Differences between populations

7 December 2016.

Joe Felsenstein

Biology 550D

Evolutionary forces acting on populations

- Natural selection toward optima (which can differ between populations)

Evolutionary forces acting on populations

- Natural selection toward optima (which can differ between populations)
- Genetic drift, making populations differ

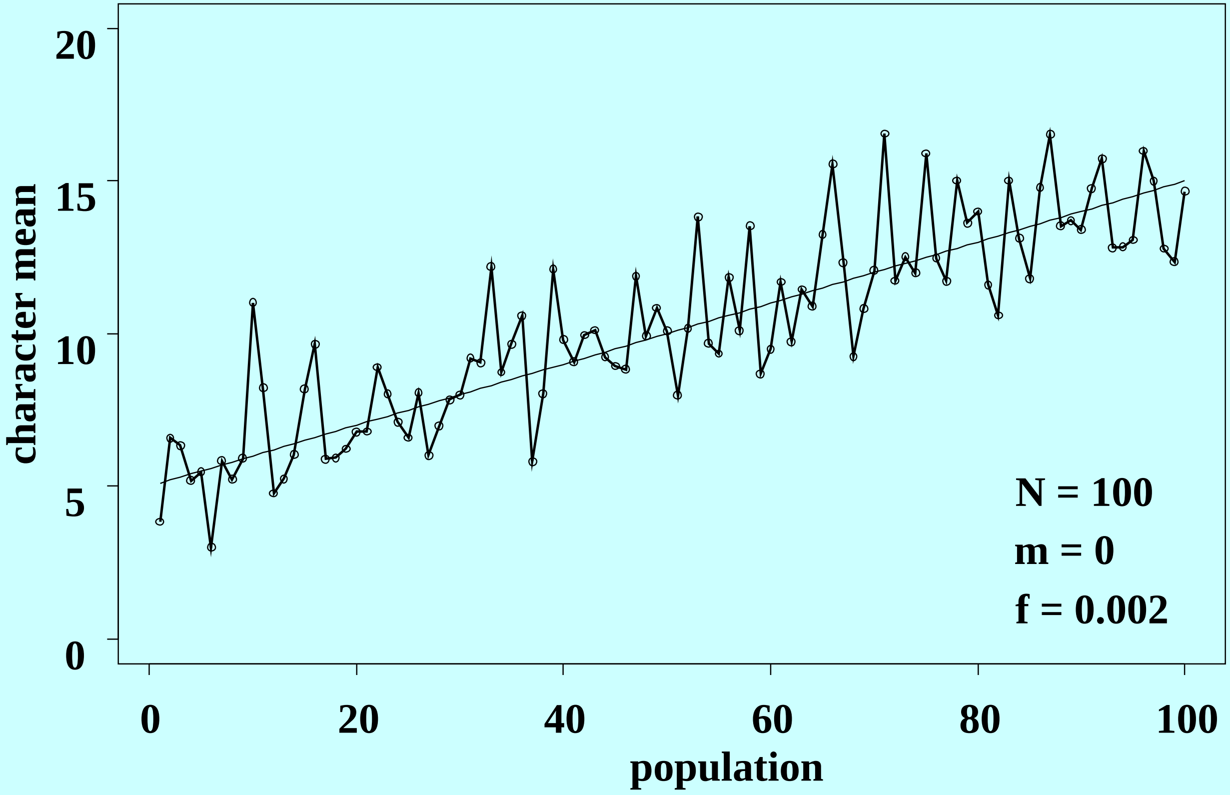
Evolutionary forces acting on populations

- Natural selection toward optima (which can differ between populations)
- Genetic drift, making populations differ
- Migration, making adjacent populations more similar

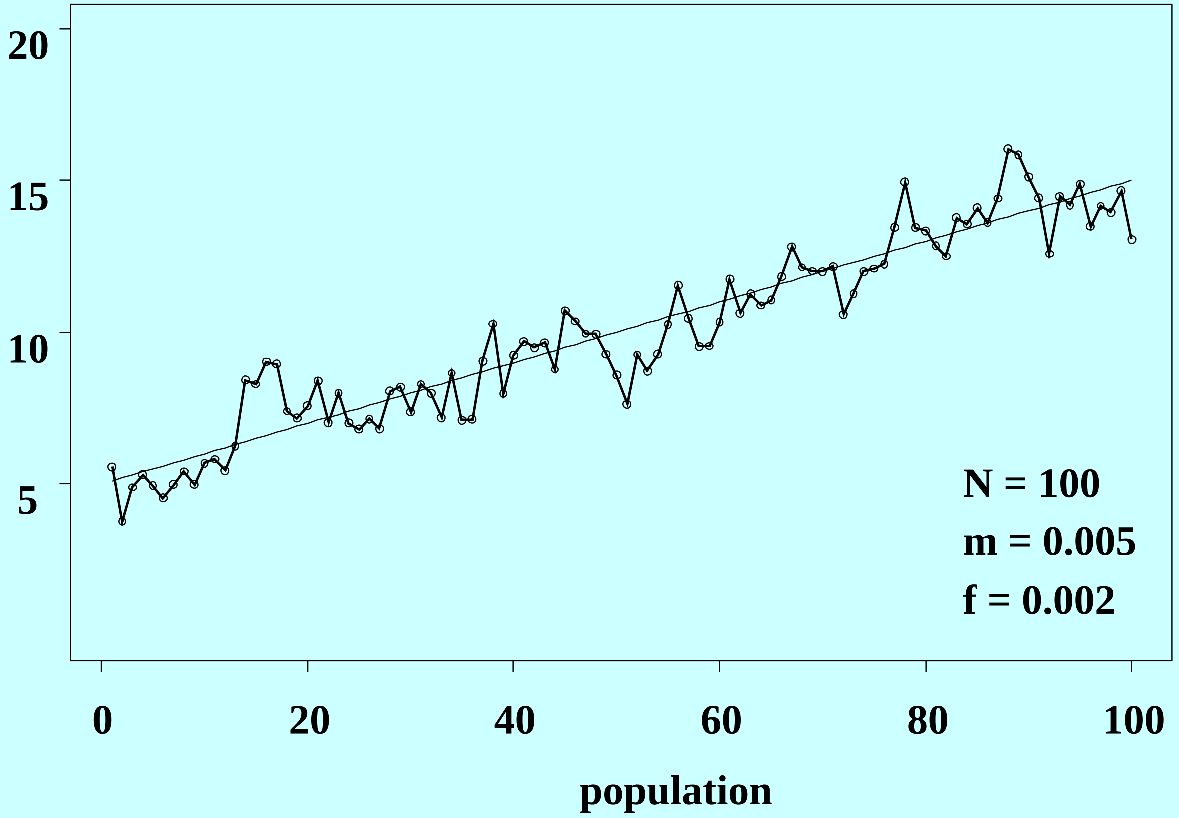
Evolutionary forces acting on populations

