Complete genome sequence and recombination analysis of infectious bronchitis virus attenuated vaccine strain H120

Yi Zhang · Hong-Ning Wang · Ting Wang ·
Wen-Qiao Fan · An-Yun Zhang · Kun Wei ·
Guo-Bao Tian · Xin Yang

Abstract The strain H120 of infectious bronchitis virus (IBV) is one of the earliest and representative attenuated live Infectious Bronchitis vaccine strains. To investigate the genomic feature of H120 and further understand its role in the epidemiology of IBV, complete genome of H120 was sequenced and compared with sequences of other IBV strains by phylogenetic and recombination analysis. The complete genome of H120 is 27631 nucleotides in length and has a similar structure with that of Beaudette strain. We found that strain ZJ971 is probably a virulence revertant of H120. Nine amino acids changes and a three-nucleotide deletion were identified in ZJ971. Besides, potential recombination events associated with H120 were found in five IBV strains including H52, KQ6, SAIBK, Ark DPI 11, and Ark DPI 101. This study suggested that H120 might have contributed to the emergence of new IBV variants through both virulence reversion and recombination.

Keywords Infectious bronchitis virus ·
Attenuated vaccine · H120 · Complete genome ·
Recombination

Introduction

Avian infectious bronchitis virus (IBV) is the causal agent of the Infectious Bronchitis, an acute and highly contagious disease of chickens. It is categorized to the group 3 of genus *Coronavirus*, together with genetically related coronavirus of turkey [1]. Numerous IBV strains existed, causing pathological changes in different organs varying from respiratory tract to kidney and gonads [2–4].

IBV is an enveloped virus replicates in the cell cytoplasm with a single-stranded, positive-sense RNA genome of 27.6 kb in size [5]. Five 5′- and 3′-co-terminal subgenomic mRNAs are transcribed from the IBV genome in virus-infected cells [6]. The Gene1 (replicase gene) of IBV contains two overlapping open reading frames (ORFs) ORF 1a and ORF 1b, of which the latter is translated to polyprotein pp 1ab by −1 frameshift translation [7]. These polyproteins are proteolytically processed into smaller products, e.g., papain-like proteinase (PL^{pro}), main proteinase (M^{pro}), RNA-dependent RNA polymerase (RdRp), and RNA helicase, et al., which are involved in RNA synthesis and other aspects of viral pathogenesis. Recently, the replicase gene was demonstrated to be a determinant of pathogenicity in IBV by reverse genetics [8]. Four structural proteins including spike (S) glycoprotein, small envelope (E) protein, membrane (M) glycoprotein, and nucleocapsid (N) protein are encoded by mRNA 2, 3, 4, and 6, respectively [7–9]. The S protein is cleaved into S1 and S2, of which S1 produces neutralizing and serotype-specific antibodies [3, 9].

Presently, vaccination is the major way for prevention of IBV, but the protection efficiency for IBV infection is not satisfactory because of the continual emergence of new serotypes and poor cross-protection between them [10–14]. H120, an attenuated live vaccine strain of Massachusetts (Mass) serotype, was originally obtained by serial passage of strain H (isolated in the Netherlands [15] in 1956) in embryonated chicken eggs up to the 120th passage. It has been used worldwide as a primary vaccine in broilers,

Y. Zhang · H.-N. Wang (✉) · T. Wang · W.-Q. Fan ·
A.-Y. Zhang · K. Wei · G.-B. Tian · X. Yang
Animal Disease Prevention and Food Safety Key Laboratory
of Sichuan Province, "985 Project" Science Innovative Platform
for Resource and Environment Protection of Southwestern China,
School of Life Science, Sichuan University, 29# Wangjiang Road,
Chengdu, Sichuan Province 610064, China
e-mail: whongning@163.com

breeders, and future layers for more than 50 years. Even in recent years, H120 is still popular used and commonly considered to conform to the rigorous safety standards required for avian vaccines (review by Bijlenga et al. [15]). However, the protection of H120 was also reported to be poor against some Chinese field strains [12], raising the issues of further evaluating the antigenicity and security of H120 from genome level.

To investigate the genomic feature of H120 and further understand its role in the epidemiology of IBV, the complete genome of IBV strain H120 was sequenced and compared with sequences of several widely used vaccine strains, recently reported field strains, and other typical IBV strains by phylogenetic and recombination analysis.

Materials and methods

Amplification, cloning, and sequencing of the H120 virus

Virus propagation

IBV strain H120 was obtained from Chinese Institute of Veterinary Drug Control. The H120 virus was inoculated into the allantoic cavity of 9-day-old embryonated specific pathogen-free (SPF) chicken eggs. The embryos were incubated at 37°C and examined twice daily for their viability. The allantoic fluid was collected at 72 h post-inoculation and stored at −80°C.

Viral RNA extraction, RACE, and RT-PCR

Genomic RNA was extracted from virus-infected allantoic fluid with TRIzol (Invitrogen, USA) following manufacturer's instructions. Primers were designed from the consensus sequence deduced from the alignment of 16 IBV complete sequences in GenBank (Table 1). Five primers were used for RACE PCR to amplify 5′ and 3′ terminal sequences, and 16 pairs of primers were used to amplify the remaining parts of H120 genome (Table 2). 5′-RACE, 3′-RACE, and reverse transcription were performed using 5′-full RACE kit (TaKaRa, Japan) and Superscript III First-Strand Synthesis System (Invitrogen, USA) following the manufacturer's instructions. PCR was performed in PCR Machine (Bio-Rad, USA) with 2 µl cDNA as template in a total of 25 µl reaction volume containing 2.5 µl of 10× LA buffer, 5 µl of 2.5 mM dNTPs, 2 µl of 25 mM MgCl$_2$, 1 µl of 10 pmol/µl each primer, and 0.3 µl of LA Taq polymerase (TaKaRa, Japan). The PCR parameters included an initial denaturation for 5 min at 94°C followed by 30 cycles of denaturation at 94°C for 50 s, annealing at 53°C for 50 s, and extension at 72°C for 1–3 min depending on

the sizes of the products and a final extension at 72°C for 10 min.

Genome sequencing

All PCR products were purified from 0.8% agarose gel using QIAquick PCR Purification Kit (QIAGEN Inc., Valencia, CA). The purified products were cloned into the pMD19-T Vector (TaKaRa, Japan). The nucleotide sequences of the positive clones were determined by Sangon Biological Engineering Technology & Services Co., Ltd. Each nucleotide was determined from three identical results. Sequences were assembled into complete genome sequence using SeqMan II program of DNAStar software package (DNAStar, Madison, WI).

Genome sequence analysis

Sequence comparison and phylogenetic analysis

Full or partial sequences of 30 IBV strains and four bird coronavirus strains (as outgroup) were retrieved from GenBank with accession numbers listed in Table 1. MegAlign program of DNAStar software package was used to generate multiple sequence alignments and determine the nucleotide identity with clustal V method. Phylogenetic analyses were performed with the neighbor-joining method using MEGA version 4 [16]. The bootstrap values were determined from 1000 replicates of the original data.

