Testing for Adaptive Divergence: Q_{ST}/F_{ST} Comparisons in Populations with Complicated Histories

Jeremy Berg and Graham Coop

University of California, Davis

June 22, 2014

Testing for adaptive differentiation

- Having observed phenotypic differences between populations, we'd like to know if they are adaptive

Testing for adaptive differentiation

- Having observed phenotypic differences between populations, we'd like to know if they are adaptive
- Raise individuals from different populations in a common environment, measure their traits, and genotype them at a (preferably large) number of genetic markers

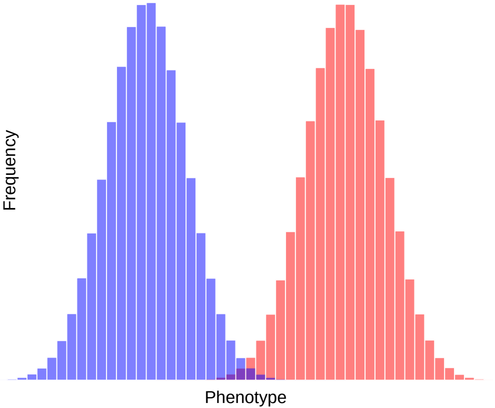
Testing for adaptive differentiation

- Having observed phenotypic differences between populations, we'd like to know if they are adaptive
- Raise individuals from different populations in a common environment, measure their traits, and genotype them at a (preferably large) number of genetic markers
- Compute Q_{ST} and F_{ST}; compare them

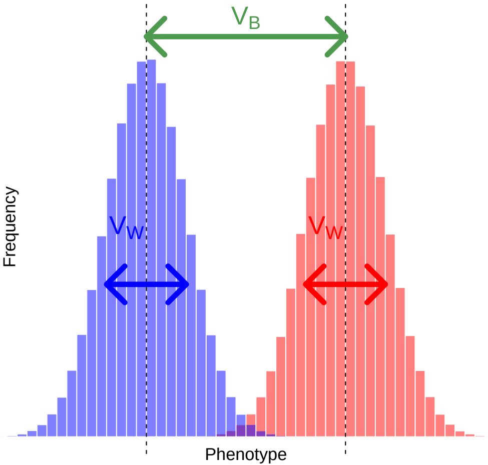
Testing for adaptive differentiation

- Having observed phenotypic differences between populations, we'd like to know if they are adaptive
- Raise individuals from different populations in a common environment, measure their traits, and genotype them at a (preferably large) number of genetic markers
- Compute Q_{ST} and F_{ST}; compare them
- **What if the individuals in our sample can't be fit neatly into a discrete set of populations?**

