Hardy-Weinberg Equilibrium – The Basic Model of Population Genetics

Tiago da Silva Ribeiro
Evolutionary Biology - 2016
Summary

• Hardy-Weinberg Equilibrium
• Deviations from HW-Equilibrium
A Null Model for Evolutionary Biology

What do we need to create a model to study evolutionary biology?

-Wait... what is evolution again?

The change in allele frequencies (or the heritable expression of those alleles) in a population across generations.

So, first thing: we need a model to describe the behavior of allele frequencies in a population through time when evolution is not happening!

-Wait again... what is a population? What is an allele?
Definitions

**Population:** A group of individuals within a species that is capable of interbreeding and producing fertile offspring (definition for sexual species).

**Gene:** A region of the genome, a sequence of DNA (or RNA), that is the unit of inheritance, the product of which contributes to phenotype.

**Allele:** Variant forms of a gene (e.g. alleles for wet or dry earwax).

**Genotype:** The combination of alleles at a gene.
- A diploid organism has two copies of its genome.
- Each copy will carry a copy of an allele for any given autosomal gene.
- These copies can be the same (homozygous) or distinct (heterozygous).
A Null Model for Evolutionary Biology

Back to modeling!

Our model has to keep track of the allele frequencies in a population through time, when evolution is \textit{not} happening!
A Null Model for Evolutionary Biology

Back to modeling!

Gametes → Zygotes → Adults → Juveniles
A Null Model for Evolutionary Biology

Back to modeling!

Gametes

Now let's add some genetics!

Adults

Zygotes

Juveniles
A Null Model for Evolutionary Biology

Back to modeling!

# of gametes???

Gametes

# of adults:
AA = 36
Aa = 48
aa = 16

Adults

# of juveniles:
AA = 36
Aa = 48
aa = 16

Juveniles

Zygotes
A Null Model for Evolutionary Biology

Back to modeling!

- Assumption: infinite number of gametes produced.
- So we will calculate FREQUENCY of gametes carrying the different alleles!
- Which is the same as the frequency of alleles in the population of adults.
A Null Model for Evolutionary Biology

Back to modeling!

### Allele Frequencies in the Population of Adults:

\[
\begin{align*}
    f(A) &= \frac{2 \times \#AA + \#Aa}{2 \times \# \text{ of adults}} \\
    f(a) &= \frac{2 \times \#aa + \#Aa}{2 \times \# \text{ of adults}}
\end{align*}
\]

- \# of adults: 
  - AA = 36
  - Aa = 48
  - aa = 16
A Null Model for Evolutionary Biology

Back to modeling!

Gametes

- f(A) = 0.6
- f(a) = 0.4

# of adults:
- AA = 36
- Aa = 48
- aa = 16

Zygotes

- How many zygotes with each genotype do we expect?

Juveniles

- # of juveniles:
  - AA = 36
  - Aa = 48
  - aa = 16
Expected Genotype Frequencies

How many AA, Aa, and aa zygotes are expected in the next generation?

- The allele frequencies are:
  - $f(A) = p = 0.6$
  - $f(a) = q = 0.4$

Each zygote is a combination of 2 alleles. So, to make a:

- AA zygote pick A twice = $0.6 \times 0.6 = 0.36$ ($p^2$)
- aa zygote pick a twice = $0.4 \times 0.4 = 0.16$ ($q^2$)
- Aa zygote pick a and A = $0.6 \times 0.4 + 0.4 \times 0.6 = 0.48$ (2pq)

Then multiply the frequency by total number of zygotes.
Expected Genotype Frequencies

Generalizing: If $p$ and $q$ represent the relative frequencies of the only two possible alleles in a population at a particular gene, then for a diploid organism (2 chromosomes),

$$(p + q)^2 = 1$$

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

where $p^2$ and $q^2$ represent the frequencies of the homozygous genotypes and $2pq$ represents the frequency of the heterozygous genotype.
A Null Model for Evolutionary Biology

Back to modeling!

Gametes

- f(A) = 0.6
- f(a) = 0.4

# of adults:
- AA = 36
- Aa = 48
- aa = 16

Zygotes

EXPECTED # of zygotes:
- AA = 36
- Aa = 48
- aa = 16

# of juveniles:
- AA = 36
- Aa = 48
- aa = 16
 Allele frequencies remain the same! This is the Hardy-Weinberg equilibrium

G. H. Hardy (1877 - 1947)
Hardy-Weinberg Equilibrium

If $p$ and $q$ represent the relative frequencies of the only two possible alleles in a population at a particular gene, then for a diploid organism (2 chromosomes),

$$(p + q)^2 = 1$$

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

where $p^2$ and $q^2$ represent the frequencies of the homozygous genotypes and $2pq$ represents the frequency of the heterozygous genotype.
Hardy-Weinberg Equilibrium

Back to modeling!

