass/65						
IBV/Mass/72	0.17					
IBV/Mass/79	0.11	0.09				
IBV/Mass/85	0.16	0.08	0.08			
IBV/Mass/06	1.00	1.00	1.00	1.00	1.00	

^a P values <0.05 (boldface) indicates that the null hypothesis of strict-neutrality (dN=dS) is rejected in favor of the alternative hypothesis of dN>dS indicating positive selection.

(EU095850); IBV/Mass 41/41 (AY851295); IBV/Ark/Arkp11 (EU418976); IBV/Ark/Arkp101 (EU418975); IBV/CK/Beaudette (NC001451); IBV/CK/H120 (FJ888351); IBV/CAL99/CAL99/99 (AY514485); IBV/CK/CH/LSD/051 (EU637854); IBV/CK/SAIBK (DQ288927); IBV/CK/BJ (AY319651); PartridgeCoV/GD/S14/03 (AY646283); PeafowlCoV/GD/KQ6/03 (AY641576); IBV/CK/A2 (EU526388); Beluga Whale/CoV/SW1/08 (NC010646); Thrush-CoV/HKU12/09 (FJ376621); MuniaCoV/HKU13/09 (FJ376622); Bulbul/CoV/HKU11/09 (FJ376619).

3.2. Full-length genome sequence and organization

The full-length genome sequence of 11 IBV isolates from the field obtained over a 41 year period (5 Mass serotypes), 25 year period (4 Conn serotypes) and 8-year period (2 CA serotypes) were sequenced and the virus genome lengths are; IBV/Mass/65, 27, 474 nt; IBV/Mass/72, 27, 474 nt; IBV/Mass/79, 27, 614 nt; IBV/Mass/85, 27, 474 nt; IBV/Mass/06, 27, 660 nt; IBV/Conn/66, 27, 629 nt; IBV/Conn/72, 27, 632 nt; IBV/Conn/83, 27, 632 nt; IBV/Conn/91, 27, 632 nt; Cal99/NE15172/95, 27, 664 nt; CA/CA557/03, 27, 615 nt (Supplementary Table 1). Genomic libraries and RT-PCR fragments from overlapping specific primer pairs were used to assemble genomic sequences with SeqMan Pro and genome annotation was performed using SeqBuilder (DNASTAR, Inc., v.8.0.2, Madison, WI), to reveal a typical gamma coronavirus gene order for all of the viruses; 5' UTR, ORF1a/ab, spike glycoprotein, ORF3a, ORF3b, envelope protein, membrane protein, ORF 4b, ORF 5a, ORF 5b, nucleocapsid protein, ORF 6b, 3' UTR and a poly(A) tail.

The full-length genome sequences were aligned using ClustalW and all the phylogenetic trees constructed with the Neighbor-Joining method, Minimum Evolution method, Maximum Parsimony method, and UPGMA using MEGA 4.0.2 had similar topography (Tamura et al., 2007). A representative tree with the 11 viral genomes sequenced herein and other IBV full-length genomes available in GenBank is presented in Fig. 1. The nucleotide sequence similarities between the genomes of the viruses sequenced herein are from 93.8% to 94.9% for the CAL isolates, 99.8% to 99.9% between the Conn isolates and 92.4 to 100% between the Mass isolates.

3.3. Phylogenetic analysis of untranslated regions (UTRs) and individual ORFs

To determine the relationship between different virus types for sequences in specific areas across the genome, phylogenetic analysis of the nucleotide sequences of the 5' and 3' UTRs and the amino acid sequences of individual ORFs was conducted (Fig. 2). The spike glycoprotein, which is used to define IBV isolate type, shows individual groupings for Conn, Mass and CAL viruses with the exception of the IBV/Mass/06 virus, which is positioned as an offshoot of the Conn group. Similar groupings were observed for the envelope and nsp 3 (PL2) protein trees, whereas IBV/Mass/06 clearly grouped with Conn type viruses in the membrane, nucleocapsid, nsp 5 (Mpro) and nsp 11/12 (RdRp) protein trees (Fig. 2). A BLASTp search showed the IBV/Mass/06 virus spike glycoprotein to have 98% similarity to H120 (Acc. No. ACQ55230), which is a commonly used Mass type vaccine. A BLASTp search and phylogenetic analysis of the other Mass spike glycoproteins identified Mass type attenuated viruses (Acc. Nos. EU283075, EU283081, EU283082), which grouped together (data not shown). A BLASTp search and phylogenetic analysis of the Conn spike glycoproteins identified Conn attenuated viruses (Acc. Nos. EU283059, EU283061 and EU283062), which also grouped together (data not shown).

The phylogenetic analysis of the 5' UTR showed individual groupings for Conn, Mass, and CAL viruses, however the IBV/Mass/06 virus grouped with the CAL viruses (Fig. 2). In the 3' UTR, the CAL/CA557/03 is an offshoot of the Mass group, the IBV/Mass/06 virus is an offshoot of the Conn group, and Cal99/NE15172/95 was clearly distinct from the other viruses. A BLAST analysis of the Cal99/NE15172/95 virus 3' UTR showed that it was 99% identical to the DE072 strain of IBV (Acc. No. AF203002), a Delaware type virus. Although the 3' UTRs sequence divergence was as high as 13.5% among the viruses, sequence alignment of the stem-loop (s2m) region in the 3' UTR revealed identical sequences among Cal99/NE15172/95, IBV/Conn/66, IBV/Conn/72, IBV/Conn/83, and IBV/Conn/91, one nucleotide difference (nt. position 9: A to G) in the CAL/CA557/03 and IBV/Mass/79 viruses, and two nucleotide differences (nt. positions 9: A to G and 26: A to C) in the IBV/Mass/65, IBV/Mass/72 and IBV/Mass/85 viruses while IBV/Mass/06 had an major A and a minor G at nt. position 36 (data not shown).

