

GDW Workshop: Landscape genomics



Elizabeth “Liz” Hemming-Schroeder

Assistant Professor

Center for Vector-borne Infectious Diseases (CVID)

Department of Microbiology, Immunology and Pathology

June 9, 2023

Outline

- Background on landscape genomics
- Case study: malaria vectors in Kenya
- Development of a cost-effective genomic assay

What is landscape genetics?

Research that combines

- *population genetics*
- *landscape ecology*
- *spatial analytical techniques*

*to explicitly quantify the effects of landscape composition,
configuration, or matrix quality*

on micro-evolutionary processes, such as gene flow, drift, and selection,

using neutral and/or adaptive genetic data.

Landscape genomics vs landscape genetics

- Landscape genomics uses genome-wide sequencing approaches
- The ubiquity of NGS, access to remotely-sensed data, and computing power has facilitated the rise in landscape genomic studies
- Since genomic approaches enables detection of loci under selection, landscape genomics is often used to refer to adaptive studies

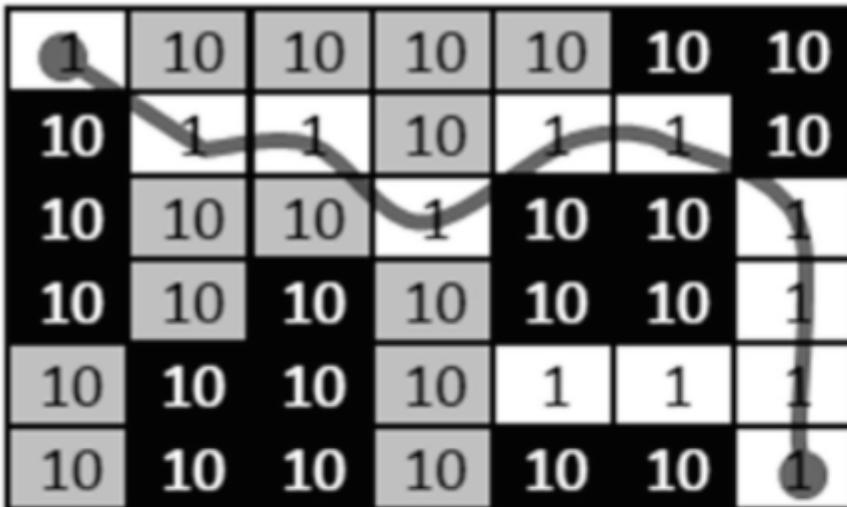
	Adaptive	Neutral
Main process of interest	Selection, local adaptation	Gene flow , genetic drift
Genetic markers	Non-synonymous SNPs, genomic approaches enables detection of loci under selection	Microsatellites, SNPs that are likely not under selection
Environmental influence	At-site (local habitat, climate, disturbances)	In-between (barriers, corridors, landscape permeability/resistance) and at- site
Analytical approach	Identify loci under selection, statistically relate adaptive genetic data to environmental heterogeneity, environmental association analysis	Genetic similarity/dissimilarity estimated and compared to functional landscape connectivity, isolation-by- barrier, isolation-by-resistance

Applications to wildlife diseases

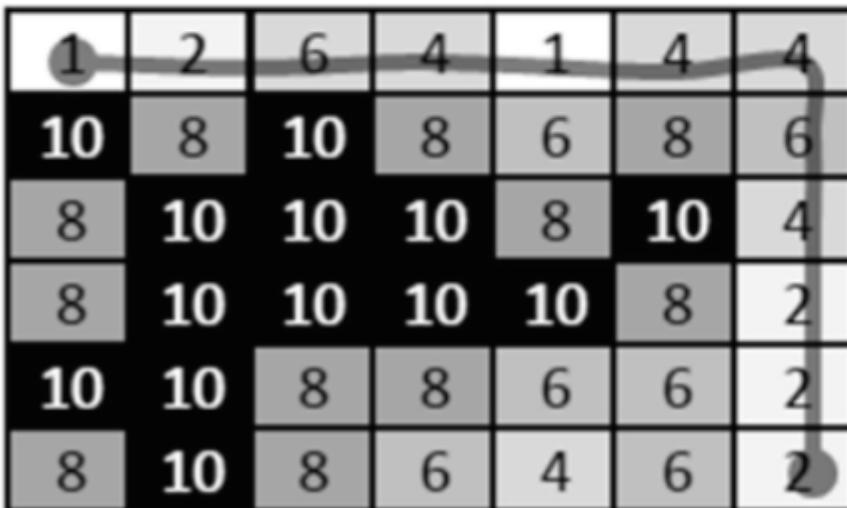
- Identify factors that influence wildlife dispersal and disease distribution
 - Robinson et al. ([2013](#)) reported that including landscape barriers to deer dispersal, identified through landscape genetics, significantly improved their predictive power to model the spatial distribution of CWD in southern Wisconsin and northern Illinois
- How genetic variation across the landscape is directly related to disease risk, susceptibility, or immune competence
- Assess environmental heterogeneity related to selection acting on both hosts and pathogens

What landscape features facilitate/impede gene flow?

- Individual or population-based approach
- Infer relative gene flow
 - Inversely related to genetic distance (e.g. F_{ST})
 - Identity-by-descent or identity-by-state
- Isolation-by-distance is typically the null model
- Between-site characteristics are primary focus
- Test associations with between-site landscape resistance to genetic distance/relatedness



+



Landcover Resistance

Grassland		10
Meadow		10
Forest		1

Elevation

Low \longleftrightarrow High



1 \longleftrightarrow 10
Resistance

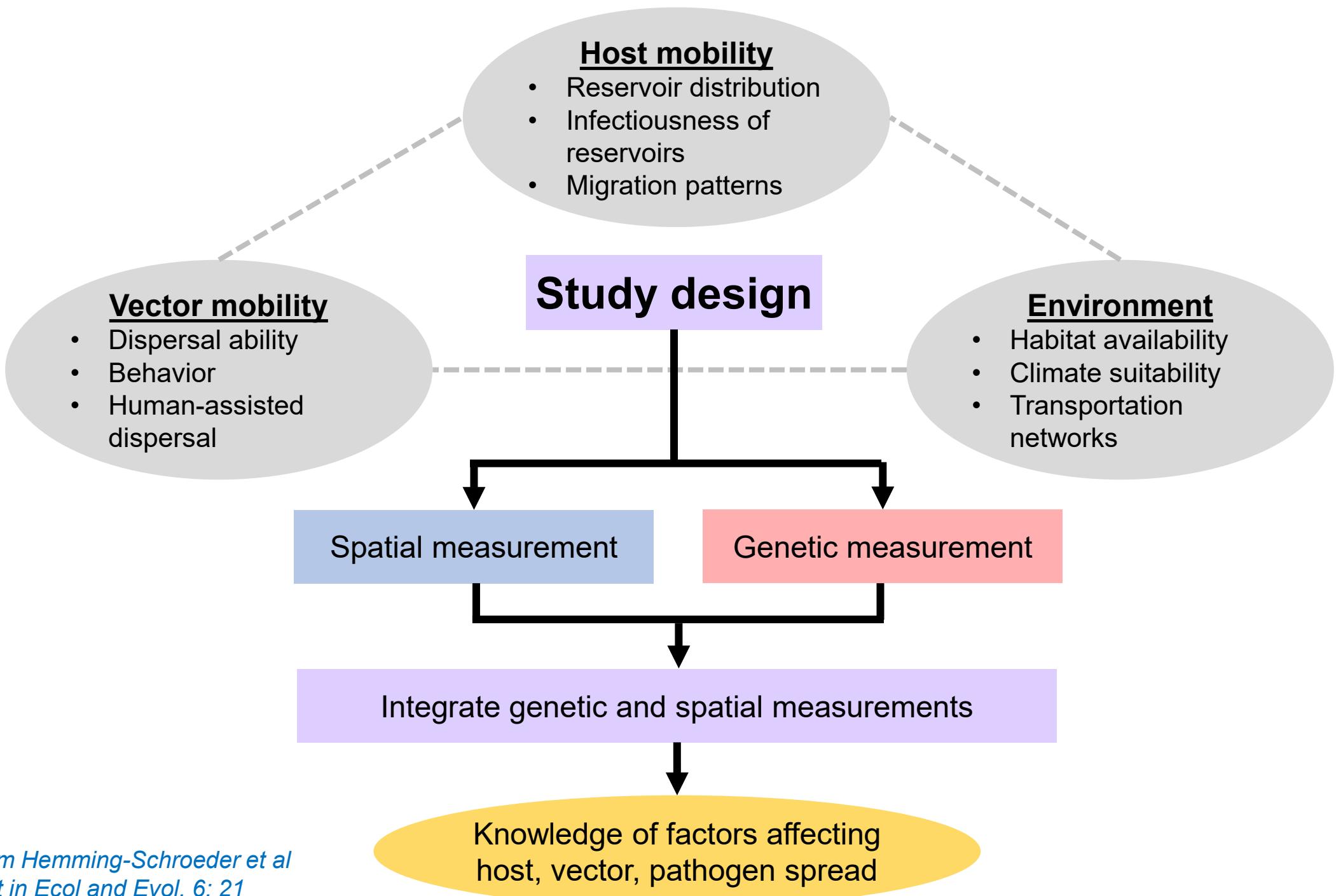
Popular satellite sources

- MODIS: high temporal, low spatial resolution
- Landsat-8: lower temporal, higher spatial resolution
- Sentinel-2: highest spatial resolution, least history

	MODIS	Landsat-8	Sentinel-2
Launch date	1999	2013	2015
Temporal resolution	Daily	8 day	5 day
Spatial resolution	250 m	30 m	10 m
Spectral bands	36	11	13

Where do I get remote sensing data?