Putative recombination analysis

The complete genome sequences based on the multiple alignment result were used for recombination analysis. Potential recombination events were detected using the RDP [17], GENECONV [18], and MaxChi [19] recombination detection methods implemented in RDP3.41 [20]. The highest acceptable P value and the sliding window size were set to 0.05 and 30 bp, respectively.

Results and discussion

Complete genome of H120

The complete genome sequence of H120 obtained in this study was submitted to GenBank database under the accession number of FJ888351. The complete genome of H120 consists of 27631 nucleotides (nt), which encodes six genes. The gene 1 consists of two ORFs, ORF 1a and 1b, which are 11802 and 8064 nt long, respectively. Gene 2, encoding S protein, has a single ORF of 3489 nt (encoding

Table 1 Coronavirus sequences analyzed in this study

Strains	Geographic origin	GenBank accession no.
M41[abc]	Massachusetts, USA	AY851295[d]
HK[ac]	China	AY761141
Jilin[ac]	Jilin, China	AY839144, EF602442
ArkDPI 11[ac]	Delmarva, USA	EU418976[d]
ArkDPI 101[ac]	Delmarva, USA	EU418975[d]
Beaudette[bc]	New Jersey, USA	NC_001451[d]
Cal99[c]	California, USA	AY514485[d]
Vic[c]	Australia	DQ490221
KB8523[c]	Japan	M21515
BJ[c]	Beijing, China	AY319651[d]
LX4[c]	Xinjiang, China	AY338732[d]
SAIBK[c]	Sichuan, China	DQ288927[d]
TW2575/98[c]	Taiwan, China	DQ646405[d]
ZJ971[c]	Zhejiang, China	EU714028[d]
CK/CH/LSD/05I[c]	Shandong, China	EU637854[d]
HKU11-796[f]	Hong Kong, China	NC_011548[d]
HKU12-600[g]	Hong Kong, China	NC_011549[d]
H52[abc]	China and USA	China1: EU817497[d]; China2: AY044184, AY044185; USA1: AY392068, AY392058, AY392048, AY392078; USA2: AF286185, AF317210
H120[abc]	China and UK	FJ888351[de], FJ807652[d], AJ278336, AY028296, AY028295, EU822341
DE072[ac]	Delmarva, USA	AY392054, AY392074, AY392064, AY392084, AF274435, AY024337, AF202998, AF202999, AF203000, AF203001, AF203002
Conn[ac]	Connecticut, USA	AY392049, AY392069, AY392059, AY392079, L18990, AY942752, AY942741, AY942746
Ma5[abc]	USA	AY561720, AY561740, AY561720, AY561730, AY561750, AY561713
Ark99[c]	Arkansas, USA	AY392051, AY392071, AY392061, AY392081, L10384, L06251
ArkDPI[ac]	Delmarva, USA	AY392052, AY392072, AY392062, AY392082, AF006624, AF335555, AF363597, AY942745
CU-T2[c]	New York, USA	AY561724, AY561744, AY561734, AY561754, AIU49858
Gray[c]	Delmarva, USA	AY392076, AY392056, AY392066, AY392086, L18989, AF394180, AF318282, AF286180, AF469011, AF469011, L06252
GA98[c]	Georgia, USA	AY392073, AY392053, AY392063, AY392083
Florida[c]	Florida, USA	AY392070, AY392050, AY392060, AY392080
Partridge/GD/S14/2003[c]	Guangdong, China	AY646283[d]
Peafowl/GD/KQ6/2003[bc]	Guangdong, China	AY641576[d]
A2[c]	Beijing, China	EU526388[d]
SC021202[c]	Sichuan, China	EU714029[d]
HKU11-934[f]	Hong Kong, China	FJ376619[d]
HKU13-3514 [h]	Hong Kong, China	NC_011550[d]

[a] Vaccine strain

[b] Massachusetts serotype strain

[c] Infectious Bronchitis coronavirus

[d] Complete genome sequences involved in alignment, phylogenetic and RDP analysis

[e] The complete sequence of H120 obtained in this study

[f] Bulbul coronavirus

[g] Thrush coronavirus

[h] Munia coronavirus

Table 2 Primers used for H120 genome amplification

Fragments	Location[a]	Forward primer (5′–3′)	Reverse primer (5′–3′)
P1	487–1912	TGGTGGTGTTTAGTGAGCAG	TCCAGAATTCAAAAGACTTT
P2	1648–3093	GGTGTAACACCGGAGATAAATG	ATTCACAACATCAAGAGCGTC
P3	2981–5478	TGTTGGCTCTTATACAAGAC	ATTAAGGTGTTCTGCACG
P4	5261–6504	TATGGGTAGACGGCAGTG	GCAAGCCTCATCCACAAC
P5	6469–8279	ATGTCTCCTGAAGTTGCT	CCTTGAAACTTTATAACC
P6	8133–9635	TGGTTCTACTGTTAGAGAAC	GCACTTAATTTAGTAATAGC
P7	9367–11056	GCATTACACACTGGAACTGA	TATTGGCAGCTTTACGAT
P8	10728–13179	TCTTGCTTCTGATGATGT	CTTGTTTCTCCTCAGTAT
P9	12813–15041	AATAAGGATTGGTACGAC	GTCCATAAGCATATTCTGAG
P10	14894–16541	GGCTGTTATGGAGCGTTAT	ACACTCACGTGATTCTGGG
P11	16410–18062	CTTATGGTTTGTGTTAAACC	ACAAAGACCATCAGCAAAC
P12	17784–20111	TGTCAAGATGTCAACTGG	GCATTCACTGCTGTACAA
P13	19860–21580	ACAGAGCACAAGTTTGATC	CGCTCTTAGTAACATAAAC
P14	20873–23171	TAAATGGTGATCTTGTTT	AACACTATACCATTAGGTGC
P15	22755–24748	TGCTGCTTCCTTTAATAAG	CTGCGACAAGACCTCCTG
P16	24685–27010	CTTAACATTGCAGTAGGTG	CTGAGGTCAATGCTTTATC
5′ region[b]	1–525	CATGGCTACATGCTGACAGCCTA	GTCACTGTCTATTGTATG
3′ region[c]	26970	AAAGAAGCAGGATGATGAAG	AAGCAGTGGTATCAACGCAGAGT
3′ RT[d]	Poly A	AAGCAGTGGTATCAACGCAGAGTAC(T)$_{30}$	

[a] The location of primers is given relative to nucleotide sequence of the Beaudette IBV strain

