Evolution 2014

Jeremy Berg and Graham Coop

University of California, Davis

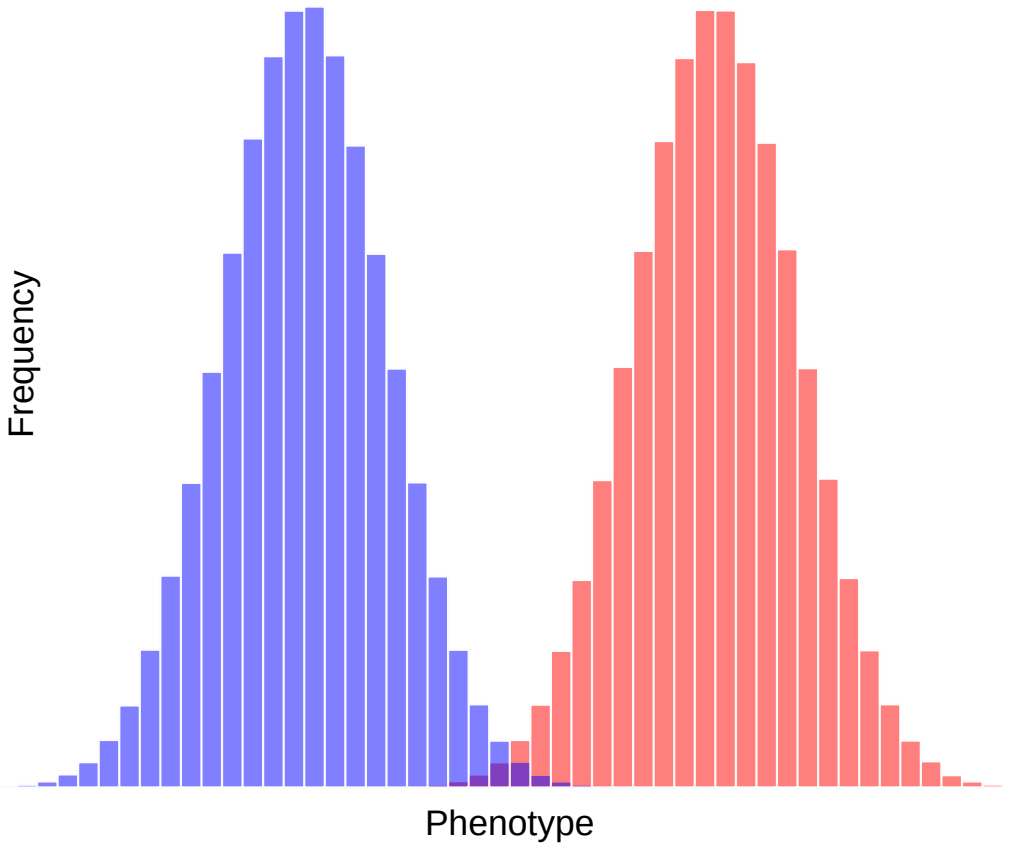
June 20, 2014

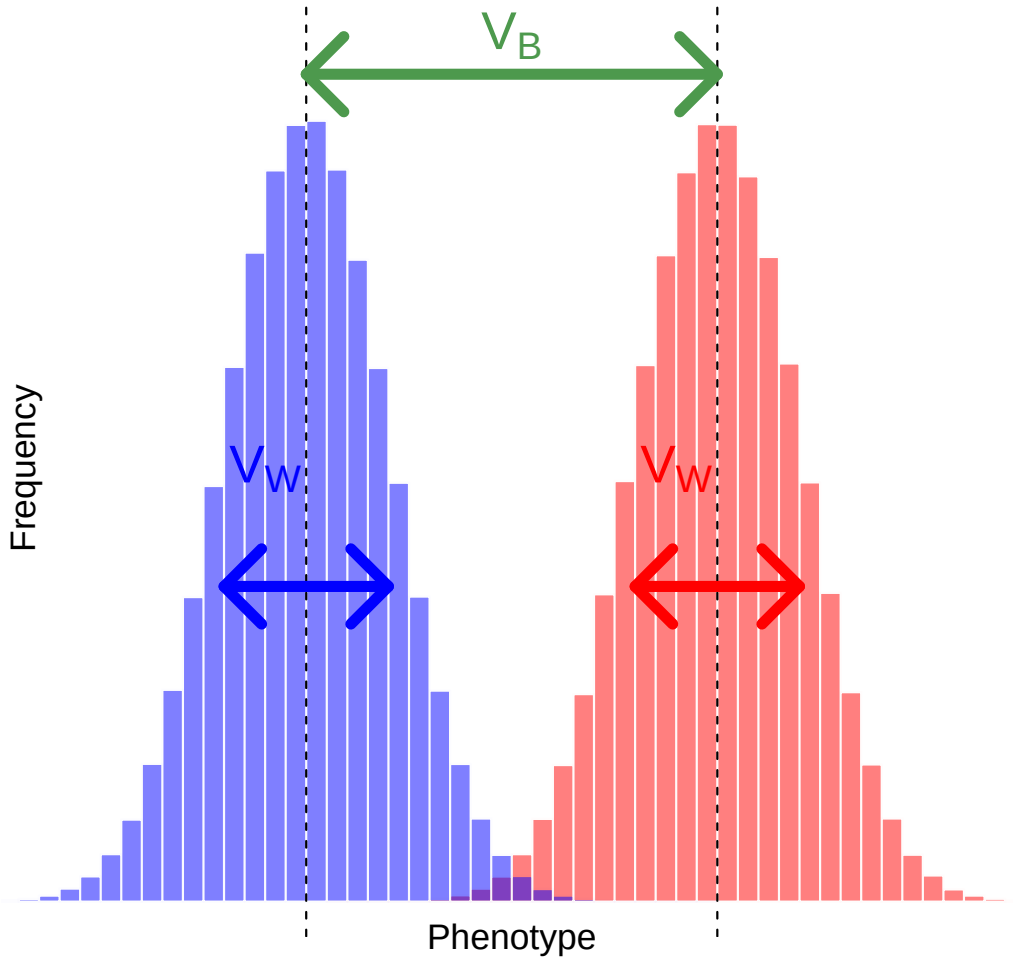
Background Slide(s)

- We want to know whether a particular trait is adaptively differentiated among populations

