Evolution

Outline
1. Forces of evolution
2. Genetic drift
3. Mutations
4. Mutations + drift

1) Forces of evolution

**Definitions:** Genotype, locus, allele, phenotype

AAGCATTAA  #p 1A, 2C, 3G  #A, C, T

Diploid: Two chromosomes for each locus

A₁, A₂ → alleles

p 1 - p = allele frequency

2) Genetic drift

Consider a population of N organisms (2N alleles)

Math:
1) Choose an allele at random
2) Make its copy (exact)
3) Place into a new generation

Let’s calculate change in p(2Np₁) in one generation.

\[ p_{j} = p_{t}^{j}(1 - p_{t})^{2N-j}(2N) \]

\[ \langle j \rangle = p_{t} \times 2N \Rightarrow \langle p_{t} \rangle = \frac{j}{2N} = p_{t} \]

\[ \text{Var}(j) = \langle j^2 \rangle - \langle j \rangle^2 = p_{t}(1-p_{t}) \times 2N \]

\[ \text{Var}(p_{t}) = \text{Var}(j)/2N = p_{t}(1-p_{t})/2N \]

Next generation:

\[ \text{P}_{t+1} = \frac{j}{2N} \]

April 9, 2010
\[ \langle P_{t+1} \rangle = P_t \]

\[ \text{Var}(P_{t+1}) = \frac{P_t(1-P_t)}{2N} \]

Conclusions:

1) Random mating doesn't change the mean frequency
2) The variance of \( p \) increases

\[
\text{C diffusion!} \quad \text{But } D \text{ is not the same for all } p
\]

\[
\text{Smaller variance} \quad \text{Greater variance}
\]
Probability of fixation, starting at the frequency $p = p_0$

$$\Pi(p_0) = \sum_{\Delta p} \text{Prob}(\Delta p) \Pi(p_0 + \Delta p)$$

Expanded:

$$\Pi(p_0 + \Delta p) = \Pi(p_0) + \Pi'(p_0) \Delta p + \Pi''(p_0) \frac{\Delta p^2}{2}$$

By plugging into $\Pi(p_0)$ we obtain

$$\Pi(p_0) = \sum_{\Delta p} \text{Prob}(\Delta p) \left[ \Pi(p_0) + \Pi'(p_0) \Delta p + \Pi''(p_0) \frac{\Delta p^2}{2} \right]$$

$$= \Pi(p_0) + \Pi'(p_0) \sum_{\Delta p} \text{Prob}(\Delta p) \Delta p + \Pi''(p_0) \sum_{\Delta p} \text{Prob}(\Delta p) \frac{\Delta p^2}{2}$$

$$= \Pi(p_0) + \Pi'(p_0) \langle \Delta p \rangle + \Pi''(p_0) \frac{\langle \Delta p^2 \rangle}{2}$$

If $\Pi'(p_0) \langle \Delta p \rangle + \Pi''(p_0) \frac{\langle \Delta p^2 \rangle}{2} = 0$

$$\Pi(0) = 0 ; \quad \Pi(1) = 1$$

Kolmogorov Backward equation

Let's obtain probability of fixation under drift only (no mutations, no selection)

$$\langle \Delta p \rangle = 0$$

$$\langle \Delta p^2 \rangle = \text{Var}(\Delta p) = \frac{p(1-p)}{2N}$$

$$\Pi''(p_0) p_0 (1-p_0) = 0$$

$$\Pi''(p_0) = 0$$

$$\Pi'(p_0) = \Delta$$

$$\Pi(p_0) = a p_0 + b$$

$$\Pi(p_0) = p_0$$

using boundary conditions
3. Estimating the timescale of genetic drift

Consider a randomly mating population of size $N$

1. Choose an allele at random among $2N$ alleles
2. Make its exact copy
3. Place into new generation
4. Repeat until $2N$ alleles are in the generation

