The Perfect Storm of 2019: An immunological and phylodynamic analysis of Cambodia's unprecedented dengue outbreak

Cara E. Brook[1], Yimei Li[1], Christina Yek[2], Graham R. Northrup[3], Sreyngim Lay[4,5], Sophana Chea[4,5], Vida Ahyong[6], Daniel M. Parker[7], Somnang Man[4,5], Andrea R. Pacheco[4], Oum Mengheng[4,5], Fabiano Oliveira[2], Liz Fahsbender[8], Rithea Leang[5], Rekol Huy[5], Chea Huch[5], Chanthap Lon[4,5], Cristina M. Tato[6], Joseph L. DeRisi[6], Michael Boots[9], Jennifer A. Bohl[2], Jessica E. Manning[2,4*]

[1] Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA
[2] Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland, USA
[3] Center for Computational Biology, University of California, Berkeley, California, USA
[4] International Center of Excellence in Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Phnom Penh, Cambodia
[5] National Center for Parasitology, Entomology, and Malaria Control, Phnom Penh, Cambodia
[6] Chan Zuckerberg Biohub, San Francisco, California, USA
[7] Department of Population Health and Disease Prevention, University of California, Irvine, California, USA
[8] Chan Zuckerberg Initiative, San Francisco, California, USA
[9] Integrative Biology, University of California, Berkeley, California, USA

***Corresponding author:**

Jessica E. Manning

Email: Jessica.manning@nih.gov

Mailing address: 1 Christopher Howes Place, Phnom Penh, CAMBODIA

Abstract

The year 2019 witnessed the highest number of dengue cases ever reported globally. We analyzed epidemiological, serological, and phylogenomic data to investigate the drivers of the 2019 epidemic in Cambodia. Using epidemiological models fit to a 19-year national dataset, we identified an overall trend of declining annual force of infection (FOI) for dengue virus (DENV) in Cambodia, interspersed with FOI spikes corresponding to epidemic year caseloads that exceeded demographic predictions. We constructed time-resolved phylogenetic trees with 105 DENV genomes sequenced from the 2019 Cambodian epidemic, paired with historical Southeast Asian data, to document the first-recorded introduction of DENV-2 Cosmopolitan genotype into Cambodia. This introduction yielded highly localized transmission and decreased genomic diversity when compared to endemic DENV-1, supporting the hypothesis of epidemic invasion. Introduction of this genetically distinct lineage into a population with limited prior immunity—paired with a spike in FOI—was a key driver of the 2019 Cambodian epidemic.

These studies were registered at clinicaltrials.gov under NCT04034264 and NCT03534245.

Introduction

Dengue virus (DENV) transmission has increased dramatically over the past two decades, culminating in the year 2019 with the highest number of global cases ever reported (>5.2 million) to the World Health Organization (WHO)[1]. Since nearly three-quarters of DENV infections are estimated to be clinically inapparent (thereby unreported), these counts represent a vast underestimate of the true scale of dengue burden on public health systems[2]. DENV is a flavivirus primarily transmitted by the *Aedes aegypti* mosquito, a ubiquitous arthropod vector in tropical and subtropical regions[2]. DENV is comprised of four antigenically distinct serotypes—each of which is further subdivided among four to seven distinct genotypes; infection with one serotype is thought to result in lifelong homotypic immunity to that same serotype (homotypic immunity) but only temporary (up to two-year) protection against different serotypes (heterotypic immunity)[3]. Heterotypic secondary infections are often more clinically severe due to an interaction with pre-existing flavivirus-specific antibodies known as antibody-dependent enhancement (ADE)[4,5]. In regions with multiple circulating endemic flaviviruses, this interplay makes dengue outbreaks difficult to predict at local, subnational, or national levels. Outbreak

forecasting is further challenged by the multifaceted disease ecology characteristic of any arbovirus—for which infections are also impacted by changes in human land use, behavior, or movement, as well as the distribution and abundance of arthropod vectors. Nonetheless, national health systems in DENV-endemic regions rely on forecasting models to inform resource allocation on a year-to-year basis—ranging from vector control to hospital capacity for supportive care, given no available treatments for dengue. The important consequences of forecasting predications emphasize the need for improved understanding of outbreak drivers, particularly in resource-scarce settings.

Southeast Asia (SEA) represents ~70% of global reported dengue cases[1]. Surprisingly, while dengue is largely classified as an urban disease, only 46% of the SEA population is considered urban. For most SEA countries, roughly 35-55% of the total population is concentrated in urban areas co-localizing with the highest dengue burdens; Singapore reports the largest urbanized population percentage in SEA (100%), while Cambodia—with 16 million people—reports the lowest (24%)[6]. Like many SEA countries, Cambodia nonetheless exhibits significant peri-urban sprawl surrounding major metropolitan centers like Phnom Penh[7,8]. As part of the global dengue phenomenon in 2019, Cambodia also suffered its worst dengue epidemic on record, with approximately 40,000 total cases reported across all 26 provinces—a likely drastic underestimation of the true disease burden[9]. Similar to other resource-scarce countries in SEA, Cambodia hosts a national disease surveillance system that is limited to clinicosyndromic diagnoses, and reporting is highly variable outside of the capital city. Clinical diagnoses often correspond to more severe heterotypic secondary DENV infections[10], and in Cambodia, these are reported collectively without the ability to systematically distinguish between serotypes.

The past two decades have witnessed explosive dengue outbreaks alongside extensive, rapid, heterogenous development across much of SEA, with significant peri-urban sprawl expanding beyond central mega-city limits and high-density population centers springing up around industry hubs (often factories) in otherwise rural areas[7,8]. Longitudinal phylodynamic studies of georeferenced DENV sequences from Thailand have demonstrated the importance of microscale transmission—particularly at the household level—in generating DENV diversity in urban Bangkok, predicting that DENV transmission will intensify as peri-urban settings in SEA become better connected[11]. Previously, dengue epidemics have been linked to turnover in the dominant, regional serotype, or—more recently—to replacement of a dominant viral genotype

with a serologically homotypic but nonetheless divergent viral lineage[12,13]. Indeed, recent work links the magnitude of periodic dengue epidemics to antigenic evolution; large epidemics tend to result from the takeover of those DENV lineages most antigenically distinct from previously circulating strains of the same serotype or most closely-related to sequences of a different serotype[13]. Investigations into the dynamical drivers of dengue transmission in rural and peri-urban settings have been limited to date. As climatic changes drive increases in the population at-risk for dengue infection in SEA and elsewhere[14], better-designed active surveillance programs and improved targeting and evaluation of vector control interventions are greatly needed[15].

Here, we explore the dynamics of dengue transmission in peri-urban Cambodia, investigating potential mechanisms underlying the 2019 epidemic, the highest year of total reported dengue cases in the 19-year duration of the Cambodian national surveillance program. We hypothesized that the 2019 epidemic spike in dengue case incidence could be due to either: (a) a heightened force of infection (FOI) resulting from improved conditions for and elevated abundance of the DENV mosquito vector, *Ae. aegypti* or (b) a constant or declining FOI coupled with a higher proportion of susceptible individuals lacking prior immunity to a recently introduced and antigenically novel strain of the virus or (c) some combination of the two. To investigate these hypotheses, we queried a 19-year dataset (2002-2020) of serotype-agnostic dengue case counts from the national surveillance program to estimate the annual FOI across this period, then compared results at the national scale with our own febrile surveillance data in peri-urban Kampong Speu province in 2019 and 2020. Whole genome sequencing of DENV from serum samples during the 2019 outbreak identified the first record of the DENV-2 Cosmopolitan lineage ever documented in Cambodia. Phylodynamic analysis of the spread of the DENV-2 Cosmopolitan lineage in this region suggests that this new genotype likely invaded a population with limited prior immunity, driving the largest documented dengue outbreak in Cambodia concurrent with the 2019 global epidemic.

Results

Epidemic year (2007, 2012, 2019) dengue case counts exceeded demographic predictions in Cambodia.

Serotype-agnostic clinicosyndromic surveillance data for Cambodian dengue from 2002-2020 demonstrated cyclic annual spikes in infection which peaked during the wet season, typically May to October, and demonstrated three major epidemic spikes in 2007, 2012, and 2019 (Fig. 1a, 1b). To query the mechanistic drivers underpinning each annual epidemic, we first fit a simple time series Susceptible-Infected-Recovered (TSIR) model[16–18] to the national data (Fig. 1c, 1d)(Supplementary Text 1). The TSIR model can successfully capture the transmission dynamics of perfectly immunizing childhood infections, such as measles, for which the population of naïve, susceptible hosts can be reconstructed from year to year based on the annual host population birth rate[17,18]. Though the TSIR model has been applied to dengue dynamics previously[19–22], the dynamics of DENV infection—in particular, the role of reinfection with secondary serotypes—are sufficiently complex such that the model is only expected to be predictive over very short time horizons. We first sought to identify the extent to which the TSIR model failed to predict dengue dynamics in epidemic years, highlighting a need for either a higher FOI or a larger susceptible population to explain the data.

