



# Detecting adaptive differentiation in structured populations with genomic data and common gardens

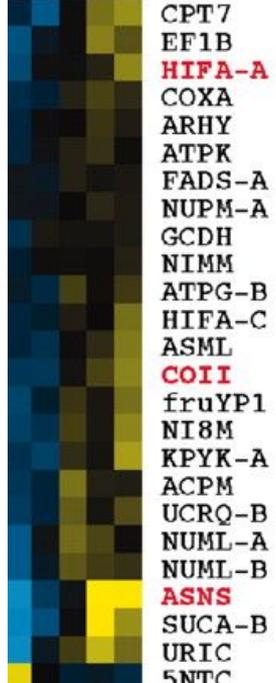
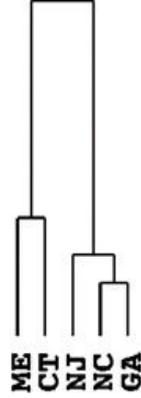
Emily Josephs

 @emjosephs

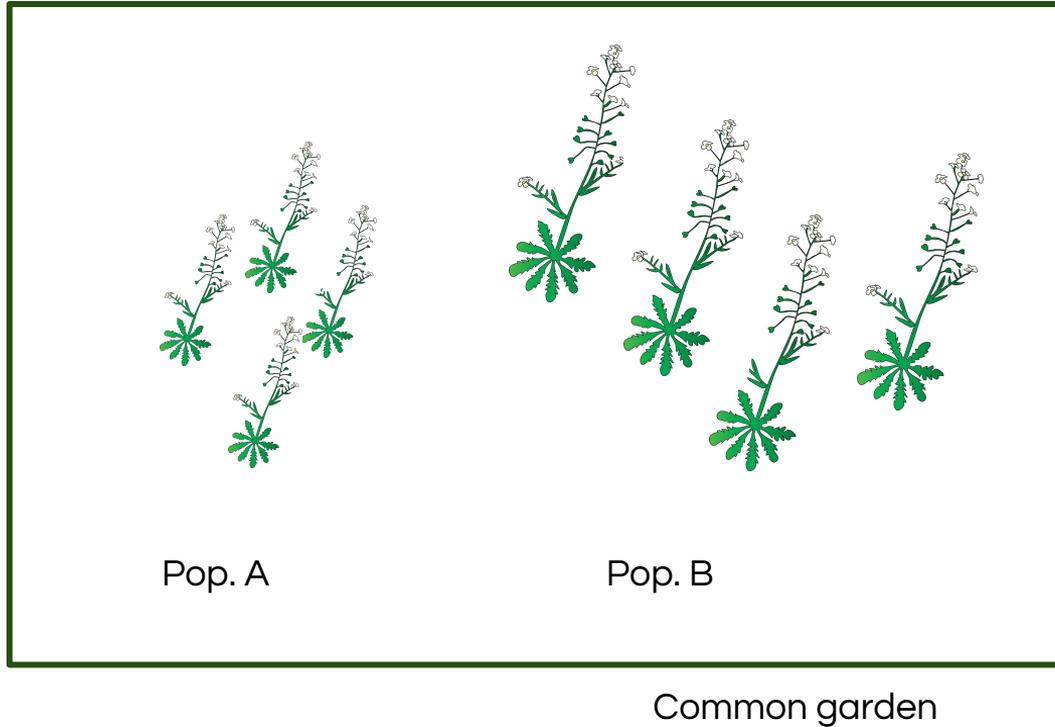


MICHIGAN STATE  
UNIVERSITY

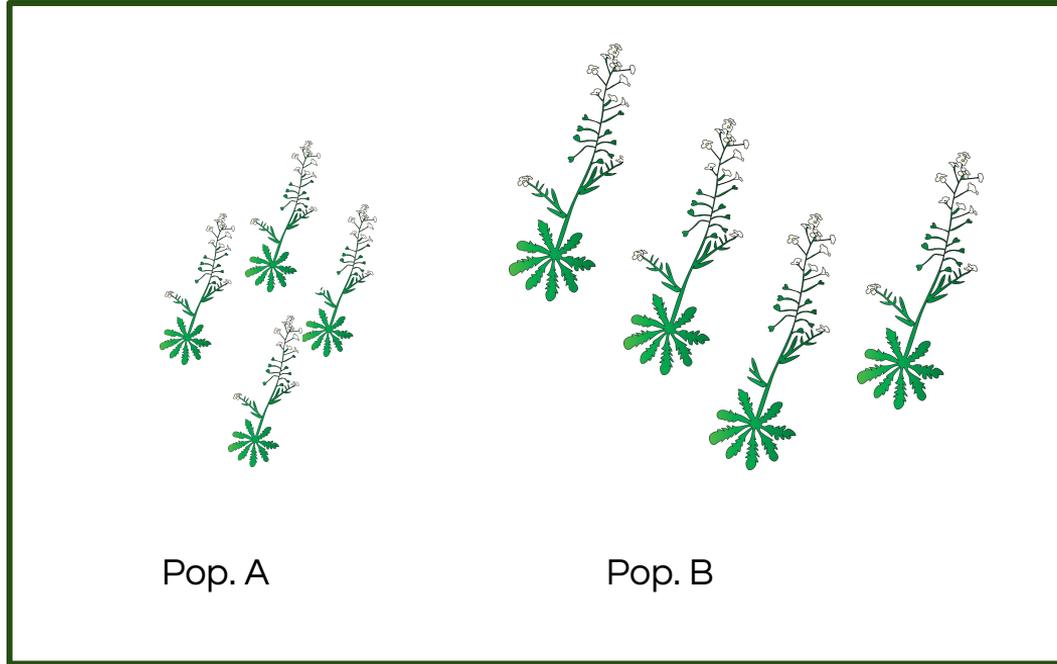




Detecting local adaptation requires knowing about  
(1) genetic variation in traits



Detecting local adaptation requires knowing about  
(1) genetic variation in traits ( $V_A$ )

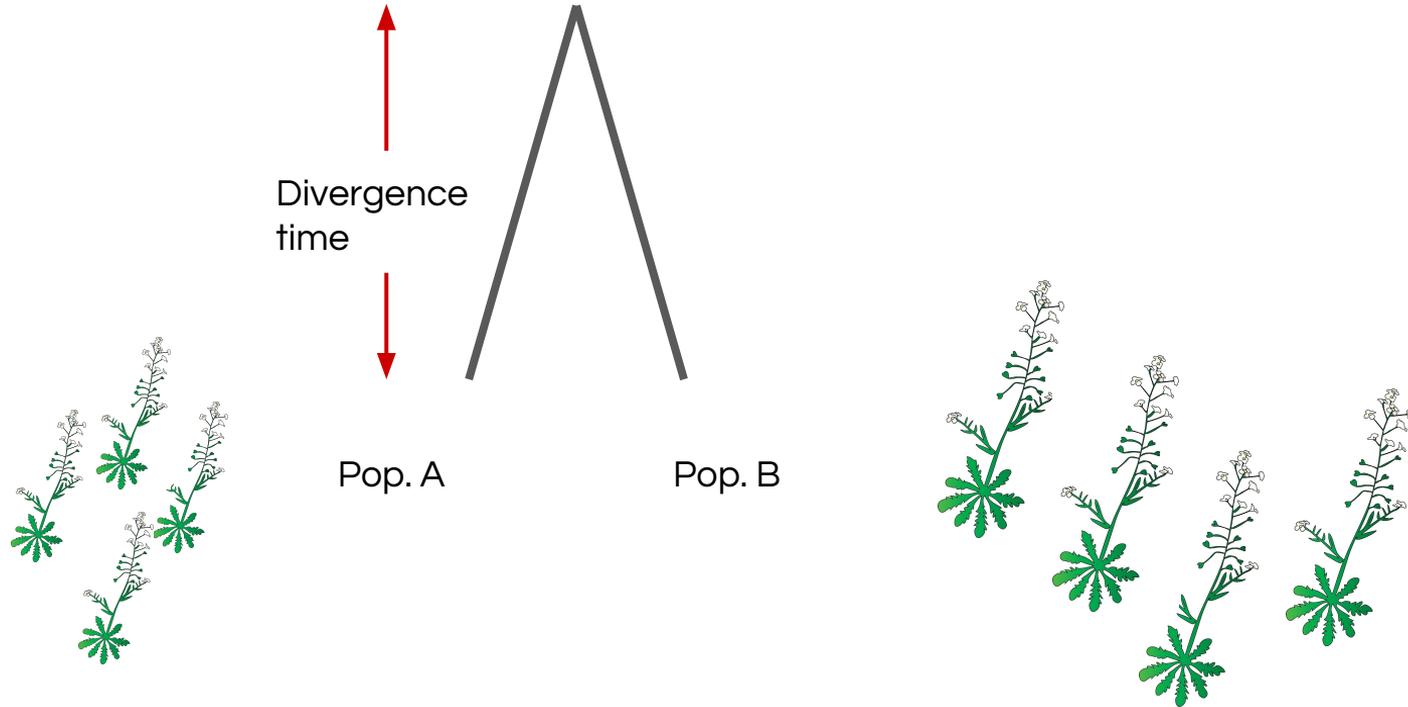


Common garden

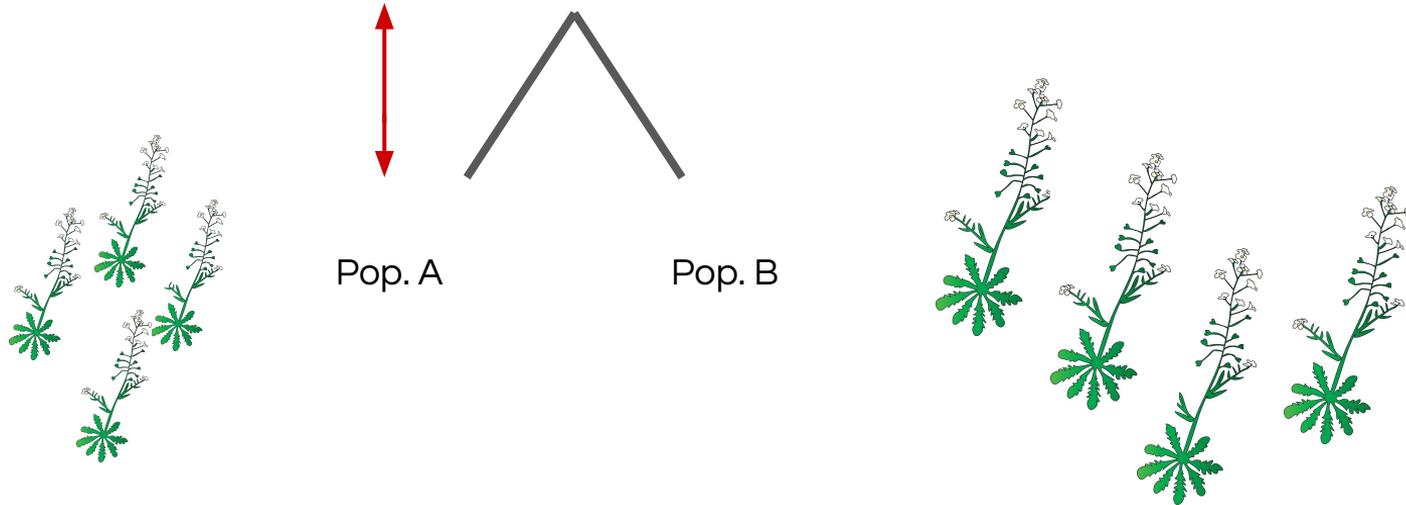
Crosses to get

$V_A$

Detecting local adaptation requires knowing about (1) genetic variation and (2) relatedness



Detecting local adaptation requires knowing about (1) genetic variation and (2) relatedness



# $Q_{st}$ - $F_{st}$ comparisons test for excess trait divergence

$Q_{st}$

vs.

$F_{st}$

btw-pop genetic var. for trait  
total genetic var. for trait

btw-pop neutral genetic var.  
total neutral genetic var.

# $Q_{st}$ - $F_{st}$ comparisons test for excess trait divergence

$$Q_{st} = F_{st}$$

btw-pop genetic var. for trait  
total genetic var. for trait

btw-pop neutral genetic var.  
total neutral genetic var.

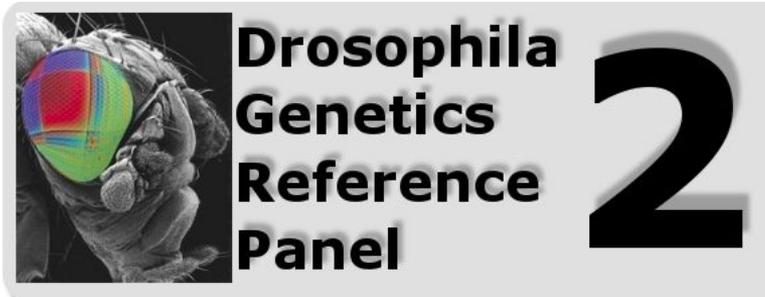
# $Q_{st}$ - $F_{st}$ comparisons test for excess trait divergence

$$Q_{st} > F_{st}$$

btw-pop genetic var. for trait  
total genetic var. for trait

btw-pop neutral genetic var.  
total neutral genetic var.

Lots of datasets that have genomes + phenotypes from diversity panels



**Drosophila  
Genetics  
Reference  
Panel** **2**

The image shows a close-up of a Drosophila melanogaster head with a colorful heatmap overlaid on its eye, representing genetic data. The heatmap has a color scale from red to blue. To the right of the image is a large black number '2'.



1001 Genomes

### 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*

The 1001 Genomes Consortium<sup>1,\*</sup>

<sup>1</sup>Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany

\*Correspondence: [magnus.nordborg@gmi.oeaw.ac.at](mailto:magnus.nordborg@gmi.oeaw.ac.at) (Magnus Nordborg), [weigel@weigelworld.org](mailto:weigel@weigelworld.org) (Detlef Weigel)

<http://dx.doi.org/10.1016/j.cell.2016.05.063>

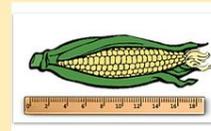
### Panzea Data Sets and Genotype Search Tools

From here you can obtain the genotypic and phenotypic datasets produced by this project as flat files. You can also perform custom searches on our large genotypic data sets, to obtain smaller subsets of data for chromosomal regions or taxa subsets of interest. Two different genotype searches are available: one for GBS genotypes, and another for other types of genotypes (from maize HapMapVI and VZ, the MaizeSNP50 Bead Chip, and traditional SNPs).