- Natural selection toward optima (which can differ between populations)
- Genetic drift, making populations differ
- Migration, making adjacent populations more similar

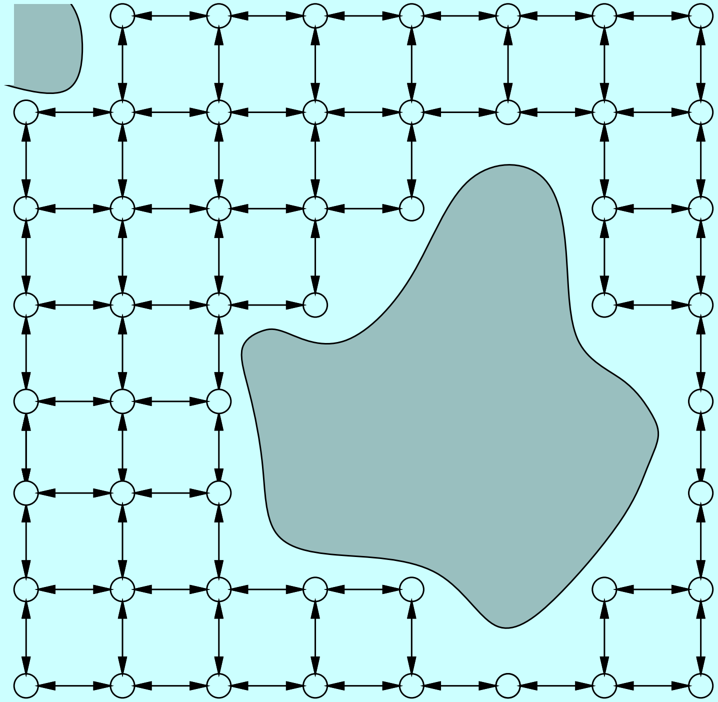