Gametes

f(A) = 0.6
f(a) = 0.4

# of adults:
AA = 36
Aa = 48
aa = 16

Zygotes

EXPECTED # of zygotes:
AA = 36
Aa = 48
aa = 16

# of juveniles:
AA = 36
Aa = 48
aa = 16

What are the assumptions behind the model?
Hardy-Weinberg Equilibrium

Assumptions:

- No migration
- No mutation
- No natural selection
- Random mating
- No genetic drift *(infinite population size)*

Violations:

- Changes due to the acquisition of new variants.
- Unequal survival and reproductive success among genotypes.
- Distinct probabilities of gametes finding each other.
- Changes in the allele frequency due to chance.
Hardy-Weinberg Equilibrium

Conclusions:
- The allele frequencies remain the same generation after generation.
- We can predict the genotype frequencies from the allele frequencies, and vice-versa.
- Rare alleles will usually be found in the heterozygous.
Hardy-Weinberg Equilibrium

\[ q^2 \quad 2pq \quad p^2 \]
Hardy-Weinberg Equilibrium

The conclusions shown here also apply for genes with more than 2 alleles and organisms with ploidy greater than 2.
Hardy-Weinberg Equilibrium

The conclusions shown here also apply for genes with more than 2 alleles and organisms with ploidy greater than 2.

3 alleles (diploid organism):

\[(p + q + r)^2 = p^2 + q^2 + r^2 + 2pq + 2pr + 2qr\]
Hardy-Weinberg Equilibrium

The conclusions shown here also apply for genes with more than 2 alleles and organisms with more than 2 chromosomes.

Triploid organisms (with two alleles):

\[(p + q)^3 = 1\]

\[= p^3 + 3p^2q + 3pq^2 + q^3 = 1\]
DEVIATION from Hardy-Weinberg Equilibrium Indicates that EVOLUTION Is happening
Hardy-Weinberg across a Genome

• In natural populations, some genes might be out of HW equilibrium, while being in Hardy-Weinberg equilibrium at other genes

• For example, some genes might be undergoing natural selection and become out of HW equilibrium, while the rest of the genome remains in HW equilibrium
How can you tell whether a population is out of HW Equilibrium?

1. When allele frequencies are changing across generations

2. When you cannot predict genotype frequencies from allele frequencies (means there is an excess or deficit of genotypes than what would be expected given the allele frequencies)
## Example 1

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>20</td>
<td>$p^2 \times 100$</td>
</tr>
<tr>
<td>Aa</td>
<td>80</td>
<td>$2pq \times 100$</td>
</tr>
<tr>
<td>aa</td>
<td>0</td>
<td>$q^2 \times 100$</td>
</tr>
<tr>
<td>p</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>q</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Total:</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Genotype</td>
<td>Observed</td>
<td>Expected</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>AA</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>Aa</td>
<td>80</td>
<td>48</td>
</tr>
<tr>
<td>aa</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>p</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>q</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Total:</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
### Example 2

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>35</td>
<td>$p^2 \times 100 = 36$</td>
</tr>
<tr>
<td>Aa</td>
<td>50</td>
<td>$2pq \times 100 = 48$</td>
</tr>
<tr>
<td>aa</td>
<td>15</td>
<td>$q^2 \times 100 = 16$</td>
</tr>
<tr>
<td>p</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>q</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Total:</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Testing for Deviation from Hardy-Weinberg Expectations

• A $\chi^2$ test can be used to determine if a population is significantly different from the expectations of Hardy-Weinberg equilibrium.

• If we have a series of genotype counts from a population, then we can compare these counts to the ones predicted by the Hardy-Weinberg model.

• $O =$ observed counts, $E =$ expected counts, sum across genotypes

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}.$$
Example 1

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Observed</th>
<th>Expected</th>
<th>((O-E)^2/E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>20</td>
<td>36</td>
<td>7.11</td>
</tr>
<tr>
<td>Aa</td>
<td>80</td>
<td>48</td>
<td>21.33</td>
</tr>
<tr>
<td>aa</td>
<td>0</td>
<td>16</td>
<td>16</td>
</tr>
</tbody>
</table>

\[ \text{Total: } 100 \quad 100 \quad 44.44 \]

\(chi-square > 3.841 \Rightarrow \text{observed genotype frequencies are significantly different from the expected genotype frequencies.}\)
### Example 2

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Observed</th>
<th>Expected</th>
<th>$(O-E)^2/E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>35</td>
<td>36</td>
<td>0.028</td>
</tr>
<tr>
<td>Aa</td>
<td>50</td>
<td>48</td>
<td>0.083</td>
</tr>
<tr>
<td>aa</td>
<td>15</td>
<td>16</td>
<td>0.062</td>
</tr>
</tbody>
</table>

$p = 0.6$  
$q = 0.4$  

*Total: 100  100  0.173*

**chi-square < 3.841 => observed genotype frequencies are NOT significantly different from the expected genotype frequencies.**
Violations of HW Eq. Assumptions

Natural Selection

p = ?
q = ?

