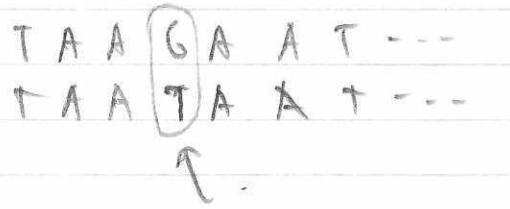
Gillespie Ch 1:

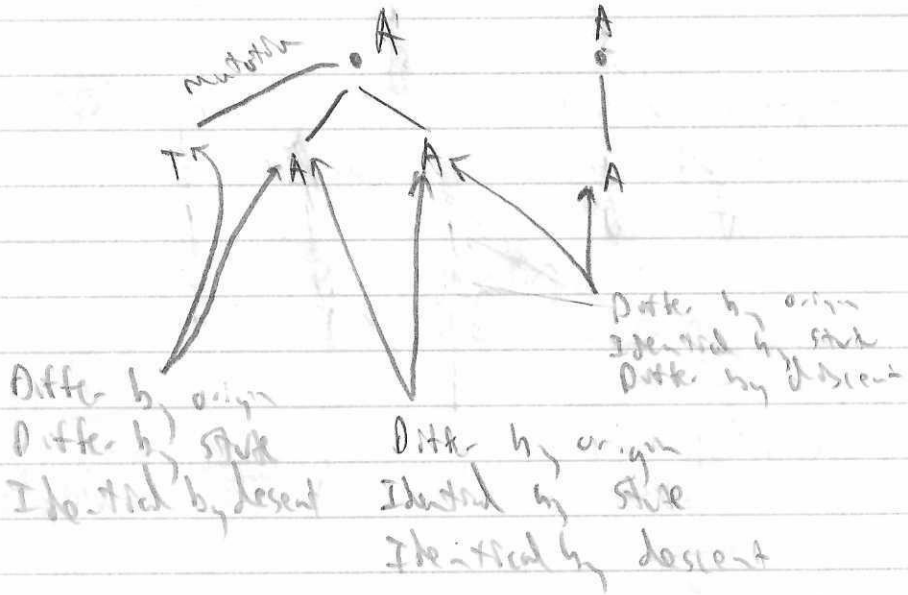
- **Nucleotides** -- A T G C --
- Most DNA similar b/t species



each are also a double helix

- **Variant** = "segregating site" = "polymorphic site" = location where differences occur
- **SNP** = single nucleotide polymorphism = location where difference is a single nucleotide
- **Allele** = type of difference
  - Ex. above we have "G" allele at "T" allele
- **Genotype** = all copies of alleles
  - Eg. "G/T" for that individual
- **Note:** Sometimes Gillespie uses "Allele" to mean a different sequence at Alcohol Dehydrogenase (ADH) b/c many different SNPs together is just a type of difference "Allele". Some folks would call this a **haplotype**.
- **locus** = location on genome (plural, loci)

- Allele differs by <u>origin</u> if on different chromosomes
- Differ by <u>state</u> if different sequence
- Differ by <u>descent</u> if allele does not from common ancestor.
- Ex.)



Identical by origin $\Rightarrow$ Identical by state and descent

Identical by state $\overset{\text{(usually)}}{\Rightarrow}$ identical by descent

↑ mutation is rare

Identical by descent $\not\Rightarrow$ identical by state

Identical by state $\not\Rightarrow$ identical by origin

Diploid

- Homozygous: Individual has both alleles identical by state

Heterozygous; Diploid individual's alleles differ by state.

- Allele frequency: Proportion of genomes that have an allele

- genotype frequency: Proportion of individuals that have a genotype.

Ex) AA, AT, TT, AT, AA

A allele frequency = $\frac{6}{10}$

AT genotype frequency = $\frac{2}{5}$

- Let's call one allele A_1 and the other A_2 that differ by state

Genotypes	A_1A_1	A_1A_2	A_2A_2
Genotype frequency	x_{11}	x_{12}	x_{22}

$x_{11} + x_{12} + x_{22} = 1$

- Allele frequency

of A_1: $p = x_{11} + \frac{1}{2} x_{12}$

of A_2: $q = 1 - p = x_{22} + \frac{1}{2} x_{12}$

pick A_1A_1 pick A_1 pick A_1A_2 pick A_1 pick A_2A_2 pick A_1

$$p = x_{11} \cdot 1 + x_{12} \cdot \frac{1}{2} + x_{22} \cdot 0$$

- 1.1.) $\binom{n}{2} + n = \frac{n(n-1)}{2} + n = \frac{n(n+1)}{2}$

1.2.) S: $\frac{2 \cdot 141 \cdot 111 + 32}{2 \cdot 332} = 0.6401$

F: $\frac{2 \cdot 28 + 111 + 15}{2 \cdot 332} = 0.2741$

I: $\frac{2 \cdot 5 + 15 + 32}{2 \cdot 332} = 0.08584$

- HWE: Consider A_1 to be a success
  X = # of A_1's in a randomly sampled individual

