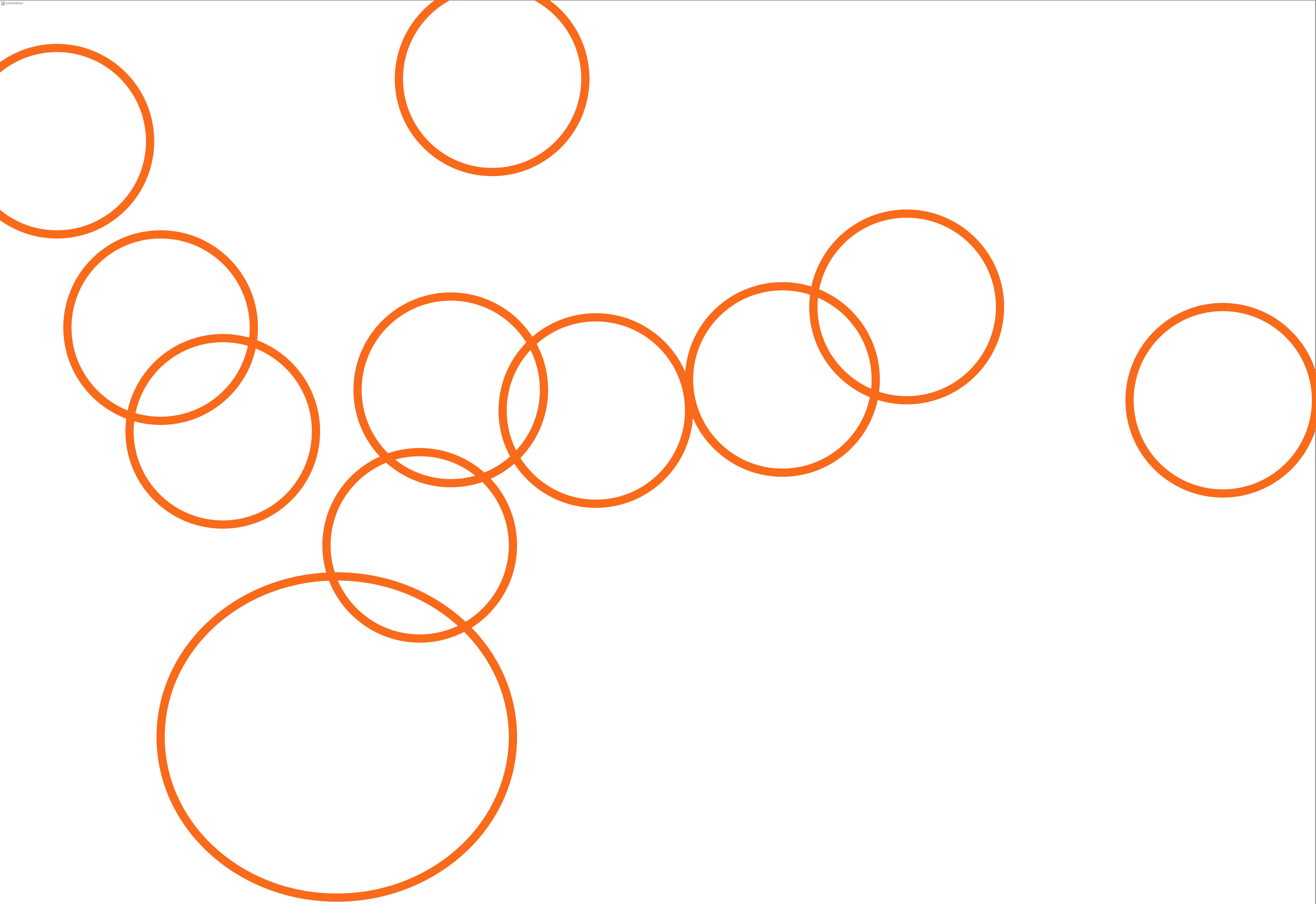


Plant of the day!

