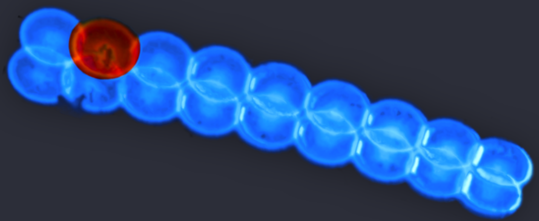
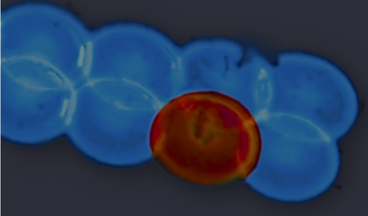
Project PREMONITION

CPS for Disease Surveillance

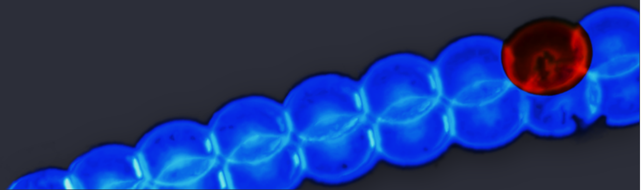
Ethan Jackson – Microsoft Research





Emerging Infectious Diseases

And technologies to stop them.



The Threat of Disease Emergence

Emerging infectious diseases pose significant health and economic threats.

SARS in China (2003):

$6.2 billion / 8000 = $775,000
cost cases

H5N1 (2003-2009):

$20 billion / 468 = $43 million
cost cases

Ebola in US (2014):

$2.8 billion / 4 = $700 million
cost cases

Need Better Surveillance of Diseases

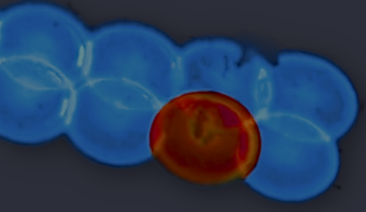
But surveillance of EIDs is particularly challenging because:

- Majority of threats caused by animal pathogens
  Over 60% of emergence events
- Best existing systems detect outbreaks too late
  Median delay of 13.5 days
- Many emerging diseases were previously unknown
  Over 70% of viruses in wild are unknown

Actually Need Surveillance in The

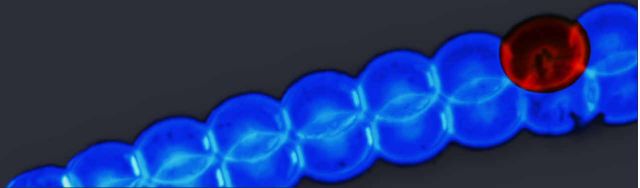
To see which pathogens are circulating in animals before they impact humans

Spoiler: Put us on a path towards an ambitious, massive, self-adapting CPS



Project PREMONITION

Preventative Monitoring of **I**nfecti**o**us Age**n**ts



Mosquito-as-a-Device

Use mosquitoes as devices that collect blood samples from animals in the wild

[illegible]
[illegible]
continent but Antarctica.

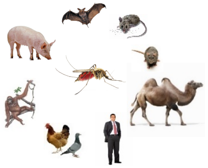
[illegible]
[illegible] plants.
[illegible] scale.

Mosquito-as-a-Device

Use mosquitoes as devices that collect blood samples from animals in the wild

- Mosquitoes grow naturally in rural and urban environments

On average live 20 days, consume 2.5 µl per blood meal, can fly several miles, and are geographically widely distributed. They have advanced olfactory systems to locate hidden prey.



- Mosquitoes sample the genes of animals and their pathogens

In studies, over 70% of viral pathogens in mosquitoes are unknown to science and come from a wide range of hosts, including humans, ducks, geese, cows, and plants.

- Can leverage classical entomological methods

Classical field entomology has been used by researchers, governments, and militaries to efficaciously surveil pathogens carried by mosquitoes. Must re-invent these methods to achieve scaled required for PREMONITION.

Key Innovations

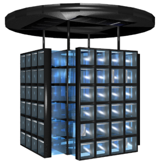
A number of technological innovations are required for scalability

[illegible] dics to
[illegible] tubes

[illegible] to
[illegible] require manual processing.

Key Innovations

A number of technological innovations are required for scalability



- Smarter Traps that automate field biology

  Existing mosquito traps are too low-throughput. They need to be placed by experts, remain in the environment for 12 – 18 hours, are heavy, and require manual processing.

- Drone-based autonomous deployment for high-throughput

  It takes too long to place traps. Ideal places to deploy traps are difficult for humans to reach. It make take several hours to place and collect a trap.

- Better metagenomics analytics to automatically identify threats

  Today humans must sift through lots of biological data to pick out candidate threats. This would not scale for high volumes of field collections.

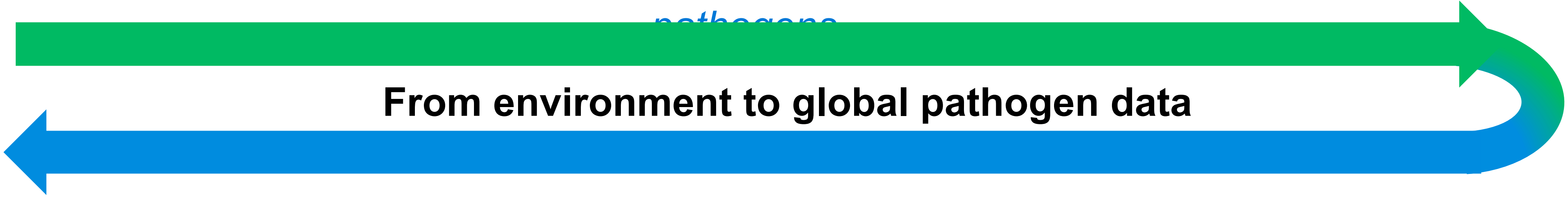
PREMONITION

High-throughput and low-cost monitoring known and unknown pathogens in the environment via autonomous collection of mosquitoes.

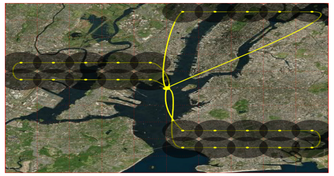
Animals maintain pathogens in environment

Mosquitoes collect blood, sampling many pathogens

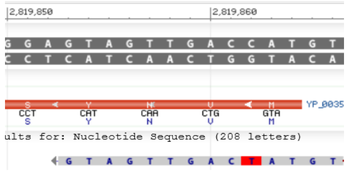
Drones deploy autonomous mosquito traps



From environment to global pathogen data

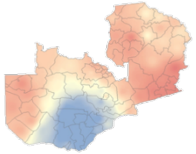


Long-range aerial or ground systems deploy short-range drones



ults for: Nucleotide Sequence (208 letters)

Collected mosquitoes are gene sequenced and pathogens detected

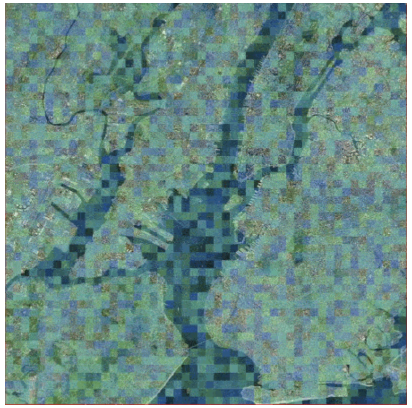


Global pathogen map of genes in space × time

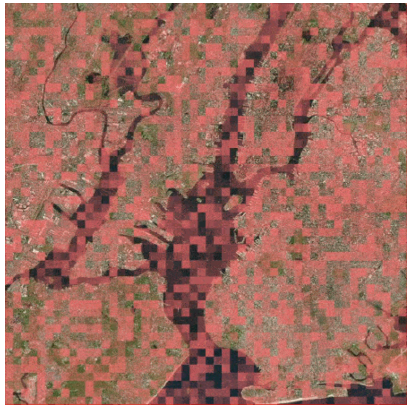
PREMONITION

Cloud-based genomic data analytics model epidemiology of emerging threats.

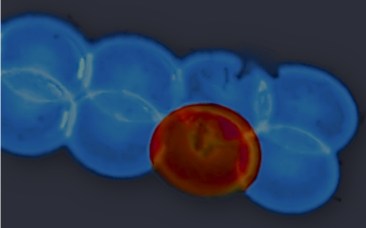
Example: Movement of potential pathogen and host observed through mosquitoes



Spatial/temporal gene data of hosts and vectors

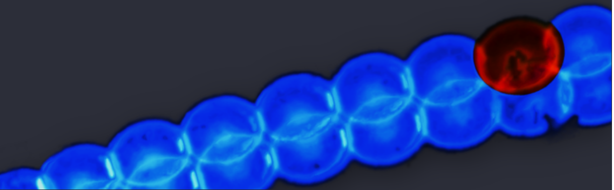


Spatial/temporal gene data of potential pathogens



The Grenada Experiment

Data and requirements from the field

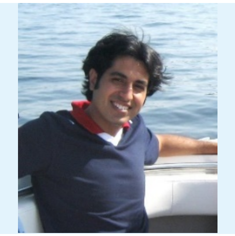


The Team

Amy Baldwin
St. George's University
Department of Microbiology

Jonathan Carlson
Microsoft Research
Computational Genomics

Ethan Jackson
Microsoft Research
Research in Software Engineering Group



Ashish Kapoor
Microsoft Research
Adaptive Systems and Interaction Group

Eamonn Keogh
University of California Riverside
Department of Computer Science and Engineering

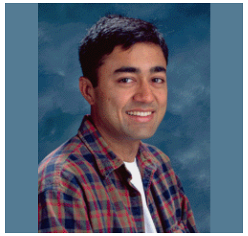
Shawn Keshmiri
University of Kansas
Department of Aerospace Engineering

Vijay Kumar
University of Pennsylvania
GRASP Laboratory

Douglas Norris
Johns Hopkins University
Department of Microbiology and Immunology

