

## UC Davis

### UC Davis Previously Published Works

**Title**

Genetic diversity of avian infectious bronchitis virus California variants isolated between 1988 and 2001 based on the S1 subunit of the spike glycoprotein

**Permalink**

<https://escholarship.org/uc/item/5dj7318k>

**Journal**

Archives of Virology, 148(1)

**ISSN**

0304-8608

**Authors**

Schikora, BM

Shih, LM

Hietala, SK

**Publication Date**

2003

**DOI**

10.1007/s00705-002-0904-5

Peer reviewed

**Genetic diversity of avian infectious bronchitis virus  
California variants isolated between 1988 and 2001  
based on the S1 subunit of the spike glycoprotein**

**B. M. Schikora, L. M. Shih, and S. K. Hietala**

California Animal Health and Food Safety Laboratory, University of California,  
Davis, U.S.A.

Received March 14, 2002; accepted August 9, 2002  
Published online October 24, 2002 © Springer-Verlag 2002

**Summary.** Twenty-nine isolates of avian infectious bronchitis virus (IBV) recovered from commercial chicken flocks in California between 1988 and 2001 and identified as California variants by serotype and direct automated cycle sequencing of the IBV spike glycoprotein S1 subunit, were further characterized phylogenetically and by nucleotide sequence comparison.

California variants were grouped according to production type of chicken, by comparison with public access sequence databases (NCBI GenBank and EMBL), or based on phylogenetic analysis. Fisher's Exact test was used to compare mutations per year, purifying and positive selection, predictive antigenicity, and a  $\geq 6$  bp deletion between California variant groups.

A high number of mutations at the nucleotide level ( $p = 0.013$ ) and a  $\geq 6$  bp deletion in the nucleotide sequence ( $p = 0.006$ ) was significantly associated with broiler-type chickens. However, 88% of significant comparisons at the amino acid level such as purifying and positive selection were seen in layer-type chickens. A pronounced predictive antigenicity in the HVR2 region was also associated with layer-type chickens ( $p = 0.001$ ). The study indicates that IBV in California is in a phase of slow evolution with different evolutionary patterns being associated with the production type of chicken.

### **Introduction**

Avian infectious bronchitis virus (IBV) is an economically important disease of chickens causing reduced performance, reduced egg quality and quantity, increased susceptibility to infections, decreased profit at slaughter, and costly vaccination programs (for review see [7]). IBV, a species in the genus *Coronavirus*,

family *Coronaviridae* is distributed worldwide [11]. Because of the international economic importance of one of the most important avian pathogens worldwide, the virus is included among the Office International des Epizooties list B diseases.

Vaccines have generally been effective in controlling the clinical disease associated with IBV infections, however escape mutants or variants continue to cause clinical disease and production problems in vaccinated flocks [38], and unique IBV variants may continue to circulate among poultry in geographically isolated areas [36]. The continuing appearance of new IBV variants is associated with the high rate of evolution, expressed as the accelerated rate at which viable mutations accumulate in the genome, which is due to a lack of proofreading enzymes seen in RNA viruses, nucleotide insertions, deletions, or point mutations in S1 as the result of the viral polymerase [15, 16], as well as the high rate of replication of the virus. Additionally, there is evidence that some IBV strains may have arisen by genetic recombination [4, 30, 32], mutation [31] or a combination of mutation and recombination [21].

Mutations in the antigenically important spike glycoprotein S1 subunit have been shown to lead to the emergence and proliferation of variant serotypes [36] associated with new disease outbreaks. Within the S1 subunit gene most of the antigenic epitopes are formed from a few amino acids encoded in 2 hypervariable regions, HVR1 [aa 54 to 68] and HVR2 [aa 116 to 141]. Numbering is based on the Mass41 S1 protein sequence (accession# M21883, GenBank [25]).

The focus of this retrospective study was the molecular epidemiology of the spike glycoprotein S1 subunit in California variants [2, 14] of IBV. The objective of the study was to compare the number of nucleotide mutations per year, codons, and amino acid sequences, to evaluate which mutations led to altered amino acids, and to predict antigenicity values in the S1 subunit among representative isolates recovered from different production types of chickens, between January 1988 and May 2001. Several statistical approaches were utilized in the analysis; including the Jukes and Cantor method [22] to estimate the number of nucleotide substitutions assuming equal frequencies of the nucleotides and accounting for the fact that more than one substitution can happen at one site, the Nei-Gojobory calculation to estimate the synonymous changes (nucleotide changes not leading to aa changes) and the number of non-synonymous changes (nucleotide changes leading to aa changes), a  $z$ -test for Darwinian selection to identify sequences favoring changes in aa sequence, a  $z$ -test for purifying selection to identify nucleotide changes not causing a change in the resulting aa sequence, and the Poisson correction distance to compare sequences on the aa level accounting for the fact that multiple substitutions can occur at one site.

## Material and methods

### *Sampling*

In a retrospective study, a convenience sampling of 125 avian infectious bronchitis virus isolates obtained from diagnostic case submissions to the California Animal Health and Food Safety Laboratory System in San Bernadino, Fresno, or Turlock, California were evaluated.

Viral isolates were serotyped by monoclonal antisera directed against the classical serotypes IBV-Conn, Mass41, and Ark99 [24], and by monospecific polyclonal antisera for the above viruses plus the more recent California variant [2, 14]. Genetic sequences of the IBV S1 subunit of each isolate were additionally compared to IBV S1 sequences in GenBank and EMBL databases. In addition to the above serotypes, isolates with nucleotide identity in the S1 region to the “Arkansas-like,” Holland, Pennsylvania and Delaware serotypes were identified. Isolates showing an S1 subunit sequence similarity of > 80% to the prototype California variants identified in GenBank (group 1 = GB# AF027510 and GB# AF027511; group 2 = GB# AF004738 and GB# AF0027508) were selected for further study. Twenty-nine isolates obtained from clinically affected broiler or broiler-breeder flocks ( $n = 12$ ) or layer flocks ( $n = 17$ ) between 1988 and 2001, were utilized in the cDNA sequencing and phylogenetic analysis.

#### *Virus isolation and propagation*

Avian infectious bronchitis viruses were isolated by standard egg inoculation techniques [1], and chorioallantoic membranes harvested 48–72 hr post inoculation at second passage were examined for the presence of IBV by a fluorescent antibody test [1]. Allantoic fluid harvested from second passage eggs described above was stored at  $-70^{\circ}\text{C}$  for further analysis or virus propagation.

#### *Rapid RNA extraction and RT-PCR amplification*

A rapid RNA extraction protocol was developed for IBV in allantoic fluid. Two hundred  $\mu\text{l}$  allantoic fluid was added to 1 ml Catrimox-14 surfactant<sup>TM</sup> (Qiagen, Valencia, CA) and 15  $\mu\text{l}$  of 1 mg/ml yeast t-RNA (Life Technologies), vortexed for 2 min, incubated at room temperature for 20 min, and the nucleic acid/surfactant complex was centrifuged at 14,000 rpm for 5 min. After the supernatant was discarded, pellets were carefully washed with 1 ml of sterile water, followed by addition of 1 ml of LiCl (2 M), vortexed for 2 min, and centrifuged for 5 min at 14,000 rpm. The pelleted RNA was washed with 1 ml of ice-cold 70% ethanol and vacuum-dried for 10 min at room temperature. RNA pellets were dissolved in 15  $\mu\text{l}$  of sterile distilled water, and kept at  $-20^{\circ}\text{C}$  until used. Reverse transcriptase-PCR was performed using a degenerate primer set CK4 and CK2 as described previously [25] to obtain the forward and the reverse strand sequence of the partial S1 spike glycoprotein (aa 43–236, based on the S-1 protein sequence of the Mass 41 strain (GB# 21883)). Standard PCR was performed in duplicate for each isolate using AmpliTaq Gold DNA Polymerase<sup>TM</sup> (PE Applied Biosystems Inc.) and 45 amplification cycles in a GeneAmp PCR System 9600 thermal cycler (PE Applied Biosystems Inc. Foster City, CA). Final elongation was performed by incubating samples at  $70^{\circ}\text{C}$  for 15 min. To verify amplification of the approximately 600 bp sequence, 10  $\mu\text{l}$  of each PCR product was electrophoresed on a 1.8% agarose gel and stained with ethidium bromide.

#### *Purification and direct cycle sequencing*

RT-PCR amplification products were purified using a Microcon-PCR column (Millipore Corp., Bedford, MA) according to the manufacturer’s recommendation. cDNA was quantified by optical density at 260 nm and 12 ng/ $\mu\text{l}$  was used as the sequencing template. Direct automated cycle sequencing was performed using an ABI Prism<sup>TM</sup> Big Dye Terminator Cycle Sequencing Ready Reaction kit (PE Applied Biosystems Inc.) on a 377 DNA sequencer, by a commercial service (Davis Sequencing Service, Davis, CA).

