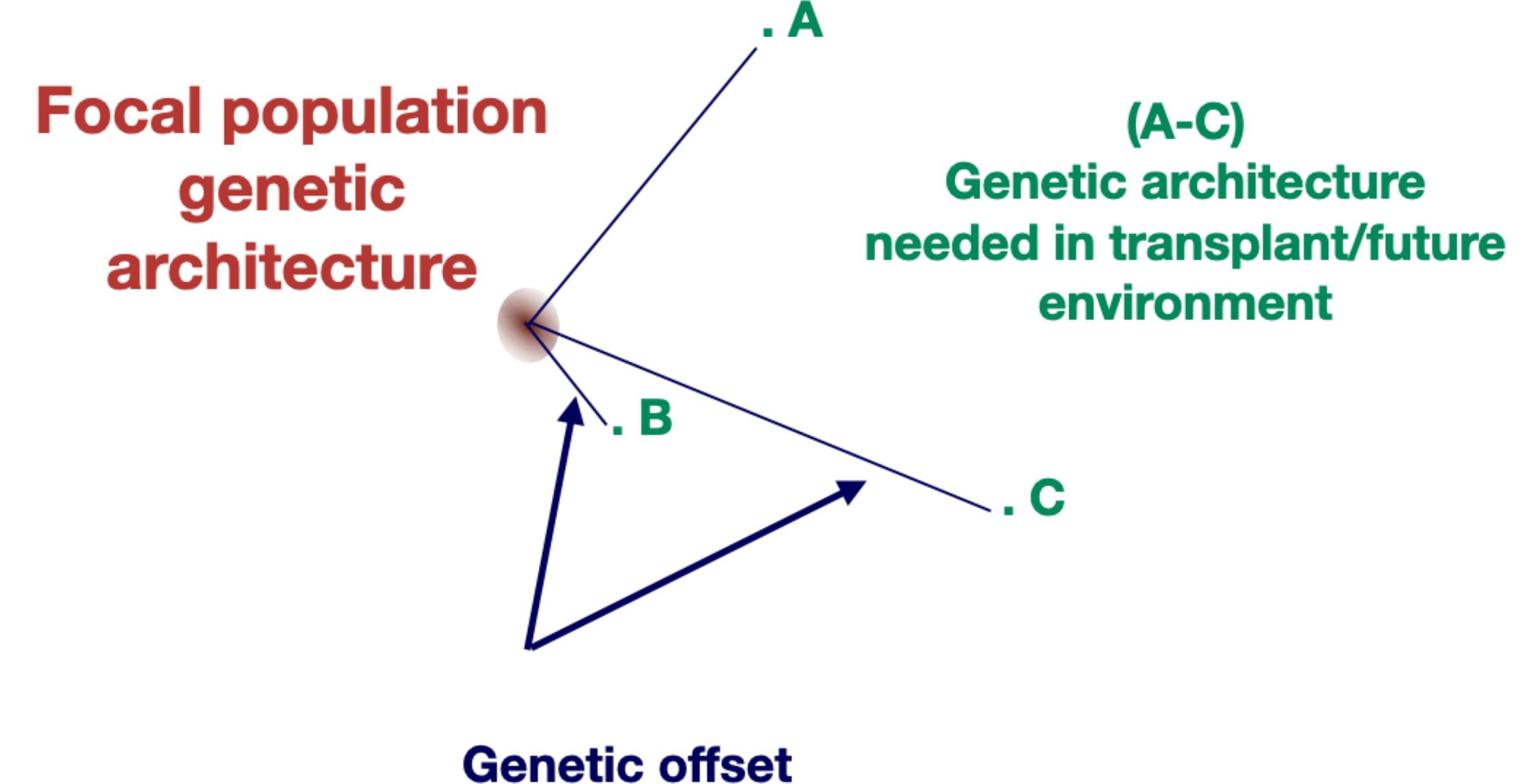


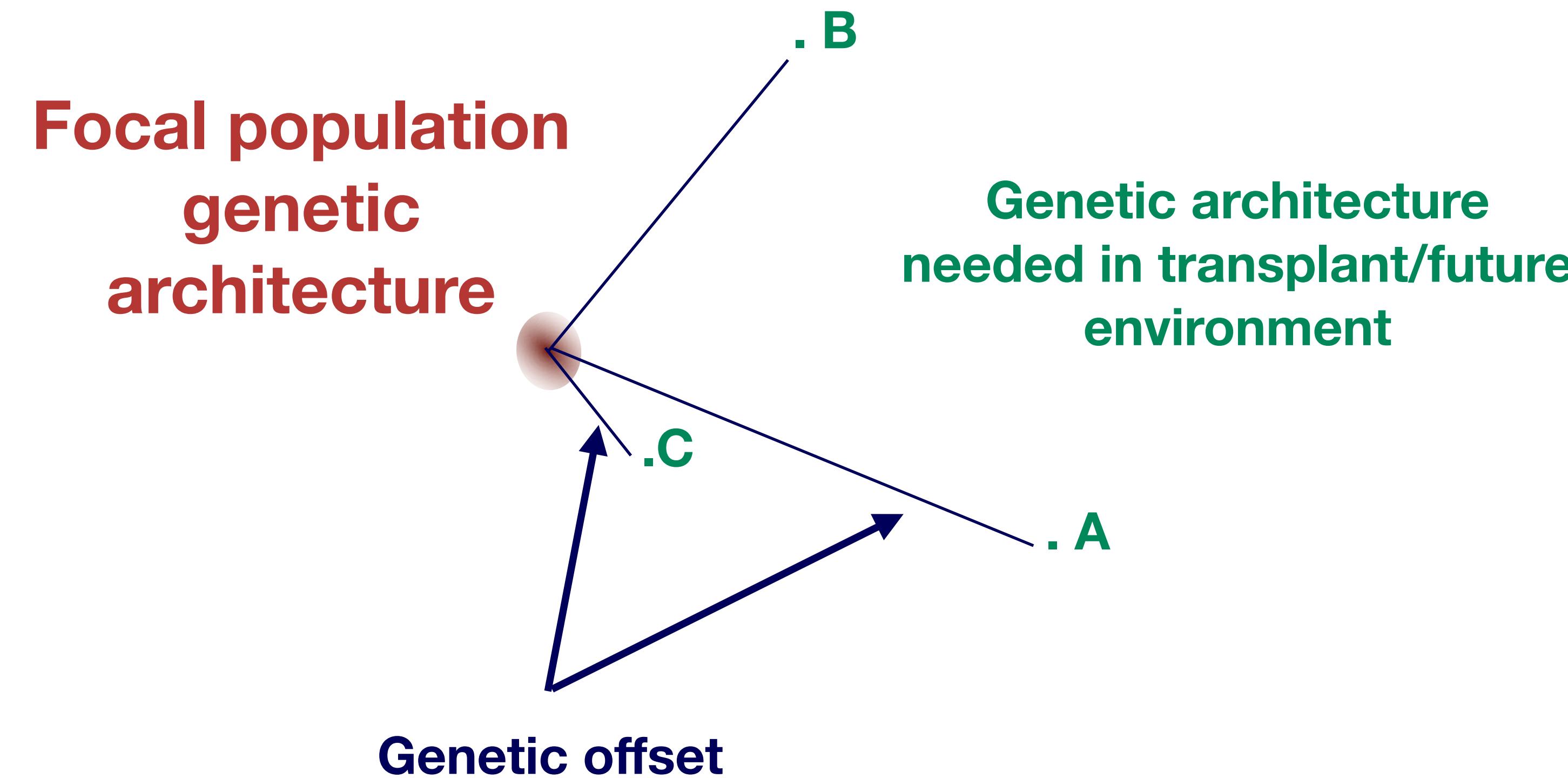
Considerations for applying genomic offsets: from experimental design to interpretation

Dr. Katie Lotterhos
Northeastern University Marine Science Center

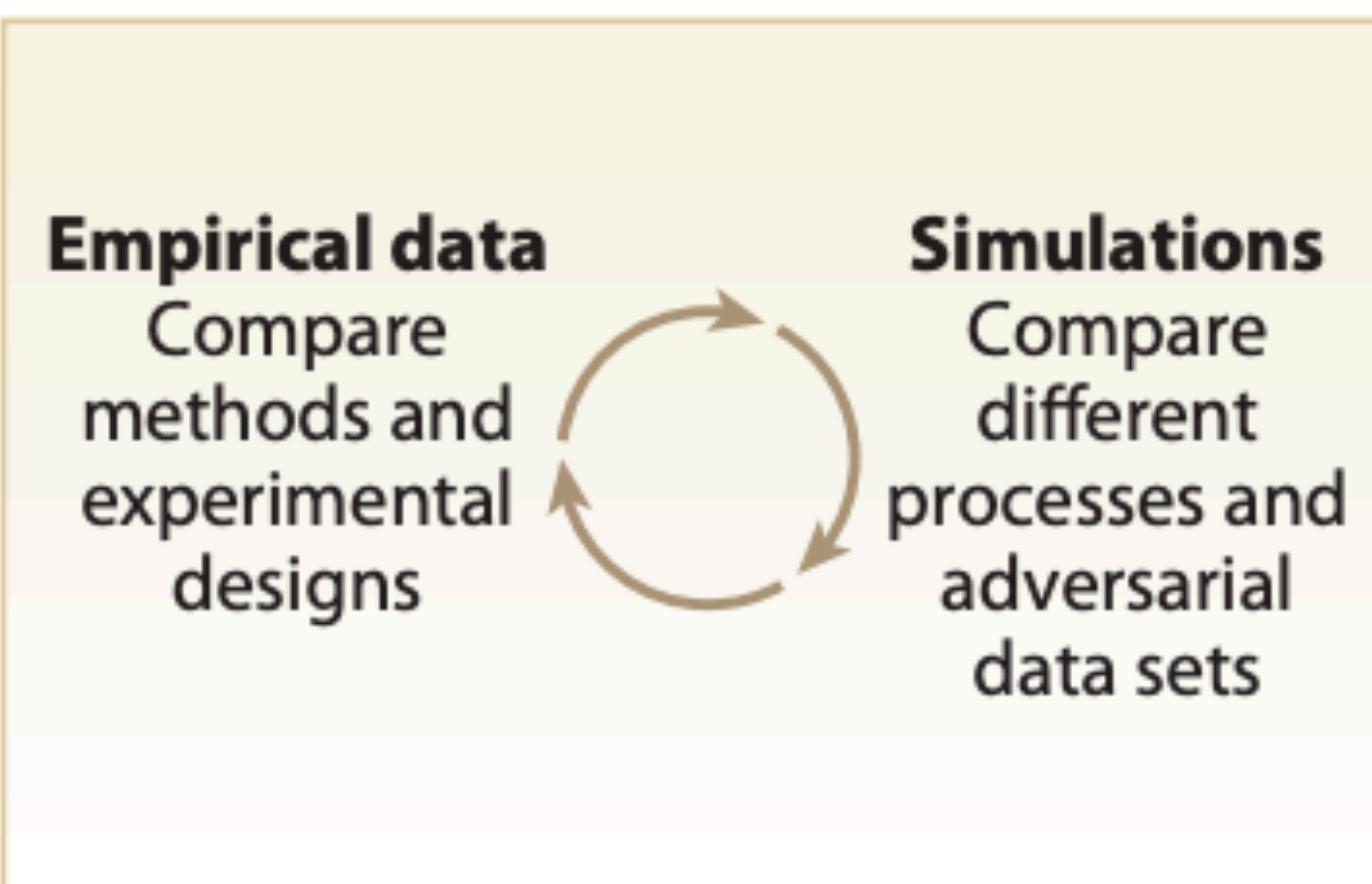
k.lotterhos@northeastern.edu
@DrK_Lo



Genetic/Genomic offset: The degree of maladaptation to a future/transplant environment