James Pipas
University of Pittsburgh
Department of Molecular Biology

The Team



Shaz Qadeer
Microsoft Research
Research in Software
Engineering Group

Anandansankar Ray
University of California
Riverside
Department of Entomology

Alex Ching
Microsoft Research
Hardware Laboratory

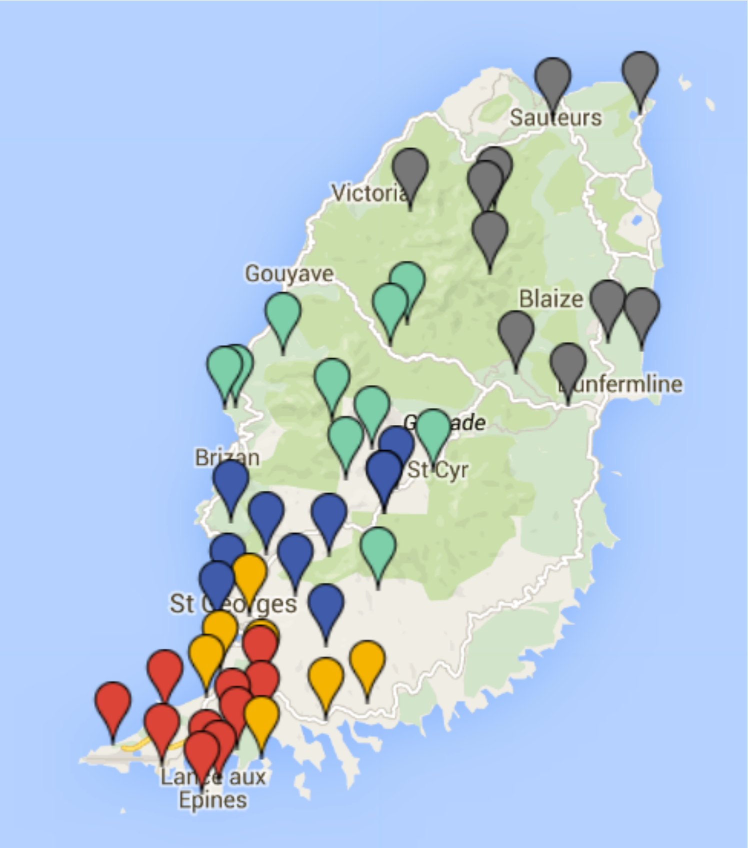
Janos Sztipanovits
Vanderbilt University
Institute for Software
Integrated Systems

Michael Zyskowski
Microsoft Research
Outreach

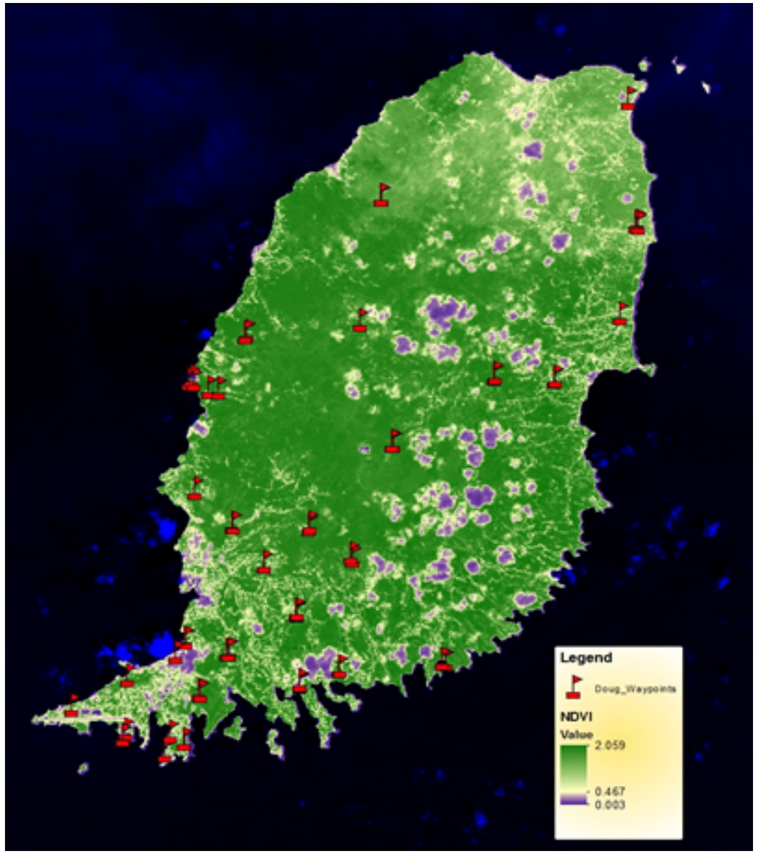
Patrick Therien
Microsoft Research
Hardware Laboratory

Sampling Grenada

Manual pilot in Grenada to understand feasibility and engineering constraints.
50 collections at 30 sites in urban, peri-urban, beach, and forest habitats



Planned



Actual (overlaid on 30m NDVI satellite data, courtesy Hopkins School of Public


SURVIVOR

Protocol – CO2 Baited CDC UV Trap

Placed during the day and collected the following morning.
Time in field: ~18 hours.

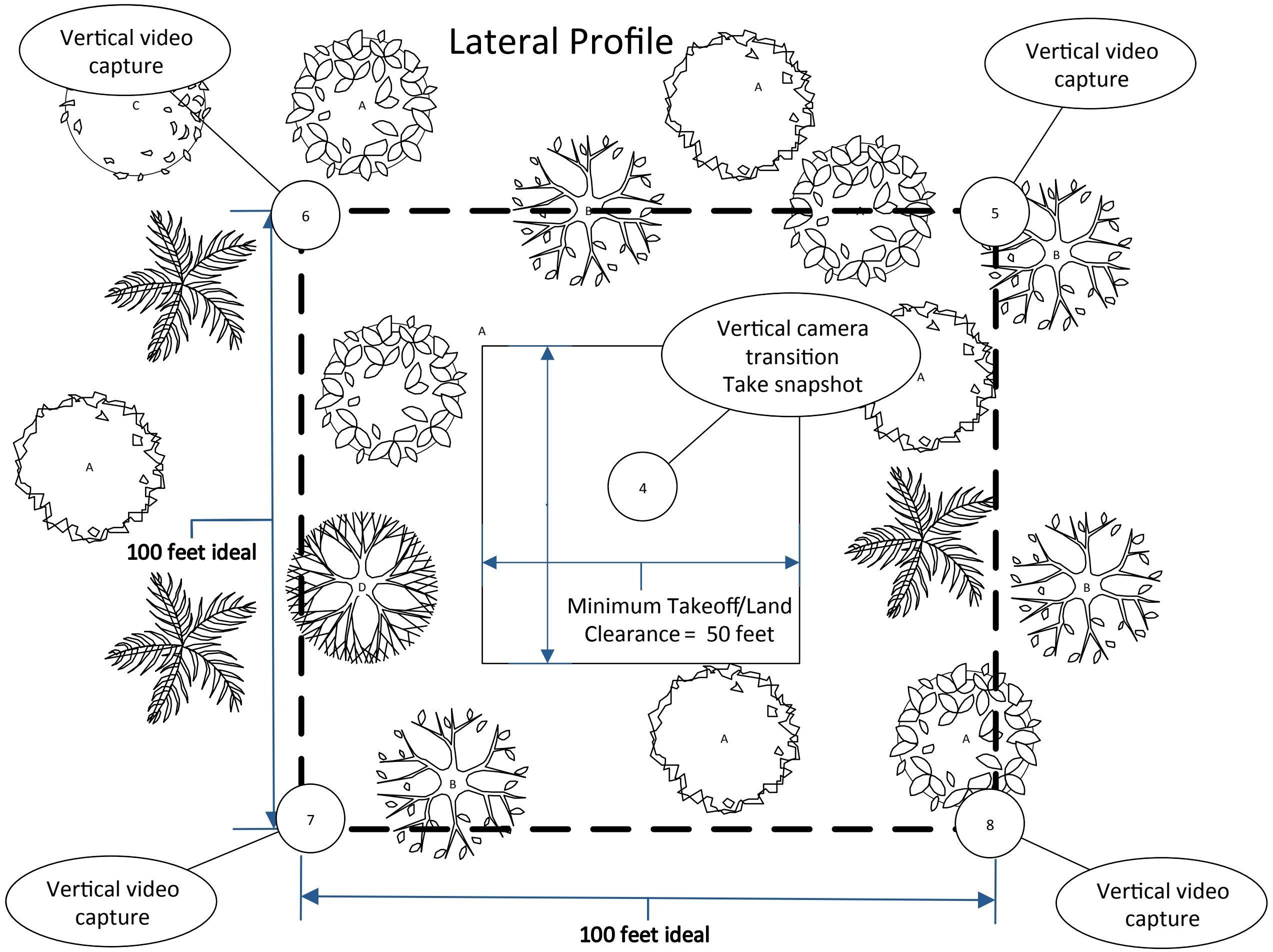
Urban

Beach

Jungle

Protocol – Drone Imagery

Manually collected drone imagery around the site to build mosquito hotspot classifiers, tree line detectors, and landing site detectors

