Journal of General Virology (2010), **91**, 759–764 DOI 10.1099/vir.0.016436-0

Short Communication

A history estimate and evolutionary analysis of rabies virus variants in China

Pinggang Ming,[1] Jiaxin Yan,[1] Simon Rayner,[2] Shengli Meng,[1] Gelin Xu,[1] Qing Tang,[3] Jie Wu,[1] Jing Luo[1] and Xiaoming Yang[1]

[1]Department of Genetic Engineering, Wuhan Institute of Biological Product, 9 Linjiang Street, Wu Chang District, Wuhan 430060, PR China

[2]Department of Bioinformatics, Wuhan Institute of Virology, Chinese Academy of Sciences, 44 Xiaohongshan, Wuhan 430071, PR China

[3]State Key Laboratory for Molecular Virology and Genetic Engineering, National Institute for Viral Disease Control and Prevention, China CDC, 100 Yingxin Street, Xuan Wu District, Beijing 100052, PR China

Correspondence
Pinggang Ming
pgming2002@sina.com

Received 7 September 2009
Accepted 2 November 2009

To investigate the evolutionary dynamics of rabies virus (RABV) in China, we collected and sequenced 55 isolates sampled from 14 Chinese provinces over the last 40 years and performed a coalescent-based analysis of the G gene. This revealed that the RABV currently circulating in China is composed of three main groups. Bayesian coalescent analysis estimated the date of the most recent common ancestor for the current RABV Chinese strains to be 1412 (with a 95 % confidence interval of 1006–1736). The estimated mean substitution rate for the G gene sequences (3.961×10^{-4} substitutions per site per year) was in accordance with previous reports for RABV.

Rabies has become an important public health concern in countries in Asia and Africa. It is estimated that this disease is responsible for over 55 000 human deaths annually, of which approximately 56 % occur in Asia and 44 % in Africa (WHO, 2005). Rabies was first described in China in about 556 BC and has been continually reported for the following 2500 years (Wang & Huang 2001). Between 1950 and 2004, a total of 108 412 human rabies cases were reported (Zhang *et al.*, 2005). Now China has the second highest number of human HBV cases in the world after India (Tang & Li, 2005); in 2007 alone, the total case numbers were 3302, with particularly high incidence in southern and eastern regions.

Zhang *et al.* (2006) and Meng *et al.* (2007) previously studied the molecular epidemiology of rabies virus (RABV) in China and identified three main phylogenetic groups of street viruses that clustered with strains from Indonesia and Thailand/Malaysia and the cosmopolitan lineage that includes many vaccine strains. Although closely related viruses were often recovered from particular regions, in some provinces, such as Hunan, Guizhou and Guangxi (Zhang *et al.*, 2006; Liu *et al.*, 2007), samples yielded isolates from two or more main groupings, indicating that substantial movement and mixing of strains was occurring in these regions. Extensive migration of humans and their associated animals in recent times is likely to have contributed to this pattern. However, little is known about the evolutionary dynamics of either grouping.

In recent years, many computational methods have been developed to estimate the rate of nucleotide substitution and apply non-clock-like models to viral nucleotide sequence data. Many of the most popular methods adopt Bayesian techniques using Markov Chain Monte Carlo (MCMC) methods to estimate substitution rates and dates of ancestors for groups of sequences when date information is available (Drummond *et al.*, 2002, 2003). These methods have already been successfully applied to the analysis of RABV sequences (Davis *et al.*, 2006; Hughes *et al.*, 2004, 2005).

The G gene, which is the major target for neutralizing antibody (Badrane *et al.*, 2001), tends to exhibit higher rates of amino acid substitutions (Holmes *et al.*, 2002; Davis *et al.*, 2005). In this study, we used Bayesian MCMC methods to analyse sequences from the highly variable G gene of RABV to determine phylogenetic relationships. By using G gene data collected over several years from almost all regions that had experienced cases of rabies, we investigated the evolution of RABV variants and determined the population history of RABV circulating in China.

