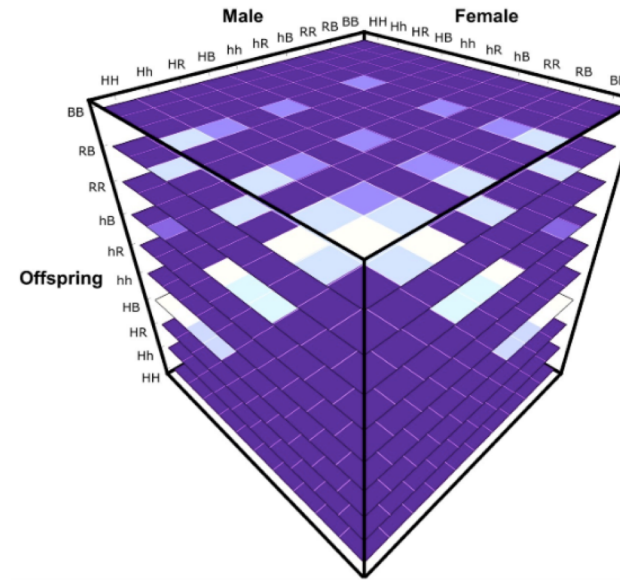
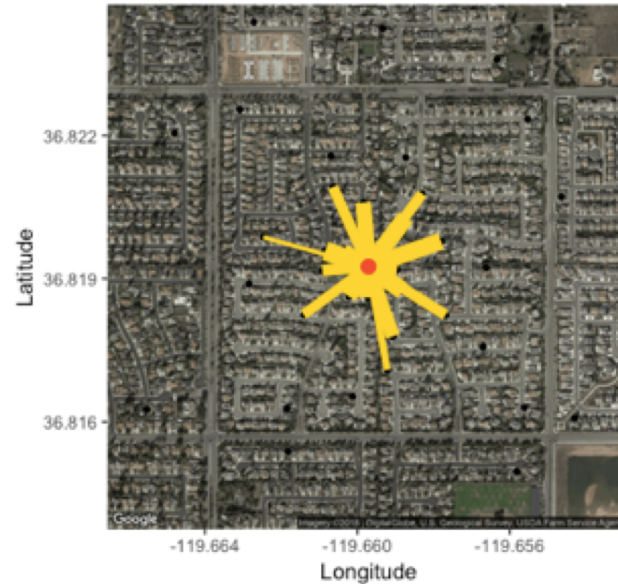
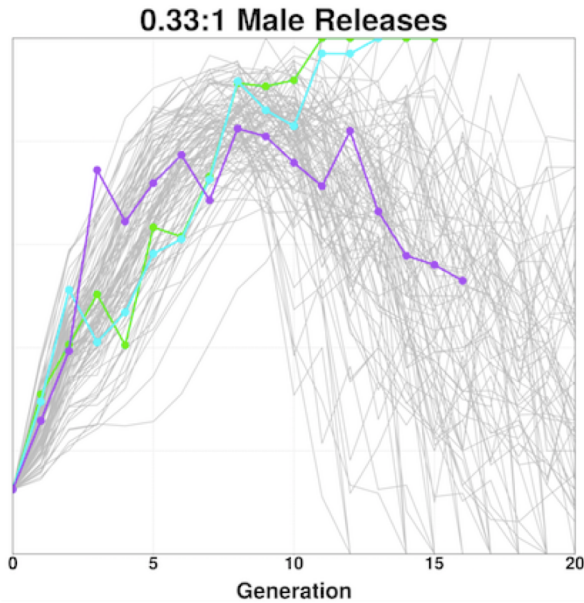


# Data sharing in gene drive research: Experiences of a mathematical modeler

Gene Drive Research Forum, Addis Ababa, Ethiopia, November 1<sup>st</sup>, 2019



**TACC**  
TEXAS ADVANCED COMPUTING CENTER

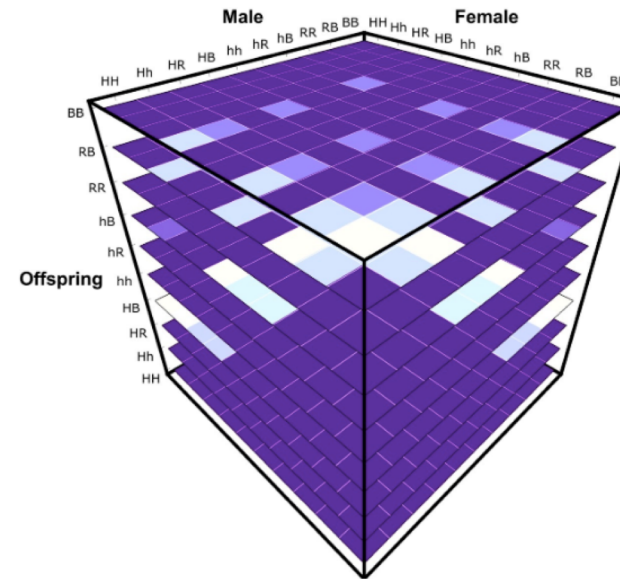
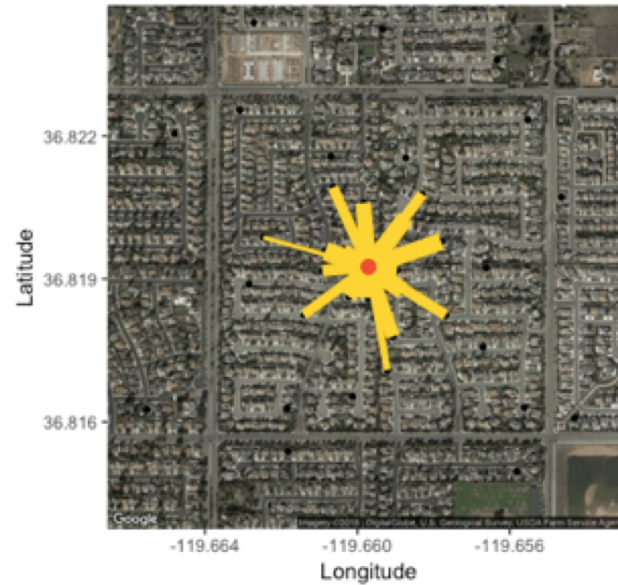
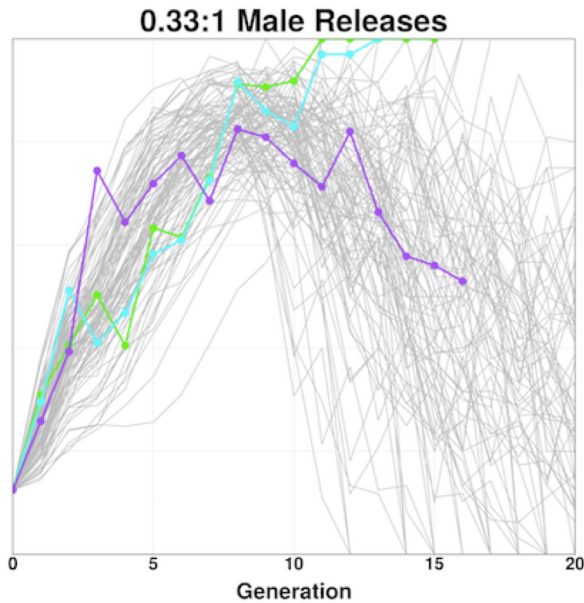


