Journal of General Virology (2007), **88**, 3094–3099 **DOI** 10.1099/vir.0.83129-0

Short Communication

Origin of highly pathogenic H5N1 avian influenza virus in China and genetic characterization of donor and recipient viruses

Muhammad Mahmood Mukhtar,[1] Sahibzada T. Rasool,[1] Degui Song,[2] Chengliang Zhu,[1] Qian Hao,[1] Ying Zhu[1] and Jianguo Wu[1]

[1]State Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan 430072, P.R. China

[2]College of Life Sciences, Guangxi Normal University, Guilin 541004, P.R. China

Correspondence
Jianguo Wu
jwu@whu.edu.cn
or
wu9988@vip.sina.com

Ying Zhu
yingzhu@whu.edu.cn

Received 1 May 2007
Accepted 6 July 2007

Genetic analysis of all eight genes of two Nanchang avian influenza viruses, A/Duck/Nanchang/1681/92 (H3N8-1681) and A/Duck/Nanchang/1904/92 (H7N1-1904), isolated from Jiangxi province, China, in 1992, showed that six internal genes of H3N8-1681 virus and five internal (except NS gene) genes of H7N1-1904 virus were closely similar to A/Goose/Guangdong/1/96 (H5N1) virus, the first highly pathogenic avian influenza (HPAI) virus of subtype H5N1 isolated in Asia. The neuraminidase (NA) gene of Gs/Gd/1/96 had the highest genetic similarity with A/Duck/Hokkaido/55/96 (H1N1-55) virus. The haemagglutinin (HA) gene of Gs/Gd/1/96 virus might have originated as a result of mutation of H5 HA gene from A/Swan/Hokkaido/51/96 (H5N3-51)-like viruses. The PA gene of H5N3-51 virus had the highest similarity with Gs/Gd/1/96. This study explains the origin of first Asian HPAI H5N1 virus in Guangdong by the reassortment of Nanchang (close to Guangdong) and Hokkaido (Japan) (H1N1-55 and H5N3-51) viruses. Genetic characteristics of donor and recipient viruses were also studied.

Highly pathogenic avian influenza (HPAI) has been recognized as a serious viral disease of poultry since 1878. Its number of outbreaks has increased globally in the past 10–15 years. In December 2003 and January 2004, outbreaks of HPAI caused by H5N1 viruses were reported almost simultaneously in many Asian countries. The emergence of H5N1 viruses in Asia since 1996 was a key determinant of the 2004 epidemic. The mutation and spread of the viruses between 1999 and 2002 apparently set the stage for the simultaneous emergence of the acute disease in many countries (Sims *et al.*, 2005).

Genetic characterization of the gene segments of viruses circulating in southern China indicated that the H5N1 viruses were generated by reassortment. Guan *et al.* (1999) suggested that H9N2 influenza viruses were the donors of internal genes for H5N1 viruses in Hong Kong. Hoffmann *et al.* (2000) reported that H6N1 viruses contribute the neuraminidase 1 (N1) and internal genes of Hong Kong H5N1 viruses, while Xu *et al.* (1999) proposed that the H5 haemagglutinin (HA) of Hong Kong viruses was derived from A/Goose/Guangdong/1/96 (Gs/Gd/1/96) virus, based on 99 % nucleotide homology between HA sequences of these H5N1 viruses. They also discussed the similarity of neuraminidase and internal genes of Gs/Gd/1/96 virus with a variety of other viruses, but didn't suggest any specific subtype to be the precursor of this first HPAI H5N1 virus in Asia. The origin of Gs/Gd/1/96 virus is not known, but presumably, like all HPAI viruses, it came from an unidentified precursor virus of low pathogenicity circulating in wild aquatic birds (Alexander, 2000). To determine the origin of HPAI H5N1 in China, we analysed the influenza virus sequences available in influenza genome databases to find the progenitor of HPAI Gs/Gd/1/96 virus. We also studied the genetic characteristics of these viruses.

