MIGRATION

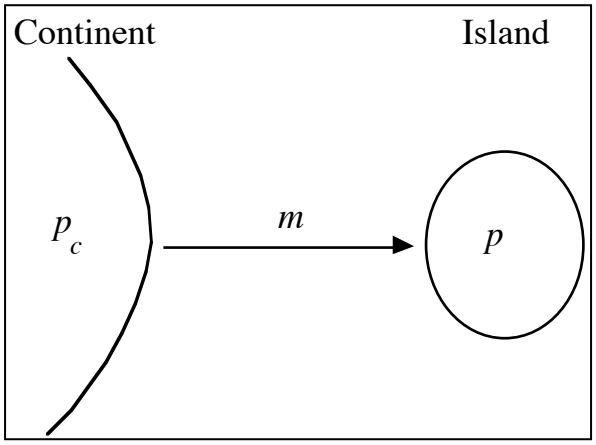
ROLES OF MIGRATION IN EVOLUTION

READING: Hedrick pp. 403-436.

- Introduces novel genetic variation into populations.
- Tends to homogenize gene frequencies in different populations.
- Sets the spatial scale for evolution.
- Opposes local adaptation.
- Migration with an evolutionary impact: Gene Flow
  - Migration introduces *individuals* and *genotypes* ("dispersal").
  - Migrants have no effect on evolution unless their genes are incorporated into a population.

- **A One-Island Model**
  - The simplest model of migration.
  - Two alleles A and a. Let p = frequency of A on island.
  - A fraction m of the island gene pool emigrates from the continent where the frequency of A is p_c.

    $\Rightarrow$ A fraction $(1-m)$ of alleles on the island originated on the island.



  - The continent is too vast to be influenced by migration from the island $\Rightarrow$ p_c is constant.
  - Then the frequency of A on the island changes according to $p' = (1-m)p + mp_c$.

– At equilibrium, set $p' = p$.

• Solving for p gives $\hat{p} = p_c$

– Rate of approach to equilibrium:

• Rewrite evolutionary equation as

$$p' - \hat{p} = p' - p_c = (1-m)p + mp_c - p_c = (1-m)(p - p_c)$$
$$= (1-m)(p - \hat{p})$$

– <u>Conclusions</u>

(1) At equilibrium, both populations have the same allele frequencies.

(2) Rate of approach to equilibrium ($\hat{p} = p_c$) is determined by the migration rate m.

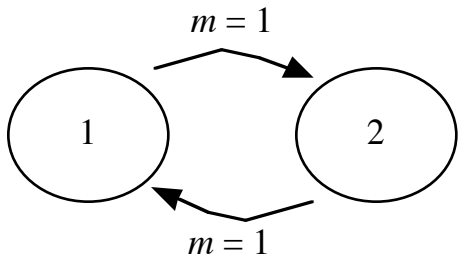
• General Models of Migration

– Same conclusions as one-island model hold.

– Exceptions, however, do exist

• For example, consider two populations with different allele frequencies that switch locations each generation.

• The populations will obviously never homogenize (because there's no real exchange of genes).



– Remark: Have implicitly assumed gene frequencies differ in different locations.

– How could this be?
– "History."
– Genetic drift.
– Selection favors different alleles in different locations.

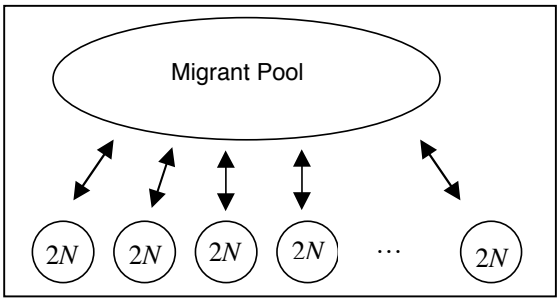
MIGRATION AND DRIFT

- Migration introduces novel genetic variation into local populations.
- Drift removes local genetic variation.

Which for dominates?

One answer…

- **Wright's "Island Model"**
- Consider a large number of "islands" each with a population of size N ($2N$ alleles per locus)
- Each generation, every island exchanges a fraction m of its gametes with a ∞-sized "migrant pool" to which all islands contribute gamates.



- Assume infinite-isoalleles model.
- Let f_t = Pr(pair or randomly drawn gametes on a typical island are IBD in generation t) = average within-island homozygosity
- By the same logic used when studying mutation-drift balance:

$$f_{t+1} = (1-m)^2\left[\frac{1}{2N} + \left(1-\frac{1}{2N}\right)f_t\right]$$

- At equilibrium, $f_{t+1} = f_t = \hat{f} \approx \dfrac{1}{1+4Nm}$

  – expression resembles that describing diversity maintained by mutation & drift, with $\theta = 4Nu$ replaced by $4Nm$.

