Evolutionary Genetics

LV 25600-01 | Lecture with exercises | 4KP

Genetic differentiation refers to the variation in genetic composition between populations of a species. When populations are isolated or exposed to different selective pressures, genetic differences can accumulate over time. This can result in distinct genetic traits or variations between populations. Genetic differentiation can be measured using different methods that analyse genetic variation within and between populations.

F-statistics (F_{ST}) - It's calculated based on differences in allele frequencies and provides a quantitative measure of population differentiation.

Analysis of Molecular Variance (AMOVA) - This method partitions genetic variance at different hierarchical levels (e.g. within individuals, between individuals within populations, between populations) to estimate the extent of genetic differentiation.

Principal component analysis (PCA) or multidimensional scaling (MDS) - These methods visualise genetic variation by reducing the dimensions of the data, allowing you to see patterns and clusters that represent genetic differences between populations.

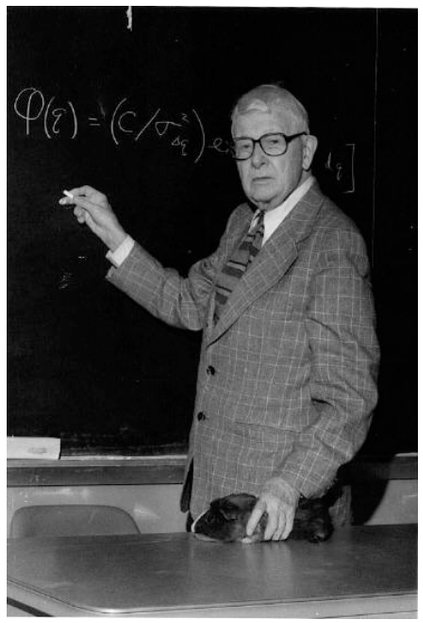
Admixture and structure analysis - These tools assess population structure and admixture by inferring individual ancestry based on genetic markers, providing insight into the degree of genetic differentiation and admixture between populations.

Phylogenetic analysis - The construction of phylogenetic trees based on genetic data helps to visualise evolutionary relationships and distances between populations.

Genome-wide association studies (GWAS) - These studies can identify genetic variants associated with population differences in traits or diseases, indirectly revealing genetic differentiation.

The choice of method often depends on the genetic data available, the specific research question and the scale of genetic differentiation being studied (e.g. within a region, between continents). These methods are often used together to gain a comprehensive understanding of genetic differentiation between populations.

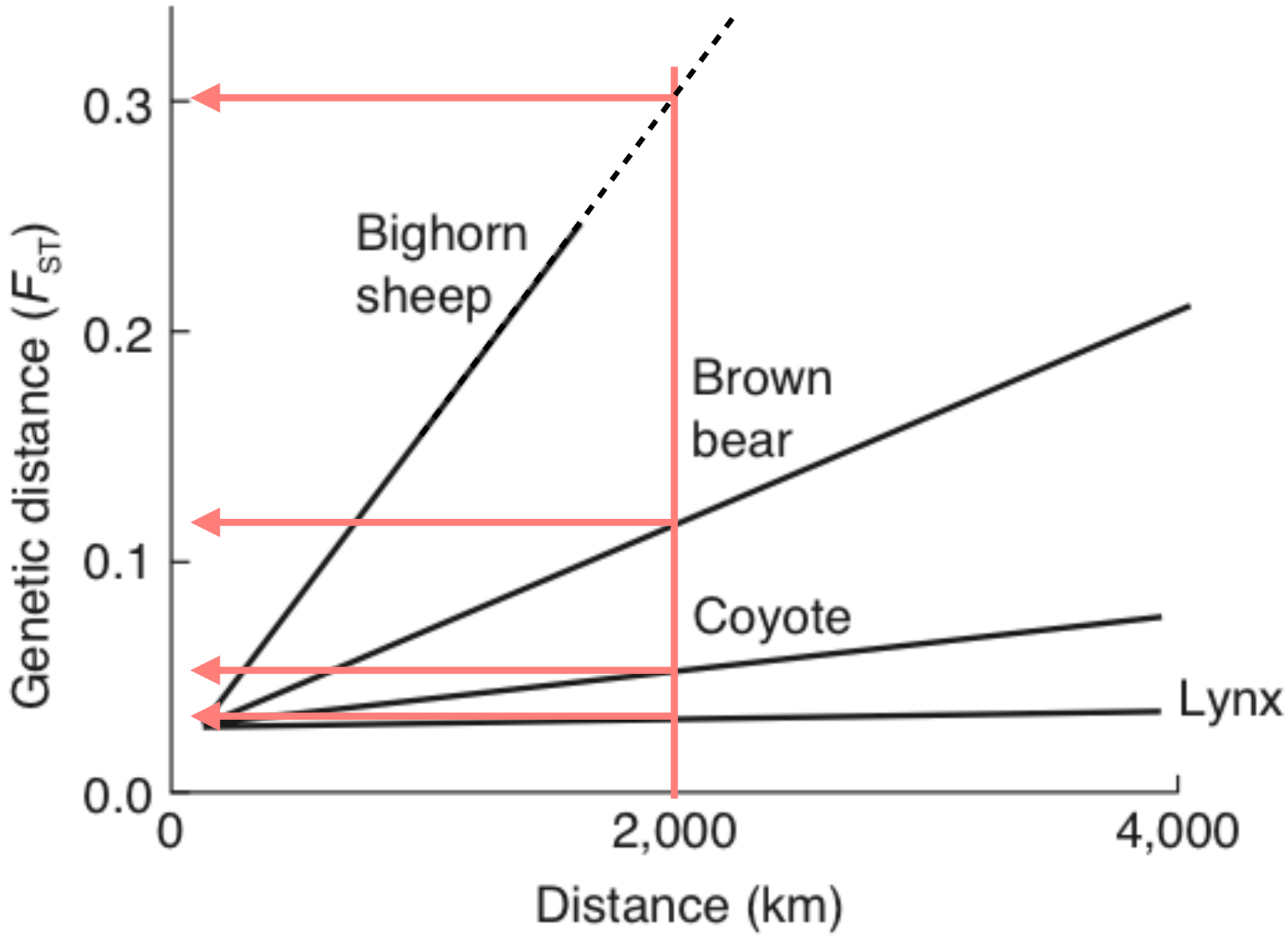
The oldest and most widely used metrics of **genetic differentiation** are **F--statistics**. Sewall Wright (1931, 1951) developed a conceptual and mathematical framework to describe the distribution of genetic variation within a species that used a series of inbreeding coefficients: F_{IS}, F_{ST}, and F_{IT}.



Sewall Wright
1889-1988

Masatoshi Nei
1931-2023

F-statistics were initially defined by Wright for loci with just two alleles. They were extended to three or more alleles by Nei (1977), who used the parameters G_{IS}, G_{ST}, and G_{IT} in what he termed the analysis of gene diversity. F- and G-statistics are often used interchangeably in the literature – see Chakraborty and Leimar (1987) for a comprehensive discussion of F- and G-statistics.



Source: Forbes and Hogg (1999)

The inbreeding coefficient of an individual (F) is the probability that an individual has two alleles at a locus that are identical by descent. It measures the amount of inbreeding by comparing the frequency of heterozygotes (H_o) in the population to the frequency expected under random mating (H_e).