John M. Marshall  
[john.marshall@berkeley.edu](mailto:john.marshall@berkeley.edu)  
Division of Epidemiology & Biostatistics  
University of California, Berkeley

Berkeley

School of  
Public Health

# Talk outline



1. Data sharing experiences with molecular biologists

- Laboratory drive experiments
- Scientific publishing model

2. Data sharing experiences with ecologists

- Longstanding academic collaborations
- Experience with Verily Life Sciences

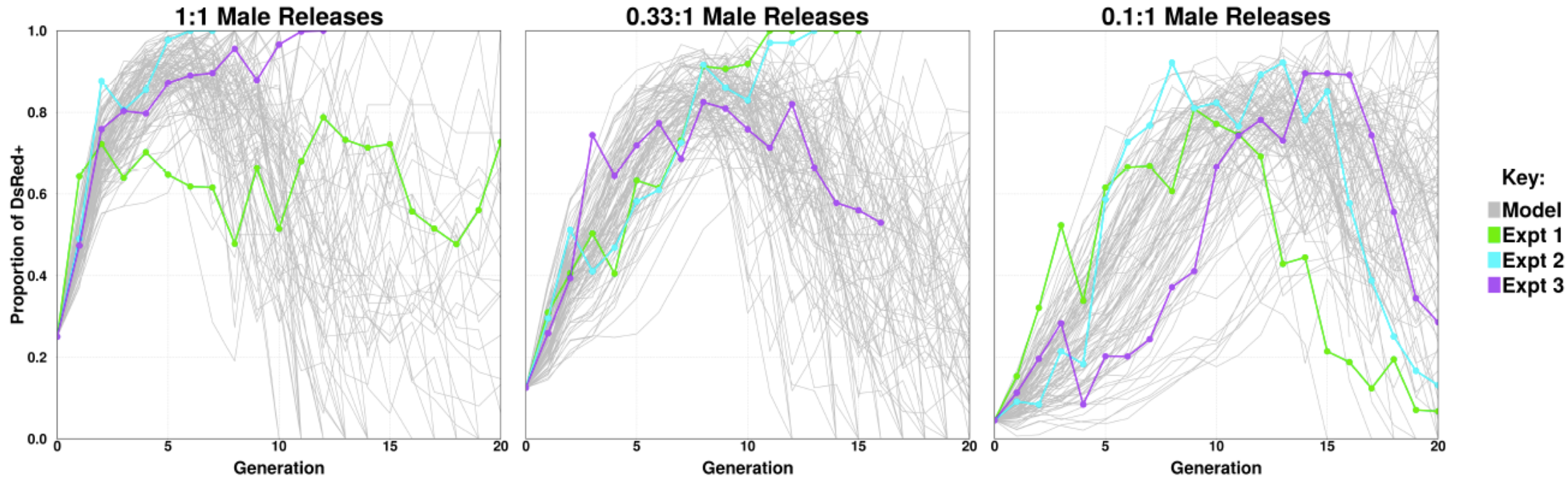
3. Open-source software development

- Mosquito Gene Drive Explorer
- Open-source software movement

4. Computing platforms for data sharing

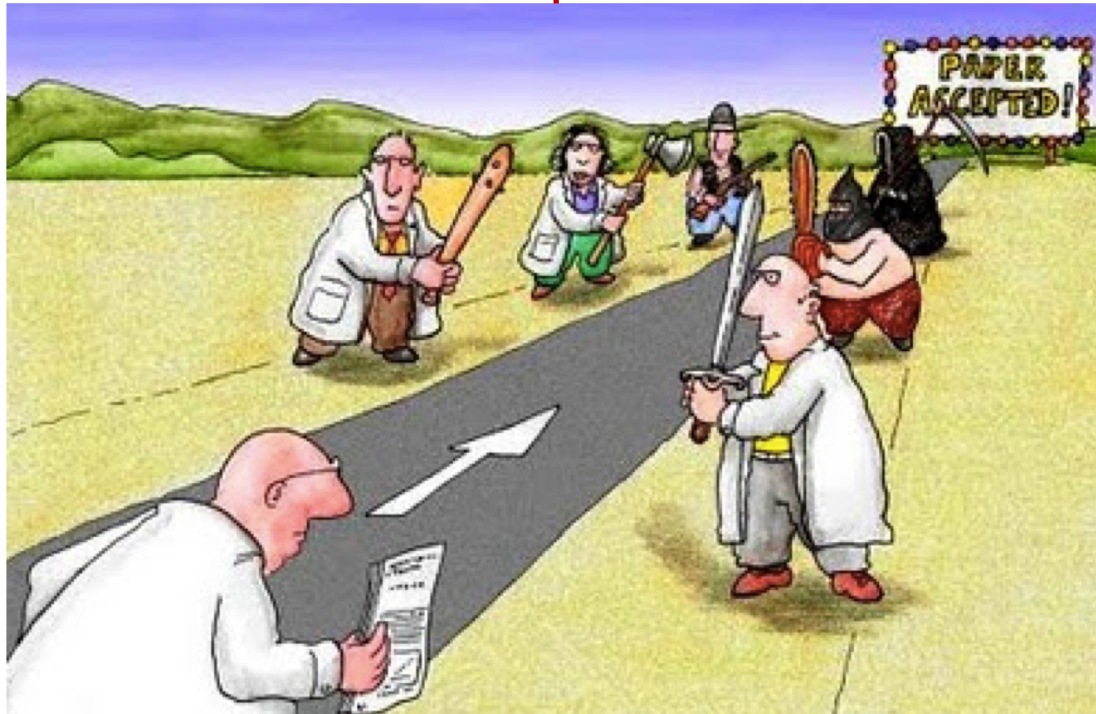
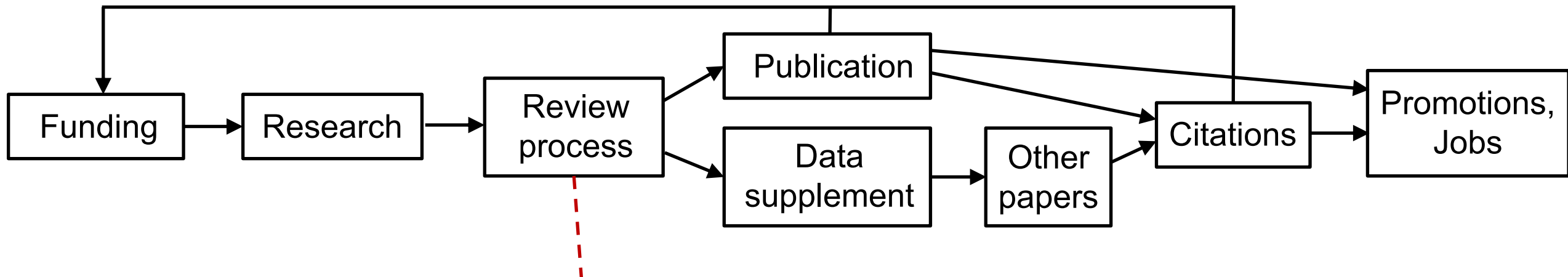
- Texas Advanced Computing Center
- Sharing permission
- Genomic data

