Rescuing Hidden Ecological Data to Tackle Emerging Mosquito-Borne Diseases

Samuel S.C. Rund[1*] and Micaela Elvira Martinez[2*]

[1]University of Notre Dame, Indiana, USA
[2]Princeton University, New Jersey, USA

[*]The authors contributed equally to this article.
Correspondence: srund@nd.edu & micaelam@princeton.edu

Worldwide, mosquito transmitted infectious diseases result in approximately 679,000 deaths annually. In the Americas, mosquitoes transmit Yellow Fever, malaria, West Nile virus, Dengue virus, Chikungunya virus, and most recently Zika virus. Municipalities throughout the US have independently created mosquito control agencies which aim to limit mosquito populations and carefully monitor mosquito abundance. Here we reveal $>$ 1000 mosquito control agencies in the US and demonstrate that collectively, their high spatiotemporal resolution—historical and real-time—data comprise the richest long-term ecological surveillance of any animal taxon. These agencies are geographically dispersed and house an enormous trove of data: weekly/monthly snapshots of mosquito abundance for up to eight genera simultaneously. Unfortunately, much of these data are hidden among many independent data silos. Here we propose a simple low-cost policy solution: the creation of a nationally-coordinated open-access database to collate mosquito data. The proposed

database would, for the first time, provide scientific and public health communities with readily accessible data on arboviral disease vectors. This would empower new interventions and insights by leveraging pre-existing human efforts and operational infrastructure already being paid-for by taxpayers.

Mosquito Control & Surveillance

Mosquito transmitted infectious diseases pose a threat to public health worldwide. In the Americas, mosquitoes, both historically and presently, transmit pathogenic viruses including, but not limited to, Yellow Fever virus, West Nile Virus (WNV), Dengue virus (DENV), Chikungunya virus (CHIKV), and most recently Zika virus (ZIKV) (*1*). Over the past century, municipalities across the US have launched mosquito control agencies, periodically spurred as efforts to tackle existing or emerging disease threats (Fig 1a), including the *Aedes aegypti* eradication initiative in the Americas (1947–1970), the malaria eradication program in the southern US (1947–1951), and the introduction of WNV to the US (1999). Importantly, the primary role of mosquito control agencies is to respond to mosquito-borne diseases by mitigating mosquito populations and carefully trapping and taxonomically identifying mosquitoes to monitor temporal changes in abundance. Mosquito control agencies, therefore, generate long-term ecological time series of mosquito populations, which collectively represent unprecedented ecological monitoring that is spatially, temporally, and taxonomically rich.

Given the public health threat of arboviral disease in the US, there is great incentive to invest in studies of mosquito distribution, ecology, and abundance. The US CDC has guidelines for standardized and repeated collection of arboviral disease-vectors. Mosquito control agencies, thus, perform routine surveillance by trapping mosquitoes to measure their population size and inform control efforts (*2, 3*). This routine surveillance—being done independently, yet simultaneously—by agencies which may also perform regular mosquito phenotypic assays (such as insecticide resistance and arbovirus-presense assays), has generated an enormous wealth of long-term ecological data that has yet to be collated and leveraged for the public good.

In the US, mosquito-borne diseases are nationally notifiable with state-level reporting to the CDC National Notifiable Disease Surveillance System and published in the Morbidity and Mortality Weekly Report (*4*). We propose a similar system be created for bringing together existing mosquito population surveillance that is: standardized, centralized, and regular. The creation of such a surveillance system would provide a repository for the vast amount of mosquito surveillance data that have been collected for decades. Here we (1) identify mosquito control agencies throughout the US that could contribute to the surveillance system, revealing the unrealized potential of coordinated national mosquito surveillance, and (2) collate the sizable amount of open access abundance data from 2009–present made available by approximately $1/10^{th}$ of the agencies we identified.

The Tip of the Iceberg of Rich Ecological Data

We identified approximately 1000 mosquito control agencies scattered throughout the contiguous US (Fig 1b), with a mosquito control agency broadly defined as the local government authority responsible for the surveillance and control of mosquitoes. Of the 1054 agencies we identified (Table S1), we attempted to locate an online presence for each. A total of 152 agencies had publically available mosquito abundance time series data online (i.e., live weekly/monthly updates and/or archived data spanning at least one year from 2009-2015). Publicly available data included mineable formats such as tables, graphs, spreadsheets, or GIS maps. Importantly, we only collated data collected using fixed mosquito traps (i.e., traps set in specific geographic locations and repeatedly sampled).

