

Towards Translational Epidemiology: Next-Generation Sequencing and Phylogeography as Epidemiological Mainstays

Crystal M. Hepp^{a,b}

^aSchool of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA

^bThe Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, Arizona, USA

ABSTRACT Next-generation sequencing, coupled with the development of user-friendly software, has achieved a level of accessibility that is revolutionizing the way we approach epidemiological investigations. We can sequence pathogen genomes and conduct phylogenetic analyses to assess transmission, identify from which country or city a pathogen originated, or which contaminated potluck item resulted in widespread foodborne illness. However, until recently, these types of studies have been rarities, limited to specific investigations usually conducted over the short term. Given the feasibility and realized public health benefits of ascertaining pathogen relationships, federal, state, and county agencies are building their sequencing capacities, either through acquisition of equipment or collaborative activities. In this perspective, I detail research projects that our group collaborates on with county and state public health agencies, where the objective is to identify pathogen source locations with the longer-term goal of implementing proactive interventions.

KEYWORDS public health, translational epidemiology, vector-borne diseases, zoonotic disease

The identification of West Nile virus (WNV) in the United States in 1999 stimulated vector-borne zoonotic disease surveillance through appropriated federal funding disseminated nationwide (1). For many vector control agencies, this funding allowed for the development of large-scale search and destroy programs, where high vector prevalence or arbovirus-positive samples collected from traps trigger reactive intervention measures. Additionally, surveillance data are relayed back to federal, state, and county public health agencies at higher geographic scales through ArboNET (2), which provides county level resolution of reportable arboviral presence in mosquito pools, humans, and animals. While regional presence and relative prevalence of pathogens has been leveraged to predict human risk (3), little can be ascertained from these traditional practices in the way of disease spread or reemergence. However, vector-borne and other zoonotic infectious disease epidemiology at the agency level is approaching a paradigm shift, largely propelled by the advent and feasibility of next-generation sequencing, innovative and user-friendly phylogenomic software packages, and motivation of the One Health Movement. Briefly, the One Health concept is a realization that human, animal, and environmental health are inextricably intertwined, largely due to overlapping habitats, activities, food sources, etc. Approaches to address One Health-related outbreaks require meticulous record-keeping and cooperation among stakeholders at all levels, from federal agencies to small farm owners, which is often challenging. A recent national investigation of the largest *Escherichia coli* O157:H7 outbreak since 2006, resulting in 210 cases and five deaths, demonstrated the ability of a One Health approach to locate pathogen sources of human illness (4). The


Citation Hepp CM. 2019. Towards translational epidemiology: next-generation sequencing and phylogeography as epidemiological mainstays. *mSystems* 4:e00119-19. <https://doi.org/10.1128/mSystems.00119-19>.

Copyright © 2019 Hepp. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to CRYSTAL.HEPP@nau.edu.

Conflict of Interest Disclosures: C.M.H. reports grants from the Arizona Biomedical Research Commission and the Pacific Southwest Regional Center of Excellence in Vector-Borne Diseases during the conduct of the study.

mSystems® vol. 4, no. 3, is a special issue sponsored by Illumina.

 Perspective: Towards Translational Epidemiology - next generation sequencing and user-friendly phylogenetic tools are paving the way for a paradigm shift in epidemiological practices.

Received 15 March 2019

Accepted 24 April 2019

Published 11 June 2019

Food and Drug Administration, Centers for Disease Control and Prevention (CDC), and state partners worked with leafy greens farms, processing facilities, cattle feeding operations, and water districts to conduct an investigation that would ultimately identify a regional source of contaminated romaine lettuce. Of the numerous samples collected, only three samples from an irrigation canal were positive for the outbreak strain, as determined by whole-genome sequencing. While a direct link could not be made to environmental contamination by animals, a large cattle feeding operation is located along the sampled segment of the canal, and canal water was applied to some lettuce crops. This investigation is an exemplary model for One Health approaches moving forward, and with the CDC making recent technological investments in public health laboratories as part of the Advanced Molecular Detection Initiative, we can expect that these investigations will form epidemiological foundations.