Genotypic Data  
Sets

From this page you can obtain the latest public, "flat file" versions of prepackaged, genotypic data sets. All of these can be opened with the TASSEL 5 (GUI or command line). The genotypes were obtained by whole genome sequencing ("HapMap"), GBS, or with MaizeSNP50 chip or traditional SNP assays.



Phenotypic  
Data Sets

From this page you can obtain the latest public, "flat file" versions of prepackaged, phenotypic data sets. These phenotypes are from the maize 282 association panel, the NAM population, the USDA Ames inbreds, and custom teosinte populations for studying domestication traits.

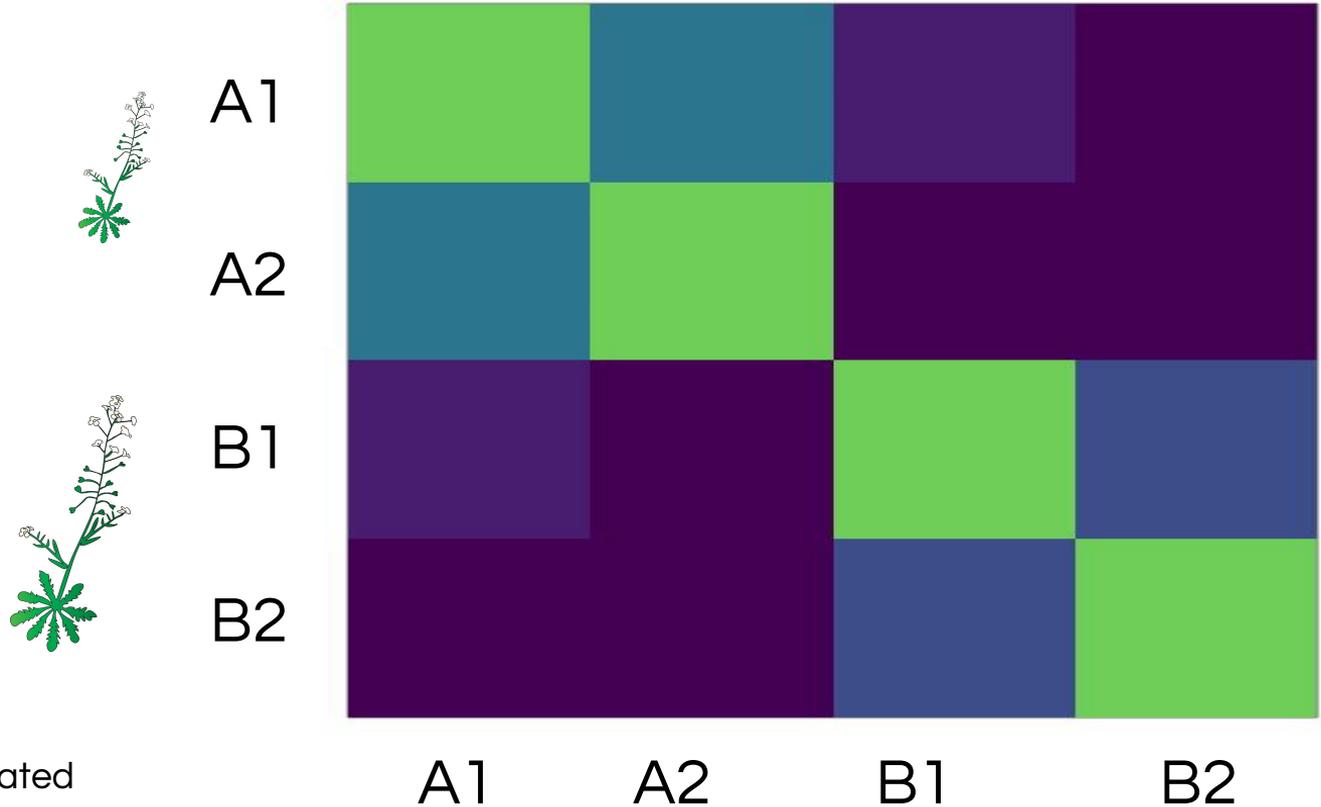
How can we systematically detect local adaptation...

In quantitative traits?

In gene expression?

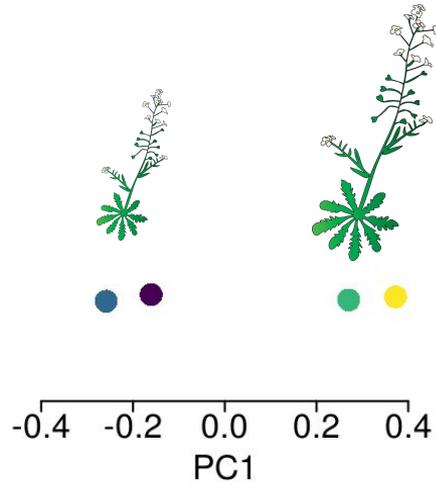
For GxE?

Relatedness between populations can be summarized with a kinship matrix.

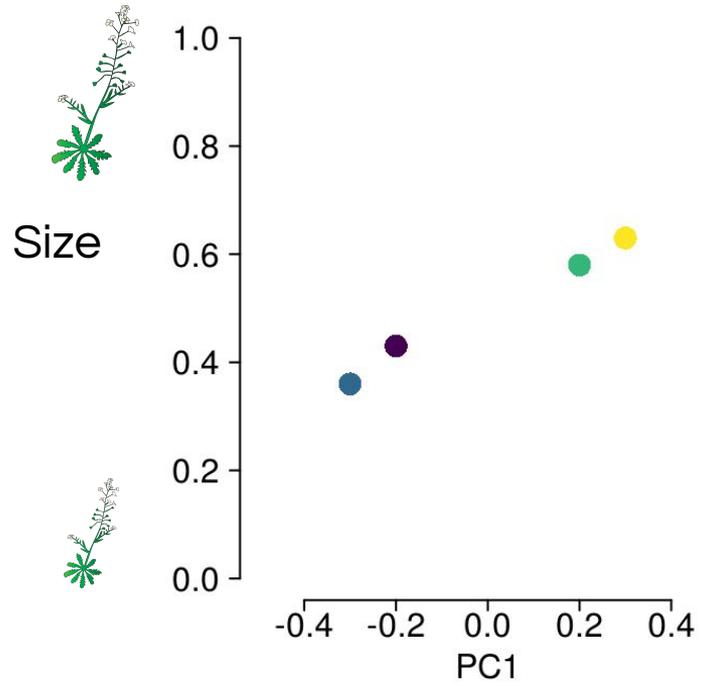


Lighter colors = more related

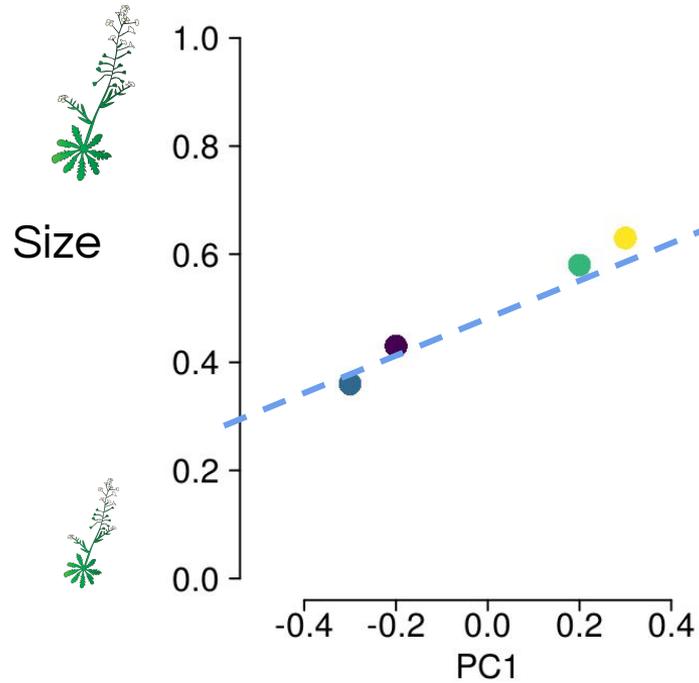
The eigenvectors of the kinship matrix (PCs) summarize relatedness



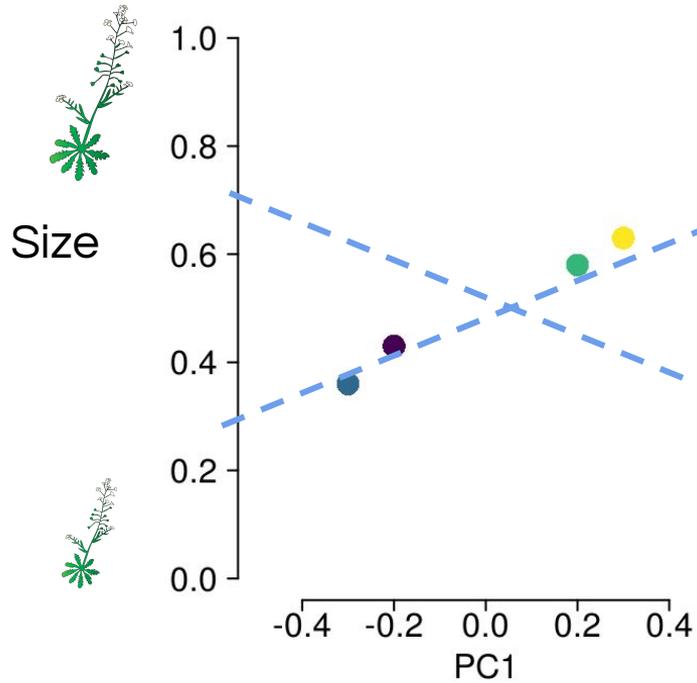
# How do PCs relate to traits?



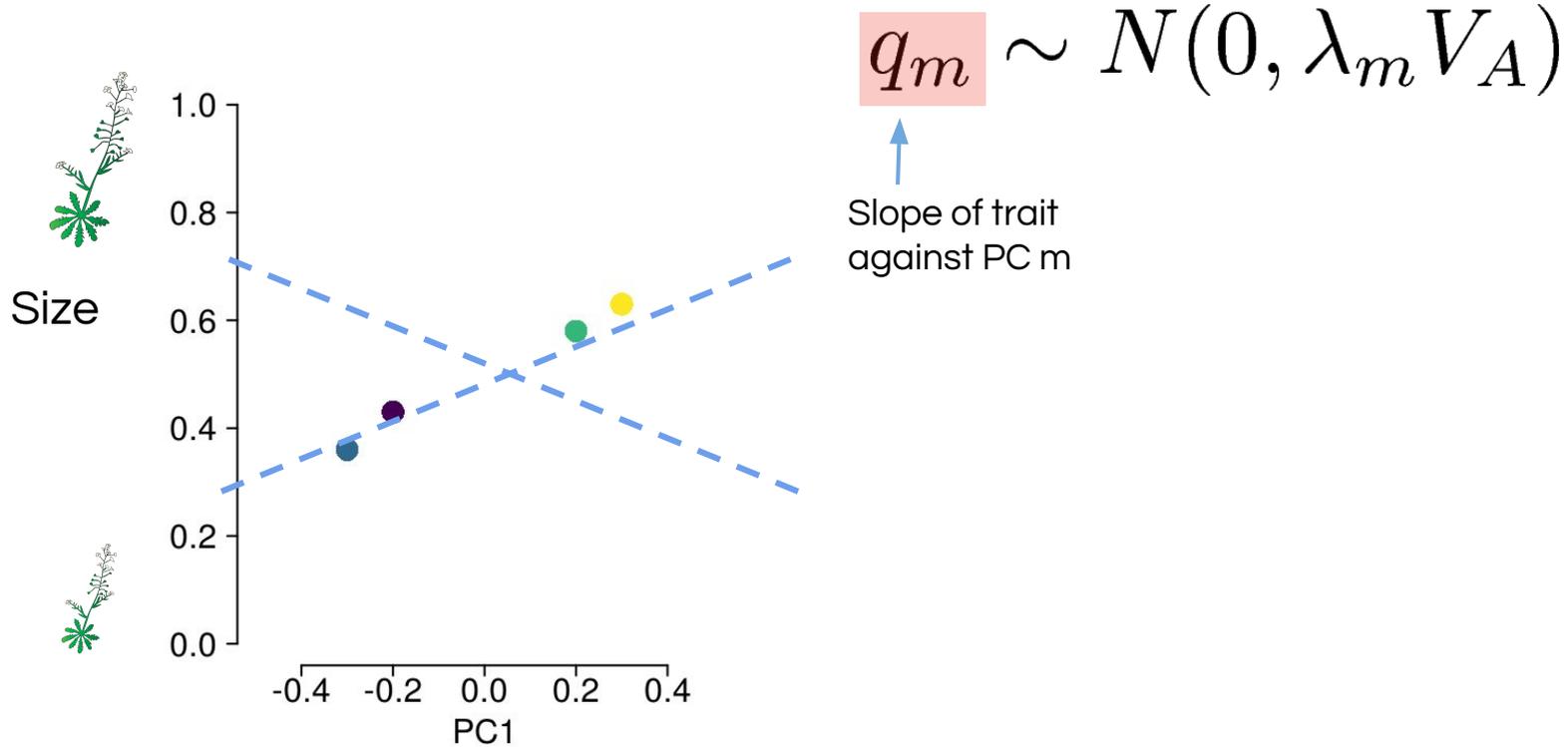
# A correlation btw PC & trait can be consistent with drift



