Chapter 4 – Population Structure & Gene Flow

Genetic Populations

~~Direct Measures of Gene Flow~~

Fixation indices (F_{ST})

Population Subdivision

Models of Population Structure

Population Structure and the Coalescent

Population Structure

❖ most natural populations exist across a landscape (or seascape) that is more or less divided into areas of suitable habitat

❖ to the extent that populations are isolated, they will become genetically differentiated due to genetic drift, selection, and eventually mutation

❖ genetic differentiation among populations is relevant to conservation biology as well as fundamental questions about how adaptive evolution proceeds

Definitions

- ❖ panmixia
- ❖ population structure
- ❖ subpopulation
- ❖ gene flow
- ❖ isolation by distance
- ❖ vicariance (vicariant event)

Structure Results in Inbreeding

- ❖ given finite population size, autozygosity gradually increases because the members of a population share common ancestors
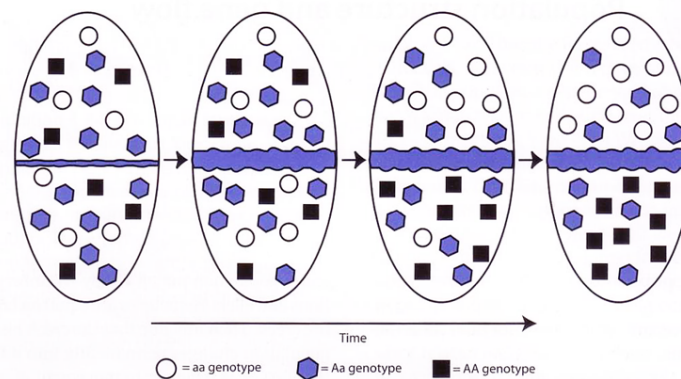 - ✧ even when there is no **close** inbreeding

Figure 4.1 An example of population structure and allele-frequency divergence produced by limited gene flow. The total population (large ovals) is initially in panmixia and has Hardy–Weinberg expected genotype frequencies. Then the stream that runs through the population grows into a large river, restricting gene flow between the two sides of the total population. Over time allele frequencies diverge in the two subpopulations through genetic drift. In this example, you can imagine that the two subpopulations drift toward fixation for different alleles but neither reaches fixation due to an occasional individual that is able to cross the river and mate. Note that there is random mating (panmixia) within each subpopulation so that Hardy–Weinberg expected genotype frequencies are maintained within subpopulations. However, after the initial time period genotype frequencies in the total population do not meet Hardy–Weinberg expectations.

"Identical by Descent"

❖ what is the probability that two randomly sampled alleles are identical by descent (i.e., "replicas of a gene present in a previous generation")?

  ◇ Wright's "fixation index" F

❖ at the start of the process (time 0), "declare" all alleles in the population to be unique or unrelated, $F_t = 0$ at $t = 0$

❖ in the next generation, the probability of two randomly sampled alleles being copies of the same allele from a single parent = $1/(2N)$, so...

"Identical by Descent"

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F_{t-1}$$

= probability that alleles are copies of the same gene from the immediately preceding generation *plus* the probability that the alleles are copies of the same gene from an earlier generation

