

Population Subdivision

Anthro/Biol 5221

Nov. 21, 2008

Population Subdivision

- Many species have physical or other barriers to migration and gene flow
- This prevents random mating and leads to departures from Hardy-Weinberg (HW) frequencies for the entire species
- F_{ST} is an index that measures the departure from HW equilibrium due to subdivision
- A species with limited migration appears to be inbred. This is called the **Wahlund Effect**
- Wright's **Island Model** describes the effect of migration on gene frequencies
- Combining these models help us understand the genetic structure of a species

Fixation Index (F)

- Nonrandom mating causes departures from Hardy-Weinberg

Genotype	AA	Aa	aa
Frequency	$p^2(1-F)+pF$	$2pq(1-F)$	$q^2(1-F)+qF$

- F describes the deviation from HW genotype frequencies
 - If $0 < F \leq 1$, there is an excess of homozygotes (compared to HW)
 - If $F < 0$, there is an excess of heterozygotes
- This should look familiar (F_I , the inbreeding coefficient)
- A special case of F , F_{ST} , is the **Fixation index in the Subpopulation relative to the Total population**.
- F_{ST} acts on genotype frequency in the same way as F_I by reducing heterozygosity.

Wahlund Effect

Genotype:	p	AA	Aa	aa
Frequency in subpopulation 1:	1/4	1/16	3/8	9/16
Frequency in subpopulation 2:	3/4	9/16	3/8	1/16
Frequency in species*:	1/2	5/16	3/8	5/16
Hardy-Weinberg frequencies:	1/2	1/4	1/2	1/4
Deviation from Hardy-Weinberg:	N/A	+1/16	-1/8	+1/16

(* for now, we are assuming that each subpopulation contributes equally)

Notice that Aa heterozygotes are *lower* than expected (similar to inbreeding)
 This is because of barriers to gene flow, physical or otherwise
 Looking at the Aa heterozygotes, let us calculate F_{ST} :

$$2pq(1-F_{ST}) = 2(1/2)(1/2)(1-F_{ST}) = 3/8, \text{ so } F_{ST} = 1/4$$

This decrease in heterozygosity is called the *Wahlund Effect*, and is a function of $\text{Var}[p]$. Let's look at that now...

Generalized form of F_{ST}

(see Gillespie, p. 132-133)

Genotype	AA	Aa	aa
In i^{th} subpop.:	p_i^2	$2p_iq_i$	q_i^2
In species: (c_i =relative contribution of i^{th} subpop to species)	$\sum c_i p_i^2$	$\sum c_i 2p_iq_i$	$\sum c_i q_i^2$
Frequency:	$p^2(1-F_{ST})+pF_{ST}$	$2pq(1-F_{ST})$	$q^2(1-F_{ST})+qF_{ST}$

Again, looking at the heterozygotes: $2pq(1-F_{ST}) = \sum c_i 2p_iq_i$

When we solve for F_{ST} , we get: $F_{ST} = (2pq - \sum c_i 2p_iq_i)/2pq$

Using: $2pq = 1 - p^2 - q^2$, and $\sum c_i 2p_iq_i = 1 - \sum c_i p_i^2 - \sum c_i q_i^2$,

The equation can be rewritten as: $F_{ST} = (\sum c_i (p_i^2 + q_i^2) - p^2 - q^2)/(1 - p^2 - q^2)$

Or: $F_{ST} = (G_S - G_T)/(1 - G_T)$ where $G_S = \sum c_i (p_i^2 + q_i^2)$ and $G_T = p^2 + q^2$

Notice that the numerator in the above equation looks a lot like $E[x^2] - (E[x])^2$

Indeed, this equation is often written as: $F_{ST} = (2\text{Var}[p_i])/(1 - G_T)$

What happens to F_{ST} when the variance of p gets small?

