INSTITUTE FOR DISEASE MODELING

INTELLECTUAL VENTURES®
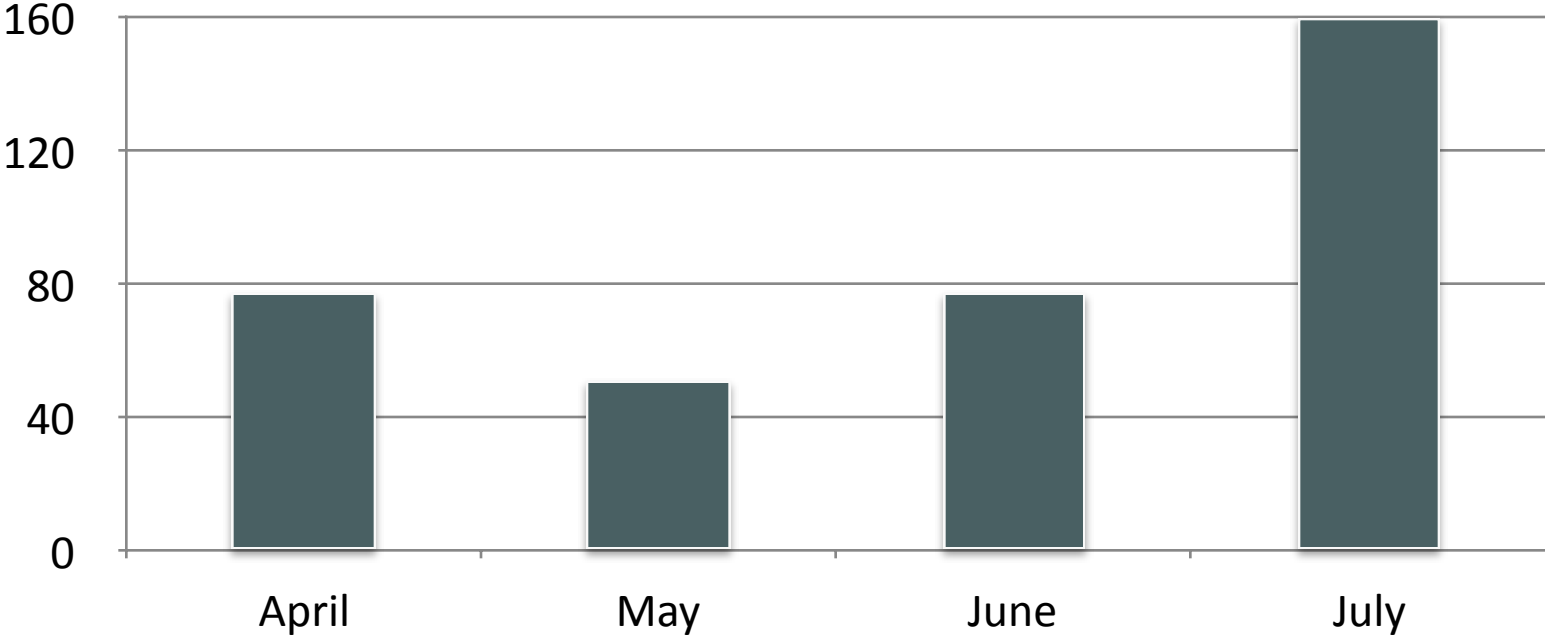
The information value of malaria parasite genetics

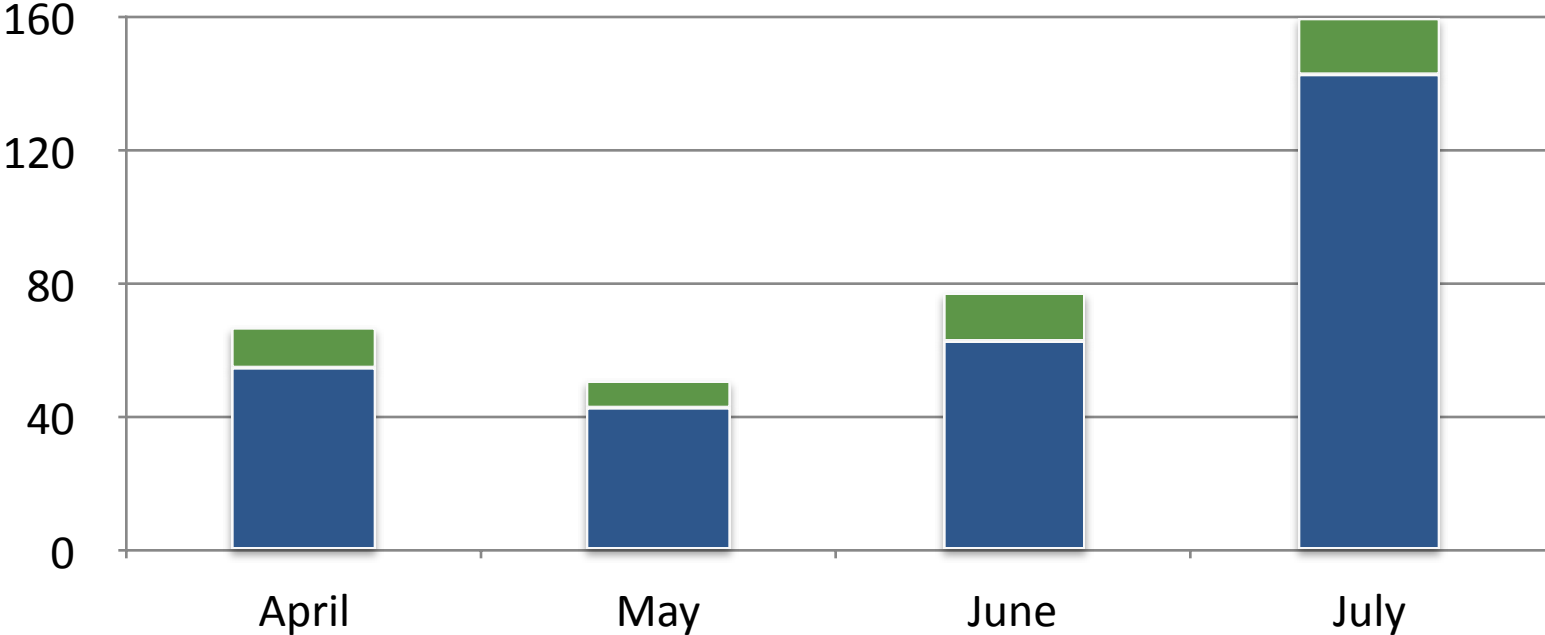Edward Wenger

4th Annual IDM Symposium — 19 April 2016

globalgood

Why should anyone get excited about parasite genetics?