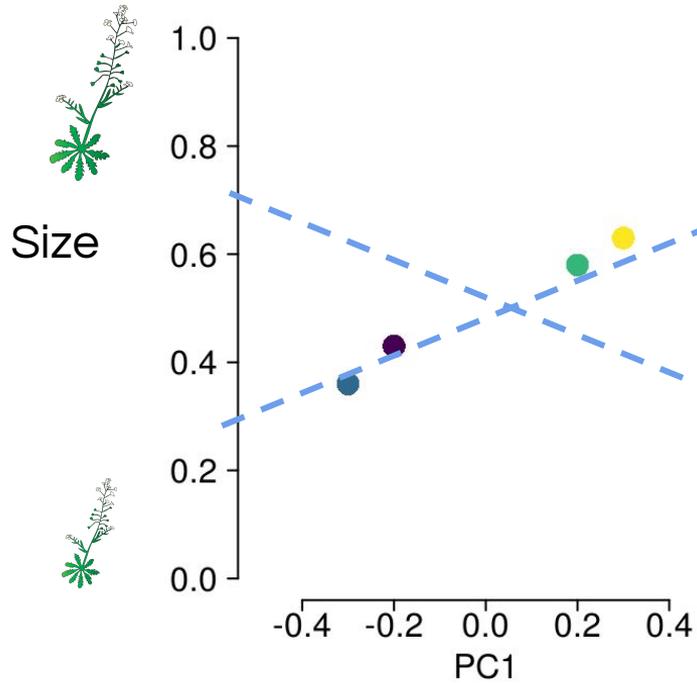
# A correlation btw PC & trait can be consistent with drift



# Modelling the slope expected due to drift



# Modelling the slope expected due to drift

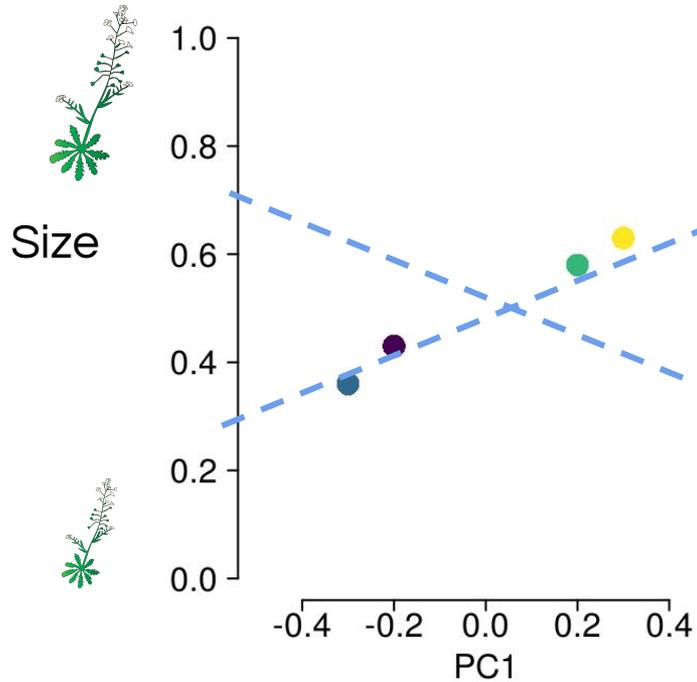


$$q_m \sim N(0, \lambda_m V_A)$$

↑  
Slope of trait  
against PC m

↑  
Mean slope

# Modelling the slope expected due to drift



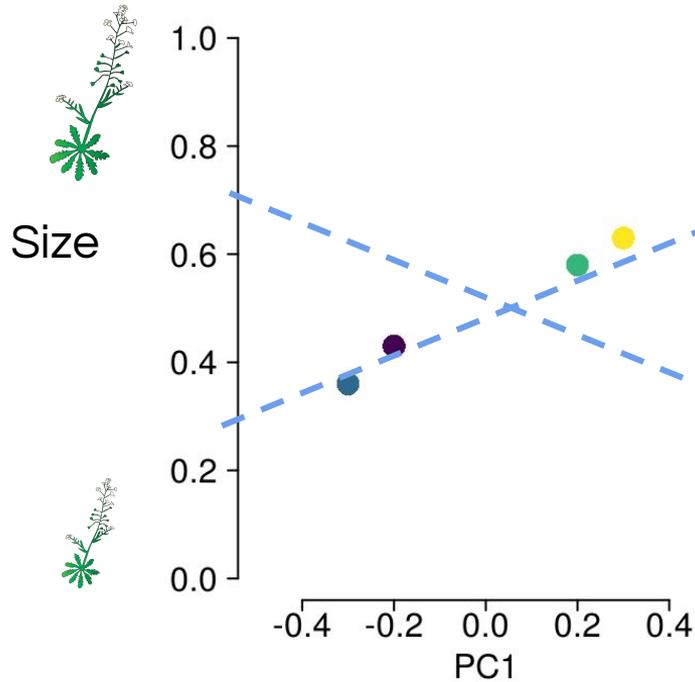
$$q_m \sim N(0, \lambda_m V_A)$$

Slope of trait against PC m

Mean slope

Amount of relatedness explained by PC m

# Modelling the slope expected due to drift

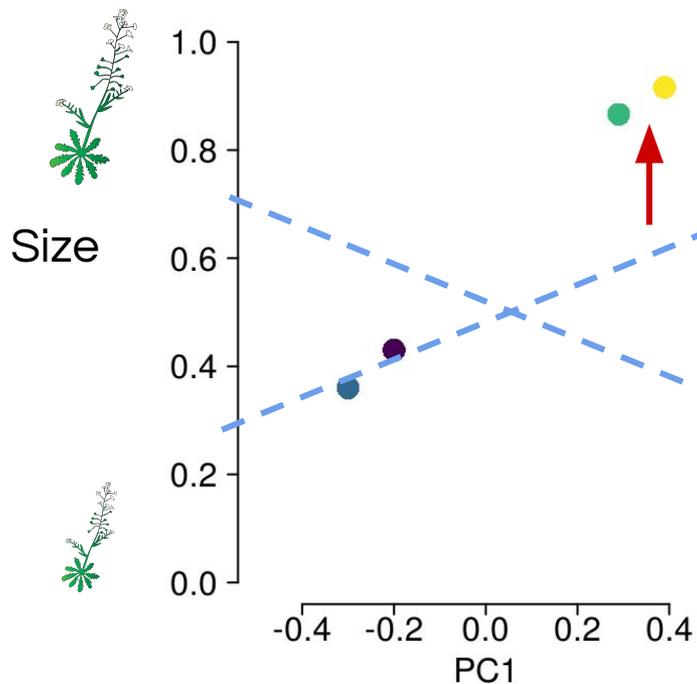


$$q_m \sim N(0, \lambda_m V_A)$$

Annotations for the equation:

- Slope of trait against PC m (points to  $q_m$ )
- Mean slope (points to 0)
- Amount of relatedness explained by PC m (points to  $\lambda_m$ )
- Can estimate  $V_A$  from a subset of PCs (points to  $V_A$ )

# Selection can increase trait divergence



$$q_m \sim N(0, \lambda_m V_A)$$

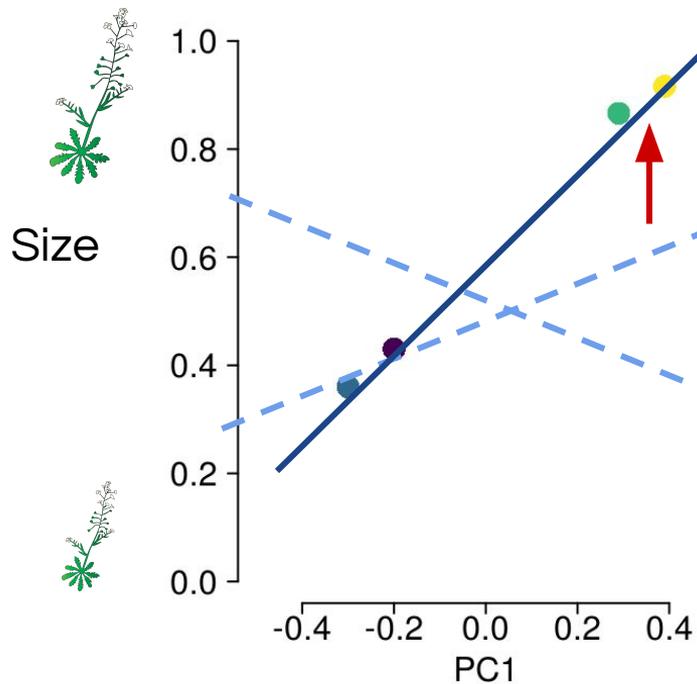
Slope of trait  
against PC m

Mean slope

Amount of  
relatedness  
explained by PC m

Can estimate  $V_A$   
from a subset of PCs

# Selection can increase trait divergence



$$q_m \sim N(0, \lambda_m V_A)$$

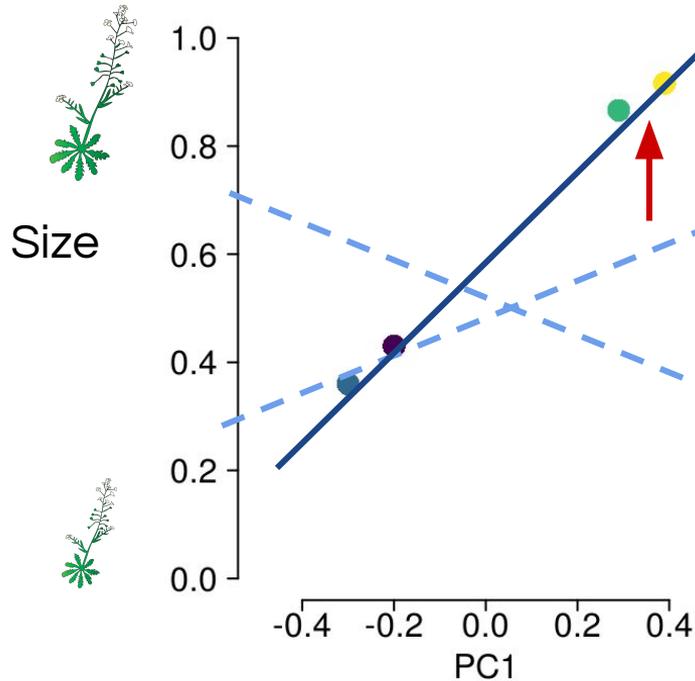
Slope of trait  
against PC m

Mean slope

Amount of  
relatedness  
explained by PC m

Can estimate  $V_A$   
from a subset of PCs

# Testing for diversifying selection



$$q_m \sim N(0, \lambda_m V_A)$$

Slope of trait  
against PC m

Mean slope

Amount of  
relatedness  
explained by PC m

Can estimate  $V_A$   
from a subset of PCs

$$\frac{q_m^2}{V_A \lambda_m} \sim F(1, N)$$

Number of PCs used  
to estimate  $V_A$

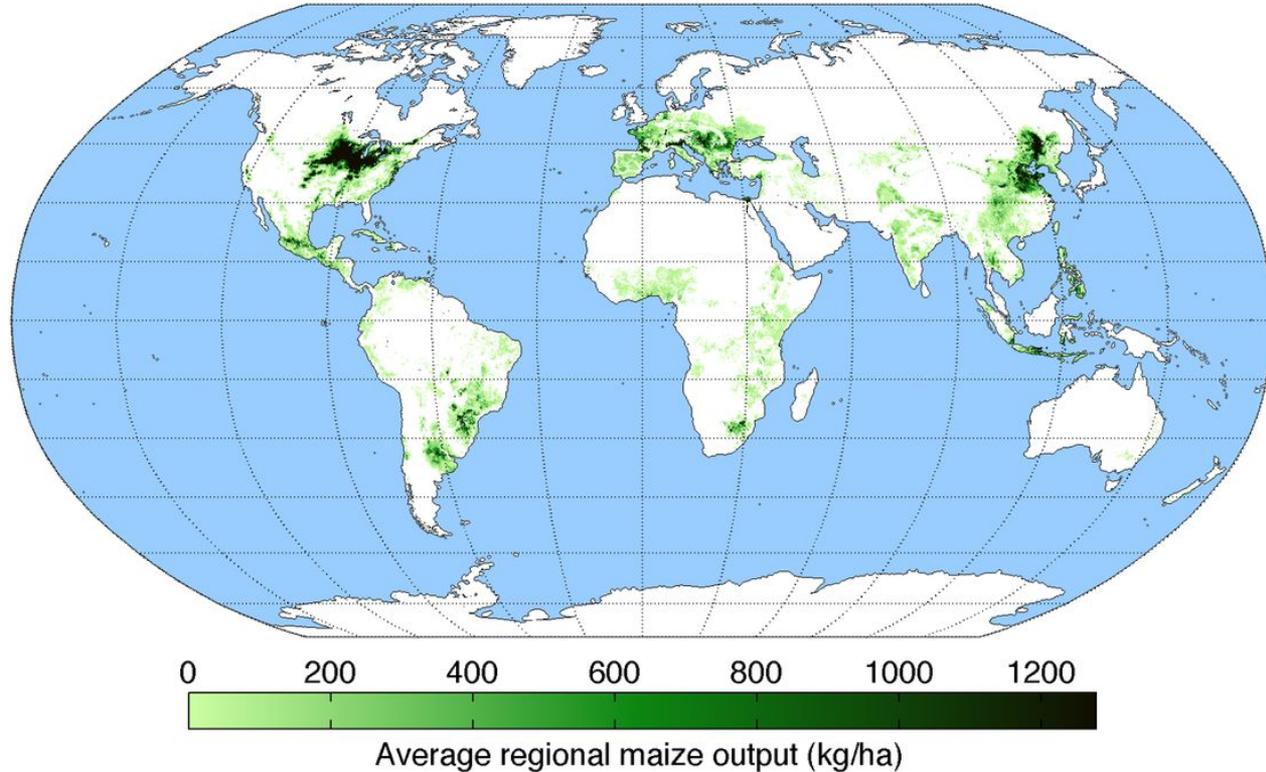
How can we systematically detect local adaptation...

In quantitative traits?

In gene expression?

For GxE?