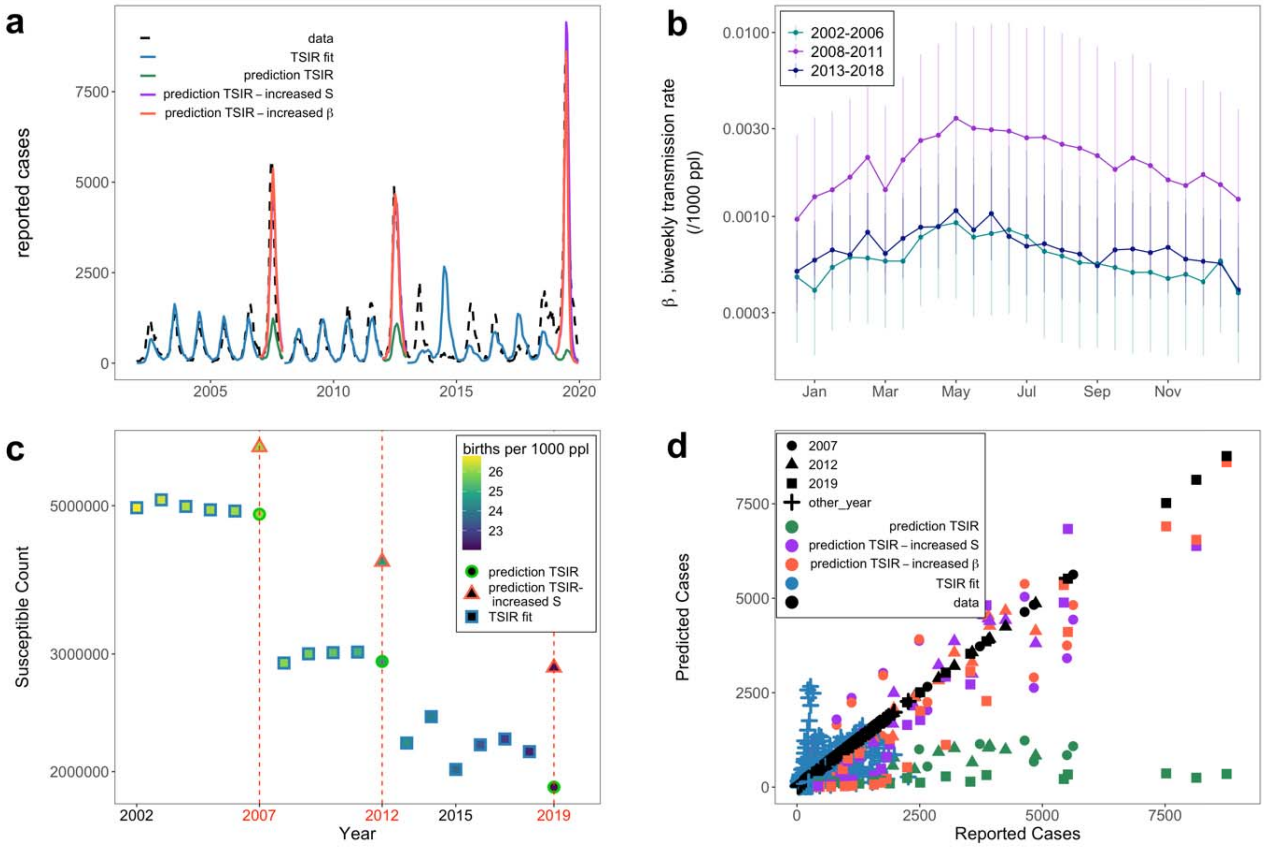

Fig. 1 TSIR modeling insights into epidemic dynamics of DENV in Cambodia. a National-level clinicosyndromic reported DENV cases (*dashed line*) with fitted TSIR output to the three inter-epidemic periods (2002-2006, 2008-2011, 2013-2018; *blue lines*) and TSIR projections for epidemic years (2007, 2012, 2019) under trained parameter estimates (*green lines*) and incorporating a factorial increase in the susceptible population (*purple line*) or, conversely, the FOI (*red line*). Parameter estimates are available for viewing in Table S1. **b** Biweekly transmission rate estimates across the calendar year, as recovered from TSIR fits to the three inter-epidemic periods. Solid line corresponds to mean estimated biweekly transmission, thin segments to 95% confidence intervals. **c** Annual susceptible population reconstructions for Cambodia, colored by World Bank reported birth rates[23], highlighting susceptible population increases needed to recover epidemic year case counts, assuming no factorial increase in FOI in epidemic years. **d** Reported cases vs. predicted cases from TSIR-fitted biweeks and epidemic year biweeks under trained parameter assumptions (*green*) and assumptions of either increased S (*purple*) or increased FOI (*red*).

We fit the TSIR model[16–18] separately to three discrete inter-epidemic intervals within our 19-year national time series, assuming circulation of a single major serotype in the years preceding each outbreak (respectively, 2002-2006, 2008-2011, and 2013-2018), then assessed the model's efficacy in predicting infections in the outbreak year (2007, 2012, 2019). The TSIR model predicted the timing of most annual outbreaks, though it struggled to fit the idiosyncratic dynamics witnessed in the 2013-2018 inter-epidemic period, and as expected, grossly underestimated the magnitude of the epidemic years (Fig. 1a). Transmission rate estimates for the three inter-epidemic training periods all recovered comparable seasonal dynamics, with transmission peaking annually in May, preceding the June-July annual spike in cases (Fig. 1b). We estimated the average biweekly TSIR transmission rate to be largely similar across the three inter-epidemic periods in the dataset and marginally lower for the most recent (2013-2018) inter-epidemic period than for the 2008-2011 period that preceded it, contradicting hypotheses of steadily increasing FOI driving more explosive epidemics with time.

The TSIR model effectively captured epidemic year incidence when the demographic reconstruction of the susceptible population was enhanced by a multiplicative factor of increase in the susceptible population of 1.25, 1.42, and 1.50, respectively, for the epidemic years of 2007, 2012, and 2019—or when the susceptible population was held constant as predicted from birth rates but the time-varying annual FOI was elevated by a corresponding factorial increase of 1.27, 1.45, and 1.54 (Fig. 1a). We hypothesized that a higher FOI might result from an increase in mosquito vector density and/or an increase in mosquito biting rates. While a larger susceptible population could in principle result from a boom in births[24] or a flux in immigration[25], there is no demographic support for either of these assumptions in Cambodia across the time period of study; indeed, on average, births have declined across the past two decades in Cambodia[23]—though concurrently decreasing death rates resulting from improved sanitation, better healthcare, and overall rapid socioeconomic development yielded a steadily increasing total population size

(Fig. S1)[23]. Additionally, invasion by a genetically distinct DENV lineage to which a larger cohort of the resident population lacks prior immunity can manifest as an expansion in the susceptible population[22]. While these dynamics might result from the introduction of a previously unseen serotype in a given region, recent work on DENV antigenic evolution suggests that, if sufficiently genetically distinct from prior circulating strains, a newly introduced viral genotype could yield secondary infections of hosts previously exposed to distantly related homotypic lineages, as well[13]. Thus, we considered the hypothesis that new serotype invasion or genotype substitution into a cohort of "susceptible" hosts lacking prior immunity to the novel lineage could explain the epidemic years. Prior reports from serotype-specific molecular analyses of the 2007 and 2012 dengue epidemics in Cambodia suggest that these outbreaks can be at least partially attributed to invasion and expansion of DENV-3 and DENV-1 lineages, respectively, in the corresponding epidemic years[26–29].

Increasingly explosive DENV epidemics despite declining Force of Infection through time.

TSIR modeling of longitudinal DENV dynamics crystallized our two major hypotheses for the possible factors driving recent epidemics in our system: those of elevated FOI or an increased susceptible population due to novel antigen invasion. Clinicosyndromic dengue surveillance data from the Cambodian national time series reported age at time of infection as metadata, offering an opportunity to quantify changes in the mean annual FOI. We observed that, on average, since 2010, the annual proportion of dengue case incidence in Cambodia has been concentrated in a progressively older patient population with each passing year (Fig. 2a), consistent with reports from Thailand, Singapore, and other SEA countries undergoing rapid socioeconomic development and urbanization[30]. Indeed, the mean age of DENV infection in Cambodia has increased by >.5 years per year since 2018 (Fig. 2b, Table S2).

