“Big Data” in infectious disease research
Viruses and diseases we study

Lassa

Ebola

Zika

West Nile
‘Classical’ Epidemiology
‘Genomic’ Epidemiology

Origins
Spread
Evolution
Function
Virus mutations accumulate over time

Cumulative total

- Patients
- Unique lineages

Date in 2014

May 26  Jun 02  Jun 09  Jun 16

Gire et al., Science 2014
Bayesian Evolutionary Analysis

Credit: Philippe Lemey
How
Spread of Zika virus across the world

Jan 2016
Zika virus outbreak in Florida
Zika virus outbreak in Florida

Was it mosquito-borne?
When did it arrive?
How many introductions?
Where did it come from? And how?
Why Miami? Why southern Florida?

Grubaugh et al., Nature 2017
Killing mosquitos really works

Transmission zones
- Little River
- Miami Beach
- Wynwood

ZIKV cases
- Travel
- Local

Month of 2016

Aedes aegypti abundance

ZIKV cases
- Wynwood
- Miami Beach
- Little River

Month of 2016

Seqs
Florida transmission started in Spring, 2016
Undetected transmissions across the Americas

Grubaugh et al., Nature 2017
Metsky et al., Nature 2017
Faria et al., Nature 2017

Figure from M. Worobey, Nature 2017
Zika virus established itself repeatedly in Florida
The Caribbean is the main source of Zika virus in Florida.

![Map showing Zika virus incidence rate in Central and South America with a focus on the Caribbean region. The map highlights the high incidence rate in the Caribbean and the Caribbean basin.](image-url)
The Caribbean is the main source of Zika virus in Florida.
A lot of travel from the Caribbean
Mostly cruise ships

Estimated proportion of infected travelers

- Caribbean (cruises)
- Caribbean (flights)
- South America (combined travel)
- Central America (combined travel)

Month of 2016
Miami is unique
... and so is Florida
Why
But what are the real determinants of spread?

• Cruise ships?
• Behaviors?
• Tourists vs Business vs Family?

*Phylogeographic analysis with covariates*

Credit: Philippe Lemey
Spread of Ebola during outbreak

Dudas et al., Nature 2017
Holmes et al., Nature 2016
### Factors associated with spread

<table>
<thead>
<tr>
<th>Factor</th>
<th>Association</th>
<th>Bayes Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature Seasonality</td>
<td>👎</td>
<td>3.79</td>
</tr>
<tr>
<td>Short Distances</td>
<td>👍</td>
<td>&gt;50</td>
</tr>
<tr>
<td>Large Population</td>
<td>👍</td>
<td>&gt;50</td>
</tr>
<tr>
<td>Borders</td>
<td>👎</td>
<td>&gt;50</td>
</tr>
</tbody>
</table>

### Factors associated with size

<table>
<thead>
<tr>
<th>Factor</th>
<th>Coefficient</th>
<th>Bayes Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature Seasonality</td>
<td>👎</td>
<td>&gt;50</td>
</tr>
<tr>
<td>Short Distances</td>
<td>👍</td>
<td>32.4</td>
</tr>
<tr>
<td>Large Population</td>
<td>👍</td>
<td>29.6</td>
</tr>
<tr>
<td>Precipitation</td>
<td>👍</td>
<td>4.4</td>
</tr>
</tbody>
</table>

Dudas et al., Nature 2017
Why microcephaly?

Distribution of cumulative confirmed cases of microcephaly, Brazil as of 02 April 2016

Source: Epidemiological report of cases of microcephaly in Brazil, Ministry of Health, Brazil.

This cumulative confirmed data includes both microbiological confirmation of Zika and radiological criteria of congenital infections. The latter probably includes other infections than Zika.

R21, Analysis of Zika virus emergence and functional evolution in the Americas
Could viral genetics explain microcephaly?

Collaboration with Lark Coffey
Function
A single mutation in Ebola rose to high frequency
A82V mutation is located in the binding pocket
A82V mutation leads to increased infectivity

Diehl et al., Cell 2016

[Diagram showing amino acid positions and sequences of different species for EBOV interacting loop 2]
... and is also associated with lower survival
Selection in *PIEZO1* gene in African populations

---

**Ancestry**
- African
- Non-African

**Allele**
- Wild-type
- E756 insertion
- E756 deletion (Plasmodium resistant)

---

### PIEZO1 codon position (minus strand)

<table>
<thead>
<tr>
<th></th>
<th>748</th>
<th>750</th>
<th>752</th>
<th>754</th>
<th>756</th>
<th>758</th>
<th>760</th>
<th>762</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neanderthal</td>
<td>CAG</td>
<td>CAG</td>
<td>AGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>ACT</td>
</tr>
<tr>
<td>Denisovan</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>ACT</td>
</tr>
<tr>
<td>Modern human</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>ACT</td>
</tr>
<tr>
<td>Human E756del</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>GGA</td>
<td>CT</td>
</tr>
</tbody>
</table>

**Amino acid**
- Q
- Q
- E
- E
- E
- E
- E
- E
- D
- S
- R
- D
- E
- G

Deletion derived in African populations

Collaboration with Ardem Patapoutian
PIEZ01 mutation protects against malaria

Graph showing % survival from malaria over days for Wildtype and Piezo1 mutant.
Translation
Center for Viral Systems Biology (CViSB)

We are a group of scientists interested in understanding what genetic, immunological, and physiological factors determine the outcome of human disease from viral infections.

