SCIENTIFIC DATA

The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence

Moritz U.G. Kraemer[1], Marianne E. Sinka[2,3], Kirsten A. Duda[1], Adrian Mylne[2,3], Freya M. Shearer[2,3], Oliver J. Brady[2,3], Jane P. Messina[1], Christopher M. Barker[4,5,6], Chester G. Moore[7], Roberta G. Carvalho[8], Giovanini E. Coelho[8], Wim Van Bortel[9], Guy Hendrickx[10], Francis Schaffner[10], G.R. William Wint[1,11], Iqbal R.F. Elyazar[12], Hwa-Jen Teng[13] & Simon I. Hay[1,6]

Aedes aegypti and *Ae. albopictus* are the main vectors transmitting dengue and chikungunya viruses. Despite being pathogens of global public health importance, knowledge of their vectors' global distribution remains patchy and sparse. A global geographic database of known occurrences of *Ae. aegypti* and *Ae. albopictus* between 1960 and 2014 was compiled. Herein we present the database, which comprises occurrence data linked to point or polygon locations, derived from peer-reviewed literature and unpublished studies including national entomological surveys and expert networks. We describe all data collection processes, as well as geo-positioning methods, database management and quality-control procedures. This is the first comprehensive global database of *Ae. aegypti* and *Ae. albopictus* occurrence, consisting of 19,930 and 22,137 geo-positioned occurrence records respectively. Both datasets can be used for a variety of mapping and spatial analyses of the vectors and, by inference, the diseases they transmit.

Design Type(s)	observation design ● epidemiological study ● data integration objective
Measurement Type(s)	viral host epidemiology
Technology Type(s)	data collection method
Factor Type(s)	
Sample Characteristic(s)	Aedes aegypti ● anthropogenic habitat ● Aedes albopictus

[1]Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. [2]Wellcome Trust Centre for Human Genetics,University of Oxford, Oxford, UK. [3]Institute for Health Metrics and Evaluation, University of Washington, Seattle, USA. [4]Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, CA, USA. [5]Center for Vectorborne Diseases, University of California, Davis, CA, USA. [6]Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA. [7]Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO, USA. [8]National Dengue Control Program, Ministry of Health, Brasilia, DF, Brazil. [9]European Centre for Disease Prevention and Control, Stockholm, Sweden. [10]Avia-GIS, Zoersel, Belgium. [11]Environmental Research Group Oxford Ltd, Department of Zoology, South Parks Road, Oxford OX1 3PS, UK. [12]Eijkman-Oxford Clinical Research Unit, Jakarta, Indonesia. [13]Center for Research, Diagnostics and Vaccine Development, Centers for Disease Control, Taipei, Taiwan (ROC). Correspondence and requests for materials should be addressed to S.I.H. (email: simon.i.hay@gmail.com).

Background & Summary

Aedes aegypti [= *Stegomyia aegypti*[1]] and *Ae. albopictus* [= *Stegomyia albopicta*[1]] are disease vectors for many important viral human diseases such as dengue, chikungunya and yellow fever[2–4]. Dengue is the most prevalent human arboviral infection causing approximately 100 million apparent annual infections with almost half of the world's population at risk[5]. Dengue transmission now occurs in over 120 countries[6], mostly in the tropics and sub-tropics. Chikungunya, another arthropod-borne virus, has caused over 2.5 million infections over the past decade and has more recently been spreading in the Americas and emerging in Europe, posing new challenges to health systems as it spreads into new areas, infecting naïve populations and consequently causing large outbreaks[7–10]. The disease burden of yellow fever was significantly reduced due to large-scale vaccination programs in the twentieth century but current estimates of 51,000–380,000 severe cases in Africa per year point to the continuing difficulty in fully controlling this virus[11]. As a result, there is growing interest in describing the global geographic distribution of both vector species to better understand the risk of transmission of these viruses.

Aedes aegypti is a predominantly urban vector, utilising the abundance of artificial containers as larval sites and feeding almost exclusively on humans[12]. *Aedes albopictus* can more often be found in peri-urban and rural environments, feeding readily on a variety of mammalian (including humans) and avian species[13].

