Supplementary Materials for

Antigenic evolution of dengue viruses over 20 years

Leah C. Katzelnick, Ana Coello Escoto, Angkana T. Huang,

Bernardo Garcia-Carreras, Nayeem Chowdhury, Irina Maljkovic Berry, Chris Chavez, Philippe

Buchy, Veasna Duong, Philippe Dussart, Gregory Gromowski, Louis Macareo,

Butsaya Thaisomboonsuk, Stefan Fernandez, Derek Smith, Richard Jarman,

Stephen S. Whitehead*, Henrik Salje*, Derek A.T. Cummings*

Correspondence to: Stephen Whitehead (swhitehead@niaid.nih.gov), Henrik Salje (hs743@cam.ac.uk), and Derek Cummings (datc@ufl.edu).

This PDF file includes:

Materials and Methods
Figs. S1 to S6
Table S1 to S3
References 46 to 50

Other Supplementary Materials for this manuscript include the following:

File S1
Movie S1

Materials and Methods

Ethics statement. This study was approved by the ethical review boards of the Queen Sirikit National Institute of Child Health, Walter Reed Army Institute of Research, and Johns Hopkins Bloomberg School of Public Health (former location of DATC) and University of Florida. The work of NIH and WRAIR was deemed non-human subjects research by their respective IRBs. Because researchers at UF, Cambridge and QSNICH can link these data to identifiers (age and location, though not used in this study), IRB approval was obtained at these institutions. These IRB approvals (Queen Sirikit National Institute of Children's Health (QSNICH 61-062), University of Florida (UF IRB201701844) and the University of Cambridge (HBREC.2019.36)) include a waiver of consent. We followed the National Institutes of Health guidelines for the humane treatment of laboratory animals. The NIAID Animal Care and Use Committee approved these protocols (11DEN33 and 14DEN34, parent protocol NIAID ASP LID 9).

Study population and selection of Thailand DENV1-4 isolates for full genome sequencing. Suspected hospitalized dengue cases presenting to the Queen Sirikit National Institute of Child Health (QSNICH) (*29*) were confirmed by RT-PCR or IgM/IgG serology in collaboration with the Armed Forces Research Institute of Medical Sciences (AFRIMS). RT-PCR positive samples were banked for future characterization and viral isolation was attempted on *Toxorhynchites splendens* mosquito cells. Strains were selected to ensure a minimum of 16 sequences/DENV type/year, with denser sequencing of DENV1 in 2006 (*28*). Further sampling was conducted to fill in years from 1994-2006 and 2007-2014. A small set of sequences from before 1994 were also available. In total 1,944 strains underwent full-genome sequencing at WRAIR as described

previously, using either Illumina MiSeq or Roche 454 sequencing (*28*). The genomes were assembled as previously described (*28*). Final consensus sequences were estimated at 80% confidence. Phylogenetic relationships among all strains of each serotype were evaluated with PhyML (*46*, *47*) using maximum likelihood estimation of a general time-reversible model with gamma-distributed rate variation and invariant sites (GTR+G4+I). Trees are midpoint rooted (**Fig. 1**, **Fig. S1)**.

Selection of strains for antigenic characterization. Our initial analysis plan was to sequence 3,110 clinical isolates and conduct in-depth antigenic characterization on n=374 DENV1-4 strains from Thailand isolated between 1994-2014. In total, we were able to titrate n=348 strains before the dataset was locked in Dec. 2019 for analysis. The virus set was selected to include at least 4 clinical isolates per serotype per year over the 20-year period (**Table S1**) as well as chemically distinct amino acid variants in the E and prM proteins present in 1-5% of sequences of each DENV type. Year of isolation and positions with ≥1% amino acid variation in the E and prM proteins are shown alongside trees in **Fig. S1**. To place the Thailand strains in a regional, temporal, and global context, we antigenically characterized 18 Cambodia DENV1-4 strains from a similar time period (2003-2011) (*48*) that were of the same dominant genotypes circulating in Thailand, 6 strains isolated in Thailand before 1994, and 40 DENV1-4 strains isolated in 18 distinct countries between 1994-2012 (*9*).

Antisera used for antigenic characterization. To define antigenic variation among the panel of 412 DENV strains, we titrated each virus against a panel of 20 sera from African green monkeys (*Chlorocebus sabaeus*, abbreviated as NHP for non-human primate, St. Kitts) selected from a

larger set of 36 sera described previously (*9*). Each NHP in this full set had been inoculated with a genetically distinct DENV strain. Prior to initiating titrations of the Thailand DENV1-4 strains, we selected five NHP sera per serotype, capturing the range of serum positions observed on the original antigenic map and that exhibited moderate to high neutralizing activity (*9*) (**Table S2**). Ten NHPs were infected with strains of the dominant genotypes circulating in Thailand, 8 with strains of minor genotypes isolated in Thailand, and 2 from genotypes not observed among the Thailand strains.