[b] PCR primers used in 5′-RACE

[c] PCR primers used in 3′-RACE

[d] RT primer used in 3′-RACE

1162 amino acids (aa)). The S1 and S2 genes are 1596 and 1893 nt long, respectively. Gene 3 consists of 709 nt with 5′-end 32nt overlaps with S gene. There were three ORFs in gene 3 including 3a (174nt), 3b (195nt), and 3c (330nt), encoding proteins of sizes 57, 64, and 109 aa in length, respectively. Gene 4, encoding M protein, has a single ORF of 678 nt (encoding 225 aa). There is a 342 nt non-coding region between the 3′-end of the M protein gene and 5′-end of gene 5. Gene 5 contains two ORFs: 5a and 5b with 198 and 249 nt, respectively (encoding 65 and 82 aa, respectively). Gene 6 has a single ORF of 1230 nt, encoding the N protein of 409 aa. Untranslated regions (UTRs) at 5′- and

Table 3 Comparison between H120 (FJ888351) and other H120 sequences

Accession	Sequence coverage	Nucleotide identity (%)	Nucleotide differences[a]
FJ807652	Complete genome	99.8	^{179}G(A) ^{422}G(A) ^{1120}G(A) ^{1335}A(C) ^{1493}A(G) ^{1952}T(C) ^{2034}A(G) ^{2480}A(G) ^{3904}G(A) ^{4242}C(T) ^{4993}G(A) ^{6195}T(C) ^{6257}A(G) ^{8213}A(G) ^{8286}T(C) ^{8727}T(C) ^{8843}T(G) ^{9163}A(G) ^{9814}C(T) ^{10630}A(G) ^{10780}G(A) ^{11668}A(G) ^{13990}A(C) ^{14993}C(T) ^{15282}T(C) ^{15949}C(T) ^{16184}C(T) ^{18914}G(A) ^{20369}T(C) ^{20666}T(G) ^{20766}T(G) ^{21376}C(G) ^{21868}T(C) ^{22098}C(A) ^{22942}A(C) ^{23214}A(T) ^{23453}T(C) ^{23720}A(G) ^{25936}C(T)
EU822341	3′-terminal 7 kb	99.9	^{20369}T(C) ^{20666}T(G) ^{21376}C(G) ^{23214}A(T) ^{23453}T(C) ^{23720}A(G) ^{25936}C(T)
M21970	S1	99.8	^{20658}C(T) ^{20666}T(G) ^{21376}C(G)
GQ154655	S1	99.7	^{20369}T(C) ^{20558}T(C) ^{20666}T(G) ^{20811}T(C) ^{21376}C(G) ^{22098}C(A)
AY028295	M	100	
AY028296	N	99.6	^{25936}C(T) ^{26123}A(G) ^{26614}G(T) ^{26678}T(C) ^{27075}T(G)
AM260960	N	100	
AJ278336	3′-UTR	100	

[a] Nucleotides that differ from H120 (FJ888351) and their locations, corresponding nucleotide of H120 (FJ888351) was given in parenthesis

3'-end of the genome are both 528nt in length. Transcription-regulating sequence (TRS), CT(T/G)AACAA, exists at the start of each gene, and the distances between TRS and the initiation codons of gene 1 to gene 6 are 464, 52, 23, 77, 9, and 93 nt, respectively.

The genomic features of H120, including location of TRS and Hypervariable regions (HVRs), size of ORFs, 5'-UTR and overlaps between neighboring genes, are similar to those of ZJ971, H52, Ark DPI, M41, and Beaudette. Specifically, H120 and H52 were found to possess a more intact and longer 3'-UTR in contrast to all the other IBV strains compared, making 3'-UTR a potential molecular marker for the H strains in IBV detection and phylogenetic studies.

Genome sequence analysis

Sequence comparison

Different sequences of the H lineage were compared with each other before compared with other IBV lineages. H120 sequences were found to be highly consistent with each other with no deletion or insertion. Corresponding regions of H120 (FJ888351) were 99.7–100% identical with other H120 sequences in GenBank (Table 3). Comparatively, the H52 sequences are more variable. Sequences of the US H52 isolates have huge differences from those of Chinese isolates though they were all originated in the Netherlands and introduced to the market of each country at about 1980 (USA) and 1990 (China). The US H52 isolates exhibit high nucleotide identities with H120. The four sequences of H52-USA1 (located at nt 1–1263, 4186–5496, 8881–9300, and 14046–14744) and the two sequences of H52-USA2 (located at nt 24150–24479 and 24451–25128) are 99.7, 99.8, 99.8, 100, 99.7, and 99.9% identical with H120 at the nucleotide level, respectively. Whereas the Chinese H52 isolates are less similar to H120. Especially, three regions of the complete sequence of H52-China1 (located at nt 14046–14744, 25682–25930, 25873–27102) and the two regions of H52-China2 (located at nt 24451–25128 and 25873–27102) exhibit comparatively low identities of 91.8, 96, 95.9, 96, and 89.9%, respectively.

Table 4 Percentage of nucleotide identity of different regions of H120 genome compared with other IBV strains (complete genome and 5'-terminal 20 kb)

Strains	Complete genome	5'-UTR	Gene1			
			Gene1	Main^{pro}	PL^{pro}	RdRp
ArkDPI 11	94.5	**99.4**	96.4	89.8(94.3)	**99.7(99.3)**	90.6(**97**)
ArkDPI 101	94.4	**99.4**	96.4	89.8(94.3)	**99.7(99.1)**	90.6(**97**)
Beaudette	90.9	95.9	90.3	89.8(94.3)	85.2(85.6)	91.7(**97**)
BJ	85.1	94.9	85.5	86.4(92.9)	82.5(81.1)	89.4(**97.9**)
Cal99	92.3	**98.9**	93.1	91.2(96.4)	**98.6(97.7)**	91(96.6)
KQ6	92.9	96	92.3	87.6(95.0)	86.8(86.3)	**99.4(100)**
LX4	85.4	NA	86	84.8(91.4)	82.4(79.9)	89.4(**97**)
M41	90.5	96.6	90	87.9(93.6)	86.7(86.3)	91.6(**97**)
S14	84.6	NA	85.3	85(90.0)	82.9(82)	86.4(93.6)
SAIBK	87.7	91.5	89.2	92.4(96.4)	85.2(84.5)	90.7(**97**)
TW2575-98	86.4	93.8	87	87.9(92.1)	83.5(83.8)	91(**97**)
A2	85.2	94.4	85.4	84.5(91.4)	82.2(81.3)	89.3(**97**)
CK/CH/LSD/05I	88	96.2	87.9	85.2(92.1)	83.4(81.5)	90.6(**97**)
H52-USA1	NA	**99.4**	NA	**99.8(99.3)**	**99.8(99.5)**	**100(100)**
H52-China1	96.8	97.6	96.3	**99.5(98.6)**	**98.6(98.9)**	91.8(**97.4**)
ZJ971	**99.8**	**99.6**	**99.8**	**99.8(99.3)**	**99.9(99.8)**	**99.9(100)**
SC021202	87.1	90.1	87.6	88.6(94.3)	82.2(82.9)	88.7(**97.4**)
Ma5	NA	**99.1**	NA	**99.3(97.9)**	**99.7(99.3)**	**99.9(99.6)**
Ark99	NA	**97**	NA	89.8(93.6)	86.9(85.4)	90.8(**97.4**)
ArkDPI	NA	**99.4**	NA	89.8(94.3)	**99.6(99.1)**	90.6(**97**)
Conn	NA	96.6	NA	90.2(94.3)	86.8(85.2)	91.1(**97.4**)
CU-T2	NA	**98.3**	NA	90(94.3)	94.4(94.1)	91.1(**97.4**)
DE072	NA	**99.1**	NA	**99.5(98.6)**	**99.6(99.1)**	**99.7(99.6)**
Florida	NA	96.4	NA	90.2(94.3)	86.8(85.2)	91.3(**97.4**)
GA98	NA	**99.2**	NA	89.8(94.3)	**99.5(98.9)**	90.6(**97**)
Gray	NA	**99.1**	NA	89.8(94.3)	86.8(84.9)	90.6(96.1)