*Sequence and phylogenetic analysis*

Sequencing data were edited using commercial software (Chromas, version 1.45, for PC, Conor McCarthy, Griffith University, Australia). A consensus sequence was obtained for each isolate using data from duplicate runs including both sequencing directions. Sequences internal to the primers, having less than 99% reproducibility were not analyzed. Mutants were clustered into 2 groups. Group 1 contained all variants with > 80% sequence similarity to prior GenBank entries CV-9437 (GB# AF027510) and CV-1686 (GB# AF027511), and included the first IBV California variant isolated in this study (CA/2/88). Group 2 contained variants with > 80% sequence similarity to the original IBV California variant submitted to GenBank, CA 85-633 (GB# AF004738) and 1013 (GB# AF027508), and included most of the more recent IBV variants and an isolate also first identified in 1988 (prototype CA/1215/88). Computer analyses of nucleic acid and amino acid sequences were performed using the MacDNASIS Pro version 3.5 (Hitachi) and Sequencher 3.0 (Gene Codes Corp.) programs. Aligned sequences were cropped to a length of 426 bp (starting at 135 bp; numbering is based on the Mass41 S1 protein sequence (GB# M21883) [25]) to insure a consistent fragment analyzed within the genome segment sequenced. Two sequences (CA/96/01 and CA/627/01) were both included in the analysis despite an identical 426 bp sequence, because they showed distinct differences within the cropped portion of their sequences. Multiple sequence alignment was performed using the computer program Clustal W version 1.7 [40]. Molecular evolutionary analyses were conducted using computer program MEGA™ version 2.1 [29]. All evolutionary calculations were performed using the complete deletion option, due to sites involved in gaps that may evolve differently in different sequences. The sequence of CA/1215/88 was used as the prototype in this study because it was believed to be the first isolate identified as a California variant [2, 14] of IBV within the sampling period and also not showing the  $\geq 6$  bp deletion. Nucleotide mutations were calculated by dividing the number of nucleotide differences by the number of years between the isolation of the prototype CA/1215/88 and the individual variant noting that counted mutations/year is a non-linear relationship with time as some mutation sites may undergo repeated substitutions and mutation rate would be underestimated. The number of nucleotide substitutions was also estimated using the Jukes and Cantor method, which corrects for multiple substitutions at the same site  $[-3/4\ln(1-4/3p)]$ , where  $p$  = proportion of sites with different nucleotides [22]. The R-ratio of nucleotide substitution was defined as the number of transitions (a substitution of a purine for another purine, or a substitution of a pyrimidine for another pyrimidine) to the number of transversions (substitutions between pyrimidines and purines).

To get an impression of the percentage of silent mutations (synonymous) compared to the percentage of mutations leading to aa changes (non-synonymous) and therefore functional or structural differences, evolutionary differences based on the number of synonymous or non-synonymous sites were computed by comparing the prototype sequence with variant sequences on a codon-by-codon basis using the modified Nei-Gojobori method [19]. After counting the number of differences ( $S_d$ ,  $N_d$ ), these counts were adjusted by the number of potential synonymous/non-synonymous sites ( $p$ -distance;  $p_S$ ,  $p_N$ ), and  $p$ -distances then were corrected by the Jukes Cantor method to account for multiple substitutions at the same site ( $d_S$ ,  $d_N$ ). Positive Darwinian selection, defined as selection favoring non-synonymous substitutions, or a change in the aa, was tested using a one tailed  $z$ -test for large sample sizes [ $z = (d_N - d_S)/\text{Sqrt}(\text{Var}(d_S) + \text{Var}(d_N))$ ] with a level of significance of 5% testing the hypothesis of  $d_N > d_S$  ( $d_N$  = number of non-synonymous substitutions per non-synonymous sites,  $d_S$  = number of synonymous substitutions per synonymous sites). Purifying selection, defined as selection in favor of synonymous substitutions or mutations that do not result in a different aa, was performed testing the hypothesis of  $d_S > d_N$ . At the amino acid (aa) level the number of mutations per year was calculated by dividing the number of aa differences by

the number of years between the isolation of the prototype CA/1215/88 and the individual variant. For a more accurate impression of the aa sequence changes, evolutionary distances were computed on a residue-by-residue level. To adjust for multiple substitutions at the same site assuming equality of substitution rates and equal amino acid frequencies, the Poisson correction distance [ $d = -\ln(1 - p)$ , where  $d$  = number of amino acids per site, and  $p$  = proportion of different amino acids between 2 sequences] was used.

A rooted phylogenetic tree was constructed using a non-poultry coronavirus (Feline Infectious Peritonitis virus (FIPV) GenBank acc# X06170) to visualize relationship among CA isolates using the neighbor-joining method [39] based on the aa sequence with the Poisson correction distance. Bootstrap interior branch test was carried out on 1,000 replicate data sets to assess the confidence limits of the branch pattern [9]. A bootstrap support value of  $\geq 70\%$  was considered significant. A rooted tree assumes a direction that corresponds to evolutionary time; the closer a node is to the root of the tree, the earlier it appeared.

For comparison an unrooted tree was constructed to show the relationship between the CA variants themselves using the neighbor-joining method based on the aa sequence with the Poisson correction distance (data not shown). In unrooted trees, adjacent sequences are not assumed to be evolutionarily closely related.

#### *Predictive antigenicity*

Sequences were tested for predictive antigenicity (<http://www.embl-heidelberg.de/JaMBW/3/1/7/>) to evaluate possible associations between virus mutations at specific sites and antigenic alterations. The method of Hopp and Wood [18] was used to design plots of hydrophilicity value versus sequence position. The point of highest local hydrophilicity is believed to be part of or neighboring an antigenic determinant [18]. Predictive antigenicity was tested in 4 areas of the sequence. A sequence containing at least 2 peaks displaying a hydrophilicity value of 0.5 [23] on the EMBL-measurement scale, or one hydrophilicity peak of 0.75 in the HVR1 area [14 aa] was considered to be positive for predictive antigenicity. Sequences showing at least 2 peaks of hydrophilicity of 0.75 height, or one hydrophilicity peak of 0.75 plus 2 hydrophilicity peaks of 0.5, or 4 peaks of hydrophilicity of 0.5 in the HVR2 area [32 aa] were also considered as being positive. One hydrophilicity peak of 0.5 within nine aa upstream of HVR1 or ten aa upstream of HVR2 was counted positive for predictive antigenicity.

#### *Glycosylation analysis*

Sequences were tested for putative O- and N-glycosylation sites [13] to evaluate possible differences among the sequences (<http://www.cbs.dtu.dk/services/NetOGlyc/>; <http://www.cbs.dtu.dk/services/NetNGlyc/>).

#### *Statistical analysis*

Statistical analyses of proportions, other than the  $z$ -test performed with MEGA™ 2.1, were performed with SPSS (Standard Version) using a Fisher's exact test. Results with a  $p$ -value of  $\leq 0.05$  were considered significant.

### **Nucleotide sequence accession numbers**

The nucleotide sequences reported here have been deposited with the public access database GenBank. The accession numbers are shown in Table 1. A standard nomenclature was adopted for submission of isolates to GenBank that included a two-digit alphabetic identifying geographical origin of the isolate followed by

**Table 1.** California variant GenBank accession numbers

Identification number	IBV variants	GenBank accession numbers	Production type <sup>1</sup>	No. of nt-differences <sup>2</sup>	% of aa-differences <sup>2</sup>
1	CA/1215/88	AF488301	L	0	0
2	CA/2/88	AF488294	B	65	29
3	CA/179/94	AF488285	B	67	29
4	CA/1601/94	AF488286	B	94	39
5	CA/4065/95	AF488284	B	68	28
6	CA/100/97	AF488282	B	68	28
7	CA/157/97	AF488287	B	106	42
8	CA/452/97	AF488283	B	67	28
9	CA/1489/97	AF488281	B	66	28
10	CA/5069/97	AF488288	B	107	47
11	CA/376/98	AF488277	L	44	26
12	CA/516A/98	AF488278	L	44	25
13	CA/561/99	AF488293	L	96	31
14	CA/588/99	AF488275	B	86	29
15	CA/3402/99	AF488296	L	39	23
16	CA/3760/00	AF488289	B	113	48
17	CA/3761/99	AF488290	B	61	30
18	CA/5442/99	AF488291	L	44	25
19	CA/37/00	AF488292	L	96	41
20	CA/354/00	AF488295	L	89	35
21	CA/1030/00	AF488297	L	118	49
22	CA/1031/00	AF488298	L	97	43
23	CA/1032/00	AF488299	L	63	28
24	CA/1033/00	AF488300	L	101	43
25	CA/96/01	AF469167	L	73	29
26	CA/291/01	AF488279	L	60	25
27	CA/389/01	AY038997	L	72	30
28	CA/541/01	AF488280	L	77	29
29	CA/627/01	AF488276	L	77	29

<sup>1</sup>Production type of the chicken: L = layer, B = broiler

<sup>2</sup>Number of differences compared to the prototype CA/1215/88

a numeric or alphanumeric unique laboratory identifier, and a two digit numeric identifying the year of isolation, taking the form CA/1215/88.