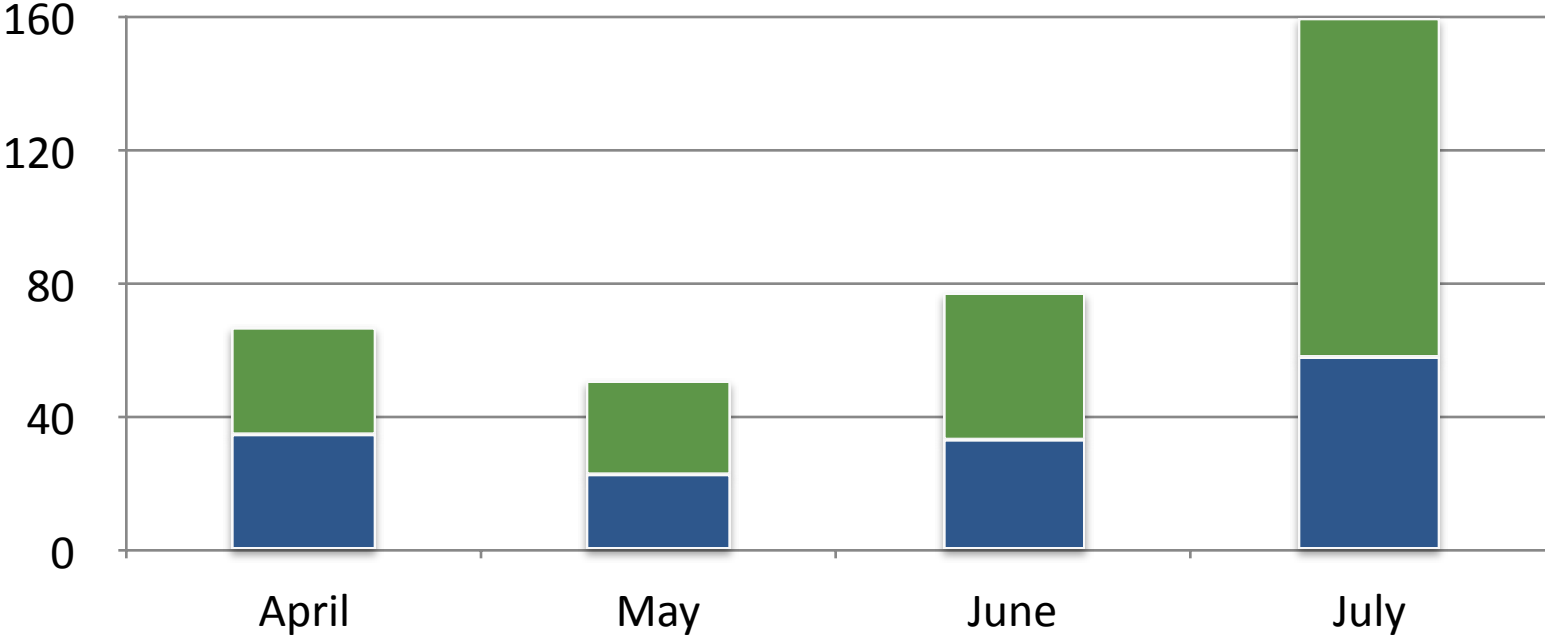# Parasite genetics can distinguish between **local** and **imported**

Parasite genetics can distinguish between **local** and **imported**
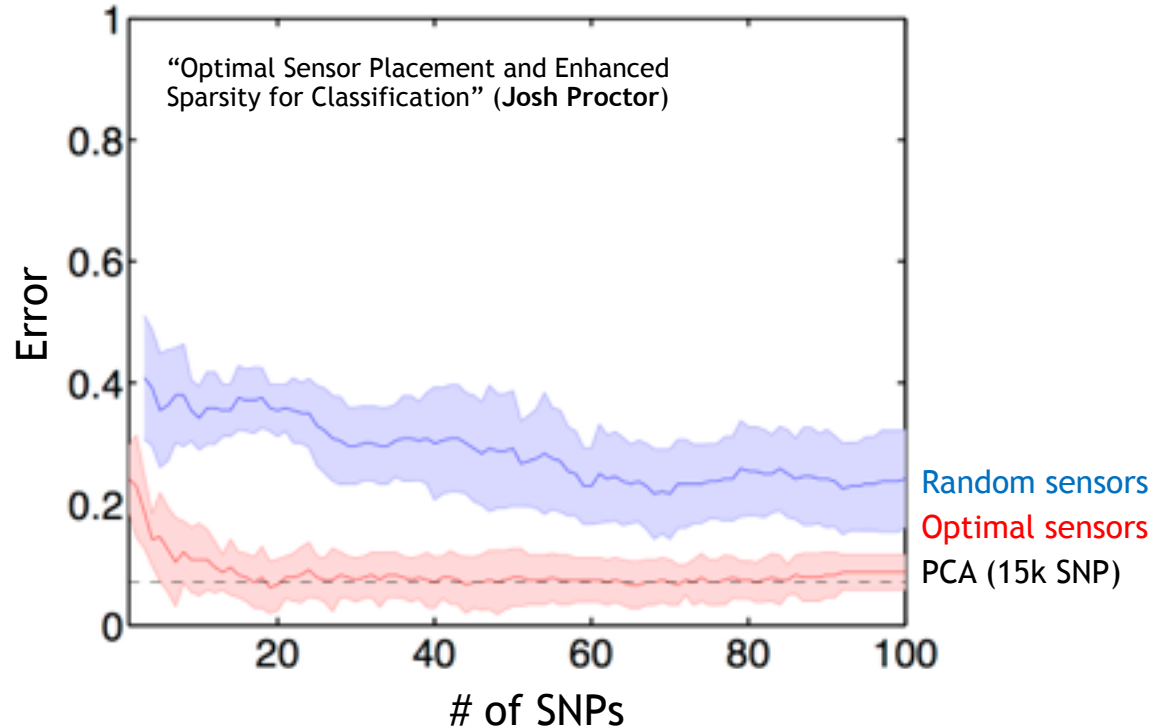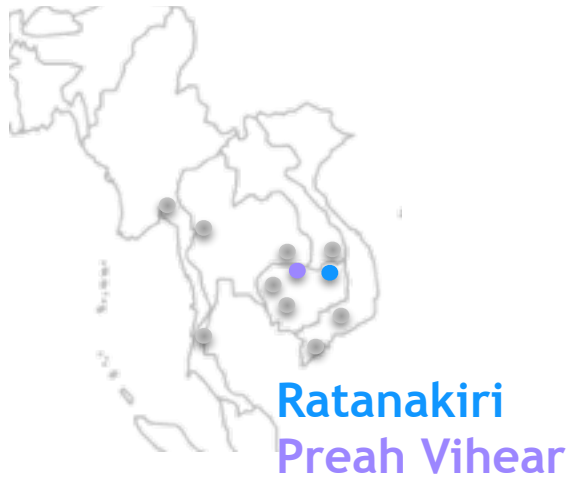

A local epidemic getting out of control?

# Parasite genetics can distinguish between **local** and **imported**

Increase in importation-driven malaria?

In specific settings, how many SNPs and which positions are optimal in determining geographic origin?