# Detecting local adaptation in maize



# Detecting local adaptation in maize

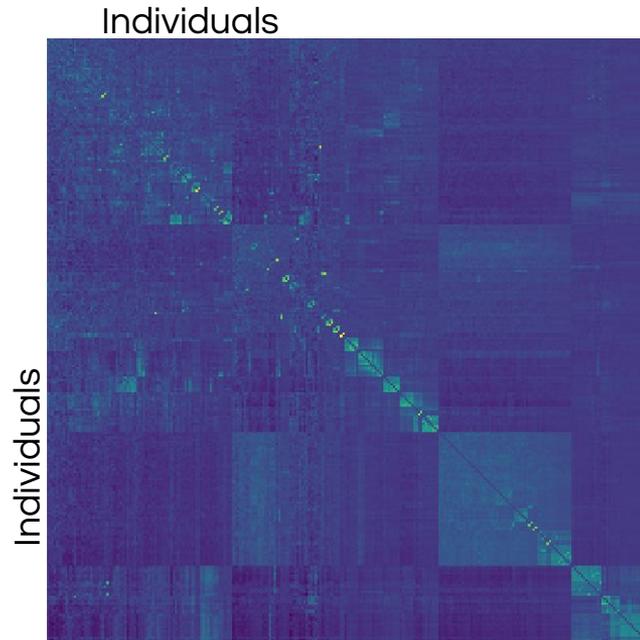
240 domesticated inbred maize lines (Flint-Garcia et al. 2005).

Whole genome sequence data (Panzea).

22 traits measured in common garden.

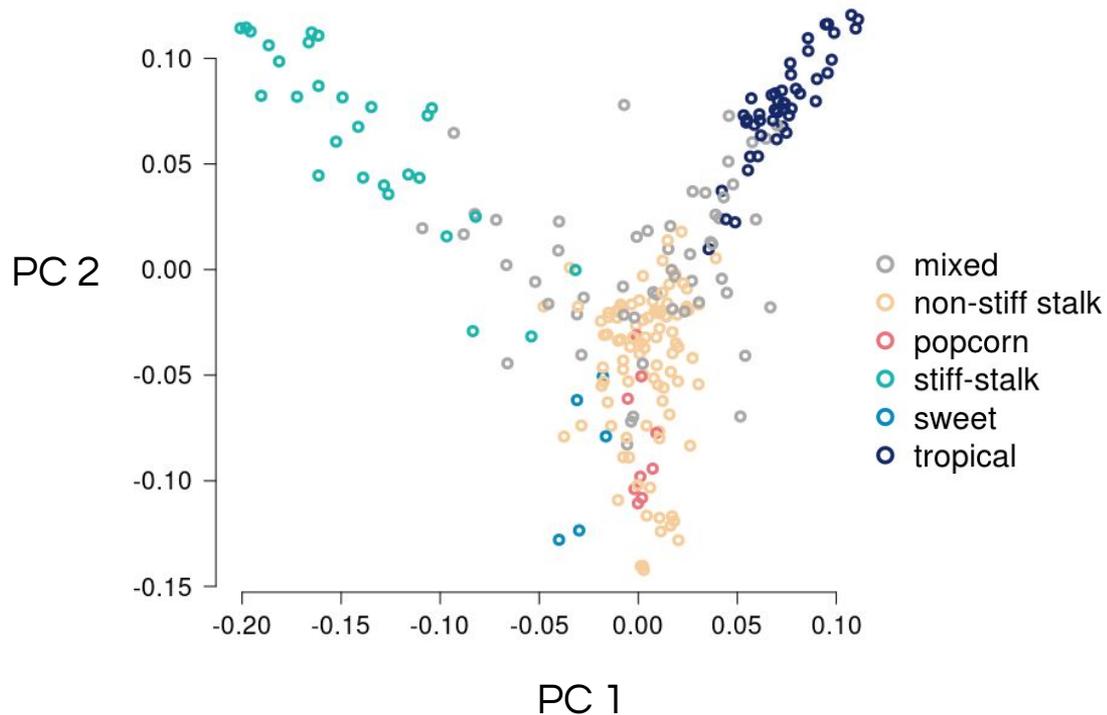
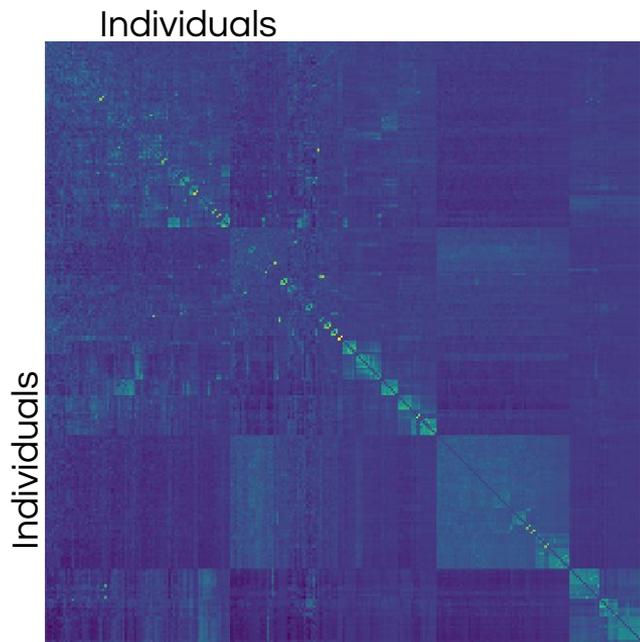