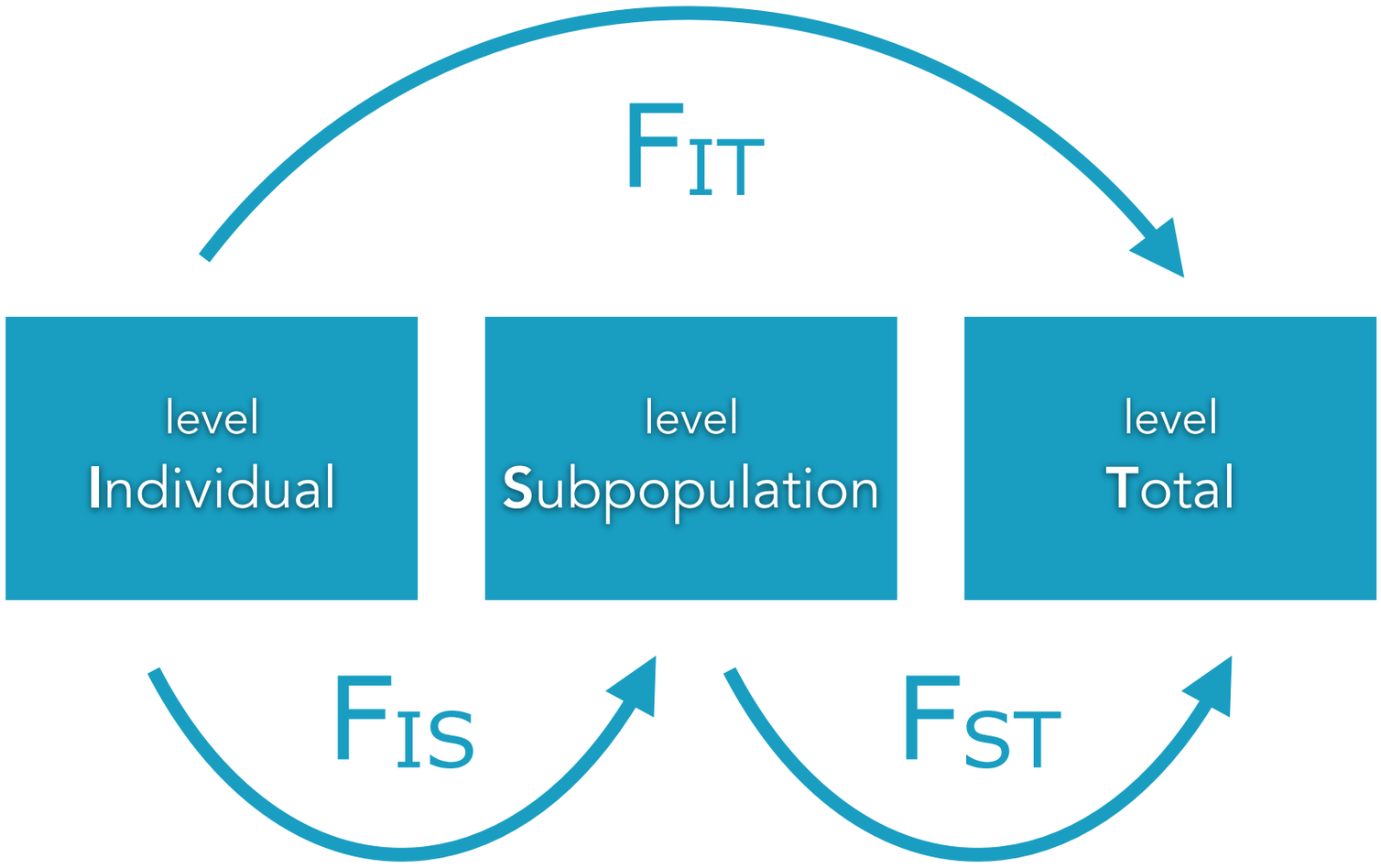
"...Sewall Wright used this similarity between genetic drift and inbreeding to create **F-statistics**, which provide an integrated view of genetic variation at three hierarchical levels of population structure..."

Source: Conner and Hartl 2004

***F* is the proportion by which heterozygosity is reduced relative to heterozygosity in a random mating population with the same allele frequencies:**

$$F = 1 - \frac{H_o}{H_e} = 1 - \frac{H_o}{2pq} = \frac{H_e - H_o}{H_e}$$

when population in HWE $H_e=H_o \rightarrow F=0$ (random mating)



The different F-statistics look at different levels of population structure. F_{IT} is the inbreeding coefficient of an individual (I) relative to the total population (T) - F_{IS} is the inbreeding coefficient of an individual (I) relative to the subpopulation (S) - and F_{ST} is the effect of the subpopulations (S) relative to the total population (T).