A total of 148 agencies (14%) across 20 states had archived data. Of these, 107 also provided live data updated at least monthly in 2016. We found that some states were coordinating across municipalities and already had state-wide surveillance networks with public data. These included North Dakota, New Jersey, and Iowa. We took existing data collation efforts a step further and combined data from 64 data silos nationwide (i.e., agencies/state repositories). Some agencies provided trap-level data, while others reported mosquito counts aggregated from a geographic feature, for example a park name, a jurisdiction, or an ecological region. In total, our dataset contained reports from > 600 unique agency-defined locations, $> 39{,}000$ instances when traps were checked, approx. 200,000 records, and documented > 15 million individual mosquitoes. Records were each time-stamped with a collection date and geo-referenced. These data represent a great deal of human effort. Each trap instance required one or more biologists to count and identify mosquitoes, often to the genus level, and sometimes species. Data provided in Table S2.

For each mosquito control agency from which we obtained data, Fig 1 shows a time series of the number of genera whose abundance they reported each week or month, depending on their trapping frequency and reporting. Fig 1 also provides an example abundance time series for mosquitoes in the Culex genus—the genus containing WNV vectors—at five sample locations in Massachusetts. While Fig 1 demonstrates an enormous amount of data, it is merely the tip of the iceberg because we only captured the small fraction of extant data that were publicly available online in a mineable format. We also focused collation efforts on data from 2009 onward; yet, large historical data collections exist, many of which may be undigitized. For example, we found that Iowa, New Jersey, and North Dakota have state-wide surveillance systems with weekly data dating back to 1969, 2003, and 2006, respectively. Many agencies in our study indicated they were in possession of finer-scale and longer-term data than presently posted online. Substantial temporal depth, above that provided by control agencies, could be found in US military installations that conducted mosquito surveillance dating back to 1947 (*5*). Finally, control agencys mosquito insecticide resistance and arbovirus-presence assays could also be collated.

Taxonomic identification and resolution of reporting varied among districts. Often, mosquitoes were identified or aggregated to the genus level, but 56 data sources report species for at least a subset of mosquitoes. This reflects local knowledge of disease vector and nuisance species. Many mosquito control agencies have species-level identification but aggregate this information on their websites. The majority of mosquito data are hidden because the majority of agencies do not post their data. For instance, a recent National Association of County and City Health Officials (NACCHO) report found 126 agencies in 9 southern-border states reported they "conduct routine surveillance for mosquitoes through standardized trapping and species identification" (*3*); yet, we were only able to access data from 12 of these 126 agencies. At least 114 of these border state agencies are, therefore, generating data presently unavailable to the public.

Surveillance Network Design

Our proposed national surveillance system would facilitate the flow of historical and ongoing mosquito abundance data into a centralized repository at regular weekly or monthly intervals. In our collation efforts, we found that the format of existing data varied widely among agencies. Data standardization, therefore, would be required in order to create a national mosquito surveillance system. We do not suggest significant changes in agency operational protocols, but it is important that agencies standardize reporting and metadata, which is necessary for data interpretation and integration. We propose, the most valuable data include (1) trap type, (2) GPS location, (3) attractants used, (4) taxonomic identification, (5) the date the trap was counted, and (6) how long the trap was set. Finally, the units for reporting mosquito counts must be standardized. We recommend agencies report the number of mosquitoes per trap, rather than summary statistics for an aggregation of traps. Of particular importance is reporting whether both male and female mosquitoes are counted, or females only. Whenever available, all historical data should also be made available. Such a national mosquito surveillance system could utilize Darwin Core standards and ensure deposition within The Global Biological Information Facility (GBIF) (*3*).

Unrealized Potential

Mathematical models are relied upon to evaluate risk of mosquito-transmitted infectious diseases. Historical data from geographically independent regions are critical for model calibiration and validation (*6–8*); however, these data are lacking for arboviral disease vectors. Acheson and Kerr 2015 reviewed 29 human disease-vector modeling publications, and found a lack of mosquito data. At least six of the publications they reviewed explicitly claimed insufficient data for model validation (*8*). During the 2016 Zika outbreak, the need for mosquito data was particularly acute. Publications inferring the range of *Ae. aegytpi* (the Zika/dengue/yellow fever mosquito) need reliable, publically available data (*6, 9*).