Plaque reduction neutralization test (PRNT). All DENV1-4 strains were titrated at least once against the entire NHP serum panel using the PRNT as described previously (*9*, *35*). Briefly, two-fold serial dilutions of antisera (1:20 to 1:40,960 in diluent consisting of OptiMEM, with 2% fetal bovine serum [FBS], plus gentamicin and albumin) were prepared in 96 well plates for a final volume of 0.04 mL per well. DENV strains were diluted to a concentration of $10^{3.43}$/mL and 0.04 mL were added to serum dilutions yielding a final concentration of 40 plaques per well. The serum-virus mixture was incubated at 37°C for 90 minutes. Culture medium was removed from confluent flat-bottom 96-well plates of C6/36 cells (*Aedes albopictus,* CRL-1660, ATCC, Manassas, VA) and 0.03 mL of the serum-virus mixture was added in technical duplicate to cells. Plates were incubated for 1 hour at 32°C, after which 0.150 mL of warmed overlay medium (1% methylcellulose) was added and plates were incubated for 3 days at 32°C. After infection, plates were washed with phosphate-buffer saline (PBS), fixed with 0.1 mL of 80% methanol (30 minutes), and blocked with antibody dilution buffer (1X PBS and 5% non-fat dry milk) for 10 minutes. Anti-flavivirus mouse monoclonal antibody 2H2 (HB114 D3-2H2-9-21, ATCC) or 4G2 (HB112, ATCC) was added at a concentration of 0.0004mg/mL in 0.1 mL

blocking buffer to each well and plates were shaken and incubated at 37°C for 2 hours. Plates were washed two times (0.18 mL PBS) and 0.1 mL peroxidase-labeled goat anti-mouse IgG (Kirkegaard & Perry Laboratories, KPL, Gaithersburg, MD, diluted 1:3000 in blocking buffer), was added to plates and incubated at 37°C for 1 hour. Plates were washed (2x 0.18 mL PBS) and TrueBlue Peroxidase substrate (KPL, 0.03 mL) was added. Plates were stained until plaques were clearly distinguishable (5-40 minutes) and scanned with the ImmunoSpot Analyzer (CTL, Shaker Heights, OH). Plaques were counted using the Viridot package in R, as described previously (*35*). The $PRNT_{50}$ titer, measured as the reciprocal serum dilution at which 50% inhibition of plaque formation was observed, was estimated using two-parameter logistic regression for all titrations. Strains for which multiple titrations had 95% confidence intervals of ≥8-fold were re-titrated. The majority of strains (86.7%) had average 95% $PRNT_{50}$ confidence intervals of less than 4-fold. As controls, each strain was titrated against a serum with strong multitypic neutralization to ensure titers could be detected as well as a placebo NHP serum and virus only control to test for non-specific neutralization. In each experiment, a serum that was highly DENV3-specific was titrated against a control DENV3 strain, with an average titer of 449 (95% confidence interval: 219 to 864). Experiments were repeated if this control had a titer that differed by 4-fold from the average titer.

Adjustment of titers for experimental effects. During our study, we identified two experimental conditions observed across sampling year and serotype that systematically increased PRNT titers: 1) whether virus was incubated with antiserum for 30 or 90 minutes before addition to cells for infection and 2) whether the virus was re-aliquoted from a stock prepared multiple years earlier or re-amplified within 1-5 months prior to titration. To evaluate

how to adjust the titers for strains incubated for 30 minutes, we selected 14 strains that captured a range of antigenic variation and re-titrated them using the 90-minute incubation period. We then used a linear model to estimate the increase in titers with the extended incubation period (median: 2.19-fold greater, IQR: 2.17 to 2.20), and increased all of the 30-minute titers by the median amount. Titers that had originally been undetectable (<10, encoded in the database as half of the threshold, e.g. 5) were left as undetectable, but the threshold was increased to double the value of the adjusted titer (e.g. <22). We estimated a similar magnitude of adjustment (2.3-fold) using an alternative approach where we made maps with the same viruses titrated using both 30- and 90-minute protocols and adjusted the 30-minute protocol titers until the viruses were as close as possible on the map to the 90-minute protocol viruses. To evaluate how to adjust the titers for the thawed and aliquoting virus (n=132 strains) versus the re-amplified virus (n=216 strains) we amplified and titrated eight strains against the serum panel that had previously only been evaluated using re-aliquoted virus. The median neutralization titer rose by 2.43-fold (2.35 to 2.49). As a separate approach, we built a simple linear model of titers as a function of whether they had been re-amplified, building 100 separate models using only 90% of the data each time. The median neutralization titer for regrown viruses was 2.43-fold higher than the aliquoted viruses (IQR: 2.42 to 2.46). Thus, the titers for all re-aliquoted viruses were increased by 2.43-fold.