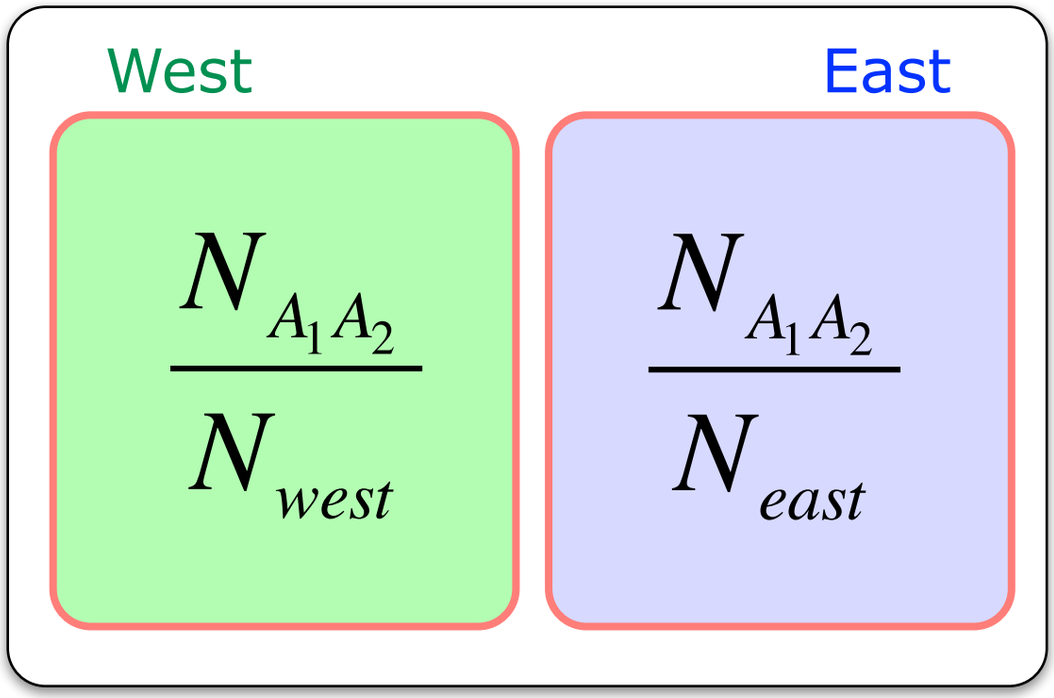

Population

Population

Expected heterozygosity without subdivision

$$H_T = 2pq_{total}$$

Population - Subpopulations



Expected heterozygosity without subdivision

$$H_T = 2pq_{total}$$

Observed average heterozygosity in subpopulations

$$H_I = \frac{(\#A_1A_{2west}/N_{west}) + (\#A_1A_{2east}/N_{east})}{2}$$

Population - Subpopulations

West

$$\frac{N_{A_1A_2}}{N_{west}}$$

East

$$\frac{N_{A_1A_2}}{N_{east}}$$

Expected heterozygosity without subdivision

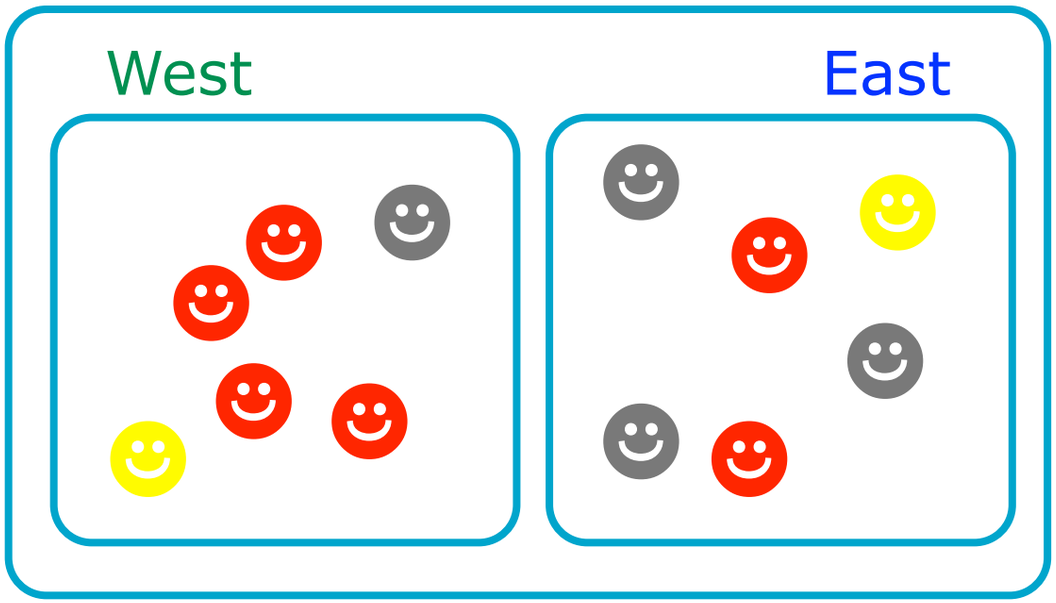
$$H_T = 2pq_{total}$$

Average observed heterozygosity in subpopulations

$$H_I = \frac{(\#A_1A_{2west}/N_{west}) + (\#A_1A_{2east}/N_{east})}{2}$$

Expected average heterozygosity assuming HWE

$$H_S = \frac{(2pq_{west}) + (2pq_{east})}{2}$$



Observed average heterozygosity in subpopulations

$$H_I = \frac{(\#A_1A_{2west}/N_{west}) + (\#A_1A_{2east}/N_{east})}{2}$$

Expected average heterozygosity assuming HWE

$$H_S = \frac{(2pq_{west}) + (2pq_{east})}{2}$$

Heterozygosity expected without subdivision

$$H_T = 2pq_{total}$$

among sub-populations

$$F_{IS} = 1-(H_I/H_S)$$

within populations

$$F_{ST} = 1-(H_S/H_T)$$

among populations/within sub-population

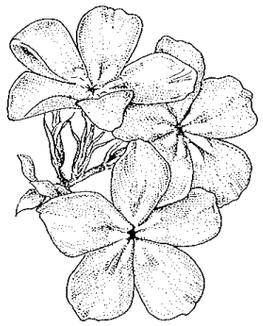
$$F_{IT} = 1-(H_I/H_T)$$

$$F_{ST} = (F_{IT}-F_{IS})/(1-F_{IS})$$

BOX 3.1 *Calculation of F-statistics*

Levin (1978) scored allele frequencies at the *Pgm-2* locus in 43 Texas subpopulations of *Phlox cuspidata*. Forty of these subpopulations were fixed for the *b* allele (listed together in the first row of the table below). In the other three subpopulations the frequencies of *b* were 0.49, 0.83, and 0.91, with observed heterozygote frequencies of 0.17, 0.06, and 0.06, respectively:

Subpopulation	p_i	H_I
1–40	1	0
41	0.49	0.17
42	0.83	0.06
43	0.91	0.06



Phlox cuspidata

From these data we can calculate the three hierarchical *F*-statistics and their components as follows:

$\overline{H}_I$ = the observed proportion (frequency, not numbers) of heterozygotes within subpopulations, averaged over all subpopulations:

$$\overline{H}_I = \frac{(40\times 0)+0.17+0.06+0.06}{43} = 0.0067$$

$\overline{H}_S$ = the expected proportion of heterozygotes within subpopulations, assuming random mating (= $2p_iq_i$), averaged over all *n* subpopulations:

$$\overline{H}_S = \frac{\sum_{i=1}^{n} 2p_iq_i}{n} =$$

$$\frac{(40\times 0)+2(0.49\times 0.51)+2(0.83\times 0.17)+2(0.91\times 0.09)}{43} = 0.0220$$

H_T = the expected proportion of heterozygotes over the entire metapopulation (=$2\bar{p}\,\bar{q}$). H_T is not itself an average because there is only one metapopulation, but since there is a unique value for *p* and *q* for each subpopulation, the average allele frequencies across all subpopulations are used:

$$\bar{p} = \frac{(40\times 1)+0.49+0.83+0.91}{43} = 0.9821$$

$$1-\bar{p} = \bar{q} = 0.0179$$

$$H_T = 2\bar{p}\bar{q} = 2\times 0.9821\times 0.0179 = 0.0352$$

Box 3.1 continued

We can now calculate the three *F*-statistics using the definitions given in the text:

$$F_{IS} = \frac{\overline{H}_S - \overline{H}_I}{\overline{H}_S} = \frac{0.0220-0.0067}{0.0220} = 0.70$$

$$F_{ST} = \frac{H_T - \overline{H}_S}{H_T} = \frac{0.0352-0.0220}{0.0352} = 0.38$$

$$F_{IT} = \frac{H_T - \overline{H}_I}{H_T} = \frac{0.0352-0.0067}{0.0352} = 0.81$$

Check these using equation 3.9:

$$(1-F_{IS})(1-F_{ST}) = (1-F_{IT})$$

$$(1-0.7)(1-0.38) = (1-0.81)$$

$$(0.3)(0.62) = 0.19$$

$$0.19 = 0.19$$

F_{ST} or F_{IT} are metapopulation level measures of population structure, quantifying the degree of subpopulation differentiation within the total population, and the overall amount of reduction in heterozygosity. Therefore, it doesn't make sense to calculate separate F_{ST} or F_{IT} for each subpopulation. Each subpopulation can have its own value for F_{IS}, however, because this is just the inbreeding coefficient we calculated before. To compare the reduction in heterozygosity across the three hierarchical levels, we use the average F_{IS} calculated previously.

In this example F_{IS} is very large, demonstrating a high level of inbreeding in these self-fertilizing plants. This high F_{IS} comes entirely from the three unfixed subpopulations. The forty fixed subpopulations do not contribute to this estimate of inbreeding, because there is no genetic variation and thus no heterozygotes to be reduced in frequency. Mathematically, these 40 subpopulations are represented by zeros in both H_I and H_S, so they do not affect F_{IS}. F_{ST} is also quite large, which is also likely due to the high level of self-fertilization. Much of gene flow in plants is through movement of pollen by wind or animal pollinators, so when most pollen stays on the same plant, as it does in highly selfing species, it greatly reduces gene flow from pollen movement among subpopulations. Note that this high differentiation is mainly due to subpopulation #41, which is the only one with a lower frequency of the *b* allele.

Phlox cuspidata, the pointed phlox, is a species of flowering plant in the family Polemoniaceae, native to the US states of Oklahoma, Texas, and Louisiana.

Phlox cuspidata is a perennial herbaceous plant that typically grows in clumps. It features narrow, lance-shaped leaves and produces showy clusters of flowers at the tips of its stems. The flowers can vary in color, including shades of pink, purple, blue, and white.