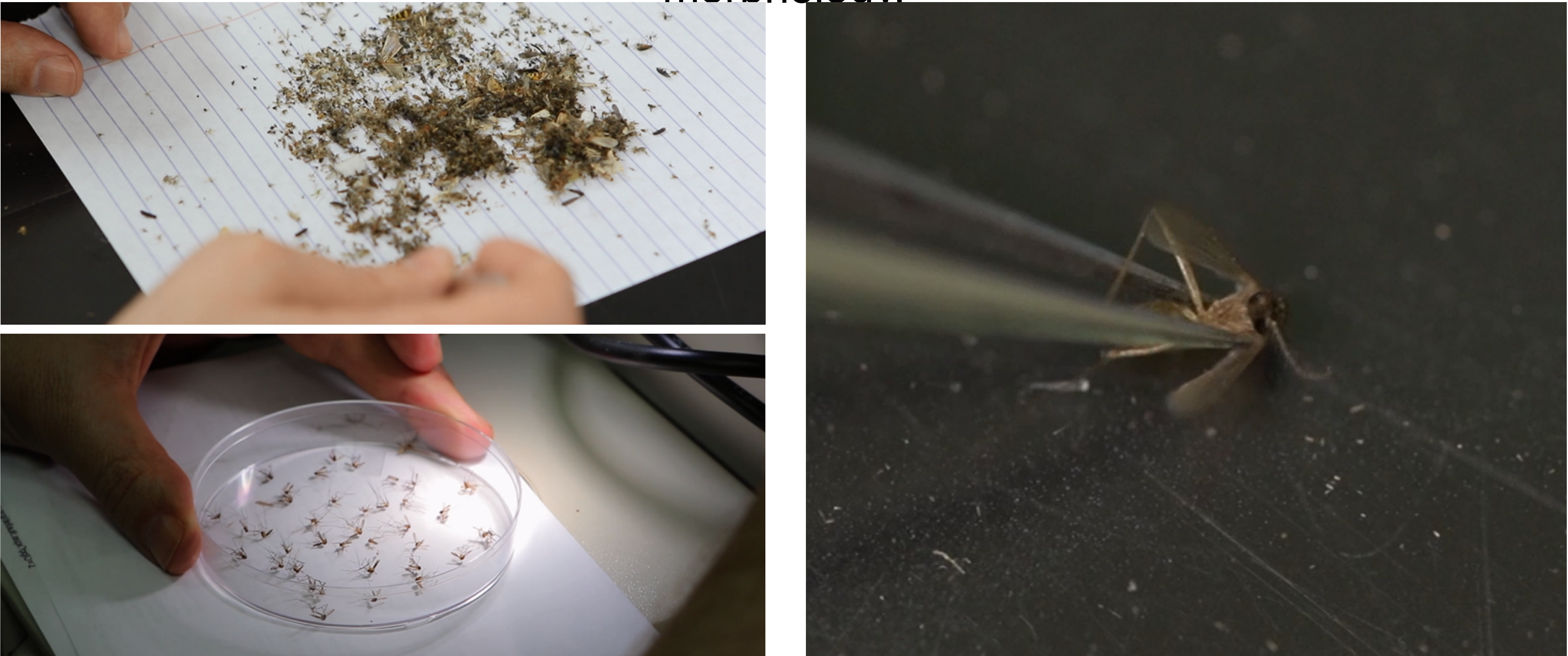


Planned flight plans

Actual flights

Protocol – Sorting

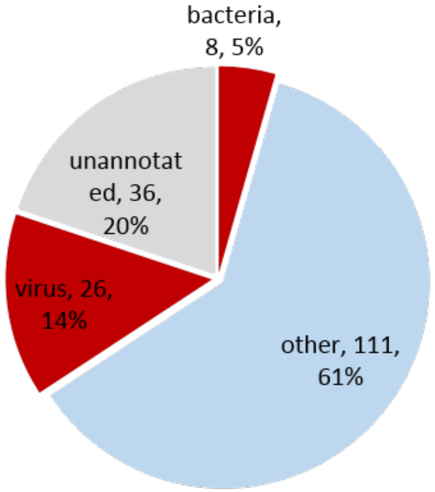
Samples sorted by genus; subset of Culex and Deinocerites sorted by morphology.

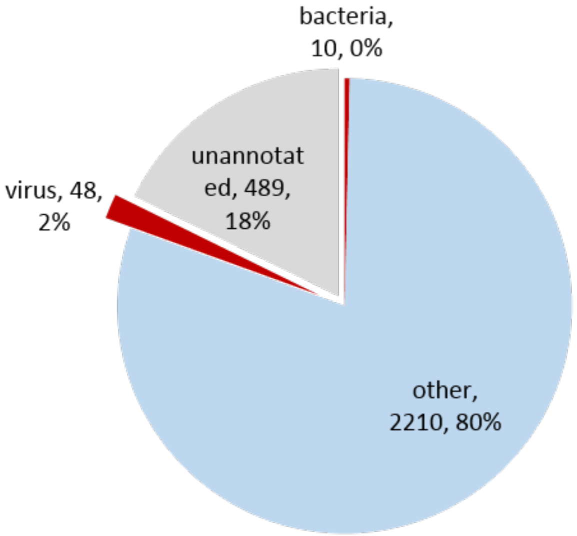


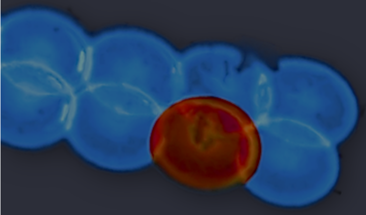
St. George's University

Protocol – Nucleic Acid Extraction

DNA/RNA aliquots deep sequenced to be computationally analyzed.

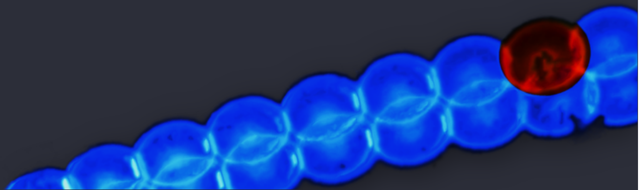






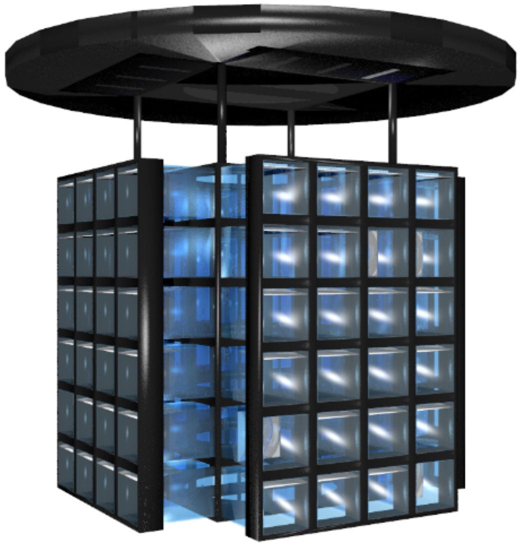
Building PREMONITION

Autonomous Traps

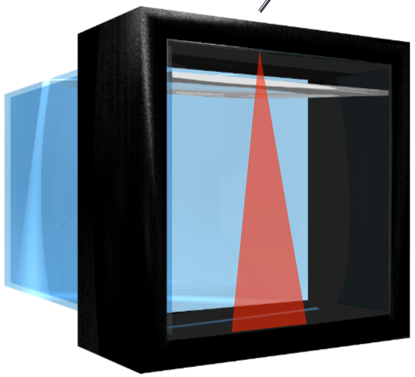


A Smarter Trap

Real-time speciation and selective trapping of insects; in-situ stabilization of nucleic acids; low-cost and low-power sensor and actuators.



Honeycomb design simplifies automated processing



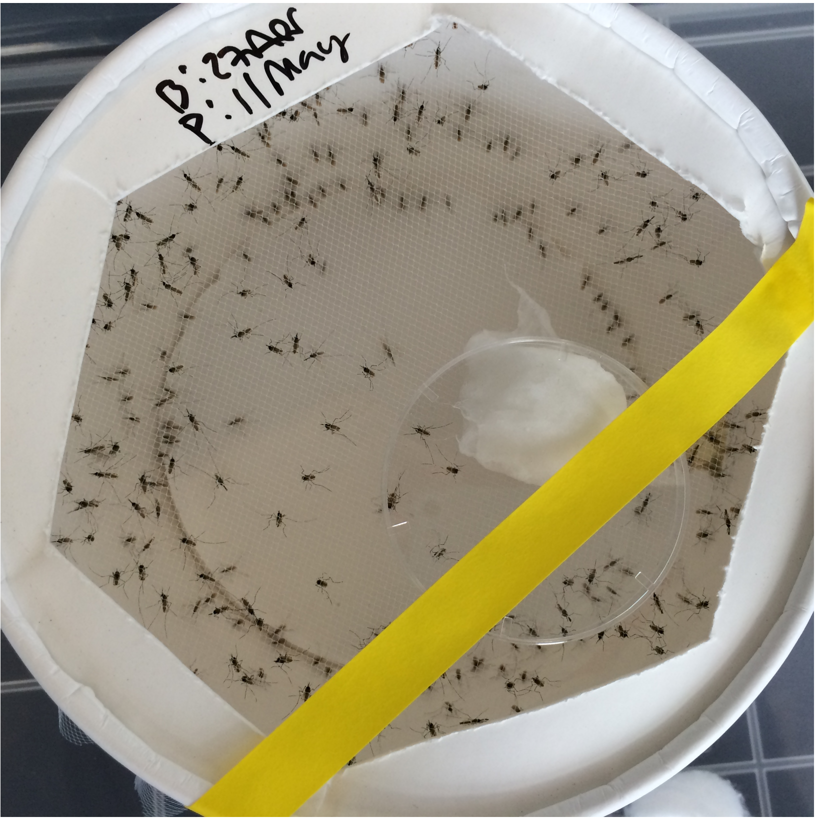
Low-cost IR emitters / photo diodes listen to insect wings; light outside of mosquito visible spectrum (>650nm)

