Infection, Genetics and Evolution 55 (2017) 48–55

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

From discovery to spread: The evolution and phylogeny of Getah virus

Yuan-Yuan Li[a,c,d,1], Hong Liu[b,1], Shi-Hong Fu[c,d,1], Xiao-Long Li[c,d], Xiao-Fang Guo[e], Ming-Hua Li[c,d], Yun Feng[f], Wei-Xin Chen[c,d], Li-Hua Wang[c,d], Wen-Wen Lei[c,d], Xiao-Yan Gao[c,d], Zhi Lv[c,d], Ying He[c,d], Huan-Yu Wang[c,d], Hong-Ning Zhou[e], Gui-Qin Wang[g,*], Guo-Dong Liang[c,d,**]

[a] National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Key Laboratory of Parasite and Vector Biology, Ministry of Health, WHO Collaborating Centre for Tropical Dieases, Shanghai 200025, People's Republic of China
[b] School of Life Sciences, Shandong University of Technology, Zibo, Shandong, People's Republic of China
[c] State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, People's Republic of China.
[d] Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou 310058, People's Republic of China
[e] Yunnan Provincial Center of Arborvirus Research, Yunnan Provincial Key Laboratory of Vector-borne Diseases Control and Research, Yunnan Institute of Parasitic Diseases, Pu'er 665000, People's Republic of China
[f] Yunnan Provincial Key Laboratory for Zoonosis Control and Prevention, Yunnan Institute of Endemic Disease Control and Prevention, Dali, Yunnan 671000, People's Republic of China
[g] Department of Immunology and Microbiology, Shanxi Medical University, Taiyuan 030001, People's Republic of China

ARTICLE INFO

Keywords:
Getah virus
Origin
Evolution
Phylogenetic analysis

ABSTRACT

Getah virus (GETV) was first isolated in Malaysia in 1955. Since then, epidemics in horses and pigs caused by GETV have resulted in huge economic losses. At present, GETV has spread across Eurasia and Southeast Asia, including mainland China, Korea, Japan, Mongolia, and Russia. Data show that the Most Recent Common Ancestor (MRCA) of GETV existed about 145 years ago (95% HPD: 75–244) and gradually evolved into four distinct evolutionary populations: Groups I–IV. The MRCA of GETVs in Group III, which includes all GETVs isolated from mosquitoes, pigs, horses, and other animals since the 1960s (from latitude 19°N to 60°N), existed about 51 years ago (95% HPD: 51–72). Group III is responsible for most viral epidemics among domestic animals. An analysis of the GETV E2 protein sequence and structure revealed seven common amino acid mutation sites. These sites are responsible for the structural and electrostatic differences detected between widespread Group III isolates and the prototype strain MM2021. These differences may account for the recent geographical radiation of the virus. Considering the economic significance of GETV infection in pigs and horses, we recommend the implementation of strict viral screening and monitoring programs.

1. Introduction

Getah virus (GETV), of the family *Togaviridae*, genus *Alphavirus*, is assigned to the Semliki Forest virus complex (Weaver et al., 2005) based on viral antigens. The prototype strain, MM2021, was first isolated from *Culex gelidus* collected in Malaysia in 1955 (Karabatsos, 1985). GETV infection can cause domestic animal diseases, including fever, rashes, edema of the hindlegs and lymph node enlargement in horses, and abortion in pigs (Kamada et al., 1980; Kumanomido et al., 1988). Thus, GETV is an important pathogen causing animal diseases. Several epidemics caused by GETV in horses and pigs occurred in Japan in 1970 and the 1980s (Sugiura et al., 1981; Sentsui and Kono, 1985; Yago et al., 1987). In 2014 and 2015, GETV infections in racehorses were reported in Japan (Nemoto et al., 2014; Bannai et al., 2016); similarly, a large number of horses were infected with GETV in India in 1990 (Broen and Timoney, 1998).

GETV is an enveloped, single-stranded, positive-sense RNA virus. The viral genome has a methylated (7-methylguanosine) cap structure at the 5′ end, and an unequal poly(A) tail at the 3′ end. The viral genome has two open reading frames that encode nonstructural and structural genes. The first two-thirds of the viral genome encode nonstructural proteins (including *nsp1* to *nsp4*), which are responsible for viral RNA transcription, replication, polyprotein cleavage, and RNA capping. This region is followed by the 26S RNA junction region, which

* Correspondence to: G. Wang, Shanxi Medical University, 56 # Xinjiannanlu, Taiyuan 030001, People's Republic of China.
** Correspondence to: G. Liang, Department of Viral Encephalitis and Arbovirus, National Institute for Viral Disease Control and Prevention, China CDC, 155 # Changbailu, Changping District, Beijing 102206, People's Republic of China.
E-mail addresses: guiqinwang321@163.com (G.-Q. Wang), gdliang@hotmail.com (G.-D. Liang).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.meegid.2017.08.016
Received 17 March 2017; Received in revised form 10 August 2017; Accepted 18 August 2017
Available online 19 August 2017

