Creating a National Vector Surveillance System: Integrated mosquito trap data and digital epidemiology

Samuel S.C. Rund[1,2,3] (Samuel.Rund@ed.ac.uk) and
Micaela Elvira Martinez[1,2,4] (micaelam@princeton.edu)

[1]These authors contributed equally to this article.
[2]These authors were co–principal investigators.

[3]Centre for Immunology, Infection and Evolution, University of Edinburgh, Edinburgh, Scotland, UK

[4]Princeton University, New Jersey, USA

Major Subject Areas

- Ecology
- Epidemiology and Global Health

Abstract

Fluctuating population dynamics and shifting ranges, behavior, and phenology make monitoring of mosquito populations essential in controlling emerging pathogens. Despite ~1000 US mosquito control agencies, there is no centralized collation of their data. We provide a roadmap for the creation of a National Vector Surveillance System, for mosquito control agencies to routinely report standardized data. We characterized the extent of current monitoring, and collated mosquito abundance between 2009-2016. Despite a minority of agencies publically reporting data, our data set consists of records on >12 million mosquitoes. We demonstrate the

utility of a National Vector Surveillance System by providing novel insight into nationwide mosquito ecology and show that Digital Epidemiology could provide indirect mosquito monitoring, with Google queries for "mosquito" covarying with mosquito abundance. We demonstrate a National Vector Surveillance System would provide a cost effective means to combat vector borne disease emergence and can be integrated with wide-scale Digital Epidemiology.

Introduction

The emergence of arboviruses ("arthropod-borne viruses") —dengue virus (DENV), West Nile Virus (WNV), chikungunya virus (CHIKV), and Zika virus (ZIKV)—is a global problem[1]. There are multiple hypotheses regarding the epidemic emergence of arboviruses. For example, increasing globalization and climate change could be driving geographic expansion of vectors and increased transmission from vectors to humans[2,3]. The occurrence of an arbovirus epidemic requires the presence of vectors in sufficient abundance to sustain transmission. Mosquito monitoring has, therefore, long been recognized as an important component of disease surveillance and control efforts[4].

Large scale surveillance efforts have provided insight into the distribution of disease vectors worldwide, including Aedes, Anopheles, and Culex mosquitoes which are primary vectors for DENV, malaria, and WNV, respectively[5–7]. Maps of vector geographic distribution (*i.e.*, presence/absence) are often used to assess disease risk[2,5,8]. Most recently, vector mapping has been used in the US to infer risk of ZIKV transmission[9]. It was predicted that the risk of ZIKV transmission in the US is relatively localized to southern states. As predicted[9,10], ZIKV transmission established in Florida, specifically, Miami-Dade County. As of October 12, 2016 there have been

128 reports of locally-acquired ZIKV cases[11]. Given the public health threat of emerging (*e.g.*, ZIKV, CHIKV) and endemic (*e.g.*, WNV) arboviruses in the US, there is great incentive to invest in studies of vector distribution and abundance.

In the US, mosquito control and surveillance efforts are ongoing at state- and local-levels[12]. Public health and environmental agencies are tasked with (1) systematic monitoring of vector abundance and mosquito infection status (*i.e.,* using adult mosquito trapping and mosquito larval pool testing for viral presence), and (2) reducing disease transmission risk by controlling vector population size using larvicide, larval-site management, and/or insecticidal fogging[13]. Many agencies have active surveillance programs that entail daily, weekly, and/or monthly trapping and counting of mosquitoes, with taxonomic identification of the genus or species level in many instances. Some vector surveillance programs were started in the 1940s at US military installations to combat malaria transmission, and have since continued[4,14,15], resulting in the generation of long-term ecological data. Time series data of vector abundance that result from such long-term monitoring are invaluable for the study of vector population dynamics and quantifying the risk vectors pose to human health.

There are ~1000 mosquito control agencies in the US, with a mosquito control agency broadly defined as the local government authority responsible for the surveillance and control of mosquitoes. Surprisingly, the exact number of agencies is unknown, (*personal communication*, Joseph M Conlon, Technical Advisor, American Mosquito Control Association). Mosquito control responsibilities may lie with the local health department, or separate entities such as mosquito abatement districts. The trapping and identification of mosquitoes requires considerable economic investment in the form of personnel, equipment, and infrastructure. Despite these efforts, there is no centralized reporting or repository in the US to facilitate the integration of such publicly funded data. At present, in the US, the majority of mosquito surveillance data remain isolated at local and/or state agencies and are inaccessible to the wider mosquito surveillance and research communities[3,16].

In the US, mosquito-borne diseases are nationally notifiable with state-level reporting to the CDC National Notifiable Disease Surveillance System[17]. We propose a parallel system be created for mosquito population surveillance that is: standardized, centralized, and regular. More specifically, data of a pre-specified type (standardized) should be reported to a centralized system (*e.g.*, such as one hosted by the CDC, PAHO, or WHO), and reporting should be done at regular intervals (weekly or monthly). Having such a surveillance system in place would not only facilitate the creation of Big Data on mosquito populations, but would also serve as a repository for the wealth of surveillance data that exist to date. The creation and deployment of a National Vector Surveillance System would be a relatively simple and low cost means of establishing the risk of current and future arbovirus threats, and inform future ecological and epidemiological studies. As proof of principle, here we summarize the breath of mosquito surveillance efforts in the US, curate the limited data currently accessible from mosquito control boards, and analyze these data to provide a glimpse into vector population ecology, highlighting the potential power of a National Vector Surveillance System data repository.

Results

Mosquito Abundance Data

We compiled a list and identified a web presence for 997 US mosquito control agencies in 48 US states (Figure 1a, Supplemental File 6). We found 91 agencies

(9%) in 22 states provided historical data for at least one year in 2009-2015. Further, 54 agencies (5%) in 18 states provided live data (*i.e.*, 2016 data *reported* at least monthly), which are indications of present and ongoing efforts to generate and make abundance data publicly available online. 50 agencies (5%) provide both live and historical data. While long, multi-year observations in the data were scarce, a total of 95 agencies provide live and/or historical data (Supplemental File 6).

We collated data from a subset of these agencies (76 of 91 agencies) where data was presented in a mineable format such as online tables, graphs, and GIS maps where population abundance could be followed over time. All data were limited to what was publicly available online. The full data set can be found in Supplemental File 1. Our final data set contains 144,234 records covering mosquito trapping data from 2009-2015 (as well as a further 5235 records from 2016). We standardized dates and taxonomic identifiers. The data covers 551 location identifiers, some trap level, some a geographic feature (*e.g.,* a particular park name), and some a whole jurisdiction or multi-agency region (*e.g.,* 'Northwestern Rural'). Our final data set represented over 12 million individual mosquitoes counted (some data were provided as averages, this is a minimum estimate of total mosquitoes trapped). From 2009-2015, approximately 80% of the mosquitoes have genus level identification, and ~11% have species level identification.

Temporal resolution of the data was variable, with some data reported as weekly, some monthly, and some a combination of the two. We consistently found a lack of continuous sampling throughout the year, as surveillance efforts were targeted to seasonal windows that varied between agencies. This frequently resulted in the onset and/or offsets of the mosquito season being unobserved.

Taxonomic identification and reporting varied between agencies. While we were able to standardize subtle differences in taxonomic reporting (*e.g., Aedes* vs. *Ochlerotatus* or *Aedes* vs. *Ae.* there are some differences in reporting, however, that could not be standardized. For example, in some districts, taxonomic descriptors such as 'summer floodwater', 'other', 'non-culex', and 'nuisance species', were used to identify at least some mosquitoes. However, we note that 56 agencies report species level identification of at least some of the mosquitoes they collected. This reflects local knowledge of the important vector and nuisance mosquito species in a given control agency's jurisdiction.

