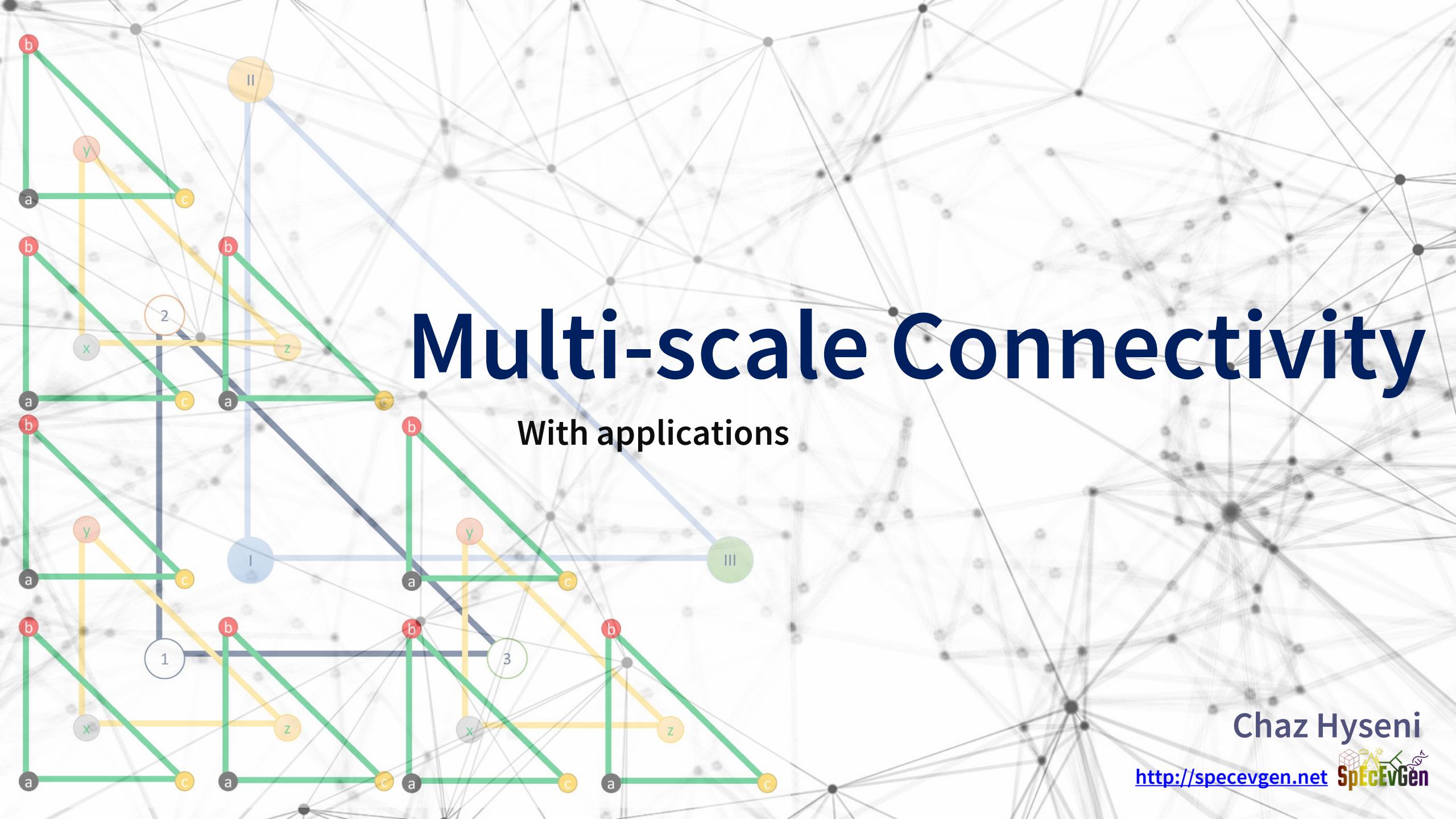


Multi-scale Connectivity

With applications



Chaz Hyseni

<http://specgen.net> SpecGen

Connectivity: the spatial arrangement of suitable habitat determines dispersal of organisms

Scales of Connectivity:

Among individuals = populations (Hardy-Weinberg eq./diseq.)

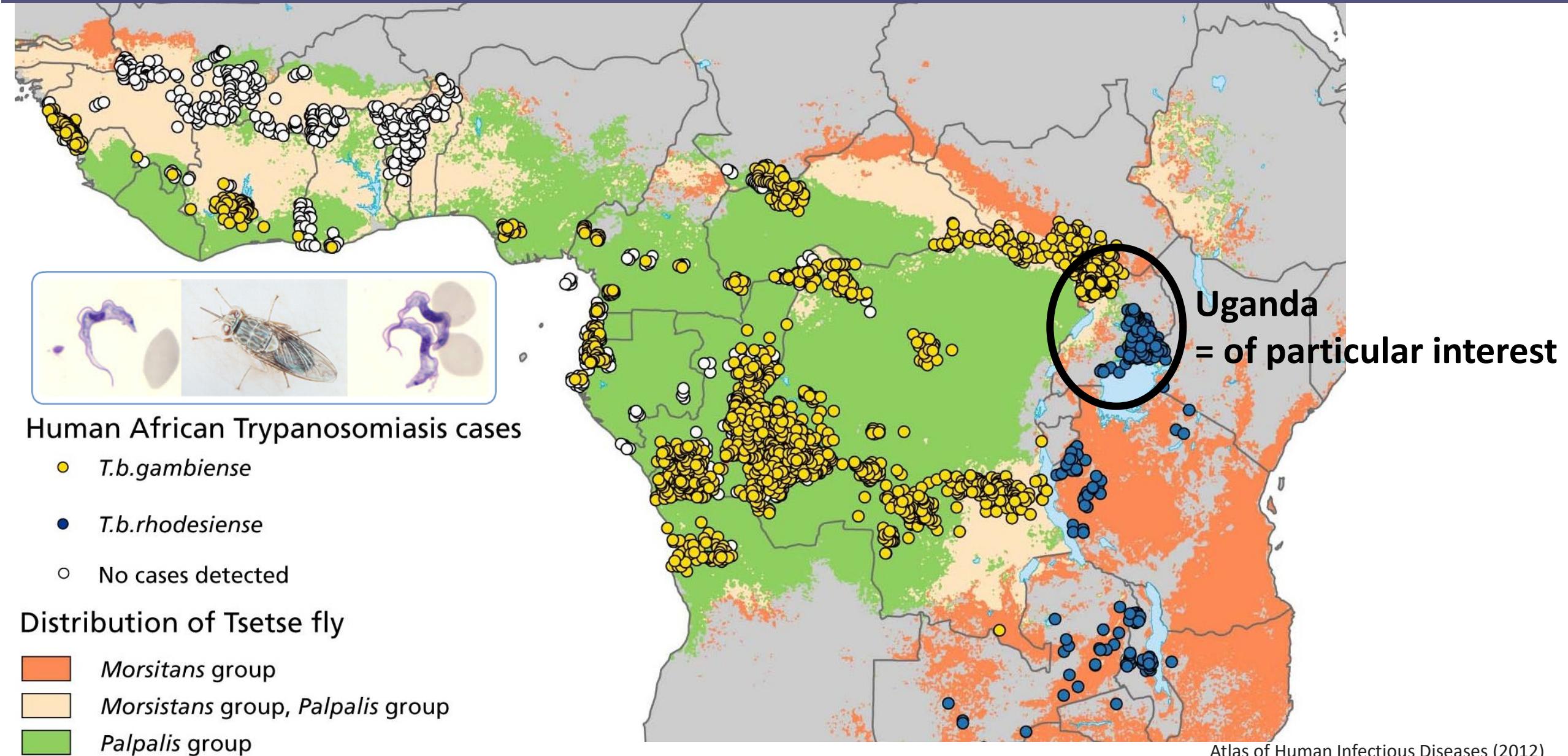
Among populations = gene flow

Among species = communities

Among communities = meta-communities

Motivation: a scalable measure of connectivity, to facilitate analyses in an eco-evo framework

Inspiration: Tsetse, Vectors of African Trypanosomiasis



Multi-Scale Connectivity = MS_{Conn}

$$\text{MS}_{\text{Conn}} = \frac{1}{n} \frac{(\sum_{g=1}^n P_g)^2}{\sum_{g=1}^n P_g^2} = \frac{1}{n \sum_{g=1}^n P_g^2}$$

Inverse of probability that any two ...
are found in the same ...

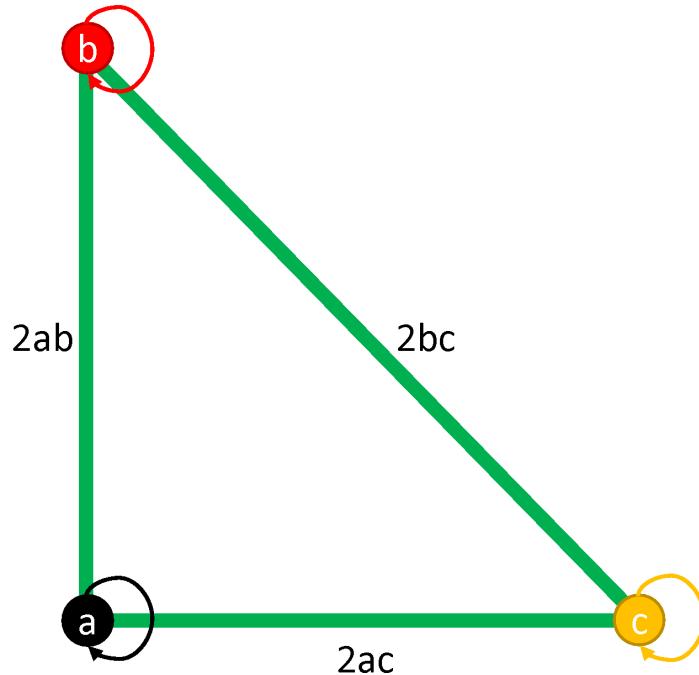
Multi-Scale Connectivity = MS_{Conn}

$$MS_{Conn} = \frac{1}{n} \frac{(\sum_{g=1}^n P_g)^2}{\sum_{g=1}^n P_g^2} = \frac{1}{n \sum_{g=1}^n P_g^2}$$

Inverse of probability that any two
IDENTICAL ALLELES are found in the same
INDIVIDUAL

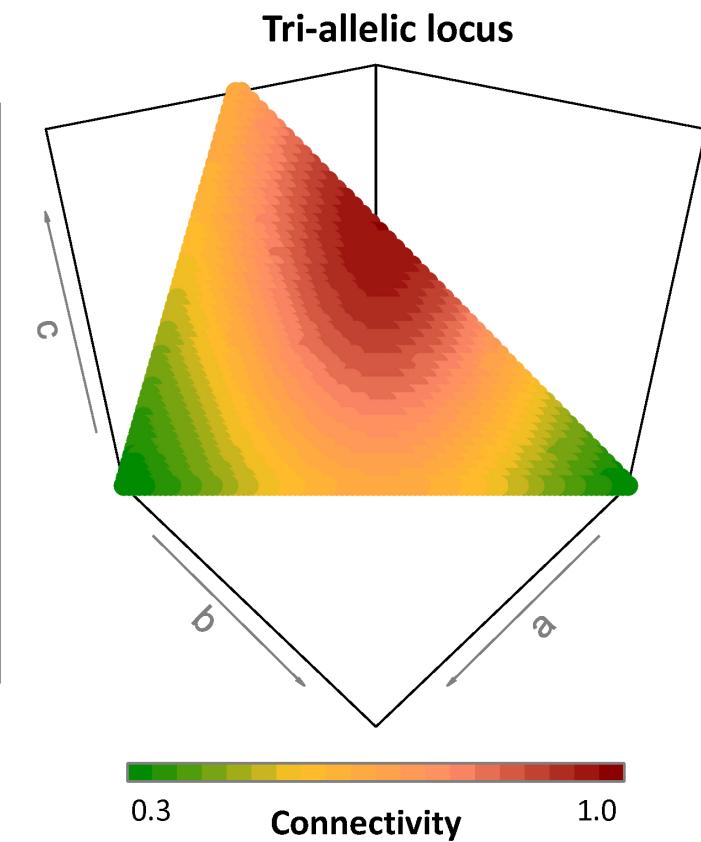
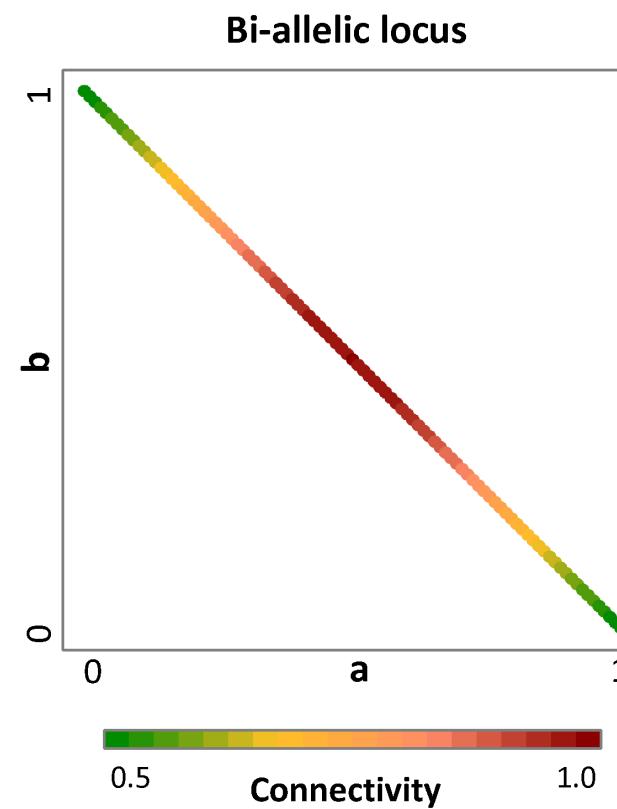
Connectivity among Individuals

Effect of allele frequencies on within-population connectivity (assuming Hardy-Weinberg equilibrium)



Maximum Connectivity:
When frequencies $a = b = c$

$$\frac{1}{n \sum_{a=1}^n p_a^2} = \frac{1}{3 (0.33^2 + 0.33^2 + 0.33^2)} = 1$$



Minimum Connectivity:
When one allele goes toward fixation

Multi-Scale Connectivity = MS_{Conn}

$$MS_{Conn} = \frac{1}{n} \frac{(\sum_{g=1}^n P_g)^2}{\sum_{g=1}^n P_g^2} = \frac{1}{n \sum_{g=1}^n P_g^2}$$

Inverse of probability that any two INDIVIDUALS (from the same “group”) are found in the same POPULATION

Multi-Scale Connectivity = MS_{Conn}

$$MS_{Conn} = \frac{1}{n} \frac{(\sum_{g=1}^n P_g)^2}{\sum_{g=1}^n P_g^2} = \frac{1}{n \sum_{g=1}^n P_g^2}$$

Inverse of probability that any two INDIVIDUALS (from the same “group”) are found in the same POPULATION

EXAMPLE:

$$\frac{1}{3 \times 0.1^2 \times 0.8^2 \times 0.1^2} = \frac{1}{3 \times 0.66} = \frac{1}{1.98} = 0.51$$

Among-Population Connectivity

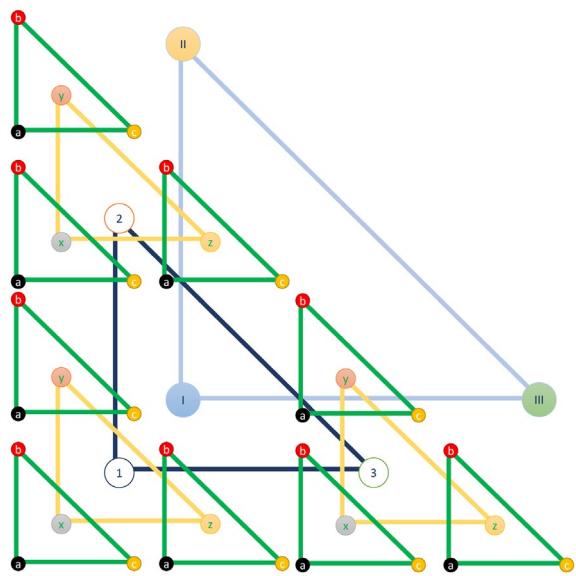
populations categorized into groups (g)

e.g., genetic lineages (Gen.)

	Pop. 1	Pop. 2	Pop. 3
Gen. 1	0.1	0.3	0
Gen. 2	0.8	0.3	0.1
Gen. 3	0.1	0.4	0.9
Connectivity:	0.51	0.98	0.41

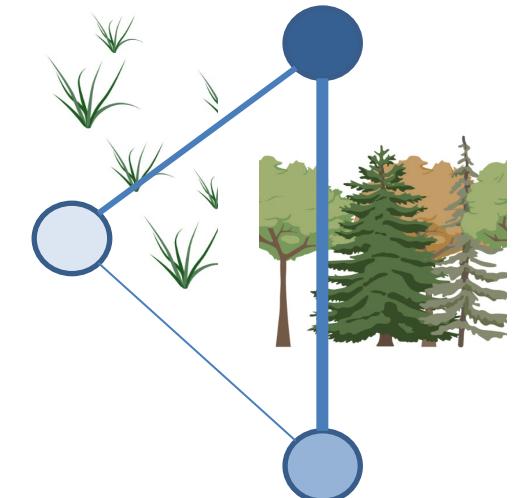
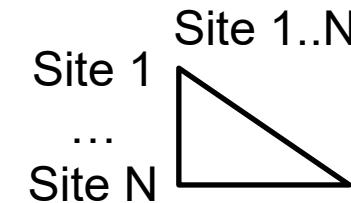
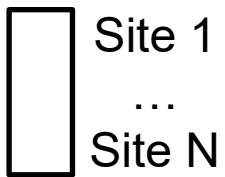
(SOME) BENEFITS

1. Scalable



(SOME) BENEFITS

2. Connectivity is an attribute of each entity (rather than a pairwise metric)



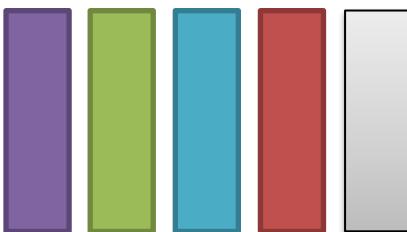
(SOME) BENEFITS

3. Makes interpolation more straightforward

Response:

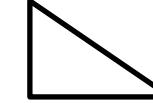


Predictors:

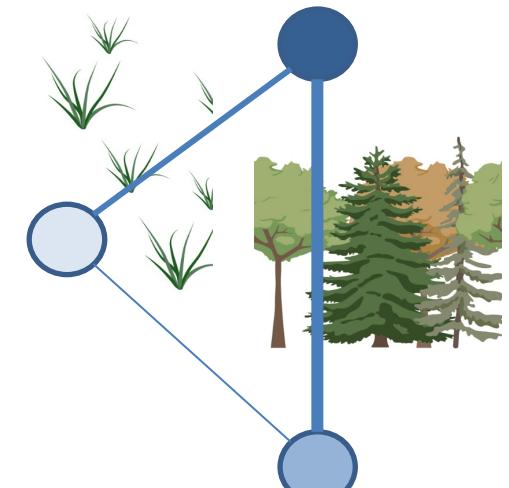
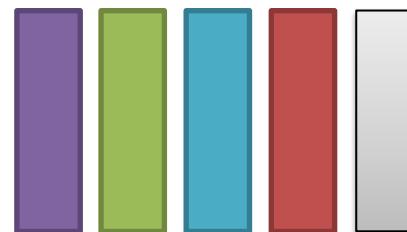


compared to:

Response:

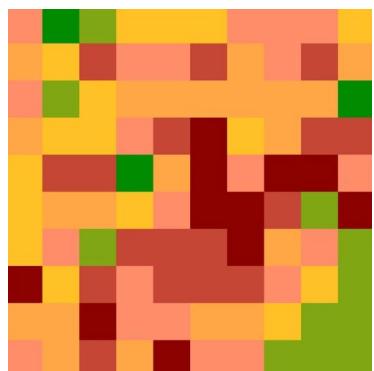


Predictors:



SIMULATED DATA

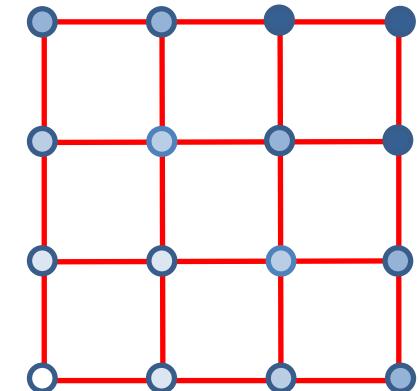
1. Generate landscape with heterogeneous carrying capacity
2. Simulate genotypes based on the generated landscapes
3. Calculate connectivity
4. Evaluate performance of model



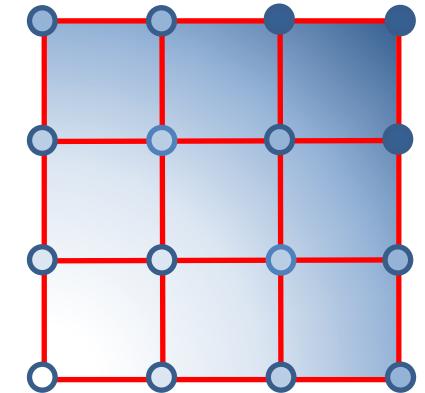
EMPIRICAL DATA

1. Calculate connectivity (discrete points)

a. *Information about population structure used to calculate population-scale connectivity*

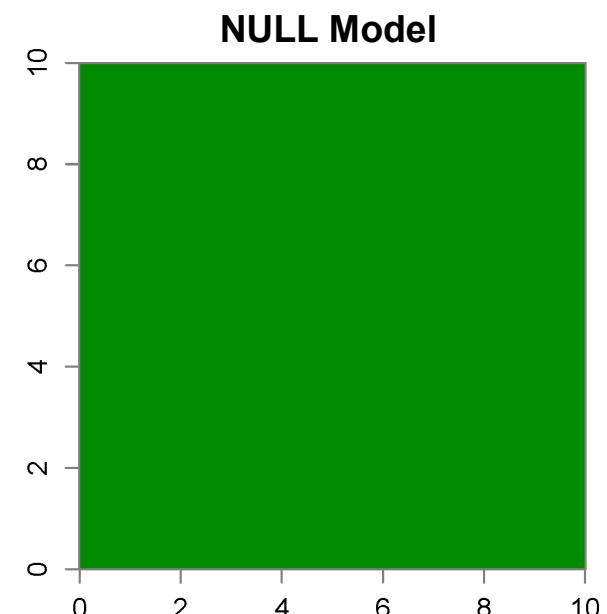
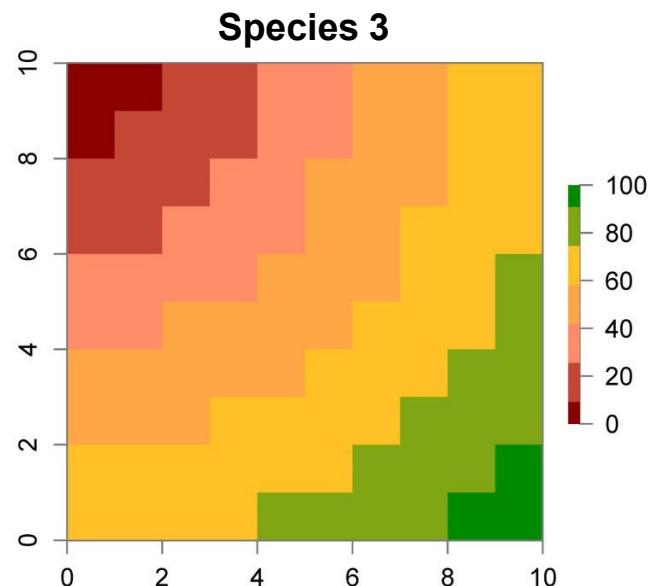
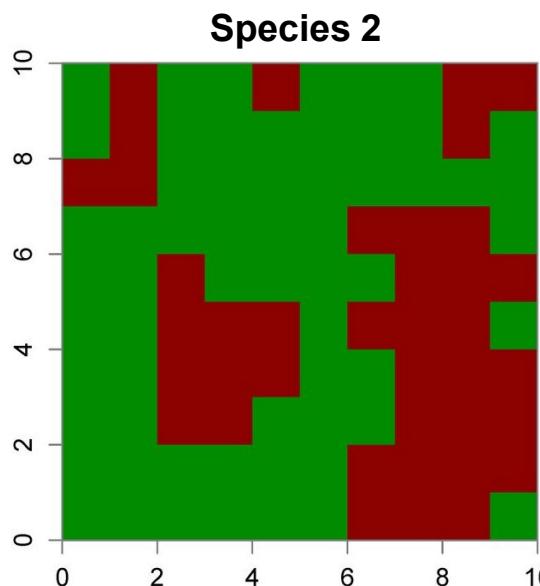
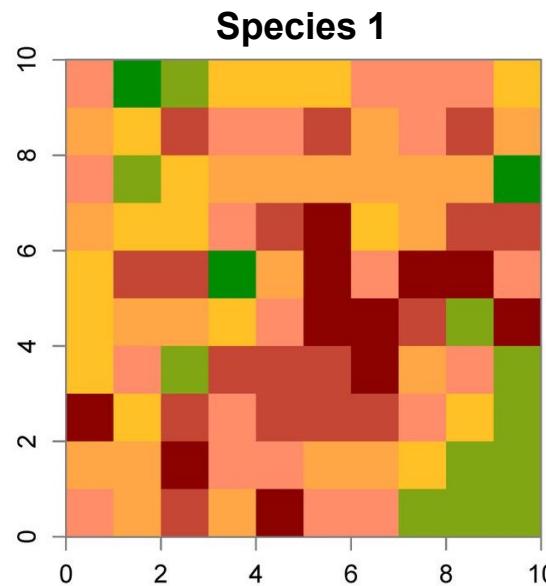


EMPIRICAL DATA

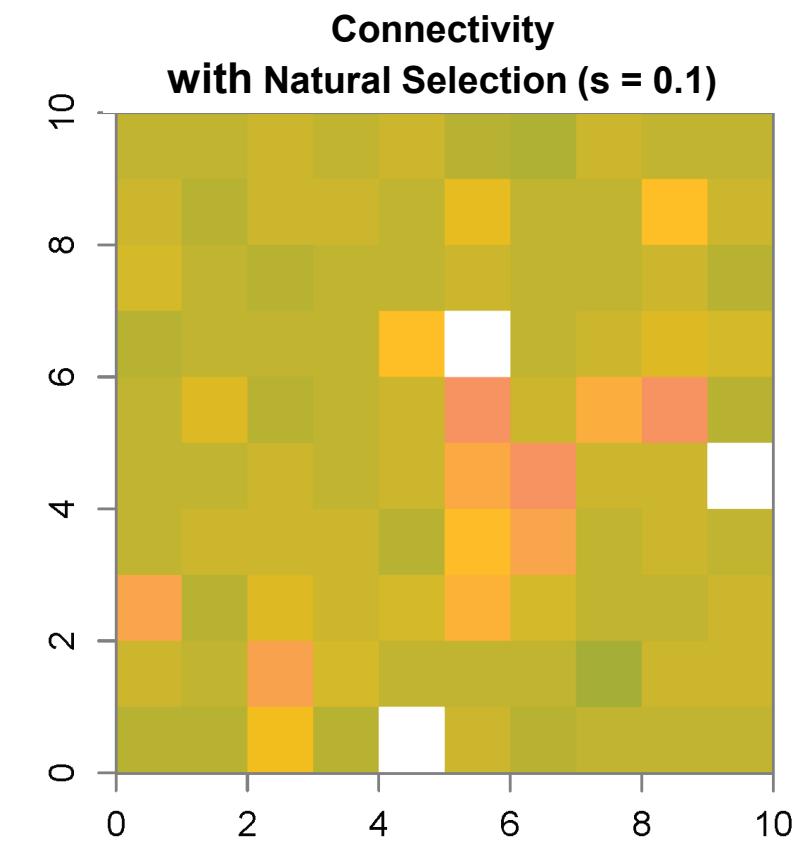
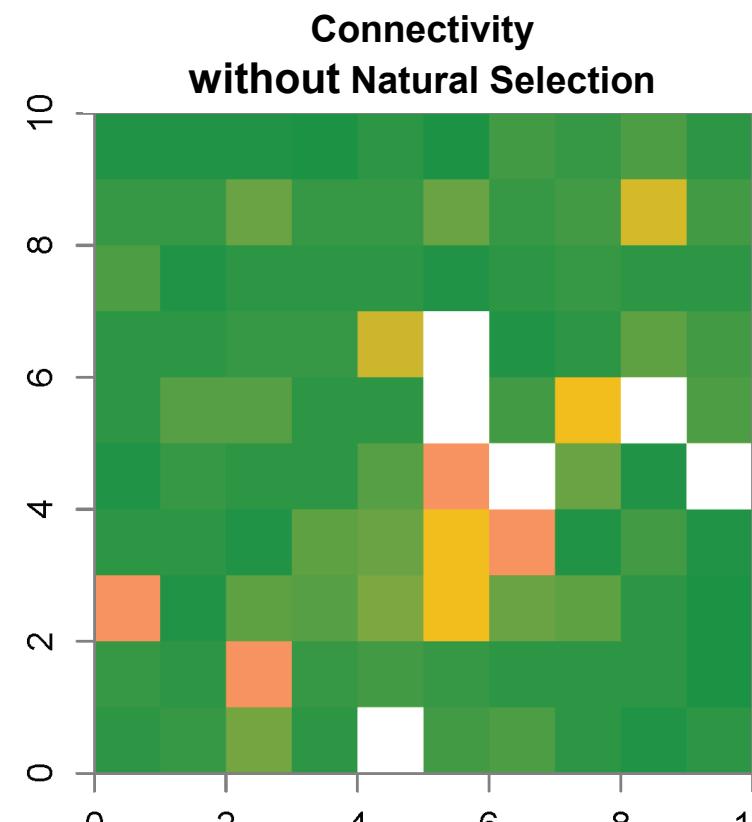
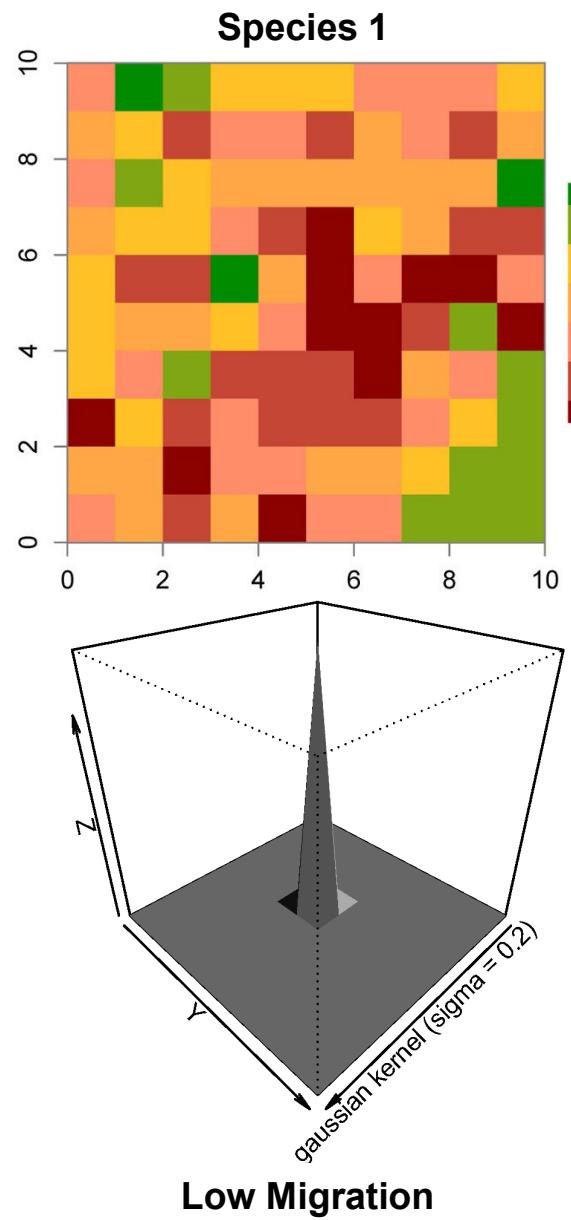


2. Go from discrete to continuous: model connectivity as a function of space and environment
 - a. *Response variable: connectivity among populations*
 - b. *Fixed effect: transformed environmental data (transformation maximizes covariance between genetic and environmental data)*
 - c. *Random effect: spatial covariance function*

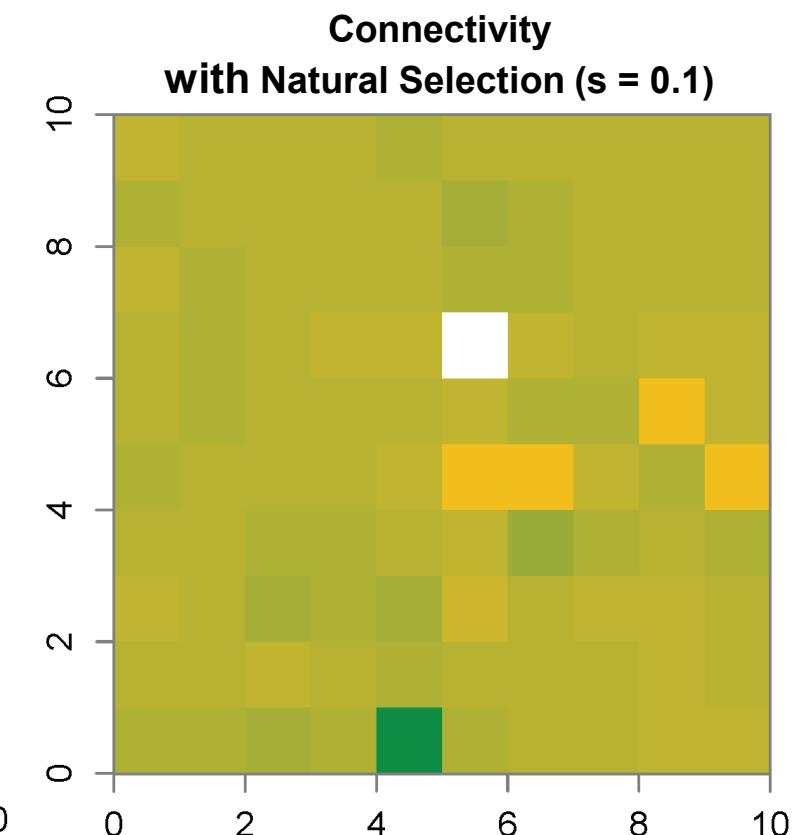
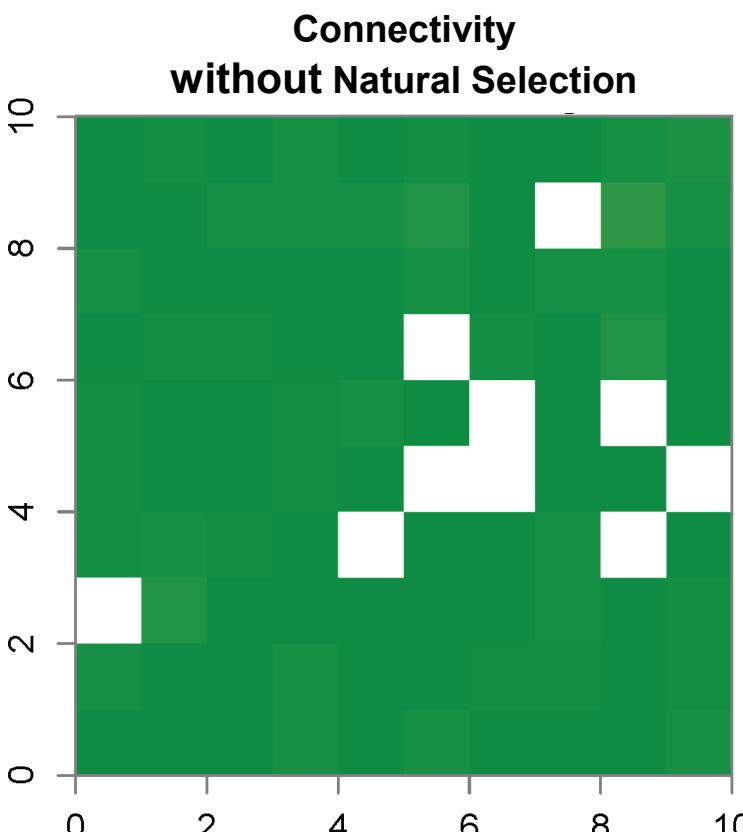
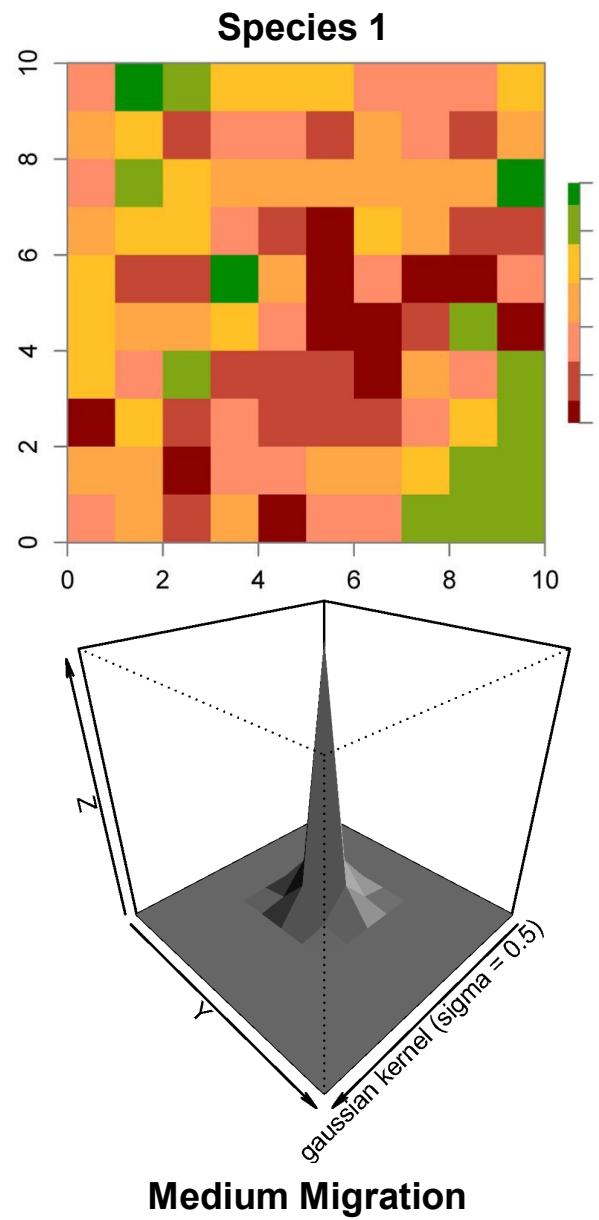
Landscape = Carrying Capacity for Different Species