Aedes mosquito surveys are performed to better understand ecological and epidemiological aspects of the vectors as well as to assist disease surveillance and control[14–16]. Surveillance of *Aedes* can involve; (i) systematic household surveys that involve searching water-filled containers for larvae and pupae, (ii) the use of backpack aspirators and suction traps baited with a chemical lure and/or CO_2 for the collection of adult mosquitoes or, (iii) using ovitraps placed strategically around a neighbourhood to collect mosquito eggs that can then be reared back in the laboratory for morphological identification or directly processed for molecular identification[17–19].

The database described here contains information on the known global occurrences of the adults, pupae, larvae or eggs of *Ae. aegypti* and *Ae. albopictus* globally from 1960 and 2014.

By including data from a variety of sources we were able to create the largest currently available standardised up-to-date global dataset for both *Ae. aegypti* and *Ae. albopictus* (Fig. 1), containing 42,067 geo-positioned occurrences.

Methods

Data collection

PubMed (http://www.ncbi.nlm.nih.gov/nlmcatalog/journals) was searched using the term '*Aedes*' OR '*aegypti*' OR '*albopictus*' for the years 1960 to 2013. The Medical Subject Headings (MeSH) term technology used in the PubMed citation archive ensured all pseudonyms were automatically included (http://www.nlm.nih.gov/mesh) in the searches. The same process was repeated for ISI Web of Science (http://wok.mimas.ac.uk) and ProMED (http://www.promedmail.org). The searches were last updated on 15th November 2013. No language restrictions were placed on these searches; however, only those citations with a full title and abstract were retrieved. This resulted in a collection of 8,597 references, of which 2,804 unique articles were identified from their abstracts as potentially containing useable location data. In-house language skills allowed processing of all English, French, Portuguese and Spanish articles. Confirmed *Aedes* occurrences within these articles were entered into the database. Occurrences were classified as confirmed when the article clearly stated the presence of the vector at a specific time in a specific location. This includes transient populations, i.e., found in ports or only during the summer months. Only for Europe were we able to include information of transient versus established populations using expert opinions. Laboratory studies were included if the mosquito/larvae were collected from the wild specifically for the purpose of the study. Occurrences were recorded separately for both species. More specific information about 'sub-species' or 'genetic characteristics' for example, were recorded where available but not included in the final database. This information can be obtained from the authors upon request. Data Citation 1 lists full references for each published record in the database.

In addition to the data directly sourced from published literature, primary and unpublished occurrence data from national entomological surveys were obtained through contact with administrators of these surveys when possible (Fig. 1). This includes primary data for *Ae. albopictus* provided from an earlier published article by Carvalho *et al.*[20] Collections have been part of the Levantamento Rapido de Indice para *Aedes aegypti* (LIRAa) in Brazil and are described in full elsewhere[20,21]. Similarly, *Ae. aegypti* primary data with geographic locations were also provided by an entomological survey directly from the Ministry of Health of Brazil for 2013. All occurrences for Brazil were classified as polygons (see geo-positioning methods) as they represent surveys conducted in Brazilian municipalities with their respective centroids being used as geographic information in the database. Primary and unpublished occurrence points for both species were provided by Elyazar (2014) for Indonesia. Moore and Barker (2014) provided summary data from the United States, collected between 1985–2014 from published and unpublished sources, including geographic location at the USA county administrative level which were included as polygons in the database. In addition, Teng (2014) added occurrence records for both species based on national entomological surveys performed in Taiwan between 2004–2013. European *Ae. albopictus* records were derived from the European Centre for Disease Prevention and Control (ECDC) funded TIGERMAPS
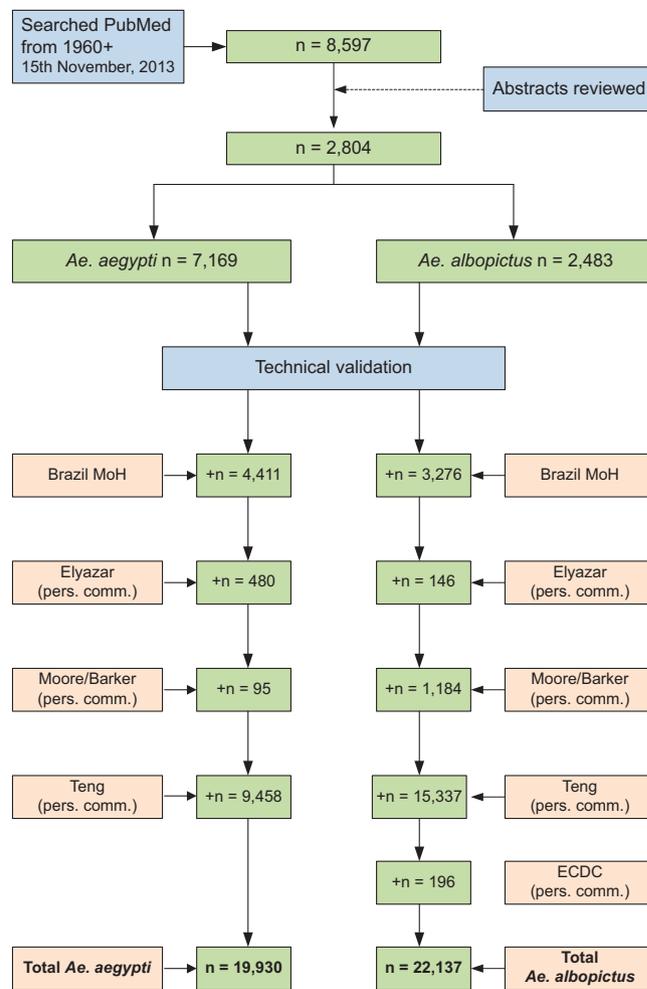
Figure 1. The numbers of occurrences for *Ae. aegypti and Ae. albopictus* by source.