A total of 55 complete G gene sequences (1572 nt) isolated from four hosts (dog, human, cattle and deer) from 14

The GenBank/EMBL/DDBJ accession numbers for the 25 RABV G gene sequences isolated in this study are FJ602447–FJ602459 and FJ418876–FJ418887.

provinces, administrative municipalities and autonomous regions of China were analysed. The four hosts from which the viruses were isolated represent reservoirs and spillover hosts. Of these isolates, 25 were newly sequenced as part of this study (GenBank accession nos FJ602447–FJ602459) by using methods described previously (Meng *et al.*, 2007); the remaining 30 isolates were obtained from GenBank. The year of isolation was available for all the isolates; full details are provided in Fig. 1 and Table 1.

Estimates of the rate of molecular evolution, μ (substitutions per site per year), and the time to most recent common ancestor (TMRCA) for the complete G gene alignment were obtained using a Bayesian MCMC method which was implemented in the BEAST program (available at http://beast.bio.ed.ac.uk/Main_Page). The best-fit model of nucleotide substitution for Bayesian analysis was selected with the MODELTEST software using Akaike's Information Criterion (AIC) (Posada & Crandall, 1998). The general time reversible (GTR) substitution model, incorporating a proportion of invariable sites (I) and a gamma distribution of rate variation among sites (Γ4) was used for the BEAST analysis. We employed both strict and relaxed (uncorrelated exponential and lognormal) molecular clocks (Drummond *et al.*, 2006) to explore the extent of variation in the rate of nucleotide substitution and from which we could estimate the TMRCA. According to the estimated Bayes factor (calculated by using the TRACER program), the relaxed-uncorrelated lognormal molecular clock was the best supported model. For each clock model, four population dynamic models were tested: constant population size, exponential population growth, logistic population growth and Bayesian skyline. The likelihoods of these demographic models were compared using AIC (Suchard *et al.*, 2001). The statistical uncertainty in the data for each parameter estimate is reflected by the value of the 95 % highest probability density (HPD). For this analysis, an input file for BEAST was generated by using the BEAUti program with sequences dated according to the year of isolation. Sequences with the same year of isolation were removed from the analysis. For each estimate, three replicate BEAST runs were performed to test the repeatability of the analysis. The BEAST output was assessed using the TRACER program. Each analysis comprised 50 million steps, of which the first 5 million were discarded as the burn-in period. This chain length resulted in effective sampling sizes of at least 200 for all estimated parameters. Samples of trees and parameters were recorded every 10 000 steps. The trees obtained from BEAST were used as input for the TREEANNOTATOR program to find the maximum clade credibility (MCC) tree.

The MCC tree of 55 complete G gene sequences is shown in Fig. 2. The topology is similar to that of earlier phylogenetic analysis of the G gene (Meng *et al.*, 2007). Similarly, these sequences can clearly be classified into three major groups (designated groups I–III). Each of these groups was supported with posterior probability values >0.95. Group I included the majority of the virus isolates. This group could be subdivided further into five subclades, a–e, most of which were clustered by region: most of the samples in subclade Ia were from the three neighbouring regions of Jiangsu, Zhejiang and Shanghai, and the subclade also contained three strains from Hebei (BD06), Yunnan (Md06) and Chongqing (CQH); subclade Ib contained samples from Hubei Province; subclade Ic contained two samples from Hunan Province; and subclade Id contained samples from Anhui Province with the exception of Zt07 (from Yunnan province) and LuoH (from Henan Province which neighbours Anhui). The above four subclades correspond to the results from the study by Meng *et al.* (2007). Clade Ie included samples

Fig. 1. Map of China showing the Provinces from which all RABV strains used for evolutionary analysis were isolated (underlined). The inset box shows islands of the South China Sea.