Sequences with ≥97% identity are indicated in bold letters; identity rates of amino acid sequences within the parenthesis

NA not analyzed

Compared with other complete sequences of IBV, H120 exhibited identities ranging from 84.6 to 99.8% (Table 4). H120 shows the highest similarity with ZJ971 (a Chinese strain causes swollen proventriculus of chickens [21]) with identity rate of 99.8% and the lowest with most of other Chinese strains (BJ, S14, LX4, SAIBK, SC021202, and TW2575/98) with identities ranging from 84.6 to 87.7%.

Between H120 and other IBV strains, 5′-UTR is the most conserved region with 91.5–99.4% identity, while the 3′-UTR is the most variable region with identities ranging from 49.9 to 80.9% (Tables 4, 5). Overall, H120 has higher nucleotide identities with other IBV strains in replicase gene (85.3–96.4%), M^{pro} (84.5–99.5%), PL^{pro} (82–99.7%),

RdRp (86.4–99.9%), M (82.3–97.4%), 5b (91.2–100%), and N (86.1–94.2%) than in S1 (52.2–99.8%), S2 (74.5–98.3%), 3a (76.1–96%), 3b (77.1–99.5%), E (73.9–97.3%), and 5a (79.3–100%). In particular, the frequency of silent mutation was found to be high in RdRp and M^{pro}, leading their amino acid sequences more conserved (90–99.3% and 93.6–100%, respectively) than nucleotide sequences.

Sequence differences between H120 and ZJ971

As shown in Table 6, when compared with H120, there were 23 point mutations detected in coding regions of

Table 5 Percentage of nucleotide identity of different regions of H120 genome compared with other IBV strains (3′-terminal 7 kb)

Strains	Gene 2		Gene 3			Gene 4	Gene 5		Gene 6	3′-UTR
	S1	S2	3a	3b	E	M	5a	5b	N	Identity/length (nt)[a]
ArkDPI 11	78.9(76.6)	95(94.2)	86.4	84.6	86.5(80.2)	91.6(93.9)	94.4	96.8	93.4(95.8)	72.4/506
ArkDPI 101	78.9(76.3)	94.8(94.2)	86.4	84.6	86.5(80.2)	91.6(93.9)	94.4	96.4	93.3(95.6)	72.2/506
Beaudette	**97.6**(96.4)	**97.4**(96.6)	96	**98.5**	**97.3**(91.9)	95.9(**97.4**)	92.4	95.2	90.7(91.7)	73.1/505
BJ	77(74.7)	84(87.8)	82.4	77.1	82.6(78.4)	88.6(92.6)	87.9	92	87.1(91.2)	73.7/504
Cal99	78(76.3)	96.7(96.6)	83.5	84.1	88.9(82.9)	93.2(96.9)	94.4	96	96.9(**97**)	74.4/511
KQ6	**97.6**(95.9)	**97.4**(95.6)	94.9	**99.5**	**97.3**(93.7)	96.2(**99.1**)	**100**	**100**	94.2(95.1)	56.8/321
LX4	77.4(76.1)	84.6(86.7)	82.4	78.5	88.6(83.8)	90.7(91.7)	84.8	89.6	89.1(91.7)	80.9/427
M41	**97.8**(96.4)	**97.8**(**97.7**)	94.9	**99.5**	**97.3**(94.6)	96.2(**99.5**)	90.4	95.2	90.6(92.2)	58.1/320
S14	81.4(80.4)	86.5(88.5)	83	78.5	86.8(80.2)	89.4(93)	88.9	91.2	86.1(90.4)	73.5/504
SAIBK	80.7(79.3)	86.4(88)	80.7	83.6	86.5(82.9)	89.4(93.4)	87.4	94.4	86.2(90.2)	61.4/364
TW2575/98	81.8(82.2)	84.7(88)	84.7	87.2	87.4(88.3)	89.6(92.6)	84.8	94.8	88(90.4)	67.2/475
A2	81.2(81.1)	85.3(85.8)	81.2	79	87.1(80.2)	82.3(93)	79.3	92.4	86.2(89.2)	73.5/500
CK/CH/LSD/05I	82(80.9)	86.6(89.3)	76.1	83.1	88(83.8)	92(95.2)	93.9	97.2	93.3(95.1)	70.5/503
H52-USA2	NA	NA	NA	NA	**99.7**(99.1)	**99.9**(**99.6**)	NA	NA	NA	NA
H52-China1	**99.1**(98.4)	**98.2**(97.6)	**99.4**	**99.5**	97(92.7)	97(**99.1**)	**99.5**	96	95.9(96.5)	**99.4**/528
H52-China2	NA	NA	NA	NA	NA	96(**99.1**)	NA	NA	89.9(92.4)	NA
ZJ971	**99.4**(98.7)	**99.7**(99.5)	**100**	**100**	**99.7**(99.1)	**100**(100)	**100**	**100**	**99.8**(99.8)	99.3/525
SAIBK	80.7(79.3)	86.4(88)	80.7	83.6	86.5(82.9)	89.4(93.4)	87.4	94.4	86.2(90.2)	61.4/364
SC021202	78.2/(77.1)	84.2(88.9)	78.2	81	86.9(86)	88.9(93.8)	82.8	93.2	86(91.9)	72.4/503
Ma5	**99.8**(99.6)	NA	NA	NA	NA	NA	NA	NA	NA	NA
Ark99	77.5(75.9)	96(94.2)	83.5	85.6	84(76.6)	NA	NA	NA	NA	72.9/507
ArkDPI	78.8(76.5)	94.8(94.2)	87.4	84.6	86.5(80.2)	94.4(93.4)	93.4	96.8	96.6(96.8)	NA
Conn	91.5(89.6)	NA	86.4	84.1	86.2(79.3)	91.4(93.4)	NA	NA	93.2(94.8)	NA
CU-T2	77.7 (75.4)	90.6(89.7)	79.5	84.6	73.9(76.4)	85.8(84.7)	92.9	96.4	95.9(95.6)	72.5/504
DE072	52.2(51.3)	74.5(74.7)	83	85.6	89.5(82)	**97.4**(**98.7**)	93.4	96.4	92.8(94.1)	72.4/506
Gray	81.7(78.4)	94.9(93.7)	NA	NA	86.2(79.3)	91.2(94.3)	92.9	96.4	92.6(93.6)	72.4/504
HK	**99.8**(99.3)	**98.3**(98.1)	85.8	84.6	86.2(80.2)	91.2(93.4)	94.4	96.4	93.3(95.3)	NA
JILIN	78.9(76.3)	94.8(94)	86.4	84.6	86.5(80.2)	91.6(93.9)	94.4	96.4	93.4(95.8)	54/405
KB8523	**99.4**(99.1)	94(95)	83.5	87.7	91(88.3)	95.2(**97.4**)	92.9	96	92(94.3)	75/503
Vic	NA	NA	NA	NA	85.3(82)	86.5(90.8)	89.9	92	87.9(91.7)	49.9/347

