Journal of General Virology (2005), **86**, 423–433 **DOI** 10.1099/vir.0.80368-0

Molecular characterization of the complete genome of human influenza H5N1 virus isolates from Thailand

Pilaipan Puthavathana,[1] Prasert Auewarakul,[1] Pakapak Chor Charoenying,[1] Kantima Sangsiriwut,[1] Phisanu Pooruk,[1] Kobporn Boonnak,[1] Raweewan Khanyok,[1] Pranee Thawachsupa,[2] Rungrueng Kijphati[2] and Pathom Sawanpanyalert[2]

[1]Department of Microbiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

[2]National Institute of Health, Department of Medical Sciences, Ministry of Public Health, Nonthaburi 11000, Thailand

Correspondence
Pilaipan Puthavathana
siput@mahidol.ac.th

Received 16 June 2004
Accepted 25 October 2004

The complete genomes of three human H5N1 influenza isolates were characterized, together with the haemagglutinin (HA) and neuraminidase (NA) genes from two additional human isolates and one chicken isolate. These six influenza isolates were obtained from four different provinces of Thailand during the avian influenza outbreak in Asia from late 2003 to May 2004. All six Thailand isolates contained multiple basic amino acids at the cleavage site in the HA gene. Amino acid residues at the receptor-binding site of the five human viruses were similar to those of the chicken virus and other H5N1 viruses from Hong Kong. The presence of amantadine resistance in the Thailand viruses isolated during this outbreak was suggested by a fixed mutation in M2 and confirmed by a phenotypic assay. All genomic segments of the Thailand viruses clustered with the recently described genotype Z. The Thailand viruses contained more avian-specific residues than the 1997 Hong Kong H5N1 viruses, suggesting that the virus may have adapted to allow a more efficient spread in avian species.

INTRODUCTION

Influenza viruses of aquatic birds have been proposed as the ancestors of all influenza virus subtypes existing in humans and other animals. They are classified into subtypes based on antigenic differences in the two surface glycoproteins, the haemagglutinin (HA) and neuraminidase (NA). Currently, 15 HA subtypes and nine NA subtypes of influenza A virus are known, and all of these subtypes can be found in aquatic birds. Genetic variation of the HA and/or NA genes resulting in the emergence of new influenza virus strains has frequently been recognized. The mechanism of variation can be either ‘genetic drift’, which occurs as a result of point mutations, or ‘genetic shift’, which occurs as a result of genetic reassortment between two strains of influenza A virus. Influenza A virus contains eight segments of a single-stranded RNA genome with negative polarity. The segmentation of the genome thus makes genomic reassortment possible when two strains co-infect a permissive cell (Nicholson *et al.*, 2003; Zambon, 2001).

The H2N2 and H3N2 influenza A viruses that caused the human pandemics of 1957 and 1968, respectively, were suggested to be new reassortants that had acquired genetic material from avian viruses. However, it was proposed that influenza virus could not transmit directly from avian species to humans, but required the pig as an intermediate host or a ‘mixing vessel’ to recombine the avian virus with the pre-existing human influenza virus and generate a new reassortant, which was subsequently transmitted to the human population (Zambon, 2001). Direct transmission of avian influenza virus from domestic poultry to humans was demonstrated for the first time by the outbreak of the highly pathogenic avian H5N1 virus in Hong Kong in 1997 (Claas *et al.*, 1998; Yuen *et al.*, 1998) and subsequently confirmed by human infection caused by avian H9N2 influenza viruses in Hong Kong in 1999 (Peiris *et al.*, 1999), by avian H7N7 viruses in the UK in 1996 (Kurtz *et al.*, 1996) and in The Netherlands in 2003 (Koopmans *et al.*, 2004), and again by avian H5N1 viruses in Hong Kong in February 2003 (Peiris *et al.*, 2004).

The Hong Kong influenza H5N1 virus that infected humans in 1997 was believed to be a reassortant virus that had acquired the HA gene from A/goose/Guangdong/1/96

The GenBank accession numbers of the sequences determined in this work are AY626143–AY626149, AY627885–AY627898, AY555150–AY555153, AY577314–AY577316, AY649382 and AY649383.

(H5N1)-like viruses, the NA gene from A/teal/Hong Kong/W312/97(H6N1)-like viruses and the internal genes from A/quail/Hong Kong/G1/97 (H9N2)-like or A/teal/Hong Kong/W312/97 (H6N1)-like viruses (Hatta & Kawaoka, 2002; Subbarao & Shaw, 2000). Even though the Hong Kong/1997 (H5N1) viruses were eliminated by the culling of millions of poultry, its ancestors remained and have gone through several reassortment events in avian species and generated new genotypes, with reemergence of the disease from time to time thereafter (Guan *et al.*, 2002; Hatta & Kawaoka, 2002). In 2001, avian H5N1 isolates were classified into five genotypes designated A–E (Guan *et al.*, 2002). After 2002, these genotypes were not detected and eight new genotypes – V, W, X1, X2, X3, Y, Z and Z^+ – emerged (Li *et al.*, 2004).

The reemergence of H5N1 influenza A viruses affected a large area of East and Southeast Asia for the first time from late 2003 to May 2004. At least eight countries were involved. According to official reports, the outbreak began in South Korea and then spread to Vietnam, Japan, Thailand, Cambodia, China, Laos and Indonesia (CDC/WHO Avian Influenza Response Team, 2004). However, only Vietnam and Thailand reported the disease in humans (Chotpitayasunondh *et al.*, 2004; Hien *et al.*, 2004).

It was presumed that the H5N1 viruses responsible for the outbreaks in these countries in 2003–2004 had the same origin, and they were antigenically distinguishable from the 1997 and early 2003 viruses (CDC/WHO Avian Influenza Response Team, 2004). Nevertheless, no solid information has yet been published. The present study performed nucleotide and amino acid sequence analyses of all eight genomic segments of three human H5N1 viruses, as well as the HA and NA of two additional human viruses and a chicken virus from Thailand. These data provide an insight into the origin of the reemergence and evolution of the H5N1 viruses.

METHODS

H5N1 virus isolation and identification. Throat swab or nasopharyngeal wash/aspirate samples from patients with severe pneumonia and with a history of exposure to poultry during the avian influenza outbreak of 2003–2004 were collected in virus transport medium. Sample tubes were mixed thoroughly before spinning at 1000 ***g*** for 15 min at 4 °C. Supernatants were then inoculated onto rhesus monkey kidney (LLC-MK2) and/or Madin–Darby canine kidney (MDCK) cells for virus isolation. Virus isolates were identified by staining the inoculated cultures with an influenza A-specific monoclonal antibody test set (Chemicon), followed by FITC-conjugated goat anti-mouse IgG (Chemicon) as the second antibody in an indirect immunofluorescence test. For H5 subtyping, the influenza viruses present in culture supernates were extracted for RNA using the QIAamp Viral RNA mini kit (Qiagen). The extracted RNA was tested by RT-PCR using the type-specific primers HA-1144 and H5-1735R (Webster *et al.*, 2002), which generated amplified products of 591 bp. N1 subtyping was performed by nucleotide sequencing as described below.