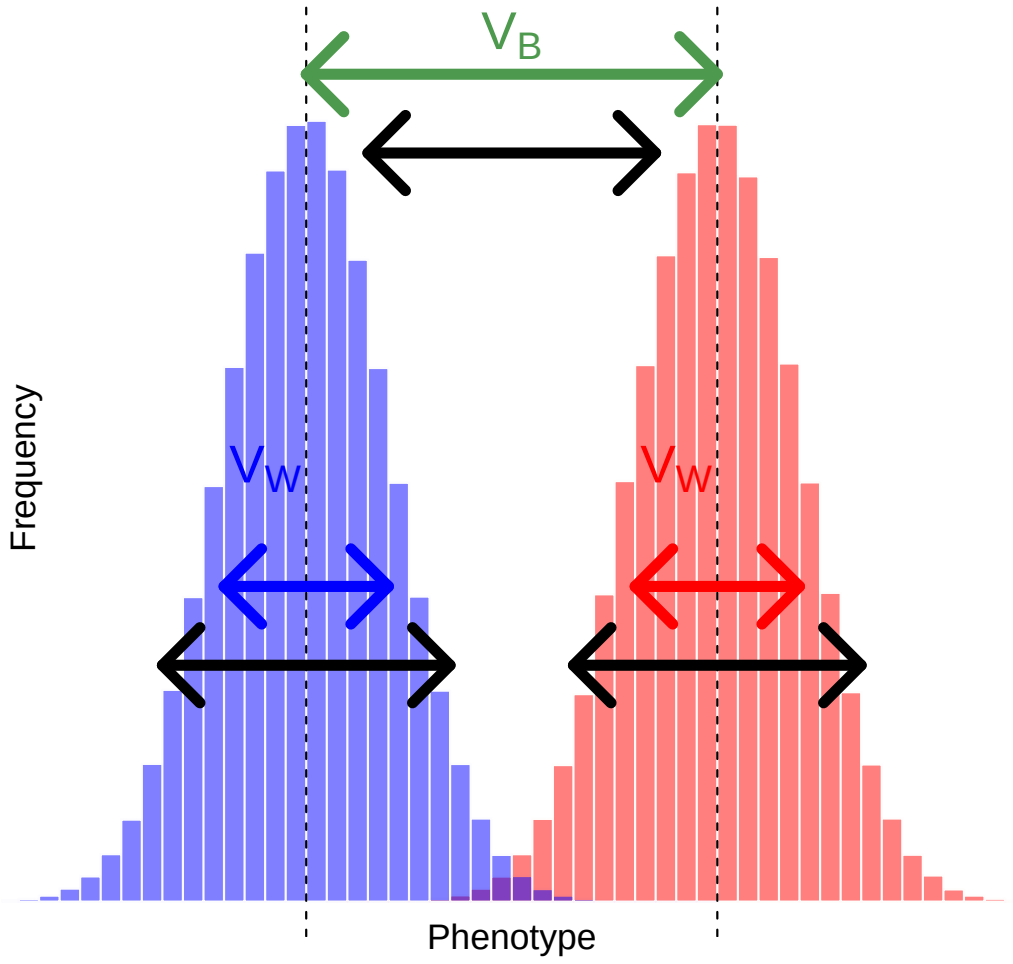
Background Slide(s)

- We want to know whether a particular trait is adaptively differentiated among populations
  - Put them in a common garden and measure them


Frequency
Phenotype

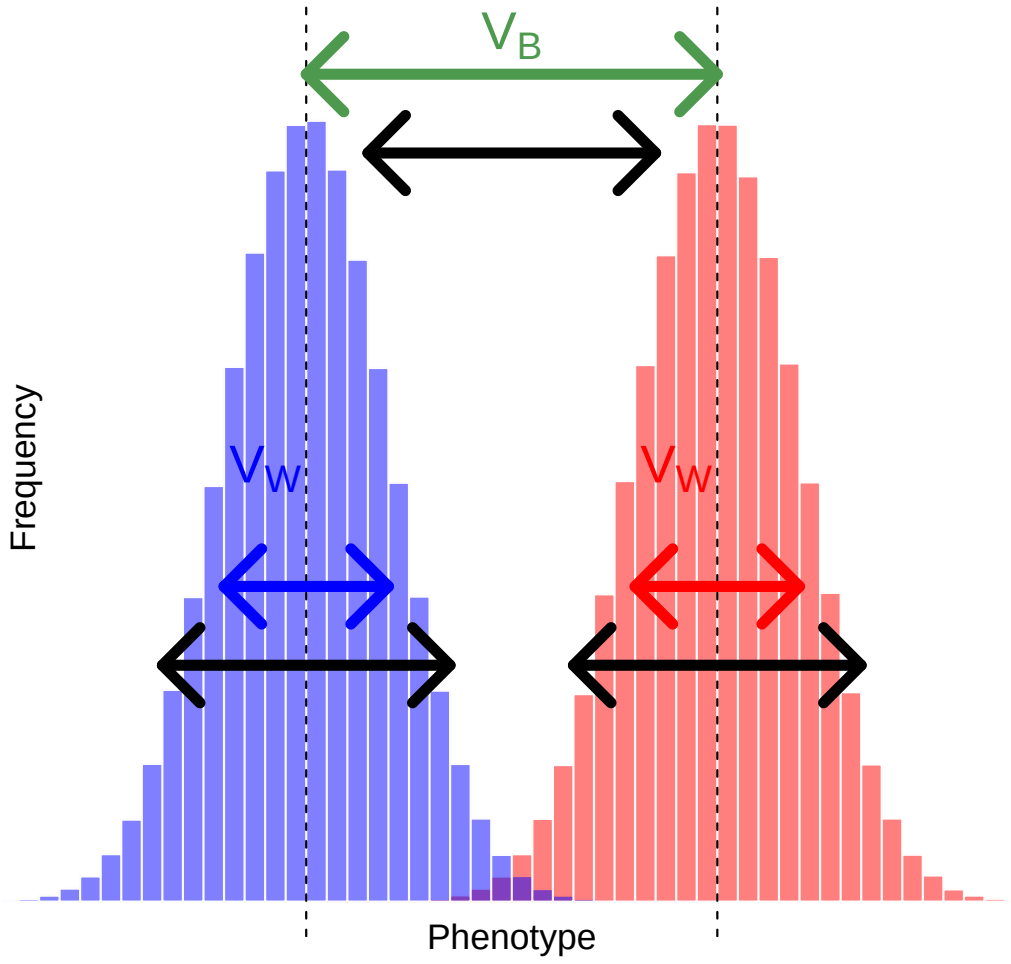


$$Q_{ST} = \frac{V_B}{V_B + V_W} = \frac{V_B}{V_T}$$



$$Q_{ST} = \frac{V_B}{V_B + V_W} = \frac{V_B}{V_T}$$

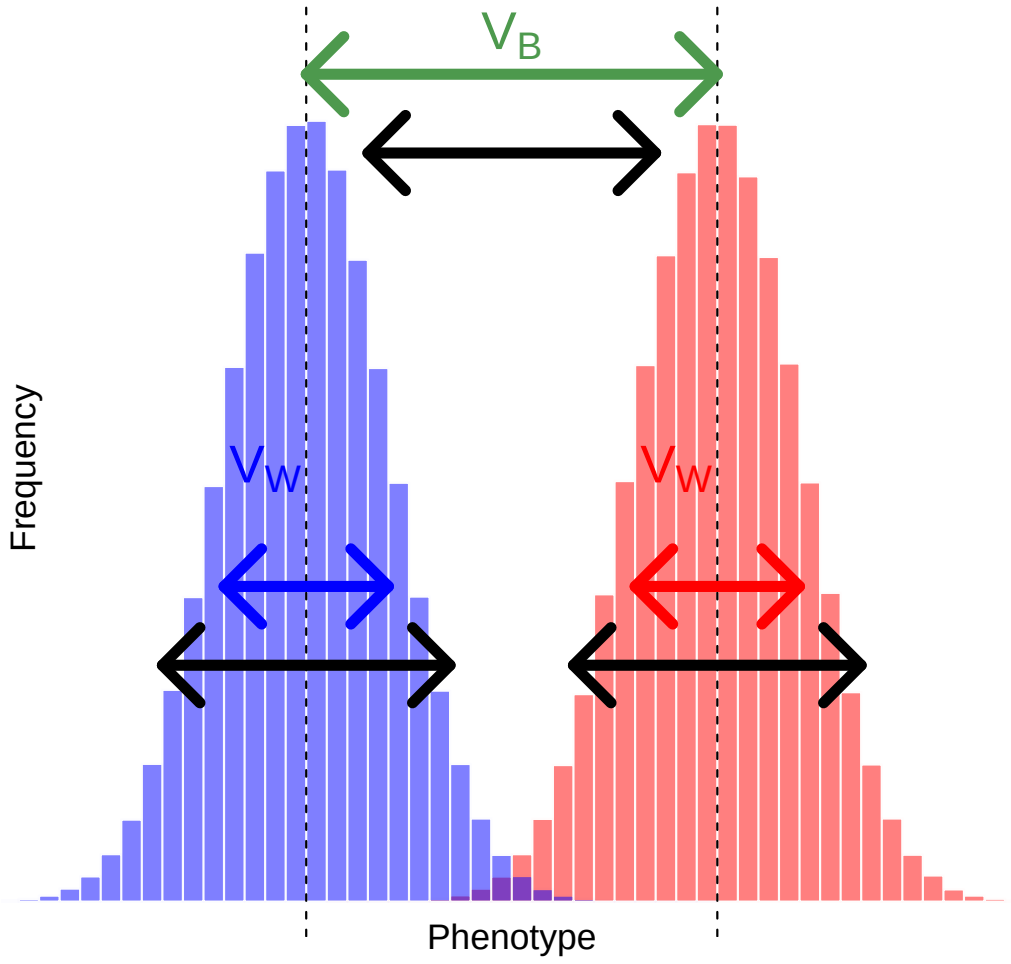
$$F_{ST} = \mathbb{E}\left[\frac{V_B}{V_B + V_W}\right]$$