Sequences with ≥97% identity are indicated in bold letters; identity rates of amino acid sequences within the parenthesis

NA not analyzed

[a] Length of 3′-UTR is represented

Table 6 Nucleotide and deduced amino acid difference between H120 and ZJ971

	Nucleotide position	H120	ZJ971	Amino acid substitutions	Gene/protein involved
1	897	T	A	Silent	Nsp2
2	1519	T	C	S → P[ab]	Nsp2
3	1600	G	A	A → T	Nsp2
4	1884	A	G	Silent	Nsp2
5	1993	C	T	Silent	Nsp2
6	2029	G	A	V → I	Nsp2
7	3951	T	C	Silent	Nsp3
8	5832	C	T	Silent	Nsp3
9	6051	C	T	Silent	Nsp3
10	7545	A	G	Silent	Nsp3
11	7572	A	G	I → M	Nsp4
12	8021	A	C	N → T	Nsp4
13	12980	T	C	Silent	Nsp11
14	14522	T	C	Silent	Nsp11
15	15470	A	G	Silent	Nsp12
16	20467	A	G	I → V	S1
17	20677	A	T	I → L	S1
18	20702	C	T	S → F	S1
19	20715	T	C	Silent	S1
20	20793	T	C	Silent	S1
21	21632	C	T	T → I	S1
22	22689	T	C	Silent	S2
23	26103	C	T	Silent	N
24	27114	C	A	Non-coding	3′-UTR
25	27175–27177	ATT	Deletion	Non-coding	3′-UTR

Silent silent mutation

[a] Single-letter codes of corresponding amino acid

[b] →: Amino acid change from H120 to ZJ971

ZJ971 resulting in 9 amino acid changes. Three of the mutations were in nsp2, two in nsp4, and four in S1 including three are located in HVRs ([52]I to V, [122]I to L, [130]S to F). One point mutation and a 3-nt deletion were identified at the 3′-UTR of ZJ971. Which of these mutations are responsible for the virulence reversion from H120 to ZJ971 remains further investigation.

Phylogenetic trees

Phylogenetic trees were constructed based on the sequence alignment of complete or partial genome of 30 IBV strains and four bird coronavirus strains (Table 1). Trees based on different genomic regions exhibited considerably different topologies (Fig. 1). In the phylogenetic tree of complete genome, IBV strains were grouped into five clades (Fig. 1a). H120 was distributed into clade I with H52, ZJ971, Ark DPI 11, Ark DPI 101, and Cal99. Clade II was formed by Beaudette, Peafowl/GD/KQ6/2003 (abbr. to KQ6), and M41. Clades III, IV, and V were exclusively formed by the Chinese strains. Phylogenetic trees based on

different regions of genome were presented in Fig. 1b–i. In the PL[pro]-based tree, H120 clustered with the Ark strains. Whereas in the trees based on structural genes of S1, S2, E, M, H120 grouped with the Mass strains into one clade. In trees based on M[pro] and RdRp, H120 formed one clade with ZJ971, H52, Ma5, and DE072, distinct from clades of Ark and Mass strains. In the N-based tree, H120 clustered with Cal99 and Ark DPI.

Recombination analysis

Potential recombination events

Supported by all the employed recombination detection methods, five strains of IBV including H52-China1, KQ6, SAIBK, Ark DPI 11, and Ark DPI 101 were identified to be possible mosaics with H120 as one of their putative parents (Fig. 2).

H52-China1 was found to be a mosaic between H120 and M41 (Fig. 2a). As indicated in Fig. 2a, the regions of recombination were located at position 3859–4571, 10954–20511,
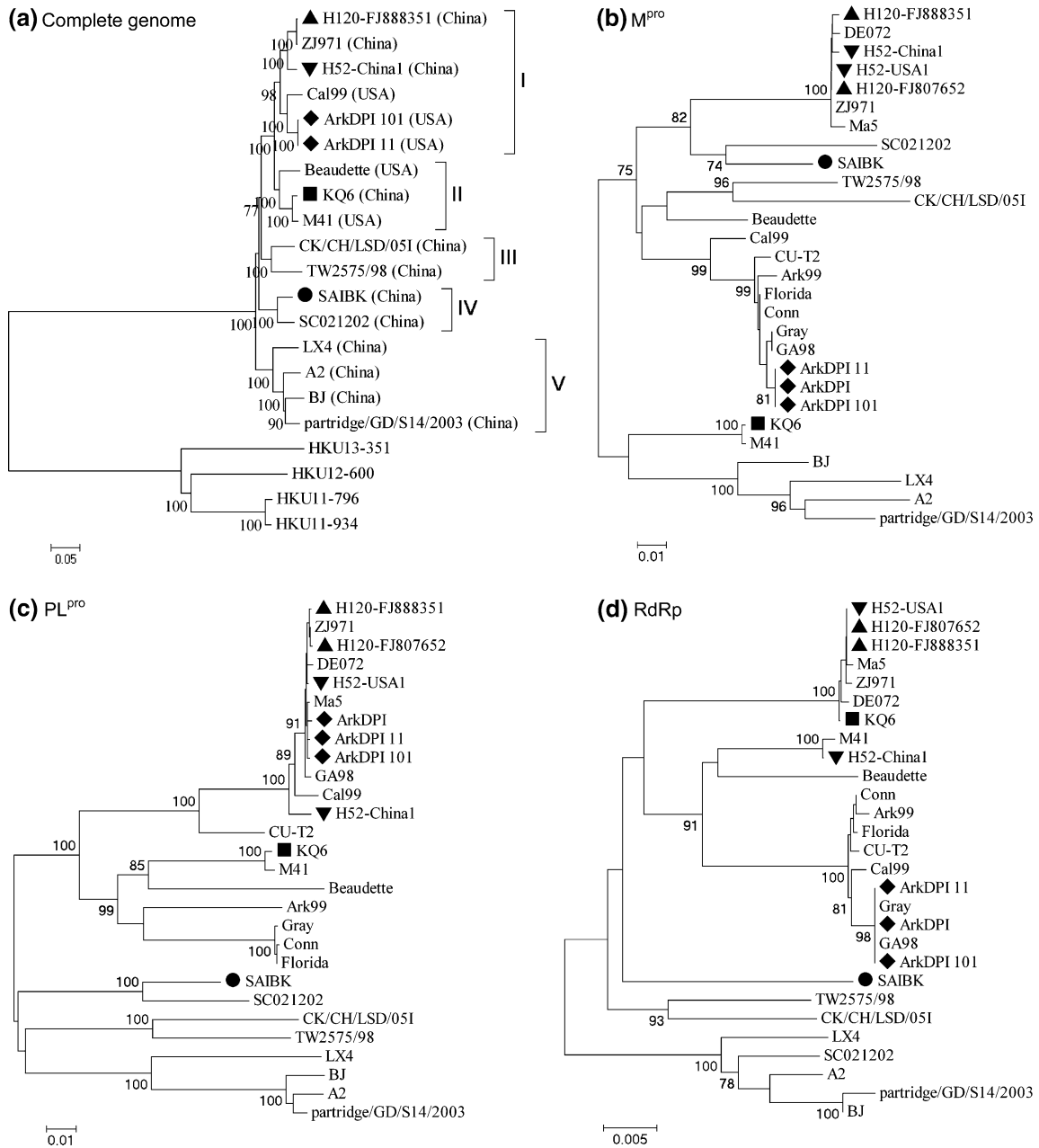
Fig. 1 Phylogenetic trees by neighbor-joining method (bootstrapping for 1000 replicates with its value >70%) based on complete genome and different regions of genome. Sequences of H120 strains were labeled by *filled upright triangle*, sequences of H52 strains were labeled by *filled inverted triangle*, sequences of Ark DPI strains were labeled by *filled diamond*, sequences of KQ6 were labeled by *filled square* and sequences of SAIBK were labeled by *filled circle*