Ratanakiri
Preah Vihear

"Optimal Sensor Placement and Enhanced Sparsity for Classification" (**Josh Proctor**)

Random sensors
Optimal sensors
PCA (15k SNP)

Parasite genetics can assist in categorizing areas with varying **capacities** for transmission and **connectivity**
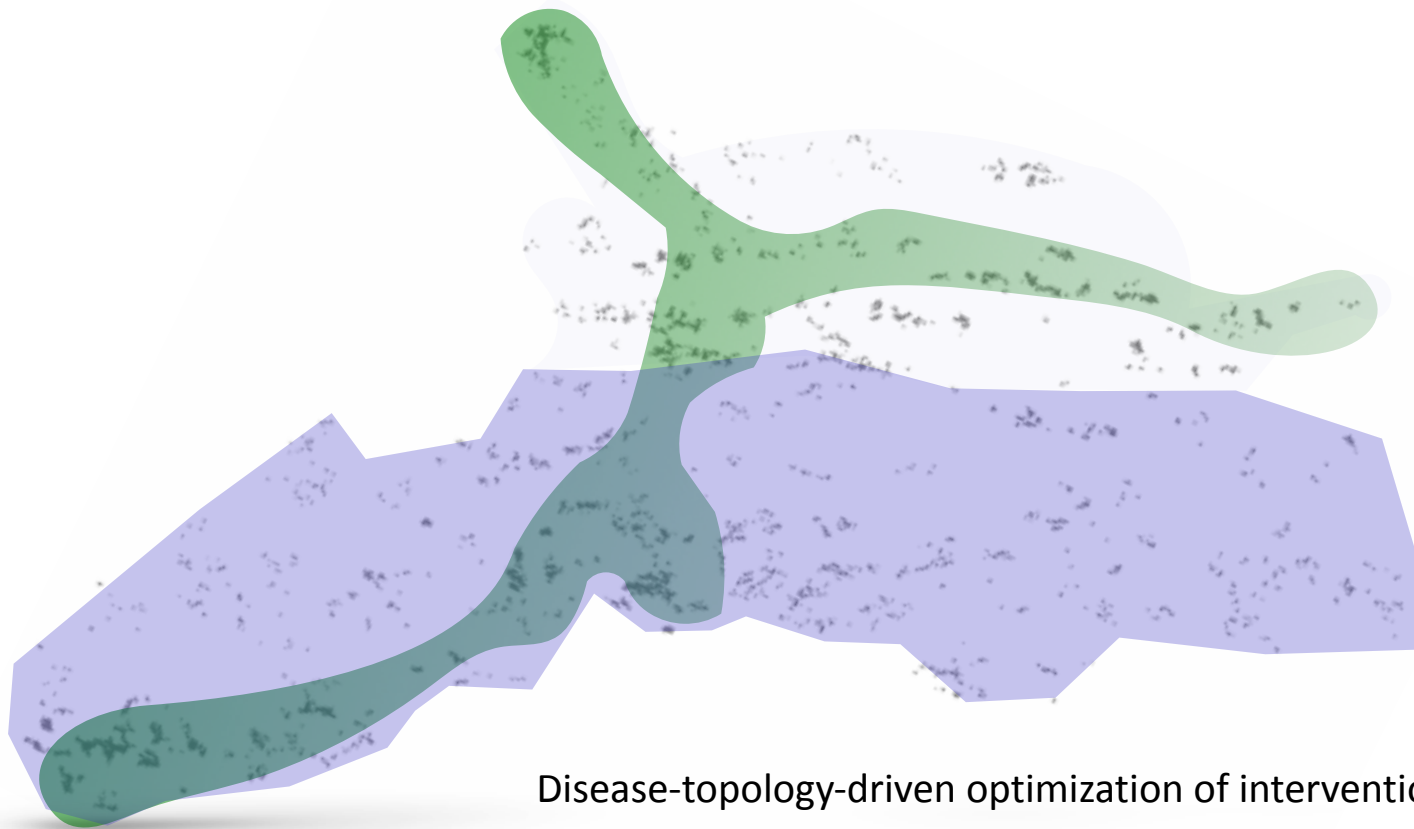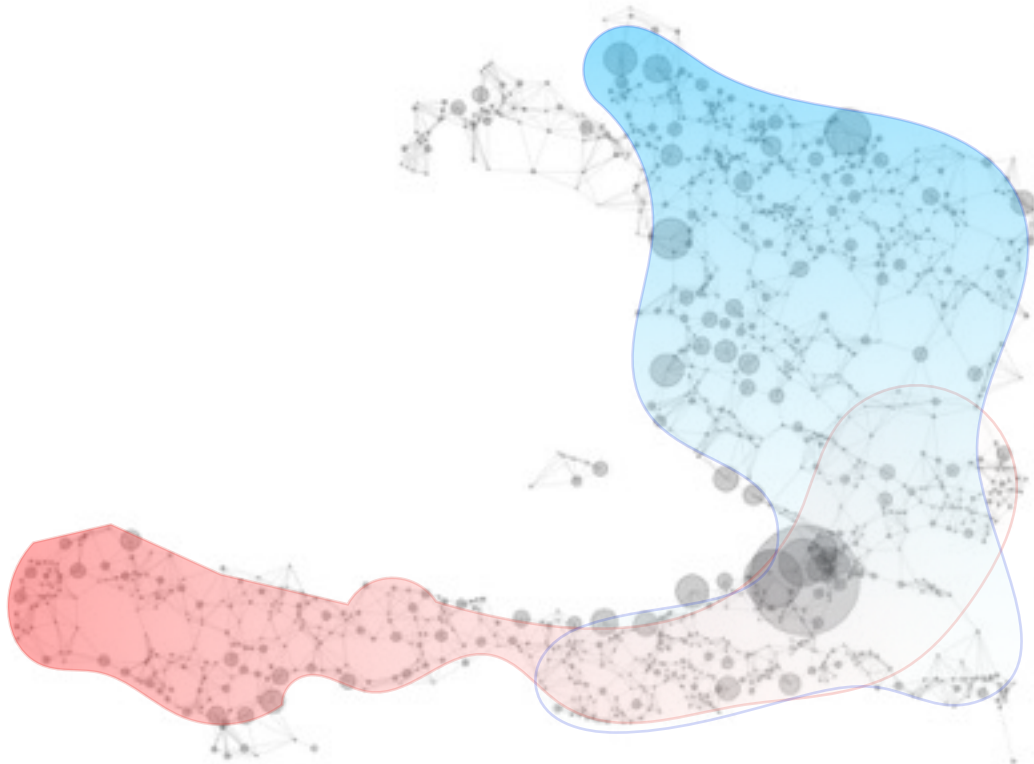

Disease-topology-driven optimization of intervention strategies?