- Pebble plants, *Lithops*, dwarf xerophytes
- Aizoaceae
- South African
- Plants consist of one or more pairs of bulbous leaves – almost no stem
- Leaf markings appear to help plant match its background and be less vulnerable to herbivory

Lithops lesliei

Genomics of Adaptation

Questions

- What are the genetic changes that underlie adaptation?
- What are the population genetic or genomic signatures of adaptation?
- How do non-adaptive processes affect tests of selection?
- Can we detect evidence of maladaptation in the genome?

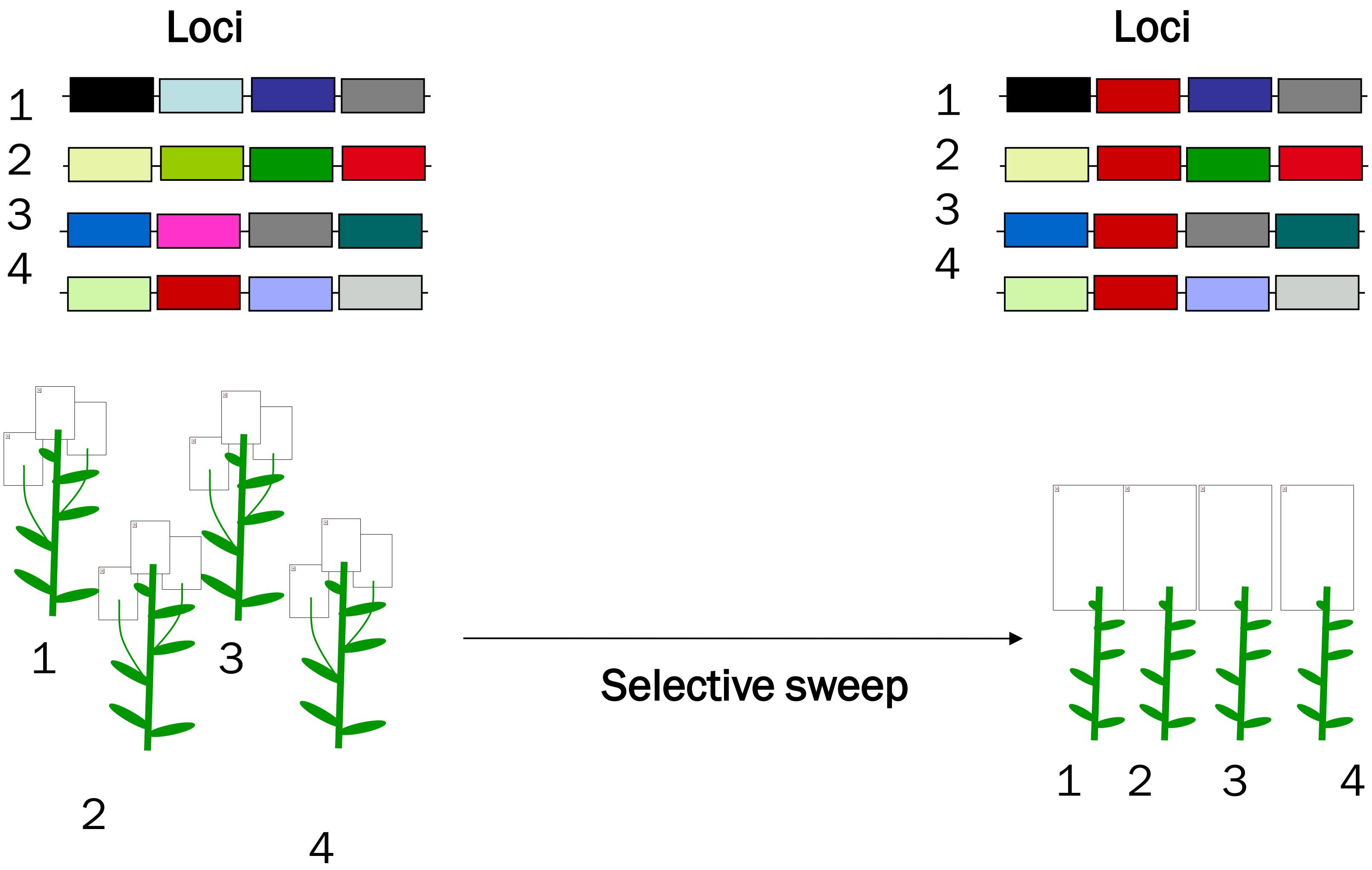
Goals

- Understand some top down and bottom up approaches used to identify genes responsible for adaptation
- Explain patterns of sequence variation expected with directional and balancing selection
- Understand the principles of population genetic tests of selection
- Understand the principles of population genetics tests of maladaptation

The genetic basis of adaptation

- Phenotype to genotype (Top down)
  - Identify important trait then find loci associated with it
  - QTL, association mapping
  - Genotype to phenotype (Bottom up)
  - Identify loci under selection, then find trait associated with loci
  - Population genetics

Which locus is likely involved in the change in floral phenotype?



Selective sweep

Quantitative trait loci (QTL)

-Genomic regions associated with trait variation

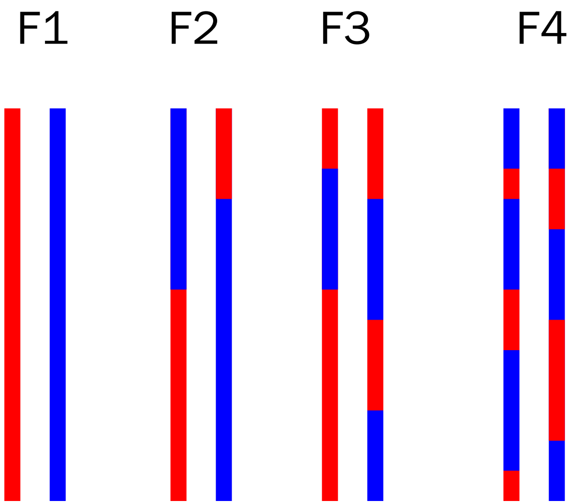
-Loci detected may differ across individuals/environments

-Statistical issues (sample size, genes of small effect, epistasis)

-Can be large regions of a chromosome (further mapping in region needed)

-Can't perform in all species

Quantitative trait loci (QTL)



-Precision limited by density of markers and number of recombination events