- NASA Earth Data: <https://search.earthdata.nasa.gov/search>
- USGS Earth Explorer: <https://earthexplorer.usgs.gov/>
- ESA's sentinel data at the Copernicus Open Access Hub:
<https://scihub.copernicus.eu/dhus/#/home>
- Google Earth Engine: <https://earthengine.google.com/>



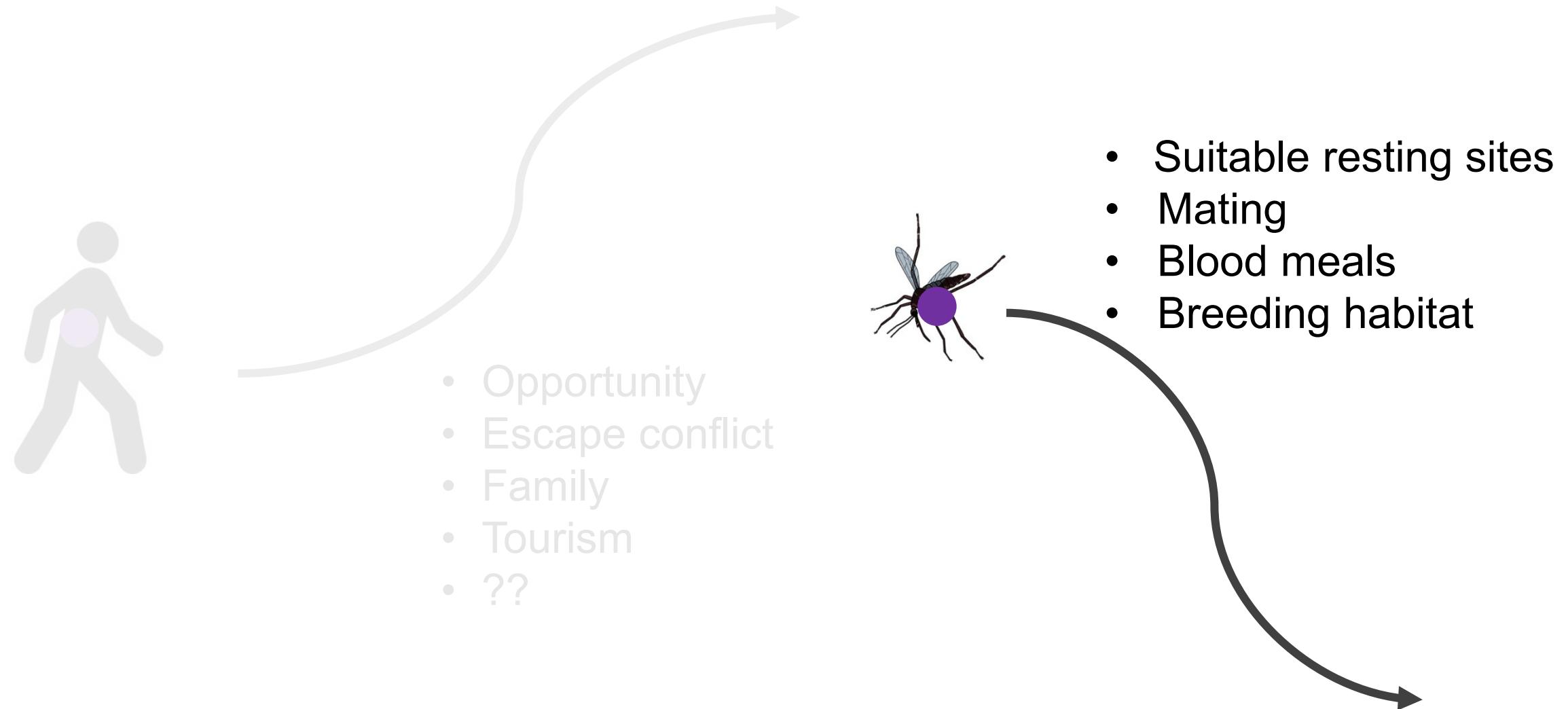
Case study

Ecological drivers of genetic connectivity for African malaria vectors *Anopheles gambiae* and *An. arabiensis*

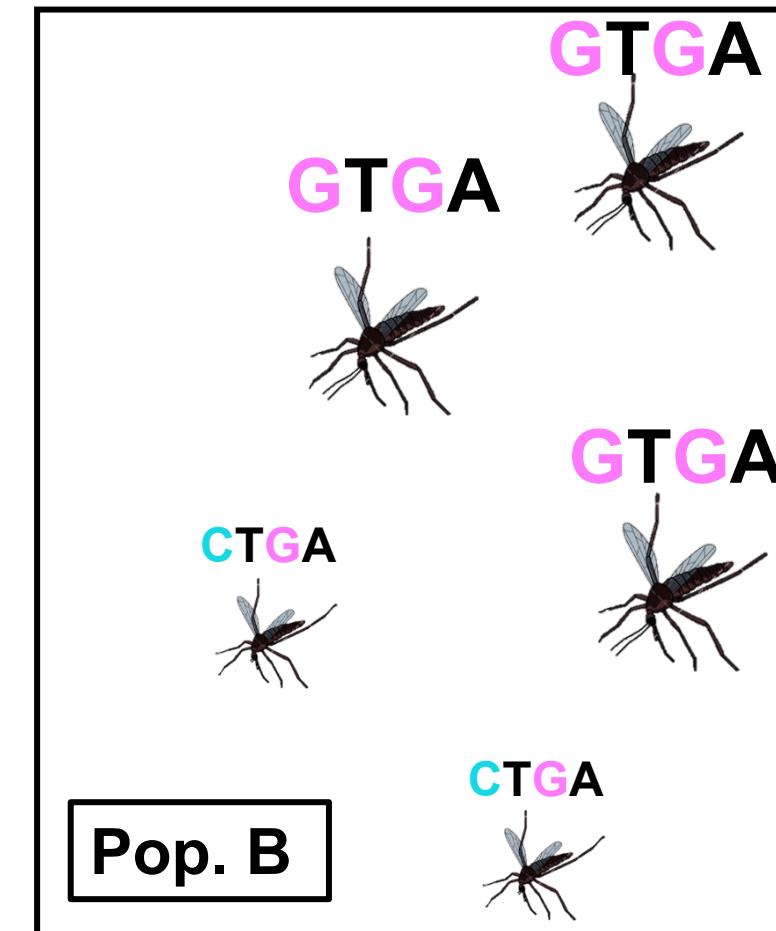
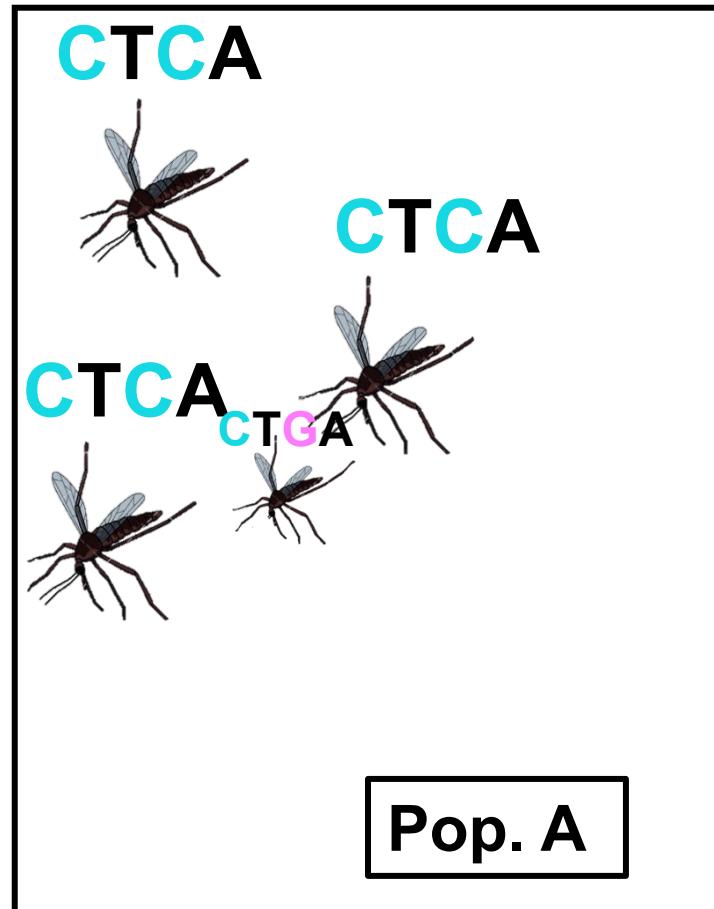
[Elizabeth Hemming-Schroeder](#), [Daibin Zhong](#), [Maxwell Machani](#), [Hoan Nguyen](#), [Sarah Thong](#),
[Samuel Kahindi](#), [Charles Mbogo](#), [Harrysone Atieli](#), [Andrew Githeko](#), [Tovi Lehmann](#), [James W. Kazura](#)
& [Guiyun Yan](#) 