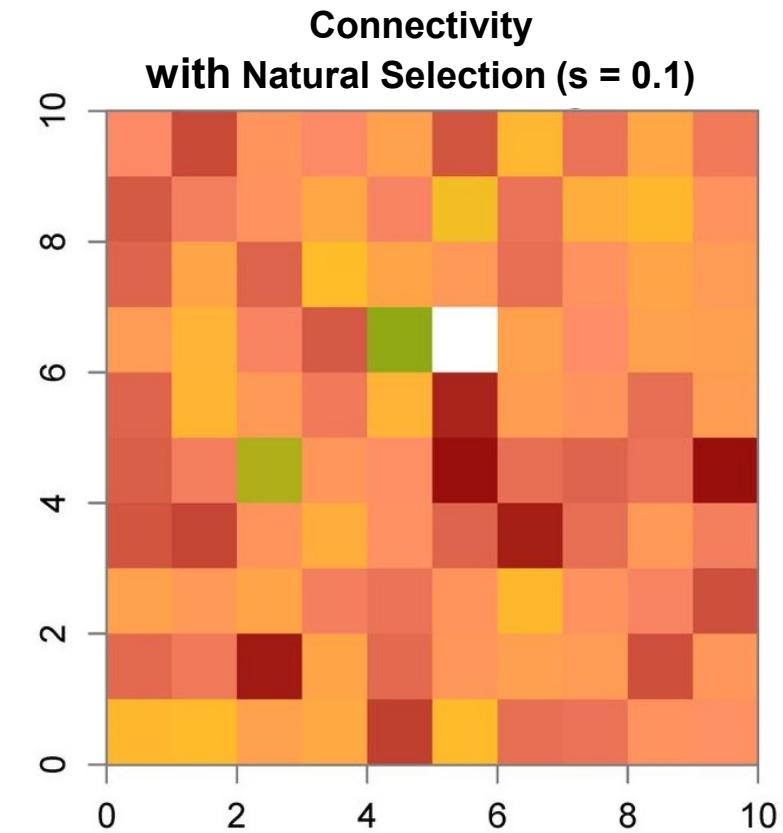
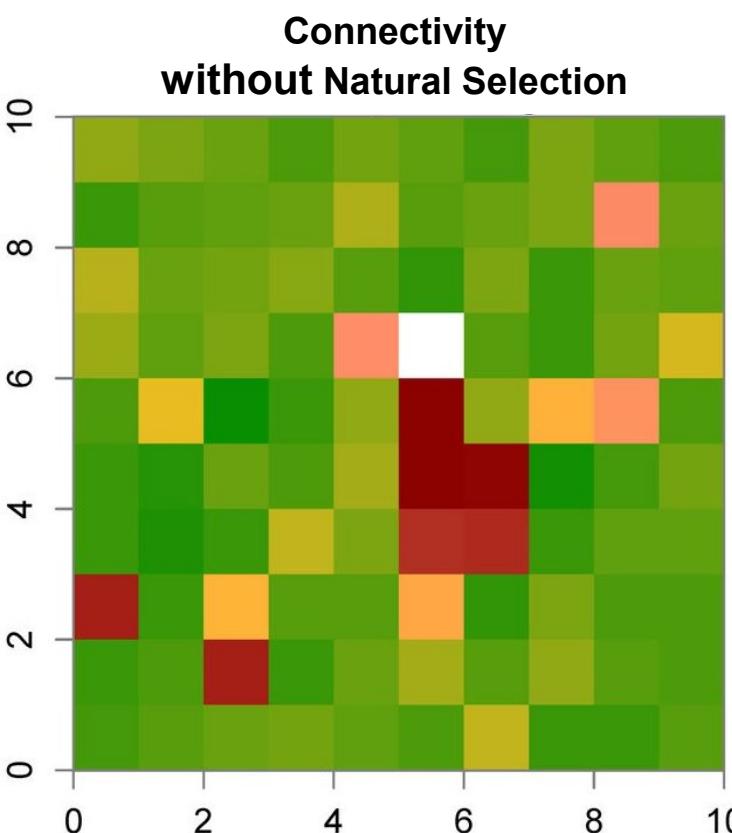
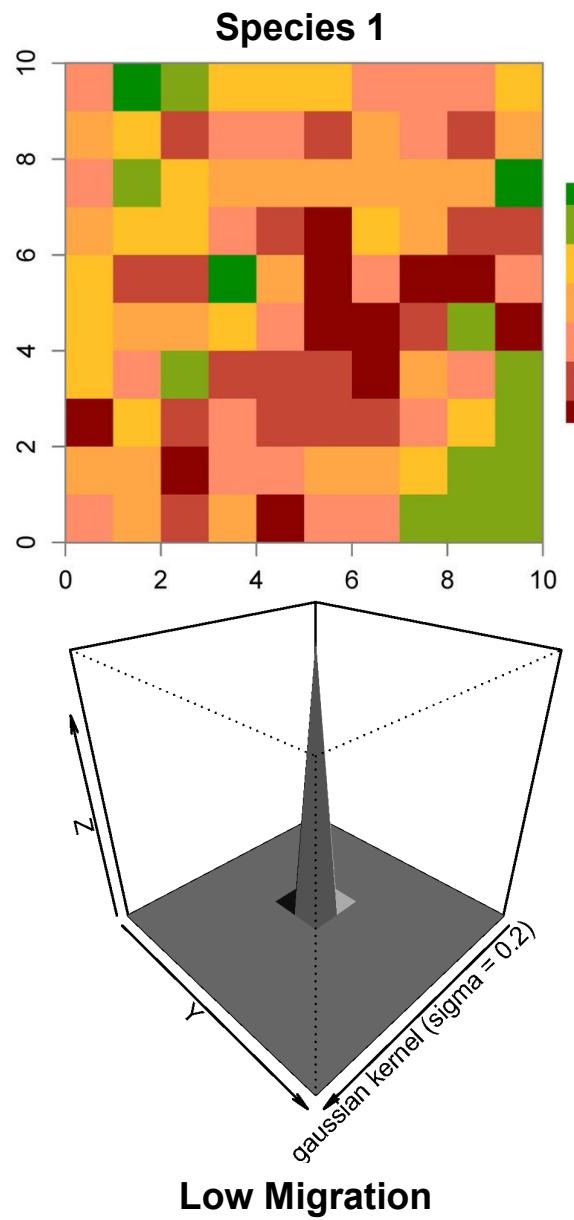
Connectivity among Individuals



Connectivity among Individuals

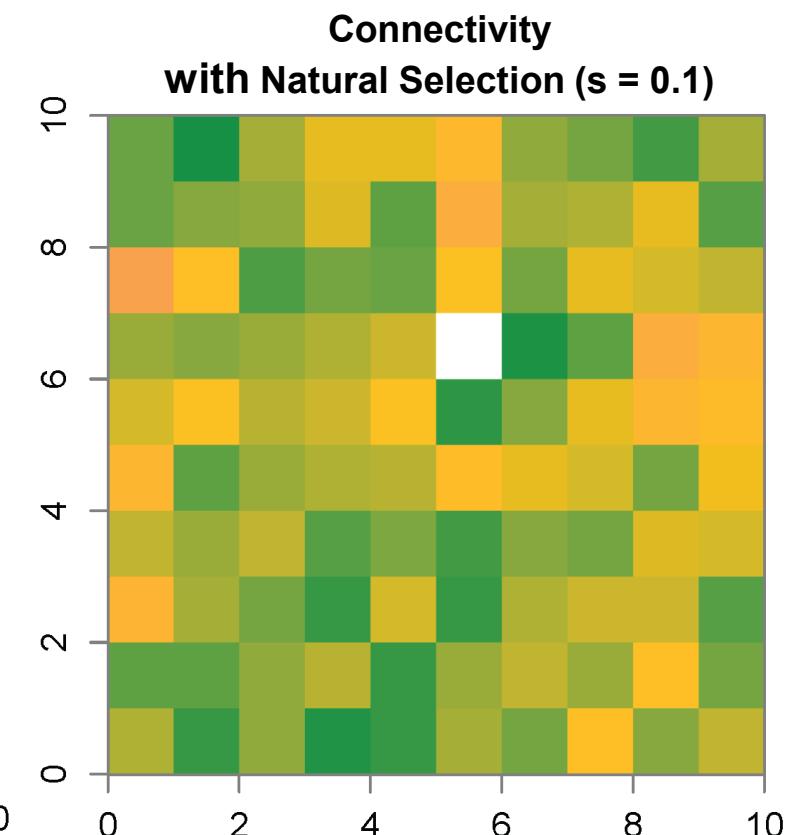
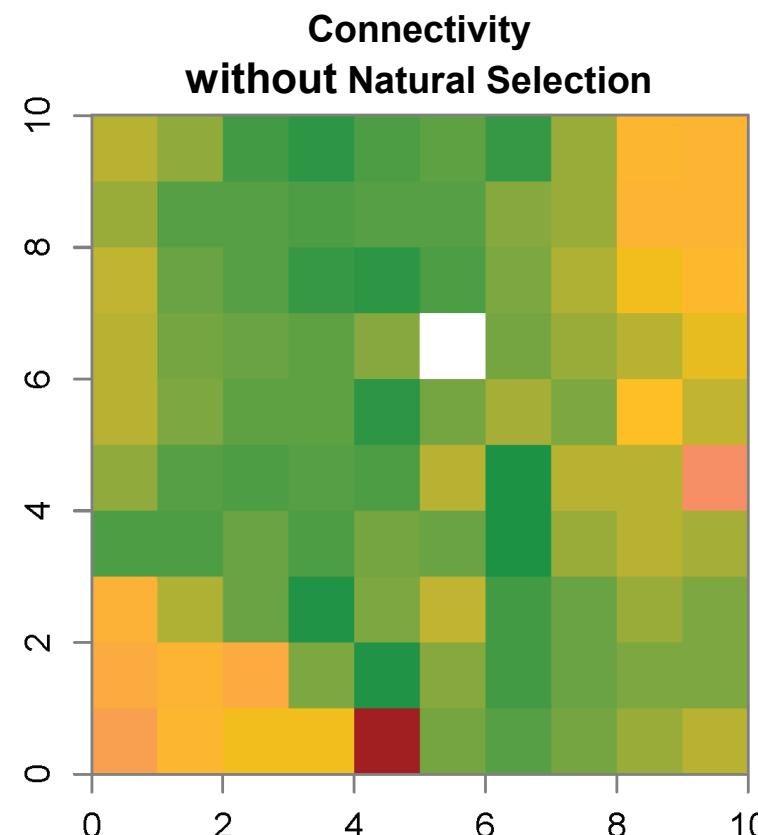
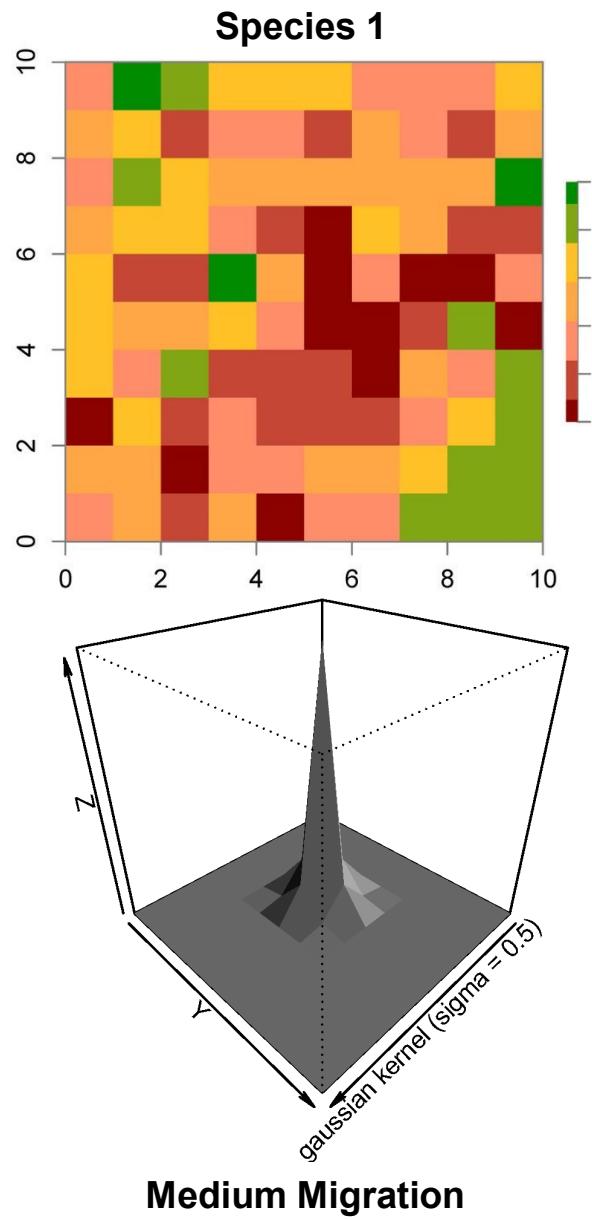


Connectivity among Populations

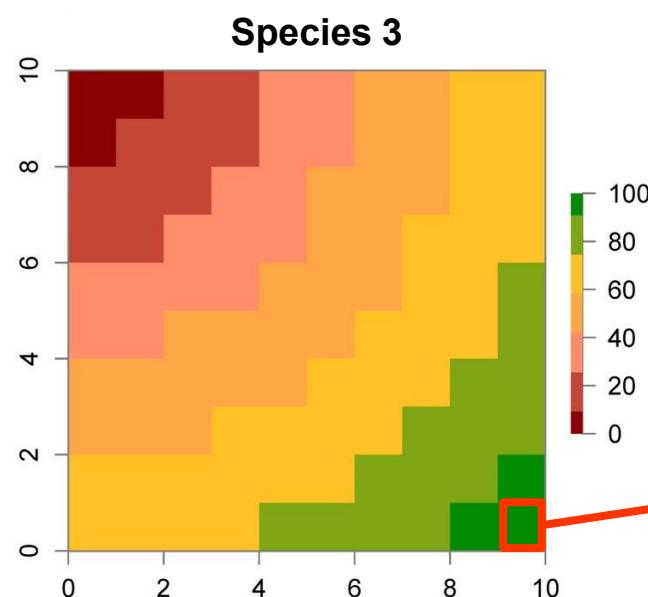
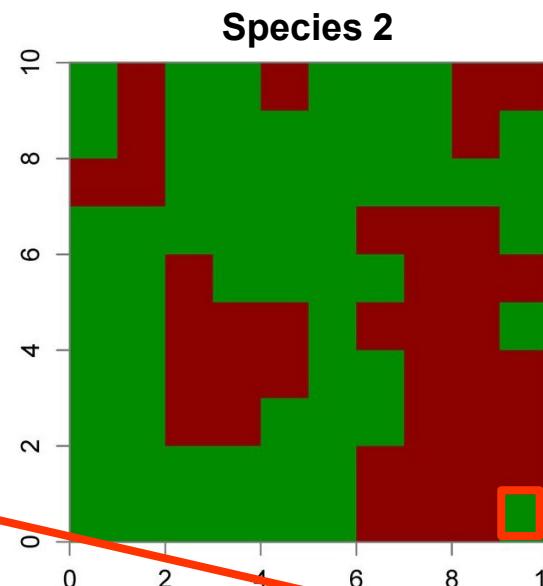
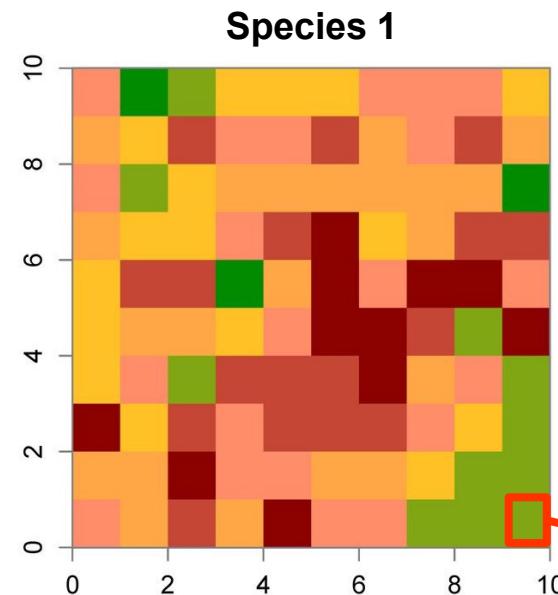