22390–23497, 24791–25919, and 26684–27251 (the light areas). These regions exhibit higher identities with M41 (99.7, 99.7, 99.5, 99.5, and 99.4%, given 5′ to 3′ in turn, similarly hereinafter) and lower identities with H120 (85.8, 93.3, 96.6, 97.2, and 90.2%). Whereas other regions of genome exhibit higher identities with H120 (99.5, 99.7, 99.4, 99.7, 99.7, and 99.8%) and lower identities with M41 (88.8, 88.3, 96.9, 98.1,

87.6, and 77.6%). However, although H120 was identified to be the parent strain, it is more convincing if H52-China1 was developed by recombination between M41 and an ancestral H52 highly similar with H120 based on two reasons: First, as indicated by a previous study on the sequence changes responsible for the attenuation of Ark DPI [22], similarities between the ancestral strains of H lineage should be extremely
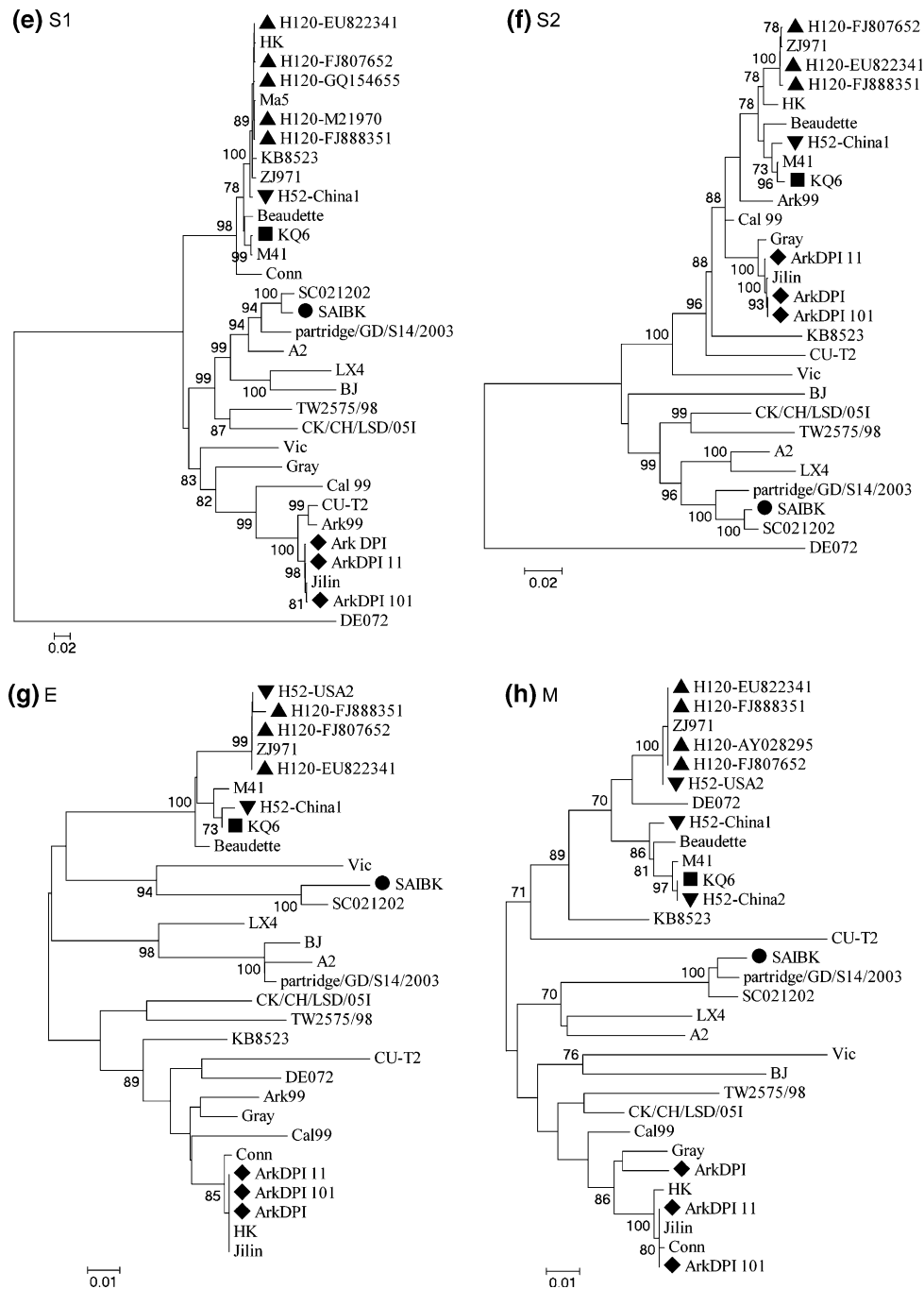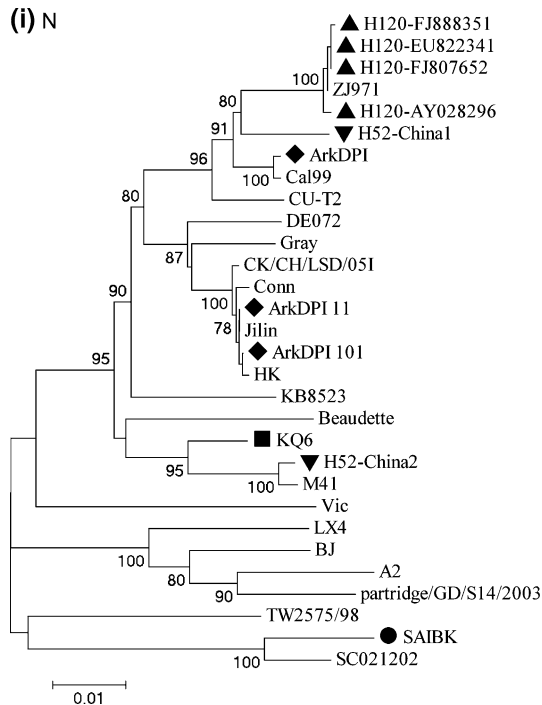
Fig. 1 continued