Let's do an example:

$$F_{ST} = (G_S - G_T)/(1 - G_T) \text{ where } G_S = \sum c_i (p_i^2 + q_i^2) \text{ and } G_T = p^2 + q^2$$

Subpop. 1	$N_1=800$	$p_1=1/2$
Subpop. 2	$N_2=100$	$p_2=1/4$
Subpop. 3	$N_3=100$	$p_3=3/4$

Now using variance:

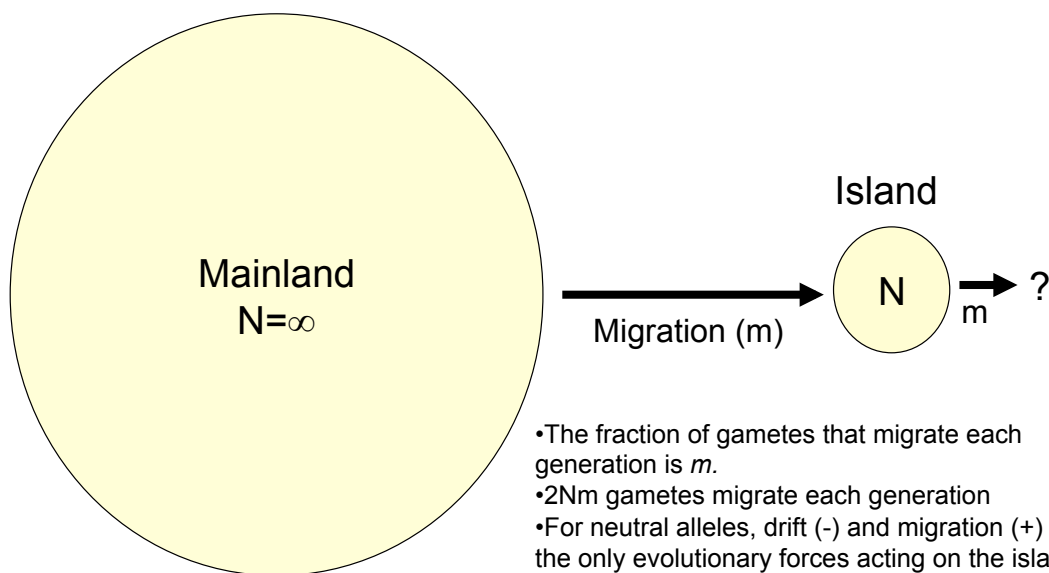
$$F_{ST} = (2\text{Var}\{p_i\})/(1-G_T) \text{ where } G_T = p^2 + q^2$$

Subpop. 1	$N_1=800$	$p_1=1/2$
Subpop. 2	$N_2=100$	$p_2=1/4$
Subpop. 3	$N_3=100$	$p_3=3/4$

...and that is how you calculate the Wahlund effect.

Island Model

(See Gillespie, p. 134)



- The fraction of gametes that migrate each generation is m .
- $2Nm$ gametes migrate each generation
- For neutral alleles, drift (-) and migration (+) are the only evolutionary forces acting on the island population.
- Eventually, equilibrium is reached, as a function of N and m (much like N and u before in calculating the mutation-drift balance).

Migration and Drift

Recall that under mutation-drift equilibrium (where G =homozygosity):

$$G=1/(1+4N\mu)$$

Equilibrium under the island migration-drift model is:

$$G=1/(1+4Nm)$$

Where m is the fraction of a population migrating per generation.

If we assume $N \rightarrow \infty$ (and $G=0$), change in heterozygosity on the island due to migration is:

$$\Delta_m H = 2m(1-H)$$

And change due to drift is still:

$$\Delta_N H = -H/2N$$

At equilibrium, $\Delta_m H + \Delta_N H = 0$, giving G is shown above.

What happens to homozygosity (G) when $4Nm \gg 1$?

How about when $m < 1/4N$?