Surveillance System

Reviewing the data across the many agencies provided insights into how a nation-wide surveillance program could be designed. First, states could coordinate with control agencies to place sentinel sites for trapping with particular emphasis on ecological conditions in terms of urban/rural or climate zones within a state. New Jersey, for example, divides their state into ten ecologically relevant zones[18]. The traps in a given location should be of fixed type and attractant, and these traps should ideally be checked at least weekly and reported weekly during the mosquito season. We propose continuous trapping throughout the year be performed. This could be as simple as adding monthly trapping in the winter to characterize the onset of mosquito season more precisely. Given climate projections of earlier springs in the northern hemisphere[2,3], there is a need to test for phenology shifts in disease vectors.

The data, units, traps, and metadata associated with the records varied widely between agencies in our study. We do not suggest significant changes in operational protocols, but it is important that agencies report full metadata, including trapping strategy so their data can be interpreted. The most useable data will include trap location name and/or GPS location, attractants used, the date the trap was counted, and the duration of time the trap was set/reset. Finally, it is critically important that the unit used for reporting mosquito count is standardized. Control boards should submit trap-level data, and not aggregations over areas, and it must be made clear if the data reported is male and female mosquitoes, or only female mosquitoes.

We found that some agencies attempted to identify each mosquito collected to the species level. This effort is certain to yield valuable insights, but it is not feasible for nation-wide standardized surveillance. Most districts, however, did identify mosquitoes to the genus level. Due to their status as viral vectors of ZIKV, CHIKV, and DENV, counts of *Ae. aegypti* and *Ae. albopictus* should always be reported. Importantly, their absence should also be recorded. A similar system would be desirable, but likely not feasible on a national scale, for members of the morphologically similar invasive *Culex pipiens/quinquefasciatus* group which vector West Nile and St. Louis Encephalitis viruses

Reporting of collected data would be submitted electronically, in a standardized spreadsheet or online web form. We provide a mock-up reporting form (Figure 1b) to illustrate the format of clean, unambiguous data that is ideal for analyzing across agencies and states. Whenever possible and feasible, corresponding historical data should also be represented in the database/national surveillance system.

Mosquito Seasonality

Due to the large amount of variation in the unstandardized data reported from mosquito control agencies, we did not analyze the full data set. Instead, we searched the compiled data for time series that would allow us to highlight the types of analyses that could be done if there were a standardized national surveillance system. Specifically, a national surveillance system would allow researchers to study (1) geographic variation in mosquito populations, (2) interspecific/inter-genus variation in phenology, and (3) interannual variation in mosquito abundance.

For any given taxonomic group, seasonal phenology may vary geographically due to variation in environmental conditions. We searched the compiled data for a taxonomic group for which we had 3+ years of data from multiple states. Data from *Culex pipiens* in Minnesota, Iowa, and California fit these criteria and were of particular interest because *C. pipiens* is the vector of West Nile Virus. We found that the *C. pipiens* season was restricted to a narrow time frame (late summer) in Minnesota (Figure 2a, d), the most northern of the three states. In Iowa and California, the *C. pipiens* season extended later into the fall (Figure 2b-c, e-f), which might be due to more mild autumnal weather in Iowa and California, relative to Minnesota. Interestingly, in California the *C. pipiens* season began in the early spring. This could be due to springtime environmental conditions being favorable throughout the state. California, however, spans a large latitudinal range, and the springtime presence of *C. pipiens* may be localized within the state, but by aggregating data from two districts in California, we may have masked geographic variation within the state.

The benefit of genus or species level identification within a state is that it will allow mosquito control agencies to identify differences in phenology among different genus/species. The Culex genus was of interest because of the public health risk posed by *C. pipiens*' transmission of WNV. We therefore searched the compiled data for a state with species identification of Culex mosquitoes. Minnesota reported data for *C. pipiens*, *C. restuans*, and *C. tarsalis*. We found that the seasonal timing of peak abundance varied for these species, with *C. tarsalis* having the earliest peak, *C. restuans* intermediate, and *C. pipiens* the latest (

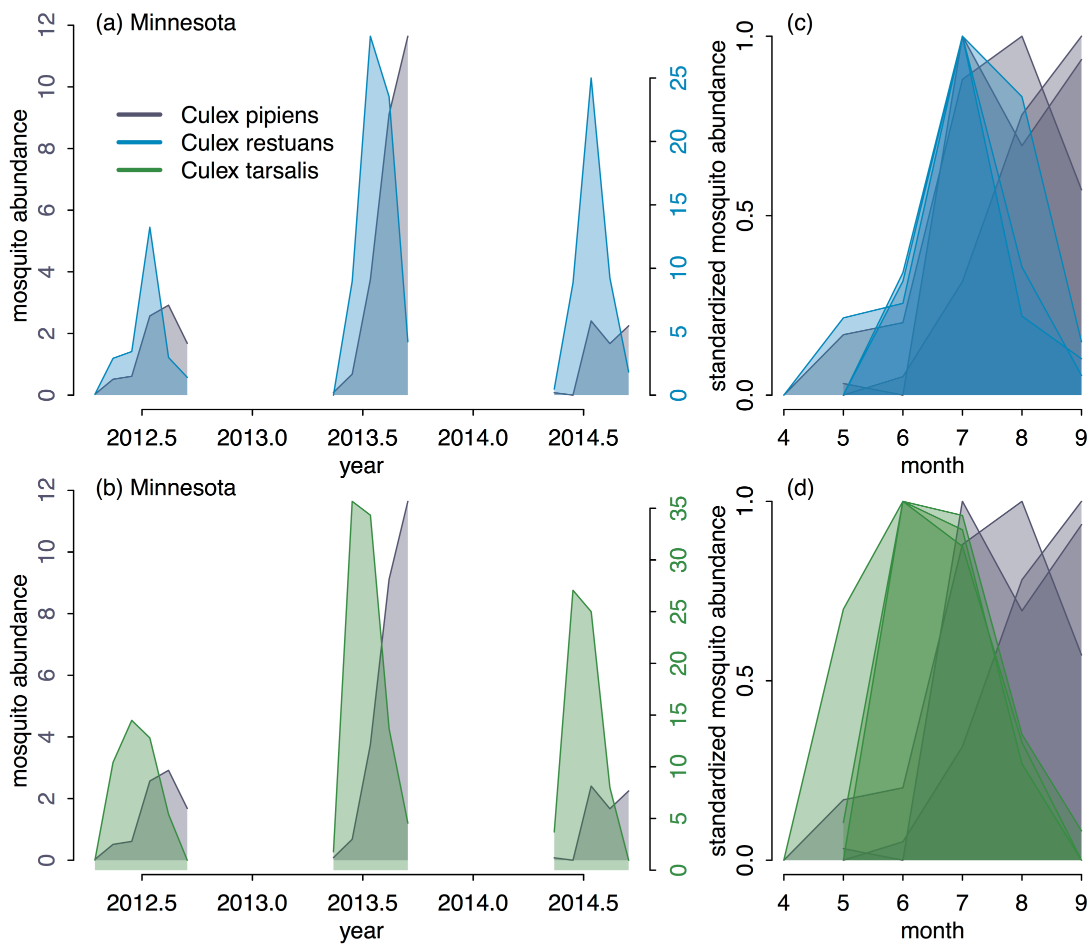


Figure 3). The seasonal staggering between taxonomic groups is important from a public health perspective, because, if this pattern is consistent year-upon-year and applies across states, then *C. tarselis* could be used as an early warning of the *C. pipiens* season.