Published nucleotide and amino acid sequences of influenza viruses, used for genetic comparison in this study, were obtained from GenBank (NCBI, 2007; http://www.ncbi.nlm.nih.gov/genomes/FLU/), the Influenza Sequence Database (ISD; http://www.flu.lanl.gov) (Macken *et al.*, 2001) and the Influenza Virus DataBase (IVDB; http://influenza.genomics.org.cn) (Chang *et al.*, 2007). Virus abbreviations for animals: Ck, chicken; Dk, duck; Eq, equine; Gs, goose; Md, mallard; Pt, parrot; Rt, ruddy turnstone; Sw, swine; Sn, swan; Ty, turkey; Tn, tern; Tl, teal; Qa, quail. Abreviations for places: Gd, Guangdong; HK, Hong Kong; Jl, Jilin; Hb, Hubei; Hok, Hokkaido; Hd, Hidalgo; Nc, Nanchang; Eng, England; Ire, Ireland; Mex, Mexico; NZ, New Zealand; OT, Ontario; Pak, Pakistan; Pd,

Supplementary material is available with the online version of this paper.

Potsdam; SA, South Africa and official abbreviations for different states of the USA are used for the nomenclature of various viruses.

Sequence editing and multiple sequence alignment were performed with CLUSTAL_X. Neighbour-joining (NJ), maximum-parsimony (MP) and minimum-evolution (ME) trees were constructed by using MEGA version 3.1 (Kumar *et al.*, 2004). Maximum-likelihood (ML) trees were constructed by using PHYML (Guindon & Gascuel, 2003). Estimates of the phylogenies were calculated by performing 500 bootstrap replicates. Each tree is mid-point rooted. Nucleotide distance matrices were estimated by Kimura two-parameter algorithm based on the number of total nucleotide substitutions and evolutionary trees for PB2, NS, NA and HA genes (Fig. 1), and PB1, PA, NP and M genes (Supplementary Fig. S1, available with the online version of this paper) were constructed. Phylogenetic trees constructed by the ML, MP

