

# Landscape Genomics

Amy Zyck

February 7, 2019

# Outline

- Genetic Drift vs. Natural Selection
- Detecting Adaptive Evolution
- Local Adaptation
- Detecting signs of natural selection and genes involved in local adaptation
  - Landscape genomics
    - Environmental association analysis

# Population Genetics

- Changes in allele frequencies within & between populations are due to:

Deterministic forces



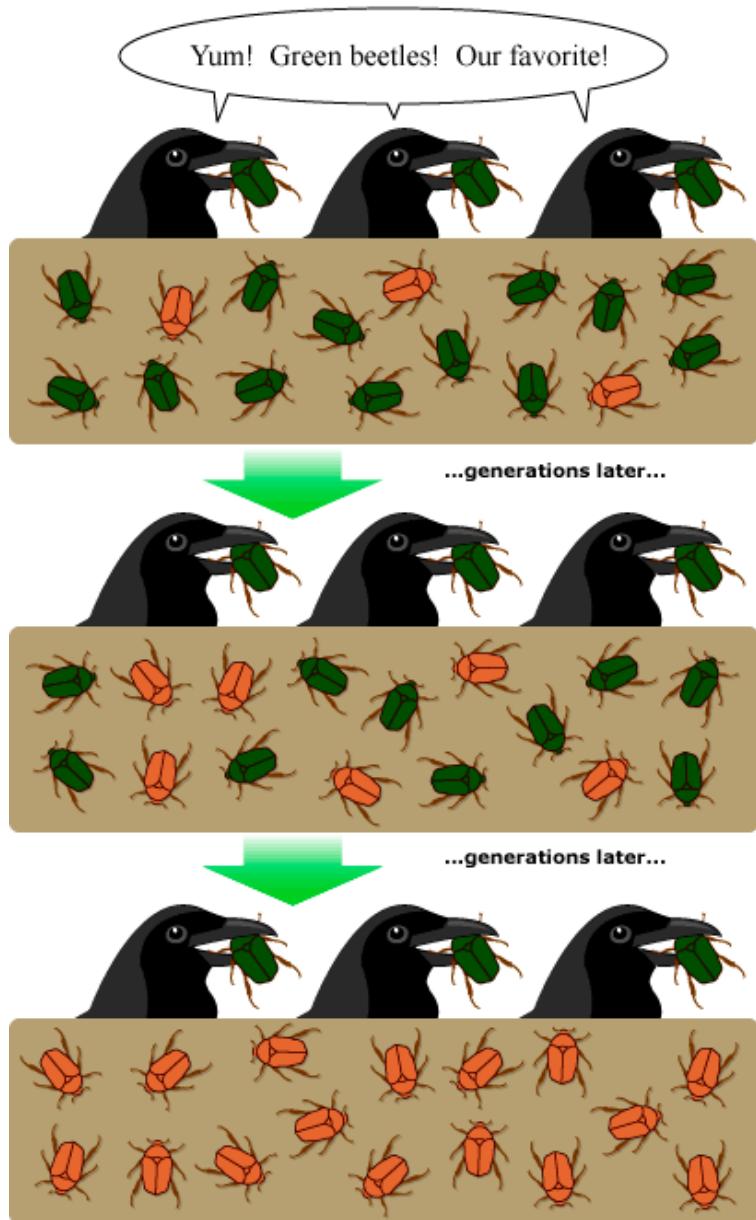
natural selection

Random forces



genetic drift

## Natural Selection



## Genetic Drift



# Population Genetics

- Changes in allele frequencies within & between populations are due to:

Deterministic forces



natural selection

Random forces



genetic drift

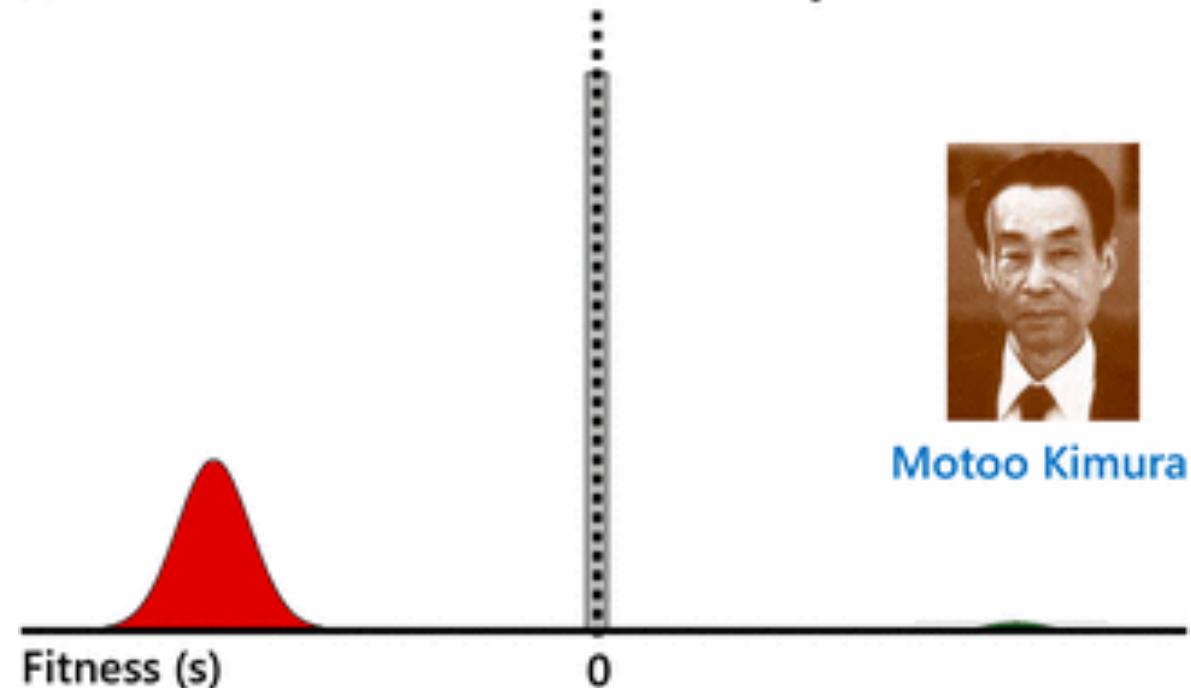
- **Neutral Theory**

(Kimura 1968)

## Neutral Model

- Most of the variation within and between species is caused by genetic drift of mutant alleles that are neutral
- A neutral mutation does not affect an organism's ability to survive and reproduce

### A – 1960s, Kimura's Neutral Theory

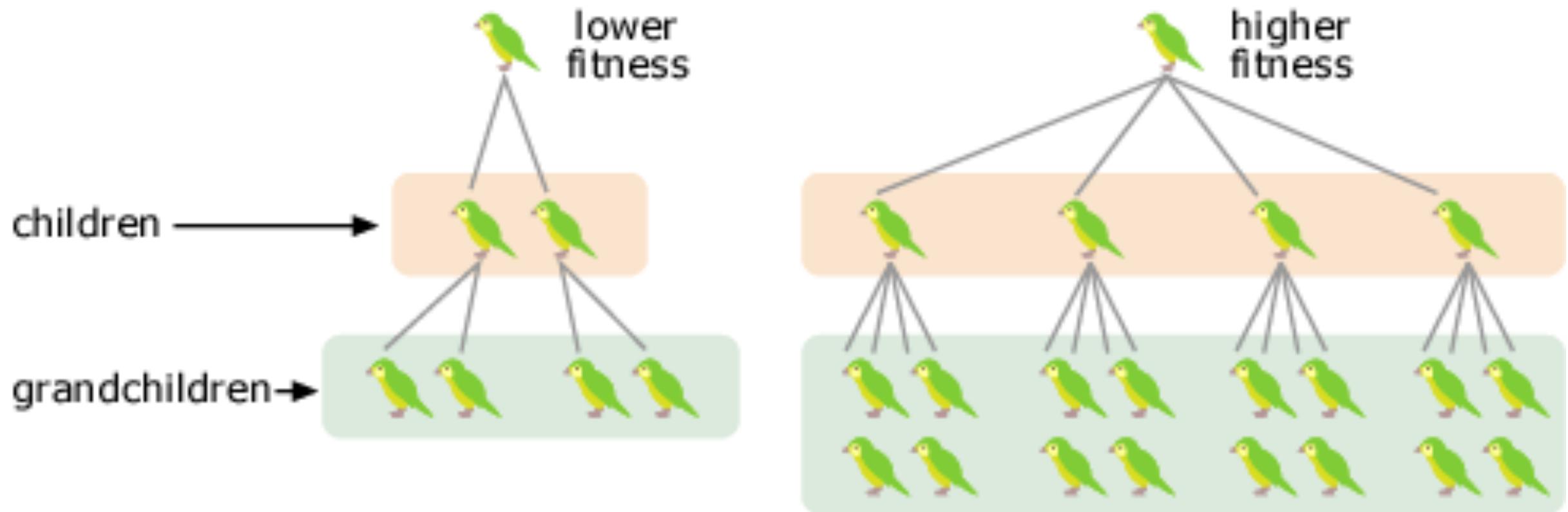


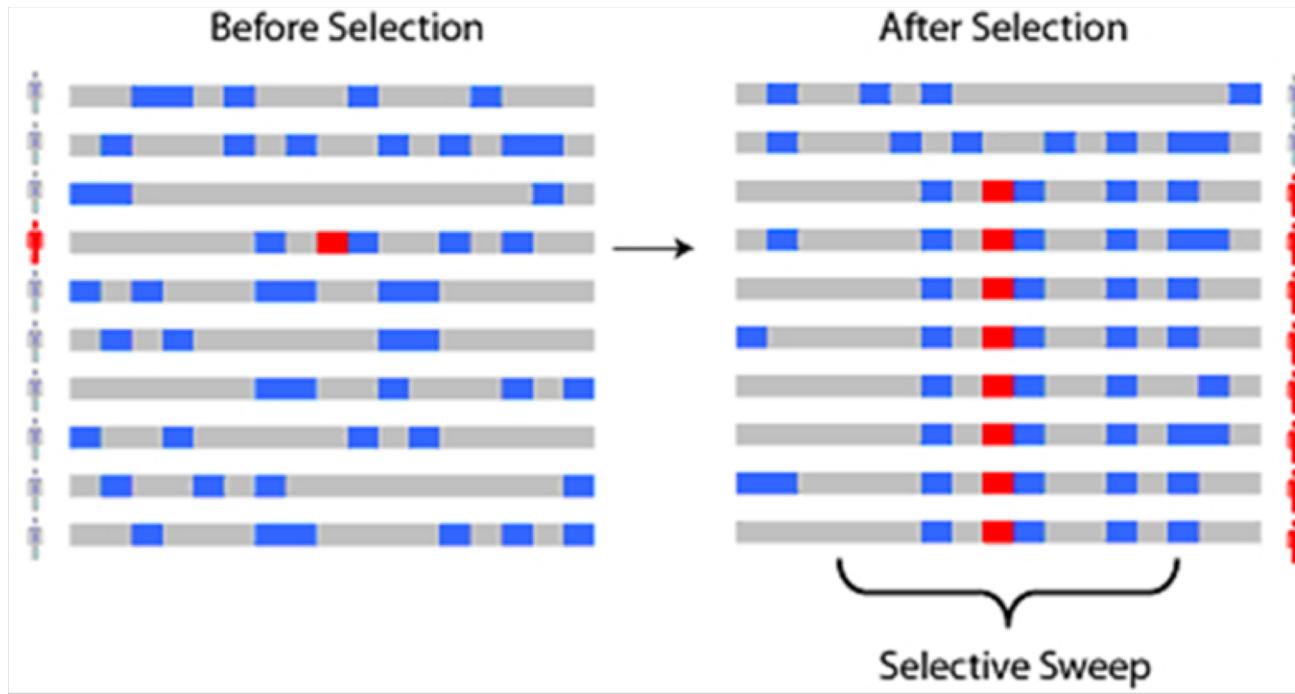
Motoo Kimura

Casillas & Barbadilla (2017) *Genetics*

# Natural selection

Differences in *fitness* among individuals → differential reproduction and survival of genotypes