or

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

assuming $F_0 = 0$

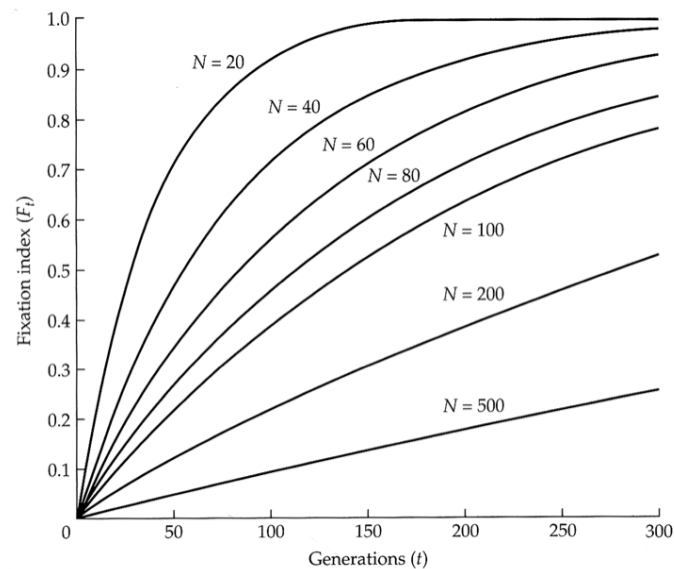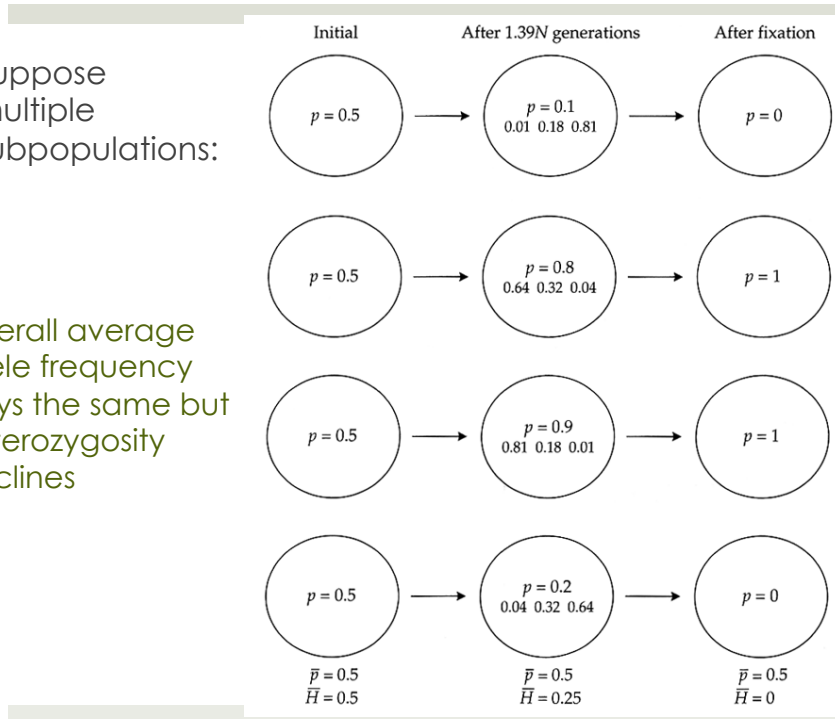compare to: mean time to fixation for new mutant = ~4N

FIGURE 3.11 Increase of F_t in ideal populations as a function of time and effective population size N.
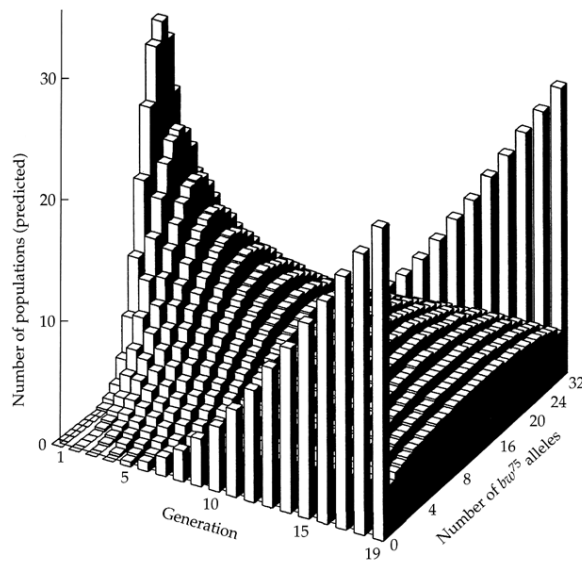
❖ Suppose multiple subpopulations:

Overall average allele frequency stays the same but heterozygosity declines

Predicted distributions of allele frequencies in replicate populations of *N* = 16

same process as in this figure...