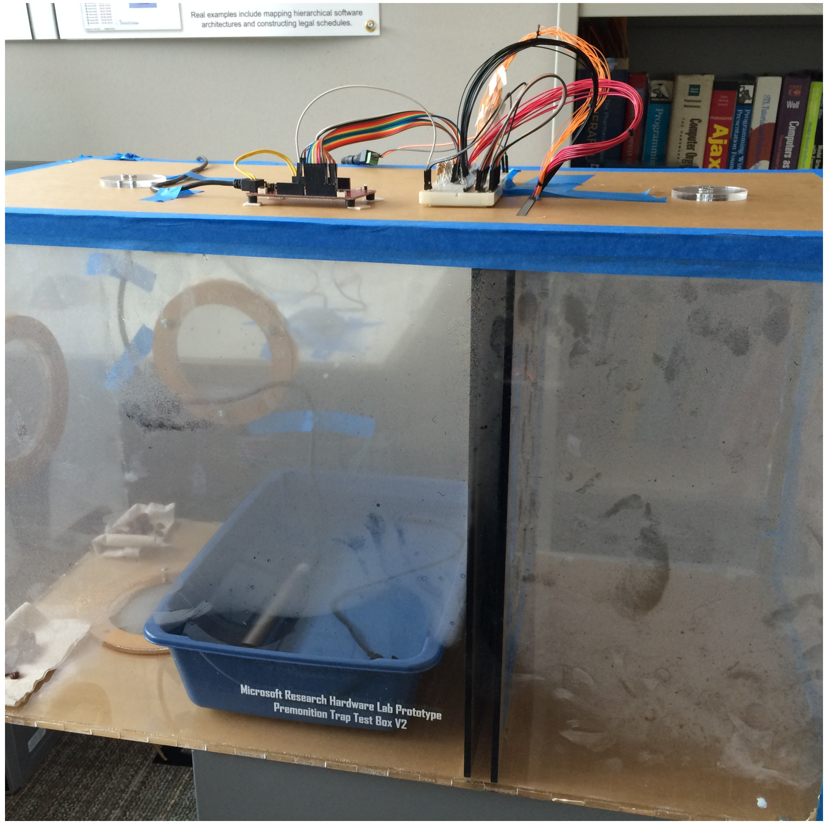
In-situ chemical stabilization of insect preserves sample for up to one week

Door closes when species of interest enters cell. Signal processing delay ~100ms

Designed and tested sensors and actuators

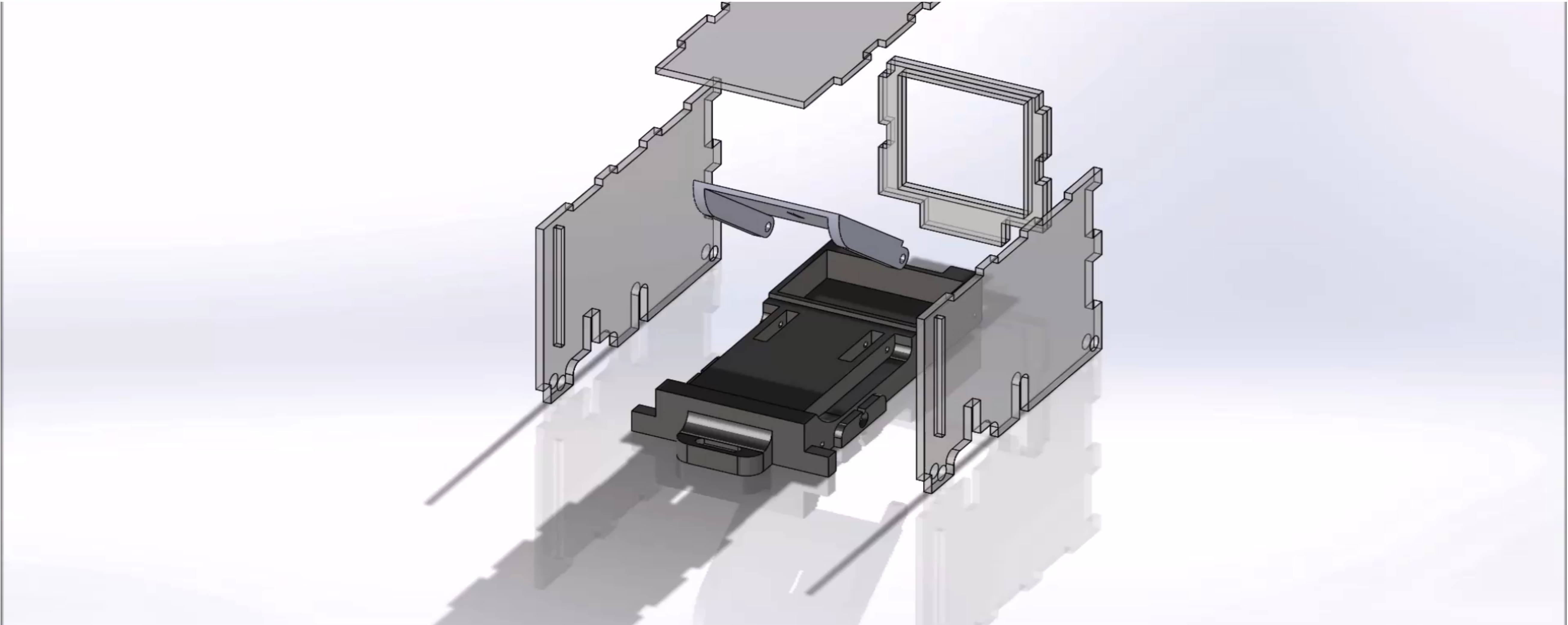
In laboratory setting on lab-raised anopheles (malaria) mosquitoes





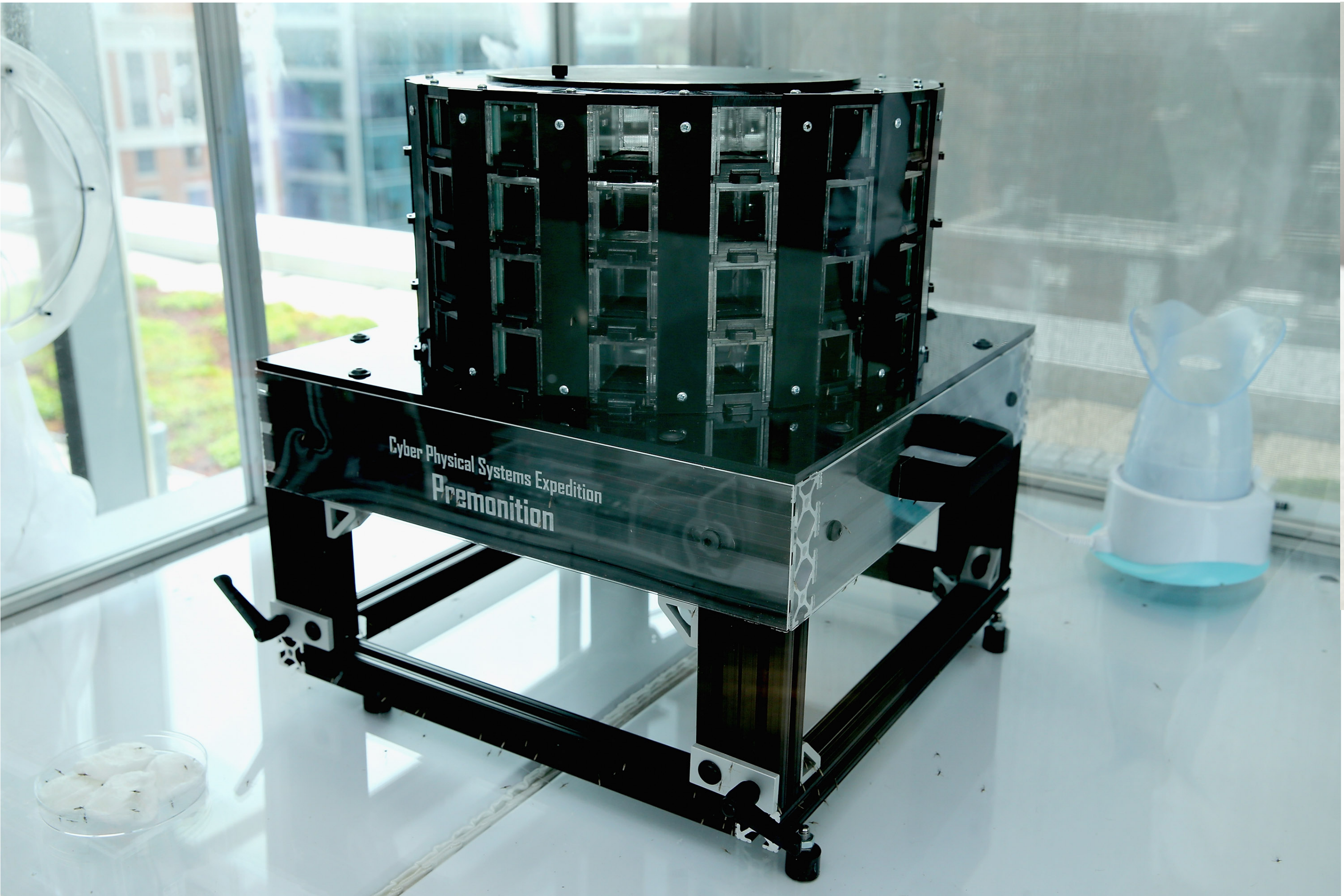
A Challenging CPS Design Problem

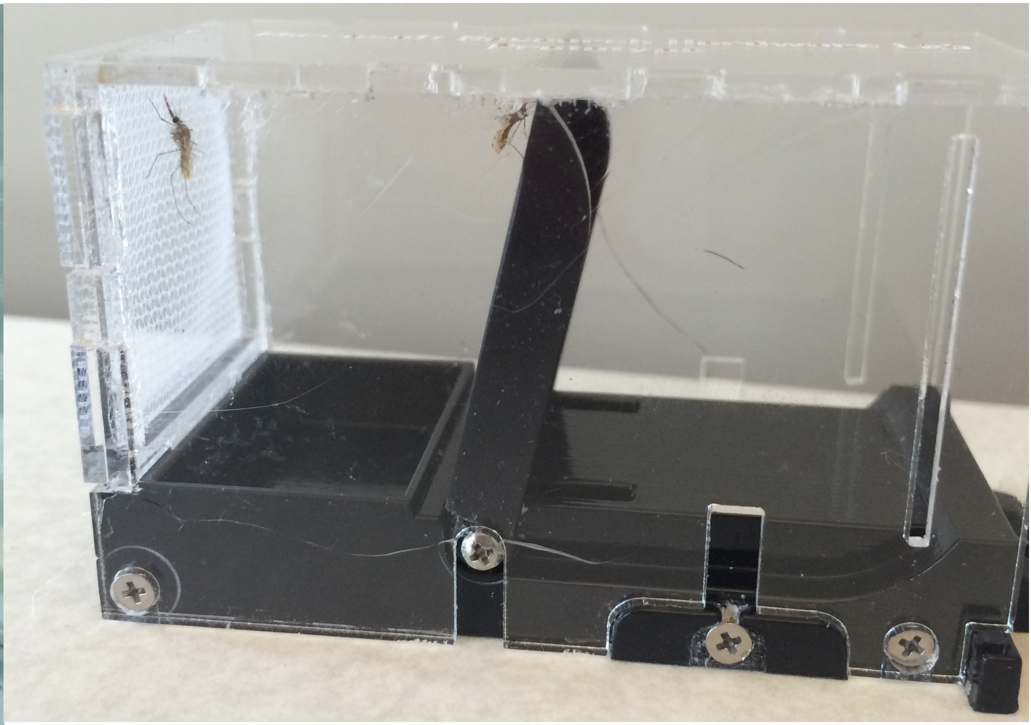
Co-design problem involving: cost, manufacturability, harsh environments, computational power, energy constraints, and size and weight