Geographical distribution of the H5N1 strains. This study included five human and one chicken H5N1 isolates designated A/Thailand/1(KAN-1)/2004 (H5N1), A/Thailand/2(SP-33)/2004 (H5N1), A/Thailand/3(SP-83)/2004 (H5N1), A/Thailand/4(SP-528)/2004 (H5N1), A/Thailand/5(KK-494)/2004 (H5N1) and A/Chicken/Thailand/CH-2/2004 (H5N1). A/Thailand/1(KAN-1) virus was isolated from a 6-year-old boy from Kanchanaburi, a province located in the west of Thailand, approximately 128 km west of Bangkok. A/Thailand/2(SP-33) virus was isolated from a 7-year-old boy, A/Thailand/3(SP-83) was from a 58-year-old woman and A/Thailand/4(SP-528) was from a 1-year-old boy. These three patients lived in different areas of the Supanburi Province located in central Thailand at a distance of about 100 km northwest of Bangkok. A/Thailand/5(KK-494) virus was isolated from a 4-year-old boy in Khon Kaen, a province in the northeast of the country, about 449 km northeast of Bangkok. A/Chicken/Thailand/CH-2 virus was isolated from a chicken in a poultry farm in Chachoengsao, a province in the central part of the country, approximately 82 km east of Bangkok. The locations of these provinces are shown in Fig. 1.

Amantadine-sensitivity assay. Resistance of the Thailand viruses to amantadine was assayed in MDCK cell cultures. The three H5N1 Thailand viruses A/Thailand/1(KAN-1), A/Thailand/2(SP-33) and A/Thailand/3(KK-494), together with A/New Caledonia/20/99(H1N1)-like and A/Sydney/05/97(H3N2)-like viruses, each at titres of 25, 50 and 100 $TCID_{50}$, were tested against concentrations of 0·3, 1·6 and 8·0 μg amantadine ml^{-1} (Sigma-Aldrich) in microtitre plates by checkerboard titration. The MDCK cell monolayer was grown in Eagle's minimal essential medium [Grand Island Biological Company (Gibco)] supplemented with 10 % fetal bovine serum (Hyclone) plus penicillin and streptomycin. Cells were washed and pre-incubated with each drug concentration for 1 h. The experiments were run in triplicate. In a parallel plate, each titre of each virus isolate was incubated with the three different drug concentrations for 1 h at 4 °C. The drug-containing medium in the

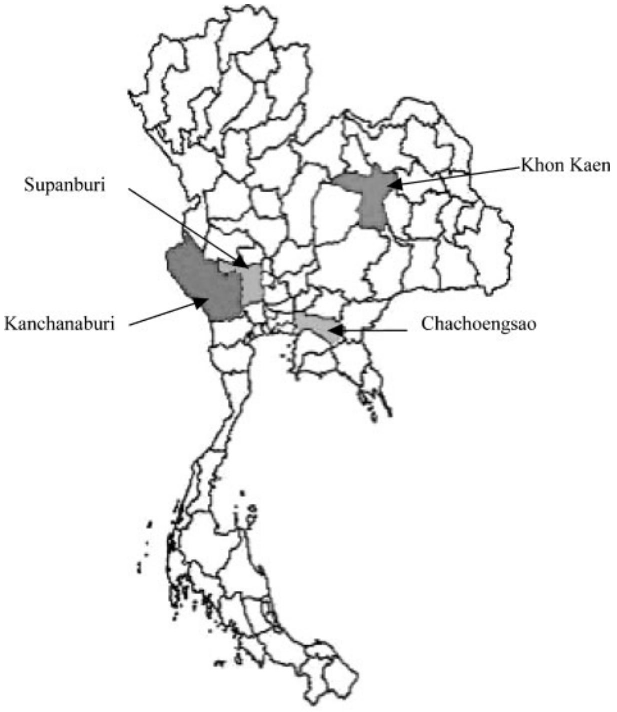


Fig. 1. Location of the provinces of origin of the Thailand H5N1 viruses.

pre-incubated cell-culture plate was replaced by the drug-treated virus solution from a parallel plate and further incubated for 1 h at 37 °C. At the end of the incubation time, the inoculated plates were washed and growth medium containing trypsin-TPCK (Gibco) at concentration of 2 μg ml^{-1} was added, together with the desired concentration of drug. The plates were incubated at 37 °C for 18 h.

After infection for 18 h, the amount of influenza virus present in the infected-cell monolayer was determined by measurement of the amount of viral nucleoprotein (NP) by ELISA. Plates of infected-cell monolayers were fixed with 80 % acetone overnight and endogenous peroxidase activity was blocked by incubation with 3 % H_2O_2. The infected-cell monolayers were then incubated with monoclonal antibody to influenza A virus NP (BV European Veterinary Laboratory) for 1 h, before washing and incubating with goat anti-mouse Ig conjugated to horseradish peroxidase (Southern Biotechnology Associates) for 1 h at 37 °C. 3,3′,5,5′-Tetramethylbenzidine peroxidase substrate (KPL) was added for 15 min at room temperature before stopping the reaction with 1 M H_2SO_4. The OD_{450} was measured. The mean OD of the infected cultures without drug was equal to the total amount of NP protein (100 %). Cultures infected with amantadine-sensitive virus should produce less than 50 % of the total NP protein.

RNA extraction. Virus propagated in LLC-MK2 or MDCK cells for one or two passages was used as the source of viral RNA for nucleotide sequencing. RNA was extracted from the culture supernatants using the QIAamp Viral RNA mini kit (Qiagen).

Gene amplification. RT-PCR was performed using the One-step RNA PCR kit (AMV) (Takara Bio) to amplify the PB2, PB1, PA and HA genes, while the Qiagen One-step RT-PCR kit was used to amplify the NP, NA, M and NS genes. PCR primers in this study were as described by Hoffmann *et al.* (2001) as follows: Ba-PB2-1 and Ba-PB2-2341R for the PB2 gene; Bm-PB1-1 and Bm-PB1-2341R for the PB1 gene; Bm-PA-1 and Bm-PA-2233R for the PA gene; Bm-HA-1 and Bm-NS-890R for the HA gene (the HA reverse primer was identical to the NS reverse primer); Bm-NP-1 and Bm-NP-1565R for the NP gene; Ba-NA-1 and Ba-NA-1413R for the NA gene; Bm-M-1 and Bm-M-1027R for the M gene; and Bm-NS-1 and Bm-NS-890R for the NS gene. The sizes of the amplified products for PB2, PB1, PA, HA, NP, NA, M and NS were 2341, 2341, 2233, 1778, 1565, 1413, 1027 and 890 bp, respectively, plus 29 mer oligonucleotide primer for all products.

Nucleotide sequencing. The amplified DNA products were electrophoresed in 1 % agarose gel and pieces of gel containing DNA bands of the expected sizes were purified using the Qiagen Gel Extraction kit. The purified DNA was used as the target for direct nucleotide sequencing using a Cycle Sequencing kit (BigDye Terminator version 3.1; Applied Biosystems), followed by analysis in an ABI PRISM version 310 or 377 DNA sequencer (PE Applied Biosystems).

Sequence analysis. Nucleotide sequences were edited using the Chromas version 2.23 program (Technelysium Pty Ltd; available at http://www.technelysium.com.au/chromas.html). Alignment and homology analysis were performed using BioEdit v. 5.0.6 (T. Hall, Ibis Therapeutics, Isis Pharmaceuticals; available at http://www.mbio.ncsu.edu/BioEdit/bioedit.html), CLUSTAL X v. 1.8 (Thompson *et al.*, 1997; available at ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) and GeneDoc v. 2.6.02 (Nicholas *et al.*, 1997; available at http://www.psc.edu/biomed/genedoc). The phylogram was drawn using PHYLIP v. 3.5c (Felsenstein, 1993; available at http://evolution.genetics.washington.edu/phylip.html) and TreeView v. 1.6.6 (Page, 1996; available at http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). All sequences have been deposited in GenBank. The accession numbers are provided in Table 1.