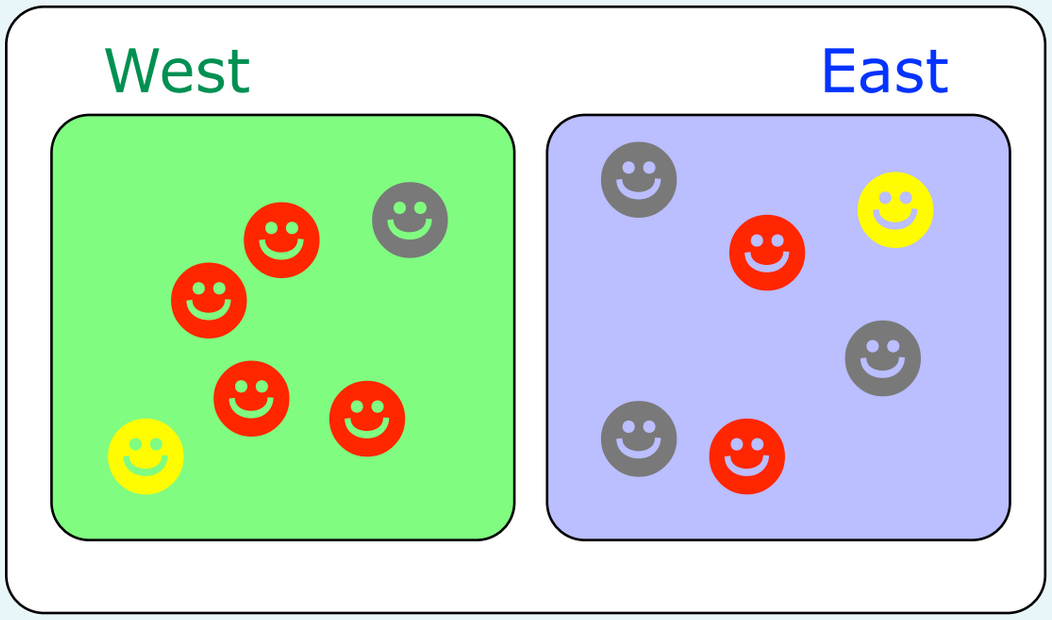
Phlox cuspidata, like many species within the Phlox genus, is capable of **self-pollination**, which means it has the ability to fertilize its own flowers. However, the extent to which it self-pollinates can vary among individual plants and populations.

Self-pollination can occur in plants through various mechanisms, such as the proximity of male and female reproductive parts within the same flower (self-fertilization) or through mechanisms that prevent or limit cross-pollination with other plants (selfing).

In the case of *Phlox cuspidata*, while it has the potential for self-pollination, its breeding system may also involve outcrossing, where pollen from one plant fertilizes another plant. The degree of selfing versus outcrossing can be influenced by factors such as the availability of pollinators, the structure of the flowers, and the genetic diversity within a population.

Fixation Index

$$F_{ST} = \frac{F_{IT} - F_{IS}}{1 - F_{IS}}$$



F_{ST} [0.05-0.15]
sub-populations are very similar

F_{ST} [0.15-0.25]
sub-populations are similar

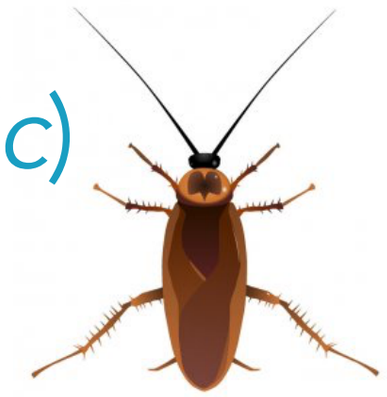
F_{ST} [>0.25]
sub-populations are distinct

Bed bugs (*Cimex lectulariu*)

$F_{ST} = 0.68$

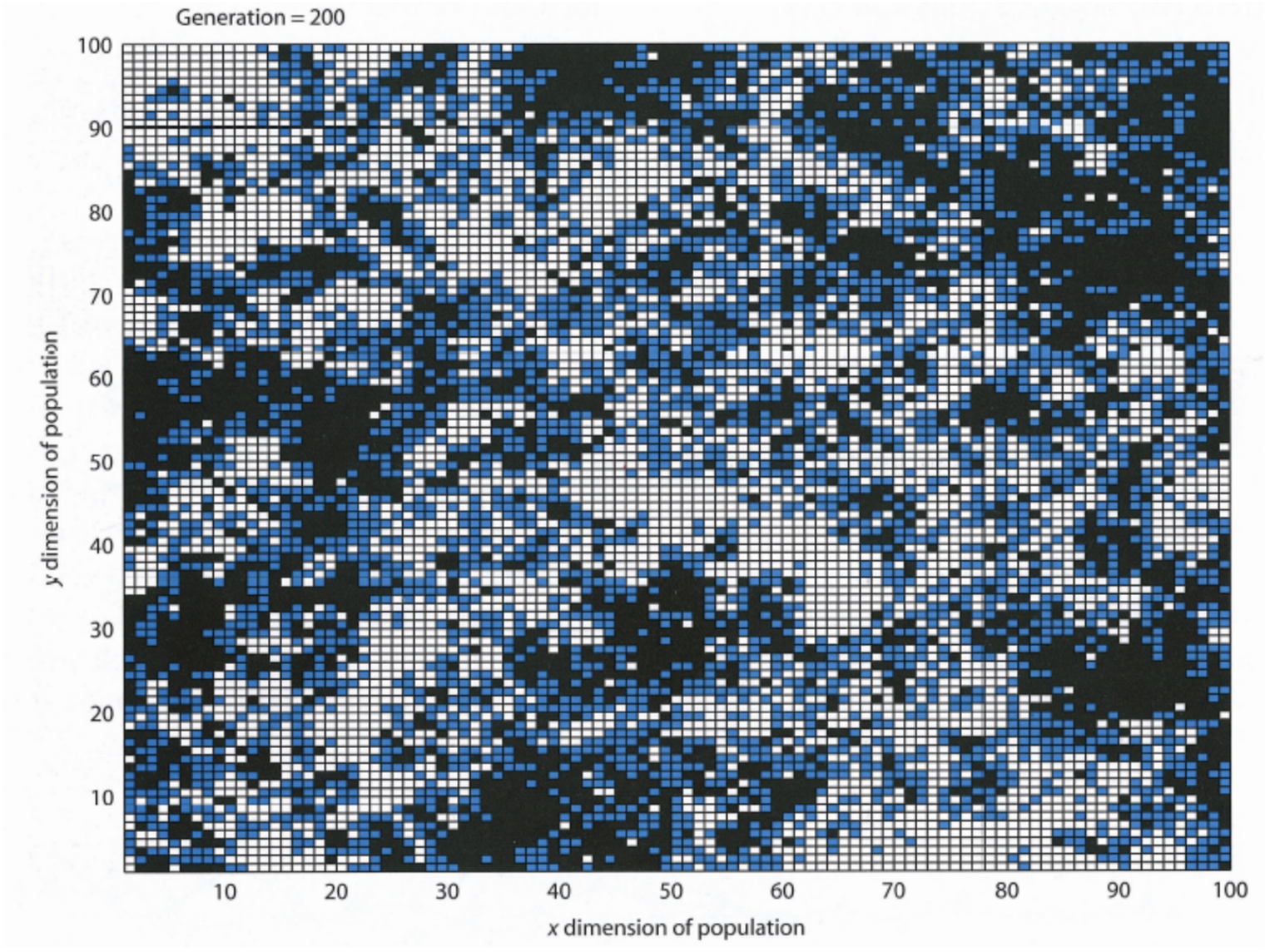
German cockroach (*Blattella germanic*)

$F_{ST} = 0.099$



Saenz et al. (2012) Genetic Analysis of Bed Bug Populations Reveals Small Propagule Size Within Individual Infestations but High Genetic Diversity Across Infestations From the Eastern United States.