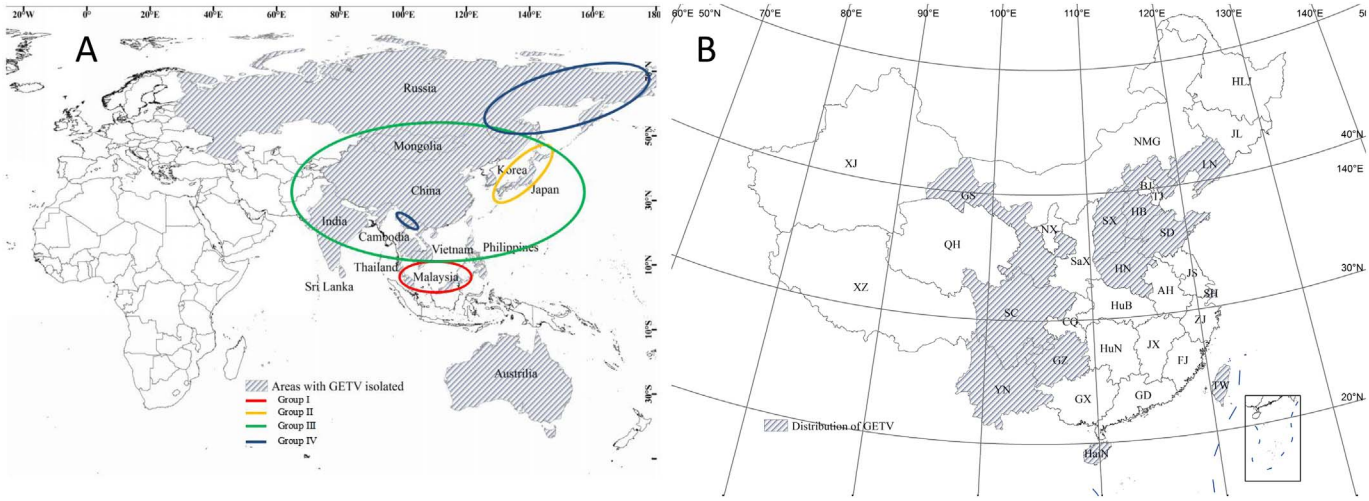

Fig. 1. Distribution of Getah virus (A) The distribution of GETV around the world. The Groups I–IV of GETV are labeled in red, yellow, green and blue, respectively. (B) The distribution of GETV in China's areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

promotes transcription of the intracellular subgenomic 26S RNA. The latter part of the viral genome encodes many structural proteins, including those of the capsid, E3, E2, 6K, and E1. E2 is the main protein that mediates viral entry into host cells during infection (Weaver et al., 2005; Karabatsos, 1985).

GETV was first isolated in Malaysia, and later isolated from a variety of mosquito species in Southeast Asia (Kono, 1988), Australia (Fukunaga et al., 2000), China (Li et al., 1992), Japan (Kamada et al., 1980; Yago et al., 1987), Mongolia, Russia (L'vov et al., 2000), and other regions (shown in Fig. 1A). Thus, the virus gradually spread from Malaysia (latitude 3°N) through mainland China to Russia (latitude 60°N). GETV has become an emerging mosquito-borne virus in the Eurasian region.

In this study, we sequenced the entire genome of a GETV strain (YN12042) isolated from the border of the China-Laos region in southern Yunnan Province in 2012. Furthermore, we conducted a phylogenetic analysis of GETV isolates obtained in China over the past 20 years, and used GenBank data to draw inferences about the molecular phylogenetics and transmission dynamics of GETV.

2. Materials and methods

2.1. GETV and genome sequencing

The virus strain (YN12042) was isolated from *Culex tritaeniorhynchus* Giles in a livestock shed in Menghai County, Xishuangbanna Dai Autonomous Prefecture, Yunnan Province, China, in 2012. cDNA was prepared from total extracted viral RNA. Whole-genome amplification was performed using the primers listed in Table 1 (Liu et al., 2010). The PCR products were identified by 1% agarose gel electrophoresis, and positive products were inserted into pGEM-Teasy vector (Promega Corp., Madison, WI, USA) for sequencing after purification using a QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The sequencing results were assembled, edited, and corrected using SeqMan in the DNASTAR software package (Lei et al., 2015). The whole genome nucleotide sequence of the virus has been deposited in GenBank (accession number: KY450683).

2.2. The genome sequences and origins of the GETV strains used in the present study

The GETV genome sequences used for molecular genetics analyses were from GETVs isolated from blood-sucking insects and host animals in various countries since 1955. Among these strains, 16 (including YN12042) were isolated and identified in our laboratory in China and were registered in GenBank for the first time. We also downloaded 26 GETV sequences for the E2 gene from GenBank before the end of September 2016. These sequences originated from GETVs isolated in Malaysia, Korea, Japan, Russia, and China. The GETV sequence data used in our study are shown in Table 2.

2.3. Bioinformatics analysis

The GETV E2 gene sequence dataset was constructed and aligned with ClustalX 1.8. Homology analysis and alignments of nucleotide and amino acid sequences were conducted using MegAlign in the software packages DNASTAR (version 5.0) and BioEdit (version 7.0.5.3), respectively. An analysis of differences among the nucleotide and amino acid sequences was conducted using GeneDOC (Lei et al., 2015).

The GETV sequence database was analyzed using the Bayesian Markov chain Monte Carlo method. The nucleotide substitution rate, divergence time of the most recent common ancestor (MRCA), and model and rate of population growth were estimated using BEAST software (Brummond and Rambaut, 2007). The GTR + I + G substitution model was selected as the best-fit nucleotide substitution model by MrModeltest (Posada, 2003). A relaxed clock was selected to match the four population growth models to set the calculation parameters. The calculated chain lengths were adjusted according to the pre-test results for a final effective sample size > 200, and the best models were selected by means of a Bayes factor test based on marginal likelihood values (2lnBF N 2) and 95% highest posterior density (HPD) intervals. The analysis was run for 200,000,000 generations to ensure sufficient mixing. Finally, a maximum clade tree was built using TreeAnnotator with 10% burn-in (http://beast.bio.ed.ac.uk/).

2.4. Protein structure analysis

The structural model of GETV E2 was built using SWISS-MODEL (http://www.swissmodel.expasy.org/). The crystal structure of CHIKV (PDB 2xfc) was selected as the best template for the model. VMD (Humphrey et al., 1996) and YASARA (Krieger and Vriend, 2002) softwares were used to analyze the structures of E2 among GETV isolates.

Table 1
Primers used in the present study.