high. Second, all the sequences of the H isolates except those of the Chinese H52 isolates share high identities, suggesting that genetic changes have occurred in the ancestry of these Chinese H52 isolates.

The IBV strain KQ6, isolated from wild peafowl in Guangdong province of China, was described as a Mass prototype like IBV strain which had existed in peafowl for a long period and evolved independently away from the influence of currently used vaccine in the previous study [23]. However, KQ6 is more likely to be a mosaic in our current study (Fig. 2b). Interestingly, its putative parents were also identified to be H120 and M41, indicating that

Fig. 1 continued

H52-China1 and KQ6 might be developed in a similar way related with vaccine manufacturing. As indicated in Fig. 2b, five H120-like regions in KQ6 were located at position 683–1473, 2775–3571, 14220–15159, 15873–16248, and 26002–27237 (the light areas). These regions exhibit higher identities with H120 (99.6, 99.7, 99.6, 100, and 99.5%) and lower identities with M41 (91.5, 77.4, 92.2, 91.7, and 88.8%). Whereas other regions of KQ6 genome exhibit higher identities with M41 (99.3, 99.7, 99.3, 99.3, 99.6, and 99.4%) and lower identities with H120 (95.8, 88.8, 90.1, 93.1, 95.4, and 77.7%).

In addition, H120-like regions were also found in SAIBK and Ark DPI 11/Ark DPI 101. The H120-like region of SAIBK, which is 99.1% identical to H120, was located at position 7431–9451 (the light area in Fig. 2c). While other regions of genome showed lower identities with H120 (86.1 and 87.7%). As shown in Fig. 2d, the H120-like regions of the Ark DPI strains were located at position 1–8832 with 99.6% identity and the other part of genome exhibit 92.1% with H120. The other parents of SAIBK and the Ark DPI strains were identified to be unknown IBV strains, which have comparatively close relationships with the strains of SC021202 and M41, respectively.

Specifically, several regions of H52-China1, including the RdRp domain, 5′-terminal 826 nt of S2, 5′-terminal 612 nt of M and E gene, were found to be involved in the

regions of recombination (Fig. 2a). As a result, in the phylogenetic trees based on RdRp, S2, M and E, H52-China1 clustered away from other isolates of H strains and clustered with M41 instead. Besides, 5′-terminal 692 nt of RdRp domain of KQ6 and the PL^pro domain of Ark DPI 11/Ark DPI 101 are also located in the regions of recombination, and with a similar pattern to H52-China1, their phylogenies in trees of corresponding regions have been affected by the recombination events.

In previous studies, vaccine strains H52 [10], D1466 [11], Conn [24], and Ark DPI [25] were reported to act as the "heterologous RNA donor template" in recombination events. However, those studies focused only on the recombination in 3′-terminal 7 kb of the genome. Our present study firstly analyzed potential recombination events in the entire genome of IBV and found proofs demonstrating that recombination events also occurred frequently in the 5′-terminal 20 kb of the genome.

Recombination hot spots

Recombination in coronavirus is generally believed to occur via template switching mechanism [26, 27], and it is predicted that recombination occurs more frequently at RNA sites of strong secondary structure, since these structures promote transcriptional pausing [28]. Potential recombination "hot spots" are reported to be located adjacent to the putative recombinant junctions, and all of these "hot spots" are described to be AT-rich motifs, e.g., CT(T/G)AACAA [10], (A/T/G)TTTTG, CTTTTG [26, 29]. However, a previous report indicated that the crossover sites were distributed almost randomly when recombination of coronavirus was examined under non-selective conditions [30]. In current study, AT-rich motifs (TTTTG or T(T/A)(G/T)AACAA) were found in regions adjacent to nearly all the putative crossover sites. Even though, it is still difficult to determine whether these motifs are related to the recombination or they just appear around the recombinant sites by chance, because these motifs are detected frequently (119 times for TTTTG and 50 times for AACAA) in the genome of H120.

In summary, this study described the features of the genome of H120 and showed the putative evidences of virulence reversion and recombination associated with H120, indicating that IBV attenuated vaccine H120 might contribute to the emergence of new IBV variants through both virulence reversion and recombination. Although more proofs are needed to verify the putative results of this study, great attention should be paid to supervise the use of IBV attenuated live vaccines.