Antigenic cartography

We used antigenic cartography, a modified multi-dimensional scaling method, to interpret all neutralizing antibody titers as Euclidean distances between the NHP antisera and DENV1-4 strains (*1*). Briefly, we first identified the neutralizing antibody titer for the virus best neutralized

by each antiserum j, defined as b_j. The antigenic distance, D_{ij}, between any given virus i and antiserum j was defined relative to b_j for that antiserum: $D_{ij} = \log_2(b_j) - \log_2(N_{ij})$, where N_{ij}, is the neutralization titer for virus i against antiserum j. Map coordinates for each antiserum and virus were sought such that the Euclidean distance d_{ij} for each virus and serum best fit the measured antigenic distance D_{ij} in a given number of dimensions. We identified the optimal set of map coordinates by minimizing the error function $E = \sum_{ij} e(D_{ij}, d_{ij})$ from random starting coordinates using a conjugate gradient optimization. One unit of distance on the antigenic map is directly proportional to a two-fold difference in neutralizing antibody titers.

To identify the optimal number of dimensions for our antigenic maps, we made maps each in 2-10 dimensions to identify the minimum error map from 10,000 independent optimizations. The fit of measured titers, measured as the map 'error', improved up to 5 dimensions, then plateaued (**Fig. S2A**). We used cross-validation (100 maps, each with a random 75% of titers, each optimized 1000 times) to identify the coordinates and dimensions (2-5 dimensions) for which map antigenic distances most closely predicted the excluded titers (**Fig. S2B**). All maps had strong correlation between measured and predicted titers with minimal additional improvement in fit above 3D (**Fig. S2C**). The 3D map was better at predicting excluded titers than the 2D map, but there was no difference between 3D, 4D, and 5D. For ease of interpretation, the 2D map is shown to visualize antigenic differences (**Fig. 1G**), while the 3D map was used for estimating all antigenic distances and is shown in a separate file (**File S1**).

Measures of antigenic distance derived from antigenic maps. All statistical analyses were conducted using R 3.6.3 (R Foundation for Statistical Computing) (*49*). We used three separate

methods to measure antigenic distance from the maps, as shown in the cartoons in **Figs. 3 and 4**. The antigenic map center was estimated by taking the average of coordinates in 3D for Thailand strains of each serotype, and then taking the average of all four serotypes. Antigenic distance to the map center for each strain isolated each year was measured relative to this central point and was normalized across serotypes by subtracting the closest strain to the map center within each serotype. As a separate measure of antigenic change between serotypes, we also estimated all pairwise distances between strains of any two given serotypes (e.g. DENV1 and DENV2) isolated within a given year. This process was repeated for all serotype pairs and in all years, enabling estimation of changes in pairwise distance between serotypes over time. Within serotype antigenic change was measured as the pairwise distance between strains isolated in 1994-1995 and strains of that serotype isolated in later years.

Generalized additive models (GAMs) to estimate antigenic dynamics over time. Each of the three measures of antigenic distance were modeled as a function of year of virus isolation using GAMs (mgcv package in R (*50*)), with year included either as a linear term (linear model) or as a smooth (s term) to allow for any non-linear trends in antigenic dynamics (**Table S3**). For each measure of antigenic distance, we built models that included all serotypes simultaneously (with serotype modeled as a categorical variable) and for each serotype separately. We then pulled 100 bootstrap samples for each antigenic distance measure and estimated the antigenic dynamics (all bootstrap samples are shown in the manuscript figures). The printed slope and p-value (linear models) and the estimated degrees of freedom (EDF) and p-value (for GAM models) correspond to the model fit including all data. All models were estimated using restricted maximum

likelihood (REML). The detrended antigenic time series, estimated by subtracting the predicted antigenic distance from the linear models from the GAM models, are shown in **Fig. S4**.

Epidemic data. We acquired the annual serotype-specific case counts of dengue cases treated at QSNICH from 1994-2014 in Bangkok, Thailand (**Fig. S6**) (*29*). For analyses comparing antigenic distance for all serotypes to the epidemic time series, we used the sum count of all dengue cases in a year, combining counts of different serotypes. We took the natural log of each epidemic time series before comparison with antigenic data. Epidemic data were detrended by subtracting the linear trend in the epidemic data from the incidence data.

Time-series correlation analyses. We estimated the Pearson correlation coefficient for each of n=1000 bootstrapped GAM fits of within-serotype and between-serotype antigenic time series using the ccf function with time lag 0 in the stats package in R (*49*). These estimates were computed for all serotypes combined as well as for each serotype separately. Within-serotype dynamics were cross-correlated with each of the two metrics of between-serotype dynamics (distance from the map center and pairwise distance between serotypes). We estimated the correlations directly on the overall antigenic time series (GAMs) as well as on the detrended time series (GAM with linear model subtracted). Results are summarized as the median Pearson correlation coefficient across bootstrap estimates with corresponding p-values computed from the bootstrap samples. Individual serotype antigenic time series were centered for visualization purposes in **Fig. 3** and **Fig. S4**.