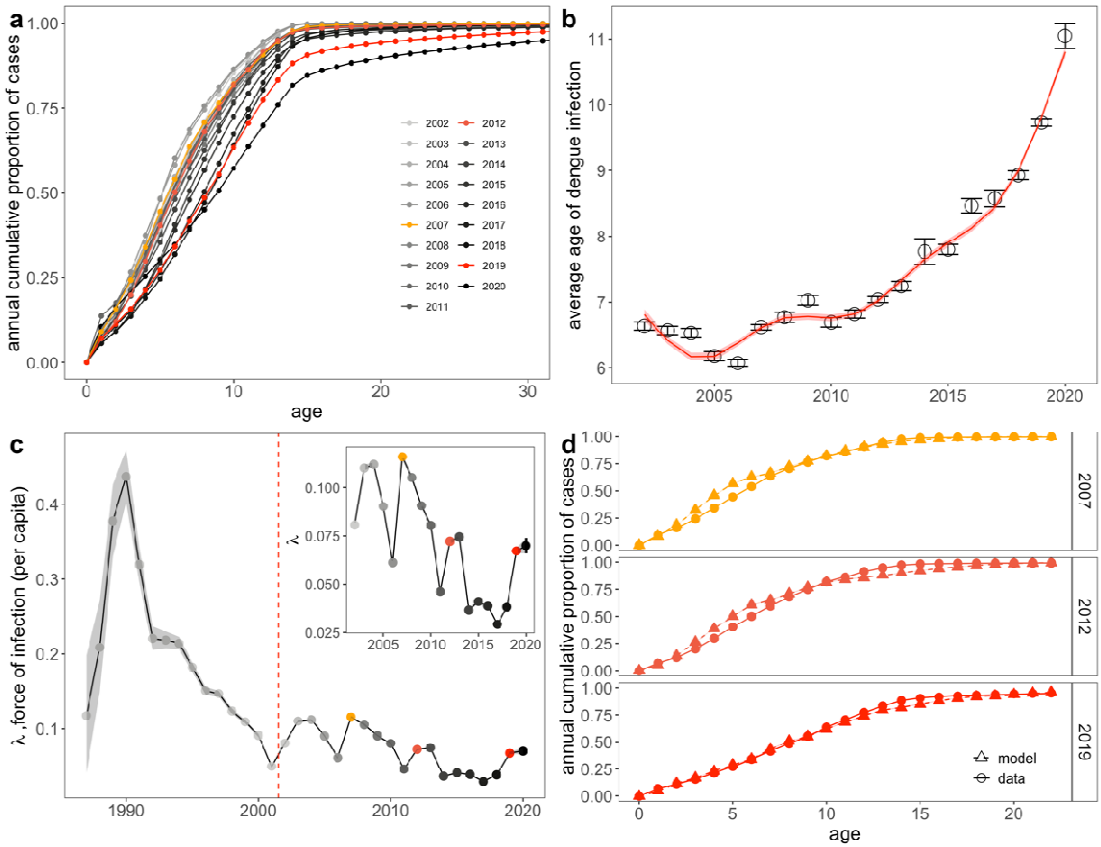

Fig. 2 Trends in age-structured incidence and insights from FOI models. a Annual cumulative proportion of incidence by age of patient from national-level reported dengue cases in Cambodia, binned by year and colored in progressively darker shades of gray while highlighting epidemic years in 2007 (*orange*), 2012 (*red*), and 2019 (*brown*). **b** Mean age of DENV infection by year across our 2002-2020 dataset with 95% confidence intervals by standard error. Data are depicted as open circles and GAM fits to the monthly annual data are shown as red line (mean) with 95% CI in red shading. GAM results and slope derivatives by year can be viewed in Table S2. **c** National annual per capita force of infection (FOI,) for Cambodia DENV from two-serotype, age-structured incidence model fitted to 2002-2020 time series surveillance data. The model estimates back in time 22 years prior to the onset of the data series, corresponding to the cohort of the oldest-age individual in the first (2002) year of the dataset. 95% CIs by standard error are shown in gray shading, based on profile likelihood of each annual FOI while holding all other parameters constant. FOI estimates are depicted from 1987, the first year for which reasonable confidence was established (full time series of FOI estimates back to 1981 can be viewed in Table S3). Vertical dashed line highlights the onset of the data series. Inset panel depicts across the data-driven period. Epidemic years are highlighted by color as in panel a. **d** Model projections (*triangles, dashed lines*) fitted to age-cumulative incidence data (*circles, solid lines*) for each epidemic year. The full time series of model fits can be viewed in Fig. S2.

Building beyond the simple TSIR, we next followed previously developed methods[30,31] to construct an FOI model recapitulating the age-structured cumulative incidence for each year in our dataset and for the 22 years preceding our time series, which corresponded to the onset of exposure for the oldest individuals in our dataset (Supplementary Text 2). Our FOI model assumes that reported dengue cases were derived from secondary infections (which typically exhibit increased clinical severity prompting healthcare-seeking behavior) and that individuals eventually become exposed to at least two serotypes in a lifetime. We used this model to

estimate the annual FOI separately for each year in the dataset under assumptions of two simultaneously circulating DENV serotypes in the system, consistent with prior reporting of dengue incidence in Cambodia. Historically, all four DENV serotypes have been identified in Cambodia, but annual circulation is typically dominated by co-circulation of two serotypes and a near-negligent proportion of cases derived from a third serotype in any given year[26].

On average, the estimated annual FOI for Cambodia DENV decreased steadily across the 19-year time series—and compared to the 22 years preceding it—but short-term, one-to-three-year spikes were observed corresponding to the 2007, 2012, and 2019 epidemic spikes (Fig. 2c). These results supported the hypothesis of elevated FOI—perhaps linked to changes in vector abundance or competency—as one possible driver of the 2019 epidemic in Cambodia. Broadly, our model effectively recaptured the data across the entire time series (Fig. S2), including in epidemic years (Fig. 2c); the steadily decreasing annual FOI resulted in a higher proportion of cases captured in older age classes and a slower accumulation of proportional case load with age over time (Fig. 2d).

Regional active surveillance in the 2019 Cambodia dengue epidemic.

To better understand the drivers of the 2019 dengue outbreak in Cambodia, we enrolled febrile patients and performed DENV RT-PCR and rapid diagnostic tests for confirmation of DENV viremia and DENV-specific IgM/IgG antibodies, respectively. This work was carried out in part with our ongoing active febrile surveillance study in Kampong Speu province, adjacent to the Cambodian capital city of Phnom Penh, between July 2018 and December 2020. In total, we screened 760 samples collected from 697 unique individuals who reported to our clinic with fever during the study period, identifying 106 PCR-positive DENV infections. All patients reported within a self-identified window of five days since fever onset; as such, we assumed that any positive infections with DENV-specific IgG antibodies represented a secondary heterotypic infection; in total, we identified 46 IgG-positive DENV infections at presentation and 57 IgG-negative infections, which likely represented primary infections in previously DENV-naïve individuals (no IgG test was administered at point of care for the remaining 3 positive dengue cases). We next applied the same methods adopted for FOI estimation from age-structured cumulative incidence data at the national level to our 2019-2020 regional data, this time fitting our model to both the age-structured cumulative proportion of secondary infections (our prior

assumption for all nationally reported cases), in addition to the age-structured proportion of primary cases. Resulting model fits reliably recapitulated the 2019 regional data for both secondary and primary infections (Fig. 3a). We then compared our resulting estimates for the 2019 FOI with our prior estimates derived from the national data (Fig. 3b). Estimated FOIs for 2019 were remarkably similar between regional and national data, again showing evidence of elevated FOI in the 2019 epidemic year. Though our regional data did not predate 2019, we were nonetheless able to estimate FOI back in time (following the same methods applied to the national level[30,31]) from the age profile of infections in 2019 and 2020. Amazingly, despite an absence of data in these years, epidemic year peaks corresponding to the 2007 and 2012 epidemic years were perceptible in the regional dataset as well, though they tended to peak in the year following the epidemic—rather than coincident with the year itself (Fig. 3b). These trends may reflect delayed reporting to the national system, regional differences in the extent of each epidemic across all of Cambodia, or uncertainty generated by a limited regional dataset.

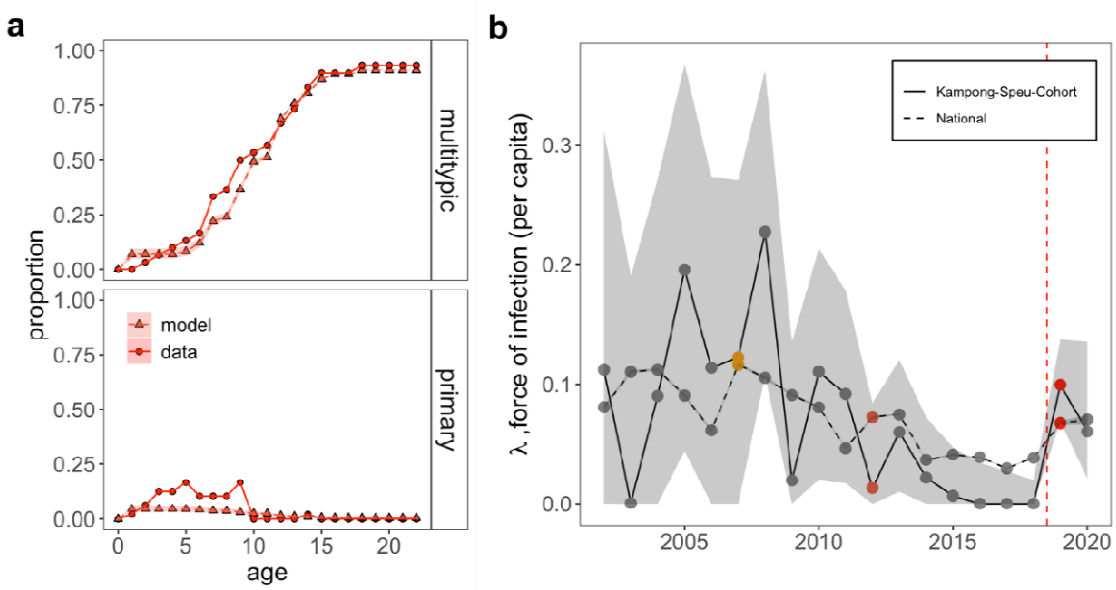

Fig. 3 Fits of FOI model to 2019-2020 age-structured incidence data for DENV from our own active febrile surveillance study in Kampong Speu province, Cambodia. a Top: cumulative proportion of multitypic infections by age of patient, binned by year, with data shown as solid lines and circles and model projections as triangles and dashed lines; 95% confidence intervals by standard error from profile likelihood of the 2019 force of infection ( ) estimate are shown in translucent shading around mean projections. Bottom: proportion of primary infections by age of patient, binned by year, with data and model projections colored as with multitypic infections. **b** Estimated Cambodia dengue per capita from 2002-2020 based on fitting of age-cumulative incidence model to 2019-2020 age-structured incidence data from our Kampong Speu active surveillance study (*solid line*). 95% confidence intervals are shown by translucent shading; *dashed line* indicates national fit for comparison. Epidemic years are highlighted in colored points after Fig. 2. Red vertical line indicates the onset of data series for Kampong Speu cohort model fitting (2019 and 2020 only). All FOI estimates from model fitting to 2019-2020 active febrile surveillance study can be viewed in Table S4.