- If $4Nm < 1$: Local homozygosity is substantial
  – drift dominates migration
- If $4Nm > 1$: Local diversity (heterozygosity) is substantial
  – migration dominates drift

Note 1

– $4Nm > 1$ same as $2Nm > 1/2$

⇒ Migration dominates drift if at least *one migrant gamete is exchanged every other generation!*

– Conclusion is independent of m, the rate of gene flow. (Why?)

Note 2

– Recall from discussion of F statistics: $\overline{H}_S = \underset{i}{\text{Avg}}\left(H_{S,i}\right) \approx 1 - \hat{f}$, since $\hat{f}$ is the average local homozygosity and there is no additional inbreeding

– Also, H_T = 1 – Pr(pair of randomly chosen gametes from entire population are IBD) = 1 – 0 = 1

$$\Rightarrow F_{ST} = \frac{H_T - \overline{H}_S}{H_T} = \frac{1 - \left(1 - \hat{f}\right)}{1} = \hat{f} = \frac{1}{1 + M}, \text{ where } M = 4Nm.$$

– Suggests way to estimate rate of migration from F_{ST}:

$$\hat{M} = \frac{1 - F_{ST}}{F_{ST}}.$$

– Careful: estimate requires lots of assumptions (island model, equilibrium, etc.) to be valid.

MIGRATION AND SELECTION

- **One-island model with selection**
  - *A* favored on island.
  - *a* fixed on continent: $p_c = 0$.
  - *A* is dominant.

    - Fitnesses on island:

Genotype	*AA*	*Aa*	*aa*
Fitness	1	1	$1 - s$

    - Life Cycle: $\underset{p}{\text{zygotes}} \xrightarrow{\text{selection}} \underset{p^*}{\text{adults}} \xrightarrow{\text{migration}} \underset{p^{**}}{\text{gametes}} \xrightarrow{\text{random union}} \underset{p'}{\text{zygotes}}$

    - After selection (before migration): $p^* = p\dfrac{1}{1 - q^2 s}$

• After migration & reproduction: $p' = (1-m)p^* + m(0) = \dfrac{p(1-m)}{1-q^2 s}$

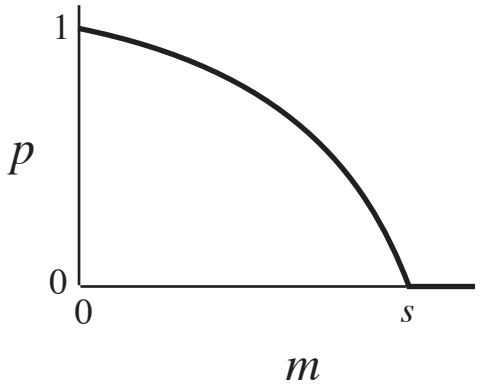
• To find any equilibria, set $p' = p$.

– Solving for p gives $\boxed{\hat{p} = 1 - \sqrt{m/s}}$.

– Require $0 \le \hat{p} \le 1$.

• This occurs only when $m < s$.

• Otherwise $\hat{p} = 0$.



– Now assume <u>A is recessive.</u>

• Fitnesses on island:

Genotype	*AA*	*Aa*	*aa*
Fitness	1	$1-s$	$1-s$

• After selection (before migration): $p^* = \dfrac{p(1-qs)}{1-sq(1+p)}$

• After migration & reproduction: $p' = (1-m)p^* + m(0) = \dfrac{p(1-m)(1-qs)}{1-sq(1+p)}$

• To find equilibria, set $p' = p$ and solve for p.

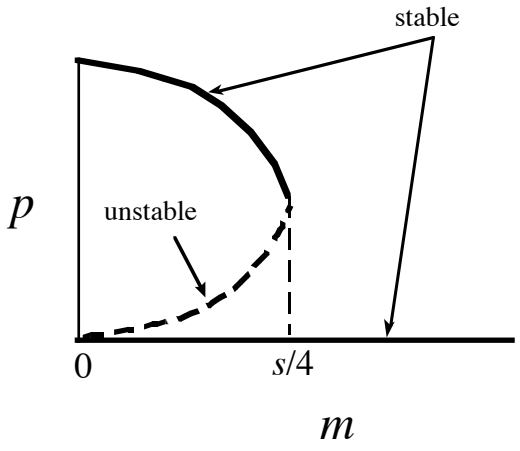
– Get cubic equation for $\hat{p}$'s (up to 3 possible solutions).