An simulated example with no migration



An simulated example with migration

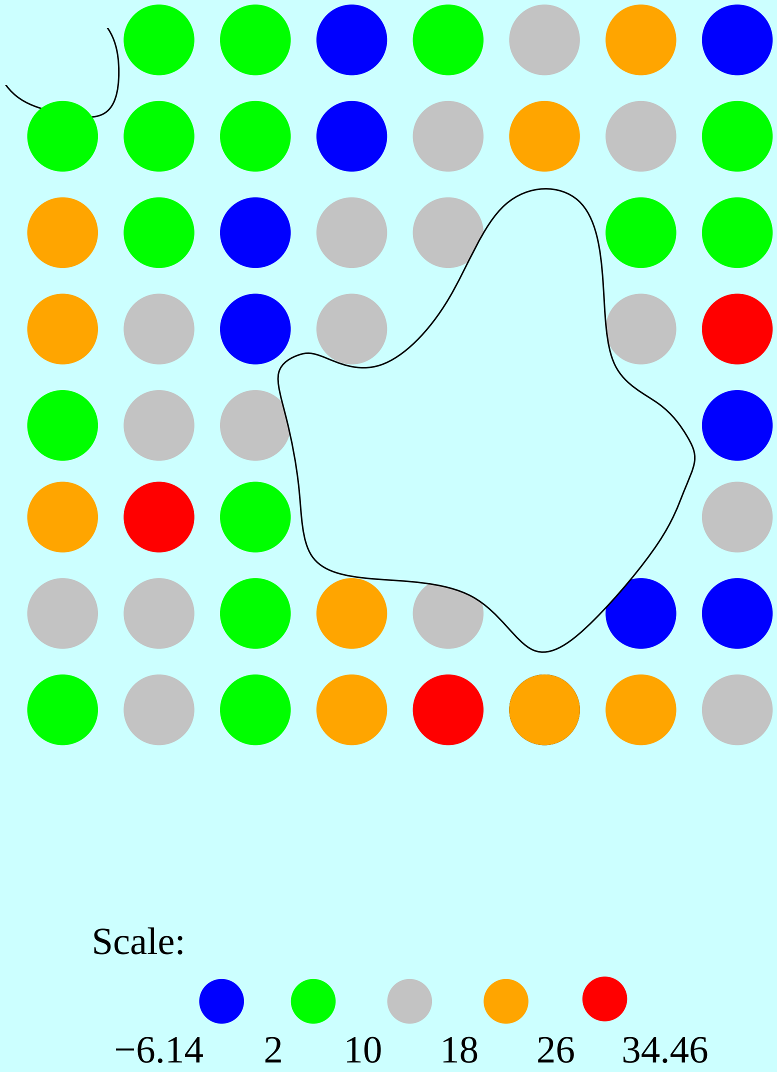


A geographic region with local migration

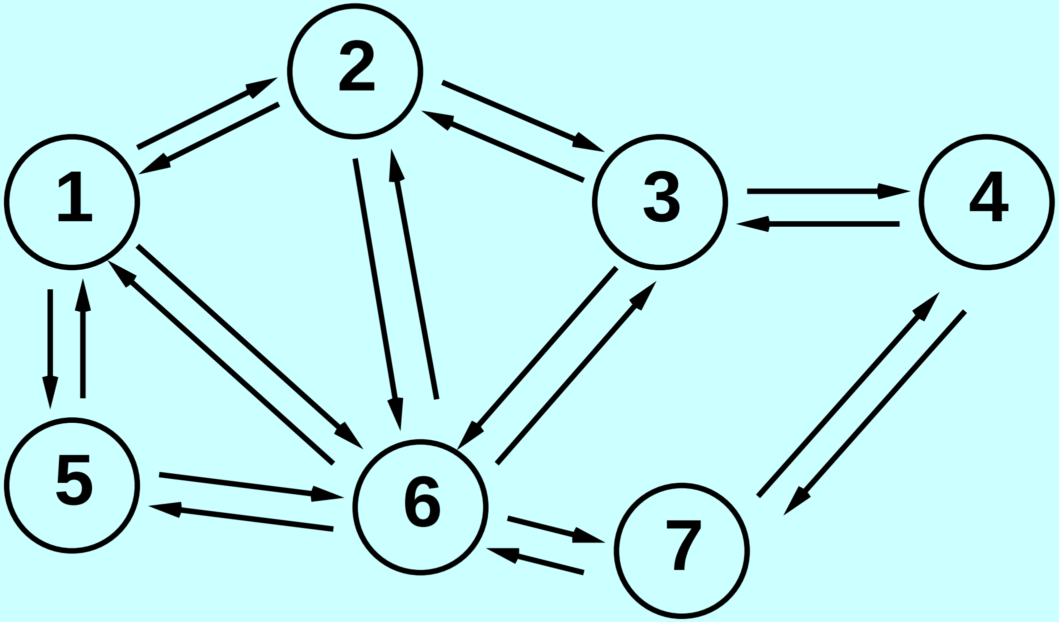


Phenotypes resulting from genetic drift alone