## Results

### *Nucleotide sequence comparison*

Twenty-nine sequences of 426 bp each including the 2 HVRs in the S1 subunit gene of IBV were aligned and analyzed. The GC content in the sequences ranged from 35.5% to 39.5% with a mean of 37.3%. Because AT richness in the 3<sup>rd</sup> codon position might increase the apparent rate of aa changes [20], the relative

frequency of the four nucleotides in the third codon position was counted and resulted in a mean AT richness of 76.8% (range 65.9% to 81.8%). The average number of mutations in the 426 bp nucleotide sequence per year ranged from 3.55 (CA/3402/99: 39/11 [nt differences compared with the prototype]/[number of years between isolation of prototype and variant]) to 15.67 with a mean of 7.92. The number of mutations/year for a single isolate, CA/2/88, could not be calculated, because the virus was recovered in the same year as the prototype. Sixty-five nucleotide differences were counted between these two sequences. The average number of nucleotide mutations for isolates obtained from layer birds was 6.23 (CA/1215/88 not included), whereas the mean mutation number in broiler and broiler-breeder birds was calculated as 9.65 per year (CA/2/88 not included). The isolates within GenBank group 1 showed an average mutation number of 8.38 per year, whereas those in group 2 showed 6.52 mutations per year. The Jukes Cantor evolutionary distance ranged between the individual variants and prototype CA/1215/88 from 0.109 to 0.395 and a mean value of 0.238 (Fig. 1), with the larger distance values between two sequences corresponding to less similarity. Figure 1 displays the evolutionary distances among all variants studied. The R-ratio (transition/transversion) ranged from 0.47 to 1.0 between the prototype and the variants with a mean ratio of 0.738 with a low R-ratio indicating an overall slow evolutionary rate. Sequence alignment of the variants revealed a 6, or greater than or equal to 9 basepair deletion starting at position 174. At the corresponding location, AATACA, AATGCA, AATTCA, AATGCG, or AATCCA were seen in sequences not showing the deletion, except in CA/3760/99, which showed CGGGCT (Table 2). A second, 3 bp deletion site was detected at position 234 in CA/1601/94, CA/157/97, CA/5069/97, CA/1030/00, CA/1031/00, and CA/1033/00. All other sequences had the nucleotides GCA in the described location, except CA/3760/99 with TTA, and CA/37/00 with GGC.

#### *Positive selection*

To detect a positive Darwinian selection at the amino acid level, it is necessary to show that the number of aa changes is significantly greater than that of silent mutations for the whole segment studied. Nine variants CA/376/98, CA/516A/98, CA/1601/94, CA/157/97, CA/5096/97, CA/37/00, CA/3402/99, CA/1030/00, and CA/10/31/00 were found to have excess of non-synonymous changes, leading to replacement of aa. The large sample  $z$ -test ( $d_N > d_S$ ) was significant for CA/1030/00 ( $p = 0.039$ ), which was the only IBV variant revealing a predisposition to altered aa composition compared with the prototype CA/1215/88. Layers and broilers, if considered separately, showed the same evolutionary distance (mean value for aa changes compared to silent mutations) of 0.047. GenBank group 1 variants showed a mean value of 0.079 (indicating excess of mutations leading to aa changes), whereas GenBank group 2 variants displayed a mean distance value of  $-0.009$  (indicating excess of silent mutations).

The  $z$ -test calculation testing the positive selection hypothesis in the HVR1 region was significant only for a single isolate, CA/354/00 ( $p < 0.035$ ). The



	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]
[11]	0.19	0.20	0.30	0.20	0.20	0.20	0.34	0.20	0.20	0.35	0.12	0.12	0.30	0.27	0.11	0.37	0.18	0.12	0.30	0.28	0.40	0.31	0.19	0.32	0.22	0.18	0.22	0.23	0.23
[2]	0.27	0.03	0.17	0.03	0.05	0.23	0.05	0.04	0.30	0.18	0.19	0.25	0.22	0.19	0.35	0.17	0.18	0.24	0.22	0.22	0.32	0.20	0.04	0.17	0.17	0.20	0.16	0.18	0.18
[3]	0.27	0.05	0.16	0.01	0.02	0.20	0.02	0.03	0.30	0.19	0.19	0.25	0.21	0.18	0.34	0.16	0.19	0.23	0.21	0.32	0.20	0.03	0.16	0.14	0.19	0.14	0.17	0.16	0.16
[4]	0.38	0.26	0.22	0.17	0.19	0.34	0.18	0.18	0.28	0.29	0.30	0.32	0.34	0.30	0.37	0.31	0.29	0.29	0.37	0.14	0.26	0.19	0.32	0.30	0.30	0.30	0.30	0.32	0.32
[5]	0.26	0.05	0.02	0.24	0.02	0.20	0.02	0.03	0.29	0.20	0.20	0.20	0.25	0.21	0.19	0.35	0.17	0.20	0.23	0.20	0.32	0.20	0.03	0.15	0.14	0.18	0.14	0.16	0.15
[6]	0.26	0.07	0.03	0.24	0.03	0.19	0.01	0.02	0.31	0.19	0.20	0.20	0.25	0.21	0.18	0.35	0.17	0.19	0.23	0.21	0.34	0.22	0.05	0.17	0.14	0.19	0.14	0.16	0.15
[7]	0.42	0.29	0.25	0.39	0.24	0.22	0.18	0.19	0.20	0.37	0.38	0.41	0.38	0.41	0.38	0.36	0.34	0.37	0.32	0.40	0.25	0.13	0.21	0.23	0.32	0.34	0.32	0.32	0.33
[8]	0.26	0.07	0.03	0.24	0.03	0.01	0.21	0.02	0.30	0.19	0.19	0.25	0.21	0.17	0.34	0.17	0.19	0.23	0.20	0.33	0.22	0.04	0.17	0.14	0.18	0.14	0.14	0.16	0.15
[9]	0.26	0.07	0.03	0.24	0.03	0.02	0.22	0.02	0.29	0.18	0.19	0.24	0.21	0.18	0.35	0.17	0.18	0.22	0.21	0.32	0.20	0.04	0.18	0.15	0.19	0.15	0.17	0.17	0.17
[10]	0.48	0.44	0.42	0.42	0.41	0.41	0.27	0.41	0.41	0.42	0.43	0.44	0.42	0.43	0.40	0.40	0.40	0.41	0.29	0.48	0.16	0.18	0.30	0.31	0.40	0.35	0.39	0.41	0.40
[11]	0.24	0.26	0.24	0.41	0.25	0.25	0.47	0.25	0.25	0.56	0.01	0.30	0.27	0.07	0.39	0.17	0.01	0.27	0.24	0.41	0.33	0.19	0.33	0.20	0.23	0.20	0.21	0.21	0.21
[12]	0.23	0.26	0.24	0.41	0.25	0.25	0.47	0.25	0.25	0.56	0.03	0.30	0.26	0.08	0.39	0.18	0.02	0.28	0.24	0.41	0.34	0.19	0.33	0.20	0.23	0.20	0.21	0.22	0.22
[13]	0.29	0.26	0.26	0.33	0.25	0.25	0.41	0.25	0.25	0.48	0.31	0.30	0.09	0.31	0.43	0.32	0.30	0.38	0.33	0.43	0.37	0.27	0.40	0.28	0.27	0.28	0.30	0.29	0.29
[14]	0.27	0.21	0.21	0.33	0.20	0.20	0.36	0.20	0.20	0.46	0.27	0.27	0.11	0.27	0.42	0.28	0.26	0.33	0.27	0.46	0.39	0.22	0.34	0.23	0.24	0.23	0.25	0.25	0.25
[15]	0.21	0.24	0.24	0.42	0.23	0.23	0.44	0.23	0.23	0.55	0.09	0.10	0.30	0.27	0.38	0.18	0.08	0.28	0.24	0.41	0.34	0.18	0.32	0.19	0.23	0.18	0.20	0.20	0.20
[16]	0.49	0.48	0.46	0.46	0.47	0.47	0.48	0.47	0.47	0.51	0.48	0.49	0.48	0.48	0.51	0.22	0.38	0.40	0.46	0.37	0.36	0.34	0.40	0.39	0.40	0.39	0.42	0.40	0.40
[17]	0.28	0.22	0.20	0.39	0.21	0.21	0.38	0.21	0.21	0.52	0.26	0.27	0.33	0.28	0.26	0.28	0.17	0.29	0.25	0.42	0.32	0.16	0.28	0.20	0.24	0.20	0.22	0.21	0.21
[18]	0.23	0.26	0.24	0.39	0.25	0.25	0.47	0.25	0.25	0.55	0.02	0.04	0.31	0.27	0.10	0.47	0.26	0.28	0.24	0.40	0.40	0.32	0.18	0.32	0.20	0.24	0.20	0.22	0.22
[19]	0.41	0.32	0.30	0.42	0.29	0.30	0.42	0.30	0.30	0.43	0.35	0.35	0.42	0.35	0.36	0.52	0.36	0.37	0.17	0.31	0.30	0.24	0.30	0.10	0.14	0.10	0.12	0.12	0.12
[20]	0.33	0.26	0.26	0.48	0.25	0.26	0.48	0.26	0.26	0.61	0.28	0.28	0.36	0.29	0.29	0.59	0.30	0.30	0.17	0.51	0.40	0.22	0.35	0.07	0.11	0.07	0.08	0.08	0.08
[21]	0.51	0.44	0.43	0.22	0.42	0.43	0.29	0.43	0.42	0.23	0.55	0.55	0.44	0.44	0.54	0.49	0.54	0.54	0.46	0.62	0.12	0.30	0.31	0.42	0.42	0.41	0.44	0.44	0.44
[22]	0.43	0.31	0.30	0.37	0.29	0.31	0.17	0.31	0.30	0.28	0.46	0.46	0.38	0.39	0.43	0.51	0.46	0.46	0.44	0.51	0.16	0.18	0.21	0.32	0.32	0.32	0.34	0.34	0.34
[23]	0.26	0.07	0.06	0.25	0.06	0.07	0.27	0.07	0.05	0.42	0.24	0.24	0.24	0.19	0.22	0.46	0.22	0.24	0.32	0.28	0.38	0.25	0.13	0.15	0.18	0.15	0.17	0.16	0.16
[24]	0.43	0.24	0.23	0.43	0.21	0.24	0.30	0.24	0.23	0.46	0.41	0.41	0.39	0.33	0.38	0.54	0.36	0.41	0.43	0.46	0.42	0.28	0.17	0.27	0.31	0.26	0.29	0.28	0.28
[25]	0.27	0.21	0.19	0.39	0.18	0.19	0.39	0.19	0.19	0.51	0.22	0.22	0.29	0.23	0.23	0.51	0.24	0.24	0.11	0.07	0.54	0.43	0.21	0.37	0.05	0.01	0.03	0.01	0.01
[26]	0.23	0.28	0.26	0.39	0.25	0.25	0.42	0.25	0.25	0.46	0.27	0.27	0.28	0.25	0.28	0.52	0.31	0.29	0.17	0.12	0.52	0.42	0.25	0.42	0.07	0.05	0.07	0.06	0.06
[27]	0.28	0.22	0.20	0.38	0.19	0.20	0.39	0.20	0.20	0.49	0.23	0.23	0.30	0.24	0.24	0.52	0.26	0.25	0.09	0.07	0.52	0.43	0.22	0.37	0.02	0.07	0.03	0.02	0.02
[28]	0.27	0.21	0.21	0.38	0.20	0.20	0.41	0.20	0.20	0.52	0.22	0.22	0.29	0.23	0.23	0.51	0.24	0.24	0.13	0.08	0.54	0.43	0.21	0.37	0.03	0.08	0.03	0.03	0.03
[29]	0.27	0.21	0.19	0.39	0.18	0.19	0.39	0.19	0.19	0.51	0.22	0.22	0.29	0.23	0.23	0.51	0.24	0.24	0.11	0.07	0.54	0.43	0.21	0.37	0.00	0.07	0.02	0.02	0.03

