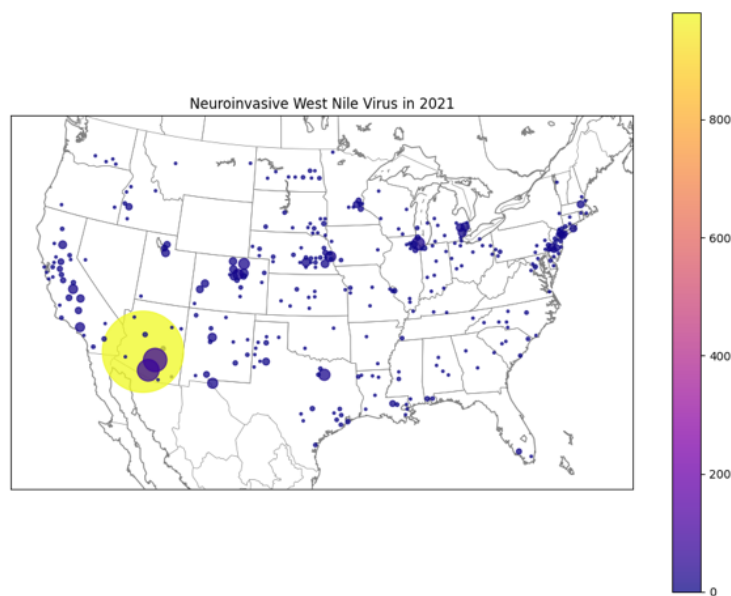


# WNV Prediction Proposal

## Introduction

Insect-borne diseases significantly impact global health, and mosquitoes from the *Culex* family are key vectors in transmitting viruses such as encephalitis and West Nile Virus (WNV) (Colpitts *et al.*, 2012) to both humans and animals (Gubler, 1998). *Culex tarsalis*, also known as the Western Encephalitis Mosquito, is of particular concern in the United States, where it has been linked to the majority of WNV cases in western regions where the disease is most prevalent (Goddard *et al.*, 2002; Evans *et al.*, 2017)(Fig. 1). This connection underscores the critical need for comprehensive research into the ecology and genetics of this mosquito species to effectively manage and predict outbreaks. Despite its significant role in disease transmission, there remain several gaps in our understanding of *Cx. tarsalis*'s genetic makeup and unique population dynamics (Venkatesan and Rasgon, 2010; Pfeiler *et al.*, 2013) which have enabled its expansion and adaptation to diverse environments across much of the continent and likely contribute to the species' capacity to spread dangerous diseases like WNV.



**Fig 1: Distribution of Human West Nile Virus cases in 2021: CDC WNV Surveillance Data.**

My research project addresses the pressing public health and economic challenges posed by WNV and its primary vector, *Cx. tarsalis* (Hadfield *et al.*, 2019). Given the prevalence of WNV in the western United States and the risk of increased outbreaks due to climate change, it is crucial that we understand the genetic landscape of this host mosquito species and the environmental factors that have shaped its landscape. By investigating *Cx. tarsalis*' genetic diversity, ecological adaptations, and their connection to past WNV outbreaks, this project aims

to bridge knowledge gaps to predict and manage future outbreaks, which will ultimately inform targeted strategies to curb virus transmission. Enhanced understanding in these areas enables more effective risk assessment and implementation of control measures, ultimately reducing WNV's incidence and societal impact.

## Gaps in knowledge and questions/hypotheses

1. What genetic variations contribute to the adaptability of *Cx. tarsalis* in diverse environmental conditions across its expanded range?
  - This question aims to uncover the genetic basis of the mosquito's ability to thrive in varied climates and geographical areas, particularly the adaptations that have allowed it to cross significant natural barriers like the Rocky Mountains.
2. Which environmental factors most strongly influence the population structure of *Cx. tarsalis* and correlate to past outbreaks of WNV?
  - This seeks to identify the specific climatic or ecological variables that correlate with the distribution and density of *Cx. tarsalis* populations, contributing to their capacity to transmit diseases over a wide area.
3. Can we develop a model to predict future outbreaks of WNV based on current *Cx. tarsalis* genetic data and publicly available environmental data?
  - This hypothesis posits that by incorporating what we have learned about the genetics of the vector species and the climatic factors that affect it, we can develop models to correlate this information to past recorded WNV outbreaks, and ultimately to predict future ones.