# The kinship matrix of 240 maize lines



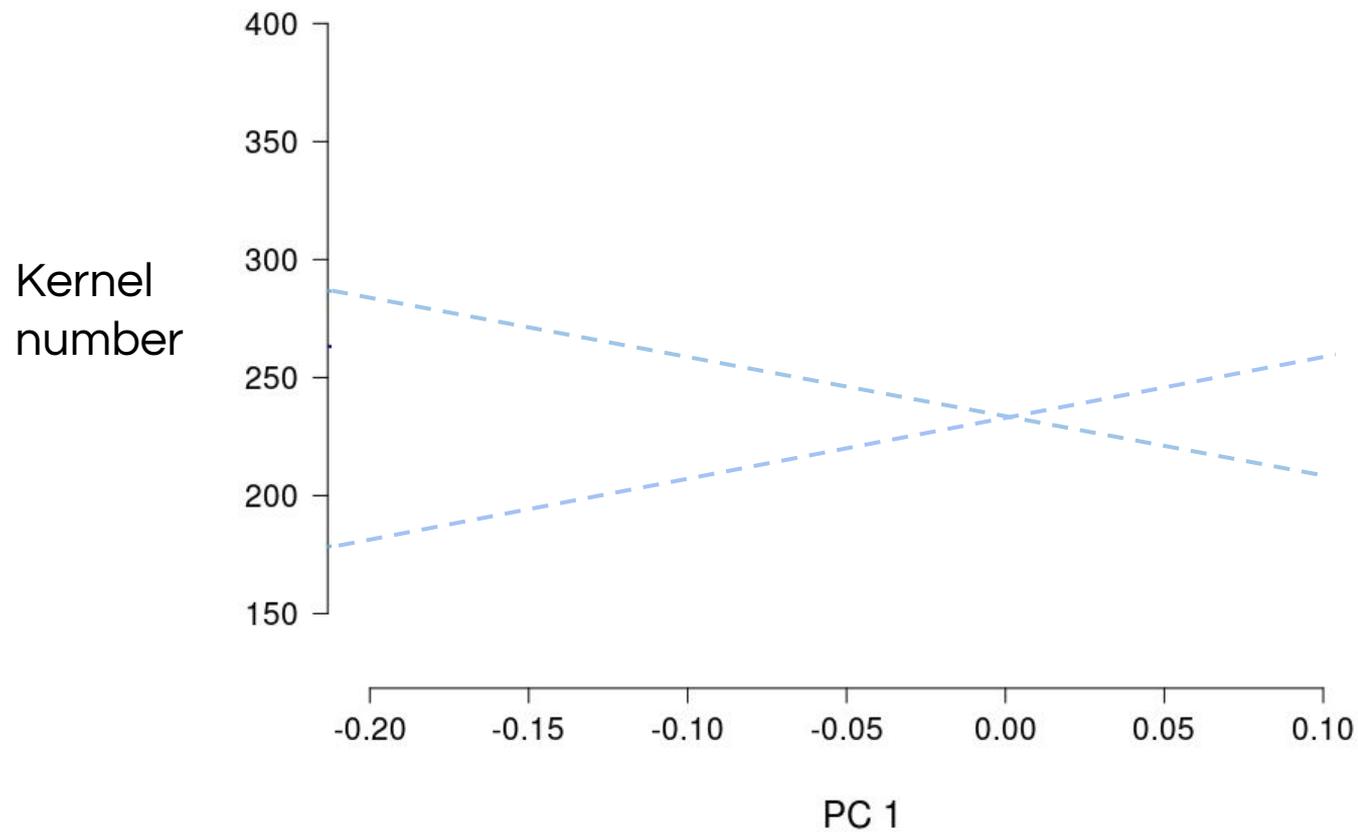
Lighter colors = more related

# The kinship matrix of 240 maize lines

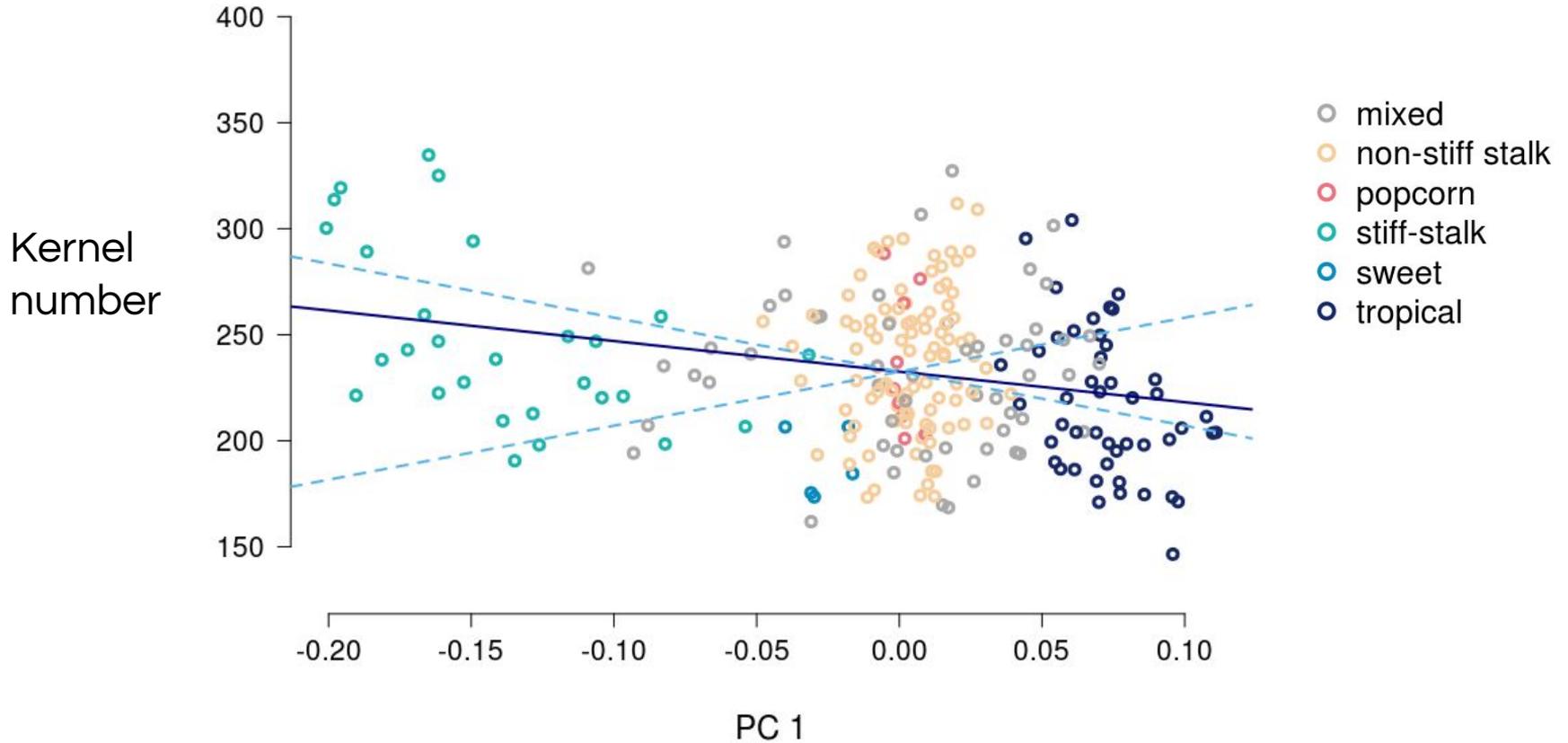


Lighter colors = more related

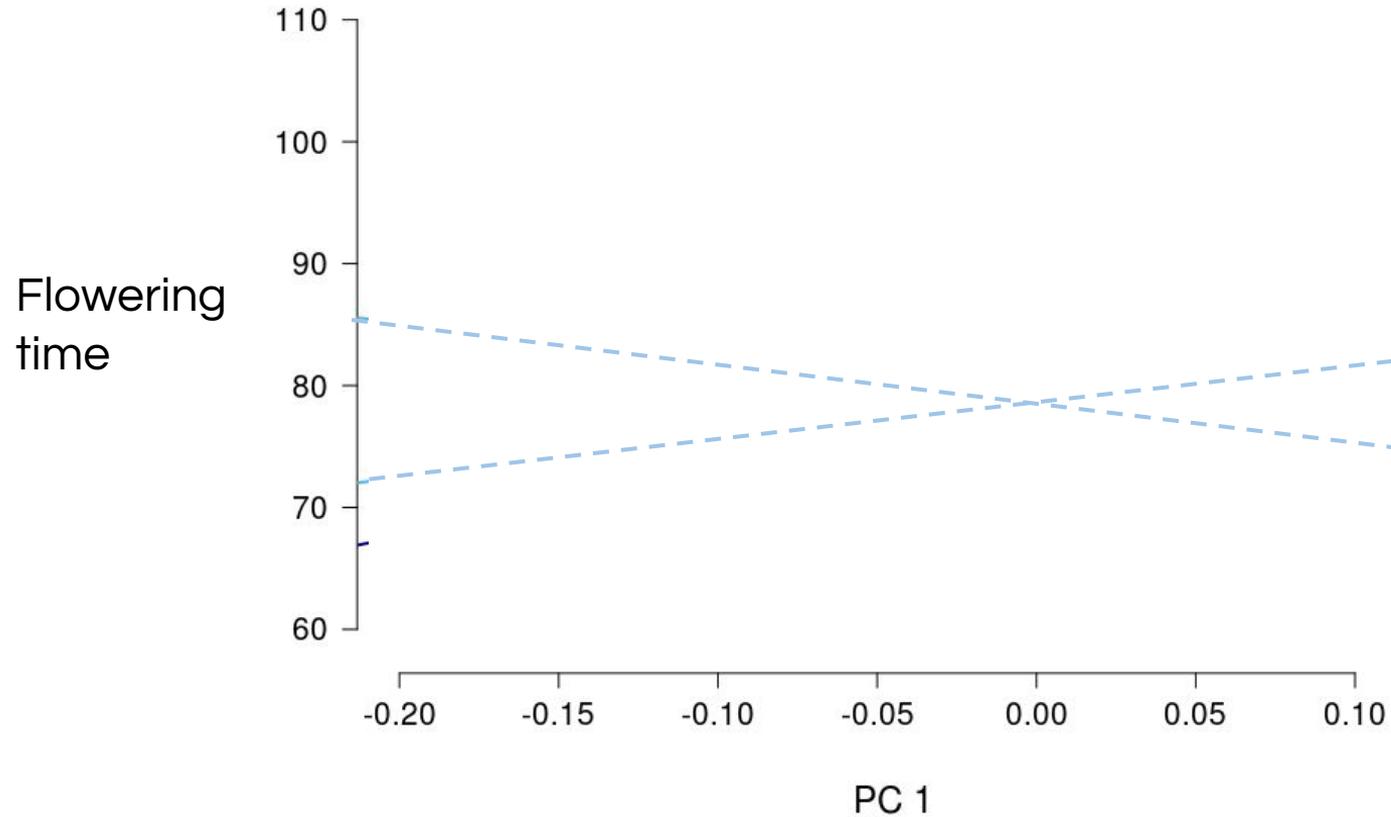
# Neutral expectations of relationship between PC and trait



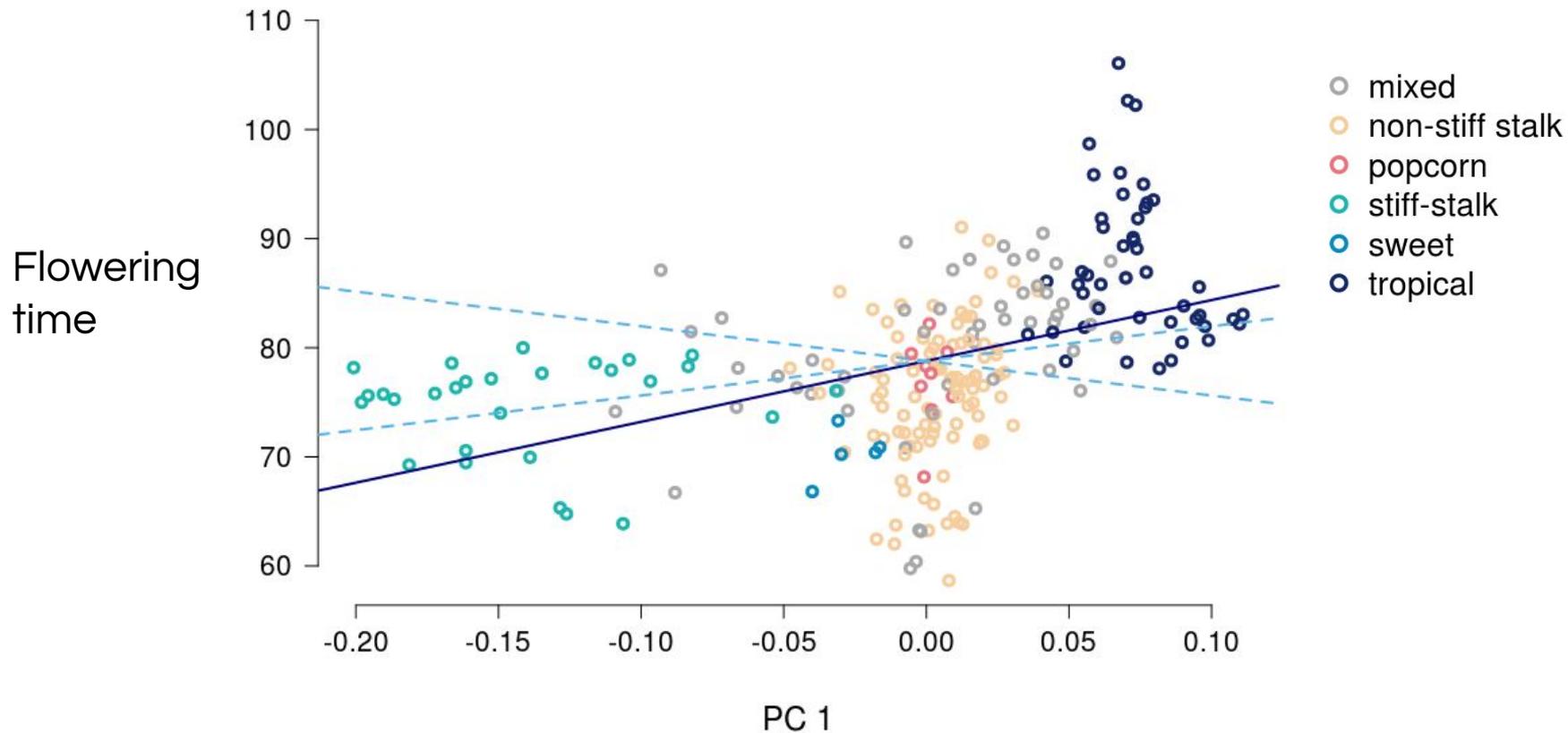
# Divergence along PC1 consistent with drift



# Neutral expectations of relationship between PC and trait

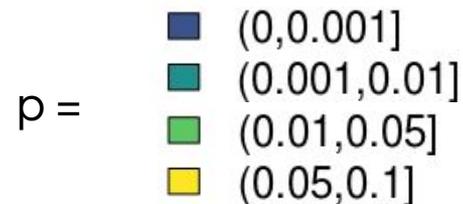
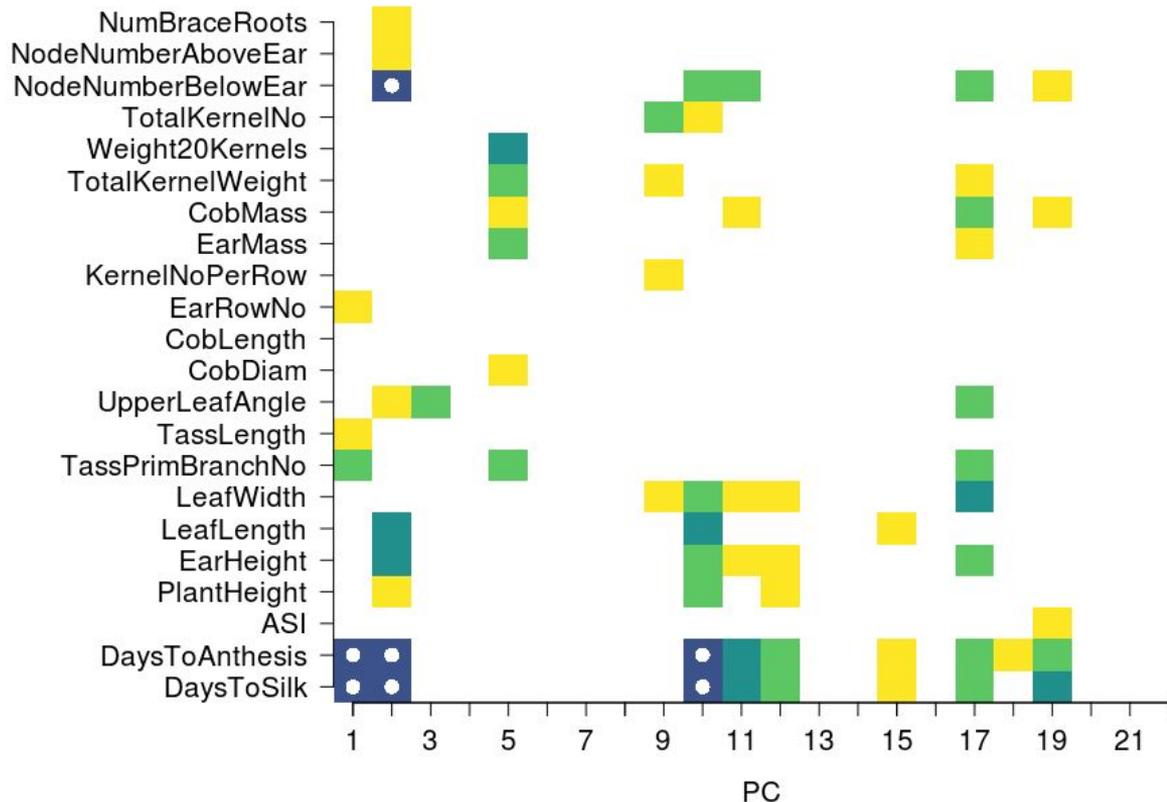


# Adaptive divergence in flowering time along PC 1



FDR < 0.05

# Signatures of adaptation across traits



White circles mean  
FDR < 0.05

How can we systematically detect local adaptation...

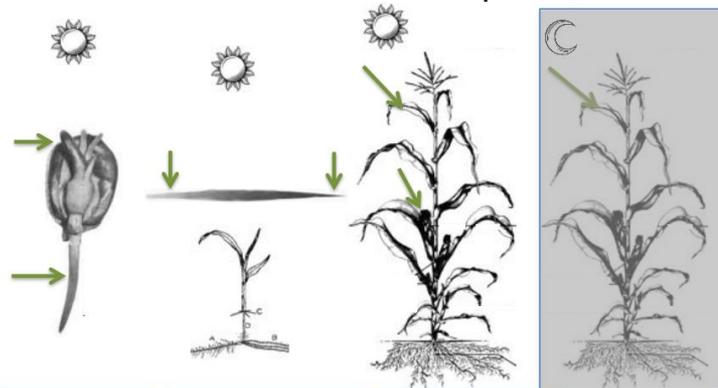
In quantitative traits?

In gene expression?

For GxE?

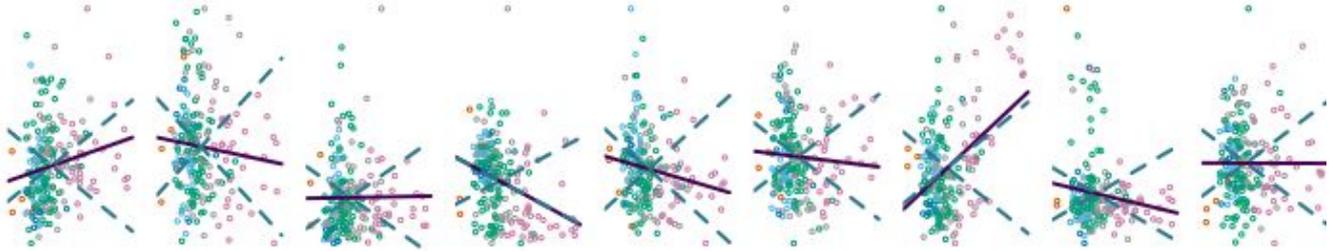
# Testing for selection on gene expression