Parasite genetics can validate the relevant time and space scales that define effectively disconnected regions
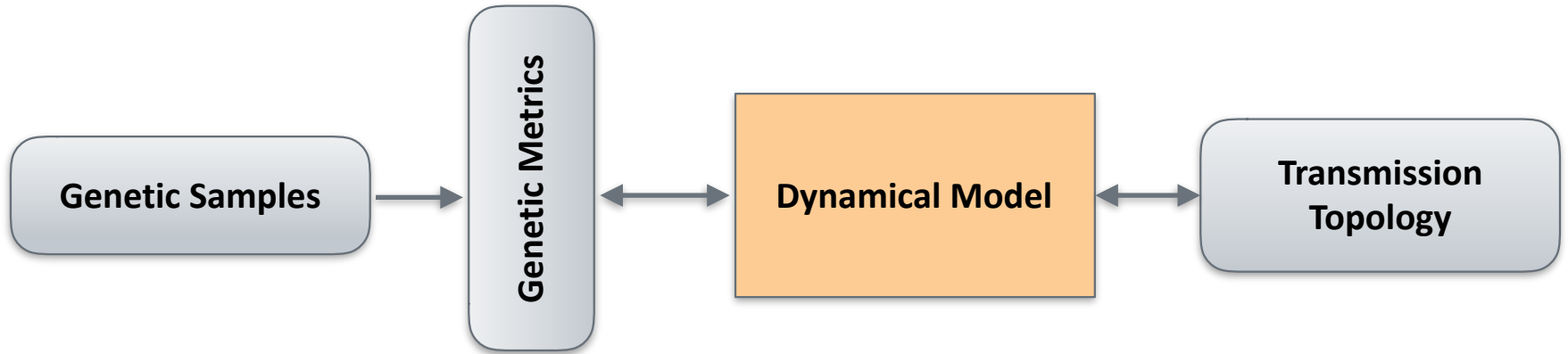

Over a short time interval?

Parasite genetics can validate the relevant time and space scales that define effectively disconnected regions
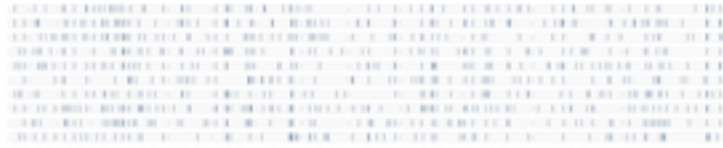

Over a longer time interval?
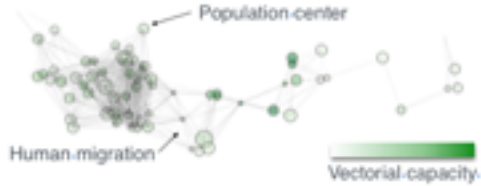
How does modeling fit in?

Dynamical Model

Genome Representation:
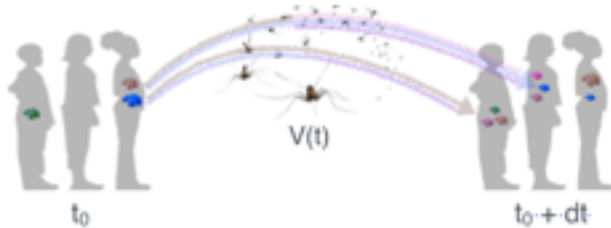Discretization of binary SNPs on chromosomes

Propagation of Infection State:
Human infectiousness, strain recombination

Spatial Topology:
Heterogeneous transmission potential
Coupling through human migration rates

Seasonality of Vectorial Capacity:
Time-varying effects of vector control

- Flexible spatial configuration and connectedness
- Full genome representation with chromosomes/outcrossing/meiosis
- Integration with full-disease model: infectiousness, immunity, symptoms, drug PkPd

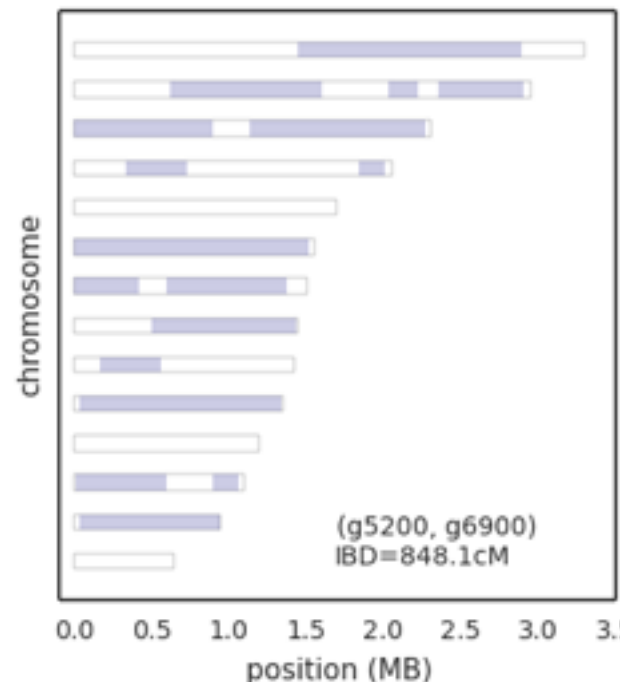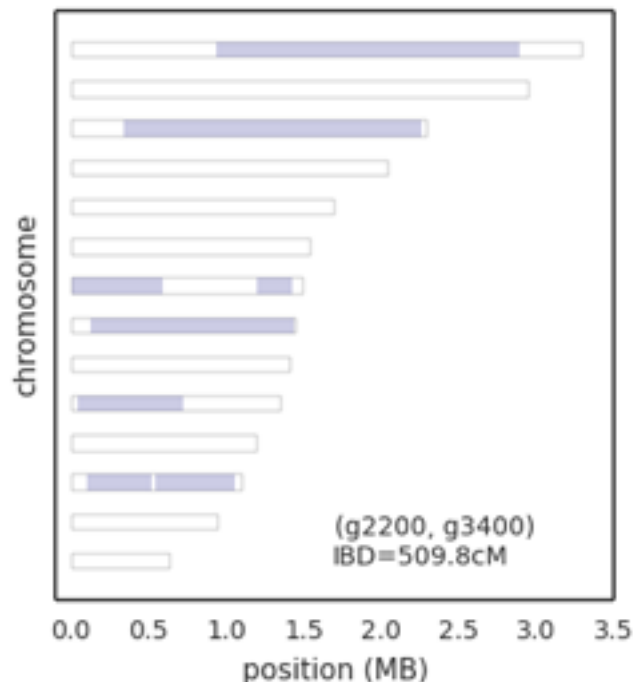Open-source Python genetic-epidemiological modeling package