Following the elevated FOI which both national and regional data identified in epidemic years, we next tested for the possible role of increased vector abundance or activity as a driver of the 2019 dengue epidemic. Because the traditional entomological indicators for *Aedes* spp. (e.g. Bretau Index, larval surveys) are not reliable predictors of dengue risk[32], we instead measured host-vector contact via quantification of antibodies to *Ae. aegypti* salivary extract in all 760 samples at time of presentation to hospital[33]. We found no evidence of association of *Ae. aegypti* saliva antibodies at point of care with DENV infection (Table S5), though, recently, increased levels of *Ae. aegypti* saliva antibodies (prior to infection) were associated with increased dengue risk in a DENV-naïve population[33]. Here, in a population with a mixed history of past exposure to DENV, this effect may be attenuated by the presentation of more severe secondary DENV infections. At time of infection, clear vector-associated risk factors—including habits of mosquito net use, larvicide use, and mosquito coil use—demonstrated no association with DENV infection status either. Though vector-driven selection has a significant role in overall transmission potential[34], epidemic years may be more intensely driven by changes in serotype or genotype that underpin viral evasion of existing host immunity[13]. In our global predictor model, patient age and upper/middle vs. lower socioeconomic status were negatively correlated with DENV positivity, highlighting the preponderance of infections in children and those in poorer income brackets (Table S6). Household living vs. apartment living was also positively correlated with DENV positivity, potentially indicative of the predominance of peri-urban infections associated with more favorable *Ae. aegypti* habitats. On the whole, statistical risk factor analysis demonstrated only limited support for the hypothesis of heightened FOI from improved vector conditions as a driver of the 2019 dengue outbreak, despite age-structured model estimations. These findings thus indirectly supported our alternative hypothesis that the epidemic could be partly attributable to the introduction of a new, antigenically distinct DENV lineage in a susceptible population. This alternate hypothesis inspired our subsequent genomics work.

Identification of a DENV-2 Cosmopolitan genotype introduction in Cambodia.

To interrogate the hypothesis of an antigenically novel DENV lineage introduction in Cambodia as a potential driver of the 2019 epidemic, we sequenced and published 105 whole DENV genomes (51 DENV-1, 51 DENV-2, and 3 DENV-4), representing one-quarter of full genome DENV-1 sequences and nearly half of DENV-2 sequences currently available for

Cambodia—thus emphasizing the considerable challenges faced in undertaking genomic epidemiology in resource-scarce settings[35]. Maximum likelihood phylogenetic analysis of the resulting genomes demonstrated that most DENV-2 sequences recovered in 2019 and 2020 belonged to the DENV-2 Cosmopolitan III lineage, the first record of this genotype reported in Cambodia; all DENV-1 sequences clustered in the Genotype 1 lineage, consistent with previously reported genotype records in Cambodia (Fig. S3). Sequences collected from 2019 were largely split between DENV-1 and DENV-2 serotypes (60.3% vs. 39.4% respectively, N=78), while sequences collected in 2020 were dominated by DENV-2 (20/24, 83.3%). In 2019 and 2020, respectively, 23/31 (74.2%) and 18/20 (90%) of DENV-2 sequences belonged to the Cosmopolitan III lineage, with the remaining sequences clustering in the Asian-1 DENV-2 lineage previously reported in Cambodia.

Recent analyses of the Thailand dengue system link transitions in serotype dominance and clade replacement of genotype sublineages within a single serotype to epidemic magnitude[13]. This work suggests that large epidemics typically result from the invasion, expansion, and evolution of viral lineages that are very distinct from previously resident lineages of the same serotype but (in the majority of cases) more closely related to co-circulating viruses of endemic heterotypic serotypes, reflecting the viral fitness advantage afforded by ADE[36]. In Thailand, the most severe epidemics were linked to clade replacement of a resident viral genotype by a more evolutionarily fit DENV lineage within the same serotype, resulting in a ‘selective sweep’, which subsequently reduced overall viral diversity and, consequently, diminished between-serotype antigenic differences[13]. In the Cambodian epidemic spike of 2019, the Cosmopolitan genotype introduction, followed by serotypic and genotypic homogeneity in 2020, are consistent with the dynamics of clade replacement for an endemic DENV serotype. Clade replacement dynamics have been previously witnessed in conjunction with epidemic outbreaks in Cambodia[12,37] and elsewhere[22,34,38–40].

We next constructed serotype-specific Bayesian timetrees[41] from DENV sequences to assess the divergence time of 2019-2020 lineages from sequences previously reported from Cambodia and from neighboring SEA countries across our 2002-2020 study period (Fig. 4). The majority of DENV-1 lineages (detailed in Fig. S4) diverged relatively recently from 2015 and 2016 Cambodian sequences recovered for this serotype (Fig. 4b), with a time to Most Recent Common Ancestor (tMRCA) of approximately 7.6 years (MRCA at November 2012, 95% HPD:

August 2013 – April 2012). BLAST analysis indicated that this cluster of DENV-1 Cambodia sequences demonstrated the highest identity to sequences recovered from China in 2019, while another, rarer subset of DENV-1 sequences more closely resembled those recovered from Thailand in 2019 and 2020[42]. By contrast, the majority of DENV-2 sequences in the Cosmopolitan III lineage demonstrated high divergence from previously reported sequences (only in the Asian-1 lineage) for Cambodia, with a tMRCA of approximately 83.8 years (MRCA at 1936, 95% HPD: 1932-1941). DENV-2 Cosmopolitan sequences recovered from Cambodia in 2019 and 2020 showed the highest similarity to sequences previously derived from recent DENV-2 Cosmopolitan outbreaks in Singapore, Malaysia, Sri Lanka, and Thailand[22,39,42]. Locally, within our regional data subset for SEA, sequences tightly clustered in both geographic space and time demonstrated high phylogenetic relatedness (Fig. S4), consistent with previous studies emphasizing the importance of microscale transmission dynamics for DENV[11].

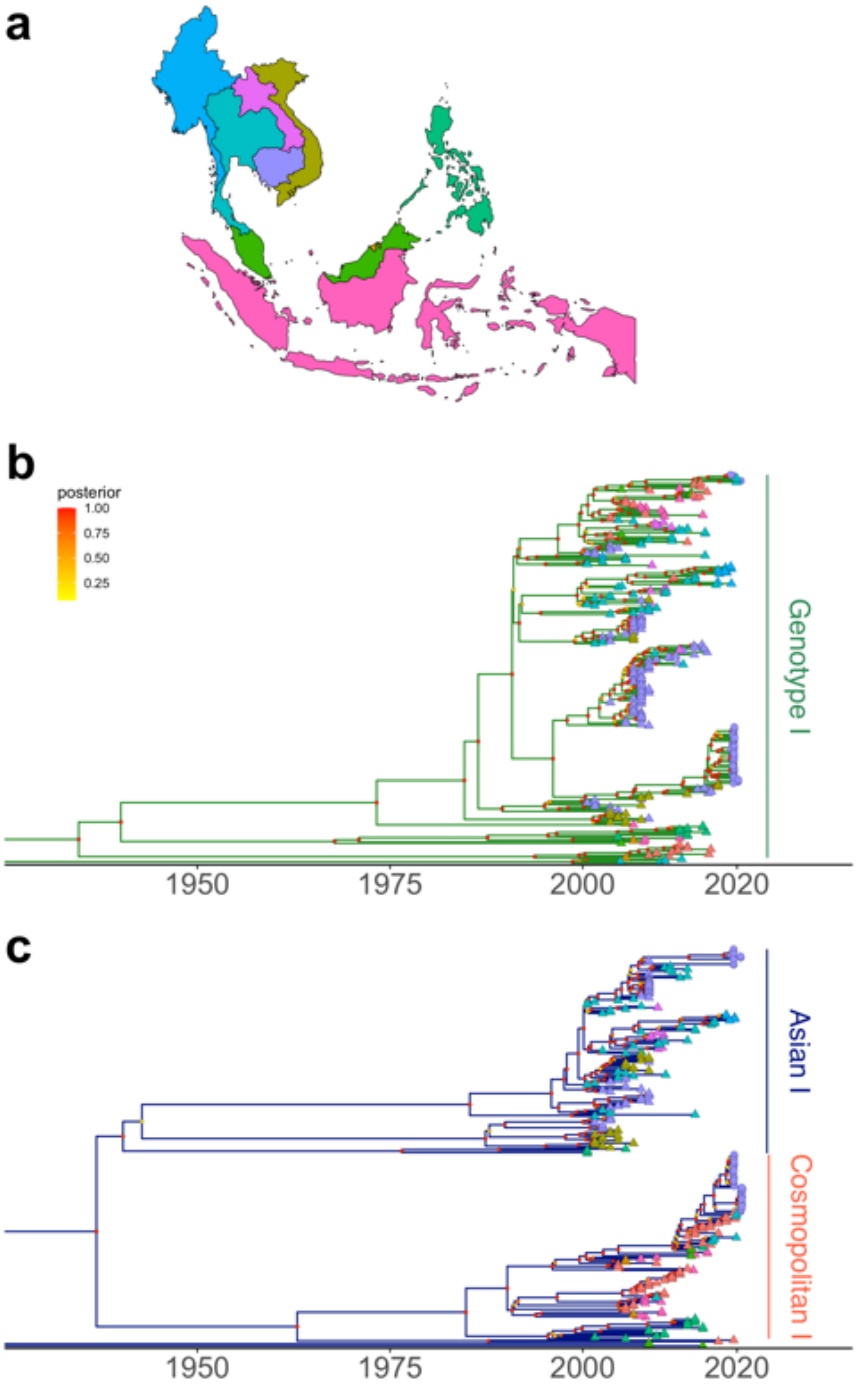

Fig. 4 Bayesian timetrees highlight geospatial structuring in evolutionary relationships for Cambodian dengue. a Map of Southeast Asia with countries colored corresponding to sequences derived from each country, as shown in tip points on phylogenetic timetrees constructed using BEAST 2 for **b** DENV-1 and **c** DENV-2. X-axis highlights divergence times between corresponding sequences. Historical sequences are represented as triangle tips and sequences contributed by active febrile surveillance in this study as circles. Cambodia and corresponding sequences are shaded purple. Clade bars indicate the genotype of corresponding sequences within each serotype: genotype-1 for DENV-1 and Asian-1 and Cosmopolitan for DENV-2. A detailed inset of geographic localities for 2019-2020 Cambodia sequences can be viewed in Fig.S4.