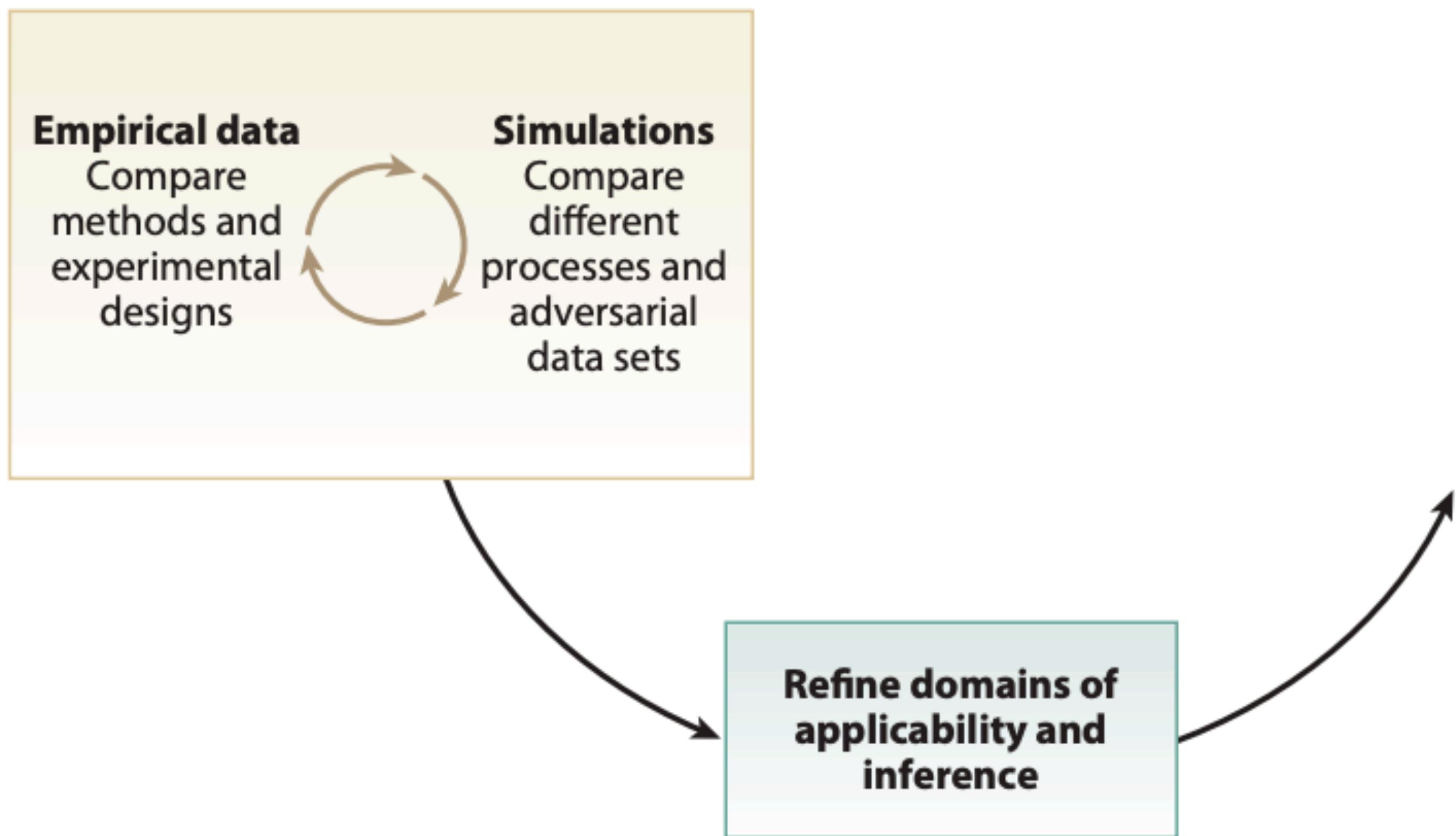


b Method Evaluation



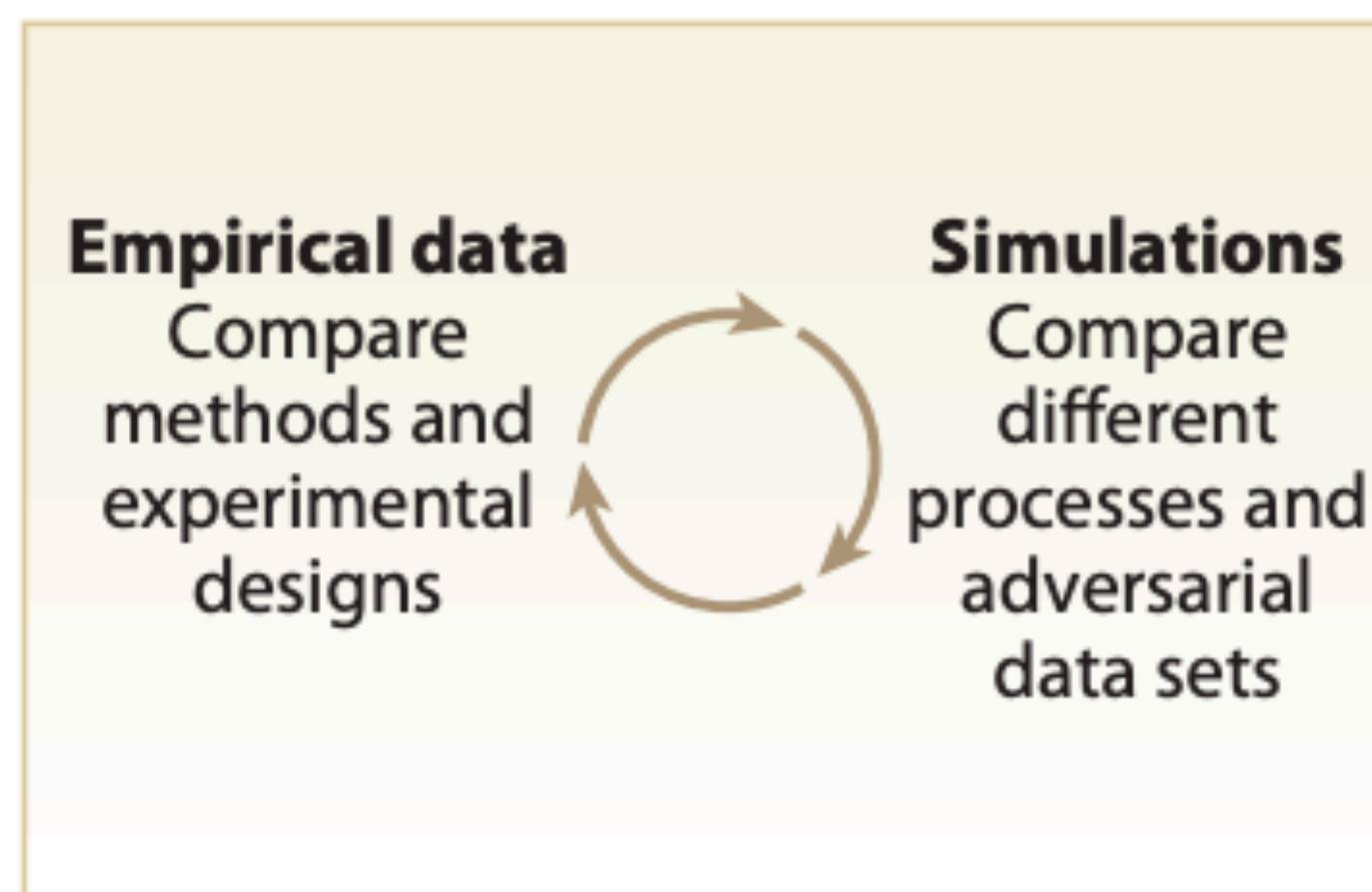
**Evaluation vs.
validation**

b Method Evaluation

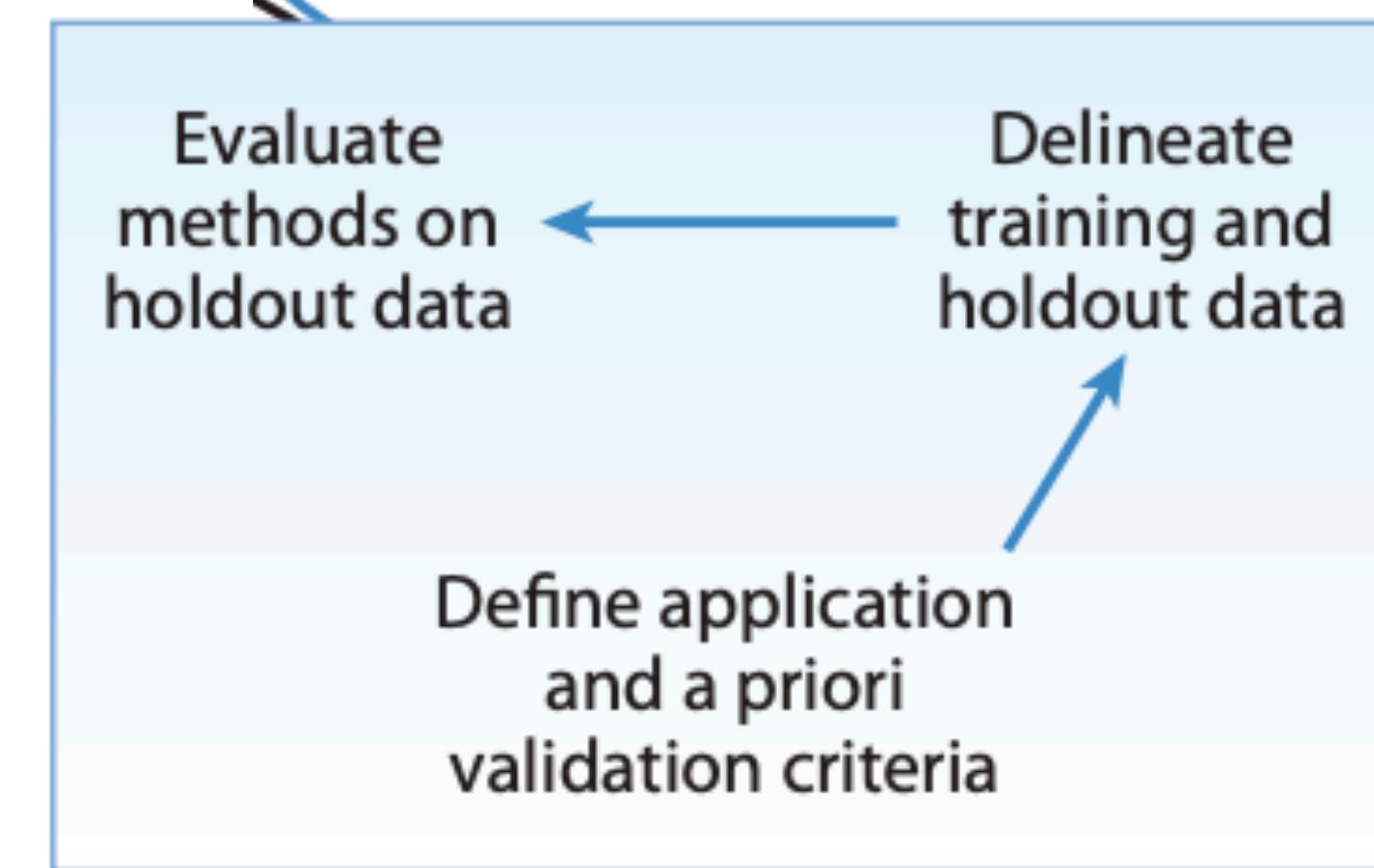


**Evaluation vs.
validation**

b Method Evaluation



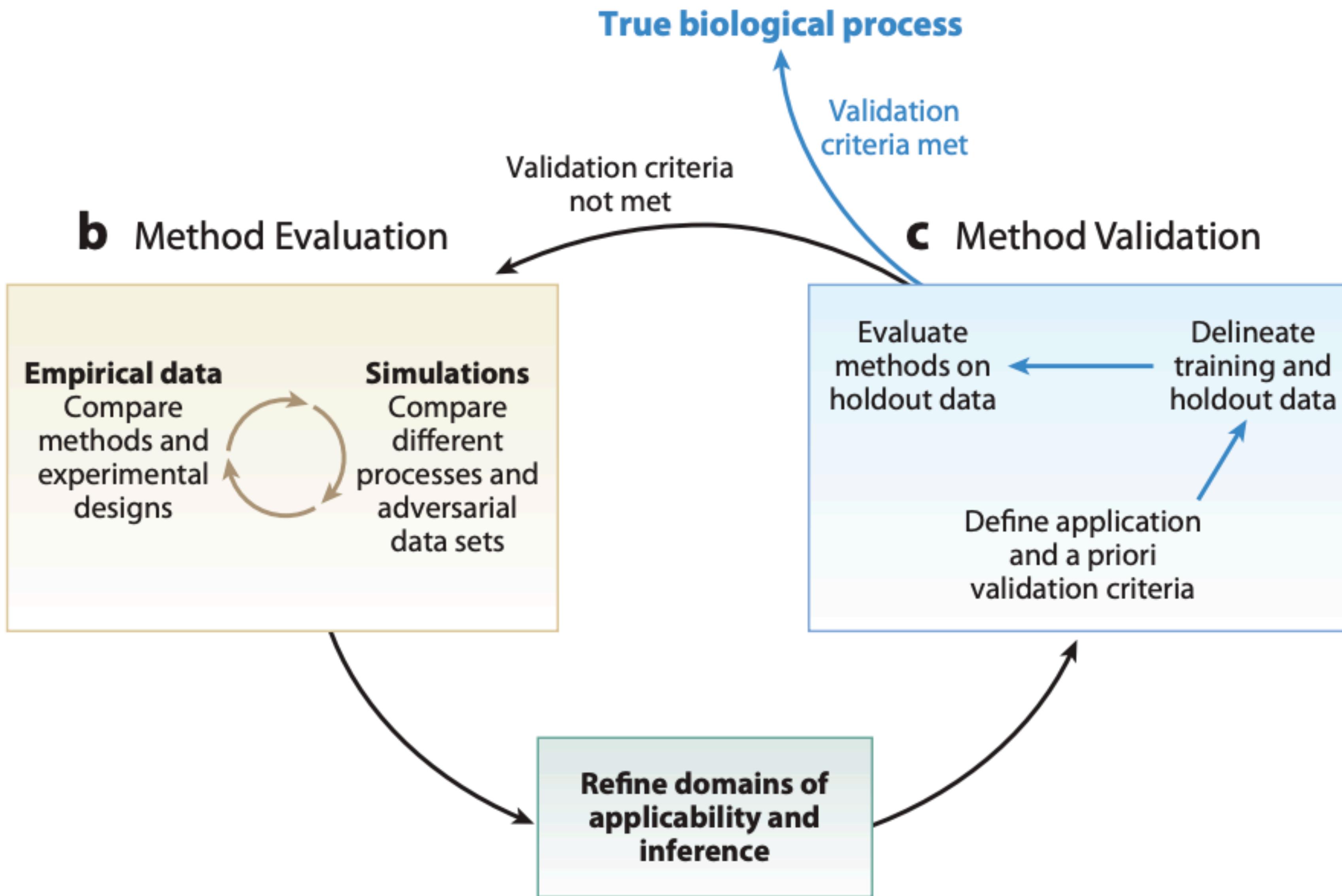
c Method Validation



Refine domains of applicability and inference

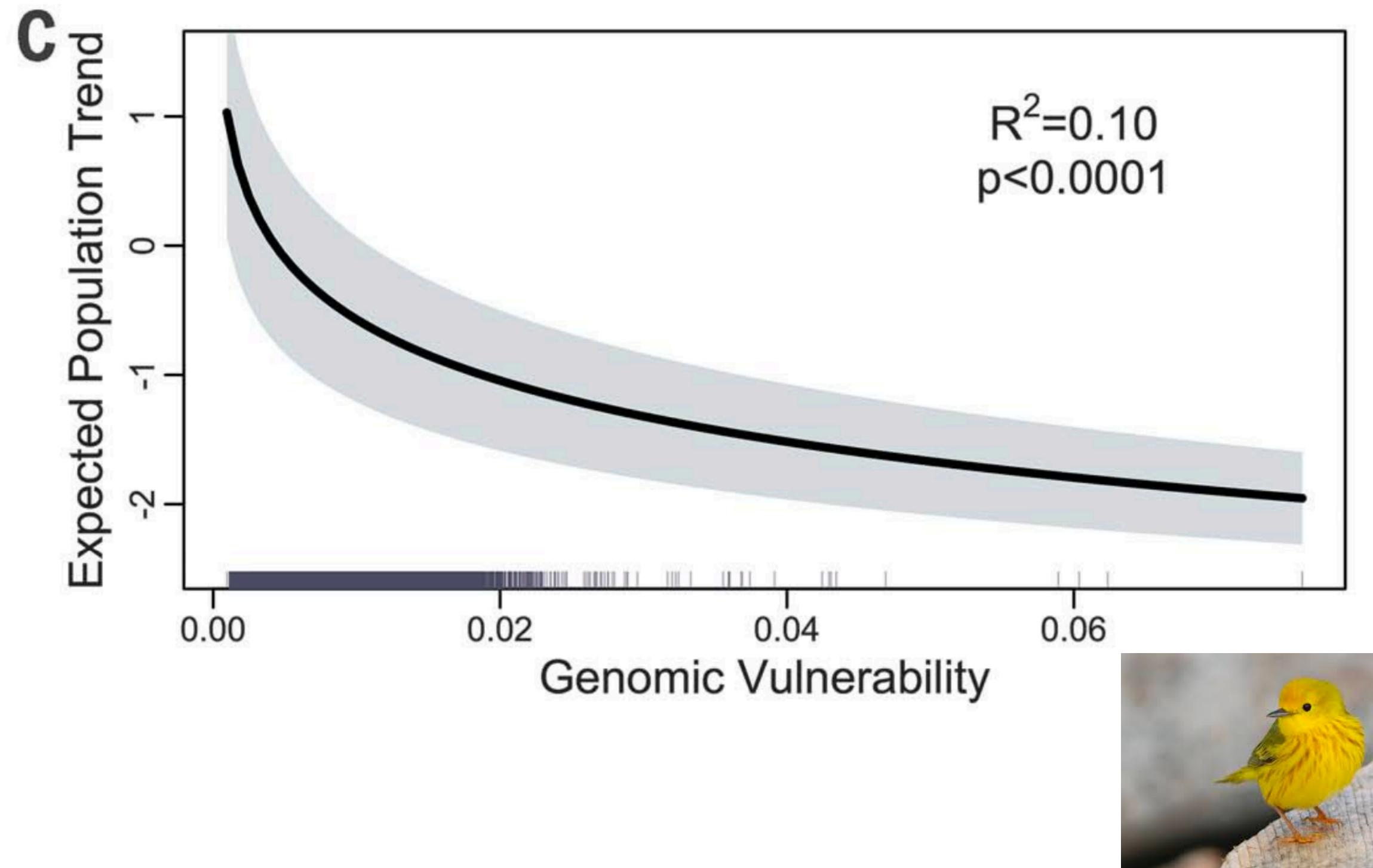
**Evaluation vs.
validation**

Evaluation vs. validation



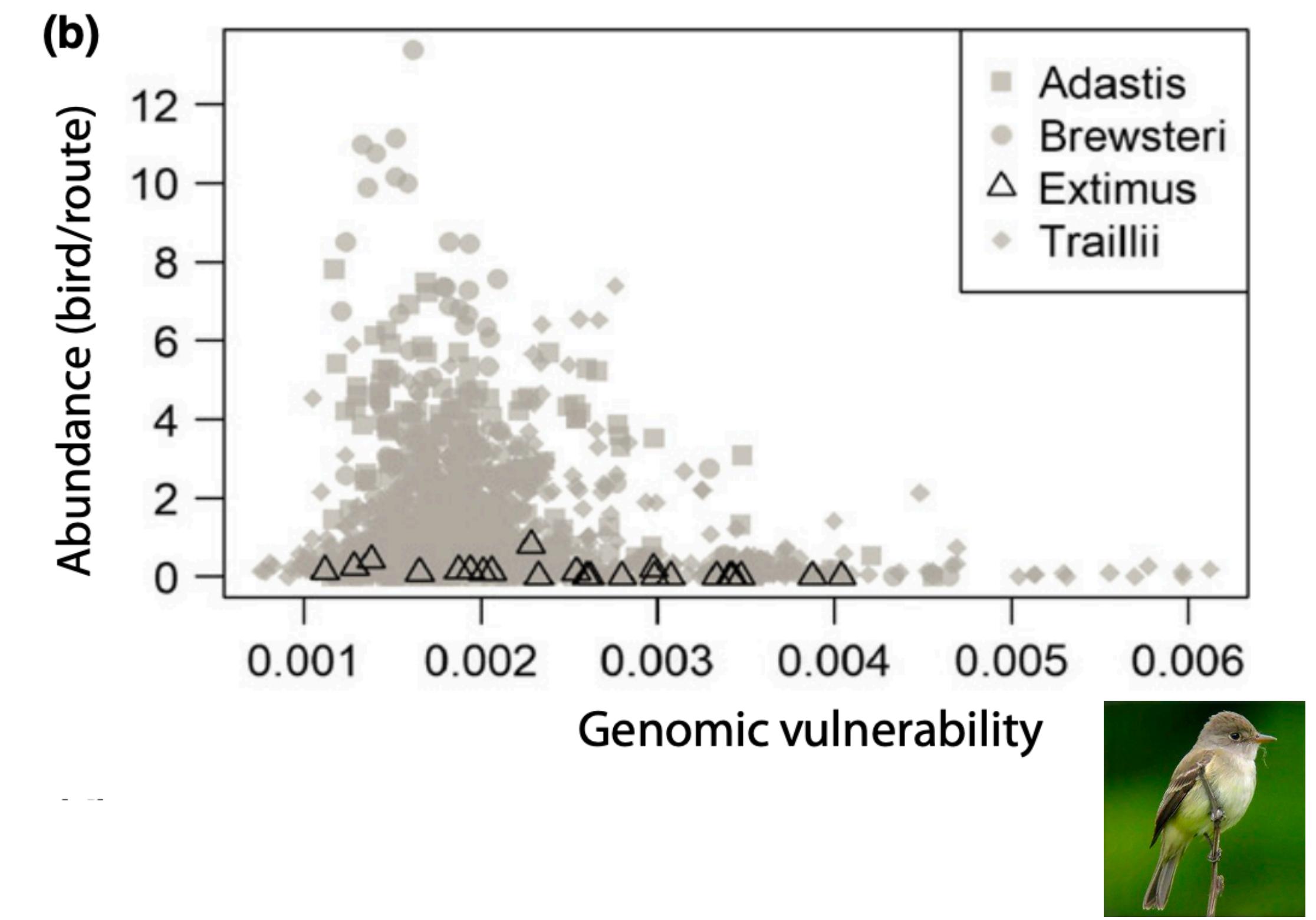
How do we validate genetic offset metrics? Are these studies an example of validation?