1.0
0.8
0.6
0.4
0.2

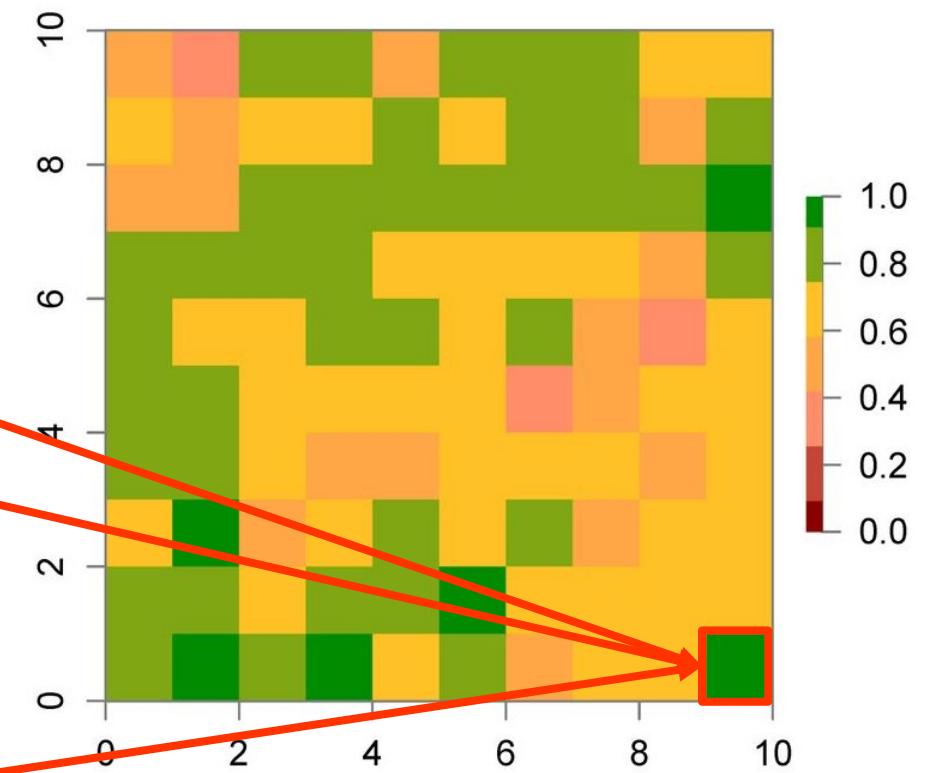
Connectivity among Populations



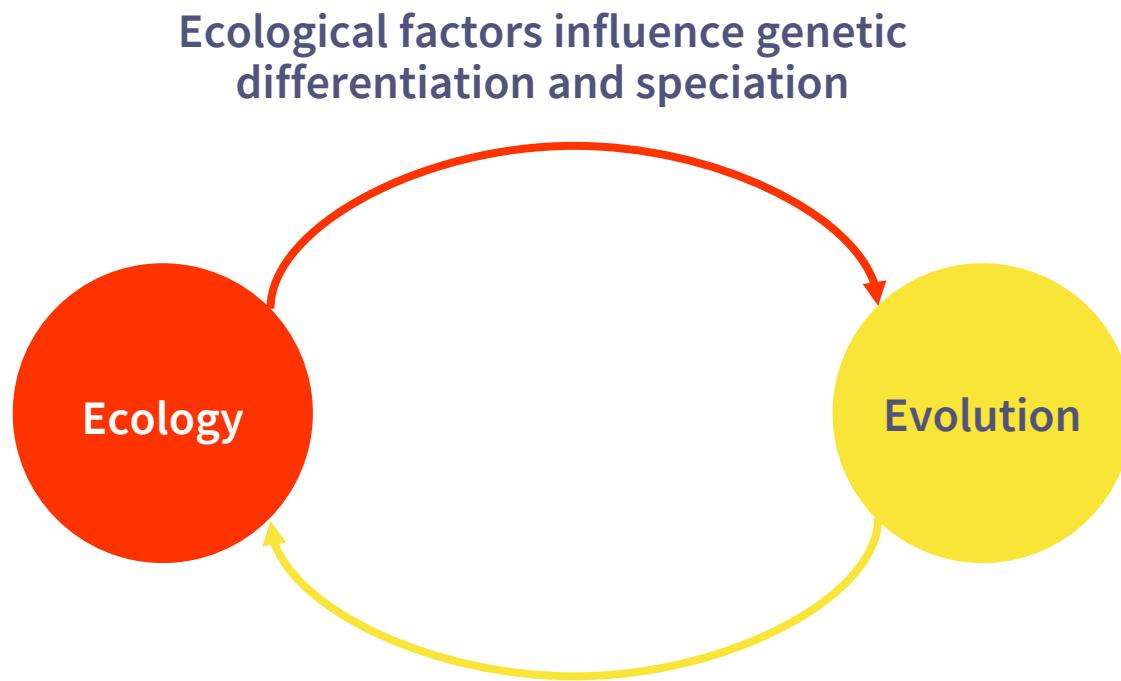
Connectivity among Species



Connectivity is highest
where the three species are found in equal proportions



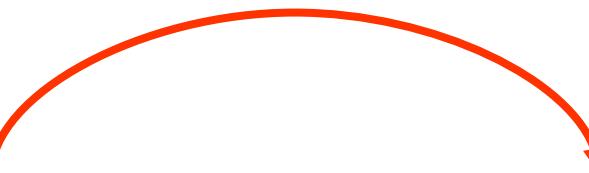
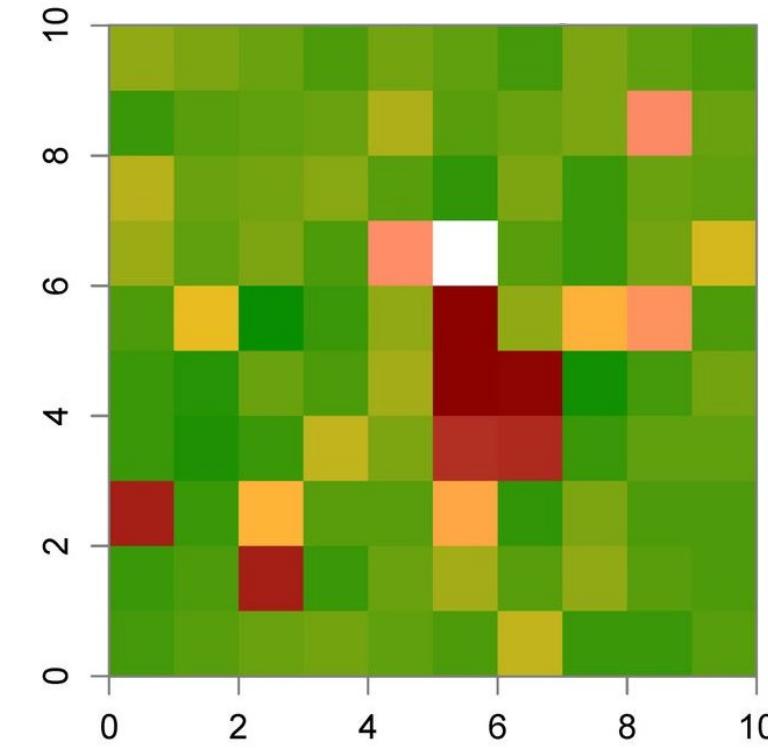
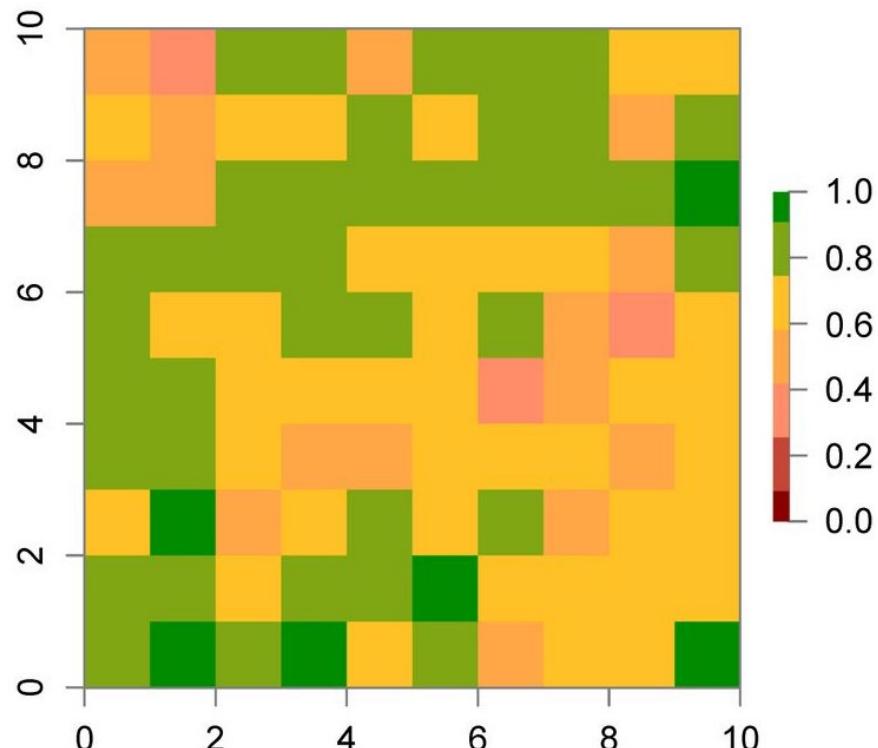
Reciprocal Impact: Eco-Evo



Ecological factors influence genetic differentiation and speciation

Evolutionary change acts on species interactions, community structure, and ecosystem function

Scalability of Metric → Can be Integrated in Eco-Evo Framework



E.g.: Connectivity among Aquatic Invertebrates

AQUATIC INVERTEBRATES

IN STILL AND FLOWING WATERS

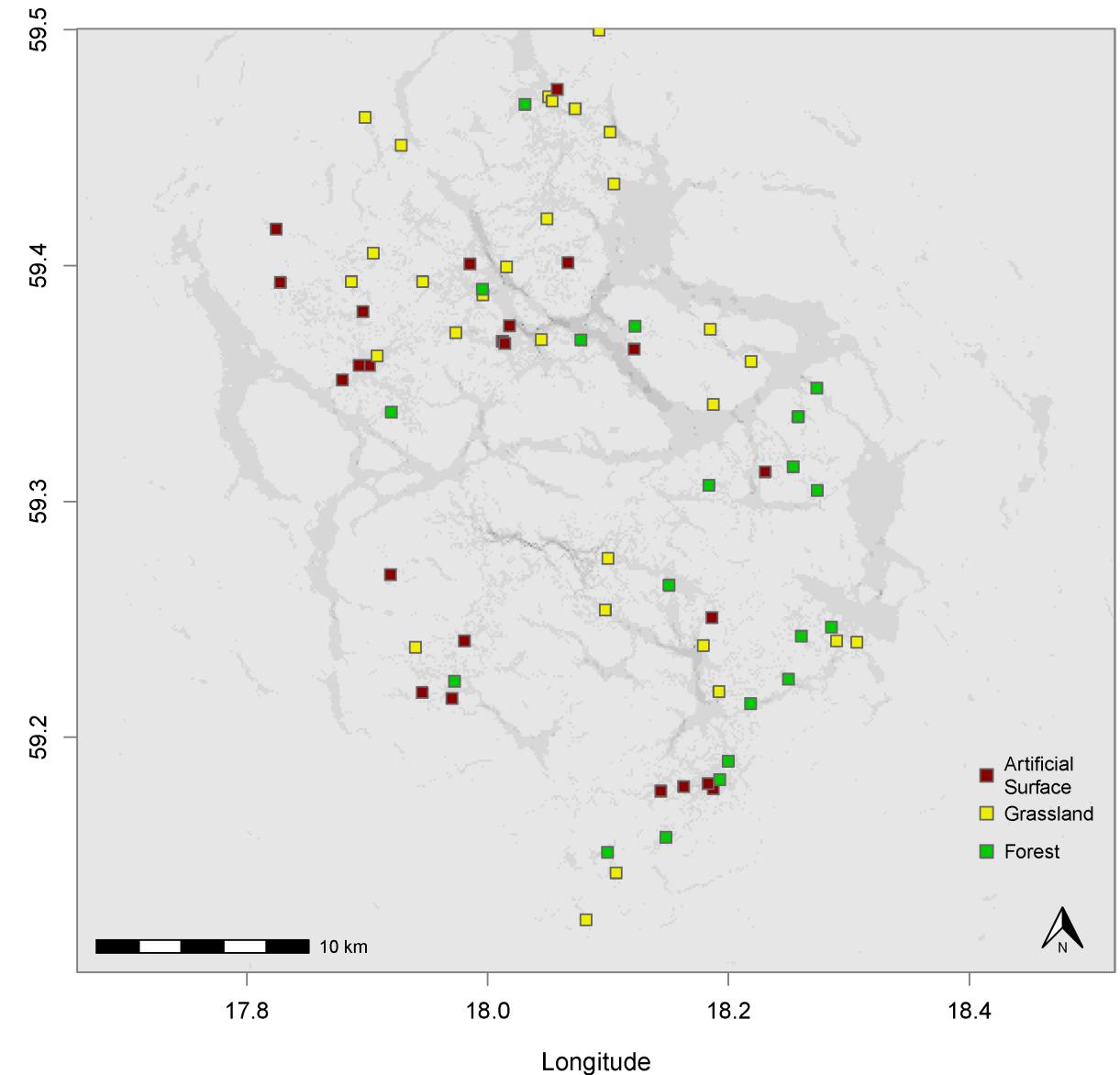
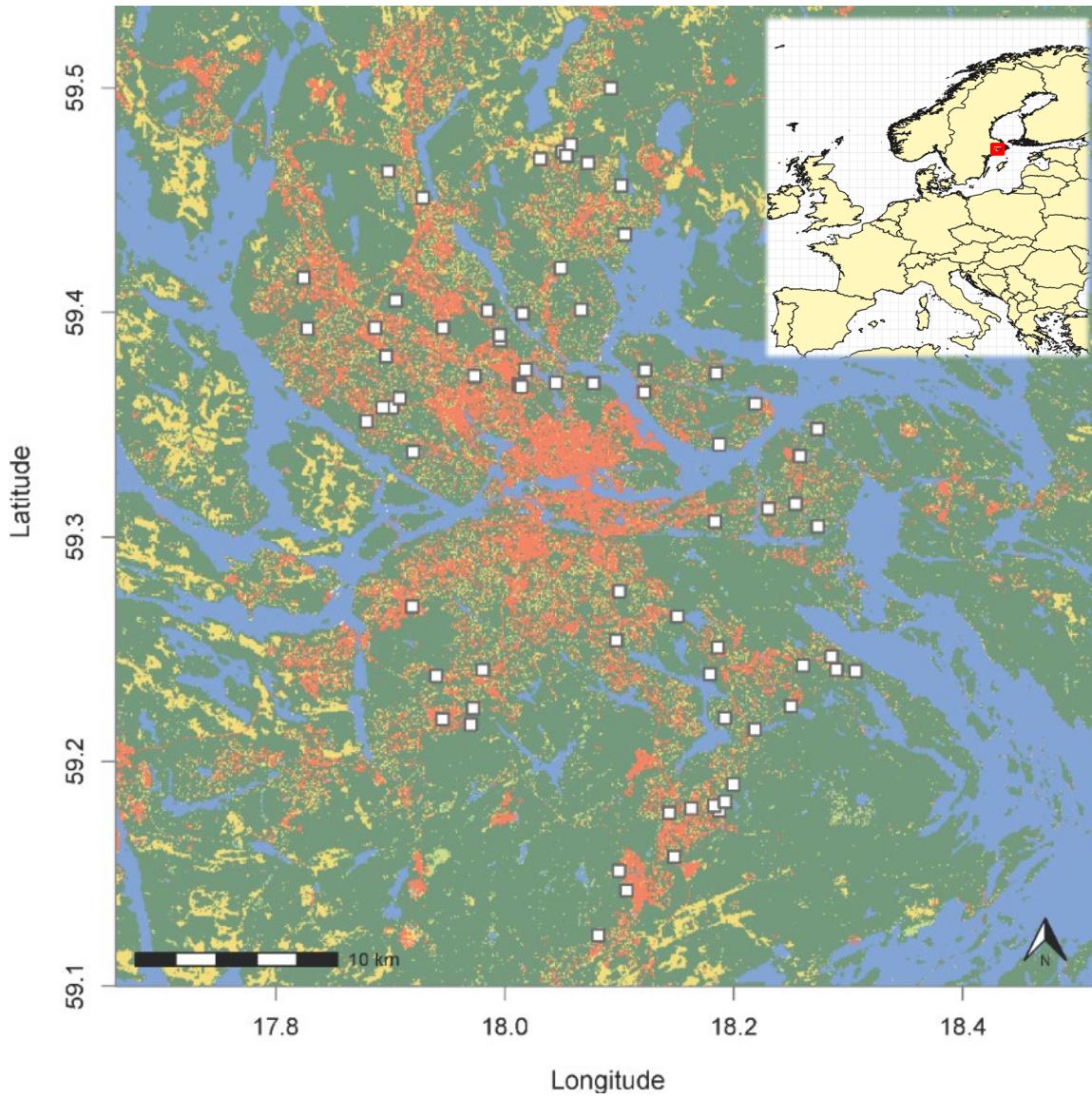


Coleoptera
(beetles)
Hemiptera
(true bugs)
Odonata

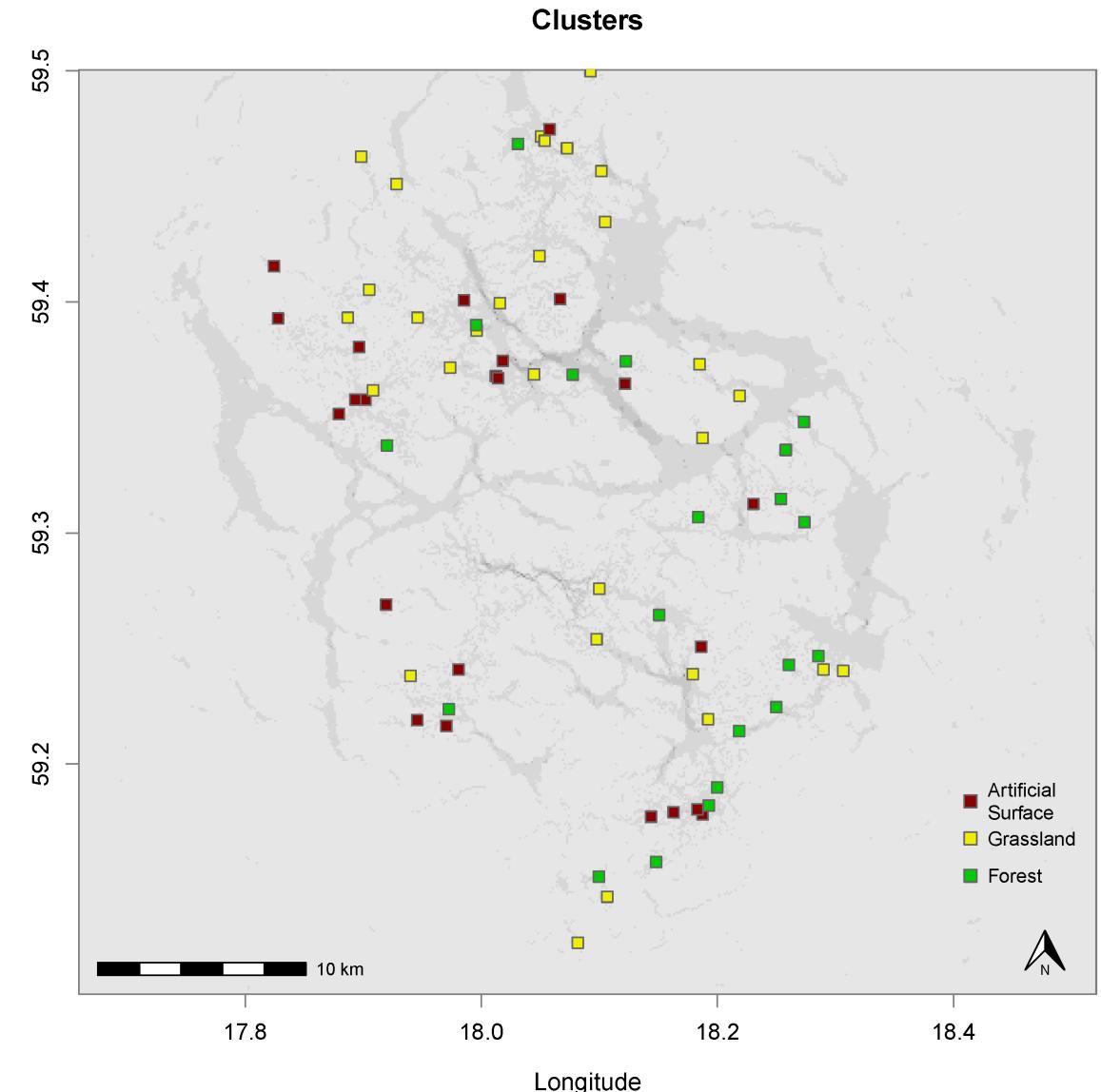
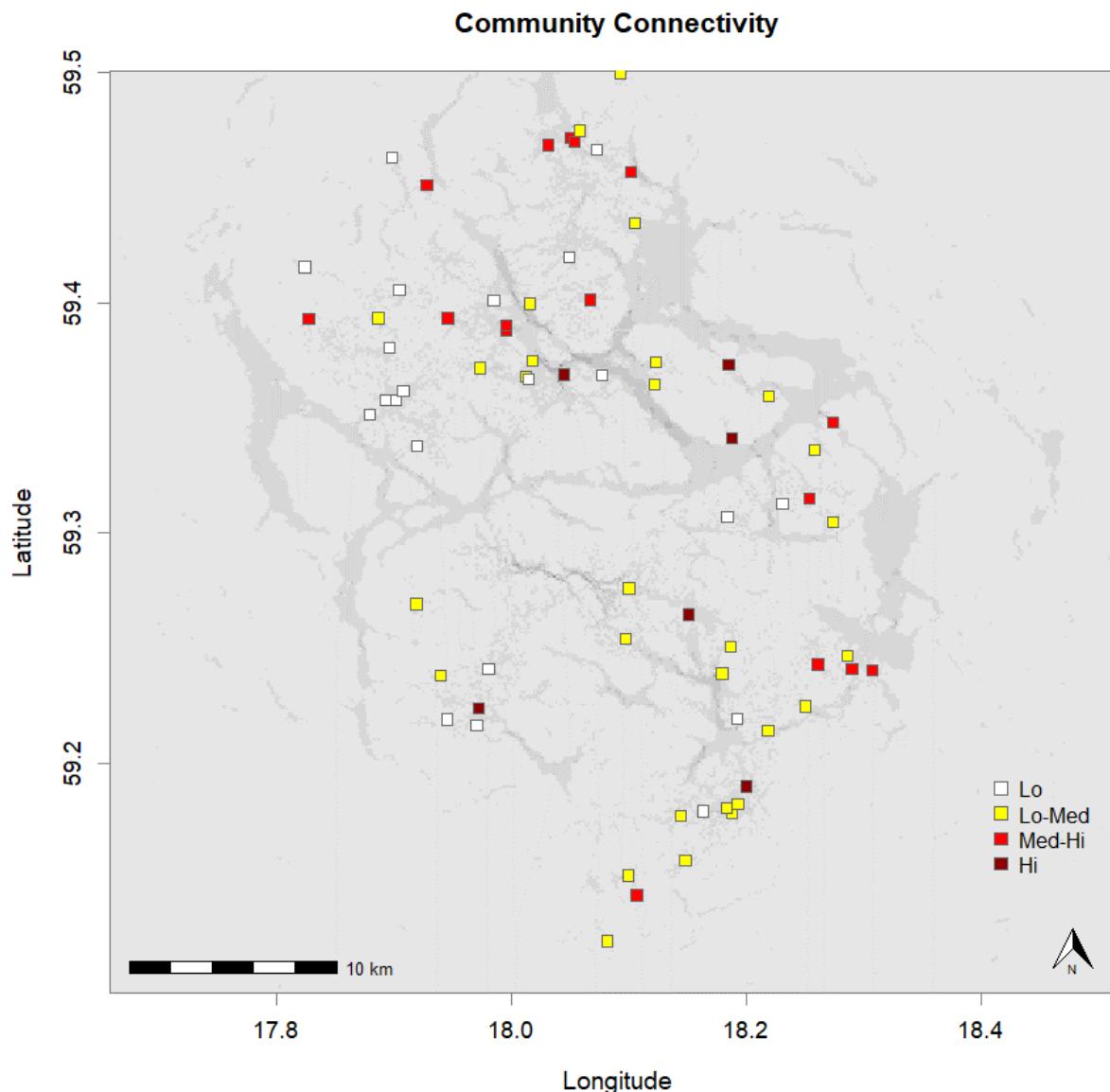
(damsel- and dragonflies)
Trichoptera
(caddisflies)

Gastropoda
(freshwater snails)