Also: Integration of explicit genomes into existing EMOD individual-based model

We can characterize shared regions from the outputs of these models
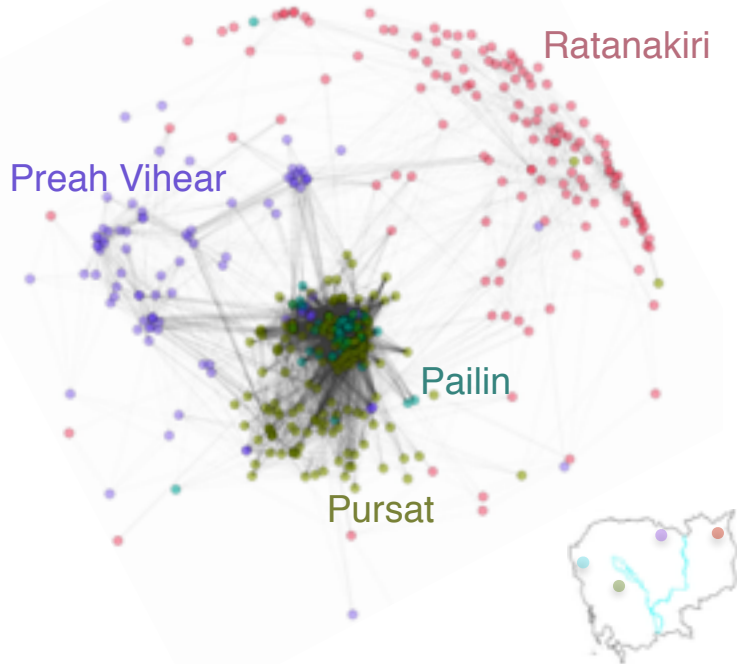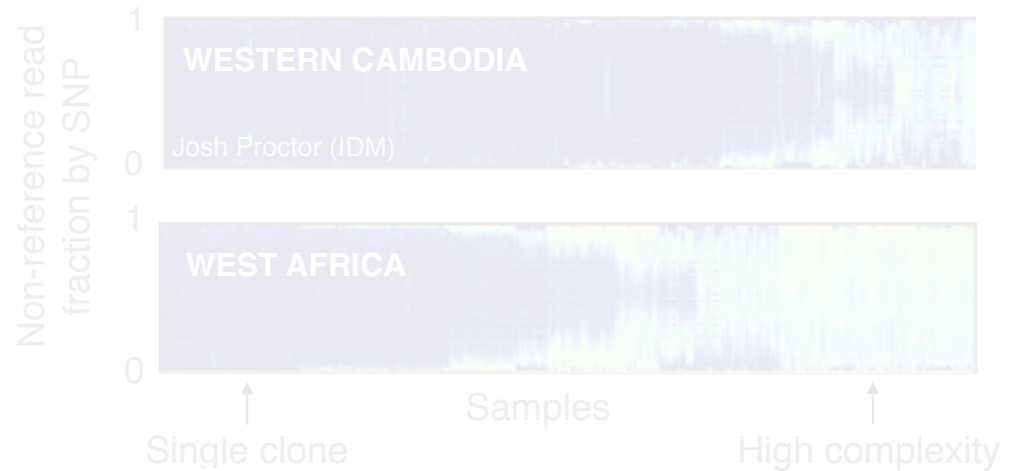
Low IBD sharing
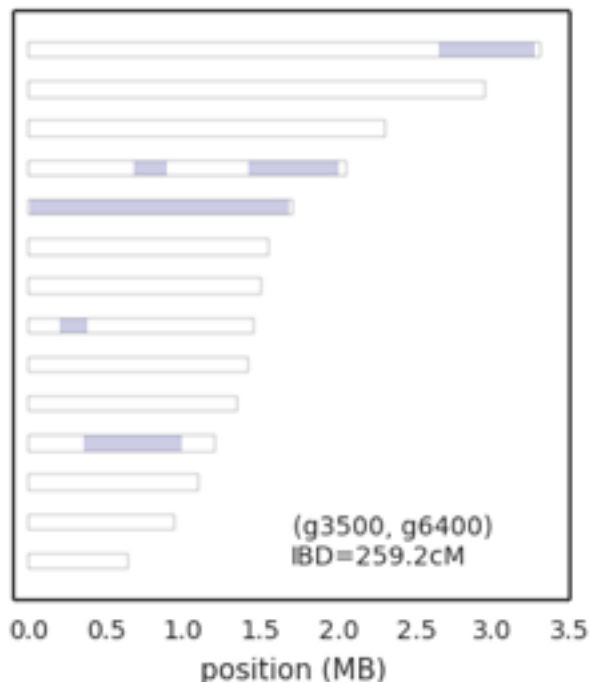
High IBD sharing

The distribution of multi-strain infections and genetic relatedness encodes information on transmission topology

Strain Relatedness
Identity-by-descent network

Ratanakiri

Preah Vihear

Pailin

Pursat

Multi-Strain Infections
Number, density, similarity

Non-reference read fraction by SNP

WESTERN CAMBODIA

Josh Proctor (IDM)

WEST AFRICA

1

0

1

0

Single clone

Samples

High complexity

Data replotted from Pf3k project (Release 3.1)
previously published in Miotto *et al.* (2013, 2015)

The distribution of multi-strain infections and genetic relatedness encodes information on transmission topology
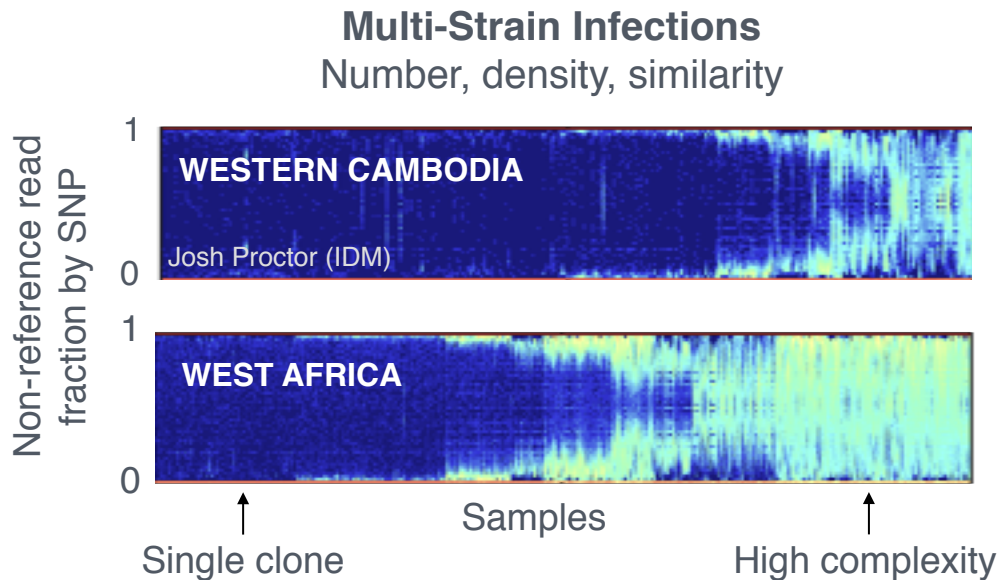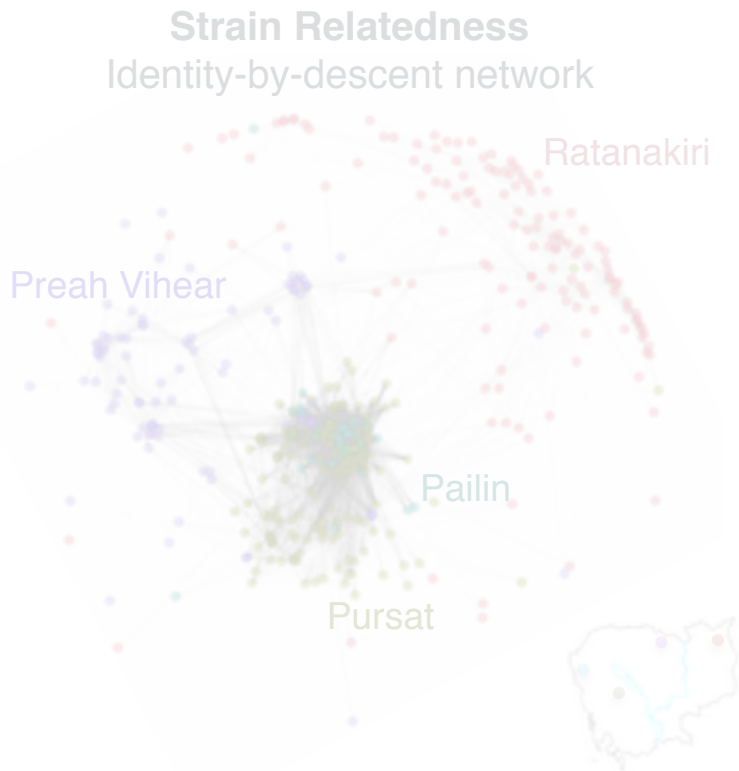