RESULTS AND DISCUSSION

Amino acid sequences indicating pathogenicity of the Thailand viruses

All of the five human H5N1 isolates and one chicken isolate from Thailand contained the amino acid sequence PQRERRRKKR ↓ G at the cleavage site in the HA molecule, indicating their high virulence (Fig. 2) (Horimoto & Kawaoka, 1994). This cleavage sequence was as the same as that of A/Hong Kong/156/97(H5N1) virus (Subbarao *et al.*, 1998).

Cleavage of the HA molecule (HA0), by host-cell proteases, into two disulphide-linked HA1 and HA2 subunits is essential for viral infectivity. Avian influenza viruses with high and low levels of pathogenicity differ in their cleavage sequence: the former possess multiple basic amino acid residues, while the latter do not. A cleavage sequence containing several basic amino acids is more readily activated by cellular proteases present in a variety of cells distributed throughout the body compared with a cleavage sequence containing only a single basic amino acid, which can be cleaved by a limited range of cellular proteases. It is well accepted that influenza viruses containing multiple basic amino acids have multiple sites of virus replication and produce more severe infection in birds and mammals (Zambon, 2001).

The Hong Kong/1997 H5N1 viruses isolated from humans were found to be highly lethal in birds and ferrets, but were

Table 1. GenBank accession numbers of the H5N1 influenza A viruses from Thailand

Virus strain	Gene accession no.							
	HA	NA	PB2	PB1	PA	NP	M	NS
A/Thailand/1(KAN-1)/2004 (H5N1)	AY555150	AY555151	AY626149	AY626148	AY626147	AY626145	AY626144	AY626146
A/Thailand/2(SP-33)/2004 (H5N1)	AY555153	AY555152	AY627898	AY627897	AY627896	AY627895	AY627893	AY627894
A/Thailand/3(SP-83)/2004 (H5N1)	AY577314	AY577315	–	–	–	–	–	–
A/Thailand/4(SP-528)/2004 (H5N1)	AY626143	AY577316	–	–	–	–	–	–
A/Thailand/5(KK-494)/2004 (H5N1)	AY627885	AY627886	AY627892	AY627891	AY267890	AY627889	AY627887	AY627888
A/Chicken/Thailand/CH-2/2004 (H5N1)	AY649382	AY649383	–	–	–	–	–	–

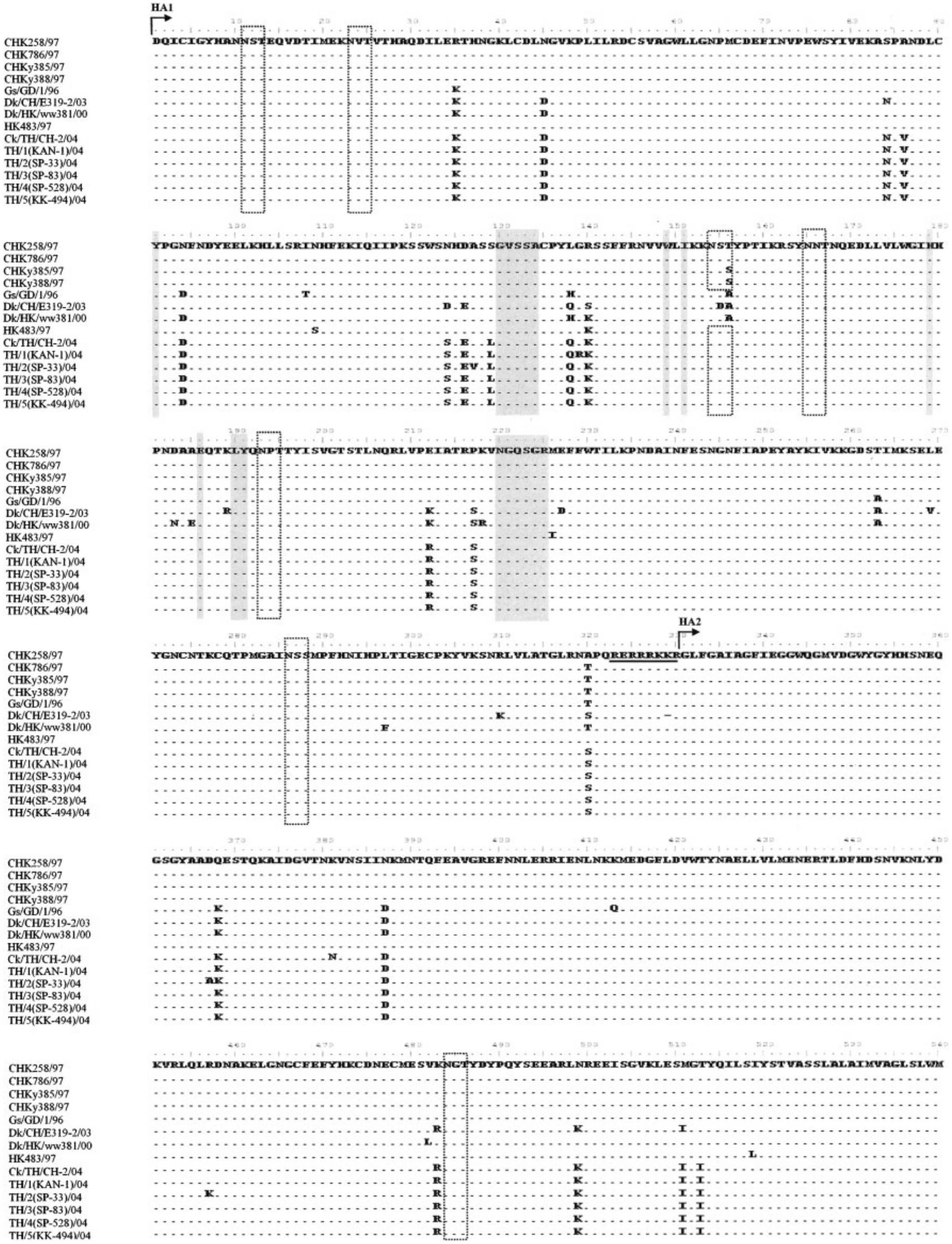
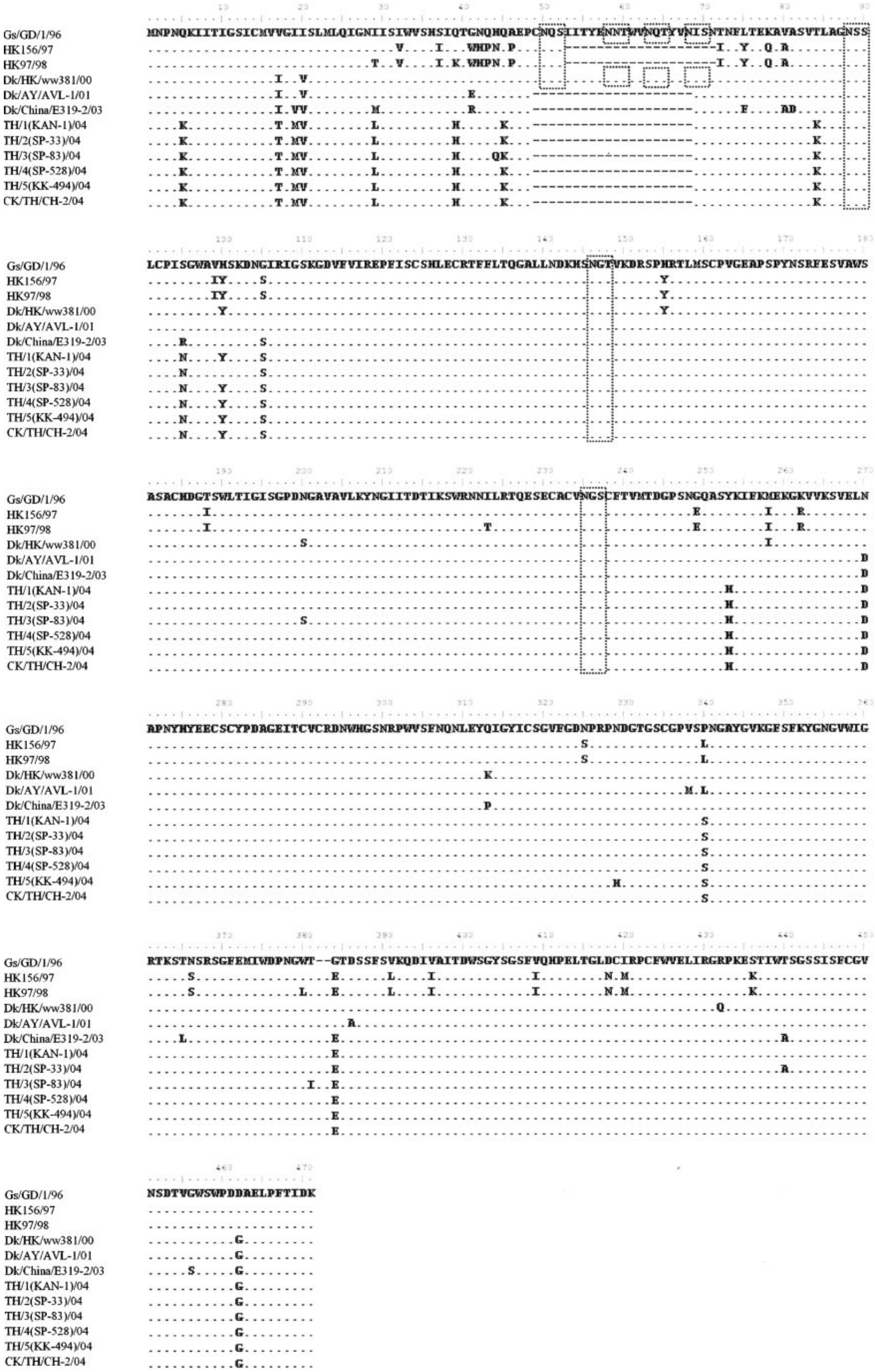