[Scientific Reports](#) **10**, Article number: 19946 (2020) | [Cite this article](#)

Malaria parasites travel by people and mosquitoes



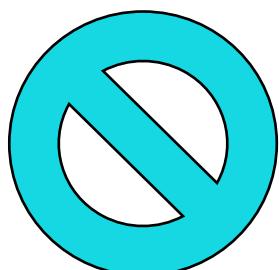
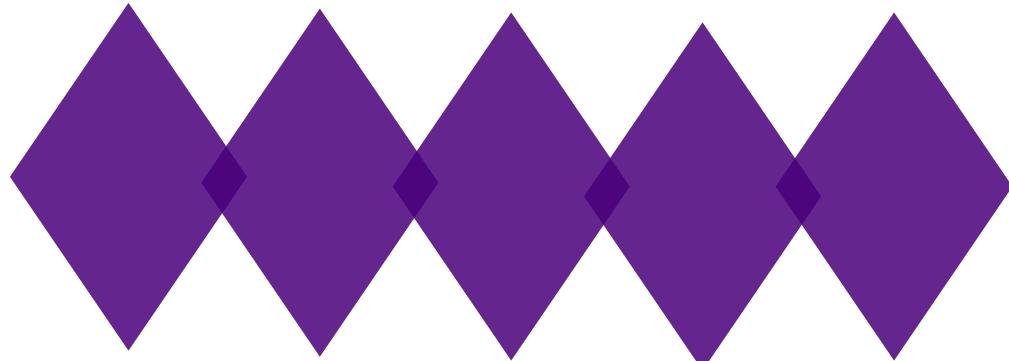
Population genetics reveals migration patterns



Understand how mosquitoes migrate



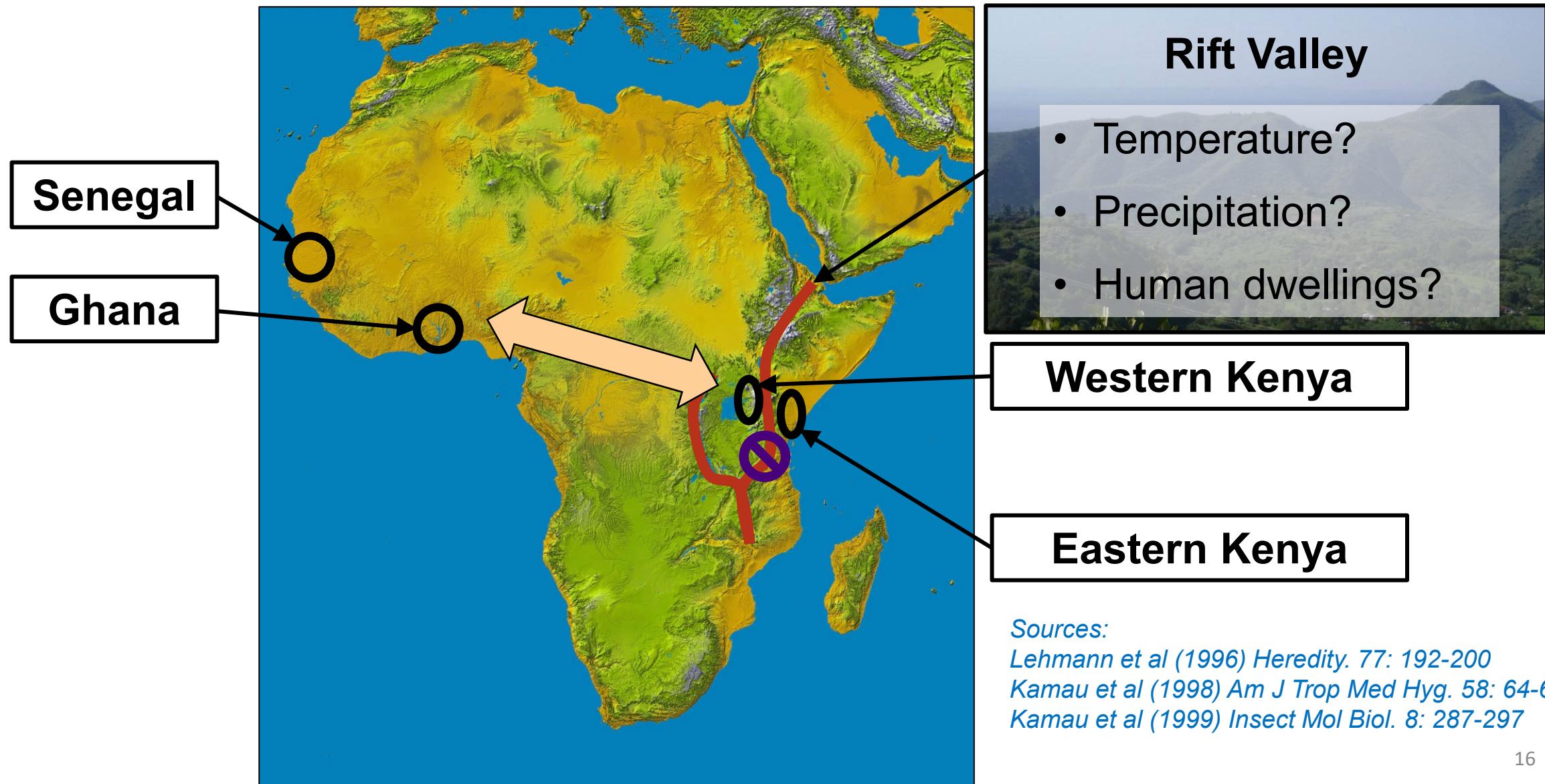
• A



• B

- Help inform interventions:
 - insecticide resistance spread
 - vector population control
 - reintroduction of malaria parasites