We used the same approach to estimate the Pearson correlation coefficient for each of n=1000 bootstrapped GAM fits of the time series for each antigenic distance metric with the corresponding logged epidemic time series. Estimates were computed for all serotypes combined, each serotype separately, for overall dynamics, and the detrended dynamics. For visualization purposes, antigenic and epidemic data were centered and scaled in **Fig. 4** and **Fig. S5**.

Code and data availability. Plaque (immunofocus) reduction neutralization tests were quantified using the Viridot program (code available on Github: https://github.com/leahkatzelnick/Viridot). Antigenic maps were made using ACMACS (https://acmacs-web.antigenic-cartography.org). Phylogenetic trees were made with PhyML (http://www.atgc-montpellier.fr/phyml/). R code and raw antigenic and epidemic data used in the figures in this manuscript are available on Zenodo (*33*). All sequence data is publicly available on GenBank (Accession #s KY586306 to KY586946, MW881266, MW945425 to MW945427, MW945430, MW945433 to MW945437, MW945454 to MW945763, MW945772 to MW946604, MW946607 to MW946985). All data are summarized in the manuscript.

Biological Materials. The viruses and antisera used in this study are covered by standard material transfer agreements between the institutions in this research (WRAIR, NIH, UF). Reasonable requests can be accommodated by emailing the corresponding authors, subject to institutional and regulatory approvals: Derek Cummings (datc@ufl.edu), Henrik Salje (hs743@cam.ac.uk), and Steve Whitehead (swhitehead@niaid.nih.gov).

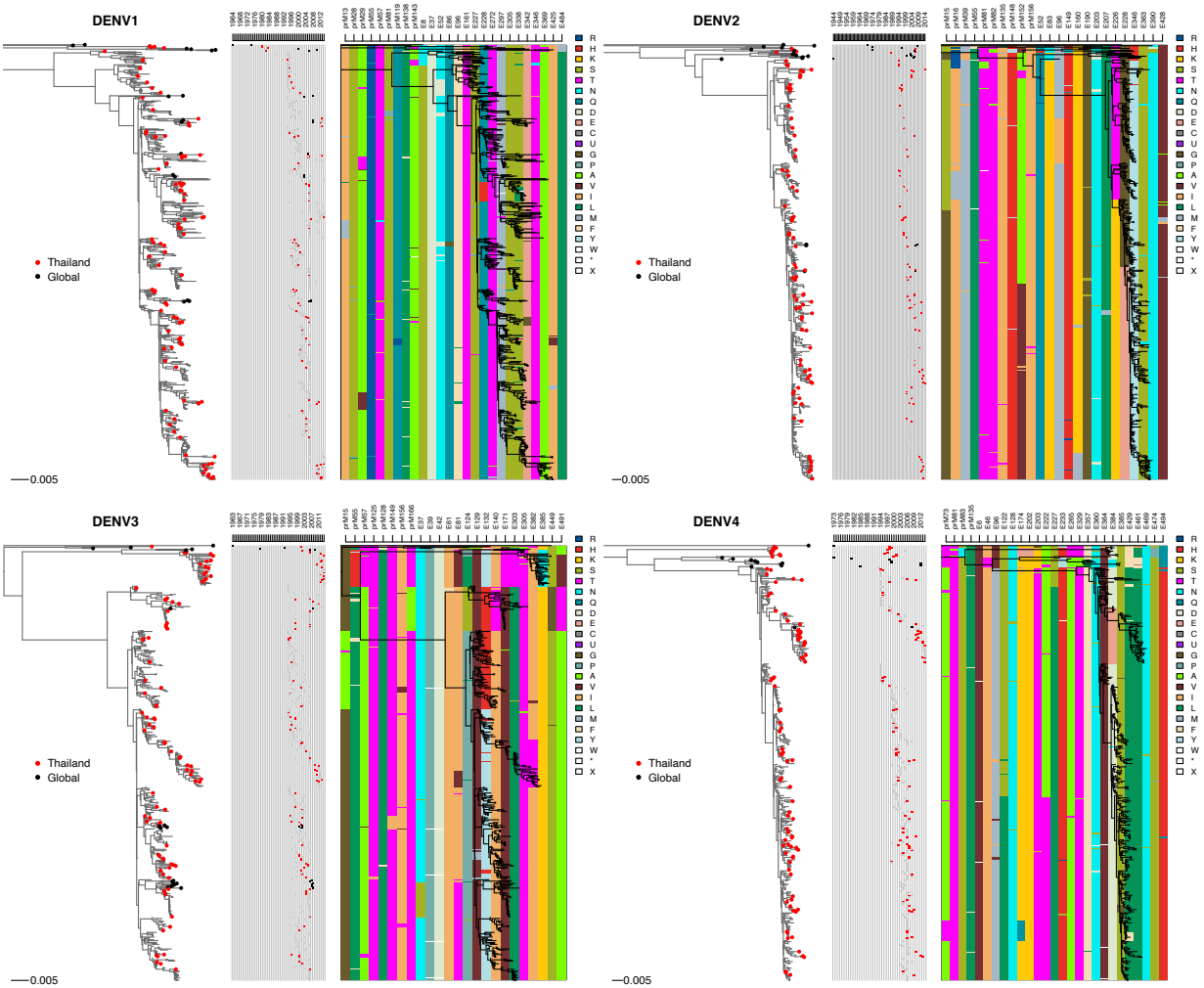
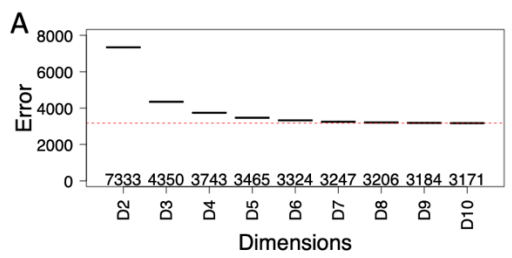