Fig. 2. Alignment of the predicted amino acid sequences of the HA protein. Glycosylation sites are marked by open boxes, residues involved in receptor binding are shaded and the cleavage site between HA1 and HA2 is underlined.

Fig. 3. Alignment of the predicted amino acid sequence of the NA protein. Glycosylation sites are marked by open boxes.


Gs/GD/1/96
HK156/97
HK97/98
Dk/HK/ww381/00
Dk/AY/AVL-1/01
Dk/China/E319-2/03
TH/1(KAN-1)/04
TH/2(SP-33)/04
TH/3(SP-83)/04
TH/4(SP-528)/04
TH/5(KK-494)/04
CK/TH/CH-2/04

either of high or low pathogenicity in mice (Gao *et al.*, 1999). This information suggested that factors other than the presence of multiple basic amino acids in the cleavage sequence contributed to the severity of H5N1 viruses in mice. A reverse genetics study has shown that a single amino acid substitution, from glutamic acid to lysine at aa 627 on the PB2 molecule (RNA polymerase subunit), was associated with virulence of H5N1 viruses in mice (Hatta *et al.*, 2001; Shinya *et al.*, 2004). H5N1 viruses with lysine at this position (HK483RG and HK6PB2-627K) showed systemic spreading and neurovirulence in mice, while viruses with glutamic acid at this position (HK3PB2-627E and HK486RG) showed non-lethal infection that was limited to the lung (Shinya *et al.*, 2004). The same mutation was found in an H7N7 virus from a patient who died from severe pneumonia, but not in patients with conjunctivitis during an outbreak in Europe in 2003 (Fouchier *et al.*, 2004). Among the three Thailand viruses that were fully sequenced, glutamic acid at aa 627 in the PB2 protein was found in one sequence, while lysine was found in the other two. Interestingly, all of these viruses were from fatal cases. In Thailand, there was a total of 12 laboratory-confirmed cases with eight deaths occurring during this outbreak. The fatality rate in the Thai patients was 66·7 %, while it was 80 % (8 of 10 cases) for the Vietnamese patients (Hien *et al.*, 2004). These figures were much higher than those reported from the Hong Kong outbreak in 1997 (33 %, 6 of 18 cases) (Nicholson *et al.*, 2003). The reason for this higher level of fatality is unclear.

One additional glycosylation site in the HA protein and a deletion of 20 aa in the NA protein are present in the Thailand viruses

The Thailand viruses, as well as other genotype Z viruses, contained a 20 aa deletion in the stalk of the NA molecule. The position of the deletion was close but not identical to the 19 aa deletion found in the Hong Kong/1997 viruses. As a result of this difference, genotype Z viruses have lost an *N*-linked glycosylation site that was located just upstream of the deletion site in the 1997 viruses (Fig. 3). The NA protein plays an essential role in the release of viral progeny from infected cells, probably through the cleavage of glycosidic linkages to sialic acids on HA molecules of the progeny virus particles. It has been proposed that deletion of amino acid residues may impair the ability of NA to release the progeny virus particles (Matrosovich *et al.*, 1999). In eggs, the length of the NA stalk correlates with the efficiency of virus replication: the longer the stalk, the more efficient the replication (Castrucci & Kawaoka, 1993). Despite a predicted inefficiency due to its short NA stalk, the H5N1 viruses studied here replicated and spread efficiently, suggesting that there might be some additional compensatory changes resulting in the short NA stalk being functional. It was previously suggested that an additional glycosylation site at the head of the HA molecule at aa 156 might compensate for the shortening of the NA stalk. Furthermore, the fact that these two changes coincided with the spread of these strains in land-based poultry raises a possibility that these two changes may have resulted in adaptation of the viruses for more efficient spread in terrestrial domestic poultry (Matrosovich *et al.*, 1999). Both the 1997 viruses and the genotype Z viruses of the present outbreak have these two mutations, while both the parental goose/Guangdong (Gs/Gd)-like viruses and most viruses isolated between the two outbreaks do not.

The Thailand viruses are amantadine resistant

The H5N1 viruses that caused the 2003–2004 outbreak in Southeast Asia were previously reported to be resistant to amantadine and rimantadine (CDC/WHO Avian Influenza Response Team, 2004). Collectively, there were four critical sites associated with amantadine resistance. Single amino acid substitutions at aa 26 (Leu→Phe), 27 (Val→Ala or Thr), 30 (Ala→Thr or Val) and 31 (Ser→Asn or Arg) in the transmembrane region of the M2 protein may confer drug resistance (Suzuki *et al.*, 2003). In accordance with this finding, analysis of M2 protein amino acid sequences showed that the Thailand isolates contained asparagine instead of serine at aa 31, indicating resistance to amantadine and rimantadine. The other mutation found in M2 was the substitution of isoleucine for leucine at aa 26, which may not play a critical role in drug resistance because of the similarity of the two amino acids.

