

The global distribution and burden of dengue

Samir Bhatt¹, Peter W. Gething¹, Oliver J. Brady^{1,2}, Jane P. Messina¹, Andrew W. Farlow¹, Catherine L. Moyes¹, John M. Drake^{1,3}, John S. Brownstein⁴, Anne G. Hoen⁵, Osman Sankoh^{6,7,8}, Monica F. Myers¹, Dylan B. George⁹, Thomas Jaenisch¹⁰, G. R. William Wint^{1,11}, Cameron P. Simmons^{12,13}, Thomas W. Scott^{9,14}, Jeremy J. Farrar^{12,13,15} & Simon I. Hay^{1,9}

Dengue is a systemic viral infection transmitted between humans by *Aedes* mosquitoes¹. For some patients, dengue is a life-threatening illness². There are currently no licensed vaccines or specific therapeutics, and substantial vector control efforts have not stopped its rapid emergence and global spread³. The contemporary worldwide distribution of the risk of dengue virus infection⁴ and its public health burden are poorly known^{2,5}. Here we undertake an exhaustive assembly of known records of dengue occurrence worldwide, and use a formal modelling framework to map the global distribution of dengue risk. We then pair the resulting risk map with detailed longitudinal information from dengue cohort studies and population surfaces to infer the public health burden of dengue in 2010. We predict dengue to be ubiquitous throughout the tropics, with local spatial variations in risk influenced strongly by rainfall, temperature and the degree of urbanization. Using cartographic approaches, we estimate there to be 390 million (95% credible interval 284–528) dengue infections per year, of which 96 million (67–136) manifest apparently (any level of clinical or sub-clinical severity). This infection total is more than three times the dengue burden estimate of the World Health Organization². Stratification of our estimates by country allows comparison with national dengue reporting, after taking into account the probability of an apparent infection being formally reported. The most notable differences are discussed. These new risk maps and infection estimates provide novel insights into the global, regional and national public health burden imposed by dengue. We anticipate that they will provide a starting point for a wider discussion about the global impact of this disease and will help to guide improvements in disease control strategies using vaccine, drug and vector control methods, and in their economic evaluation.

Dengue is an acute systemic viral disease that has established itself globally in both endemic and epidemic transmission cycles. Dengue virus infection in humans is often inapparent^{1,6} but can lead to a wide range of clinical manifestations, from mild fever to potentially fatal dengue shock syndrome². The lifelong immunity developed after infection with one of the four virus types is type-specific¹, and progression to more serious disease is frequently, but not exclusively, associated with secondary infection by heterologous types^{2,5}. No effective antiviral agents yet exist to treat dengue infection and treatment therefore remains supportive². Furthermore, no licensed vaccine against dengue infection is available, and the most advanced dengue vaccine candidate did not meet expectations in a recent large trial^{7,8}. Current efforts to curb dengue transmission focus on the vector, using combinations of chemical and biological targeting of *Aedes* mosquitoes and management of breeding sites². These control efforts have failed to stem the increasing incidence of dengue fever epidemics and expansion of the

geographical range of endemic transmission⁹. Although the historical expansion of this disease is well documented, the potentially large burden of ill-health attributable to dengue across much of the tropical and subtropical world remains poorly enumerated.

Knowledge of the geographical distribution and burden of dengue is essential for understanding its contribution to global morbidity and mortality burdens, in determining how to allocate optimally the limited resources available for dengue control, and in evaluating the impact of such activities internationally. Additionally, estimates of both apparent and inapparent infection distributions form a key requirement for assessing clinical surveillance and for scoping reliably future vaccine demand and delivery strategies. Previous maps of dengue risk have used various approaches combining historical occurrence records and expert opinion to demarcate areas at endemic risk^{10–12}. More sophisticated risk-mapping techniques have also been implemented^{13,14}, but the empirical evidence base has since been improved, alongside advances in disease modelling approaches. Furthermore, no studies have used a continuous global risk map as the foundation for dengue burden estimation.

The first global estimates of total dengue virus infections were based on an assumed constant annual infection rate among a crude approximation of the population at risk (10% in 1 billion (ref. 5) or 4% in 2 billion (ref. 15)), yielding figures of 80–100 million infections per year worldwide in 1988 (refs 5, 15). As more information was collated on the ratio of dengue haemorrhagic fever to dengue fever cases, and the ratio of deaths to dengue haemorrhagic fever cases, the global figure was revised to 50–100 million infections^{16,17}, although larger estimates of 100–200 million have also been made¹⁰ (Fig. 1). These estimates were intended solely as approximations but, in the absence of better evidence, the resulting figure of 50–100 million infections per year is widely cited and currently used by the World Health Organization (WHO). As the methods used were informal, these estimates were presented without confidence intervals, and no attempt was made to assess geographical or temporal variation in incidence or the inapparent infection reservoir.

