Molecular Evolution
Genes in populations

Dr. Erica Bree Rosenblum

Genes in context
Genes in genome

Gene in individuals

Genes in genomes

Locus: position on a chromosome of gene (or genes) (plural: loci)

Allele: variant of DNA sequence at a particular locus

Genes in individuals

Homozygous: same allele at a locus for diploid (or polyploid) individuals

Heterozygous: different allele at a locus for diploid (or polyploid) individuals
Genes in populations
Gene pool: the set of all alleles at all loci in a population

Population size: generally denoted as N for haploid and 2N for diploid

N = 5
2N = 10

Genes in populations
Polymorphic: >1 genetically distinct type in a population
Monomorphic: a single genetic type in a population

Polymorphism can refer to:
Genetic scale: i.e., Single Nucleotide Polymorphism (SNP)

Population scale: i.e., color morphs of jaguars in South America

Genes in populations
Allele frequency (gene frequency):
relative proportion of allele in population

For a 2 allele locus:
p = frequency of A allele
q = frequency of a allele (1 - p)
p + q = 1

This is called Hardy Weinberg Equilibrium (HWE).
Genes in populations

What is an idealized population at equilibrium?

infinite population size, random mating, no mutation, no selection, no gene flow...

Is this realistic???

Generally NO, but HWE provides an important "null hypothesis". Deviations from HWE tell us that something interesting is happening (e.g., selection, small pop size, assortative mating, etc).

For a 2 allele locus:

\[ p = \text{frequency of A allele} \]
\[ q = \text{frequency of a allele (1} - \text{p)} \]
\[ p + q = 1 \]

If a population is in HWE, what will genotype frequencies be?

Genotype:

\[ AA \]
\[ Aa \]
\[ aa \]

Frequency:

\[ p^2 \]
\[ 2pq \]
\[ q^2 \]

This tells us allele but not genotype frequencies

With 2 alleles, we have 3 possible genotypes:

\[ AA \]
\[ Aa \]
\[ aa \]

HW can be generalized for more than 2 alleles.

\[ (p_1 + \cdots + p_n)^n = \sum_{k_1 + \cdots + k_n = n} \binom{n}{k_1, \ldots, k_n} p_1^{k_1} \cdots p_n^{k_n} \]

You don’t need to know how to derive this!

Example: expected genotype frequencies for tetraploidy

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAA</td>
<td>p^4</td>
</tr>
<tr>
<td>AAa</td>
<td>4p^3q</td>
</tr>
<tr>
<td>Aaa</td>
<td>6p^2q^2</td>
</tr>
<tr>
<td>Aaaa</td>
<td>4pq^3</td>
</tr>
<tr>
<td>aaaa</td>
<td>q^4</td>
</tr>
</tbody>
</table>

HW can be used to:

- Calculate genotype frequencies for a population in HWE
- Test for statistical deviations from HWE
Genes in populations

HW can be used to:
- Calculate genotype frequencies for a population in HWE
- Test for statistical deviations from HWE

Example calculation of HWE freq:

1 in 1700 newborns have cystic fibrosis

What proportion of the population are asymptomatic carriers?

\[ q^2 = \frac{1}{1700} = 0.00059 = 0.059\% \]

\[ q = 0.024 \]

\[ p = 0.976 \]

\[ p + q = 1 \]

\[ 2pq = 0.468 \]

4.68% of the population are carriers

Test for HWE deviation

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Genotype</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>White-spotted</td>
<td>AA</td>
<td>1469</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Aa</td>
<td>138</td>
</tr>
<tr>
<td>Little spotting</td>
<td>aa</td>
<td>5</td>
</tr>
</tbody>
</table>

Null hypothesis: population is in HWE.
Test for HWE deviation

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>1469</td>
<td>1467.4</td>
</tr>
<tr>
<td>Aa</td>
<td>138</td>
<td>141.2</td>
</tr>
<tr>
<td>aa</td>
<td>5</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Pearson’s chi-square test:

\[
X^2 = \sum \frac{(O - E)^2}{E}
\]

\[
= \frac{(1469 - 1467.4)^2}{1467.4} + \frac{(138 - 141.2)^2}{141.2} + \frac{(5 - 3.4)^2}{3.4}
\]

\[
X^2 = 0.83
\]

df = # phenotypes - # alleles = 1

So we don’t reject null hypothesis.

Some natural populations DO conform to HWE!

Effective population size

Census population size: denoted as “N”, total # individuals in population

Effective population size: denoted as “Ne”, reproductively active # individuals in population

Why are N and Ne different?

Ne is generally less than N

Can be due to:

- Unequal sex ratio (polygamous species or non reproductive caste)
- Spatial population structure
- Population size fluctuations
- Overlapping generations

Effective population size

Census population size: denoted as “N”, total # individuals in population

Effective population size: denoted as “Ne”, reproductively active # individuals in population

One traditional approach to calculating effective pop size:

\[
N = N_m + N_f
\]

\[
Ne = \frac{4N_m N_f}{N_m + N_f}
\]

\[
N = N \text{ for equal sex ratios}
\]

Calculating long term effective pop size:

Harmonic mean of N values

\[
Ne = \frac{\# \text{ generations}}{1/N_1 + 1/N_2 + \ldots + 1/N_n}
\]
Effective population size
Census population size: denoted as “N”, total # individuals in population
Effective population size: denoted as “N_e”, reproductively active # individuals in population

\[ N_e = \frac{\# \text{ generations}}{\frac{1}{N_1} + \frac{1}{N_2} + \ldots + \frac{1}{N_n}} \]

Northern Elephant Seal
Reduced to 20 inds in 1896
Now 30,000 inds

There are lots of ways to calculate \( N_e \)
(and it is controversial)

Traditional methods
- \( N_e \) based on inbreeding coefficients
- \( N_e \) based on variances
  - with variable sex ratio (4NmNf way)
  - with variable pop size through time (harmonic mean way)
  - with variable offspring number

Coalescent methods
- Sjödin paper

On the Meaning and Existence of an Effective Population Size

P. Sjödin, 1, 2, 3, A. Kaj, 4, S. Bonen, 1, M. Lacoursière, 5, 6 and M. Nordhagen 7