Amantadine resistance of the Thailand viruses was confirmed by a drug-sensitivity assay in MDCK cell cultures. All three H5N1 Thailand viruses were resistant to amantadine. The highest concentration of 8·0 $\mu g\ ml^{-1}$ could not inhibit the H5N1 viruses at all titres assayed: 25, 50 and 100 $TCID_{50}$ (Fig. 4 and data not shown). In contrast, A/New Caledonia/20/1999 (H1N1)-like viruses at a titre of 100 $TCID_{50}$ could be inhibited almost completely by

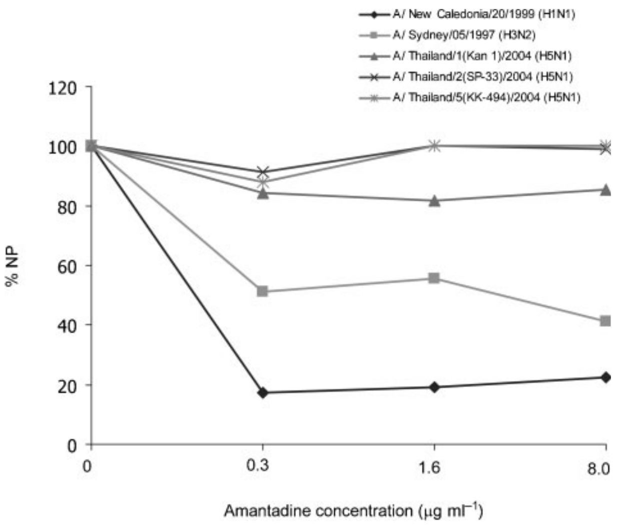


Fig. 4. Amantadine sensitivity assay in MDCK cell cultures. Viruses were used at a titre of 100 $TCID_{50}$.

Table 2. Host-specific amino acid residues in influenza proteins

Protein	Amino acid position	Predicted amino acid*		Amino acid present in Thailand isolates		
		Avian	Human	TH/1(KAN-1)	TH/2(SP-33)	TH/5(KK-494)
M1	137	T	A	T	T	T
M2	16	E	G	E	E	E
	20	S/N	N	S	S	S
	28	I	I/V	V	V	V
	55	L	F	L	L	L
	78	Q	K	Q	Q	Q
NP	33	V	I	V	V	V
	61	I	L	I	I	I
	100	R	V	R	R	R
	136	L	M	L	L	L
	214	R	K	R	R	R
	283	L	P	L	L	L
	293	R	K	R	R	R
	313	F	Y	F	F	F
	375	D	G/E	D	D	D
PA	28	P	L	P	P	P
	55	D	N	D	D	D
	65	S	L	S	S	S
	100	V	A	V	V	V
	382	E	D	E	E	E
	400	Q/T/S	L	S	S	S
	409	S	N	S	S	S
	552	T	S	T	T	T
PB2	44	A	S	A	A	A
	81	T	M	T	T	T
	199	A	S	A	A	A
	271	T	A	T	T	T
	588	A	I	A	A	A
	613	V	T	V	V	V
	661	A	T	A	A	A
	674	A/S	T	A	A	A
	702	K	R	K	K	K

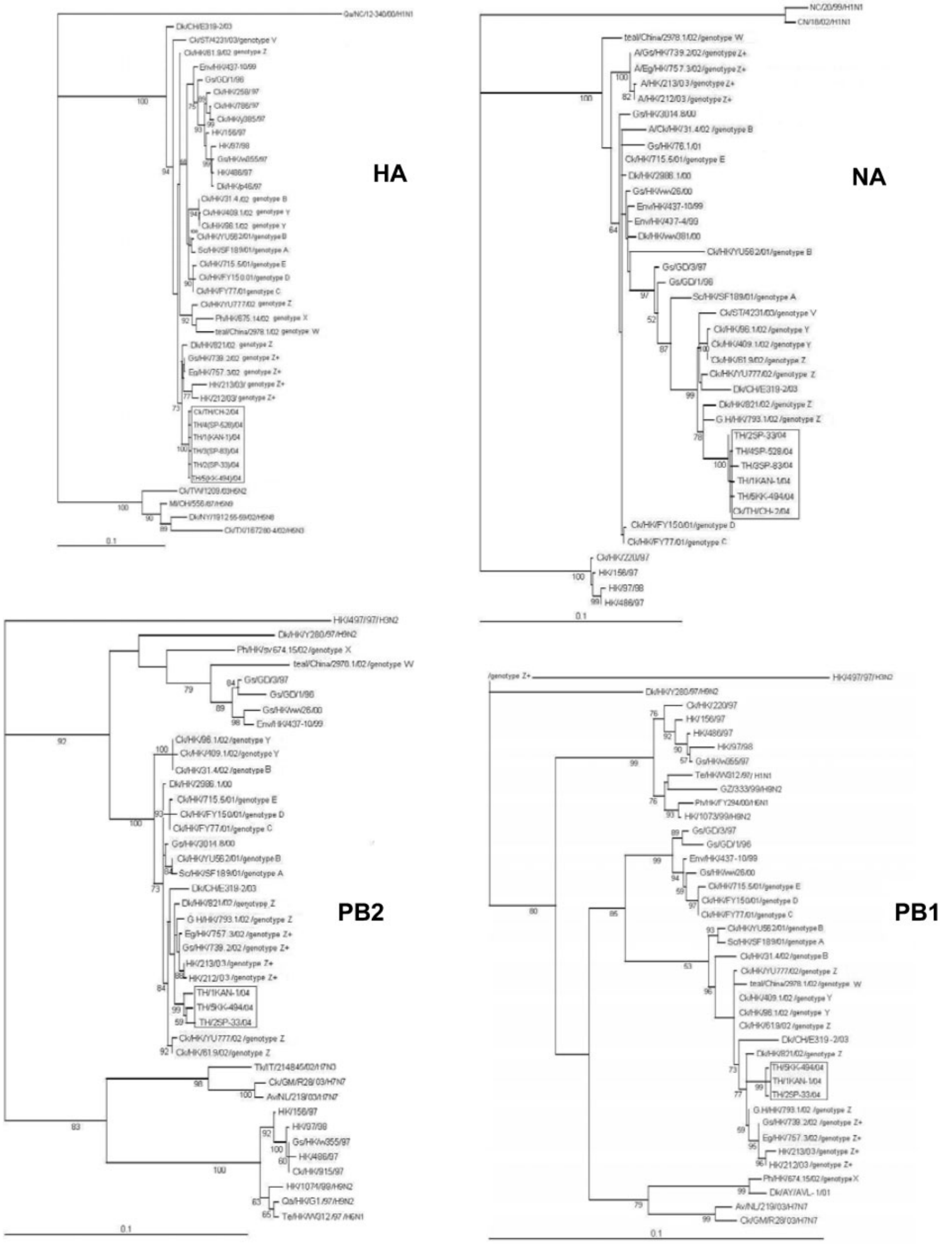
*Predicted amino acids are as described by Shaw *et al.* (2002).

amantadine at a concentration as low as 0·3 μg ml^{-1} (Fig. 4). With A/Sydney/05/1997 (H3N2)-like viruses at a titre of 100 TCID_{50}, a 50 % reduction in the amount of NP protein was seen with a drug concentration of 8 μg ml^{-1} (Fig. 4). It was not clear how amantadine resistance was selected and maintained in the Thailand viruses without selection pressure from the antiviral drug.

Human and chicken viruses from Thailand share the same receptor-binding sites

Host tropism of the Thailand viruses was determined by characterization of amino acid residues at the receptor-binding site. Our sequence analysis showed 100 % identity of amino acid residues at the receptor-binding site among the five human and one chicken isolates from Thailand (Fig. 2). This identity was also shared with the other previously reported H5 viruses including the Hong Kong/1997 H5N1 viruses (Claas *et al.*, 1998; Nobusawa *et al.*, 1991; Weis *et al.*, 1988). In addition, six out of seven *N*-linked glycosylation sites in the HA molecule of the Thailand viruses were conserved compared with the Hong Kong/1997 H5N1 viruses (Fig. 2).