Hyper-localized transmission of DENV-2 Cosmopolitan sequences in Cambodia in the 2019 outbreak.

Previous genomic analysis of geolocated DENV sequences recovered over a 16-year period in Thailand (1994 – 2010) demonstrated the importance of microscale, household-level transmission in driving DENV dynamics, particularly in the highly urbanized setting of Bangkok[11]. In Bangkok, population size was shown to be tightly, positively correlated with the number of distinct transmission chains—a measure of phylogenetic diversity—driving DENV cases for the region (Fig. 5a). In more rural settings in Thailand, this relationship was less pronounced, but the authors predicted that patterns of increasing urbanization would drive DENV diversity to more closely approximate dynamics witnessed in Bangkok. We compared patterns in DENV genomic diversity captured in geolocated sequences from 2019 and 2020 in peri-urban Cambodia with those previously investigated in Thailand (Fig. 5), building transmission chains from sequence pairs which shared a MRCA within the past six months. The number of transmission chains recovered for endemic DENV-1 in our study region closely matched that predicted by population size in urban Bangkok. For DENV-2, which showed less overall diversity than DENV-1—in keeping with a hypothesis of recent invasion—the lower number of transmission chains per population size better approximated that recovered from rural regions in Thailand (Fig. 5a). We further demonstrated a tighter coupling between phylogenetic relatedness and physical, geographic distance for DENV-2 vs. DENV-1 sequences, consistent with epidemic invasion behavior and the tight clustering of a novel antigenic introduction in space and time (Fig. 5b). Finally, prior analysis of Bangkok DENV sequences demonstrated a tighter coupling of phylogenetic relatedness and spatial distance in less mobile populations, particularly young children and women[11]. These patterns held constant for DENV-1 sequences in our peri-urban setting in Cambodia but were reversed for DENV-2 sequences, consistent with a hypothesis of a novel sequence introduction that generated secondary infections and heightened local transmission in older individuals (Fig. 5c-f).

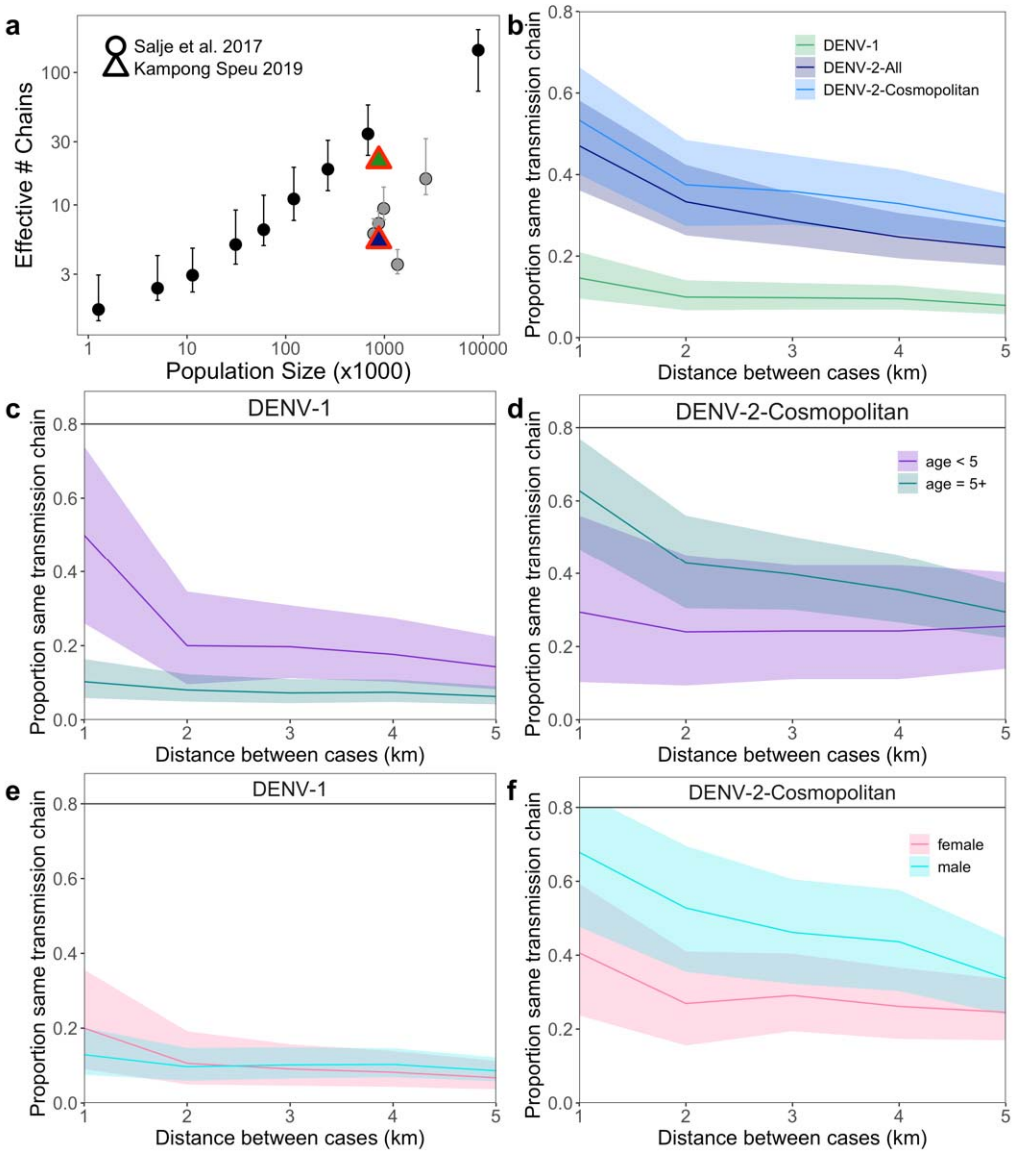

Fig. 5 Geographic proximity structures evolutionary relatedness for Cambodian DENV. a Number of effective transmission chains for circulating DENV estimated across populations of varying densities. Black (urban) and gray (rural) circles with corresponding 95% confidence intervals depict estimates for Thailand from Salje et al. 2017[11], while triangles depict estimates from our Kampong Speu active febrile surveillance study for DENV-1 (*green*) and DENV-2 (*blue*). **b** Proportion of sequence pairs for Kampong Speu DENV-1 sequences (*green*), all DENV-2 (*dark blue*), and DENV-2 Cosmopolitan sequences (*light blue*) which are derived from the same transmission chain (*i.e.* share a MRCA <6 months since the earlier sequence in the pair) across progressively longer Euclidean distances separating the localities from which the sequences were recovered. Proportion of sequence pairs within the same transmission chain for children <5 (*purple*) vs. 5+ (*green*) years of age for **c** DENV-1 and **d** DENV-2 Cosmopolitan sequence pairs across progressively longer distances between cases. Proportion of sequence pairs within the same transmission chain for female (*pink*) vs. male (*blue*) hosts of **e** DENV-1 and **f** DENV-2 Cosmopolitan sequence pairs across progressively longer distances between cases.

Discussion

To date, few studies have been published on the worldwide 2019 dengue surge; further studies of this global phenomenon are needed to forecast and mitigate future dengue epidemics.

Here, we report the novel introduction of the DENV-2 Cosmopolitan genotype, coupled with elevated FOI, as one potential explanation for Cambodia's largest recorded dengue epidemic to date. By contrast, recent analyses of the 2019 dengue epidemic from Brazil suggest that DENV-1 and DENV-2 circulated cryptically in-country for five years before resurging in 2019[43]. Consistent with our findings in Cambodia, the study in Brazil identified no change in mosquito virus transmission potential to explain the 2019 resurgence, but rather attributed the DENV epidemic to transient immune protection, new public health interventions, and behavioral modifications arising in the wake of the Zika virus outbreak[43]. Reports from other 2019 dengue epidemics in Bhutan and Bangladesh point to increased monsoon activity coincident with mass movement of people for school and religious holidays as key drivers of the surge in South Asia[44,45]—while others still highlight a role for reintroduction of previously extinct serotypes, combined with high rainfall and insecticide resistance in these same regions[46]. Global climactic factors may play a role in these coincident epidemics but have not yet been systematically studied in the context of the 2019 global dengue surge.