Combining F_{ST} and the Island Model

Recall, that:

$$F_{ST} = (G_S - G_T) / (1 - G_T) \text{ where } G_S = \sum c_i (p_i^2 + q_i^2) \text{ and } G_T = p^2 + q^2$$

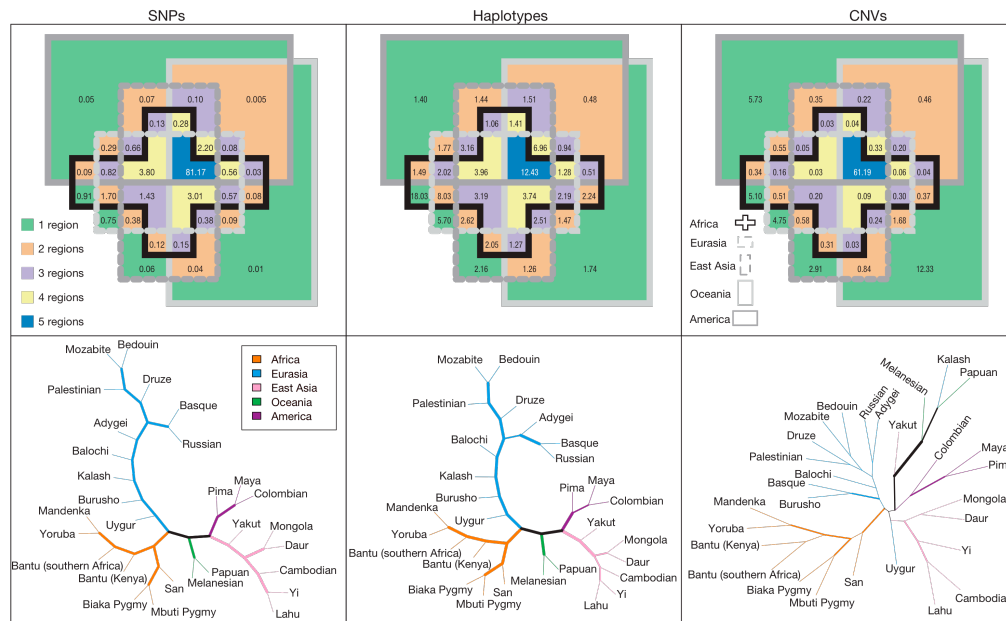
Notice that G_T is the probability that two alleles drawn at random from the *entire species* are identical, and G_S is the probability that two alleles drawn at random from a *randomly chosen subdivision* are identical.

If we assume that G_T is 0 ($N=\infty$), and G_S is about the same as the homozygosity of the island ($1/(1+4Nm)$), then:

$$F_{ST} = 1/(1+4Nm)$$

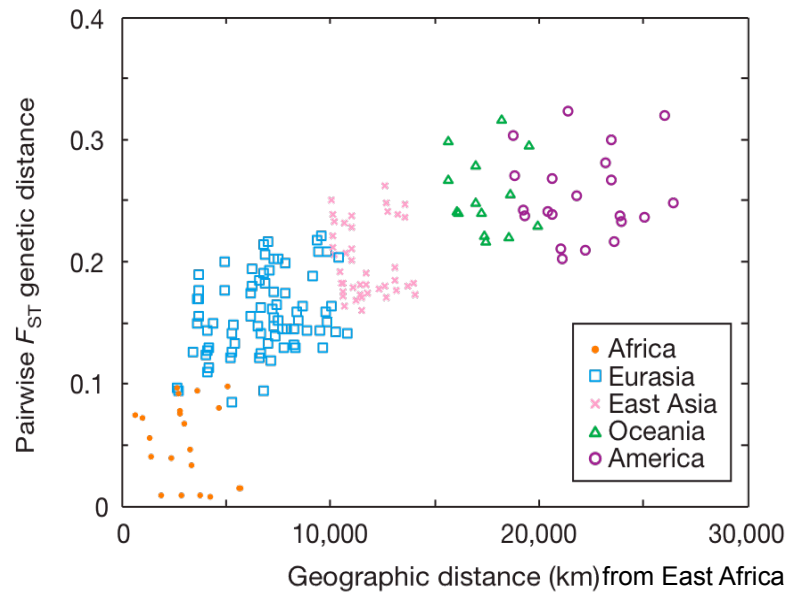
Therefore, if we know an island's population (N) and its F_{ST} , we can calculate its migration rate. However, this assumes that variation at the locus is selectively neutral, that the population is at equilibrium, and that migration pattern follows the island model (a lot of assumptions!). Nevertheless, F_{ST} can still be (and is) used to study the genetic structure of a species.