Fig. S1. Phylogenetic relationships among full genome sequences of DENV1-4 isolates from Bangkok, Thailand and global DENV strains. Phylogenetic relationships among DENV1, DENV2, DENV3, and DENV4 sequences were evaluated using maximum likelihood estimation of a general time-reversible model (GTR+G4+I). Trees are cropped to the left to enable visualization. Time series plots show year of virus isolation. Colored circles on tree tips and in the time-series indicate strains selected for antigenic characterization. To the right, positions with <99% conservation of the amino acid in the E and prM proteins are shown with a superimposed phylogenetic tree.



B

	Numeric titers (correlation)					Threshold
	Min.	1st Qu.	**Median**	3rd Qu.	Max.	Prop. >2-fold
D2	0.82	0.85	**0.86**	0.86	0.87	0.06
D3	0.80	0.89	**0.90**	0.91	0.92	0.02
D4	0.75	0.89	**0.90**	0.92	0.93	0.02
D5	0.79	0.89	**0.90**	0.92	0.93	0.02

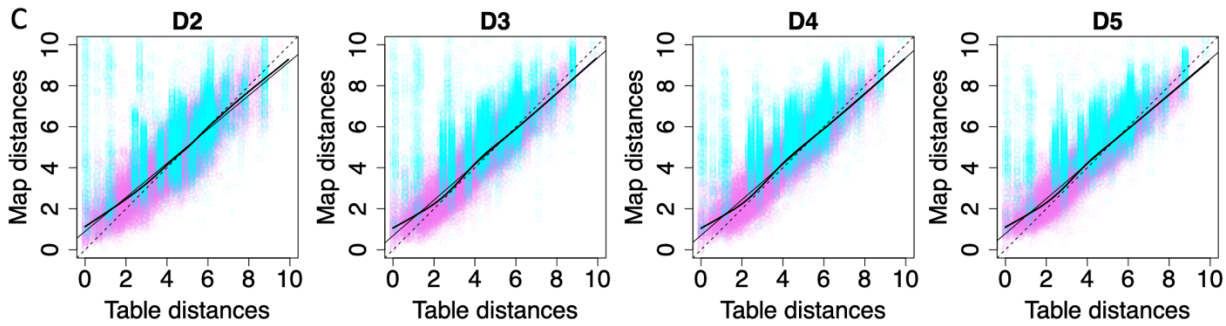


Fig. S2. Antigenic map dimensionality and cross-validation. (A) Elbow plot of the number of dimensions (2-10) versus the distribution of absolute map error (total sum of squared errors) for lowest error 10 maps of 10,000 optimizations made in each dimension. **(B)** Cross-validation of 100 datasets each made with a random 75% of $PRNT_{50}$ titers in 2-5 dimensions (1000 optimizations per map). For each dimension, summary statistics of cross-validation maps are listed: Pearson correlation coefficients for numerically measured titers and proportion with >2-fold error for titers below limit of detection (e.g. <1:20, 'threshold'). **(C)** Plots of the measured versus predicted antigenic distances (pink: titers measured within assay range of detection, cyan: threshold titers), in 2-5 dimensions. Thin black line shows linear regression of titers, thick black line is non-linear locally weighted regression line, dotted line is shown as a guide to indicate perfect correspondence.

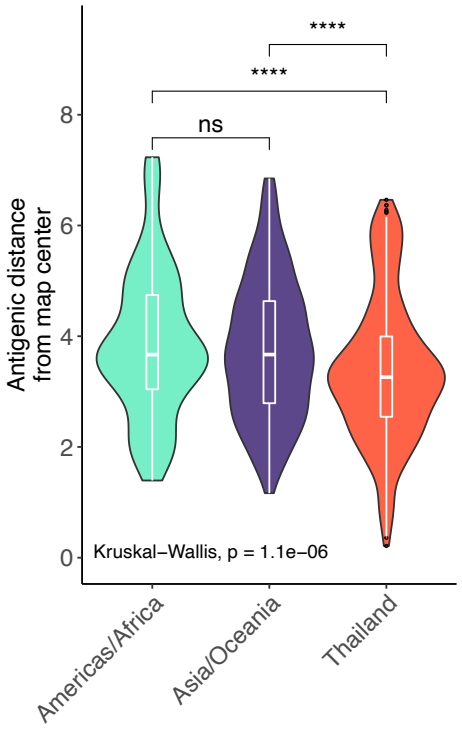


Fig. S3. Violin plot of antigenic distances of each virus from the 3D map center by geographic location, estimated from 10 down-sampled antigenic maps. Each map included all global viruses and 1/10th of the Thailand strains.