Continuous (throughout the year), long term times series of mosquito abundance, such as would be generated by a National Vector Surveillance System, will allow the research community to not only characterize mosquito seasonality, but also study variation in mosquito abundance at larger temporal scales. California provided continuous data for *C. pipiens* for 2011-2013. This time series, although it only represents three years, revealed that *C. pipiens* abundance can vary greatly between years. This demonstrates that mosquito abundance can be characterized on two temporal scales (i) seasonal, characterized by the mosquito season vs. the off-season, and (ii) interannual, characterized by high abundance years vs. low years (Figure 4a). High abundance years might be tied to climate conditions or other environmental variables. When continuous long-term time series are available, they can be combined with data on environmental or other ecological variables, and population dynamic models, to test for mechanistic drivers of variation in abundance. A mechanistic understanding of vector population dynamics could allow for the forecasting of high abundance years and allow for interventions to mitigate vector borne diseases. Importantly, the data from California highlight that vector abundance does not necessarily translate to higher disease risk. Although the abundance of female *C. pipiens* was higher in 2011 than 2012 or 2013, there were fewer human WNV cases in California in 2011 than in 2012 or 2013. There were 158, 479, 379 human cases in 2011, 2012, and 2013, respectively[19]. To study disease risk, mosquito population dynamics would need to be factored into disease transmission models[20].

Google Queries as a Proxy for Mosquito Abundance

For the states for which we had both mosquito trap data and Google Trends data during the same time period, we found that the abundance of Google queries covaried with direct measures of mosquito abundance. Google queries for "mosquito" in New York showed the same seasonal cycle (*i.e.*, high mosquito abundance in the summer) as that observed in trap data (Figure 4b). Not only did the Google query data capture seasonal variation in mosquito abundance, but in Illinois, a state with low Culex abundance in 2011 and 2012, followed by high abundance in 2013, 2014, and 2015, Google queries captured both the seasonal variation in abundance and the variation in abundance in high years vs. low years (Figure 4c). These data suggest digital epidemiology can be used to supplement on-the-ground mosquito surveillance, help inform early warning systems, and afford more surveillance coverage of the population. We submit, however, that more mosquito trap data are needed in order to perform a formal statistical evaluation of the digital epidemiology for mosquito surveillance.

As the Google query data covaried with mosquito trap data in the US, we therefore used Google query data of the language-specific search term "mosquito" as an indirect measure of mosquito abundance in 17 countries, representing 39 locations with sufficient data for testing seasonality (Figure **5**). There were multiple Google trends time series for some countries with high search volume. Of the 39 locations, wavelet analysis detected significant seasonality in 38 of them, with 2 countries (Thailand and Taiwan) having a biannual signature. For those locations with significant seasonality, GAMM model predictions demonstrated that the seasonal variation in mosquito abundance was geographically structured.

Discussion

CDC guidelines call for standardized and repeated collection of mosquito vectors[13] – a point recently reaffirmed by others[21,22]. We argue that nation-wide aggregation and sharing of standardized data would allow the research community to better understand the ecology of vectors, and this in turn would inform disease control efforts. Clearly many states and agencies recognize the need to generate and disseminate mosquito abundance data, as some states already operate repositories[18,23,24]. We identified 49 agencies that provided abundance data on their websites, independent of a state repository, indicating these agencies recognize the importance of open data.

By collating publicly available data from 76 mosquito control agencies in the US, we provide an extensive set of high temporal resolution mosquito abundance data, comprised of >144,000 records. Importantly, these data represent only a small fraction of the data that exist. The majority of mosquito abundance data are not readily available to the scientific community. We demonstrate that mosquito abundance data, collated from multiple data-silos, can give insight into the spatiotemporal ecology of disease vectors. As we discovered, there is currently a great deal of variation in how mosquito trap data are collected and reported. The lack of data standardization presents hurdles to data integration. The mosquito control community, however, could come together to decide upon the most efficient and effective way to collect and report data in a standardized manner that would facilitate real-time data release and data integration. Given the need to respond to emerging mosquito-borne diseases in the Americas, the integration of mosquito abundance data should be made a priority.

Enormous human effort and public expense has gone into the collection and identification of the more than 12 million mosquitoes in our data set. Since the generation of such data is publicly funded, future and past data should be made publicly available in the spirit of The White House's Open Government Initiative[25,26]. A central repository is the most efficient way to promote data reuse and accessibility of publicly funded data[16] across agencies/states - or even international and language boundaries. Examples of similar interjurisdictional databases for the reporting and dissemination of disease incidence include ArboNET[27] (US arbovirus incidence), the CDC National Notifiable Disease Surveillance System[17], and the PAHO database[28] (dengue surveillance in the Americas).

A significant hurdle to data integration is that the units of reporting varied among agencies. As to be expected, agencies deploy a variety of traps, with a variety of attractants. Comparing absolute mosquito trap numbers between localities is unlikely to be useful to researchers or control professionals – it is the relative change in abundance that is of interest, regardless of units. However, successful interpretation of species composition is dependent on knowing trapping methodology – as traps have species bias[29].

A significant shortfall we identified is the lack of historical reporting of the important invasive disease vectors, *Ae. aegypti* and *Ae. albopictus*. Only Iowa and New Jersey report live *Ae. albopictus* abundance data, and no state reported live *Ae. aegypti* abundance data. Further, it is often unclear if "0" or no data represents true zeros (surveillance was performed, with zero members of the genus/species trapped) versus a lack of specific surveillance for these critical species. Reliable presence/absence data are necessary for distribution maps and modeling efforts[3,21,22].

We only used data generated by US mosquito control agencies. Substantial data, primarily historical species surveillance from US military installations going back to 1947, are held by VectorMap[15]. Further, NSF funded NEON projects sites are coming online[30]. These twenty nationwide sites will record highly standardized, year-round mosquito surveillance, identifications, and virus-status[30]. Our proposed surveillance system, along with other research based data sets and international governmental and non-governmental data sources are promising data sources for integration.

Considering the caveats of our collated data, and the fact that it represents only a small fraction of the mosquito abundance data that exists, we were nevertheless able to uncover spatiotemporal patterns of mosquito populations that are relevant to public health. This is enticing evidence that the data generated by centralized, regular, standardized surveillance would be a powerful method for anticipating and responding to arboviral threats. High temporal resolution time series data (weekly/monthly) are powerful tools. As we have highlighted with our analyses, such time series would allow for the characterization of spatial and temporal mosquito population ecology. Characterizing mosquito population cycles and understanding the mechanisms generating seasonal and interannual variation would enable us to:

(1) Infer when to anticipate epidemics/outbreaks;

(2) Quantify geographic variation in risk (how bad can epidemics become), especially at geographic margins of vectors;

(3) Monitor virus persistence in source and sink populations and determine if vectors overwinter in sufficient numbers to sustain transmission;

(4) Identify windows of opportunity for spraying, larval site management, ramping up surveillance for pathogen (testing for virus), public outreach campaigns;

(5) Study vector-borne disease transmission dynamics; and

(6) Coordinate responses across states/counties, with the goal of controlling epidemics and creating early warning systems.