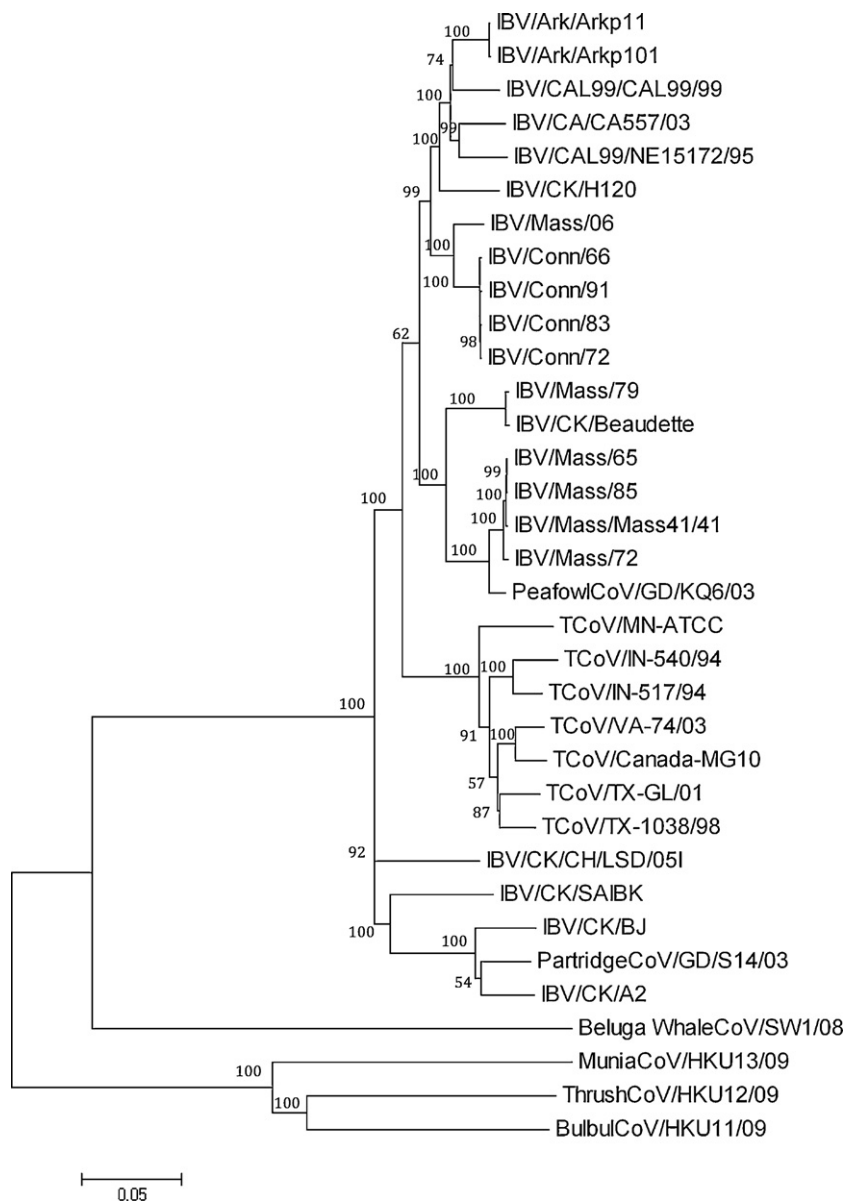


Fig. 1. Infectious bronchitis virus full-length genomic sequences were aligned with ClustalW and the phylogenetic trees were constructed with the Neighbor-Joining method. The optimal tree with the sum of branch lengths is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances. A bootstrap consensus tree was constructed from 1000 replicates (percentage of replicate trees in which associated strains clustered together are presented at nodes). The *p*-distance scale is presented at the bottom of the figure. The analyses were constructed using MEGA4 (Tamura et al., 2007).

3.4. Recombination events detected in the Mass06 isolate

To further examine the relationship of the IBV/Mass/06 virus with the other viruses examined in this study as well as selected IBV genomes available in GenBank we conducted a Recombination Detection Program (RDP) analysis, which indicated that the IBV/Mass/06 isolate was a recombinant with Conn and CAL viruses as the potential major parents (data not shown). The recombination events in IBV/Mass/06 were identified using Simplot, which showed that the IBV/Mass/06 genome shared a high sequence similarity, 97–100%, with CAL and Conn isolates in different regions of ORF1ab, with Mass in the S1 gene and with Conn isolates in the remainder of the 3' end of the genome (Fig. 3A). At the extreme 3' end of ORF1ab, between nucleotides 15,500 and 19,800, the IBV/Mass/06 virus did not show any significant sequence similarity with any of the genomes in this study. A BLAST analysis of that region showed 99% similarity with IBV strain H120 (Acc. No. FJ807652). The Simplot bootscan analysis revealed a total

of seven breakpoints located at approximate nucleotide positions 1530, 5895, 8188, 11,200, 12,266, 19,918 and 21,975 in the IBV/Mass/06 virus (Fig. 3B). The sequences located between the breakpoints were designated regions I–VIII, respectively. Regions one through five were located in nsp2, nsp3, nsp4, nsp9 and nsp11/12 respectively. Region VII contained the entire S1 glycoprotein gene, whereas the S2 glycoprotein gene was in region VIII.