In addition to sequencing accessibility, the development of user-friendly software that allows for the incorporation of geographic coordinates, timing information, and other metadata (e.g., BEAST [5]) has paved the way for better understanding regional circulation of infectious diseases, identifying probable source locations and reservoirs from which these diseases emerge, and factors that encourage or mitigate spread. Such methods have been recently employed to understand the circulation of Ebola virus during the 2013–2016 epidemic in West Africa (6) and Zika virus in the 2015–2016 epidemic in the Americas (7). In collaboration with county vector control agencies, state public health departments, the Pathogen and Microbiome Institute at Northern Arizona University, and the Pathogen Genomics Division of the Translational Genomics Research Institute, our research team employs the described technological approaches within a One Health framework toward characterizing the circulation of three annually reemerging zoonotic viruses, at various geographic scales, to identify locations that are critical in long-term local maintenance.

WEST NILE VIRUS

WNV was first identified in New York City in 1999 and successfully migrated across the continental United States by 2004. Nearly 2 decades after the initial introduction, WNV is still the most important arbovirus nationwide, causing 95% of arboviral diseases reported to the Centers for Disease Control and Prevention (CDC) (8). While WNV has been variably present in counties across the United States from year to year, positive mosquito pools and human clinical cases have occurred in Maricopa County in Arizona every year after the first detection in September 2003. In response, the Maricopa County Environmental Services Vector Control Division has developed what is likely one of the most intricate arbovirus surveillance systems in the United States. With more than 800 carbon dioxide traps distributed throughout the Phoenix metropolitan area, collecting mosquitoes that are tested for arboviral activity each week, the county consistently detects the environmental threat of WNV each year prior to the occurrence of human clinical cases—the hallmark of an effective surveillance system.

We have initiated an extensive collaboration with the county, which involves sequencing WNV genomes from all positive mosquito pools they collect each year, layered with geographic and timing information, to better understand whether WNV is maintained in Maricopa County each year or is annually imported, and to identify source locations. Our efforts, which now include more than 300 sequenced genomes from 2014 to 2018, have shown that the currently circulating population of WNV has been endemic in Maricopa County since 2013 (9). We have additionally incorporated other southwestern locations, and a preliminary analysis resulting from our collaborations with Yuma County (Arizona) Pest Abatement District and Coachella (California) Valley Mosquito and Vector Control District has revealed that the endemic population of WNV in Maricopa County, at least during 2017, acted as a source for WNV strains that emerged and thrived in Yuma County, AZ, and Riverside County, CA. We just received samples from these locations for 2018, and we are in the process of investigating whether or not the same patterns reemerge, which would suggest that Maricopa

County is an important source location for WNV maintenance in the southwestern United States.

ST. LOUIS ENCEPHALITIS VIRUS

St. Louis encephalitis virus (SLEV) is a second arbovirus that has caused disease in Arizona since 2015. While the virus was first detected in St. Louis, Missouri, in 1933, the strain currently circulating in the southwestern United States, causing human disease primarily in Arizona, California, and Nevada, likely originated in Argentina (10). In a similar fashion to the WNV study mentioned above, we have partnered with the Coachella Valley Mosquito and Vector Control District and Maricopa County Environmental Services Vector Control Division to determine whether the SLEV is locally maintained or is annually imported to Riverside County, CA, and Maricopa County, AZ. First, our studies have led us to question whether or not this recently imported strain of SLEV is competitively excluded by WNV, as has been suggested for previous strains (11), given geographic proximity and often overlap of the two viruses in positive mosquito pools in Maricopa County. With only 1 of every 100 mosquito pools being positive for WNV, the vector component of the mosquito-avian enzootic cycle does not appear to be consumed such that either virus would be competitively excluded. However, less is known about current susceptibility and immunity of competent avian hosts. Additionally, our preliminary results based on 65 sequenced genomes suggest that SLEV may also be establishing an endemic population in Arizona, alongside WNV, as strains sequenced from 2018 are polyphyletically nested within their 2017 counterparts. While SLEV infection does not result in clinical manifestations as often as WNV infections, in 2015, 23 cases of SLEV were reported in Arizona, and of those, 19 resulted in neuroinvasive disease from which 2 people succumbed (12).