Having rich data sets of mosquito abundance over time will allow for analyses that have not yet been feasible, and the combination of data from multiple sources would allow for analyses that can only be done in aggregate[31]. Modelers, for example, require high quality data for model training and validation[3,22]. Data for aggregation, need not only be mosquito abundance data, but truly new and powerful analyses would be possible if abundance time series are combined with other sources of Big Data. These include arbovirus disease incidence data, mosquito pool testing for virus (ArboNET[27]); land use data (National Land Cover Database[32]); meteorological data (Land Data Assimilation Systems, LDAS[33]), and climate change predictions (NCAR-CCSM 4.0[34]). The data could be used to motivate action when neighboring communities have vectors of concern[22], and tools could be developed to assist local control agencies, such as the automated generation of up-to-date disease risk maps, and other publicly disseminated reports. Finally, the data could be integrated into biodiversity databases such as the Global Biodiversity Information Facility (GBIF)[31,35].

The problems and opportunities we addressed here are not unique to the United States. Emerging arboviruses are a global problem. Here we used the United States as an operational example, but the data curation and application applies worldwide. Furthermore, this is the appropriate time for the US and other countries to initiate country-wide mosquito surveillance efforts due to the recent emergence of ZIKV and CHIKV, and the continued threat that DENV poses in >100 countries across six continents, causing 96 million cases per year[36]. The United States is currently allocating resources for emerging vector-borne diseases and the Americas are battling the ongoing Zika outbreak. We propose standardized reporting by existing mosquito surveillance programs would be a low cost but powerful means of empowering control efforts for vector borne diseases.

Methods

Mosquito Abundance Data

We systematically searched online to identify mosquito control agencies within the US. Specifically, we used Google searches, the American Mosquito Control Association website[37], state government websites (*e.g.*, Florida Department of Agriculture & Consumer Services[38]), and state mosquito control association web pages (*e.g.,* Texas, Michigan, and California[39–41]) to find the names/locations of mosquito control agencies. We then supplemented our list of mosquito control agencies by cross-referencing it with lists of mosquito control agencies generated by other researchers. For each agency identified (Supplemental File 6), we searched for a web presence, either an agency website or agency representation in a state-wide mosquito abundance repository (*e.g.,* such as those that exist for Iowa, New Jersey, and South Dakota[18,23,24]). The mosquito control agency websites/repositories were searched for the presence of mosquito abundance surveillance data at a temporal resolution of daily, weekly, and/or monthly from 2009 to mid-2016. We used data from fixed traps. Data from temporary traps were excluded because they cannot be adequately used to reconstruct time series of abundance. We collated data from 76 of the 91 agencies presented in a mineable format such as online tables, graphs, and GIS maps where population abundance could be followed over time. All data were limited to what was publically available online.

Mosquito Abundance Data Analysis

In order to obtain state-level time series of mosquito abundance, the mosquito trap data were aggregated within each state. We worked with states with data that spanned multiple years: California, Colorado, Florida, Illinois, Iowa, Massachusetts, Michigan, Minnesota, New Jersey, New York, and Ohio. The data were subset by state, taxonomic group (*i.e.*, genus, with the exception of Culex which was analyzed to the species-level), temporal resolution (*i.e.*, weekly or monthly data), and unit of measure. The data were then aggregated within each subset by either summing or averaging—depending on which was appropriate for each data type—across traps collected at the same time. Data aggregation resulted in time series of mosquito abundance for each combination of state and taxonomic group represented in our data. The taxonomic groups represented in the aggregated data included *Culex*, *Culiseta*, *Aedes*, *Anopheles*, *Mansonia*, *Psorophora*, *Coquillettidia*, *Uranotaenia*, and

Orthopodomyia. All aggregated data can be found in the Supplemental File 2. We analyzed the time series for seasonal fluctuations in mosquito abundance and characterized differences in seasonal phenology across states and taxonomic groups.

The unit of measure for reporting mosquito abundance differed among mosquito control boards. Some mosquito control boards reported total mosquitoes trapped, while others reported females per trap night, or mean females per trap. When comparing seasonality across time series, we standardized the data so each time series took values from 0-1. For time series observation, $y_{i,j}$, observation i in year j, the following was used to standardized the data $s_{i,j} = (y_{i,j} - \min(y_j))/(\max(y_j)-\min(y_j))$.

Digital Epidemiology

Google Trends[42] were downloaded for queries of "mosquito" in the US, as well as language-specific queries of "mosquito" in other countries (Supplemental Files 3-5 contain the data and search term used for each country). Google Trends data were downloaded for countries in each hemisphere, for all years for which data were available, typically 2004-2015/2016, depending on the date of download. Google Trends data represent the relative abundance of Google queries reported as values from 0-100. In order to test for seasonality in Google queries of mosquito, we ran a wavelet analysis on each time series, using the R package biwavelet[43]. Google Trends time series with a significant 1-year period (significant when tested against time series with the same lag-1 autocorrelation) were $\log_{10}$ transformed and detrended to make them more sinusoidal and remove long term trends (which we assumed were due to changes in internet query habits or Google algorithms, rather than changes in mosquito ecology). Time series were standardized to values between 0-1 and a GAMM model was fit to each standardized time series with the explanatory variable being the time of the year. Predictions from the fitted GAMM models were used to characterize mosquito seasonality on a global scale.

Acknowledgements

This work would not be possible without the efforts of the 95 local mosquito control agencies highlighted in Supplemental File 6 who collected mosquito population abundance data and made their data available online, as well as Iowa State University Entomology Department, the Rutgers Center for Vector Biology, and South Dakota State University for their state-level repositories[18,23,24].