Table 1. Chinese RABV isolates analysed in this study

Strain	Isolate source			Source/reference	GenBank accession no.
	City/Province	Species	Year		
CQH	Congqing	Human	2007	This study	FJ602447
CXs	Cixi/Zhejiang	Dog	2007	This study	FJ602448
HByx	Huangshi/Hubei	Dog	2008	This study	FJ602449
HeFei	Hefei/Anhui	Dog	1989	This study	FJ602450
HNC	Wuhan/Hubei	Cattle	2007	This study	FJ602451
LU	Henan	Deer	1993	This study	FJ602452
LuoH	Luohe/Henan	Human	2007	This study	FJ602453
WHqs	Wuhan/Hubei	Dog	2007	This study	FJ602454
WHWD	Wuhan/Hubei	Dog	2007	This study	FJ602455
WHyf	Wuhan/Hubei	Dog	2007	This study	FJ602456
ZJhz	Huzhou/Zhejiang	Dog	2007	This study	FJ602457
ZJzj	Zhuji/Zhejiang	Dog	2007	This study	FJ602458
HeX	He/Anhui	Dog	1989	This study	FJ602459
J	Ningxia	Human	1985	This study	FJ418876
nu	Ningxia	Human	1986	This study	FJ418877
Yue2	Guangxi	Dog	1997	This study	FJ418878
FS	Fengshan/Guangxi	Dog	1998	This study	FJ418879
GX2	Guangxi	Dog	1994	This study	FJ418880
H69	Anhui	Dog	1969	This study	FJ418881
H	He/Anhui	Dog	1989	This study	FJ418882
SBH	Shanghai	Human	1992	This study	FJ418883
SBD	Shanghai	Dog	1992	This study	FJ418884
gk5	Kai/Guizhou	Dog	2006	This study	FJ418885
SH06	Shanghai	Dog	2006	This study	FJ418886
gg4	Guiyang/Guizhou	Dog	2006	This study	FJ418887
FY1	Fuyang/Anhui	Dog	2004	Meng *et al.* (2007)	DQ849044
FY4	Funan/Anhui	Dog	2004	Meng *et al.* (2007)	DQ849047
FY7	Lixin/Anhui	Dog	2004	Meng *et al.* (2007)	DQ849050
FY10	Yingshang/Anhui	Dog	2004	Meng *et al.* (2007)	DQ849053
FY15	Taihe/Anhui	Dog	2005	Meng *et al.* (2007)	DQ849057
FY16	Fuyang/Anhui	Dog	2005	Meng *et al.* (2007)	DQ849058
WG432	Wugang/Hunan	Dog	2005	Meng *et al.* (2007)	DQ849059
WG430	Wugang/Hunan	Dog	2005	Meng *et al.* (2007)	DQ849060
WH5	Wuhan/Hubei	Dog	2005	Meng *et al.* (2007)	DQ849061
HN06	Wuhan/Hubei	Dog	2005	Meng *et al.* (2007)	DQ849062
QC	Qichun/Hubei	Human	2006	Meng *et al.* (2007)	DQ849063
NC	Nanchang/Jiangxi	Dog	2004	Meng *et al.* (2007)	DQ849064
JSS62	Suqian/Jiangsu	Dog	2005	Meng *et al.* (2007)	DQ849065
JSL26	Ganyu/Jiangsu	Dog	2005	Meng *et al.* (2007)	DQ849066
JSL27	Ganyu/Jiangsu	Dog	2005	Meng *et al.* (2007)	DQ849067
JSL29	Ganyu/Jiangsu	Dog	2005	Meng *et al.* (2007)	DQ849068
N11	Ningming/Guangxi	Dog	1997	Meng *et al.* (2007)	DQ849069
YUE1	Liuzhou/Guangxi	Dog	1997	Meng *et al.* (2007)	DQ849070
GX4	Guanxi	Dog	1994	Meng *et al.* (2007)	DQ849071
CQ92	Chongqing	Dog	1992	Meng *et al.* (2007)	DQ849072
FEIDONG	Feidong/Anhui	Dog	1989	Meng *et al.* (2007)	DQ849073
HN10	Yongzhou/Hunan	Human	2006	Ming *et al.* (2009)	EU643590
BD06	Baoding/Hebei	Dog	2006	Shoufeng Zhang, personal communication	EU549783
CGX89-1	Guangxi	Dog	1993	Bai *et al.* (1994)	L04523
CNX8601	Ningxia	Human	1986	Tang *et al.* (2000)	AY009098
CNX8511	Ningxia	Human	1985	Tang *et al.* (2000)	AY009099
Yunnan_Md06	Mouding/Yunnan	Dog	2006	Huanyun Zhao, personal communication	EU253477
Yunnan_Qj07	Yunnan	Dog	2007	Huanyun Zhao, personal communication	EU275240
Yunnan_Zt07	Yunnan	Dog	2007	Huanyun Zhao, personal communication	EU275241
Yunnan_Tc06	Yunnan	Dog	2006	Huanyun Zhao, personal communication	EU275242