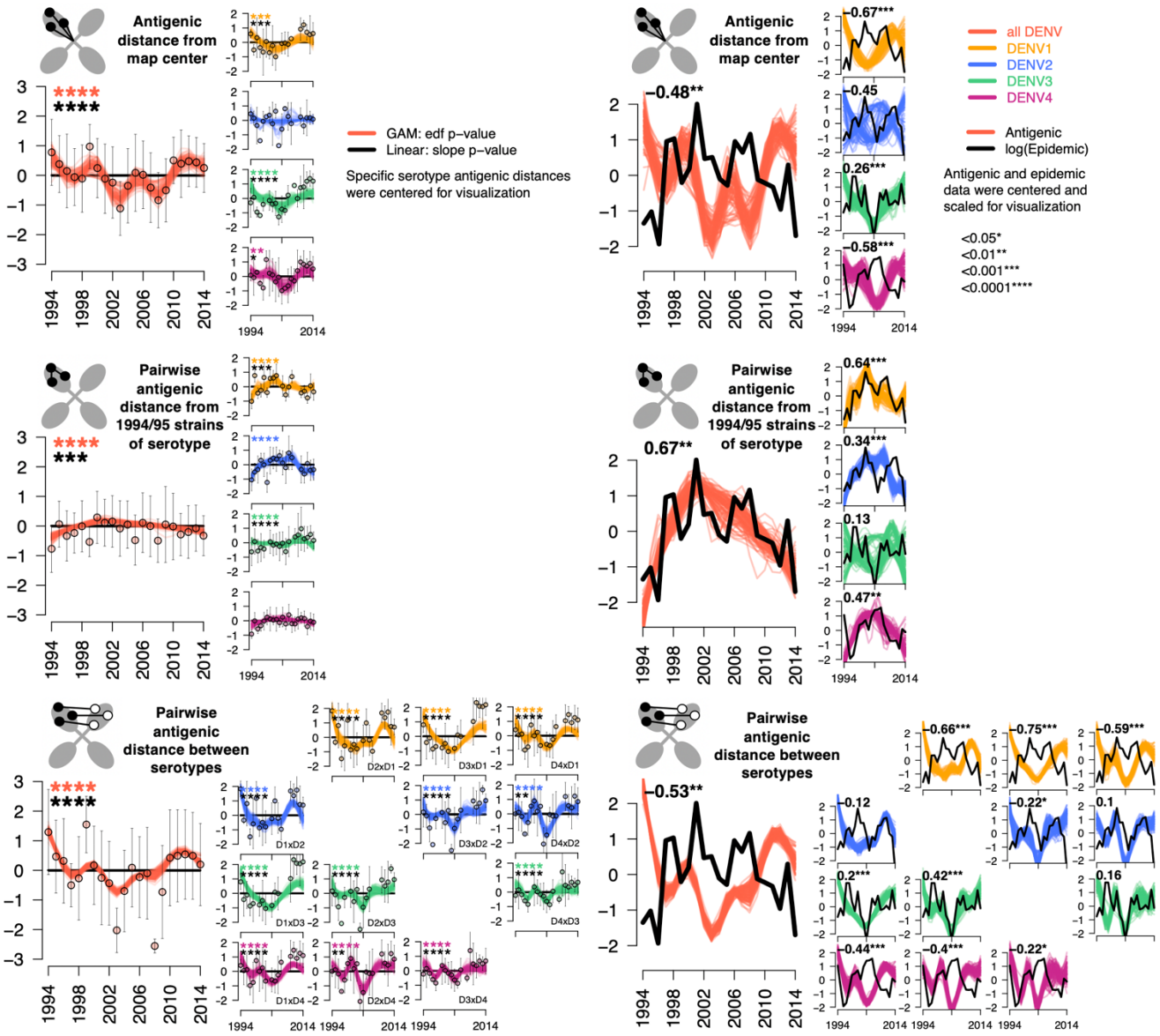


Fig. S4. Detrended antigenic time series and epidemic time series for Thailand DENV1-4 strains. Same as Fig. 3 and 4, except each time series was detrended (linear fit of data subtracted from the GAM fit).