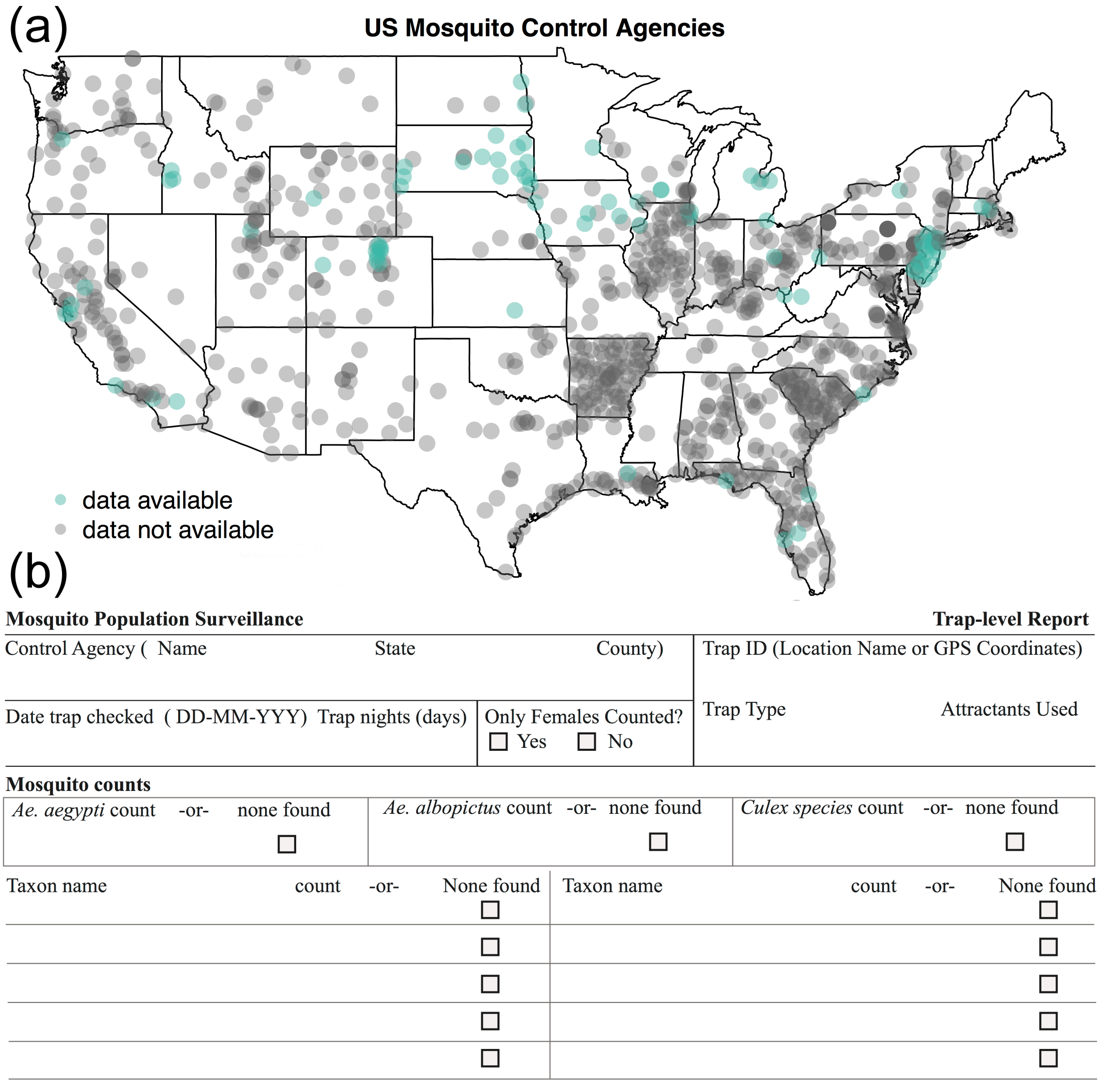
We thank Frank Collins, Cynthia Lord, Paul Hickner, Lauren Cator, Luke McNally, and Ryan Smith for their helpful discussions during the development of this project.

We are incredibly grateful for Nathan Nixon, Ben Weise, Ben Chan, Jason Cleland, Ollie Lee, and Emma Stotter for their work in locating and digitizing the mosquito population abundance data.

Micaela Martinez was funded by a US National Science Foundation Postdoctoral Fellowship in Biology Award Number 1523757, and is currently funded by the NIH Director's Early Independence Award. Research reported in this publication was supported by the Office of the Director, National Institutes of Health, under Award Number DP5OD023 100. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Samuel S.C. Rund is funded by the Royal Society (NF140517) and a strategic award from the Wellcome Trust (No. 095831).

Figures



Mosquito Population Surveillance | **Trap-level Report**

Control Agency (Name | State | County) | Trap ID (Location Name or GPS Coordinates)

Date trap checked (DD-MM-YYY) | Trap nights (days) | Only Females Counted? ☐ Yes ☐ No | Trap Type | Attractants Used

Mosquito counts

Ae. aegypti count -or- none found	*Ae. albopictus* count -or- none found	*Culex species* count -or- none found
☐	☐	☐

Taxon name	count	-or-	None found	Taxon name	count	-or-	None found
			☐				☐
			☐				☐
			☐				☐
			☐				☐
			☐				☐

Figure 1: (a) Location of the 997 US mosquito control agencies with an identified web presence and the 91 agencies from which historical data, from any year between 2009 – 2015, were publicly available online. South Dakota, Iowa, and New Jersey had aggregated data available online. (b) Example report form that could be used to standardize reporting of mosquito trap data to a national surveillance system.

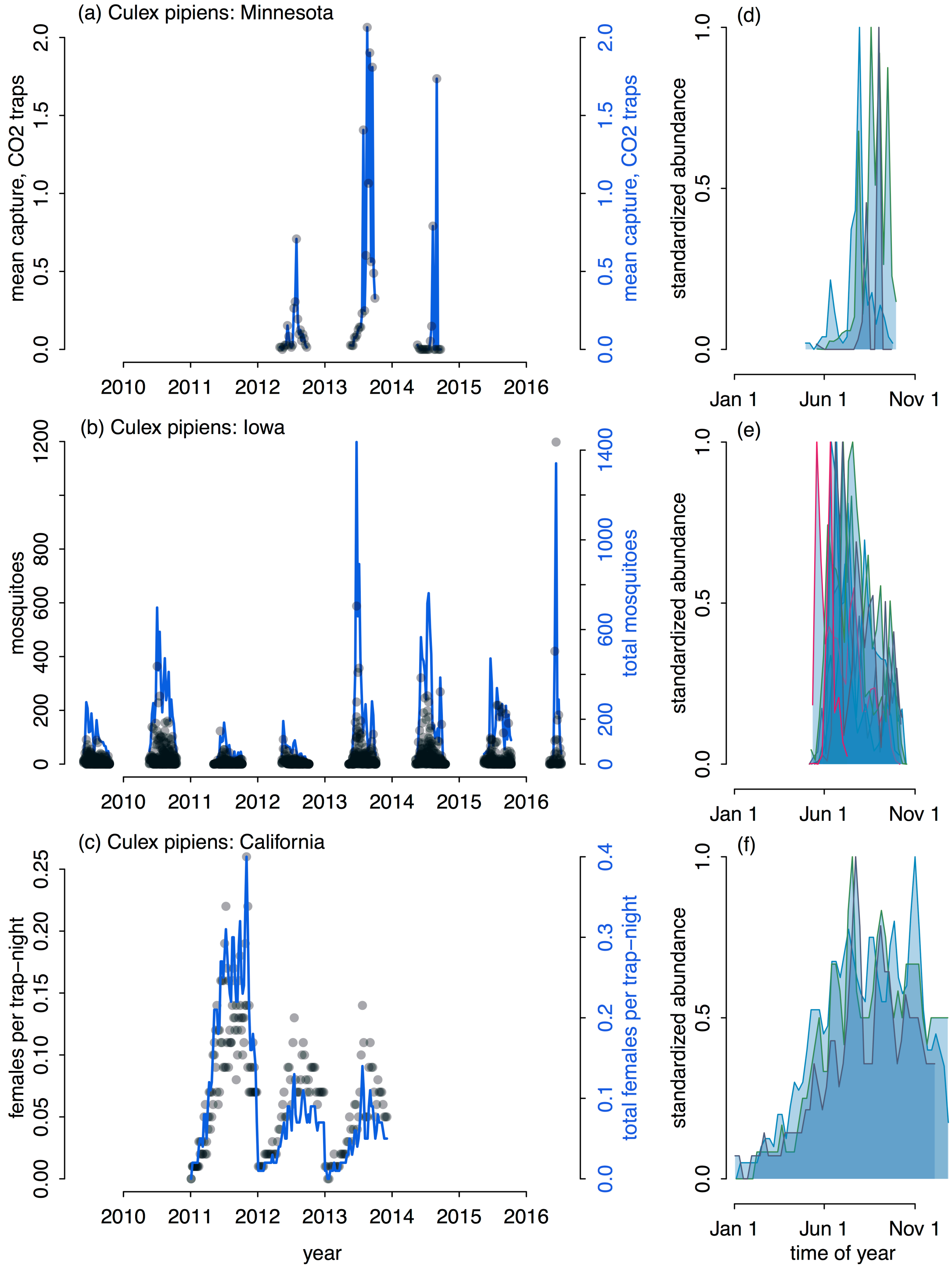


Figure 2: ***Culex pipiens*** **seasonality. (a-c) Black points show the raw data from each trap within the state and correspond to the left y-axis. Blue time series are aggregated data across traps within the state, corresponding to the right y-axis. (d-f) Standardized data corresponding to the time series in a-c. The standardized time series were plotted individually for each year to show how mosquito abundance varied by time-of-year and among states.**