from Guizhou, Yunnan and Guangxi. This subclade of the Chinese rabies street virus has been identified here for the first time. Group II comprised strains from Ningxia and Chongqing. Group III comprised samples from Guangxi, Henan and Hunan. It was notable that the samples from Yunnan Province were not located together but were found in a different clade or subclade: Md06 in Ia, Zt07 in Id, Qj07 in Ie and Tc06 in Group III. Our phylogenetic analysis based on the Bayesian method (Fig. 2) showed results similar to those of Meng *et al.* (2007). A total of five subclades were identified within group I, of which four corresponded with the data from Meng *et al.* (2007), and an additional new fifth subclade was found in this study. Notably, within group I, two samples each occupied a separate branch: N11 from Guangxi Province and NC from Jiangxi Province. We considered that these two samples may be two other subclades in group I. Based on these analyses, we could cautiously conclude that there were at

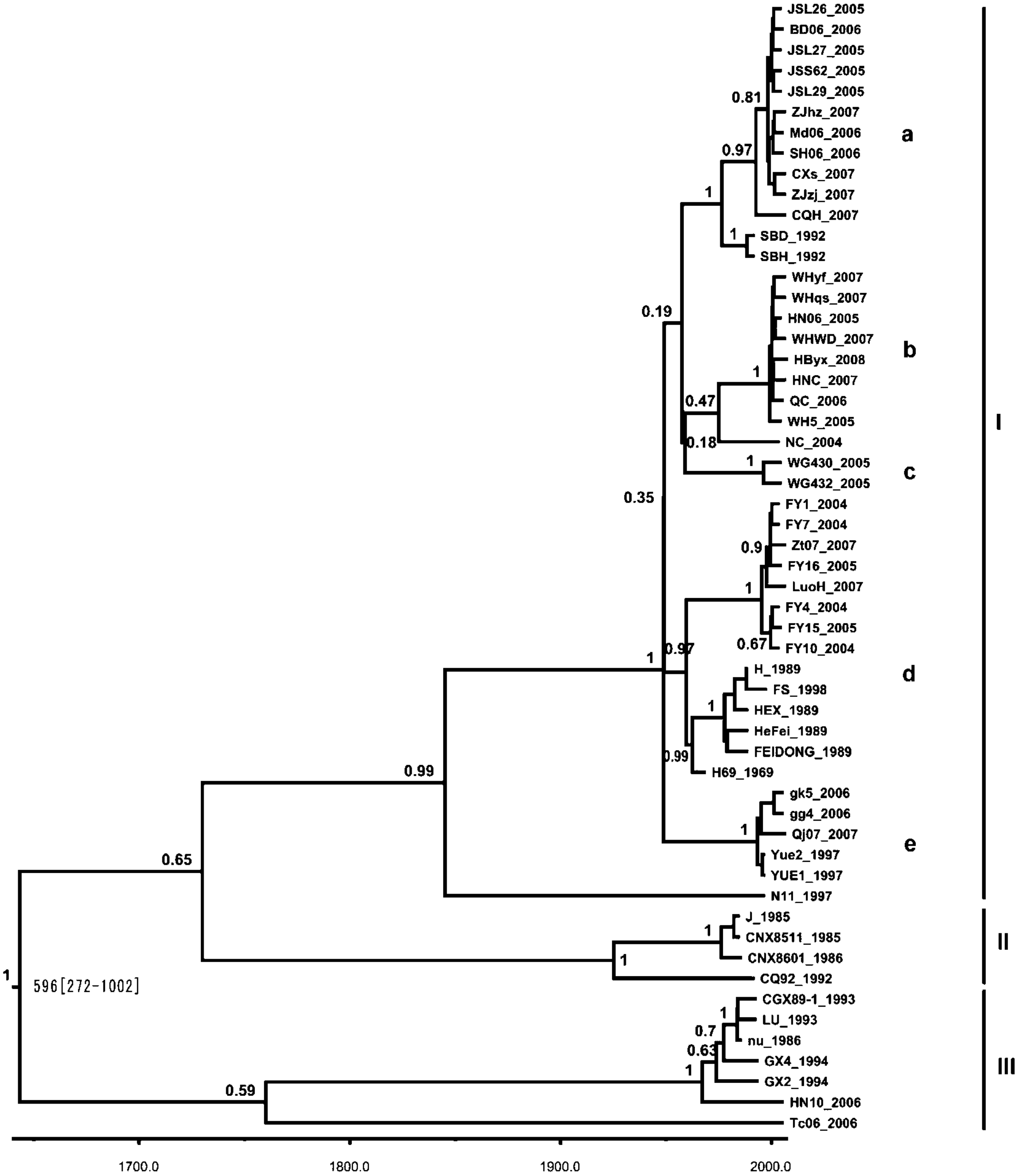

Fig. 2. MCC tree from Bayesian coalescent analysis of 55 RABV street strains isolated from China, based on the entire G gene sequence. The estimated TMRCA for this dataset and its 95 % HPD values are indicated. Isolate names are given according to Table 1. Three lineages and the subclades are indicated. Horizontal branches are drawn to a scale of estimated year of divergence, with tip times reflecting sampling date (year). Posterior probability values for all of the major nodes are shown.

least seven subclades in group I of Chinese RABV. Notably, RABV strains from the neighbouring provinces of Guangxi and Yunnan were found on different groups or subclades. The complicated situation of RABV in Guangxi was described previously. Liu *et al.* (2007) collected 42 samples from suspected rabid animals from different areas of Guangxi from 2000 to 2005 and amplified partial N gene sequences; these sequencing data were used for phylogenetic analysis. Nucleotide homology comparisons and phylogenetic tree analysis based on these sequences indicated that all the RABV isolates could be divided into four groups. The situation was reflected in the present analysis, with the RABV samples distributed in three groups or subclades. The province of Guangxi acts as a major commercial hub, with both goods and animals being transported through the province to different parts of China (Liu *et al.*, 2007). This may explain in part the diversity of viruses circulating there. Neighbouring provinces such as Yunnan Province may also be affected by this phenomenon, with four samples located on four distant groups or subclades.