**Genomic signals of selection predict
climate-driven population declines
in a migratory bird**



Bay et al. 2018, *Science*

**Ecological genomics predicts climate vulnerability in an
endangered southwestern songbird**

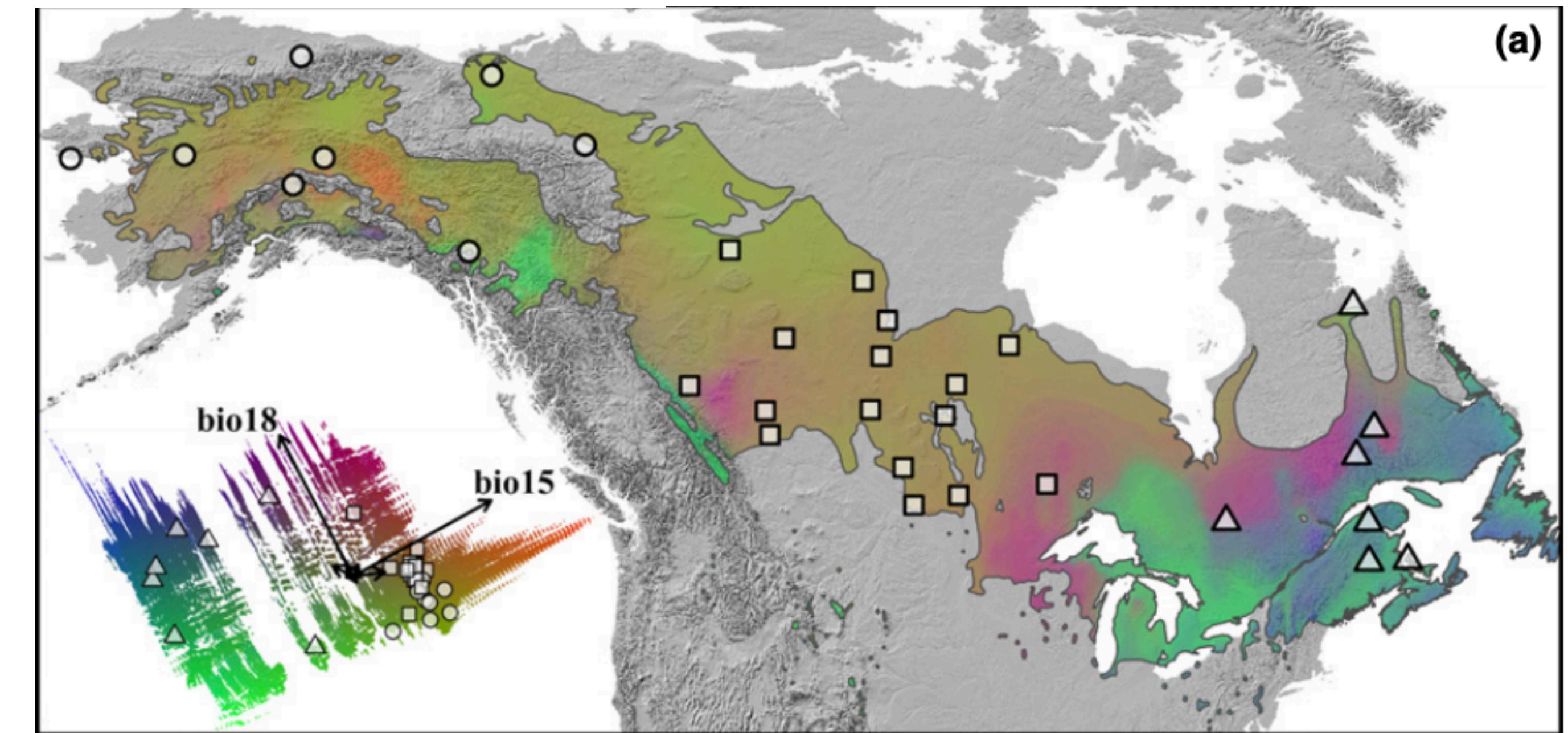
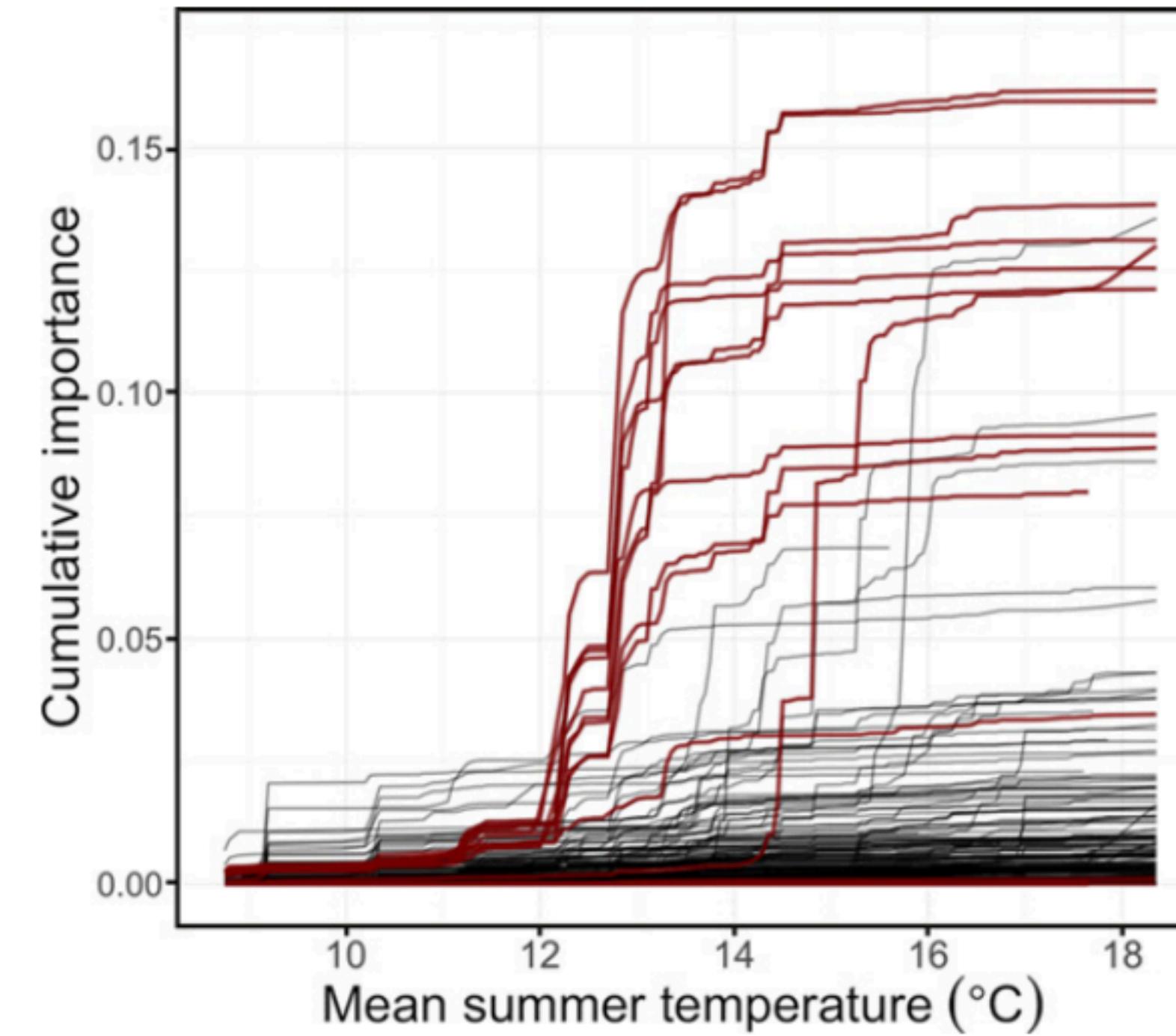


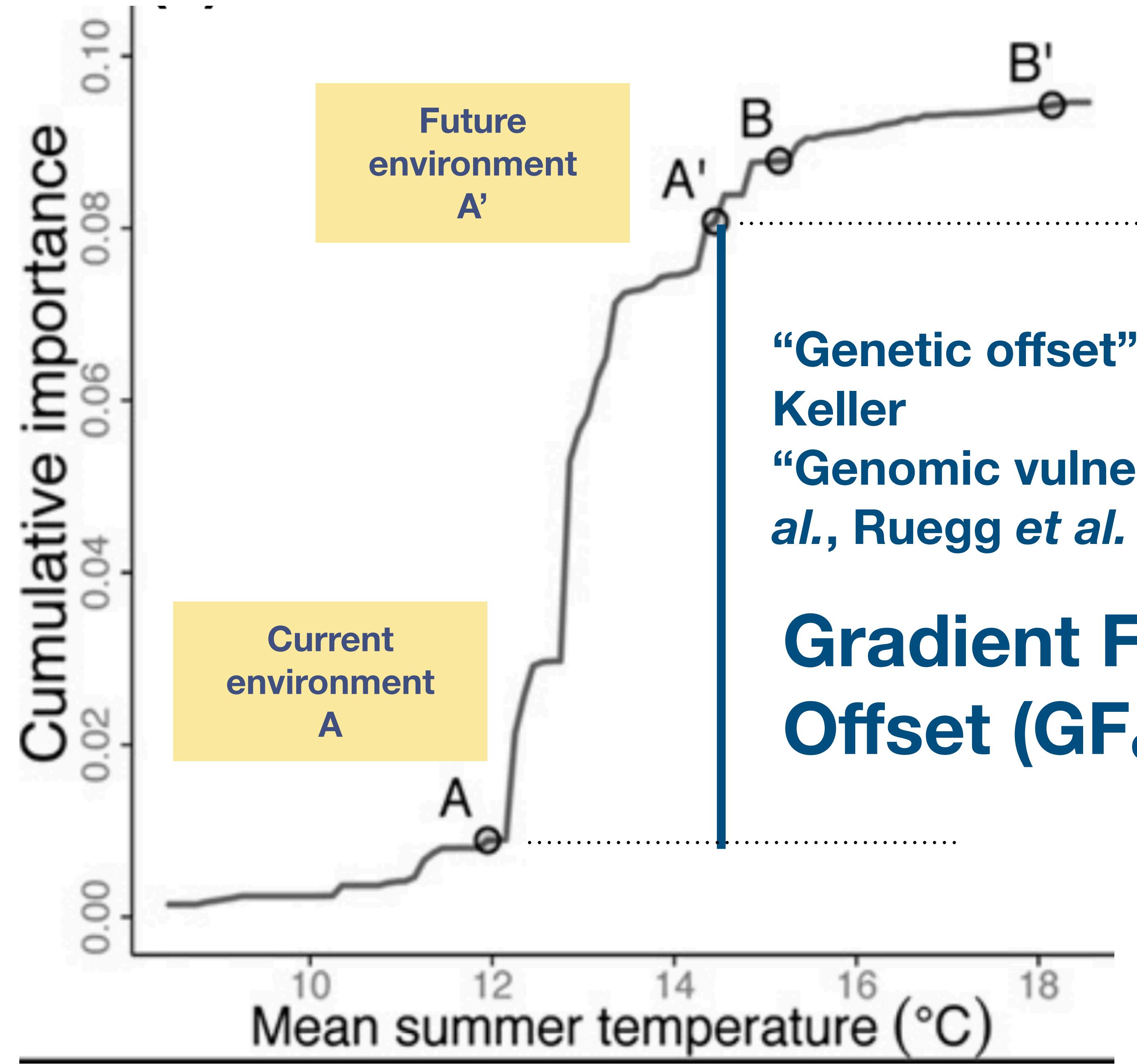
Ruegg et al. 2018, *Ecology Letters*

Gradient Forests is used to calculate genetic offset

Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation

Fitzpatrick and Keller 2015





“Genetic offset” Fitzpatrick and Keller
“Genomic vulnerability” of Bay et al., Ruegg et al.