Evidence of recombination in the IBV/Mass/06 isolate was further confirmed when phylogenetic trees constructed from the regions delineated by the breakpoints showed incongruence (Fig. 4). The IBV/Mass/06 isolate grouped with Mass type viruses in region VII, which contained the S1 gene. However, IBV/Mass/06 grouped with CAL type viruses in regions I, III and V, and with Conn type viruses in regions II, IV and VIII. In region VI, IBV/Mass/06 did not group with any of the genomes included in this study. As indicated above, a BLAST search of the IBV/Mass/06 virus sequence in that region showed that it was most closely related to the H120 Mass type vaccine virus.

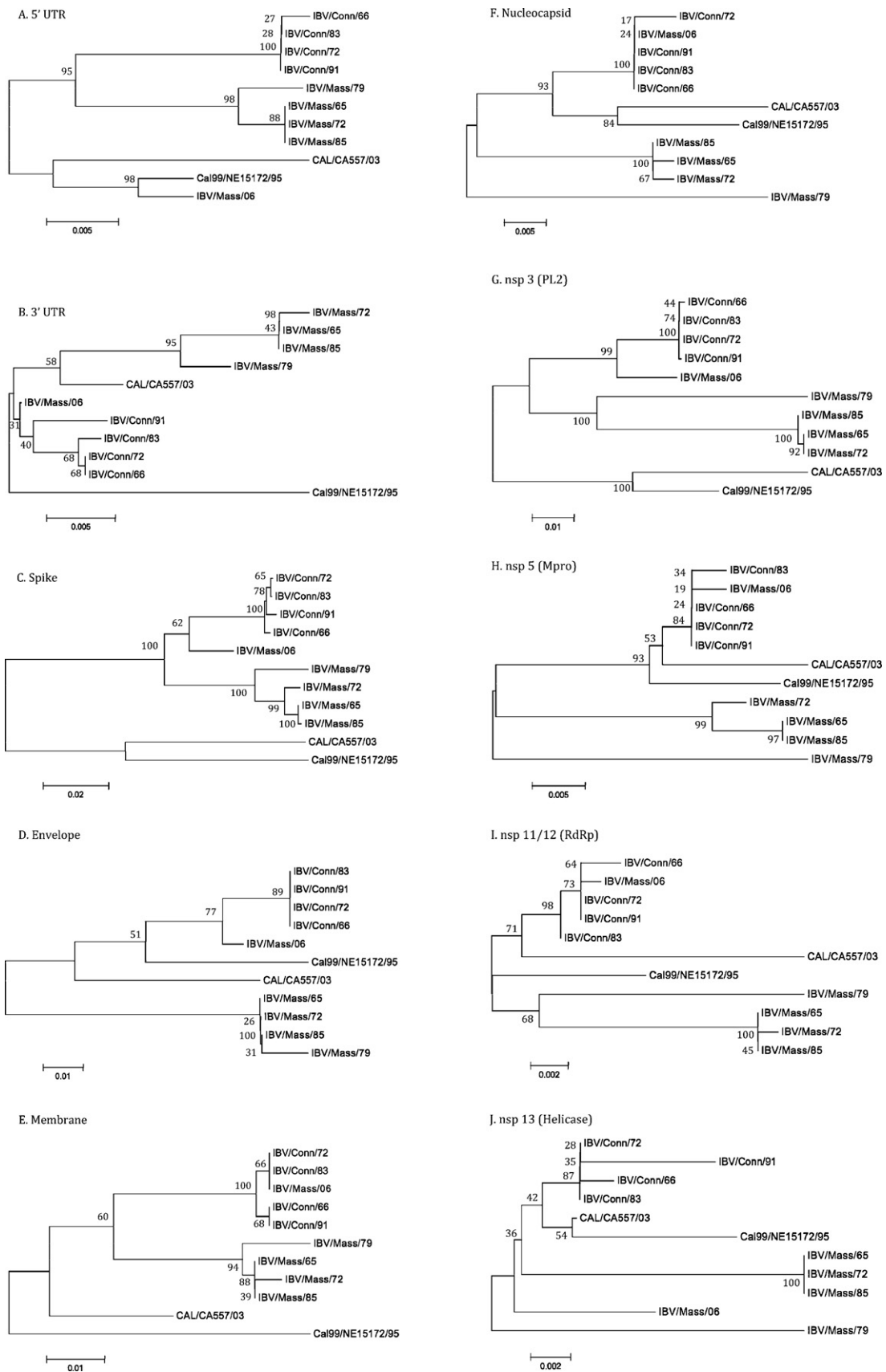


Fig. 2. Infectious bronchitis virus amino acid sequences A through J as indicated in the figure, were aligned with ClustalW and the phylogenetic trees were constructed using the Neighbor-Joining method. The optimal trees with the sum of branch lengths are shown. The trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances (the *p*-distance scale is presented at the bottom of each tree). Bootstrap consensus trees were constructed from 1000 replicates and the percentage of replicate trees in which associated strains clustered together are presented at the nodes. The analyses were conducted using MEGA4 (Tamura et al., 2007).