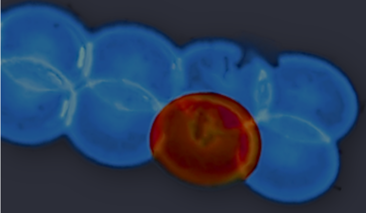
Most Advanced Mosquito Trap in World

Real-time detection and preservation using low-cost sensors, embedded software, and machine learning



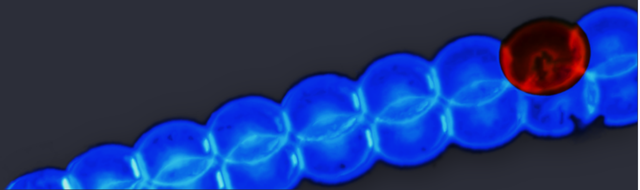


Cells have no electronics or motors



Building PREMONITION

Autonomous Deployment



Autonomous Deployment

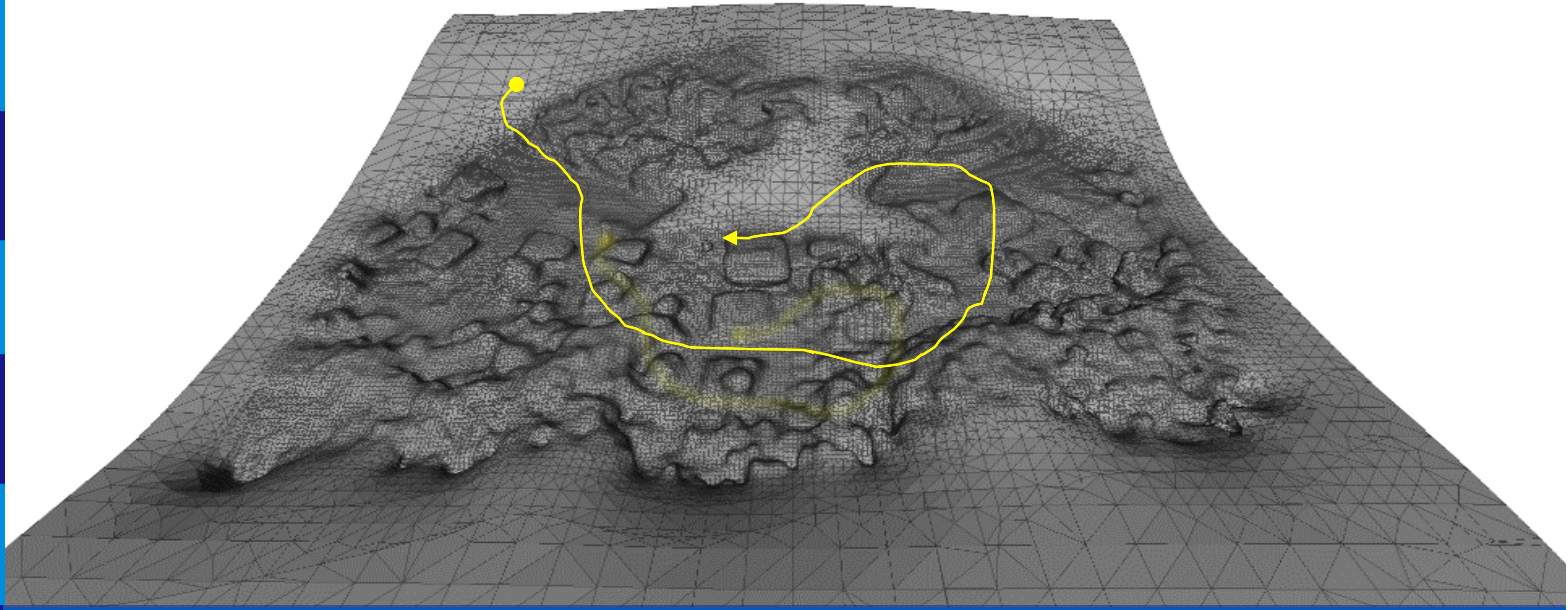
Developing a drone control stack using advanced operating systems, program verification technologies, vision, and machine learning

High-level Planning

Correct Control

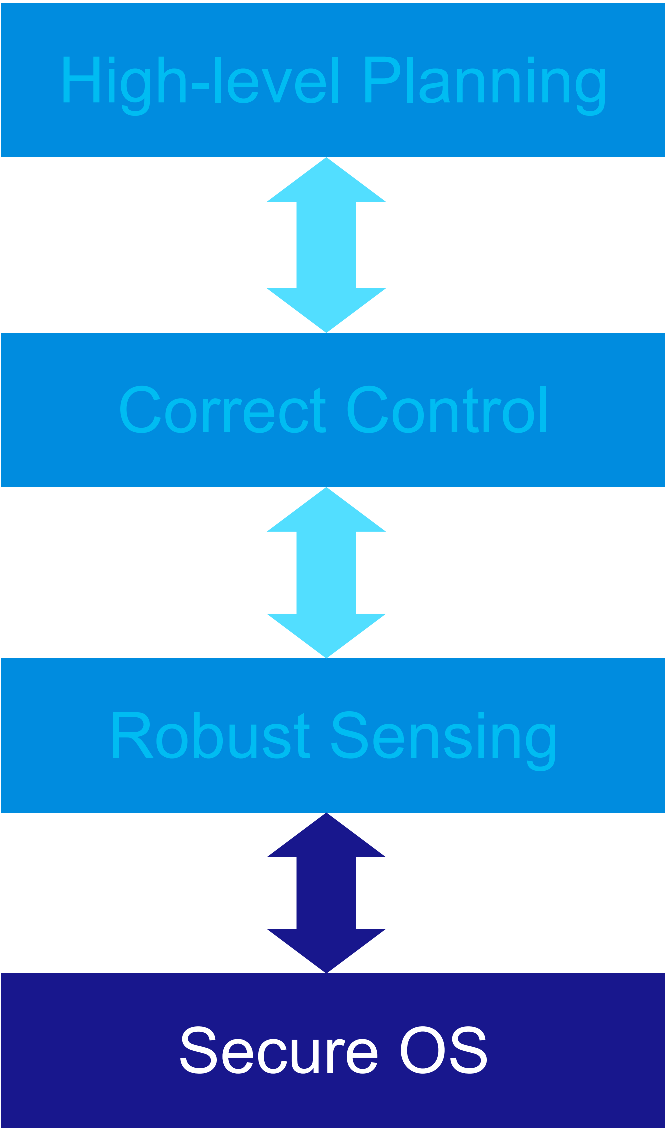
Robust Sensing

Secure OS



Safe despite limited power, external disturbances, sensor noise, and complex missions

Autonomous Deployment



Drone OS:

Provides secure and robust infrastructure for embedded software

Guarantees:

Full functional correctness of OS components and protocols

Methods:

Built from Verve OS; mechanically verified using Z3, Boogie

- Verified **Garbage Collector**
- Verified **Threads**
- Verified **Interrupt Handlers**
- Verified **Device Interface**
- Verified **Startup**

Autonomous Deployment

High-level Planning

Correct Control

Robust Sensing

Secure OS

Drone Vision:

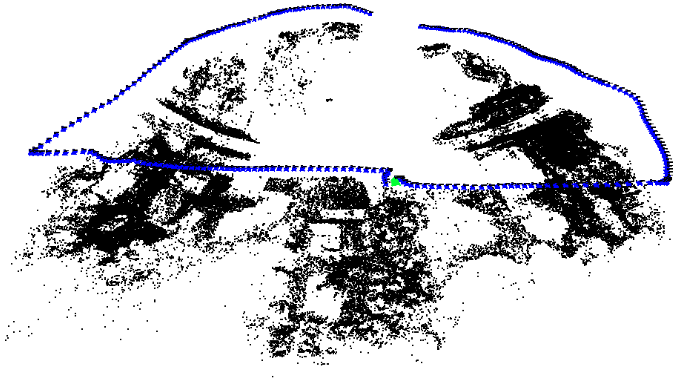
Reconstructs 3D environment and camera poses

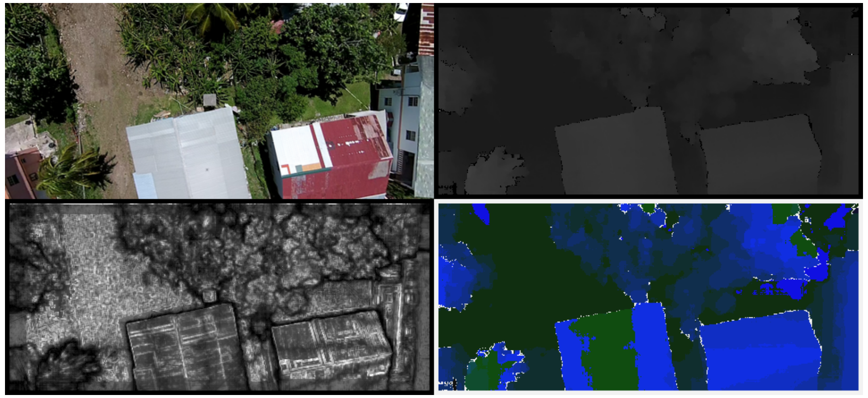
Guarantees:

Goodness of classifications (Bayesian posterior intervals)

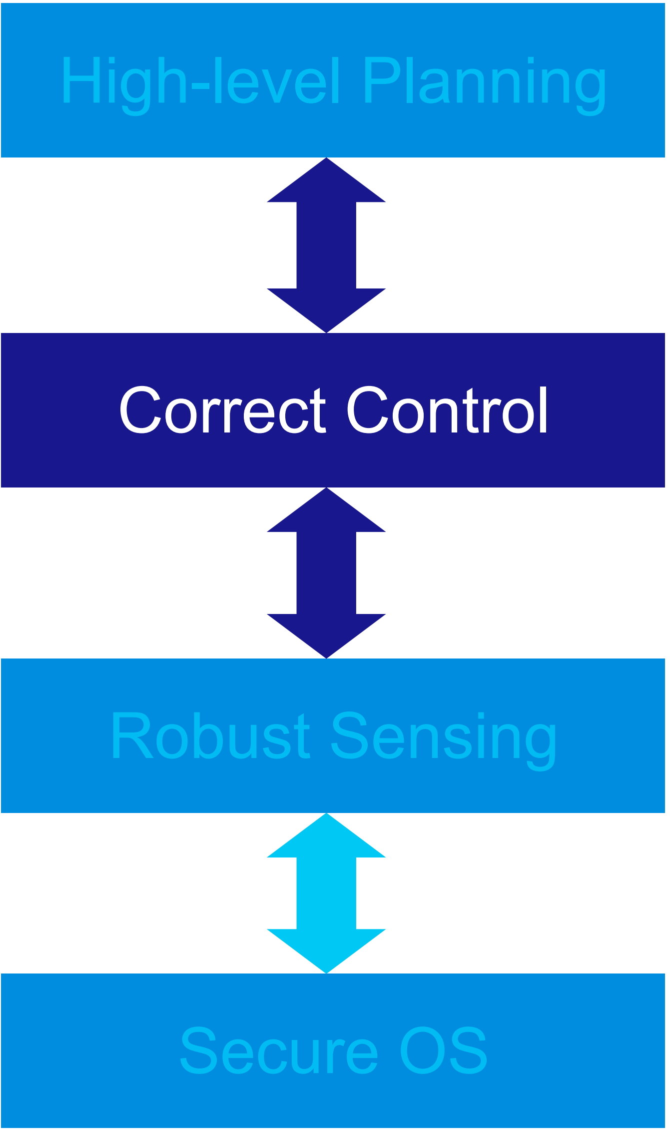
Methods:

Built from MSR vision libs and Bayesian ML





Autonomous Deployment


High-level Planning

Correct Control

Robust Sensing

Secure OS

Controller synthesis and validation:

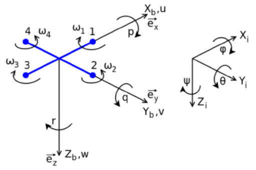
Designs safe control laws handling probabilistic sensor data

Guarantees:

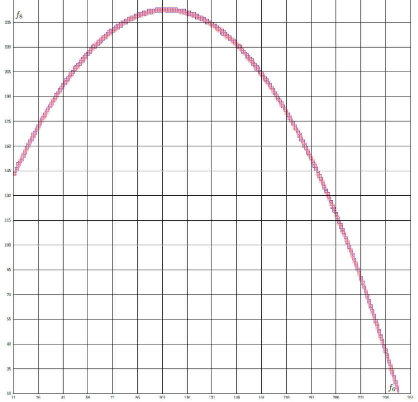
Satisfaction of (probabilistic) safety constraints – extensions of LTL

Methods:

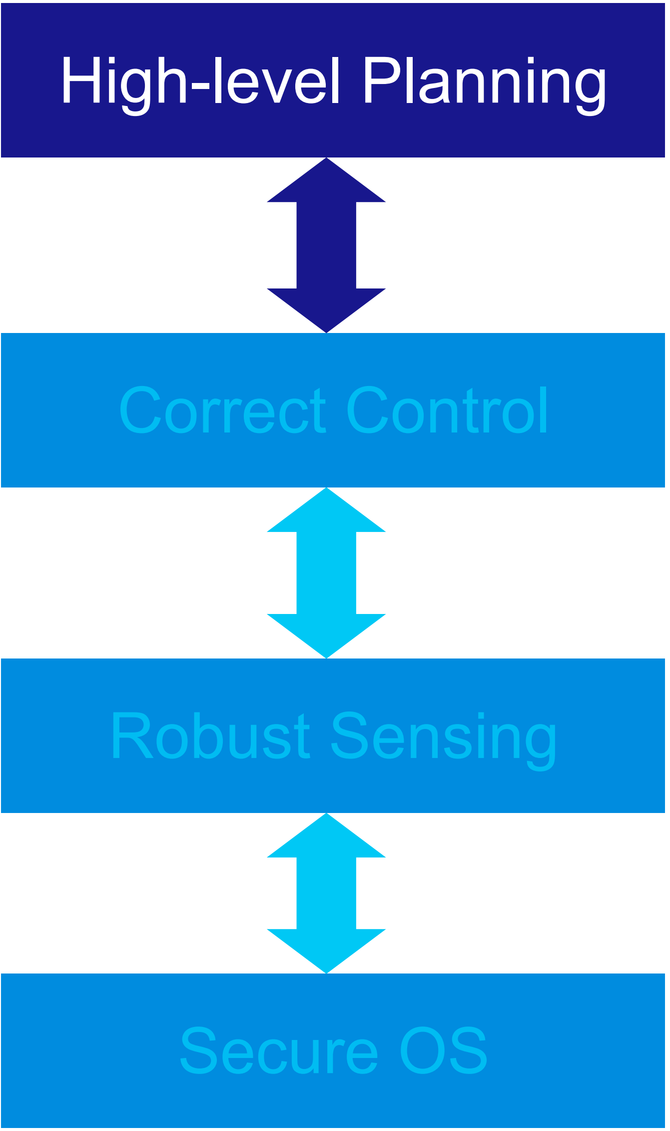
Optimal control, Taylor models and Z3 for hybrid validation



$$
\begin{aligned}
\dot{p} &= \frac{I_a}{I_x}\sum_{i=0}^{3}\gamma_i(\omega_2^i-\omega_4^i)+\frac{I_G}{I_x}q\sum_{j=1}^{4}\omega_j(-1)^j+\frac{I_y-I_z}{I_x}qr\\
\dot{q} &= \frac{I_a}{I_y}\sum_{i=0}^{3}\gamma_i(\omega_3^i-\omega_1^i)+\frac{I_G}{I_y}q\sum_{j=1}^{4}\omega_j(-1)^j+\frac{I_x-I_z}{I_y}pr\\
\dot{r} &= \frac{1}{I_z}\sum_{j=1}^{4}(I_G\dot{\omega}_j+k_D\omega_j^2+B_a\omega_j)(-1)^j\\
\dot{u} &= vr-wq-g\sin(\theta)\\
\dot{v} &= wp-ur+g\sin(\phi)\cos(\theta)\\
\dot{w} &= uq-vp+g\cos(\phi)\cos(\theta)-\frac{1}{m}\sum_{j=1}^{4}\sum_{i=0}^{3}\gamma_i\omega_j^i
\end{aligned}
$$



Autonomous Deployment



High-level PL for planners:

DSL for asynchronous event-driven programming of planners

Guarantees:

Safety and liveness properties over event streams

Methods:

P language, Zing model checker, FORMULA DSL tools

Motion Planner

Plan Executor

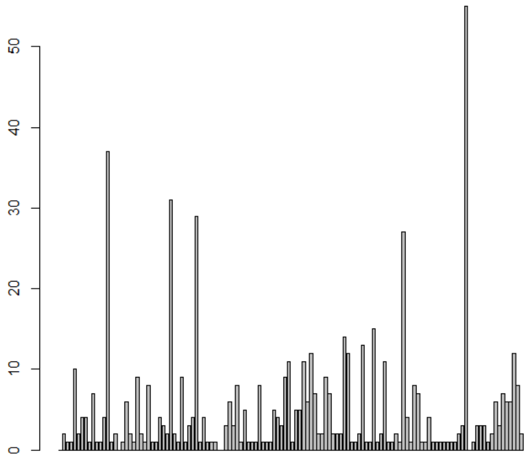
Sense and Infer

State Manager

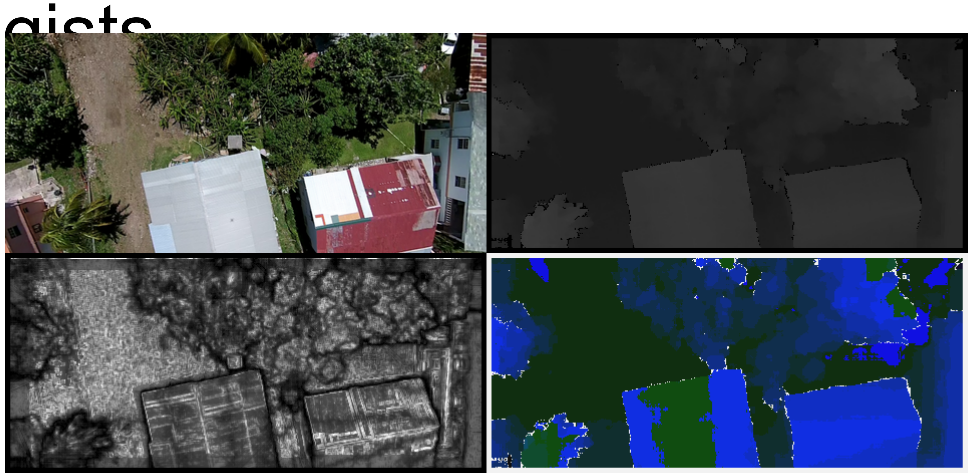
P code

Learning Mosquito Hotspots

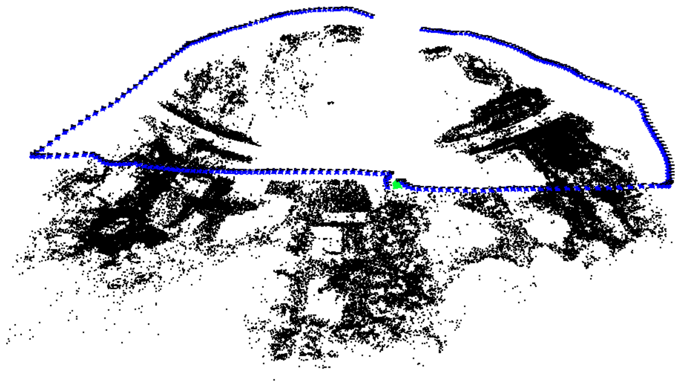
So drones can make deployment decisions that normally require field biologists