The Standard Q_{ST}/F_{ST} Comparison

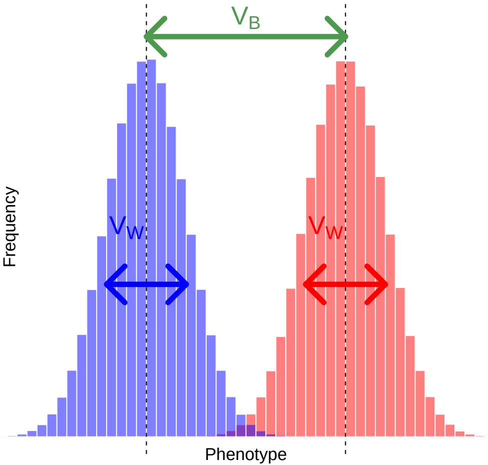


The Standard Q_{ST}/F_{ST} Comparison



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

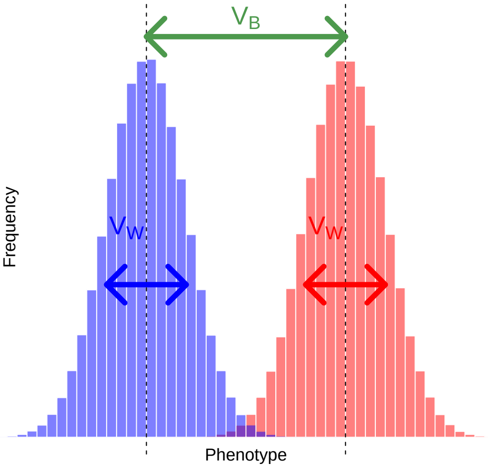
The Standard Q_{ST}/F_{ST} Comparison



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

$$\mathbb{E}\left[Q_{ST}\right] = F_{ST}$$

The Standard Q_{ST}/F_{ST} Comparison

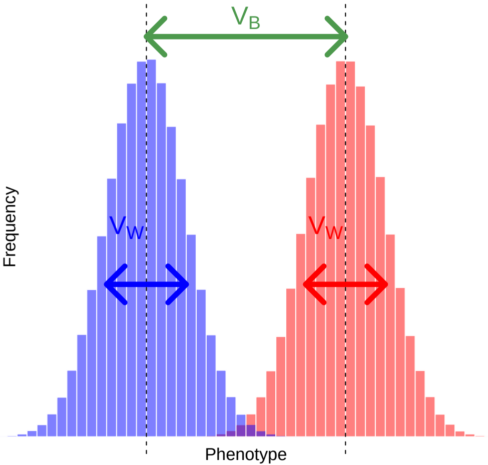


$$Q_{ST} = \frac{V_B}{V_B + V_W} = \frac{V_B}{V_T}$$

$$\mathbb{E}\left[Q_{ST}\right] = F_{ST}$$

$$\mathbb{E}[V_B] = V_T F_{ST}$$

The Standard Q_{ST}/F_{ST} Comparison



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

$$\mathbb{E}\left[Q_{ST}\right] = F_{ST}$$

$$\mathbb{E}[V_B] = V_T F_{ST}$$

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

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

$$Z_k = \sum_{\ell=1}^{L} \alpha_\ell p_{k\ell}$$

$\alpha =$ effect size

$p =$ allele count

Using GWAS hits to estimate genetic values

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

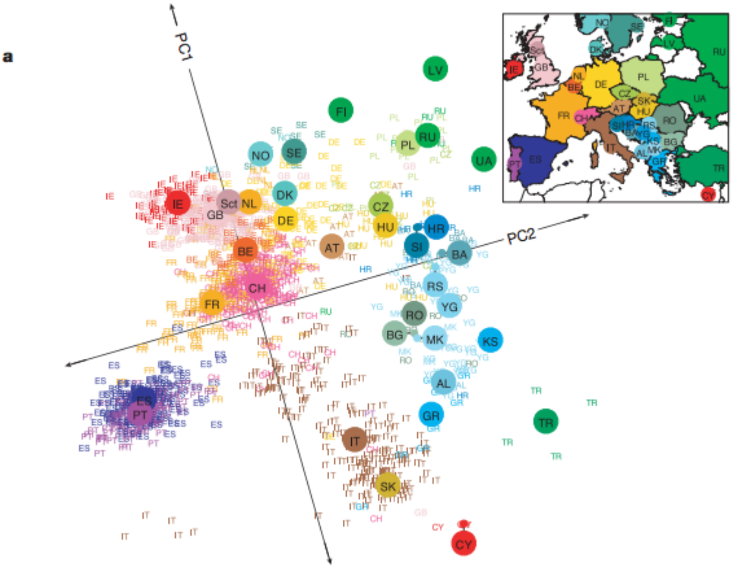
Michael C Turchin[1–5,8], Charleston WK Chiang[1–6,8], Cameron D Palmer[1–5], Sriram Sankararaman[5,6], David Reich[5,6], Genetic Investigation of ANthropometric Traits (GIANT) Consortium[7] & Joel N Hirschhorn[1–6]

$$Z_k = \sum_{\ell=1}^{L} \alpha_\ell p_{k\ell}$$

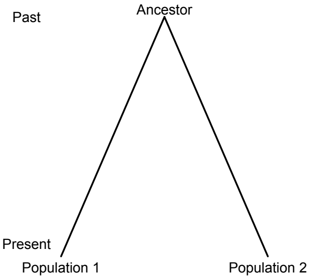
$\alpha =$ effect size

$p =$ allele count

What about continuously sampled populations?