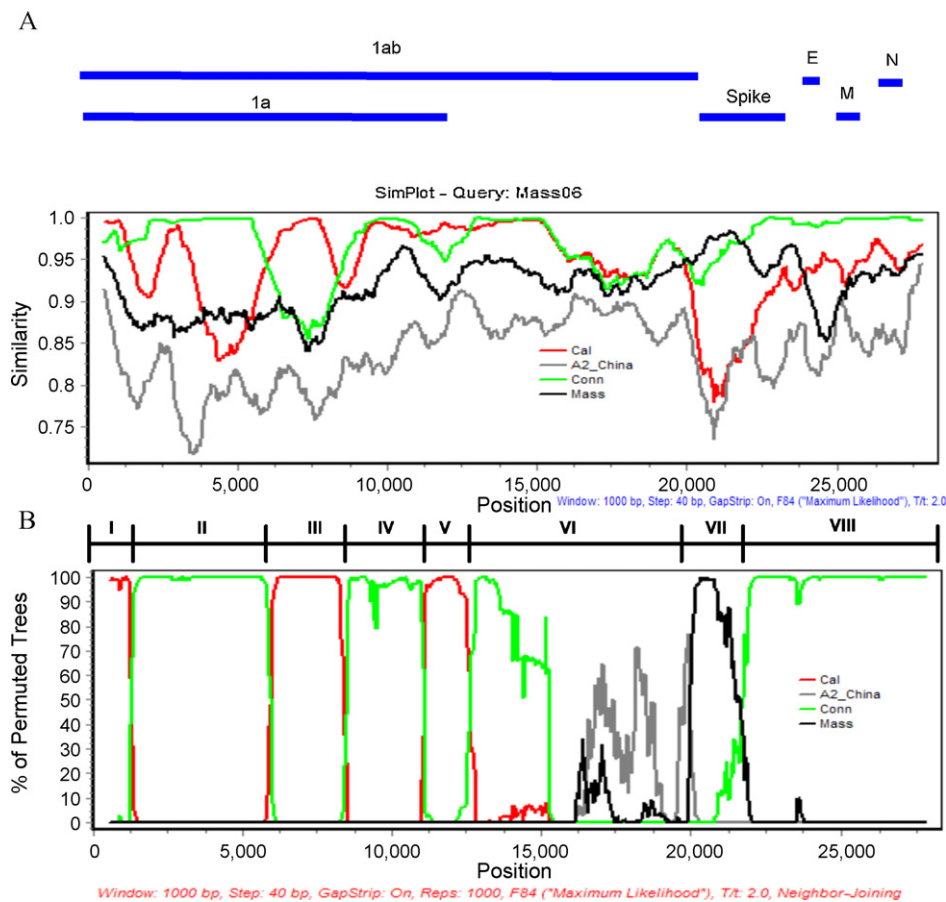


Fig. 3. Simplot analysis to detect recombination and estimate breakpoint within the IBV/Mass/06 genome. (A) Relative gene location and similarity plot using IBV/Mass/06 as the query sequence. (B) Bootscan analysis showing multiple breakpoints designated as regions I–VIII. Complete genome nucleotide sequences used in the analysis were Cal = Cal99/NE15172/95, IBV/CK/CA99/99, CA/CA557/03, Conn = IBV/Conn/66, IBV/Conn/72, IBV/Conn/83, IBV/Conn/91, and Mass = IBV/Mass/65, IBV/Mass/72, IBV/Mass/79, IBV/Mass/85. The IBV/CK/A2 China strain was used in the analysis as an out-group. Both analyses were performed with the F84 distance model, window size of 1000 bp and step size of 40 bp.

3.5. Mutation rates

A codon-based test of positive selection (Z-test, MEGA4), which measures the numbers of non-synonymous and synonymous substitutions per site (dN/dS ratio) showed that positive selection was occurring in the CAL viruses and between IBV/Conn/66 and other Conn viruses but not in Mass type viruses (Table 1). The BEAST v1.4.1 software, used to estimate the mutation rates for individual coding regions within each virus type, indicated that the CAL isolates had mutation rates that were 2–4 orders of magnitude higher than the Mass and Conn isolates (data not shown). Data for the full-length genomes and the S1 glycoprotein are presented in Table 2. The mutation rate for the ORF1ab, S1 and N gene was calculated to be 10^{-3} substitutions/site/year and 10^{-2} for the E and M genes from the CAL viruses, whereas the Conn and Mass viruses had mutation rates that were 10 fold lower and ranged from 10^{-4} to 10^{-6} in all of the coding regions except for the E gene in the Conn viruses, which did not show any polymorphisms.