**Table 2.** IBV variants not showing a six or  $\geq 9$  bp deletion in their nucleotide sequence alignment with their corresponding amino acids. The GenBank group assignment of the IBV variants are displayed in column 2

IBV variants	CA group assignment	nt 174–176	nt 177–179	nt 180–182	aa 58	aa 59	aa 60
CA/1215/88	group 2	AAT	GCA	GGT	asparagine	alanine	glycine
CA/376/98	group 2	AAT	GCA	GGT	asparagine	alanine	glycine
CA/516A/98	group 2	AAT	GCA	GGT	asparagine	alanine	glycine
CA/561/99	group 1	deletion	deletion	AGG	deletion	deletion	arginine
CA/588/99	group 1	deletion	deletion	AGG	deletion	deletion	arginine
CA/3402/99	group 2	AAT	GCA	GGT	asparagine	alanine	glycine
CA/3760/99	group 2	CGG	GCT	CAG	arginine	alanine	glutamine
CA/3761/99	group 2	AAT	GCG	GGC	asparagine	alanine	glycine
CA/5442/99	group 2	AAT	GCA	GGT	asparagine	alanine	glycine
CA/354/00	group 1	AAT	GCA	GGT	asparagine	alanine	glycine
CA/37/00	group 1	AAT	GCA	GGT	asparagine	alanine	glycine
CA/96/01	group 1	AAT	ACA	GGT	asparagine	threonine	glycine
CA/291/01	group 1	AAT	GCA	GGT	asparagine	alanine	glycine
CA/389/01	group 1	AAT	TCA	GGT	asparagine	serine	glycine
CA/541/01	group 1	AAT	GCA	GGT	asparagine	alanine	glycine
CA/627/01	group 1	AAT	ACA	GGT	asparagine	threonine	glycine

z-test calculations testing the predisposition to altered aa in only the HVR2 region (located between aa 116 and 141) were significant for 4 variants ( $p < 0.006$ ). Testing the same hypothesis for HVR1 and HVR2 together without omitting the genetically stable portion of the S1 sequence between the two regions, demonstrated significant results in 4 variants ( $p < 0.003$ ). There was a statistically significant association between the predisposition to altered aa in HVR1 and HVR2 combined (HVR1 + 2) and the grouping according to GenBank ( $p = 0.021$ ). Variants in GenBank group 2 were more likely to show predisposition to altered aa in HVR1 + 2, however there was no statistical association between predisposition to altered aa found in HVR1 + 2 and the production type of the chicken ( $p = 0.1$ ).

#### *Purifying selection*

Testing for a purifying selection, where number of silent mutations exceeds number of mutations leading to aa changes, no significant results were seen at the 5% level within only HVR1, only HVR2, or HVR1 + HVR2. Calculation of purifying



**Fig. 1.** S1 amino acid sequence alignment pair distances (Poisson correction) are displayed in the lower-left matrix, and S1 nucleic acid sequence alignment pair distances (Jukes-Cantor) are printed in the upper-right matrix. Larger distances between two sequences represent less similarity. Code numbers are displayed in Table 1

selection of the entire 426 bp segment, identified two IBV variants (CA/588/99 and CA/561/99) at the 5% level of significance. There was no statistical association between a variants showing a predominance of silent mutations and the production type of the chicken ( $p = 0.665$ ).

#### *Amino acid sequence comparison*

Amino acid sequences in the IBV variants had an average content of 59.05% non-polar residues and 40.95% polar residues, indicative of membrane bound proteins (proteins immersed in a non-polar environment requiring a hydrophobic exterior). Of all amino acids, glutamic acid was detected least frequently, with a mean of 0.99% (range 0% to 2.14%). Valine was detected most commonly with a mean of 9.33% (range 7.19% to 11.35%). The number of differences in the amino acid composition compared with the study prototype CA/1215/88, ranged from 23 (Poisson corrected distance: 0.207) to 49 (Poisson corrected distance: 0.508) with a mean value of 32.7 (Poisson corrected distance: 0.313). The average number of aa differences per year between the prototype and each variant ranged from 1.92 (CA/291/01) to 6.50 (CA/1601/94) with a mean value of 3.24 aa changes per year, and was not related to the time between isolation of the prototype and isolation of the variant. Evolutionary distances between the prototype and the variants as well as among the variants themselves are displayed in Fig. 1, where a larger distance between two sequences indicates less similarity. IBV isolates from broilers averaged 34 amino acid differences from the prototype, whereas IBV mutants from layers averaged 32 amino acid differences from the prototype. The average number of amino acid differences between the prototype and phylogenetic group was 32 (Poisson corrected distance: 0.298) for GenBank group 1 and 35 (Poisson corrected distance: 0.337) for GenBank group 2.

Though not part of the consensus sequence analyzed, it was noted that part of the 'HGGAY' aa motif within the reportedly conserved CK4 primer region described by Keeler et al. [25] was only observed in 18 of 29 isolates studied (Fig. 2). All replacements led to a larger total size of the amino acid in the motif, except the replacement combination found in CA/1215/88, suggesting a possible change in the three dimensional structure of the protein at this site. Two additional deletions were detected during the nucleotide alignment: one deletion ( $\geq 2$ aa) starting at aa position 78 and one deletion at aa position 121 (Fig. 2).