RABIES

Over the past 3 years, there have been more than 450 wild animal cases of rabies in Arizona (13). During the same time frame, there were 88 human exposures. Although human deaths are infrequent in the United States, public health costs range between \$245 and 510 million annually, which includes an estimated 40,000 to 50,000 postexposure prophylaxis treatments (14). These treatments tend to be concentrated in locations where the rabies virus (RABV) has become established in reservoir populations. The Arizona Department of Health Services (ADHS) has been diligently testing, variant typing, and tracking RABV-positive animals and exposed humans throughout the state for several years. We have recently initiated a collaboration with this agency with the overarching goal of characterizing spatiotemporal circulation of the virus. In a similar manner to the SLEV and WNV projects, we will sequence RABV from positive brain stem samples provided by ADHS and overlay the resulting genomes with geographic and timing information. Campaigns to reduce RABV in raccoon populations of the southeastern United States have been successful in locations where oral RABV vaccine baits have been placed (15), and identification of contemporary RABV source populations in Arizona may pave the way for similar intervention campaigns.

FUTURE DIRECTIONS

We anticipate that these studies, which have been heavily focused on Arizona up to this point, will have major implications regarding the public health-focused utility of incorporating intensely sampled pathogen genomes over local geographic scales for several years. With many state public health labs now gaining access to next-generation sequencing technologies, I anticipate that projects like those that our team has embarked upon will become a mainstay in epidemiological practices. Our hope for each project is that the identification of reemergent circulation patterns will provide proactive intervention opportunities that can reduce pathogen load in the environment.

ACKNOWLEDGMENTS

The research discussed in this perspective would not be possible without the contributions of the county vector control (Maricopa County Environmental Services Vector Control Division, Coachella Valley Mosquito and Vector Control District, Yuma County Pest Abatement District, and Southwest Mosquito Abatement and Control District) and state public health agencies (Arizona Department of Health Services, Utah Department of Health, Washington State Department of Health, and Montana Department of Public Health & Human Services) involved in this work. Many thanks to Paul Keim at the Pathogen and Microbiome Institute and David Engelthaler at the Translational Genomics Research Institute for valuable discussions regarding the topics presented in this perspective. Finally, I thank Greg Caporaso at the Pathogen and Microbiome Institute for encouraging me to write this piece.

This work has been supported by the following funds awarded to Crystal Hepp: New Investigator Award from the Arizona Biomedical Research Center, start-up funds from the Arizona Technology Research and Initiative Fund, and a training grant through the Pacific Southwest Regional Center of Excellence for Vector-Borne Diseases funded by the U.S. Centers for Disease Control and Prevention (Cooperative Agreement 1U01CK000516).