Traditional approach: Isolation-by-distance



Concept: Landscape genetics

CTCA GTGA

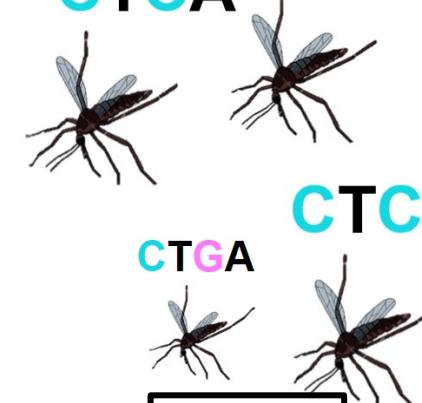


CTGA

Pop. A



GTGA
CTCA



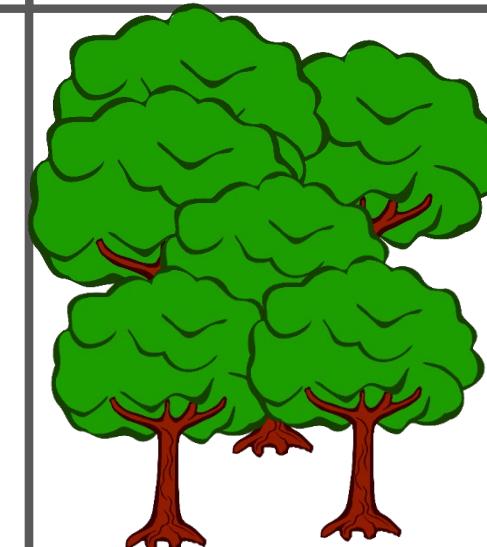
Pop. B



GGGC



Pop. C



Habitat and behavioral differences

Anopheles gambiae

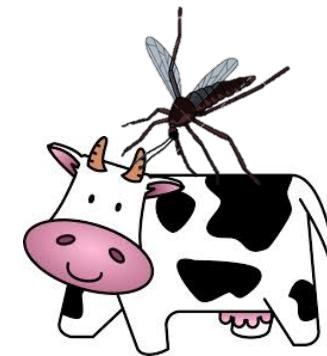


Feed indoors,
mostly on people

Prefers humid
environments



Anopheles arabiensis



Feed outdoors,
on animals & people

Adapted to arid climate



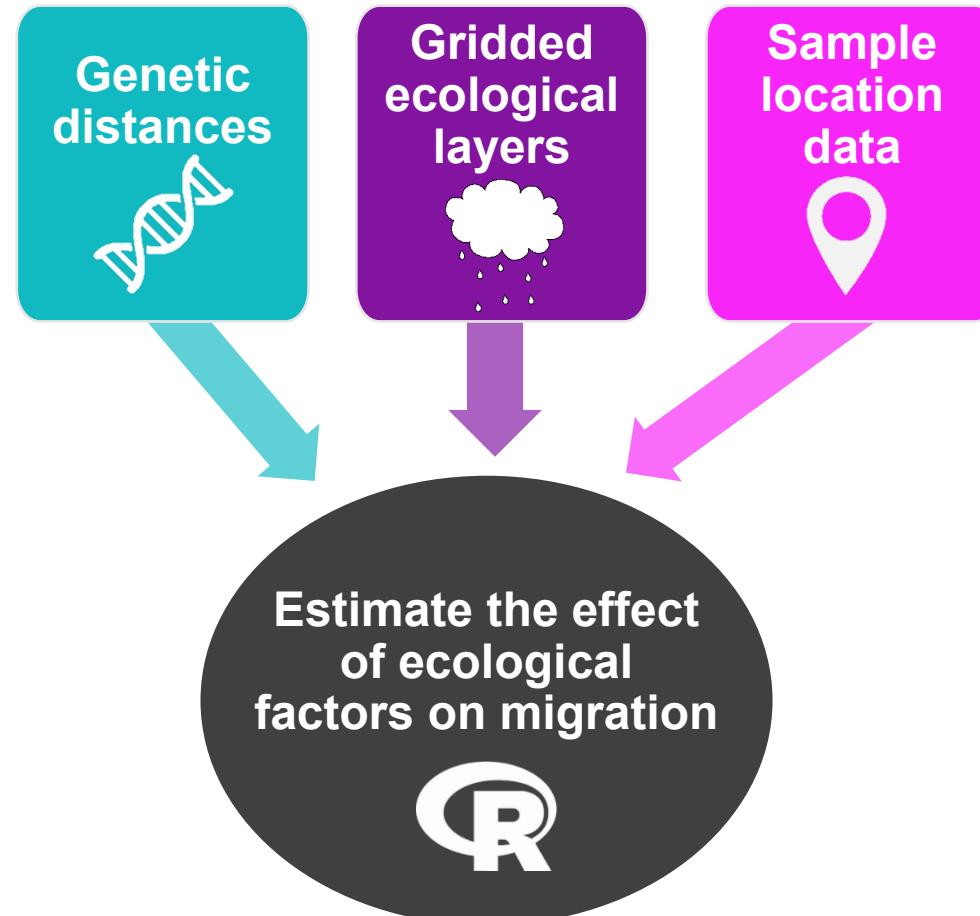




Mosquito larvae collection



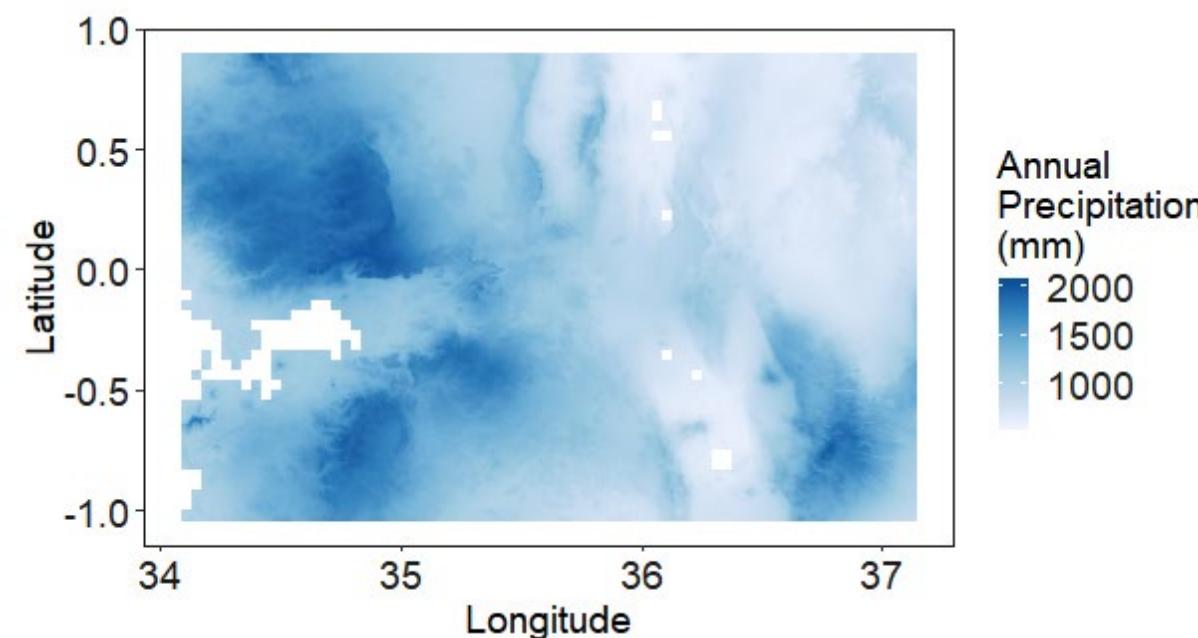
Estimate the effect of ecological factors on migration



Category	Variable	Source
Environmental	Average temperature	WorldClim BIO1
	Annual Precipitation	WorldClim BIO12
Landscape	Percent tree cover	NASA MOD44B
	Cropland	NASA MOD12Q
Social	Human population density	Worldpop
	Road proximity	OpenStreetMap

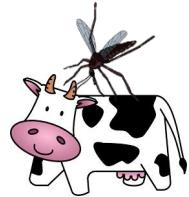
Estimate the effect of ecological factors on migration

Precipitation ecological layer

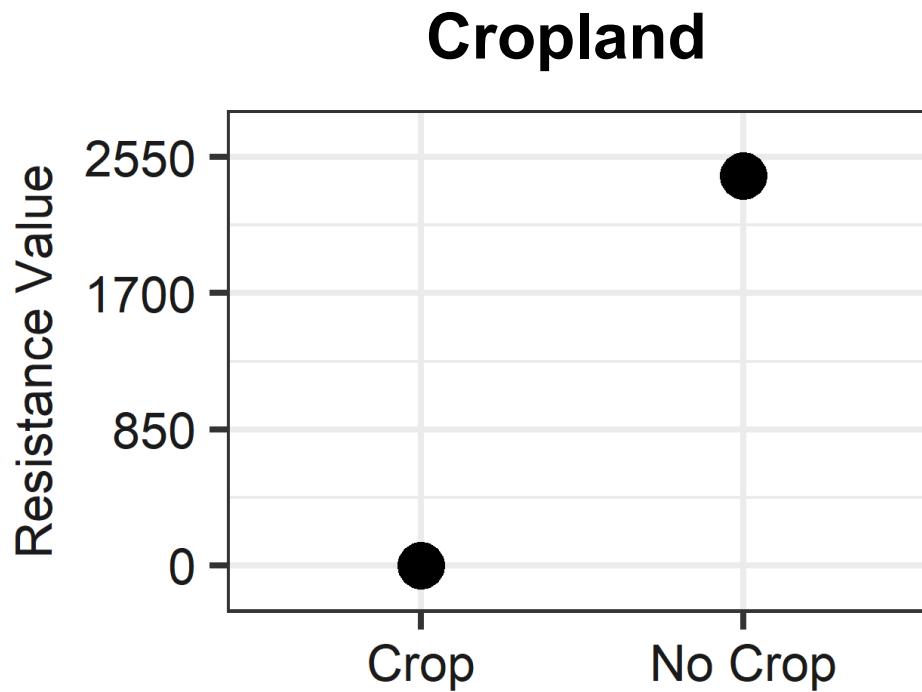


Category	Variable	Source
Environmental	Average temperature	WorldClim BIO1
	Annual Precipitation	WorldClim BIO12
Landscape	Percent tree cover	NASA MOD44B
	Cropland	NASA MOD12Q
Social	Human population density	Worldpop
	Road proximity	OpenStreetMap

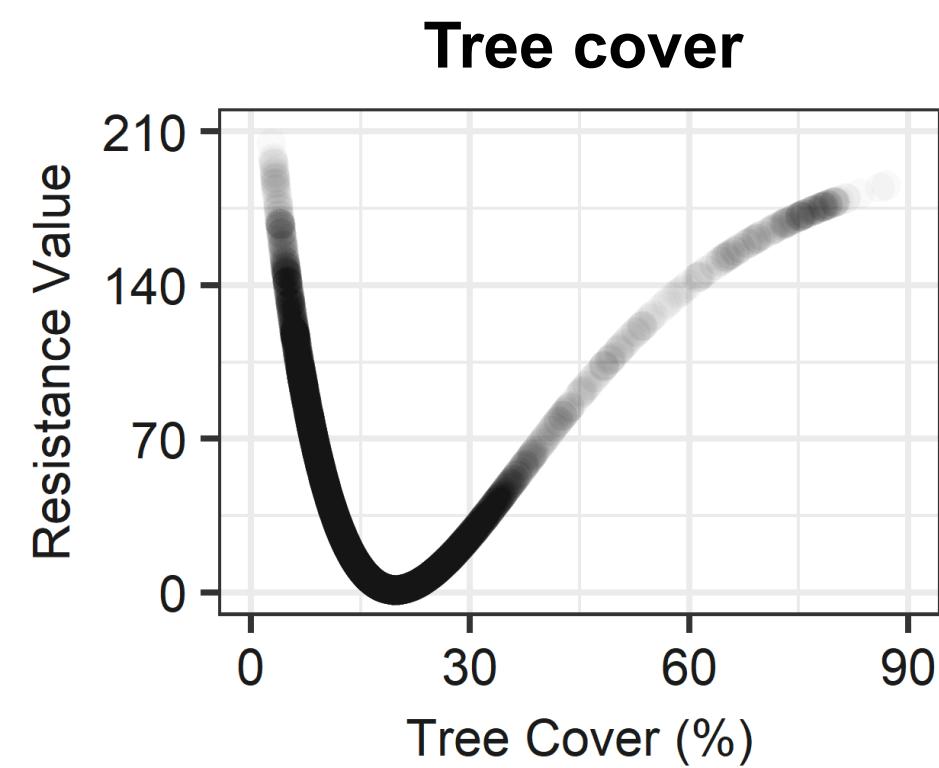
Relationship between ecological variables and landscape resistance to gene flow