Gradient Forests Offset (GF_{offset})

Gradient Forests Analysis Validation Project

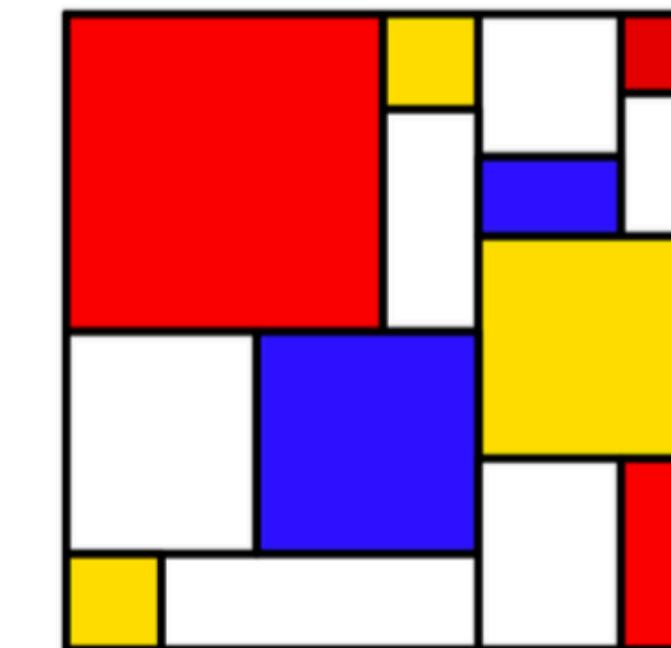


**Áki
Láruson**



**Ben
Haller**

SLiM: An Evolutionary Simulation Framework
Selection on Linked Mutations



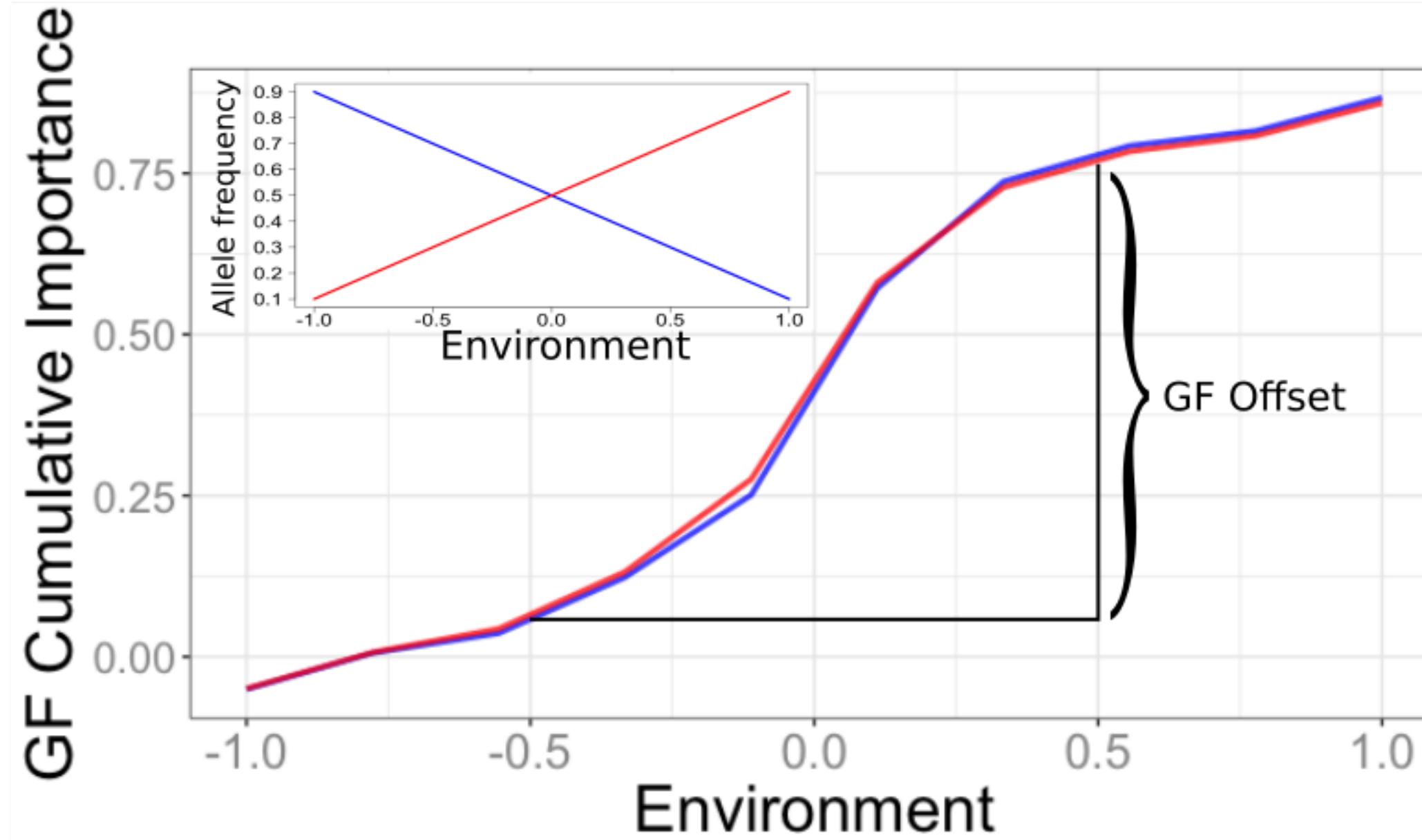
**Matt
Fitzpatrick**



**Steve
Keller**



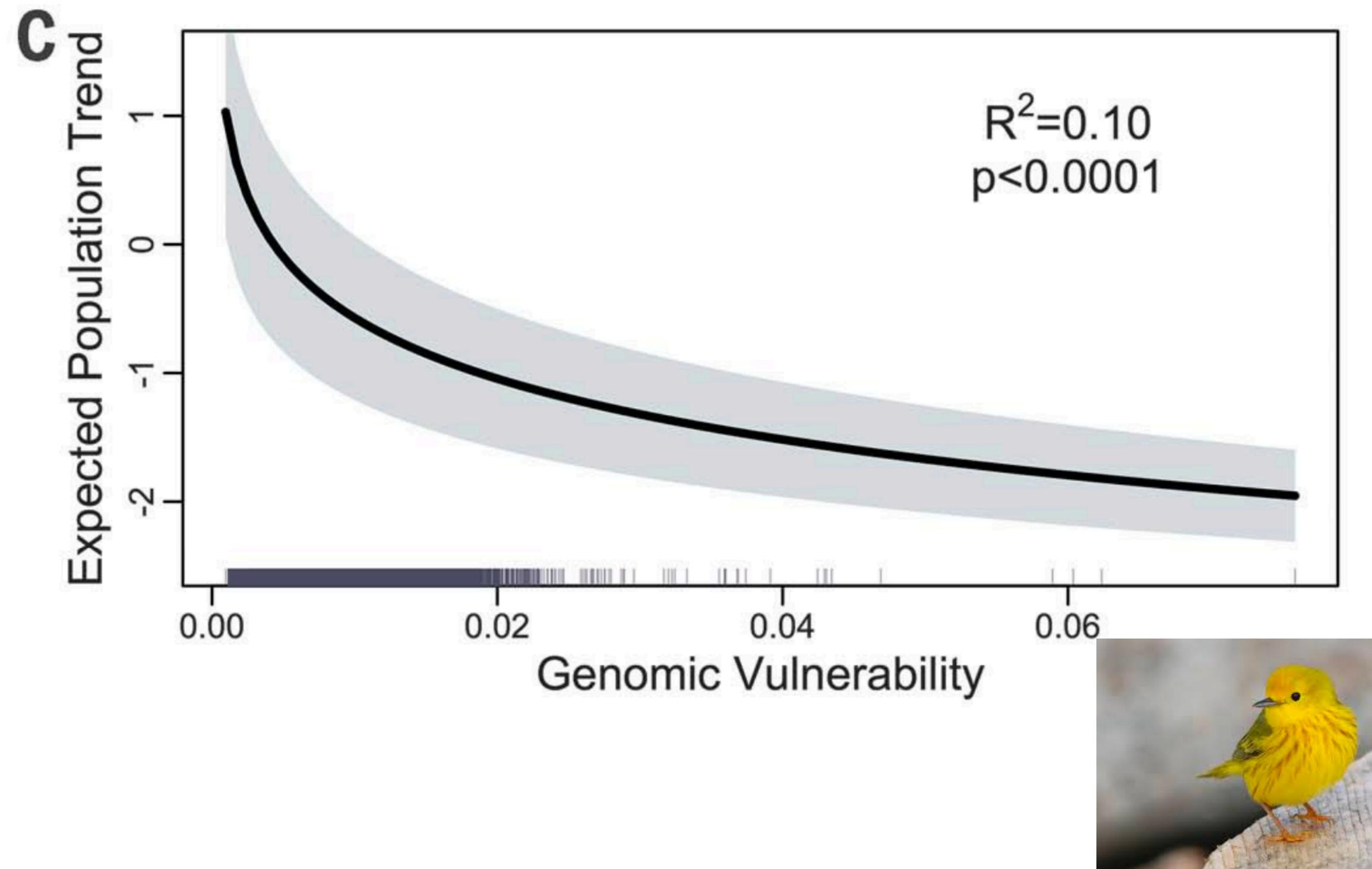
Láruson, Haller, Keller, Fitzpatrick, and Lotterhos *in prep*



Models turnover in allele frequencies on an environmental gradient

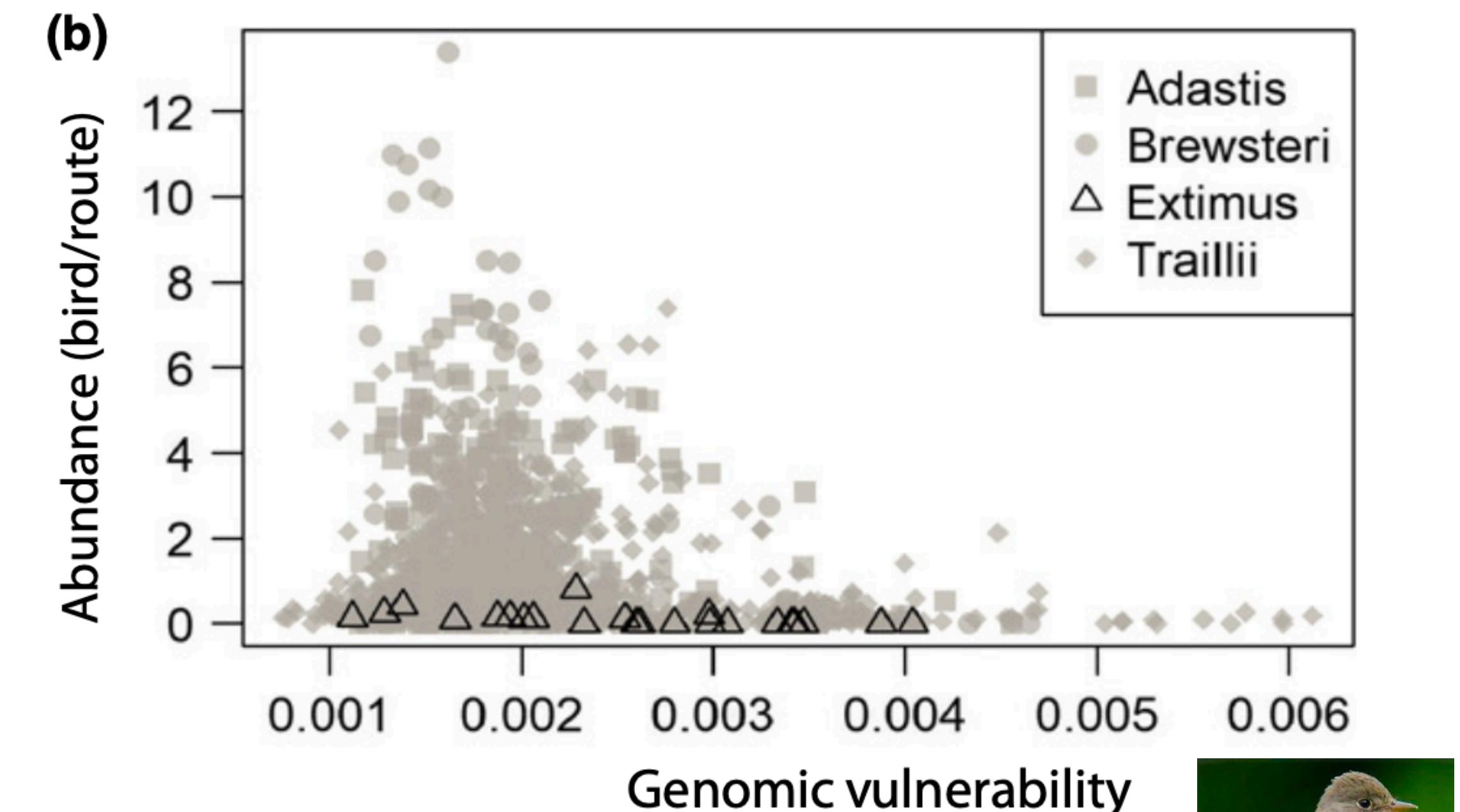
Could these patterns arise because of genetic drift and not because of signals of selection?

**Genomic signals of selection predict
climate-driven population declines
in a migratory bird**



Bay et al. 2018, *Science*

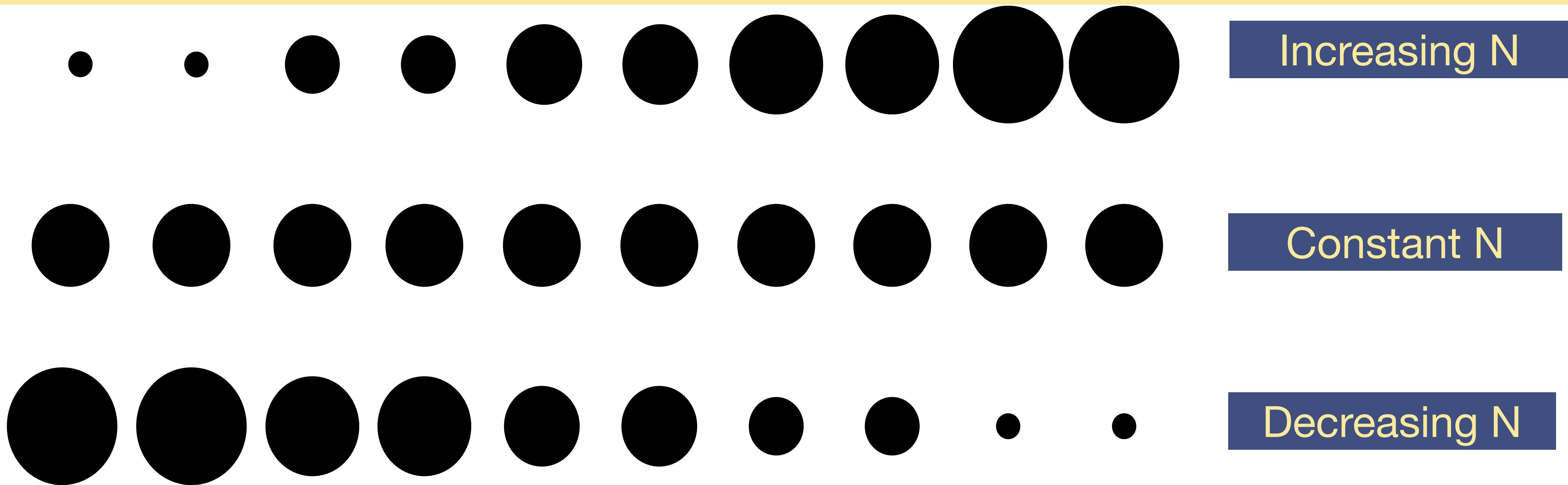
**Ecological genomics predicts climate vulnerability in an
endangered southwestern songbird**