If alleles are independent, then $X \sim Bin(2, p)$

Independent: Large population size
Random mating
No selection
No migration

genotype frequencies

$$Pr(X=2) = p^2 = x_{22}$$
$$Pr(X=1) = 2p(1-p) = x_{12}$$
$$Pr(X=0) = (1-p)^2 = x_{11}$$

- HWE is reached in one generation in hermaphrodites

$$Pr(A_1A_1) = \underbrace{Pr(Egg=A_1)}_{p} \cdot \underbrace{Pr(Sperm=A_1)}_{p} = p^2$$

$$Pr(A_2A_2) = (1-p)^2$$

$$Pr(A_1A_2) = \underbrace{Pr(Egg=A_1)}_{p} \cdot \underbrace{Pr(Sperm=A_2)}_{1-p} + \underbrace{Pr(Egg=A_2)}_{1-p} \cdot \underbrace{Pr(Sperm=A_1)}_{p}$$

$$= 2p(1-p)$$

- Note: ① Allele frequencies do not change under random mating
  ② In 2-sex species, it takes 2 generations
  ③ Only need allele freq to get genotype freq.

1.3.) Female genotype freqs: x_{11}, x_{12}, x_{22}
Male genotype freqs: y_{11}, y_{12}, y_{22}

Let $p = x_{11} + \frac{1}{2}x_{12}$ $\quad q = y_{11} + \frac{1}{2}y_{12}$

$P_r(A_1A_1) = pq$
$P_r(A_1A_2) = p(1-q) + q(1-p)$
$P_r(A_2A_2) = (1-p)(1-q)$

↕ Genotype freq for all individuals, regardless of sex
↕ so by previous argument, sex generation is HWE

1.4.) See R

- Ratio of A_1A_2 freq to A_2A_2 freq

$$\frac{P_r(A_1A_2)}{P_r(A_1A_1)} = \frac{2pq}{q^2} = \frac{2p}{q} \approx \frac{2}{q}$$

↑ for small q

- If have alleles $A_1, \ldots, A_k$ w/ allele freqs $p_1, \ldots, p_k$

Under HWE: $P_r(A_iA_i) = p_i^2$
$P_r(A_iA_j) = 2p_ip_j$

Homozygosity → $G = P_r(\text{homo}) = \sum_{i=1}^{k} p_i^2$

Heterozygosity → $H = P_r(\text{hetero}) = 1 - G = 1 - \sum_{i=1}^{k} p_i^2$

only under HW

- Heterozygosity = frequency of heterozygote
  - but this is used more generally

$$\sum \frac{(Obs - Exp)^2}{Obs}$$

Obs	Exp
191	
111	
28	
32	
15	...
5	

- 1.7.) p_f = A_1 frequency for females
  p_m = A_1 frequency males

males only have one copy

$A_1 A_1$ × A_1 Y X

Men: $P_r(A_1) = p_f$

females: $P_r(A_1A_1) = p_f p_m$
$P_r(A_1A_2) = p_f(1-p_m) + (1-p_f)p_m$
$P_r(A_2A_2) = (1-p_f)(1-p_m)$

Next $p_f = p_f p_m + \frac{1}{2} p_f - \frac{1}{2} p_f p_m + \frac{1}{2} p_m - \frac{1}{2} p_f p_m$
$= \frac{1}{2}(p_f + p_m)$

Each generation: $p_m^{(new)} = p_f^{(old)}$

$p_f^{(new)} = \frac{1}{2}\left(p_m^{(old)} + p_f^{old}\right)$

$p_f^{(1)} = p_f$ $\qquad p_m^{(1)} = p_m$

$p_f^{(2)} = \frac{1}{2}(p_m + p_f)$ $\qquad p_m^{(2)} = p_f$

$p_f^{(3)} = \frac{1}{2}\left(p_f + \frac{1}{2}(p_m + p_f)\right)$ $\qquad p_m^{(3)} = \frac{1}{2}(p_m + p_f)$

$= \frac{3}{4}p_f + \frac{1}{4}p_m$

$p_f^{(4)} = \frac{1}{2}\left(\frac{1}{2}p_m + \frac{1}{2}p_f + \frac{3}{4}p_f + \frac{1}{4}p_m\right)$ $\qquad p_m^{(4)} = \left(\frac{3}{4}p_f + \frac{1}{4}p_m\right)$

$= \frac{5}{8}p_f + \frac{3}{8}p_m$

$x_i = \frac{1}{2}\left(x_{i-2} + x_{i-1}\right)$ $\qquad x_0 = p_m$

$x_1 = p_f$