$$Q_{ST} = \frac{V_B}{V_B + V_W} = \frac{V_B}{V_T}$$

$$F_{ST} = \mathbb{E}\left[\frac{V_B}{V_B + V_W}\right]$$

$$\mathbb{E}[V_B] = (V_B + V_W)\, F_{ST}$$



$$Q_{ST} = \frac{V_B}{V_B + V_W} = \frac{V_B}{V_T}$$

$$F_{ST} = \mathbb{E}\left[\frac{V_B}{V_B + V_W}\right]$$

$$\mathbb{E}[V_B] = (V_B + V_W)\, F_{ST}$$

$$\frac{Q_{ST}}{F_{ST}} = \frac{V_B}{\mathbb{E}[V_B]} \sim \chi^2$$

Generalizing Q_{ST}/F_{ST}

Q_{ST}/F_{ST} with hierarchical relatedness

$$\vec{Z} \sim MVN\left(\vec{\mu}, V_T \mathbf{F}\right)$$

$$\frac{\vec{Z}^T \mathbf{F}^{-1} \vec{Z}}{V_T} \sim \chi^2$$

- **Ovaskainen et al 2011**, Berg and Coop 2014
- $\propto$ to negative log likelihood of the data under the null model
- Reduces to $\frac{Q_{ST}}{F_{ST}}$ when all populations equally related

Using GWAS hits to estimate genetic values

LETTER

doi:10.1038/nature09410

Hundreds of variants clustered in genomic loci and biological pathways affect human height

$$Z_m = \sum_{\ell=1}^{L} \alpha_\ell p_{m\ell}$$

α = effect size

p = allele frequency

Using GWAS hits to estimate genetic values

Evidence of widespread selection on standing variation in Europe at height-associated SNPs

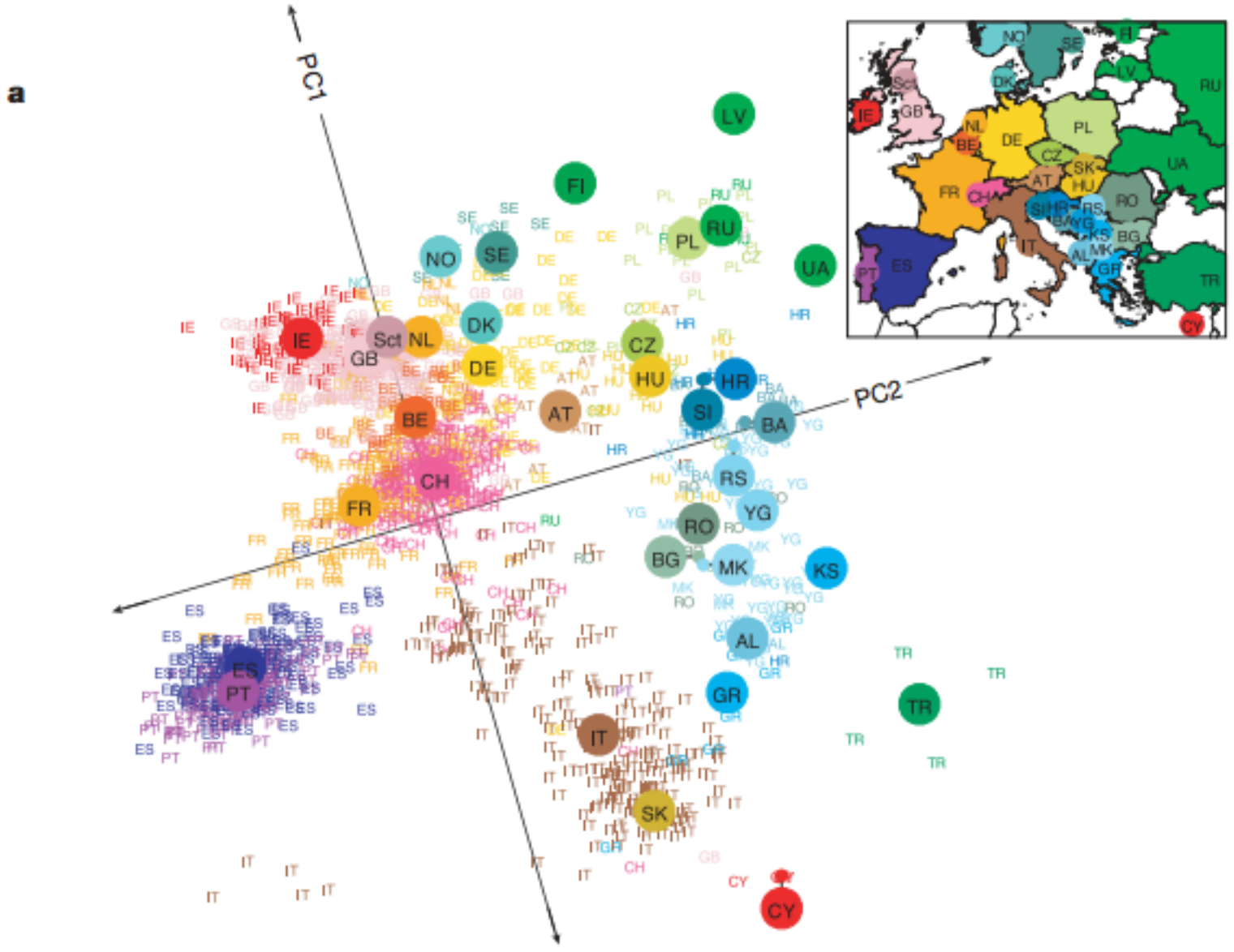
Michael C Turchin[1–5,8], Charleston WK Chiang[1–6,8], Cameron D Palmer[1–5], Sriram Sankararaman[5,6], David Reich[5,6], Genetic Investigation of ANthropometric Traits (GIANT) Consortium[7] & Joel N Hirschhorn[1–6]

$$Z_m = \sum_{\ell=1}^{L} \alpha_\ell p_{m\ell}$$

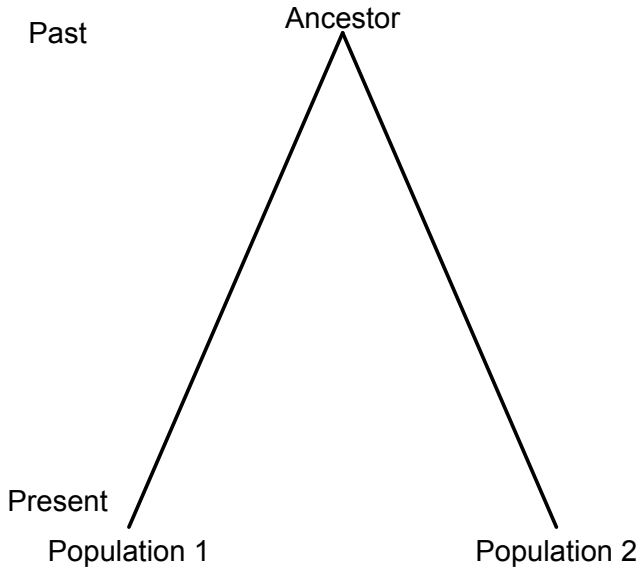
$\alpha =$ effect size

$p =$ allele frequency

What about continuously sampled populations?