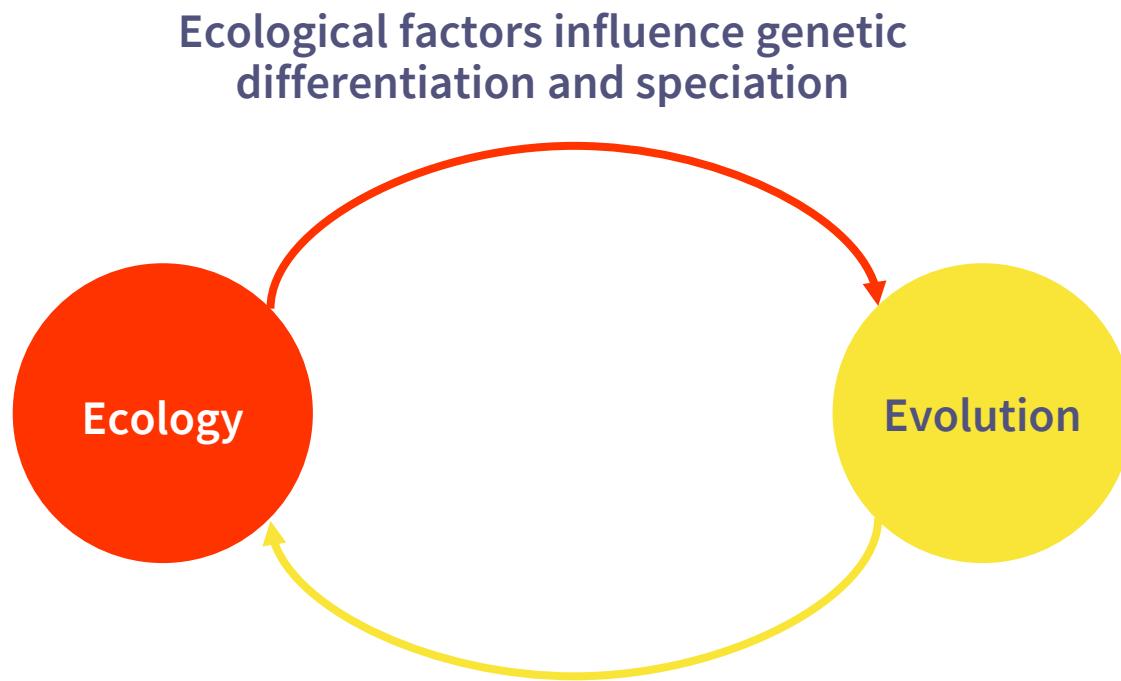
E.g.: Connectivity among Aquatic Invertebrates



E.g.: Connectivity among Aquatic Invertebrates



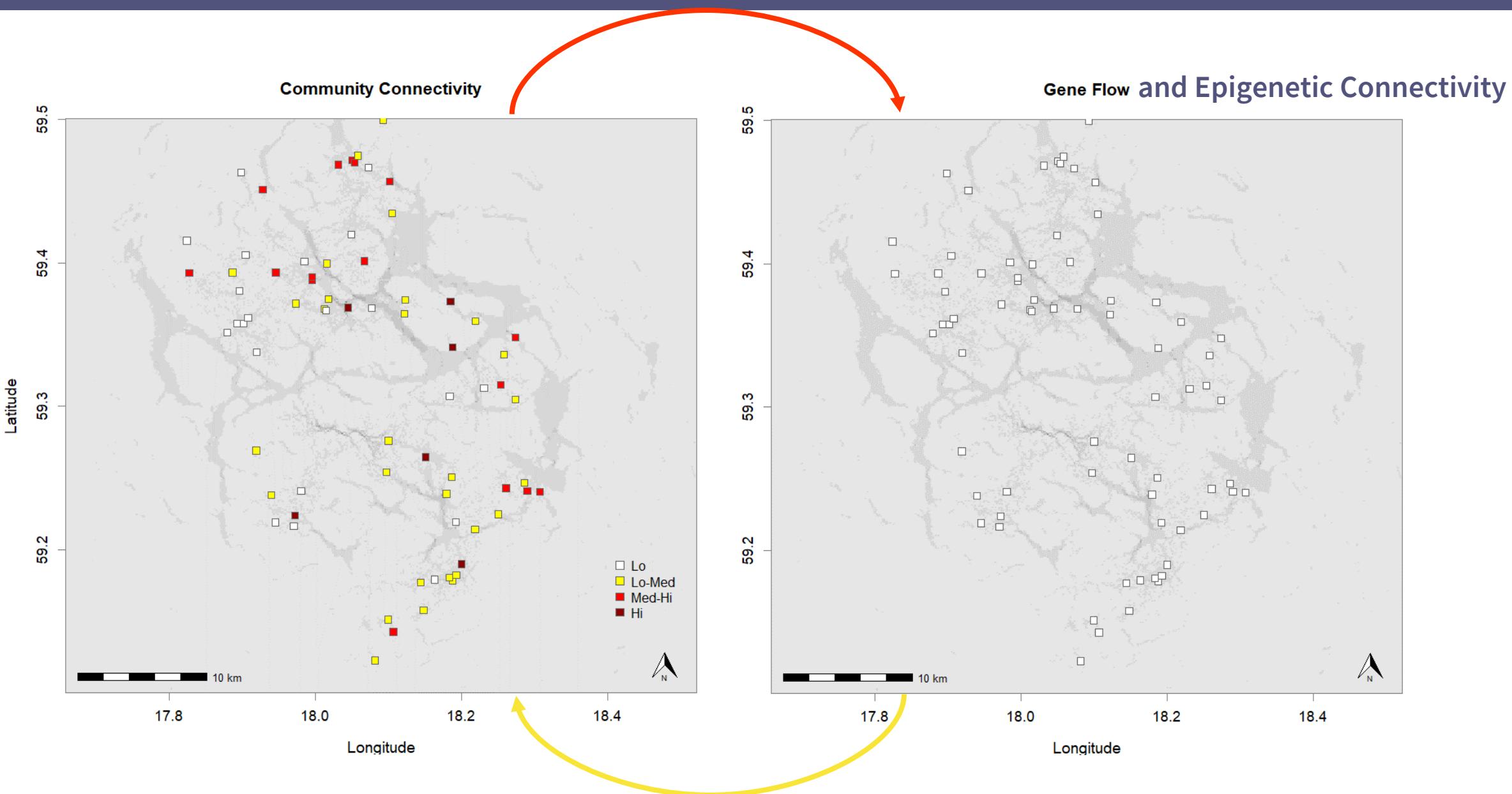
Reciprocal Impact: Eco-Evo



Ecological factors influence genetic differentiation and speciation

Evolutionary change acts on species interactions, community structure, and ecosystem function

Community connectivity \leftrightarrow Within-species connectivity



E.g.: Tsetse = Vectors of African Trypanosomiasis



Tsetse (*Glossina fuscipes*) in Uganda

Motivation

Finding out how populations of tsetse are connected, which has practical (epidemiological) implications

Goal

Model continuous gene flow (i.e., CONNECTIVITY)

Unconstrained (neutral)

Constrained (by selection; adaptive)

Aims

Calculate metric in sampled locations

Interpolate metric in unsampled locations

Modeling Gene Flow

Response variable: Gene flow

Fixed effects: Environmental variables

Random effect: Spatial locations

Gene Flow at Discrete Points

Neutral flow: $F_p = \frac{(\sum_{k=1}^n N_{pk})^2}{\sum_{k=1}^n N_{pk}^2}; F_{ip} = F_p$

Interpolation: $F_i \sim \beta_{0_i} + \beta_{x_i} x_i + \beta_{y_i} y_i + \beta_{ij}^T C I_{ij} + f(s_i)$

Presence/Absence Selected Alleles: $A_i | \pi_i \sim Binom(n_i, \pi_i)$

Interpolated Log-Odds: $logit(\pi_i) \sim \beta_{0_i} + \beta_{x_i} x_i + \beta_{y_i} y_i + \beta_{ij}^T C I_{ij} + f(s_i)$

Adaptive flow: $AF_i = \frac{(\sum_{j=1}^n A_{ij})^F}{\sum_{j=1}^n A_{ij}^F}$

Continuous Gene Flow

Neutral flow: $F_p = \frac{(\sum_{k=1}^n N_{pk})^2}{\sum_{k=1}^n N_{pk}^2}; F_{ip} = F_p$

Interpolation: $F_i \sim \beta_{0_i} + \beta_{x_i} x_i + \beta_{y_i} y_i + \beta_{ij}^T C I_{ij} + f(s_i)$

Presence/Absence Selected Alleles: $A_i | \pi_i \sim \text{Binom}(n_i, \pi_i)$

Interpolated Log-Odds: $\text{logit}(\pi_i) \sim \beta_{0_i} + \beta_{x_i} x_i + \beta_{y_i} y_i + \beta_{ij}^T C I_{ij} + f(s_i)$

Adaptive flow: $AF_i = \frac{(\sum_{j=1}^n A_{ij})^F}{\sum_{j=1}^n A_{ij}^F}$

(Bayesian) Interpolation

Skipping complex mathematical details, including:

Covariance function to model spatial dependence (discrete space)

Continuous-space approximation (using stochastic partial differential equations)

Interpolation

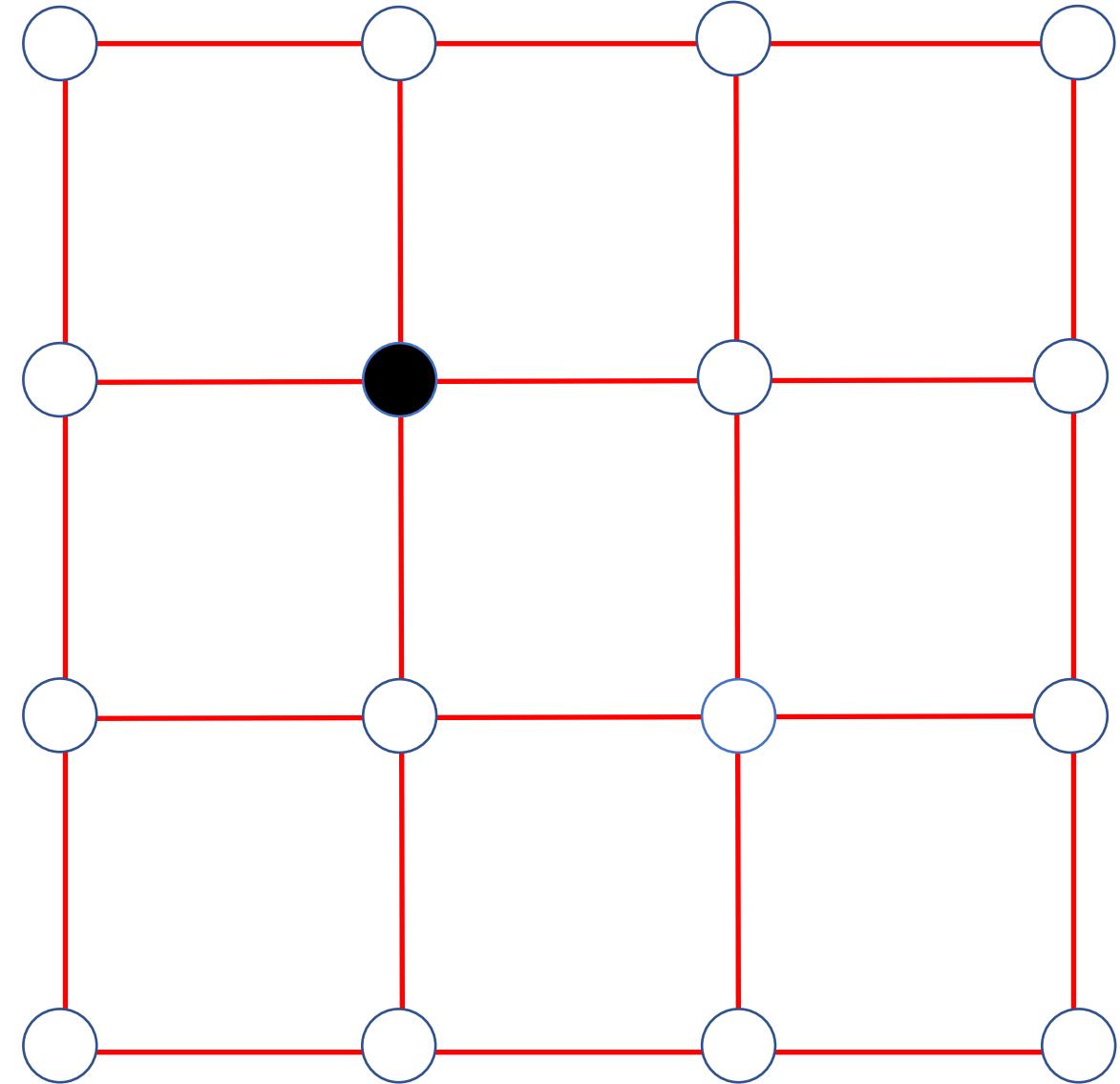
Finding neighboring locations



Sampled locations



Unsampled locations

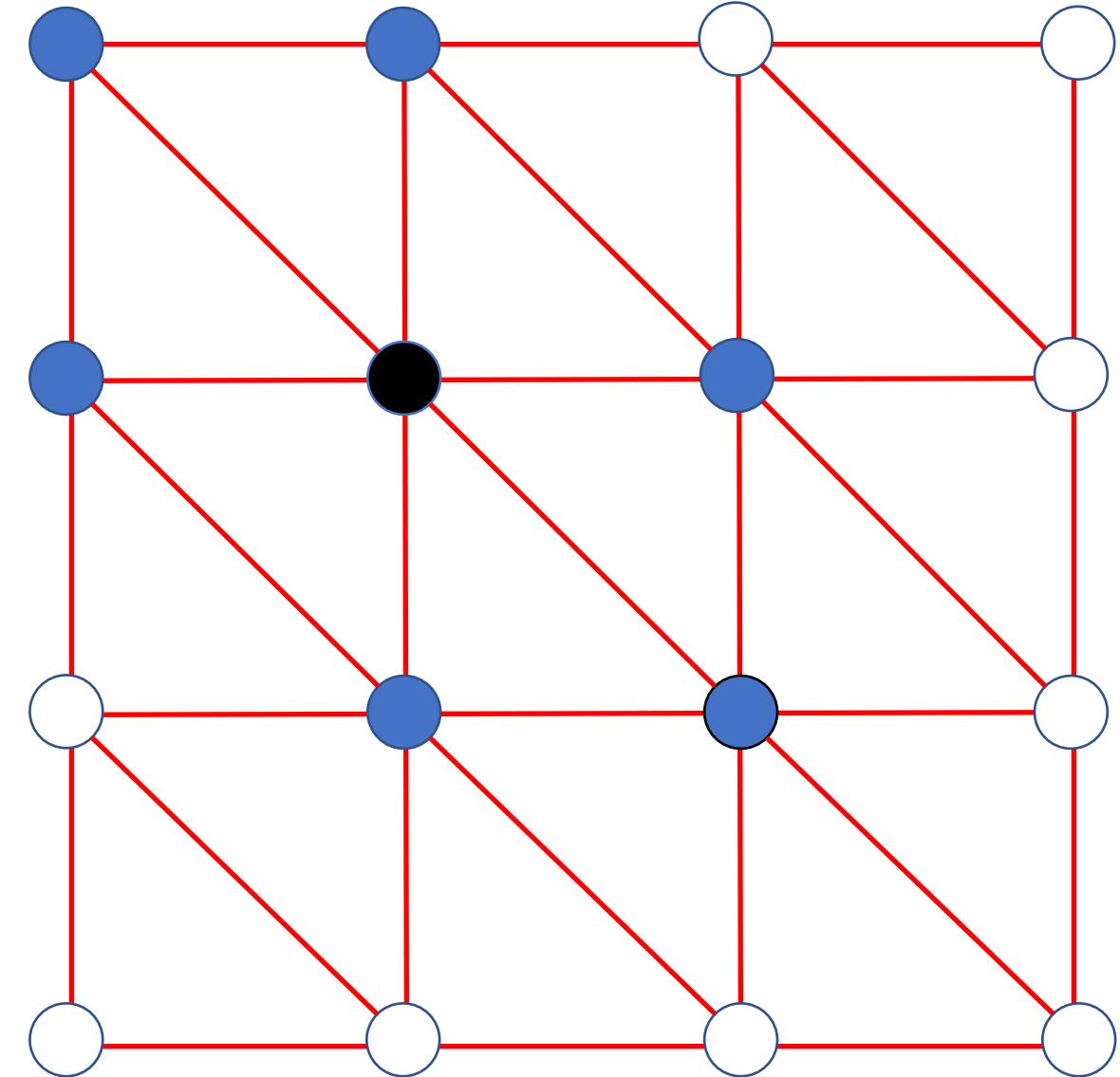


Interpolation

Mesh (triangulation)