Here we present the outcome of a new project to derive an evidence-based map of dengue risk and estimates of apparent and inapparent infections worldwide on the basis of the global population in 2010. We compiled a database of 8,309 geo-located records of dengue occurrence from a systematic search, resulting from 2,838 published literature sources as well as newer online resources¹⁸ (see Supplementary Information, section A; the full bibliography⁴ and occurrence data are available from authors on request). Using these occurrence records we: chose a set of gridded environmental and socioeconomic covariates known, or proposed, to affect dengue transmission (see Supplementary Information, section B); incorporated recent work assessing the strength of evidence on national and subnational-level dengue

¹Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. ²Oxitec Limited, Milton Park, Abingdon OX14 4RX, UK. ³Odum School of Ecology, University of Georgia, Athens, Georgia 30602, USA. ⁴Department of Pediatrics, Harvard Medical School and Children's Hospital Informatics Program, Boston Children's Hospital, Boston, Massachusetts 02115, USA. ⁵Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire 03755, USA. ⁶INDEPTH Network Secretariat, East Legon, PO Box KD 213, Accra, Ghana. ⁷School of Public Health, University of the Witwatersrand, Braamfontein 2000, Johannesburg, South Africa. ⁸Institute of Public Health, University of Heidelberg, 69120 Heidelberg, Germany. ⁹Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹⁰Section Clinical Tropical Medicine, Department of Infectious Diseases, Heidelberg University Hospital, INF 324, D 69120 Heidelberg, Germany. ¹¹Environmental Research Group Oxford (ERGO), Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. ¹²Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam. ¹³Centre for Tropical Medicine, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK. ¹⁴Department of Entomology, University of California Davis, Davis, California 95616, USA. ¹⁵Department of Medicine, National University of Singapore, 119228 Singapore.

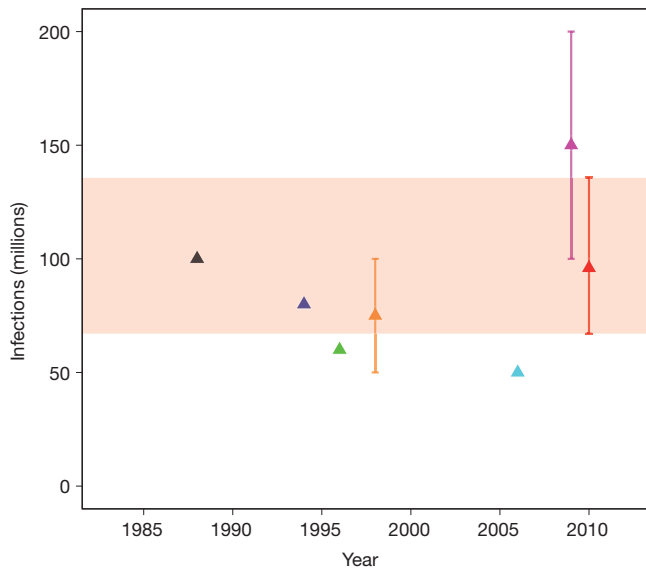


Figure 1 | Global estimates of total dengue infections. Comparison of previous estimates of total global dengue infections in individuals of all ages, 1985–2010. Black triangle, ref. 5; dark blue triangle, ref. 15; green triangle, ref. 17; orange triangle, ref. 16; light blue triangle, ref. 30; pink triangle, ref. 10; red triangle, apparent infections from this study. Estimates are aligned to the year of estimate and, if not stated, aligned to the publication date. Red shading marks the credible interval of our current estimate, for comparison. Error bars from ref. 10 and ref. 16 replicated the confidence intervals provided in these publications.

present/absent status⁴ (Fig. 2a); and built a boosted regression tree (BRT) statistical model of dengue risk that addressed the limitations of previous risk maps (see Supplementary Information, section C) to define the probability of occurrence of dengue infection (dengue risk) within each 5 km × 5 km pixel globally (Fig. 2b). The model was run 336 times to reflect parameter uncertainty and an ensemble mean map was created (see Supplementary Information, section C). We then combined this ensemble map with detailed longitudinal information on dengue infection incidence from cohort studies and built a non-parametric Bayesian hierarchical model to describe the relationship between dengue risk and incidence (see Supplementary Information, section D). Finally, we used the estimated relationship to predict the number of apparent and inapparent dengue infections in 2010 (see Supplementary Information, section E). Our definition of an apparent infection is consistent with that used by the cohort studies: an infection with sufficient severity to modify a person's regular schedule, such as attending school. This definition encompasses any level of severity of the disease, including both clinical and subclinical manifestations.