“Genotype, haplotype and copy-number variation in worldwide human populations” (Jakobsson, *Nature* 2008)



Top: Venn diagram of 29 populations used in this study grouped into 5 regions
Bottom: Unrooted NJ tree of population subgroups

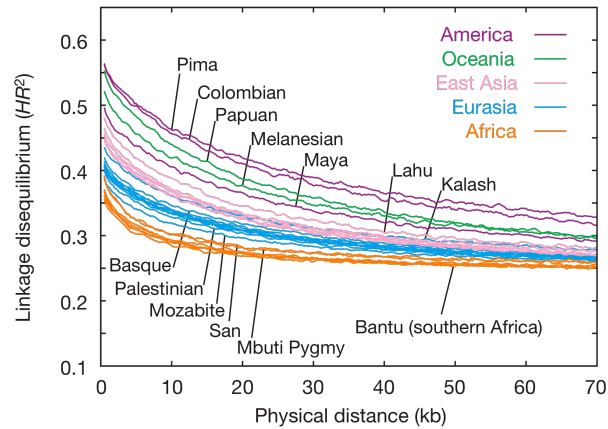
F_{ST} in Humans (a teaser for Alan’s lecture on Monday)



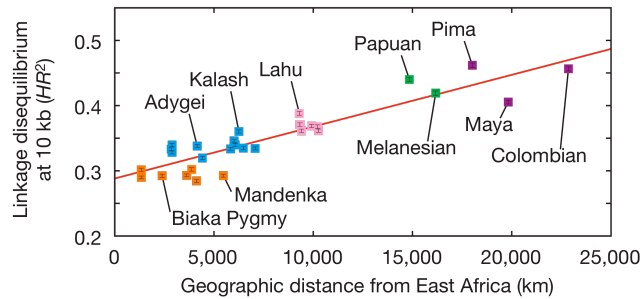
In the case where there is random mating within each segregated population, F_{ST} is a crude measure of the geographic structure of a species.
What other factors (besides migration) could explain these data?

Linkage Disequilibrium and Geography

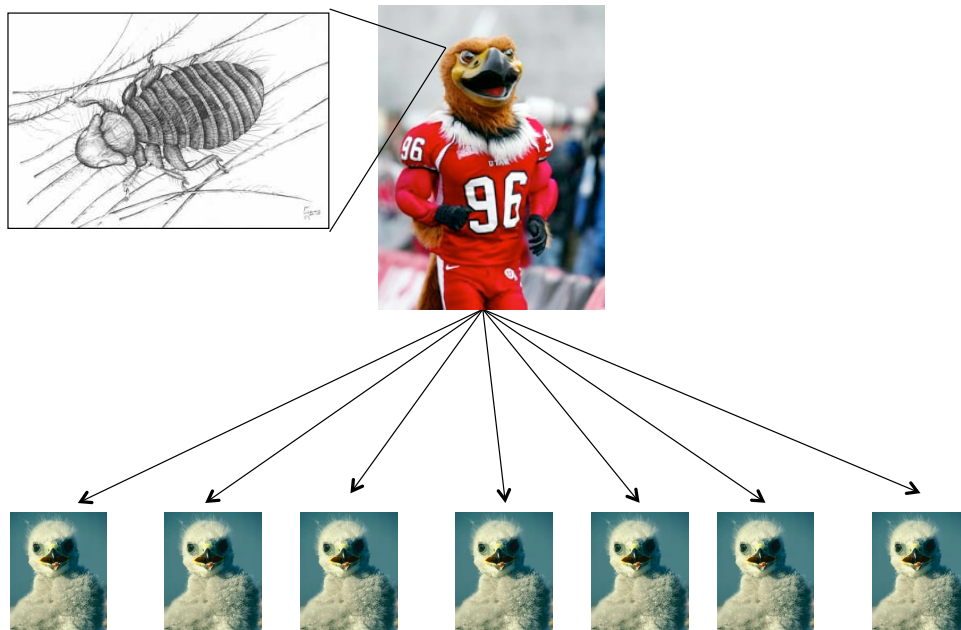
- Top graph shows LD declining as a function of physical distance.
- What can explain this decline?
- Why are the curves at the top decline more gradually?



