Mutation Models, Neutral Theory
TABLE 3

Proportion of loci, out of 18, polymorphic and proportion of the genome estimated to be heterozygous in an average individual for each population studied

<table>
<thead>
<tr>
<th>Population</th>
<th>No. of loci polymorphic</th>
<th>Proportion of loci polymorphic</th>
<th>Proportion of genome heterozygous per individual</th>
<th>Maximum proportion of genome heterozygous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strawberry Canyon</td>
<td>6</td>
<td>.33</td>
<td>.148</td>
<td>.173</td>
</tr>
<tr>
<td>Wildrose</td>
<td>5</td>
<td>.28</td>
<td>.106</td>
<td>.156</td>
</tr>
<tr>
<td>Cimarron</td>
<td>5</td>
<td>.28</td>
<td>.099</td>
<td>.153</td>
</tr>
<tr>
<td>Mather</td>
<td>6</td>
<td>.33</td>
<td>.143</td>
<td>.173</td>
</tr>
<tr>
<td>Flagstaff</td>
<td>5</td>
<td>.28</td>
<td>.081</td>
<td>.120</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td><strong>.30</strong></td>
<td><strong>.115</strong></td>
<td><strong>.155</strong></td>
</tr>
</tbody>
</table>
Table 1.3
The heterozygosity for 71 allozyme loci in humans (Harris and Hopkinson, 1972).

<table>
<thead>
<tr>
<th>Locus</th>
<th>Heterozygosity (H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>51 monomorphic loci</td>
<td>0.00</td>
</tr>
<tr>
<td>Peptidase C</td>
<td>0.02</td>
</tr>
<tr>
<td>Peptidase D</td>
<td>0.02</td>
</tr>
<tr>
<td>Glutamate-oxaloacetate transaminase</td>
<td>0.03</td>
</tr>
<tr>
<td>Leucocyte hexokinase</td>
<td>0.05</td>
</tr>
<tr>
<td>6-Phosphogluconate dehydrogenase</td>
<td>0.05</td>
</tr>
<tr>
<td>Alcohol dehydrogenase-2</td>
<td>0.07</td>
</tr>
<tr>
<td>Adenylate kinase</td>
<td>0.09</td>
</tr>
<tr>
<td>Pancreatic amylase</td>
<td>0.09</td>
</tr>
<tr>
<td>Adenosine deaminase</td>
<td>0.11</td>
</tr>
<tr>
<td>Galactase-1-phosphate uridyl transferase</td>
<td>0.11</td>
</tr>
<tr>
<td>Acetyl cholinesterase</td>
<td>0.23</td>
</tr>
<tr>
<td>Mitochondrial malic enzyme</td>
<td>0.30</td>
</tr>
<tr>
<td>Phosphoglucomutase-1</td>
<td>0.36</td>
</tr>
<tr>
<td>Peptidase A</td>
<td>0.37</td>
</tr>
<tr>
<td>Phosphoglucomutase-3</td>
<td>0.38</td>
</tr>
<tr>
<td>Pepsinogen</td>
<td>0.47</td>
</tr>
<tr>
<td>Alcohol dehydrogenase-3</td>
<td>0.48</td>
</tr>
<tr>
<td>Glutamate-pyruvate transaminase</td>
<td>0.50</td>
</tr>
<tr>
<td>RBC acid phosphatase</td>
<td>0.52</td>
</tr>
<tr>
<td>Placental alkaline phosphatase</td>
<td>0.53</td>
</tr>
</tbody>
</table>
Irreversible Mutation

• 1 locus, 2 alleles
  – A, a (frequencies p, q)

• Let $\mu = A$ to a mutation rate (per generation)
  – $\Pr(A \text{ mutates to } a) = \mu$

\[
p_t = p_{t-1}(1 - \mu)
\]
\[
p_t = p_{t-2}(1 - \mu)^2
\]
\[
p_t = p_0(1 - \mu)^t
\]

What does $p_t$ approach as $t \rightarrow \infty$?
Irreversible Mutation

Hartl and Clark
Reversible Mutation

- 1 locus, 2 alleles
  - A, a (frequencies \( p, q \))
- Let \( \mu = A \) to a mutation rate (per generation)
- Let \( \nu = a \) to A mutation rate (per generation)

\[
p_t = p_{t-1}(1-\mu) + (1-p_{t-1})\nu
\]

\[
\hat{p} = \frac{\nu}{\mu + \nu}
\]

Hartl and Clark
Summary so far

- Random Mating
- Discrete Generations
- Hardy Weinberg
  - Infinite pop size
- Wright Fisher
- Mutational model
  - Neutral model

<table>
<thead>
<tr>
<th>Model</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele frequency constant, genetic variation maintained</td>
<td>Hardy Weinberg - Infinite pop size</td>
</tr>
<tr>
<td>Allele frequency changes, genetic variation lost</td>
<td>Wright Fisher</td>
</tr>
<tr>
<td>Allele frequency changes SLOWLY, genetic variation lost, maintained</td>
<td>Mutational model - Neutral model</td>
</tr>
</tbody>
</table>
Neutral Theory