MORE
THINGS
CONSIDERED
1



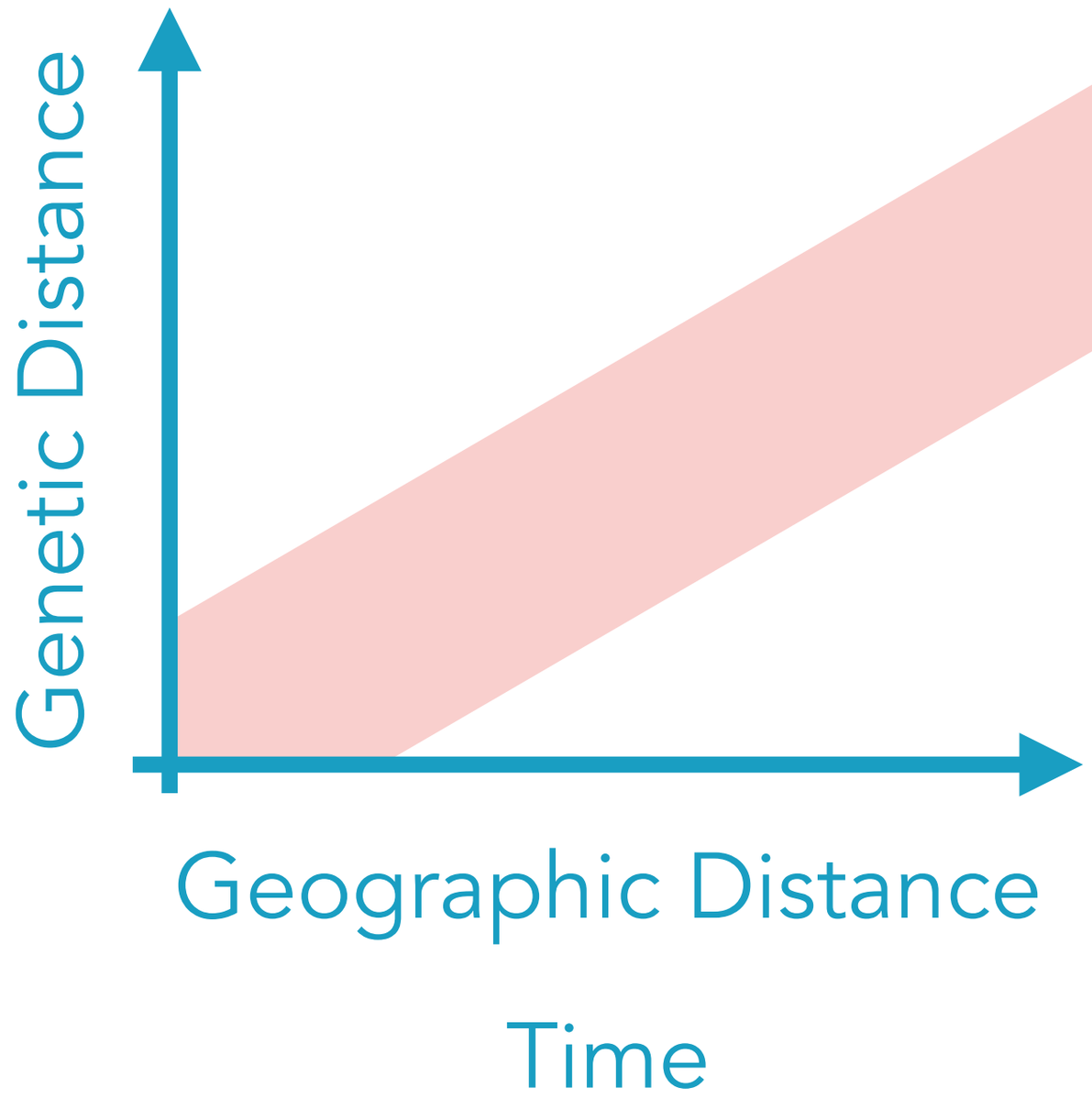
Non-Random Mating > Isolation by Distance > Subdivision

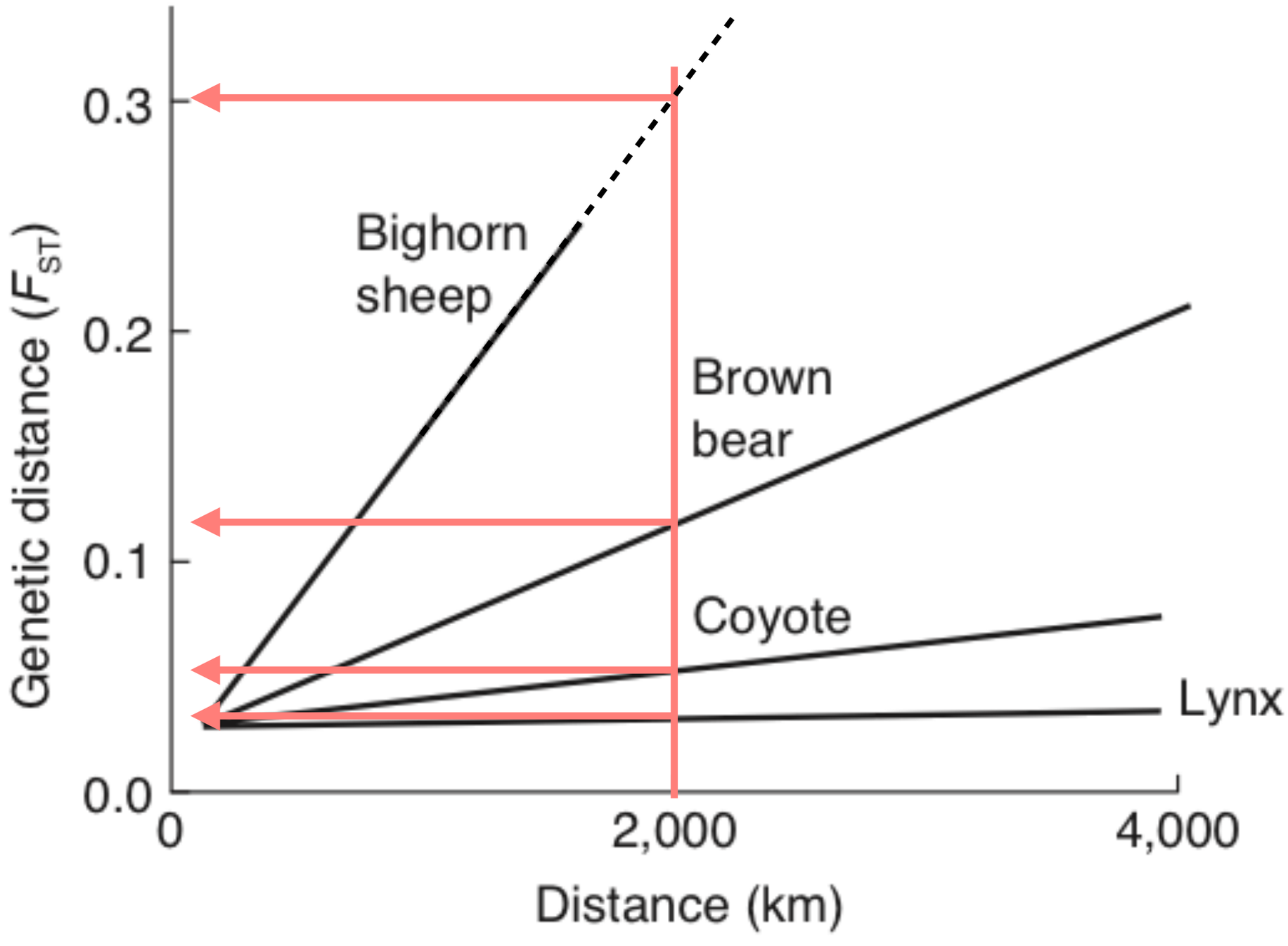
ISOLATION BY DISTANCE*

SEWALL WRIGHT
The University of Chicago[1]

Received November 9, 1942

STUDY of statistical differences among local populations is an important line of attack on the evolutionary problem. While such differences can only rarely represent first steps toward speciation in the sense of the splitting of the species, they are important for the evolution of the species as a whole. They provide a possible basis for intergroup selection of genetic systems, a process that provides a more effective mechanism for adaptive advance of the species as a whole than does the mass selection which is all that can occur under panmixia.