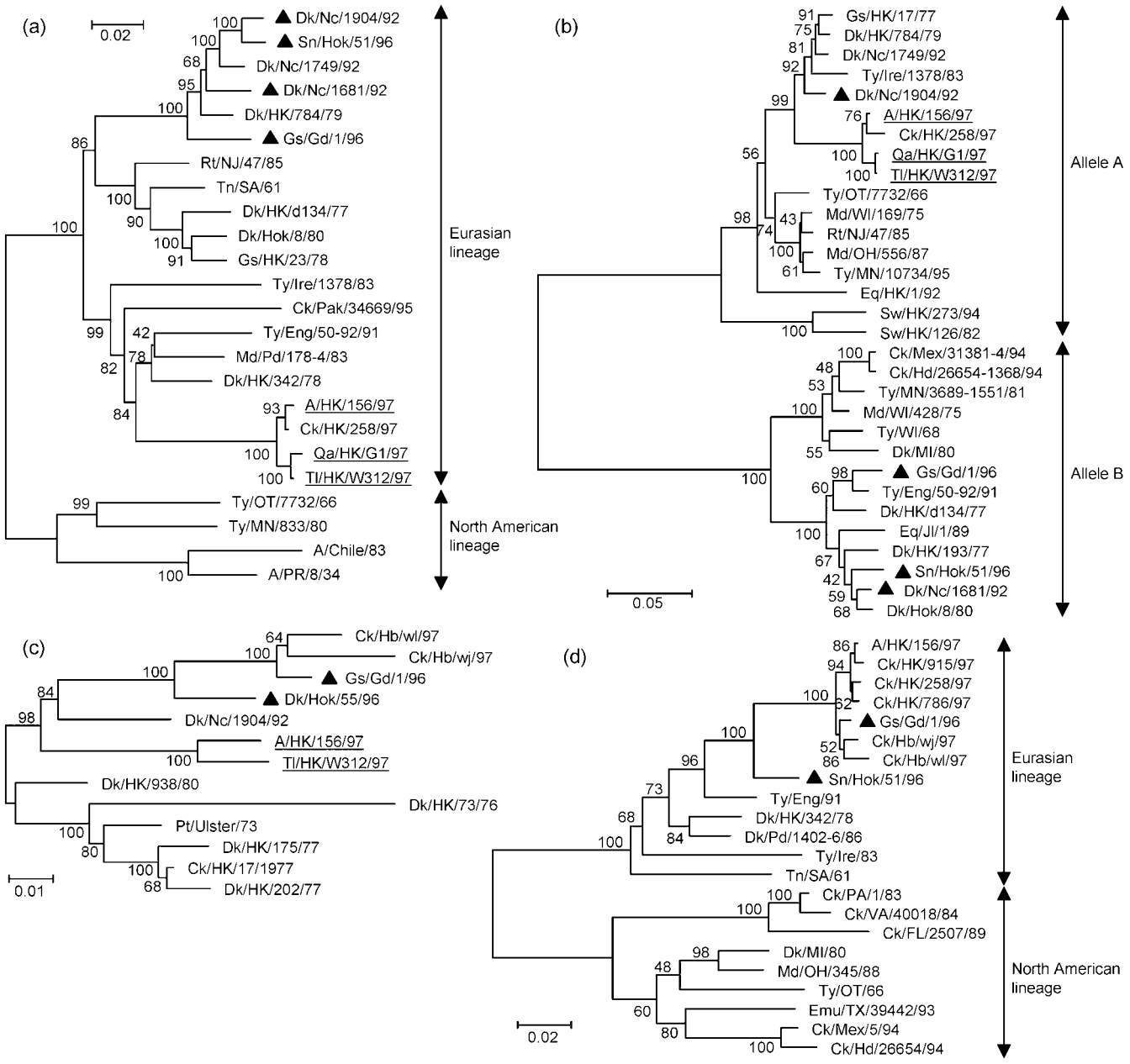

Fig. 1. Phylogenetic trees are based on the nucleotide sequences of influenza A virus PB2 (a), NS (b), NA (c) and HA (d) genes. Analyses were based on nucleotides 34–2275 (2242 bp) of the PB2, 17–840 (824 bp) of the NS, 21–1430 (1410 bp) of the NA and 77–1031 (955 bp) of the HA genes. Numbers at the nodes indicate confidence levels of bootstrap analyses as percentages. Donor and recipient viruses of this study are indicated by ▲.

and ME analyses revealed similar relationships (not shown). Various strains included in different phylogenetic analyses depend upon their close similarity among different genes, and absence of these strains in some trees is due to non-availability of gene sequences.

In Asia, highly pathogenic H5N1 viruses have been circulating since the discovery of Gs/Gd/1/96 virus in 1996. Sequences of all eight genes of influenza viruses, from influenza virus databases, were analysed and compared with Gs/Gd/1/96 virus to find the donor virus or viruses. Phylogenetic trees are based on the nucleotide sequences of the influenza virus genes that were isolated before or in 1996 (year of isolation of first HPAI H5N1 in China). Nanchang viruses (H3N8-1681 and H7N1-1904) isolated from Jiangxi province in 1992 (reported in 2006) and Hokkaido viruses (H5N3-51 and H1N1-55) isolated from Hokkaido (Japan) in 1996 (reported in 2005 and 2006, respectively) are also part of the current study. We also analysed the sequences of H5N1 viruses isolated in 1997 in Hong Kong and Hubei (central China).

The influenza A viruses with the greatest nucleotide and amino acid sequence similarity to the gene segments of Gs/Gd/1/96 virus are presented in Table 1. Xu *et al.* (1999) reported that PB2 of Gs/Gd/1/96 had highest identity of nucleotide and amino acid sequences with ruddy turnstone/Newjersey/47/85 (Rt/NJ/47/85), while our study shows that PB2 of Nanchang and Gs/Gd/1/96 viruses have higher nucleotide and amino acid identity than previously reported (Table 1).

The phylogenetic analysis of PB2 sequences (Fig. 1a) revealed two main lineages, North American and Eurasian. PB2 of Rt/NJ/47/85 falls into the Eurasian lineage which shows its higher nucleotide similarity to this lineage instead of the North American lineage. The Eurasian lineage is further divided into four subgroups. Among these, H3N8-1681 and H7N1-1904 viruses were present in the same subgroup as Gs/Gd/1/96 virus (distinguished by ▲) with high statistical confidence (100 % bootstrap value).