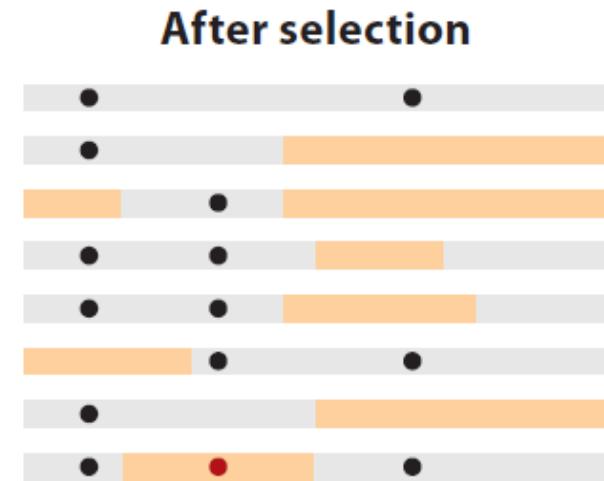
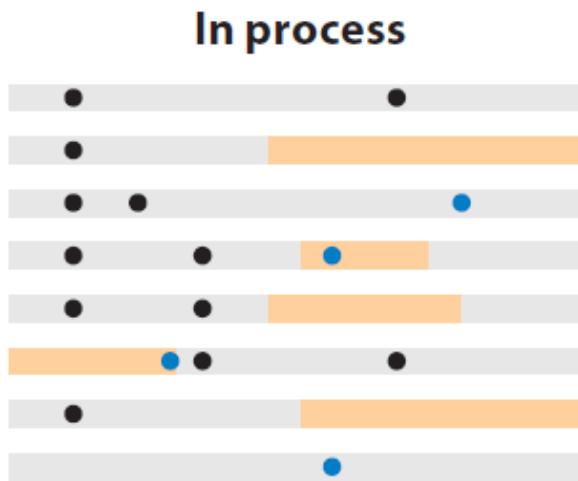
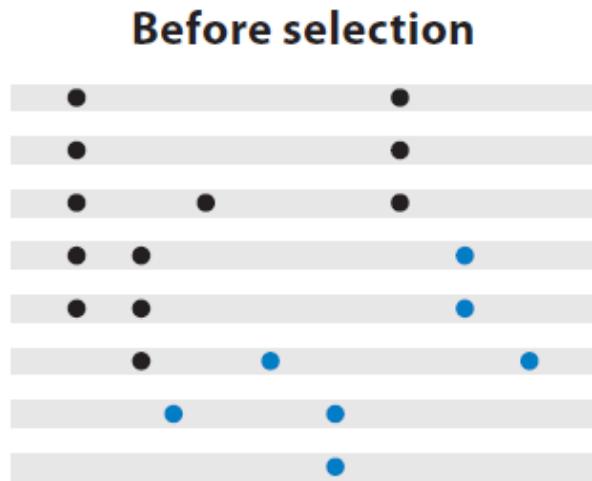


Schaffner & Sabeti (2008) *Nature*

Positive selection

Selection for advantageous variants

#### d Purifying selection



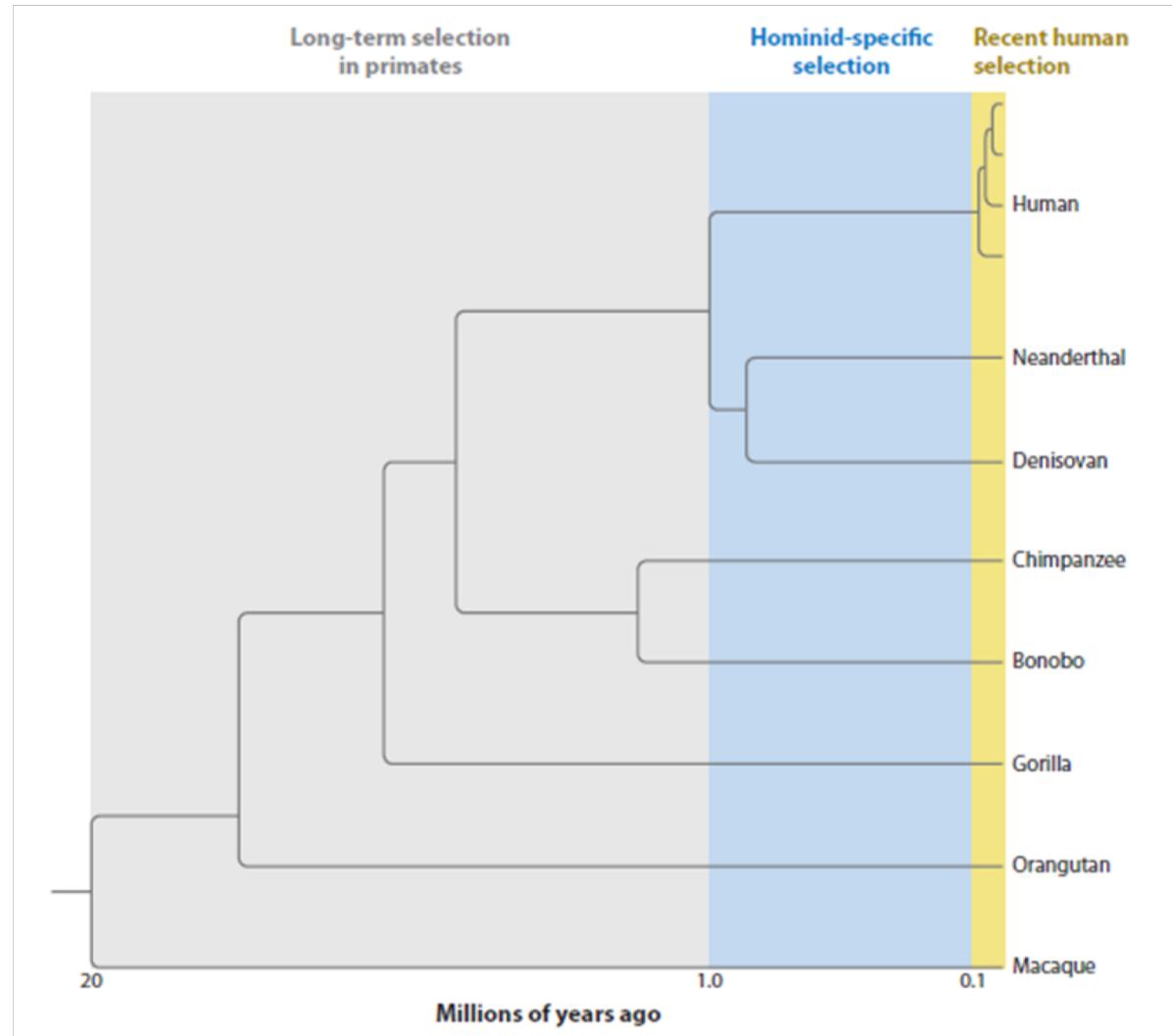
Fu & Akey (2013) *Annu. Rev. Genomics Hum. Genet.*

Purifying selection

Selection against deleterious variants

# Adaptive Evolution

- Quantifying adaptive evolution can provide insights into the genetic basis of human-specific phenotypes and provide a window into human history
- Interspecific divergence
- Intraspecific polymorphisms



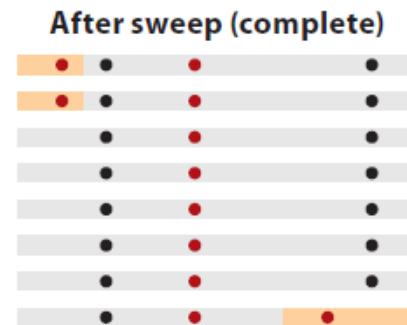
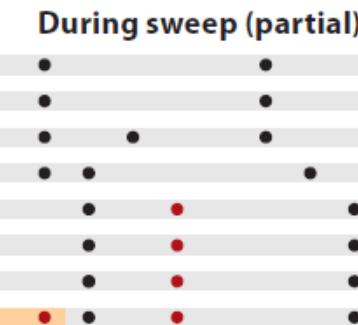
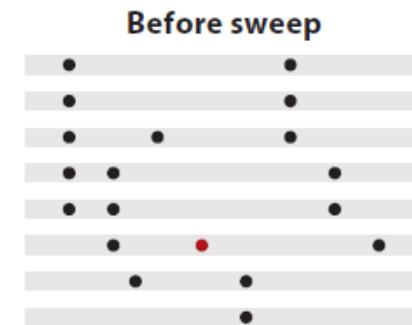
# Intraspecific Polymorphism

Classic selection sweeps

## Classic Selective Sweep (Hard Sweep)

- Rapid fixation of a favorable mutation
  - Hitchhiking effect

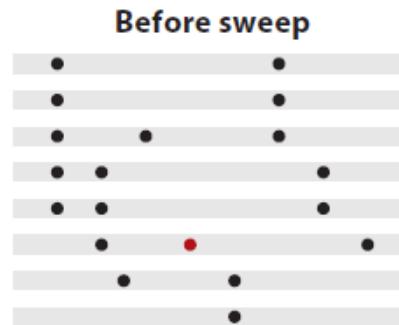
**a Hard sweep**



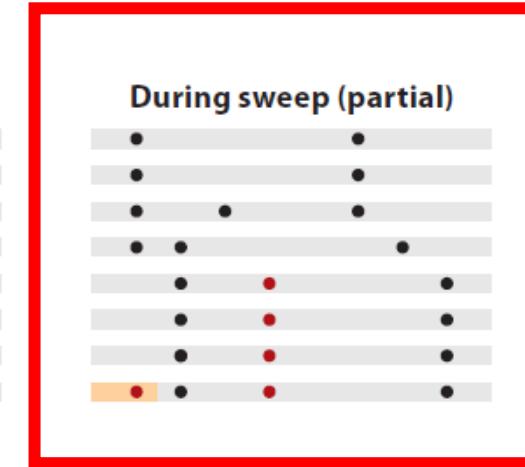
## Partial Sweep

- Rapid increase of advantageous mutation that has not yet reached fixation

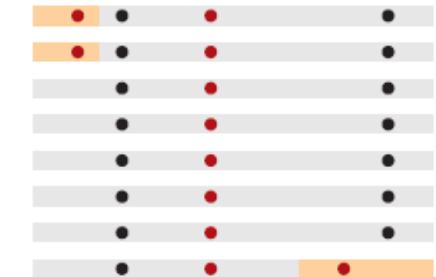
### a Hard sweep



### During sweep (partial)



### After sweep (complete)



# Nonclassic sweeps

Selection on Standing Variation

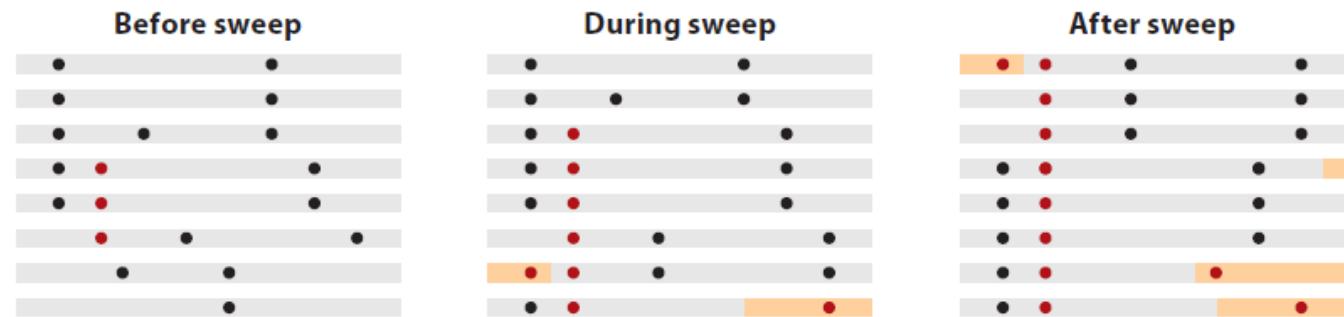
# Soft Sweeps

Selection on standing variation

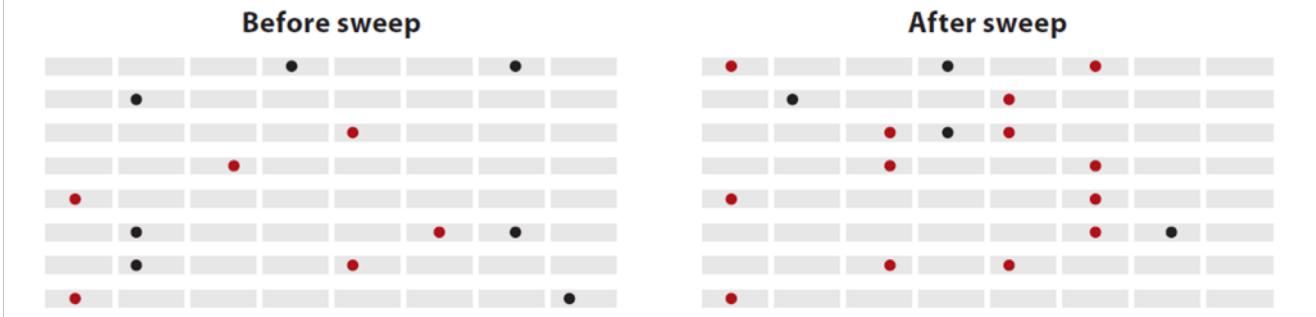
OR

Selection acting simultaneously on multiple independent loci (polygenic selection)