# Data sharing experiences with molecular biologists



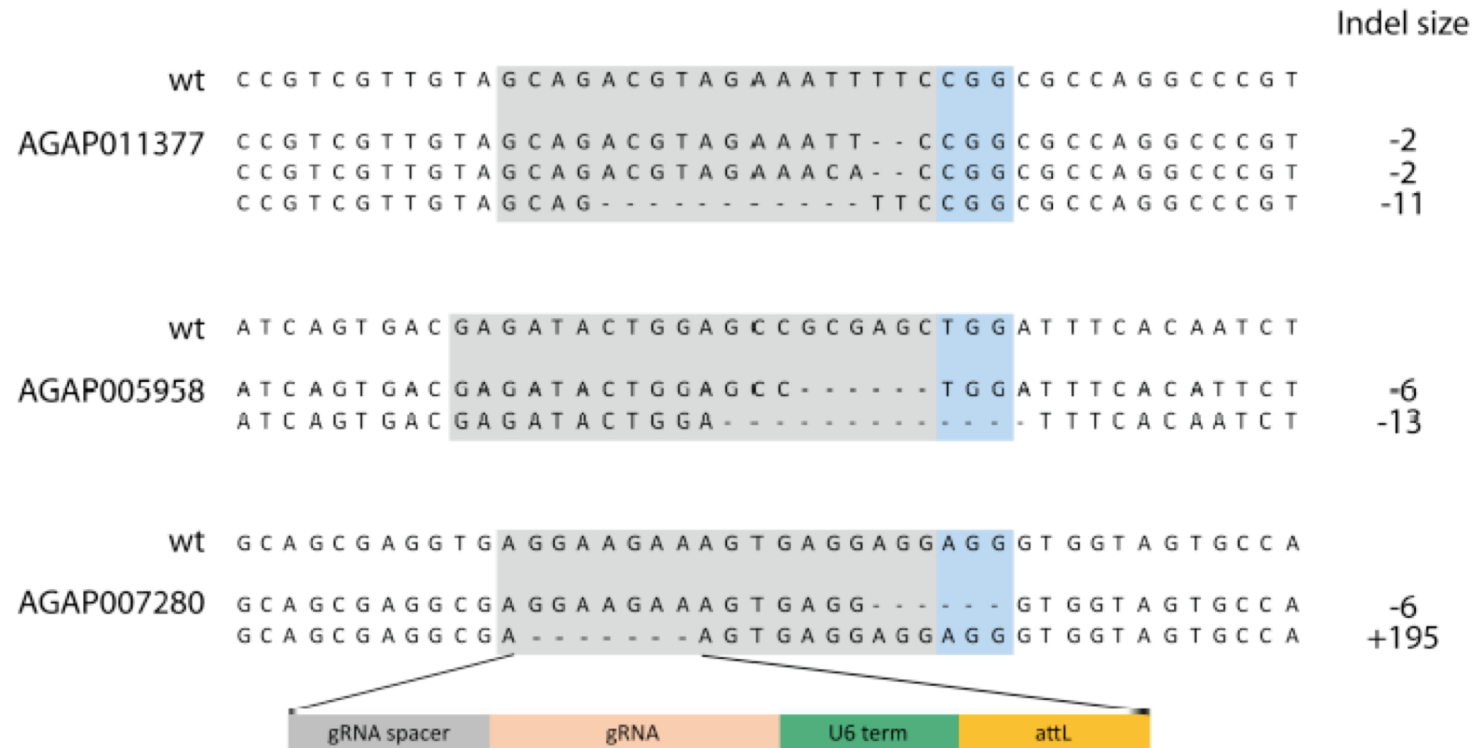
- Laboratory gene drive experiments provide some of the most useful data for validating our models of gene drive at this stage of technology development.
- By comparing model predictions to observed marker phenotypes over time, we can infer rates of resistant allele generation & relative fitness costs of allele variants.
- This is essential preceding sensitivity analyses for projecting expected dynamics in the field.