Evolution of $P$

Evolution of $H$

$G_{t+1}$ given $G_t$

$G = \text{Probability to pick two identical alleles in the population}$

$$G_{t+1} = \frac{1}{2N} + (1-\frac{1}{2N})G_t$$
4. Mutations and drift

Infinite allele model = no backward mutations

\[ M = \text{mutation rate} = \text{mutation probability per locus per generation} \]

\[ A_1 \rightarrow A_2 \rightarrow A_3 \rightarrow A_4 \text{ etc.} \]

\[ g_{t+1} = (1 - \mu)^3 \left( \frac{1}{2N} + (1 - \frac{1}{2N}) g_t \right) \]

\[ g_t \text{ and } g_{t+1} \text{ are frequencies of allele 1 in generation } t \text{ and } t+1 \]

\[ H_t = 1 - g_t \quad \Rightarrow \quad H_{t+1} = (1 - \frac{1}{2N}) - (1 - \frac{1}{2N})(1 - H_t) - 2\mu H_t + 2\mu \]

\[ \Delta H = -\frac{H_t}{2N} \]

\[ \tau = 2N \lambda \text{ characteristic decay time} = 2N \text{ generations} \]

\[ N \approx 10^5 \text{ gen. time 20 years} \]

\[ \tau = 40 \cdot 10^5 \text{ years} = 4 \text{ million years} \]

\[ \Rightarrow \text{Drift is very slow in large populations.} \]

\[ \text{common ancestor with chimp.} \]
\[ H_{t+1} = \left(1 - \frac{1}{2N} \right) H_t + 2\mu (1 - H_t) \]
\[ \Delta H = H_{t+1} - H_t = - \frac{H_t}{2N} + 2\mu (1 - H_t) \]

**Steady state** \( \Delta H < 0 \)

\[ \frac{H_t}{2N} = 2\mu (1 - H_t) \]
\[ \frac{H_t}{1 - H_t} = 4N\mu \]

\[ \mathcal{E}_h = \frac{4N\mu}{1 + 4N\mu} \]

\[ G = \frac{1}{1 + 4N\mu} \]

1. **Above**
   - **drift** (decreases \( H \))
   - **mutation** (increases \( H \))
   - Doesn't depend on \( N \)

2. \[ 4N\mu = \Theta \]

   - \( \Theta \ll 1 \quad H \to 0 \quad \text{drift wins} \)
   - \( \Theta \gg 1 \quad H \to 1 \quad \text{mutations win} \)

> Conclusion:
Balance between mutations and drift leads to a constant (steady state) diversity of the population.
Selection and drifts

Haldein's equation

\[
\begin{align*}
\text{Selection (frequency independent selection)} & \\
A_1A_1 & A_1A_2 & A_2A_2 & \text{fitness before selection} \\
\frac{w_1^2}{2} & w_1w_2 & w_2^2 & q^2 & 1 - s & \text{after selection} \\
\frac{1}{w} & & & & & 1 - s & \text{relative fitness} \\
\text{Where} & w &= p^2 w_1 + 2p(1-p)w_1w_2 + (1-p)^2 w_2^2 \geq \frac{1}{w} & \text{mean fitness} \\
& & & & & & \text{fitness of the population} \\
\end{align*}
\]

- selection coefficient (strength of selection), \( s > 0 \)

\( h \) determines the type of selection

* **Types of selection**

\( h = 0 \) A1 dominant

\( h = 1 \) A2 dominant

\( 0 < h < 1 \) incomplete dominance --> directional selection

\( h < 0 \) overdominance --> balancing selection

\( h > 1 \) underdominance --> disruptive selection

* **Special type of balancing selection**

- frequency-dependent selection.