– $\hat{p} = 0$ is always an equilibrium (since $p_c = 0$).

– There are two *polymorphic* equilibria when $s > 4m$ (assuming m is small).

• one equilibrium is stable, the other is unstable.

– Graphically ⟶



• Implications

– If recessive selection is strong enough to maintain *A* in the face of migration, *A* will spread only if it's initially sufficiently frequent enough. Otherwise, it will be lost.

– In general, unless locally advantageous allele is completely dominant, it must reach a threshold frequency to persist.

– If an allele persists, it won't be found at a low frequency.

– Historical "accidents" play a role.

• Identical patches will evolve differently if they differ in initial allele frequency.

• The Levene Model

Q: What happens when a population is made up of a group of distinct subpopulation patches, with different selection pressures occurring in each and migration between locations?

A: Depends on geography (population structure).

– Natural populations fall somewhere between the following two extremes:

• Unrestricted migration.

• Restricted migration.

– A simple model of unrestricted migration was presented in 1953 by H. Levene.

– Assumptions of Levene's 1953 model:

• n patches in which different patterns of selection occur.

• Frequency of *A* among gametes is p.

• After fertilization, (diploid) zygotes colonize the different patches (at random).
– Important: this implies that the zygotes *within* patches are in H-W proportions.

• ith patch makes up a fraction c_i of the environment.

• Fitnesses in the ith patch:

Genotype	*AA*	*Aa*	*aa*
Fitness	$w_{AA}(i)$	$w_{Aa}(i)$	$w_{aa}(i)$

• Random mating between patches.

– Individuals from different localities form a single mating (gamete) pool.

– Why study the Levene model?

• Captures essential features of spatially subdivided population and is mathematcially tractable.

• Is a reasonable representation of certain natural systems as well.

– Back to model...How many gametes does each patch contribute to the gamete pool?

• Two extremes:

(1) *Hard selection* (due to Dempster, 1955)

• Patch contributes gametes in proportion to the fraction of survivors.
– i.e., patches with higher fitness contribute disproportionately more.

• Implies population size is not regulated within patches.

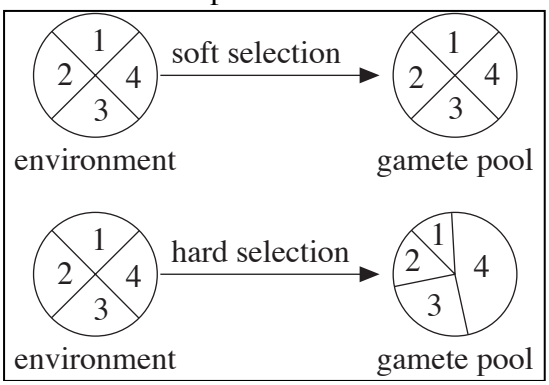
(2) *Soft selection*

• Each patch contributes fixed number of gametes to the mating pool *regardless* of local fitnesses.

• Number of reproducing adults from each patch is the same from one generation to the next.

• Implies population size is regulated within each patch.

• A schematic comparison between soft and hard selection, assuming $c_1 = c_2 = c_3 = c_4$ and $\overline{w}_4 > \overline{w}_3 > \overline{w}_2 > \overline{w}_1$ ⟶



– Levene model with hard selection ("constant number of zygotes"):

• Assumes contribution of genotype from patch i to the gamete pool is proportional to it's fitness in that patch [$w_{\text{genotype}}(i)$] × frequency of i patch in environment (c_i):
– i.e., total number of survivors of that genotype in patch $i \propto c_i w_{\text{genotype}}(i)$

• Overall fitness of genotype in population is its average fitness over patches:

– For example, mean fitness of *AA*: $\overline{w}_{AA} = \sum_{i=1}^{n} c_i w_{AA}(i)$

– Likewise for *Aa* and *aa*.

• Consider changes in the frequency p of A in the gamete pool.

• $p' = p\frac{p\overline{w}_{AA} + q\overline{w}_{Aa}}{\overline{\overline{w}}} = p\frac{\overline{\overline{w}}_A}{\overline{\overline{w}}}$ where $\overline{\overline{w}}_A = p\overline{w}_{AA} + q\overline{w}_{Aa}$ and $\overline{\overline{w}} = p^2\overline{w}_{AA} + 2pq\overline{w}_{Aa} + q^2\overline{w}_{aa}$.