Q_{ST}/F_{ST} on a natural scale



Variance Partitions

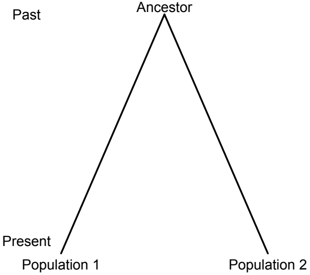
$$Q_{ST} = \frac{V_B}{V_B + V_W}$$

$$F_{ST} = \mathbb{E}\left[Q_{ST}\right]$$

The Kinship Matrix

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

Q_{ST}/F_{ST} on a natural scale



Variance Partitions

$$V_B = \left(\vec{U}_1 \cdot \vec{Z}\right)^2$$

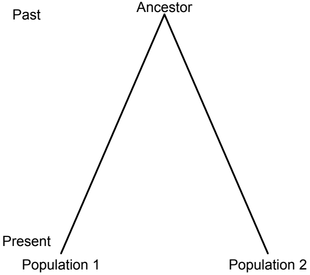
$$V_W = \sum_{i=2}^{K-1} \left(\vec{U}_i \cdot \vec{Z}\right)^2$$

$$F_{ST} = \frac{\lambda_1}{\sum_{i=1}^{K-1} \lambda_i}$$

The Eigenvalue Decomposition (Principal Components)

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

Q_{ST}/F_{ST} on a natural scale



Variance Partitions

$$V_B = \left(\vec{U}_1 \cdot \vec{Z}\right)^2$$

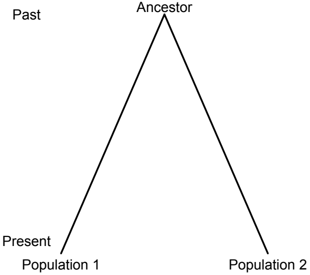
$$V_W = \sum_{i=2}^{K-1} \left(\vec{U}_i \cdot \vec{Z}\right)^2$$

$$F_{ST} = \frac{\lambda_1}{\sum_{i=1}^{K-1} \lambda_i}$$

The Eigenvalue Decomposition (Principal Components)

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

Q_{ST}/F_{ST} on a natural scale



Variance Partitions

$$V_B = \left(\vec{U}_1 \cdot \vec{Z}\right)^2$$

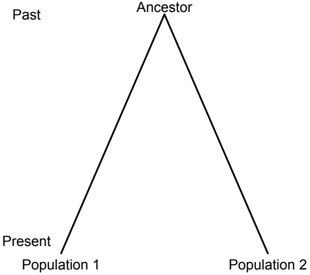
$$V_W = \sum_{i=2}^{K-1} \left(\vec{U}_i \cdot \vec{Z}\right)^2$$

$$F_{ST} = \frac{\lambda_1}{\sum_{i=1}^{K-1} \lambda_i}$$

The Eigenvalue Decomposition (Principal Components)

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

Q_{ST}/F_{ST} on a natural scale



Variance Partitions

$$V_B = \left(\vec{U}_1 \cdot \vec{Z}\right)^2$$

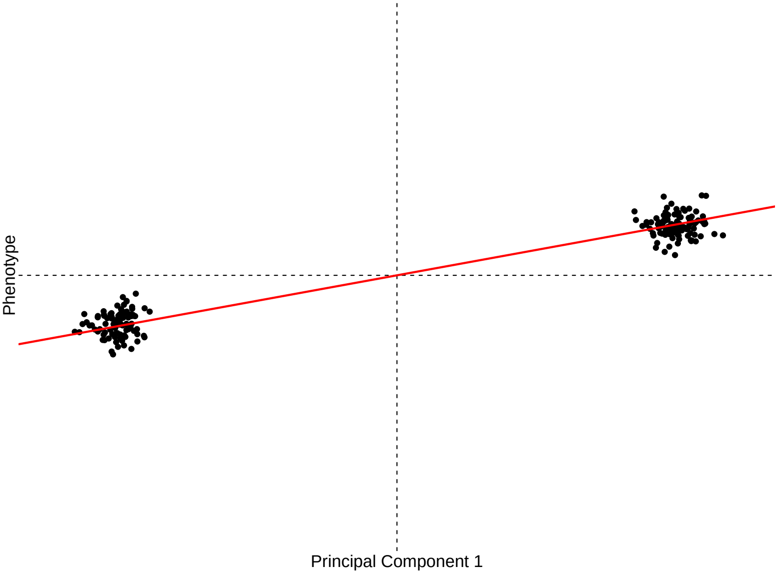
$$V_W = \sum_{i=2}^{K-1} \left(\vec{U}_i \cdot \vec{Z}\right)^2$$