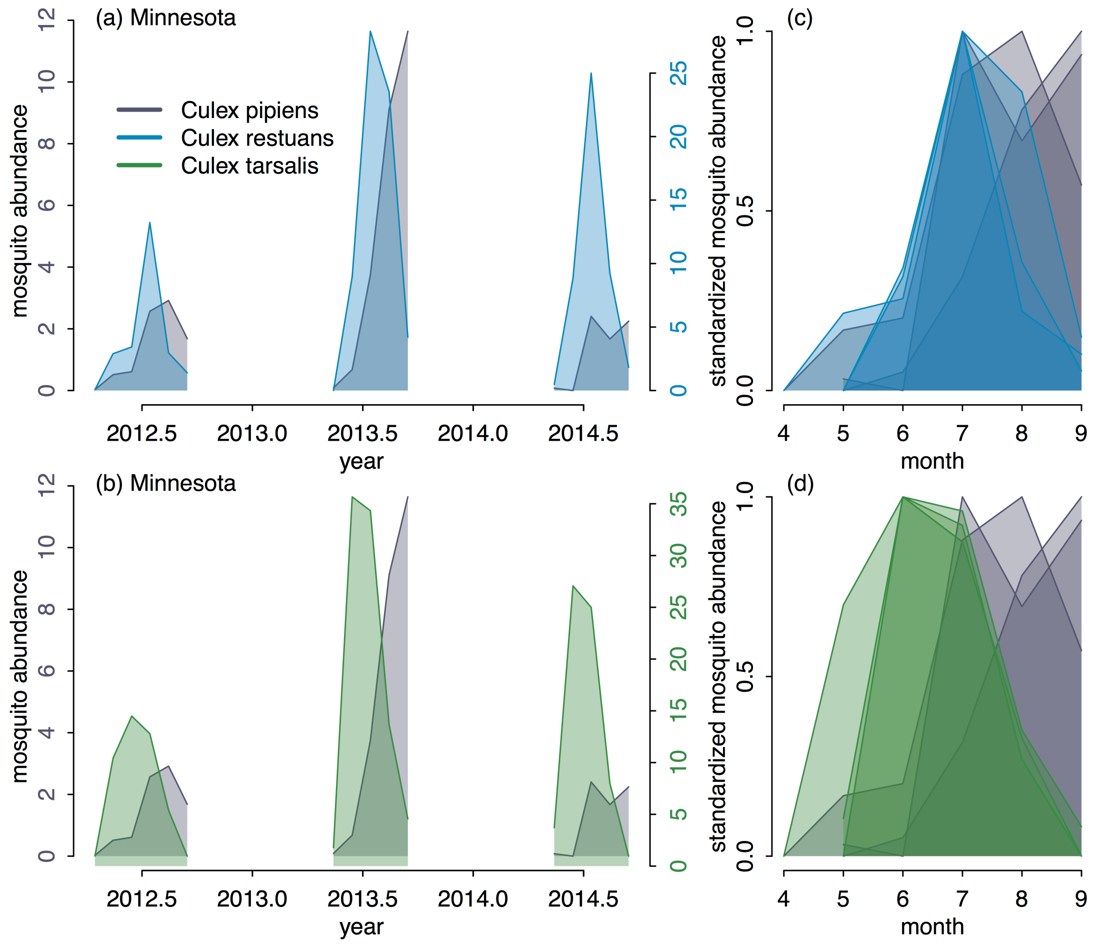


Figure 3: Variation in seasonal phenology among species. (a) *C. restuans* and *C. pipiens* abundance as reported in Minnesota. (b) *C. tarsalis* and *C. pipiens* abundance as reported in Minnesota. (c-d) Standardized data corresponding to the time series in a-b. The standardized time series are plotted individually for each year to show how mosquito abundance varied by month and among species.