We predict that dengue transmission is ubiquitous throughout the tropics, with the highest risk zones in the Americas and Asia (Fig. 2b). Validation statistics indicated high predictive performance of the BRT ensemble mean map with area under the receiver operating characteristic (AUC) of 0.81 (± 0.02 s.d., $n = 336$) (see Supplementary Information, section C). Predicted risk in Africa, although more unevenly distributed than in other tropical endemic regions, is much more widespread than suggested previously. Africa has the poorest record of occurrence data and, as such, increased information from this continent would help to define better the spatial distribution of dengue within it and to improve its derivative burden estimates. We found high levels of precipitation and temperature suitability for dengue transmission to be most strongly associated among the variables considered with elevated dengue risk, although low precipitation was not found to limit transmission strongly (see Supplementary Information, section C). Proximity to low-income urban and peri-urban centres was also linked to greater risk, particularly in highly connected areas, indicating that human movement between population centres is an important facilitator of dengue spread. These associations have previously

been cited⁹, but have not been demonstrated at the global scale and highlight the importance of including socioeconomic covariates when assessing dengue risk.

We estimate that there were 96 million apparent dengue infections globally in 2010 (Table 1). Asia bore 70% (67 (47–94) million infections) of this burden, and is characterized by large swathes of densely populated regions coinciding with very high suitability for disease transmission. India^{19,20} alone contributed 34% (33 (24–44) million infections) of the global total. The disproportionate infection burden borne by Asian countries is emphasized in the cartogram shown in Fig. 2c. The Americas contributed 14% (13 (9–18) million infections) of apparent infections worldwide, of which over half occurred in Brazil and Mexico. Our results indicate that Africa's dengue burden is nearly equivalent to that of the Americas (16 (11–22) million infections, or 16% of the global total), representing a significantly larger burden than previously estimated. This disparity supports the notion of a largely hidden African dengue burden, being masked by symptomatically similar illnesses, under-reporting and highly variable treatment-seeking behaviour^{6,9,20}. The countries of Oceania contributed less than 0.2% of global apparent infections.

We estimate that an additional 294 (217–392) million inapparent infections occurred worldwide in 2010. These mild ambulatory or asymptomatic infections are not detected by the public health surveillance system and have no immediate implications for clinical management. However, the presence of this huge potential reservoir of infection has profound implications for: (1) correctly enumerating economic impact (for example, how many vaccinations are needed to avert an apparent infection) and triangulating with independent assessments of disability adjusted life years (DALYs)²¹; (2) elucidating the population dynamics of dengue viruses²²; and (3) making hypotheses about population effects of future vaccine programmes²³ (volume, targeting efficacy, impacts in combination with vector control), which will need to be administered to maximize cross-protection and minimize post-vaccination susceptibility.

The absolute uncertainties in the national burden estimates are inevitably a function of population size, with the greatest uncertainties in India, Indonesia, Brazil and China (see full rankings in Supplementary Table 4). In addition, comparing the ratio of the mean to the width of the confidence interval²⁴ revealed the greatest contributors to relative uncertainty (see full rankings in Supplementary Table 4). These were countries with sparse occurrence points and low evidence consensus on dengue presence, such as Afghanistan or Rwanda (see Fig. 2a), or those with ubiquitous high risk, such as Singapore or Djibouti, for which our burden prediction confidence interval is at its widest (see Supplementary Information, section D, Fig. 2). Therefore, increasing evidence consensus and occurrence data availability in low consensus countries and assembling new cohort studies, particularly in areas of high transmission, will reduce uncertainty in future burden estimates. Our approach, uniquely, provides new evidence to help maximize the value and cost-effectiveness of surveillance efforts, by indicating where limited resources can be targeted to have their maximum possible impact in improving our knowledge of the global burden and distribution of dengue.

Our estimates of total infection burden (apparent and inapparent) are more than three times higher than the WHO predicted figure (Supplementary Information, section E). Our definition of an apparent infection is broad, encompassing any disruption to the daily routine of the infected individual, and consequently is an inclusive measurement of the total population affected adversely by the disease. Within this broad class, the severity of symptoms will affect treatment-seeking behaviours and the probability of a correct diagnosis in response to a given infection. Our definition is therefore more comprehensive than those of traditional surveillance systems which, even in the most efficient system, report a much narrower range of dengue infections. By reviewing our database of longitudinal cohort studies, in which total infections in the community were documented exhaustively, we find