Strain Relatedness
Identity-by-descent network

Ratanakiri

Preah Vihear

Pailin

Pursat

Multi-Strain Infections
Number, density, similarity

Non-reference read fraction by SNP

1

WESTERN CAMBODIA

Josh Proctor (IDM)

0

1

WEST AFRICA

0

Samples

Single clone

High complexity

Data replotted from Pf3k project (Release 3.1)
previously published in Miotto *et al.* (2013, 2015)

From modeled features we can infer the characteristic transmission properties and the migration-based coupling.
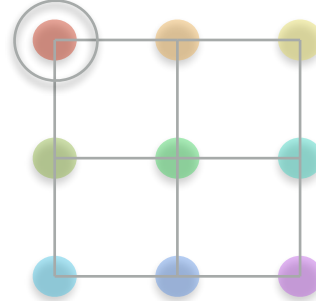

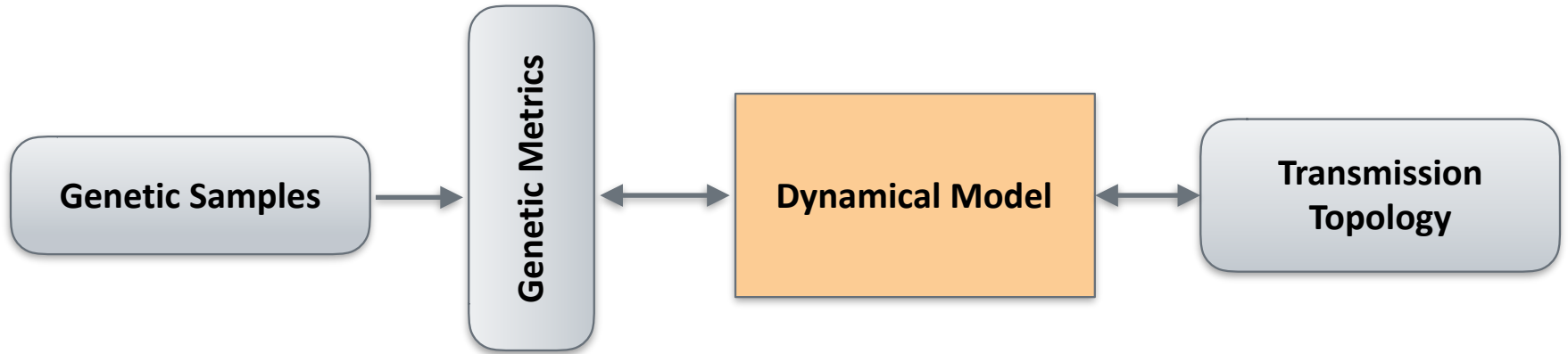Simulated IBD network example on 3x3 grid
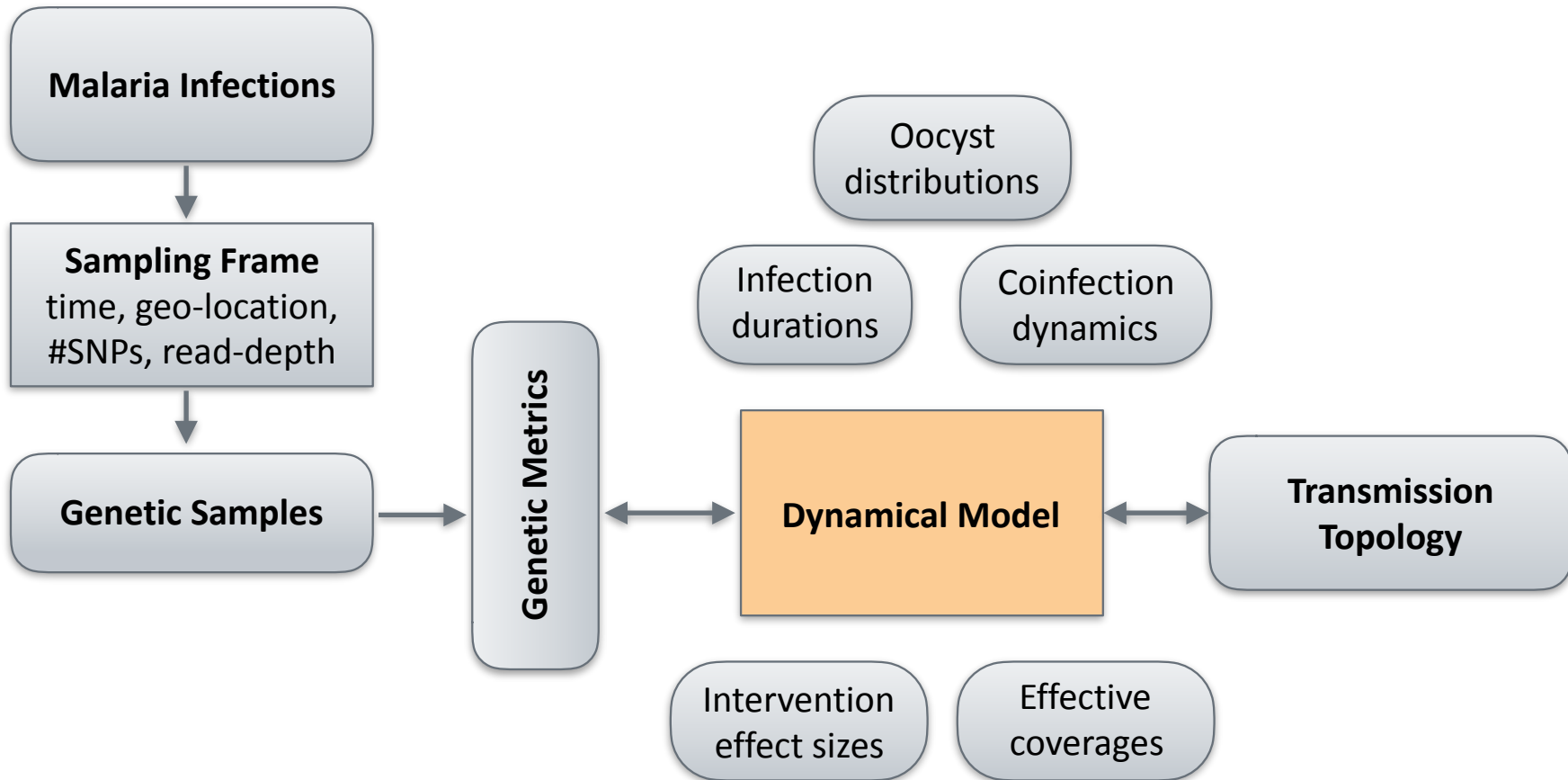
Emergence