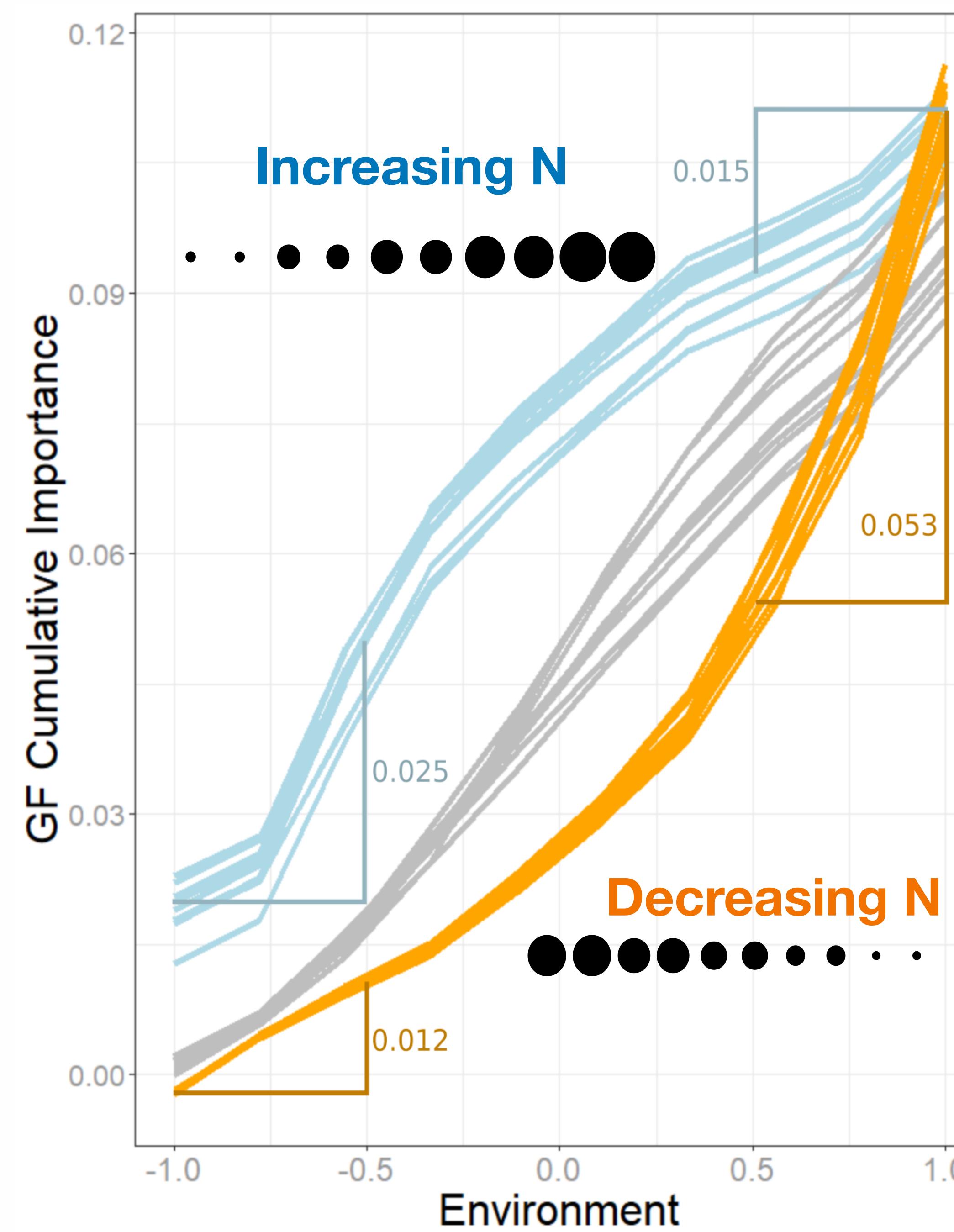


Ruegg et al. 2018, *Ecology Letters*



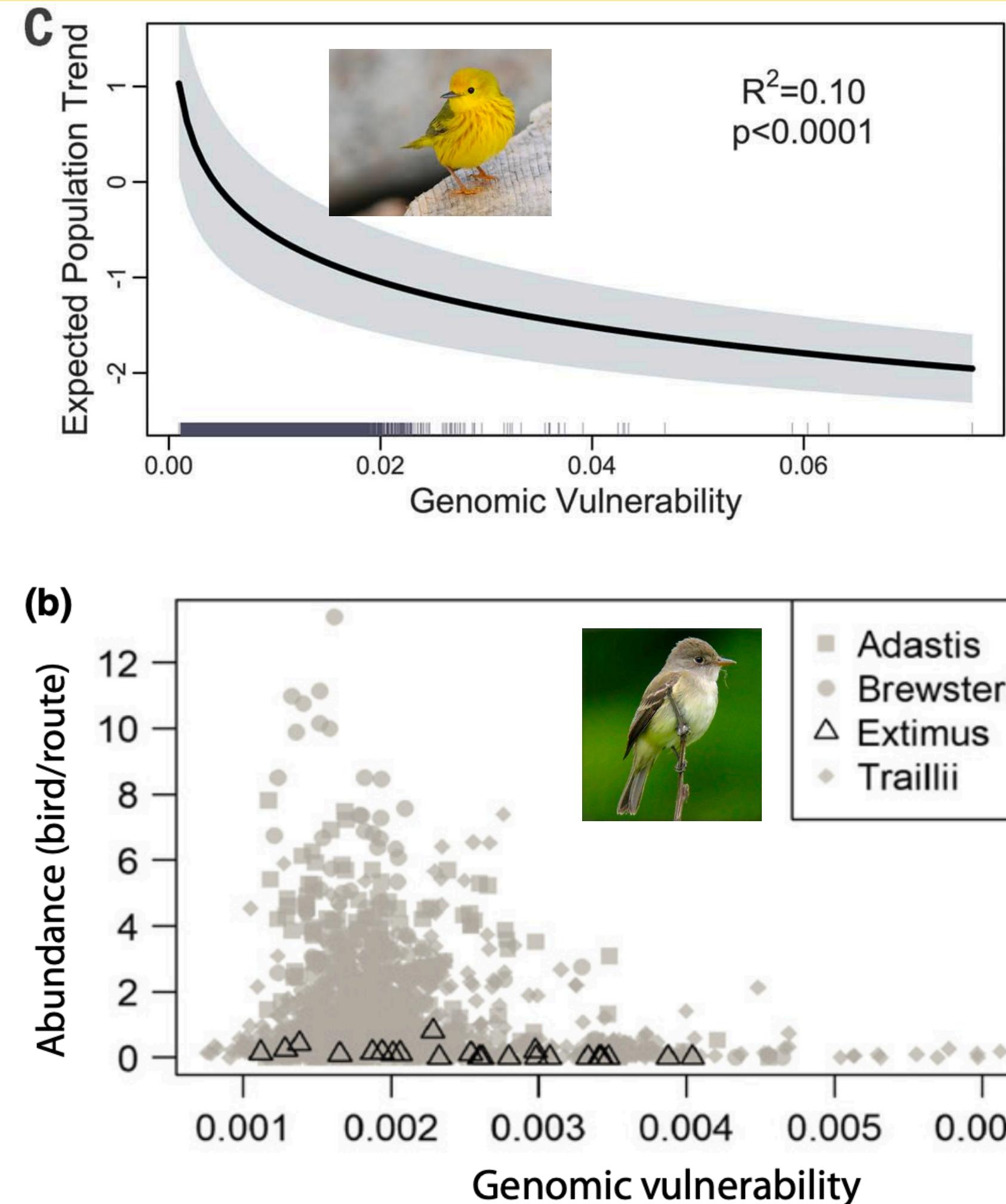
Could patterns between $G F_{\text{offset}}$ and population size arise because of genetic drift and not selection?



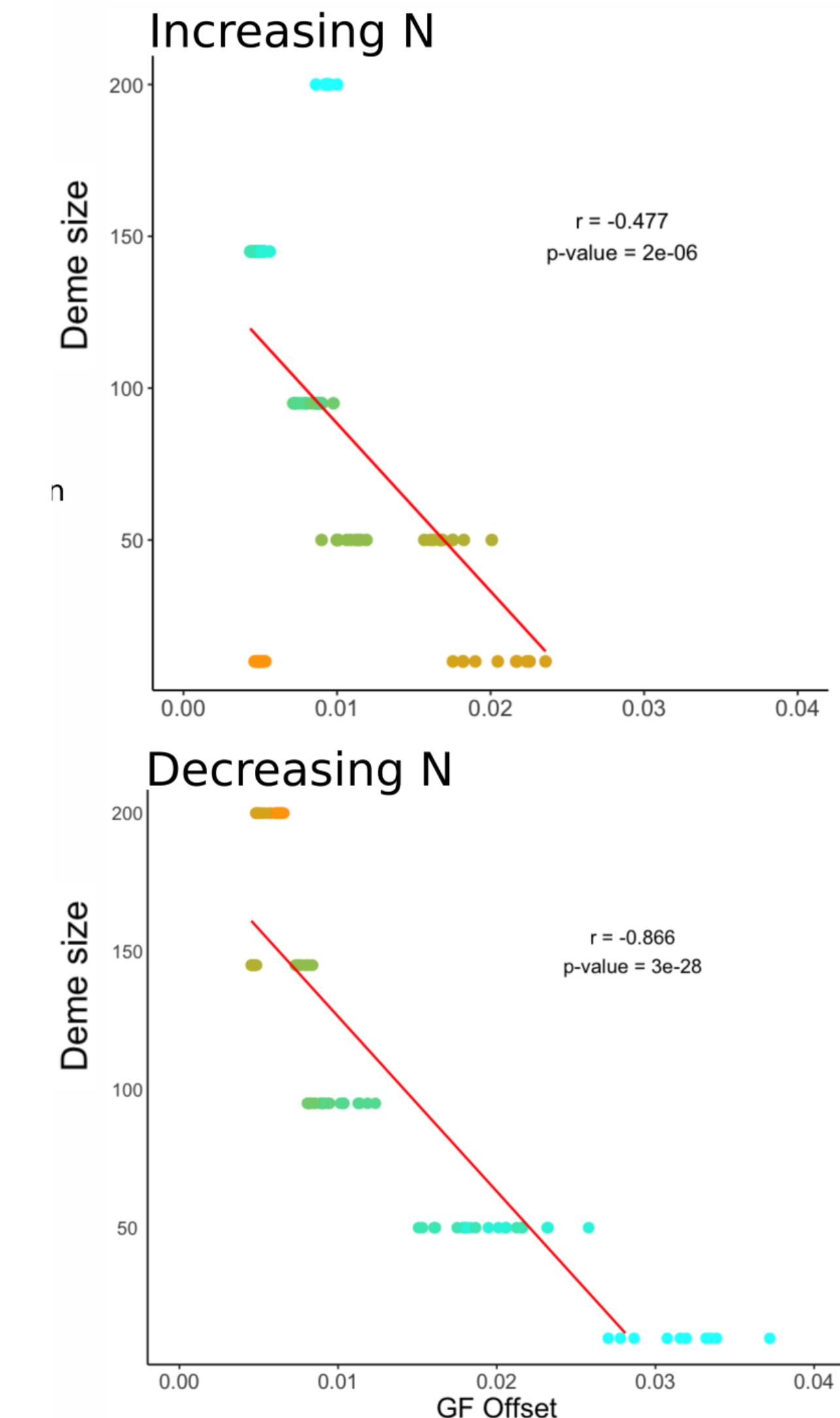


Neutral simulations

Could patterns between GF_{offset} and population size arise because of genetic drift and not selection?