- Intersection of mutation with drift
- Most mutations selectively neutral
- Drift determines allele frequencies
Infinite Alleles Model

• Each mutation creates **new** allele
  – 2 alleles with identical sequence MUST be IBD

• To measure homozygosity, we can measure $\Pr(\text{IBD})$

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

$$
\Pr(\text{IBD}_t) = F_t = \frac{1}{2N}(1-\mu)^2 + \left(1 - \frac{1}{2N}\right)(1-\mu)^2F_{t-1}
$$
Infinite Alleles Model

• Each mutation creates **new** allele
  – 2 alleles with identical sequence MUST be IBD
• To measure homozygosity, we can measure Pr(IBD)

\[
\text{Pr} \left( \text{IBD}_t \right) = F_t = \frac{1}{2N} + (1 - \frac{1}{2N}) F_{t-1}
\]

\[
\text{Pr} \left( \text{IBD}_t \right) = F_t = \frac{1}{2N} (1 - \mu)^2 + (1 - \frac{1}{2N})(1 - \mu)^2 F_{t-1}
\]

\[
\hat{F} = \frac{1}{1 + 4N_e \mu} \quad \hat{H} = 1 - \hat{F} = \frac{4N_e \mu}{1 + 4N_e \mu} \quad \theta = 4N_e \mu
\]
Infinite Alleles Model

- Each mutation creates **new** allele
  - 2 alleles with identical sequence MUST be IBD
- To measure homozygosity, we can measure $\Pr(\text{IBD})$

$$\Pr(\text{IBD}_t) = F_t = \frac{1}{2N} + (1 - \frac{1}{2N})F_{t-1}$$

$$\Pr(\text{IBD}_t) = F_t = \frac{1}{2N}(1 - \mu)^2 + (1 - \frac{1}{2N})(1 - \mu)^2F_{t-1}$$

$$\hat{F} = \frac{1}{1 + \theta} \quad \hat{H} = 1 - \hat{F} = \frac{\theta}{1 + \theta} \quad \theta = 4N_e\mu$$
At equilibrium:

- Steady-State under infinite alleles:
  - $H = \frac{\theta}{1+\theta}$
  - # alleles stationary

\[ E(k) = \sum_{i=1}^{n} \frac{\theta}{\theta + (i - 1)} \]
FIGURE 4.8 Relation between $\theta$, the expected number of alleles, and the sample size according to the Ewens sampling theory of a population in steady state under the infinite-alleles model of neutral mutation.

Hartl and Clark
At equilibrium

- Stationary distribution of allele frequencies
  - Allele frequency spectrum
    - (Unique) alleles 1\ldots k
    - Allelic configuration (frequencies $p_1, p_2\ldots p_k$)
      - Allele frequency spectrum
Allele Frequency Spectrum

- Sample size $n = 20$, $k = 10$ unique alleles
  - $p_1 = 6$
  - $p_2 = 4$
  - $p_3 = p_4 = p_5 = 2$
  - $p_6 = p_7 = p_8 = p_9 = p_{10} = 1$
Implications

• If we know $n, \theta$, we can write down $E(k)$
• If we know $n, k$, we can generate expected allele frequency distribution under neutrality
• We can use neutral expectations as null models to test for deviations from neutrality
Summary

• Neutral model is intersection of mutation, drift
• Mutations introduced through a population
• Once there, alleles are subject to drift and are ultimately fixed or lost
• At equilibrium there is a balance between drift and mutation
  – Every allele introduced by mutation is exactly balanced by allelic loss through drift
Controversial implications

• Allele frequency changes driven by drift, not selection
• Most polymorphisms have nothing to do with adaptation
Molecular Evolution

\[ k = 2N_e \mu \Pr(\text{fixation}) \]

\[ k = 2N_e \mu \left( \frac{1}{2N} \right) \]

\[ k = \mu \]
Molecular Clock
Molecular Evolution

• $k = \mu$
• Expected time b/t substitutions is $1/\mu$
• $K = 2\mu t$
• For $p = 1/2N$, $t_{fix} \approx 4N_e$
• For $p = 1/2N$, $t_{loss} \approx 2\ln(2N_e)$
Nearly Neutral Theory

• Considers ‘slightly deleterious’ mutations
  – $0 < |N_e s| < 1$

• Nearly neutral mutations
  – $|N_e s| < 1$
The distributions of fitness effects modelled by Ohta (1977) (exponential or gamma with $\beta=1$, dashed curve) and Kimura (1979) (gamma with $\beta=0.5$, solid curve).

Molecular Clock

The diagram illustrates the number of amino acid changes per 100 sites over millions of years since divergence for various proteins and peptides. The x-axis represents millions of years since divergence, while the y-axis shows the number of amino acid changes per 100 sites. The graph shows different rates of molecular evolution for Fibrinopeptides, Hemoglobin, Cytochrome c, and Histone H4.