#### *Phylogenetic tree*

Based on the amino acid sequences with elimination of indels, a phylogenetic tree rooted to the coronavirus FIPV was established (Fig. 3). To show the relationship among the CA variants themselves, an unrooted tree based on the aa sequences was constructed and resulted in a bifurcating tree displaying two main branches (subgroup A and B), each of them further dividing. Subgroup A consists of all 16 variants having small evolutionary distances from the prototype as shown in the rooted tree. Subgroup B includes 13 variants with larger evolutionary

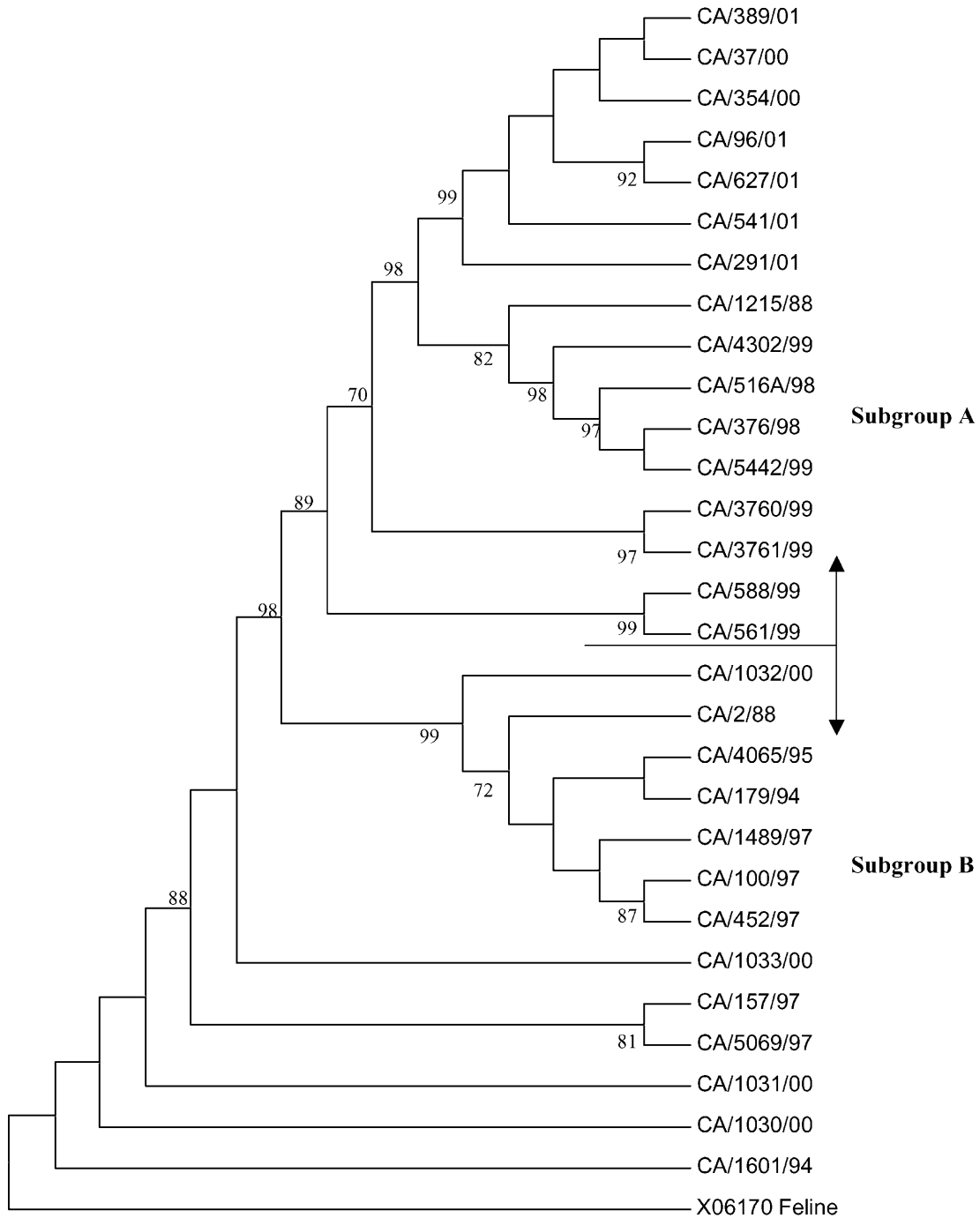
IBV variant	group <sup>a</sup>	antigenicity upstream HVR-1				HVR-1																													
		GAY		HVR-1																															
		upstream HVR-1 <sup>b</sup>	HVR-1 <sup>c</sup>	upstream HVR-2 <sup>d</sup>	HVR-2 <sup>e</sup>																														
CA/1215/88	2	1	2	2	2	68	C	T	A	G	A	I	V	W	S	K	N	F	S	A	A	S	V	A	M	T									
CA/2/88	1	1	2	2	1	2	47	E	S	N	A	G	S	A	S	Q	R	N	A	G	S	A	S	Q	R	N	A	G	S	A	S	Q			
CA/179/94	1	1	2	2	1	2	45	T	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
CA/1601/94	1	1	2	2	2	2	45	R	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
CA/4065/95	1	1	2	2	2	2	45	T	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
CA/100/97	1	2	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
CA/157/97	1	2	2	2	2	2	45	V	C	R	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
CA/452/97	1	2	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
CA/1489/97	1	2	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
CA/5069/97	1	1	2	2	2	2	45	V	H	M	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
CA/376/98	2	2	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CA/516A/98	2	2	2	2	2	2	45	W	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CA/561/99	1	2	2	2	2	2	45	L	Q	G	G	L	C	S	S	C	V	F	I	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CA/588/99	1	1	1	1	1	1	45	A	N	P	E	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CA/3402/99	2	2	2	2	2	2	45	R	M	R	L	M	F	L	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CA/3760/99	2	2	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA/3761/99	2	2	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA/5442/99	2	2	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA/37/00	2	2	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA/354/00	1	1	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA/1030/00	2	2	2	2	2	2	45	G	V	F	S	S	C	V	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CA/1031/00	2	2	2	2	2	2	45	G	V	F	S	S	W	S	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CA/1032/00	2	2	2	2	2	2	45	A	Y	S	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA/1033/00	2	2	2	2	2	2	45	A	Y	S	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA/96/01	1	1	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA/291/01	1	1	2	2	2	2	45	V	-	M	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CA/389/01	1	2	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA/541/01	1	1	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA/627/01	1	1	2	2	2	2	45	A	Y	A	V	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Fig. 2 (continued)