Gene expression data (RNAseq)  
for 208 maize lines in 7 tissues

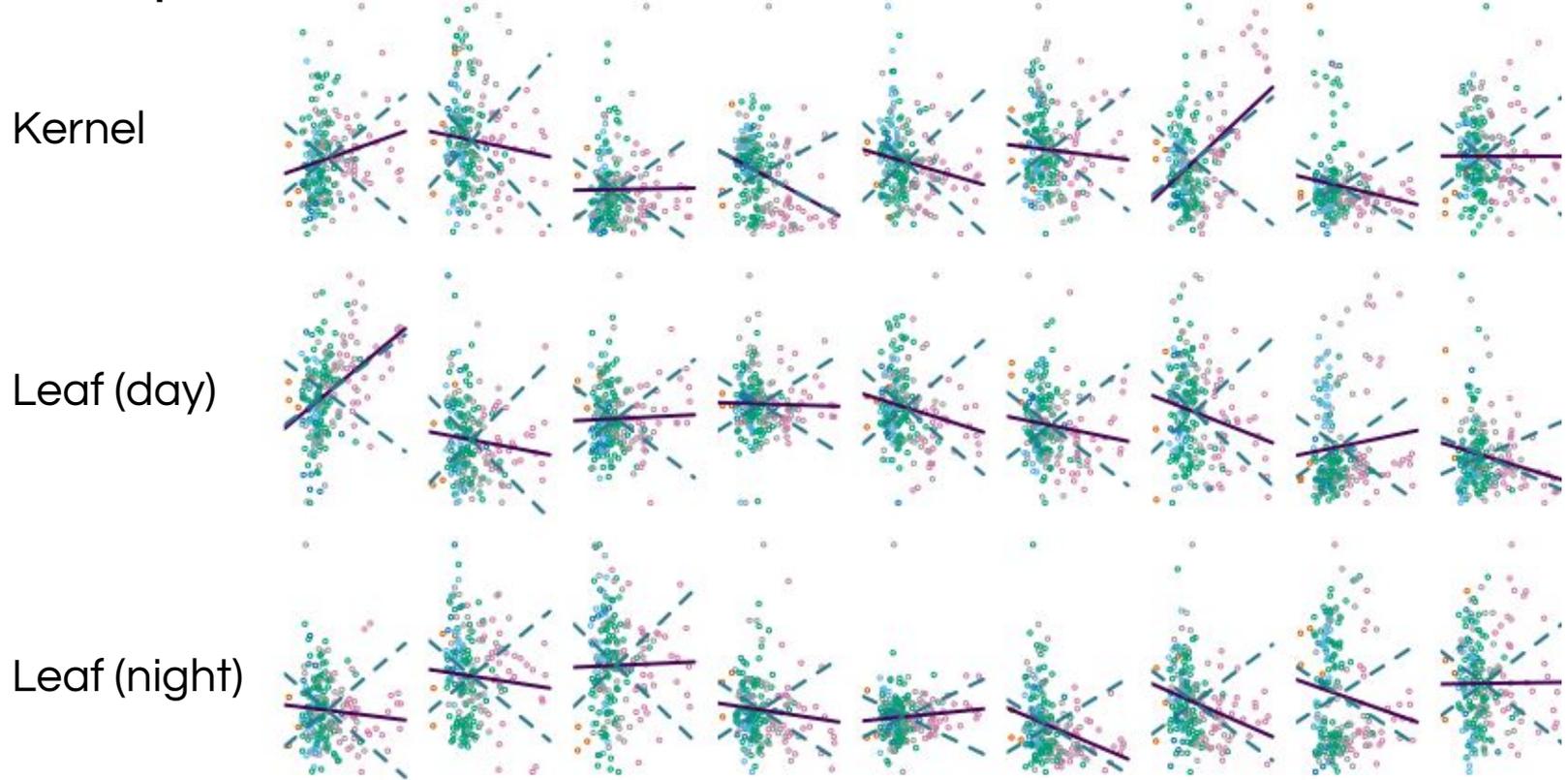


# How many genes show evidence of selection on expression?

Kernel



# Which tissues show strongest evidence of adaptation?



# Evidence of local adaptation for gene expression

|                     |     |     |     |
|---------------------|-----|-----|-----|
| Leaf night (day 8)  | 1   |     |     |
| Leaf night (day 26) | 4   |     | 2   |
| Leaf day (day 8)    | 1   | 1   | 3   |
| Leaf day (day 26)   | 1   | 6   | 2   |
| 3rd leaf tip        |     |     |     |
| 3rd leaf base       |     | 15  |     |
| Kernel              | 9   | 10  | 21  |
| Seedling shoot      |     | 2   | 2   |
| Seedling root       |     |     |     |
|                     | PC1 | PC2 | PC3 |

Number of genes that show evidence of selection on expression



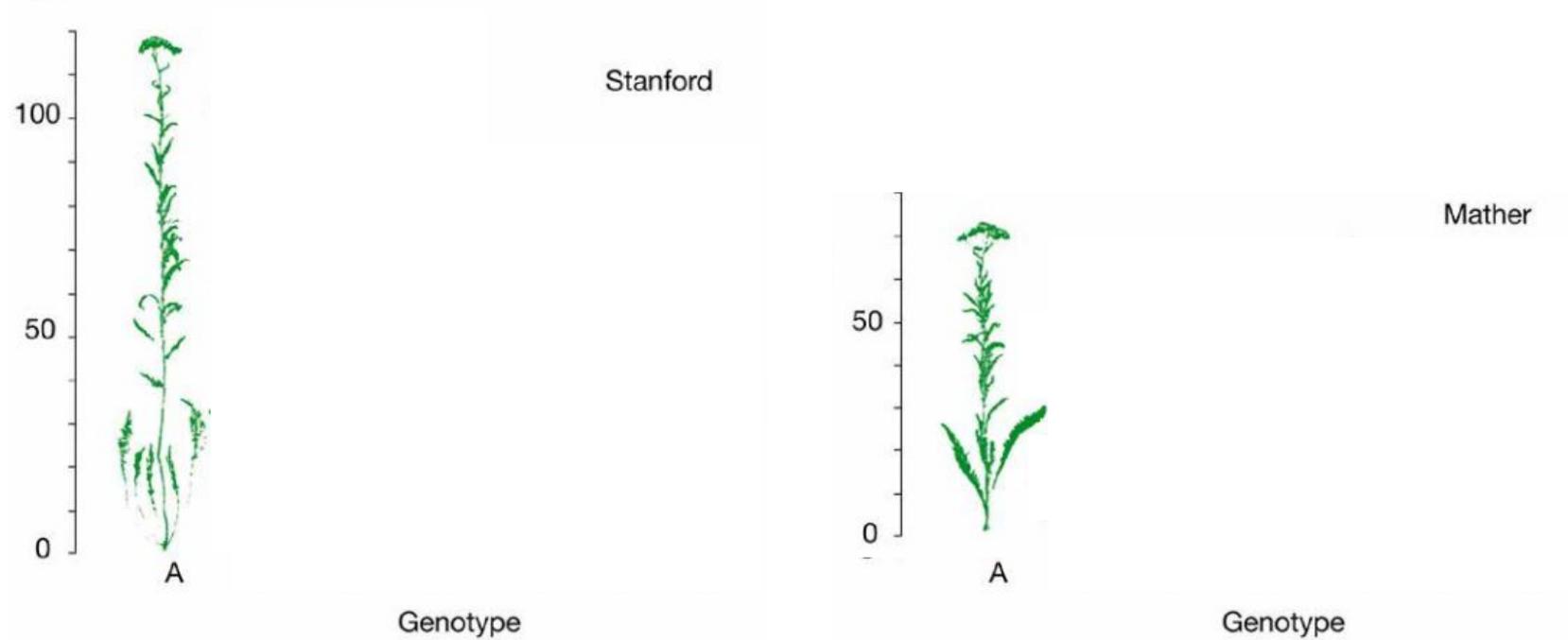
How can we systematically detect local adaptation...

In quantitative traits?

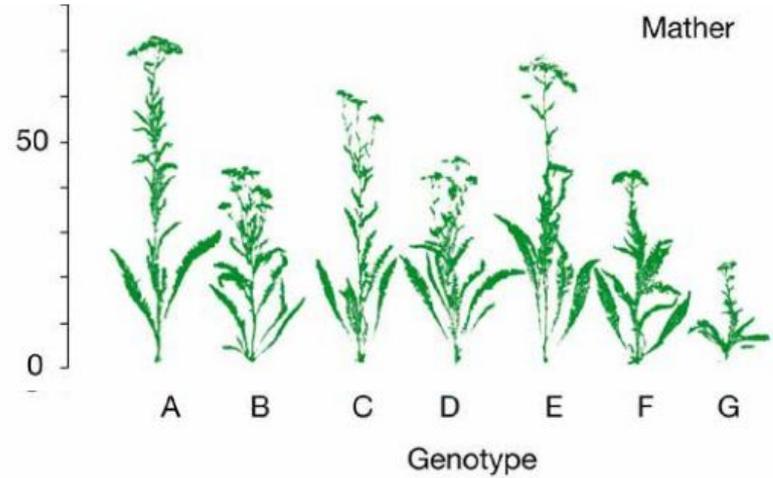
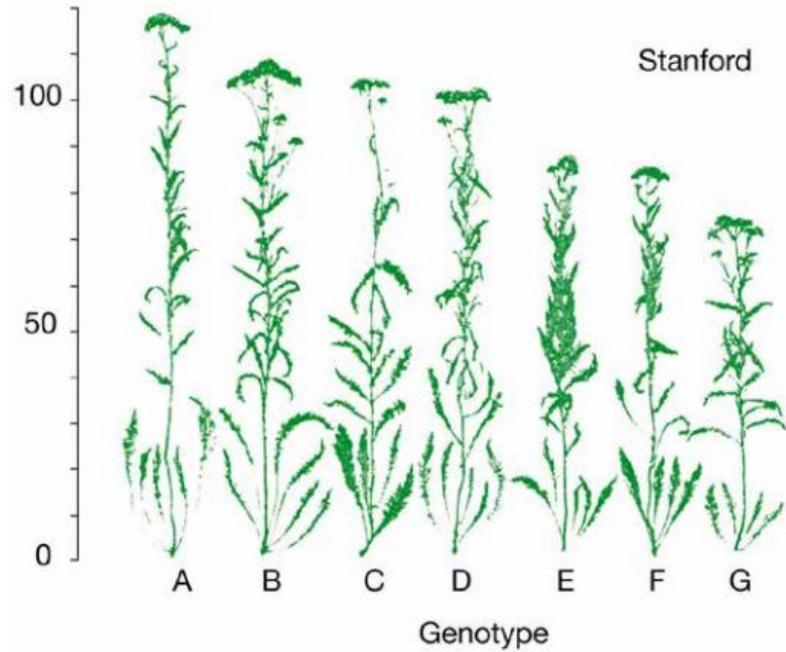
In gene expression?

For GxE?

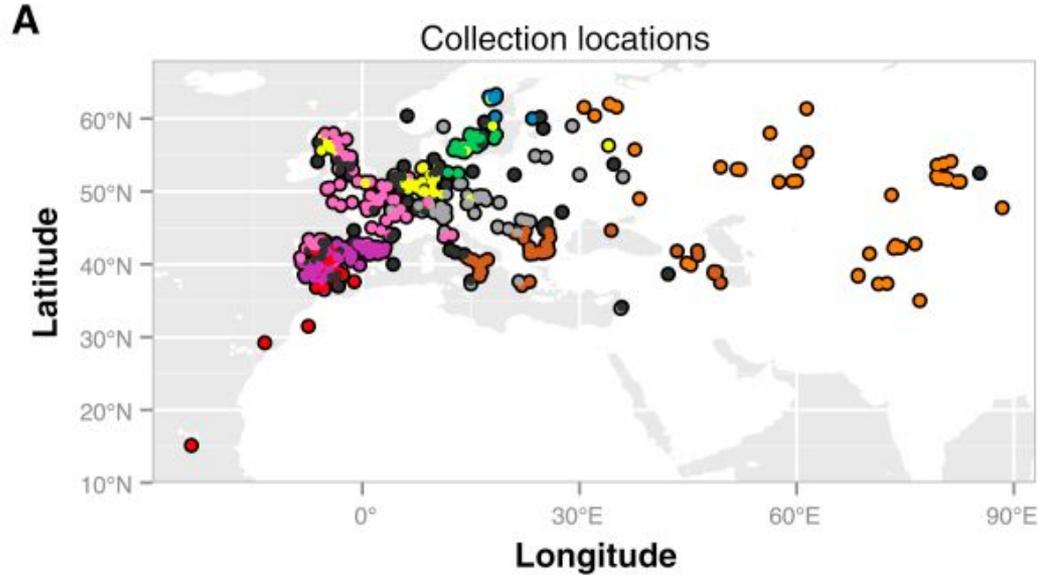
# The environment shapes traits



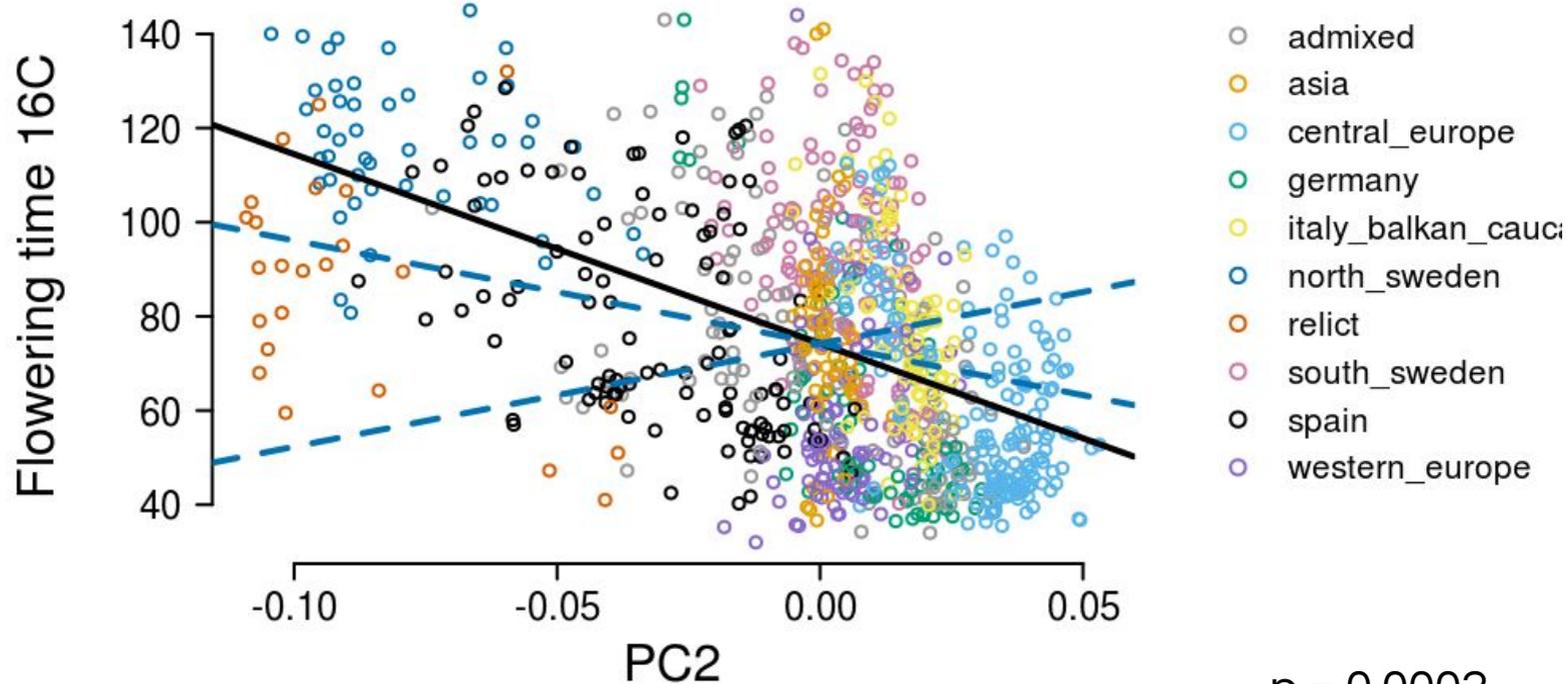
# Genetic variation for plasticity ('GxE')