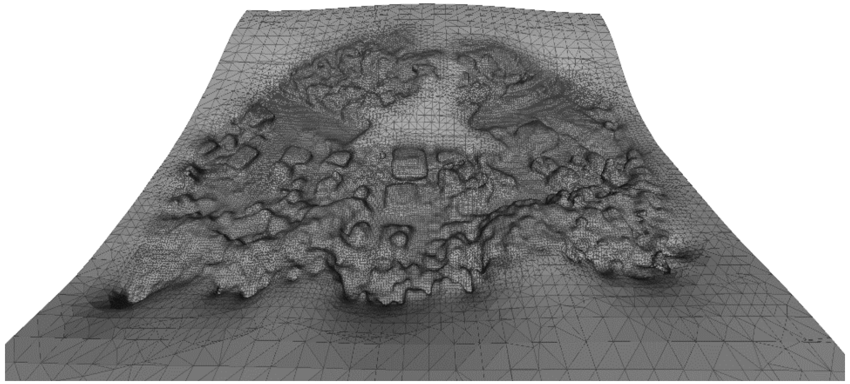
Mosquito counts by site



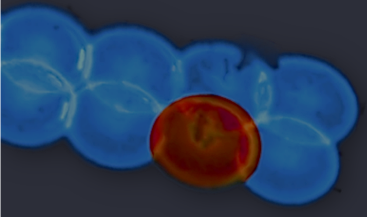
Mosquito hotspot



Reconstruction of camera pose and drone path

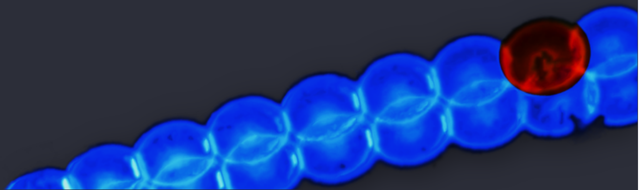


Reconstruction of fined-grained 3D environment



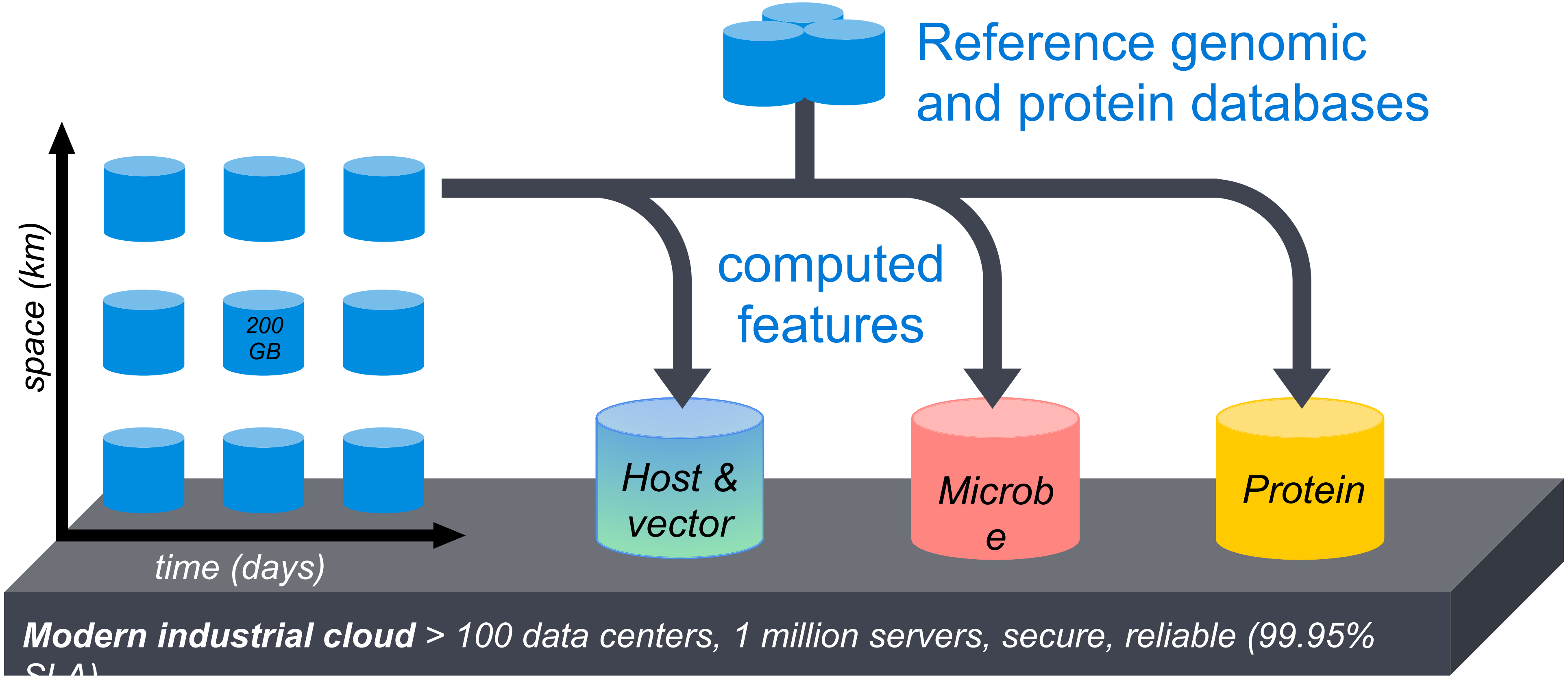
Building PREMONITION

Autonomous threat detection



Better genomic data analytics

Feature extraction and indexing from spatial/temporal genetic data.



Reference genomic and protein databases

computed features

Modern industrial cloud *> 100 data centers, 1 million servers, secure, reliable (99.95% SLA)*

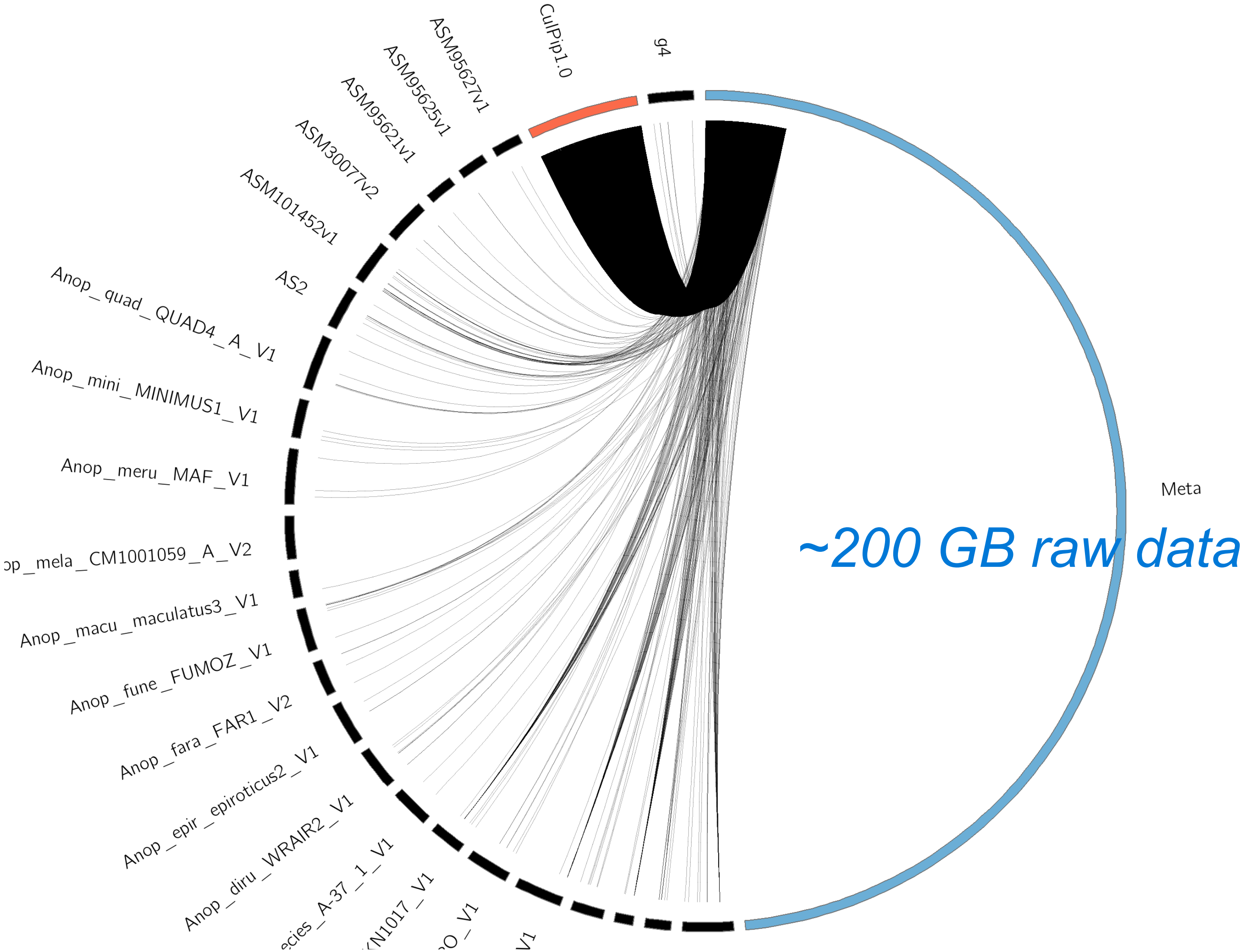
Data cleansing and similarity to reference genes