$$F_{ST} = \frac{\lambda_1}{\sum_{i=1}^{K-1} \lambda_i}$$

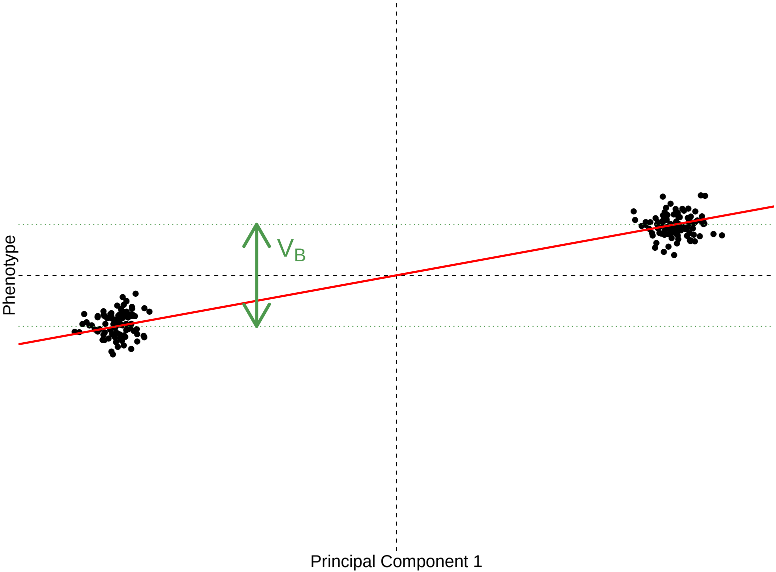
The Eigenvalue Decomposition (Principal Components)

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

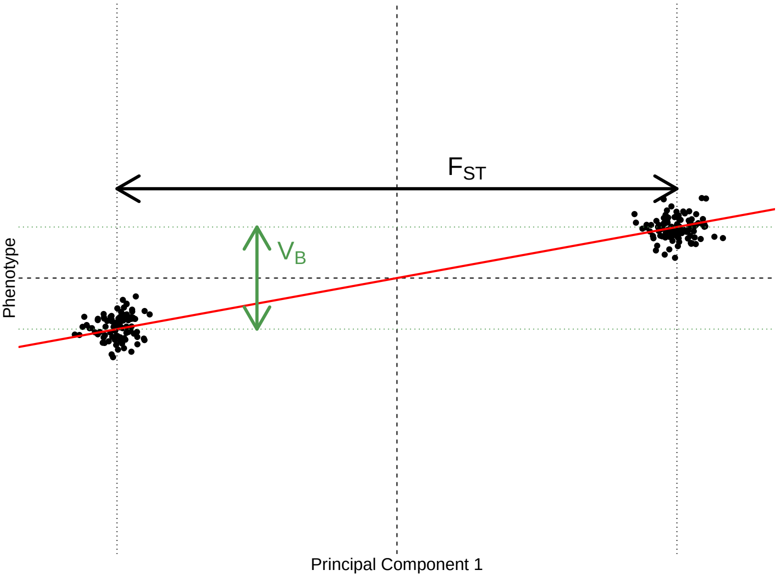
Positive result if PC predicts phenotype better than expected