and VBORNET datasets provided by Schaffner and Hendrickx (2013). Data were based on Nomenclature of territorial units for statistics 3 (NUTS3) level for Europe from 1979–2013. NUTS3 centroids were calculated in ArcGIS, occurrences classified as polygons and added to the database.

Geo-positioning of data from published sources
All available location information was extracted for each occurrence from the relevant primary research article. The site name was used together with all contextual information provided about the site position to determine its latitudinal and longitudinal coordinates using Google Maps (https://www.maps.google.co.uk), Google Earth (http://www.google.co.uk/intl/en_uk/earth), or other online geo-positional databases including Geonames (http://www.geonames.org), Fallingrain (http://www.fallingrain.com/world/index.html) or as a last resort, using simple Google searches. Place names are often duplicated within a country, so contextual information was used to ensure the right site was selected. When the site name was not found, information from the text, was also used to scan sites in the approximate area to check for alternate spelling of the site name. If the study site could be geo-positioned to a specific latitude and longitude within a 5 km × 5 km pixel, it was termed a 'point location'. For each occurrence that could not be assigned a single 5 km × 5 km pixel, e.g., a large city, the occurrence was entered as a polygon data type. Polygon occurrences were subsequently classified based on the polygon size they correspond to as either between 5–10 km^2, 10–25 km^2, 25–100 km^2 or >100 km^2. All locations were then linked to administrative units as recognised by the FAO Global Administrative Unit Layer (GAUL) system[22]. This initial database then underwent spatial and temporal standardisation and finally technical validation.

Occurrence database management: spatial and temporal standardisation
As the database was compiled from many different sources and several institutions, it was first necessary to standardise the data entries such that identical locations which may have been geo-positioned slightly differently were given the same unique identifier. Point records were given the same unique identifier if they lay within the same 5 km × 5 km pixel within a global grid. Finally, any record associated with a

polygon measuring larger than $111\,km \times 111\,km$ at the equator (1 degree) was removed from the database ($n = 475$, $n = 54$ for *Ae. aegypti* and *Ae. albopictus* respectively).

Similarly, it was necessary to temporally standardise the database to avoid duplicates. We chose to define a single occurrence at a given unique location (as identified above) within one calendar year. This was particularly important for oversampled regions that undergo multiple yearly surveys such as Taiwan and involved a procedure which: (i) disaggregated any records which were in the same location but spanning multiple years into individual records for each respective year and then (ii) aggregated all records with the same unique location identifier and occurring within the same year to form a single occurrence record. This led to 1,112 and 370 records being removed for *Ae. aegypti* and *Ae. albopictus,* respectively.