Primer	Amplify region	Sequence(5′–3′)	Site in genome	Length of amplification
GETV-F1	5′UTR–Nsp1	ATGGCGGACGTGTGACATCAC	1–21	930 bp
GETV-R1		GTAACCTTCGCATGACACCACC	909–930	
GETV-F2	Nsp1–Nsp2	GGCATTTACCTTCCGTGTTTC	848–868	883 bp
GETV-R2		TGTGCTTGCGGTGTAACCTTC	1710–1730	
GETV-F3	Nsp2	TAGTGAGCGGCTCTTGTGCTG	1610–1630	907 bp
GETV-R3		CCGCACAGTACTACCTTACCTGAC	2493–2516	
GETV-F4	Nsp2	GATGAGGCGTTCGCGTGTCACT	2434–2455	1070 bp
GETV-R4		GGTAACCGACGATTGGATGGGACT	3480–3503	
GETV-F5	Nsp2–Nsp3	TCTACGTGGCAACATGAACTCG	3399–3420	1066 bp
GETV-R5		CGTGAATAAGTGGTTCAAGGACTGC	4440–4464	
GETV-F6	Nsp3	GCTGTGGCTAGCATAATTAGTACC	4345–4368	984 bp
GETV-R6		TGGGATAGCGCGTATGTCTGT	5308–5328	
GETV-F7	Nsp3–Nsp4	GTCGCCCAACTTAGACAGG	5212–5230	978 bp
GETV-R7		TGGTTGGTGGTATGCGTGG	6171–6189	
GETV-F8	Nsp4	CCGATGAGTATGACGCTTATCTGG	6071–6094	1048 bp
GETV-R8		ACTTCCATGTTGACCCAACTC	7098–7118	
GETV-F9	Nsp4–C protein	CGCTGCTGAACATTGTCATAG	6965–6985	1022 bp
GETV-R9		GGTTGTCGATCACACCTTTG	7967–7986	
GETV-F10	C protein–E2	GATTGCATCTTCGAGGTCAAGC	7881–7922	1039 bp
GETV-R10		GTGCGTGTTTGTACTGCACTTTG	8897–8919	
GETV-F11	E2–6K protein	TGCGCTATTCGAGGCACGAT	8793–8812	1088 bp
GETV-R11		ATGATTATGGCAGCGAGCGG	9861–9880	
GETV-F12	E2–E1	CCGGTAACACTAGGAGTACTATGC	9744–9767	989 bp
GETV-R12		TTGTCATTCAGCGACGTGCCT	10,712–10,732	
GETV-F13	E1–3′UTR	CCTCAAGTTGTCAAGACCTTCGTC	10,625–10,648	1066 bp
GETV-R13		GTAAAATATTAAAAAAACAAATTAGACGCC	11,661–11,690	

3. Results

3.1. Isolation and sequencing of GETV strain YN12042

The strain YN12042 was isolated from *C. tritaeniorhynchus* specimens collected along the border between Yunnan Province and Laos (latitude 21°69′N, longitude 100°5′E) in 2012. This virus can have cytopathic effects on BHK cells and is capable of continuous passage. A whole-genome sequence analysis revealed that the virus was GETV (GenBank accession number: KY450683).

3.2. Molecular characterization of GETV

3.2.1. Similarity of nucleotide and amino acid sequences

Because the E2 gene of GETV encodes the outer membrane protein and determines the antigenic relationship between virus and host (Jin, 2001), a molecular evolution analysis based on E2 may reveal the evolutionary dynamics between GETV and its hosts. A homology analysis of GETV E2 nucleotide and amino acid sequences revealed that the degree of homology among GETVs isolated in the past 60 years (1955–2014) was 93.8–100% (average: 98.4%) and 96.0–100% (average: 99.3%), respectively, compared with the prototype strain MM2021. This suggested that GETV has a conservative gene structure.

Further analysis revealed that the nucleotide homology between the newly isolated GETV strain (YN12042) and other GETVs was 94.6% (MM2021) to 99.8% (YN0540), and that the amino acid homology was 97.4% (MM2021) to 100% (South Korea).

3.2.2. Amino acid substitutions in E2

To characterize mutations in E2, we analyzed the amino acids encoded by E2 genes derived from the previously mentioned GETV strains. We found that E2 was 1266 nucleotides in length and encoded 422 amino acids in all GETV isolates. Compared with MM2021, seven common amino acid differences were identified in E27 (serine [Ser] → phenylalanine), E90 (threonine → valine), E102 (alanine → valine), E122 (isoleucine → threonine), E213 (arginine [Arg] → Ser), E314 (alanine → valine), and E323 (glutamic acid → aspartic acid, except in Gansu isolate GS10-2). Furthermore, the GETV isolates had different amino acid mutations located at various positions within E2 (Table 3). For example, there were seven amino acid differences between MM2021 and YN12031, but only one amino acid difference between MM2021 and YN12042.

3.3. Three-dimensional structural analysis of E2 of GETVs

We sought to determine whether the seven common mutations in E2 affected the structure of the encoded protein. To this end, we used SWISS-MODEL and YASARA and VMD software to model the three-dimensional structure and electrostatic potential of E2 from MM2021 (Fig. 3A). We applied the same methods to GETV strains isolated from 1956 to 2014 (Fig. 3B). Three of the seven divergent amino acids were located in the loop region (residues 323, 90, and 213), while the remaining four were located in the β-sheet (residues 27, 102, 122, and 314). These mutations had some effect on the structure and electrostatic properties of E2. In particular, the residue at position 213 in the loop was altered from Arg in the original strain to Ser in the epidemic strain. Notably, the R group structure and electrical charge of the two amino acids differed greatly, resulting in significant differences in structure and charge at this site (Fig. 3) between the two strains.

3.4. Molecular phylogenetic and population dynamics analyses of GETVs

3.4.1. Molecular phylogenetic analysis of GETVs

Fig. 2A presents the GETV maximum confidence tree results from our phylogenetic analysis. The MRCA of GETV existed about 145 years ago (95% HPD: 75–244) and gradually evolved into four major evolutionary populations. Group I is derived from the original GETV strain (MM2021) and is located at the root of the phylogenetic tree (red branch), suggesting that it is the oldest strain. Group II appeared 60 years ago (95% HPD: 59–73) and includes two GETVs isolated in Japan in 1956 (SAGE, Japan, 1956, and SAGE-original, Japan, 1956) (yellow branch). The MRCA of the GETVs in Group III (green branch) existed about 51 years ago (95% HPD: 51–72); this group includes all of the strains isolated from mosquitoes, pigs, horses, and other animals from 1964 to 2014 in China, Japan, Korea, and Mongolia. Group IV emerged about 28 years ago (95% HPD: 16–55) and includes YN12031

Table 2
GETV isolates analyzed in this study.