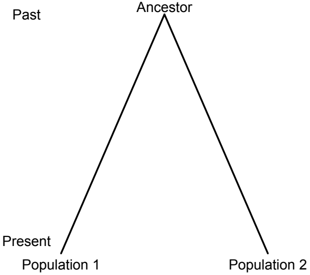
Positive result if PC predicts phenotype better than expected



Positive result if PC predicts phenotype better than expected



Q_{ST}/F_{ST} on a natural scale



Variance Partitions

$$V_B = \left(\vec{U}_1 \cdot \vec{Z}\right)^2$$

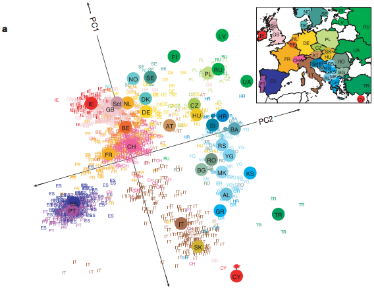
$$V_W = \sum_{i=2}^{K-1} \left(\vec{U}_i \cdot \vec{Z}\right)^2$$

$$F_{ST} = \frac{\lambda_1}{\sum_{i=1}^{K-1} \lambda_i}$$

The Eigenvalue Decomposition (Principal Components)

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

Q_{ST}/F_{ST} on a natural scale



Variance Partitions

$$V_B = \left(\vec{U}_1 \cdot \vec{Z}\right)^2 + \left(\vec{U}_2 \cdot \vec{Z}\right)^2$$

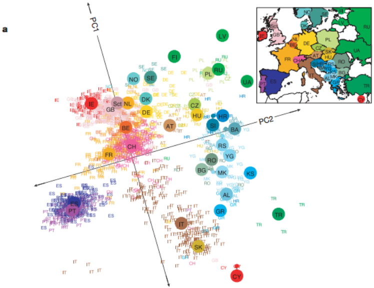
$$V_W = \sum_{i=3}^{K-1} \left(\vec{U}_i \cdot \vec{Z}\right)^2$$

$$F_{PC1:2} = \frac{\lambda_1 + \lambda_2}{\sum_{i=1}^{K-1} \lambda_i}$$

The Eigenvalue Decomposition (Principal Components)

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

Q_{ST}/F_{ST} on a natural scale



Variance Partitions

$$V_B = \left(\vec{U}_1 \cdot \vec{Z}\right)^2 + \left(\vec{U}_2 \cdot \vec{Z}\right)^2$$

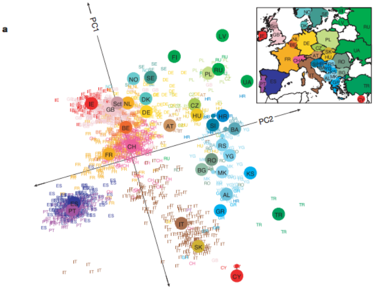
$$V_W = \sum_{i=3}^{K-1} \left(\vec{U}_i \cdot \vec{Z}\right)^2$$

$$F_{PC1:2} = \frac{\lambda_1 + \lambda_2}{\sum_{i=1}^{K-1} \lambda_i}$$

The Eigenvalue Decomposition (Principal Components)

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

Q_{ST}/F_{ST} on a natural scale



Variance Partitions

$$V_B = \left(\vec{U}_1 \cdot \vec{Z}\right)^2 + \left(\vec{U}_2 \cdot \vec{Z}\right)^2$$

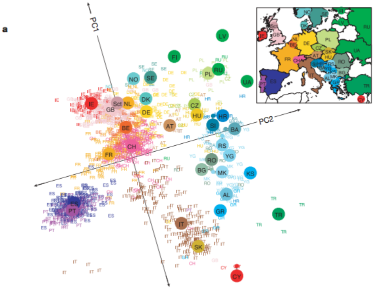
$$V_W = \sum_{i=3}^{K-1} \left(\vec{U}_i \cdot \vec{Z}\right)^2$$

$$F_{PC1:2} = \frac{\lambda_1 + \lambda_2}{\sum_{i=1}^{K-1} \lambda_i}$$

The Eigenvalue Decomposition (Principal Components)