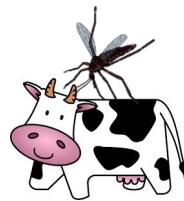
An. arabiensis



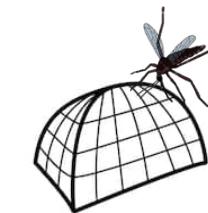
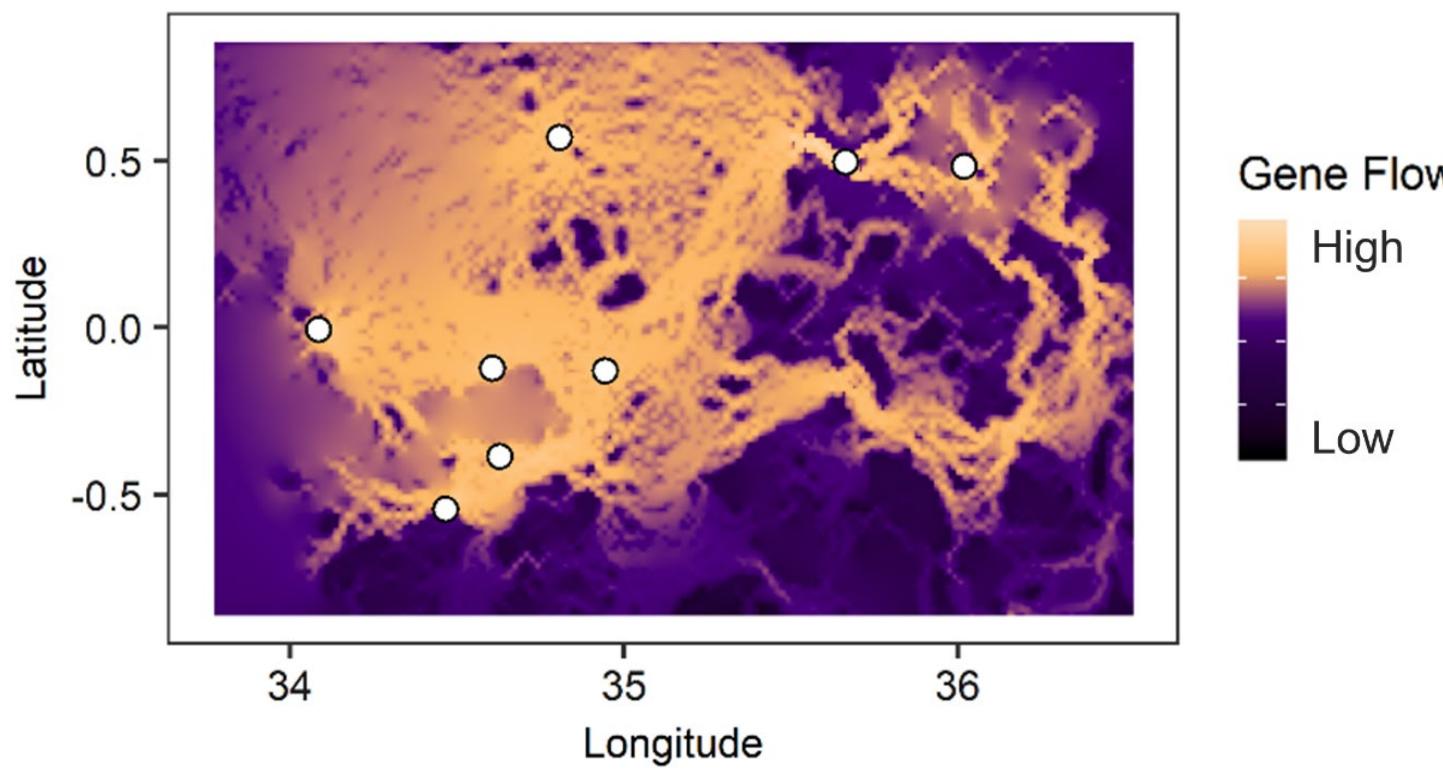
An. gambiae



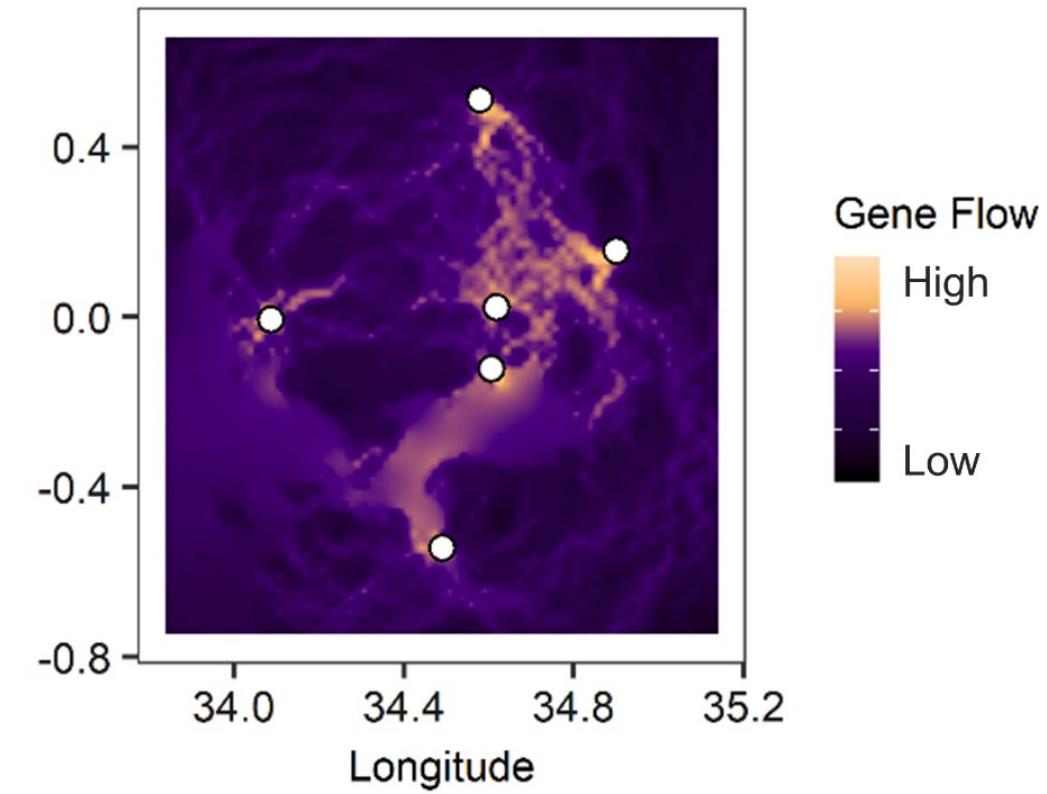
Potential migration pathways between populations



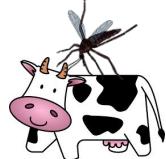
An. arabiensis



An. gambiae



Summary of case study

- Corridors for *An. arabiensis* are areas with cropland 
- Corridors for *An. gambiae* are areas of moderate tree cover 
- Human induced land cover and land use modifications may enhance connectivity of these species
- This knowledge can help us more effectively deploy interventions