Q_{ST}/F_{ST} in continuously sampled populations



Q_{ST}

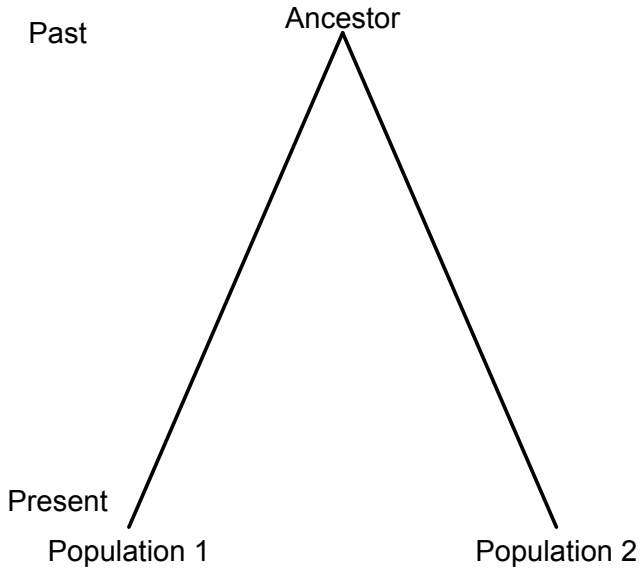
$$Q_{ST} = \frac{V_B}{V_B + V_W}$$

$$F_{ST} = \mathbb{E}\left[\frac{V_B}{V_B + V_W}\right]$$

The Principal Components View

$$\mathbf{F} = \begin{bmatrix} \mathbf{F_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{F_2} \end{bmatrix}$$

Q_{ST}/F_{ST} in continuously sampled populations



Q_{ST}

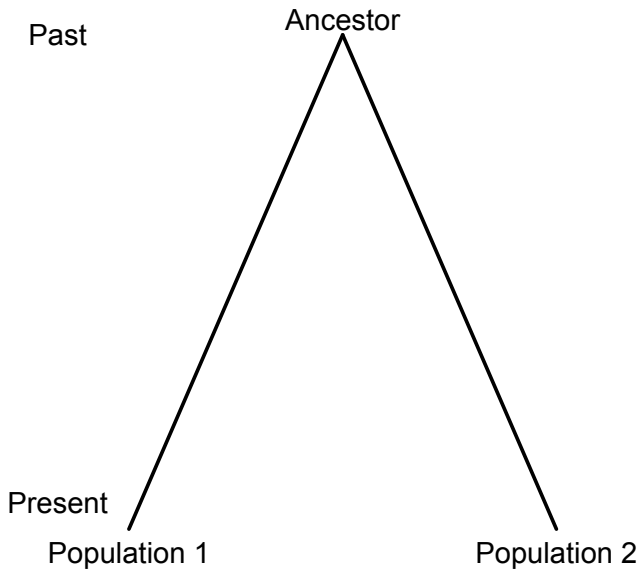
$$Q_{ST} = \frac{V_B}{V_B + V_W}$$

$$F_{ST} = \mathbb{E}\left[\frac{V_B}{V_B + V_W}\right]$$

The Principal Components View

$$\mathbf{F} = \begin{bmatrix} \mathbf{F_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{F_2} \end{bmatrix}$$

Q_{ST}/F_{ST} in continuously sampled populations



Q_{ST}

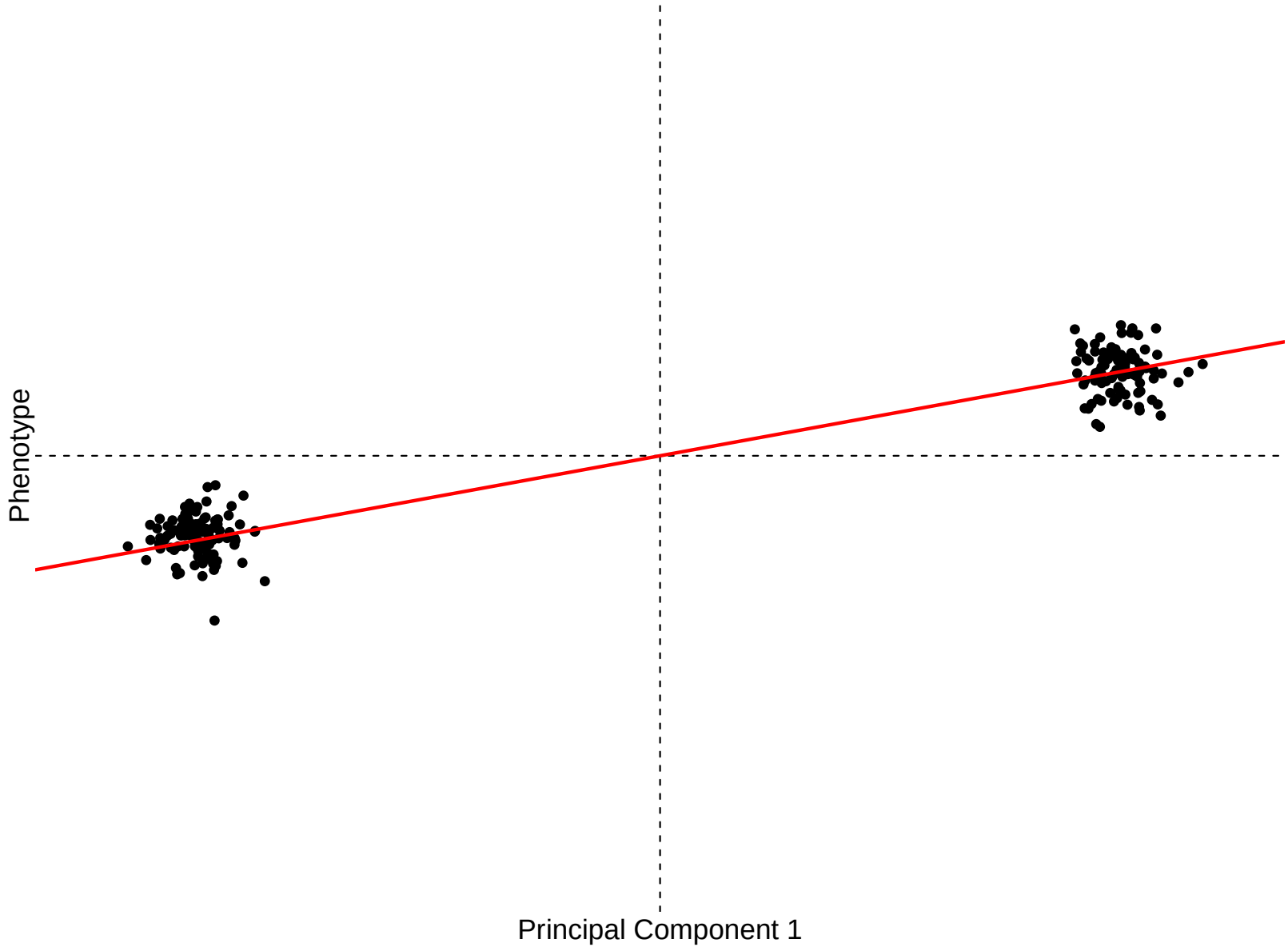
$$Q_{ST} = \frac{\left(\vec{U}_1 \cdot \vec{Z}\right)^2}{\left(\vec{U}_1 \cdot \vec{Z}\right)^2 + \sum_{j=2}^{K} \left(\vec{U}_j \cdot \vec{Z}\right)^2}$$