4. Discussion

The majority of emerging infectious diseases are caused by RNA viruses. Presumably, high rates of mutation and rapid replication drive their evolution. Infectious bronchitis virus, like most RNA viruses, undergoes a high rate of genetic change that contributes to diverse subpopulations of the virus that continually emerge to form new variants of existing strains as well as completely new types of the virus. New variants and types of IBV are

defined by the sequence of the spike glycoprotein and/or by specific antibodies against spike, and for many IBV types, only spike sequences are available. To examine sequence changes not only in spike but also over the entire length of the genome, we sequenced the full-length genomes of 5 Mass type, 4 Conn type and 2 CAL type IBV field viruses that were isolated over a 41, 25 and 8 year period respectively. All of the viruses were obtained from chickens with clinical signs of the disease. Alignments of the full-length genome sequences showed that the Mass viruses had $\geq 92.4\%$ identity ($\geq 93.7\%$ if the recombined IBV/Mass/06 is not included), the CAL viruses had $\geq 93.8\%$ identity and the Conn viruses varied the least, at $\geq 99.8\%$ identity. Examination of the mutation rates for the ORF1ab, S1, E, M and N genes showed, in general, that the CAL viruses had rates between 10^{-4} and 10^{-2} substitutions/site/year, whereas the Conn and Mass viruses had lower mutation rates that ranged from 10^{-4} to 10^{-6} substitutions/site/year. In addition, the E gene in the Conn viruses showed no signs of change. Although the egg passage level of all the viruses used in this study was relatively low (≤ 15 passages) it is possible some genetic changes could have been introduced as a result of propagation in embryonated eggs. However phenotypic change is likely to be minimal since it typically takes 75 or more egg passages to attenuate IBV and in a study that examined genetic changes in the highly variable S1 gene following 10 passages in embryonated eggs, five viruses had no changes, two viruses had only 2 nucleotide changes one virus had only 1 change (McKinley et al., 2008).

It has been shown that vaccine usage can result in faster evolutionary rates in IBV field viruses (Lee and Jackwood, 2001; Wang

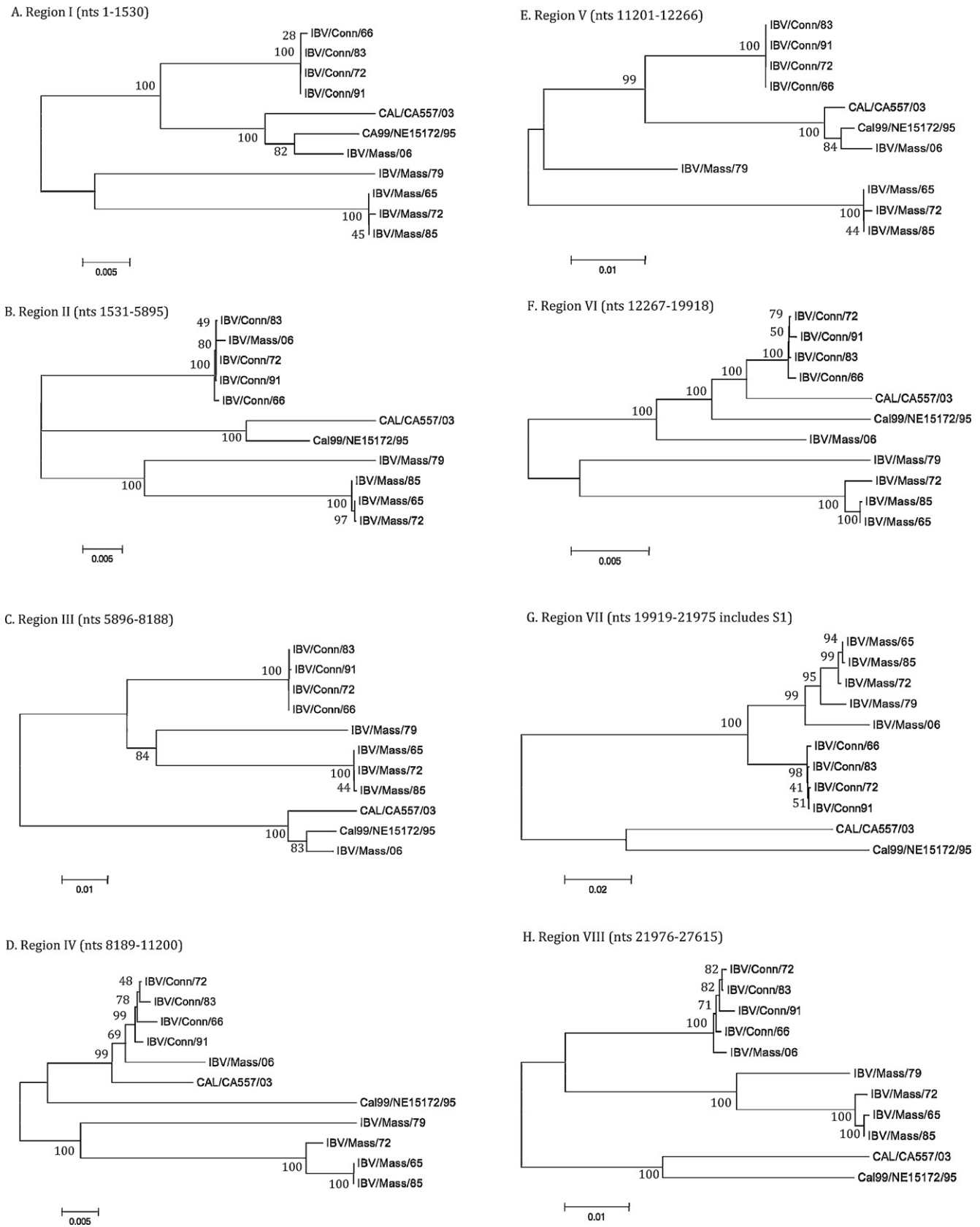


Fig. 4. The nucleotide sequences A through H as indicated in the figure, were aligned with ClustalW and the phylogenetic trees were constructed using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal trees with the sum of branch lengths are shown. The trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances (the *p*-distance scale is presented at the bottom of each tree). Bootstrap consensus trees were constructed from 1000 replicates and the percentage of replicate trees in which associated strains clustered together are presented at the nodes. The analyses were conducted using MEGA4 (Tamura et al., 2007).