* Let's calculate the effect of selection on allele freq. \( p \)

(in one generation)

\[
p_{t+1} = p_t^2 \left( \frac{w_1}{w} \right) + 2p_t(1-p_t) \left( \frac{w_1w_2}{w} \right) + (1-p_t)^2 \left( \frac{w_2}{w} \right)
\]
\[ p' = \frac{p^2 w_{11}}{2} + 2 \left( 1 - p \right) \frac{p w_{12}}{3} - p \left( 1 - p \right) \frac{w_{22}}{3} \]

\[ w = p^2 w_{11} + 2p(1-p)w_{12} + w_{22}(1-p)^2 \]

\[ \frac{d\bar{w}}{dp} = 2pw_{11} + 2(1-2p)w_{12} - 2w_{22}(1-p) \]

\[ \Delta p = \frac{1}{w} \left[ p^2 w_{11} + (1-p)p w_{12} - p [p^2 w_{11}] - 2p^2 (1-p) w_{12} - p (1-p)^2 w_{22} \right] \]

\[ = \frac{1}{w} \left[ w_{11} (1-p) p^2 + w_{12} (1-p) p (1-2p) - w_{22} p (1-p)^2 \right] \]

\[ = \frac{p(1-p)}{2w} \left[ 2w_{11} p + 2w_{12} (1-2p) - 2w_{22} (1-p) \right] \]

\[ \Delta p = \frac{p(1-p)}{2w} \left( \frac{d\bar{w}}{dp} \right) \]

Wright Equation

1. Direction of evolution is determined by \( \frac{d\log \bar{w}}{dp} \)

2. Stable equilibria are points where \( \frac{d\log \bar{w}}{dp} = 0 \).

\( \frac{d\log \bar{w}}{dp} = 0 \) is a local maximum of \( \bar{w} \).
1) New mutations
\[ P = \frac{1}{2N} \]
\[ \Pi \left( \frac{1}{2N} \right) = \frac{1 - e^{-S}}{1 - e^{-2NS}} \approx \frac{S}{1 - e^{-2NS}} \approx \frac{S}{2NS} \quad \text{for } 2NS \gg 1 \]
\[ \frac{S}{2NS} \quad \text{for } 2NS = 1 \]
\[ \Pi(s) \]
\[ \text{s} \]
\[ \sqrt{2N} \]

2) For \( h \neq \frac{1}{2} \)
\[ \Pi \left( \frac{1}{2N} \right) \approx 2(1-h)S \]

3) Deleterious
\[ \Pi_1 = 1 - \Pi_0 \]
\[ \Pi_2 = 1 - \Pi_1 (1 - q) = e^{-2NS} - 1 \]
\[ e^{-2NS} - 1 \]
\[ 1 - e^{-2NS} - 2NS(1-q) \]

New mutation
\[ q = \frac{1}{2N} \]
\[ \Pi = \frac{S}{e^{2NS} - 1} \]
\[ \frac{1}{2N} \quad \text{for } 2NS < e \]
\[ \frac{S}{2N} \]

Haldane's equation
\[ S > 0 \quad \Rightarrow \quad A2 \text{ disadvantageous mutation} \]
\[ k = \frac{\frac{2N}{4N_s}}{\text{rate of fixation}} - 1 \]

rate of a q. subst = rate of neutral subst.
Evolution, lecture 2

1. Selection (Wright's equation)

\[ p, \quad p_0 \quad \text{frequency} \]
\[ 1, \quad 1-s \quad \text{fitness} \]

In the next generation, \( s \ll 1 \Rightarrow \text{small disadvantage} \)

\[ p \cdot \frac{1}{w} \quad (1-p) \cdot \frac{1-s}{w} \quad \text{frequency} \]

\[ \frac{p}{w} + \frac{(1-p)(1-s)}{w} = 1 \]

\[ \bar{w} = p + (1-p)(1-s) = p + 1 - p - s + ps = 1 - s + ps \]

\[ \Delta p = p \cdot \frac{1}{\bar{w}} - p = p \cdot \frac{1}{w} = p \cdot \frac{1}{\bar{w}} \]

\[ \Delta p = p(1-p) \cdot \frac{s}{w} \quad \text{notice} \quad \frac{d\bar{w}}{dp} = s \]

\[ \Rightarrow \Delta p = p(1-p) \frac{d\bar{w}}{dp} \cdot \frac{1}{\bar{w}} = p(1-p) \frac{d\log \bar{w}}{dp} \]