Genetic Distance
Geographic Distance
Time



Forbes and Hogg (1999)

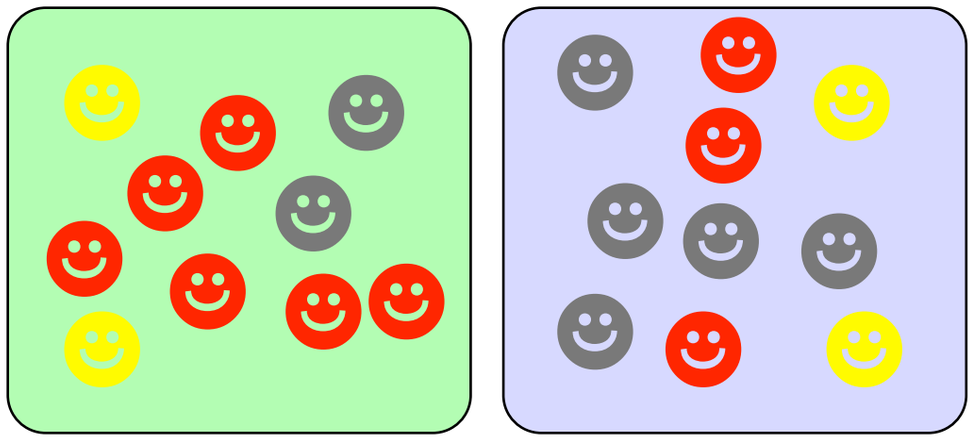

Population

Population

Expected heterozygosity without subdivision

$$H_T = 2pq_{total}$$

Population - Subpopulations

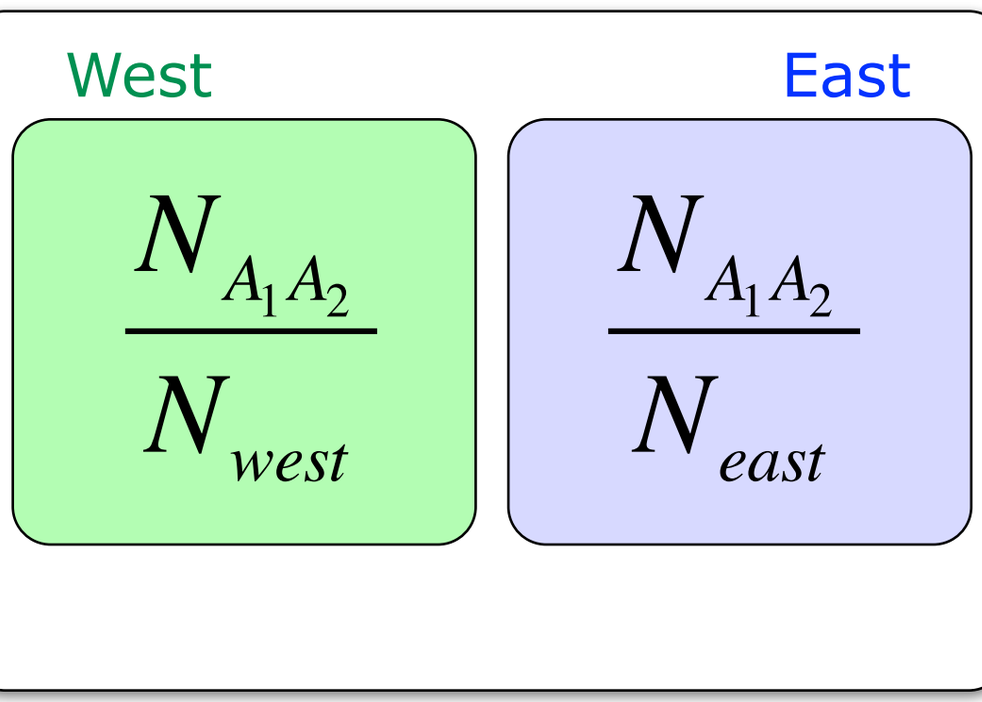


Expected heterozygosity without subdivision

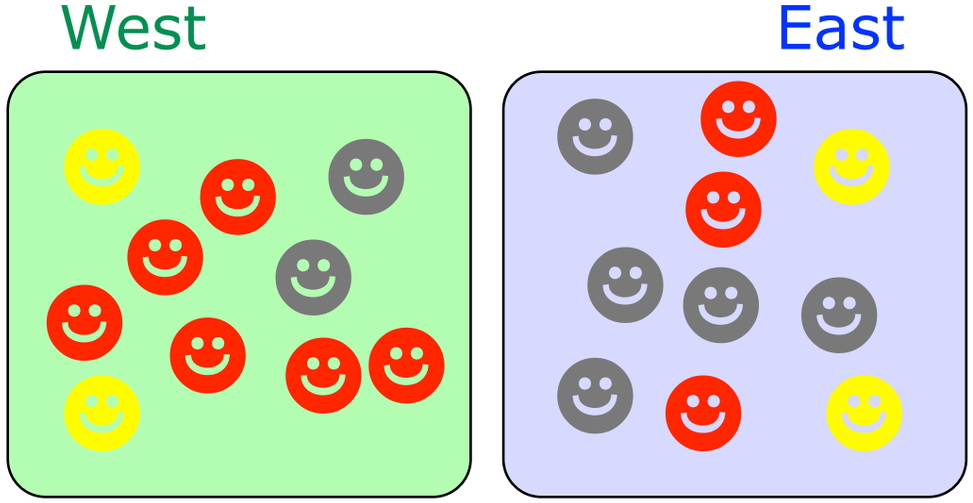
$$H_T = 2pq_{total}$$

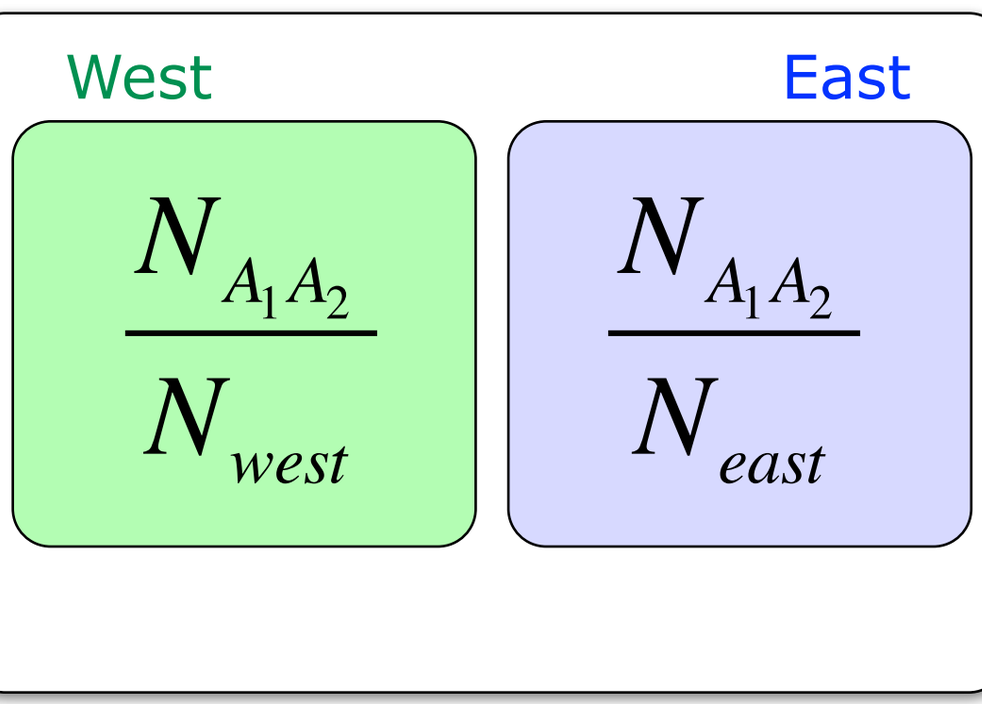
Average observed heterozygosity in subpopulations

$$H_I = \frac{(\#A_1A_{2west}/N_{west}) + (\#A_1A_{2east}/N_{east})}{2}$$



Population - Subpopulations





Expected heterozygosity without subdivision

$$H_T = 2pq_{total}$$

Average observed heterozygosity in subpopulations

$$H_I = \frac{(\#A_1A_{2west}/N_{west}) + (\#A_1A_{2east}/N_{east})}{2}$$

Average expected heterozygosity assuming HWE

$$H_S = \frac{(2pq_{west}) + (2pq_{east})}{2}$$

The F statistics can be calculated using the relationship between heterozygosity and inbreeding. This allows F statistics to be determined from genetic markers (Nei 1977; de Jong et al. 1994).