Comparison of potential genotyping assays for population genetics

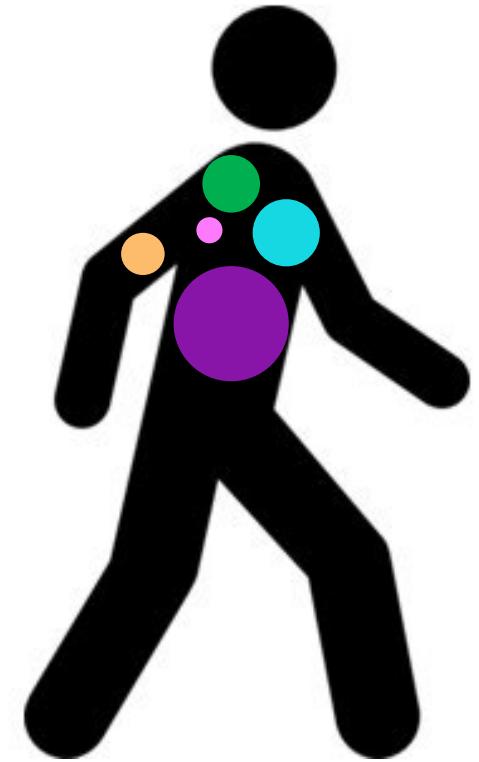
	Proportion of genome sampled	Throughput	Ease to start	Genotype point mutations of interest	Cost per sample
Microsatellite panel	X	X	●	X	~\$20
SNP array	●	✓	●	✓	~\$30
Highly multiplexed deep amplicon sequencing	●	✓	●	✓	~\$6
RAD-seq	✓	●	✓	X	~\$30
Whole genome sequencing	✓	●	✓	✓	~\$100

Highly multiplexed amplicon sequencing panel for pathogens

- **Advantages**

- Cost effective
- Possible to resolve haplotypes in complex infections
- Sensitive to detect changes/differences in multiplicity of infection
- Can also sequence specific genetic markers of interest

- **Design:** 200-300bp genome-wide, highly polymorphic targets (n=130)



Assessing microhaplotypes from amplicon sequences

- SNPs are often biallelic

A T G **C** T A A (=C)

A T G **G** T A A (=G)

A T G **G** T A A (=G)

- Relative to biallelic SNPs, microhaplotypes provide increased power for relationship inference (Baetscher *et al Mol Ecol* 2017. doi:10.1111/1755-0998.12737)

- Microsatellites are often multiallelic

AGT (=1 repeat)

AGT AGT (=2 repeats)

AGT AGT AGT AGT (=4 repeats)

- In *P. falciparum*, microhaplotypes provided higher resolution to discriminate related from unrelated polyclonal infections than biallelic SNP barcodes (Tessema *et al J Infect Dis* 2020. doi:10.1093/infdis/jiaa527)

- Microhaplotypes are multiallelic

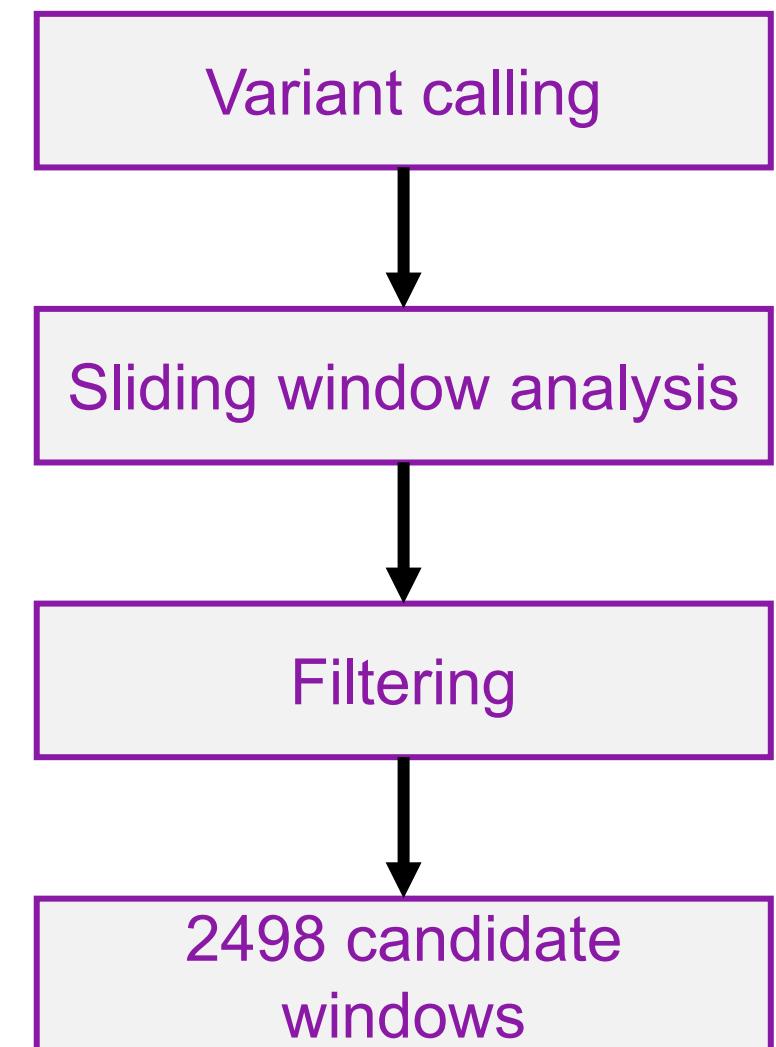
A **T** A A G **C** C A G T **T** A G A (=TCT)

A **A** A A G **G** C A G T **A** A G A (=AGA)

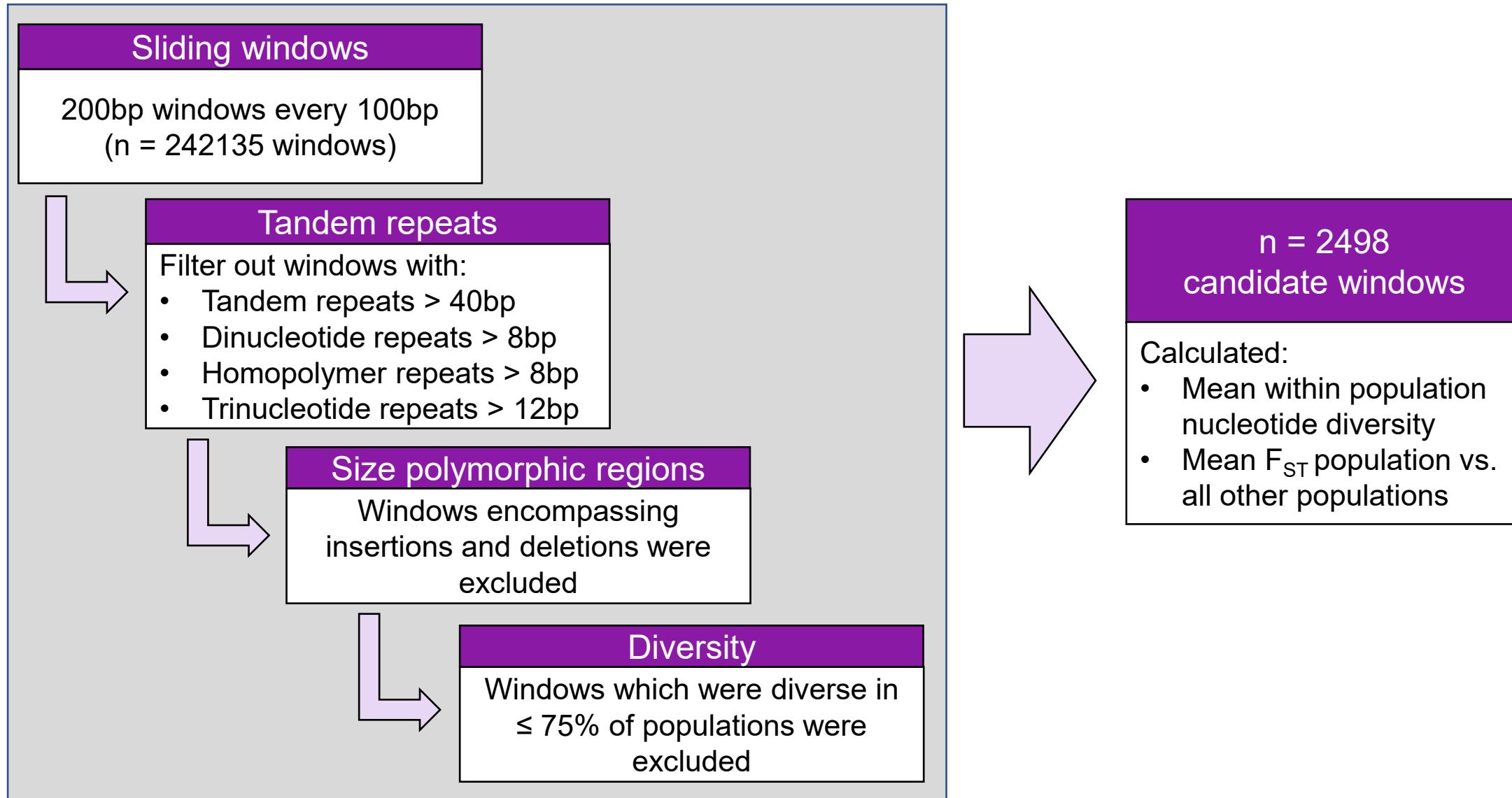
A **T** A A G **C** C A G T **A** A G A (=TCA)