# Traditional scientific publishing model



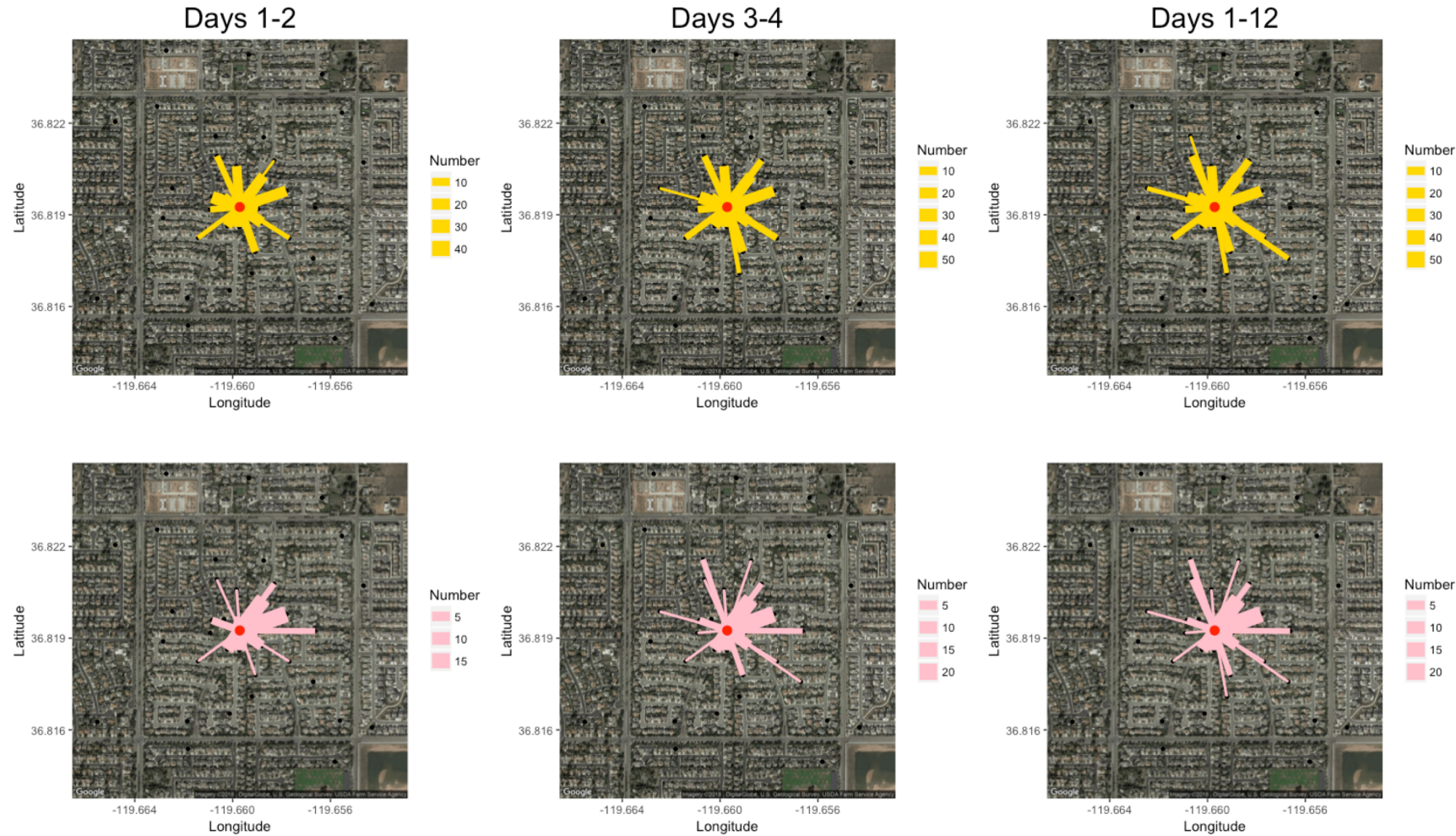


# Importance for CRISPR-based gene drive modeling



- CRISPR-based homing gene drive systems produce a range of allelic variants through mechanisms such as non-homologous end-joining (NHEJ).
- Current models don't characterize the full spectrum of these alleles that can emerge.
- Understanding allelic variants & their characteristics (impact on homing, fitness, etc.) is desirable prior to implementation & would benefit greatly from access to collated data sets.

# Data sharing experiences with mosquito ecologists



- Ecological data is essential for developing models of mosquito dispersal & habitat distribution.
- In combination with models of gene drive from the lab, these data & models can help predict how constructs should behave in the wild, which is of use to planning field trials & surveillance, etc.
- Many of the same data sharing considerations apply here as for molecular biology data.

# Additional issues for ecological data & field trials

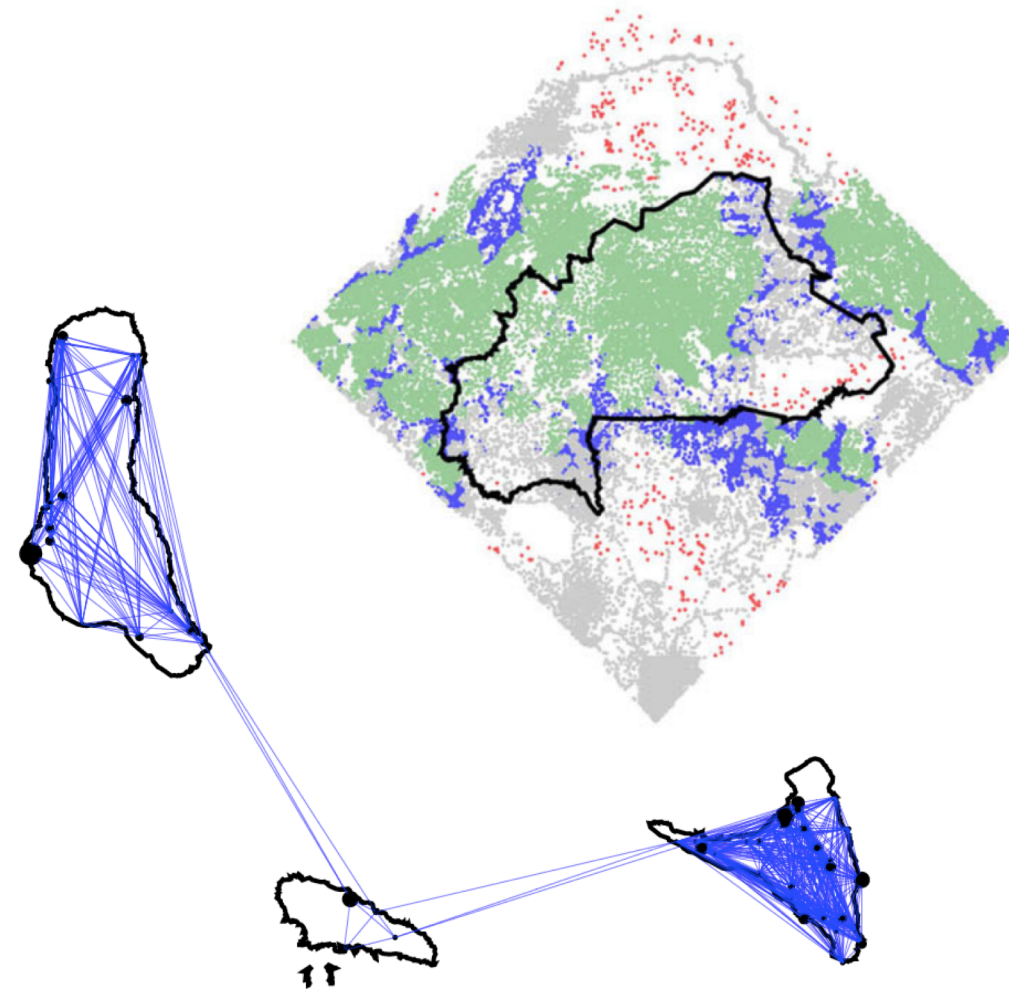
Am. J. Trop. Med. Hyg., 98(Suppl 6), 2018, pp. 1–49  
doi:10.4269/ajtmh.18-0083  
Copyright © 2018 by The American Society of Tropical Medicine and Hygiene

## Pathway to Deployment of Gene Drive Mosquitoes as a Potential Biocontrol Tool for Elimination of Malaria in Sub-Saharan Africa: Recommendations of a Scientific Working Group†

Stephanie James,<sup>1\*</sup> Frank H. Collins,<sup>2</sup> Philip A. Welkhoff,<sup>3</sup> Claudia Emerson,<sup>4</sup> H. Charles J. Godfray,<sup>5</sup> Michael Gottlieb,<sup>1</sup> Brian Greenwood,<sup>6</sup> Steve W. Lindsay,<sup>7</sup> Charles M. Mbogo,<sup>8</sup> Fredros O. Okumu,<sup>9,10,11</sup> Hector Quemada,<sup>12</sup> Moussa Savadogo,<sup>13</sup> Jerome A. Singh,<sup>14</sup> Karen H. Tountas,<sup>1</sup> and Yeya T. Touré<sup>15</sup>