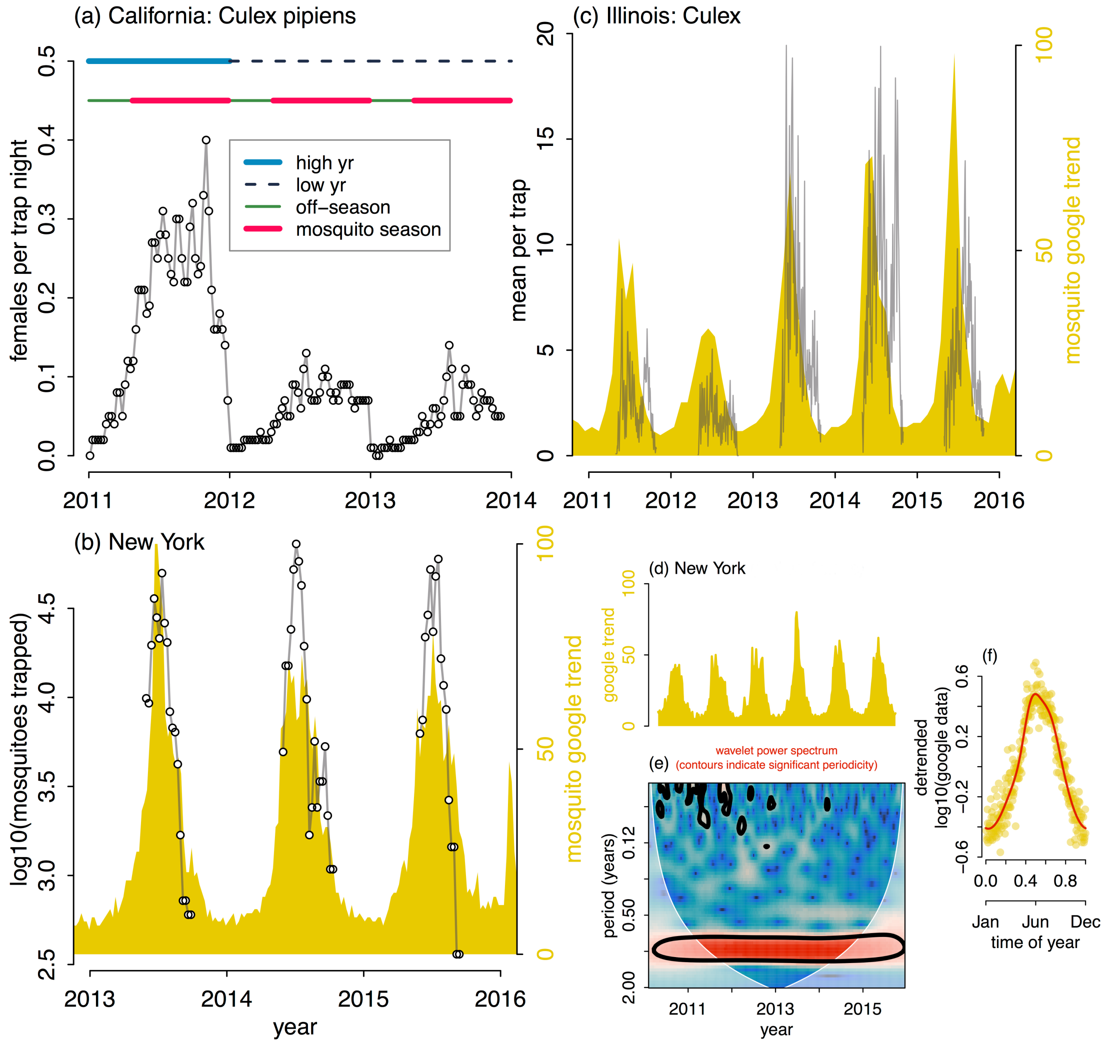
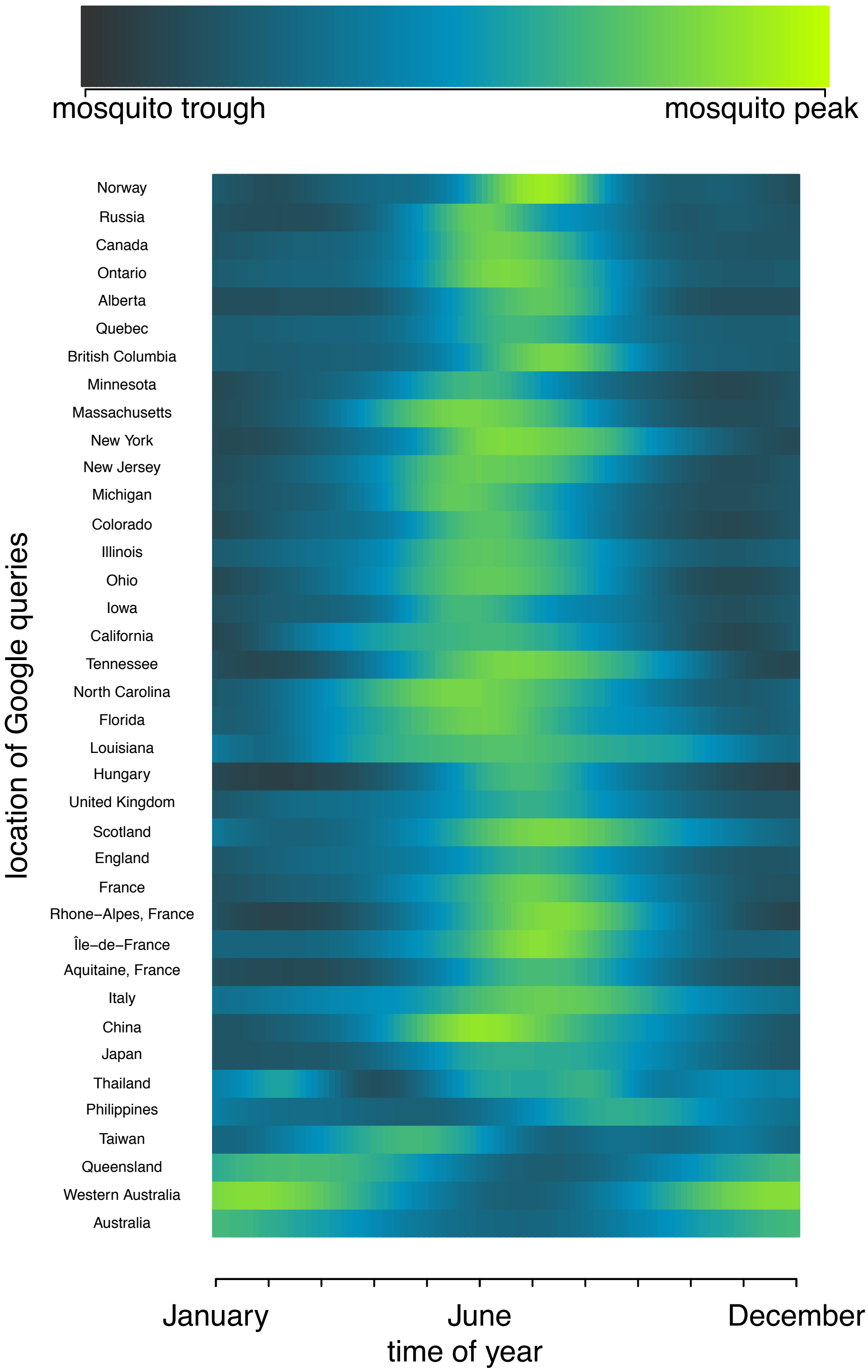


Figure 4: (a) Time series of female *C. pipiens* abundance in California. Top line above the time series indicates the high abundance year vs. the low abundance years. Bottom line above the time series indicates the mosquito season vs. the off season. (b-c) Mosquitoes trapped in New York state and Illinois (black time series corresponding to the left y-axis), along with Google queries for the search term "mosquito" in each state during the same time period (yellow time series corresponding to the right y-axis). (d) Google queries for "mosquito" from 2010-2015 used for a wavelet analysis with power spectrum shown in (e) where the red area indicates significant 1-year periodicity of the time series. (f) Yellow points indicate the Google queries from New York, detrended and $\log_{10}$ transformed. Red curve indicates the predicted seasonal abundance as predicted by a GAMM model with time-of-year as the independent variable.



466
467 **Figure 5: Mosquito seasonality around the world. The seasonal structure of mosquito abundance as**
468 **measured indirectly via Google Trends using the language-specific search term "mosquito". Each row of the**
469 **heatmap shows mosquito seasonality as predicted by the GAMM model for the location listed on the y-axis.**
470 **The x-axis is the time-of-year. Dark blue represents the mosquito off-season and the bright green indicates**
471 **the mosquito season.**

Supplementary Material

Supplemental File 1: CSV file containing compiled mosquito population data from 2009-2016. See Mosquito Abundance Data in the Methods section for more details.

Supplemental File 2: CSV file containing time series of mosquito abundance created by aggregating data across traps within each state. The data are from California, Colorado, Florida, Illinois, Massachusetts, Minnesota, New Jersey, Iowa, Michigan, New York, and Ohio.

Supplemental File 3: CSV file containing monthly Google Trends data from multiple countries.

Supplemental File 4: CSV file containing weekly Google Trends data from multiple countries.

Supplemental File 5: CSV file containing Google Trends data from US with multiple years of mosquito trap data in Supplemental File 1.

Supplemental File 6: CSV file containing identified mosquito control agencies. The name and state of each mosquito control agency is provided. We provide a representative postcode and GPS coordinate in the geographic vicinity of each agency. Each agency is assigned a unique ID (AgencyID) that corresponds with population data recorded in Supplemental File 1.

References

1. Fauci A, Morens D. Serving two masters. *N Engl J Med*. 2016;374(7):601-604. doi:10.1056/NEJMp1002530.
2. Proestos Y, Christophides GK, Ergüler K, Tanarhte M, Waldock J, Lelieveld J. Present and future projections of habitat suitability of the Asian tiger mosquito, a vector of viral pathogens, from global climate simulation. *Philos Trans R Soc Lond B Biol Sci*. 2015;370(1665):20130554-. doi:10.1098/rstb.2013.0554.
3. Campbell LP, Luther C, Moo-Llanes D, Ramsey JM, Danis-Lozano R, Peterson AT. Climate change influences on global distributions of dengue and chikungunya virus vectors. *Philos Trans R Soc B*. 2015;370:20140135-. doi:10.1098/rstb.2014.0135.
4. Carpenter SJ. Insect control at army training camps. *J Econ Entomol*. 1944;37(4):470-474. http://jee.oxfordjournals.org/content/37/4/470. Accessed October 19, 2016.
5. Hay SI, Sinka ME, Okara RM, et al. Developing global maps of the dominant Anopheles vectors of human malaria. *PLoS Med*. 2010;7(2):e1000209. doi:10.1371/journal.pmed.1000209.