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

Q_{ST}/F_{ST} on a natural scale



Variance Partitions

$$V_B = \left(\vec{U}_1 \cdot \vec{Z}\right)^2 + \left(\vec{U}_2 \cdot \vec{Z}\right)^2$$

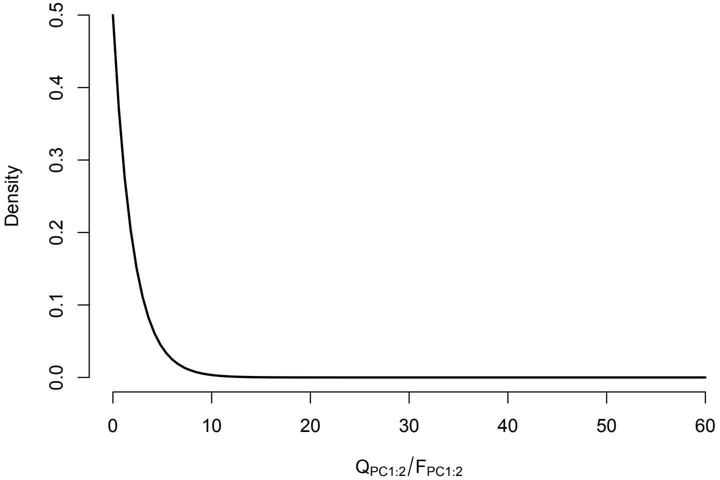
$$V_W = \sum_{i=3}^{K-1} \left(\vec{U}_i \cdot \vec{Z}\right)^2$$

$$F_{PC1:2} = \frac{\lambda_1 + \lambda_2}{\sum_{i=1}^{K-1} \lambda_i}$$

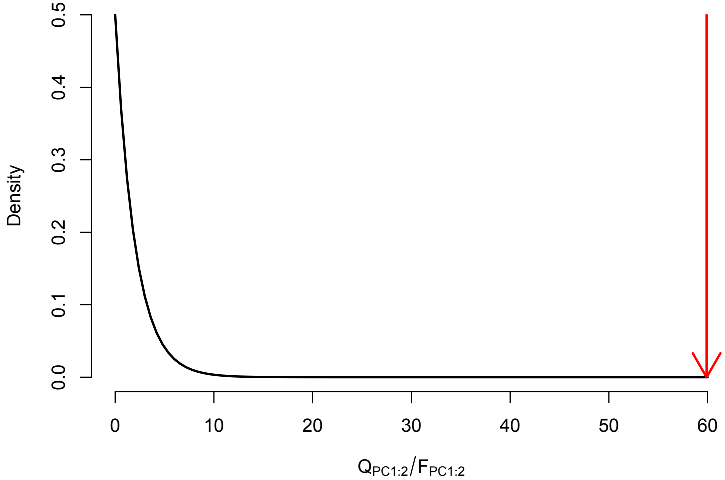
The Eigenvalue Decomposition (Principal Components)