<sup>1</sup>Foundation for the National Institutes of Health, Bethesda, Maryland; <sup>2</sup>University of Notre Dame, Notre Dame, Indiana; <sup>3</sup>Institute for Disease Modeling, Bellevue, Washington; <sup>4</sup>McMaster University, Hamilton, Canada; <sup>5</sup>Oxford University, Oxford, United Kingdom; <sup>6</sup>London School of Hygiene & Tropical Medicine, London, United Kingdom; <sup>7</sup>Durham University, Durham, United Kingdom; <sup>8</sup>Kenya Medical Research Institute, Nairobi, Kenya; <sup>9</sup>Ifakara Health Institute, Ifakara, Tanzania; <sup>10</sup>University of Glasgow, Glasgow, Scotland; <sup>11</sup>University of the Witwatersrand, Johannesburg, South Africa; <sup>12</sup>Donald Danforth Plant Science Center, Saint Louis, Missouri; <sup>13</sup>New Partnership for Africa's Development, Ouagadougou, Burkina Faso; <sup>14</sup>Centre for the AIDS Programme of Research in South Africa, Durban, KwaZulu-Natal, South Africa; <sup>15</sup>University of Sciences, Techniques and Technologies of Bamako, Bamako, Mali

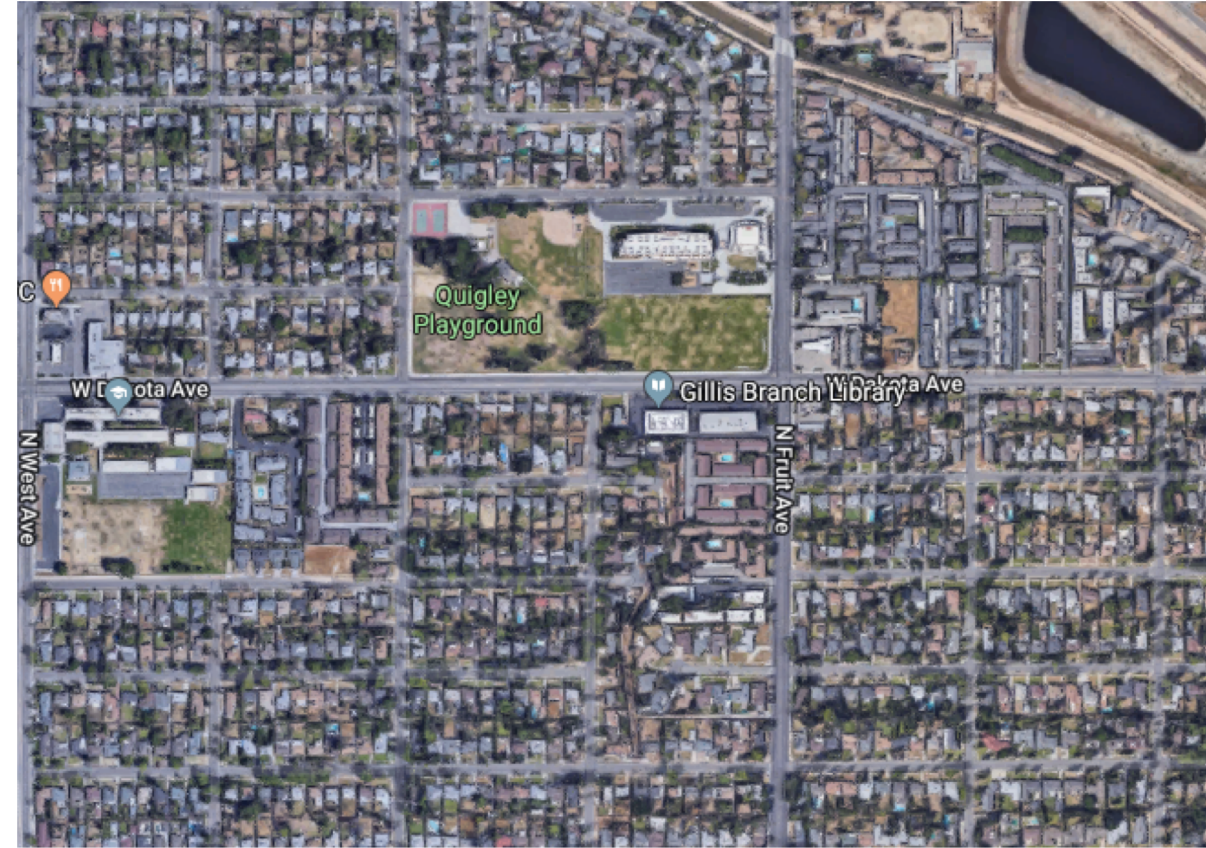
**Abstract.** Gene drive technology offers the promise for a high-impact, cost-effective, and durable method to control malaria transmission that would make a significant contribution to elimination. Gene drive systems, such as those based on clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated protein, have the potential to spread beneficial traits through interbreeding populations of malaria mosquitoes. However, the characteristics of this technology have raised concerns that necessitate careful consideration of the product development pathway. A multi-disciplinary working group considered the implications of low-threshold gene drive systems on the development pathway described in the World Health Organization *Guidance Framework for testing genetically modified (GM) mosquitoes*, focusing on reduction of malaria transmission by *Anopheles gambiae* s.l. mosquitoes in Africa as a case study. The group developed recommendations for the safe and ethical testing of gene drive mosquitoes, drawing on prior experience with other vector control tools, GM organisms, and biocontrol agents. These recommendations are organized according to a testing plan that seeks to maximize safety by incrementally increasing the degree of human and environmental exposure to the investigational product. As with biocontrol agents, emphasis is placed on safety evaluation at the end of physically confined laboratory testing as a major decision point for whether to enter field testing. Progression through the testing pathway is based on fulfillment of safety and efficacy criteria, and is subject to regulatory and ethical approvals, as well as social acceptance. The working group identified several resources that were considered important to support responsible field testing of gene drive mosquitoes.



1. Data sharing between projects may enhance our understanding of relevant ecosystem features.
2. Openness with field data, as appropriate, can build trust & accountability with local communities.



# Data sharing experiences with Verily Life Sciences



- In working with a private company, data sharing is more restricted, but productive collaborations are still possible.
- Verily's MRR data on mass released, *Wolbachia*-infected mosquitoes provides important information on how other varieties of mass released *Aedes aegypti* may disperse.
- We have a pre-publication data sharing agreement through a DARPA Safe Genes project.