An example from simulation

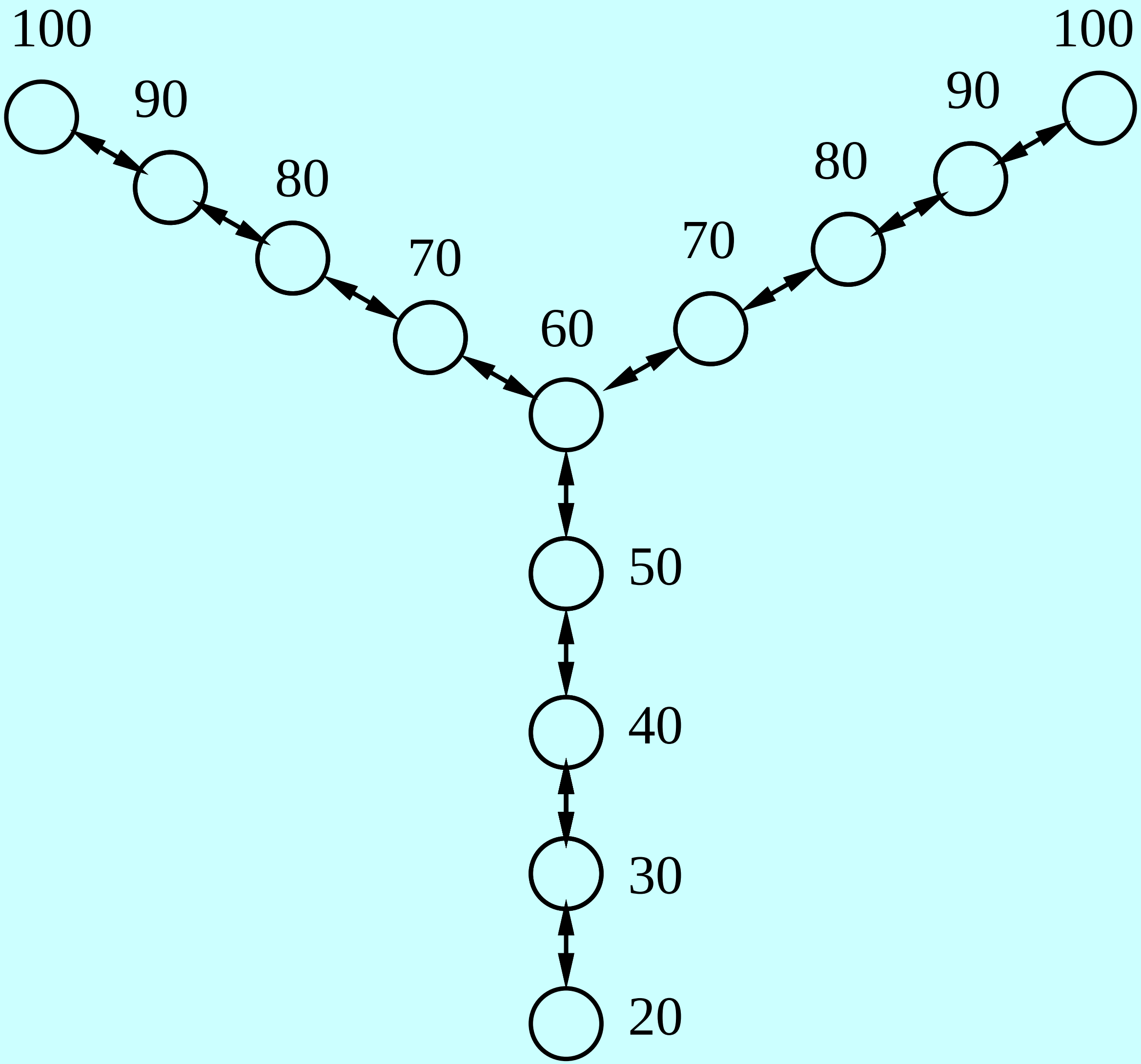


A more general migration matrix



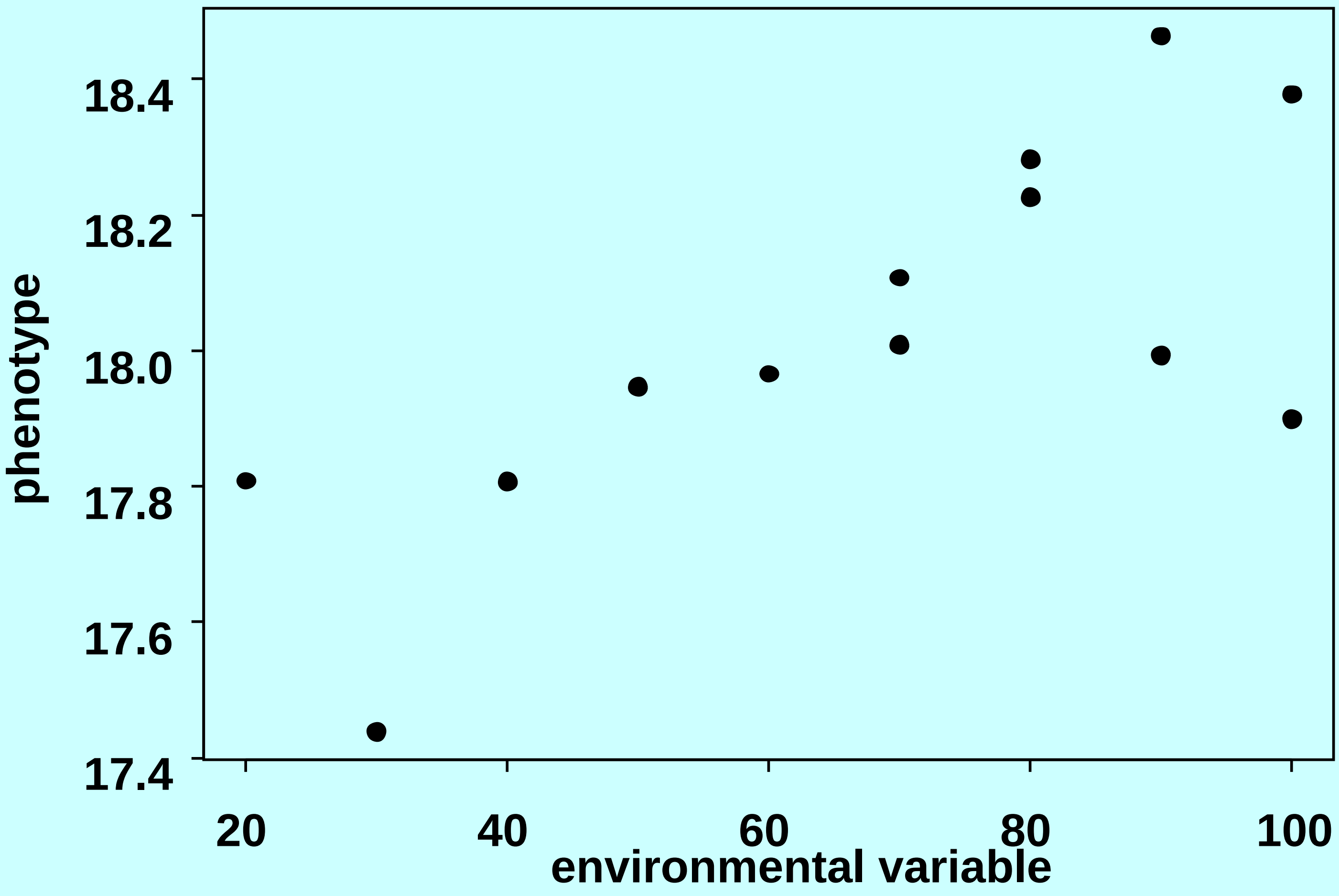
a separate m_{ij} between each pair of connected populations