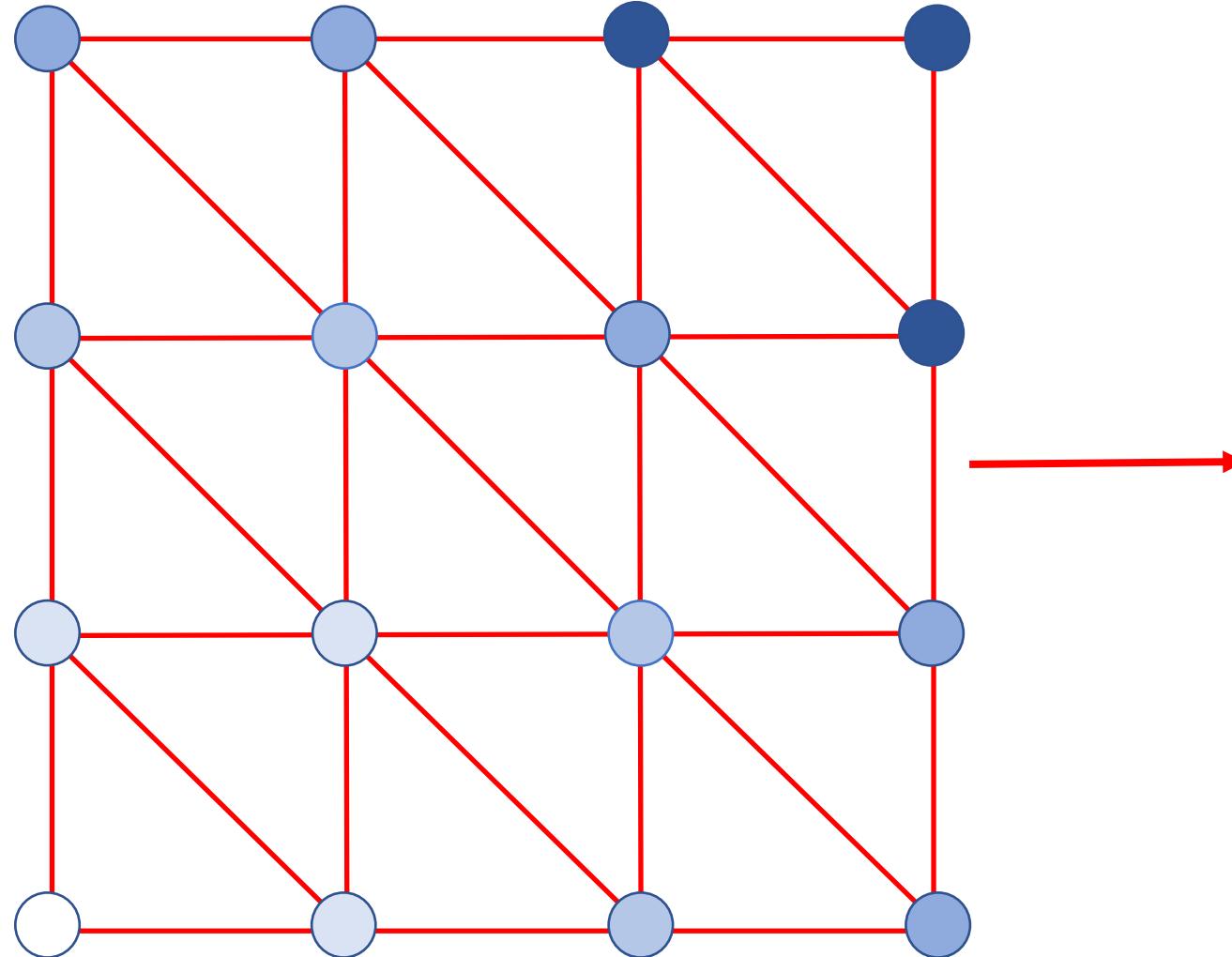
● Sampled locations

● Neighboring locations

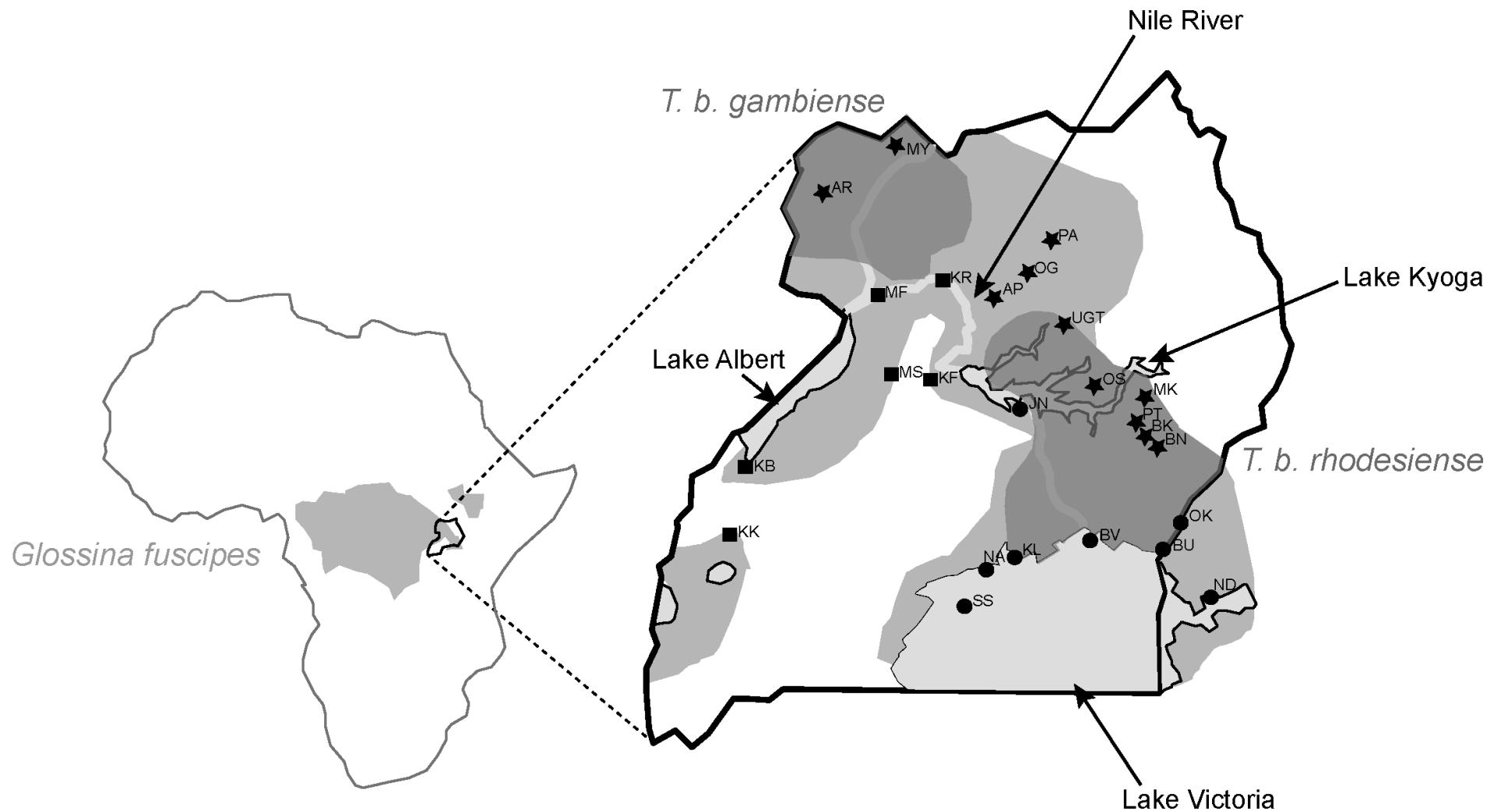


Interpolation

From Discrete to Continuous Gene Flow

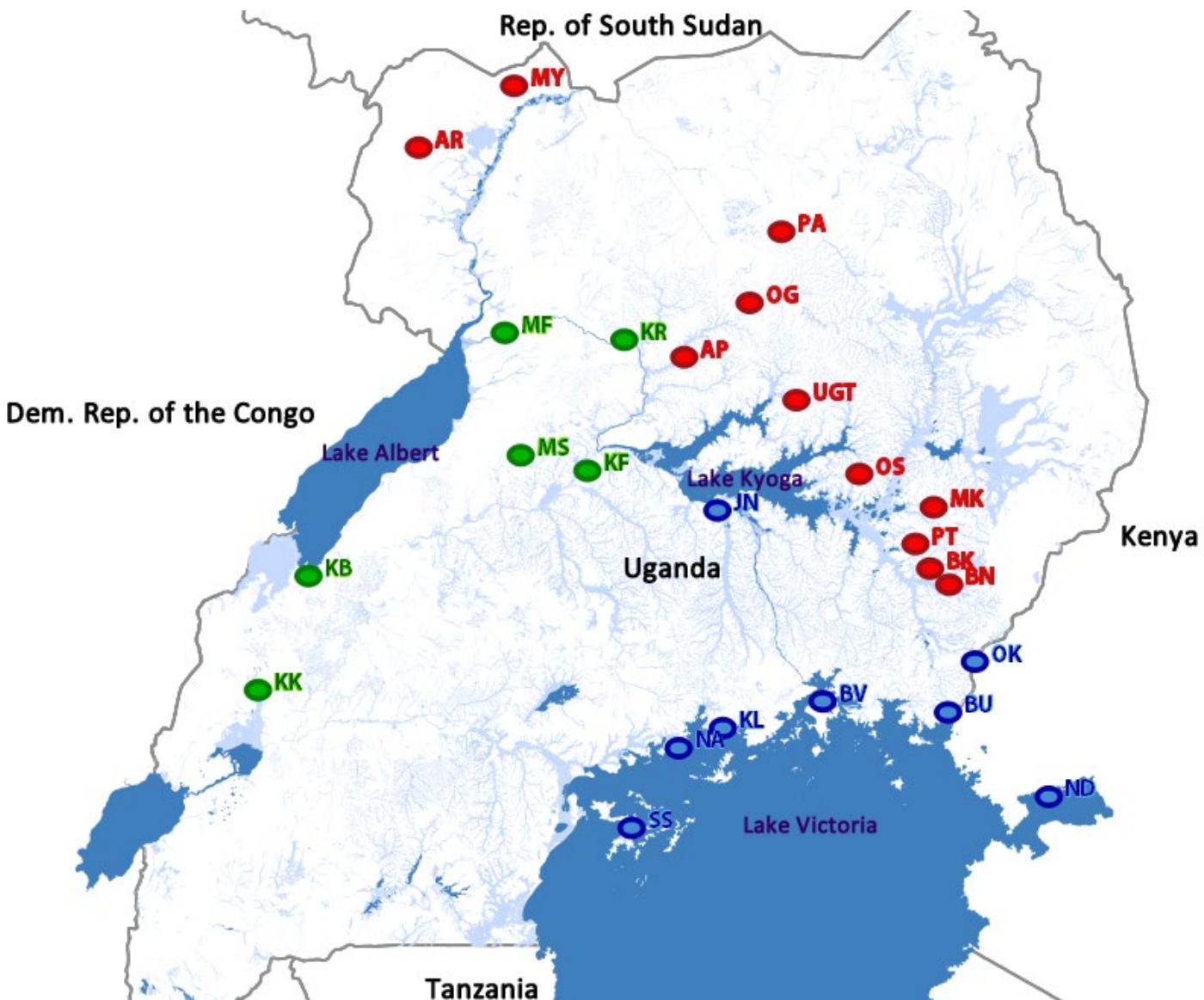


Sampling Sites



Tsetse in Uganda: 25 sites

Genetic Populations

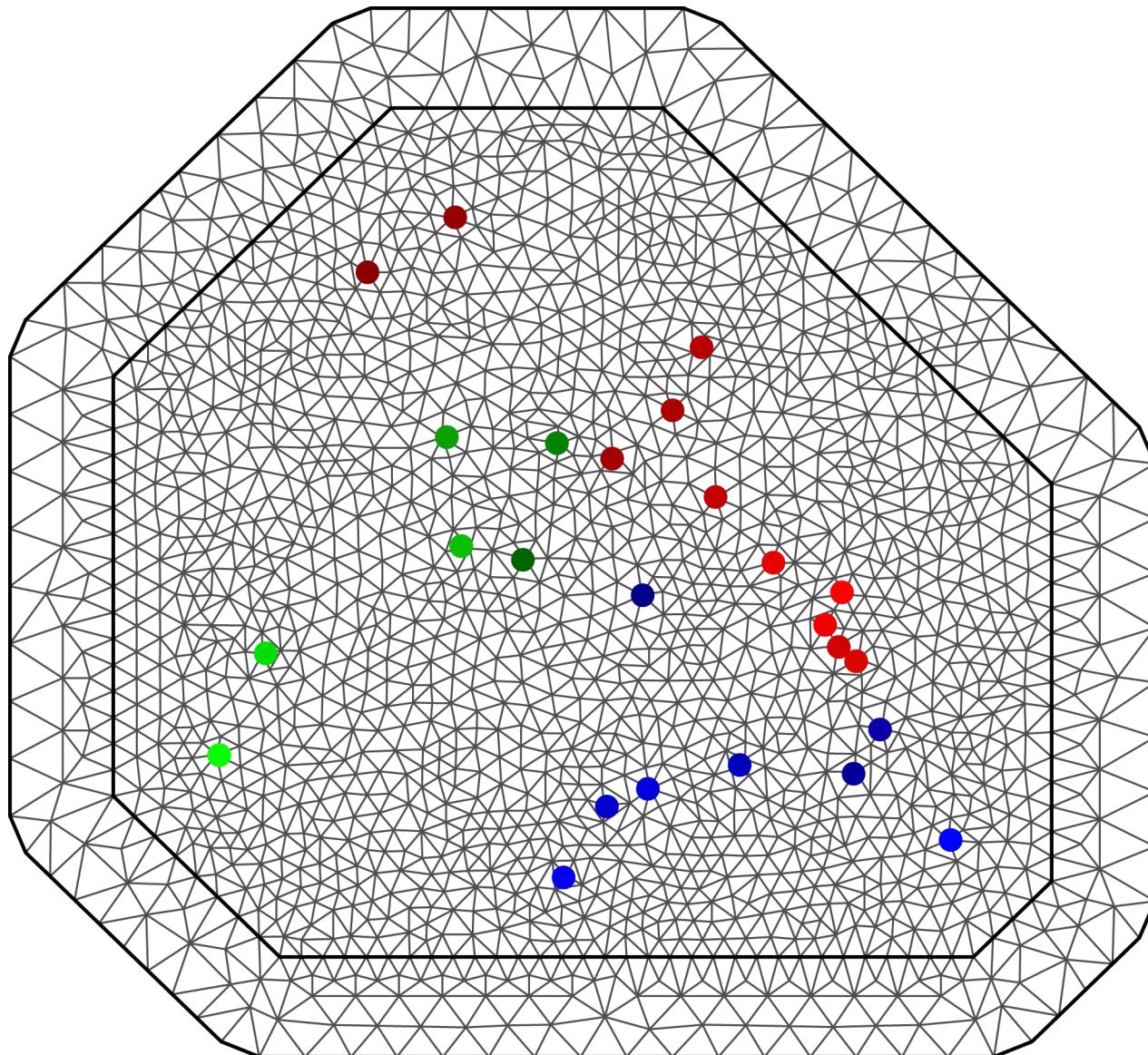


11 microsatellite loci

965 individuals, 25 sites, 3 populations (9 sub-populations)