Leadership
What determines outcome from disease?

Antibody evolution
Antibody function
T-cell dynamics
HLAs

Physiological factors
Symptoms
Co-infections
Virus genetics

Lassa
Ebola
Measure all things | Overview of CViSB

CViSB U19 Center Information and Data Flow

**Admin Core**
- Scientific management
- Oversight: NIH/NIH and NIH/FDA
- Public engagement

**Cohorts**
- Lassa survivors
- Lassa non-survivors
- Lassa contacts
- Ebola survivors
- Ebola non-survivors
- Ebola contacts
- Population controls

**Questions**
- Survival vs Death
- Development of sequelae
- Symptomatic vs Asymptomatic

**Project 1**
- Host determinants of patient outcomes
  - Aim 1: Physiological and metabolic attributes
  - Aim 2: Antibody networks
  - Aim 3: T cell dynamics and HLA-HLA
  - Aim 4: Predictive modeling of host factors

**Technology Core**
- Technologies
  - Wireless sensing
  - AI sequencing
  - Virus sequencing
  - Proteomics
  - Mass Spec
  - Single cell sorting
  - HTS: Next generation sequencing
  - mAb isolation
  - mAb reactivity
  - HLA typing

**Data Core**
- Data
  - ELISA
  - Cytokines
  - HLA
  - Virus sequence
  - Metagenomics
  - Ab sequences
  - TCR sequences
  - Viral titers
  - mAb function
  - mAb neutralization
  - mAb reactivity

**Capture**
- Patient data
  - Symptomatic data
  - Metabolic data
  - Physiological data

**Processing pipelines**
- Assembly
  - TCR classification
  - Ab classification
  - Taxonomy classification
  - HLA haplotypes

**Modeling Core**
- Evolutionary
  - Survival analysis
- Network analysis

**Data release**
- NCBI Genomes
- NCBI SRA
- CViSB Database
- Github
- Amazon Cloud

**Visualization**
- AREX
- Annotation
- Trees
- Posterior

**Pipelines**
- Assembly
- Alignments
- Classification
- Variant calling
- Ab profiling
- Ab specificity
- TCR profiling

**Software**
- AbStar
- Clonify
- Variant Caller
- REDCap
- BWA
- MAF
- Centrifuge
- DIAMOND
- metaSPAdes

**Public**
- User data upload

**Legend**
- Dotted box = longitudinal data
- Blue = raw data
- Green = analysis-ready data
- Tan = public/cohort partners
- Bold text = CViSB software/pipelines
- Italics = currently, no new cases

**Conclusion**
- Features associated with:
  - Survival
  - Death
  - Symptomatic infection
  - 'Successful' immunity
- Functional validation
- Predictive modeling

**NIH BRCs**
- CViSB.org
- Github
- Docker
- CRAN
HLA associations in Lassa and Ebola

Collaboration with Michael Oldstone
HLA associations in Lassa and Ebola

<table>
<thead>
<tr>
<th></th>
<th>Patient Allele Frequency</th>
<th>Control Allele Frequency</th>
<th>Uncorrected P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ebola</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQB1*031901</td>
<td>0.2551</td>
<td>0.0952</td>
<td>0.007</td>
</tr>
<tr>
<td>DRB1*070101</td>
<td>0.0204</td>
<td>0.1190</td>
<td>0.013</td>
</tr>
<tr>
<td>DQA1*020101</td>
<td>0.0204</td>
<td>0.1190</td>
<td>0.013</td>
</tr>
<tr>
<td>DRB4*010301</td>
<td>0.0510</td>
<td>0.1667</td>
<td>0.014</td>
</tr>
<tr>
<td>DQA1*050501</td>
<td>0.2653</td>
<td>0.1190</td>
<td>0.015</td>
</tr>
<tr>
<td><strong>Lassa</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRB3*020101</td>
<td>0.0000</td>
<td>0.0714</td>
<td>0.008</td>
</tr>
<tr>
<td>C*021001</td>
<td>0.0648</td>
<td>0.0000</td>
<td>0.022</td>
</tr>
</tbody>
</table>
Immunity as a driver of viral evolution (Lassa)

- Glycoprotein (GPC)
- Zinc-binding protein (Z)
- Polymerase (L)
- Nucleoprotein (NP)
- Bi-segmented RNA genome
Lassa glycoprotein is under selection

Andersen et al., Cell 2015

Glycoprotein (GPC)  Zinc-binding protein (Z)
Polymerase (L)  Nucleoprotein (NP)
Bi-segmented RNA genome

![Diagram showing viral elements and dN/dS ratio]

**dN/dS**

- **Within**
- **Between**

Andersen et al., Cell 2015
Immune escape in viral hemorrhagic fevers?