Our identification of the invasion of the DENV-2 Cosmopolitan genotype in Cambodia, coincident with the 2019 epidemic, echoes studies from the late 1990s in the Americas, which demonstrated that the introduction and subsequent takeover of the DENV-2 Asian-1 genotype were associated with higher frequencies of dengue hemorrhagic fever than had been previously witnessed under the endemic American genotype. Indeed, the invading DENV-2 Asian-1 genotype exhibited a higher viral replication rate within the *Ae. aegypti* mosquito than did previous American genotypes. This replication increase, combined with a shorter extrinsic incubation period, resulted in a 65-fold increase in vectorial capacity that drove dengue outbreaks in the Americas throughout the late 1990s[47]. Previous modeling studies have suggested that novel genotypes may circulate at low prevalence in a host population for many years prior to detection, expansion, and displacement of resident genotypes[48]. Our study emphasizes the extraordinary dearth of publicly available DENV sequence data for Southeast Asia; indeed, DENV is sequenced so infrequently in Cambodia that it is impossible to know whether 2019 truly marked the first year of DENV-2 Cosmopolitan introduction to the country, or simply the year of intensified expansion and consequential epidemic dynamics. More broadly, our work illustrates the importance of the combined forces of pre-existing immunity and viral genetics in driving increasingly severe dengue epidemics. As the global burden of dengue continues to

expand, ongoing serological and genomic surveillance is needed to improve epidemic forecasting in Southeast Asia and around the world.

Methods

Ethics

This study was approved by the National Ethics Committee on Human Research and the National Institutes of Health (NIH) Institutional Research Board. Written informed consent was obtained from the individual participant or the parent or guardian of the child participants enrolled in this study. This study was registered at clinicaltrials.gov as NCT04034264 and NCT03534245.

Enrollment and sample collection

Nursing staff at the Kampong Speu (KPS) Referral Hospital identified, consented, enrolled, and collected demographic data from study participants. Participants, aged 6 months to 65 years of age, presented to the outpatient department with a documented fever of 38°C or greater in the previous 24 hours. DENV infection was first screened using SD Bioline DengueDuo rapid tests for NS1 antigen, pan-dengue IgM and IgG. Sera was collected and processed as described elsewhere for RNA extraction and confirmatory qRT-PCR testing for DENV-1 – 4[33]. To assess direct evidence of vector-host contact in patient sera, quantification of total IgG to *Ae. aegypti* whole salivary gland homogenate was performed via ELISA in 96-well Immulon® plates in duplicate via antibody binding to whole *Ae. aegypti* SGH 2ul/ug concentration diluted in carbonate-bicarbonate buffer, reported as arbitrary ELISA units as described in Manning et al. 2021[33].

Cambodia national data

We obtained a 2002-2020 time series of age-structured dengue cases reported at the national level from clinicosyndromic surveillance efforts administered by the National Dengue Control Program of the Cambodian Ministry of Health. Each of the 26 provinces in Cambodia reports cases on a monthly basis to the national authorities who compile these reports for public health use and reporting to the World Health Organization since inception of the national surveillance system in 2002. Daily cases and the corresponding age and gender of each patient

were reported. We binned cases to every-two-week (biweekly) intervals for application to TSIR modeling and summarized by annual age class within each year for application to FOI models.

TSIR modeling

We first fit a time series Susceptible-Infected-Recovered model[16–18] to the three inter-epidemic periods (2002-2006, 2008-2011, and 2013-2018) of the national time series, in order to highlight the extent to which the case load in each epidemic year deviated from that projected from birth rates alone. The TSIR model leverages an input time series of case counts, births, and total population to estimate the susceptible population and disease transmission rate at intervals corresponding to the generation time of the pathogen in question (here, roughly two weeks[49]). Under TSIR assumptions, the transmission rate is held constant from year to year, though allowed to vary intra-annually to reflect seasonal dynamics corresponding to the pathogen transmission rate. Using the R package tsiR[16], we fit the TSIR model to an input time series of dengue case counts from the national surveillance data and an evenly distributed input vector of birth and total population size per biweek derived from publicly available demographic data for Cambodia published by the World Bank[23]. For each inter-epidemic period, we reconstructed the Susceptible population from the regression of cumulative cases on cumulative births, allowing the transmission rate (β) to vary across 26 biweekly intervals in a single year and the homogeneity parameter (α) to vary with each model fit. Using the model trained on each inter-epidemic period, we then projected cases in the subsequent epidemic years (2007, 2012, 2019) while reconstructing the Susceptible population from births alone.

This first exercise demonstrated the ineffectiveness with which these assumptions were able to recover case counts reported from epidemic years; therefore, we next sought to quantify the increase in the assumed Susceptible count needed to capture reported cases. Previously, Wagner et al. 2020[22] demonstrated analytically that, under short time horizons, the Susceptible population (S') in a two-strain DENV system, in which a new secondary serotype invades a population previously dominated by a single endemic serotype, can be approximated by the equation:

$$S' \approx S + \frac{R_1 \rho_2 N}{I_1} \approx S^* \left[1 + \frac{\rho_2 \gamma \beta}{\mu(\gamma + \mu)}\right]$$

in which S corresponds to individuals naïve to all prior infection, while R_1 and I_1 correspond to individuals recovered from or infected with the endemic strain only, and N and ρ_2 respectively represent the total population size and the rate of exogenous importation of a new serotype. S^* then corresponds to the steady-state value of the Susceptible population under assumptions of a single endemic serotype only, while μ, γ, and β encompass the rates of birth, death, and transmission of exogenous infections of the new serotype. Building on these assumptions, we first calculated the factorial increase in the reconstructed Susceptible population needed to recover case counts for the three epidemic years. We then estimated the factorial increase in FOI that would be needed to recover the same result if the Susceptible population was held constant, thereby highlighting the two major hypotheses under consideration.

Quantifying mean age of infection

We quantified annual variation in the mean age of DENV infection across our 2002-2020 time series using a generalized additive model (GAM) in the 'mgcv' package in R[50]. We fit a GAM with one thinplate smoothing spline for year and one cyclic cubic smoothing spline for month to the response variable of numerical age across our time series. This approach allowed us to distinguish intra- and interannual seasonal patterns in our data and project model predictions by year excluding the seasonal effects of month. To quantify the rate of change in the age of DENV infection in Cambodia across our time series, we then calculated the derivative of our model projection of age at annual intervals in the R package 'gratia'.

Force of Infection estimation

We next used the age-stratified, national-level surveillance data to estimate the annual FOI for DENV in Cambodia across the 19-year time series from 2002-2020. Methods for estimating FOI from age-stratified serological data for single-strain pathogens are well established[51–55], and prior work has adapted these methods to account for the role of pre-existing heterotypic immunity in DENV infection[31] and modified them for application to age-structured incidence data, in lieu of serology[30]. We applied the model developed by Cummings et al. 2009[30] to age-structured incidence recovered from the Cambodian national time series, assuming reported cases to represent secondary infections and all individuals in the dataset to eventually experience exposure to multiple DENV serotypes across their lifetimes. We allowed for a unique

FOI across each year in the time series but assumed a constant FOI across all age cohorts within a single year; as in Cummings et al. 2009[30] and Ferguson et al. 1999[31], this method additionally supported the estimation of FOIs that predated our data time series, albeit with diminished confidence in inference as compared to our data-associated years. Specifically, we estimated one FOI per year for all 22 years predating the onset of the 2002 time series and corresponding to the year of birth cohort for the oldest individual in the first year of our dataset. Subsequently, we applied this same model to age-structured incidence data recovered through our own active surveillance study in Kampong Speu province in 2019 and 2020. For the latter estimation, we computed FOI estimates 37 years back from 2019 (again, corresponding to the oldest infection in our 2019 data), this time simultaneously fitting model projections of both age-stratified primary and secondary cases to those identified in our datasets from IgG antibody tests at point of care.

Risk factor analysis of DENV positivity

We next sought to identify predictive variables associated with DENV positivity from our 2019-2020 active febrile surveillance study in Kampong Speu province. We first tested for a relationship between *Ae. aegypti* saliva antibodies and DENV positivity (based on rapid test at point of care with follow-up PCR confirmation) in 621 fevers from 580 unique patients (a subset of individuals reported with fevers multiple times). We used a generalized linear mixed model in the binomial family to test the association between the response variable of DENV-positive infection (by PCR) and the predictor variable of saliva antibody titer, controlling for individual ID. We subsequently queried the entire dataset of 760 fevers (106 of which were DENV-positive) from 697 unique patients for any association between DENV-positive infection and a suite of potential risk factors. We again used a generalized linear mixed model in the binomial family, controlling for individual ID, to test the association between the response variable of DENV-positive infection and the predictor variables of: age (numerical), sex (categorical: male/female), housing type (categorical: house/apartment/other), socioeconomic status (categorical: upper-middle/lower), bed net use (categorical: rarely/regularly), insecticide use (categorical: yes/no), and mosquito coil use (categorical: yes/no).

Viral sequencing

Metagenomic Next-Generation Sequencing (mNGS) was applied to serum samples collected from all patients reporting with symptoms in our febrile cohort study, whether or not PCR or rapid testing determined these individuals to be DENV-positive. Briefly, pathogen mNGS libraries were prepared from isolated pathogen RNA and converted to cDNA Illumina libraries using the NEBNext Ultra II DNA Library Prep Kit (E7645) according to manufacturer's instructions. Library size and concentration were determined using the 4150 Tapestation system, Agilent, and Qubit 4 Flurometer, Invitrogen (for quantitation only). External RNA Controls 103 Consortium collection, ERCC, ThermoFisher, were used as indicators of potential library preparation errors and for input RNA mass calculation. Samples were sequenced on a NovaSeq (Illumina) instrument and an iSeq100 (Illumina) instrument using 150 nucleotide paired-end sequencing. A water ("no template") control was included in each library preparation.

Raw fastq files were uploaded to the CZID portal, a cloud-based, open-source bioinformatics platform, to identify microbes from metagenomic data (https://czid.org)[56]. Potential pathogens were distinguished from commensal flora and contaminating microbial sequences from the environment by establishing a Z-score metric based on a background distribution derived from 16 non-templated "water-only" control libraries. Data were normalized to reads mapped per million input reads for each microbe at both species and genus levels. Taxa with Z-score less than 1, base pair alignment less than 50 base pairs, NT log(1/e) less than 10 and reads per million (rpM) less than 10 were removed from analysis. Microbial sequences from the samples are located in the National Center for Biotechnology Information (NCBI) Sequence Read Archive.