**b Positive selection on standing variation**



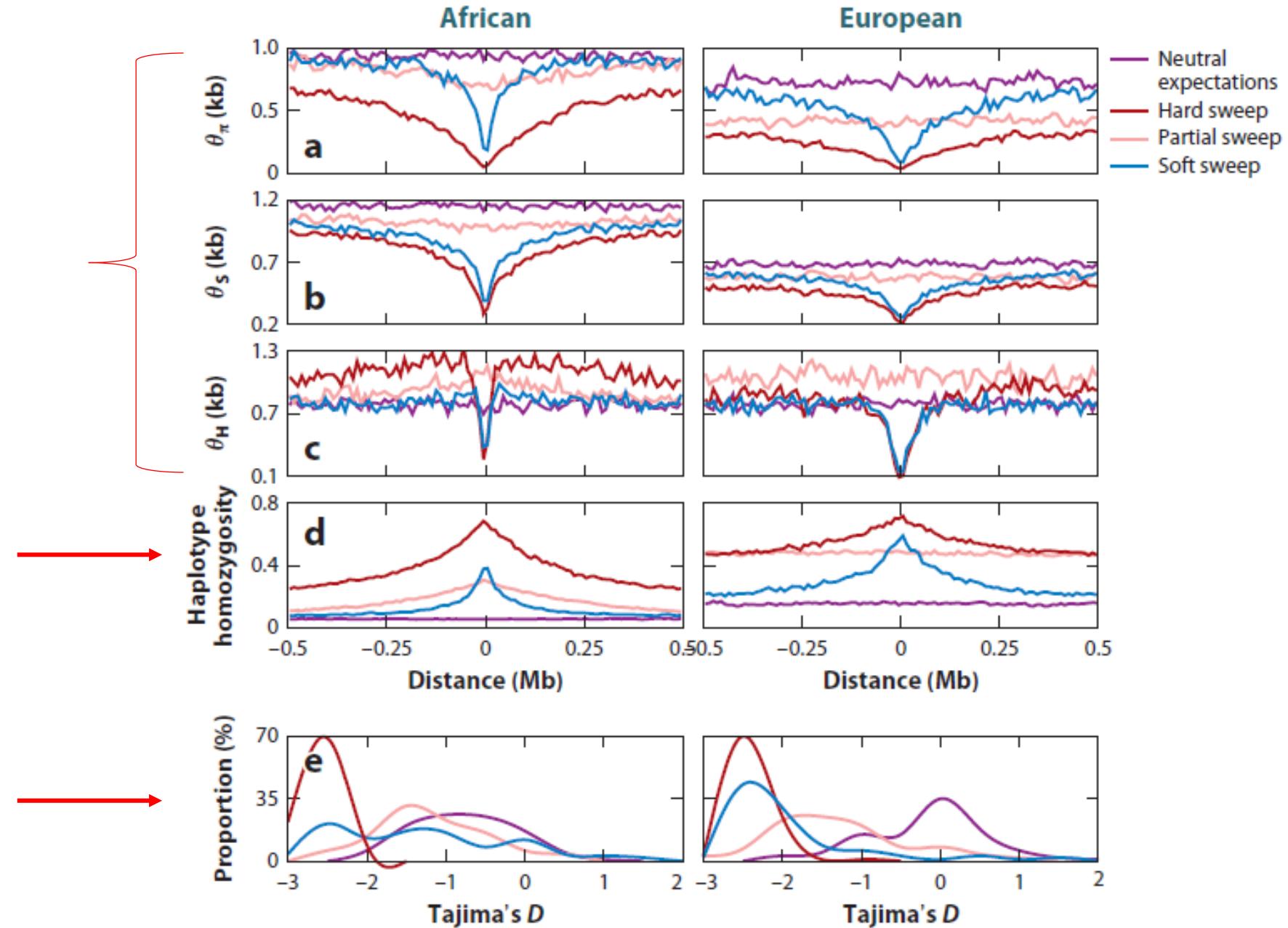
**c Polygenic selection (adaptation)**



Greater reduction of genetic diversity in hard sweeps

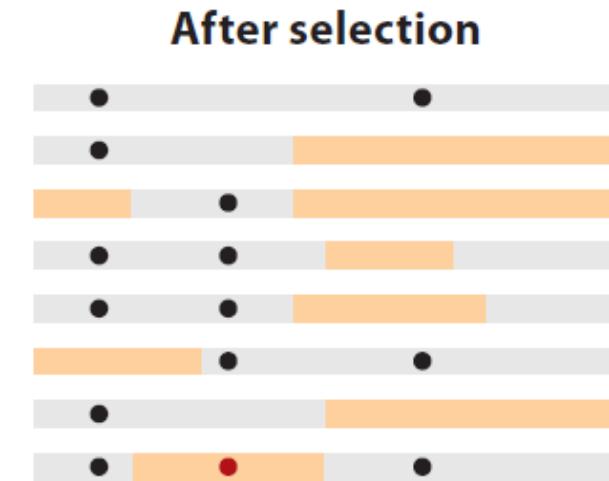
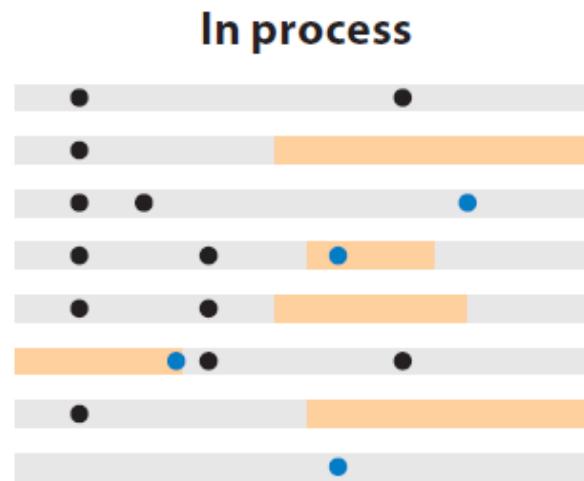
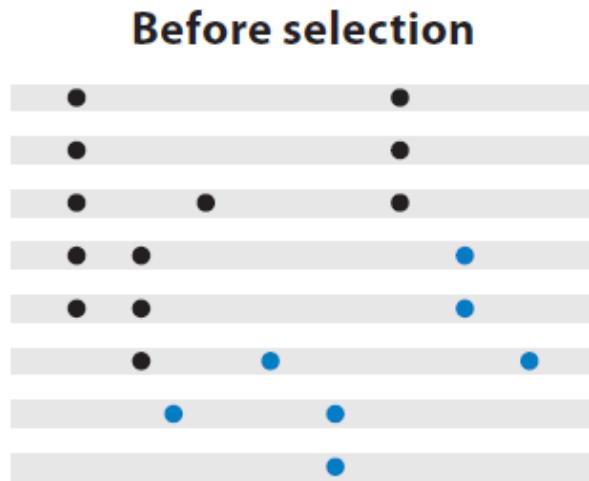
Weaker levels of linkage disequilibrium in soft sweeps

Greater variance in allele frequency distribution in soft sweeps



# Deleterious Variants and Background Selection

#### d Purifying selection



Fu & Akey (2013) *Annu. Rev. Genomics Hum. Genet.*

Purifying selection

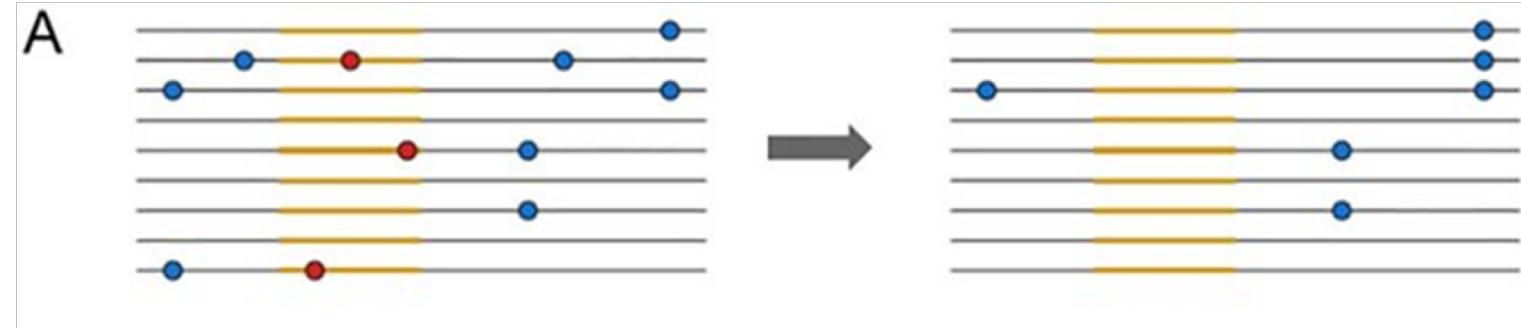
Selection against deleterious variants

## Background selection

Deleterious mutations drift up to low frequencies and are then purged from the population, with an occasional removal of linked variation



Reduction in diversity

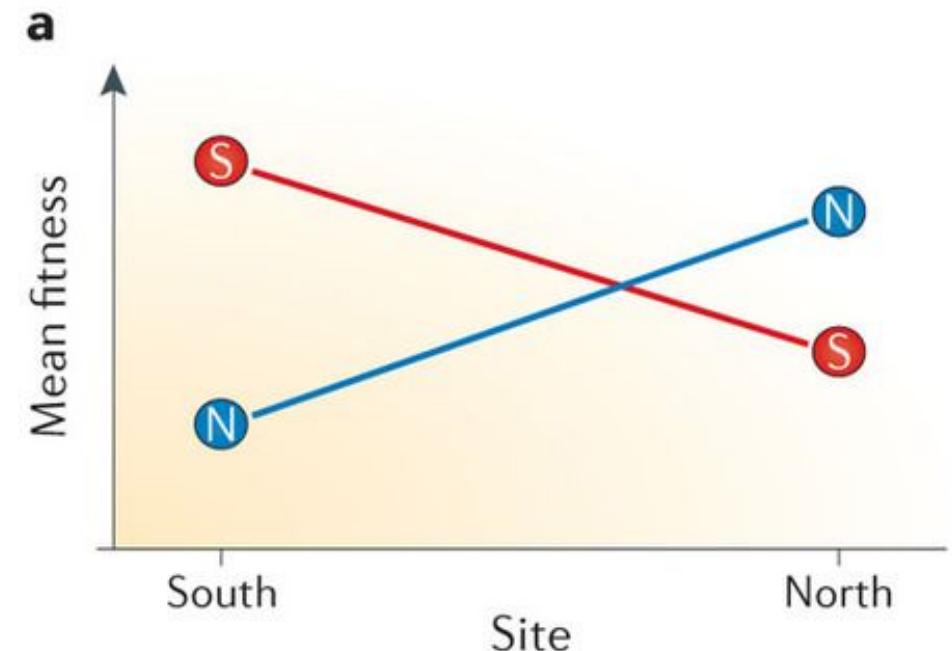


Understanding how natural selection generates and maintains adaptive genetic diversity in heterogeneous environments is key to predicting the evolutionary response of populations to rapid environmental change