Drug-resistant selective sweep on simulated network
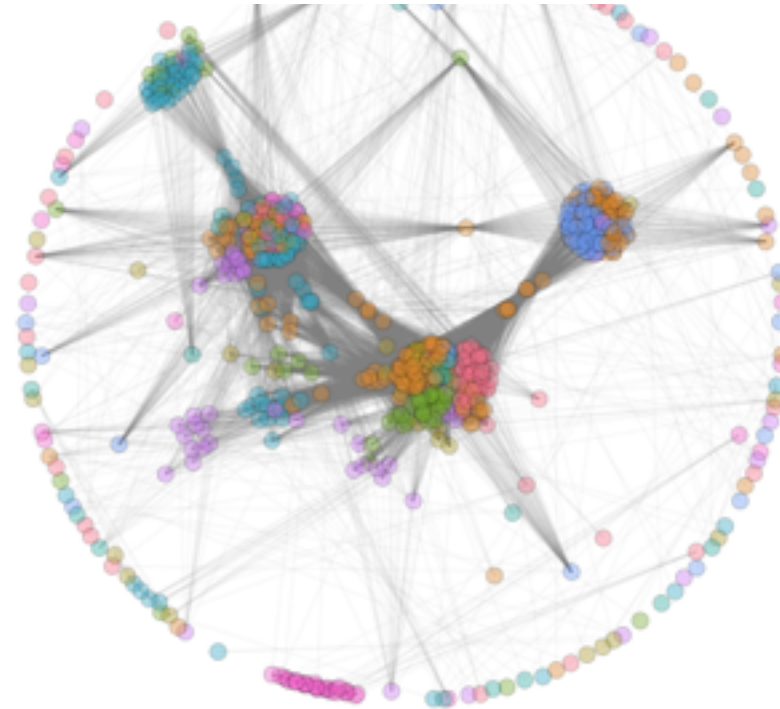
Specific challenges for interpretation

Malaria Infections

Sampling Frame
time, geo-location,
#SNPs, read-depth

Genetic Samples

Genetic Metrics

Transmission Topology

Intervention effect sizes

Effective coverages

household RDT+ prevalence

0.0   1.0
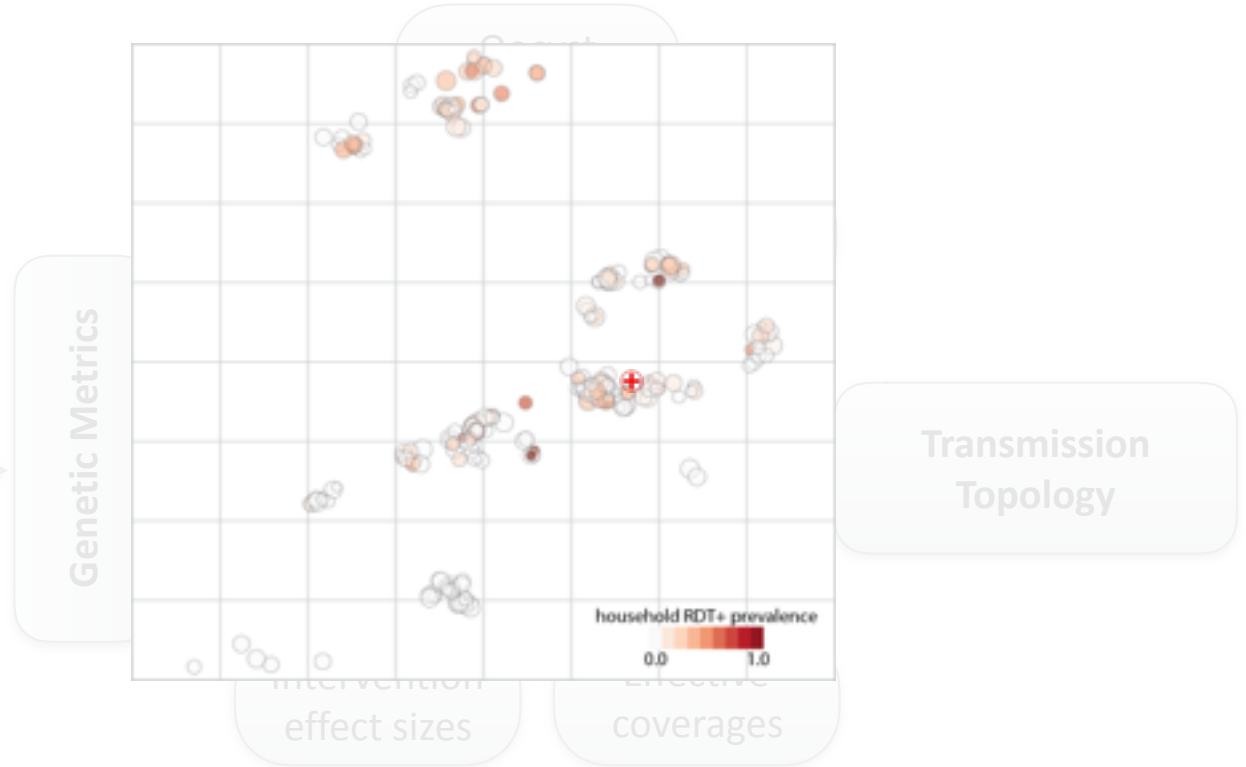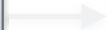
Specific challenges for interpretation
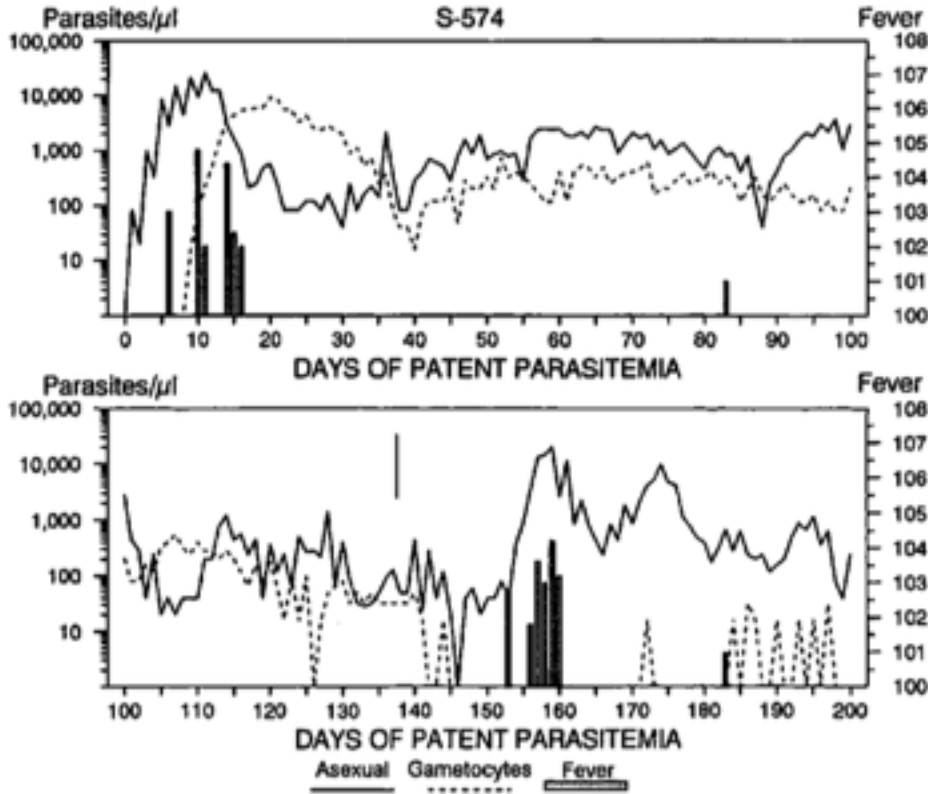
Collins and Jeffery (1999)