$$F_{ST} = \frac{\lambda_1}{\lambda_1 + \sum_{j=2}^{K} \lambda_j}$$

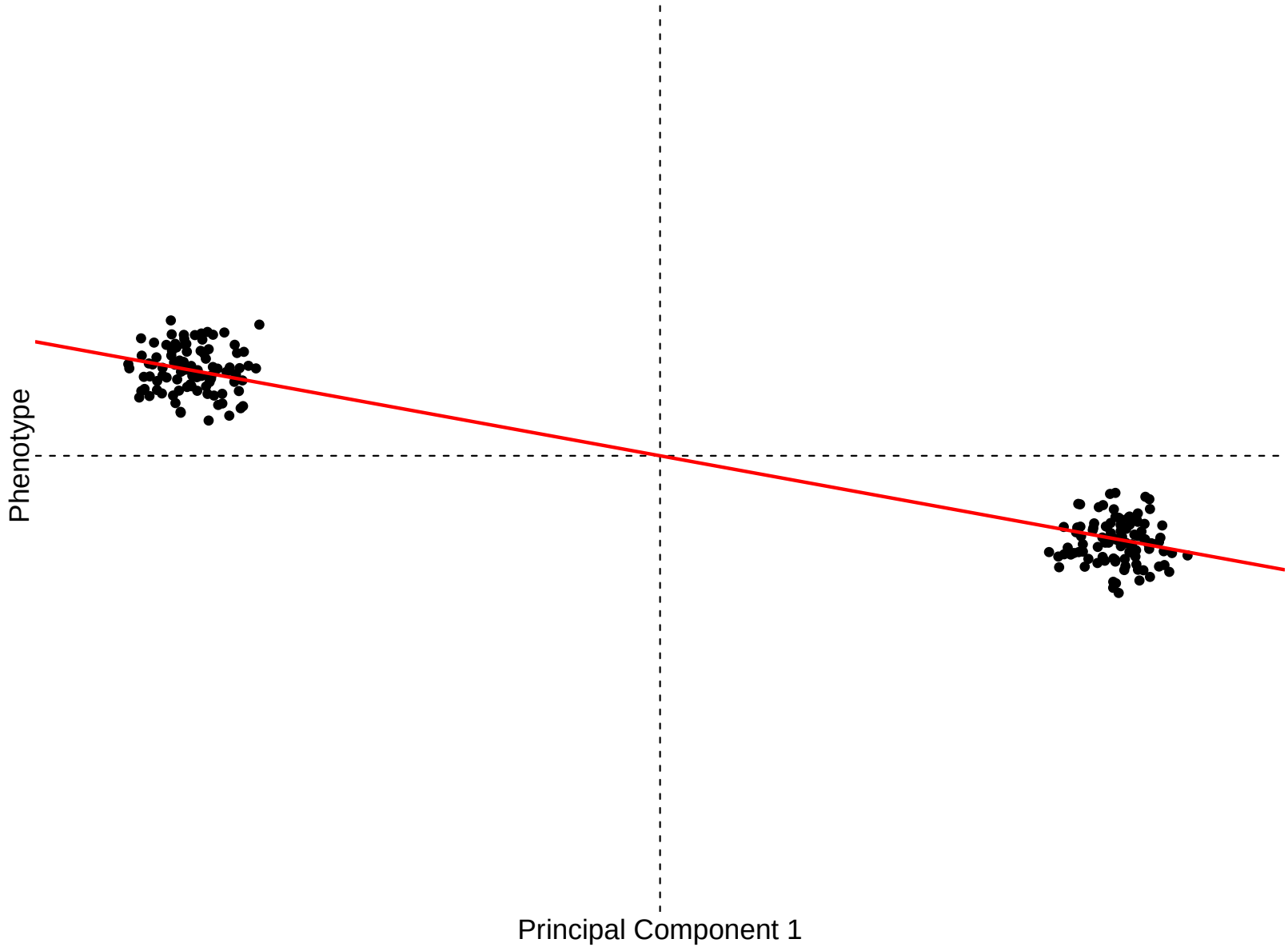
The Principal Components View

$$\mathbf{F} = \begin{bmatrix} \vec{U}_1 & \vec{U}_2 & \dots & \vec{U}_K \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_K \end{bmatrix} \begin{bmatrix} \vec{U}_1^T \\ \vec{U}_2^T \\ \dots \\ \vec{U}_K^T \end{bmatrix}$$

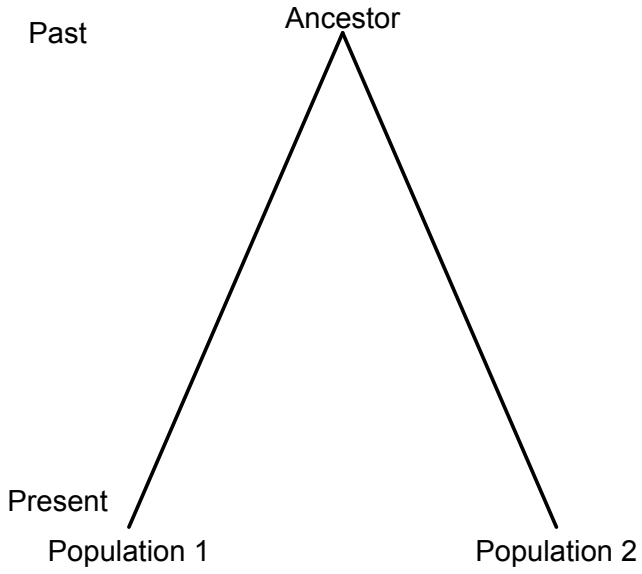
Positive result if population predicts phenotype better than expected



Positive result if population predicts phenotype better than expected



Q_{ST}/F_{ST} in continuously sampled populations



Q_{ST}

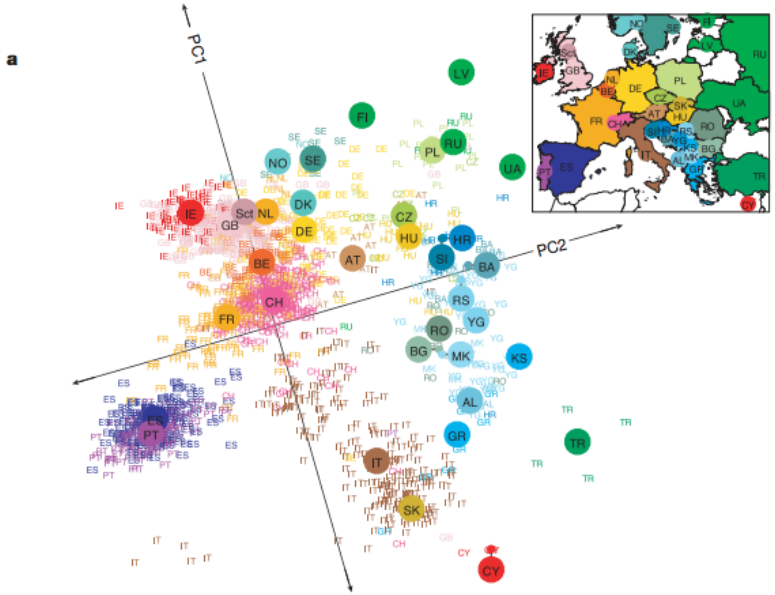
$$Q_{ST} = \frac{\left(\vec{U}_1 \cdot \vec{Z}\right)^2}{\left(\vec{U}_i \cdot \vec{Z}\right)^2 + \sum_{j=2}^{K-1} \left(\vec{U}_j \cdot \vec{Z}\right)^2}$$

$$F_{ST} = \frac{\lambda_1}{\lambda_1 + \sum_{j=2}^{K-1} \lambda_j}$$

The Principal Components View

$$\mathbf{F} = \begin{bmatrix} \vec{U}_1 & \vec{U}_2 & \dots & \vec{U}_{K-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{K-1} \end{bmatrix} \begin{bmatrix} \vec{U}_1^T \\ \vec{U}_2^T \\ \dots \\ \vec{U}_{K-1}^T \end{bmatrix}$$