Population Structure

❖ F_t for a single population is essentially the same thing as F_{ST}
  ✧ a measure of genetic differentiation among populations
❖ due to autozygosity, structured populations have lower heterozygosity than expected if all were combined into a single random breeding population
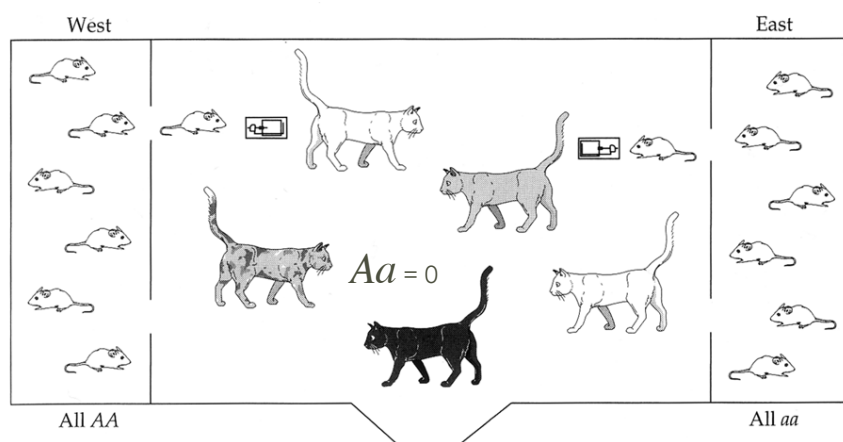
West   East

$Aa = 0$

All AA All aa

FIGURE 6.12 An extreme example of the general principle that a difference in allele frequency among subpopulations results in a deficiency of heterozygotes. The floor plan is that of a hypothetical barn. The mouse subpopulations in the east and west enclaves are completely isolated because of the cats in the middle. The west subpopulation is fixed for the A allele and the east subpopulation for the a allele. Trapping mice at random in the area patrolled by the cats would yield an overall allele frequency of $\frac{1}{2}$, but no heterozygous genotypes.

F_{ST}

❖ measures the deficiency of heterozygotes in the total population relative to the expected level (assuming HWE)

❖ in the simplest case, one can calculate F_{ST} for a comparison of two populations…

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Two population, two allele F_{ST}

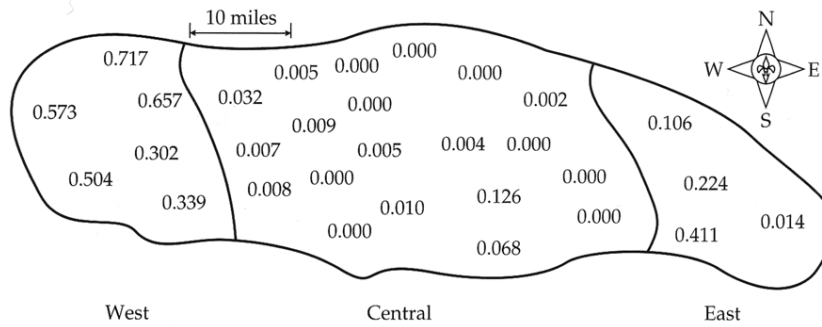| Frequency of "A"				
Population 1	**Population 2**	H_T	H_S	F_{ST}
0.5	0.5	0.5	0.5	0
0.4	0.6	0.5	0.48	0.04
0.3	0.7	0.5	0.42	0.16
0.2	0.8	0.5	0.32	0.36
0.1	0.9	0.5	0.18	0.64
0.0	1.0	0.5	0	1
0.3	0.35	0.43875	0.4375	0.002849
0.65	0.95	0.32	0.275	0.140625

FIGURE 6.13 Estimated frequency of a recessive allele for blue flower color in populations of *Linanthus parryae* in an area of approximately 900 square miles in the Mohave desert. Each allele frequency is based on an examination of approximately 4000 plants over an area of about 30 square miles. (After Wright 1943a.)