Data Records

This database is publicly available online as a comma-delimited file for both species independently for ease of use and the ability to import it into a variety of software programs (Data Citation 1). Each of the rows represents a single occurrence record (one or more *Aedes* cases in the same unique location within a single calendar year). The fields contained in the database are as follows:

1. **VECTOR:** Identifying the species; *Ae. aegypti* or *Ae. albopictus*
2. **OCCURRENCE_ID:** Unique identifier for each occurrence in the database after temporal and spatial standardisation.
3. **SOURCE_TYPE:** Published literature or unpublished sources with reference ID that corresponds to the full list of references in Data Citation 1.
4. **LOCATION_TYPE:** Whether the record represents a point or a polygon location.
5. **POLYGON_ADMIN:** Admin level or polygon size which the record represents when the location type is a polygon. −999 when the location type is a point ($5\,km \times 5\,km$).
6. **X:** The longitudinal coordinate of the point or polygon centroid (WGS1984 Datum).
7. **Y:** The latitudinal coordinate of the point or polygon centroid (WGS1984 Datum).
8. **YEAR:** The year of the occurrence.
9. **COUNTRY:** The name of the country within which the occurrence lies.
10. **COUNTRY_ID:** ISO alpha-3 country codes.
11. **GAUL_AD0:** The country-level global administrative unit layer (GAUL) code (see http://www.fao.org/geonetwork) which identifies the Admin-0 polygon within which any smaller polygons and points lie.
12. **STATUS:** Established versus transient populations.

Technical Validation

The following procedures were carried out on the final database to ensure the accuracy and validity of the occurrence records.

1. A raster distinguishing land from water[22] was created at a $5\,km \times 5\,km$ resolution and was used to ensure all occurrences were positioned on a valid land pixel ($n = 95$ and $n = 64$ records were removed for *Ae. aegypti* and *Ae. albopictus* respectively).
2. We cross-validated all of the unique occurrence locations against temperature-based *Aedes* population persistence metrics developed by Brady *et al.*[23] In brief, this classification was determined by modelling the effect of temperature on adult *Ae. aegypti* and *Ae. albopictus* survival and length of first gonotrophic cycle, the interaction of which determines whether the population can persist. Population
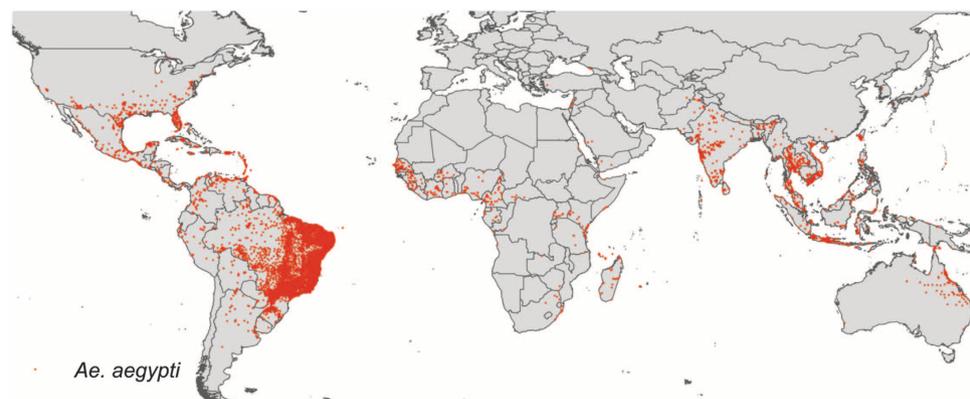
Figure 2. Map of occurrence points for *Ae. aegypti*.

persistence was then predicted on a global scale using interpolated meteorological data[24]. Occurrences that fell outside this range were re-checked to ensure the quality of the occurrence records.

The result is a database consisting of 19,930 and 22,137 geo-positioned occurrences in total worldwide for *Ae. aegypti* and *Ae. albopictus* respectively, broken down by region, location type and source type in Fig. 1. In Figs 2 and 3 the global geographic distribution of both species is displayed.
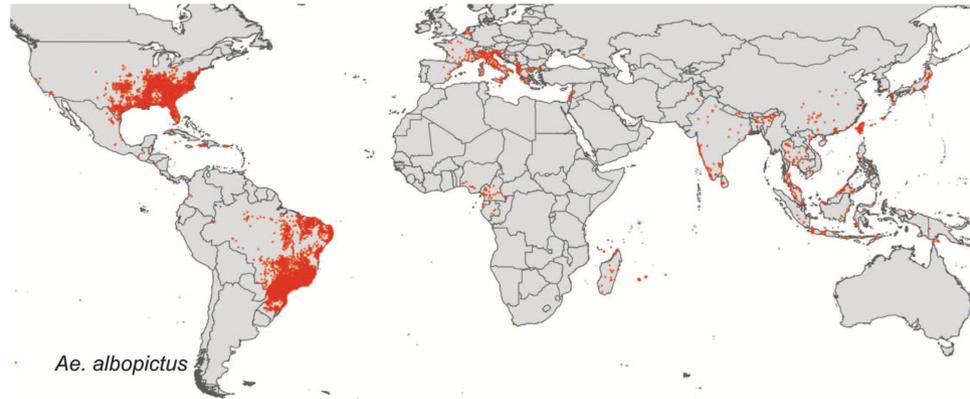