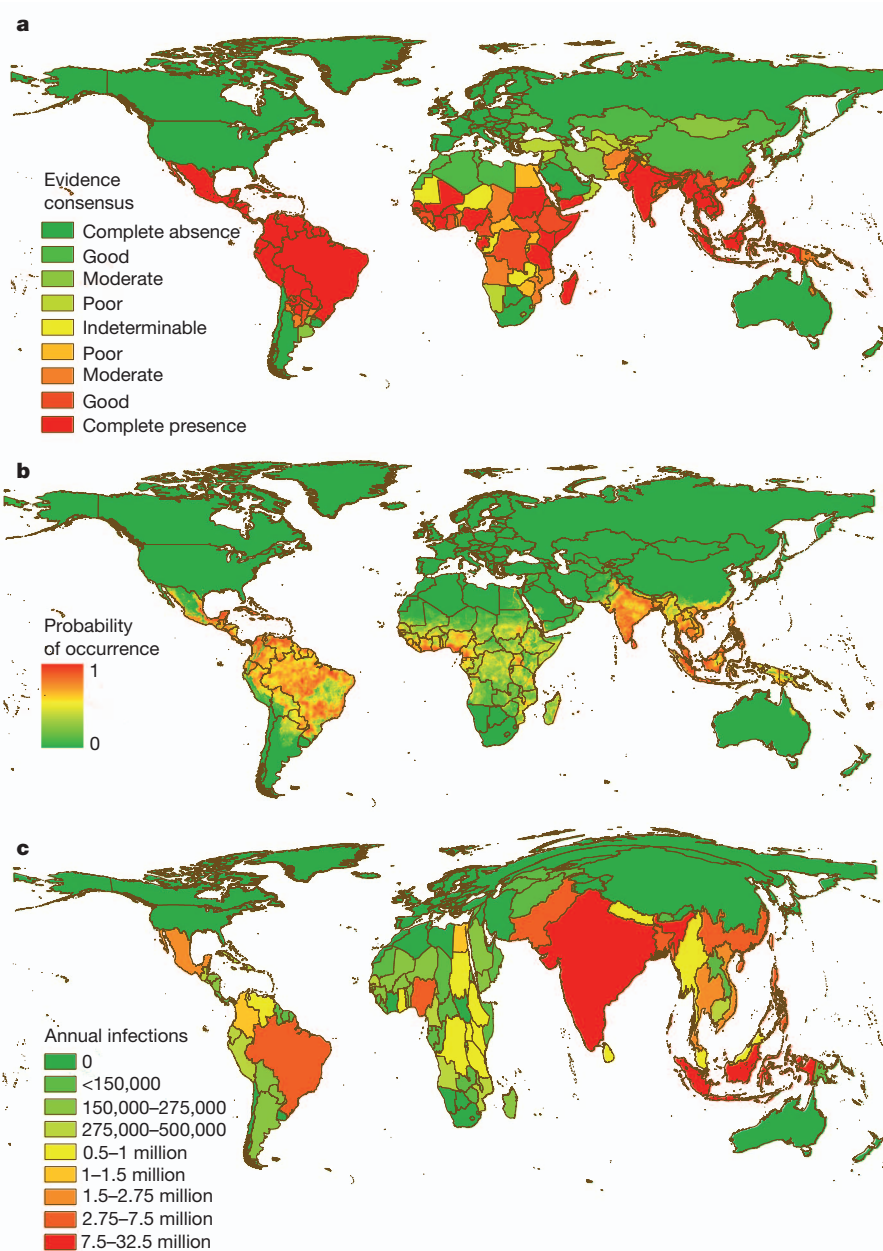


Figure 2 | Global evidence consensus, risk and burden of dengue in 2010. **a**, National and subnational evidence consensus on complete absence (green) through to complete presence (red) of dengue⁴. **b**, Probability of dengue occurrence at 5 km × 5 km spatial resolution of the mean predicted map (area under the receiver operator curve of 0.81 (± 0.02 s.d., $n = 336$)) from 336

boosted regression tree models. Areas with a high probability of dengue occurrence are shown in red and areas with a low probability in green. **c**, Cartogram of the annual number of infections for all ages as a proportion of national or subnational (China) geographical area.

that the biggest source of disparity between actual and reported infection numbers is the low proportion of individuals with apparent infections seeking care from formal health facilities (see Supplementary Information, section E, Fig. 5 for full analysis). Additional biases are

introduced by misdiagnosis and the systematic failure of health management information systems to capture and report presenting dengue cases. By extracting the average magnitude of each of these sequential disparities from published cohort and clinical studies, we can recreate a hypothetical reporting chain with idealized reporting and arrive at estimates that are broadly comparable to those countries reported to the WHO. This is most clear in more reliable reporting regions such as the Americas. Systemic under-reporting and low hospitalization rates have important implications, for example, in the evaluation of vaccine efficacy based on reduced hospitalized caseloads. Inferences about these biases may be made from the comparison of estimated versus reported infection burdens in 2010, highlighting areas where particularly poor reporting might be strengthened (see Supplementary Information, section E).