Strain	Date	Country	Host	GenBank accession no
MM2021	1955	Malaysia	*C. gelidus*	AF339484
GETV-SAGV	1956	Japan	Mosquito	AB032553
GETV-SAGV-Original	1956	Japan	-	AF339483
MI-110-C2	1978	Japan	*Equus caballus*	LC079087
MI-110-C1	1978	Japan	*Equus caballus*	LC079086
Kochi/01/2005	2005	Japan	*Sus scrofa*	AB859822
14-I-605-C2	2014	Japan	*Equus caballus*	LC079089
14-I-605-C1	2014	Japan	*Equus caballus*	LC079088
QIAG9303	1993	South Korea	Swine	KR081240
QIAG9302	1993	South Korea	Swine	KR081239
QIAG9301	1993	South Korea	Swine	KR081238
South Korea	2004	South Korea	Swine	AY702913
LEIV 16275 Mag	2000	Russia	*Aedes* sp.	EF631998
LEIV 17741 MPR	2000	Mongolia	*Culex* sp.	EF631999
M1	1964	China, Hainan	*Culex* sp.	EU015061
HB0234	2002	China, Hebei	*Culex tritaeniorhynchus* Giles	EU015062
HB0215-3	2002	China, Hebei	*Culex tritaeniorhynchus* Giles	EU015065
YN0540	2005	China, Yunnan	*Armigeres subalbatus*	EU015063
YN0542	2005	China, Yunnan	*Armigeres subalbatus*	EU015064
TC07180[a]	2007	China, Yunnan	*Culex pseudovishnui*	KY450684
LH07012[a]	2007	China, Yunnan	*Culex tritaeniorhynchus*	KY450685
DH10M1102[a]	2010	China, Yunnan	*Culex fuscocephala*	KY450686
DH10M1106[a]	2010	China, Yunnan	*Culex annulus*	KY450687
DH10M390[a]	2010	China, Yunnan	*Culex tritaeniorhynchus*	KY450688
DH10M1105[a]	2010	China, Yunnan	*Anopheles sinensis*	KY450689
YN12042[b]	2012	China, Yunnan	*Culex tritaeniorhynchus* Giles	KY450683
YN12031	2012	China, Yunnan	*Armigeres subalbatus*	KY434327
SH05-17	2005	China, Shanghai	*Culex tritaeniorhynchus*	EU015069
SH05-16	2005	China, Shanghai	*Culex tritaeniorhynchus*	EU015068
SH05-15	2005	China, Shanghai	*Culex tritaeniorhynchus*	EU015067
SH05-6	2005	China, Shanghai	*Culex tritaeniorhynchus*	EU015066
GS10-2	2006	China, Gansu	*Armigeres subalbatus*	EU015070
DY0824[a]	2008	China, Shandong	*Culex tritaeniorhynchus* Giles	KY434328
GZ0809[a]	2008	China, Guizhou	*Armigeres subalbatus*	KY457545
GZ0862[a]	2008	China, Guizhou	*Armigeres subalbatus*	KY457546
GZ0867[a]	2008	China, Guizhou	*Armigeres subalbatus*	KY457547
GZ0881[a]	2008	China, Guizhou	*Armigeres subalbatus*	KY457548
GZ0885[a]	2008	China, Guizhou	*Armigeres subalbatus*	KY457549
GZ0896[a]	2008	China, Guizhou	*Armigeres subalbatus*	KY457550
GZ08133[a]	2008	China, Guizhou	*Armigeres subalbatus*	KY457551
GZ08142[a]	2008	China, Guizhou	*Culex tritaeniorhynchus*	KY457552
SC1210	2012	China, Sichuan	*Armigeres subalbatus*	LC107870

-: not available in GenBank.
[a] Getah virus strains were isolated and identified in our laboratory in China and were registered in GenBank for the first time.
[b] First report of the Getah virus in this study.

(isolated in Yunnan Province, China) and LEIV/16275/Mag (isolated in Russia in 2000) (blue branch). The nucleotide substitution rate for E2 was $3.47 * 10^{-4}$ (95% HPD: $1.96 * 10^{-4}$ to $4.98 * 10^{-4}$).

3.4.2. Population dynamics analysis of GETV

The results of a population dynamics analysis showed that the GETV population dynamics remained stable from 1955 to 1970 (Fig. 2B). In 1970, the population diversity of GETV began to increase, peaking around 1990. From 1990 to 2005, virus population diversity declined, and was stable from 2005 to 2012.

4. Discussion

By phylogenetic analysis, we determined that GETV first appeared about 145 years ago and subsequently diverged into four distinct populations (Fig. 2A). The GETV strain isolated from mosquitoes in Malaysia in 1955 belongs to one branch, representing the oldest group (Group I). Two GETVs isolated in Japan in 1956 constitute Group II, while Group IV includes YN12031 (isolated in Yunnan Province in 2012; Li et al., 2017) and LEIV/16275/Mag (isolated in Russia in 2000). The strains apart from those in the groups above constitute Group III (green branch in the figure). These GETVs were isolated from 1964 to 2014, from areas including southern China (Hainan Province, latitude 19°N; Yunnan Province, latitude 21°69′N, longitude 100°5′E), northern China (Liaoning Province, latitude 42°N), Japan, South Korea, and Mongolia (latitude 48°N). The associated hosts include a variety of mosquito vectors, pigs, horses, and other animals. Furthermore, epidemic strains of GETV that afflict animals such as pigs and horses appear only in Group III (Fig. 2A).

In summary, Group III includes most of the GETV strains that cause animal diseases. Group III has completely replaced the original GETV populations (Groups I and II) in terms of its geographical distribution and range of mosquito vectors and host animals.

The SAGE virus was isolated from mosquitoes in Japan in 1956, and was initially considered an independent virus species (Wekesa et al., 2001). Further study suggested that SAGE virus is a type of GETV rather than an independent species of alphavirus (Shirako and Yamaguchi, 2000). In the present study, we found that the phylogenetic position of SAGE was intermediate between the original GETV strain (MM2021) and subsequently isolated strains, indicating that SAGE virus belongs within the GETV group. SAGE virus may therefore represent a transitional population between Groups I and III (Fig. 2A).

Notably, YN12031, isolated in Yunnan in 2012 (Li et al., 2017), was placed on the same branch as LEIV/16275/Mag (isolated in Russia in 2000). These two strains comprise Group IV, the GETV population that diverged most recently (about 30 years ago). Furthermore, a homology

Table 3
Comparison of E2 amino acid sequences between the Malaysian prototype strain MM2021 and other GETV strains.