Rich longitudinal data of mosquito abundance would facilitate analyses that are not currently feasible. Data from existing mosquito control agencies represents the breadth of urban and ecological conditions in the US, and could be used to gain mechanistic insight into the underpinnings of mosquito population biology (*10*). Truly new and powerful analyses would be possible if abundance time series were combined with other sources of Big Data, such as arbovirus disease incidence data (e.g., ArboNET (*11*)); land use data (e.g., National Land Cover Database (*12*)); meteorological data (e.g., Land Data Assimilation Systems, LDAS (*13*)), climate change predictions (e.g., NCAR-CCSM 4.0 (*14*)), and other mosquito phenotypic data such as insecticide resistance and arbovirus presence. Mosquito surveillance data could be used to motivate action when neighboring communities have vectors of concern (*6*), and tools could be developed to assist local control agencies, such as the automated generation of up-to-date disease risk maps and reports. The great potential of mosquito surveillance data would be unlocked if they were collated and used to:

1. Predict epidemic timing for mosquito-borne diseases;
2. Quantify geographic variation in risk and epidemic size, especially at geographic edges of each vector's range;
3. Identify sources for pathogen persistence across the US by determining where vectors overwinter in sufficient numbers to sustain transmission;
4. Define windows of opportunity for efficient mosquito control, including spraying, larval site management, testing for virus, and public outreach campaigns;
5. Coordinate responses across states/counties, e.g, for mitigating epidemics and creating early warning systems, or responding preemptively based on the situation in neighboring communities;
6. Monitor invasive mosquito species for changing geographical ranges under climate change.

The Way Forward

The increasing interconnectedness of human populations, global climate change, and the emergence of new vector-borne diseases necessitates diligent vector control. One simple low-cost solution is the creation of a nationally-coordinated open-access database to collate mosquito data. The creation of such a system would be a means of empowering vector research and disease control efforts. Current mosquito surveillance is happening with (1) a tremendous taxpayer cost and (2) large amounts human labor and expert knowledge, which has been ongoing for decades. Substantial financial and human resources have already been invested into mosquito trapping, the public health and research benefits of these data should be maximized through data dissemination.

References

1. World Health Organization, A Global Brief on Vector-borne Diseases, *Tech. rep.* (2014).
2. Centers for Disease Control and Prevention, West Nile Virus in the United States: Guidelines for Surveillance, Prevention, and Control, *Tech. rep.* (2013).
3. National Association of County and City Health Officials, Mosquito Surveillance and Control Assessment in Zika Virus Priority Jurisdictions, *Tech. Rep. December* (2016).
4. D. Adams, *et al.*, *MMWR Morbidity and Mortality Weekly Report* **63** (2016).
5. The Walter Reed Biosystematics Unit, VectorMap.
6. A. J. Monaghan, *et al.*, *PLoS Current Outbreaks* **1**, 31 (2016).
7. L. P. Campbell, *et al.*, *Philosophical Transactions of the Royal Society, B.* **370**, 20140135 (2015).
8. E. Acheson, J. Kerr, *Vector-Borne Zoonotic Diseases* **15**, 173 (2015).
9. M. B. Hahn, *et al.*, *Journal of Medical Entomology* **ePub ahead**, 1 (2016).
10. M. J. Costello, W. K. Michener, M. Gahegan, Z. Q. Zhang, P. E. Bourne, *Trends in Ecology and Evolution* **28**, 454 (2013).
11. CDC ArboNET, Disease Maps - Dynamic Map (2016).
12. C. Homer, *et al.*, *Photogrammetric Engineering & Remote Sensing* pp. 345–354 (2015).
13. NASA Goddard, Land Data Assimilation Systems (LDAS).
14. National Center for Atmospheric Research, Community Climate System Model (CCSM).

Acknowledgments

This work was possible due to the efforts of the 148 mosquito control agencies who collected mosquito population abundance data and made their data available online. We thank research students Nathan Nixon, Ben Weise, Ben Chan, Jason Cleland, Ollie Lee, and Emma Stotter for their work in locating and digitizing 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 Directors Early Independence Award. Research reported in this publication was supported by the Office of the Director, National Institutes of Health, under Award Number DP5OD023l00. Samuel S.C. Rund was funded by the Royal Society (NF140517) and a strategic award from the Wellcome Trust (No. 095831).