...for individuals (I) within sub-population (S):

$$F_{IS} = 1-(H_I/H_S)$$

F_{IS}: That proportion of the total inbreeding within a population due to inbreeding within sub-populations.

The F statistics can be calculated using the relationship between heterozygosity and inbreeding. This allows F statistics to be determined from genetic markers (Nei 1977; de Jong et al. 1994).

...for individuals (I) within sub-population (S):

$$F_{IS} = 1-(H_I/H_S)$$

F_{IS}: That proportion of the total inbreeding within a population due to inbreeding within sub-populations.

...for sub-population (S) relative to metapopulation (T):

$$F_{ST} = 1-(H_S/H_T)$$

F_{ST}: That proportion of the total inbreeding in a population due to differentiation among sub-populations.

The F statistics can be calculated using the relationship between heterozygosity and inbreeding. This allows F statistics to be determined from genetic markers (Nei 1977; de Jong et al. 1994).

...for individuals (I) within sub-population (S):

$$F_{IS} = 1-(H_I/H_S)$$

F_{IS}: That proportion of the total inbreeding within a population due to inbreeding within sub-populations.

...for sub-population (S) relative to metapopulation (T):

$$F_{ST} = 1-(H_S/H_T)$$

F_{ST}: That proportion of the total inbreeding in a population due to differentiation among sub-populations.

...for individuals (I) within the metapopulation (T):

$$F_{IT} = 1-(H_I/H_T)$$

F_{IT}: The total inbreeding in a population due to both inbreeding within sub-populations, and differentiation among sub-populations.

$H_I=\sum H_{Ii}/n$: observed heterozygosity within subpopulations , $H_S=\sum 2p_iq_i/n$: is the expected herterozygosity with random mating, and $H_T=2\bar{p}\bar{q}$: is the expected heterozygosity of individuals based on allele frequencies averaged with random mating.

Sewall Wright (1969) used inbreeding coefficient to describe the distribution of genetic diversity within and among population fragments; he partitioned total inbreeding of individuals (I) relative to the total (T) population (F_{IT}) into that inbreeding of individuals relative to their sub-population (S), F_{IS} and that dues to differentiation among sub-populations, relative to the total population F_{ST}.

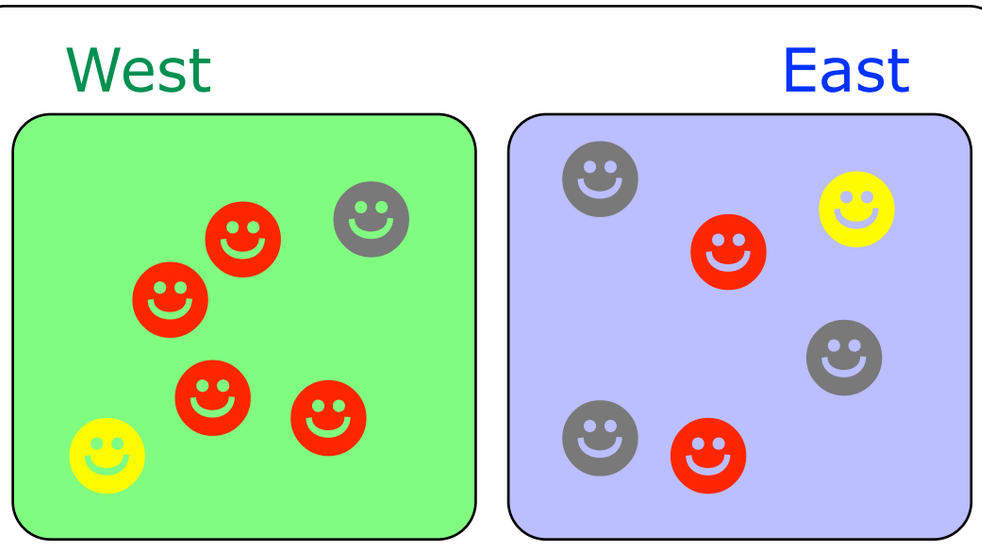
$$F_{IT} = F_{IS}+F_{ST}-(F_{IS}F_{ST})$$

$$> F_{IT} = F_{ST}+F_{IS}-F_{IS}F_{ST}$$
$$> 1-F_{IT} = 1-F_{ST}-F_{IS}+F_{IS}F_{ST}$$
$$> (1-F_{IT}) = (1-F_{IS})(1-F_{ST})$$

In words, the total inbreeding is the probability of identity by descent within fragments (F_{IS}) plus the probability of identity by descent to subdivision (F_{ST}) minus the probability of identity by descent due to both.

$$F_{ST} = (F_{IT}-F_{IS})/(1-F_{IS})$$

$$> F_{IT} = F_{ST}+F_{IS}-F_{IS}F_{ST}$$
$$> F_{IT} = F_{IS}+F_{ST}(1-F_{IS})$$
$$> F_{ST} = (F_{IT}-F_{IS})/(1-F_{IS})$$



$2pq_{total}$

Observed average heterozygosity in subpopulations

$$H_I = \frac{(\#Aa_{west}/N_{west}) + (\#Aa_{east}/N_{east})}{2}$$

Expected average heterozygosity assuming HWE

$$H_S = \frac{(2pq_{west}) + (2pq_{east})}{2}$$

Heterozygosity expected without subdivision

$H_T = 2pq_{total}$

---

among sub-populations

$F_{IS} = 1-(H_I/H_S)$

within populations

$F_{ST} = 1-(H_S/H_T)$

among populations/within sub-population

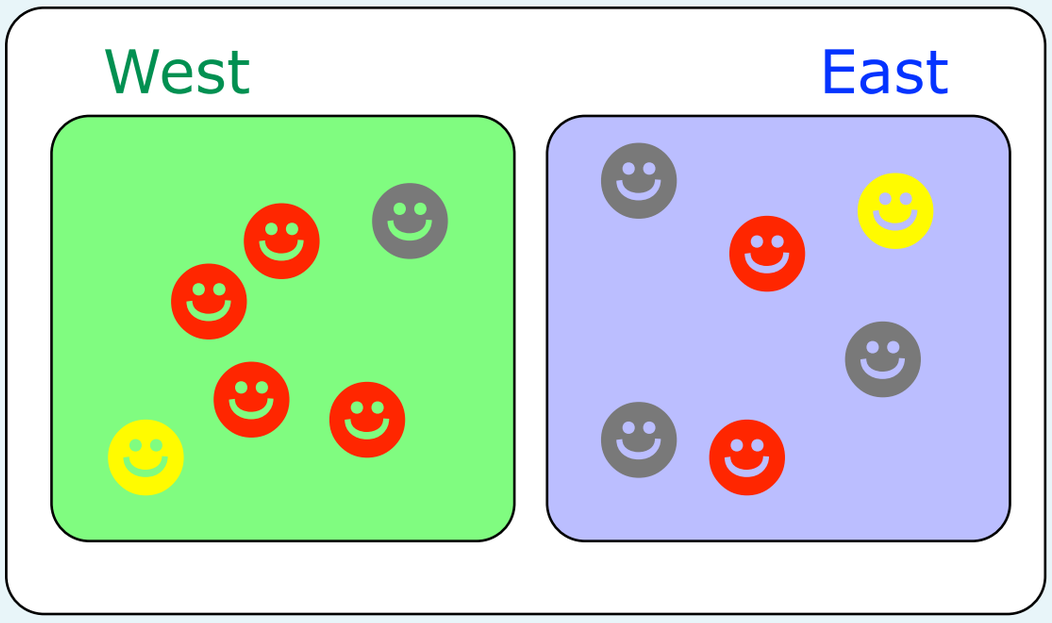
$F_{IT} = 1-(H_I/H_T)$

$F_{IT} = F_{IS}+F_{ST}-(F_{IS}F_{ST})$

$F_{ST} = (F_{IT}-F_{IS})/(1-F_{IS})$

Fixation Index

$$F_{ST} = (F_{IT}-F_{IS})/(1-F_{IS})$$



F_{ST} [0.05-0.15]
sub-populations are very similar

F_{ST} [0.15-0.25]
sub-populations are similar

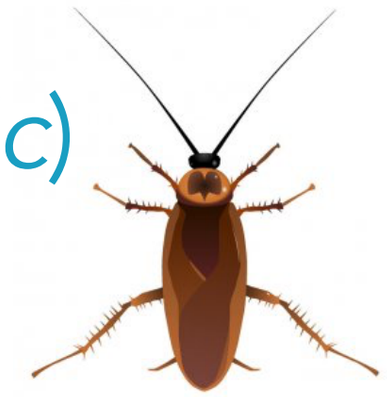
F_{ST} [>0.25]
sub-populations are distinct