Example: development of amplicon panel for *P. vivax*

Country	n	Reference
Cambodia	31	Accession: PRJNA295233
China-Myanmar	28	Chen <i>et al.</i> <i>BMC Genomics</i> (2017) 18:131
Colombia	8	Winter <i>et al.</i> <i>PLoS NTD</i> (2015) 9:e0004252
Ethiopia	33	Auburn <i>et al.</i> <i>J Infect Dis</i> (2019) 220: 1738 Lo <i>et al.</i> <i>PLoS NTD</i> (2019) 13: e0007222
Madagascar	9	Menard <i>et al.</i> <i>PLoS NTD</i> (2013) 7:e2489
Malaysia	30	Auburn <i>et al.</i> <i>Nature Commun</i> (2018) 9:1
Panama	29	Accession: PRJNA655141
Peru	30	Cowell <i>et al.</i> <i>Genome Medicine</i> (2018) 10:52

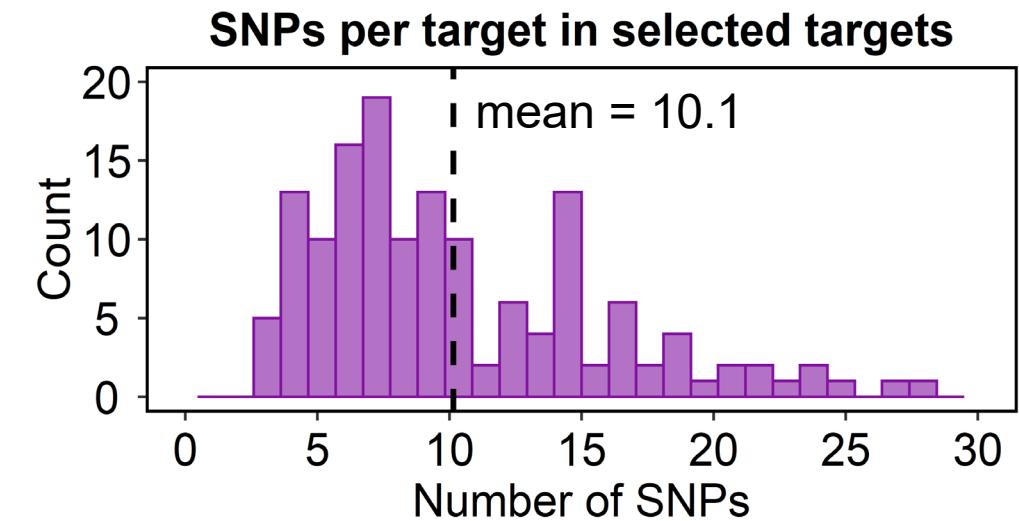
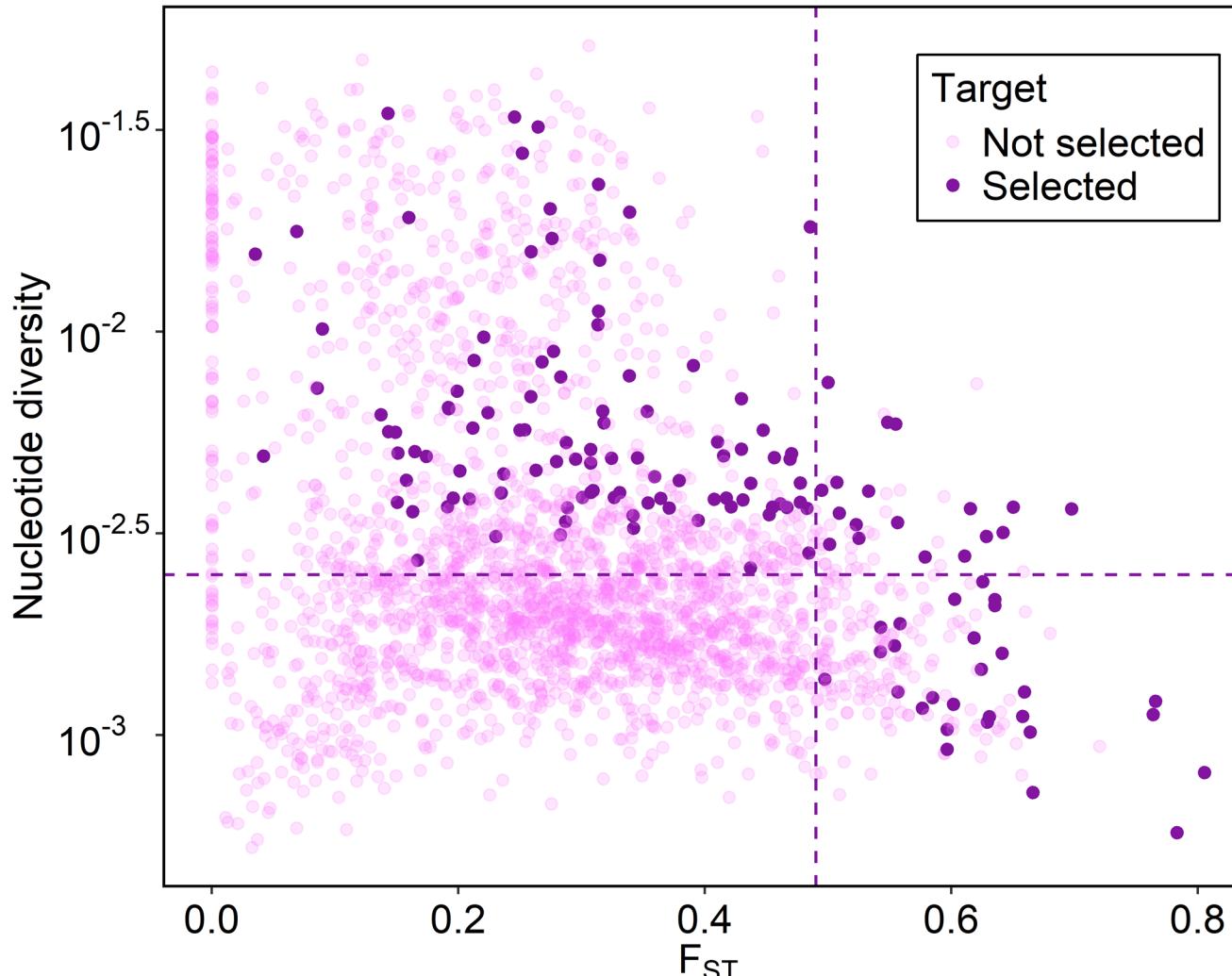