REFERENCES

- Hadler JL, Patel D, Nasci RS, Petersen LR, Hughes JM, Bradley K, Etkind P, Kan L, Engel J. 2015. Assessment of arbovirus surveillance 13 years after introduction of West Nile virus, United States. *Emerg Infect Dis* 21: 1159–1166. <https://doi.org/10.3201/eid2107.140858>.
- Marfin AA, Petersen LR, Eidson M, Miller J, Hadler J, Farello C, Werner B, Campbell GL, Layton M, Smith P, Bresnitz E, Cartter M, Scaletta J, Obiri G, Bunning M, Craven RC, Roehrig JT, Julian KG, Hinten SR, Gubler DJ. 2001. Widespread West Nile virus activity, eastern United States, 2000. *Emerg Infect Dis* 7:730–735. <https://doi.org/10.3201/eid0704.010423>.
- Marini G, Rosa R, Pugliese A, Rizzoli A, Rizzo C, Russo F, Montarsi F, Capelli G. 2018. West Nile virus transmission and human infection risk in Veneto (Italy): a modelling analysis. *Sci Rep* 8:14005. <https://doi.org/10.1038/s41598-018-32401-6>.
- US Food and Drug Administration. 2018. Environmental assessment of factors potentially contributing to the contamination of romaine lettuce implicated in a multi-state outbreak of *E. coli* O157:H7. US Food and Drug Administration, Silver Spring, MD. <https://www.fda.gov/Food/RecallsOutbreaksEmergencies/Outbreaks/ucm624546.htm#II>.
- Suchard MA, Baele G, Lemey P, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 4:vey016. <https://doi.org/10.1093/ve/vey016>.
- Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, Park DJ, Ladner JT, Arias A, Asogun D, Bielejec F, Caddy SL, Cotten M, D'Ambrozio J, Dellicour S, Di Caro A, DiCiaro JW, Duraffour S, Elmore MJ, Fakoli LS, Faye O, Gilbert ML, Gevaou SM, Gire S, Gladden-Young A, Gnirke A, Goba A, Grant DS, Haagmans BL, Hiscox JA, Jah U, Kugelman WK, Liu D, Lu J, Malboeuf CM, Mate S, Matthews DA, Matranga CB, Meredith LW, Qu J, Quick J, Pas SD, Phan MVT, Pollakis G, Reusken CB, Sanchez-Lockhart M, Schaffner SF, Schieffelin JS, Sealfon RS, Simon-Loriere E, Smits SL, et al. 2017. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* 544:309–315. <https://doi.org/10.1038/nature22040>.
- Faria NR, Quick J, Claro IM, Theze J, de Jesus JG, Giovanetti M, Kraemer MUG, Hill SC, Black A, da Costa AC, Franco LC, Silva SP, Wu CH, Raghwani J, Cauchemez S, Du Plessis L, Verotti MP, de Oliveira WK, Carmo EH, Coelho GE, Santelli A, Vinhal LC, Henriques CM, Simpson JT, Loose M, Andersen KG, Grubaugh ND, Somasekar S, Chiu CY, Munoz-Medina JE, Gonzalez-Bonilla CR, Arias CF, Lewis-Ximenez LL, Baylis SA, Chieppe AO, Aguiar SF, Fernandes CA, Lemos PS, Nascimento BLS, Monteiro HAO, Siqueira IC, de Queiroz MG, de Souza TR, Bezerra JF, Lemos MR, Pereira GF, Loudal D, Moura LC, Dhalia R, Franca RF, Magalhaes T, et al. 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 546:406–410. <https://doi.org/10.1038/nature22401>.
- Krow-Lucal E, Lindsey NP, Lehman J, Fischer M, Staples JE. 2017. West Nile virus and other nationally notifiable arboviral diseases - United States, 2015. *MMWR Morb Mortal Wkly Rep* 66:51–55. <https://doi.org/10.15585/mmwr.mm6602a3>.
- Hepp CM, Cocking JH, Valentine M, Young SJ, Damian D, Samuels-Crow KE, Sheridan K, Fofanov VY, Furstenuau TN, Busch JD, Erickson DE, Lancione RC, Smith K, Will J, Townsend J, Keim PS, Engelthaler DM. 2018. Phylogenetic analysis of West Nile virus in Maricopa County, Arizona: evidence for dynamic behavior of strains in two major lineages in the American Southwest. *PLoS One* 13:e0205801. <https://doi.org/10.1371/journal.pone.0205801>.
- White GS, Symmes K, Sun P, Fang Y, Garcia S, Steiner C, Smith K, Reisen WK, Coffey LL. 2016. Reemergence of St. Louis encephalitis virus, California, 2015. *Emerg Infect Dis* 22:2185–2188. <https://doi.org/10.3201/eid2212.160805>.
- Fang Y, Reisen WK. 2006. Previous infection with West Nile or St. Louis encephalitis viruses provides cross protection during reinfection in house finches. *Am J Trop Med Hyg* 75:480–485. <https://doi.org/10.4269/ajtmh.2006.75.480>.
- Centers for Disease Control and Prevention. 2018. St. Louis encephalitis statistics and maps. Centers for Disease Control and Prevention, Atlanta, GA. <https://www.cdc.gov/sle/technical/epi.html#casebyyear>.
- Arizona Department of Health Sciences. 2019. Rabies data. Arizona Department of Health Sciences, Phoenix, AZ. <https://www.azdhs.gov/preparedness/epidemiology-disease-control/rabies/#data-publications-maps>.
- Centers for Disease Control and Prevention. 2015. Cost of rabies prevention. Centers for Disease Control and Prevention, Atlanta, GA. <https://www.cdc.gov/rabies/location/usa/cost.html>.
- Plants KB, Wen S, Wimsatt J, Knox S. 2018. Longitudinal analysis of raccoon rabies in West Virginia, 2000–2015: a preliminary investigation. *PeerJ* 6:e4574. <https://doi.org/10.7717/peerj.4574>.