IBV variant	antigenicity upstream HVR-2										HVR-2																																						
	A	P	V	P	G	-	M	Q	W	S	F	T	N	L	V	V	F	V	T	H	C	F	V	P	G	N	G	H	C	P	L	T	G	L	V	P	Q	G	Y	I									
CA/1215/88	A	P	V	P	G	-	M	Q	W	S	K	S	Q	F	C	T	A	H	C	N	C	F	V	P	G	N	G	H	C	P	L	T	G	L	V	P	Q	G	Y	I									
CA/2/88	.	A	N	.	.	T	.	.	.	.	D	A	.	.	.	.	.	.	.	N	.	.	K	N	R	L	D	Y	.	.	.	.	.	.	.	.	.	.	.	I	.	N	.						
CA/179/94	.	A	N	.	G	.	T	.	.	.	D	A	.	.	.	.	.	.	.	N	.	.	K	N	R	P	N	Y	.	.	.	.	.	.	.	.	.	.	.	I	.	N	.						
CA/1601/94	.	A	N	.	G	.	T	.	.	.	D	A	.	.	.	.	.	.	.	N	.	.	K	K	L	N	Y	.	.	.	.	.	.	.	.	.	.	.	.	I	.	N	.						
CA/4065/95	.	A	N	.	.	T	.	.	.	.	D	V	.	.	.	.	.	.	.	N	.	.	K	N	.	P	N	Y	.	.	.	.	.	.	.	.	.	.	.	.	I	.	N	.					
CA/100/97	.	A	N	.	.	T	.	.	.	.	D	A	.	.	.	.	.	.	.	N	.	.	K	Y	R	P	N	Y	.	.	.	.	.	.	.	.	.	.	.	.	I	.	N	.					
CA/157/97	C	T	S	-	W	Y	D	M	V	.	G	C	T	I	L	.	.	.	.	N	.	.	K	Y	.	P	N	Y	.	.	.	.	.	.	.	.	.	.	.	.	I	.	Y	.	H	.			
CA/452/97	.	A	N	.	.	T	.	.	.	.	D	A	.	.	.	.	.	.	.	N	.	.	K	Y	.	P	N	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	H	.				
CA/1489/97	.	A	N	.	.	T	.	.	.	.	S	A	.	.	.	.	.	.	.	N	.	.	K	Y	.	P	N	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	H	.				
CA/5069/97	C	T	S	-	W	Y	D	M	V	.	G	C	T	I	L	A	.	.	.	N	.	.	K	Y	.	P	N	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	H	.				
CA/376/98	.	G	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	Y	A	S	.	D	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	K	.	S	.		
CA/516A/98	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	Y	A	S	.	D	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Q	I	.	K	.	S	.	
CA/561/99	T	.	P	D	.	.	S	.	.	.	A	.	.	.	.	.	.	.	.	N	.	.	Y	A	S	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Q	I	.	.	.	N	.	
CA/588/99	T	.	P	D	.	.	S	.	.	.	V	.	.	.	.	.	.	.	.	N	.	.	K	S	.	T	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Q	I	.	.	.	N	.
CA/3402/99	.	G	.	.	.	R	.	.	.	.	V	.	.	.	.	.	.	.	.	N	.	.	Y	A	S	R	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Q	I	.	.	.	N	.
CA/3760/99	L	R	L	W	-	Y	A	M	A	.	T	K	.	.	.	.	.	.	.	N	.	.	.	K	D	L	Q	.	A	.	.	.	.	.	.	.	.	.	.	.	.	I	.	R	.	.	N	.	
CA/3761/99	.	G	.	.	.	L	.	.	.	.	T	K	.	.	.	.	.	.	.	N	.	.	.	K	D	R	Q	.	A	.	.	.	.	.	.	.	.	.	.	.	.	I	.	R	.	.	N	.	
CA/5442/99	.	G	.	.	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	Y	A	S	.	D	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Q	I	.	.	.	N	.
CA/37/00	C	.	Q	S	-	.	S	.	.	.	V	Q	.	.	.	.	.	.	.	N	.	.	.	K	N	.	P	N	T	.	.	.	.	.	.	.	.	.	.	.	.	.	R	I	D	.	.	N	.
CA/354/00	.	Q	S	-	.	S	.	.	.	.	V	Q	.	.	.	.	.	.	.	N	.	.	.	K	N	.	P	N	T	.	.	.	.	.	.	.	.	.	.	.	.	.	R	I	D	.	.	N	.
CA/1030/00	C	T	S	-	W	Y	D	M	V	.	.	C	T	I	L	A	.	.	.	N	.	.	.	K	N	.	P	N	T	.	.	.	.	.	.	.	.	.	.	.	.	.	R	I	D	.	.	N	.
CA/1031/00	C	T	S	-	W	Y	D	M	V	.	.	C	T	I	L	.	.	.	.	N	.	.	.	K	S	.	H	N	-	.	.	.	.	.	.	.	.	.	.	.	.	.	R	I	.	.	N	.	
CA/1032/00	.	A	N	.	.	T	.	.	.	.	S	A	.	.	.	.	.	.	.	N	.	.	.	K	S	.	H	N	-	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	.	N	.		
CA/1033/00	W	.	L	H	Q	L	.	V	H	G	Q	V	H	N	F	D	G	S	L	N	.	.	.	K	S	.	H	N	-	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	.	N	.		
CA/96/01	.	Q	S	-	.	S	.	.	.	.	V	Q	.	.	.	.	.	.	.	N	.	.	.	K	N	.	P	N	T	.	.	.	.	.	.	.	.	.	.	.	.	.	R	I	D	.	.	N	.
CA/291/01	.	Q	S	-	.	S	.	.	.	.	V	Q	.	.	.	.	.	.	.	N	.	.	.	K	S	.	P	N	T	.	.	.	.	.	.	.	.	.	.	.	.	.	R	I	D	.	.	N	.
CA/389/01	.	Q	S	-	.	S	.	.	.	.	V	Q	.	.	.	.	.	.	.	N	.	.	.	K	N	.	P	N	T	.	.	.	.	.	.	.	.	.	.	.	.	.	R	I	D	.	.	N	.
CA/541/01	.	Q	S	-	.	S	.	.	.	.	V	R	.	.	.	.	.	.	.	N	.	.	.	K	D	.	L	N	T	.	.	.	.	.	.	.	.	.	.	.	.	.	R	I	D	.	.	N	.
CA/627/01	.	Q	S	-	.	S	.	.	.	.	V	Q	.	.	.	.	.	.	.	N	.	.	.	K	N	.	P	N	T	.	.	.	.	.	.	.	.	.	.	.	.	.	R	I	D	.	.	N	.

Fig. 2 (continued)





**Fig. 3.** Rooted phylogenetic tree describing the ancestor-descendant relationships between aa sequences of IBV variants isolated between 1988 and 2001 in California and the distant related FIP virus. The closer a node is to the root of the tree, the more similar to the prototype. Horizontal and vertical branch lengths are not significant. Interior branch values (shown when  $\geq 70\%$ ) represent the percent occurrence of that clade per 1,000 bootstrap replicates

distances (Fig. 3). Subgroup A with the shorter evolutionary distances contains predominantly variants isolated from layer chickens (13 layer versus 3 broiler [CA/3760/99, CA/3761/99, CA/588/99]), whereas subgroup B displays mostly isolates from broiler chickens (9 broiler versus 4 layer [CA/1030/00, CA/1031/00, CA/1032/00, CA/1033/00]). IBV variants with  $\geq 9$  bp nucleotide deletion were located in subgroup B and were found with the 2 isolates having a 6 bp nucleotide deletion in the upper main branch of the phylogenetic tree. Most of the IBV variants (15/16) included in subgroup A were isolated after 1997, except CA/1215/88 (the prototype for this study), which was isolated in 1988 from layer birds. Ten of 13 IBV variants displayed in subgroup B were isolated between 1995–1997 and 1999/2000. There was no statistical association between the grouping according to GenBank and the location in either subgroup of the phylogenetic tree ( $p = 0.244$ ), however a significant association between the production type and the location of the variants in the subgroups (0.008) was seen. Significant associations were also found between the location of the variants in the “phylogenetic tree and a high number of nucleotide mutations/year ( $p = 0.005$ ), with positive antigenicity values in HVR-2 ( $p \leq 0.001$ ), as well as with the occurrence of a deletion in HVR-1 ( $p \leq 0.001$ ).

#### *Predictive antigenicity*

Predictive antigenicity was evaluated within four areas of each sequence within the HVR and at least 9 aa upstream in conserved regions of the genome (Fig. 2) and reported as either positive or negative for antigenicity as measured by values expressing the relative hydrophilicity of each aa averaged over the length of the polypeptide chain. Sixty-five percent of the evaluated layer-type sequences and 71% of the broiler sequences, 68% of the total sequences, showed positive predictive antigenicity within 9 amino acids upstream of HVR1. Only 10% of the sequences displayed predictive antigenicity within the HVR1. Six percent within the layer-type sequences, and 17% within the broiler sequences were positive for predictive antigenicity in HVR1. Twelve percent of the evaluated layer-type sequences and 50% of the broiler sequences, 28% of the total sequences, showed positive predictive antigenicity within 10 amino acids upstream of HVR2. In HVR2 itself 59% of the layer type sequences had a positive predictive antigenicity value, whereas no positive sequences were found among the broiler sequences.

#### *Glycosylation analysis*

O- and N-glycosylation sites are shown in Fig. 2. All sequences show a N-glycosylation site at aa position 103 and 145 within the conserved part of the sequence. Sixty-nine % (20/29) of the sequences display a N-glycosylation site at aa position 51 within the HVR-1, and 62% of the sequences (18/29) show a putative N-glycosylation site at aa position 135 within the HVR-2 region.



### *Statistical evaluation*

Fisher's Exact test showed an association between the production type of chicken and the number of nucleotide mutations/year at a chosen cut off point of 7 mutations per year ( $p = 0.013$ ). IBV variants isolated from broilers demonstrated a higher number of mutations/year on the nucleotide level compared with IBV variants isolated from layers. There was no association between a high number of mutations/year and the GenBank group of the variants ( $p = 0.198$ ). IBV variants isolated from broilers were also more likely to have the  $\geq 6$  bp deletion in their nucleotide sequence compared with IBV variants isolated from layers ( $p = 0.006$ ). There was no association between the appearance of the deletion and the GenBank group of the variant ( $p = 0.099$ ). An association was identified between the production type of the chicken and the GenBank grouping of the IBV variant ( $p = 0.028$ ). IBV variants isolated from broilers were more likely to belong in the original or group 1 California variant group. The production type of chicken was also associated with predictive antigenicity in ten amino acids upstream of HVR2 ( $p = 0.033$ ), and in HVR2 ( $p = 0.001$ ). Predictive antigenicity was more likely to be found in variants isolated from layer-type birds, and those located in the subgroup A (closer to the prototype) of the phylogenetic tree ( $p < 0.001$ ). Predictive antigenicity in HVR2 was associated with a high number of mutations/year on the nucleic acid level ( $p = 0.008$ ), showing that sequences with more than 7 mutations per year were more likely to have a positive predictive antigenicity value. A  $p$ -value of  $< 0.001$  was also found testing the association between the presence of the  $\geq 6$  bp deletion and predictive antigenicity in HVR-2. Variants showing predictive antigenicity in HVR2 did not display the deletion. A positive predictive antigenicity in HVR2 was also associated with being located in the Subgroup A (closer to the prototype) of the phylogenetic tree ( $p < 0.001$ ). A significant association between antigenicity and changing of aa sequences in HVR-2 was not found ( $p = 0.105$ ).