Strain	Amino acid difference site																																				
	4	5	7	8	27	30	59	71	74	86	90	102	109	116	122	132	134	178	187	193	194	205	207	213	248	249	259	304	312	314	323	331	340	350	368	374	383
MM2021	E	H	N	V	**S**	S	G	A	D	H	**T**	**A**	D	Q	**I**	A	A	I	N	G	G	S	N	**R**	L	S	P	R	R	**A**	**E**	N	Q	W	A	G	A
GETV-SAGV	E	H	N	V	F	S	**R**	**V**	**Y**	H	V	V	D	Q	T	**D**	A	I	N	G	G	S	N	S	L	**P**	P	R	R	V	D	N	Q	W	A	G	A
GETV-SAGV -Original	E	H	N	V	F	S	**R**	A	**Y**	H	V	V	D	Q	T	**D**	A	I	N	G	G	S	N	S	L	**P**	P	R	R	V	D	N	Q	W	A	G	A
MI-110-C2	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
MI-110-C1	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	**E**	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
Kochi/01/2005	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	**H**	S	L	S	P	R	R	V	D	N	Q	W	A	G	**T**
14-I-605-C2	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
14-I-605-C1	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
QIAG9303	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	**E**	G	S	N	S	L	S	P	R	R	V	D	N	**R**	W	A	G	A
QIAG9302	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
QIAG9301	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
South Korea	**K**	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
LEIV/16275/ Mag	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
LEIV/17741/ MPR	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
M1	E	H	N	V	F	S	G	A	D	H	V	V	**G**	Q	T	A	A	I	N	G	G	**R**	N	S	L	S	P	R	R	V	D	N	Q	W	**V**	**C**	A
HB0234	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	**T**	N	G	G	S	N	S	L	S	**L**	R	R	V	D	N	Q	W	**T**	G	A
HB0215-3	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	**T**	N	G	G	S	N	S	L	S	**L**	R	R	V	D	N	Q	W	**T**	G	A
YN0540	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
YN0542	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
TC07180	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
LH07012	E	**D**	**I**	V	F	S	G	A	D	H	V	V	D	Q	T	**T**	A	I	N	G	G	S	N	S	**S**	S	P	R	R	V	D	N	Q	W	A	G	A
DH10M1102	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
DH10M1106	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
DH10M390	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
DH10M1105	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
YN12042	**K**	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
YN12031	E	H	N	V	F	**R**	G	A	D	**Y**	V	V	D	**K**	T	A	**V**	I	N	G	G	**N**	N	S	L	S	P	R	**Q**	V	D	N	Q	W	A	G	A
SH05-17	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	**T**	Q	**L**	A	G	A
SH05-16	E	H	N	**A**	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
SH05-15	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
SH05-6	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
GS10-2	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	**E**	N	Q	W	A	G	A
DY0824	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
GZ0809	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
GZ0862	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
GZ0867	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
GZ0881	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
GZ0885	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
GZ0896	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
GZ08133	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
GZ08142	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	N	G	G	S	N	S	L	S	P	R	R	V	D	N	Q	W	A	G	A
SC1210	E	H	N	V	F	S	G	A	D	H	V	V	D	Q	T	A	A	I	**K**	G	G	S	N	S	L	S	P	**L**	R	V	D	N	Q	W	A	G	A

Note: The shaded portion of the table represents seven co-mutated amino acids; the bold font represents an amino acid different from that of the other strain, and the ungrazed font represents the same amino acid as the other strain.

analysis of GETV E2 showed that the nucleotide and amino acid sequences of YN12031 and LEIV/16275/Mag had the highest identities (97.4 and 98.1%, respectively; results not shown). This is striking, given that these strains are remote from one another, at a distance of thousands of kilometers and a latitude difference of about 38° (Yunnan Province is located at 22°N; Russia is located at 60°N). This suggests that GETV can exist not only in tropical regions but also at high latitudes, and is capable of adapting to a range of environments.

Mosquito-borne viruses have a global distribution. Dengue virus, Japanese encephalitis virus, West Nile virus, Chikungunya virus, and others are the most harmful arboviruses in the world (Weaver and Reisen, 2010). Studies have shown that the MRCA for Japanese encephalitis virus, West Nile virus, Yellow fever virus, and Chikungunya virus existed about 1695 (Pan et al., 2011), 200 (May et al., 2011), 306 (Bryant et al., 2007), and 500 (Volk et al., 2010) years ago, respectively. While the MRCA of GETV (145 years ago) appeared significantly more recently than did those of most of the viruses mentioned above, it is similar to that of Zika virus (166 years ago), which has been widely distributed throughout the South American continent since 2015 (Liu et al., 2016). Thus, GETV should be considered an "emerging mosquito-borne virus."

GETV was first isolated from *Culex* collected in Malaysia (latitude 3°N) in southern tropical regions of the globe, and was later isolated from various mosquito specimens and animals (horses and pigs in Japan and India). The geographic distribution of mosquito-borne GETV now extends from the tropics to Russia (Fig. 1A). At present, GETVs have been isolated from mosquitoes belonging to nine species within four genera, including five species of *Culex* (excluding unidentified *Culex* sp.; Table 2). More GETVs have been isolated from mosquito specimens collected in mainland China than in any other country. These GETVs