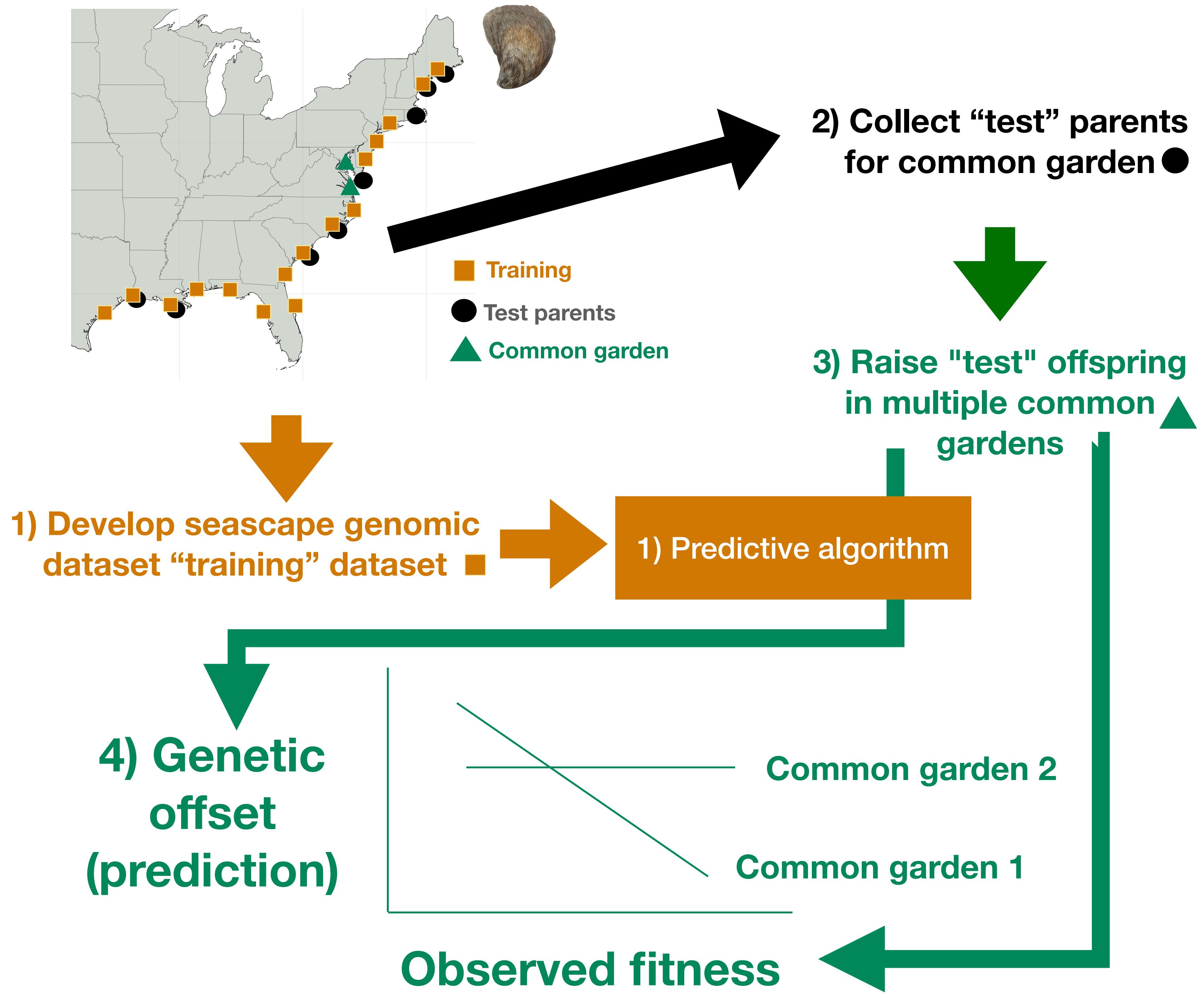


Yes - Empirical studies find exactly what we expect based on neutrality

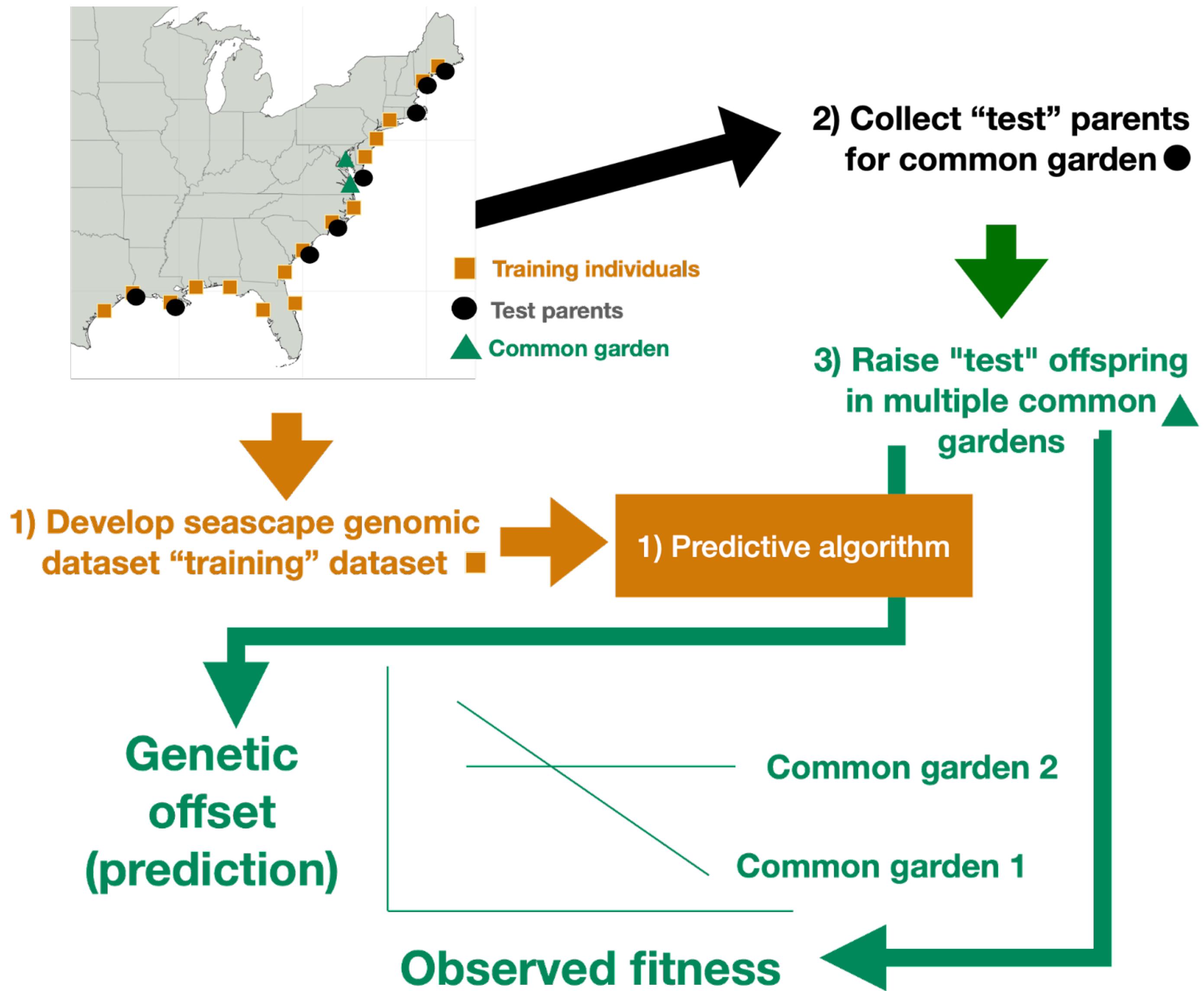


Required design elements for evaluation studies

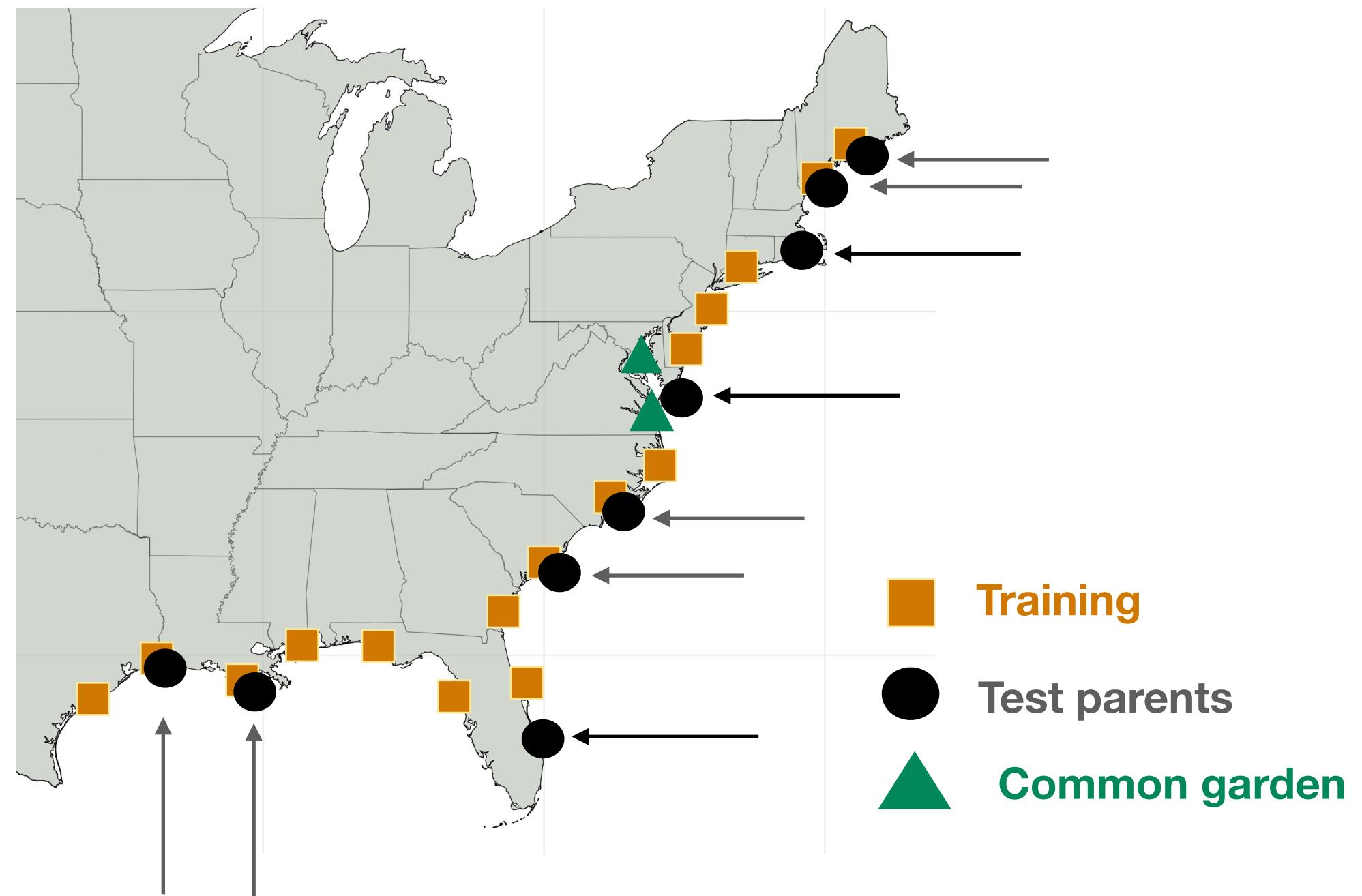
Recommended approach (simulations or experiment)



Required design element 1: Training and testing are independent phases.



Required design element 2: Training samples are different individuals from test samples.



- 1) Develop seascape genomic dataset “training” dataset
- 2) Collect “test” parents for common garden

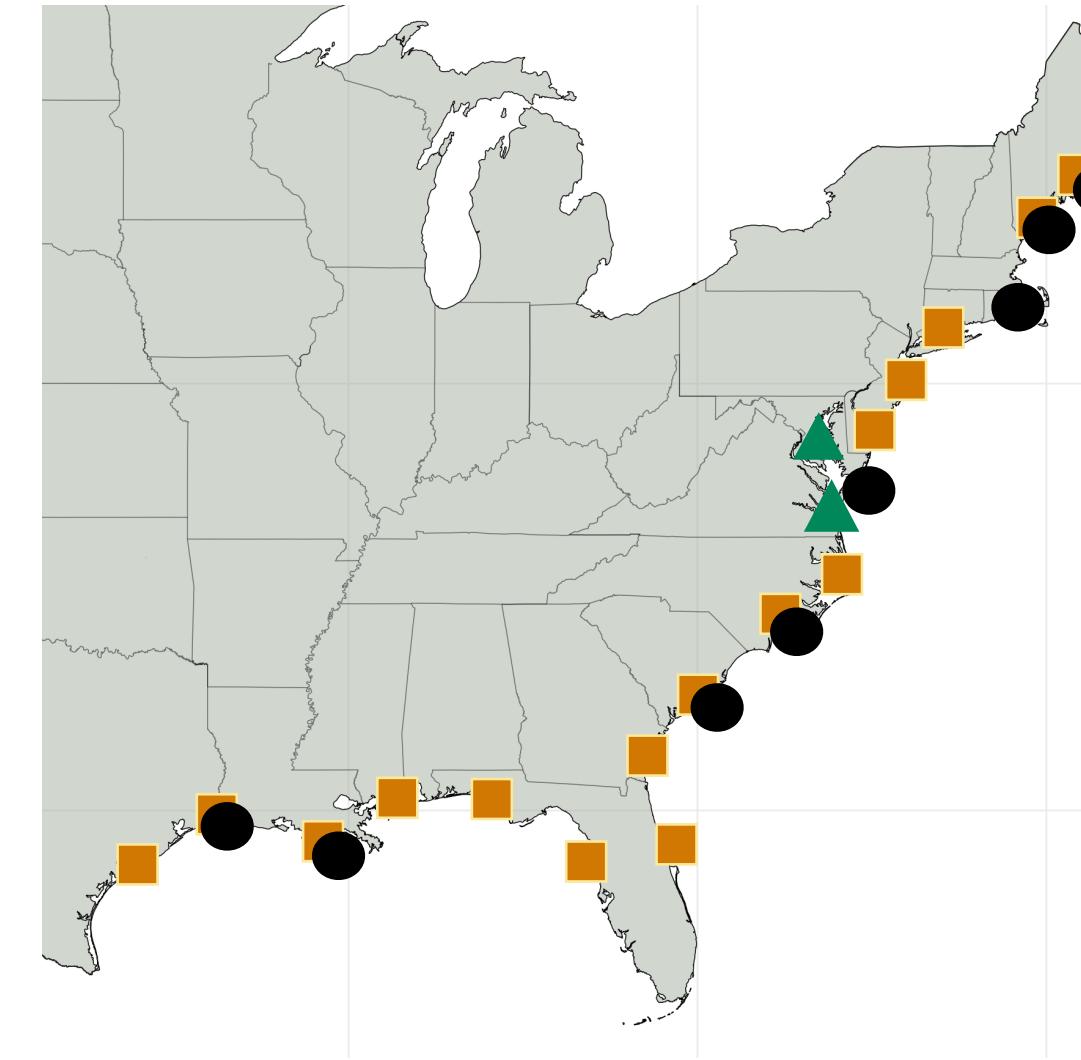
Design consideration:
Whether test individuals come from populations that were included in the set of training populations...
...or from populations that were not included in the set of training populations

Required design element 3:

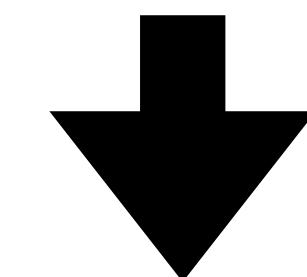
The multivariate environment is measured the same way in the test and training environments.

The novelty of the multivariate test environment(s) relative to the training environments is quantified.

Consideration:
Only quantify the test environment during testing



2) Collect "test" parents for common garden



3) Raise "test" offspring in multiple common gardens

Best practice

	Env 1	Env 2	Env 3	Env 4	Env 5
Training-1	19	12.1	1	6.5	33
Training-2	20	15.4	2	6.8	21
Test-1	21	9.2	3	5.3	24
Test-2	21	9.2	3	5.3	24

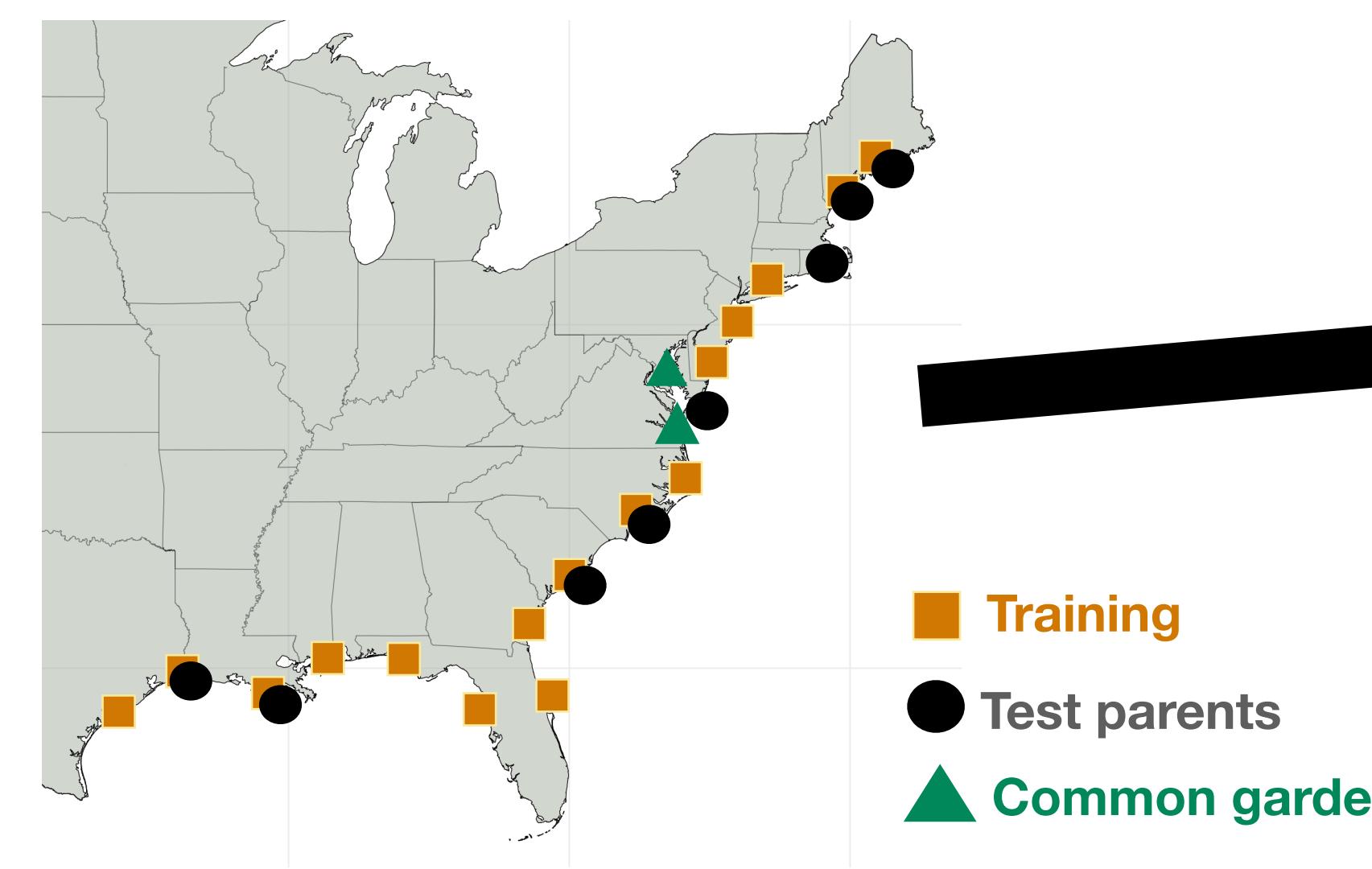
Not useful

	Env 1	Env 2	Env 3	Env 4	Env 5
Training-1	19	12.1	1		33
Training-2	20	15.4	2		21
Test-1	21		3	5.3	24
Test-2	21		3	5.3	24

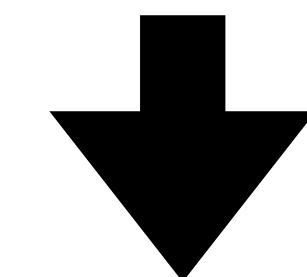
Required design element 3:

The multivariate environment is measured the same way in the test and training environments.

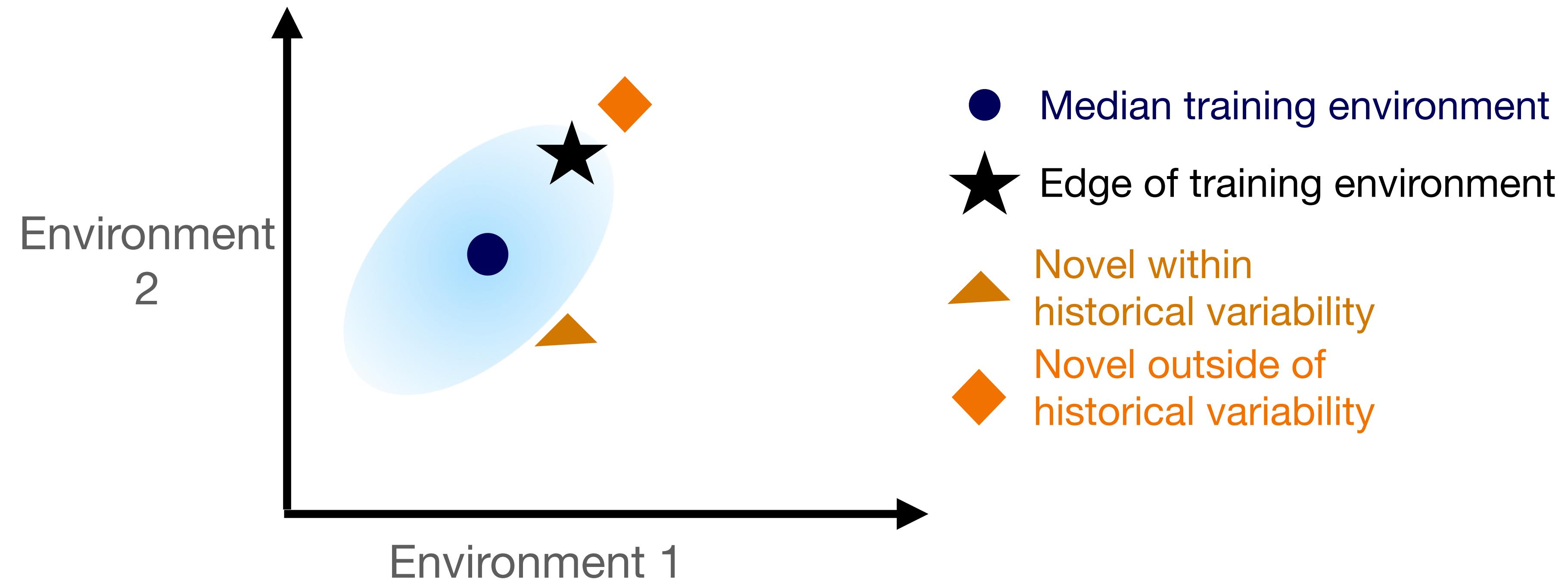
The novelty of the multivariate test environment(s) relative to the training environments is quantified.