13 populations and an environmental variable

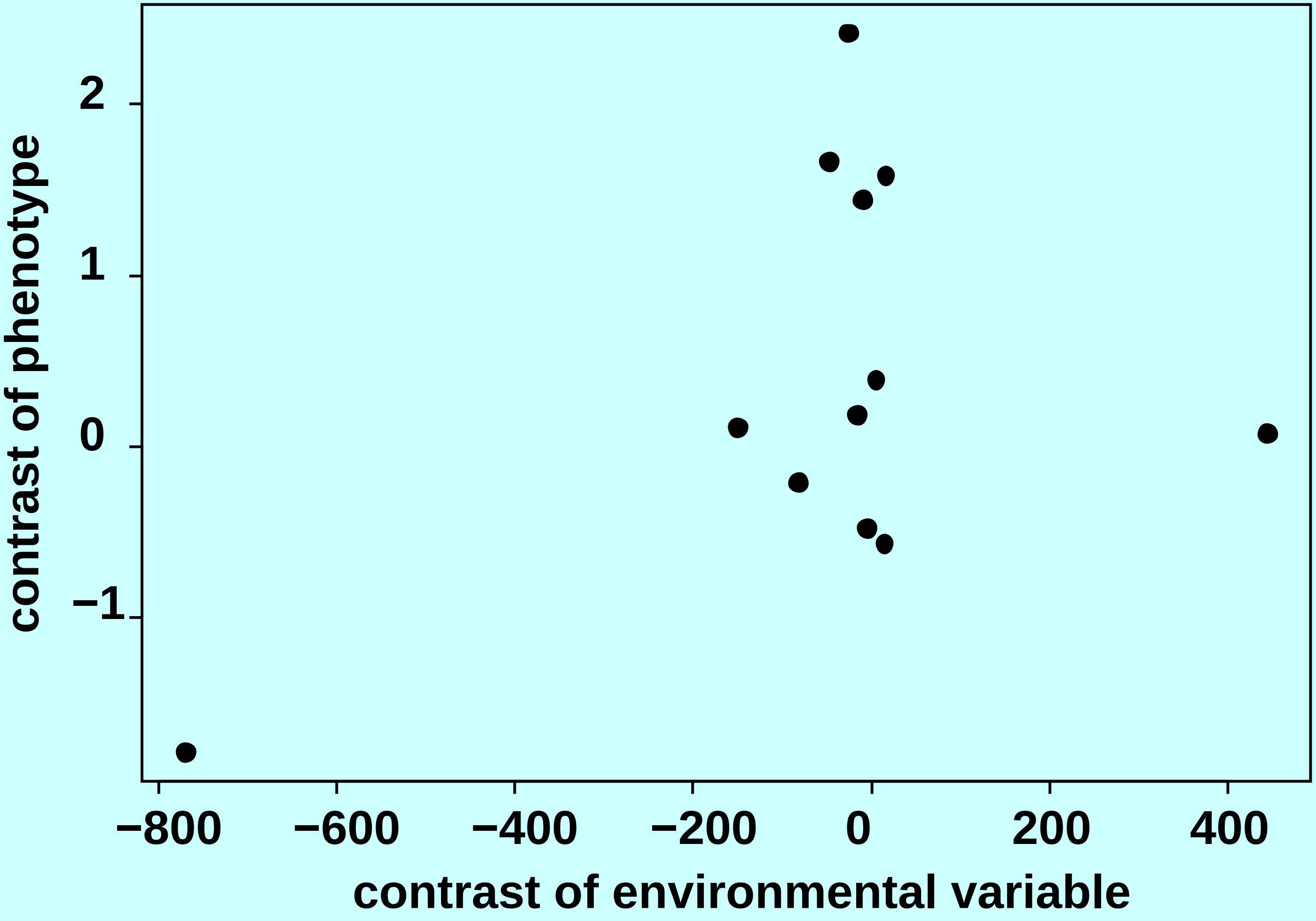


Simulated with $N = 100, \quad m = 0.2, \quad a = 10, \quad b = 0.1, \quad \sigma = 0.0001$