Table 2
Mutation and evolutionary rates of the genome and S1 genes for California (CAL), Connecticut (Conn) and Massachusetts (Mass) isolates.

Virus type	Mutation rate (nucleotide changes/year)		Evolutionary rate (amino acid changes/year)	
	Genome ^a	S1 gene ^b	Genome ^a	S1 gene ^b
CAL	1.0×10^{-2}	9.4×10^{-4}	2.0×10^{-2}	2.4×10^{-3}
Conn	1.5×10^{-4}	1.7×10^{-4}	2.6×10^{-4}	3.5×10^{-4}
Mass	1.9×10^{-3}	2.4×10^{-4}	3.3×10^{-3}	5.8×10^{-4}

^a Viruses used in the genome analysis were the CAL, Conn and Mass viruses shown in Fig. 1.

^b Viruses used in the S1 gene analysis are the same viruses shown in Table 1.

et al., 1993). Thus, it was surprising that mutation rates for Mass and Conn type viruses were lower than mutation rates for CAL type viruses because live attenuated vaccines have been used against the Mass and Conn type viruses in poultry for many years, whereas no vaccine exists for the CAL type viruses. One possible explanation for the differences in mutation rates is that the Mass and Conn viruses were possibly re-isolated vaccine viruses that have been circulating in the field for an unknown but relatively short period of time, with the exception of the IBV/Conn/66 virus, which did show positive selection when compared to the other Conn viruses. In contrast, because no vaccine exists for CAL viruses, those viruses represent true field isolates that have been circulating for known periods of time.

The possibility that the origin of the Mass and Conn isolated viruses is vaccine, is supported by lower levels of genetic diversity observed for many of the genes within each virus type (Fig. 2) as well as the genetic similarity of the viruses, previously characterized as attenuated viruses. It has been shown that IBV can be shed for long periods of time and that vaccine viruses can persist in the field and revert to cause disease (Farsang et al., 2002; Jackwood et al., 2009; Matthijs et al., 2008; Naqi et al., 2003). Thus, it is possible that vaccine viruses are more likely to be isolated from the field because they may be more abundant having displaced field viruses of the same type or because they are more easily isolated since they are adapted to grow in embryonated eggs. Because the Mass and Conn type viruses were isolated from clinically diseased chickens suggests that if they were originally vaccine viruses then they have likely acquired the sequence changes necessary for pathogenicity. Further studies would need to be conducted on vaccine viruses and pathogenic strains to gain a better understanding of the molecular changes necessary for coronavirus reversion to pathogenicity.

Another explanation for lower mutation rates in Mass and Conn viruses could be that some of the mutations observed are incidental, meaning that they are not deleterious and provide no selective advantage to the virus rather than due to selection pressure. This possibility was suggested in another study where no evolution was detected in the 5' hypervariable region of the S1 gene for a number of 793/B type viruses isolated over a 11 year period (Cavanagh et al., 1998). This however does not explain why polymorphisms were observed for the CAL type viruses.

Recombination contributes to the genetic diversity of coronaviruses and can lead to the emergence of new viruses and outbreaks of new diseases (Holmes, 2009; Jia et al., 1995; Woo et al., 2009). Although the IBV/Mass/06 virus is identified as a Mass type virus based on characterization of the spike glycoprotein, it was shown herein to be a recombinant virus with Conn and CAL viruses as major parents. Eight regions defined by 7 breakpoints distributed across the entire genome were identified. Region VII was similar to other Mass type viruses, which was expected because that region includes the entire S1 gene. However, CAL like sequences were found in regions I, III and V, Conn sequences were found in regions II, IV and VIII, and region VI was found to be similar to the H120 Mass type vaccine strain. High frequencies of recombination including recombination events between vaccine and field strains have been reported for IBV (Brooks et al., 2004; Cavanagh

et al., 1992; Fang et al., 2005; Jackwood et al., 2010; Kottier et al., 1995b; Lee and Jackwood, 2000; Yu et al., 2001). In addition, IBV pathogenicity was shown to be polygenic, involving spike as well as replicase proteins (Ammayappan et al., 2009; Armesto et al., 2009; Cavanagh, 2007; Hodgson et al., 2004). Although IBV type is defined by the spike glycoprotein, our data shows that viruses within the same type can have vastly divergent sequences in other regions of their genome including regions that contain genes important for pathogenicity. If recombination occurs outside of the spike gene where a pathogenic field virus acquires attenuated vaccine virus sequences, the outcome would likely be a relatively benign virus regardless of the virus types (spike sequences) involved. Furthermore, if a vaccine virus circulating in a flock recombines with a pathogenic virus outside of the spike gene, resulting in the vaccine virus becoming pathogenic, the outcome is still likely to be of little or no consequence since the birds would presumably be immune to the unchanging spike. However, in the event that a pathogenic field virus recombines with either a vaccine virus or another pathogenic virus and acquires a new or unique spike glycoprotein gene, the outcome could be the emergence of a new virus capable of causing disease. This was recently shown for TCoV (Jackwood et al., 2010). Although new coronaviruses can emerge and cause disease as a result of recombination it appears that the emergence of most variant or new IBV types is due to accumulation of mutations in the spike glycoprotein over time (Cavanagh et al., 2005; Jackwood et al., 2005, 2007; Lee and Jackwood, 2001; Nix et al., 2000). Thus, until pathogenicity genes can be specifically identified, monitoring the spike gene in IBV appears to be the most reliable measure of genetic change leading to the emergence of new viruses capable of causing disease.