Figure 3. Map of occurrence points for *Ae. albopictus*.
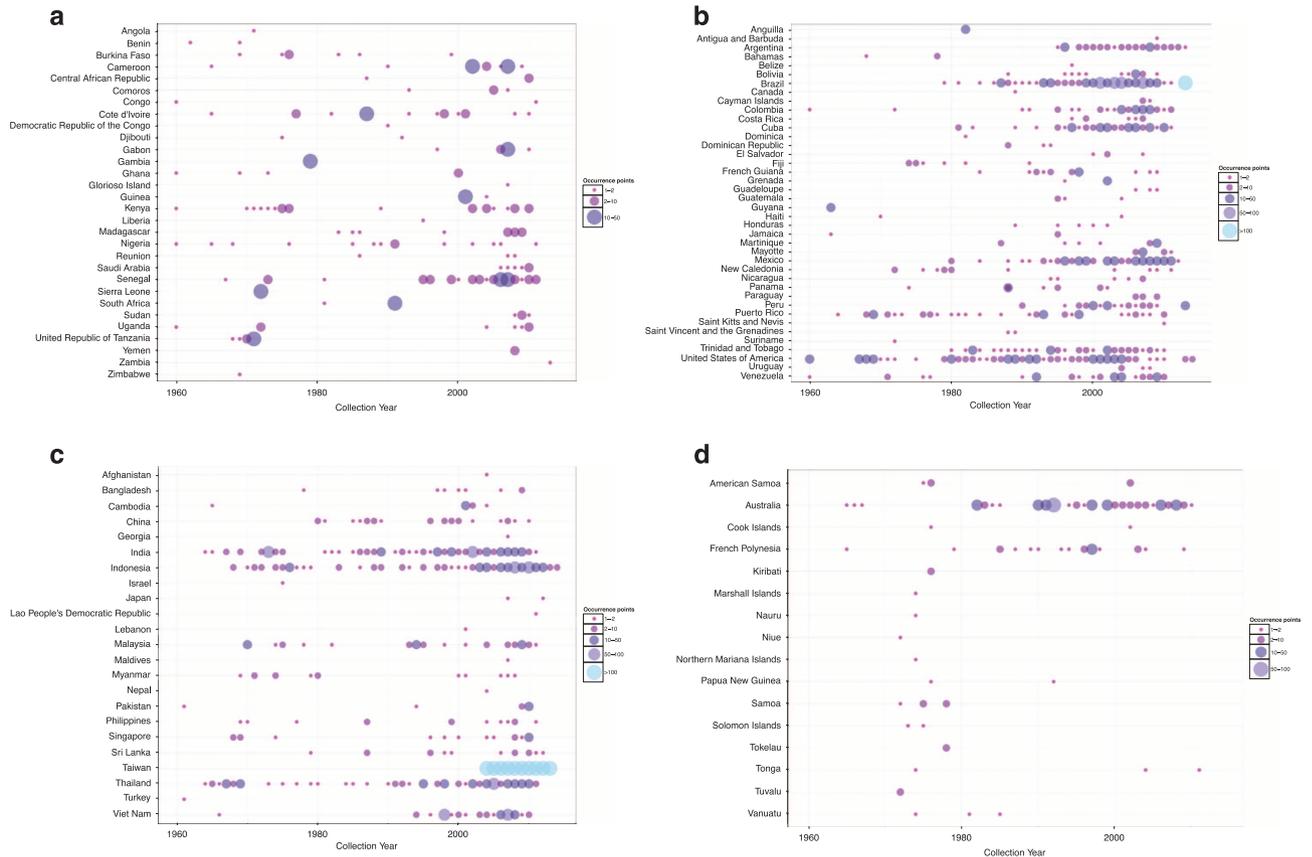
Figure 4. The numbers of occurrences for *Ae. aegypti* with locations per year separated by region, with panels (**a**–**d**) representing regions. Colours indicate the number of mosquitoes recorded in each year with the size of the circles scaled respectively.