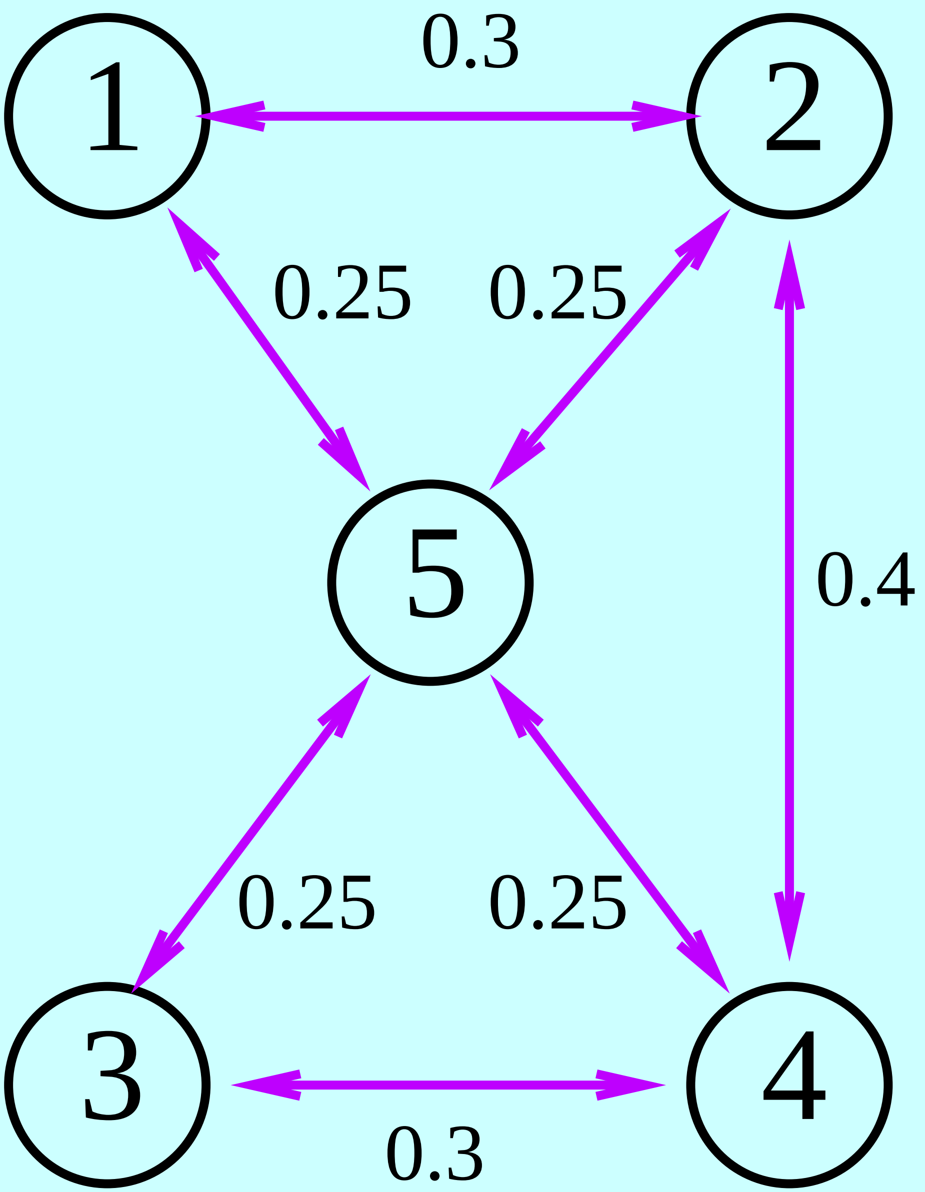
phenotype plotted against the environmental variable