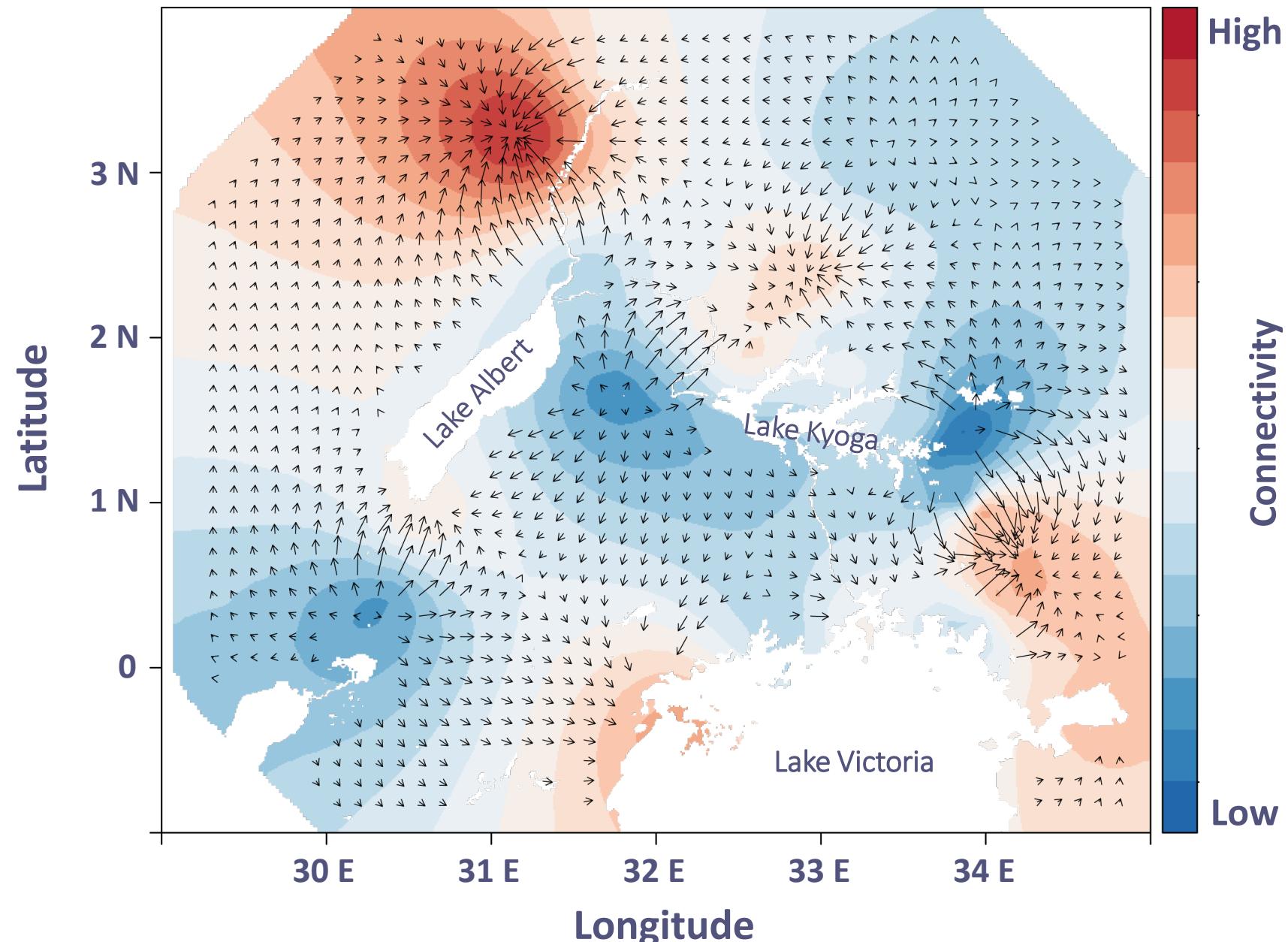
Mesh

Going from
D I S C R E T E
to
CONTINUOUS

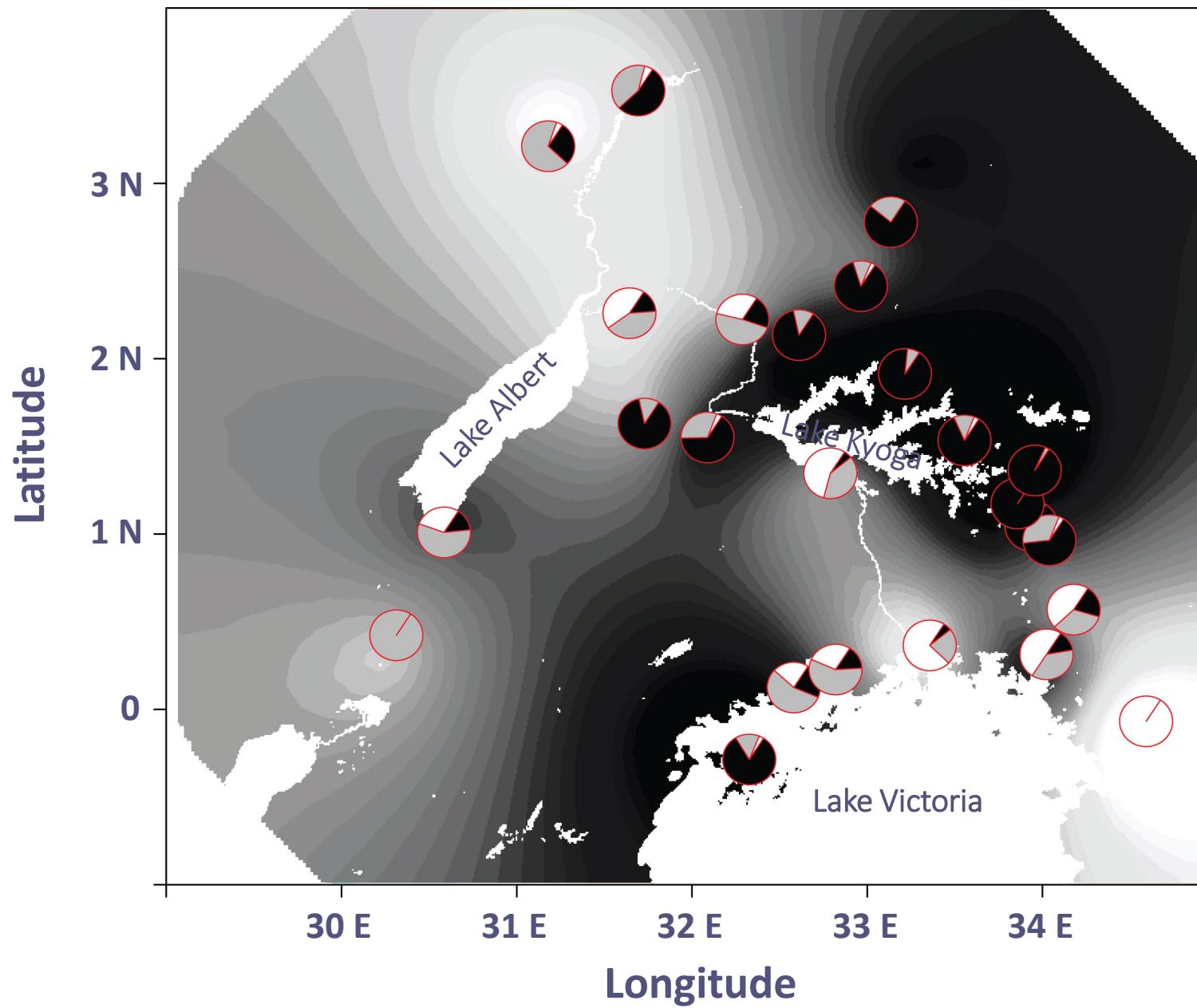


Neutral Gene Flow

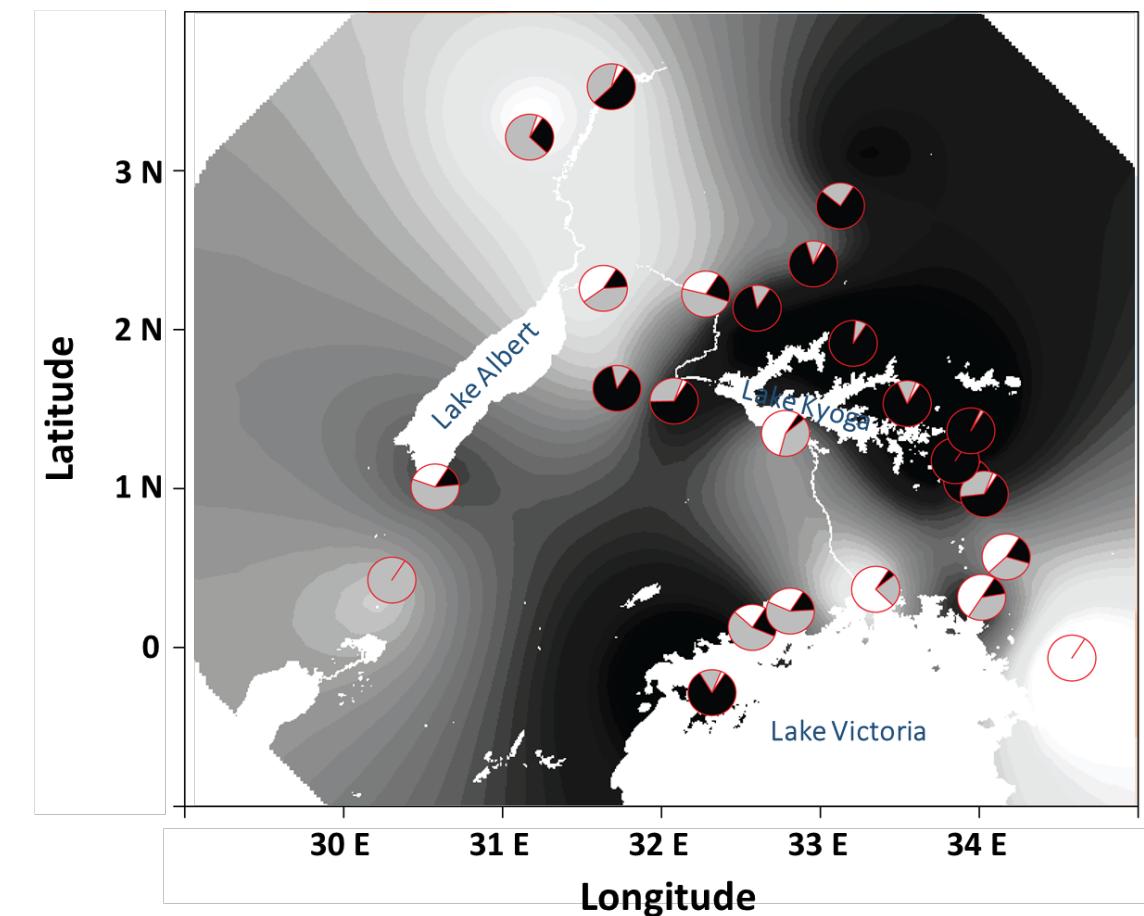
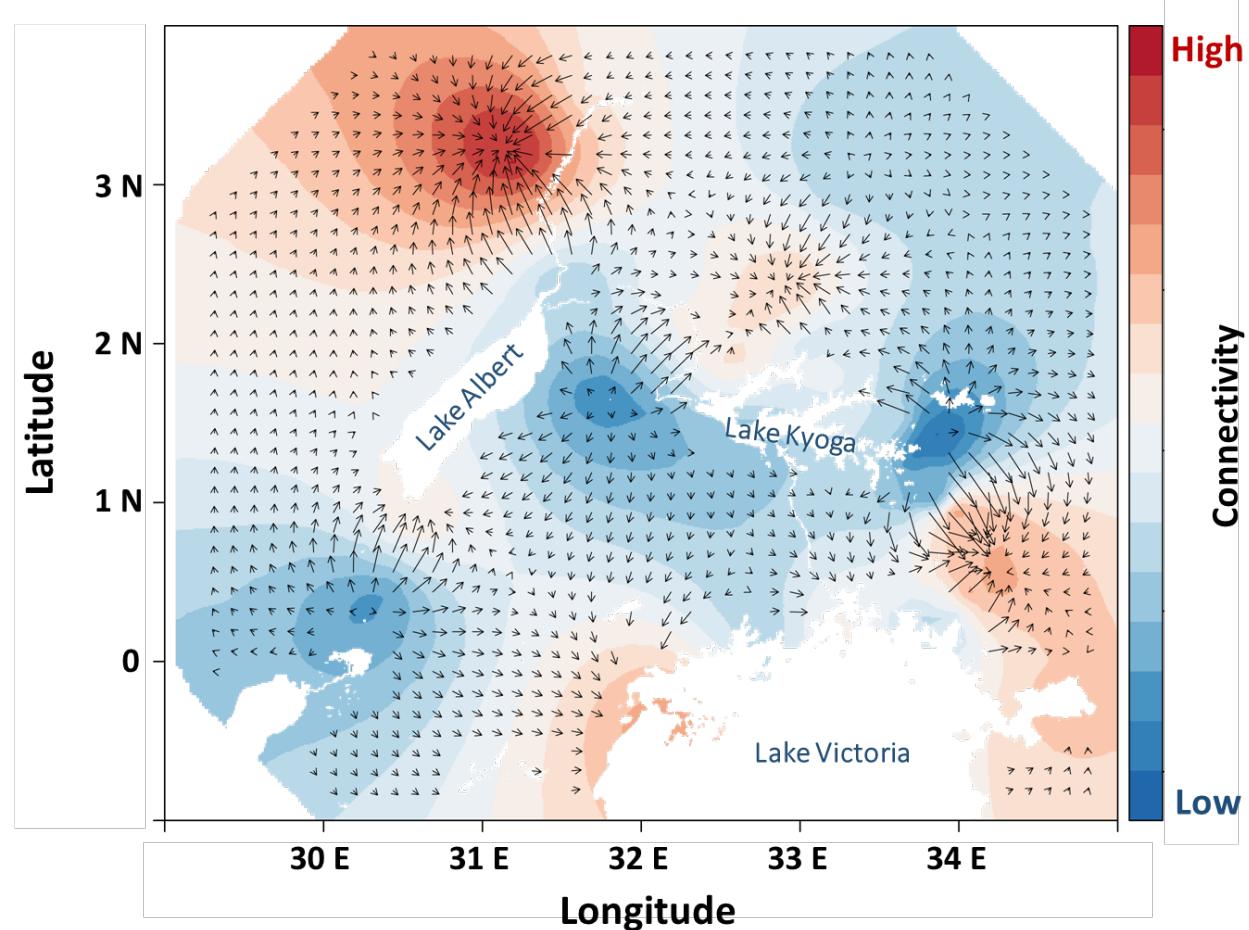
Vector Field:
Magnitude and Direction
of Gene Flow



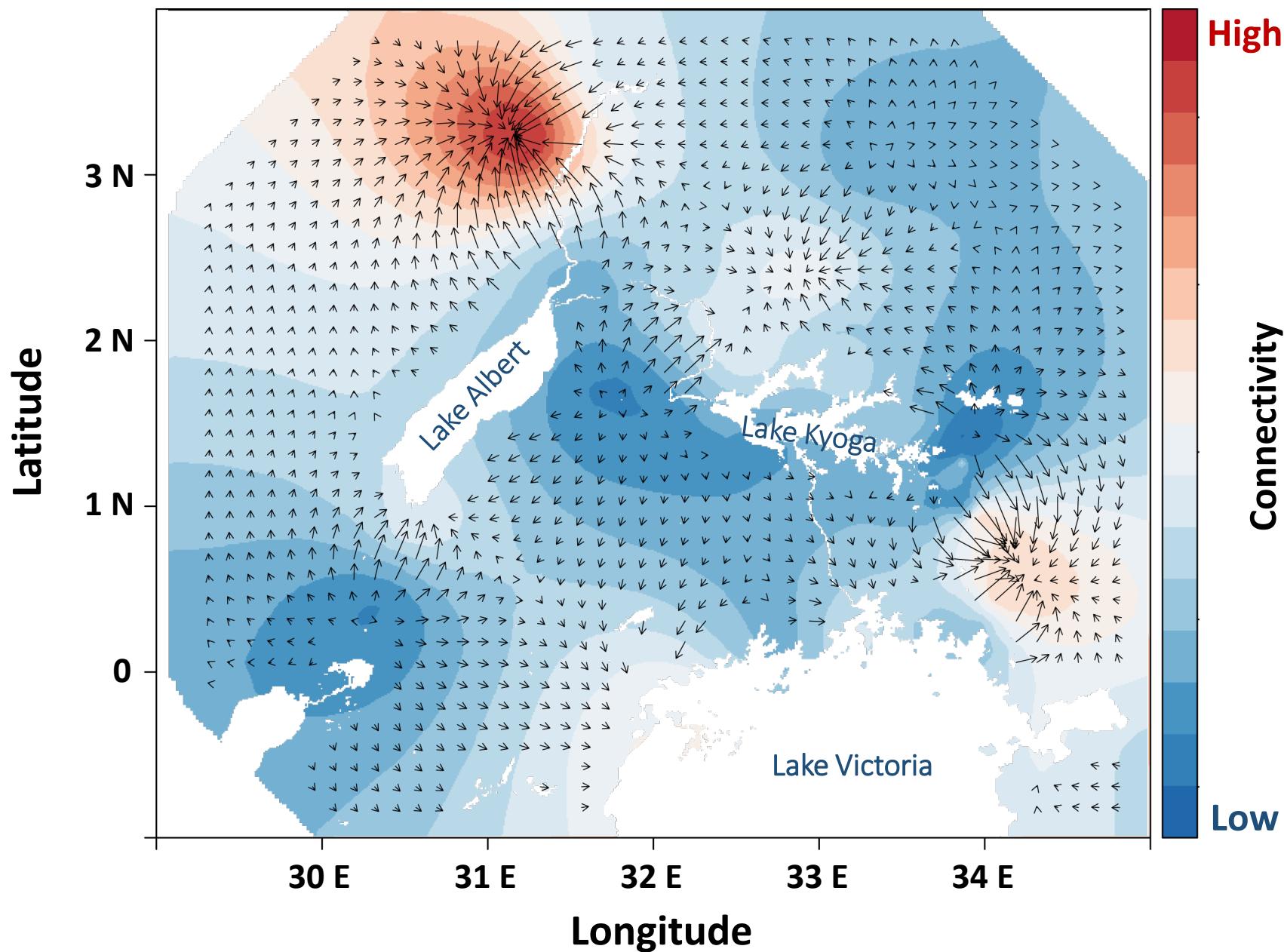
(Positive) Selection



Neutral Gene Flow + Selection



Long-term Connectivity



Results: Neutral and Adaptive Models

<i>Fixed Effects</i>	<i>Neutral Gene Flow</i>	<i>Adaptive Allele L</i>	<i>Adaptive Allele H</i>
Intercept	4.464	-0.066	0.653
x	-0.087	-0.047	-0.047
y	0.380	0.617	0.255
Precip. 1	0.015	-0.988	-0.009
Precip. 2	0.019	0.370	0.056
Temp. 1	-0.014	0.255	0.141
Temp. 2	0.008	0.052	-0.063
Veg. 1	0.001	0.051	0.032
Veg. 2	0.003	-0.040	-0.023
Dist. Wat.	-0.008	-0.278	0.135
Land Cover	-0.002	0.008	-0.020

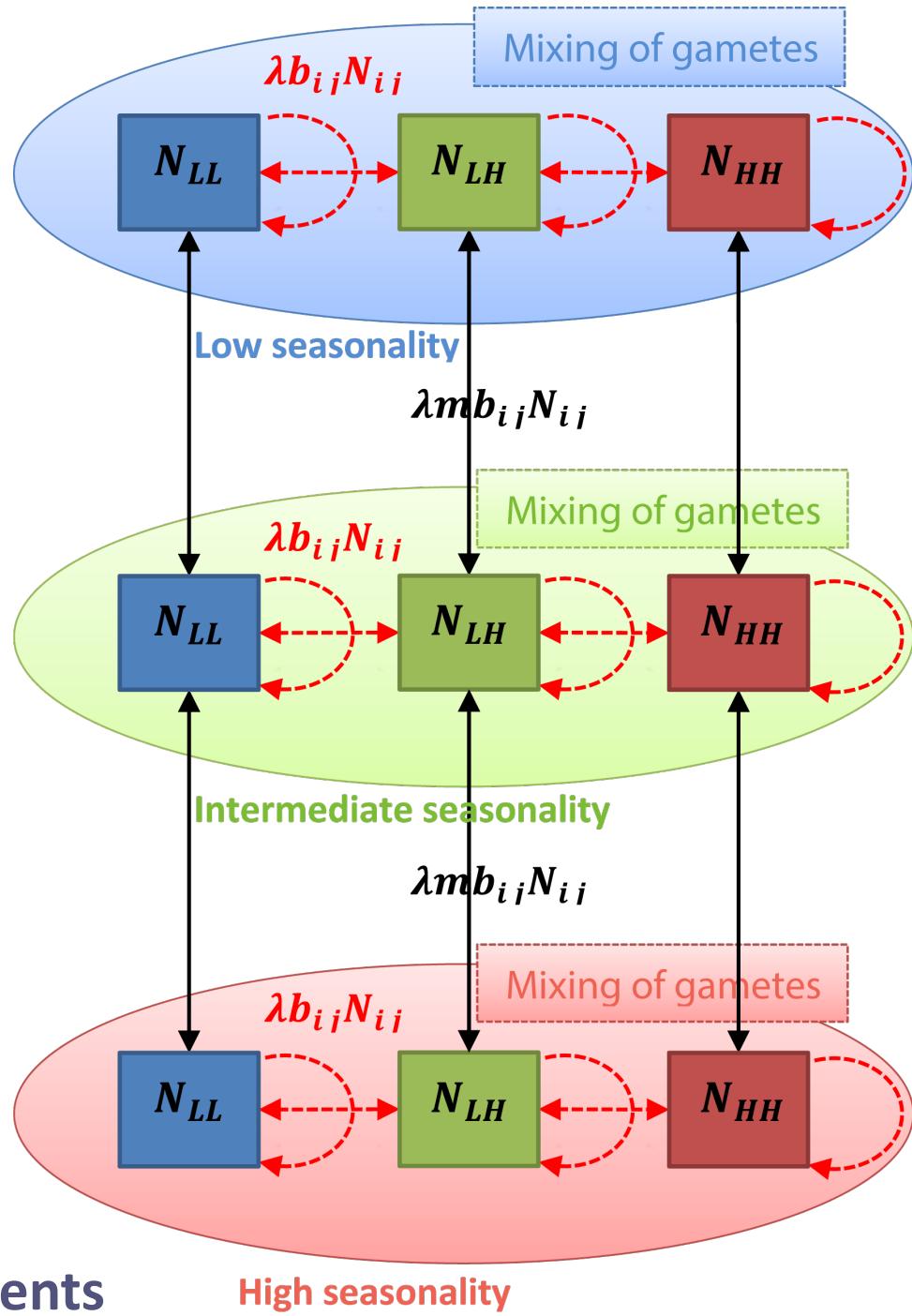
Divergence: Forward Simulation

b = survivorship:

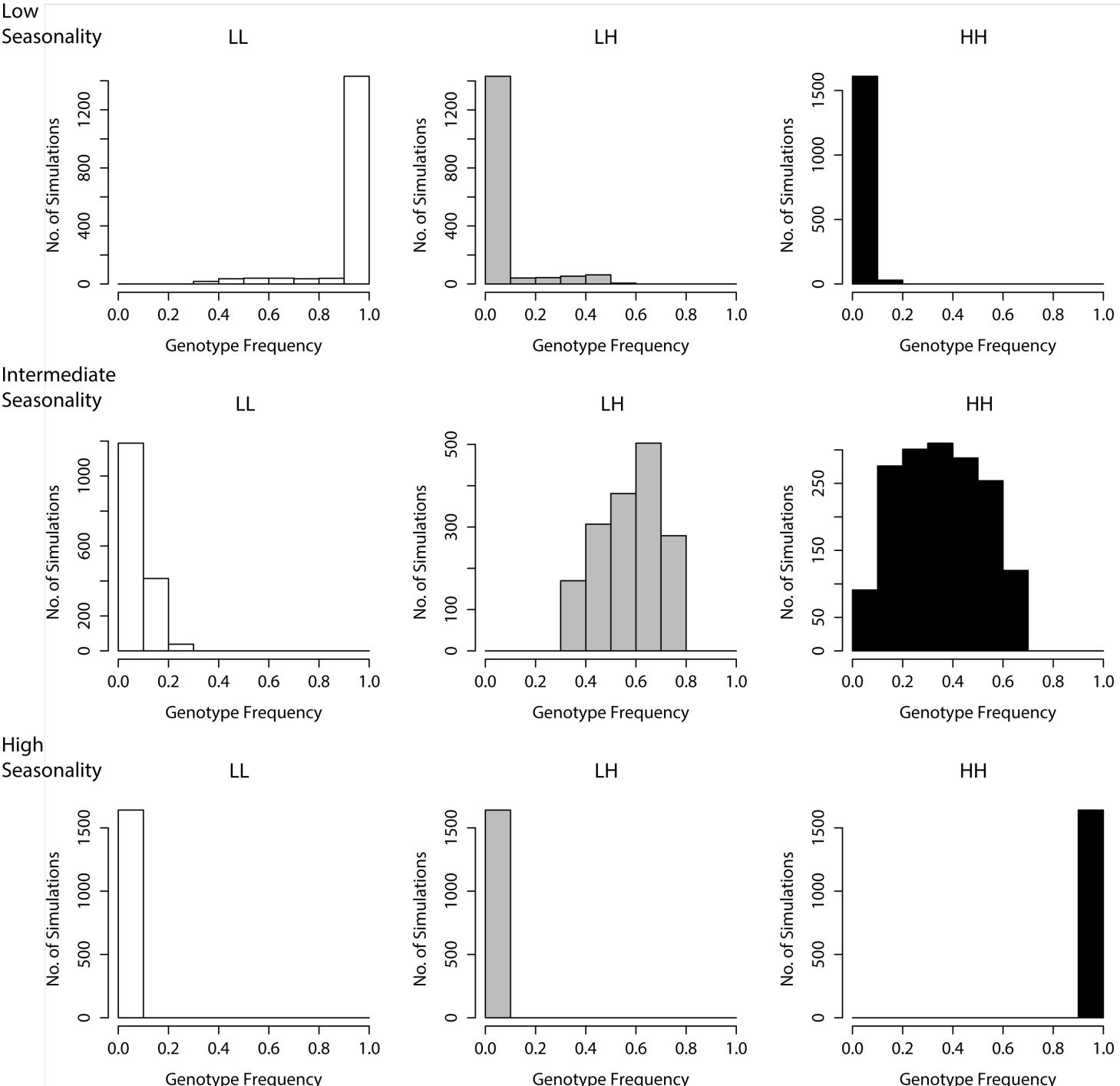
	LL	LH	HH
Low Seasonality	0.48	0.35	0.14
Intermediate Seasonality	0.16	0.44	0.18
High Seasonality	0.03	0.13	0.84

Spatially-varying selection:

Different genotypes selected in different environments



Divergence: Forward Simulation

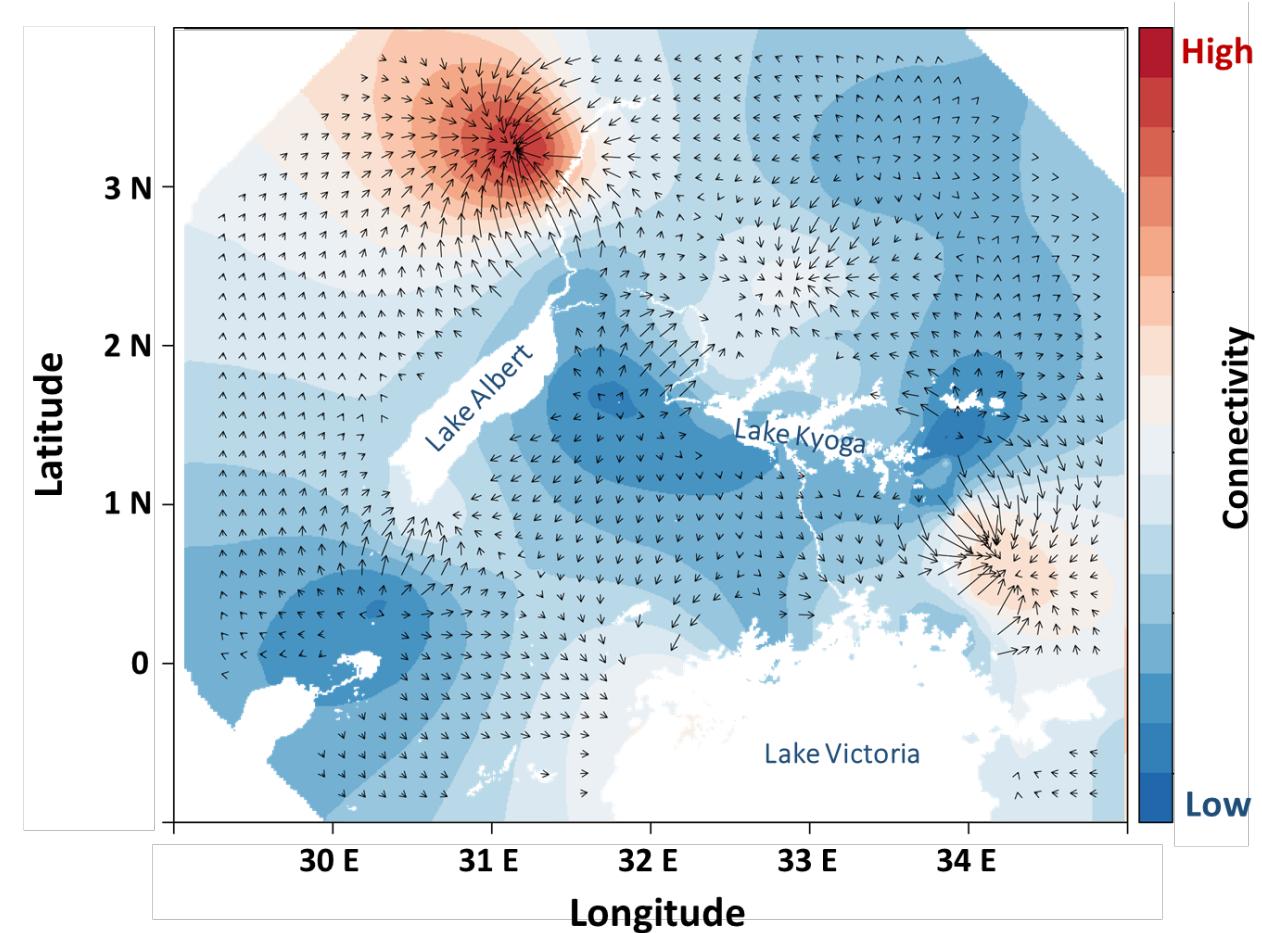
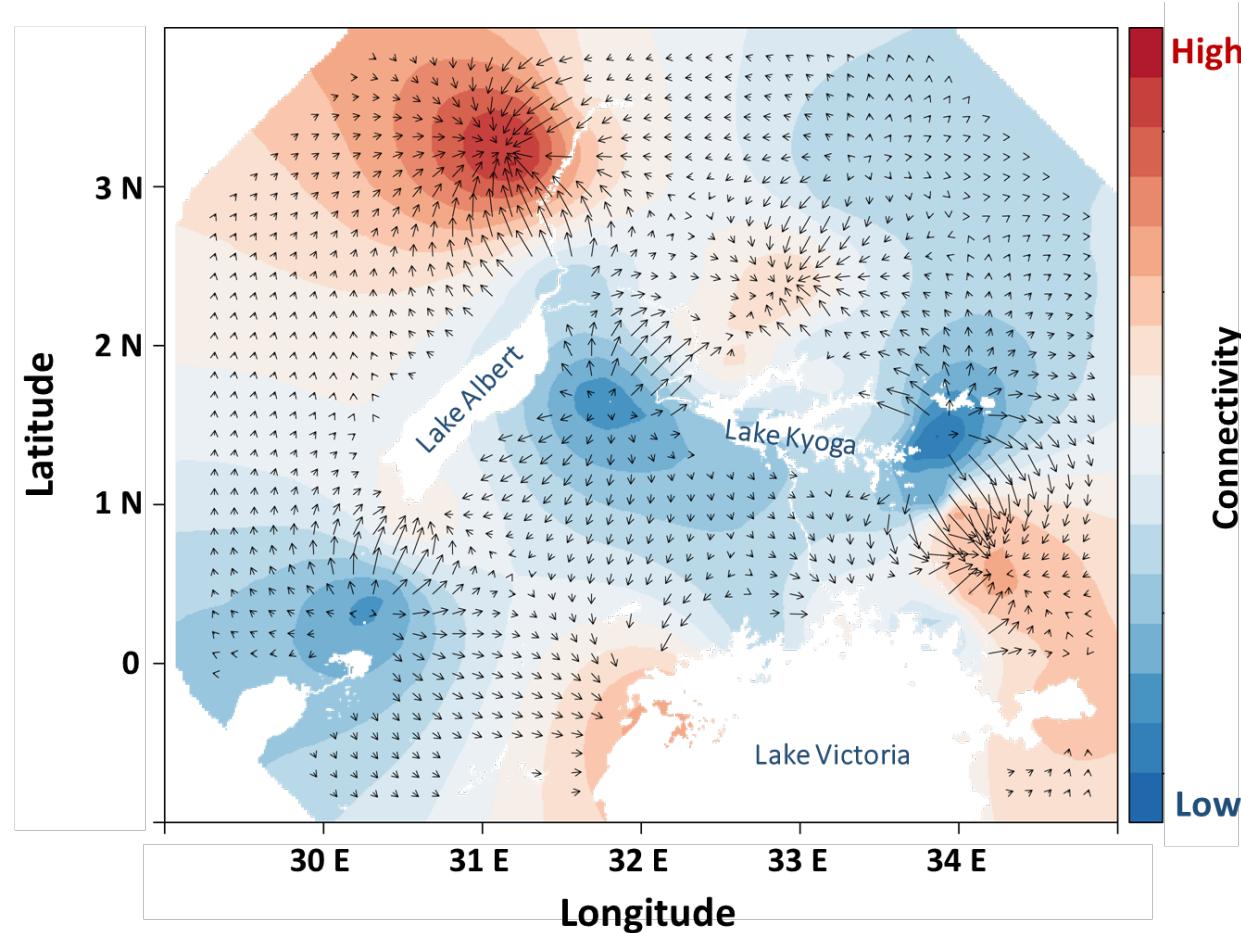


Spatially-varying selection:
Persistence of heterozygotes

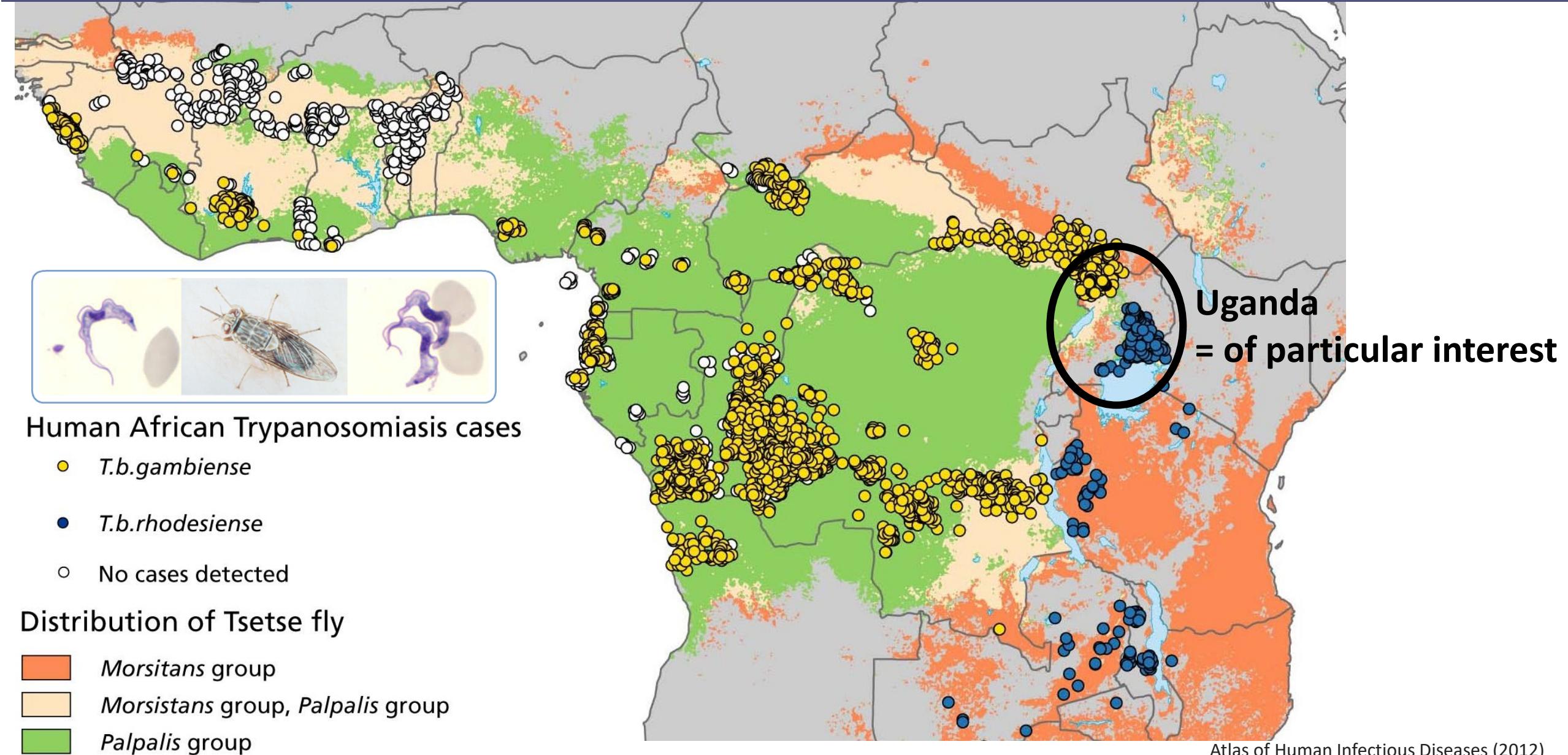
Landscape Connectivity

Which areas to target?

Importance of accounting for selection



Separation of Trypanosome Strains Maintained?



Other Applications



“And now for something completely different.”

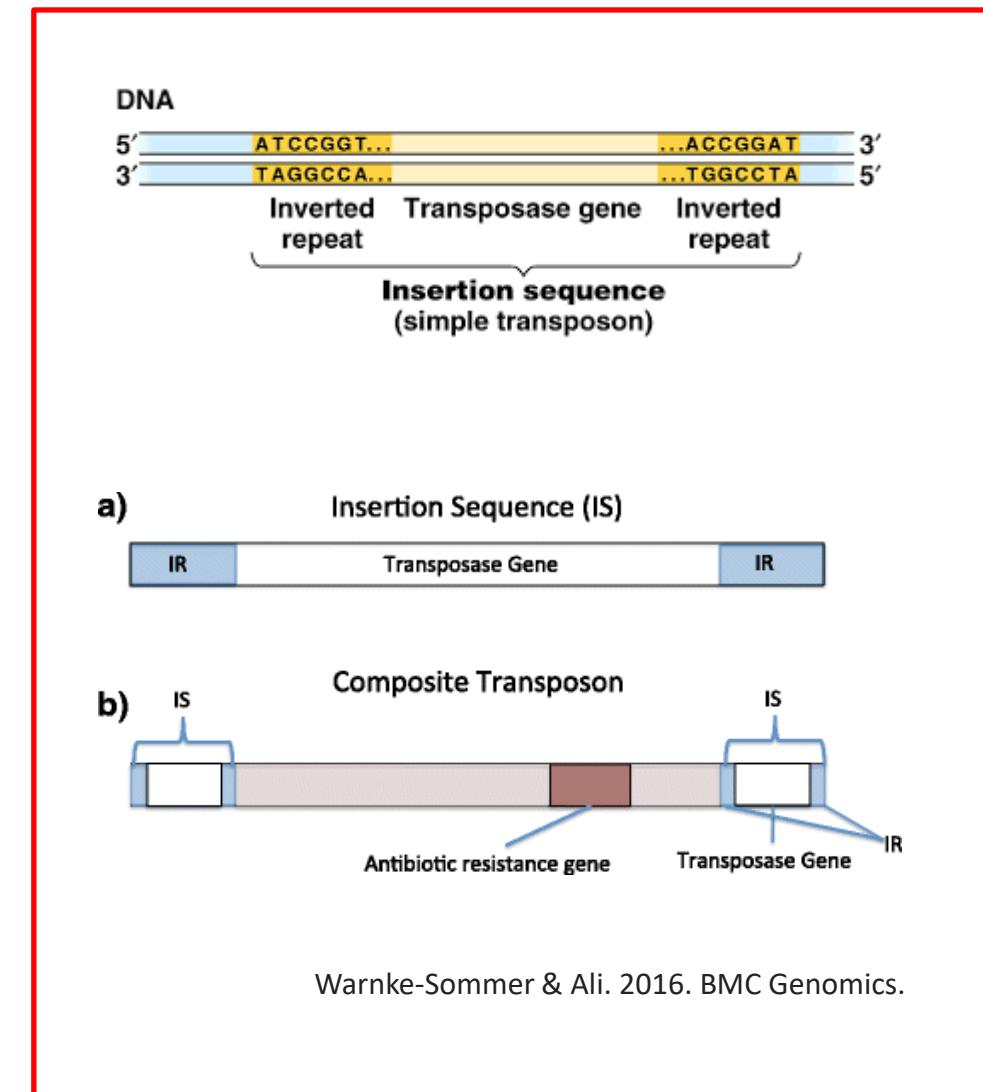
Monty Python

E.g.: Transposable Element Connectivity

How Antibiotic Resistance Spreads

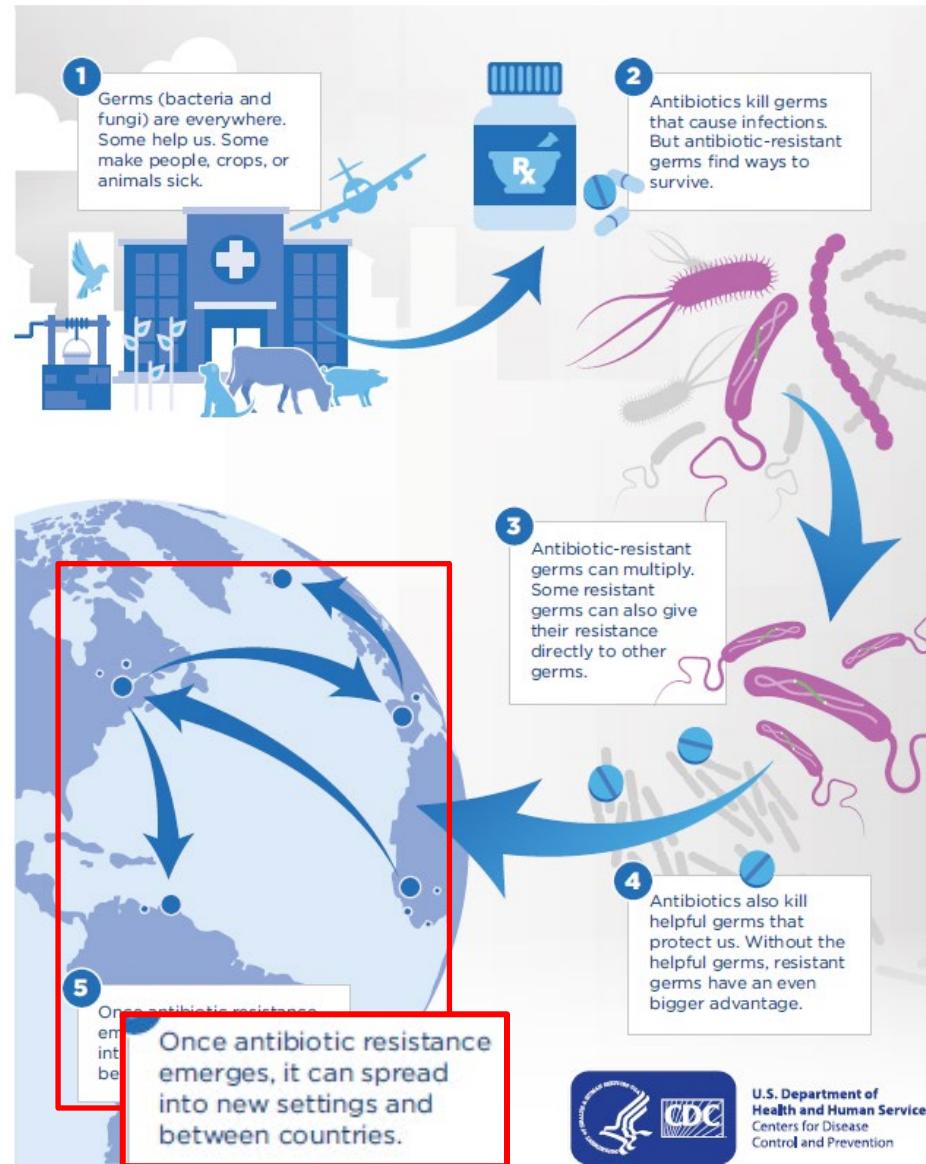


Insertion Sequences = Simplest Mobile Elements

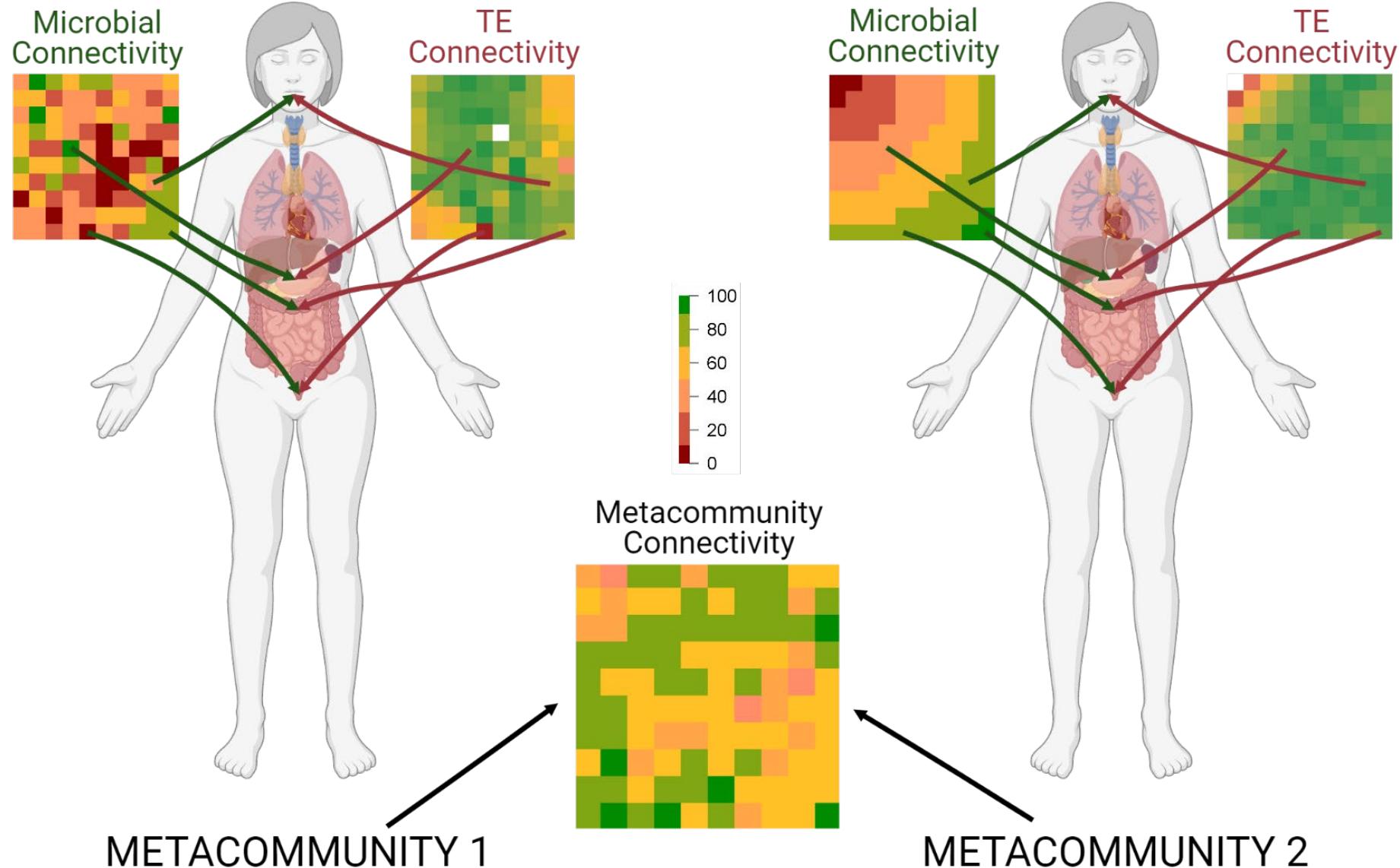


E.g.: Microbial Community Connectivity

How Antibiotic Resistance Spreads



E.g.: Microbial and TE Community Connectivity Among Hosts



A scalable metric that captures connectivity at multiple scales, from alleles to communities

Chaz Hyseni