Q_{ST}/F_{ST} in continuously sampled populations



Q_{ST}

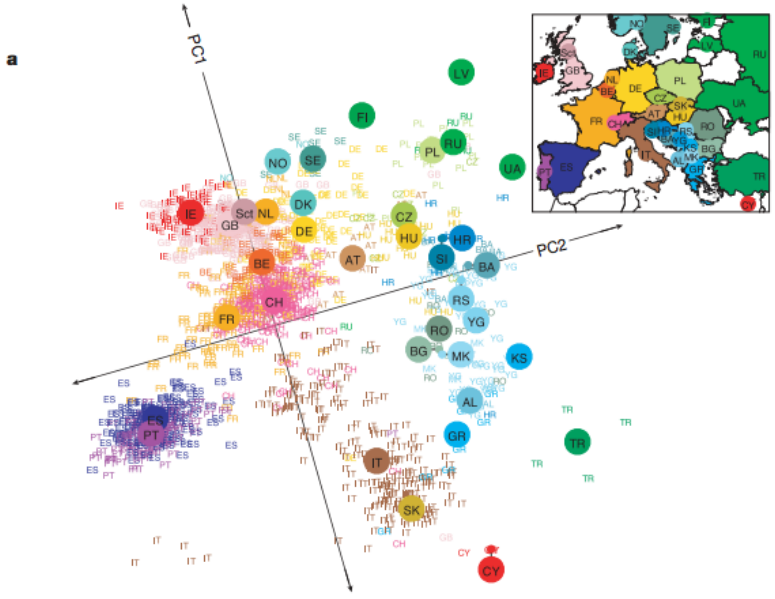
$$Q_{PC1:2} = \frac{V_{PC1:2}}{V_{PC1:2}+V_{PC3:K-1}}$$

$$F_{PC1:2} = \mathbb{E}\left[\frac{V_{PC1:2}}{V_{PC1:2}+V_{PC3:K-1}}\right]$$

The Principal Components View

$$\mathbf{F} = \begin{bmatrix}\vec{U}_1 & \vec{U}_2 & \dots & \vec{U}_{K-1}\end{bmatrix} \begin{bmatrix}\lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{K-1}\end{bmatrix} \begin{bmatrix}\vec{U}_1^T \\ \vec{U}_2^T \\ \dots \\ \vec{U}_{K-1}^T\end{bmatrix}$$

Q_{ST}/F_{ST} in continuously sampled populations



Q_{ST}

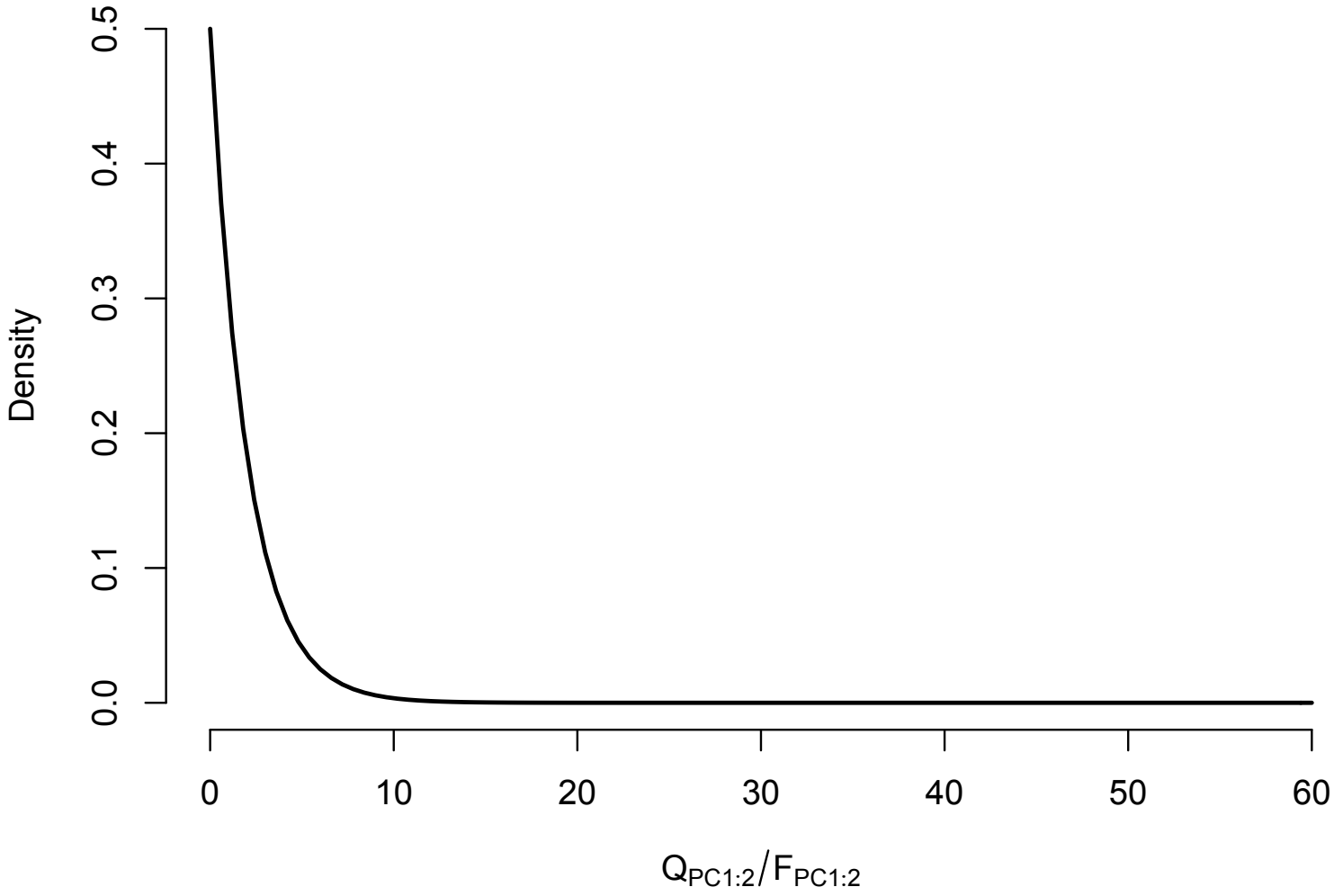
$$Q_{PC1} = \frac{\sum_{i=1}^{2} \left(\vec{U}_i \cdot \vec{Z}\right)^2}{\sum_{i=1}^{2} \left(\vec{U}_i \cdot \vec{Z}\right)^2 + \sum_{j=3}^{K-1} \left(\vec{U}_j \cdot \vec{Z}\right)^2}$$

$$F_{PC1} = \frac{\sum_{i=1}^{2} \lambda_i}{\sum_{i=1}^{2} \lambda_i + \sum_{j=3}^{K-1} \lambda_j}$$