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

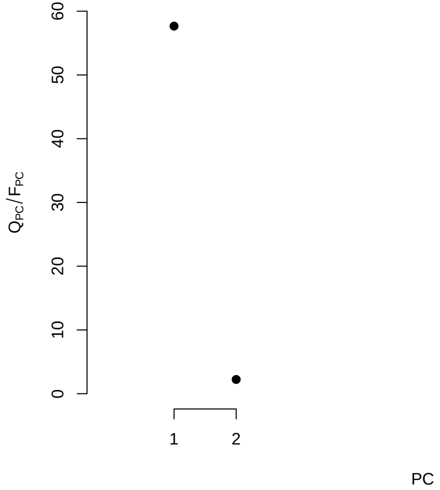
Genetic Divergence for Height in Europe



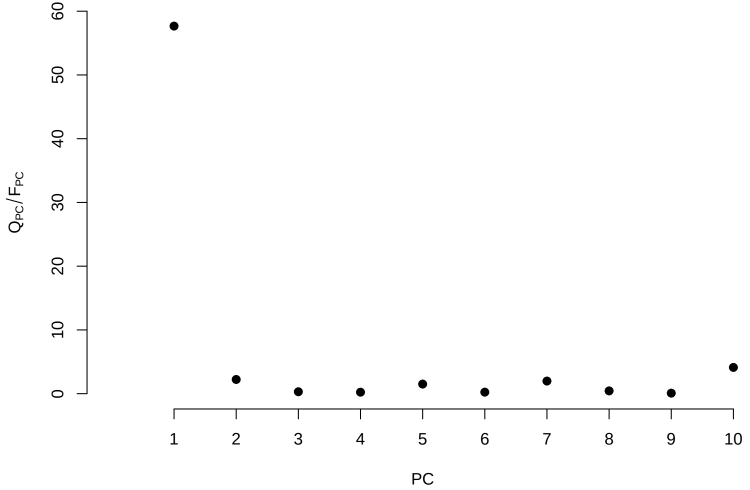
Genetic Divergence for Height in Europe



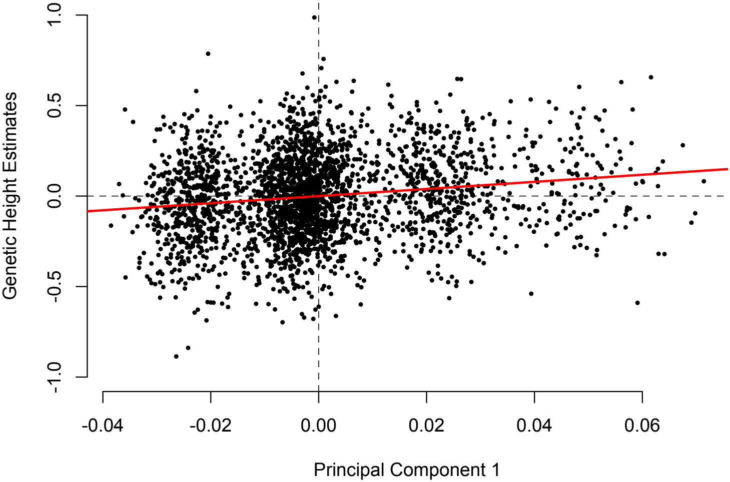
Genetic Divergence in Height Along PC1 in Europe



Genetic Divergence in Height Along PC1 in Europe



Genetic Divergence in Height Along PC1 in Europe



Take Aways

- Significant but subtle correlation between genetic values for human height and PC1/N-S axis in Europe (but we already knew that)

Take Aways

- Significant but subtle correlation between genetic values for human height and PC1/N-S axis in Europe (but we already knew that)
- Q_{ST}/F_{ST} can be formulated in terms of projections onto reduced rank factorizations of the individual-by-individual kinship matrix
  - Relationship to PCA, *structure*, factor analysis in general
    - Engelhardt and Stephens 2010
  - Every word of caution applicable to PCA/*structure* etc. applies here as well

Things I Didn't Mention

- Builds on generalized Q_{ST}/F_{ST} framework of Ovaskainen et al 2011

Things I Didn't Mention

- Builds on generalized Q_{ST}/F_{ST} framework of Ovaskainen et al 2011
- Using G matrix information, can include multiple correlated traits in a single analysis, but interpretation potentially trickier
  - Chenoweth and Blows (2008)
  - Martin et al (2008)

Things I Didn't Mention

- Builds on generalized Q_{ST}/F_{ST} framework of Ovaskainen et al 2011
- Using G matrix information, can include multiple correlated traits in a single analysis, but interpretation potentially trickier
  - Chenoweth and Blows (2008)
  - Martin et al (2008)
- Breeding designs easily be incorporated to help with estimation of genetic variance parameters

Thanks!

NSF Graduate Research Fellowship Program

The Center for Population Biology aims to advance understanding of the origins & maintenance of diversity in biological systems through ecological & evolutionary research.

- Coop Lab
  - Alisa
  - Gideon
  - Simon
  - Kristin
  - Chenling
  - Graham
- Annie Schmitt Lab
- Jeff Ross-Ibarra Lab
- Simon Myers