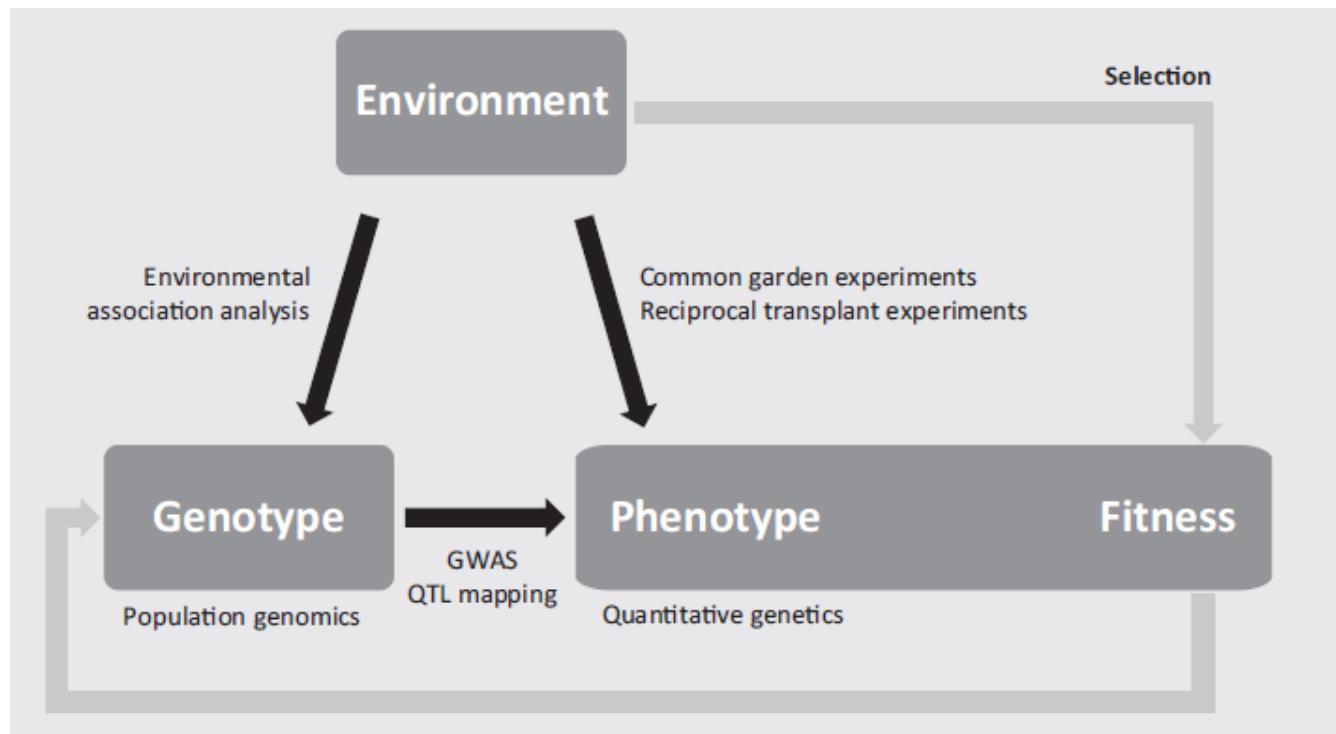


# Local adaptation

- Differential pressures of natural selection
- Interested in *genes underlying heritable phenotypic variation*
  - To identify these genes, we use:
    - Top-down approaches
      - Genome-wide association studies (GWAS)
      - Quantitative trait locus (QTL) mapping
    - Bottom-up approaches
      - Population & landscape genomics



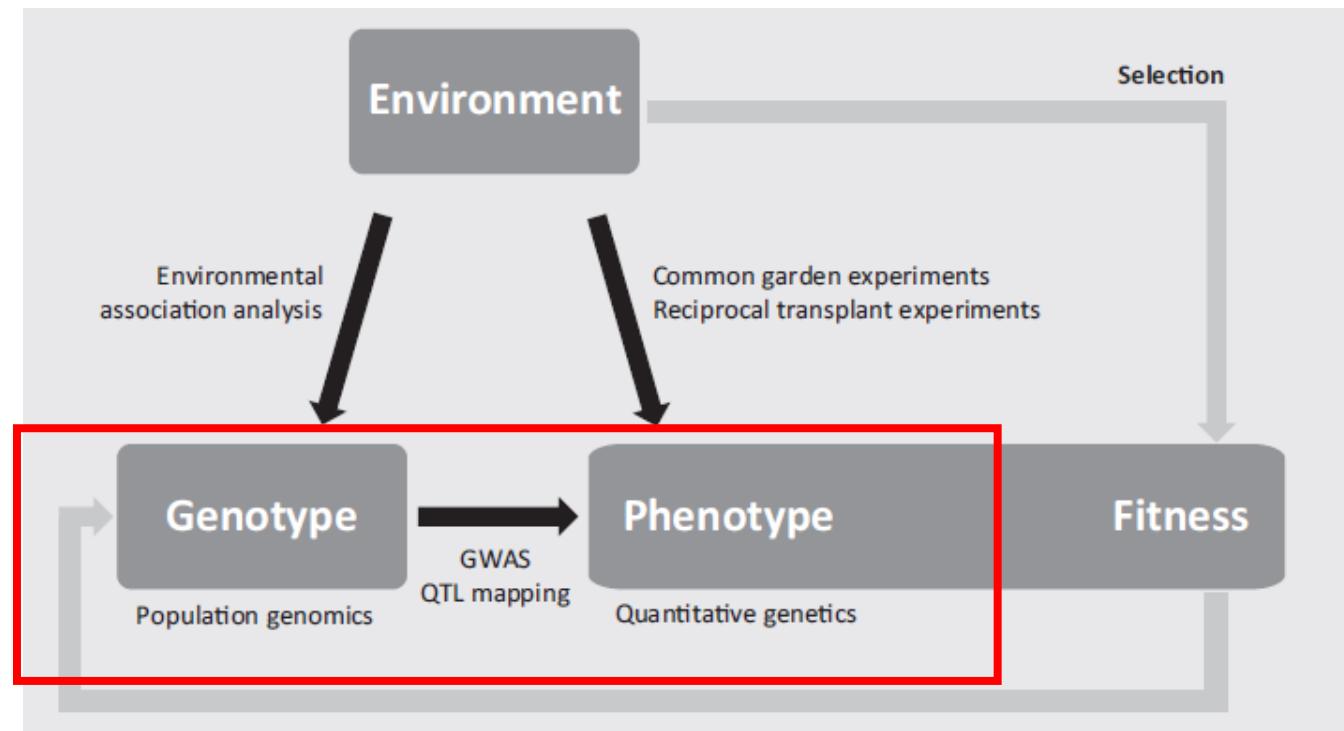
Blanquart et al. (2013) *Ecology Letters*



Rellstab et al. (2015) *Mol. Ecol.*

# Top-down approaches

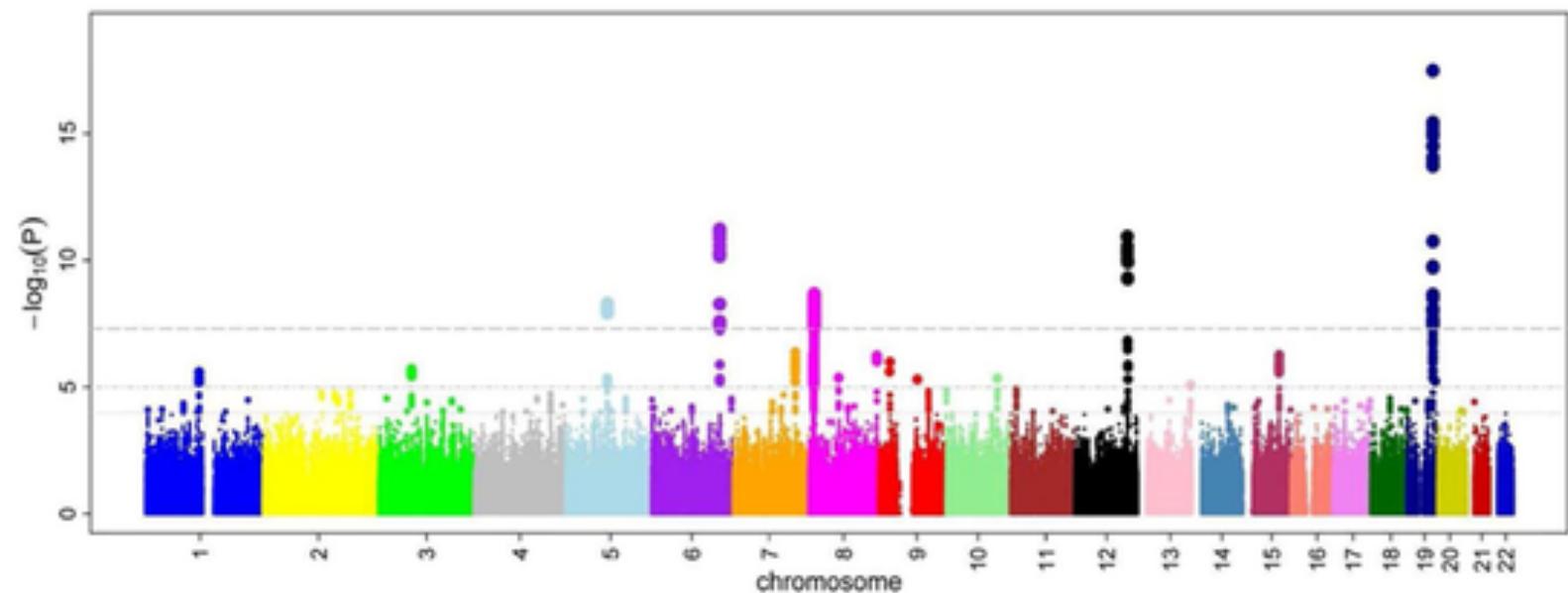
Take advantage of phenotypic measurements and relate them to genotypic data



# Genome-wide association studies (GWAS)

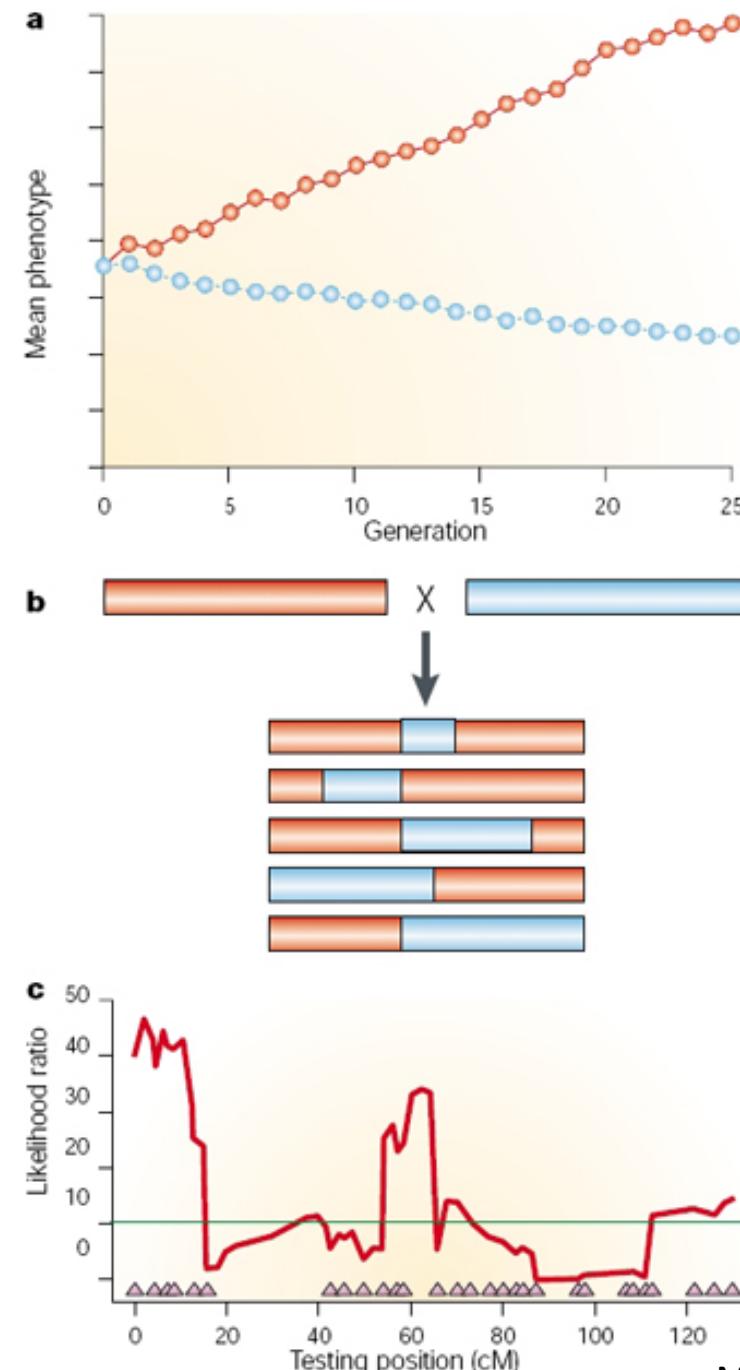
An observational study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait

- Typically focus on associations between SNPs and traits (BayeScan)
- Can also be used in studies with environmental factors instead of traits (Bayenv)