The mean rate of nucleotide substitution for the G gene of RABV isolates in China, estimated by using a Bayesian MCMC approach (assuming an uncorrelated lognormal molecular clock), was 3.961×10^{-4} substitutions per site per year (95 % HPD values, $2.126–5.992 \times 10^{-4}$ substitutions per site per year). The rate is strongly concordant with previous estimates of substitution rates in the analysis of dog RABV sampled worldwide (Bourhy *et al.*, 2008) and dog RABV in western and central Africa (Talbi *et al.*, 2009). By using the same approach, we were able to estimate the TMRCA of the Chinese RABV to be 596 years (95 % HPD, 272–1002 years), or sometime around 1412 (1006–1736). More recently, Bourhy *et al.* (2008) determined a common origin of all dog RABV circulating globally and proposed that the ancestor of these viruses existed <1500 years ago, perhaps in the Indian subcontinent. Another recent study based on high-resolution temporal and spatial data suggested that RABV spread rapidly and continually from endemic rabies regions in Africa (Hampson *et al.*, 2007). According to our analysis, the TMRCA of all viruses sampled from China was estimated to be 596 years (95 % HPD, 272–1002 years), which was in accordance with the estimate from the study by Bourhy *et al.* (2008). Although there have been several examples illustrating long-distance transmission of RABV due to human-mediated animal movements (Jenkins & Winkler, 1987; Windiyaningsih *et al.*, 2004; Fèvre *et al.*, 2006), this analysis could not provide direct evidence to prove that the RABV in China were spread from Africa or the Indian subcontinent.

Our study used coalescent-based methods to reconstruct the molecular epidemiology of RABV in China. For the first time, we have provided evidence that RABVs circulating in China nowadays emerged several hundred years ago. The phylogenetic analysis constructed in this study provided additional data and enrich previous molecular epidemiology studies of RABV in China.

Acknowledgements

This work was financially supported by the National High Technology Research and Development Program of China (2007AA02Z402).

References

Badrane, H., Bahloul, C., Perrin, P. & Tordo, N. (2001). Evidence of two Lyssavirus phylogroups with distinct pathogenicity and immunogenicity. *J Virol* **75**, 3268–3276.

Bai, X. H., Gu, L., Cynthia, K., Warner, C. K. & Makonnen, F. (1994). Sequencing and analysis of nucleotides of the glycoprotein of Chinese street strain (CGX89–1) of rabies virus. *Zhonghua Min Guo Wei Sheng Wu Ji Mian Yi Xue Za Zhi* **14**, 10–15 (in Chinese).

Bourhy, H., Reynes, J. M., Dunham, E. J., Dacheux, L., Larrous, F., Huong, V. T., Xu, G., Yan, J., Miranda, M. E. & Holmes, E. C. (2008). The origin and phylogeography of dog rabies virus. *J Gen Virol* **89**, 2673–2681.

Davis, P. L., Holmes, E. C., Larrous, F., Van Poel, W., Tjornehoj, K., Alonso, W. J. & Bourhy, H. (2005). Phylogeography, population dynamics, and molecular evolution of European bat lyssaviruses. *J Virol* **79**, 10487–10497.

Davis, P. L., Bourhy, H. & Holmes, E. C. (2006). The evolutionary history and dynamics of bat rabies virus. *Infect Genet Evol* **6**, 464–473.

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. & Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**, 1307–1320.

Drummond, A., Pybus, O. G. & Rambaut, A. (2003). Inference of viral evolutionary rates from molecular sequences. *Adv Parasitol* **54**, 331–358.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88.

Fèvre, E. M., Bronsvoort, B. M., Hamilton, K. A. & Cleaveland, S. (2006). Animal movements and the spread of infectious diseases. *Trends Microbiol* **14**, 125–131.

Hampson, K., Dushoff, J., Bingham, J., Bruckner, G., Ali, Y. H. & Dobson, A. (2007). Synchronous cycles of domestic dog rabies in sub-Saharan Africa and the impact of control efforts. *Proc Natl Acad Sci U S A* **104**, 7717–7722.

Holmes, E. C., Woelk, C. H., Kassis, R. & Bourhy, H. (2002). Genetic constraints and the adaptive evolution of rabies virus in nature. *Virology* **292**, 247–257.

Hughes, G. J., Páez, A., Boshell, J. & Rupprecht, C. E. (2004). A phylogenetic reconstruction of the epidemiological history of canine rabies virus variants in Colombia. *Infect Genet Evol* **4**, 45–51.