# MGDrivE as an open-source software package



Mosquitos + Tensors + Genetics + CS +  
Networks + Math + Coffee

View on  
CRAN

Browse  
Documentation

View on  
Youtube

Fork on  
Github

Developed in [John Marshall's Lab](#) by:

Lead: [Héctor M. Sánchez C.](#)

Core Dev: [Sean L. Wu](#), [Jared Bennett](#)

Spatial Analysis: [Gillian Chu](#), [Maya Shen](#), [Yunwen Ji](#)

Environmental Factors: [Tomás León](#)

Releases Optimization: [Valeri Vasquez](#)

Aux Dev: [Chase Violet](#)

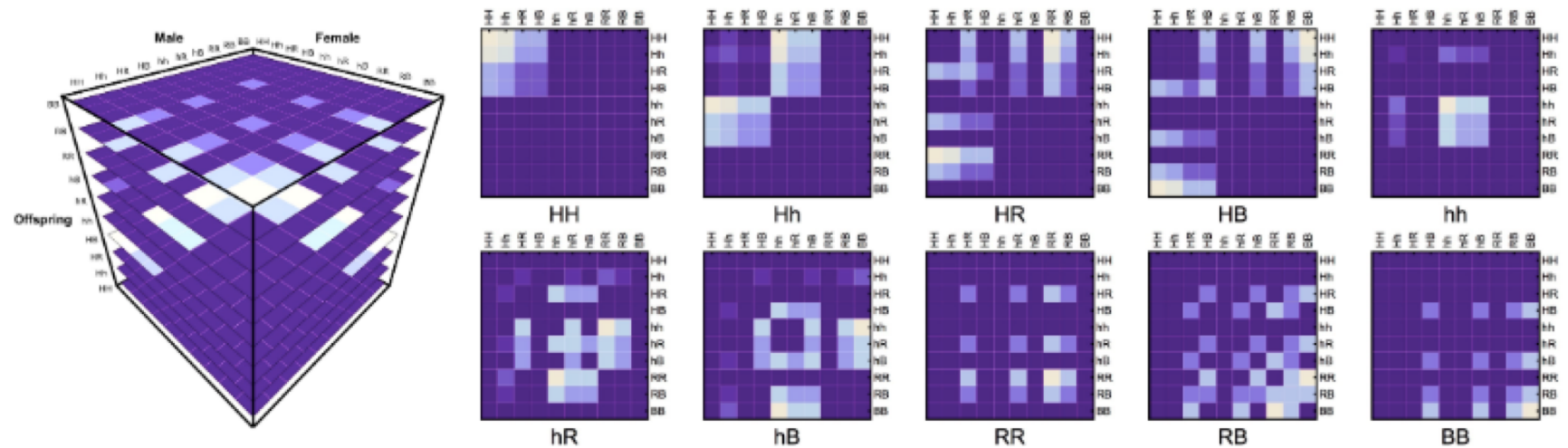
Former Members: [Víctor Ferman](#),

[Biyonka Liang](#), [Sarafina Smith](#), [Sabrina Wong](#)

## Mosquito Gene Drive Explorer

### Brief Description

MGDrivE is a framework designed to serve as a testbed in which gene-drive releases for mosquito-borne diseases control can be tested. It is being developed to accommodate various mosquito-specific gene drive systems within a population dynamics model that allows migration of individuals between nodes in a spatial landscape.



### Demonstration

# MGDrivE as an open-source software package

New Results

[View current version of this article](#)

[← Previous](#)

## **MGDrivE: A modular simulation framework for the spread of gene drives through spatially-explicit mosquito populations**

[Comment on this paper](#)

Posted June 19, 2018.

Héctor M. Sánchez C., Sean L. Wu, Jared B. Bennett, John M. Marshall

doi: <https://doi.org/10.1101/350488>

[Download PDF](#)

[Supplementary Material](#)

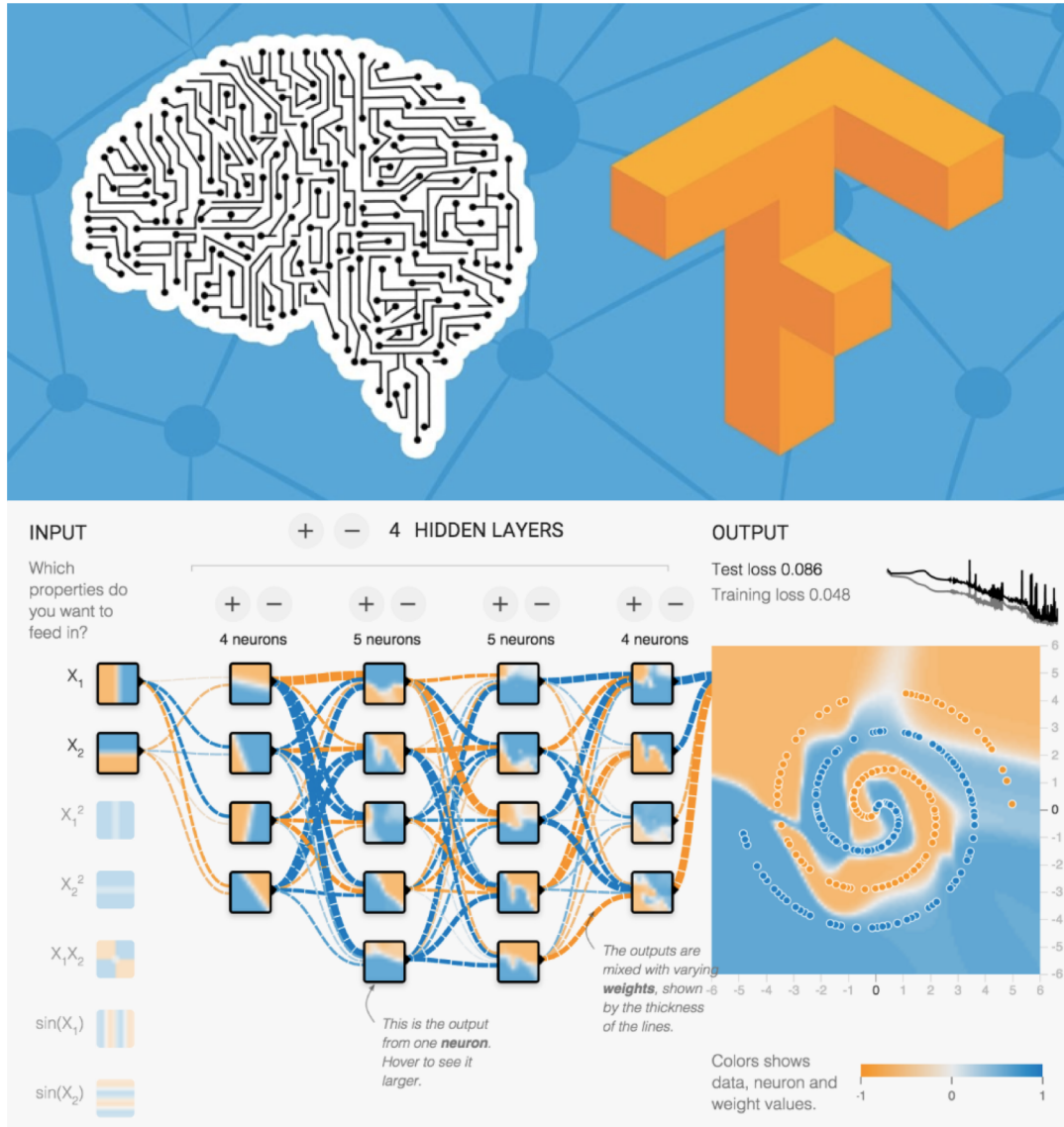
Now published in *Methods in Ecology and Evolution* doi: [10.1111/2041-210X.13318](https://doi.org/10.1111/2041-210X.13318)