Fig. 2 RDP screenshots displaying the possible recombination events associated with H120. Each *panel* displays the pairwise identities among the possible mosaic and its putative parents. Pairwise identity refers to the average pairwise sequence identity within a 30nt sliding window moved one nucleotide at a time along the alignment of the three sequences. The light area demarcates the potential recombination regions. Crossover sites were indicated by *arrows* with nt positions above. **a** Comparisons among the putative mosaic H52-China1 and its putative parents, H120 and M41. **b** Comparisons among the putative mosaic KQ6 and its putative parents, H120 and M41. **c** Comparisons among the putative mosaic SAIBK and its putative parents, H120 and SC021202. **d** Comparisons among the putative mosaics Ark DPI 11/Ark DPI 101 and their putative parents, H120 and M41. The RDP result of Ark DPI 101 was not shown independently because it was identical to that of Ark DPI 11
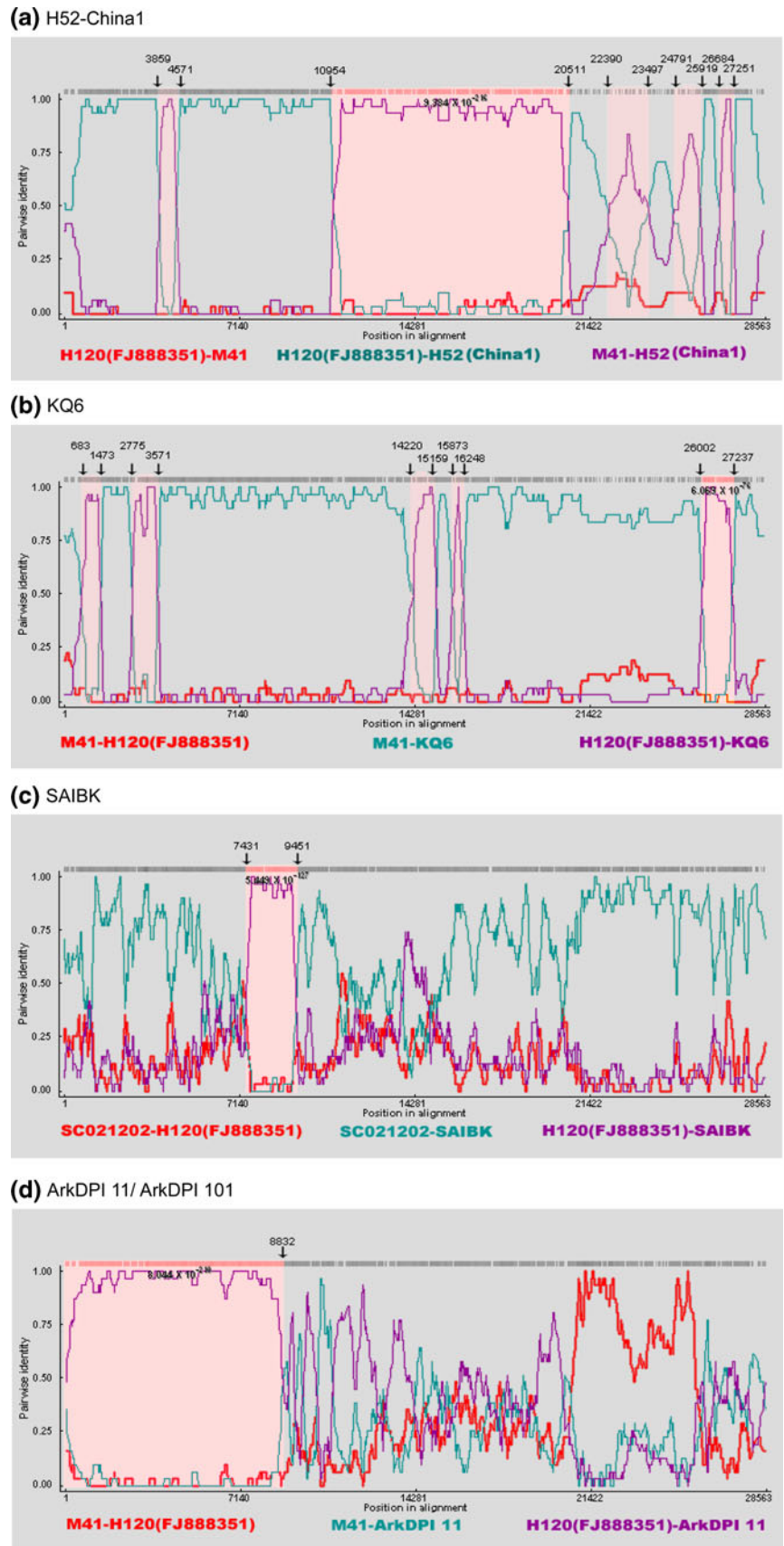
References

1. M.H. Gomaa, J.R. Barta, D. Ojkic, D. Yoo, Virus Res. **135**, 237–246 (2008)
2. B. Lucio, J. Fabricant. Avian Dis. **34**, 865–870 (1990)
3. R.L. Parr, E.W. Collisson, Arch. Virol. **133**, 369–383 (1993)
4. D.A. Boltz, M. Nakai, J.M. Bahra, Avian Dis. **48**, 909–915 (2004)
5. M.E.G. Boursnell, T.D.K. Brown, I.J. Foulds, P.F. Green, F.M. Tomely, M.M. Binns, J. Gen. Virol. **68**, 57–77 (1987)
6. D.F. Stern, B.M. Sefton, J. Virol. **50**, 22–29 (1984)
7. I. Brierley, P. Digard, S.C. Inglis, Cell **57**, 537–547 (1989)
8. M. Armesto, D. Cavanagh, P. Britton, PLoS ONE **4**, e7384 (2009)
9. D. Cavanagh, P.J. Davis, J.K.A. Cook, D. Li, A. Kant, G. Koch, Avian Pathol. **21**, 33–43 (1992)
10. W. Jia, K. Karaca, C.R. Parrish, S.A. Naqi, Arch. Virol. **140**, 259–271 (1995)
11. C.W. Lee, M.W. Jackwood, Arch. Virol. **145**, 2135–2148 (2000)
12. S. Liu, J. Chen, Z. Han, Q. Zhang, Y. Shao, X. Kong et al., Avian Pathol. **35**, 394–399 (2006)
13. S. Liu, X. Kong, Avian Pathol. **33**, 321–327 (2004)
14. J. Gelb, Y. Weisman, B.S. Ladman, R. Meir, Avian Pathol. **34**, 194–203 (2005)
15. G. Bijlenga, J.K. Cook, J. Gelb, J.J. de Wit, Avian Pathol. **33**, 550–557 (2004)
16. K. Tamura, J. Dudley, M. Nei, S. Kumar, Mol. Biol. Evol. **24**, 1596–1599 (2007)
17. D.P. Martin, E. Rybicki, Bioinformatics **16**, 562–563 (2000)
18. M. Padidam, S. Sawyer, C.M. Fauquet, Virology **265**, 218–225 (1999)
19. S.J. Maynard, J. Mol. Evol. **34**, 126–129 (1992)
20. D.P. Martin, C. Williamson, D. Posada, Bioinformatics **21**, 260–262 (2005)
21. J. Zhou, L. Cheng, X. Shen, H. Ding, J. Wu, Agric. Sci. China **1**, 101–107 (2002). (in Chinese)
22. A. Ammayappan, C. Upadhyay, J. Gelb, V.N. Vakharia, Arch. Virol. **154**, 495–499 (2009)
23. L. Sun, G. Zhang, J. Jiang, J. Fu, T. Ren, W. Cao et al., Virus Res. **130**, 121–128 (2007)
24. A. Ammayappan, C. Upadhyay, J. Gelb, V.N. Vakharia, Virol. J. **5**, 157 (2008)
25. S.P. Mondal, C.J. Cardona, Virus Genes **34**, 327–341 (2007)
26. J.G. Kusters, E.J. Jager, H.G. Niesters, B.A. van der Zeijst, Vaccine **8**, 605–608 (1990)
27. M.M.C. Lai, Curr. Top. Microbiol. Immunol. **176**, 21–32 (1992)
28. D.R. Mills, C. Dobkin, F.R. Kramer, Cell **15**, 541–550 (1978)
29. J.E. Brooks, A.C. Rainer, R.L. Parr, P. Woolcock, F. Hoerr, E.W. Collisson, Virus Res. **100**, 191–198 (2004)
30. L.R. Banner, M.M.C. Lai, Virology **185**, 441–445 (1991)