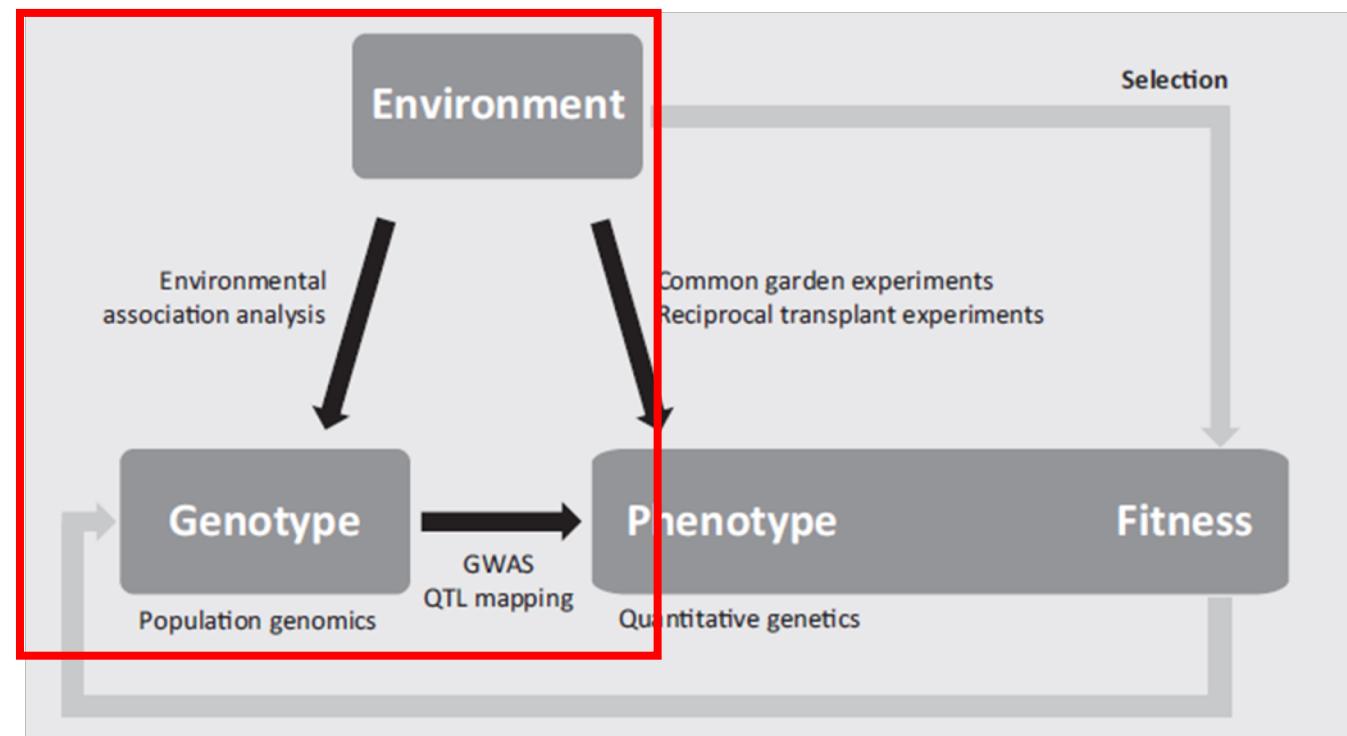
# QTL Mapping

- A quantitative trait locus (QTL): a locus that correlates with variation of a quantitative trait in the phenotype of a population
- QTLs are **mapped** by identifying which molecular markers (such as SNPs) correlate with an observed trait



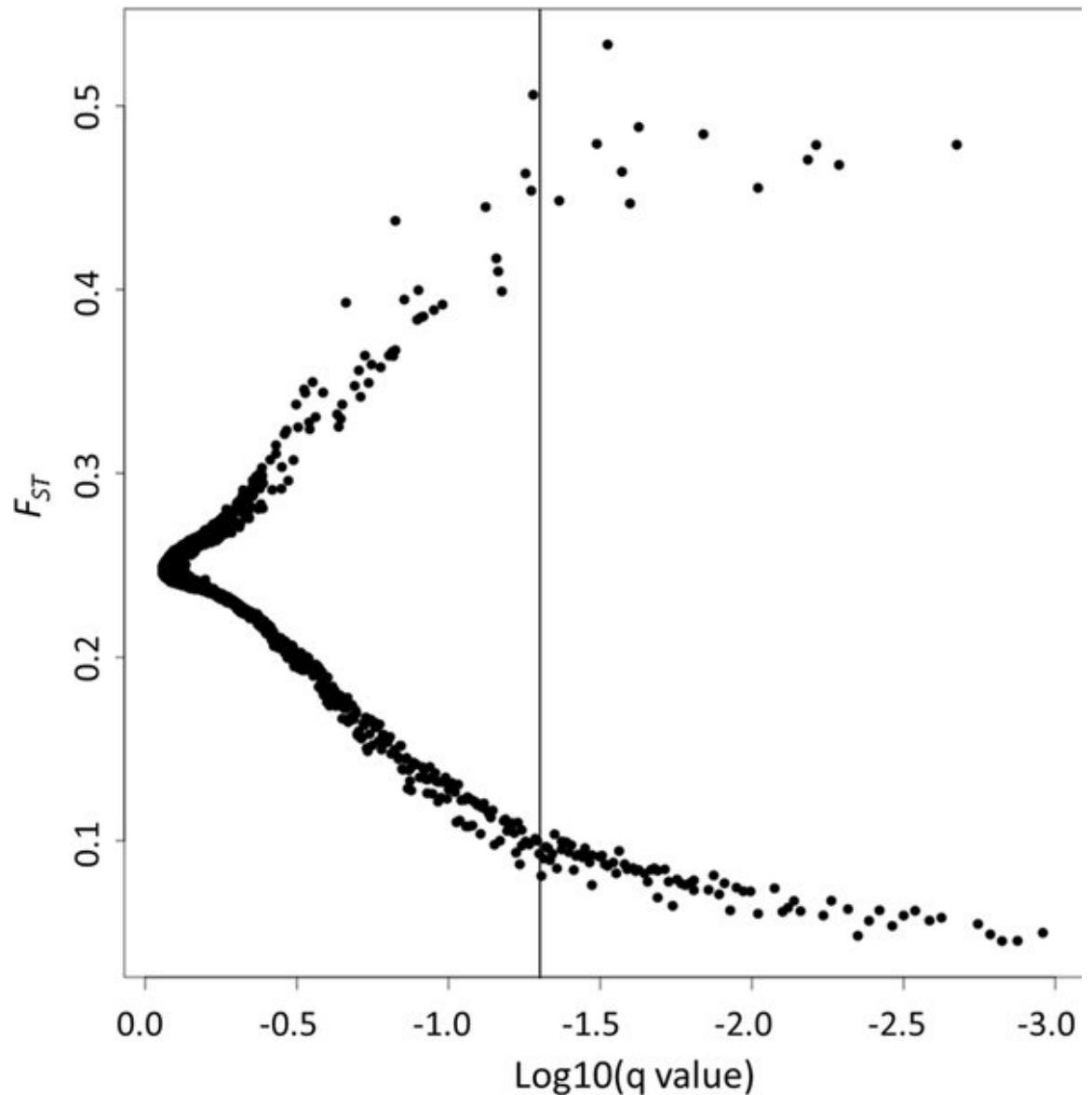
# Bottom-up approaches

Use genomic info to identify signatures of adaptive genetic variation and relate them to environmental variation



# Outlier tests

- Uses estimates of population genetic differentiation (e.g.,  $F_{ST}$ )
- Assumes neutral processes affect the genome uniformly, while selection should act on specific loci
- Loci with significantly higher or lower genetic differentiation than expected under neutrality are outliers that are considered to be under selection



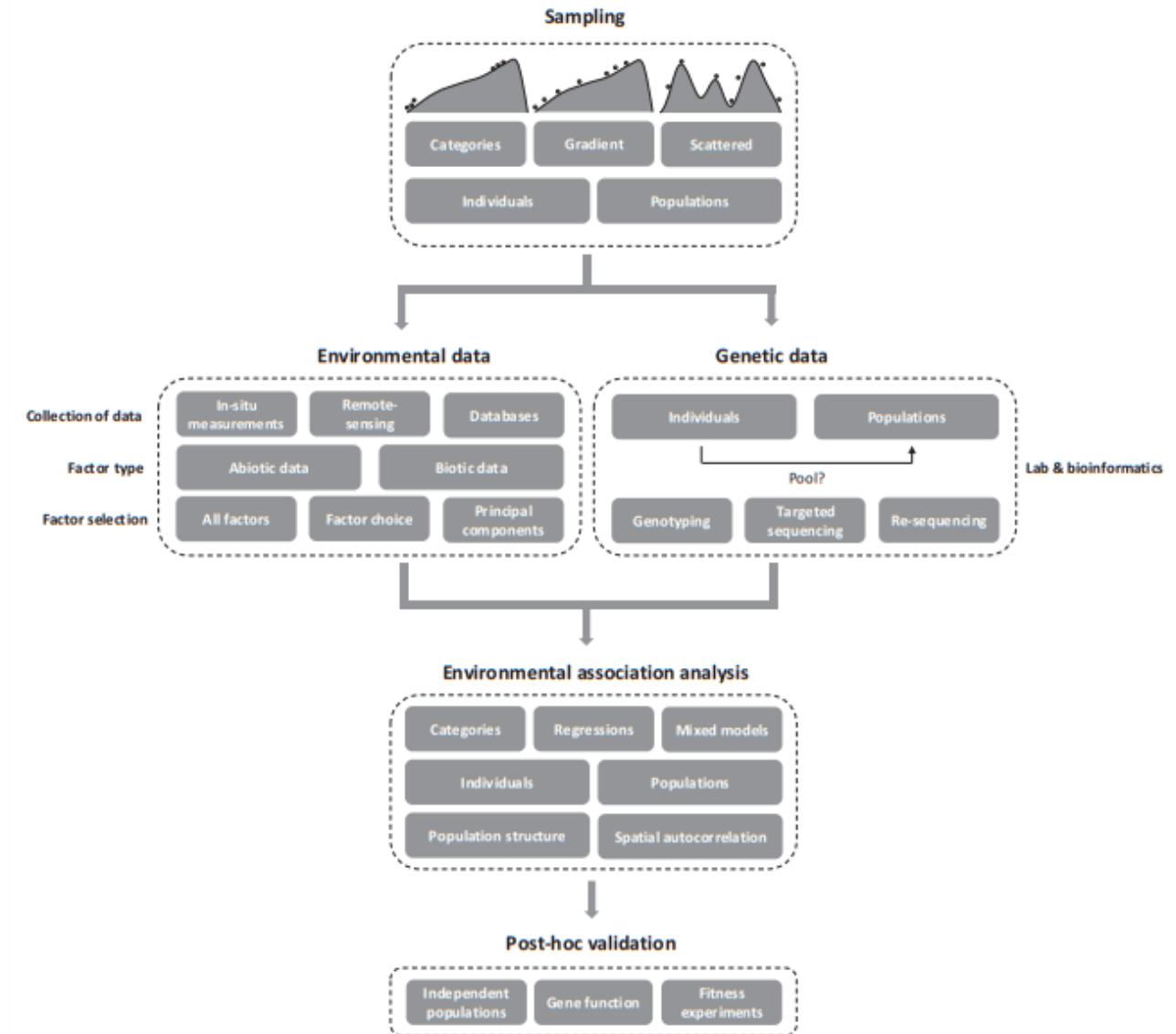
- Outlier Test Problems:
  - Aimed at identifying loci that underwent selective sweeps of beneficial alleles
  - Make the assumption that selection pressure differ among populations, but usually don't attempt to link to specific selection pressures that underlie adaptation

# Landscape Genomics

- Landscape = Seascape = Riverscape
- Aim to identify the environmental factors that shape adaptive and neutral genetic variation and identify the gene variants that drive local adaptation
- Uses environmental association analysis (EAA)

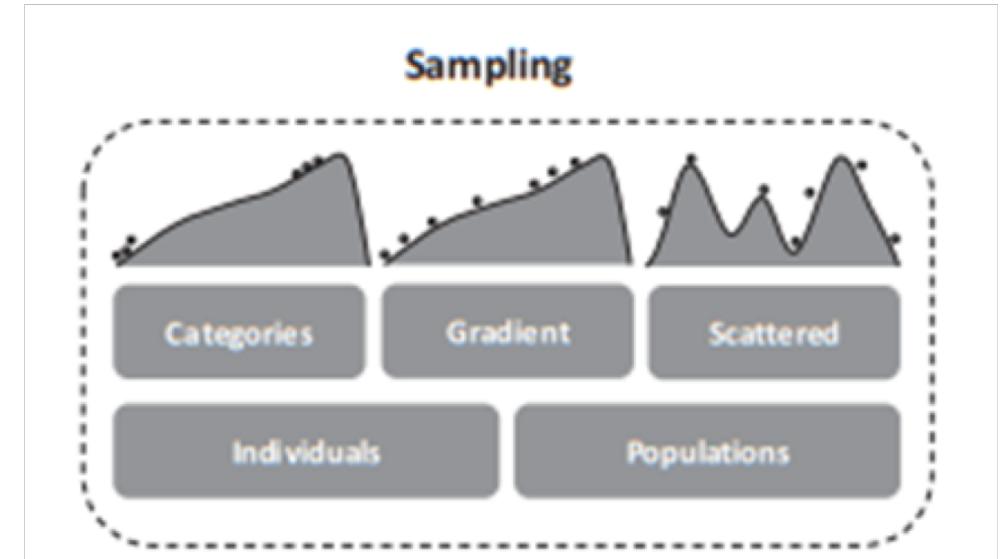
# Environmental Association Analysis

Main Goal: to test whether a specific allele is associated with a specific environmental factor, while controlling for neutral genetic structure