Phylogenetic trees of PB1, PA and NP revealed similar topology to the PB2 tree (Supplementary Fig. S1) and suggested a very close relationship between H3N8-1681 and Gs/Gd/1/96 viruses, which was supported by bootstrap probabilities of 93, 100 and 97 %, respectively. Polymerase complex genes of H3N8-1681 virus have the highest nucleotide similarity (ranging from 93 to 97 %) with Gs/Gd/1/96 virus, followed by H7N1-1904 virus (Table 1). Amino acid comparison of polymerase complex genes shows that H7N1-1904 virus has high similarity (ranging from 97.6 to 99.2 %) to Gs/Gd/1/96, followed by H3N8-1681, except that PA of Gs/Gd/1/96 has the highest nucleotide similarity (94.6 %) with H5N3-51 virus (Table 1).

The PA phylogenetic tree showed a strong relationship between Gs/Gd/1/96 and H5N3-51 viruses, supported by a

Table 1. Sequence comparison with A/Goose/Guangdong/1/96*

Gene	Region of the gene compared	Percentage identity		
		Dk/Nc/1681/92	Dk/Nc/1904/92	Reported†
PB2	34–2275 nt	95.7 %	95.1 %	Rt/NJ/47/85 (91.1 %)
	1–759 aa	98.3 %	98.4 %	Rt/NJ/47/85 (98.2 %)
PB1	31–2258 nt	96.0 %	95.9 %	A/HK/156/97 (93.4 %)
	1–757 aa	97.3 %	97.7 %	Ty/MN/833/80 (98 %)
PA	21–2185 nt	93.1 %	93.1 %	Sw/HK/126/82 (92.7 %)
	1–716 aa	97.5 %	97.6 %	Ty/MN/833/80 (97.2 %)
NP	37–1543 nt	97.0 %	97.0 %	Eq/Jilin/1/89 (94.4 %)
	1–498 aa	99.2 %	99.2 %	Eq/Jilin/1/89 (97.6 %)
M	29–1000 nt	95.7 %	94.9 %	Dk/HK/193/77 (95.4 %)
M1/M2	1–252/1–97 aa	99.2/94.7 %	98.8/96.9 %	Dk/HK/193/77 (98.4/94.4 %)
NS	17–840 nt	94.5 %	71.0 %	Ty/Eng/50-92/91 (97.5 %)
NS1/NS2	1–230/1–121 aa	94.2/96.6 %	61.8/80.9 %	Ty/Eng/50-92/91 (96.5/99.2 %)
		Sn/Hok/51/96‡	**Dk/Hok/55/96‡**	
PA	21–2185 nt	94.6 %	–	Sw/HK/126/82 (92.7 %)
	1–716 aa	97.2 %	–	Ty/MN/833/80 (97.2 %)
H5	77–1031 nt	94.7 %	–	HK/156/97 (98.6 %)
	3–346 aa	95.3 %	–	HK/156/97 (98 %)
N1	21–1398 nt	–	95.2 %	Pt/Ulster/73 (90.3 %)
	1–469 aa	–	96.2 %	Pt/Ulster/73 (94.7 %)

*Nucleotide and amino acid sequences were compared and identity was determined by BLAST (NCBI).
†Xu *et al.* (1999) already reported similarity of these viruses with Gs/Gd/1/96 virus, but the percentage identities shown here were determined again.
‡Hokkaido viral genes having the highest sequence identity with Gs/Gd/1/96 virus.

100 % bootstrap value (Supplementary Fig. S1). The NP gene has the highest identity between Nanchang and Gs/Gd/1/96 viruses in both the nucleotide (97 %) and amino acid (>99 %) comparisons, they differed by only four amino acids, whereas previously reported Equine/Jilin/1/89 virus differed by 12 amino acids (Table 1). In short, Gs/Gd/1/96 virus has the highest similarity to H3N8-1681 virus at the nucleotide level and H7N1-1904 virus at the amino acid level, for polymerase complex genes. The M gene of H3N8-1681 virus has the highest nucleotide identity (95.7 %) with Gs/Gd/1/96 virus, while amino acid comparison of M1 and M2 proteins showed 99.2 and 94.7 % identity (only two and five amino acid difference), respectively (Table 1), where phylogenetic tree of M gene (Supplementary Fig. S1) revealed the similar topology to polymerase complex gene trees, but bootstrap support is low (62 %). On the other hand, H7N1-1904 has lower similarity to Gs/Gd/1/96, both at the nucleotide and the amino acid level (for M1), than already reported similarity of Dk/HK/193/77 virus (Xu *et al.*, 1999), but a higher amino acid similarity for M2 protein (Table 1). The bootstrap support for M gene of Gs/Gd/1/96 with H7N1-1904, H5N3-51 and Dk/HK/193/77 viruses, in the phylogenetic tree, is high (95 %) (Supplementary Fig. S1).