-Recombination events limited by the number of individuals and their degree of recombination between the parental genomes (i.e. F2, F3, etc)

Parental genomes are more finely recombined with each generation of consecutive intercrosses.

Association mapping

Associations between markers (SNPs) and phenotypes in individuals collected from natural populations

- Individuals of sunflower vary in their floral UV pattern.
- Look through whole genome to find SNPs associated UV pattern variation.

Association mapping

Pros:

- Much higher resolution
- No need for crosses

Cons:

- Population structure may lead to spurious associations
- Need many many markers, more than QTL mapping.

Example: genome-wide association (GWA) mapping

n = 159

n = 563

One locus explains 62% of variation for UV patterns in wild *H. annuus*

Example: association mapping using the environmental variables as the phenotype

Genotype-Environment Association (GEA) analysis for ”continentality” in common sunflower

The genetic basis of adaptation

- Phenotype to genotype (Top down)
  - Identify important trait then find loci associated with it
  - QTL, association mapping, bulk segregant analysis
- Genotype to phenotype (Bottom up)
  - Identify loci under selection, then find trait associated with loci
  - Population genetics

Detecting natural selection

- The Neutral theory suggests that most molecular changes are neutral and are caused by random genetic drift
- This is used as a null hypothesis and deviations from neutral expectations are evidence of selection
- Important to consider how non-selective processes like population structure and linkage affect the statistics

The effect of selection on the genome

Directional selection

- Best allele(s) sweep to fixation
- Loss of variation
- Change in frequency distribution of polymorphisms
- Increase in linkage disequilibrium around the site