Hughes, G. J., Orciari, L. A. & Rupprecht, C. E. (2005). Evolutionary timescale of rabies virus adaptation to North American bats inferred from the substitution rate of the nucleoprotein gene. *J Gen Virol* **86**, 1467–1474.

Ito, N., Kakemizu, M., Ito, K. A., Yamamoto, A., Yoshida, Y., Sugiyama, M. & Minamoto, N. (2001). A comparison of complete genome sequences of the attenuated RC-HL strain of rabies virus used for production of animal vaccine in Japan, and the parental Nishigahara strain. *Microbiol Immunol* **45**, 51–58.

Jenkins, S. R. & Winkler, W. G. (1987). Descriptive epidemiology from an epizootic of raccoon rabies in the Middle Atlantic States, 1982–1983. *Am J Epidemiol* **126**, 429–437.

Liu, Q., Xiong, Y., Luo, T. R., Wei, Y. C., Nan, S. J., Liu, F., Pan, Y., Feng, L., Zhu, W. & other authors (2007). Molecular epidemiology of rabies in Guangxi Province, south of China. *J Clin Virol* **39**, 295–303.

Ming, P. G., Du, J. L., Tang, Q., Yan, J. X., Nadin-Davis, S. A., Li, H., Tao, X. Y., Huang, Y., Hu, R. L. & Liang, G. D. (2009). Molecular characterization of the complete genome of a street rabies virus isolated in China. *Virus Res* **143**, 6–14.

Meng, S. L., Yan, J. X., Xu, G. L., Nadin-Davis, S. A., Ming, P. G., Liu, S. Y., Ming, H. T., Zhu, F. C., Zhou, D. J. & other authors (2007). A molecular epidemiological study targeting the glycoprotein gene of rabies virus isolates from China. *Virus Res* **124**, 125–138.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. (2001). Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol* **18**, 1001–1013.

Talbi, C., Holmes, E. C., de Benedictis, P., Faye, O., Nakoune, E., Gamatie, D., Diarra, A., Elmamy, B. O., Sow, A., Adjogoua, E. V., Sangare, O., Dundon, W. G., Capua, I., Sall, A. A. & Bourhy, H. (2009). Evolutionary history and dynamics of dog rabies virus in western and central Africa. *J Gen Virol* **90**, 783–791.

Tang, Q. & Li, H. (2005). Epidemic situation and related factors analysis of rabies in China. *Zhonghua Liu Xing Bing Xue Za Zhi* **26**, 223–224 (in Chinese).

Tang, Q., Orciari, L. A., Rupprechti, C. E. & Zhao, X. (2000). Sequencing and positional analysis of the glycoprotein gene of four Chinese rabies viruses. *Zhongguo Bingduxue* **15**, 22–33 (in Chinese).

Wang, X. J. & Huang, J. T. (2001). Epidemiology. In *Rabies and Rabies Vaccines*, 1st edn, pp. 127–144. Edited by Y. X. Yu. Beijing, China: Chinese Medicine Technology Press.

WHO (2005). *Expert Consultation on Rabies, First Report. WHO Technical Report Series, No 931.* Geneva: World Health Organization.

Windiyaningsih, C., Wilde, H., Meslin, F. X., Suroso, T. & Widarso, H. S. (2004). The rabies epidemic on Flores Islands, Indonesia (1998–2003). *J Med Assoc Thai* **87**, 1389–1393.

Zhang, Y. Z., Xiong, C. L., Xiao, D. L., Jiang, R. J., Wang, Z. X., Zhang, L. Z. & Fu, Z. F. (2005). Human rabies in China. *Emerg Infect Dis* **11**, 1983–1984.

Zhang, Y. Z., Xiong, C. L., Zou, Y., Wang, D. M., Jiang, R. J., Xiao, Q. Y., Hao, Z. Y., Zhang, L. Z., Yu, Y. X. & Fu, Z. F. (2006). Molecular characterization of rabies virus isolates in China during 2004. *Virus Res* **121**, 179–188.