hierarchical population structure
subpopulations nested within regions

TABLE 6.3   Hierarchical Structure of *Linanthus parryae*

Region	Subpopulations Allele frequency	Subpopulations Heterozygosity	Regions Average allele frequency	Regions Heterozygosity	Total Average allele frequency	Total Heterozygosity
W	0.573	0.4893				
	0.717	0.4058				
	0.504	0.5000				
	0.657	0.4507				
	0.302	0.4216				
	0.339	0.4482				
C	9 × 0.000	0.0000				
	0.032	0.0620				
	0.007	0.0139				
	0.008	0.0159				
	0.005	0.0100				
	0.009	0.0178				
	0.005	0.0100				
	0.010	0.0198				
	0.068	0.1268				
	0.002	0.0040				
	0.004	0.0080				
	0.126	0.2202	0.0138	0.0272		
E	0.106	0.1895				
	0.224	0.3476				
	0.411	0.4842				
	0.014	0.0276	0.1888	0.3062	0.1374	0.2371
Average heterozygosity		$H_S = 0.1424$		$H_R = 0.1589$		$H_T = 0.2371$

Source: Data from Wright 1943a.

e.g., 328 blue flowers, 672 white flowers

$$b^2 = 0.328, \quad b = \sqrt{0.328} = 0.573$$

H_R is a weighted average of the regional heterozygosities

$$H_R = \left(6 \times 0.4995 + 20 \times 0.0272 + 4 \times 0.3062\right)/30$$

$$H_R = 0.1589$$

TABLE 6.3 Hierarchical Structure of *Linanthus parryae*

Region	Subpopulations		Regions		Total	
	Allele frequency	Heterozygosity	Average allele frequency	Heterozygosity	Average allele frequency	Heterozygosity
W	0.573	0.4893				
	0.717	0.4058				
	0.504	0.5000				
	0.657	0.4507				
	0.302	0.4216				
	0.339	0.4482	0.5153	0.4995		
C	9 × 0.000	0.0000				
	0.032	0.0620				
	0.007	0.0139				
	0.008	0.0159				
	0.005	0.0100				
	0.009	0.0178				
	0.005	0.0100				
	0.010	0.0198				
	0.068	0.1268				
	0.002	0.0040				
	0.004	0.0080				
	0.126	0.2202	0.0138	0.0272		
E	0.106	0.1895				
	0.224	0.3476				
	0.411	0.4842				
	0.014	0.0276	0.1888	0.3062	0.1374	0.2371
Average heterozygosity		$H_S = 0.1424$		$H_R = 0.1589$		$H_T = 0.2371$

Source: Data from Wright 1943a.

H_S - average heterozygosity in subpopulations assuming HWE within each

H_R - average heterozygosity in regions assuming HWE among all subpopulations within each region

H_T - "expected" heterozygosity assuming HWE across the total population

Hierarchical F-statistics

❖ measure the reduction in heterozygosity in subpopulations (or regions) due to differences in allele frequencies

$$F_{SR} = \frac{H_R - H_S}{H_R} = 0.1036$$

$$F_{RT} = \frac{H_T - H_R}{H_T} = 0.3299$$

$H_S = 0.1424$

$H_R = 0.1589$

$H_T = 0.2371$

$$F_{ST} = \frac{H_T - H_S}{H_T} = 0.3993$$

Hierarchical F-statistics

❖ relationship among F-statistics

$$F_{ST} = 1 - (1 - F_{SR})(1 - F_{RT})$$
$$= 1 - (0.8964 \times 0.6701)$$
$$= 0.3993$$

$H_S = 0.1424$
$H_R = 0.1589$
$H_T = 0.2371$

$F_{SR} = 0.1036$
$F_{RT} = 0.3299$
$F_{ST} = 0.3993$

Hierarchical F-statistics

❖ relationship among F-statistics

$$F_{ST} = 1 - (1 - F_{SR})(1 - F_{RT})$$

❖ hierarchical F-statistics partition the variance in allele frequencies into components corresponding to different levels of the hierarchy