Wright's equation

Conclusions:

1) Allele frequency changes to increase fitness of the population

\[ \Delta p = \frac{d\bar{w}}{dp} \]

\[ \bar{w}(p) \text{ arbitrary function of } p \]
2) The rate of change (rate of evolution) is higher for more diverse populations. 
\[ p(1-p) \propto \text{variance of the population} \]

\[ \Delta P \sim p(1-p) \]

2. Selection + drift (Haldane's equation)

**Probability of fixation \( \Pi(p) \)**

Backward Kolmogorov equation

\[ \Delta P \Pi' + \frac{\Delta P^2}{2} \Pi''(p) = 0 \]

Drift \( \Delta P^2 = \frac{p(1-p)}{2N} \)

Selection \( \Delta P = p(1-p)S \) (assuming \( W = b \))

\[ \frac{p(1-p)S \Pi'(p) + \frac{p(1-P)S}{2N} \Pi''(p)}{0} = 0 \]

\[ \Pi(0) = 0 \]
\[ \Pi(1) = 1 \]

\[ \Pi(p) = \frac{1-e^{-4NP}}{1-e^{-4NS}} \]
Selection and drift (Haldein's equation).

\[ \Pi = \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}} \]

1) New mutation (new \( A_1 \))

\[ p = \frac{1}{2N} \]

\[ \Pi(\frac{1}{2N}) = \frac{1 - e^{-S}}{1 - e^{-2N}} \approx \frac{S}{1 - e^{-2Ns}} \approx S \quad \text{for} \quad 2Ns > 1 \]

\[ \frac{S}{2} = \frac{1}{2N} \quad \text{for} \quad 2Ns < 1 \]

2) For \( h \neq \frac{1}{2} \)

\[ \Pi(\frac{1}{2N}) \approx 2(1 - h)S \]

depends on the fitness of heterozygote

3) Deleterious allele \( A_2 \)

\[ \Pi_1 = \Pi \]

\[ \Pi_2 = (1 - \Pi)(1 - q) = \frac{2Ns}{e^{2Ns} - 1} \]

\[ 1 - e^{-2Ns} - 2Ns(1 - q) \]

\[ 1 - e^{-2Ns} - 1 + e^{2Ns} \]

New mutation (new \( A_2 \))

\[ q = \frac{1}{2N} \]

\[ \Pi_2 = \frac{S}{e^{2Ns} - 1} \]

\[ 1 - \frac{\Pi}{2N} \quad \text{for} \quad 2Ns < 1 \]

\[ \frac{1}{2N} \quad \text{for} \quad 2Ns < 1 \]
Effective population size

Recall

\[ H_{t+1} = H_t \left( 1 - \frac{1}{2N_t} \right) \]

if \( N_t \), then

\[ H_{t+1} = H_t \left( 1 - \frac{1}{2N_t} \right) \]

\[ H_t = H_0 \prod_{i=0}^{t-1} \left( 1 - \frac{1}{2N_i} \right) \]

then

\[ \left( 1 - \frac{1}{2N_e} \right)^t = \prod_{i=0}^{t-1} \left( 1 - \frac{1}{2N_i} \right) \]

\[ \prod_{i=1}^{n} (1+x_i) = e^{\sum \log(1+x_i)} \]

\( x \ll 1 \)

using \( \log(1+x) \approx x \) we obtain

\[ e^{\sum x_i} \approx 1 + \sum x_i \]

\[ e^{-t/2N_e} = e^{-\sum_{i=0}^{t-1} 1/(2N_i)} \]

\[ N_e = \left( \frac{1}{t} \sum_{i=0}^{t-1} \frac{1}{N_i} \right)^{-1} \]

\( H \) - heterozygosity

(\( \frac{\text{fraction of different nucleotides}}{\text{fraction of identical nucleotides}} \))

effective pop. size

for \( x \approx 0.1 \)