Andersen et al., Cell 2015
Park et al., Cell 2015

Lassa (within host)  

Ebola (outbreak)  

<table>
<thead>
<tr>
<th>Protein</th>
<th>Lassa (within host)</th>
<th>Ebola (outbreak)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPC</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>NP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Andersen et al., Cell 2015  
Park et al., Cell 2015
How to build a great antibody?

~ 170 monoclonal antibodies

Measure lots of stuff

Test protection in animal model

Large-scale analysis

What does a protective antibody look like?

Collaboration with Erica Saphire / Galit Alter
Lots’ of diversity in antibody function

Credit: Bonnie Gunn, Galit Alter
Building correlations

Karthik Gangavarapu | apps.andersen-lab.com/cornetwork
What makes an effective antibody?
Collaborators

Scripps
Erica Saphire
Michael Oldstone
Brian Sullivan
Emily Spencer
Ardem Patapoutian
Andrew Su

Tulane University
Bob Garry
John Schieffelin
James Robinson
Luis Branco
Matt Boisen
Jess Hartnett

Kenema Government Hosp.
Humarr Khan
Donald Grant
Simbirie Jalloh
Augustine Goba
John Demby
Mambu Momoh

Redeemer’s University
Christian Happi
Onikepe Folarin
Philomena Ehiane

Ragon Institute
Galit Alter
Bonnie Gunn

UMass Medical School
Jeremy Luban
Ted Diehl

Florida Gulf Coast
Sharon Isern
Scott Michael
Amanda Tan

Broad Institute
Pardis Sabeti
Nathan Yozwiak
Bronwyn MacInnis
Danny Park
Chris Matranga
Andreas Gnirke
Stephen Schaffner
Ilya Shlyakhter
Aaron Lin
Jesse Shapiro
Sarah Winnick
Kayla Barnes
Hayden Metsky
Dylan Kotliar

USAMRIID
Gus Palacios
Jason Ladner
Mike Wiley

University of Miami
David Watkins
Diogo Magnani
Mike Ricciardi

Fred Hutch
Trevor Bedford
Gytis Dudas
Allison Black

University of Oxford
Oliver Pybus
Nuno Faria
Moritz Kraemer
Julien Thézé

University of Edinburgh
Andrew Rambaut
Luiz Carvalho

University of Birmingham
Nick Loman
Josh Quick

KU Leuven
Philippe Lemey
Simon Dellicour

UC Davis
Lark Coffey

Florida Department of Health
Leah Gillis
Marshall Cone
Andrew Cannons
Chalmers Vasquez

San Diego Vector Control
Nikos Gurfield
Saran Grewal

Illumina
Sandra Balladares
Christiane Honisch
Alex Lindell

NASA
Kate Rubins

NIAID IRF
Lisa Hensley
Jens Kuhn
Open Science & Open Data

Data
We make our data publicly available as soon as possible after data generation. You can find it all here.

Code
Analyses should be reproducible, so we make our tools, scripts and pipelines available. Come grab our code here.

Protocols
Here you'll find the various protocols that power our lab.

Talks
We give talks. Listen, don't take notes, then come download all of our slides here.

Yozwiak et al., Cell 2016

www.andersen-lab.com/secrets

www.nextstrain.org
Our dear President - on Ebola

Donald J. Trump
@realDonaldTrump

The U.S. cannot allow EBOLA infected people back. People that go to far away places to help out are great—but must suffer the consequences!

6:22 PM - 1 Aug 2014

I have been saying for weeks for President Obama to stop the flights from West Africa. So simple, but he refused. A TOTAL incompetent!

7:31 PM - 23 Oct 2014

We must immediately stop all air traffic coming from the Ebola infected areas of Africa—before it is too late.

9:04 AM - 16 Sep 2014

Donald J. Trump
@realDonaldTrump

A single Ebola carrier infects 2 others at a minimum. STOP THE FLIGHTS! NO VISAS FROM EBOLA STRICKEN COUNTRIES!

1:31 PM - 10 Nov 2014

Donald J. Trump
@realDonaldTrump

Ebola is much easier to transmit than the CDC and government representatives are admiting. Spreading all over Africa—and fast. Stop flights

2:52 AM - 2 Oct 2014

Donald J. Trump
@realDonaldTrump

The Ebola doctor who just flew to N.Y. from West Africa and went on the subway, bowling and dining is a very SELFISH man—should have known!

6:09 AM - 24 Oct 2014

Donald J. Trump
@realDonaldTrump

Ebola patient will be brought to the U.S. in a few days - now I know for sure that our leaders are incompetent. KEEP THEM OUT OF HERE!

7:04 PM - 31 Jul 2014

Donald J. Trump
@realDonaldTrump

If this doctor, who so recklessly flew into New York from West Africa, has Ebola, then Obama should apologize to the American people & resign!

4:38 PM - 23 Oct 2014

Donald J. Trump
@realDonaldTrump

Stop the EBOLA patients from entering the U.S. Treat them, at the highest level, over there. THE UNITED STATES HAS ENOUGH PROBLEMS!