Supplementary materials

Materials and Methods

Tables S1 to S2

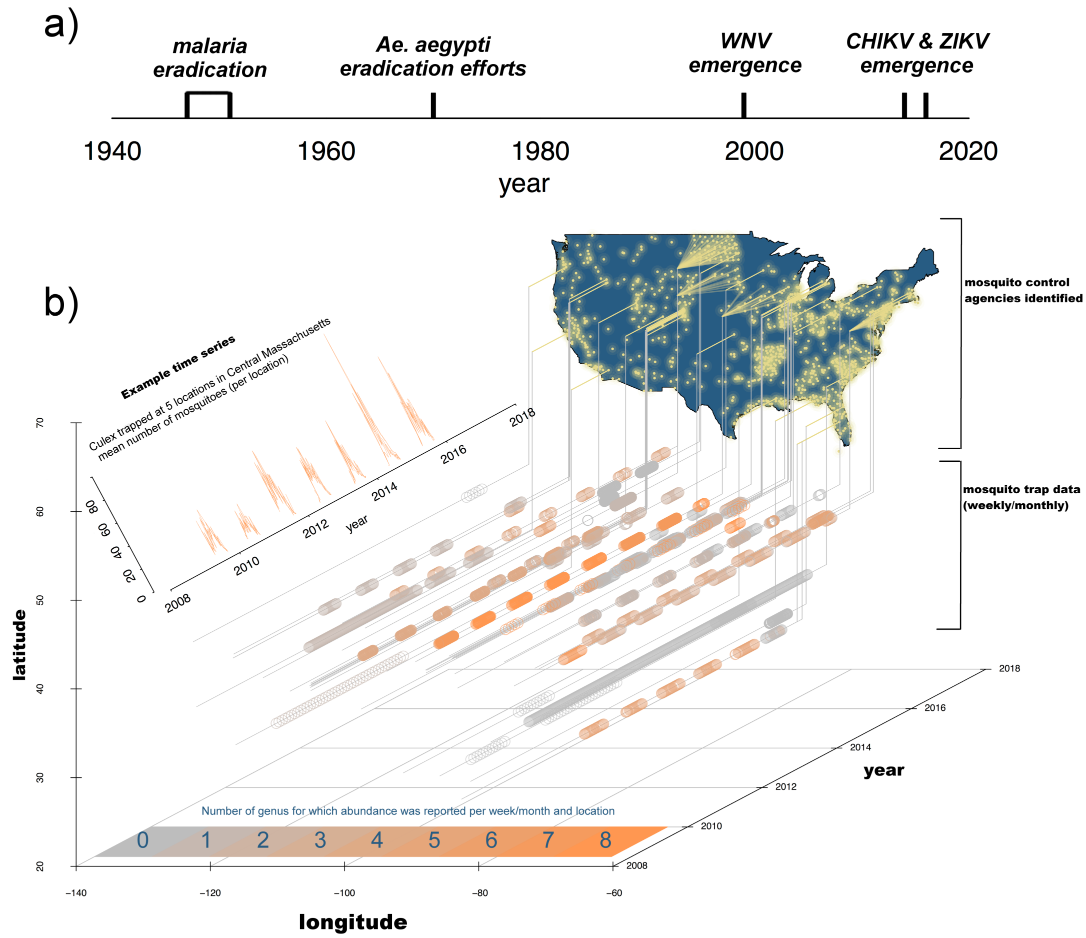

Fig. 1. (a) Key events spurring the creation of mosquito control agencies in the US. (b) Map shows the location of agencies we identified (yellow points). Some agencies made their data publicly available (those with a yellow line projection). Time series show the collection dates and the number of genus whose abundance was reported in our data. Each time series represents one data silo. Despite the enormous amount of data (shown are $>$ 39,000 trap collections), far more exist. Each agency is a potential data source. (inset) Example time series of Culex abundance from one agency, the Central Massachusetts Mosquito Control Project, which had five trap locations in their jurisdiction. Culex is of interest because it is the genus containing species that transmit WNV.