### **Discussion**

The S1 gene of IBV has serotype specific and neutralization specific epitopes [3, 27, 34, 37]. Antigenic variability and cross-protection between serotype and virus are believed to be limited [33], a phenomenon that has stimulated much research interest in the antigenic HVRs in S1 of IBV [5, 21, 33, 36, 37, 41]. This study documents different evolutionary patterns of IBV depending on the production type of chicken (layer versus broiler).

The fact that 88% of all significant, directed evolutionary changes such as the conservation of aa as measured by purifying selection, and positive selection measured by a change of the aa, was found in variants isolated from layer-type birds suggests that IBV variants isolated from layer chickens show a higher potential and/or evolutionary need to respond to external pressures, whereas a higher number of mutations/year on the nucleotide level without subsequent amino acid changes were seen in IBV variants isolated from broilers. Mutations on

the nucleotide level occur in IBV in poultry constantly, but the life expectancy of the two different production types is probably responsible for the different outcomes. In broilers, with an average life expectancy of only 6 weeks, there is not sufficient pressure placed on the virus to avoid immune surveillance. Therefore the mutations on the nucleotide level observed are most likely due to the high mutation rate of RNA viruses in general [15, 16] combined with limited transmission opportunity over the relatively short lifespan of the bird or flock. On the other hand, layers could be subject during their more than one year life span [11] to several IBV infections in addition to routinely administered IBV-vaccines, resulting in selective immune pressures toward evolutionary change expressed in aa changes and positive predictive antigenicity. Though the association between evolutionary conservation of aa in the sequenced S1 gene fragment and the production type of the chicken was not statistically significant in our analysis, the findings may be different with a larger sample size.

The production type of chicken and the location of the IBV variants in the phylogenetic tree subgroup were associated. Two IBV isolates recovered from broilers (CA/3760/99, CA/3761/99), but located with layers in subgroup A of the phylogenetic tree were recovered from black skin chickens produced for a specialty market. The same isolates showed closer evolutionary distances to the prototype in the rooted tree. These specialty birds are used as breeders and layers as well as for meat production, and therefore do not represent typical genetics or management practices seen in broilers or layers. The different genetic makeup, the multipurpose production type, and the different management type, including extended life span of this breed might explain why variants isolated from these birds were more closely linked to the IBV variants associated with layer-type birds. Interestingly those two variants form a possible independent branch in the unrooted tree, if tree reconstruction is based only on aa sequences in the HVRs (data not shown). Both variants were isolated in the same year (1999), and within the same zip-code area (data not shown), and a geographical influence such as airborne transmission [8] cannot be excluded in this case. Other possible sources of virus might include outside introductions through the use of feed mills or other shared management practices.

Some IBV variants isolated from layer type chickens were phylogenetically grouped (Subgroup B) with the variants isolated predominantly from broilers and show the 9 base-pair deletion. Despite the statistical association between the broilers and the deletion, variants containing the deletion were occasionally recovered from layer-flocks, suggesting that the deletion itself is not responsible for the association with chicken production type. Three of these variants were isolated from the same zip-code area and the zip-code for the fourth could not be retrieved. The findings are consistent with Moore et al. [36] who showed that IBV variants with unique HVRs circulate in geographically isolated areas in the United States.

A  $\geq 6$  bp deletion was observed within the HVR1 in all variants located in Subgroup B of the phylogenetic tree, placing this group evolutionarily further from the prototype. The observed deletion was consistent, so presumed not to be a

random sequencing error or phenomenon created during laboratory passage of the original IBV isolate, leading to the conclusion that IBV variants display mutations without frame shifts within the HVR1. Interestingly, CA/561/99 and CA/588/99 are the only variants with a 6 bp deletion that were grouped into Subgroup A in the phylogenetic tree. The different phylogenetic grouping of the 6 bp and the  $\geq 9$  bp deletion suggest the 2 deletions are not evolutionarily associated despite similar location in the HVR1. The two deletion variants appear to be circulating in different pools of chickens, and just coincidentally, or possibly for an evolutionary advantage, have deletions located at the same position.

Subgrouping of IBV variants using the phylogenetic tree is not necessarily consistent with GenBank grouping. The difference between the two methods is due to the small number of nucleotides per variant that the GenBank match was based on. Wang et al. showed that phylogenetic grouping of only the S1 HVR amino acid sequences provides similar results to those obtained from sequencing the entire S1 genome [41]. However, trends toward mutations or conservation of specific aa patterns found in HVRs cannot be generalized for the entire S1 gene based on our study, showing that even though HVRs are thought to function as the center of diversity, mutations in more stable parts of the sequence were also detected. Additionally, the ratio of aa changes to silent mutations and the significance of the hypothesis testing for aa changes was different for HVRs compared to the larger S1 sequence, suggesting that HVR2 is predisposed to alterations in the aa pattern compared to HVR1. We can speculate that the neutralizing epitopes associated with HVR2 might be more important for making the first virus contact with the epithelial cells in affected poultry. Predictive antigenicity in HVR2, found only in variants isolated from layer type chickens, was also significantly linked to an overall number of less than 7 mutations per year, not having a  $\geq 6$  bp deletion, and being located within the Subgroup A of the phylogenetic tree. These characteristics all distinguish sequences according to the production type of bird. This result also seems to emphasize the selective pressure on IBV isolates from layer-type chickens toward change. However, we did not find any significant associations between predisposition to altered aa in HVRs or in the entire sequenced area, and predictive antigenicity. We explain the missing association with the possibility that some antigenic determinants might not be correlated with hydrophilicity and would not be assessed with a predictive antigenicity evaluation based on the principle of the greatest local hydrophilicity [18]. The non-significant association between predisposition to altered aa and predictive antigenicity, and especially the fact that CA/516A/98 does not show a positive predictive antigenicity value despite a predisposition toward altered aa in HVR2, might be examples of evolution without increasing hydrophilicity. Further immunologic studies, including field trials are needed to confirm predicted antigenicity, and to evaluate the impact of changes in HVR1 and HVR2 on the three-dimensional structure of epitopes. Though the population of variants studied was too small to statistically evaluate clinical disease association (data not shown), there is an interesting trend toward increased severity of disease that deserves further evaluation.

GenBank variant group might be linked to an evolutionary pattern, as suggested by results showing that variants closely related to CA 85-633 and 1013 may be constantly changing in the HVR-related epitopes, whereas variants closely related to CV-9437 and CV-1686 are more likely to remain structurally constant. An observed high mean nucleotide number of 8 mutations/year in concert with the fact that only one variant tested positive for an altered aa pattern seems to be unusual for a genomic sequence including two HVRs. It would be expected, that a high number of nucleotide mutations per year does conform to higher predisposition to altered aa in the group 1 variants, which are predominantly from broiler type birds. Additionally, the AT richness in the 3<sup>rd</sup> codon position can increase the apparent rate of non-synonymous substitution [20] and would favor a predisposition to altered aa, meaning that among the IBV isolates evaluated, fewer than five mutants might show true predisposition to altered aa. In the population studied, a period of slow evolution despite a rapid replication of the virus and a high mutation frequency was observed. Eigen et al. [10] explained this phenomenon by demonstrating that in constant environments selection for fit of the master sequence occurs. Such periods of relative stasis do not imply that the viruses are incapable of rapid evolution under different circumstances [17]. High numbers of nucleotide mutations per year that do not result in altered aa, as seen in IBV variants from broilers, suggest a neutral polymorphism [26]. Cavanagh et al., reported a low mutation rate ( $10^{-4}$ /year) at the nucleotide level in British and French IBV isolates collected from both production types [6], and associated the low mutation rate with a small immune pressure on the S1 protein to change. Our results demonstrate different findings depending on the production type of the chicken, with layer birds having evidence of more immune pressure to change.

A conserved amino acid sequence ('HGGAY') on which the widely used CK4 primer was developed [25] was not observed in all IBV variants evaluated in this study, emphasizing the importance of the three wobble bases in the highly degenerate CK4 primer. Though unexpected, our results are consistent with unpublished observations (J. Gelb, personal communication, 2002). Most of the observed changes in the motif are radical changes [19], increasing the polarity and/or the hydrophobicity of the sequence. Those changes might suggest an adaptation allowing improved, membrane fusion and attachment functions as well as increased ability to avoid neutralization [35]. The observation also allows the speculation that the CK4 primer does not detect all IBV variants.

Possible sequencing errors cannot be excluded despite the clear statistical data concerning associations between production type, mutant type, and number of mutations/year. In fact, 2 sequences eliminated from this study that were very similar on the nucleotide level appeared to have had frame shift mutations, because no plausible protein-reading frame could be designated. It is also possible that other sequences might have multiple frame shift mutations that would not result in a stop codon, sequence mutations that occurred during passage of the virus in the laboratory [15], or due to errors in the techniques [28]. An imperfection in the current study that should be corrected in future work is the addition of a housekeeping gene as an assay control. Additionally an unknown variable resulting

from the choice of convenience sampling is the unknown time period flocks or the representative sentinel chickens were infected with the particular variants prior to isolation.