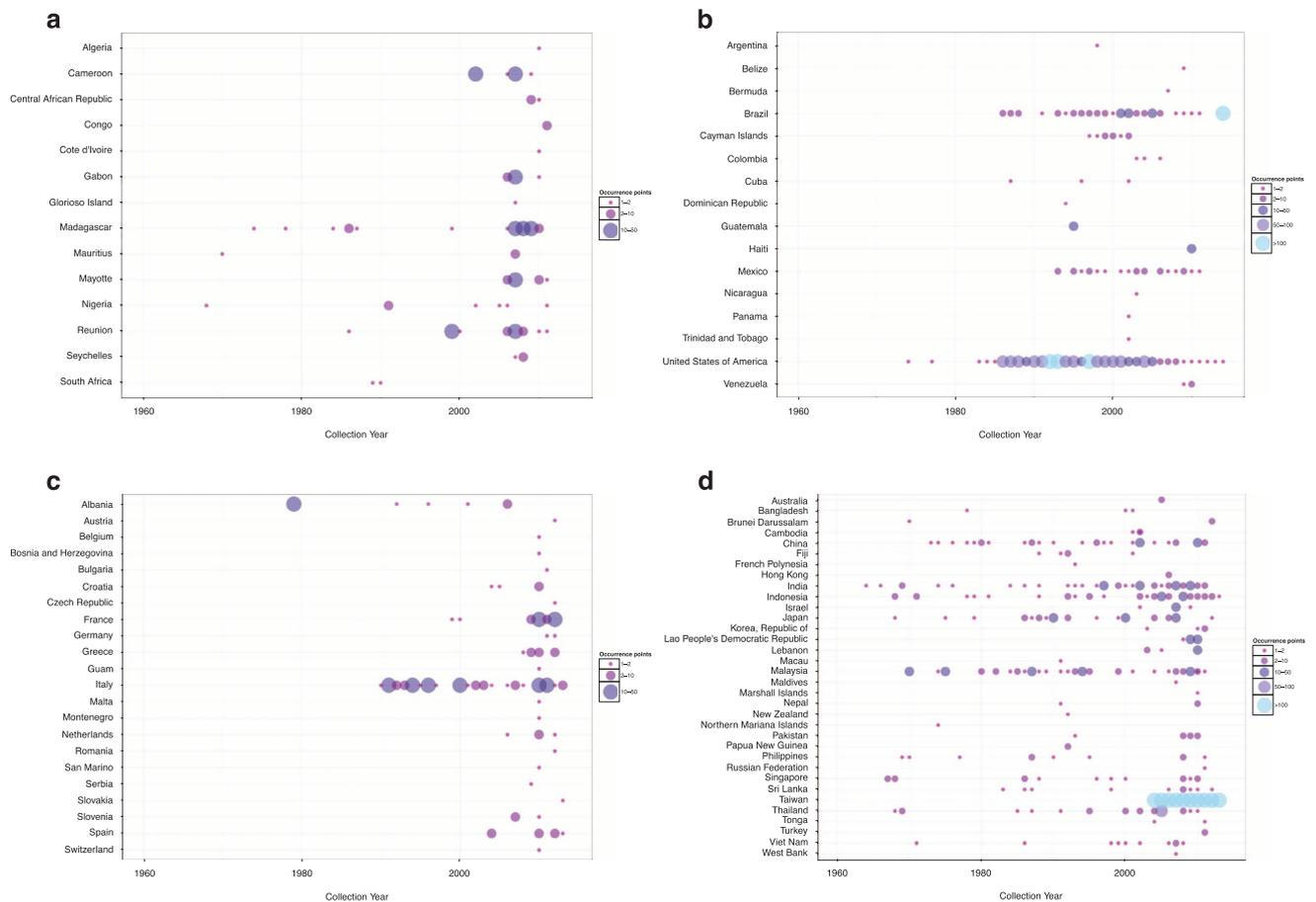
Figure 5. The numbers of occurrences for *Ae. albopictus* with locations per year separated by region, with panels (**a**–**d**) representing regions. Colours indicate the number of mosquitoes recorded in each year with the size of the circles scaled respectively.