Table 1 | Estimated burden of dengue in 2010, by continent

	Apparent	Inapparent
	Millions (credible interval)	Millions (credible interval)
Africa	15.7 (10.5–22.5)	48.4 (34.3–65.2)
Asia	66.8 (47.0–94.4)	204.4 (151.8–273.0)
Americas	13.3 (9.5–18.5)	40.5 (30.5–53.3)
Oceania	0.18 (0.11–0.28)	0.55 (0.35–0.82)
Global	96 (67.1–135.6)	293.9 (217.0–392.3)

We have strived to be exhaustive in the assembly of contemporary data on dengue occurrence and clinical incidence and have applied new modelling approaches to maximize the predictive power of these data. It remains the case, however, that the empirical evidence base for global dengue risk is more limited than that available, for example, for *Plasmodium falciparum*²⁵ and *Plasmodium vivax*²⁶ malaria. Records of disease occurrence carry less information than those of prevalence and, as databases of the latter become more widespread, future approaches should focus on assessing relationships between seroprevalence and clinical incidence as a means of assessing risk²⁷. Additional cartographic refinements are also required to help differentiate endemic from epidemic-prone areas, to determine the geographic diversity of dengue virus types and to predict the distributions of future risk under scenarios of socioeconomic and environmental change.

The global burden of dengue is formidable and represents a growing challenge to public health officials and policymakers. Success in tackling this growing global threat is, in part, contingent on strengthening the evidence base on which control planning decisions and their impact are evaluated. It is hoped that this evaluation of contemporary dengue risk distribution and burden will help to advance that goal.

METHODS SUMMARY

We compiled a database of 8,309 geo-located occurrence records for the period 1960 to 2012 from a combination of published literature and online resources¹⁸. All records were standardized annually (that is, repeat records in the same location within a year merged as one occurrence) and underwent rigorous quality control. From a suite of potential environmental and socioeconomic covariates, we chose a relevant subset including: (1) two precipitation variables interpolated from global meteorological stations; (2) an index of temperature suitability for dengue transmission adapted from an equivalent index for malaria²⁸; (3) a vegetation/moisture index; (4) demarcations of urban and peri-urban areas; (5) an urban accessibility metric; and (6) an indicator of relative poverty. We then built a disease distribution model using a boosted regression tree (BRT) framework. To compensate for the lack of absence data, we created an evidence-based probabilistic framework for generating pseudo-absences that mitigated the main biasing factors in pseudo-absence generation²⁹, namely: (1) geographical extent; (2) number; (3) contamination bias; and (4) sampling bias. We then created an ensemble of 336 BRT models using different plausible combinations of these factors and representing independent samples of possible sampling distributions. We calculated the final probability of occurrence (risk) map as the central tendency of these 336 BRT models predicted at a 5 km × 5 km resolution. Exclusion criteria were based on the definitive extents of dengue⁴ and temperature suitability for dengue transmission²⁸. Using detailed longitudinal information from 54 dengue cohort studies, we defined a relationship between the probability of dengue occurrence and inapparent and apparent incidence using a Bayesian hierarchical model. We defined a negative binomial likelihood function with constant dispersion and a rate characterized by a highly flexible data-driven Gaussian process prior. Uninformative hyperpriors were assigned hierarchically to the prior parameters and the full posterior distribution determined by Markov Chain Monte Carlo (MCMC) sampling. Using human population gridded data, estimates of dengue infections were then calculated nationally, regionally and globally for both apparent and inapparent infections.

Full Methods and any associated references are available in the online version of the paper.

Received 8 October 2012; accepted 7 March 2013.

Published online 7 April 2013.

1. Simmons, C. P., Farrar, J. J., van Vinh Chau, N. & Wills, B. Dengue. *N. Engl. J. Med.* **366**, 1423–1432 (2012).
2. World Health Organization. *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control*. WHO/HTM/NTD/DEN/2009.1 (World Health Organization, 2009).
3. Tatem, A. J., Hay, S. I. & Rogers, D. J. Global traffic and disease vector dispersal. *Proc. Natl Acad. Sci. USA* **103**, 6242–6247 (2006).
4. Brady, O. J. *et al.* Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl. Trop. Dis.* **6**, e1760 (2012).
5. Halstead, S. B. Pathogenesis of dengue: challenges to molecular biology. *Science* **239**, 476–481 (1988).
6. Endy, T. P. *et al.* Determinants of inapparent and symptomatic dengue infection in a prospective study of primary school children in Kamphaeng Phet, Thailand. *PLoS Negl. Trop. Dis.* **5**, e975 (2011).