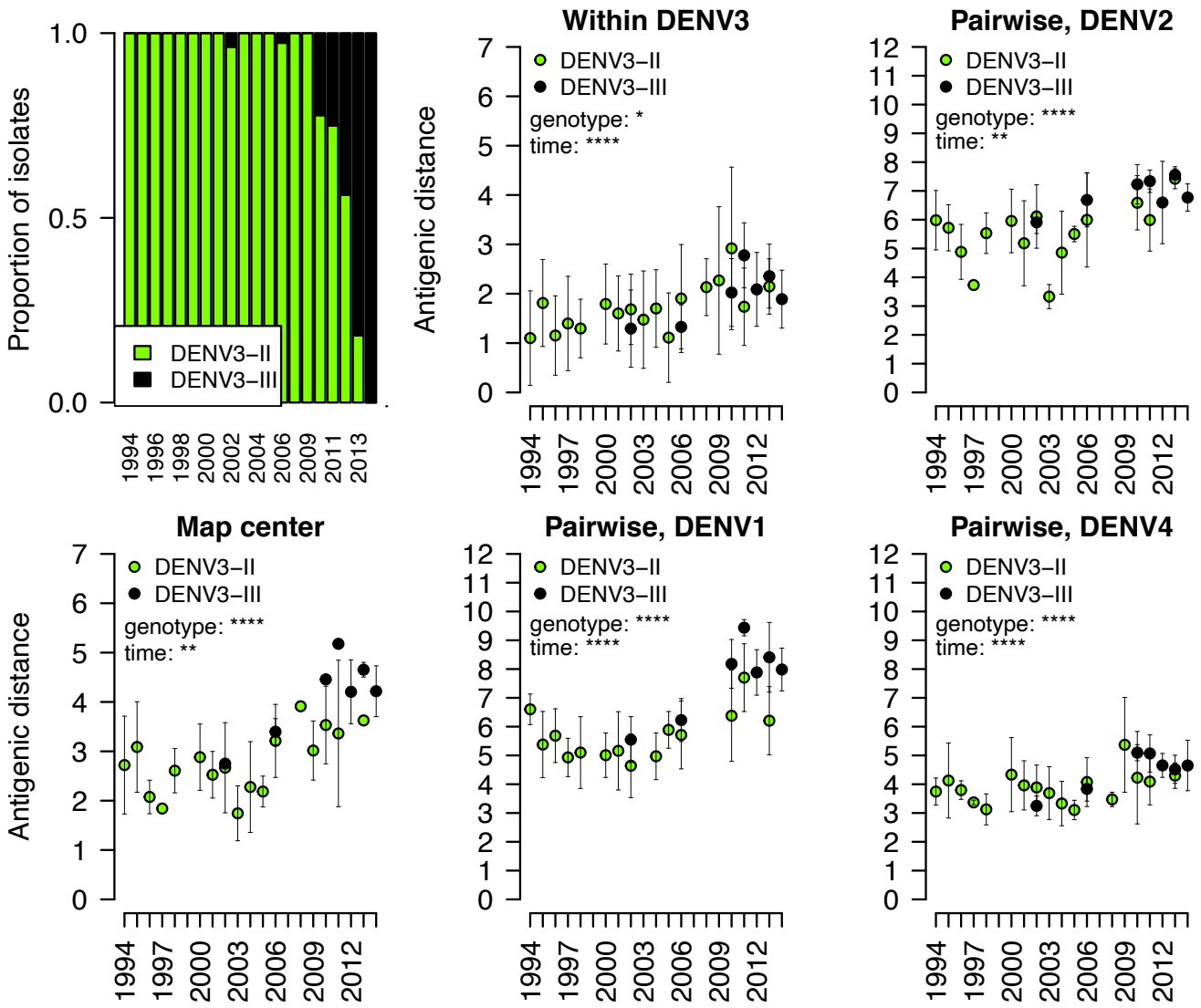


Fig. S5. Antigenic changes for DENV3 genotypes II and III in Thailand over time. Bar chart shows the proportion of isolates of each genotype by year. Antigenic distance (mean and standard deviation) for isolates of each genotype in each year for each metric of antigenic distance (indicated in the figure title).

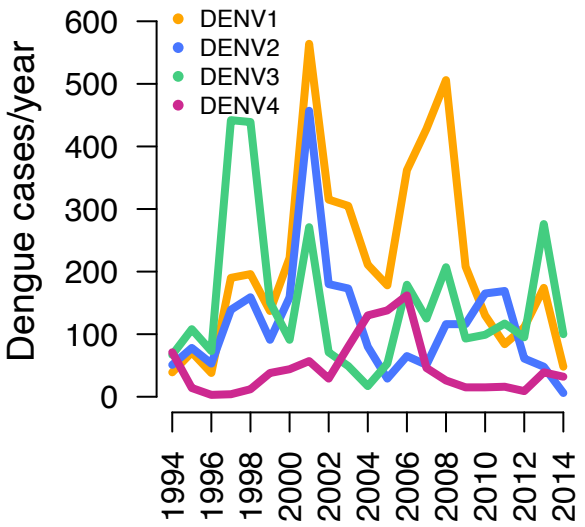


Fig. S6. Annual counts of dengue cases caused by each DENV serotype in Bangkok, Thailand collected at QSNICH from 1994-2014.

Table S1. Antigenically characterized Thailand DENV1-4 strains, by serotype and year of isolation (n=348).