[Tweet](#)

[Like 0](#)

- As an example of review process delays, the MGDrivE paper was published 2 weeks ago, but was originally uploaded to bioRxiv & GitHub a year & a half ago.
- This allowed us to be cited & to make the documentation and code openly available, while the paper was being peer-reviewed.

# Free & open-source software movements



## Strengths:

- Makes model building & coding work available to the wider research community.
- Promotes adoption of the framework.
- Offers flexibility & customizability.
- Provides a coding environment that is well suited to collaboration.
- Provides a means to outsource debugging effort.

## Weaknesses:

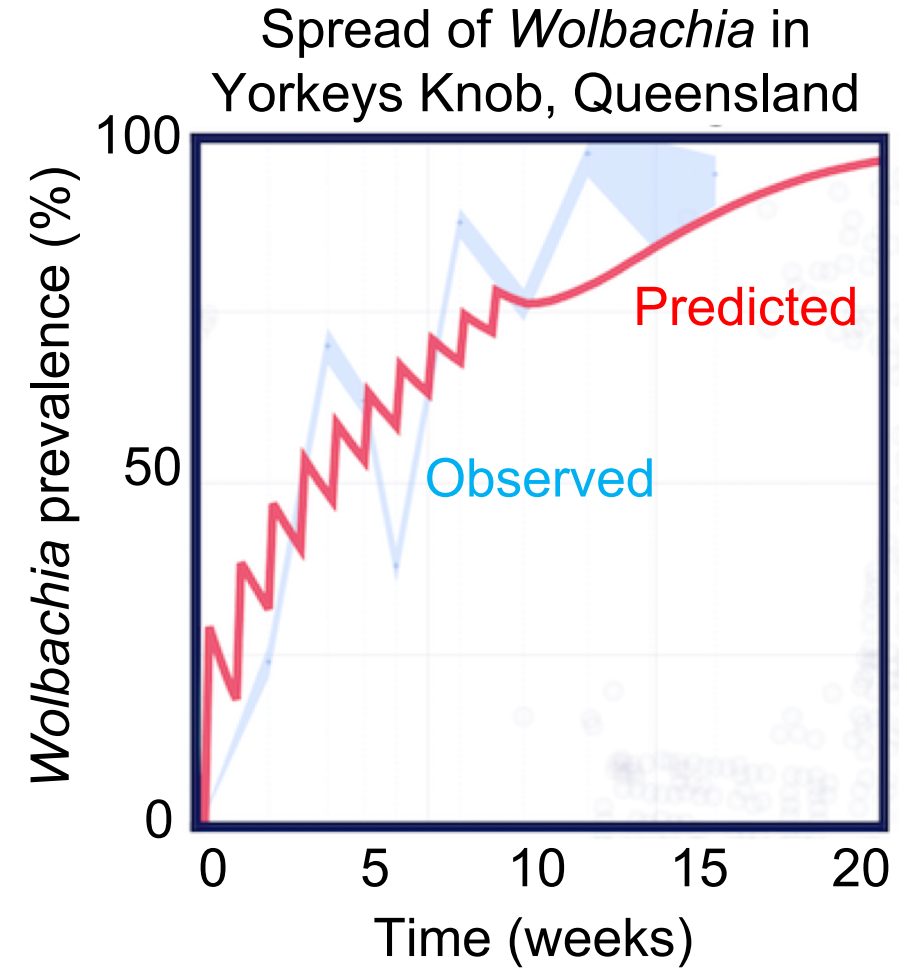
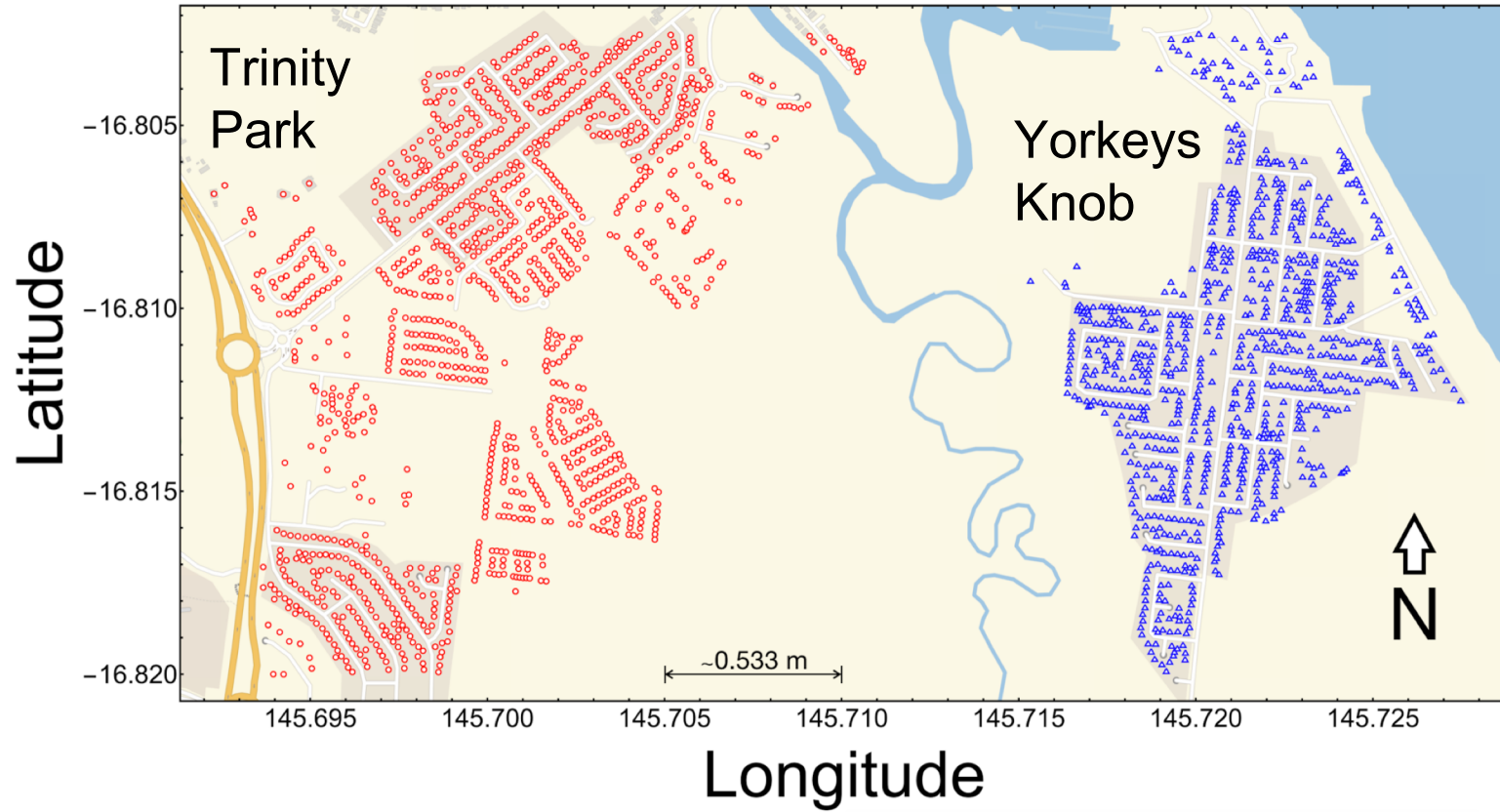
- Potential for other researchers to address research questions that you had planned to.
- Steep learning curve to begin contributing to code development.
- Core development team should be paid to avoid developer abandonment.