7. Sabchareon, A. *et al.* Protective efficacy of the recombinant, live-attenuated, CYD tetravalent dengue vaccine in Thai schoolchildren: a randomised, controlled phase 2b trial. *Lancet* **380**, 1559–1567 (2012).
8. Halstead, S. B. Dengue vaccine development: a 75% solution? *Lancet* **380**, 1535–1536 (2012).
9. Gubler, D. J. Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.* **11**, 480–496 (1998).
10. Beatty, M. E., Letson, G. W. & Margolis, H. S. Estimating the global burden of dengue. *Am. J. Trop. Med. Hyg.* **81** (Suppl. 1), 231 (2009).
11. Van Kleef, E., Bambrick, H. & Hales, S. The geographic distribution of dengue fever and the potential influence of global climate change. *TropIKA. net* http://journal.tropika.net/scielo.php?script=sci_arttext&pid=S2078-86062010005000001&lng=en&nrm=iso (2009).
12. World Health Organization. *International Travel and Health: Situation as on 1 January 2012* (World Health Organization, 2012).
13. Hales, S., de Wet, N., Maindonald, J. & Woodward, A. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *Lancet* **360**, 830–834 (2002).
14. Rogers, D. J., Wilson, A. J., Hay, S. I. & Graham, A. J. The global distribution of yellow fever and dengue. *Adv. Parasitol.* **62**, 181–220 (2006).
15. Monath, T. P. Yellow fever and dengue: the interactions of virus, vector and host in the re-emergence of epidemic disease. *Semin. Virol.* **5**, 133–145 (1994).
16. Rigau-Pérez, J. G. *et al.* Dengue and dengue haemorrhagic fever. *Lancet* **352**, 971–977 (1998).
17. Rodhain, F. La situation de la dengue dans le monde. *Bull. Soc. Pathol. Exot.* **89**, 87–90 (1996).
18. Freifeld, C. C., Mandl, K. D., Reis, B. Y. & Brownstein, J. S. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J. Am. Med. Inform. Assoc.* **15**, 150–157 (2008).
19. Chakravarti, A., Arora, R. & Luxemburger, C. Fifty years of dengue in India. *Trans. R. Soc. Trop. Med. Hyg.* **106**, 273–282 (2012).
20. Kakkar, M. Dengue fever is massively under-reported in India, hampering our response. *Br. Med. J.* **345**, e8574 (2012).
21. Murray, C. J. L. *et al.* Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2197–2223 (2012).
22. Cummings, D. A. *et al.* The impact of the demographic transition on dengue in Thailand: insights from a statistical analysis and mathematical modeling. *PLoS Med.* **6**, e1000139 (2009).
23. Johansson, M. A., Hombach, J. & Cummings, D. A. Models of the impact of dengue vaccines: a review of current research and potential approaches. *Vaccine* **29**, 5860–5868 (2011).
24. Hay, S. I. *et al.* Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. *PLoS Med.* **7**, e1000290 (2010).
25. Gething, P. W. *et al.* A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar. J.* **10**, 378 (2011).
26. Gething, P. W. *et al.* A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Negl. Trop. Dis.* **6**, e1814 (2012).
27. Anders, K. L. & Hay, S. I. Lessons from malaria control to help meet the rising challenge of dengue. *Lancet Infect. Dis.* **12**, 977–984 (2012).
28. Gething, P. W. *et al.* Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasites Vectors* **4**, 92 (2011).
29. Chefaoui, R. M. & Lobo, J. M. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecol. Modell.* **210**, 478–486 (2008).
30. TDR/World Health Organization. *Report of the Scientific Working Group on Dengue, 2006*. TDR/SWG/08 (TDR/World Health Organization, 2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements S.I.H. is funded by a Senior Research Fellowship from the Wellcome Trust (095066) which also supports S.B. and P.W.G. C.P.S. is also funded by a Senior Research Fellowship from the Wellcome Trust (084368). O.J.B. is funded by a BBSRC Industrial CASE studentship. J.P.M., A.W.F., T.J., G.R.W.W., C.P.S., T.W.S. and S.I.H. received funding from, and with S.B., P.W.G., O.J.B. and J.J.F. acknowledge the contribution of, the International Research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS, 21803, <http://www.idams.eu>). This work was funded in part by EU grant 2011-261504 EDENEXT and the paper is catalogued by the EDENEXT Steering Committee as EDENEXT. S.I.H. and T.W.S. also acknowledge funding support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health.