# *A. thaliana* flowering time as a model for GxE



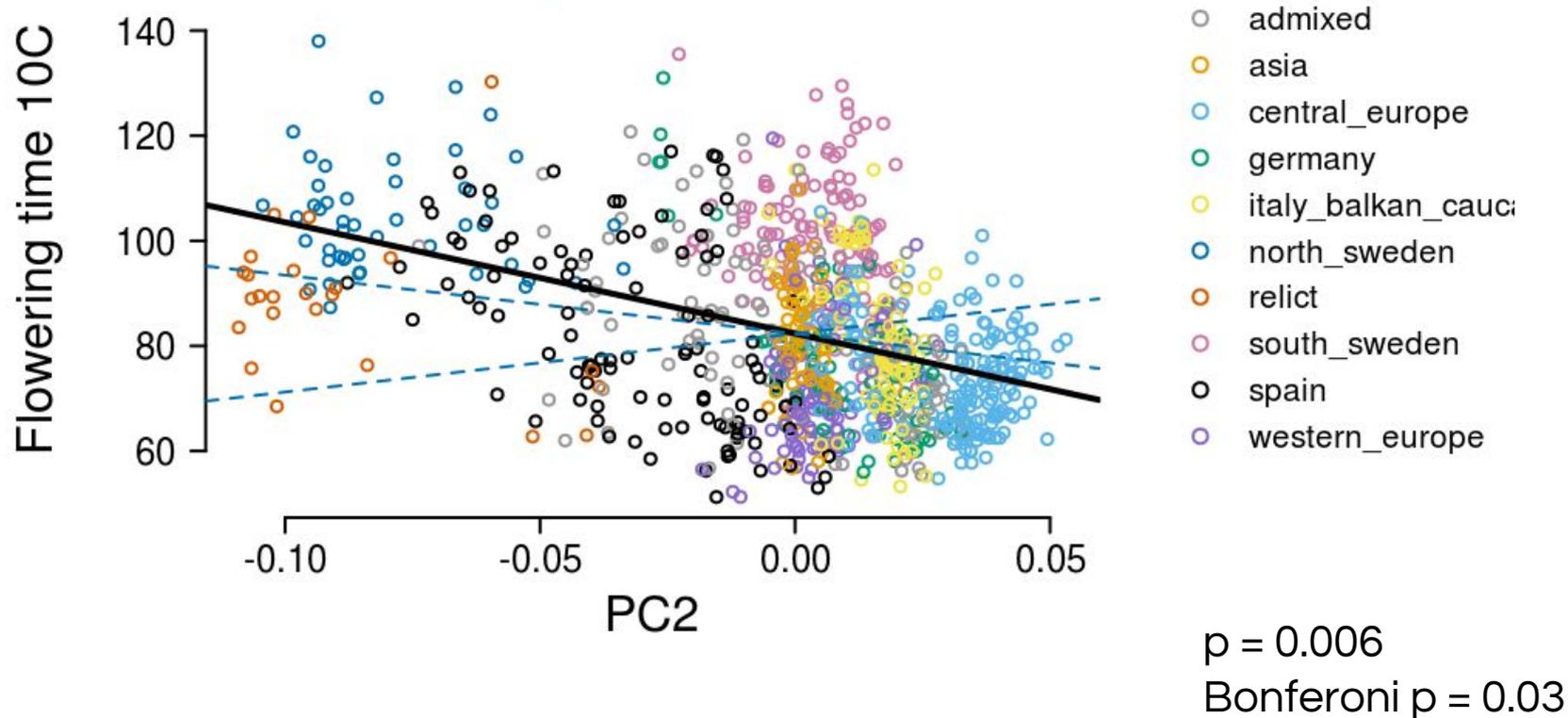
# Flowering time at 16°C is locally adapted



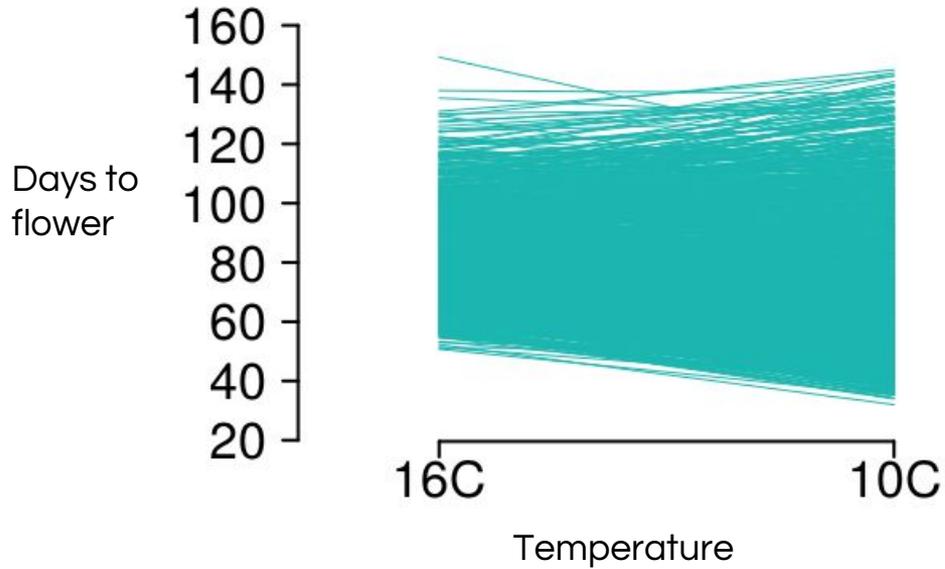
$p = 0.0002$

Bonferoni  $p = 0.0015$

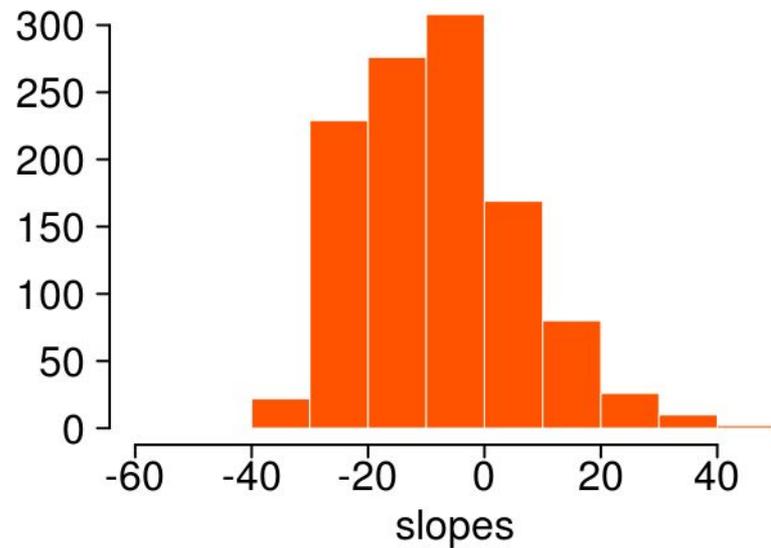
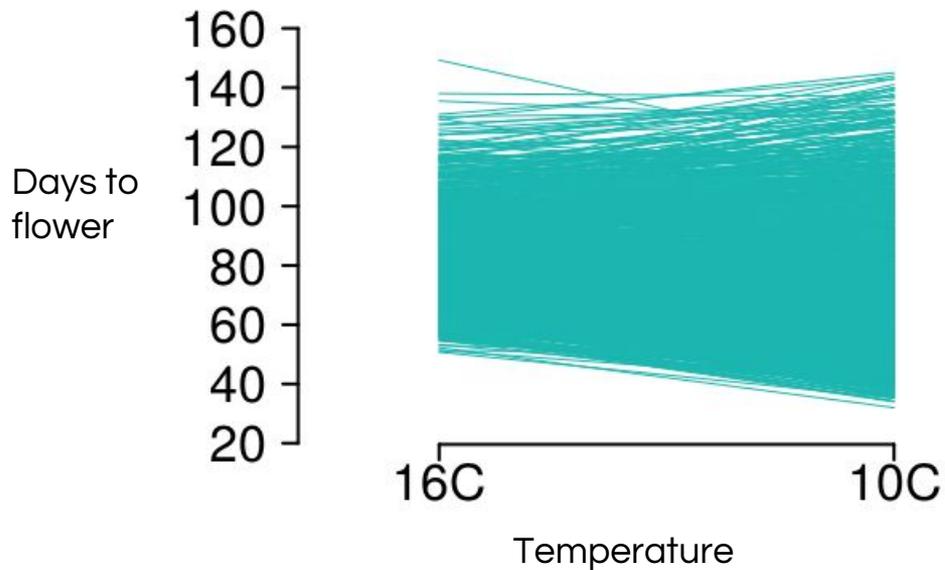
# Flowering time at 10°C is also locally adapted