The effect of selection on the genome

Directional selection

- Best allele(s) sweep to fixation
- **Loss of variation**
- **Change in frequency distribution of polymorphisms**
- Increase in linkage disequilibrium around the site

Balancing selection

- Maintains variation that otherwise would be lost to drift
- Heterozygote advantage, frequency dependent selection, fluctuating selection, (divergent selection)

Directional selection

- A beneficial allele arises
- Variants with this allele rapidly spread through the species
- Genetic diversity is reduced around this adaptive locus

ancestral

After selection

Chance of detecting natural selection

Depends on:

- Time
- Strength of selection
- Recombination, mutation
- Initial frequency

Selective sweep

Methods for detecting selection

A. MacDonald-Kreitman Type Tests

B. Site Frequency Spectrum Approaches

C. Linkage Disequilibrium (LD) and Haplotype Structure

D. Population Differentiation: Lewontin-Krakauer Methods

These tests can be applied to single genes, or across the whole genome.

A. MacDonald-Krietman type tests

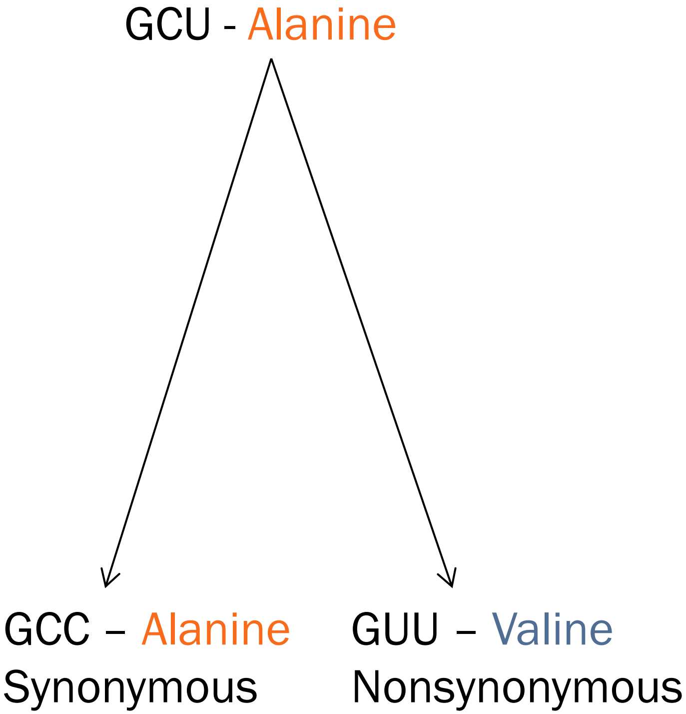
- **Synonymous** substitutions:
- Mutations that do not cause amino acid change (usually 3rd position)

“silent substitutions”

- **Nonsynonymous** substitutions:
- Mutations that cause amino acid change (1st, 2nd position)

“replacement substitutions”

A. MacDonald-Krietman type tests



GCU - Alanine

GCC – Alanine
Synonymous

GUU – Valine
Nonsynonymous

A. MacDonald-Krietman type tests

K_a/K_s Test

$$\frac{\text{Nonsynonymous substitutions}}{\text{Synonymous substitutions}} \qquad \frac{K_a}{K_s}$$

- Uses coding sequence (sequence that codes proteins)
- K_s doesn't change protein so is "neutral" and is used as baseline rate
- Important to remember that both types of mutations occur at the same rate, it is fixation rate that varies.

A. MacDonald-Krietman type tests

K_a/K_s Test

$$\frac{\text{Nonsynonymous substitutions}}{\text{Synonymous substitutions}} \qquad \frac{K_a}{K_s}$$

- $K_a/K_s = 1$ --- Neutral drift. Protein changes aren't being selected for or against.
- $K_a/K_s > 1$ --- Positive selection. Protein changes are being selected for
- $K_a/K_s < 1$ --- Purifying selection. Protein changes are being selected against.