The NS genes of influenza A virus isolates can be allocated to allele A or B on the basis of nucleotide sequence variation (Ludwig *et al.*, 1991). The phylogenetic tree of NS gene shows that H3N8-1681 and Gs/Gd/1/96 viruses belong to allele B, while NS of H7N1-1904 virus belongs to allele A (Fig. 1b). The NS gene of the Gs/Gd/1/96 virus is most closely related to that of Ty/Eng/50-92/91 (H5N1) virus on the basis of both nucleotide and amino acid similarities. However, H3N8-1681 virus is also very closely related to Gs/Gd/1/96 virus in NS gene (Table 1) and belongs to the same clade in the phylogenetic tree (Fig. 1b), with high statistical confidence (100 % bootstrap value). Although Ty/Eng/50-92/91 virus has a high identity, there are more chances of H3N8-1681 being the donor of NS gene to Gs/Gd/1/96 because of their high similarity among other internal genes and isolation of both viruses from the same region. Phylogenetic trees for internal genes (PB2, PA, M and NS) revealed that H5N3-51 virus is present in the same clade as Gs/Gd/1/96 virus with high bootstrap probabilities (Fig. 1 and Supplementary Fig. S1), which suggests that this virus may also contribute internal genes (especially PA gene) along with the HA gene. The internal genes of donor and recipient viruses of our study showed a similar relationship as already reported among the internal genes of donor (Qa/HK/G1/97 and Tl/HK/W312/97) and recipient (A/HK/156/97) viruses (underlined in phylogenetic trees) by Guan *et al.* (1999) and Hoffmann *et al.* (2000) (Fig. 1 and Supplementary Fig. S1).

The N1 NA phylogenetic tree showed a very close relationship between H1N1-55 and Gs/Gd/1/96 viruses (Fig. 1c), which is supported by 100 % bootstrap probability. The sequence comparison of NA gene of H1N1-55 and Gs/Gd/1/96 viruses showed higher similarity than the previously reported similarity between Gs/Gd/1/96 and A/Parrot/Ulster/1973 (Pt/Ulster/73) virus (Xu *et al.* 1999), both at the nucleotide and the amino acid level (Table 1). H1N1-55 differed by 68 nt and 20 aa from Gs/Gd/1/96 virus, whereas Pt/Ulster/73 differed by 133 nt and 24 aa. NA of H7N1-1904 and Gs/Gd/1/96 viruses have the highest amino acid identity (differing by only 16 aa). In the NA tree (Fig. 1c), Gs/Gd/1/96 and Hubei 1997 (Hubei/97) H5N1 viruses (A/Chicken/Hubei/wj/97 and A/Chicken/Hubei/wl/97) fall in the same clade (bootstrap value 100 %), which showed that Hubei/97 viruses obtained the NA gene from Gs/Gd/1/96 virus.

Xu *et al.* (1999) proposed that the H5 haemagglutinin gene of Hong Kong/97 viruses was derived from Gs/Gd/1/96 virus, based on 99 % nucleotide homology between HA sequences of these H5N1 viruses. The HA gene of the Gs/Gd/1/96 virus clustered phylogenetically within the Eurasian lineage of H5 HA genes (Fig. 1d). The HA gene of Gs/Gd/1/96 was different from that of Hubei/97 H5N1 viruses, based on the bootstrap value of 52 %, although they fall in the same clade (Fig. 1d). The HA of H5N3-51 virus (isolated in 1996) has a close similarity to Gs/Gd/1/96 both at the nucleotide (94.7 %) and the amino acid (95.3 %) level, followed by Turkey/England/50-92/1991 (Ty/Eng/91) virus (Fig. 1d). H5N3-51-like viruses formed an out-group relationship with H5N1 viruses of central and southern China in the HA phylogenetic tree (Fig. 1d), which shows its donor status supported by a 100 % bootstrap value. Therefore, it seems that H5N3-51-like viruses mutated locally and provided the H5 HA gene to the first HPAI H5N1 virus (Gs/Gd/1/96), Hong Kong/97 and Hubei/97 H5N1 viruses.