Figs 4 and 5 show occurrence records by year and region for both *Ae. aegypti* and *Ae. albopictus* respsectively. The increase in occurrence records since 2013 is largely attributable to the LIRAa data from Brazil.

Usage Notes

The dataset described here can be used to investigate the spatial and temporal patterns of *Aedes* distribution at multiple scales and resolutions. As *Ae. aegypti* and *Ae. albopictus* are invasive species, spreading to new areas via shipping routes and human movement[25–27], this dataset could improve predictions of locations at high-risk for importation[25]. This dataset can also be used to contribute to modelling areas at risk for dengue[28] and chikungunya[9] especially in areas in Europe[16,28] and the USA[29,30]. We aimed at building a comprehensive set of data based on occurrences ever recorded globally including their respective dates to allow researchers as well as policy makers to filter the dataset based on their respective research questions.

   This dataset was first used in an ecological niche modelling framework along with a set of environmental covariates to map the global distribution of each species[32]. A generic code to produce the global risk maps is openly available as an R software package 'seegSDM' from GitHub (https://github.com/SEEG-Oxford/seegSDM). Such maps can help to guide vector surveillance efforts in countries where the distribution of both species is not well-known, but which are at high risk for importation of related viruses.

   Regional biases in density of occurrence records are apparent and may be due to differences in the amount of regular surveillance, differences in the number of published studies and availability of routinely collected data. Use on a global scale, however, would need to take into account geographical sampling bias as done in Kraemer *et al.* using similarly biased background points in a presence-only niche modelling approach[31,32]. The method for accounting for sampling bias, however, might vary depending on the research question asked and methodology applied in subsequent analyses.