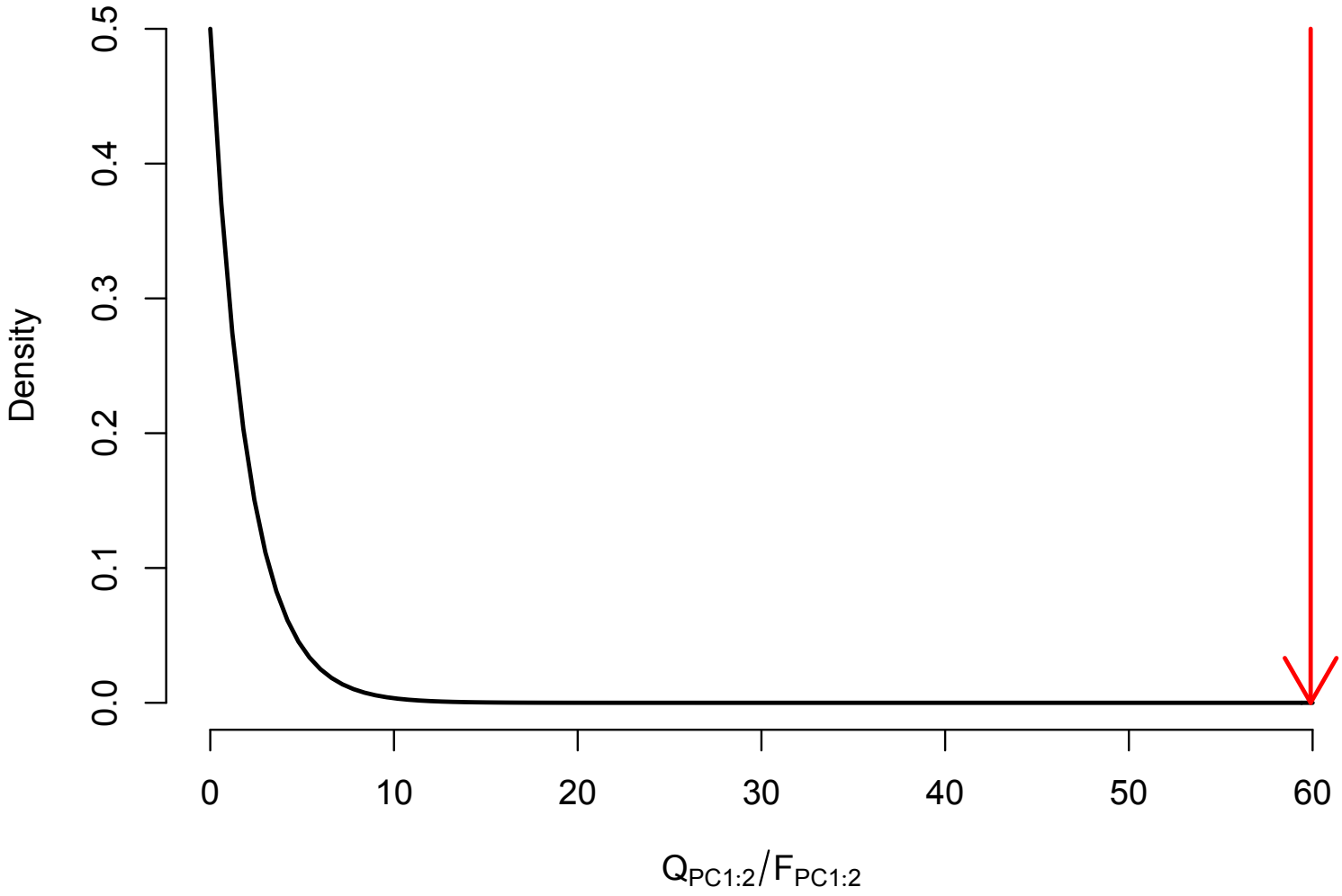
The Principal Components View

$$\mathbf{F} = \begin{bmatrix} \vec{U}_1 & \vec{U}_2 & \dots & \vec{U}_{K-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{K-1} \end{bmatrix} \begin{bmatrix} \vec{U}_1^T \\ \vec{U}_2^T \\ \dots \\ \vec{U}_{K-1}^T \end{bmatrix}$$