This is the first study to identify genetic changes and mutation rates using full-length genome sequences of IBVs isolated over many years. Our data show that mutation rates are lower and that positive selection was not occurring for IBV types where attenuated live vaccines are used compared to an IBV type where no vaccine exists, suggesting that the virus isolates examined were re-isolated vaccine viruses that have been circulating in the field for a relatively short-period of time. This observation is extremely important because it could lead to incorrect estimates of mutation rates for RNA viruses where modified live vaccines are used to control the disease. In this study, we also identified a recombinant IBV isolate with two major parents and no fewer than 7 breakpoints distributed across the entire genome. Although recombination among coronaviruses can contribute to genetic diversity of the virus and recombination has been shown to lead to the emergence of new viruses, it appears that monitoring mutations in the spike gene that accumulate over time is currently the best method of identifying potentially new genetic variants and IBV types capable of causing disease.

Acknowledgments

This research was supported by USDA, CSREES award number 2007-35600-17786.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.virusres.2011.04.006.

References

- Ammayappan, A., Upadhyay, C., Gelb Jr., J., Vakharia, V.N., 2009. Identification of sequence changes responsible for the attenuation of avian infectious bronchitis virus strain Arkansas DPL. *Arch. Virol.* 154, 495–499.
- Armesto, M., Cavanagh, D., Britton, P., 2009. The replicase gene of avian coronavirus infectious bronchitis virus is a determinant of pathogenicity. *PLoS One* 4 (10), e7384.
- Brooks, J.E., Rainer, A.C., Parr, R.L., Woolcock, P., Hoerr, F., Collisson, E.W., 2004. Comparisons of envelope through 5B sequences of infectious bronchitis coronaviruses indicates recombination occurs in the envelope and membrane genes. *Virus Res.* 100 (2), 191–198.
- Cavanagh, D., 2007. Coronavirus avian infectious bronchitis virus. *Vet. Res.* 38 (2), 281–297.
- Cavanagh, D., Davis, P.J., Cook, J., 1992. Infectious bronchitis virus – evidence for recombination within the massachusetts serotype. *Avian Pathol.* 21 (401), 401–408.
- Cavanagh, D., Gelb Jr., J., 2008. Infectious bronchitis. In: Saif, Y.M. (Ed.), *Diseases of Poultry*, 12th ed. Blackwell Publishing, Ames, Iowa, Ames, IA, pp. 117–135.
- Cavanagh, D., Mawditt, K., Adzhar, A., Gough, R.E., Picault, J.P., Naylor, C.J., Haydon, D., Shaw, K., Britton, P., 1998. Does IBV change slowly despite the capacity of the spike protein to vary greatly? *Adv. Exp. Med. Biol.* 440, 729–734.
- Cavanagh, D., Picault, J.P., Gough, R., Hess, M., Mawditt, K., Britton, P., 2005. Variation in the spike protein of the 793/B type of infectious bronchitis virus, in the field and during alternate passage in chickens and embryonated eggs. *Avian Pathol.* 34 (1), 20–25.
- Domingo, E., Holland, J.J., 1997. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51, 151–178.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Fang, S.G., Shen, S., Tay, F.P., Liu, D.X., 2005. Selection of and recombination between minor variants lead to the adaptation of an avian coronavirus to primate cells. *Biochem. Biophys. Res. Commun.* 336 (2), 417–423.
- Farsang, A., Ros, C., Renstrom, L.H., Baule, C., Soos, T., Belak, S., 2002. Molecular epidemiology of infectious bronchitis virus in Sweden indicating the involvement of a vaccine strain. *Avian Pathol.* 31 (3), 229–236.
- Gelb, J.J., Jackwood, M.W., 2008. Infectious bronchitis. In: Dufour-Zavala, L., Swayne, D.E., Glisson, J.R., Pearson, J.E., Reed, W.M., Jackwood, M.W., Woolcock, P. (Eds.), *A Laboratory Manual for the Isolation, Identification, and Characterization of Avian Pathogens*, 5th ed. American Association of Avian Pathologists, Kennett Square, PA, pp. 146–149.
- Hodgson, T., Casais, R., Dove, B., Britton, P., Cavanagh, D., 2004. Recombinant infectious bronchitis coronavirus Beaudette with the spike protein gene of the pathogenic M41 strain remains attenuated but induces protective immunity. *J. Virol.* 78 (24), 13804–13811.
- Holmes, E.C., 2009. The evolution and emergence of RNA viruses. In: Harvey, P.H., May, R.M. (Eds.), *Oxford Series in Ecology and Evolution*, 1st ed. Oxford University Press Inc., New York.