References

1. Reinert, J. F., Harbach, R. E. & Kitching, I. J. Phylogeny and classification of tribe *Aedini (Diptera: Culicidae). Zool. J. Linn. Soc.* **157,** 700–794 (2009).
2. Hay, S. I. *et al.* Global mapping of infectious disease. *Philos. Trans. R. Soc. Lond. B. Biol. Sci* **368,** 20120250 (2013).
3. Juliano, S. a. & Philip Lounibos, L. Ecology of invasive mosquitoes: effects on resident species and on human health. *Ecol. Lett.* **8,** 558–574 (2005).
4. Morrison, A. C., Zielinski-Gutierrez, E., Scott, T. W. & Rosenberg, R. Defining challenges and proposing solutions for control of the virus vector *Aedes aegypti. PLoS Med.* **5,** e68 (2008).
5. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496,** 504–507 (2013).
6. 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).
7. Staples, J. E. & Fischer, M. Chikungunya virus in the Americas—what a vectorborne pathogen can do. *N. Engl. J. Med.* **371,** 887–889 (2014).
8. Sharp, T. M. *et al.* Chikungunya Cases Identified Through Passive Surveillance and Household Investigations—Puerto Rico, May 5—August 12, 2014. *MMWR Morb Mortal Wkly Rep.* **63,** 500–501 (2014).
9. Powers, A. M. Chikungunya virus control: is a vaccine on the horizon? *Lancet* **384,** 2008–2009 (2014).
10. Schaffner, F., Medlock, J. M. & Van Bortel, W. Public health significance of invasive mosquitoes in Europe. *Clin. Microbiol. Infect.* **19,** 685–692 (2013).
11. Garske, T. *et al.* Yellow Fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. *PLoS Med.* **11,** e1001638 (2014).
12. Powell, J. R. & Tabachnick, W. J. History of domestication and spread of *Aedes aegypti*—a review. *Mem. Inst. Oswaldo Cruz* **108** (Suppl.), 11–17 (2013).
13. Delatte, H. *et al.* Geographic distribution and developmental sites of *Aedes albopictus (Diptera: Culicidae)* during a Chikungunya epidemic event. *Vector borne zoonotic Dis.* **8,** 25–34 (2008).
14. Zouache, K. *et al.* Three-way interactions between mosquito population, viral strain and temperature underlying chikungunya virus transmission potential. *Proc. R. Soc. Biol. Sci.* **281,** 20141078 (2014).
15. Gubler, D. J. Prevention and Control of *Aedes aegypti*-borne Diseases: Lesson Learned from Past Successes and Failures. *AsPac J. Mol. Biol. Biotechnol.* **19,** 111–114 (2011).
16. Schaffner, F. & Mathis, A. Dengue and dengue vectors in the WHO European region: past, present, and scenarios for the future. *Lancet. Infect. Dis.* **14,** 1271–1280 (2014).
17. Honório, N. A. *et al.* Dispersal of *Aedes aegypti* and *Aedes albopictus (Diptera: Culicidae)* in an urban endemic dengue area in the State of Rio de Janeiro, Brazil. *Mem. Inst. Oswaldo Cruz* **98,** 191–198 (2003).
18. Schaffner, F. *et al.* Development of guidelines for the surveillance of invasive mosquitoes in Europe. *Parasit. Vectors* **6,** 209 (2013).
19. Schaffner, F., Kaufmann, C., Pflüger, V. & Mathis, A. Rapid protein profiling facilitates surveillance of invasive mosquito species. *Parasit. Vectors* **7,** 142 (2014).
20. Carvalho, R. G., Lourenço-de-Oliveira, R. & Braga, I. A. Updating the geographical distribution and frequency of *Aedes albopictus* in Brazil with remarks regarding its range in the Americas. *Mem. Inst. Oswaldo Cruz* **109,** 787–796 (2014).
21. Coelho, G. E. Challenges in the control of *Aedes aegypti. Rev. Inst. Med. Trop. Sao Paolo* **54,** S13–S14 (2012).
22. Food and Agricultural Organization of the United Nations. The Global Administrative Unit Layers (GAUL): Technical Aspects. *Food Agric. Organ. United Nations, EC-FAO Food Secur. Program.* (2008).
23. Brady, O. J. *et al.* Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasit. Vectors* **7,** 338 (2014).
24. 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).
25. Tatem, A. J., Hay, S. I. & Rogers, D. J. Global traffic and disease vector dispersal. *Proc. Natl. Acad. Sci. USA* **103,** 6242–6247 (2006).
26. Hawley, W. A., Reiter, P., Copeland, R. S., Pumpuni, C. B. & Craig, G. B. *Aedes albopictus* in North America: probable introduction in used tires from northern Asia. *Science* **236,** 1114–1116 (1987).
27. Hofhuis, A. *et al.* The hidden passenger of lucky bamboo: do imported *Aedes albopictus* mosquitoes cause dengue virus transmission in the Netherlands? *Vector borne zoonotic Dis* **9,** 217–220 (2009).
28. Semenza, J. C. *et al.* International dispersal of dengue through air travel: importation risk for Europe. *PLoS Negl. Trop. Dis.* **8,** e3278 (2014).
29. Morens, D. M. & Fauci, A. S. Chikungunya at the door—deja vu all over again? *N. Engl. J. Med.* **371,** 885–887 (2014).
30. Khan, K. *et al.* Assessing the origin of and potential for international spread of Chikungunya virus from the Caribbean. *PLOS Curr. Outbreaks* Jun 6, doi: 10.1371/currents.outbreaks.2134a0a7b"f37fd8d388181539fea2da5 (2014).
31. Phillips, S. J. *et al.* Sample selection bias and presence-only distribution model: implications for background and pseudo-absence data. *Ecol. Appl.* **19,** 181–197 (2009).
32. Kraemer, M. *et al.* The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus. eLife.* **4,** e08347 (2015).

Data Citation

1. Kraemer, M. U. G. *et al. Dryad Digital Repository* http://dx.doi.org/10.5061/dryad.47v3c (2015).

Acknowledgements

Author Contributions

M.U.G.K. drafted the manuscript with editorial input from J.P.M. and S.I.H. and approval from all authors. M.E.S. coordinated and compiled the data collection. K.A.D., A.M., M.U.G.K. and F.M.S. compiled the data records. M.U.G.K. performed database standardisation and technical validation. C.M.B., C.G.M., R.G.C., G.E.C., I.R.F.E., H.J.T., W.V.B., F.S., G.H., O.J.B. and G.R.W.W. provided additional data and geo-positioning. S.I.H. conceived the database design and advised on standardisation and validation procedures.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Kraemer, M. U. G. *et al.* The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence. *Sci. Data* 2:150035 doi: 10.1038/sdata.2015.35 (2015).