contrasts of phenotype versus constrast of environment

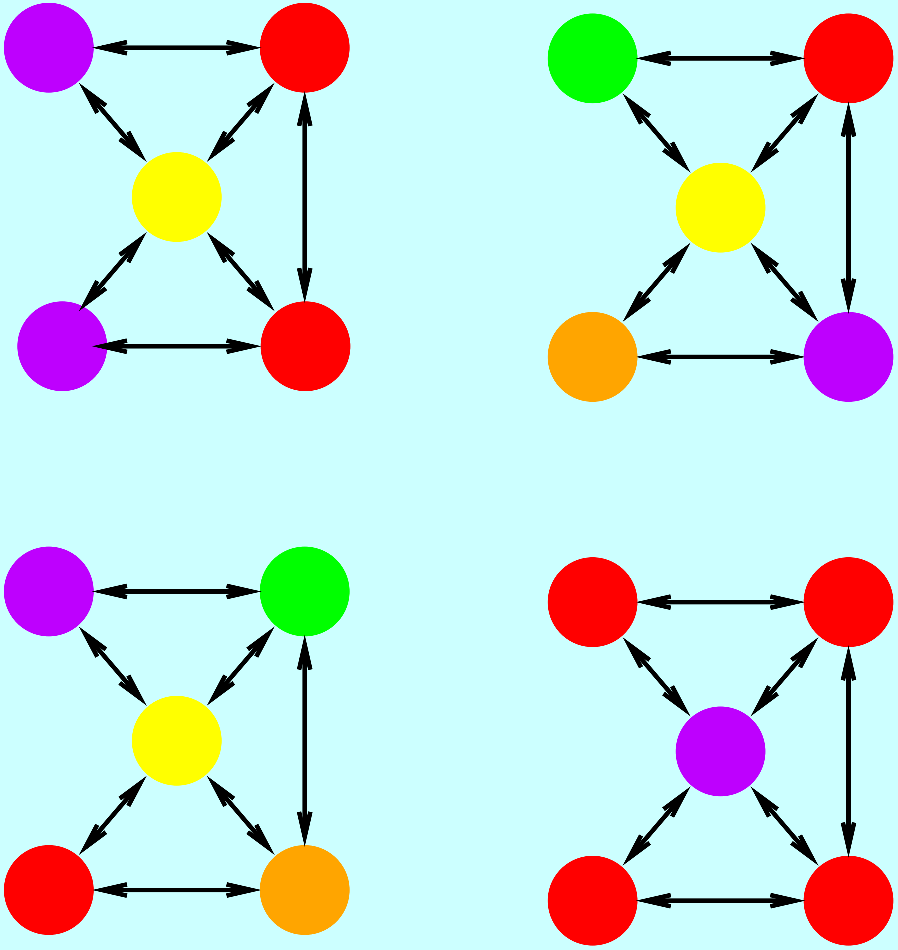


A numerical example with a 5-population matrix

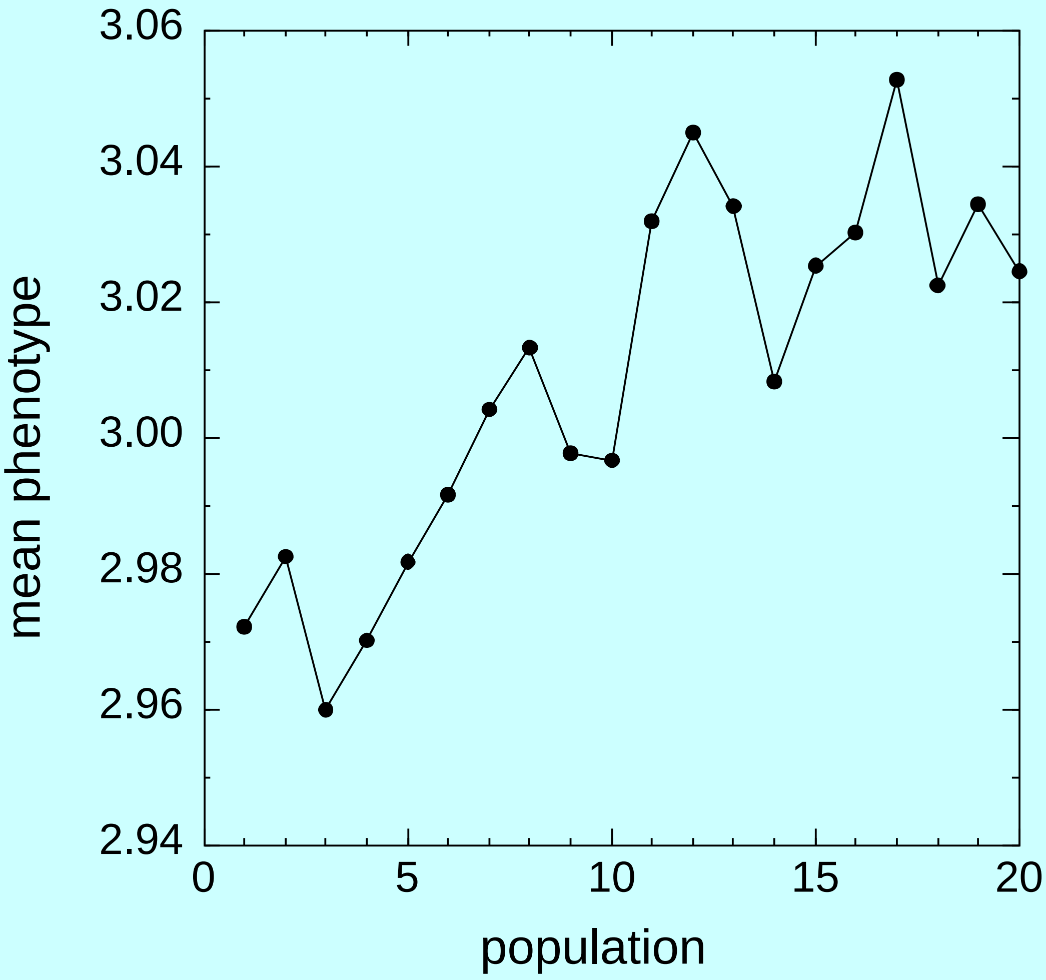


The contrasts from that 5-population matrix

key: 🟪 ++ 🟩 + 🟨 0 🟧 − 🟥 −−

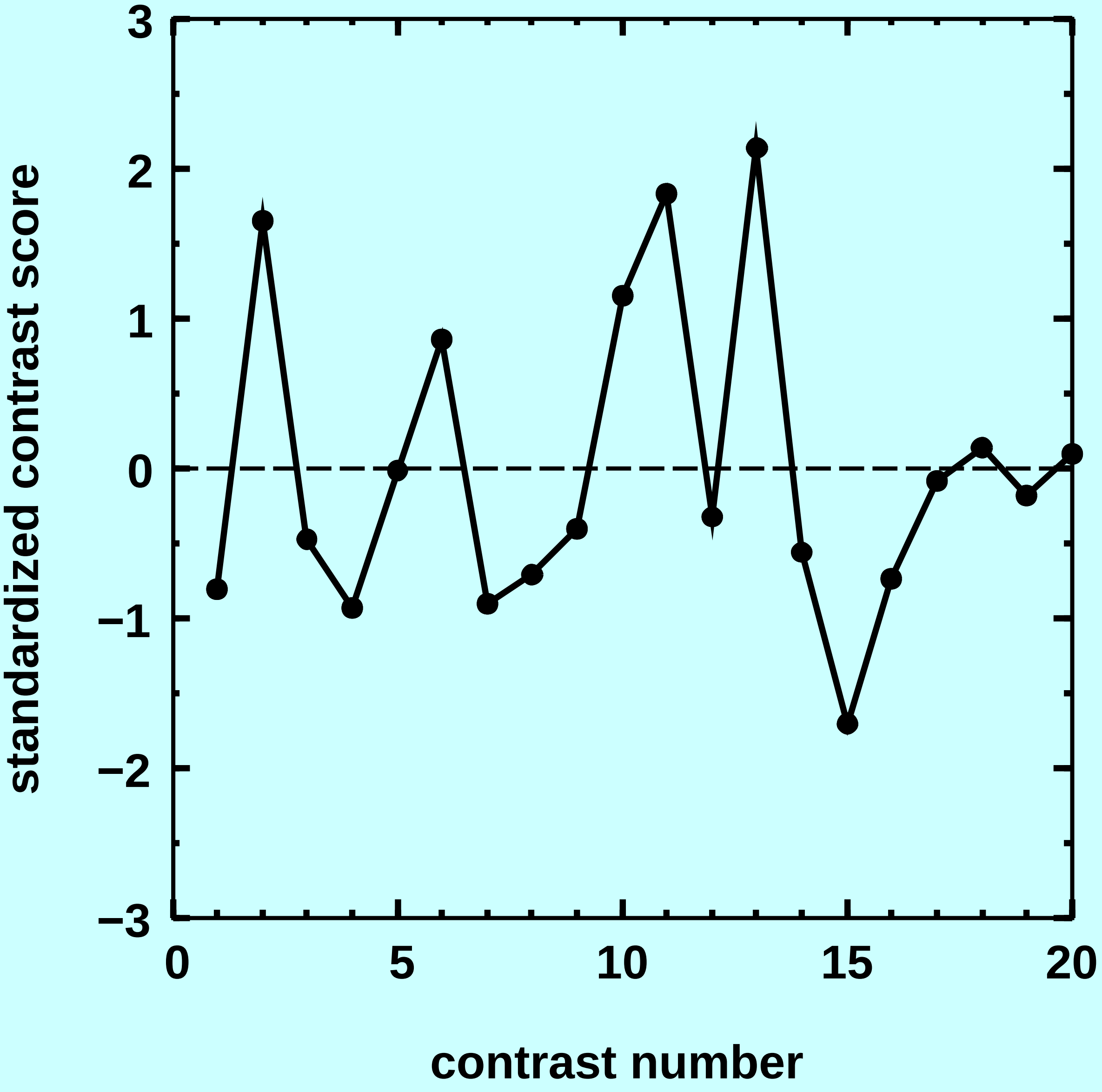


A simulated example



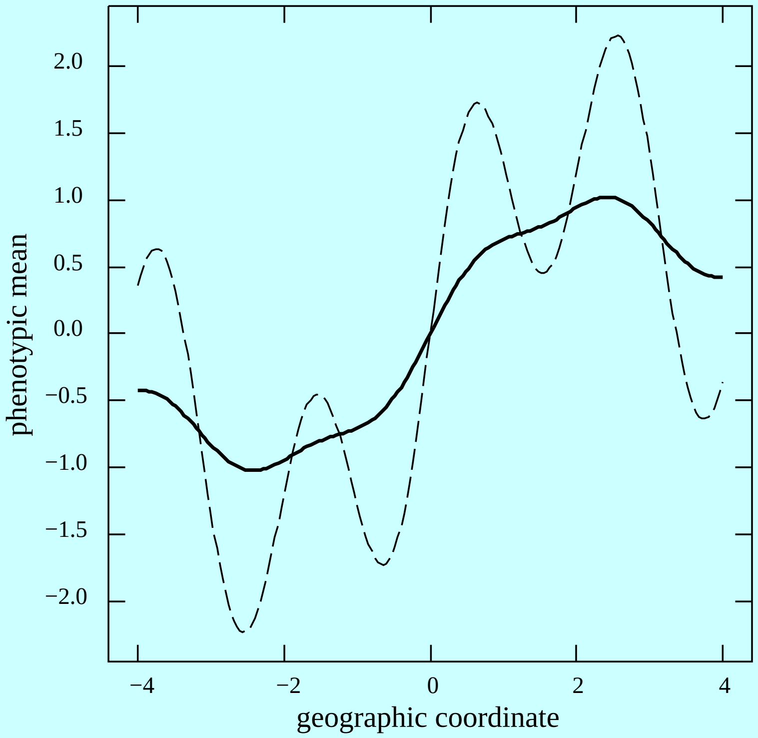
(No selection, with $m = 0.01$ and $\beta = 0.02$)

Its contrasts



Fourier transform of characters

When we have a linear “stepping-stone" model of population structure, the appropriate transform is simply the discrete Fourier transform. Migration damps the sine waves of different wavelengths differently. Here is a linear continuum with the optimum phenotype curve the sum of three sine waves, and the corresponding curve of the phenotypic means that results when migration damps them:



A worry: direct effects of the environment

If the responses to the environment are not genetic but are the direct effect on the phenotype, we might get wrong inferences about selection.

- We could escape this by “common garden” experiments but this is hard.
- There are statistical ways of detecting this – basically that migration among neighbors will not make them more similar than corresponding pairs of populations that have the same two environments but are farther away from each other. Power of inference?

Previous work

Previous papers on this:

- Malécot (1949) and Kimura and Weiss (1964) used Fourier transforms to solve for identity by descent in one- and two-dimensional infinite stepping stone models of gene frequencies.