- Holmes, K.V., 1991. Coronaviridae and their replication. In: Fields, B.N. (Ed.), *Fundamental Virology*, 2nd ed. Raven Press, New York, pp. 471–486.
- Jackwood, M.W., Boynton, T.O., Hilt, D.A., McKinley, E.T., Kissinger, J.C., Paterson, A.H., Robertson, J., Lemke, C., McCall, A.W., Williams, S.M., Jackwood, J.W., Byrd, L.A., 2010. Emergence of a group 3 coronavirus through recombination. *Virology* 398, 98–108.
- Jackwood, M.W., Hilt, D.A., Lee, C.W., Kwon, H.M., Callison, S.A., Moore, K.M., Moscoso, H., Sellers, H., Thayer, S., 2005. Data from 11 years of molecular typing infectious bronchitis virus field isolates. *Avian Dis.* 49 (4), 614–618.
- Jackwood, M.W., Hilt, D.A., McCall, A.W., Polizzi, C.N., McKinley, E.T., Williams, S.M., 2009. Infectious bronchitis virus field vaccination coverage and persistence of Arkansas type viruses in commercial broilers. *Avian Dis.* 53, 175–183.
- Jackwood, M.W., Hilt, D.A., Williams, S.M., Woolcock, P., Cardona, C., O'Connor, R., 2007. Molecular and serologic characterization, pathogenicity, and protection studies with infectious bronchitis virus field isolates from California. *Avian Dis.* 51 (2), 527–533.
- Jia, W., Karaca, K., Parrish, C.R., Naqi, S.A., 1995. A novel variant of avian infectious bronchitis virus resulting from recombination among three different strains. *Arch. Virol.* 140 (2), 259–271.
- Kottier, S.A., Cavanagh, D., Britton, P., 1995a. Experimental evidence of recombination in coronavirus infectious bronchitis virus. *Virology* 213 (2), 569–580.
- Kottier, S.A., Cavanagh, D., Britton, P., 1995b. First experimental evidence of recombination in infectious bronchitis virus. *Recombination in IBV. Adv. Exp. Med. Biol.* 380, 551–556.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., DG, H., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Lee, C.W., Jackwood, M.W., 2000. Evidence of genetic diversity generated by recombination among avian coronavirus IBV. *Arch. Virol.* 145 (10), 2135–2148.
- Lee, C.W., Jackwood, M.W., 2001. Origin and evolution of Georgia 98 (GA98), a new serotype of avian infectious bronchitis virus. *Virus Res.* 80 (1–2), 33–39.
- Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 73 (1), 152–160.
- Matthijs, M.G., Bouma, A., Velkers, F.C., van Eck, J.H., Stegeman, J.A., 2008. Transmissibility of infectious bronchitis virus H120 vaccine strain among broilers under experimental conditions. *Avian Dis.* 52 (3), 461–466.
- McKinley, E.T., Hilt, D.A., Jackwood, M.W., 2008. Avian coronavirus infectious bronchitis attenuated live vaccines undergo selection of subpopulations and mutations following vaccination. *Vaccine* 26, 1274–1284.
- Mondal, S.P., Cardona, C.J., 2007. Genetic and phenotypic characterization of the California 99 (Cal99) variant of IBV. *Virus Genes* 34, 327–341.
- Moya, A., Holmes, E.C., Gonzalez-Candelas, F., 2004. The population genetics and evolutionary epidemiology of RNA viruses. *Nat. Rev. Microbiol.* 2 (4), 279–288.
- Naqi, S., Gay, K., Patalla, P., Mondal, S., Liu, R., 2003. Establishment of persistent avian infectious bronchitis virus infection in antibody-free and antibody-positive chickens. *Avian Dis.* 47 (3), 594–601.
- Niesters, H.G.M., 1987. *Molecular epidemiology of infectious bronchitis virus*. Ph.D., Netherlands.
- Nix, W.A., Troeber, D.S., Kingham, B.F., Keeler Jr., C.L., Gelb Jr., J., 2000. Emergence of subtype strains of the Arkansas serotype of infectious bronchitis virus in Delmarva broiler chickens. *Avian Dis.* 44 (3), 568–581.
- Rodriguez, F., Oliver, J.L., Martin, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142, 485–501.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24 (8), 1596–1599.
- Vijaykrishna, D., Smith, G.J., Zhang, J.X., Peiris, J.S., Chen, H., Guan, Y., 2007. Evolutionary insights into the ecology of coronaviruses. *J. Virol.* 81 (8), 4012–4020.
- Wang, L., Junker, D., Collisson, E.W., 1993. Evidence of natural recombination within the S1 gene of infectious bronchitis virus. *Virology* 192 (2), 710–716.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., Barton, G.J., 2009. Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.
- Woo, P.C., Lau, S.K., Huang, Y., Yuen, K.Y., 2009. Coronavirus diversity, phylogeny and interspecies jumping. *Exp. Biol. Med.* (Maywood).
- Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* 39, 306–314.
- Yu, L., Wang, Z., Jiang, Y., Low, S., Kwang, J., 2001. Molecular epidemiology of infectious bronchitis virus isolates from China and Southeast Asia. *Avian Dis.* 45 (1), 201–209.