A. MacDonald-Krietman type tests

K_a/K_s Test

$$\frac{\text{Nonsynonymous substitutions}}{\text{Synonymous substitutions}} \qquad \frac{K_a}{K_s}$$

- Can be done with single sequences per species/group (don’t need population genetics data)
- Can pinpoint where selection occurred on a phylogeny
- Proteins very rarely have $K_a/K_s > 1$ for their entirely sequence, often only small pieces or single codons are under selection
  - Proteins with $K_a/K_s > 1$ are often under balancing selection, e.g. immune or self-incompatibility genes

B. Site Frequency Spectrum

- Selection affects the distribution of alleles within populations
- Method examines site frequency spectrum and compares to neutral expectations
- Could be applied to a single locus. Now used often for genomic scans for selective sweeps

B. Site Frequency Spectrum

B. Site Frequency Spectrum

Many derived alleles at high frequency because they were swept there with the positively selected site

Proportion of polymorphic sites

Low

High

Population derived allele frequency

B. Site Frequency Spectrum

Many derived alleles at high frequency because they were swept there with the positively selected site

B. Site Frequency Spectrum

Many derived alleles at low frequency are new mutations in the swept region

Proportion of polymorphic sites

Low

High

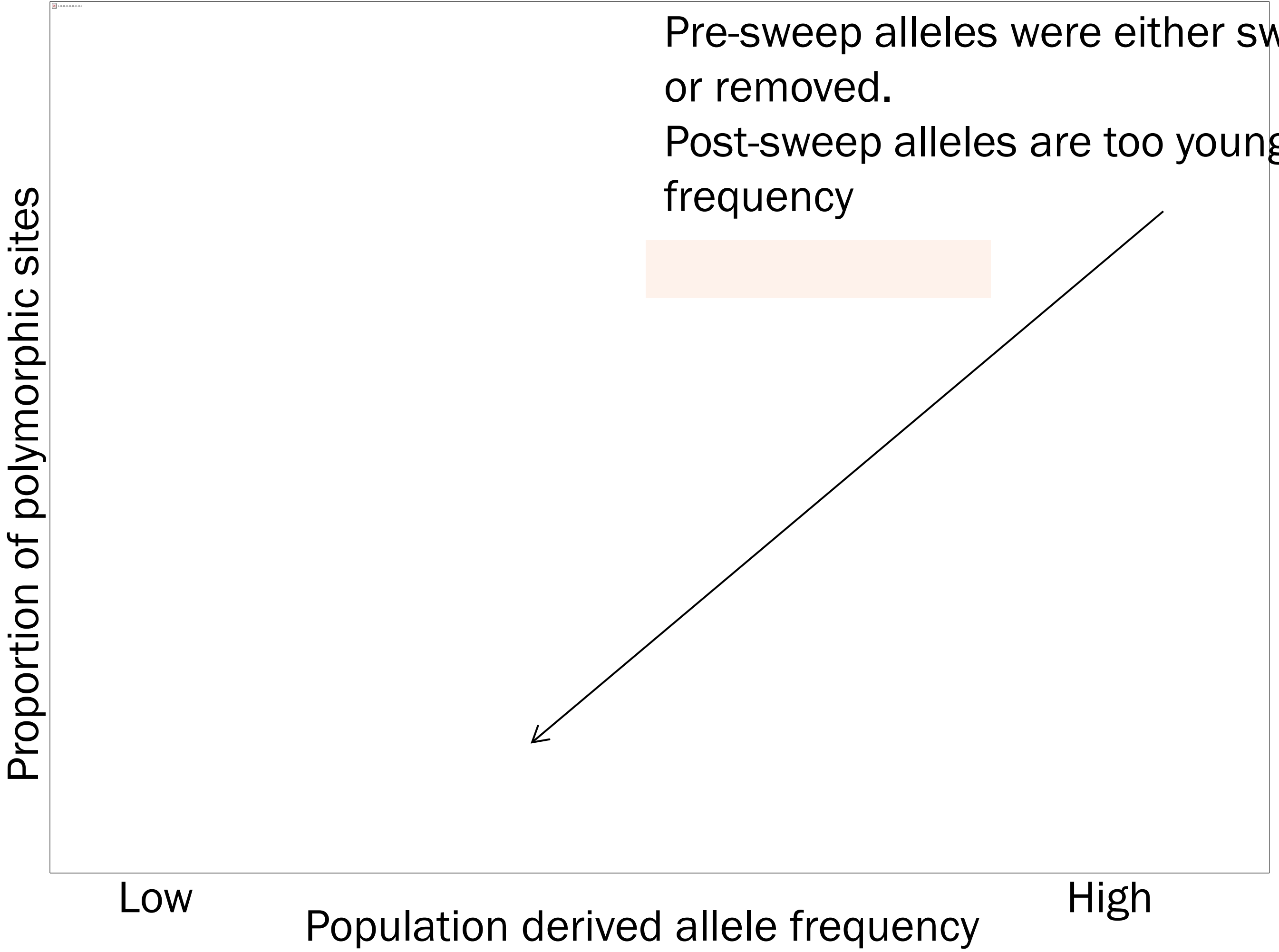
Population derived allele frequency

B. Site Frequency Spectrum

Many derived alleles at low frequency are new mutations in the swept region

B. Site Frequency Spectrum

Few medium frequency derived alleles.
Pre-sweep alleles were either swept to high frequency or removed.
Post-sweep alleles are too young to reach medium frequency



B. Site Frequency Spectrum

Few medium frequency derived alleles. Pre-sweep alleles were either swept to high frequency or removed.

Maize cupulate fruitcase genetics

Wildtype teosinte hard fruitcase

Teosinte with maize *tga1* gene

Maize cupulate fruitcase genetics

Maize has less diversity compared to teosinte

Maize cupulate fruitcase genetics

Tajima’s D looks at site frequency spectrum. Negative values suggests many rare polymorphisms, which occurs during a selective sweep.