2) Collect "test" parents for common garden

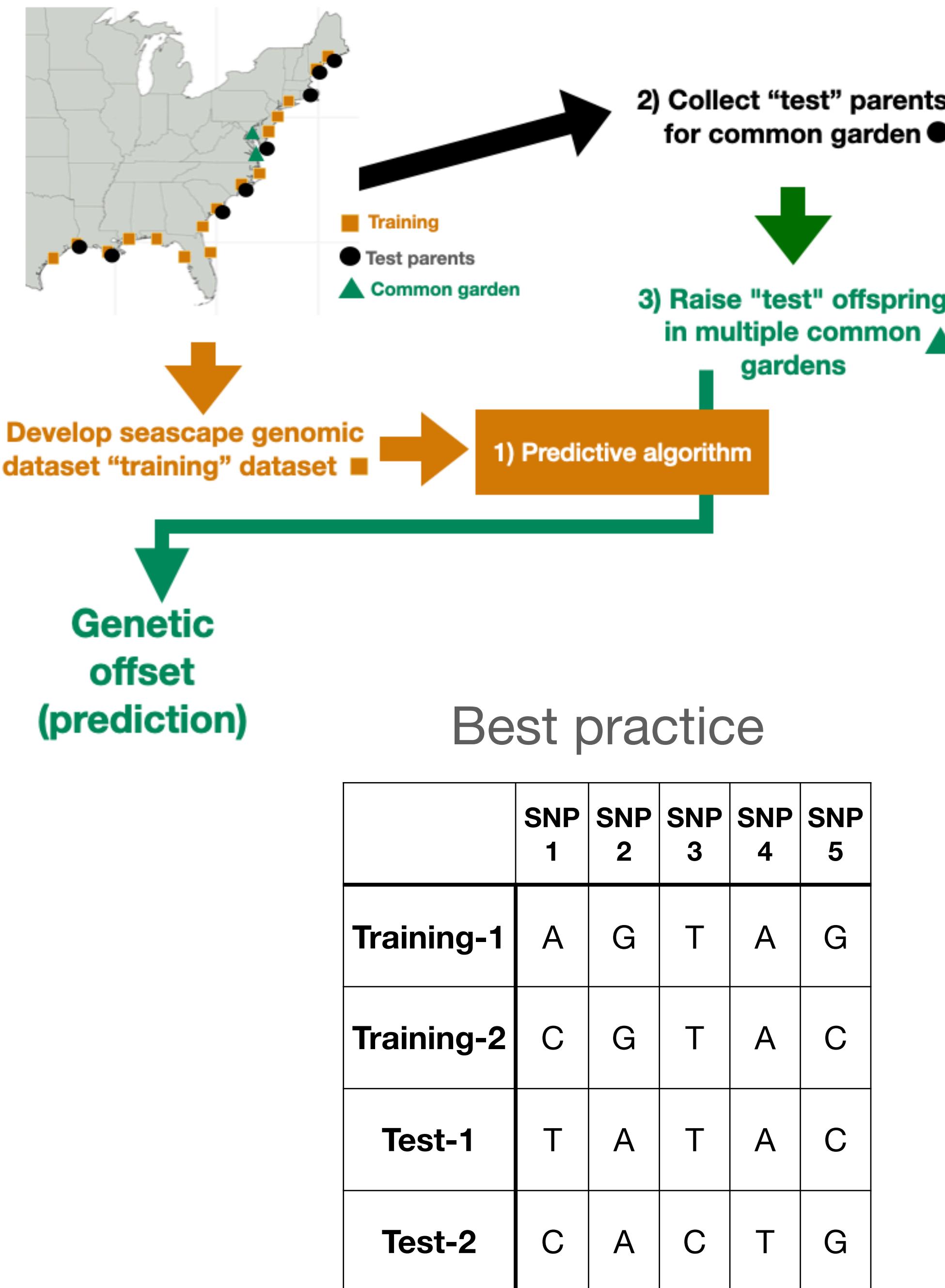


3) Raise "test" offspring in multiple common gardens

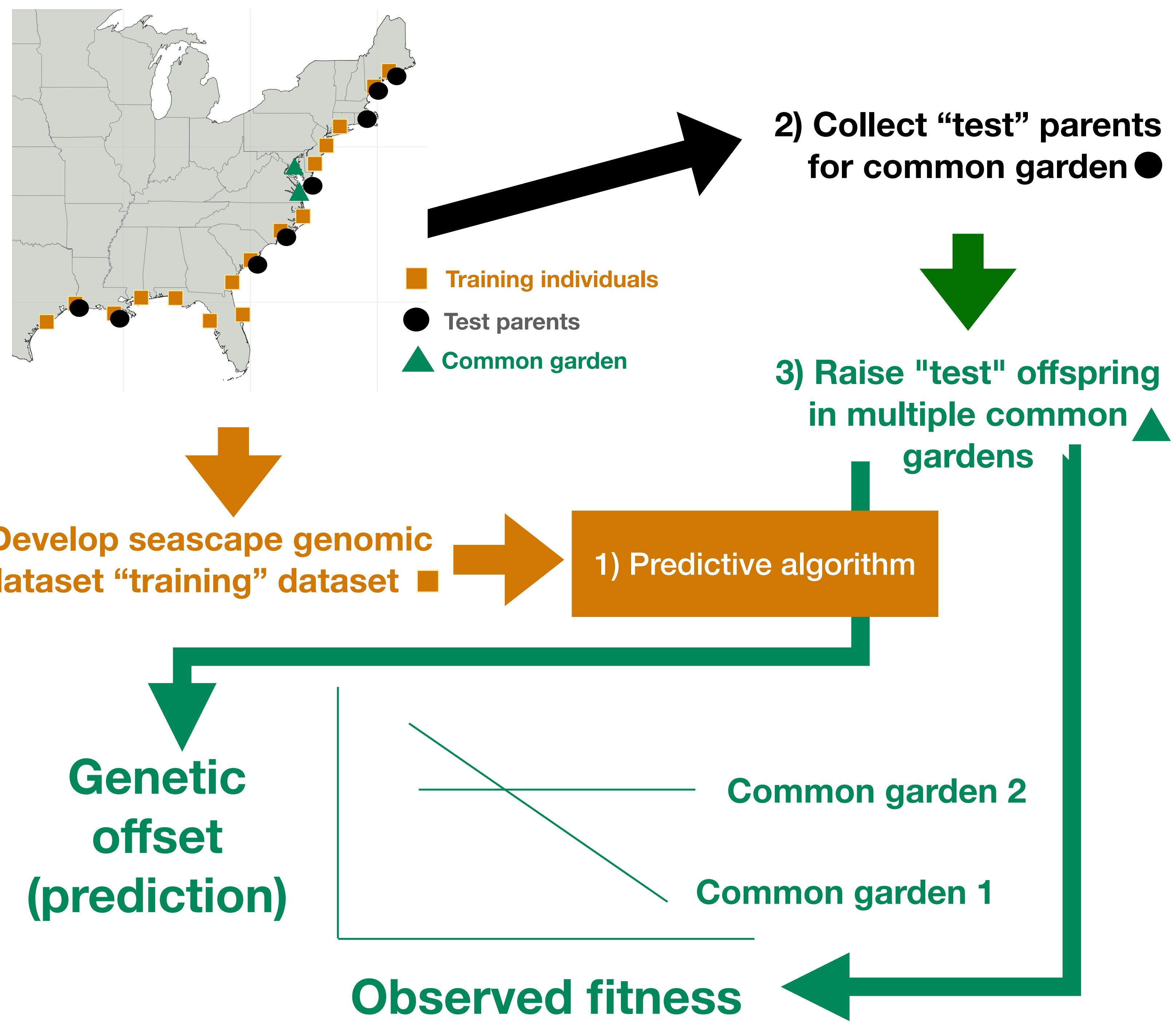


Required design element 4:

The same genetic markers are genotyped in training and test individuals.

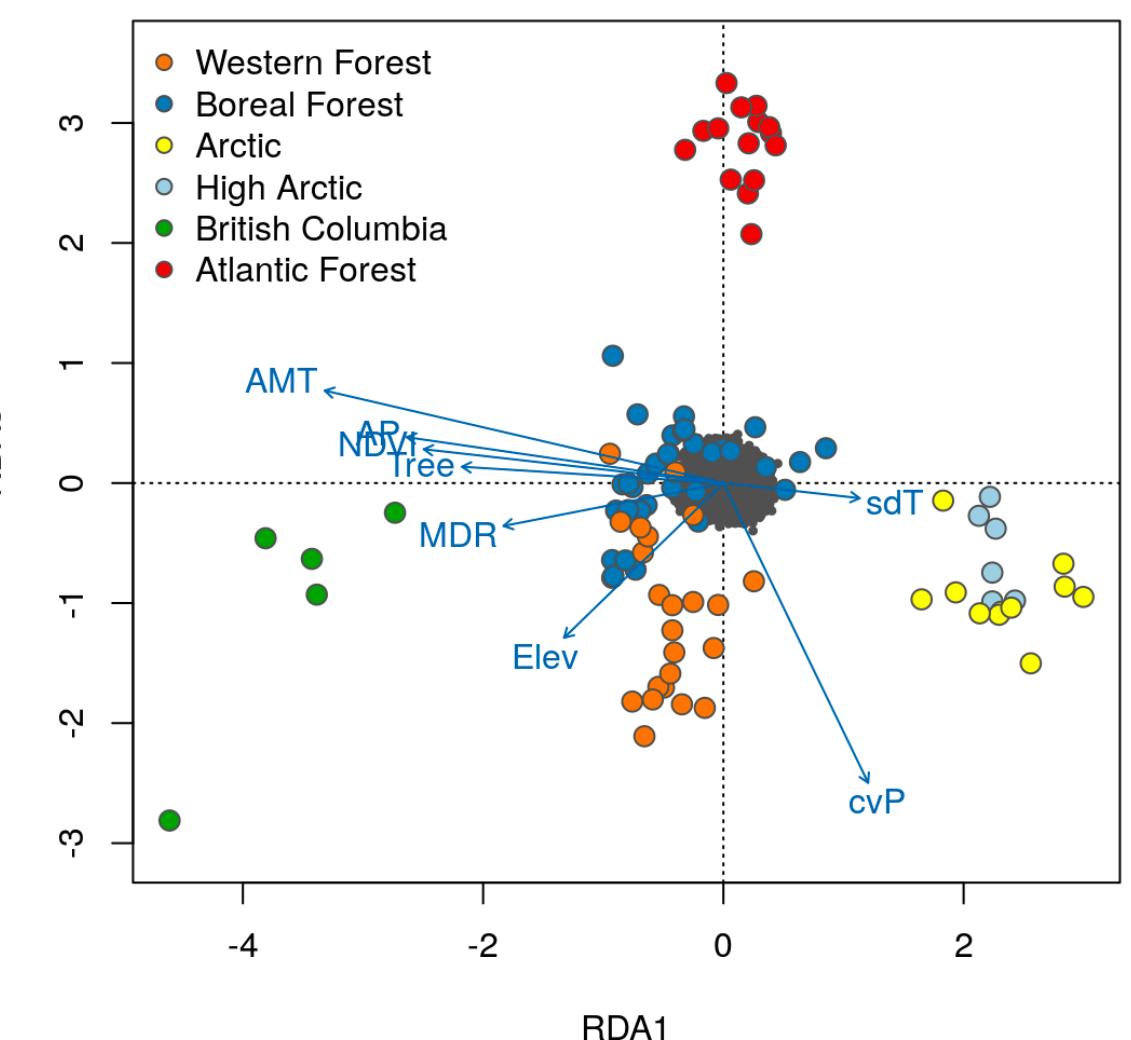


Required design element 5:
Multiple measures of fitness proxies or performance are quantified for each test individual and used as ground-truth in the evaluation

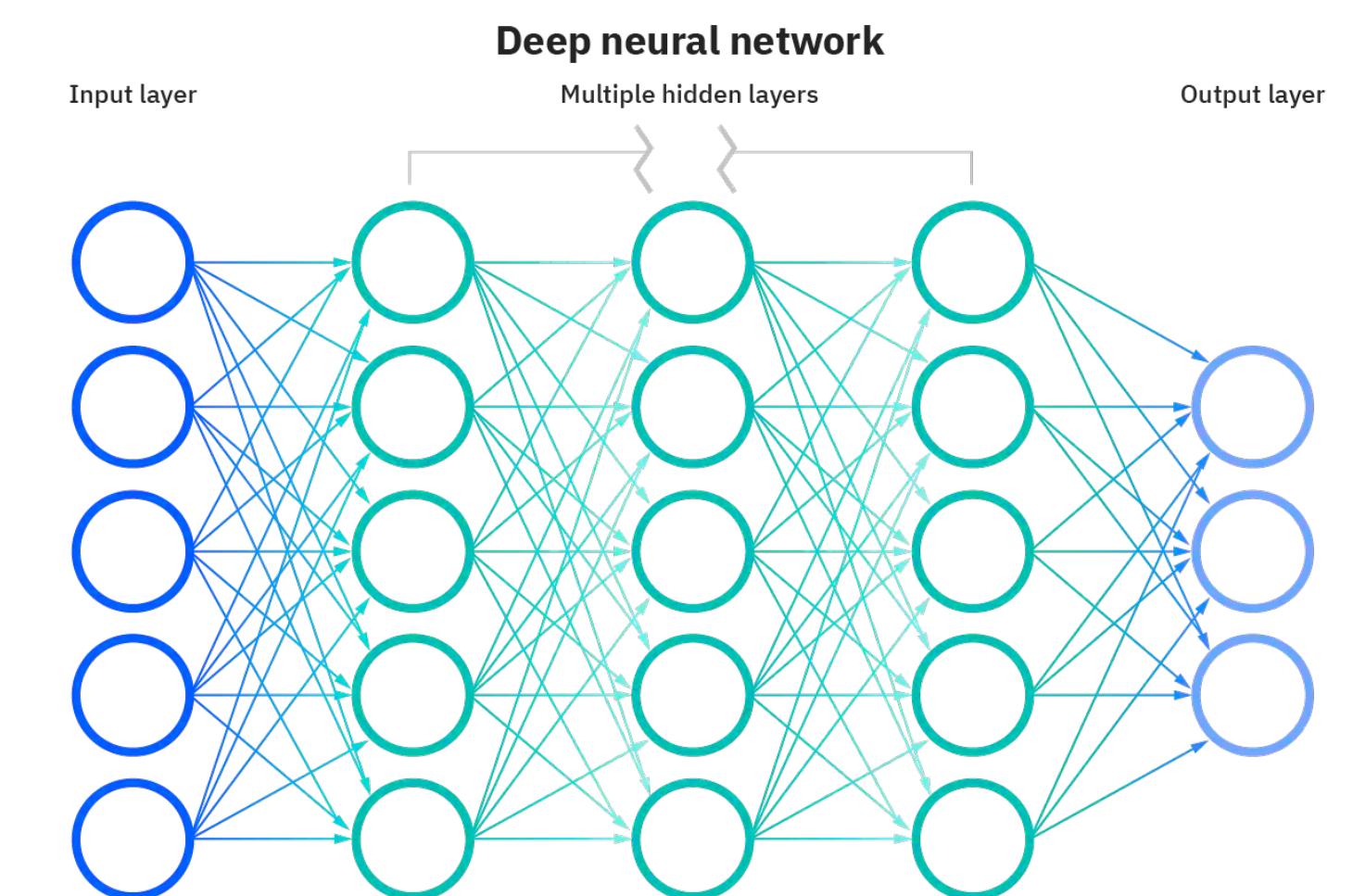


Required design element 6: Model training is clearly explained.