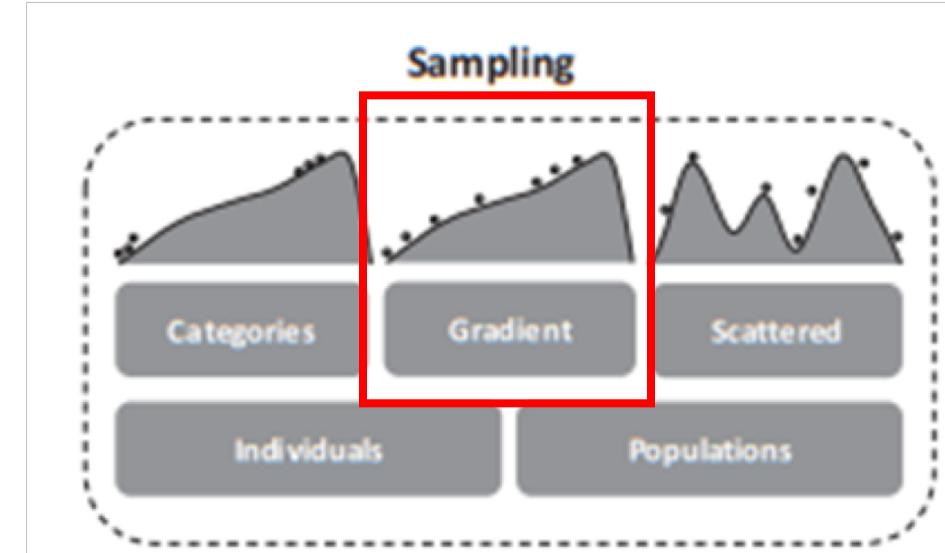
# Sampling

- Sampling along environmental gradients
- Categorical sampling
- Scattered sampling
- \*Sample scattered and random pairs of closely situated populations that exhibit differences in environmental conditions while being within gene flow distance (Lotterhos & Whitlock, 2015)



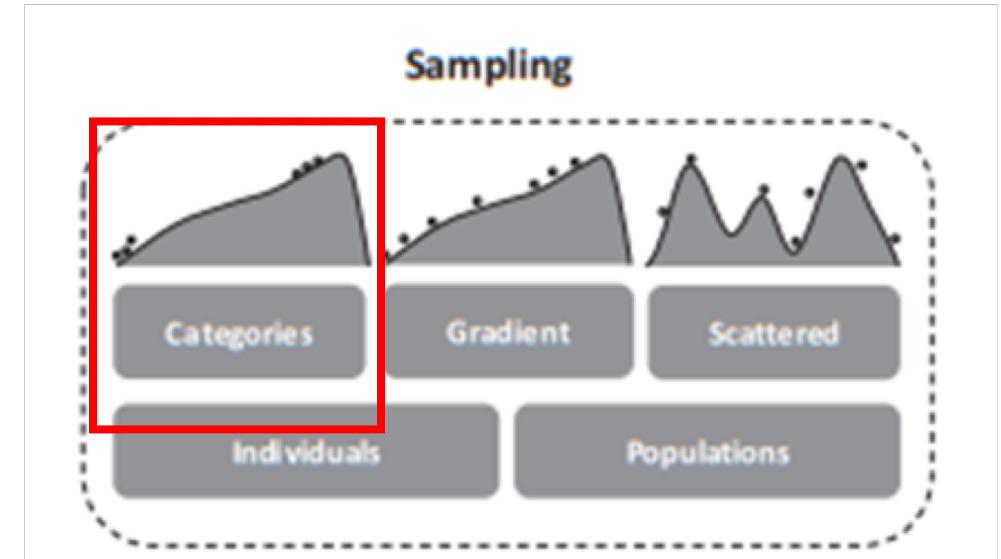
# Sampling

- **Sampling along environmental gradients**
  - Replication of gradients is necessary
- Categorical sampling
- Scattered sampling
- \*Sample scattered and random pairs of closely situated populations that exhibit differences in environmental conditions while being within gene flow distance (Lotterhos & Whitlock, 2015)



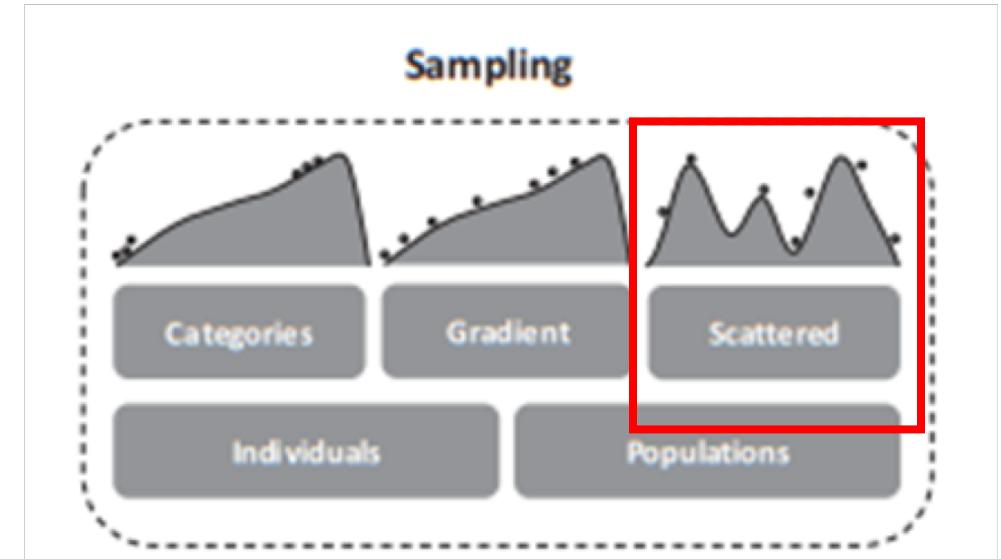
# Sampling

- Sampling along environmental gradients
- **Categorical sampling**
  - Setting up different treatments
- Scattered sampling
- \*Sample scattered and random pairs of closely situated populations that exhibit differences in environmental conditions while being within gene flow distance (Lotterhos & Whitlock, 2015)



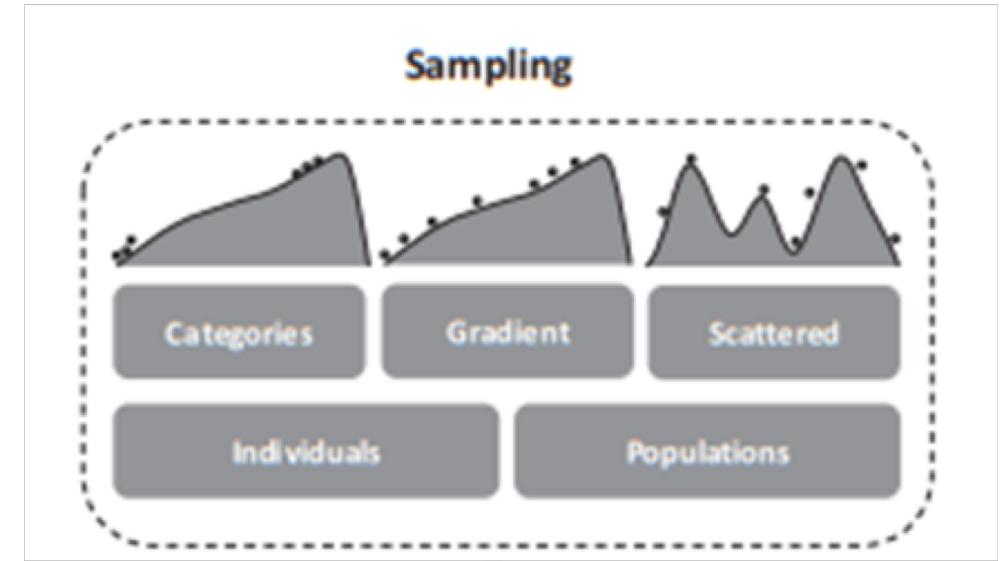
# Sampling

- Sampling along environmental gradients
- Categorical sampling
- **Scattered sampling**
  - Broad sample covering the entire environmental niche
  - \*Sample scattered and random pairs of closely situated populations that exhibit differences in environmental conditions while being within gene flow distance (Lotterhos & Whitlock, 2015)

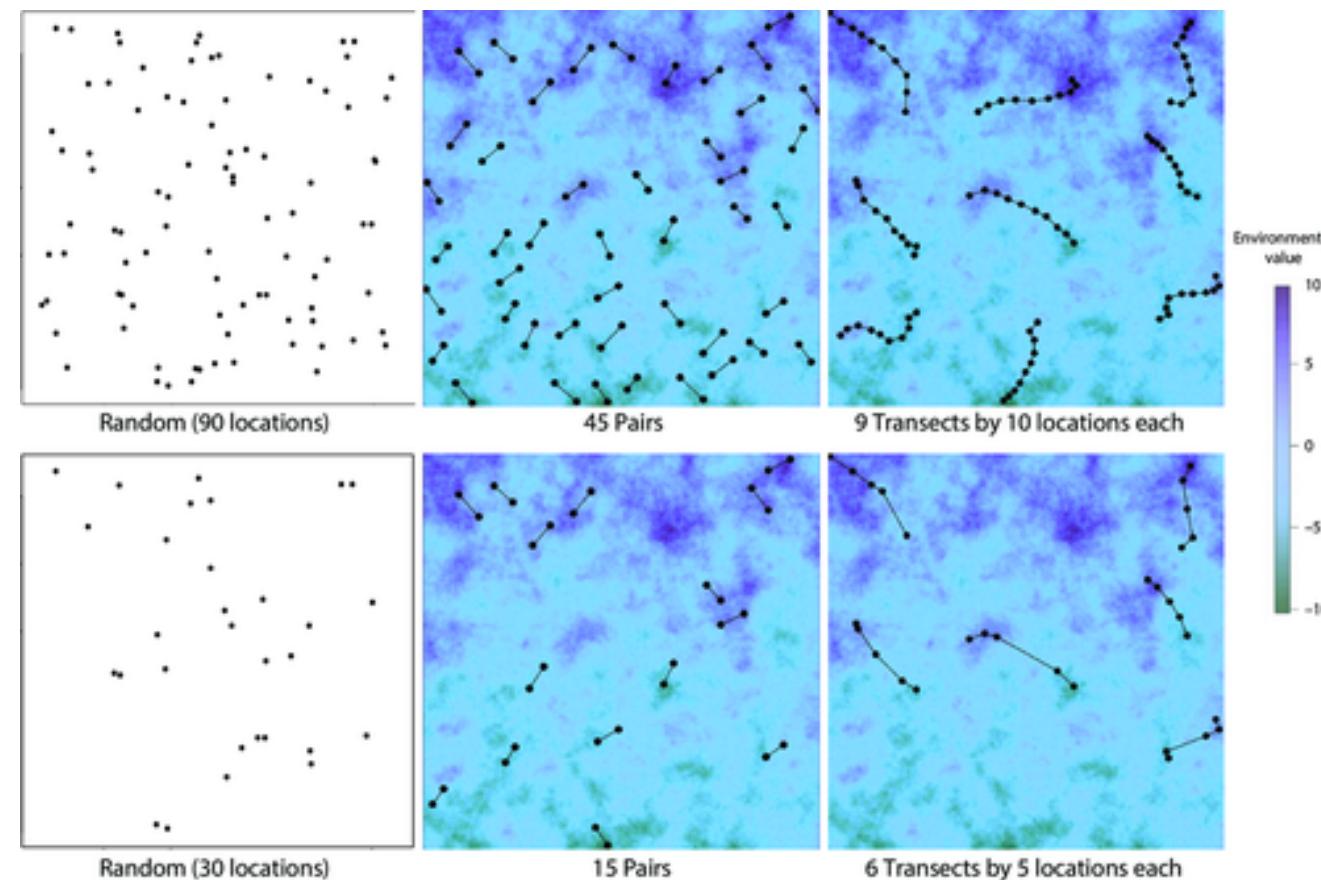


# Sampling

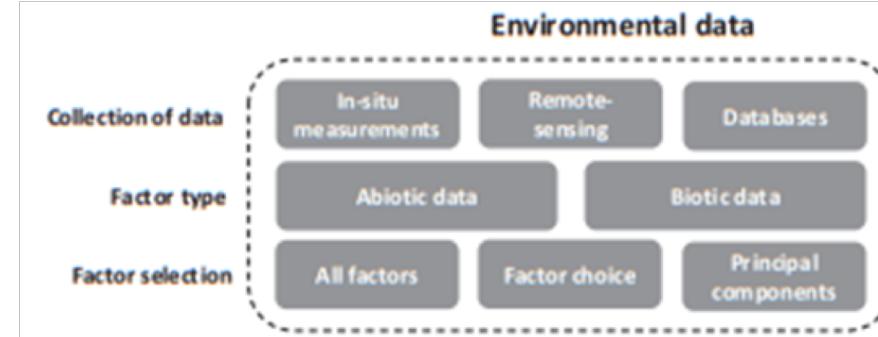
- Sampling along environmental gradients
- Categorical sampling
- Scattered sampling
- **\*Sample scattered and random pairs of closely situated populations that exhibit differences in environmental conditions while being within gene flow distance (Lotterhos & Whitlock, 2015)**