Gs/Gd/1/96 is a HPAI virus with high morbidity and mortality in geese (Xu *et al.*, 1999). It shares many common residues with other HPAI H5N1 viruses (Table 2) of high pathogenicity among humans and mice (Gao *et al.*, 1999; Hatta *et al.*, 2001; Katz *et al.*, 2000; Krug, 2006). Nanchang viruses also share many amino acid residues related to pathogenicity with Gs/Gd/1/96 virus (Table 2). PB2 of Gs/Gd/1/96 has residues lysine (K) and glutamic acid (E) at positions 355 and 627, respectively (Table 2), while Nanchang viruses have a different residue (arginine, R) at position 355. The presence of E at position 627 in Nanchang and Gs/Gd/1/96 viruses illustrates their host-range restriction (Subbarao *et al.*, 1993). PB1 of Nanchang and Gs/Gd/1/96 viruses have residues K and methionine (M) at position 198 and 317, respectively (Table 2). The role of PA in virulence and host specificity of influenza virus has not yet been established. NP plays an important role in host specificity (Gabriel *et al.*, 2005; Scholtissek *et al.*, 1985). Gs/Gd/1/96 has amino acid M at position 136 of NP, which may have a role in facilitating transfer to humans (Hiromoto *et al.*, 2000; Zhou *et al.*, 1999), whereas Nanchang viruses have leucine (L). Amino acid E at position 92 of NS1 protein confers resistance to interferon (IFN) and tumour necrosis factor α (TNF-α) (Seo *et al.*, 2002), whereas Nanchang and Gs/Gd/1/96 viruses have

Table 2. Amino acid residues that correlate with influenza virus pathogenicity*

Gene	Amino acid position	Virus†			Pathogenicity‡		References
		Gs/Gd/1/96 (H5N1)	Dk/Nc/1681/92 (H3N8)	Dk/Nc/1904/92 (H7N1)	HP	LP	
PB2	355	K	R	R	K	Q	Katz *et al.* (2000)
	627	E	E	E	K/E	E	Gao *et al.* (1999)
PB1	198	K	K	K	K	R	Katz *et al.* (2000)
	317	M	M	M	I	M	Katz *et al.* (2000)
HA	HA1 term.	RRKKR	–	–	RXBR§	Absent	Subbarao & Shaw (2000)
NA	223	I	–	I‖	I	T	Katz *et al.* (2000)
M1	15	V	V	V	I	V	Katz *et al.* (2000)
NS1	92	D	D	D	E	D	Seo *et al.* (2002)
C-term.	ESEV	ESEV	ESEV	ESEV	EPEV/ESEV	RSKV	Obenauer *et al.* (2006)

*These residues are related to influenza virus pathogenicity in humans and mice.
†Nanchang viruses were compared with Gs/Gd/1/96 virus for residues that have a role in pathogenicity.
‡HP, high pathogenicity; LP, low pathogenicity.
§Minimum motif for highly pathogenic viruses is RXBR (where R is arginine, X is any amino acid and B is a basic amino acid).
‖NA of Dk/Hok/55/96(H1N1) virus also has the same amino acid residue (I) at this position.

amino acid aspartic acid (D) at position 92 (Table 2). Nanchang and Gs/Gd/1/96 viruses have an ESEV sequence motif at the carboxyl-terminal end of NS1 protein. These residues are predicted to mediate binding to proteins bearing a region called the PDZ domain and correlates with virulence in humans (Obenauer *et al.*, 2006). Analysis of amino acid residues that correlate with viral pathogenicity showed that Gs/Gd/1/96 and Nanchang viruses have high similarity (six out of seven amino acids) for internal genes (Table 2). This fact also supports our argument that Nanchang viruses donated internal genes to Gs/Gd/1/96 virus (Supplementary Fig. S2, available with the online version of the paper).