Human influenza viruses preferentially bind to sialyl oligosaccharide receptors with α2,6 linkage to galactose, while avian viruses preferentially bind to cellular receptors with α2,3 linkages (Cross *et al.*, 2001). Despite the ability to infect humans, the Hong Kong/1997 H5N1 viruses still preferred α2,3-linkage receptors, which suggested that transmission from birds to humans did not require a change in receptor-binding preference (Matrosovich *et al.*, 1999). The preference of H5N1 viruses for avian receptors was probably the reason why human-to-human transmission was limited.


HA
NA
PB2
PB1
0.1
0.1
0.1
0.1
TH/1KAN-1/04
TH/5KK-494/04
TH/2SP-33/04
TH/3SP-83/04
TH/4SP-528/04
Ck/TH/CH-2/04

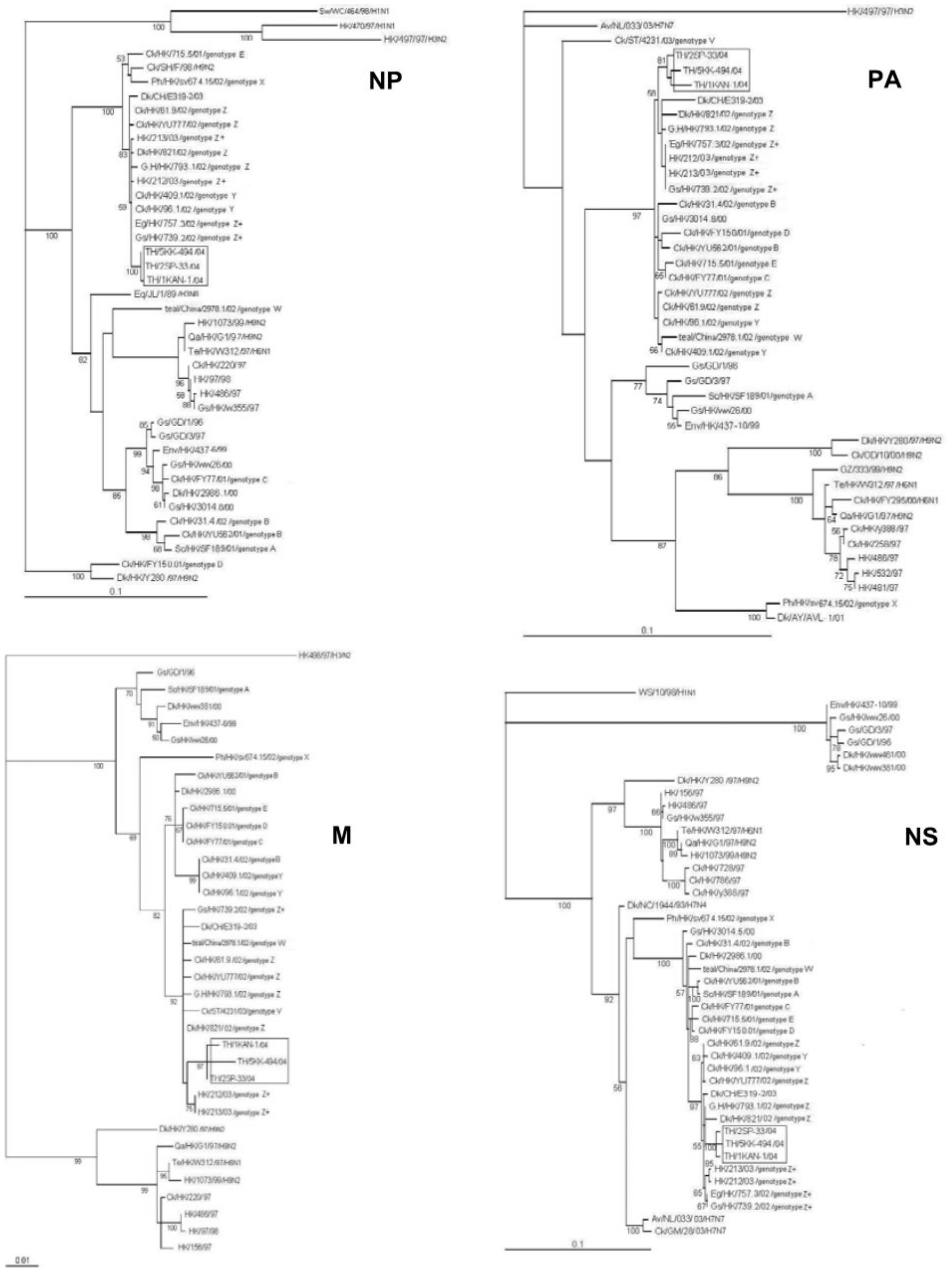

NP
Sw/WC/464/98/H1N1
100
HK/470/97/H1N1
100
HK/497/97/H3N2
53
Ck/HK/715.5/01/genotype E
Ck/SH/F/98/H9N2
Ph/HK/sv67 4.15/02/genotype X
Dk/CH/E319-2/03
100
Ck/HK/61.9/02/genotype Z
Ck/HK/YU777/02/genotype Z
HK/213/03/genotype Z+
83
Dk/HK/821/02/genotype Z
G.H/HK/793.1/02/genotype Z
HK/212/03/genotype Z+
Ck/HK/409.1/02/genotype Y
59
Ck/HK/96.1/02/genotype Y
Eg/HK/757.3/02/genotype Z+
Gs/HK/739.2/02/genotype Z+
100
TH/5KK-494/04
TH/2SP-33/04
100
TH/1KAN-1/04
Eq/JL/1/89/H3N8
teal/China/2978.1/02/genotype W
HK/1073/99/H9N2
Qa/HK/G1/97/H9N2
Te/HK/W312/97/H6N1
82
Ck/HK/220/97
96
HK/97/98
58
HK/486/97
88
Gs/HK/w355/97
85
Gs/GD/1/96
Gs/GD/3/97
99
Env/HK/437-6/99
94
Gs/HK/ww26/00
Ck/HK/FY77/01/genotype C
98
Dk/HK/2986.1/00
85
61
Gs/HK/3014.8/00
Ck/HK/31.4/02/genotype B
98
Ck/HK/YU56 2/01/genotype B
68
Sc/HK/SF189/01/genotype A
Ck/HK/FY150.01/genotype D
100
Dk/HK/Y280/97/H9N2
0.1
PA
HK/497/97/H3N2
Av/NL/033/03/H7N7
Ck/ST/4231/03/genotype V
81
TH/2SP-33/04
TH/5KK-494/04
TH/1KAN-1/04
58
Dk/CH/E319-2/03
Dk/HK/821/02/genotype Z
G.H/HK/793.1/02/genotype Z
Eg/HK/757.3/02/genotype Z+
HK/212/03/genotype Z+
HK/213/03/genotype Z+
Gs/HK/739.2/02/genotype Z+
Ck/HK/31.4/02/genotype B
97
Gs/HK/3014.8/00
Ck/HK/FY150/01/genotype D
Ck/HK/YU56 2/01/genotype B
Ck/HK/715.5/01/genotype E
65
Ck/HK/FY77/01/genotype C
Ck/HK/YU777/02/genotype Z
Ck/HK/61.9/02/genotype Z
Ck/HK/96.1/02/genotype Y
teal/China/2978.1/02/genotype W
56
Ck/HK/409.1/02/genotype Y
Gs/GD/1/96
77
Gs/GD/3/97
74
Sc/HK/SF189/01/genotype A
Gs/HK/ww26/00
56
Env/HK/437-10/99
Dk/HK/Y280/97/H9N2
100
Ck/GD/10/00/H9N2
86
GZ/333/99/H9N2
Te/HK/W312/97/H6N1
100
Ck/HK/FY295/00/H6N1
64
Qa/HK/G1/97/H9N2
56
Ck/HK/y388/97
Ck/HK/258/97
78
HK/480/97
87
72
HK/532/97
75
HK/481/97
Ph/HK/sv674.15/02/genotype X
100
Dk/AY/AVL-1/01
0.1
M
HK/498/97/H3N2
Gs/GD/1/96
70
Sc/HK/SF189/01/genotype A
Dk/HK/ww381/00
91
Env/HK/437-6/99
60
Gs/HK/ww26/00
100
Ph/HK/sv674.15/02/genotype X
Ck/HK/YU56 2/01/genotype B
Dk/HK/2986.1/00
76
Ck/HK/715.5/01/genotype E
67
Ck/HK/FY150.01/genotype D
69
Ck/HK/FY77/01/genotype C
Ck/HK/31.4/02/genotype B
99
Ck/HK/409.1/02/genotype Y
Ck/HK/96.1/02/genotype Y
82
Gs/HK/739.2/02/genotype Z+
Dk/CH/E319-2/03
teal/China/2978.1/02/genotype W
Ck/HK/61.9/02/genotype Z
Ck/HK/YU777/02/genotype Z
92
G.H/HK/793.1/02/genotype Z
Ck/ST/4231/03/genotype V
Dk/HK/821/02/genotype Z
TH/1KAN-1/04
97
TH/5KK-494/04
TH/2SP-33/04
HK/212/03/genotype Z+
75
HK/213/03/genotype Z+
Dk/HK/Y280/97/H9N2
99
Qa/HK/G1/97/H9N2
Te/HK/W312/97/H6N1
95
HK/1073/99/H9N2
99
Ck/HK/220/97
HK/486/97
100
HK/97/98
HK/156/97
0.01
NS
WS/10/98/H1N1
Env/HK/437-10/99
Gs/HK/ww26/00
100
Gs/GD/3/97
Gs/GD/1/96
78
Dk/HK/ww461/00
95
Dk/HK/ww381/00
Dk/HK/Y280/97/H9N2
HK/156/97
97
66
HK/486/97
Gs/HK/w355/97
100
Te/HK/W312/97/H6N1
100
Qa/HK/G1/97/H9N2
99
HK/1073/99/H9N2
Ck/HK/728/97
Ck/HK/786/97
100
Ck/HK/y388/97
100
Dk/NC/1944/93/H7N4
Ph/HK/sv67 4.15/02/genotype X
Gs/HK/3014.5/00
Ck/HK/31.4/02/genotype B
Dk/HK/2986.1/00
100
teal/China/2978.1/02/genotype W
Ck/HK/YU56 2/01/genotype B
57
100
Sc/HK/SF189/01/genotype A
92
Ck/HK/FY77/01 genotype C
Ck/HK/715.5/01/genotype E
Ck/HK/FY150.01/genotype D
88
Ck/HK/61.9/02/genotype Z
63
Ck/HK/409.1/02/genotype Y
Ck/HK/96.1/02/genotype Y
Ck/HK/YU777/02/genotype Z
56
Dk/CH/E319-2/03
97
G.H/HK/793.1/02/genotype Z
Dk/HK/821/02/genotype Z
TH/2SP-33/04
55
100
TH/5KK-494/04
TH/1KAN-1/04
85
HK/213/03/genotype Z+
HK/212/03/genotype Z+
65
Eg/HK/757.3/02/genotype Z+
67
Gs/HK/739.2/02/genotype Z+
Av/NL/033/03/H7N7
100
Ck/GM/28/03/H7N7
0.1