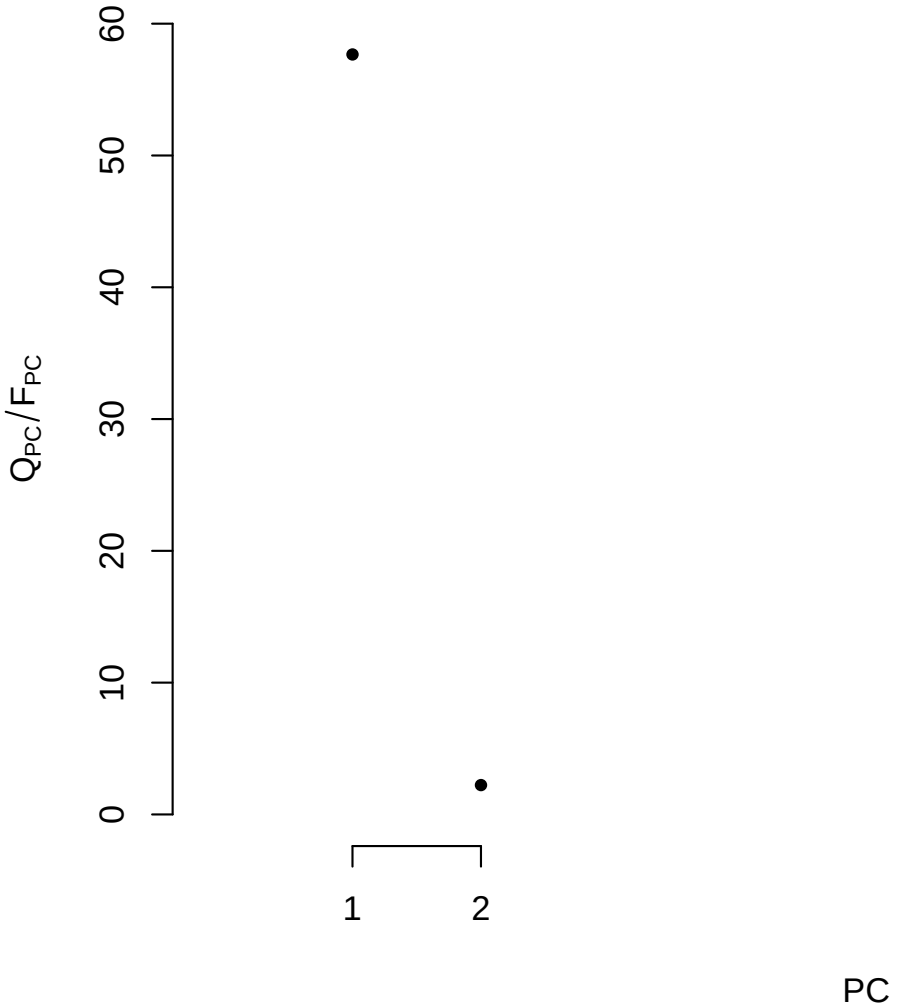
Genetic Divergence for Height in Europe



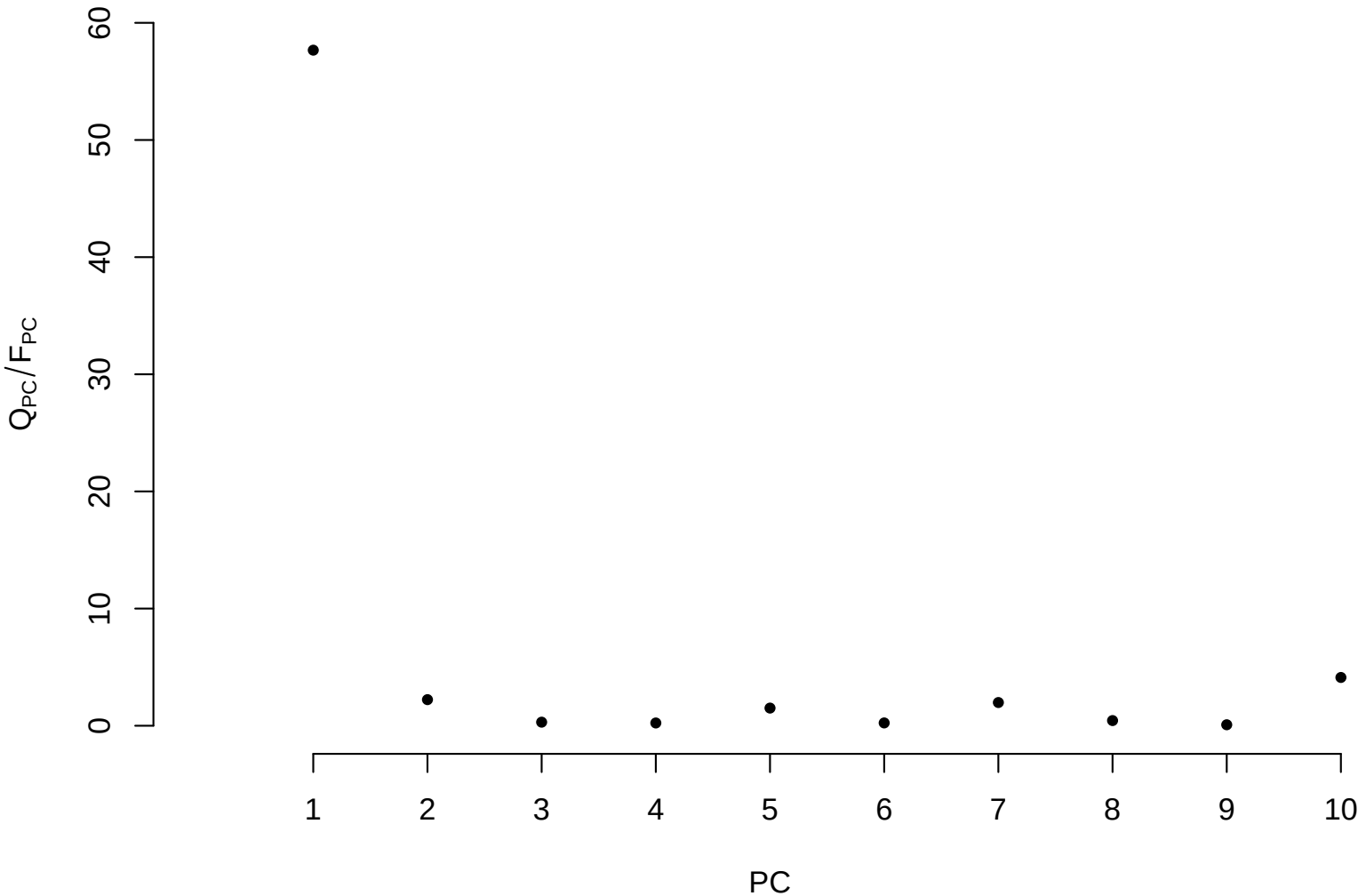
Genetic Divergence for Height in Europe



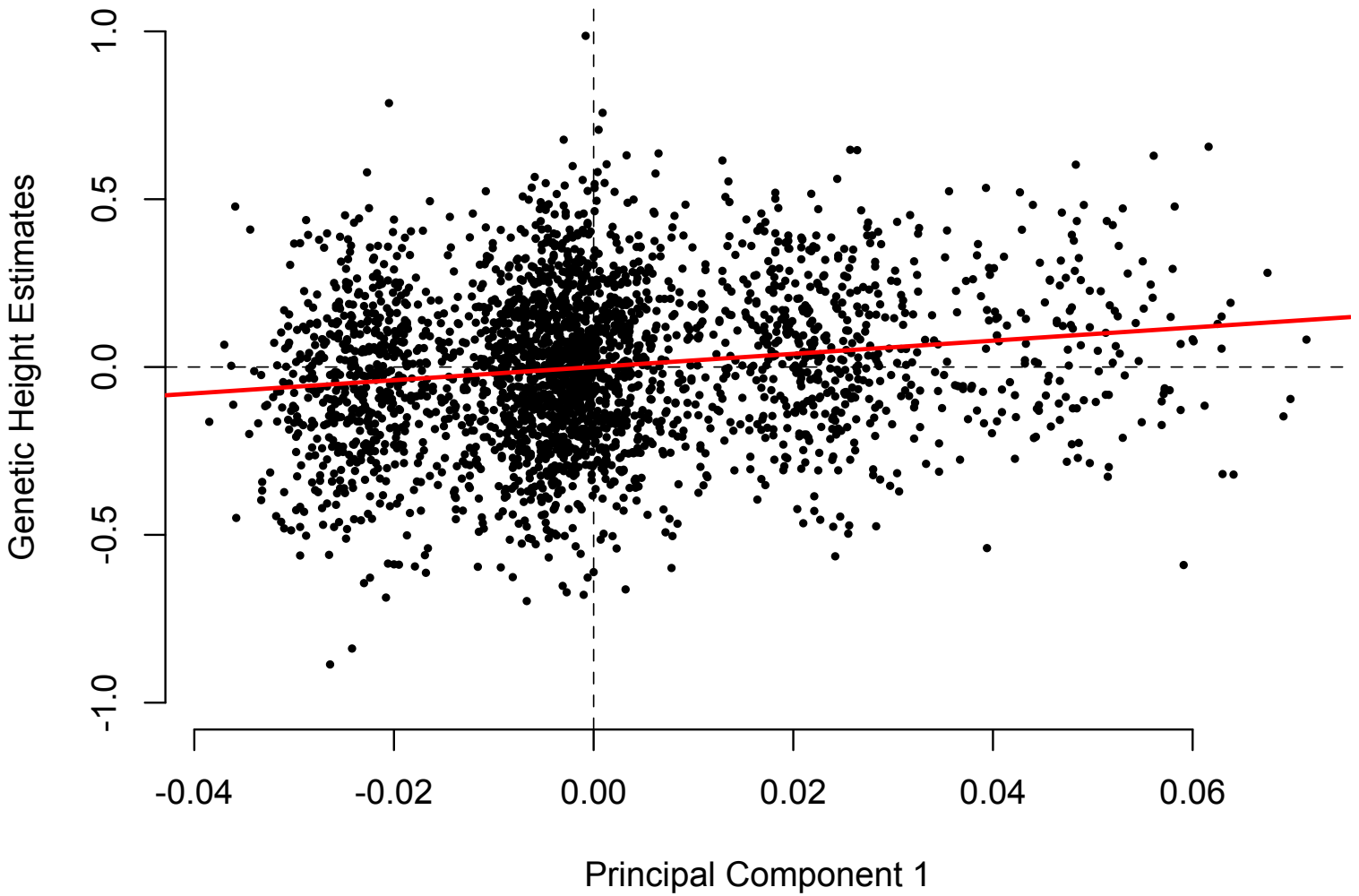
Genetic Divergence in Height Along PC1 in Europe



Genetic Divergence in Height Along PC1 in Europe



Genetic Divergence in Height Along PC1 in Europe



Take Aways

- Height associated SNPs significantly correlated with PC1/N-S axis in Europe (but we already knew that)

Take Aways

- Height associated SNPs significantly correlated with PC1/N-S axis in Europe (but we already knew that)
- Q_{ST}/F_{ST} can be formulated in terms of reduced rank factorizations of the individual-by-individual kinship matrix
  - Direct relationship to PCA, *structure*, Sparse Factor Analysis
  - Engelhardt and Stephens 2010

Things I Didn't Mention

- Can add include multiple correlated traits, but interpretation potentially trickier
  - Chenoweth and Blows (2008)
  - Martin et al (2008)
- In species where it's possible to set up breeding designs, continuous sampling is not a barrier

Acknowledgements