Construction of consensus genomes

We attempted to construct full-genome DENV sequences from any samples which were confirmed to be DENV-positive by RT-qPCR and which generated at least one reliable contig mapping to any serotype of DENV in the CZID pipeline. To generate these full genome consensus sequences, we ran the ARTIC network's Nextflow consensus genome pipeline[57], mapping each sequence to the closest GenBank accession number hit in the original mNGS run of CZID, using a cutoff of 5 reads per nucleotide site to make a consensus call (sites with <5 reads were called as "N"). Sequences were additionally run through the CZID integrated consensus genome pipeline, again mapping to the closest hit identified in GenBank from the

original mNGS assembly. Resulting consensus sequences from both assembly pipelines were then aligned with reference sequences and visually examined in Geneious Prime. Raw reads from mNGS were then mapped to each full genome contig in turn and examined manually to determine the correct call for each base pair.

Using these methods, we generated full or near-full genome sequences for 51 DENV-1, 51 DENV-2, and 3 DENV-4 samples, representing 105 of the 106 DENV-positive patients identified in our dataset. All contributed genomes were >10000 bps in length and had a maximum of 90 Ns (corresponding to <1% of the DENV genome). Resulting sequences were uploaded to NCBI as individual FASTA files under the following accession numbers: (DENV-1) OK159935-OK159976, OL411495-OL411499, and OL412140, OL412678, and OL412703; (DENV-2) OL414717-OL414765, OL412740, OL420733 and OL435143; (DENV-4): MZ976858-MZ976860.

Phylogenetic analysis

We supplemented our own Cambodia sequences with all other Cambodian sequences for DENV-1 and DENV-2 available in GenBank at the time of analysis, selecting all full or partial genome nucleotide sequences >10000 bp in length up to a collection date of December 31, 2020 (DENV-1: tax id 11053, 160 sequences, including 51 contributed by this study; DENV-2: tax id 11060, 100 sequences, including 51 contributed by this study). We further supplemented these Cambodia sequences with genomes collected from other major Southeast Asian countries (nine), which were Laos, Myanmar, Malaysia, Thailand, Vietnam, Brunei, Indonesia, the Philippines, and Singapore. All countries were represented in both the DENV-1 and DENV-2 datasets. To avoid overrepresenting certain countries outside of Cambodia, we limited sequence selection to a maximum of three randomly selected genomes collected per year from each available year per country, beginning in 2002, the year in which we began our national time series.

After selection, sequences were aligned separately by serotype in the program MAFFT[58], and the best fit nucleotide substation model for each set of data was evaluated in the program ModelTest-NG[59]. For both DENV-1 and DENV-2, a GTR+I+G4 nucleotide was determined to offer the best fit to the data. Using this best fit nucleotide substitution model, we next built a Bayesian phylogenetic tree for each set of DENV genomes in the program BEAST 2[60]. We incorporated the date of sample collection for each sequence (or the midpoint of the year of

collection if the date was not reported), and, after Salje et al. 2017[11], we specified a strict molecular clock at a rate of 7.9×10^{-4} s/s/y[61] and a Coalescent Bayesian skyline prior in our models. We ran Markov Chain Monte Carlo chains in BEAST 2 for 150 million iterations, logging results every 10,000 iterations. After chains completed, we removed the initial 10% of iterations as burn-in and evaluated parameter convergence in Tracer v1.6. We summarized resulting phylogenetic trees in TreeAnnotator and visualized summary trees in the R package ggtree[62].

Finally, we computed a Maximum Likelihood phylogenetic tree to illustrate the phylogenetic placement of our new Cambodia sequences in relation to all previously described genotypes of DENV-1 and DENV-2. For this analysis, we included all Cambodian sequences of DENV-1 and DENV-2 available in NCBI, as well as a broadly representative subset of sequences within all known genotypes of DENV-1 (genotypes I, II, and III) and DENV-2 (genotypes Asian-I, Asian-II, Asian-American, American, Cosmopolitan I, Cosmopolitan II, and Cosmopolitan III). As with Bayesian timetrees, sequences were first aligned in MAFFT[58], and the best fit nucleotide substitution model was computed in ModelTest-NG[59]. As before, we found that a GTR+I+G4 best represented both sequence subsets; using this model, corresponding phylogenetic trees were then constructed in the program RAxML[63] . Following best practices outlined in the RAxML-NG manual, 20 ML inferences were made on each original alignment and bootstrap replicate trees were inferred using Felsenstein's method[64], with the MRE-based bootstopping test applied after every 50 replicates[65]. Resulting phylogenies were then visualized in ggtree[62].

Estimating transmission chains from sequence data

Finally, we followed methods outlined in Salje et al. 2017[11] to calculate the proportion of sequences within each DENV serotype, genotype, or qualitative categorization that could be attributed to the same transmission chain on our Bayesian timetrees, defined as having a most recent common ancestor within the past six months in the same season. We then compared the resulting measures of evolutionary divergence against the Euclidean distance separating the precise GPS coordinates of the collection points of each sequence pair. The reciprocal of the proportion of sequences sharing a transmission chain corresponds to the effective number of transmission chains circulating in a given population. Thus, to compare our estimates of

transmission chain density against those previously reported for Thailand[11], we computed the total effective number of transmission chains observed in our 2019-2020 sequence dataset, separately for DENV-1 and DENV-2, at the World Bank reported population density for Kampong Speu province[23].

Data Availability

All genome sequence data from this study have been submitted to the NCBI Sequence Read Archive under Bioproject ID PRJNA681566. Consensus DENV sequences are also available in GenBank, under the following accession numbers: DENV-1: OK159935-OK159976, OL411495-OL411499, and OL412140, OL412678, and OL412703; DENV-2: OL414717-OL414765, OL412740, OL420733 and OL435143; DENV-4: MZ976858-MZ976860). All bioinformatics code for the initial mNGS assembly is available at https://github.com/chanzuckerberg/idseq-workflows and all wet lab bench protocols are updated at https://docs.google.com/document/d/1RtNQc1D4or_ys7OxCCBjh4SDIdy7JaI4IE7if8EkHgE/edit. Further information on how to use ½ reaction volumes and FastSelect® are available upon request. Code for generation of consensus genomes is available from the Artic Network at https://artic.readthedocs.io/en/latest/, and detailed instructions of all specific analyses undertaken to produce the figures and data presented here are available on our open access github repository at: https://github.com/brooklabteam/cambodia-dengue-national. All other reasonable data requests can be made to authors directly.

References

1. World Health Organization. Dengue and severe dengue. *WHO Fact Sheets* (2022).
2. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496**, 504–507 (2013).
3. Sabin, A. B. Research on dengue during World War II. *American Journal of Tropical Medicine and Hygiene* **1**, 30–50 (1952).
4. Katzelnick, L. C. *et al.* Antibody-dependent enhancement of severe dengue disease in humans. *Science* **932**, 929–932 (2017).
5. Katzelnick, L. C. *et al.* Zika virus infection enhances future risk of severe dengue disease. *Science (1979)* **369**, 1123–1128 (2020).
6. statista. Cambodia: Urbanization from 2010 to 2020. *https://www.statista.com/statistics/455789/urbanization-in-cambodia/*. (2022).

7. Peng, T. N. Southeast Asia's demographic situation, regional variations, and national challenges. in (eds. Singh, D. & Cook, M.) 55–82 (ISEAS Publishing, 2017). doi:doi:10.1355/9789814762878-007.
8. Jones, G. W. *The Population of Southeast Asia*. *Asia Research Institute Working Paper Series* www.nus.ari.edu.sg/pub/wps.htm. (2013).
9. Vong, S. *et al.* Under-recognition and reporting of dengue in Cambodia: A capture-recapture analysis of the National Dengue Surveillance System. *Epidemiology and Infection* **140**, 491–499 (2012).
10. Burke, D. S., Nisalak, A. & Johnson, D. E. A prospective study of dengue infections in Bangkok. *The American Journal of Tropical Medicine and Hygiene* **38**, 172–180 (1988).
11. Salje, H. *et al.* Dengue diversity across spatial and temporal scales: Local structure and the effect of host population size. *Science* **355**, 1302–1306 (2017).
12. O'Connor, O. *et al.* Potential role of vector-mediated natural selection in dengue virus genotype/lineage replacements in two epidemiologically contrasted settings. *Emerging Microbes and Infections* **10**, 1346–1357 (2021).
13. Katzelnick, L. C. *et al.* Antigenic evolution of dengue viruses over 20 years. *Science (1979)* (2021).
14. Messina, J. P. *et al.* The current and future global distribution and population at risk of dengue. *Nature Microbiology* **4**, 1508–1515 (2019).
15. Cattarino, L., Rodriguez-Barraquer, I., Imai, N., Cummings, D. A. T. & Ferguson, N. M. *Mapping global variation in dengue transmission intensity*. *Sci. Transl. Med* vol. 12 https://www.science.org (2020).
16. Becker, A. D. & Grenfell, B. T. TSIR: An R package for time-series susceptible-infected-recovered models of epidemics. *PLoS ONE* **12**, 1–10 (2017).
17. Grenfell, B. T., Bjørnstad, O. N., Finkenstädt, B. F., Monographs, S. E. & May, N. Dynamics of measles epidemics: Scaling noise, determinism, and predictability with the TSIR model. *Ecological Monographs* **72**, 185–202 (2002).
18. Bjornstad, O. N., Finkenstadt, B. F. & Grenfell, B. T. Dynamics of measles epidemics: Estimating scaling of transmission rates using a times series SIR model. *Ecological Monographs* **72**, 169–184 (2002).
19. Oidtman, R. J. *et al.* Inter-annual variation in seasonal dengue epidemics driven by multiple interacting factors in Guangzhou, China. *Nature Communications* **10**, (2019).
20. Rehman, N. A. *et al.* Quantifying the localized relationship between vector containment activities and dengue incidence in a real-world setting: A spatial and time series modelling analysis based on geo-located data from Pakistan. *PLoS Neglected Tropical Diseases* **14**, 1–22 (2020).
21. Kraemer, M. U. G. *et al.* Inferences about spatiotemporal variation in dengue virus transmission are sensitive to assumptions about human mobility: a case study using geolocated tweets from Lahore, Pakistan. *EPJ Data Science* **7**, (2018).
22. Wagner, C. E. *et al.* Climatological, virological and sociological drivers of current and projected dengue fever outbreak dynamics in Sri Lanka. *Journal of the Royal Society Interface* **17**, (2020).
23. World Bank. Cambodia. *World Development Indicators* https://data.worldbank.org/country/KH (2021).