Sliding window analysis

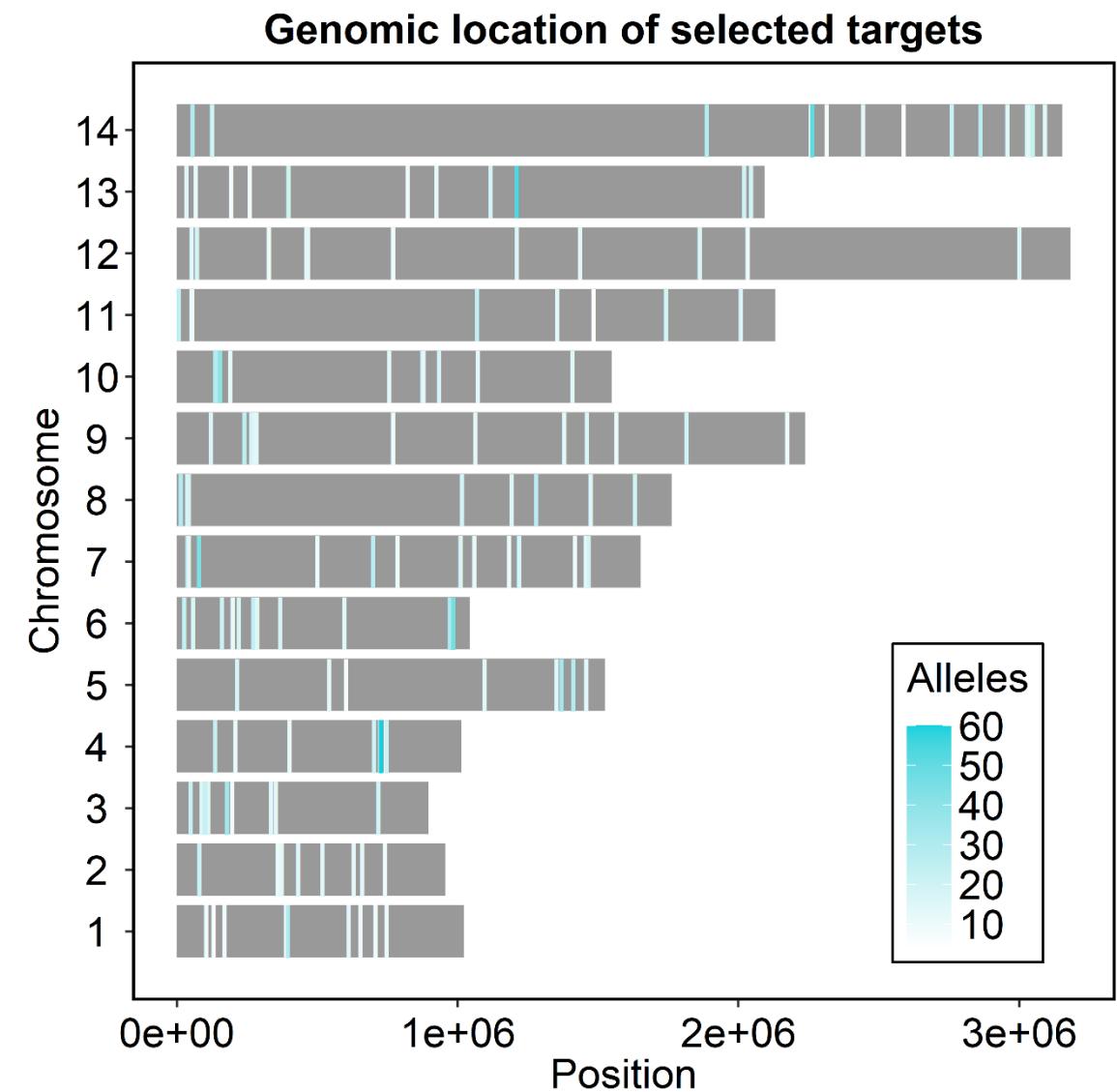
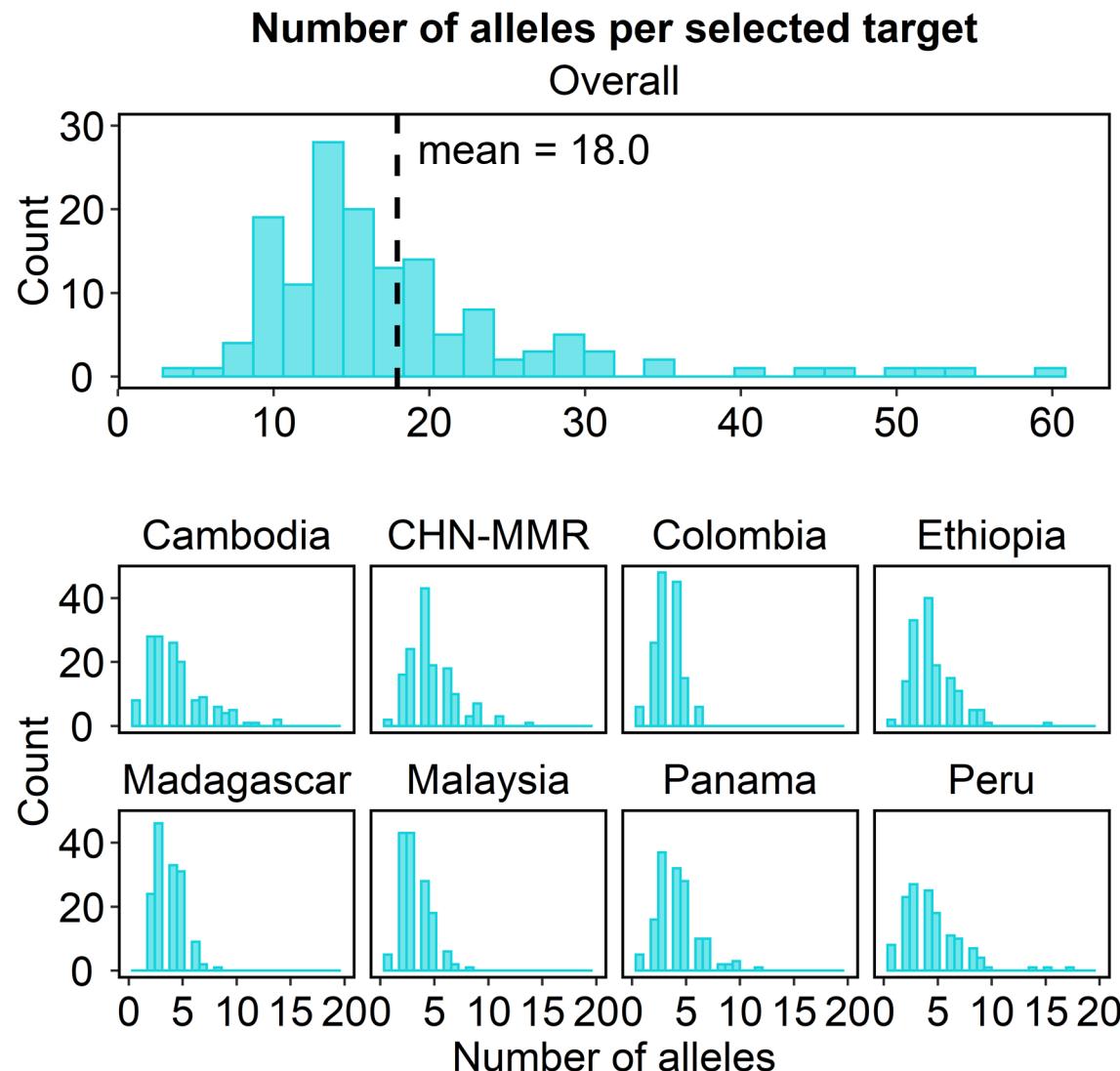


Selection of *P. vivax* targets for panel

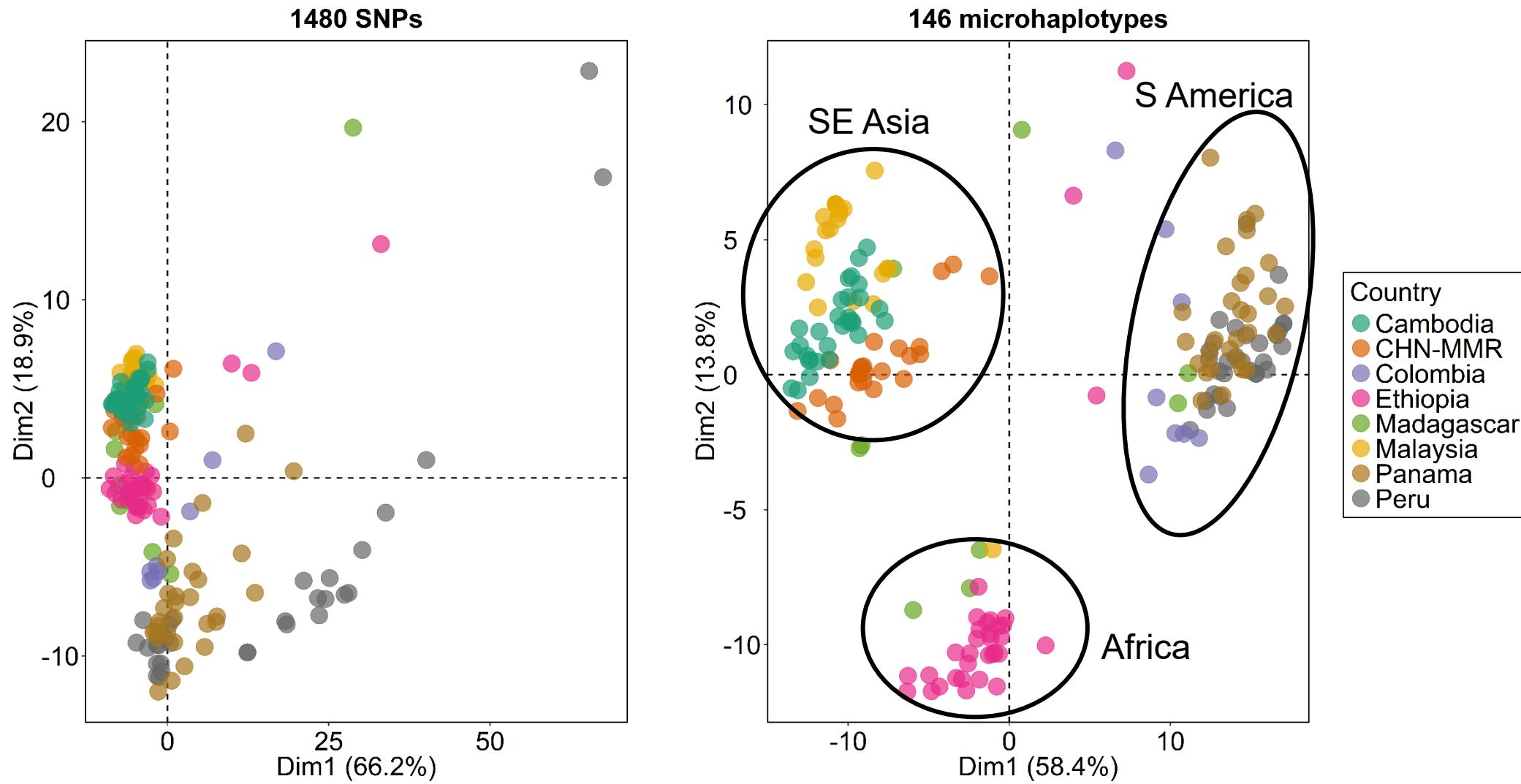
Relationship between genetic diversity and differentiation of targets



Polymorphism of *P. vivax* microhaplotype alleles



Principle component analysis based on selected *P. vivax* targets



Individual-based pairwise genetic distance metric: D_{PS} (1-proportion of shared alleles)

Questions?

ehemming@colostate.edu