HA has an important role in influenza virus virulence. Gs/Gd/1/96, Hong Kong/97 and Hubei/97 H5N1 viruses have a common multiple basic amino acid motif (RERRRKKR) at the cleavage site of the HA gene (Supplementary Table S1, available with the online version of the paper) which has been described in several HPAI H5 and H7 HA genes (Senne *et al.*, 1996). H5N3-51 virus (having a RETR sequence at the HA cleavage site) does not have the minimum motif that is required for the high pathogenicity of virus (Hatta *et al.*, 2001; Subbarao & Shaw, 2000). The seven potential glycosylation sites reported in the HA of the A/HK/156/97 virus (Subbarao *et al.*, 1998) are conserved in the HA of H5N3-51 (seventh is missing due to non availability of full-length sequence), Gs/Gd/1/96 and Hubei/97 H5N1 viruses. Avian influenza viruses acquire virulence through changes in glycosylation patterns of HA (Kawaoka *et al.*, 1984) and with addition of polybasic amino acid at the HA cleavage site (Wood *et al.*, 1994). H5N3-51 and Gs/Gd/1/96 viruses have similar glycosylation patterns, but differ in polybasic amino acid cleavage site (Supplementary Table S1). The NA gene of H1N1-55 and Gs/Gd/1/96 viruses lacks the 19 aa deletion in the stalk region seen in that of Hong Kong/97 H5N1 influenza viruses (Subbarao *et al.*, 1998). Therefore, the NA gene of H1N1-55 and Gs/Gd/1/96 viruses has three potential glycosylation sites in the stalk of NA that were deleted in the Hong Kong/97 viruses (Bender *et al.*, 1999).

Nanchang viruses were isolated from Jiangxi province (Nanchang is the capital city), which is situated between the Guangdong (located very close to Hong Kong) and Hubei Provinces of China. Nanchang viruses have the highest sequence similarity with Gs/Gd/1/96 virus for internal genes. H3N8-1681 contributes the PB2, PB1, NP and NS genes, while H7N1-1904 contributes the M gene to Gs/Gd/1/96 virus (Supplementary Fig. S2). H5N3-51 has the highest nucleotide similarity with Gs/Gd/1/96 for the PA gene, which suggests that it has contributed the PA gene (Table 1 and Supplementary Fig. S1). H5N3-51-like viruses may have contributed the HA, while H1N1-55 contributes the NA gene (Supplementary Fig. S2). Nanchang viruses were isolated from ducks in 1992 (4 years before the isolation of Gs/Gd/1/96 virus) and Hokkaido viruses were isolated in 1996, so there is a great chance that these viruses reassorted together and lead to the evolution of HPAI Gs/Gd/1/96 virus in the region (Supplementary Fig. S2). The results suggest that surveillance for influenza in avian as well as in swine and human populations must be maintained and also extended to other regions of China and Asia for more reliable understanding of influenza virus evolution and epidemiology.

Acknowledgements

We thank Khalid Abbas for his help in phylogenetic analysis. M. M. M. is supported by the HEC Pakistan. This work was supported

by research grants from the Major State Basic Research Development Program of China ('973' projects No. 2005CB522901, No. 2005CB523010 and No. 2006CB504202) and the National Natural Science Foundation of China (No. 30470087, No. 30570070 and No. 30570066) to J. W. and Y. Z.

References

Alexander, D. J. (2000). A review of avian influenza in different bird species. *Vet Microbiol* **74**, 3–13.

Bender, C., Hall, H., Huang, J., Klimov, A., Cox, N. J., Gregory, V., Cameron, K., Lim, W. & Subbarao, K. (1999). Characterization of the surface protein of influenza A (H5N1) viruses isolated from humans in 1997–1998. *Virology* **254**, 115–123.

Chang, S., Zhang, J., Liao, X., Zhu, X., Wang, D., Zhu, J., Feng, T., Zhu, B., Gao, G. F. & other authors (2007). Influenza Virus Database (IVDB): an integrated information resource and analysis platform for influenza virus research. *Nucleic Acids Res* **35** (Database issue), D376–D380.

Gabriel, G., Dauber, B., Wolff, T., Planz, O., Klenk, H. D. & Stech, J. (2005). The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc Natl Acad Sci U S A* **102**, 18590–18595.

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., Shortridge, K. F., Krauss, S. & Webster, R. G. (1999). Molecular characterization of H9N2 influenza viruses: were they the donors of the "internal" genes of H5N1 viruses in Hong Kong? *Proc Natl Acad Sci U S A* **96**, 9363–9367.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.

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.