	DENV1	DENV2	DENV3	DENV4
1994	3	4	4	2
1995	4	5	5	5
1996	2	5	4	2
1997	4	1	1	2
1998	5	7	8	5
1999	3	1	0	2
2000	5	7	7	4
2001	4	5	8	7
2002	4	4	8	5
2003	0	1	2	2
2004	9	8	6	6
2005	3	2	2	1
2006	20	7	12	9
2007	1	5	0	10
2008	0	0	1	5
2009	0	0	2	3
2010	4	4	4	6
2011	4	4	4	4
2012	4	4	4	3
2013	4	4	4	4
2014	4	2	4	4

Table S2. Identity of the strains used for producing NHP antisera.

Serotype	Location	Year	GenBank	Genotype*	Identity number
DENV1	Myanmar	2005	KT452791	I	61117
DENV1	Cambodia	2003	GQ868619	I	BID-V1991
DENV1	Peru	2000	AY780643	V	IQT-6152
DENV1	Puerto Rico	2006	EU482591	V	BID-V852
DENV1	Vietnam	2008	FJ461335	I	BID-V1937
DENV2	Cambodia	2009	KT452795	Asian I	D2T0601085_KH09_KSP
DENV2	Malaysia	2008	FJ467493	Sylvatic	DKD-811
DENV2	Nicaragua	2006	EU482684	Asian/American	BID-V571
DENV2	Tonga	1974	AY744147	American	Tonga/74
DENV2	Vietnam	2006	EU482672	Cosmopolitan	32-135
DENV3	Myanmar	2008	KT452792	II	80931
DENV3	Cambodia	2011	KT452799	III	V0907330-AC23
DENV3	Fiji	1992	L11422	I	29472
DENV3	Puerto Rico	2006	EU529698	III	429965
DENV3	Vietnam	2007	FJ432743	II	BID-V1817
DENV4	Cambodia	2010	KF543272	I	U0811386
DENV4	Cambodia	2011	KT452802	I	V0624301-AC33
DENV4	Indonesia	1973	KT452801	II	M30153-AC36
DENV4	Nicaragua	1999	KT452803	II	703
DENV4	Puerto Rico	1999	FJ882599	II	BID-V2446

***** Holmes and Twiddy genotype nomenclature.

Table S3. Effects for linear and non-linear antigenic dynamics models.

Antigenic distance metric	Serotype	Linear model AIC	Slope*	p-value	Generalized additive model AIC	EDF	p-value
Map center, all		952	0.050	3.0E-08	927	4.1	0.0E+00
Between, all		18500	0.075	2.0E-102	17805	8.5	0.0E+00
Within, all		2847	0.015	9.8E-04	2808	4.0	0.0E+00
Map center	DENV1	233	0.061	6.3E-04	227	2.8	4.1E-04
Map center	DENV2	219	0.001	9.7E-01	219	1.0	9.8E-01
Map center	DENV3	220	0.090	2.6E-08	209	2.8	0.0E+00
Map center	DENV4	271	0.046	2.4E-02	261	3.8	2.8E-03
Between, DENV1	DENV2	1501	0.068	4.7E-08	1400	6.7	0.0E+00
Between, DENV1	DENV3	1866	0.138	8.1E-30	1734	5.2	0.0E+00
Between, DENV1	DENV4	1542	0.076	9.3E-10	1479	6.2	0.0E+00
Between, DENV2	DENV1	1501	0.068	4.7E-08	1400	6.7	0.0E+00
Between, DENV2	DENV3	1504	0.072	8.1E-10	1466	7.0	0.0E+00
Between, DENV2	DENV4	1489	0.044	2.6E-03	1418	7.4	0.0E+00
Between, DENV3	DENV1	1866	0.138	8.1E-30	1734	5.2	0.0E+00
Between, DENV3	DENV2	1504	0.072	8.1E-10	1466	7.0	0.0E+00
Between, DENV3	DENV4	1236	0.048	4.9E-08	1195	7.5	0.0E+00
Between, DENV4	DENV1	1542	0.076	9.3E-10	1479	6.2	0.0E+00
Between, DENV4	DENV2	1489	0.044	2.6E-03	1418	7.4	0.0E+00
Between, DENV4	DENV3	1236	0.048	4.9E-08	1195	7.5	0.0E+00
Within	DENV1	604	-0.028	9.5E-04	582	4.5	4.1E-06
Within	DENV2	883	0.014	1.3E-01	827	5.3	0.0E+00
Within	DENV3	941	0.048	8.7E-09	941	1.0	0.0E+00
Within	DENV4	342	0.008	3.1E-01	337	2.5	6.2E-02

* Slope units: change in antigenic units/per year, where one antigenic unit corresponds to a two-fold dilution in the PRNT.

File S1. Webpage showing antigenic map of all 412 strains in 3 dimensions.

Movie S1. Movie showing the 2-dimensional antigenic map, with Thailand DENV1-4 colored sequentially by year of isolation.