Fig. 5. (*on page 430 & 431*) Phylogenetic trees of the H5N1 viruses. Nucleotide sequences were analysed with PHYLIP using the maximum-parsimony algorithm. Nt 49–1665 of the HA gene, nt 1–1104 of the NA gene, nt 1009–2226 of the PB2 gene, nt 17–1482 of the PB1 gene, nt 1432–2148 of the PA gene, nt 1–966 of the NP gene, nt 1–975 of the M gene and nt 3–833 of the NS gene were used for the analysis. The lengths of the horizontal lines are proportional to the minimum number of nucleotide difference required to join the nodes. The trees are rooted to either H1N1 or H3N2 sequences. Bootstrap values over 50 based on 100 replications are shown. Ck, Chicken; Dk, Duck; Gs, Goose; Qa, Quail; Av, Avian; Sc, Silky chicken; Ml, Mallard; Te, Teal; Tk, Turkey; Ph, Pheasant; Env, Environment; TH, Thailand; HK, Hong Kong; OH, Ohio; TW, Taiwan; CH, China; GD, Guangdong; NC, Nanchang; WS, Wisconsin; NL, The Netherlands; GM, Germany; SH, Shanghai; JL, Jilin; GZ, Guangzhou; AY, Anyang; IT, Italy; PR, Paris.

Host-specific amino acids of the Thailand viruses are mostly avian specific

There were 32 amino acid residues in M1, M2, NP, PA and PB2 that have been described as host-specific residues (Shaw *et al.*, 2002). Of these 32 residues, the Hong Kong H5N1 viruses had avian-like residues at 23 sites, human-like residues at five sites and mixture of the two at four other sites (Shaw *et al.*, 2002). Our H5N1 viruses, on the other hand, had avian-like residues at 31 of 32 sites and a human-like residue at only one site (valine at aa 28 in M2) (Table 2).

Origin of the Thailand viruses

Culling of live poultry in December 1997 led to the eradication of H5N1 viruses from chicken and other terrestrial poultry in Hong Kong. Nevertheless, their viral ancestors – goose/Guangdong/1/96 (H5N1)-, quail/Hong Kong/G1/97 (H9N2)- and teal/Hong Kong/W312/97 (H6N1)-like viruses – continued to circulate in the region. In addition, Gs/Gd-like viruses have gone through several reassortment events with other unknown aquatic avian viruses resulting in the reemergence of multiple new reassortants in terrestrial poultry in Hong Kong in 2001 (Guan *et al.*, 2002). These new reassortants were classified into five genotypes designated A–E based on the source of each genomic segment. After 2002, these genotypes were replaced by the genotypes V, W, X, Y, Z^+ and Z. These genotypes probably acquired their HA and NA genes from a Gs/Gd-like virus and internal genes from unknown avian sources (Li *et al.*, 2004).

In order to understand the evolution and origin of the Thailand viruses, we performed a phylogenetic analysis of all genomic segments of the human H5N1 isolates from Thailand (Fig. 5). All segments of the Thailand isolates were closely related to the genotype Z. The viruses had a 5 aa deletion in NS and a 20 aa deletion in NA. The Thailand viruses and other genotype Z viruses had HA and NA proteins that were in the Gs/Gd-like lineage. PB2, PB1, PA and NS of these viruses had other avian origins in the same lineage as most of the other genotypes, while the NP and M genes were probably from another lineage that was shared by genotypes Y, Z^+ and Z for NP and by genotypes V, W, Z^+ and Z for M. In all trees, the Thailand sequences formed separate subclusters distinguishable from the most closely related genotype Z (Guan *et al.*, 2004; Li *et al.*, 2004). Several genotypes or reassortants have been circulating in avian species in Hong Kong and probably Southern China since 1997. It is not clear why a large outbreak did not take place earlier. Because of the rapid turnover of chicken and other domestic avian population, immunological memory is probably not a crucial factor in determining an outbreak. The HA protein of H5N1 viruses is the most constant fragment of the genome and has not undergone multiple reassortment events. This suggests that the HA has been optimized for efficient spread in the avian population and that the virus has not evolved towards immunological escape. The new genotype that caused the widespread outbreak in 2003–2004 contained more avian-specific residues. This suggests that the virus may have become more adapted to avian species than the other genotypes and the original 1997 viruses. This may be one of the reasons why this virus was selected from its ancestral pool and has expanded enormously in avian populations.