- Much subsequent work by Takeo Maruyama (1971ff.) on this for finite one- and two-dimensional stepping stone models.
- Hansen, Armbruster and Antonsen (*American Naturalist*, 2000) used Fourier transforms to correct for migration in analyzing phenotypic data in one-dimensional stepping stone models. This work is in effect a generalization of theirs to the general migration matrix model of Bodmer and Cavalli-Sforza (1965).

Previous work, continued

- Much of the present algebra was already published by me:

  Felsenstein, J. 2002. Contrasts for a within-species comparative method. pp. 118-129 in M. Slatkin and M. Veuille, *Modern Developments in Theoretical Population Genetics.* Oxford University Press, Oxford.

- See also relevant agonizing about what we can infer about migration matrices in my paper in Journal of Theoretical Biology, 1981.
- Inference of migration matrices by sampling (IS and MCMC) methods: papers by Beerli and Felsenstein (1999, 2001) and Bahlo and Griffiths (2000).

Furthermore ...

Is there some way of thinking about how we can infer, from distributions of gene frequencies, the whole set of migration matrices and historical branching events that are not rejected? This raises some very interesting algebraic issues.

It is an example (like pedigree estimation and invariants of phylogenies) of a highly structured set of hypotheses, all with the same number of degrees of freedom, where we need to know which ones the data supports.

In this sense the problem is not so statistically simpleminded.