Author Contributions S.I.H. and J.J.F. conceived the research. S.B. and S.I.H. drafted the manuscript. S.B. drafted the Supplementary Information with significant support on sections A (O.J.B., C.L.M.), B (J.P.M., G.R.W.W.), C (P.W.G.), D (O.J.B., T.W.S.), and O.J.B. wrote section E. J.S.B. and A.G.H. provided HealthMap occurrence data and advice on its provenance. O.J.B. reviewed all the occurrence data. S.B. did the modelling and analysis with advice from J.M.D., P.W.G. and S.I.H. J.P.M. created all maps. All authors discussed the results and contributed to the revision of the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.I.H. (simon.hay@zoo.ox.ac.uk).

METHODS

Assembly of the occurrence database and its quality control. Occurrence data comprised of point or polygon locations of confirmed dengue infection presence derived from both peer-reviewed literature and HealthMap alerts^{18,31} (see Supplementary Information, section A). An occurrence was defined as one or more laboratory or clinically confirmed infection(s) of dengue occurring at a unique location (a 5 km × 5 km pixel) within one calendar year. All occurrence data underwent manual review and automatic quality control to ensure information fidelity and precise geo-positioning. In total, 9,648 and 1,622 occurrence locations were obtained from literature searches and HealthMap, respectively. After the quality control procedures, our final data set contained 8,309 occurrence locations (5,216 point locations and 3,093 small polygon centroids) spanning a period from 1960 to 2012. We assume any record of dengue occurrence, regardless of its age, represented an environment permissible for the disease, as dengue has expanded from a focal disease in Asia to a cosmopolitan disease of the tropics.

Explanatory covariates. We assembled gridded global data for a suite of eight explanatory covariates. The covariates were chosen based on factors known or hypothesized to contribute to suitability for dengue transmission (see Supplementary Information, section B). These covariates included: (1) annual maximum and minimum precipitation variables from a Fourier processed³² synoptic annual series interpolated from global meteorological stations³³; (2) a biological model combining the effects of temperature on the extrinsic incubation period of dengue virus and lifespan of the *Aedes aegypti* vector to quantify the dengue-specific temperature suitability for transmission^{28,34,35}; (3) Fourier-processed annual average normalized difference vegetation index³⁶; (4) categorical demarcations of urban and peri-urban areas³⁷; (5) an urban accessibility metric defining the travel time to nearest city of 50,000 people or more by land- or water-based travel³⁸; and (6) an indicator of relative poverty derived from the finest geographic scale data available for economic productivity and adjusted for purchasing power parity³⁹. No covariate grids were shown to be adversely affected by multicollinearity (see Supplementary Information, section B) and were standardized to ensure identical spatial resolution, extent and boundaries. For point records, covariate values corresponded to the pixel value containing the location of the point. For polygon occurrence records, covariate values were averaged across the whole polygon.

Predicting the probability of occurrence (risk) of dengue transmission. We used a boosted regression tree (BRT) approach to establish a multivariate empirical relationship between the probability of occurrence of a dengue virus infection and the environmental conditions sampled at each site from the covariate suite. The BRT method has been shown to fit complicated response functions efficiently, while guarding against overfitting, and is therefore widely used for vector and disease distribution mapping^{40,41}. The BRT approach combines regression trees⁴² with gradient boosting⁴³, whereby an initial regression tree is fitted and iteratively improved upon in a forward stage-wise manner (boosting) by minimizing the variation in the response not explained by the model at each iteration (see Supplementary Information, section C).

Like other niche mapping approaches, the BRT models require not only presence data but also absence data defining areas of disease absence and potentially unsuitable environmental conditions at unsampled locations. Because data on absence of disease are not definitive, pseudo-absence data estimate areas of disease absence instead. No consensus approach has been developed to optimize the generation of pseudo-absence data and we therefore created an evidence-based probabilistic framework for generating pseudo-absences, incorporating the main biasing factors in pseudo-absence generation, namely: (1) geographical extent; (2) number; (3) contamination bias; and (4) sampling bias. To represent areas of absence, n_a pseudo-absence points^{29,44,45} were randomly generated based on dengue presence or absence certainty measures at a national or subnational level⁴. Pseudo-absence locations were restricted to a maximum distance μ from any recorded presence site^{46,47}. Additionally, to compensate for 'contamination' of true but unobserved presences within the generated pseudo-absences⁴⁸, n_p pseudo-presence points were generated using the same procedure used to generate the pseudo-absences. Variation in the parameter set $\pi = \{\mu, n_a, n_p\}$ resulted in independent samples of the possible states of the real distribution, with all parameter combinations representing a null distribution of possible states. Therefore, rather than using an individual parameter combination from π , we created an ensemble⁴⁹ of 336 BRT models spanning reasonable ranges in π and evaluated the central tendency as the mean across all 336 BRT models (see Supplementary Information, section C). The final ensemble BRT model was used to predict a global map of the probability of occurrence of dengue virus infection at a 5 km × 5 km resolution.