• Looks just like selection with constant fitnesses: $\overline{w}_{AA}, \overline{w}_{Aa}, \overline{w}_{aa}$

• Consequences
– An allele will spread if it has the highest arithmetic mean fitness across patches.
– Selection will maintain a stable polymorphism if heterozygotes have the greatest arithmetic mean fitness across patches.

• For example, consider two equally sized patches, $c_1 = c_2 = 0.5$.

Fitness in patch:	*AA*	*Aa*	*aa*
# 1	0	0.75	1
# 2	1	0.75	0
Average:	0.5	0.75	0.5

• Selection maximizes arithmetic mean fitness across environments

– Levene model with soft selection ("constant number of adults"):

– Within each patch, selection operates as usual.

– Fitness in patch *i*:

AA	*Aa*	*aa*
w_i	1	v_i

– After selection, frequency of *A* in patch *i* is

$$p^*(i) = p\frac{pw_i + q(1)}{p^2w_i + 2pq(1) + q^2v_i} = p\frac{\overline{w}_A(i)}{\overline{w}(i)}$$

– Density regulation occurs independently in each patch.

– Survivors contribute to gamete pool in proportion to the size (= relative proportion of adults) of the patch, c_i:

$$p' = \sum_{i=1}^{n} c_i p^*(i) = \sum_{i=1}^{n} c_i p \frac{pw_i + q}{p^2 w_i + 2pq + q^2 v_i}$$

– Equilibrium: set $p' = p$ and solve for p.

• Results in polynomial of degree $2n+1$ in p.
$\Rightarrow$ as many as $2n+1$ equilibria, $\hat{p}$, are possible!

• Mathematically too difficult to find all these.

– Alternative: protected polymorphism analysis:

• Near $p = 0$, $p' \approx \sum_{i=1}^{n} \left(\frac{c_i}{v_i}\right) p = p \frac{1}{\tilde{v}}$

where $\tilde{v} = 1 \Big/ \left[\sum_{i=1}^{n} c_i \left(\frac{1}{v_i}\right)\right]$ is the <u>harmonic mean fitness</u> of *aa* homozygotes.

• Note that $p' > p$ (i.e., $\Delta p > 0$) whenever $1/\tilde{v} > 1 \Leftrightarrow \tilde{v} < 1$
– i.e., whenever the "harmonic mean fitness of *aa* homozygotes" < "mean fitness of heterozygotes"

• Likewise, near $p = 1$, $(q = 0)$, $q' > q$ whenever $\tilde{w} < 1$.

• <u>Conclude</u>: protected polymorphism occurs with soft selection whenever there is harmonic mean overdominance in fitness across patches: $\tilde{w} < 1 > \tilde{v}$.

– <u>Bottom line(s) for soft selection</u>
• Harmonic mean fitness across patches is the relevant fitness measure if $p \approx 0$ or 1.
• Turns out, however, that selection maximizes *geometric* mean fitness.

– Hard versus Soft Selection

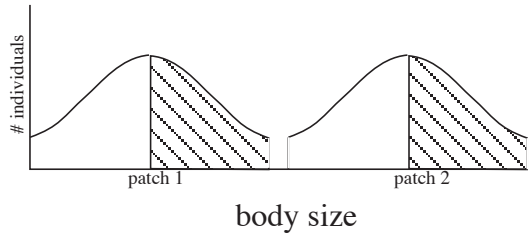
• Conditions exist in which an allele will increase under soft selection but not hard selection.
– I.e., polymorphisms can be maintained under a broader range of conditions with soft selection versus hard selection.

– Intuitively follows because under soft selection, individuals compete selectively only against "patch-mates".
• With hard selection, all compete.

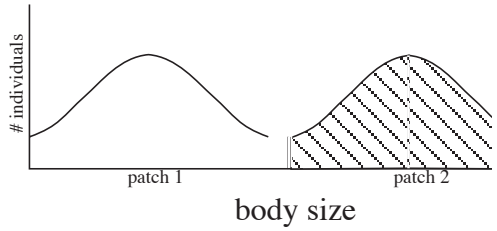
– Mathematically follows because harmonic mean is never larger than the arithmetic mean: $\tilde{v} \le \bar{v} = \sum_{i=1}^{n} c_i v_i$.

Q: Why does soft selection seem "hard" (density regulation; intense local competition) while hard selection seems "soft" (little competition; no density regulation)?

A: It all depends on your viewpoint (genetic vs. demographic).



– Soft selection: top 50% in each patch selected.



– Hard selection: top 50% selected (regardless of patch).