De novo assembly to reconstruct genes of unknown organisms

Recognize potential threat sequences by expressed proteins

Vector data

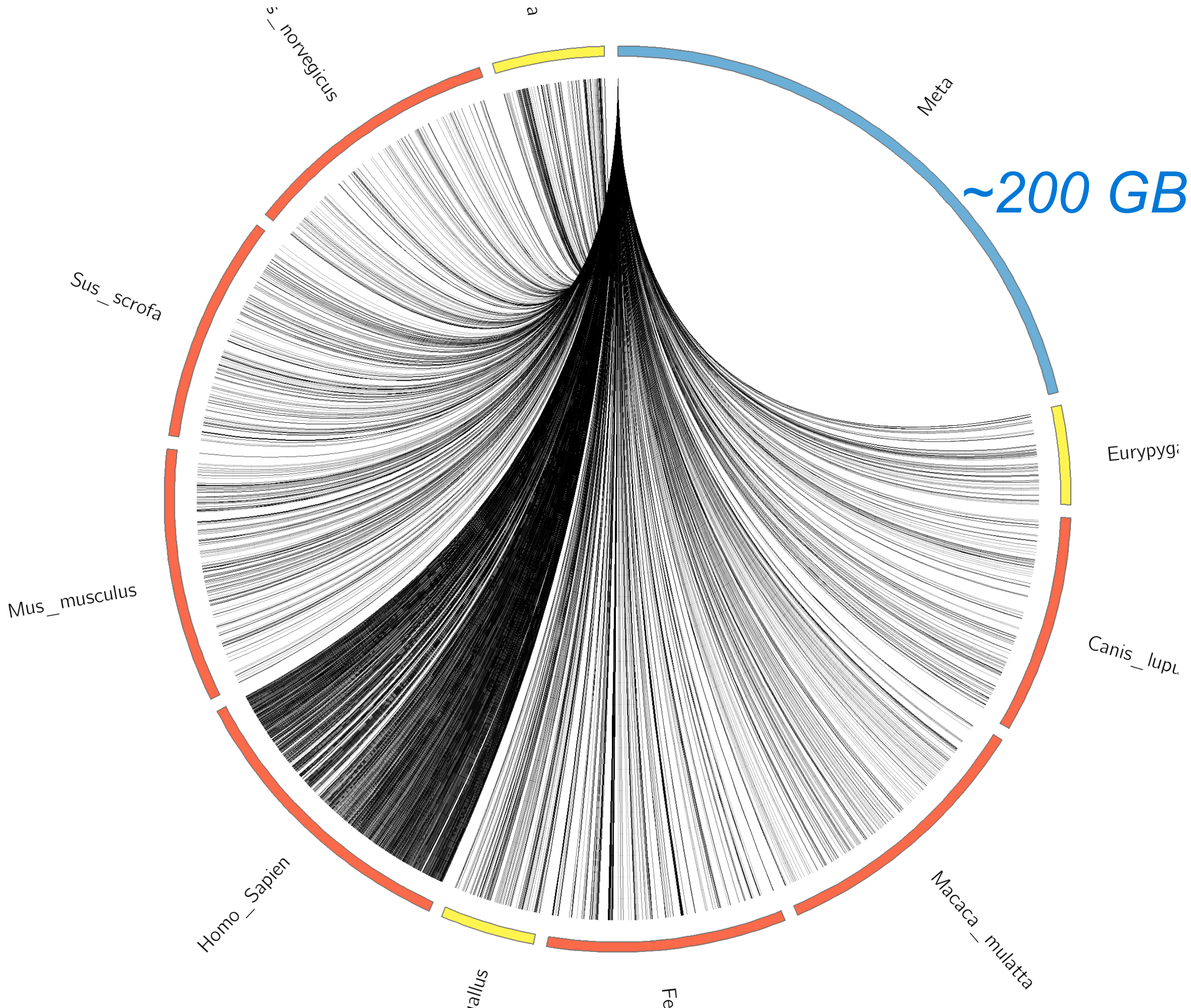
We can recover the species of mosquitoes collected in Grenada



~200 GB raw data

Host data

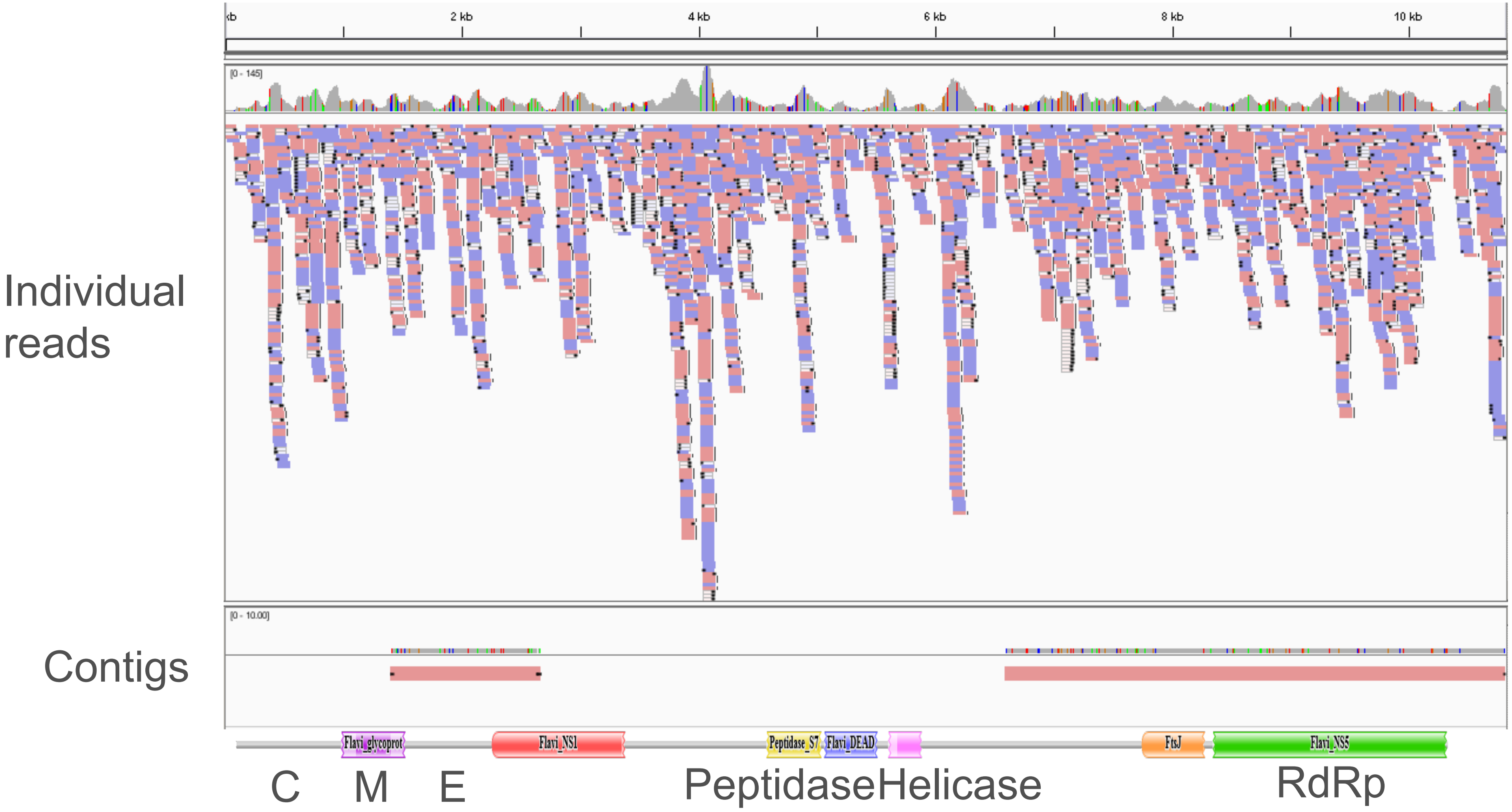
We can recover what mosquitoes bit in Grenada



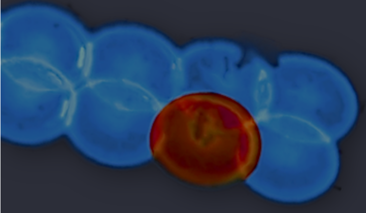
~200 GB raw data

Virus data

We can recover novel potential pathogens from mosquitoes in Grenada

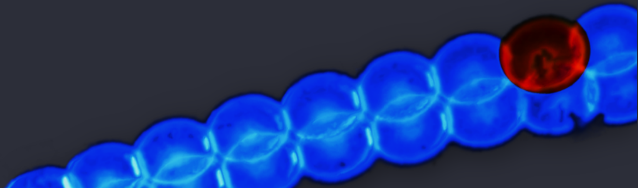


A computationally reconstructed candidate novel Culex Flavivirus



The Path Ahead

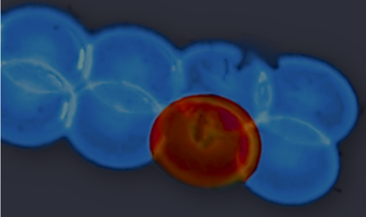
How far off is it?



How far off is it?

We estimate 5 years because the technology trends are clear:

- Moore’s law on microprocessors means field entomology can be automated at scale.
- Commodity drones and robotic platforms are here to stay. Innovations in 3D printing continue to fuel a Cambrian explosion.
- Gene sequence costs decreasing faster than Moore’s law, and cloud capacity continues to grow. Gene sequencing will enable high-throughput surveillance of (unknown) pathogens.



Thank You

Questions are welcome