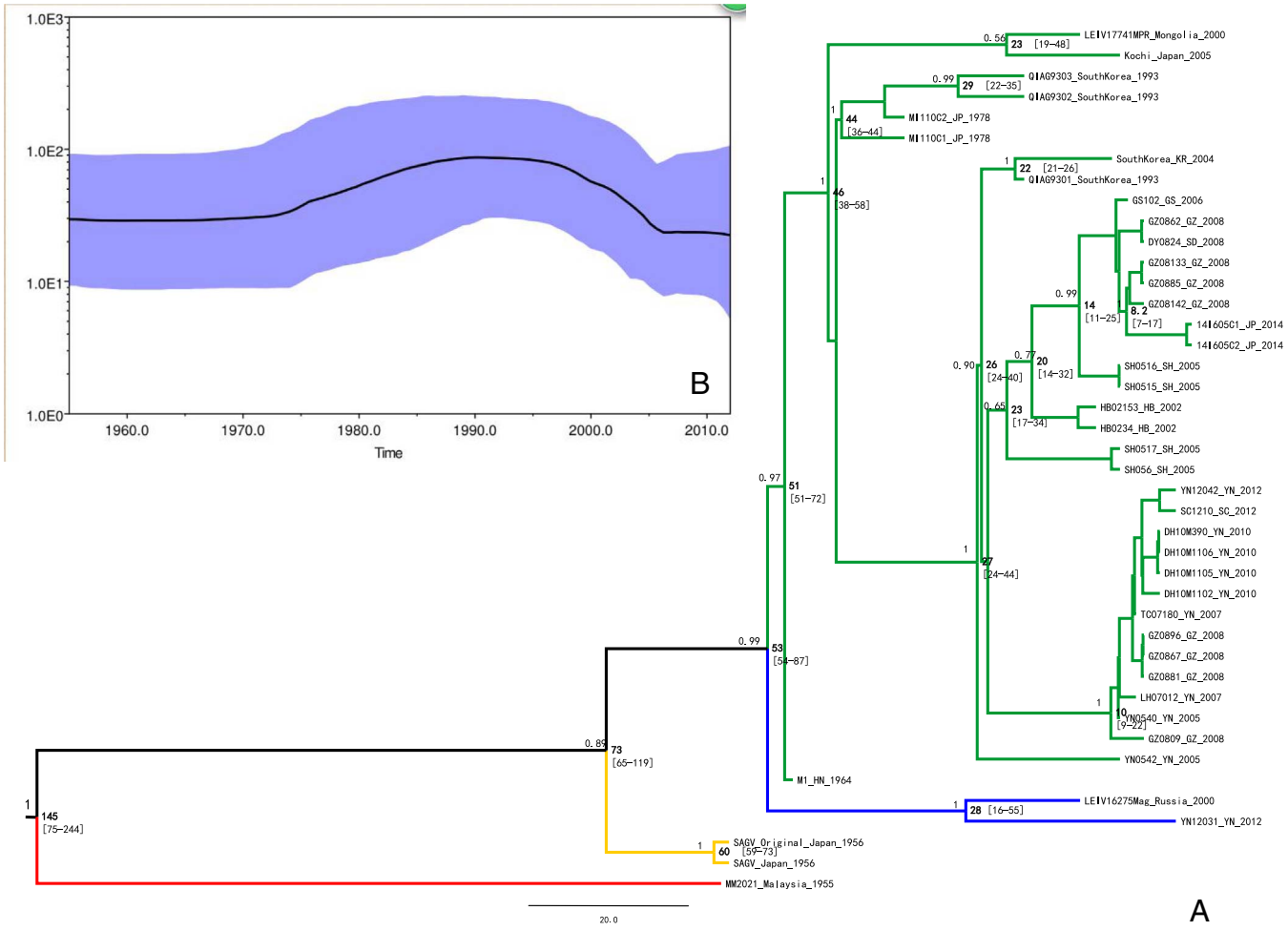

Fig. 2. A. Maximum clade credibility tree for E2 gene sequences of GETV. B. Bayesian skyride plots for GETV. The highlighted areas correspond to 95% HPD intervals.

have been collected from seven mosquito species within four genera, including *Culex* sp. (Li et al., 1992) from which strain M1 was isolated in 1965, *C. tritaeniorhynchus* (Zhai et al., 2008; Feng et al., 2014a; Li et al., 2012), *Culex pseudovishnui* (Feng et al., 2014b), *Culex fuscocephalus* (Feng et al., 2014a), *Culex annulus* (Feng et al., 2014a), *Anopheles sinensis* (Feng et al., 2014a), *Aedes albopictus* (Zhou et al., 2012), *Armigeres* (Zhai et al., 2008; Li et al., 2012), and *C. tritaeniorhynchus* from which strain YN12042 was isolated in the present study. From 1964 (when China's first GETV strain was isolated) to 2014, 28 GETV strains were isolated in mainland China, accounting for 66.7% (28/42) of all GETV isolates in the globe. The isolates were located in regions ranging in latitude from 19 to 42°N, and in longitude from 95 to 125°E. Therefore, GETV has a broad geographical distribution in China (Fig. 1B).

Our results also show that the GETVs isolated from mosquitoes and horse or swine specimens were located at different positions on our phylogenetic tree, and they did not exhibit vector or host specificity. GETV thus exhibits strong host adaptability.

The structural (E2) gene of GETV is conserved; the calculated homologies of the nucleotide and amino acid sequences were 93.8–100% and 96.0–100%, respectively. However, compared with MM2021, there were seven common amino acid differences in E2 among the isolates (Table 3B). We hypothesized that these mutations may have played a role in the gradual spread of GETV from the tropics to other areas. We first established a three-dimensional model of the protein using strain MM2021 (Fig. 3A). We then established three-dimensional structural models associated with amino acid mutations (Fig. 3B). The results showed that the seven amino acid mutations altered the structure and electrostatic presentation of E2 between the original strain and other GETVs. The latter isolates have a loop-surface interface and divergent electrostatic potential that may alter the binding properties of E2 to host factors; this may have contributed to the recent geographical radiation of the virus. However, the crystal structure of GETV E2 has not been characterized; the present results are based on simulations. Although our results suggest structural differences in E2 between the epidemic strains and MM2021, in-depth and detailed studies are required to understand the impact of specific mutations on the structure and function of this protein.

GETV has a broad geographical distribution and a strong host adaptability, and has caused several horse and pig disease outbreaks over the past 20 years. GETV is therefore of great public health significance. To prevent viral epidemics and associated economic losses, it is essential to improve detection and monitoring programs for this virus. Such programs may focus on regions where GETV has been isolated from mosquitoes but is not yet known to infect horses and pigs.

Conflict of interest

None.

Acknowledgements

This work was supported by grants from National Natural Science Foundation of China (81290342, 81301479 and 81501757); The Natural Science Foundation of Shandong Province, China (ZR2013HQ008) and the young scientist supportive program of

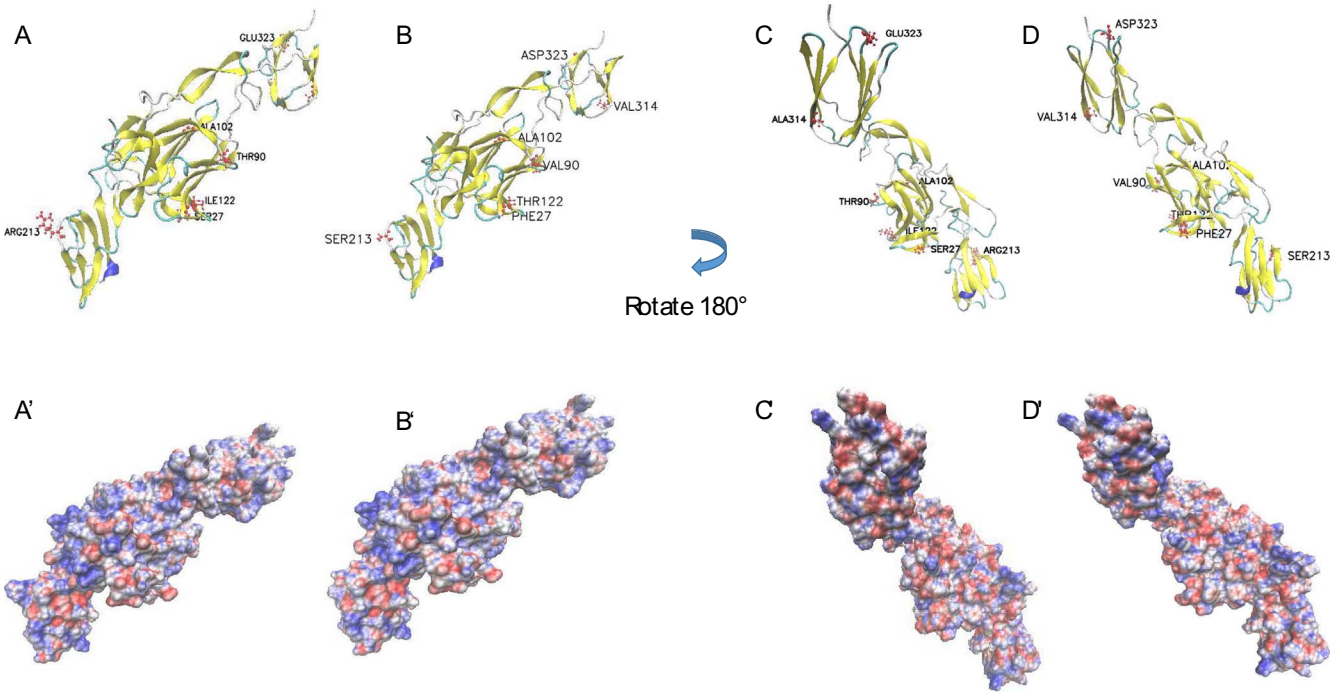

Fig. 3. Model of the three-dimensional structure and electrostatic potential in GETV E2 gene. (A) The three-dimensional structure and mutation site diagram of prototype strain MM2021. (A′) The surface charge distribution map of prototype strain MM2021. (B) The three-dimensional structure and mutation site diagram of epidemic GETV strains isolated from 1956 to 2014. (B′) The surface charge distribution map of epidemic GETV strains isolated from 1956 to 2014. (C) The three-dimensional structure and mutation site diagram of prototype strain MM2021. (C′) The surface charge distribution map of prototype strain MM2021. (D) The three-dimensional structure and mutation site diagram of epidemic GETV strains isolated from 1956 to 2014. (D′) The surface charge distribution map of epidemic GETV strains isolated from 1956 to 2014.

Shandong University of Technology; The Special National Project on Research and Development of Key Biosafety Technologies (2016YFC1201904); National Key Plan for Scientific Research and Development of China (2016YFD0500300), Development Grant of State Key Laboratory of Infectious Disease Prevention and Control (2014 SKLID103 and 2015SKLID505) and CAS Key Laboratory of Special Pathogens and Biosafety, Chinese Academy of Sciences, Funded by Open Research Fund Program of Wuhan National Bio-Safety Level 4 Lab of CAS (2017SPCAS003). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contribution

Conceived and designed the experiments: LYY, LH, FSH and LGD. Performed the experiments: LYY, LH, FSH and WLX. Isolated the virus and analyzed the data: GXF, LMH, FY, CWX, LWW,GXY, LZ, HY, WHY and ZHN. Wrote the paper: LYY, LH, LXL, WGQ and LGD. All authors have read and approved the final version of the manuscript.

References

Bannai, H., Ochi, A., Nemoto, M., Tsujimura, K., Yamanaka, T., Kondo, T., 2016. A 2015 outbreak of Getah virus infection occurring among Japanese racehorses sequentially to an outbreak in 2014 at the same site. BMC Vet. Res. 12, 98. http://dx.doi.org/10.1186/s12917-016-0741-5.

Broen, C.M., Timoney, P.J., 1998. Getah virus infection of Indian horses. Trop. Anim. Health Prod. 30, 24l–252.

Brummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214. http://dx.doi.org/10.1186/1471-2148-7-214.

Bryant, J.E., Holmes, E.C., Barrett, A.D., 2007. Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. PLoS Pathog. 3, e75.

Feng, Y., Fu, S.H., Yang, W.H., Zhang, Y.Z., Yang, J., Liu, Y.H., Dong, C.L., Li, S., Gao, X.Y., Jiang, W., Yin, X.X., Yang, S.C., Guo, T.Q., Zhang, B.S., Yin, Z.L., Liang, G.D., Zhang, H.L., 2014a. Molecular characterization of mosquito-borne viruses strains newly isolated in Western Yunnan Province along the border of China and Myanmar, China. Int. J. Virol. 21, 193–198 (in Chinese).

Feng, Y., Li, S.G., Zhang, H.L., Fu, S.H., Kang, X.H., Cun, D.Q., Guo, C., Zhang, Y.Z., Yang, W.H., Wang, H.Y., Liang, G.D., 2014b. Melecular characterization of Japanese encephalitis virus and Getah virus strains newly isolated in Tengchong County,Yunnan Province, China. Chin. J. Zoonoses 30, 353–363 (in Chinese).

Fukunaga, Y., Kumanomido, T., Kamada, M., 2000. Getah virus as an equine pathogen. Vet. Clin. North Am. Equine Pract. 16, 605–617.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD-visual molecular dynamics. J. Mol. Graph. 14, 33–38.

Jin, Q., 2001. Medical Molecular Virology. Medical Publishing Center, Science Press (ISBN 7-03-008132-3).

Kamada, M., Ando, Y., Fukunaga, Y., Kumanomido, T., Imagawa, H., Wada, R., Akiyama, Y., 1980. Equine Getah vires infection: isolation of the virus from racehorses duringan enzootic in Japan. Am. J. Trop. Med. Hyg. 29, 984–988.

Karabatsos, N., 1985. International Catalogue of Arthropod-Borne Viruses, 3rd ed. American Society for Tropical Medicine and Hygiene, San Antonio (TX).

Kono, Y., 1988. Getah virus disease. In: Monath, T.P. (Ed.), The Arboviruses. Epidemiology and Ecology. CRC Press, Boca Raton, FL, pp. 21–36.

Krieger, E., Vriend, G., 2002. Models@Home: distributed computing in bioinformatics using a screensaver based approach. Bioinformatics 18, 315–318.

Kumanomido, T., Wada, R., Kanemaru, T., Kamada, M., Hirasawa, K., Akiyama, Y., 1988. Clinical and virological observatiom on swine experimentally infected with Getah virus. Vet. Microbiol. 16, 295–301.

Lei, W.W., Guo, X.F., Fu, S.H., Feng, Y., Nie, K., Song, J.D., Li, Y., Ma, X.J., Liang, G.D., Zhou, H.N., 2015. Isolation of Tibet orbivirus, TIBOV, from *Culicoides* collected in Yunnan, China. PLoS One 10, e0136257.

Li, X.D., Qiu, F.X., Yang, H., Rao, Y.N., Calisher, C.H., 1992. Isolation of Getah virus from mosquitos collected on Hainan Island, China, and results of a serosurvey. Southeast Asian J. Trop. Med. Public Health 23, 730–734.

Li, M.H., Fu, S.H., Feng, Y., Ye, X.F., Zhao, S.Y., Liu, C.T., Liang, G.D., 2012. Investigation of arboviruses in different regions of Guizhou province, China in 2008. Chin. J. Vector. Biol. Control. 23, 417–420.

Liu, H., Li, M.H., Zhai, Y.G., Meng, W.S., Sun, X.H., Cao, Y.X., Fu, S.H., Wang, H.Y., Xu, L.H., Tang, Q., Liang, G.D., 2010. Banna virus, China, 1987–2007. Emerg. Infect. Dis. 16, 514–517.

Liu, H., Shen, L., Zhang, X.L., Li, X.L., Liang, G.D., Ji, H.F., 2016. From discovery to outbreak: the genetic evolution of the emerging Zika virus. Emerg. Microbes. Infect. 5, e111.

Li, Y.Y., Fu, S.H., Guo, X.F., Lei, W.W., Li, X.L., Song, J.D., Cao, L., Gao, X.Y., Lv, Z., He, Y., Wang, H.Y., Ren, X.J., Zhou, H.N., Wang, G.Q., Liang, G.D, 2017. Identification of a Newly Isolated Getah Virus in the China-Laos Border, China. Biomed. Environ. Sci. 30, 210–214.

L'vov, S.D., Gromashevski'i, V.L., Aristova, V.A., Morozova, T.N., Skvortsova, T.M., Gushchina, E.A., Petrova, E.S., L'vov, D.K., 2000. Isolation of Getah virus (Togaviridae, Alfavirus) strains in North-Eastern Asia. Vopr. Virusol. 45, 14–18 (in Russian).

May, F.J., Davis, C.T., Tesh, R.B., Barrett, A.D., 2011. Phylogeography of West Nile virus:

from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. J. Virol. 85, 2964–2974.
Nemoto, M., Bannai, H., Tsujimura, K., Kobayashi, M., Kikuchi, T., Yamanaka, T., Kondo, T., 2014. Getah virus infection among racehorses, Japan, 2014. Emerg. Infect. Dis. 21, 883–885.
Pan, X.L., Liu, H., Wang, H.Y., Fu, S.H., Liu, H.Z., Zhang, H.L., Li, M.H., Gao, X.Y., Wang, J.L., Sun, X.H., Lu, X.J., Zhai, Y.G., Meng, W.S., He, Y., Wang, H.Q., Han, N., Wei, B., Wu, Y.G., Feng, Y., Yang, D.J., Wang, L.H., Tang, Q., Xia, G., Kurane, I., Rayner, S., Liang, G.D., 2011. Emergence of genotype I of Japanese encephalitis virus as the dominant genotype in Asia. J. Virol. 85, 9847–9853.
Posada, D., 2003. Using MODELTEST and PAUP* to select a model of nucleotide substitution. Curr. Protoc. Bioinformatics. http://dx.doi.org/10.1002/0471250953.bi0605s00. (Chapter 6, Unit 6.5).
Sentsui, H., Kono, Y., 1985. Reappearance of Getah virus infection among horses in Japan. Jpn. J. Vet. Sci. 47, 333–335.
Shirako, Y., Yamaguchi, Y., 2000. Genome structure of Sagiyama virus and its relatedness to other alphaviruses. J. Gen. Virol. 81, 1353–1360.
Sugiura, T., Ando, Y., Imagawa, H., Kumanomido, T., Fukunaga, Y., Kamada, M., Wada, R., Hirasawa, K., Akiyama, Y., 1981. An epizootiological study of Getah virus among light horses in Japan in 1979. Bull. Equine. Res. Inst. 18, 103–109.
Volk, S.M., Chen, R., Tsetsarkin, K.A., Adams, A.P., Garcia, T.I., Sall, A.A., Nasar, F., Schuh, A.J., Holmes, E.C., Higgs, S., Maharaj, P.D., Brault, A.C., Weaver, S.C., 2010. Genome-scale phylogenetic analyses of chikungunya virus reveal independent emergences of recent epidemics and various evolutionary rates. J. Virol. 84, 497–504.
Weaver, S.C., Reisen, W.K., 2010. Present and future arboviral threats. Antivir. Res. 85, 328–345.
Weaver, S.C., Frey, T.K., Huang, H.V., Kinney, R.M., Rice, C.M., Roehrig, J.T., Shope, R.E., Strauss, E.G., 2005. Togaviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier/Academic Press, London, pp. 999–1008.
Wekesa, S.N., Inoshima, Y., Murakami, K., Sentsui, H., 2001. Genomic analysis of some Japanese isolates of Getah virus. Vet. Microbiol. 83, 137–146.
Yago, K., Hagiwara, S., Kawamura, H., Narita, M., 1987. A fatal case in newborn piglets with Getah virus infection: isolation of the virus. Nippon Juigaku Zasshi 49, 989–994.
Zhai, Y.G., Wang, H.Y., Sun, X.H., Fu, S.H., Wang, H.Q., Attoui, H., Tang, Q., Liang, G.D., 2008. Complete sequence characterization of isolates of Getah virus (genus *Alphavirus*, family Togaviridae) from China. J. Gen. Virol. 89, 1446–1456.
Zhou, X.F., Feng, Z.L., Hu, T.S., Qiu, W., Fan, Q.S., Li, L.J., Zhang, F.Q., 2012. Isolation and identification of Getah virus from *Aedes albopictus* in Yunnan. Chin. J. Heal. Lab. Tech. 22, 103–106 (in Chinese).