24. Finkenstadt, B. F., Grenfell, B. T. & Grenfell, T. Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **49**, 187–205 (2000).
25. Ferrari, M. J. *et al.* The dynamics of measles in sub-Saharan Africa. *Nature* **451**, 679–84 (2008).
26. Vong, S. *et al.* Dengue incidence in urban and rural Cambodia: Results from population-based active fever surveillance, 2006-2008. *PLoS Neglected Tropical Diseases* **4**, (2010).
27. Lover, A. A. *et al.* Spatial epidemiology and climatic predictors of paediatric dengue infections captured via sentinel site surveillance, Phnom Penh Cambodia 2011-2012. *BMC Public Health* **14**, (2014).
28. Ly, S. *et al.* Asymptomatic dengue virus infections, cambodia, 2012–2013. *Emerging Infectious Diseases* **25**, 1354–1362 (2019).
29. Arima, Y., Chiew, M. & Matsui, T. Epidemiological update on the dengue situation in the Western Pacific Region, 2012. *Western Pac Surveill Response J* **6**, 82–89 (2015).
30. Cummings, D. A. T. *et al.* The impact of the demographic transition on dengue in Thailand: Insights from a statistical analysis and mathematical modeling. *PLoS Medicine* **6**, (2009).
31. Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. Transmission dynamics and epidemiology of dengue: insights from age-stratied sero-prevalence surveys. *Proceedings of the Royal Society B* **354**, 757–768 (1999).
32. Garjito, T. A. *et al.* Stegomyia indices and risk of dengue transmission: A lack of correlation. *Frontiers in Public Health* **8**, (2020).
33. Manning, J. E. *et al.* Development of inapparent dengue associated with increased antibody levels to *Aedes aegypti* salivary proteins: A longitudinal dengue cohort in Cambodia. *The Journal of Infectious Diseases* (2021) doi:10.1093/infdis/jiab541.
34. Lambrechts, L. *et al.* Dengue-1 virus clade replacement in Thailand associated with enhanced mosquito transmission. *Journal of Virology* **86**, 1853–1861 (2012).
35. Bohl, J. A. *et al.* Discovering disease-causing pathogens in resource-scarce Southeast Asia using a global metagenomic pathogen monitoring system. *Proceedings of the National Academy of Sciences* **119**, e2115285119 (2022).
36. Cummings, D. A. T., Schwartz, I. B., Billings, L., Shaw, L. B. & Burke, D. S. Dynamic effects of antibody-dependent enhancement on the fitness of viruses. *Proceedings of the National Academy of Sciences* **102**, 15259–15264 (2005).
37. Duong, V. *et al.* Genetic diversity and lineage dynamic of dengue virus serotype 1 (DENV-1) in Cambodia. *Infection, Genetics and Evolution* **15**, 59–68 (2013).
38. de Jesus, J. G. *et al.* Genomic detection of a virus lineage replacement event of dengue virus serotype 2 in Brazil, 2019. *Memorias do Instituto Oswaldo Cruz* **115**, (2020).
39. Wijewickrama, A. *et al.* Emergence of a Dengue virus serotype 2 causing the largest ever dengue epidemic in Sri Lanka. *bioRxiv* (2018) doi:10.1101/329318.
40. Suzuki, K. *et al.* Genotype replacement of dengue virus type 3 and clade replacement of dengue virus type 2 genotype Cosmopolitan in Dhaka, Bangladesh in 2017. *Infection, Genetics and Evolution* **75**, (2019).
41. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969–1973 (2012).

42. Poltep, K. *et al.* Genetic diversity of dengue virus in clinical specimens from Bangkok, Thailand, during 2018–2020: Co-circulation of all four serotypes with multiple genotypes and/or clades. *Tropical Medicine and Infectious Disease* **6**, (2021).
43. Brito, A. F. *et al.* Lying in wait: the resurgence of dengue virus after the Zika epidemic in Brazil. *Nature Communications* **12**, 1–13 (2021).
44. Tsheten, T. *et al.* Epidemiological analysis of the 2019 Dengue epidemic in Bhutan. *International Journal of Environmental Research and Public Health* **18**, 1–13 (2021).
45. Hossain, M. S. *et al.* Dengue in a crowded megacity: Lessons learnt from 2019 outbreak in Dhaka, Bangladesh. *PLoS Neglected Tropical Diseases* **14**, 1–5 (2020).
46. Ahsan, A., Haider, N., Kock, R. & Benfield, C. Possible drivers of the 2019 Dengue outbreak in Bangladesh: the need for a robust community-level surveillance system. *Journal of Medical Entomology* **58**, 37–39 (2021).
47. Rico-Hesse, R. *et al.* Origins of Dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology* **230**, 244–251 (1997).
48. Lourenço, J. & Recker, M. Viral and epidemiological determinants of the invasion dynamics of Novel Dengue Genotypes. *PLoS Neglected Tropical Diseases* **4**, (2010).
49. Salje, H. *et al.* Reconstructing unseen transmission events to infer dengue dynamics from viral sequences. *Nature Communications* 1–10 (2021) doi:10.1038/s41467-021-21888-9.
50. Wood, S. N. mgcv: GAMs and Generalized Ridge Regression for R. *R News* **1/2**, 20–24 (2001).
51. Grenfell, B. T. & Anderson, R. M. The estimation of age-related rates of infection from case notifications and serological data. *The Journal of Hygiene* **95**, 419–36 (1985).
52. Pomeroy, L. W. *et al.* Serotype-specific transmission and waning immunity of endemic foot-and-mouth disease virus in Cameroon. *PLoS ONE* **10**, 1–16 (2015).
53. Long, G. H. *et al.* Identifying the age cohort responsible for transmission in a natural outbreak of Bordetella bronchiseptica. *PLoS Pathogens* **6**, (2010).
54. Heisey, D. M., Joly, D. O. & Messier, F. The fitting of general force-of-infection models to wildlife disease prevalence data. *Ecology* **87**, 2356–2365 (2006).
55. Muench, Hugo. *Catalytic models in epidemiology.* (Harvard University Press, 1959).
56. Kalantar, K. L. *et al.* IDseq-An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *Gigascience* **9**, 1–14 (2021).
57. ARTIC-network. The ARTIC field bioinformatics pipeline. https://github.com/artic-network/fieldbioinformatics (2021).
58. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**, 1160–1166 (2018).
59. Darriba, Di. *et al.* ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution* **37**, 291–294 (2020).
60. Bouckaert, R. *et al.* BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, (2014).
61. Sall, A. A. *et al.* Yellow fever virus exhibits slower evolutionary dynamics than dengue virus. *Journal of Virology* **84**, 765–772 (2010).
62. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. Ggtree: an R Package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, 28–36 (2017).

63. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
64. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution (N Y)* **39**, 783–791 (1985).
65. Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E. & Stamatakis, A. How many bootstrap replicates are necessary? *Journal of Computational Biology* **17**, 337–354 (2010).

Acknowledgements

This research is supported by the Division of Intramural Research at the National Institute of Allergy and Infectious Diseases at the National Institutes of Health and the Bill and Melinda Gates Foundation [grant numbers OPP1211806, OPP1211841]. We thank patients and families of Kampong Speu District Referral Hospital who participated in this study. We thank members of the National Dengue Control Program not listed in the author byline and the Provincial Health Department of Kampong Speu province in Cambodia. We thank all the other employees at the Chan Zuckerberg Biohub and Chan Zuckerberg Initiative not listed in the author byline. We thank Brian Moyer and the NIAID Office of Cyberinfrastructure and Computational Biology (OCICB) for their assistance in improving the cyberinfrastructure of our Cambodian field sites.

Author contributions

CEB and JEM conceived the study. JEM, JAB, HR, RL, CH, and SC carried out the febrile surveillance study and collected samples for mNGS in 2019. VA, CMT, JLD, JAB, SC, SL, CY, OM, SM, and JEM carried out RT-qPCR to identify dengue-positive samples, as well as library preparation and mNGS for raw sequencing. Other ancillary demographic data collection and analyses were performed by RL, KS, CY, FO, and ARP. DP, CL, and JEM set up geospatial system for patients. CEB and JAB generated consensus dengue genomes for upload to NCBI. CEB carried out TSIR and FOI modeling, as well as all phylogenetic analyses, with help from YL, GN, and MB. CEB and JEM wrote the first draft of the manuscript. All authors approved and contributed to subsequent drafts of the manuscript and agree with the results presented.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence and request for materials should be addressed to JEM.