Sample scattered and random pairs of closely situated populations that exhibit differences in environmental conditions while being within gene flow distance



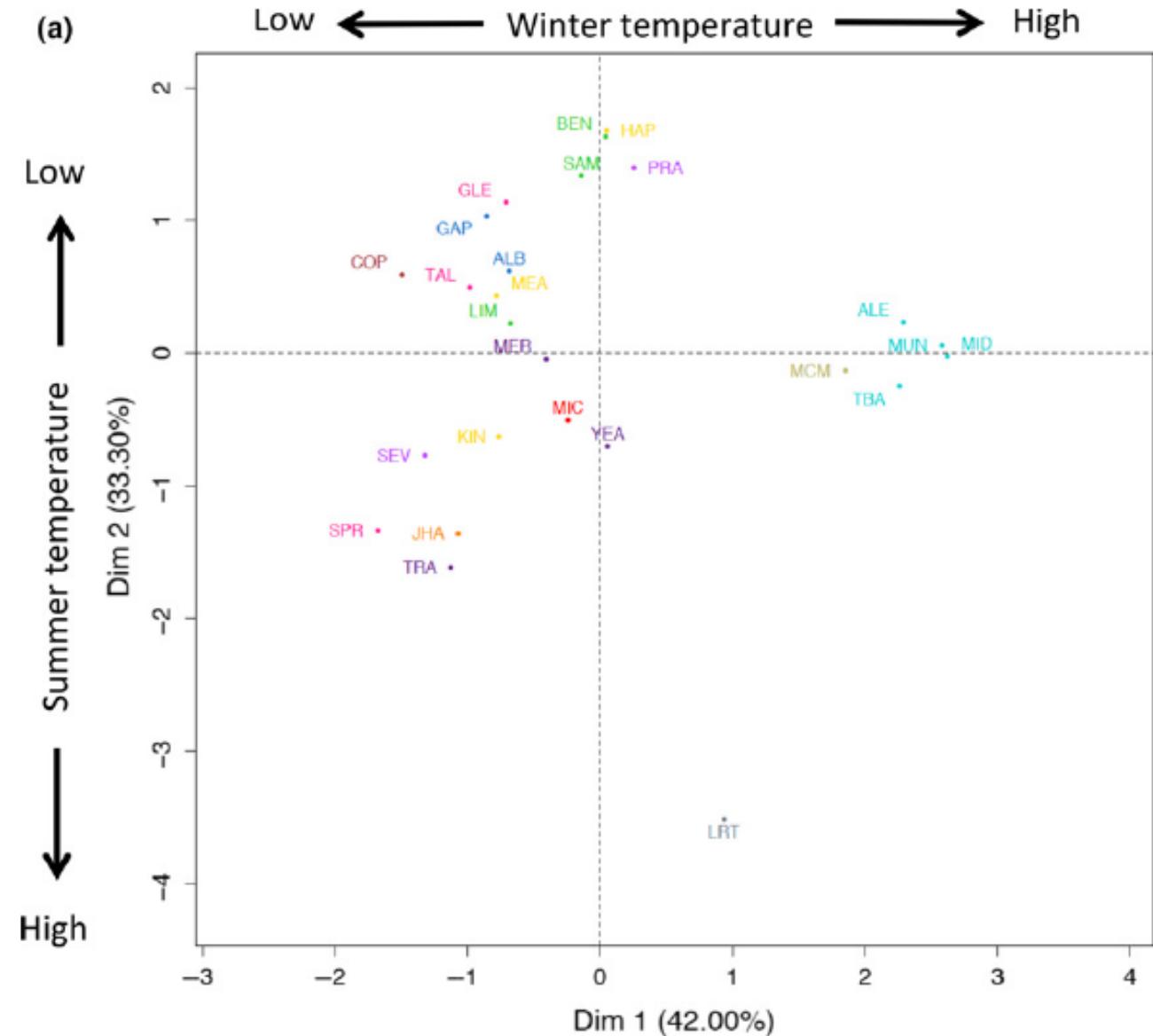
# Environmental Factors



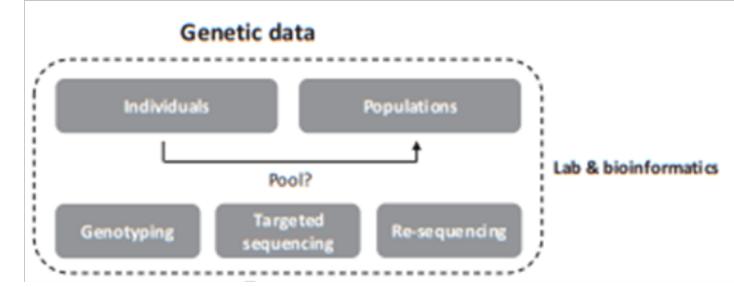
- Using all factors increases # of statistical tests
- Many abiotic and biotic factors are highly correlated → could lead to wrong conclusions
  - Reduce the # of factors
  - Select or remove high correlated factors based on their contribution to the first few axes of a principal component analysis (PCA)
  - Control for multicollinearity with Variance Inflation Factor (VIF)

# PCA

- Keeping the environmental factors that with the highest contribution to each axis



# Genetic polymorphisms



- Whole-genome sequencing

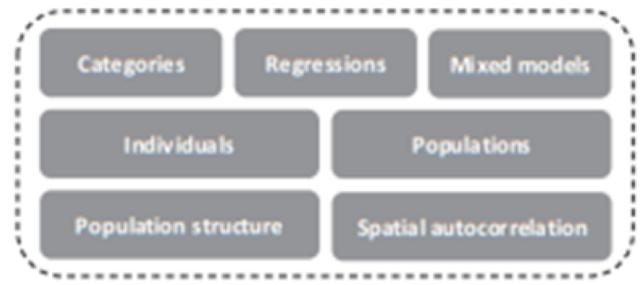
Cheaper approaches:

- Pool-Seq
- Candidate-gene approach
- Complexity reduction of the genome

# Incorporating neutral genetic structure

- Consider various types of autocorrelation
  - Associations may just be a consequence of spatial arrangement and demographic history of the individuals or populations
- Spatial autocorrelation can serve as a proxy for **neutral genetic structure**
- Neutral genetic structure can result in false positives
- Method for estimating neutral genetic structure: estimating pairwise fixation indices among populations

## Environmental association analysis



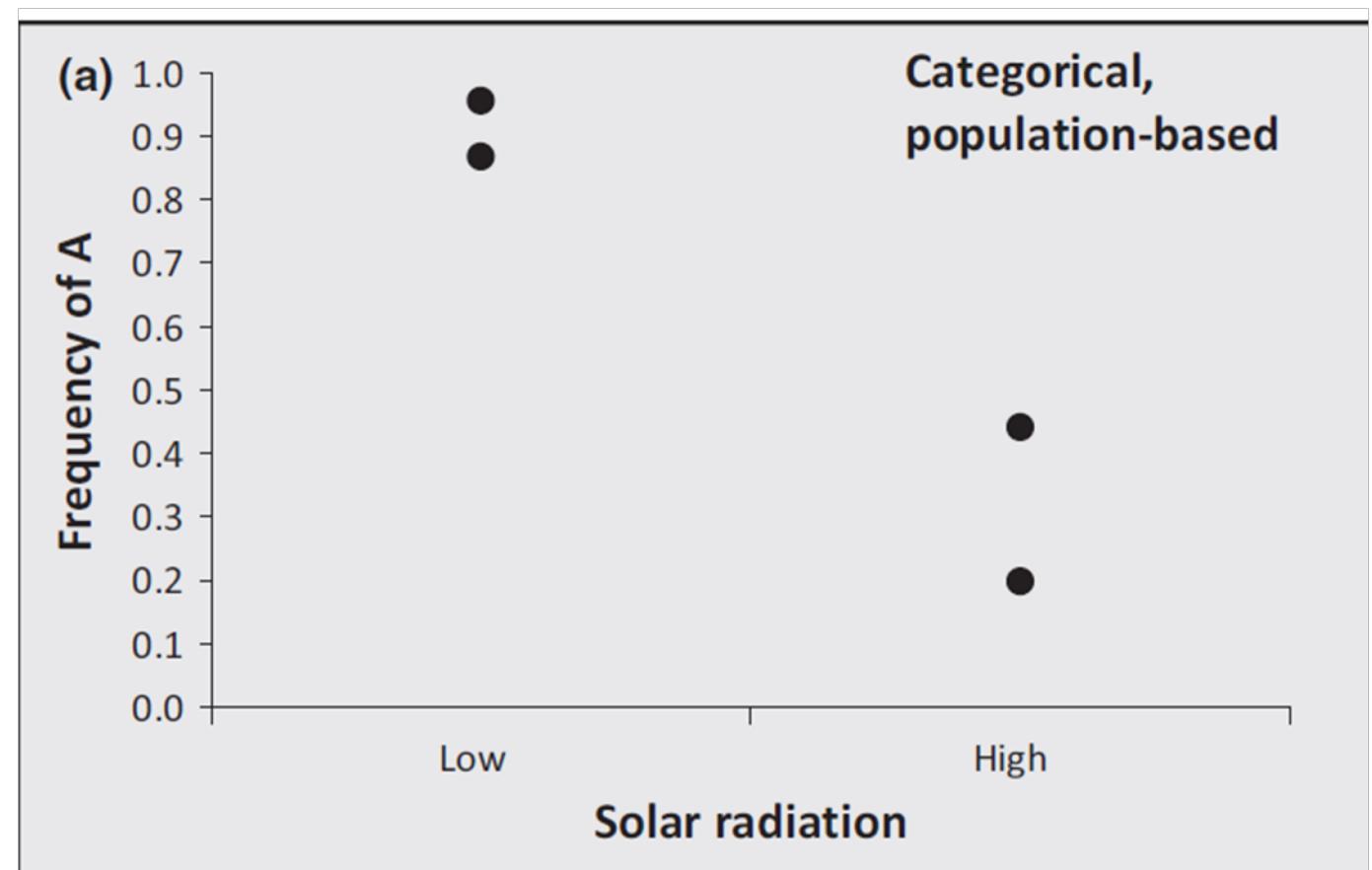
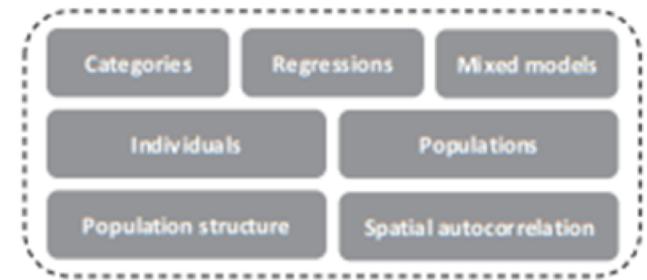
# Analysis of data

- Testing categorical factors
- Logistic regressions
- Matrix correlations
- General linear models
- Mixed effects models

# Analysis of data

- Testing categorical factors
- Logistic regressions
- Matrix correlations
- General linear models
- Mixed effects models