ACKNOWLEDGEMENTS

This project was financially supported by the Faculty of Medicine Siriraj Hospital, and the National Center for Genetic Engineering and Biotechnology. The authors would like to thank Miss Jarunee Prasertsophon and Mr Navin Horthongkham for their technical assistance.

REFERENCES

Castrucci, M. R. & Kawaoka, Y. (1993). Biologic importance of neuraminidase stalk length in influenza A virus. *J Virol* **67**, 759–764.

CDC/WHO Avian Influenza Response Team (2004). Outbreaks of avian influenza A (H5N1) in Asia and interim recommendations for evaluation and reporting of suspected cases – United States, 2004. *Morb Mortal Wkly Rep* **53**, 97–99.

Chotpitayasunondh, T., Lochindarat, S., Srisan, P., Chokepaibulkit, K., Weerakul, J., Maneerattanaporn, M., Sawanpanyalert, P., World Health Organization Thailand & CDC International Emerging Infections Program Thailand (2004). Cases of influenza A (H5N1) – Thailand, 2004. *Morb Mortal Wkly Rep* **53**, 100–102.

Claas, E. C. J., Osterhaus, A. D. M. E., Van Beek, R., De Jong, J. C., Rimmelzwaan, G. F., Senne, D. A., Krauss, S., Shortridge, K. F. & Webster, R. G. (1998). Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet* **351**, 472–477.

Cross, K. J., Burleigh, L. M. & Steinhauer, D. A. (2001). Mechanisms of cell entry by influenza virus. In *Expert Reviews in Molecular Medicine*. Available at http://www-ermm.cbcu.cam.ac.uk

Felsenstein, J. (1993). PHYLIP: Phylogeny Interference Package, version 3.5c. University of Washington, Seattle, WA, USA.

Fouchier, R. A. M., Schneeberger, P. M., Rozendaal, F. W. & 11 other authors (2004). Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc Natl Acad Sci U S A* **101**, 1356–1361.

Gao, P., Watanabe, S., Ito, T., Goto, H., Wells, K., McGregor, M., Cooley, A. J. & Kawaoka, Y. (1999). Biological heterogeneity, including systemic replication in mice, of H5N1 influenza A virus isolates from humans in Hong Kong. *J Virol* **73**, 3184–3189.

Guan, Y., Peiris, J. S. M., Lipatov, A. S., Ellis, T. M., Dyrting, K. C., Krauss, S., Zhang, L. J., Webster, R. G. & Shortridge, K. F. (2002). Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. *Proc Natl Acad Sci U S A* **99**, 8950–8955.

Guan, Y., Poon, L. L. M., Cheung, C. Y. & 13 other authors (2004). H5N1 influenza A protean pandemic threat. *Proc Natl Acad Sci U S A* **101**, 8156–8161.

Hatta, M. & Kawaoka, Y. (2002). The continued pandemic threat posed by avian influenza viruses in Hong Kong. *Trends Microbiol* **10**, 340–344.

Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y. (2001). Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science* **293**, 1840–1842.

Hien, T. T., Liem, N. T., Dung, N. T. & 26 other authors (2004). Avian influenza A (H5N1) in 10 patients in Vietnam. *N Engl J Med* **350**, 1179–1188.

Hoffmann, E., Stech, J., Guan, Y., Webster, R. G. & Perez, D. R. (2001). Universal primer set for the full-length amplification of all influenza A viruses. *Arch Virol* **146**, 2275–2289.

Horimoto, T. & Kawaoka, Y. (1994). Reverse genetics provides direct evidence for a correlation of hemagglutinin cleavability and virulence of an avian influenza A virus. *J Virol* **68**, 3120–3128.

Koopmans, M., Wilbrink, B., Conyn, M. & 8 other authors (2004). Transmission of H7N7 avian influenza A virus to human beings during a large outbreak in commercial poultry farms in the Netherlands. *Lancet* **363**, 587–593.

Kurtz, J., Manwell, R. J. & Banks, J. (1996). Avian influenza virus isolated from a woman with conjunctivitis. *Lancet* **348**, 901–902.

Li, K. S., Guan, Y., Wang, J. & 19 other authors (2004). Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* **430**, 209–213.

Matrosovich, M., Zhou, N., Kawaoka, Y. & Webster, R. (1999). The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *J Virol* **73**, 1146–1155.

Nicholas, K. B., Nicholas, H. B., Jr & Deerfield, D. W., II (1997). GeneDoc: analysis and visualization of genetic variation. *Embnet News* **4**, 1–4.

Nicholson, K. G., Wood, J. M. & Zambon, M. (2003). Influenza. *Lancet* **362**, 1733–1745.

Nobusawa, E., Aoyama, T., Kato, H., Suzuki, Y., Tateno, Y. & Nakajima, K. (1991). Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinin of influenza A viruses. *Virology* **182**, 475–485.

Page, R. D. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.

Peiris, M., Yuen, K. Y., Leung, C. W., Chan, K. H., Ip, P. L., Lai, R. W., Orr, W. K. & Shortridge, K. F. (1999). Human infection with influenza H9N2. *Lancet* **354**, 916–917.

Peiris, J. S. M., Yu, W. C., Leung, C. W. & 9 other authors (2004). Re-emergence of fatal human influenza A subtype H5N1 disease. *Lancet* **363**, 617–619.

Shaw, M., Cooper, L., Xu, X. & 10 other authors (2002). Molecular changes associated with the transmission of avian influenza A H5N1 and H9N2 viruses to humans. *J Med Virol* **66**, 107–114.

Shinya, K., Hamn, S., Hatta, M., Ito, H., Ito, T. & Kawaoka, Y. (2004). PB2 amino acid at position 627 affects replicative efficiency, but not cell tropism, of Hong Kong H5N1 influenza A viruses in mice. *Virology* **320**, 258–266.

Subbarao, K. & Shaw, M. W. (2000). Molecular aspects of avian influenza (H5N1) viruses isolated from humans. *Rev Med Virol* **10**, 337–348.

Subbarao, K., Klimov, A., Katz, J. & 13 other authors (1998). Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science* **279**, 393–396.

Suzuki, H., Saito, R., Masuda, H., Oshitani, H., Sato, M. & Sato, I. (2003). Emergence of amantadine-resistant influenza A viruses: epidemiological study. *J Infect Chemother* **9**, 195–200.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Webster, R., Cox, N. & Stohr, K. (2002). *WHO Manual on Animal Influenza Diagnosis and Surveillance.* WHO: WHO Global Influenza Programme.

Weis, W., Brown, J. H., Cusack, S., Paulson, J. C., Skehel, J. J. & Wiley, D. C. (1988). Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature* **333**, 426–431.

Yuen, K. Y., Chan, P. K. S., Peiris, M. & 8 other authors (1998). Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *Lancet* **351**, 467–471.

Zambon, M. C. (2001). The pathogenesis of influenza in humans. *Rev Med Virol* **11**, 227–241.