The results of the study provide data suggesting that different management environments play a role in the evolution of IBV, and further epidemiologic tracking of variants is warranted.

### Acknowledgments

We thank Drs. Peter Woolcock and Michael McFarland of the California Animal Health and Food Safety Laboratory System (CAHFS) Avian Virology Laboratory for providing the IBV isolates, and Helen Piontkivska, Penn State University, for instruction and explanations with the evolutionary methodology. This research was supported by California Food Animal Health Funds through the School of Veterinary Medicine, University of California, Davis. Post-graduate fellowship support was provided to Beate M. Schikora by the German Academic Exchange Service (DAAD).

### References

1. American Association of Avian Pathologists (1980) Infectious bronchitis virus. In: Hitchner SB (ed) Isolation and identification of avian pathogens, 2nd edn. College Station, TX, New York, pp 70–71
2. Case JT, Sverlow KW, Reynolds BJ (1997) A novel protein polymorphism differentiates the California serotype of infectious bronchitis from other serotypes common to California. *J Vet Diagn Invest* 9(2): 149–155
3. Cavanagh D, Davis PJ, Mockett AP (1988) Amino acids within hypervariable region 1 of avian coronavirus IBV (Massachusetts serotype) spike glycoprotein are associated with neutralization epitopes. *Virus Res* 11(2): 141–150
4. Cavanagh D, Davis PJ, Cook JKA (1992) Infectious bronchitis virus: evidence for recombination within the Massachusetts serotype. *Avian Pathology* 21: 401–408
5. Cavanagh D, Cook JKA, Li D, Kant A, Koch G (1992) Location of the amino acid differences in the S1 spike glycoprotein subunit of closely related serotypes of infectious bronchitis virus. *Avian Pathol* 21: 33–43
6. Cavanagh D, Mawditt K, Adzhar A, Gough RE, Picault J-O, Naylore CJ, Haydon D, Shaw K, Britton P (1998) Does IBV change slowly despite the capacity of the spike protein to vary greatly? 729–734. In: Enjuanes L, Siddell SG, Spaan W (eds) Coronaviruses and arteriviruses. Plenum Press, New York
7. Cavanagh DAN, Syed A (1997) Infectious bronchitis. In: Calnek BW, Barnes HJ, Beard ChW, McDougald LR, Saif YM (eds) Diseases of poultry, 10th edn. Iowa State University Press, Ames, Iowa, pp 511–526
8. Cumming RB (1970) Studies on Australian infectious bronchitis virus. IV. Apparent farm-to-farm airborne transmission of infectious bronchitis virus. *Avian Dis* 14: 191–195
9. Dopazo J (1994) Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach. *J Mol Evol* 38: 300–304
10. Eigen M, Biebricher C (1988) Sequence space and quasispecies distribution. In: Domingo E, Holland JJ, Ahlquist P (eds) RNA genetics, vol 3. CRC Press, Boca Raton, Fla
11. Estola T (1966) Studies on the infectious bronchitis virus of chickens isolated in Finland. In: Studies on the infectious bronchitis virus of chickens isolated in Finland.

- State Veterinary Medical Institute, University of Helsinki, National Veterinary Institute, Helsinki
12. Gelb J, Wolff JB, Moran CA (1991) Variant serotypes of infectious bronchitis virus isolated from commercial layer and broiler chickens. *Avian Dis* 35(1): 82–87
  13. Hansen JE, Lund O, Rapacki K, Brunak S (1997) O-GLYCBASE version 2.0: a revised database of O-glycosylated proteins. *Nucleic Acids Res* 24: 278–282
  14. Harless DH, Donahoe VA, Daft JP, BM (1987) Serologic investigation of suspect variant bronchitis isolates from California. *Proc West Poult Dis Conf* 36: 63–66
  15. Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S (1982) Rapid evolution of RNA genomes. *Science* 215(4540): 1577–1585
  16. Holland JJ, de la Torre JC, Steinhauer DH, Clarke D, Duarte E, Domingo E (1989) Virus mutation frequencies can be greatly underestimated by monoclonal antibody neutralization of virions. *J Virol* 63(12): 5030–5036
  17. Holland JJ, de la Torre JC, Steinhauer DA (1992) RNA virus populations as quasispecies. In: Holland JJ (ed) *Current topics in microbiology and immunology*, vol 176. Springer Berlin Heidelberg New York
  18. Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci U.S.A.* 78(6): 3824–3828
  19. Hughes AL, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol Biol Evol* 7(6): 515–524
  20. Hughes MK, Hughes AL (1995) Natural selection on Plasmodium surface proteins. *Mol Biochem Parasitol* 71(1): 99–113
  21. Jia W, Karaca K, Parrish CR, Naqi SA (1995) A novel variant of avian infectious bronchitis virus resulting from recombination among three different strains. *Arch Virol* 140(2): 249–271
  22. Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HNE, Allison JB (eds) *Mammalian protein metabolism III*. Academic Press, New York
  23. Kabat EA (1968) *Structural concepts in immunology and immunochemistry*. Holt, Rinehart & Winston, New York
  24. Karaca K, Naqi S, Gelb J (1992) Production and characterization of monoclonal antibodies to three infectious bronchitis virus serotypes. *Avian Dis* 36(4): 903–915
  25. Keeler CL, Reed KL, Nix WA, Gelb J (1998) Serotype identification of avian infectious bronchitis virus by RT-PCR of the peplomer (S-1) gene. *Avian Dis* 42(2): 275–284
  26. Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267: 275–276
  27. Koch G, Hartog L, Kant A, van Roozelaar DJ (1990) Antigenic domains on the peplomer protein of avian infectious bronchitis virus: correlation with biological functions. *J Gen Virol* 71(Pt 9): 1929–1935
  28. Kottier SA, Cavanagh D, Britton P (1995) Experimental evidence of recombination in coronavirus infectious bronchitis virus. *Virology* 213(2): 569–580
  29. Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17(12): 1244–1245
  30. Kusters JG, Jager EJ, Niesters HG, van der Zeijst BA (1990) Sequence evidence for RNA recombination in field isolates of avian coronavirus infectious bronchitis virus. *Vaccine* 8(6): 605–608
  31. Kusters JG, Niesters HG, Bleumink-Pluym NM, Davelaar FG, Horzinek MC, Van der Zeijst BA (1987) Molecular epidemiology of infectious bronchitis virus in The Netherlands. *J Gen Virol* 68(Pt 2): 343–352

32. Kusters JG, Niesters HG, Lenstra JA, Horzinek MC, van der Zeijst BA (1989) Phylogeny of antigenic variants of avian coronavirus IBV. *Virology* 169(1): 217–221
33. Kwon HM, Jackwood MW, Gelb J (1993) Differentiation of infectious bronchitis virus serotypes using polymerase chain reaction and restriction fragment length polymorphism analysis. *Avian Dis* 37(1): 194–202
34. Mockett AP, Cavanagh D, Brown TD (1984) Monoclonal antibodies to the S1 spike and membrane proteins of avian infectious bronchitis coronavirus strain Massachusetts M41. *J Gen Virol* 65(Pt 12): 2281–2286
35. Monos DS, Tekolf WA, Shaw S, Cooper HL (1984) Comparison of structural and functional variation in class I HLA molecules: the role of charged amino acid substitutions. *J Immunol* 132(3): 1379–1385
36. Moore KM, Bennett JD, Seal BS, Jackwood MW (1998) Sequence comparison of avian infectious bronchitis virus S1 glycoproteins of the Florida serotype and five variant isolates from Georgia and California. *Virus Genes* 17(1): 63–36
37. Moore KM, Jackwood MW, Hilt DA (1997) Identification of amino acids involved in a serotype and neutralization specific epitope within the S1 subunit of avian infectious bronchitis virus. *Arch Virol* 142(11): 2249–2256
38. Nix WA, Troeber DS, Kingham BF, Keeler BL, Gelb J (2000) Emergence of subtype strains of the Arkansas serotype of infectious bronchitis virus in Delmarva broiler chickens. *Avian Dis* 44(3): 568–581
39. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4): 406–424
40. Thompson JD, Higgins DG, Gibson TJ (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(22): 4673–4680
41. Wang CH, Huang YC (2000) Relationship between serotypes and genotypes based on the hypervariable region of the S1 gene of infectious bronchitis virus. *Arch Virol* 145(2): 291–300

Author's address: Sharon K. Hietala, California Animal Health and Food Safety Laboratory, University of California, Davis, West Health Services Drive, Davis, CA 95617-1770, U.S.A.; e-mail: [skhietala@ucdavis.edu](mailto:skhietala@ucdavis.edu)