# Using the reaction norm to measure plasticity



# Using the reaction norm to measure plasticity

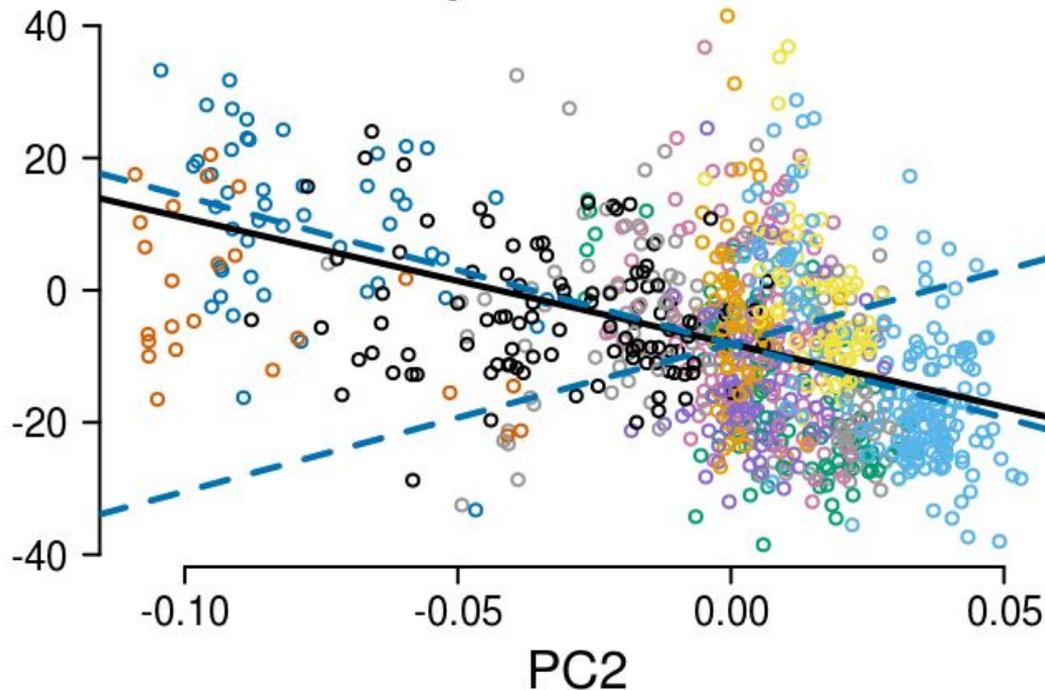


# Variation in reaction norm consistent with drift

Later flowering at  
16°C than 10°C

Reaction norm

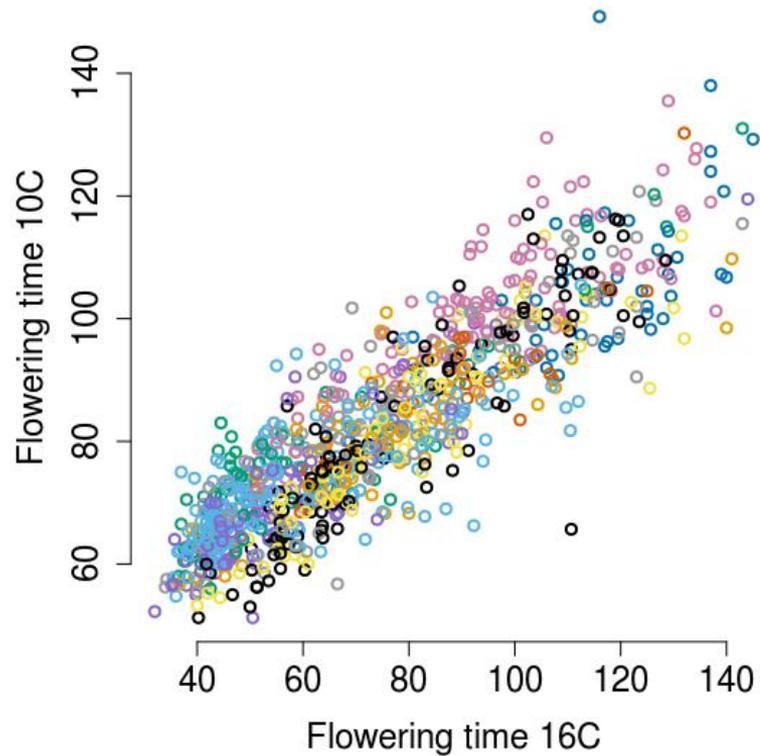
Earlier flowering at  
16°C than 10°C



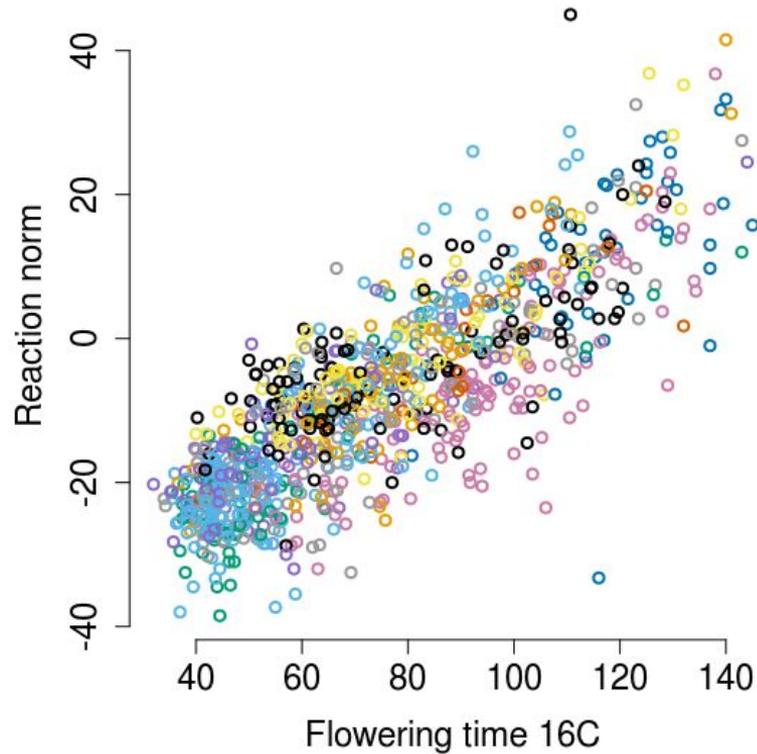
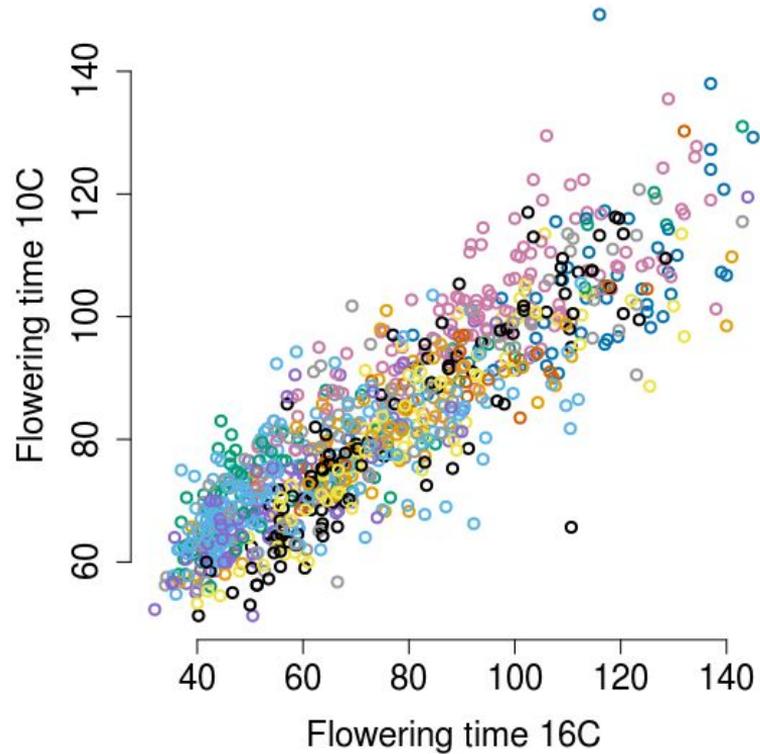
- admixed
- asia
- central\_europe
- germany
- italy\_balkan\_cauc:
- north\_sweden
- relict
- south\_sweden
- spain
- western\_europe

$p = 0.03$   
Bonferoni  $p = 0.16$

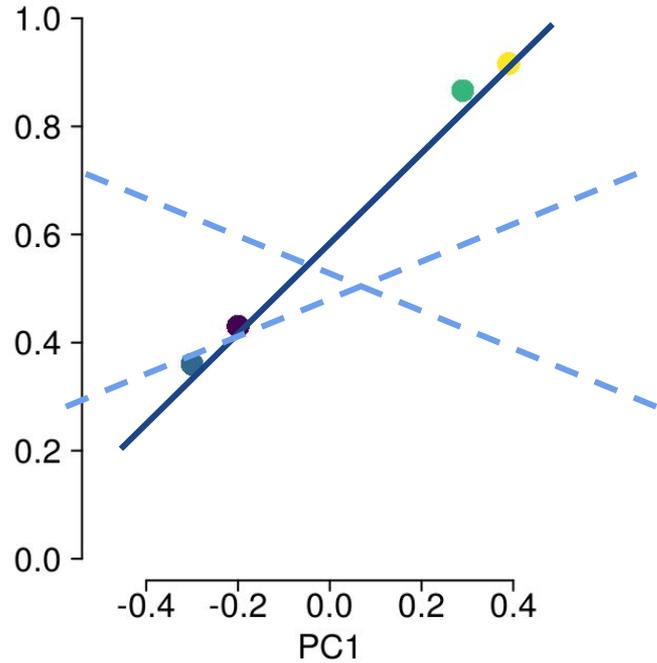
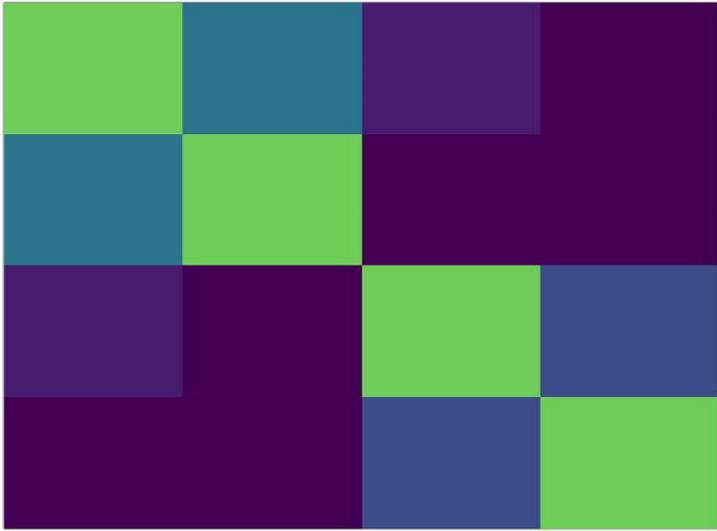
# How to deal with trait correlations?



# How to deal with trait correlations?

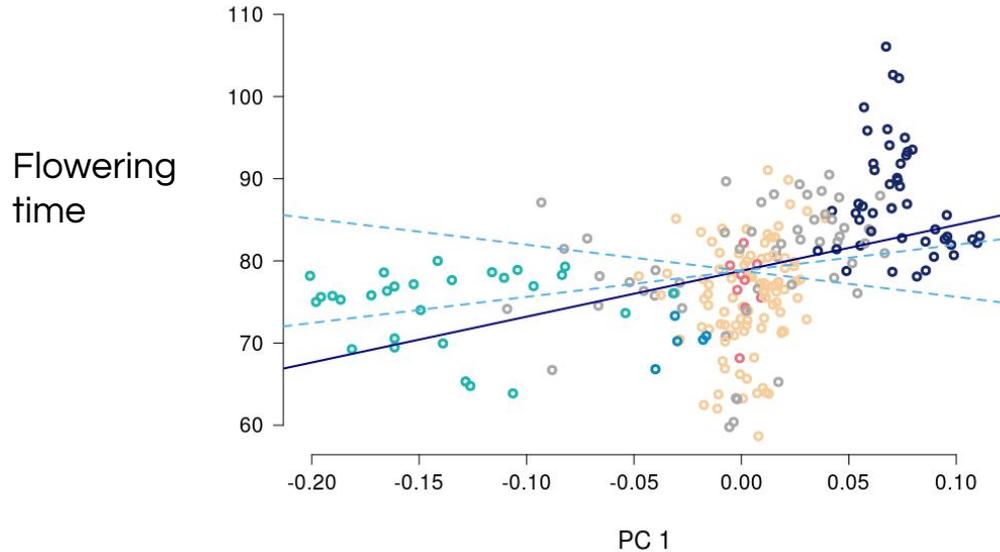


We can use PC + trait relationships to detect adaptation



We can use PC + trait relationships to detect adaptation

Local adaptation shapes trait divergence



We can use PC + trait relationships to detect adaptation

Local adaptation shapes trait divergence

These methods can be used in additional systems!!!

**Detecting Adaptive Differentiation in Structured Populations with Genomic Data and Common Gardens**

Emily B. Josephs, Jeremy J. Berg, Jeffrey Ross-Ibarra and Graham Coop  
GENETICS *Early online January 28, 2019; <https://doi.org/10.1534/genetics.118.301786>*

<https://github.com/emjosephs/quaint>

# Thanks!

Jeremy Berg  
Jennifer Blanc

Graham Coop  
Jeff Ross-Ibarra

Karl Kremling  
Ed Buckler  
Cinta Romay  
Kate Crosby



Coop Lab



Ross-Ibarra Lab

The Josephs lab at **MICHIGAN STATE**  
UNIVERSITY



Contact me! Josep993@msu.edu, <http://JosephsLab.github.io/>

We can use PC + trait relationships to detect adaptation

Local adaptation shapes trait divergence

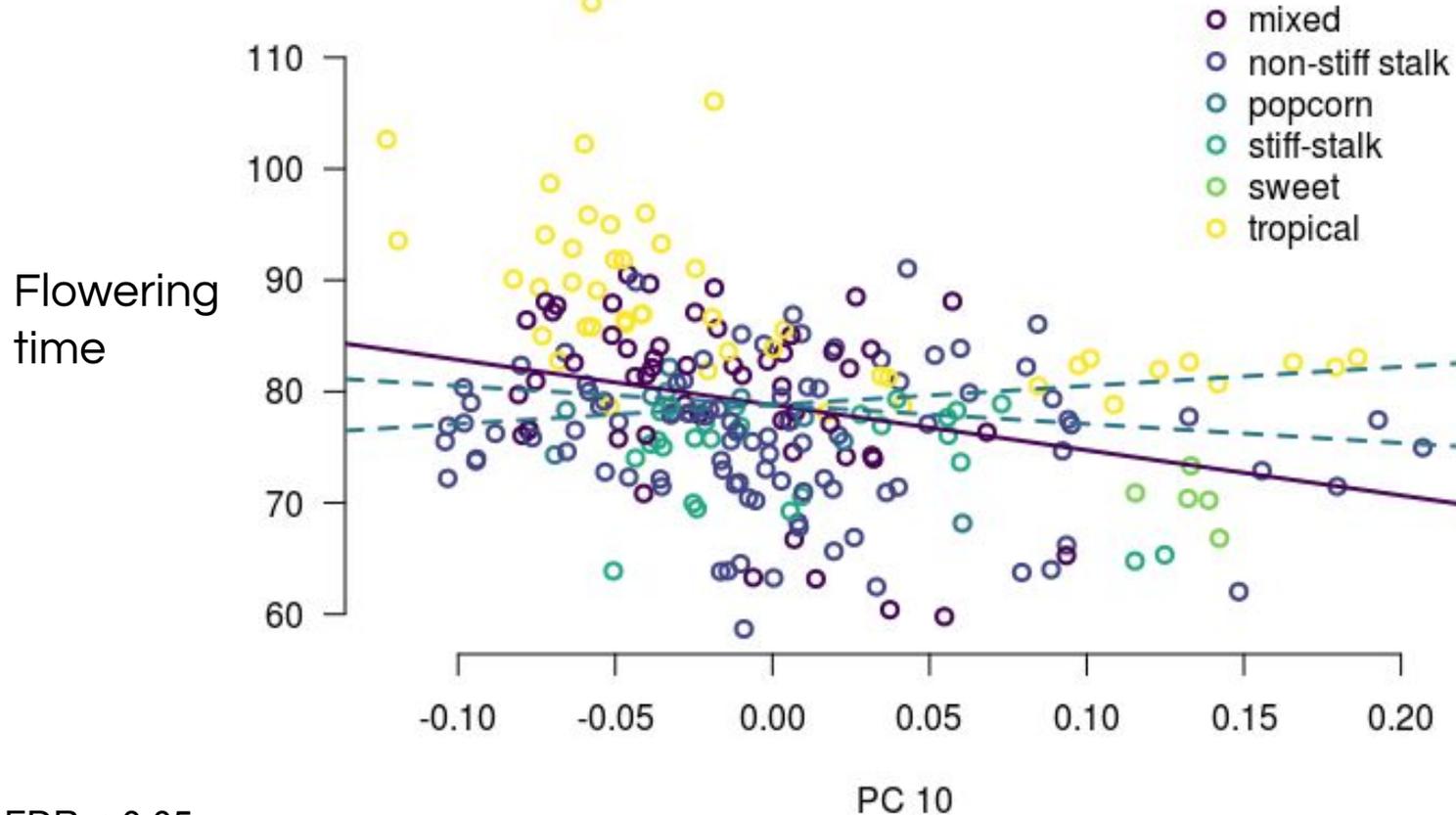
These methods can be used in additional systems!!!

**Detecting Adaptive Differentiation in Structured Populations with Genomic Data and Common Gardens**

Emily B. Josephs, Jeremy J. Berg, Jeffrey Ross-Ibarra and Graham Coop  
GENETICS *Early online January 28, 2019*; <https://doi.org/10.1534/genetics.118.301786>

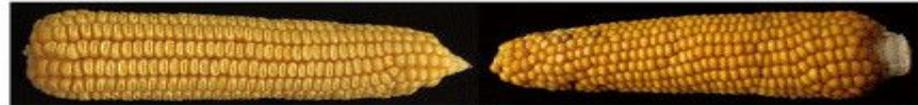
<https://github.com/emjosephs/quaint>

# Adaptive divergence along additional PCs



FDR < 0.05

What about detecting adaptation in genotypes that haven't been phenotyped?



# Testing for adaptation with polygenic scores

$$Z = 2 \sum_{l=1}^L \alpha_l p_{ml}$$

The diagram shows the equation  $Z = 2 \sum_{l=1}^L \alpha_l p_{ml}$  with four blue arrows pointing to its components:  $Z$ , the summation index  $l=1$  to  $L$ ,  $\alpha_l$ , and  $p_{ml}$ . The arrows are labeled as follows: 'Polygenic score' points to  $Z$ ; 'At all GWAS loci' points to the summation range; 'Effect size' points to  $\alpha_l$ ; and 'Allele frequency' points to  $p_{ml}$ .

Polygenic score

At all GWAS loci

Effect size

Allele frequency

# Evidence for polygenic adaptation in European maize