- Bottom graph is a cross-section of LD at 10 kb. Notice how LD correlates with geographic distance from East Africa, the widely assumed birthplace of modern humans.

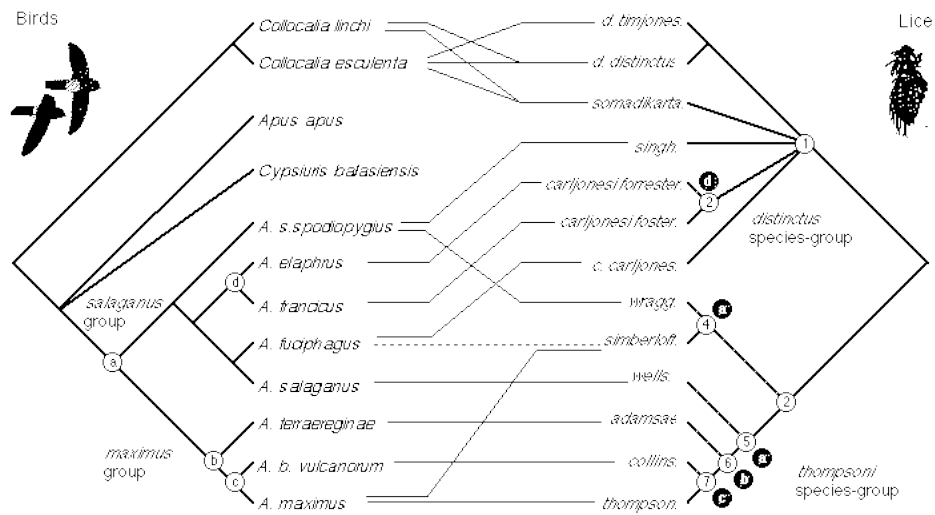


Population Subdivision and Speciation



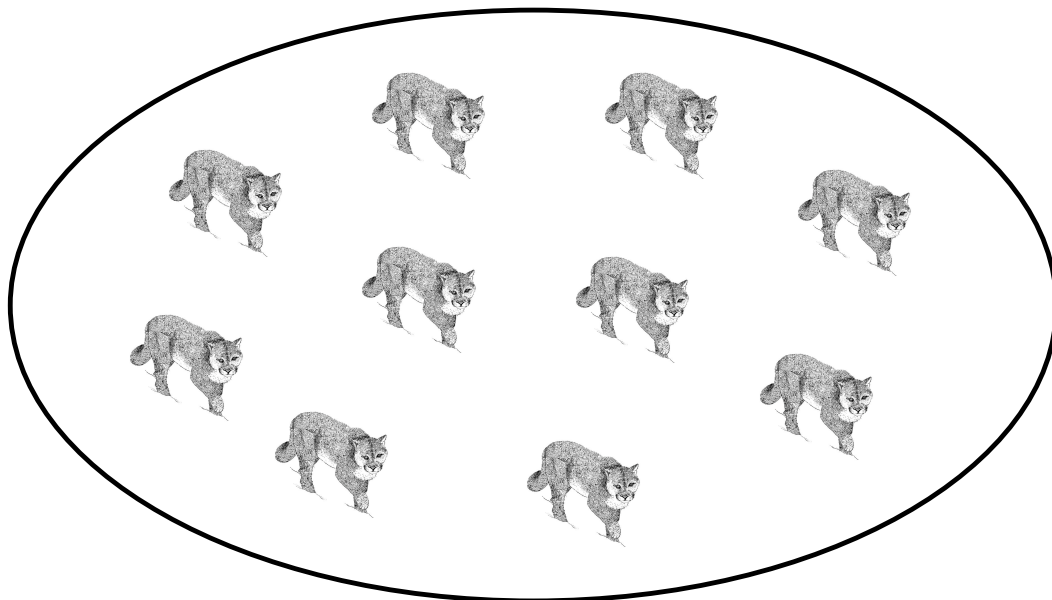
In this case, migration occurs once so $m \sim 0$. What is a result?

