Hardy-Weinberg equilibrium

$p$ and $q$ represent the two alleles at any locus, $p + q = 1$

For a population, genotype frequencies are calculated as:

$$p^2 + 2pq + q^2 = 1$$

based on the frequencies of $p$ and $q$

For recessive disorders, by convention:

$p$ is the wild-type allele
$q$ is the mutant allele

Factors affecting Hardy-Weinberg equilibrium

1. Large population (no random fluctuation)
2. Random mating within the population (no inbreeding)
3. No migration (in or out)
4. Stable mutation rate (equal gain and loss)
5. No selection (no differences in fitness based on genotype)

Migration

Example: Duffy blood group (Fy glycoprotein), the receptor on red blood cells for $P. vivax$

The West African Fy allele has a point mutation in a GATA transcription factor binding site, leading to lack of expression in red cells and resistance to malaria (Fy (a-b-))

Fy a+ = 0 in Africa
Fy a+ = 0.42 in Georgia white population
Fy a+ = 0.46 in Georgia black population

Migration led to introduction of the a+ allele into the black population

Merging two isolated populations of equal size

Population I:

\[
\begin{align*}
AA &= 0.25; \\
Aa &= 0.5; \\
aa &= 0.25
\end{align*}
\]

$p = ; q =$

Population II:

\[
\begin{align*}
AA &= 0.64; \\
Aa &= 0.32; \\
aa &= 0.04
\end{align*}
\]

$p = ; q =$

Merged Population:

\[
\begin{align*}
A &= p = \\
a &= q = \\
AA &= ; Aa &= ; aa =
\end{align*}
\]
Merging two isolated populations of unequal size

Population I, 60% of the total:
\[ AA = 0.25; \quad Aa = 0.5; \quad aa = 0.25 \]
\[ p = ; \quad q = \]

Population II, 40% of the total:
\[ AA = 0.64; \quad Aa = 0.32; \quad aa = 0.04 \]
\[ p = ; \quad q = \]

Merged Population:
\[ A = p = \]
\[ A = q = \]
\[ AA = ; \quad Aa = ; \quad aa = \]

Selection

<table>
<thead>
<tr>
<th>Genotype</th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial freq</td>
<td>( p^2 )</td>
<td>( 2pq )</td>
<td>( q^2 )</td>
</tr>
<tr>
<td>fitness (( w ))</td>
<td>1</td>
<td>1</td>
<td>1-s</td>
</tr>
</tbody>
</table>

\( p = 0.5, \quad q = 0.5, \quad s = 0.1 \) (90% fitness)

In the next generation:
\[ AA = 0.5^2 \times 1 = 0.25 \]
\[ Aa = 2 \times 0.5 \times 0.5 \times 1 = 0.5 \]
\[ aa = 0.5^2 \times (1-0.1) = 0.225 \]

Total = 0.25 + 0.5 + 0.225 = 0.975

Selection

In the next generation:
\[ \text{Genotype} \quad AA \quad Aa \quad aa \]
\[ \text{proportional} \quad 0.25/0.975 \quad 0.5/0.975 \quad 0.225/0.975 \]
\[ \text{contribution} \quad = 0.256 = 0.513 = 0.231 \]

Allele frequency in the next generation:
\[ a = q' = \]
\[ A = p' = \]
In the next generation:

<table>
<thead>
<tr>
<th>Genotype</th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>$p'^2$</td>
<td>$2p'q'$</td>
<td>$q'^2$</td>
</tr>
<tr>
<td>frequency</td>
<td>$(.513)^2$</td>
<td>$2(.513)(.487)$</td>
<td>$(.487)^2$</td>
</tr>
<tr>
<td>fitness ($w$)</td>
<td>1</td>
<td>1</td>
<td>1-s</td>
</tr>
<tr>
<td>proportion</td>
<td>.263/.976</td>
<td>.5/.976</td>
<td>.213/.976</td>
</tr>
<tr>
<td>proportion</td>
<td>.270</td>
<td>.512</td>
<td>.218</td>
</tr>
</tbody>
</table>

Allele frequency in the next generation:

- $a = q' = 0.218 + (1/2)(0.512) = 0.474$  
  
- $A = p' = 0.270 + (1/2)(0.512) = 0.526$

Inbreeding: e.g. 1st cousin mating

- $a_ia_i = (1/2)^6 = 1/64$
- $a_na_n = (1/2)^6 \times 4 = 1/16$

Risk of autosomal recessive disease in families

<table>
<thead>
<tr>
<th>mutation frequency</th>
<th>disease incidence</th>
<th>outbred population</th>
<th>first cousin mating</th>
<th>outbred/first cousin ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>q</td>
<td>$q^2$</td>
<td>$q \times (1/16)$</td>
<td>$q^2 / q \times (1/16)$</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>$0.0001$</td>
<td>$0.00063$</td>
<td>6.3:1 ↑</td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>$10^{-6}$</td>
<td>$6.3 \times 10^{-5}$</td>
<td>63:1 ↑</td>
<td></td>
</tr>
</tbody>
</table>

Chance of affected status = (chance mom is a carrier)(chance dad is a carrier)(chance of inheriting both mutant alleles)

Carrier frequency = 1/30

Integrating pedigree analysis with Hardy-Weinberg
Risk of autosomal recessive disease in families

- Chance mom is a carrier =
- Chance dad is a carrier =
- Chance of inheritance =
- Chance of affected status =

Increased frequency of recessive disorders is seen in some populations

<table>
<thead>
<tr>
<th>Disease</th>
<th>Population</th>
<th>Incidence</th>
<th>Carrier frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sickle Cell</td>
<td>African American</td>
<td>1/600</td>
<td>1/12</td>
</tr>
<tr>
<td>Tay Sachs</td>
<td>Ashkenazi Jews</td>
<td>1/3000</td>
<td>1/27</td>
</tr>
<tr>
<td>Beta thalassemia</td>
<td>Mediterranean</td>
<td>1/3600</td>
<td>1/30</td>
</tr>
<tr>
<td>Alpha thalassemia</td>
<td>Southeast Asian</td>
<td>1/2500</td>
<td>1/25</td>
</tr>
<tr>
<td>30 disorders</td>
<td>Finland</td>
<td>varies</td>
<td>high aggregate</td>
</tr>
</tbody>
</table>

How did they become so frequent?

A. Genetic drift: random (stochastic) allele transmission
   - large populations buffered against this

B. Population bottleneck
   - geographic isolation (e.g. Finland)
   - founder effect
   - reproductive isolation (e.g. Amish)
   - founder effect
   - created by disease, persecution, calamity
     - reduction in population size

C. Selection: current or former heterozygote advantage
   - e.g. malaria resistance in sickle cell carriers (HbS/HbA)

Why don’t deleterious alleles disappear?

- e.g. malaria and sickle cell

HbS/HbA are malaria resistant (heterozygote advantage in malarial areas) → gain HbS alleles
HbS/HbS generally die and don’t reproduce → lose HbS alleles
HbA/HbA leads to genetic loss from malaria → lose HbA alleles

“Balanced selection” leads to offsetting selective pressure on allele frequencies
Implications of Hardy-Weinberg

- carriers are frequent relative to affected individuals
  - e.g. Tay Sachs: carrier frequency 1/27, incidence 1/3000
- most parents are carriers as carrier rate >> mutation rate
- high carrier frequency facilitates public health carrier screening
  - e.g. Tay Sachs screening in Ashkenazi college students
  - e.g. beta thalassemia screening in Mediterranean populations
- can dramatically reduce the disease incidence

Genome variation in humans

- How many SNPs are there in an individual?
- How many differences are there between people?
- How many recessive alleles do we each carry?

Watson’s genome

- 3,320,000 SNPs, 610,000 new ones
- 227,718 indels (2 bp to 38 kb), 109,179 new ones
- 23 CNVs (26 kb to 1.6 Mb)
- 10,569 nonsynonymous single nucleotide changes
  - carrier for 10 recessive disorders
  - other changes may also be deleterious
  - population estimates of 1-3 recessive mutations are too low!

Watson vs Venter

- 3,320,000 vs. 3,540,000 SNPs
- 610,000 vs. 740,000 novel SNPs
- differ by 7,648 coding sequence changes

There is far more variation than expected
Most variation is not deleterious
Humans are robust to variation (adaptive?)

More genomes are informing migration, evolution, disease, etc.