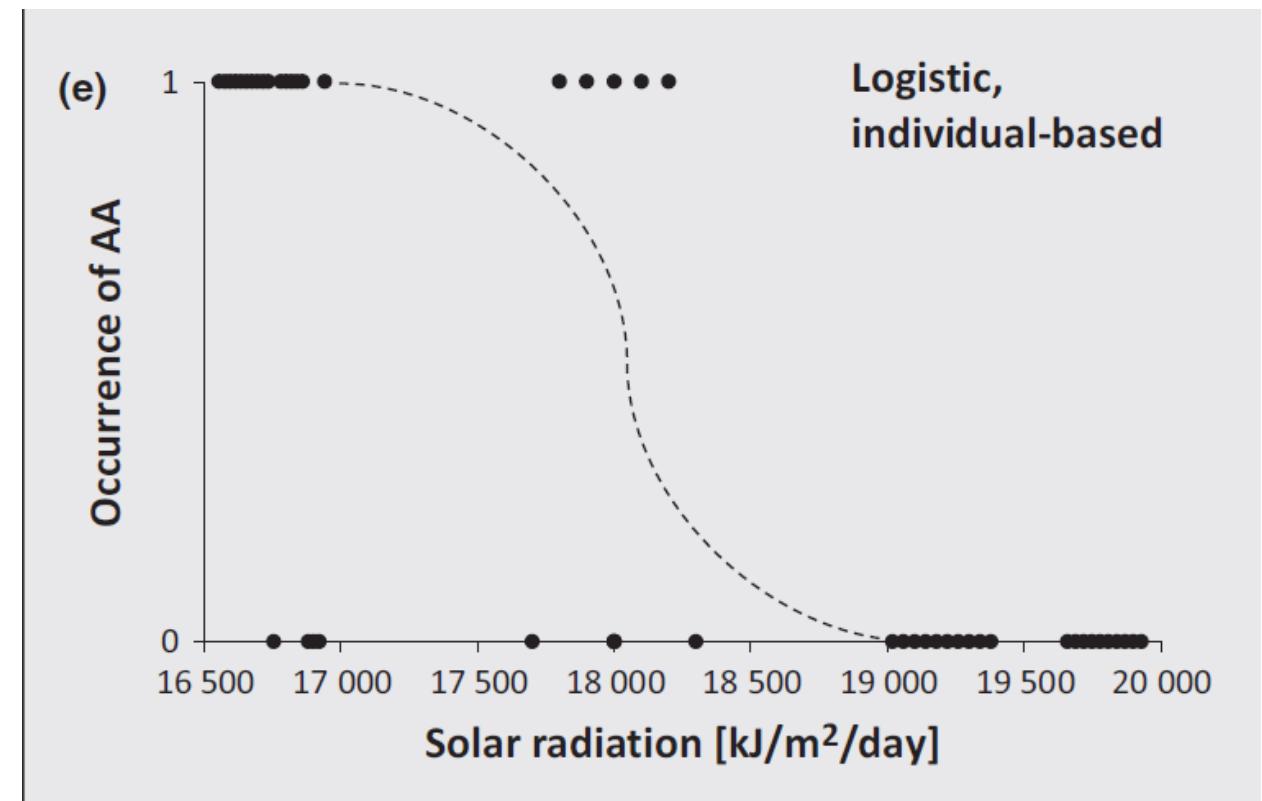
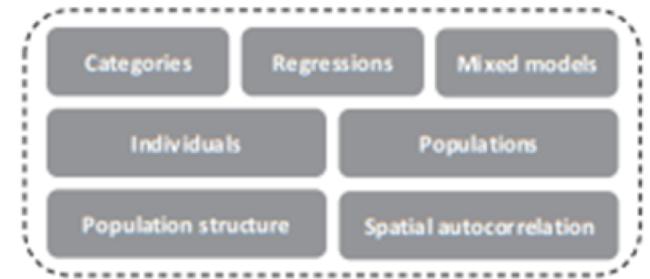
## Environmental association analysis



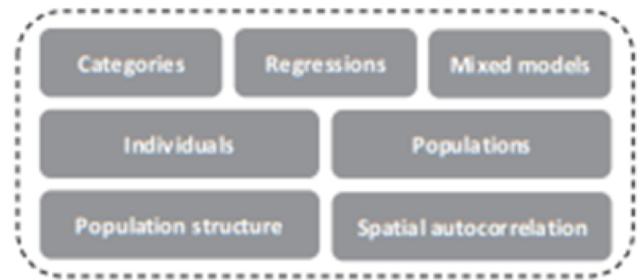
# Analysis of data

- Testing categorical factors
- **Logistic regressions**
- Matrix correlations
- General linear models
- Mixed effects models

## Environmental association analysis



## Environmental association analysis



# Analysis of data

- Testing categorical factors
- Logistic regressions
- **Matrix correlations** (Hancock et al. 2011)
- General linear models
- Mixed effects models

# Matrix correlations

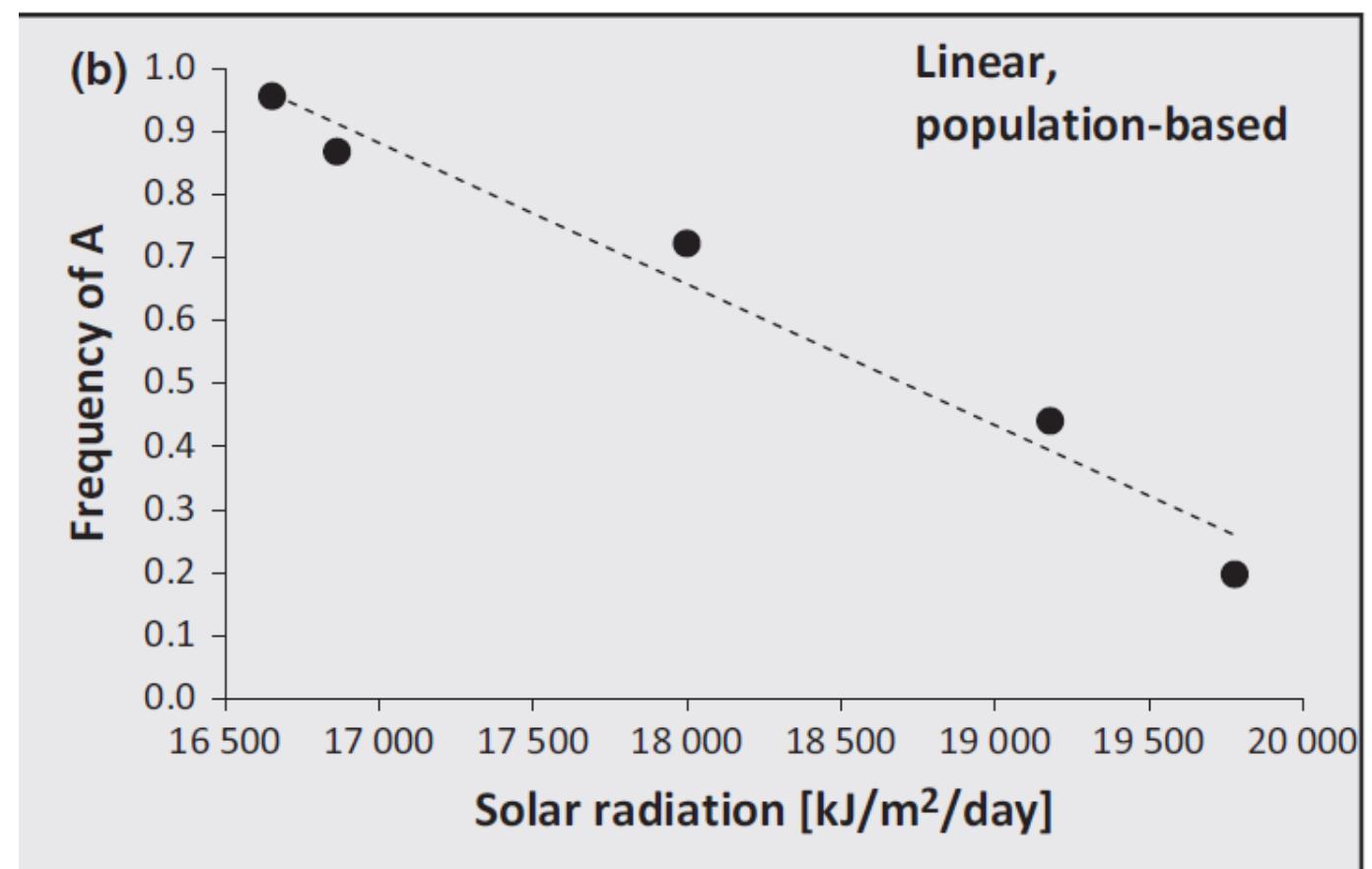
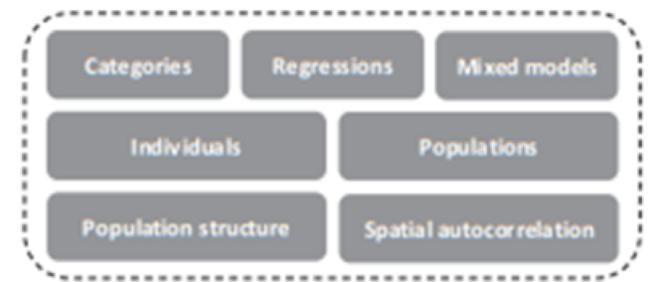
## Mantel & Partial Mantel tests

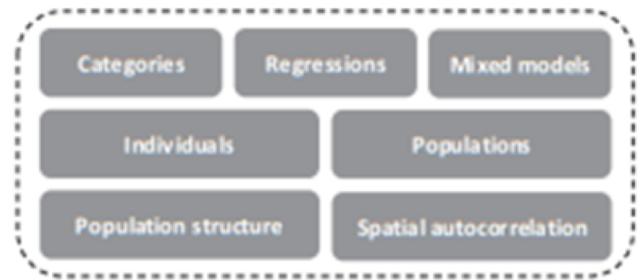
17	17	12	20	24	26	31	29	26	28	30	31	29	35	39	43	35	36	35	39	39			
0	14	-2	-9	-8	-7	-7	-4	4	-4	-4	-6	-9	-12	-12	-16	-6	-8	-10	-8	-11	-9		
10	10	-4	-8	-6	-5	-5	-2	2	-2	-3	-5	-9	-9	-10	-12	-3	-8	-6	-6	-7	-7		
6	9	-4	-2	0	1	-3	2	5	-1	0	2	-7	-6	-9	-7	0	-4	-2	-3	-1	-1		
9	21	-7	2	-4	1	-5	0	1	8	7	2	-4	-5	-7	-9	-4	-4	0	-1	-5	-4		
7	8	2	-8	-1	4	0	6	2	1	3	-1	3	-1	-1	-1	0	5	3	2	0	2		
-3	7	2	4	3	7	-1	2	4	9	2	5	4	-1	3	6	1	3	4	3	5	5		
9	25	-3	-5	-1	1	7	8	7	9	6	6	7	4	6	2	6	9	5	6	4	6		
12	22	-5	8	5	11	5	9	14	11	8	12	-2	1	-1	5	11	8	8	7	5	8		
15	18	-2	17	4	5	5	5	15	8	12	14	1	5	3	7	12	8	7	10	6	9		
14	12	0	2	6	7	13	12	10	10	10	13	10	11	14	13	13	14	16	16	15	12		
9	17	6	1	2	6	3	6	6	10	10	9	7	5	7	11	5	8	8	8	9	9		
6	8	5	0	0	3	9	6	8	5	9	9	8	9	13	13	13	7	8	12	12	10		
FLC FT Field	FRI	LN	FT	8W	FT10	LN10	SD	LN	SDV	OW	LN16	LN22	FT16	4W	LD	LDV	FT	GH	FT22	2W	OW	GH	FT
0W	GH	LN	8W	GH	FT	8W	GH	LN	SDV	OW	LN16	LN22	FT16	4W	LD	LDV	FT	GH	FT22	2W	OW	GH	FT

# Analysis of data

- Testing categorical factors
- Logistic regressions
- Matrix correlations
- **General linear models**
- Mixed effects models

## Environmental association analysis





# Analysis of data

- Testing categorical factors
- Logistic regressions
- Matrix correlations
- General linear models
- **Mixed effects models**
  - Fixed and Random effects
    - Environmental factors = fixed effects
    - Neutral genetic structure = random effects

$$\mathbf{Y}_i = \mathbf{X}_i \times \boldsymbol{\beta} + \mathbf{Z}_i \times \mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

Fixed Term      Random Term

# Important notes

- EAAs do have limitations, so be careful
  - false positives
- You can combine approaches
  - EAA + outlier detection methods
- Post hoc validation

# Conclusions

- Genetic variation due to both neutral and selective processes
- Selection comes in many different forms → signature of selection are variable
- Natural selection in the midst of environmental change
- Differential pressures of natural selection can result in local adaptation
- Landscape genomics aims to identify the environmental factors that shape adaptive and neutral genetic variation

# Discussion Readings

Brauer, Hammer & Beherregaray 2016  
&  
Hancock et al. 2011