Estimation of dengue burden and populations at risk. Formal literature searches were conducted for serological dengue virus incidence surveys. Inclusion criteria were restricted to longitudinal surveys of seroconversion to

dengue-virus-specific antibodies carried out in parallel with active symptom surveillance in a defined cohort. The surveys were abstracted, standardized and geopositioned (see Supplementary Information, section D). In total, 54 dengue incidence surveys were collected. Of these, 39 contained information about the ratio of inapparent to apparent infections.

The empirical relationship between incidence and the probability of occurrence was represented using a Bayesian hierarchical model. We defined a negative binomial likelihood function⁵⁰ with constant dispersion and a rate characterized by a highly flexible data-driven Gaussian process prior⁵¹. The Gaussian process prior was parameterized with a quadratic mean function and a squared exponential covariance function⁵¹. Uninformative hyperpriors were assigned hierarchically to the prior parameters and the full posterior distribution determined by Markov Chain Monte Carlo (MCMC) sampling⁵². The entire model was fitted separately for apparent and inapparent infection incidences, with missing inapparent to apparent ratio values imputed in the MCMC. Using human population gridded data for the year 2010 (ref. 53), estimates of apparent and inapparent dengue infections were calculated nationally, regionally and globally. These estimates were then compared to national clinical cases reported to the WHO and differences between our cartographic estimates of infections and the WHO surveillance estimates were reconciled in a comparative analysis addressing key factors in traditional surveillance under-reporting (see Supplementary Information, section E).

- Brownstein, J. S., Freifeld, C. C., Reis, B. Y. & Mandl, K. D. Surveillance sans frontières: internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med.* **5**, e151 (2008).
- Scharlemann, J. P. W. *et al.* Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data. *PLoS ONE* **3**, e1408 (2008).
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
- Focks, D. A., Haile, D. G., Daniels, E. & Mount, G. A. Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): analysis of the literature and model development. *J. Med. Entomol.* **30**, 1003–1017 (1993).
- Focks, D. A., Haile, D. G., Daniels, E. & Mount, G. A. Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): simulation and validation. *J. Med. Entomol.* **30**, 1018–1028 (1993).
- Hay, S. I., Tatem, A. J., Graham, A. J., Goetz, S. J. & Rogers, D. J. Global environmental data for mapping infectious disease distribution. *Adv. Parasitol.* **62**, 37–77 (2006).
- Hay, S. I. *et al.* A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med.* **6**, e48 (2009).
- Nelson, A. Estimated travel time to the nearest city of 50,000 or more people in year 2000. <http://bioval.jrc.ec.europa.eu/products/gam> (accessed 1 January 2012) (Global Environment Monitoring Unit – Joint Research Centre of the European Commission, 2008).
- Nordhaus, W. D. Geography and macroeconomics: new data and new findings. *Proc. Natl Acad. Sci. USA* **103**, 3510–3517 (2006).
- Elith, J. *et al.* Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151 (2006).
- Stevens, K. B. & Pfeiffer, D. U. Spatial modelling of disease using data- and knowledge-driven approaches. *Spat. Spatiotemporal Epidemiol.* **2**, 125–133 (2011).
- Breiman, L. *Classification and Regression Trees* (Chapman & Hall/CRC, 1984).
- Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
- Stokland, J. N., Halvorsen, R. & Stoa, B. Species distribution modelling. Effect of design and sample size of pseudo-absence observations. *Ecol. Modell.* **222**, 1800–1809 (2011).
- Lobo, J. M. & Tognelli, M. F. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *J. Nat. Conserv.* **19**, 1–7 (2011).
- VanDerWal, J., Shoo, L. P., Graham, C. & William, S. E. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecol. Modell.* **220**, 589–594 (2009).
- Barbet-Massin, M., Jiguet, F., Albert, C. H. & Thuiller, W. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol.* **3**, 327–338 (2012).
- Ward, G., Hastie, T., Barry, S., Elith, J. & Leathwick, J. R. Presence-only data and the EM algorithm. *Biometrics* **65**, 554–563 (2009).
- Araújo, M. B. & New, M. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**, 42–47 (2007).
- Hilbe, J. M. *Negative Binomial Regression* 2nd edn, 251 (Cambridge Univ. Press, 2011).
- Banerjee, S., Carlin, B. P. & Gelfand, A. E. *Hierarchical Modeling and Analysis for Spatial Data. Monographs on Statistics and Applied Probability 101* (Chapman & Hall/CRC, 2004).
- Patil, A., Huard, D. & Fonnesbeck, C. J. PyMC: Bayesian stochastic modelling in Python. *J. Stat. Softw.* **35**, e1000301 (2010).
- Balk, D. L. *et al.* Determining global population distribution: methods, applications and data. *Adv. Parasitol.* **62**, 119–156 (2006).