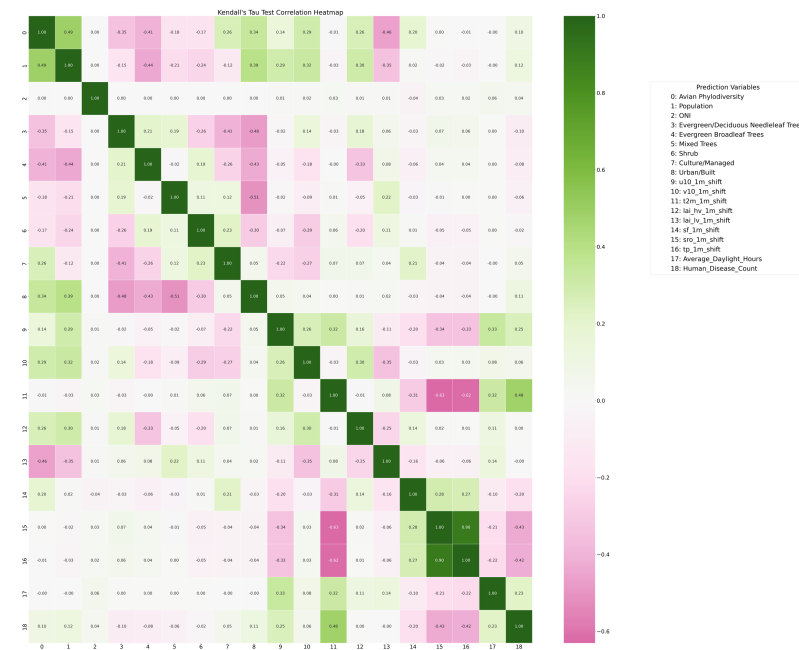
## Methods

My recent manuscript (Liao et al., *in prep*) addresses two gaps, identifying SNPs linked to local adaptation and demonstrating that temperature and birds are key to WNV spread (Fig. 2). However, the national WNV prediction model is limited by non-continuous surveillance data, which is largely dependent on hospital reports.

To enhance the accuracy and utility of our models, my summer research will focus on two main objectives. Firstly, I will develop more precise regional models by harnessing data from states with more comprehensive and continuous surveillance records. This will allow for a more nuanced understanding of the temporal and spatial patterns of disease spread.

Secondly, I intend to shift the emphasis of our predictive analysis from the mere quantification of West Nile Virus cases to a more robust assessment of WNV risk levels across counties. To achieve this, I will employ advanced machine learning techniques that have shown promise in epidemiological modeling. Models such as neural network, Random Forest and ensemble model will be trained and validated using a rich dataset that integrates environmental factors, demographic variables, and, where possible, genetic information pertaining to the *Cx. tarsalis* populations at the county level. Through this integrated machine learning approach, I anticipate

gaining deeper insights into the drivers of disease transmission and improving our predictive capabilities.



**Fig 2: Correlation Heatmap of WNV Cases and Environmental Variables**

## Potential significance of the outcome

My research on *Cx. tarsalis* genetics and adaptations aids scientific understanding and public health. We're uncovering how genetics and environment affect this mosquito's distribution, informing improved control methods and reducing mosquito-borne diseases like West Nile Virus. Our innovative use of genomics and machine learning in disease prediction advances epidemiological tools, aiding outbreak preparedness. Our integrative approach enhances health policies and outcomes, showcasing the importance of biological research in disease management.

## Reference

Colpitts, T.M. *et al.* (2012) 'West Nile Virus: biology, transmission, and human infection', *Clinical microbiology reviews*, 25(4), pp. 635–648.

Gubler, D.J. (1998) 'Resurgent vector-borne diseases as a global health problem', *Emerging infectious diseases*, 4(3), pp. 442–450.

Goddard, L.B. *et al.* (2002) 'Vector competence of California mosquitoes for West Nile virus', *Emerging infectious diseases*, 8(12), pp. 1385–1391.

Evans, M.V. *et al.* (2017) 'Data-driven identification of potential Zika virus vectors', *eLife*, 6. Available at: <https://doi.org/10.7554/eLife.22053>.

Venkatesan, M. and Rasgon, J.L. (2010) 'Population genetic data suggest a role for mosquito-mediated dispersal of West Nile virus across the western United States', *Molecular ecology*, 19(8), pp. 1573–1584.

Pfeiler, E. *et al.* (2013) 'Genetic diversity and population genetics of mosquitoes (Diptera: Culicidae: Culex spp.) from the Sonoran Desert of North America', *TheScientificWorldJournal*, 2013, p. 724609.

Hadfield, J. *et al.* (2019) 'Twenty years of West Nile virus spread and evolution in the Americas visualized by Nextstrain', *PLoS pathogens*, 15(10), p. e1008042.

Yunfei Liao. *et al.* (In prep) 'Climate Adaptation and Genetic Differentiation in the Mosquito Species Culex tarsalis'.