Hiromoto, Y., Yamazaki, Y., Fukushima, T., Saito, T., Lindstrom, S., Omoe, K., Nerome, R., Lim, W., Sugita, S. & Nerome, K. (2000). Evolutionary characterization of the six internal genes of H5N1 human influenza A virus. *J Gen Virol* **81**, 1293–1303.

Hoffmann, E., Stech, J., Leneva, I., Krauss, S., Scholtissek, C., Chin, P. S., Peiris, M., Shortridge, K. F. & Webster, R. G. (2000). Characterization of the influenza A virus gene pool in avian species in southern China: was H6N1 a derivative or a precursor of H5N1? *J Virol* **74**, 6309–6315.

Katz, J. M., Lu, X., Tumpey, T. M., Smith, C. B., Shaw, M. W. & Subbarao, K. (2000). Molecular correlates of influenza A H5N1 virus pathogenesis in mice. *J Virol* **74**, 10807–10810.

Kawaoka, Y., Naeve, C. W. & Webster, R. G. (1984). Is virulence of H5N2 influenza viruses in chickens associated with loss of carbohydrate from the hemagglutinin? *Virology* **139**, 303–316.

Krug, R. M. (2006). Clues to the virulence of H5N1 viruses in humans. *Science* **311**, 1562–1563.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.

Ludwig, S., Schultz, U., Mandler, J., Fitch, W. M. & Scholtissek, C. (1991). Phylogenetic relationship of the nonstructural (NS) genes of influenza A viruses. *Virology* **183**, 566–577.

Macken, C., Lu, H., Goodman, J. & Boykin, L. (2001). The value of a database in surveillance and vaccine selection. In *Options for the Control of Influenza IV*: proceedings of the World Congress on Options for the Control of Influenza IV, Crete, 23–28 September 2000, pp. 103–106. Edited by A. D. M. E. Osterhaus, N. Cox & A. W. Hampson. Amsterdam: Elsevier Science.

NCBI (2007). Influenza Virus Resource. Available from http://www.ncbi.nlm.nih.gov/genomes/FLU/

Obenauer, J. C., Denson, J., Mehta, P. K., Su, X., Mukatira, S., Finkelstein, D. B., Xu, X., Wang, J., Ma, J. & other authors (2006). Large-scale sequence analysis of avian influenza isolates. *Science* **311**, 1576–1580.

Scholtissek, C., Burger, H., Kistner, O. & Shortridge, K. F. (1985). The nucleoprotein as a possible major factor in determining host specificity of influenza H3N2 viruses. *Virology* **147**, 287–294.

Senne, D. A., Panigrahy, B., Kawaoka, Y., Pearson, J. E., Suss, J., Lipkind, M., Kida, H. & Webster, R. G. (1996). Survey of the hemagglutinin (HA) cleavage site sequence of H5 and H7 avian influenza viruses: amino acid sequence at the HA cleavage site as a marker of pathogenicity potential. *Avian Dis* **40**, 425–437.

Seo, S. H., Hoffmann, E. & Webster, R. G. (2002). Lethal H5N1 influenza viruses escape host anti-viral cytokine responses. *Nat Med* **8**, 950–954.

Sims, L. D., Domenech, J., Benigno, C., Kahn, S., Kamata, A., Lubroth, J., Martin, V. & Roeder, P. (2005). Origin and evolution of highly pathogenic H5N1 avian influenza in Asia. *Vet Rec* **157**, 159–164.

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

Subbarao, E. K., London, W. & Murphy, B. R. (1993). A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J Virol* **67**, 1761–1764.

Subbarao, K., Klimov, A., Katz, J., Regnery, H., Lim, W., Hall, H., Perdue, M., Swayne, D., Bender, C. & other authors (1998). Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science* **279**, 393–396.

Wood, G. W., Banks, J., McCauley, J. W. & Alexander, D. J. (1994). Deduced amino acid sequences of the hemagglutinin of H5N1 avian influenza virus isolates from an outbreak in turkeys in Norfolk, England. *Arch Virol* **134**, 185–194.

Xu, X., Subbarao, K., Cox, N. J. & Guo, Y. (1999). Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. *Virology* **261**, 15–19.

Zhou, N. N., Shortridge, K. F., Claas, E. C. J., Krauss, S. L. & Webster, R. G. (1999). Rapid evolution of H5N1 influenza viruses in chickens in Hong Kong. *J Virol* **73**, 3366–3374.