# of adults:
AA = 3600
Aa = 0
aa = 1600

# of juveniles:
AA = 3600
Aa = 4800
aa = 1600

\[ p = \frac{3600 + 3600 + 0}{10400} = 0.69 \]
\[ q = 1 - p = 0.31 \]
Violations of HW Eq. Assumptions

Natural Selection

Gametes

- p = 0.69
- q = 0.31

# of adults:
- AA = 3600
- Aa = 0
- aa = 1600

Adults

Zygotes

EXPECTED # of zygotes (total 10000):
- AA = ?
- Aa = ?
- aa = ?

Juveniles

# of juveniles:
- AA = 3600
- Aa = 4800
- aa = 1600

Death of Aa juveniles!!!
Violations of HW Eq. Assumptions

Natural Selection

Gametes

\[ p = 0.69 \]
\[ q = 0.31 \]

\# of juveniles:
\[ AA = 3600 \]
\[ Aa = 4800 \]
\[ aa = 1600 \]

\# of adults:
\[ AA = 3600 \]
\[ Aa = 0 \]
\[ aa = 1600 \]

\[ p = 0.69 \]
\[ q = 0.31 \]

\[ AA = p^2 = 0.69^2 = 0.4761 \]
\[ 0.4761 \times 10000 = 4761 \]

\[ Aa = 2pq = 2 \times 0.69 \times 0.31 = 0.4278 \]
\[ 0.4278 \times 10000 = 4278 \]

\[ aa = q^2 = 0.31^2 = 0.0961 \]
\[ 0.0961 \times 10000 = 961 \]

\[ 0.4761 + 0.4278 + 0.0961 = 1 \]

Zygotes

EXPECTED \# of zygotes (total 10000):

\[ AA = ? \]
\[ Aa = ? \]
\[ aa = ? \]
Violations of HW Eq. Assumptions

Natural Selection

Gametes

\[ p = 0.69 \]
\[ q = 0.31 \]

# of adults:
\[ AA = 3600 \]
\[ Aa = 0 \]
\[ aa = 1600 \]

# of juveniles:
\[ AA = 3600 \]
\[ Aa = 4800 \]
\[ aa = 1600 \]

Zygotes

EXPECTED # of zygotes (total 10000):
\[ AA = 4761 \]
\[ Aa = 4278 \]
\[ aa = 961 \]

Death of Aa juveniles!!!
Violations of HW Eq. Assumptions

Natural Selection

Gametes

\[ p = 0.69 \]
\[ q = 0.31 \]

# of adults:
\[ AA = 3600 \]
\[ Aa = 0 \]
\[ aa = 1600 \]

# of juveniles:
\[ AA = 3600 \]
\[ Aa = 4800 \]
\[ aa = 1600 \]

Distinct allele & genotype frequencies!

Zygotes

EXPECTED # of zygotes (total 10000):
\[ AA = 4761 \]
\[ Aa = 4278 \]
\[ aa = 961 \]

Death of Aa juveniles!!!
Violations of HW Eq. Assumptions

Natural Selection

Questions:
1) In the zygote stage, is the population in HW-Eq?
2) In the adult stage, is the population in HW-Eq?

# of juveniles:
AA = 3600
Aa = 4800
aa = 1600

# of adults:
AA = 3600
Aa = 0
aa = 1600

p = 0.69
q = 0.31

EXPECTED # of zygotes (total 10000):
AA = 4761
Aa = 4278
aa = 961

Zygotes

Gametes

Adults

Juveniles
## Violations of HW Eq. Assumptions

### Example 3

<table>
<thead>
<tr>
<th>Generation</th>
<th>A</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.64</td>
<td>0.32</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>0.63</td>
<td>0.33</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>0.64</td>
<td>0.315</td>
<td>0.045</td>
</tr>
<tr>
<td>4</td>
<td>0.65</td>
<td>0.31</td>
<td>0.04</td>
</tr>
</tbody>
</table>

What may be happening?
Hardy-Weinberg – Summary

• Allele frequencies remain constant from generation to generation.
• Genotype frequencies can be predicted from allele frequencies, and vise-versa.

• Changes in allele frequencies from one generation to the next OR mismatch between observed and expected genotype frequencies means that evolution is happening.
Questions?