Bed bugs (*Cimex lectulariu*)

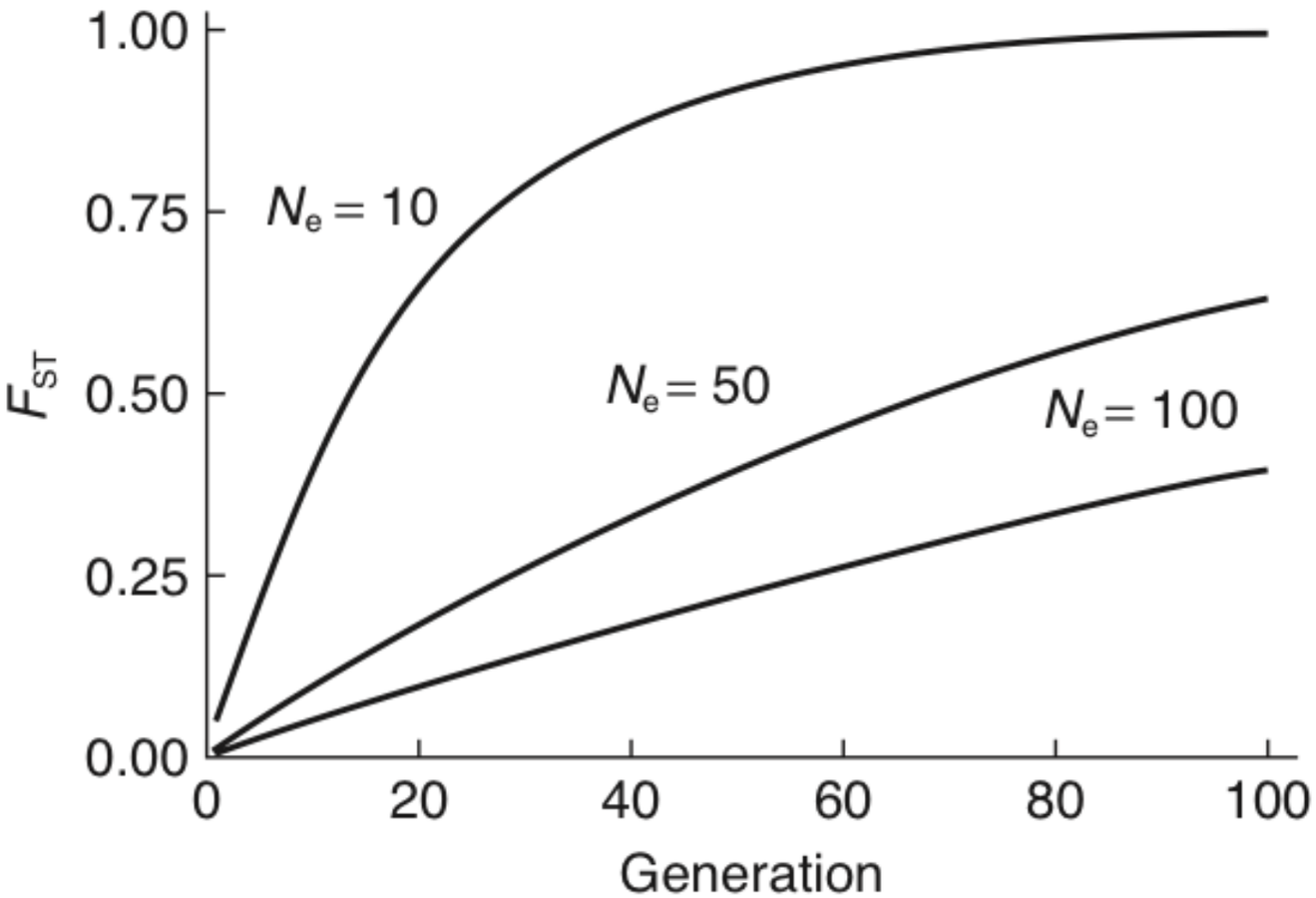
$F_{ST} = 0.68$

German cockroach (*Blattella germanic*)

$F_{ST} = 0.099$



Saenz et al. (2012) Genetic Analysis of Bed Bug Populations Reveals Small Propagule Size Within Individual Infestations but High Genetic Diversity Across Infestations From the Eastern United States.



$$F_{ST} = 1 - \left(1 - \frac{1}{2N_e}\right)^t$$

Expected increase in F_{ST} over time (generations) among completely isolated populations of different population sizes.

Allendorf, Luikart, and Aitken (2013)


MORE
THINGS
CONSIDERED
2

The deficit of heterozygotes relative to HW proportions caused by the subdivision of a population into separate demes is often referred to as the "Wahlund effect".

$$Var(q) = \frac{1}{S}\sum(q_i - \bar{q})^2$$

When Var(q)=0, all subpopulations have the same allele frequencies and the population is in H-W proportion.

Genotype	H-W	Wahlund
A_1A_1	p^2	p^2 + Var(q)
A_1A_2	2pd	2pq - 2Var(q)
A_2A_2	q2	q^2 + Var(q)

The Wahlund effect refers to reduction of heterozygosity in a population caused by **subpopulation** structure. Namely, if two or more subpopulations have different allele frequencies then the overall heterozygosity is reduced, even if the subpopulations themselves are in a Hardy-Weinberg equilibrium. The underlying causes of this population subdivision could be **geographic barriers to gene flow followed by genetic drift** in the subpopulations.

The Wahlund effect has a number of important consequences:

• We have to know about the **structure of a population** when applying the Hardy-Weinberg principle to it, otherwise there may seem to be more homozygotes than expected from the Hardy-Weinberg principle. We might then suspect that selection, or some other factor, was favouring homozygotes. In fact both sub-populations are in perfectly good Hardy-Weinberg equilibrium and the deviation is due to the unwitting pooling of the separate populations.

• A second consequence of the Wahlund effect is that when a number of previously subdivided populations merge together, the frequency of homozygotes will decrease. In humans, this can lead to a decrease in the incidence of rare recessive genetic diseases when a previously isolated population comes into contact with a larger population. The recessive disease is only expressed in the homozygous condition, and when the two populations start to interbreed, the frequency of those homozygotes goes down.

Wahlund, S. (1928). Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas 11:65–106.

9-1-2009

"Genetics in geographically structured populations: defining, estimating and interpreting FST."

Kent E. Holsinger
University of Connecticut - Storrs, kent.holsinger@uconn.edu

Bruce S. Weir
University of Washington - Seattle Campus, bsweir@u.washington.edu

http://www.evolution.unibas.ch/teaching/evol_genetics/3_Population_Genetics/reading/Holsinger_and_Weir_2009.pdf

deme (*dēm*)

Greek dēmos, *people, land*; see d- in Indo-European roots.

In biology, a deme is a term for a local population of organisms of one species that actively interbreed with one another and share a distinct gene pool. When demes are isolated for a very long time they can become distinct subspecies or species. The term deme is mainly used in evolutionary biology and is often used as a synonym for population.

In evolutionary computation a "deme" often refers to any **isolated subpopulation subjected to selection as a unit rather than as individuals**.

A deme in biological evolution is conceptually related to a meme in cultural evolution, a term suggested by Richard Dawkins' 1976 book The Selfish Gene.