6. Vazeille M, Mousson L, Rakatoarivony I, et al. Population genetic structure and competence as a vector for dengue type 2 virus of Aedes aegypti and Aedes albopictus from Madagascar. *Am J Trop Med Hyg*. 2001;65(5):491-497. http://www.ncbi.nlm.nih.gov/pubmed/11716103. Accessed October 19, 2016.
7. Turell MJ, Dohm DJ, Sardelis MR, O'guinn ML, Andreadis TG, Blow JA. An update on the potential of North American mosquitoes (Diptera: Culicidae) to transmit west Nile virus. *J Med Entomol*. 2005;42(1):57-62.
8. Kraemer MUG, Sinka ME, Duda KA, et al. The global distribution of the arbovirus vectors Aedes aegypti and Ae. Albopictus. *Elife*. 2015;4:e08347. doi:10.7554/eLife.08347.
9. Bogoch II, Brady OJ, Kraemer MUG, et al. Anticipating the international spread of Zika virus from Brazil. *Lancet*. 2016;387:335-336. doi:10.1016/S0140-6736(16)00080-5.
10. Messina JP, Kraemer MU, Brady OJ, et al. Mapping global environmental suitability for Zika virus. *Elife*. 2016;5:42-47. doi:10.7554/eLife.15272.
11. Centers for Disease Control and Prevention. Maps of Zika in the United States | Zika virus. http://www.cdc.gov/zika/intheus/maps-zika-us.html. Published 2016. Accessed October 19, 2016.
12. Council of state and territorial epidemiologists. *Assessment of Capacity in 2012 for the Surveillance, Prevention and Control of West Nile Virus and Other Mosquito-Borne Virus Infections in State and Large City/county Health Departments and How It Compares to 2004*.; 2014.
13. Centers for Disease Control and Prevention. *West Nile Virus in the United States: Guidelines for Surveillance, Prevention, and Control*.; 2013.
14. Foley DH, Maloney FAJ, Harrison FJ, Wilkerson RC, Rueda LM. Online spatial database of US Army Public Health Command Region-West 29 mosquito surveillance records: 1947-2009. *United States Army Med Dep J*. 2011;July-Sep:29-36. http://vectormap.si.edu/downloads/AMEDDJ.pdf. Accessed October 19, 2016.
15. The Walter Reed Biosystematics Unit. VectorMap. http://vectormap.si.edu/. Published 2016. Accessed October 20, 2016.
16. Sikes DS, Copas K, Hirsch T, Longino JT, Schigel D. On natural history collections, digitized and not: A response to Ferro and Flick. *Zookeys*. 2016;618:145-158. doi:10.3897/zookeys.618.9986.
17. Centers for Disease Control and Prevention. National Notifiable Diseases Surveillance System (NNDSS). https://wwwn.cdc.gov/nndss/. Published 2016. Accessed October 19, 2016.
18. Rutgers Center for Vector Biology. New Jersey Statewide Adult Mosquito Surveillance Reports. http://vectorbio.rutgers.edu/reports/mosquito/. Published 2016. Accessed October 20, 2016.
19. The California Department of Public Health. West Nile Virus Website. http://westnile.ca.gov/. Published 2016. Accessed December 6, 2016.
20. Martinez ME. Preventing zika virus infection during pregnancy using a seasonal window of opportunity for conception. Read AF, ed. *PLOS Biol*. 2016;14(7):e1002520. doi:10.1371/journal.pbio.1002520.
21. Hahn MB, Eisen RJ, Eisen L, et al. Reported Distribution of Aedes (Stegomyia) aegypti and Aedes (Stegomyia) albopictus in the United States, 1995-2016 (Diptera: Culicidae). *J Med Entomol*. 2016;ePub ahead:1-

7. doi:10.1093/jme/tjw072.

22. Monaghan AJ, Morin CW, Steinhoff DF, et al. On the seasonal occurrence and abundance of the Zika virus vector mosquito Aedes Aegypti in the contiguous United States. *PLoS Curr Outbreaks*. 2016;1. doi:10.1371/currents.outbreaks.50dfc7f46798675fc63e7d7da563da76.Authors.
23. Iowa State University Entomology Department. Iowa Mosquito Surveillance. http://mosquito.ent.iastate.edu/. Published 2016.
24. South Dakota State University. South Dakota Mosquito Information Systems. http://mosquito.sdstate.edu/. Accessed October 20, 2016.
25. Costello MJ, Michener WK, Gahegan M, Zhang ZQ, Bourne PE. Biodiversity data should be published, cited, and peer reviewed. *Trends Ecol Evol*. 2013;28(8):454-461. doi:10.1016/j.tree.2013.05.002.
26. The White House. Open Government Initiative. https://www.whitehouse.gov/open.
27. CDC ArboNET. Diseasemaps. http://diseasemaps.usgs.gov/mapviewer/. Published 2016. Accessed October 20, 2016.
28. PAHO, WHO. Dengue | PAHO/WHO Data, Maps and Statistics. http://www.paho.org/hq/index.php?option=com_topics&view=readall&cid=3273&Itemid=40734&lang=en. Published 2016. Accessed November 1, 2016.
29. Acuff VR. Trap biases influencing mosquito collecting. *Mosq News*. 1976;36(2):173-176.
30. National Science Foundation. NEON: National Ecological Observatory Network. http://www.neonscience.org/.
31. Costello MJ, Wieczorek J. Best practice for biodiversity data management and publication. *Biol Conserv*. 2014;173:68-73. doi:10.1016/j.biocon.2013.10.018.
32. Homer C, Dewitz J, Yang L, et al. Completion of the 2011 National Land Cover Database for the conterminous United States – Representing a decade of land cover change information. *Photogramm Eng Remote Sens*. 2015:345-354.
33. NASA Goddard. Land Data Assimilation Systems (LDAS).
34. National Center for Atmospheric Research. Community Climate System Model (CCSM). http://www.cesm.ucar.edu/models/ccsm4.0/. Accessed October 20, 2016.
35. Global Biodiversity Information Facility. GBIF. http://www.gbif.org/. Accessed October 20, 2016.
36. Guzman MG, Harris E. Dengue. *Lancet*. 2015;385:453-465. doi:10.1016/S0140-6736(14)60572-9.
37. American Mosquito Control Association. Links. http://www.mosquito.org/links. Published 2016. Accessed October 20, 2016.
38. Florida Department of Agriculture & Consumer Services. Mosquito Directory. http://www.freshfromflorida.com/Divisions-Offices/Agricultural-Environmental-Services/Consumer-Resources/Mosquito-Control/Mosquito-Directory. Published 2016. Accessed October 20, 2016.
39. Texas Mosquito Control Association. Organized Mosquito Control Districts

in Texas. http://www.texasmosquito.org/Districts.html. Published 2016. Accessed October 20, 2016.
40. Mosquito and vector control association of California. Member Agencies - MVCAC. http://www.mvcac.org/resources/member-agencies/. Published 2016. Accessed October 20, 2016.
41. Michigan Mosquito Control Association. Information Links. http://www.mimosq.org/links/links.htm. Accessed October 20, 2016.
42. Google. Google Trends. https://www.google.com/trends/. Published 2016. Accessed October 20, 2016.
43. Gouhier TC, Grinsted A, Simko V. Package "biwavelet." 2016. https://github.com/tgouhier/biwavelet. Accessed October 27, 2016.