Congruent Phylogenies



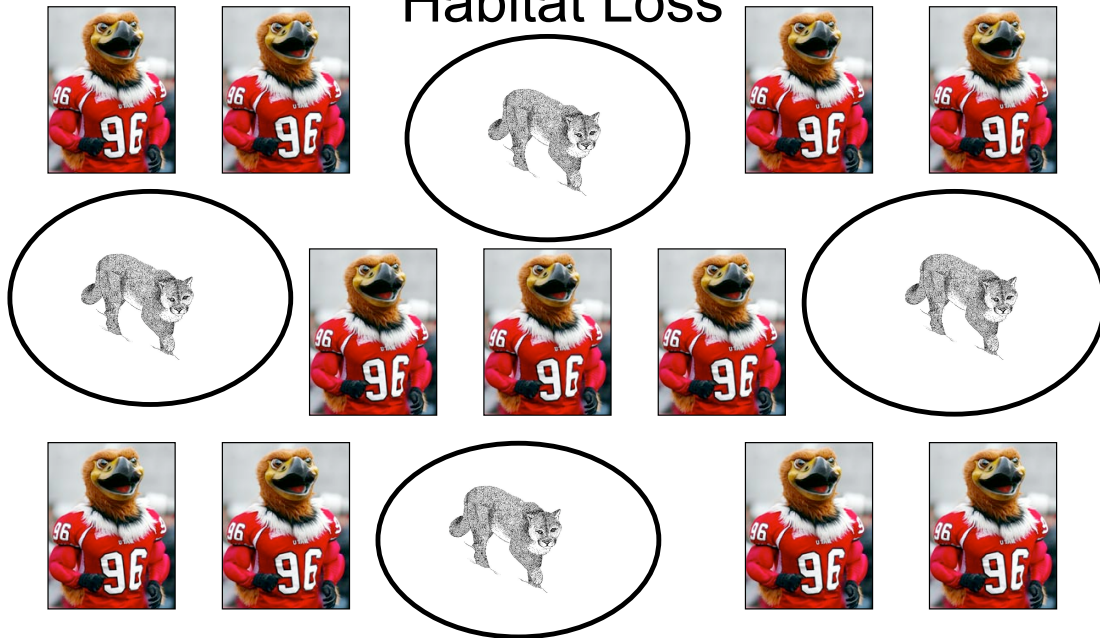
The bird host acts like an island for the lice. As the host speciates, so do the lice. How can our understanding of population subdivision help to explain this phenomenon?

Practical Application: Habitat Loss and Species Fitness



Initially, the predators are free to move throughout the entire population. This results in random mating and high heterozygosity.

Practical Application: Habitat Loss



One consequence of habitat loss is fragmentation of populations. Based on your understanding of homozygosity and population subdivision, what impact will this have on species fitness? What sorts of land use policies can we implement to mitigate this effect?



GO U'TES!