Oocyst distributions

Infection durations

Coinfection dynamics

Dynamical Model
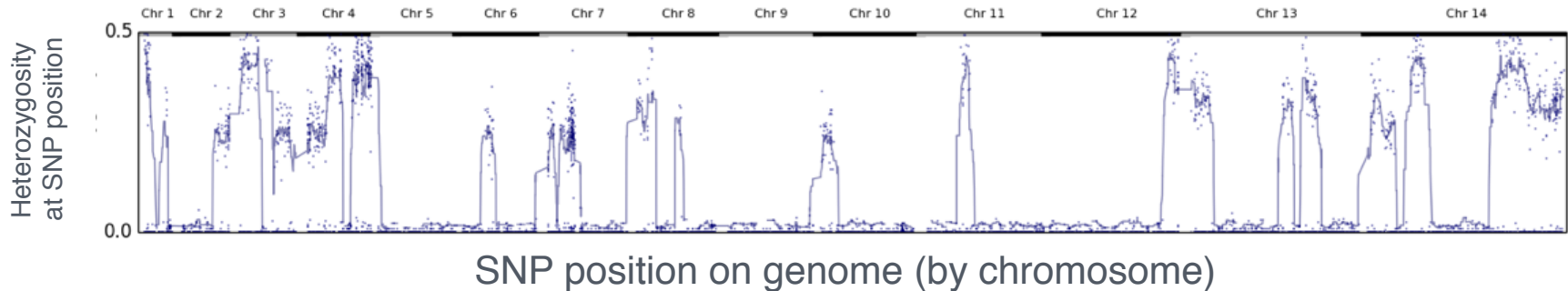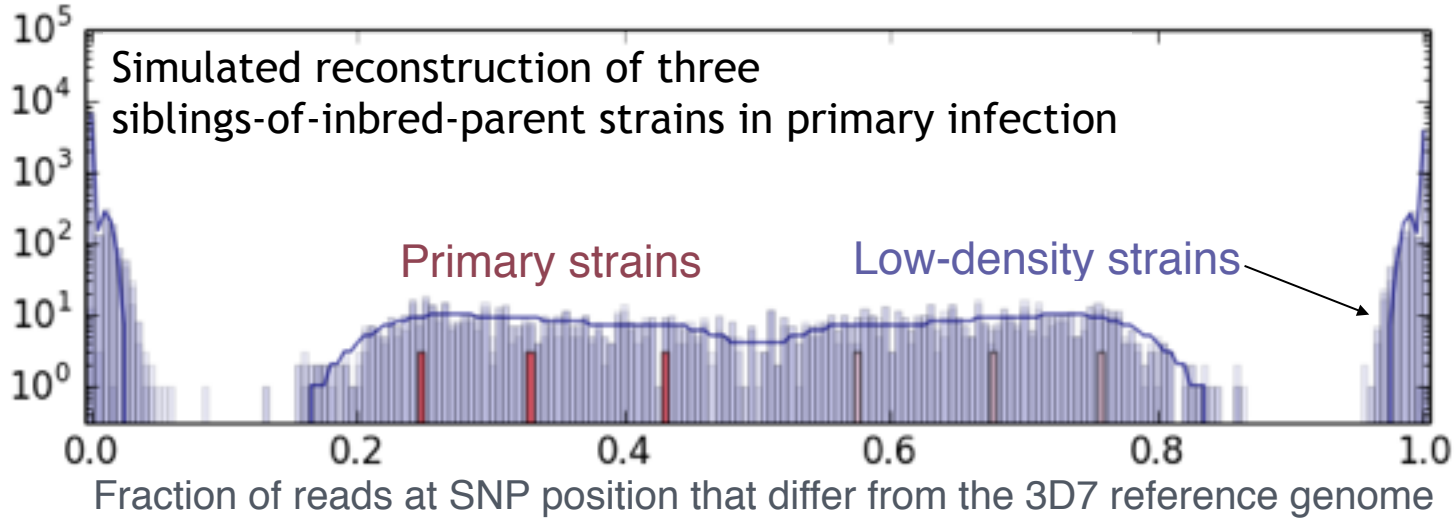
Transmission Topology

Effective coverages

Co-infection are a "headache" but also are information-rich

Simulated reconstruction of three siblings-of-inbred-parent strains in primary infection

Primary strains

Low-density strains

Coinfection dynamics

Fraction of reads at SNP position that differ from the 3D7 reference genome

Chr 1 Chr 2 Chr 3 Chr 4 Chr 5 Chr 6 Chr 7 Chr 8 Chr 9 Chr 10 Chr 11 Chr 12 Chr 13 Chr 14

Heterozygosity at SNP position

SNP position on genome (by chromosome)

Goals

- Converge on pathways for operational utility of existing (and near-term future) genetic sampling methods.

- Push the science and modeling of dynamical mechanisms forward to resolve technical challenges in quantitative interpretation of genetic samples.

Acknowledgements