# Computing platforms for data sharing



- DARPA Safe Genes teams were granted access to the TACC as a means for data sharing & collaboration.
- For DARPA, this is also a means to demonstrate synergy between the Safe Genes & SD2 programs.
- The TACC provides an excellent environment for high performance computing, data storage & analysis.
- Files & folders can be given viewing & editing permissions, as appropriate, to: i) lab members, ii) project team members, iii) members of collaborating teams, iv) the entire Safe Genes community, or v) the public.
- This provides a useful model for a data storage & analysis platform that could serve the gene drive research community.

# Running large simulations on TACC servers



# Defining sharing permissions in directories

### Test Project XYZ

slindsey

[← Back to Projects List](#)

[Request New Allocation](#)

[Add User](#)

[Add Publication](#)

[Add Grant](#)

[Edit Project Information](#)

Principal Investigator: Susan Lindsey 

Project Type: Startup

Unix Group: G-816631 

Field of Science: Classical Analysis Classical Analysis

Abstract/Descr

I. Set the group id bit. Any new files created in the shared directory will inherit the group ownership:

```
login$ chmod g+s mysharedirectory
```

II. Then edit your ".bashrc" and set the umask variable to "027". This ensures that all NEW files created will inherit the proper permissions.

```
umask 027
```

III. Last, set permissions to readable and accessible to group members:

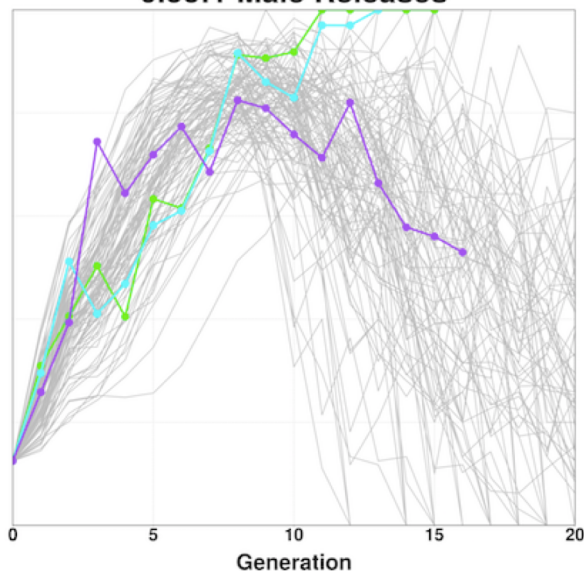
```
login$ chmod g+rX mysharedirectory
```

- Clear instructions are available in user manuals on the the TACC website.
- This can assist in data sharing between:
  - Lab members,
  - Project team, members,
  - Members of collaborating teams,
  - Safe Genes community, or
  - The public.

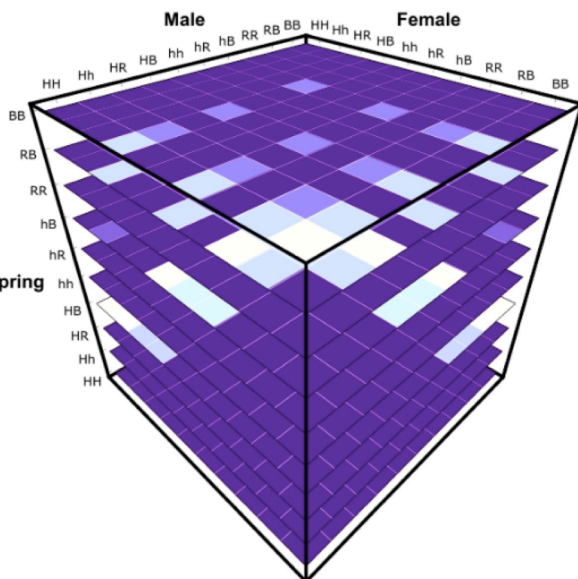
# Conclusion

- Data exchange is crucially important for advancing research.
- Exchange can be accelerated through mechanisms that bypass the publication timeline, e.g.:
  - Data papers,
  - Pre-publication data sharing.
- Infrastructure is required to incentivize these mechanisms (editorial, funding, hiring & promotions).
- Open-source software development:
  - Advances research,
  - Is well suited to collaboration,
  - Offers flexibility & customizability.
- Computing platforms like the TACC:
  - Useful for data storage & analysis,
  - Offer a means to implement & control data sharing.

0.33:1 Male Releases



**TACC**  
TEXAS ADVANCED COMPUTING CENTER





# Acknowledgements

## LAB MEMBERS:

- Héctor M. Sánchez C.
- Yogita Sharma
- Tomás León
- Gordana Rašić
- Sanjay Lamba
- Victor Ferman
- Samson Kiware
- Sean L. Wu
- Jared B. Bennett
- Valeri Vasquez
- Francois Rerolle
- Thien-An Ha
- Yi Li
- Biyonka Liang
- Gillian Chu
- Maya Shen
- Sarafina Smith
- Sejal Mohata

## COLLABORATORS:

- **Akbari Lab @ UC Riverside**
- **Lanzaro Lab @ UC Davis**
- **James Lab @ UC Irvine**
- **Bier Lab @ UCSD**
- **Hay Lab @ Caltech**
- **Malaria Elimination Initiative @ UCSF**
- **Prof David Smith @ IHME, UW**
- **School of Public Health @ UC Berkeley**

## FUNDERS:

