

# Molecular evolution and epidemiology of four serotypes of dengue virus in Thailand from 1973 to 2007

S. P. CHEN\*

State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, PR China

Received 29 November 2011; Final revision 18 March 2012; Accepted 14 April 2012;  
first published online 14 May 2012

## SUMMARY

Thailand was a hyper-endemic country for dengue with co-circulation of four serotypes and tens of thousands of infected cases annually. Taking into consideration the large number of local dengue virus (DENV) sequences available in GenBank, Thailand was the most ideal locality to study co-evolution of DENV. Therefore, we undertook a large-scale molecular epidemiological analysis of all DENV strains isolated in Thailand. In this study, we demonstrated that DENV strains of four serotypes post-1990 grouped into distinct clades, and that specific mutations in the envelope protein were first confirmed in these clades. Compared to the DENV1, DENV2 and DENV3 clades, the DENV4 clade evolved markedly more slowly ( $6.4 \times 10^{-5}$  substitutions/site per year). Our results also showed that the genetic diversity of the predominant genotype of each serotype tended to slightly increase over time with fluctuating changes, followed by a stationary phase after 2000. This suggests that the four DENV clades became the predominant strains due to DENV possessing improved fitness after long-term selection.

**Key words:** Clade replacement, dengue virus, evolution.

## INTRODUCTION

Dengue virus (DENV) belongs to family Flaviviridae, genus *Flavivirus*. Like all *Flavivirus* members, it has a single-stranded positive-sense RNA genome which encodes three structural proteins (C, prM, E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5). Humans are usually infected by mosquito (e.g. *Aedes aegypti*) bite and develop self-limited febrile disease called dengue fever (DF), or potentially fatal dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS). In recent decades, probably due to globalization, urbanization, human

population increase and lack of effective mosquito control, four serotypes of DENV (DENV1 to DENV4) have been distributed to almost all tropical and subtropical countries and regions, with increasing numbers of humans becoming infected. It is estimated that 50–100 million infections occur annually, and more than 2.5 billion people are at risk [1, 2].

Thailand is representative of countries where DENVs are endemic almost annually, with co-circulation of four serotypes. Since the viruses were first confirmed in 1958 [3], the incidence of dengue diseases has increased markedly. From 1985 to 2006, the case load has been consistently above 10 000. In addition, there are sufficient sequences of virus strains isolated from this country; therefore, Thailand is an ideal locality to address the long-term molecular evolution of DENVs. In this study, we analysed all

\* Address for correspondence: Dr S. P. Chen, Beijing Institute of Microbiology and Epidemiology, NO 20, Dongdajie Street, Fengtai District, Beijing 100071, PR China.  
(Email: shpchen@hotmail.com)

DENV sequences with known isolation date to elucidate co-evolution of DENVs in Thailand.

## MATERIALS AND METHODS

### Sequence datasets

The DENV E protein was the major antigen and was also associated with virus binding and entry. Thus, the majority of phylogenetic and evolutionary analysis of DENV had focused on the E protein [4]. All complete E gene sequences were collected from GenBank ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). The following criteria were used when choosing the sequences: (1) all sequences had a known date of isolation; (2) all putative recombinant sequences were excluded [5–7]; (3) all sylvatic strains were excluded; (4) sequences with 100% similarity were excluded. DENV strain numbers and date of isolation range were: 114 strains for DENV1 (1980–2007), 144 strains for DENV2 (1974–2006), 98 strains for DENV3 (1973–2005), and 61 strains for DENV4 (1976–2007). No recombinant DENV strains were identified in these cohorts. All sequences and alignments in this study are available from the author upon request.

### Phylogenetic analysis

The TREE-PUZZLE program [8] was used to infer maximum-likelihood trees for DENV1, DENV2, DENV3, and DENV4. Trees were rooted by using strains from different serotypes.

### Molecular clock and Bayesian skyline plot analysis

The rate of nucleotide substitution per site and the time to the most recent common ancestor (TMRCA) were estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in the BEAST 1.5.3 package (<http://beast.bio.ed.ac.uk>) [9]. The analysis was performed using the GTR + G +  $\Gamma$ 4 substitution model under a coalescent model of constant population size. In each case, the relaxed (uncorrelated lognormal) molecular clock model was used (the relaxed clock model was favoured over the strict clock model in these datasets). A final chain length of 10 million or 20 million was employed to make an effective sample size for parameter estimates >200. The resulting convergence and the relative genetic diversity were analysed using Tracer 1.4.1 (<http://evolve.zoo.ox.ac.uk>). To reveal the uncertainty

in the estimations, the 95% high probability density (HPD) interval in each case was also determined.

## RESULTS

### Phylogenetic analysis

According to our phylogenetic analysis of 114 DENV1 E gene sequences (Supplementary Fig. S1*a*), all DENV1 strains consisted of two genotypes (genotypes I and III). Almost all DENV1 strains belonged to genotype I (110 strains), with the exception of four strains isolated in 1980 (ThD1\_0673\_80, ThD1\_0442\_80, ThD1\_0127\_80, ThD1\_0038\_83). Most notably, all genotype I strains could be divided into two distinct clades. The first clade (clade 1980–1994) consisted of 40 strains from 1980 to 1994 and the second clade (clade 1990–2007) consisted of 70 strains from 1990 to 2007.

DENV2 strains (Supplementary Fig. S1*b*) consisted of three genotypes (genotype Asian I, Cosmopolitan, Asian/American). Almost all strains belonged to genotype Asian I (133 strains) except for 11 strains (CAMR14 fell into genotype Cosmopolitan and the other 10 strains fell into genotype Asian/American). One clade (clade 1983–2006), which consisted of 100 strains from 1983 to 2006, was identified in genotype Asian I. Only 21% (11/52) of strains isolated prior to 1990 fell into clade 1983–2006.

For DENV3, only two genotypes (I and II) circulated in Thailand (Supplementary Fig. S1*c*). Genotype I consisted of 11 strains isolated in 2005. The remaining 87 strains fell into genotype II. One clade (clade 1991–2003), which consisted of 56 strains post-1990, was identified in genotype II, except for three strains (98901640, 98901590, 00-27-1HuNIID2000).

For DENV4, three genotypes (I, II, III) were confirmed (Supplementary Fig. S1*d*). All strains belonged to genotype I (55 strains), with the exception of one strain (ThD4\_0734\_00) in genotype II and five strains (ThD4\_1270\_98, ThD4\_0439\_01, ThD4\_0476\_97, ThD4\_0164\_99, ThD4\_0017\_97) in genotype III. One clade (clade 1991–2007), which consisted of 37 strains from 1991 to 2007, was identified in genotype I.

### Specific mutations in each clade

In the above-mentioned clades, we found that all strains within these clades contained specific mutations in the envelope protein compared to those strains outside the corresponding clade (Table 1).

Table 1. Substitution mutations in the envelope protein of predominant genotypes

	Position in envelope protein	Strains outside clades	Strains within clades	Location
DENV1 (genotype I)	8	Asn	Ser	Domain I
	324	Val	Ile	Domain III
	351	Val	Leu	Domain III
DENV2 (genotype Asian I)	83	Asn	Lys	Domain II
DENV3 (genotype II)	479	Ala	Val	TM2
DENV4 (genotype I)	96	Val	Met	Domain II
	203	Lys	Asn, Ser	Domain II
	461	Phe	Leu	TM1

TM, Transmembrane domain.

Table 2. Evolutionary processes on E gene of dengue viruses

Dataset	Rate ( $\times 10^{-4}$ substitutions/site per year, 95% HPD interval)	Age (years) (95% HPD interval)
DENV1 Clade 1990–2007 ( $n=70$ )	8.6 (6.5–11.0)	31.4 (20.6–46.7)
DENV2 Clade 1983–2006 ( $n=100$ )	8.5 (7.0–10.0)	26.0 (24.0–28.4)
DENV3 Clade 1991–2003 ( $n=56$ )	8.7 (6.7–10.8)	17.1 (13.8–21.2)
DENV4 Clade 1991–2007 ( $n=37$ )	0.64 (0.52–0.77)	1713.4 (1699.0–1749.8)

HPD, High probability density.

Specifically, three (Asn<sup>8</sup>Ser, Val<sup>324</sup>Ile, Val<sup>351</sup>Leu), one (Asn<sup>83</sup>Lys), one (Ala<sup>479</sup>Val), and three [Val<sup>96</sup>Met, Lys<sup>203</sup>Asn(Ser), Phe<sup>461</sup>Leu] mutations were identified in DENV1–4 clades, respectively.

### Evolutionary rates and ancestral time analysis

Our Bayesian MCMC analysis (Table 2) revealed that overall substitution rates of DENV1 ( $n=70$ ), DENV2 ( $n=100$ ) and DENV3 ( $n=56$ ) clade datasets were between  $(8.5\text{--}8.7) \times 10^{-4}$  substitutions/site per year. However, the rate of DENV4 clade dataset ( $n=37$ ) was markedly lower ( $6.4 \times 10^{-5}$  substitutions/site per year). Ages of the TMRCA for each DENV clade were 31.4 (DENV1), 26.0 (DENV2), 17.1 (DENV3), and 1713.4 (DENV4) years, respectively.

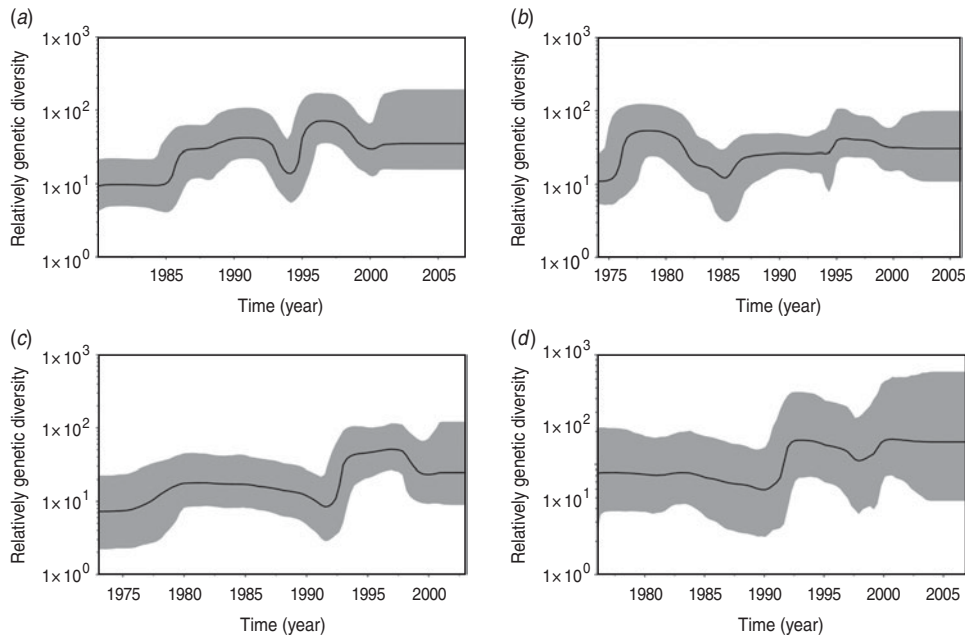
### Estimate of genetic diversity

Bayesian skyline plot analysis (Fig. 1) indicated that the relative genetic diversity of the predominant

genotype of each serotype tended to slightly increase over time followed by a stationary phase after 2000.

### Discussion

In this study we confirmed that strains of DENV2 and DENV4 post-1990 grouped into clades which were significantly different from those prior to 1990, as DENV1 and DENV3 described here and previously [10]. The origin and dispersion of *Flavivirus* in some countries has been described previously [11, 12]. In this study, we hypothesized that DENV1–3 clades could have been introduced from other countries or evolved *in situ* in the late 1970s or early 1980s (DENV1, 31.4 years old; DENV2, 26.0 years old; DENV3, 17.1 years old). However, DENV4 clade appeared 1713 years ago (DENV4, 1713 years old). Because these strains harboured specific mutations, they might have improved fitness in mosquitoes (e.g. *Aedes aegypti*) or humans, allowing them to survive better than other strains. In fact, these specific



**Fig. 1.** Bayesian skyline plot analysis of (a) DENV1, (b) DENV2, (c) DENV3, (d) DENV4. The black lines indicate the median estimates, and the shaded areas indicate the 95% high probability density intervals.

mutations were first confirmed in the envelope protein of these strains (Table 1), e.g. Asn<sup>8</sup>→Ser<sup>8</sup>, Val<sup>324</sup>→Ile<sup>324</sup>, Val<sup>351</sup>→Leu<sup>351</sup> in DENV1 genotype I, Asn<sup>83</sup>→Lys<sup>83</sup> in DENV2 genotype Asian I, Ala<sup>479</sup>→Val<sup>479</sup> in DENV3 genotype II, Val<sup>96</sup>→Met<sup>96</sup>, Lys<sup>203</sup>→Thr<sup>203</sup>(Ala<sup>203</sup>), Phe<sup>461</sup>→Leu<sup>461</sup> in DENV4 genotype I. The above-mentioned mutations were located within the important structural domains of the outer E protein (Table 1). E was believed to mediate receptor binding, membrane fusion, and it also induced protective immunity [13]. According to the crystal structure of E protein [14, 15], it consisted of three domains (domains I, II, III), followed by the stem region (H1, H2) and the transmembrane domain (TM1, TM2). Domain I was the molecular hinge region involved in low-pH-triggered conformational changes [16], domain II was involved in dimerization and membrane fusion [14], domain III played a role in receptor binding [14], and transmembrane domains were involved in virion assembly [17]. As described previously [18, 19], lineage replacement and point mutations often improve viral fitness, which might be why strains in these clades could be repeatedly endemic in Thailand under purifying selection [10, 20, 21] and/or immune selection [22]. Due to limited availability of full-length viral genomes, only the envelope protein was analysed in this study. Therefore, some specific mutations might also exist in non-structural proteins, other structural proteins

(e.g. C and prM) and non-coding regions of these clades because some virulent sites have already been identified in these regions [23–25]. Certainly, DENV mutants harbouring these mutations should be constructed to validate this deduction.

The epidemiological feature of DENVs in Thailand is considerably different from those in Singapore, a neighbouring country in South East Asia, where more genotypes were co-circulating in the last decade [26]. The high diversity of DENVs in Singapore might be due to increased globalization in Singapore, as this country was a regional and global commercial and travel centre and therefore able to frequently exchange and import DENV through human movement.

Moreover, substitution rates of DENV1, DENV2 and DENV3 clades presented here are consistent with those described previously [20, 27]. However, the DENV4 clade evolved much more slowly ( $6.4 \times 10^{-5}$  substitutions/site per year) than the other three clades described here and the DENV4 strains isolated in Thailand prior to 2002 [21]. In fact, the substitution rate of strains prior to 1990 (18 strains outside clade 1991–2007) in DENV4 genotype I was  $1.37 \times 10^{-3}$  substitutions/site per year (95% HPD interval:  $0.85 \times 10^{-3}$  to  $1.94 \times 10^{-3}$ ), which was 21.4 times higher than for DENV4 clade 1991–2007. The lower substitution rate showed a lower replication cycle of virus, particularly tropism, and transmission modes [28, 29]. This might be one reason that DENV4 was

the least frequently circulating serotype in Thailand since 1991 [9].

As shown in Figure 1, the predominant genotypes of DENV1–4 showed a similar demographic pattern in genetic diversity: several ‘invasion’ phases followed by a ‘stationary’ phase which started from 2000. The ‘invasion’ phase was characterized by an increase in the genetic diversity of DENV. Markedly, apparent ‘invasion’ phases (1994 for DENV1, 1985 for DENV2, 1992 for DENV3, 1991 for DENV4) appeared just after the time of the first isolated strain in each DENV clade (1990 for DENV1, 1983 for DENV2, 1991 for DENV3, 1991 for DENV4), which suggested that new strains in DENV1–4 clades began to be responsible for the local dengue outbreak. Consistent with such a fact: DENV1 clade 1990–2007, DENV2 clade 1983–2006, DENV3 clade 1991–2003, and DENV4 clade 1991–2007 were predominant from 1994, 1985, 1992, and 1991, respectively. However, there were only four strains in DENV1 clade 1990–2007 but 15 strains outside this clade from 1990 to 1993 (Supplementary Fig. S1a); there were only three strains in DENV2 clade 1983–2006 but eight strains outside this clade from 1983 to 1984 (Supplementary Fig. S1b); and there was one strain in DENV3 1991–2003 and one strain outside this clade in 1991 (Supplementary Fig. S1c). Because strains in DENV1–4 clades had improved fitness in hosts (humans and mosquitoes), they survived better than other strains and were predominant so that the genetic diversity remained constant after 2000.

In summary, clade replacements in DENV1–4 took place in strains isolated in Thailand. The DENV strains in the new clades harbour specific mutations in the E protein and are responsible for local dengue epidemics post-1990.

## SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S0950268812000908>.

## DECLARATION OF INTEREST

None.

## REFERENCES

1. **Gubler DJ.** Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in

- the 21st century. *Trends in Microbiology* 2002; **10**: 100–103.
2. **Gubler DJ.** The changing epidemiology of yellow fever and dengue, 1900–2003: full circle? *Comparative Immunology, Microbiology & Infectious Diseases* 2004; **27**: 319–330.
3. **Halstead SB, et al.** The Thai hemorrhagic fever epidemic of 1962 (a preliminary report). *Journal of the Medical Association of Thailand* 1963; **46**: 449–462.
4. **Holmes EC, et al.** The origin, emergence and evolutionary genetics of dengue virus. *Infection, Genetics and Evolution* 2003; **3**: 19–28.
5. **Holmes EC, et al.** Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution* 1999; **16**: 405–409.
6. **Worobey M, et al.** Widespread intra-serotype recombination in natural populations of dengue virus. *Proceedings of the National Academy of Sciences USA* 1999; **96**: 7352–7357.
7. **Toulou HJ, et al.** Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. *Journal of General Virology* 2001; **82**: 1283–1290.
8. **Schmidt HA, et al.** TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 2002; **18**: 502–504.
9. **Drummond AJ, et al.** BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 2007; **7**: 214.
10. **Zhang C, et al.** Clade replacements in dengue virus serotypes 1 and 3 are associated with changing serotype prevalence. *Journal of Virology* 2005; **79**: 15123–15130.
11. **Ospina MC, et al.** Prolonged co-circulation of two distinct dengue virus type 3 lineages in the hyperendemic area of Medellín, Colombia. *American Journal of Tropical Medicine and Hygiene* 2010; **83**: 672–678.
12. **Chen SP.** Molecular phylogenetic and evolutionary analysis of Japanese encephalitis virus in China. *Epidemiology and Infection*. Published online: 1 December 2011. doi: S095026881100255X.
13. **Whitehead SS, et al.** Prospects for a dengue virus vaccine. *Nature Reviews Microbiology* 2007; **5**: 518–528.
14. **Rey FA, et al.** The envelope glycoprotein from tick-borne encephalitis virus at 2Å resolution. *Nature* 1995; **375**: 291–298.
15. **Modis Y, et al.** A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proceedings of the National Academy Sciences* 2003; **100**: 6986–6991.
16. **Roehrig JT, et al.** Monoclonal antibody mapping of the envelope glycoprotein of the dengue 2 virus, Jamaica. *Virology* 1998; **246**: 317–328.
17. **Hsieh SC, et al.** A strong endoplasmic reticulum retention signal in the stem-anchor region of envelope glycoprotein of dengue virus type 2 affects the production of virus-like particles. *Virology* 2008; **374**: 338–350.
18. **Hang VTT, et al.** Emergence of the Asian 1 genotype of Dengue virus serotype 2 in Viet Nam: in vivo fitness

- advantage and lineage replacement in South-East Asia. *PLoS Neglected Tropical Diseases* 2010; **4**: e757.
19. **Van Slyke GA, et al.** Point mutations in the West Nile virus (Flaviviridae, Flavivirus) RNA-dependent RNA polymerase alter viral fitness in a host-dependent manner in vitro and in vivo. *Virology* 2012; **427**: 18–24.
  20. **Zhang C, et al.** Structure and age of genetic diversity of dengue virus type 2 in Thailand. *Journal of General Virology* 2006; **87**: 873–883.
  21. **Klungthong C, et al.** The molecular epidemiology of dengue virus serotype 4 in Bangkok, Thailand. *Virology* 2004; **329**: 168–179.
  22. **Twiddy SS, et al.** Phylogenetic evidence for adaptive evolution of dengue viruses in nature. *Journal of General Virology* 2002; **83**: 1679–1689.
  23. **Bordignon J, et al.** Dengue neurovirulence in mice: identification of molecular signatures in the E and NS3 helicase domains. *Journal of Medical Virology* 2007; **79**: 1506–1517.
  24. **Engel AR, et al.** The neurovirulence and neuroinvasiveness of chimeric tick-borne encephalitis/dengue virus can be attenuate by introducing defined mutations into the envelope and NS5 protein genes and the 3' non-coding region of the genome. *Virology* 2010; **405**: 243–252.
  25. **Grant D, et al.** A single amino acid in nonstructural NS4B protein confers virulence to dengue virus in AG129 mice through enhancement of viral RNA synthesis. *Journal of Virology* 2011; **85**: 7775–7787.
  26. **Lee KS, et al.** Dengue virus surveillance in Singapore reveals high viral diversity through multiple introductions and in situ evolution. *Infection Genetics and Evolution* 2012; **12**: 77–88.
  27. **Sall AA, et al.** Yellow fever virus exhibits slower evolutionary dynamics than dengue virus. *Journal of Virology* 2010; **84**: 765–772.
  28. **Gray RR, et al.** The mode and tempo of hepatitis C virus evolution within and among hosts. *BMC Evolutionary Biology* 2011; **11**: 131.
  29. **Hanada K, et al.** A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Molecular Biology and Evolution* 2004; **21**: 1074–1080.