RDA:
Model training not required



Neural net:
Model training required

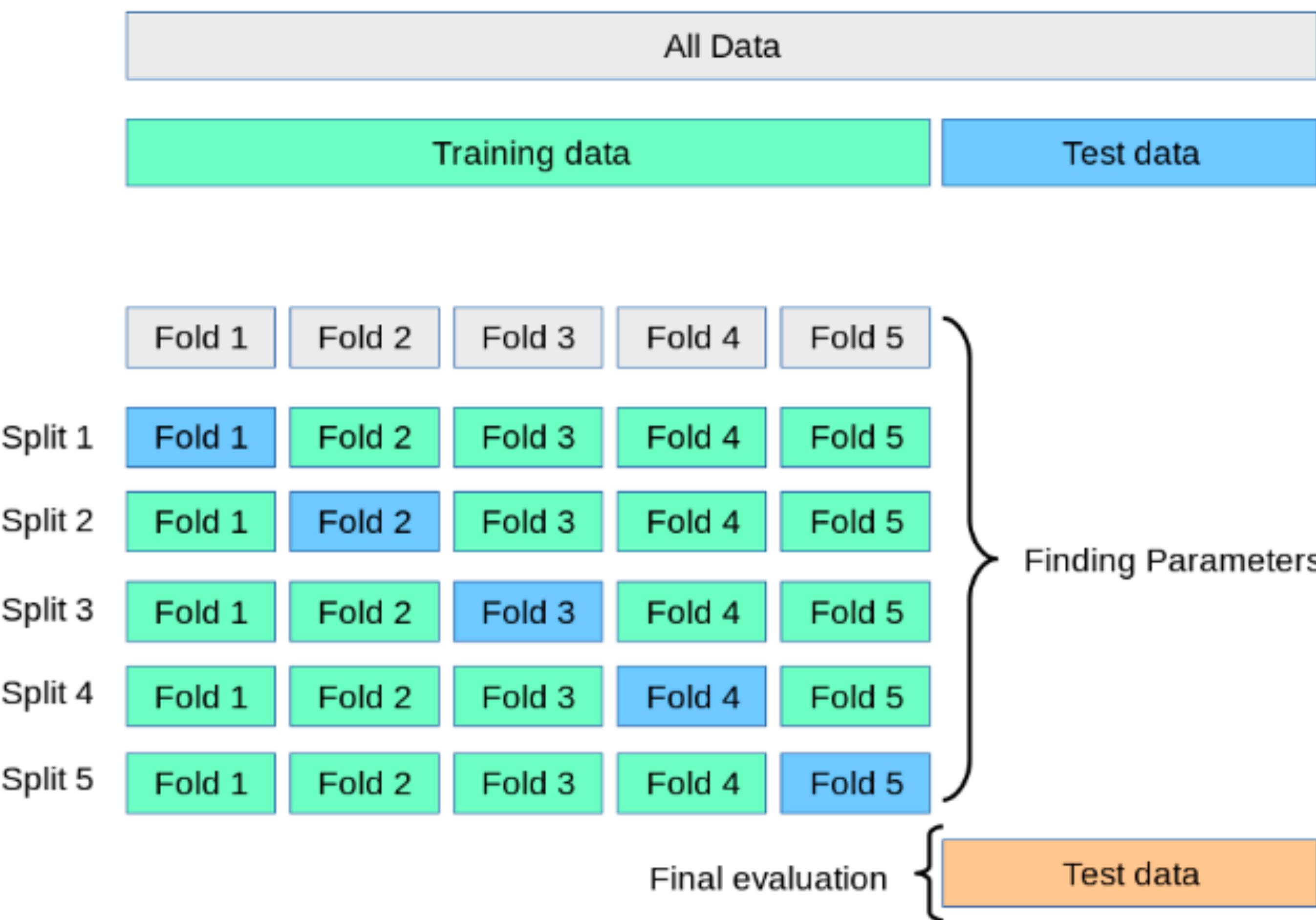


Brenna Forester - popgen.nescent.org

IBM.com

Use cross-validation for training
(Cross validation is a summary of model fit - not actual validation)

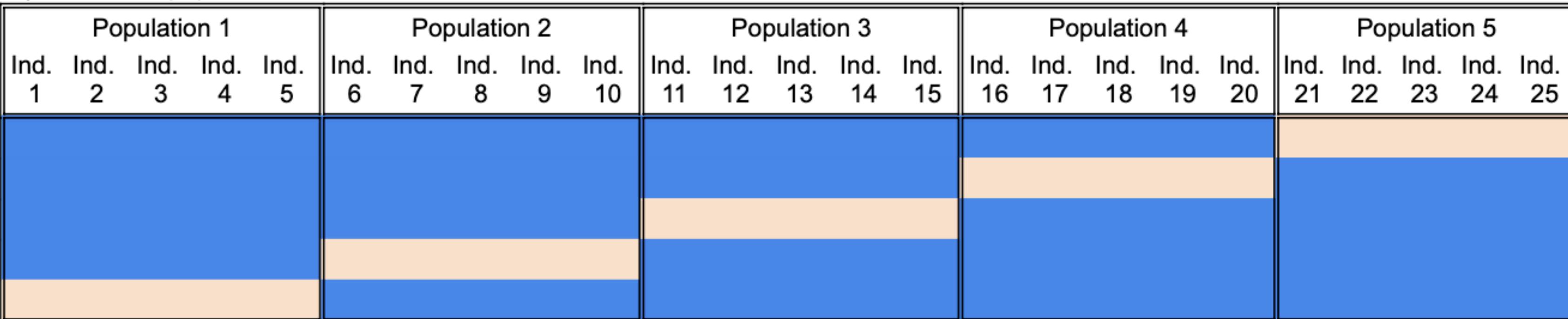
Required design element 6: Model training is clearly explained.



Required design element 6: Model training is clearly explained.

Use cross-validation for training
(Cross validation is a summary of model
fit - not actual validation)

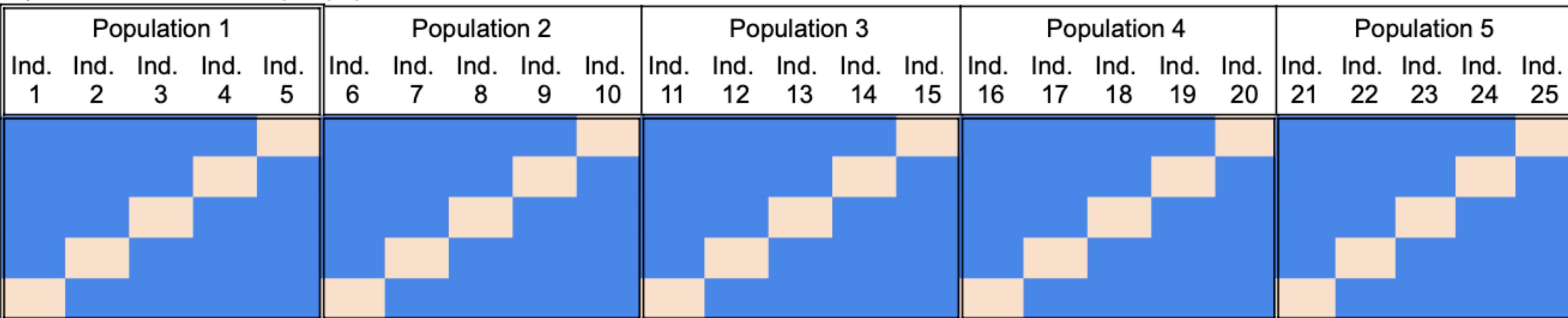
A) Leave-one-population-out cross-validation



- Split 1 accuracy
- Split 2 accuracy
- Split 3 accuracy
- Split 4 accuracy
- Split 5 accuracy

Average accuracy
reflects ability of
model to predict new
populations

B) Leave-one-individual-per-population-out cross-validation



- Split 1 accuracy
- Split 2 accuracy
- Split 3 accuracy
- Split 4 accuracy
- Split 5 accuracy

Average accuracy
reflects ability of
model to predict new
individuals from
training populations



Training Data



Cross-validation test data

Required design element 6: Model training is clearly explained.

A) Model training with cross validation

	Population 1					Population 2					Population 3					Population 4					Population 5					
	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	
Split 1																										
Split 2																										
Split 3																										
Split 4																										
Split 5																										

Parameter set 1:
least accurate

	Population 1					Population 2					Population 3					Population 4					Population 5					
	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	
Split 1																										
Split 2																										
Split 3																										
Split 4																										
Split 5																										

Parameter set 2:
somewhat
accurate

	Population 1					Population 2					Population 3					Population 4					Population 5					
	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.	
Split 1																										
Split 2																										
Split 3																										
Split 4																										
Split 5																										

Parameter set n :
most accurate

B) Model testing.
Apply
parameterized
model to test data

Test data					
Pop 4		Pop 5		Pop 6	
Ind	Ind	Ind	Ind	Ind	Ind
26	27	28	29	30	31



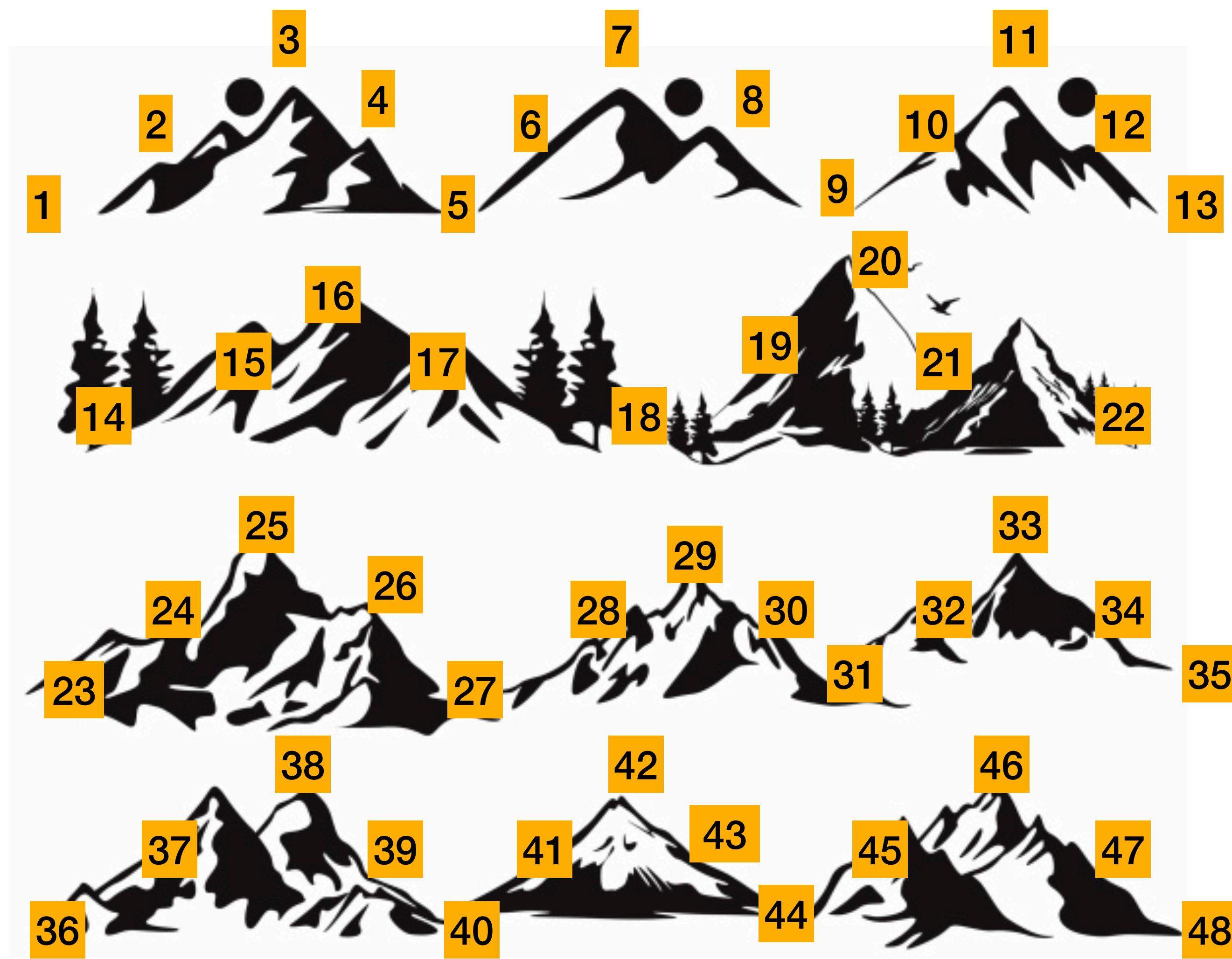
Dahu challenge

Dahu challenge:

11 mountain peaks to possibly sample

Latitude

Longitude



Summit Dahu
Really unequal legs

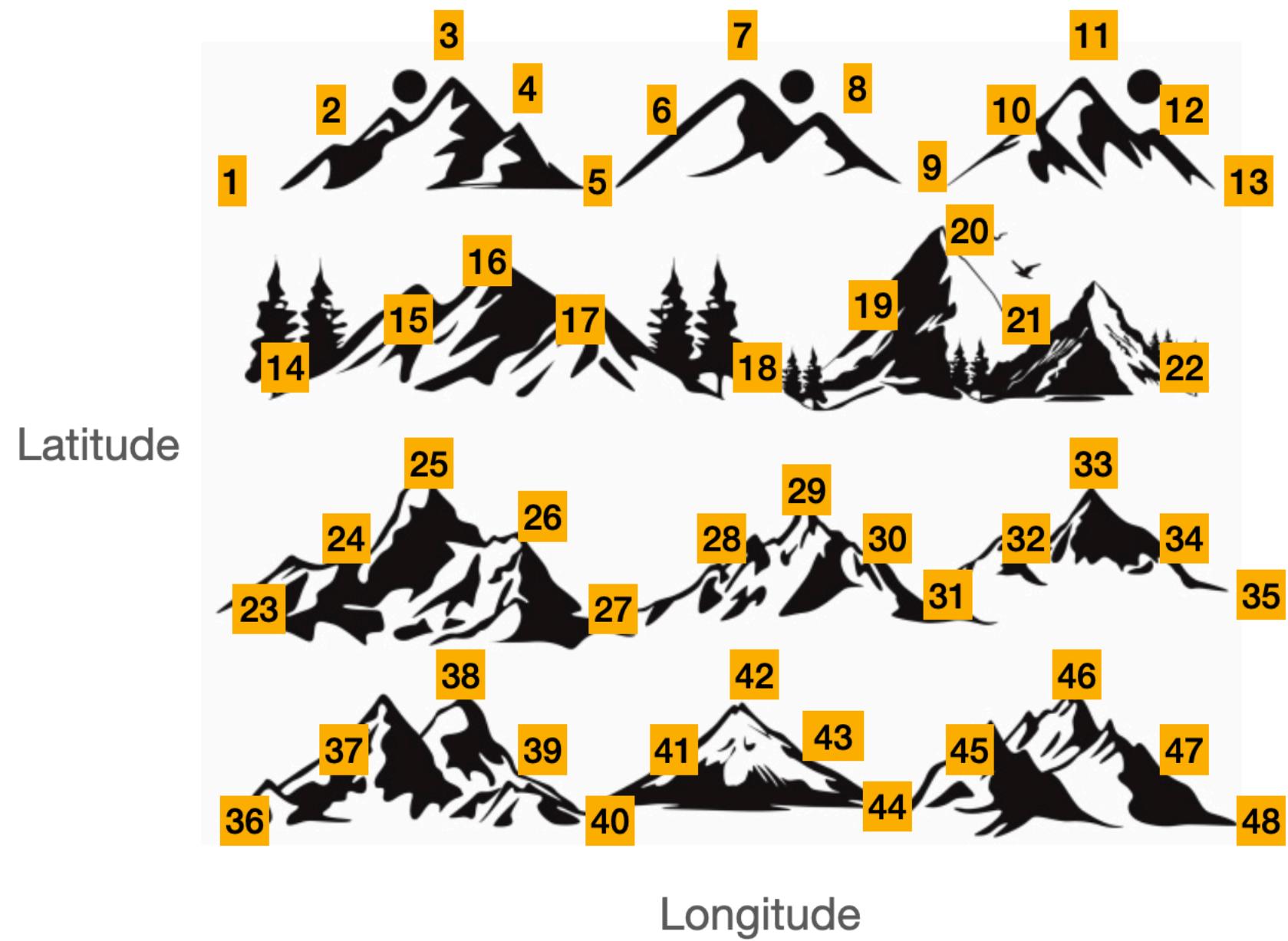


Slope Dahu
Somewhat unequal legs



Valley Dahu
Equal legs





You have the budget to sequence:

1000 individuals for SNP genotype

OR

75 populations for Pool-Seq - population-level allele frequency

Your challenge: Propose an experimental design for a genomic offset evaluation experiment. Link to share your design: https://docs.google.com/presentation/d/1CRL2wzds_InKQVko1zhd8jCTXQI-O5JNwDSHLiOHaxQ/edit#slide=id.p

Guiding questions:

Q1: Which populations do you sample for training individuals and test individuals? How do you divide the sampling for training and testing?

Q2: Where would you plan the test common garden(s)? How do you divide the test individuals among the common gardens?

Dahu challenge:

11 mountain peaks to possibly sample