Maize cupulate fruitcase genetics

HKA asks if there is more divergence between species than would be expected by the amount of polymorphism in the species

C. Linkage Disequilibrium (LD)

- The nonrandom association of alleles from different loci
- Levels of linkage disequilibrium will increase during selective sweeps
  - As a new mutation rises in frequency, it will drag along linked sites
  - This haplotype block will have high LD until recombination breaks it up over time

High LD between linked sites

D. Population Differentiation: Lewontin-Krakauer Methods

- Selection will often increase the degree of genetic distance between populations

- Compute pairwise genetic distances (*e.g.*, F_{ST}) for many loci between populations

- When a locus shows extraordinary levels of genetic distance relative to other loci, this “outlier” locus is a candidate for positive selection

Example of Fst scan – selection under improvement in canola (Brassica napus)

Lu et al. (2019; Nature Comm.)

Maladaptation

Organisms accumulate deleterious mutations, especially in small populations and in genomic regions of low recombination.

Gao et al. 2023, Mol. Plant

Detecting Deleterious Mutations

Approach 1: Predicted effects on protein function

Structure of Lac1, a DNA-binding protein

Ng and Henikoff (2001, Genome Res.)

Detecting Deleterious Mutations

Approach 1: Conservation of amino acids or indels

Sequence logo representation of the LacI DNA binding region

Ng and Henikoff
(2001, Genome Res.)

Deleterious Mutations Accumulate in Regions of Low Recombination

Example of relationship between recombination rate and deleterious load.

Chromosome (linkage group) 10 in cultivated sunflower.

Renaut and Rieseberg
(2010, Mol. Biol. Evol.)

Unanswered questions

- What are the genes that underlie adaptation?
- Is it many genes or a few?
- How repeatable is the genetics of adaptation?
- Do adaptive mutations mainly occur in coding or regulatory regions?
- What is the effect size of adaptive alleles?
- Is it possible to purge deleterious mutations from crops or endangered wild species?